

Nonredundant Precoding-Assisted Blind Channel Estimation for Single-Carrier Space–Time Block-Coded Transmission With Frequency-Domain Equalization

Jwo-Yuh Wu, *Member, IEEE*, and Ta-Sung Lee, *Senior Member, IEEE*

Abstract—Relying on nonredundant diagonal precoding and independent and identically distributed (i.i.d.) source assumption, this paper proposes a blind channel estimation scheme for single-carrier frequency-domain equalization-based space–time block coded systems. The proposed method exploits the precoding-induced linear signal structure in the conjugate cross correlation between the two temporal block received signals as well as the circulant channel matrix property and can yield exact solutions whenever the channel noise is circularly Gaussian and the receive data statistic is perfectly obtained. The channel estimation formulation builds on rearranging the set of linear equations relating the entries of conjugate cross-correlation matrix and products of channel impulse responses into one with a distinctive block-circulant with circulant-block (BCCB) structure. This allows a simple identifiability condition depending on precoder parameters alone and also provides a natural yet effective optimal precoder design framework for improving solution accuracy when imperfect data estimation occurs. We consider two models of data mismatch, from both deterministic and statistical points of view, and propose the associated design criteria. The optimization problems are formulated to take advantage of the BCCB system matrix property and are solved analytically. The proposed optimal precoder aims to optimize solution robustness against deterministic error perturbation and also minimize the mean-square error when the data mismatch is modeled as a white noise. Pairwise error probability analysis is conducted for investigating the equalization performance. Numerical examples are used to illustrate the performance of the proposed method.

Index Terms—Blind channel estimation, block-circulant matrix with circulant blocks (BCCB), circulant matrix, multiple input single output (MISO), nonredundant precoders, single-carrier frequency-domain equalization, space–time block code (STBC), transmit diversity.

Manuscript received September 8, 2005; revised May 9, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Nicholas D. Sidiropoulos. This work is sponsored by the National Science Council of Taiwan under joint Grants NSC 96-2752-E-002-009 and NSC-95-2221-E-009-047-MY2, the MoE ATU Program of the Ministry of Education, Taiwan, and the MediaTek Research Center at the National Chiao Tung University. Part of this paper was presented jointly at the Seventh IEEE International Workshop on Signal Processing Advances in Wireless Communications, Cannes, France, July 2–5, 2006 and the IEEE International Symposium on Information Theory, Seattle, WA, July 6–12, 2006.

J.-Y. Wu and T.-S. Lee are with the Department of Communication Engineering, National Chiao Tung University, Hsinchu 300, Taiwan, R.O.C. (e-mail: jywu@cc.nctu.edu.tw; tslee@mail.nctu.edu.tw).

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

Digital Object Identifier 10.1109/TSP.2006.887113

I. INTRODUCTION

A. Overview

SPACE–TIME block code (STBC) is a widely known transmit diversity technique for combating channel fading in modern wireless communications [22]. Most of the existing proposals are devised for the flat-fading channel environment, e.g., the Alamouti's scheme [1] and the related generalization by Tarokh *et al.* [34], among others. When the propagation channels are subject to frequency-selective fading, one popular STBC technique is via time-reversal block-wise encoding, either combined with orthogonal frequency-division multiplexing (OFDM) mechanism [27], [40], or resorting to time-domain equalizer [26], for removing the channel distortion. The multicarrier-related solutions, although simplifying receiver implementations, would incur a high peak-to-average power ratio (PAPR) and is sensitive to carrier frequency offset. The scheme with time-domain equalization, on the other hand, can provide additional multipath diversity at the expense of decoding complexity. To avoid the drawbacks of the multicarrier strategy and to also maintain low receiver complexity, an alternative single-carrier frequency-domain equalization (FDE)-based STBC was proposed in [2]. The aforementioned STBCs capable of mitigating dispersive channels can be cast into a general code formulation [39]; comparisons of the achievable performances and implementation costs can be found in [3].

To realize the diversity benefit of STBC, the channel state information must be known at the receiver to coherently combine the multiple temporal received signals for decoding.¹ Since STBC potentially entails low spectral efficiency and training-based channel estimation further consumes the bandwidth resource, blind approaches then become appealing candidate solutions. There has been extensive literature on blind multiple-input multiple-output (MIMO) channel estimation [14], [16]. However, only a few studies are tailored for STBC systems, typically through a multiple-input single-output (MISO) channel link. Under flat-fading assumption, several schemes were put forth for orthogonal STBC [4], [7], [32], and for a general linear code family [33]. For time-reversal

¹Differential STBC does not require channel information for decoding but incurs a 3-dB penalty in signal-to-noise ratio [22, Sec. 9.6].

STBC over frequency-selective channels, the work [5] focused on codes with time-domain equalization [26]. Through linear symbol precoding, blind schemes for OFDM-based STBC were shown in [27] and [40]. The method [27] resorts to zero-padding for removing interblock interference and is applicable only for constant-modulus sources and channel pairs without common zeros; the one in [40], instead, uses cyclic prefix (CP) as guard interval and leverages redundant precoding to relieve the source and channel-zero constraints imposed in [27]. For FDE-STBC, training-based channel estimation has recently been considered in [11]. It is known that single-carrier FDE systems fall within the class of precoded OFDM, with the fast Fourier transform (FFT) matrix as the precoder [25]. In view of this fact, the method in [40] for OFDM scenario also provides an immediate blind solution for FDE-STBC: one just chooses the FFT precoding matrix to convert the multicarrier transmission into a single-carrier scheme and then inserts certain redundancy into the symbol streams to facilitate channel identification. The price to be paid for this approach, however, would be the loss in the effective data rate.

B. Paper Contributions

This paper proposes a blind channel estimation scheme for FDE-STBC systems in a two-transmit-antennas/single-receive-antenna environment. The proposed approach relies on *nonredundant* diagonal precoding (hence, preserving the baud rate), assumes independent and identically distributed (i.i.d.) source statistics (irrespective of constellation modulus), and does not impose constraints on subchannel zero locations. It exploits the precoding-induced linear signal structure in the time-domain conjugate cross correlation between the two temporal receive branches, as well as the circulant channel matrix property after CP is discarded. Specifically, we show that the set of linear equations relating the entries of conjugate cross-correlation matrix and *products* of channel impulse responses can be rearranged into one with a block-circulant with circulant-block (BCCB) structure. The products of channel taps are first obtained by solving this linear equation set; the channel pair is then simultaneously identified, up to a 2×2 complex matrix ambiguity, as the dominant left singular vectors of an associated rank-two matrix. A similar “bilinear” estimation strategy has also been adopted in [13], [21], [24], and [37]. In our formulation, a natural sufficient condition for unique channel recovery is the nonsingularity of certain BCCB matrix with precoder coefficients as its entries. Channel identifiability is thus free from any *a priori* assumptions on subchannel characteristics and is shown to be fulfilled by almost all choices of precoders. As long as the channel noise is circularly Gaussian and the received data statistic is perfectly obtained, the resultant channel estimate is exact. In the presence of finite-sample estimation error, the proposed channel estimation framework allows a natural precoder design formulation for improving solution robustness. We consider two models of data mismatch—one as an unknown deterministic perturbation while the other statistically as a white noise—and propose the associated optimal precoder design criteria, aiming for minimizing the worst-case solution sensitivity to perturbation and mean-square errors, respectively. Both optimization problems are further formulated to take

advantage of the BCCB system matrix property and are then analytically solved; the resultant solutions are shown to be the same two-level form precoder. Pairwise error probability (PEP) analysis is conducted to investigate the equalization performance of the proposed optimal solution and characterize the associated design tradeoff. It is noted that blind channel estimation via nonredundant diagonal precoding has been considered in the single channel case [10], [24], [31], [37]; the related generalizations to MIMO single- and multicarrier spatial multiplexing systems can be found in [8] and [9]. The rest of this paper is organized as follows. Section II briefly describes the system model and the underlying assumptions. Section III presents the proposed method; the associated key features are investigated in Section IV. Section V addresses the optimal precoder design against imperfect data estimation. Section VI examines the equalization performance through PEP analysis. Section VII contains the simulation results. Finally, Section VIII is the conclusion.

Notation List: Let $\mathbb{R}^{m \times n}$ and $\mathbb{C}^{m \times n}$ be, respectively, the sets of $m \times n$ real and complex matrices. Denote by $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$, respectively, the transpose, complex conjugate, and Hermitian operations. The symbols $\mathbf{I}_m$ and $\mathbf{0}_m$ denote the $m \times m$ identity and zero matrices; $\mathbf{0}_{m \times n}$ is the $m \times n$ zero matrix. The notation $\otimes$ stands for the Kronecker product [19, p. 242]. For $\mathbf{X} \in \mathbb{C}^{m \times n}$ with $\mathbf{x}_j \in \mathbb{C}^m$ being the j th column, define $\text{vec}(\mathbf{X}) := [\mathbf{x}_1^T \dots \mathbf{x}_n^T]^T \in \mathbb{C}^{mn}$. For $\mathbf{x} \in \mathbb{C}^m$, let $\text{diag}\{\mathbf{x}\}$ be the $m \times m$ diagonal matrix with the elements of $\mathbf{x}$ on the main diagonal. The notation E_y stands for the expected value of the random variable y , and $j := \sqrt{-1}$. We denote by $\mathbf{F} \in \mathbb{C}^{N \times N}$ the FFT matrix with the k lth entry $[\mathbf{F}]_{k,l} := 1/\sqrt{N} \cdot w^{-(k-1)(l-1)}$, where $w := \exp(j2\pi/N)$, $1 \leq k, l \leq N$. We denote by $\|\cdot\|$ the two norm of a vector by $k(\mathbf{X})$ and $\text{Tr}[\mathbf{X}]$ the condition number and trace of matrix $\mathbf{X}$.

II. SYSTEM MODEL AND BASIC ASSUMPTIONS

We consider the discrete-time baseband model of an FDE-STBC system [2] over frequency-selective channels as shown in Fig. 1. Let $\mathbf{s}_k$ and $\mathbf{s}_{k+1}$ be two N -dimensional symbol blocks to be transmitted. Prior to the STBC encoder, each symbol block is precoded by an $N \times N$ diagonal matrix

$$\mathbf{P} := \text{diag}\left\{[p(0) \dots p(N-1)]^T\right\} \quad (2.1)$$

with $p(n) \in \mathbb{R}$, to obtain

$$\mathbf{x}_l = \mathbf{P}\mathbf{s}_l, \text{ for } l = k, k+1 \quad (2.2)$$

which are then spatially and temporally coded according to [2] for transmit diversity as well as for mitigating the multipath channel distortion. For $1 \leq i \leq 2$, let $h_i(n)$ be the impulse response of the channel between the i th transmit antenna and the receive antenna. In terms of block signals, the input–output relations in time domain are described as [2]

$$\mathbf{y}_k = \mathbf{G}_1\mathbf{x}_k + \mathbf{G}_2\mathbf{x}_{k+1} + \mathbf{v}_k \quad (2.3)$$

and

$$\mathbf{y}_{k+1} = \mathbf{G}_1\tilde{\mathbf{x}}_{k+1} - \mathbf{G}_2\tilde{\mathbf{x}}_k + \mathbf{v}_{k+1} \quad (2.4)$$

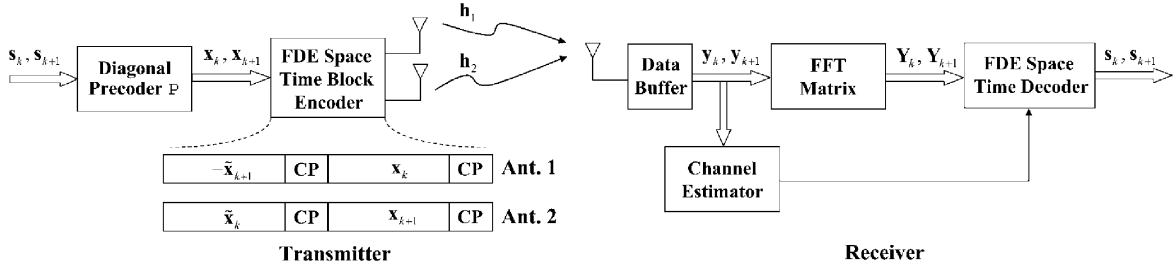


Fig. 1. FDE-STBC system diagram.

where, for $l = k, k+1$, $\mathbf{y}_l$ and $\mathbf{v}_l$ are the received signal (upon CP removal) and noise, $\tilde{\mathbf{x}}_l$ is the time-reversed and element-wise conjugated version associated with $\mathbf{x}_l$, that is

$$\tilde{\mathbf{x}}_l(n) = \mathbf{x}_l^*((-n)_N), \quad 0 \leq n \leq N-1 \quad (2.5)$$

and $\mathbf{G}_i \in \mathbb{C}^{N \times N}$ is circulant with²

$$\mathbf{g}_i := [h_i(0) \quad \dots \quad h_i(L) \quad 0 \quad \dots \quad 0]^T \in \mathbb{C}^N \quad (2.6)$$

as the first column, $1 \leq i \leq 2$. Since $\mathbf{G}_i$ is circulant, we have $\mathbf{G}_i = \mathbf{F}^H \mathbf{D}_i \mathbf{F}$, where $\mathbf{D}_i$ is diagonal with $[\mathbf{D}_i]_{mm} = \sum_{n=0}^L h_i(n) w^{-(m-1)n}$, $1 \leq m \leq N$. Let us define $\mathbf{Y}_l := \mathbf{F} \mathbf{y}_l$, $\mathbf{X}_l := \mathbf{F} \mathbf{x}_l$, and $\mathbf{V}_l := \mathbf{F} \mathbf{v}_l$, for $l = k, k+1$. Then the frequency-domain representation associated with (2.3) and (2.4) can be expressed in a compact vector-matrix form as [2]

$$\begin{bmatrix} \mathbf{Y}_k \\ \mathbf{Y}_{k+1}^* \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_2 \\ -\mathbf{D}_2^* & \mathbf{D}_1^* \end{bmatrix}}_{:=\mathbf{D}} \begin{bmatrix} \mathbf{X}_k \\ \mathbf{X}_{k+1} \end{bmatrix} + \begin{bmatrix} \mathbf{V}_k \\ \mathbf{V}_{k+1}^* \end{bmatrix} \quad (2.7)$$

through space-frequency matched filtering using the effective channel matrix $\mathbf{D}$, we get

$$\mathbf{D}^H \begin{bmatrix} \mathbf{Y}_k \\ \mathbf{Y}_{k+1}^* \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{D}} & \mathbf{0}_N \\ \mathbf{0}_N & \tilde{\mathbf{D}} \end{bmatrix} \begin{bmatrix} \mathbf{X}_k \\ \mathbf{X}_{k+1} \end{bmatrix} + \mathbf{D}^H \begin{bmatrix} \mathbf{V}_k \\ \mathbf{V}_{k+1}^* \end{bmatrix} \quad (2.8)$$

where $\tilde{\mathbf{D}} := \mathbf{D}_1^H \mathbf{D}_1 + \mathbf{D}_2^H \mathbf{D}_2 \in \mathbb{R}^{N \times N}$ is diagonal with $[\tilde{\mathbf{D}}]_{ii} = |\mathbf{D}_1]_{ii}|^2 + |\mathbf{D}_2]_{ii}|^2$, $1 \leq i \leq N$: this asserts that twofold transmit diversity is achieved in the frequency domain. To recover the source signals, per-tone frequency-domain equalizer [2], [15] can be designed based on (2.8), as long as a channel estimate is available at the receiver. Based on the time-domain signal model (2.3) and (2.4), this paper proposes a blind channel estimation scheme by using the second-order statistics of the received signal and discusses an optimal design of the precoder $p(n)$ for improving channel estimation accuracy. The following assumptions are made in the sequel.

- The source sequence $s(n)$ is i.i.d. with zero mean and $E s(k) s(l)^* = \delta(k-l)$, where $\delta(\cdot)$ is the Kronecker delta function.
- The noise $v(n)$ is white circular Gaussian with zero mean, variance σ_v^2 , and is independent of the source sequence $s(n)$.

²Without loss of generality, we may take L as a common channel order, or simply an associated upper bound, which can be determined as the maximum among the two.

III. BLIND CHANNEL ESTIMATION

To introduce the proposed method, we first assume that all the signal statistics can be perfectly obtained; the case with imperfect data estimation will be treated later. To obtain the channel matrix $\mathbf{D}$, one may focus on direct estimation of the N tones of the channel frequency response. Since the block length N could be large, this strategy would involve considerable computational efforts. Hence, we propose to instead estimate the time-domain channel impulse response $h_i(n)$, for $0 \leq n \leq L$ and $1 \leq i \leq 2$; the gains of the associated frequency tones can then be obtained by using FFT operations.

A. Identification Equations

The proposed approach exploits the imbedded linear signal structure in the time-domain conjugate cross-correlation matrix of the two received signals $\mathbf{y}_k$ and $\mathbf{y}_{k+1}$ as well as the circulant property of the channel matrix $\mathbf{G}_i$. To proceed, let us first define the matrix

$$\Gamma := \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \vdots & \dots & 1 \\ \vdots & 0 & \dots & 1 \\ 0 & 1 & \dots & 0 \end{bmatrix} \in \mathbb{R}^{N \times N}. \quad (3.1)$$

Then, from (2.5), it is easy to see

$$\tilde{\mathbf{x}}_k = \Gamma \mathbf{x}_k^* \text{ and } \tilde{\mathbf{x}}_{k+1} = \Gamma \mathbf{x}_{k+1}^*. \quad (3.2)$$

With (2.2) and (3.2), the signal models (2.3) and (2.4) then become

$$\mathbf{y}_k = \mathbf{G}_1 \mathbf{P} \mathbf{s}_k + \mathbf{G}_2 \mathbf{P} \mathbf{s}_{k+1} + \mathbf{v}_k \quad (3.3)$$

and

$$\mathbf{y}_{k+1} = \mathbf{G}_1 \Gamma \mathbf{P} \mathbf{s}_{k+1}^* - \mathbf{G}_2 \Gamma \mathbf{P} \mathbf{s}_k^* + \mathbf{v}_{k+1}. \quad (3.4)$$

From (3.3), (3.4), and by assumptions a) and b), it is easy to check the following:

$$\begin{aligned} \mathbf{R}_y(1) &:= E \mathbf{y}_k \mathbf{y}_{k+1}^T \\ &= \mathbf{G}_2 \mathbf{P}^2 \Gamma \mathbf{G}_1^T - \mathbf{G}_1 \mathbf{P}^2 \Gamma \mathbf{G}_2^T \\ &\quad + \{ \mathbf{G}_1 \mathbf{P} E \mathbf{s}_k \mathbf{v}_{k+1}^T + \mathbf{G}_2 \mathbf{P} E \mathbf{s}_{k+1} \mathbf{v}_k^T \\ &\quad + E \mathbf{v}_k \mathbf{v}_{k+1}^T \}. \end{aligned} \quad (3.5)$$

Since the noise $v(n)$ is circular, we have $E \mathbf{v}_k \mathbf{v}_{k+1}^T = \mathbf{0}_N$. Also, we assume that both the real and imaginary components of the noise process are independent of those of the source sequence

$s(n)$: this thus implies $Es_k \mathbf{v}_{k+1}^T = Es_{k+1} \mathbf{v}_{k+1}^T = \mathbf{0}_N$. Under these conditions, the noise contributions to the conjugate cross-correlation matrix $\mathbf{R}_y(1)$ in (3.5) become a zero matrix, leading to

$$\mathbf{R}_y(1) = \mathbf{G}_2 \mathbf{P}^2 \Gamma \mathbf{G}_1^T - \mathbf{G}_1 \mathbf{P}^2 \Gamma \mathbf{G}_2^T. \quad (3.6)$$

For a given $\mathbf{R}_y(1)$, the matrix (3.6) defines a set of N^2 scalar equations nonlinear in the unknowns $h_i(0), \dots, h_i(L)$, $1 \leq i \leq 2$ but is linear with respect to *product* channel coefficients $h_i(k)h_{\bar{i}}(l)$, $1 \leq i, \bar{i} \leq 2$. As a result, in lieu of directly solving for $h_i(0), \dots, h_i(L)$, we propose to exploit the imbedded linear structure in $\mathbf{R}_y(1)$ for channel estimation. This will be done by further taking into account the circulant property of the channel matrices $\mathbf{G}_i$'s.

Specifically, define the following permutation matrix:

$$\mathbf{J} := \begin{bmatrix} \mathbf{0}_{1 \times (N-1)} & 1 \\ \mathbf{I}_{N-1} & \mathbf{0}_{(N-1) \times 1} \end{bmatrix} \in \mathbb{R}^{N \times N}. \quad (3.7)$$

Since $\mathbf{G}_i$ is circulant, it can be expressed in terms of its first column (cf. (2.6)) as

$$\mathbf{G}_i = [\mathbf{g}_i \quad \mathbf{J}\mathbf{g}_i \quad \dots \quad \mathbf{J}^{N-2}\mathbf{g}_i \quad \mathbf{J}^{N-1}\mathbf{g}_i], \quad 1 \leq i \leq 2. \quad (3.8)$$

By definitions of $\mathbf{P}$ and Γ (see (2.1) and (3.1)) and from (3.8), it follows that

$$\begin{aligned} \mathbf{G}_2 \mathbf{P}^2 \Gamma \mathbf{G}_1^T &= (\mathbf{G}_2 \mathbf{P}^2) \cdot (\Gamma \mathbf{G}_1^T) \\ &= \underbrace{[p(0)^2 \mathbf{g}_2 \quad p(1)^2 \mathbf{J}\mathbf{g}_2 \quad \dots \quad p(N-1)^2 \mathbf{J}^{N-1}\mathbf{g}_2]}_{=\mathbf{G}_2 \mathbf{P}^2} \\ &\quad \cdot \underbrace{[\mathbf{g}_1 \quad \mathbf{J}^{N-1}\mathbf{g}_1 \quad \dots \quad \mathbf{J}\mathbf{g}_1]^T}_{=\Gamma \mathbf{G}_1^T} \\ &= \sum_{n=0}^{N-1} p(n)^2 \mathbf{J}^n \mathbf{g}_2 \mathbf{g}_1^T (\mathbf{J}^T)^{N-n}. \end{aligned} \quad (3.9)$$

Similarly, we have

$$\mathbf{G}_1 \mathbf{P}^2 \Gamma \mathbf{G}_2^T = \sum_{n=0}^{N-1} p(n)^2 \mathbf{J}^n \mathbf{g}_1 \mathbf{g}_2^T (\mathbf{J}^T)^{N-n}. \quad (3.10)$$

Combining (3.9) and (3.10), $\mathbf{R}_y(1)$ in (3.6) becomes

$$\mathbf{R}_y(1) = \sum_{n=0}^{N-1} p(n)^2 \mathbf{J}^n \tilde{\mathbf{G}} (\mathbf{J}^T)^{N-n} \quad (3.11)$$

where³

$$\tilde{\mathbf{G}} := \mathbf{g}_2 \mathbf{g}_1^T - \mathbf{g}_1 \mathbf{g}_2^T = [\mathbf{g}_2 \quad \mathbf{g}_1] \cdot \begin{bmatrix} \mathbf{g}_1^T \\ -\mathbf{g}_2^T \end{bmatrix} \in \mathbb{C}^{N \times N}. \quad (3.12)$$

With $\mathbf{g}_i$ given in (2.6), the matrix $\tilde{\mathbf{G}}$ is seen to contain the product channel impulse responses of the form $h_2(k)h_1(l) - h_1(k)h_2(l)$, $0 \leq k, l \leq L$, which are to be determined from (3.11). Toward a tractable procedure for computing $\tilde{\mathbf{G}}$, we observe that $\mathbf{R}_y(1)$ in (3.11) is a weighted sum of N matrices of the form $\mathbf{J}^n \tilde{\mathbf{G}} (\mathbf{J}^T)^{N-n}$, in which the unknown $\tilde{\mathbf{G}}$ are pre-, and post-, multiplied by the known matrices $\mathbf{J}^n$ and $(\mathbf{J}^T)^{N-n}$. Based on this structural property, we can further rearrange (3.11) into a standard linear equation form. This is done via the next lemma.

Lemma 3.1 [19, p. 255]: The matrix equation $\sum_{k=1}^K \mathbf{A}_k \mathbf{X} \mathbf{B}_k = \mathbf{C}$ can be equivalently expressed as $\left[\sum_{k=1}^K \mathbf{B}_k^T \otimes \mathbf{A}_k \right] \text{vec}(\mathbf{X}) = \text{vec}(\mathbf{C})$. $\square$

Based on Lemma 3.1, we can immediately rewrite (3.11) as

$$\left[\sum_{n=0}^{N-1} p(n)^2 \mathbf{J}^{N-n} \otimes \mathbf{J}^n \right] \text{vec}(\tilde{\mathbf{G}}) = \text{vec}(\mathbf{R}_y(1)). \quad (3.13)$$

By definitions of the Kronecker product and $\mathbf{J}$ in (3.7), (3.13) turns out to be (3.14), shown at the bottom of the page.

The $N^2 \times N^2$ real-valued matrix $\mathbf{Q}$ defined in (3.14), which is characterized by the N circulant matrices $\{p(0)^2 \mathbf{I}_N, p(1)^2 \mathbf{J}, \dots, p(N-1)^2 \mathbf{J}^{N-1}\}$ on the top row block, is BCCB [12, p. 184]. Equation (3.14) forms the basis of the proposed approach.

B. Identification of Channel Impulse Response

Assume that $\text{vec}(\tilde{\mathbf{G}})$, and hence the matrix $\tilde{\mathbf{G}}$, can be uniquely recovered from the linear equation (3.14); the uniqueness condition and the computational issue will be investigated in the next section. We then collect the product unknowns $h_2(k)h_1(l) - h_1(k)h_2(l)$, $0 \leq k, l \leq L$ to form the following $(L+1) \times (L+1)$ matrix:

$$\mathbf{H} := [\mathbf{H}_{k,l}]_{0 \leq k, l \leq L}, \quad \mathbf{H}_{k,l} = h_2(k)h_1(l) - h_1(k)h_2(l). \quad (3.15)$$

³We assume that the two channel impulse response vectors are linearly independent, for otherwise $\tilde{\mathbf{G}}$ is identically a zero matrix; this assumption holds whenever the environment is with sufficiently rich scattering.

$$\underbrace{\begin{bmatrix} p(0)^2 \mathbf{I}_N & p(1)^2 \mathbf{J} & \dots & p(N-2)^2 \mathbf{J}^{N-2} & p(N-1)^2 \mathbf{J}^{N-1} \\ p(N-1)^2 \mathbf{J}^{N-1} & p(0)^2 \mathbf{I}_N & \dots & p(N-3)^2 \mathbf{J}^{N-3} & p(N-2)^2 \mathbf{J}^{N-2} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ p(2)^2 \mathbf{J}^2 & p(3)^2 \mathbf{J}^3 & \dots & p(0)^2 \mathbf{I}_N & p(1)^2 \mathbf{J} \\ p(1)^2 \mathbf{J} & p(2)^2 \mathbf{J}^2 & \dots & p(N-1)^2 \mathbf{J}^{N-1} & p(0)^2 \mathbf{I}_N \end{bmatrix}}_{:=\mathbf{Q}} \text{vec}(\tilde{\mathbf{G}}) = \text{vec}(\mathbf{R}_y(1)). \quad (3.14)$$

Observe that the matrix $\mathbf{H}$ is of rank two and can be factorized as

$$\mathbf{H} = \mathbf{h}_2 \mathbf{h}_1^T - \mathbf{h}_1 \mathbf{h}_2^T = [\mathbf{h}_1 \quad \mathbf{h}_2] \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \end{bmatrix} \quad (3.16)$$

where

$$\mathbf{h}_i := [h_i(0) \quad h_i(1) \quad \dots \quad h_i(L)]^T \in \mathbb{C}^{L+1}, \quad 1 \leq i \leq 2 \quad (3.17)$$

is the desired channel impulse response vectors. Based on (3.16), the channels can thus be identified, up to a 2×2 complex matrix of the form

$$\mathbf{U} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \text{with } ad - bc = 1 \quad (3.18)$$

by computing the two dominant left singular vectors associated with $\mathbf{H}$; the inherent matrix ambiguity must satisfy (3.18) since, for any vector pair of the form $\bar{\mathbf{h}} = [\mathbf{h}_1 \quad \mathbf{h}_2] \mathbf{U}$ with $\mathbf{U} \in \mathbb{C}^{2 \times 2}$, we have

$$\bar{\mathbf{h}} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \bar{\mathbf{h}}^T = \mathbf{h}_2 \mathbf{h}_1^T - \mathbf{h}_1 \mathbf{h}_2^T \quad (3.19)$$

whenever $\mathbf{U}$ verifies (3.18). We note that a similar matrix outer-product-based approach for blind channel estimation is also adopted in [13], [21], [24], [37].

C. On Ambiguity Removal

The matrix ambiguity (3.18) can be resolved through insertion of additional pilot symbols. To see this, let $[\bar{\mathbf{h}}_1 \quad \bar{\mathbf{h}}_2]$ be a dominant left singular vector pair associated with the rank-two matrix $\mathbf{H}$ defined in (3.15). Then, we have $[\bar{\mathbf{h}}_1 \quad \bar{\mathbf{h}}_2] = [\mathbf{h}_1 \quad \mathbf{h}_2] \mathbf{U}$, with $\mathbf{U} \in \mathbb{C}^{2 \times 2}$ fulfilling (3.18); this implies

$$\begin{aligned} [\mathbf{h}_1 \quad \mathbf{h}_2] &= [\bar{\mathbf{h}}_1 \quad \bar{\mathbf{h}}_2] \mathbf{U}^{-1} \\ &= [\bar{\mathbf{h}}_1 \quad \bar{\mathbf{h}}_2] \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} \\ &= [\bar{\mathbf{h}}_1 \quad \bar{\mathbf{h}}_2] \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \end{aligned} \quad (3.20)$$

Since both $\mathbf{G}_1$ and $\mathbf{G}_2$ are circulant, the first output branch (2.3), at some $k = k_0$, can be alternatively expressed as

$$\mathbf{y}_{k_0} = \mathbf{C}_{k_0} \mathbf{g}_1 + \mathbf{C}_{k_0+1} \mathbf{g}_2 + \mathbf{v}_{k_0} \quad (3.21)$$

where $\mathbf{g}_i$ ($i = 1, 2$) is the zero-padded channel impulse response as in (2.6), and $\mathbf{C}_l \in \mathbb{C}^{N \times N}$ is circulant with the precoded symbol vector $\mathbf{x}_l$ as the first column, $l = k_0, k_0 + 1$. Let us write

$$\mathbf{g}_i = [\mathbf{h}_i^T \quad \mathbf{0}_{1 \times (N-L-1)}]^T, \quad 1 \leq i \leq 2 \quad (3.22)$$

where $\mathbf{h}_i$ is the desired channel impulse response vector defined in (3.17). With (3.22), (3.21) is then reduced to

$$\mathbf{y}_{k_0} = \tilde{\mathbf{C}}_{k_0} \mathbf{h}_1 + \tilde{\mathbf{C}}_{k_0+1} \mathbf{h}_2 + \mathbf{v}_{k_0} \quad (3.23)$$

where $\tilde{\mathbf{C}}_l \in \mathbb{C}^{N \times (L+1)}$ contains the first $L+1$ columns of $\mathbf{C}_l$. With (3.20), we can write (3.23) in terms of the scalar ambiguities as

$$\begin{aligned} \mathbf{y}_{k_0} &= \tilde{\mathbf{C}}_{k_0} (d\bar{\mathbf{h}}_1 - c\bar{\mathbf{h}}_2) + \tilde{\mathbf{C}}_{k_0+1} (-b\bar{\mathbf{h}}_1 + a\bar{\mathbf{h}}_2) + \mathbf{v}_{k_0} \\ &= \underbrace{[\tilde{\mathbf{C}}_{k_0} \bar{\mathbf{h}}_1 \quad -\tilde{\mathbf{C}}_{k_0} \bar{\mathbf{h}}_2 \quad -\tilde{\mathbf{C}}_{k_0+1} \bar{\mathbf{h}}_1 \quad \tilde{\mathbf{C}}_{k_0+1} \bar{\mathbf{h}}_2]}_{:=\mathbf{T}} \\ &\quad \times [d \quad c \quad b \quad a]^T + \mathbf{v}_{k_0}. \end{aligned} \quad (3.24)$$

It is noted that, subject to the constraint $ad - bc = 1$, there are only three independent unknowns in (3.24). One can just solve for, say (b, c, d) , from (3.24) and then determine a via the non-linear equation $a = (1 + bc)/d$; this, however, would be more prone to error propagation. Hence, we propose to instead compute (a, b, c, d) all at once from (3.24). Toward this end, pilot symbols should be appropriately inserted to produce *at least* four training components in $\mathbf{y}_{k_0}$. We observe that each column of $\mathbf{T}$ in (3.24) is a linear combination of $L+1$ circularly shifted symbol vector $\mathbf{x}_l$ for some $l \in \{k_0, k_0+1\}$. The cyclicity structural constraint implies at least $L+4$ pilot symbols are needed in both $\mathbf{x}_l$. One plausible placement, in particular, is to insert four (and L , respectively) consecutive pilots at the head (and tail) of $\mathbf{x}_l$, $l = k_0, k_0+1$; in this way, the first four components in $\mathbf{y}_{k_0}$, denoted by $\mathbf{y}_t$, then act as training data and the scalar unknowns are estimated via

$$[\hat{d} \quad \hat{c} \quad \hat{b} \quad \hat{a}]^T = \mathbf{T}_t^{-1} \mathbf{y}_t \quad (3.25)$$

where $\mathbf{T}_t \in \mathbb{C}^{4 \times 4}$ contains the first four rows of $\mathbf{T}$. Hence, even though the proposed blind method reduces the number of unknown channel parameters from $2L+2$ to three, no less than $2L+8$ pilot symbols are nonetheless required for ambiguity removal. This is due to the nonredundant precoding based channel estimation formulation as well as the circulant signal structure (the proposed channel estimation procedures are outlined in Table I).

IV. IDENTIFIABILITY AND PRODUCT UNKNOWN COMPUTATION

This section first specifies the channel identifiability condition, and then introduces two methods for computing the product channel coefficients. The presented results also lay the foundation for further investigating the optimal precoder design problem.

A. Channel Identifiability

From the previous discussions, it is easy to see that the channel can be identified if $\text{vec}(\hat{\mathbf{G}})$ is uniquely determined from (3.14): this is true if the matrix $\mathbf{Q}$ is nonsingular. By exploiting the BCCB property of $\mathbf{Q}$, the following theorem explicitly shows the associated eigenvalues, and in turn specifies the condition for $\mathbf{Q}$ to be nonsingular. Roughly speaking, if we define the vector

$$\mathbf{p} := [p(0)^2 \quad p(1)^2 \quad \dots \quad p(N-1)^2]^T \in \mathbb{R}^N \quad (4.1)$$

TABLE I
ALGORITHM SUMMARY

1) Estimate the conjugate cross-correlation matrix $\mathbf{R}_y(1)$ via
$\hat{\mathbf{R}}_y(1) := \frac{1}{K} \sum_{i=1}^K \mathbf{y}(i) \mathbf{y}(i+1)^T$
where K is the number of temporal received symbol block pairs.
2) Form the matrix equation (3.14), and select a precoder $\mathbf{p}(n)$ with which the matrix $\mathbf{Q}$ is nonsingular.
3) Compute the product channel coefficients based on (4.3) or (4.8).
4) Form the rank-two matrix $\mathbf{H}$ as in (3.15). The channel impulse response pair is then computed, up to a 2×2 matrix ambiguity, as the two dominant left singular vectors associated with $\mathbf{H}$.
5) Remove the matrix ambiguity by inserting extra training symbols as suggested in (3.25).

then the N^2 eigenvalues of $\mathbf{Q}$ are completely determined by the N eigenvalues associated with the $N \times N$ circulant matrix with $\mathbf{p}^T$ as the first row (the proof of theorem is given in Appendix A).

Theorem 4.1: Associated with the vector $\mathbf{p}$ in (4.1), we define the polynomial

$$\mathbf{p}(z) := p(0)^2 + p(1)^2 z^{-1} + \dots + p(N-1)^2 z^{-(N-1)}. \quad (4.2)$$

Then the N^2 eigenvalues of the matrix $\mathbf{Q}$ defined in (3.14) are given by the N replicas of the N -tuple $\{\mathbf{p}(1), \mathbf{p}(w), \dots, \mathbf{p}(w^{N-1})\}$. $\square$

Theorem 4.1 shows that channel identifiability is guaranteed whenever $\mathbf{p}(w^n) \neq 0$ for all $0 \leq n \leq N-1$; this condition is quite mild and can hold for almost all choices of $p(n)$. We should note that the significance of Theorem 4.1 is far above just characterizing a sufficient condition for unique channel recovery. It moreover specifies the eigenvalues associated with the matrix $\mathbf{Q}$: this result will be exploited for selecting $p(n)$ to improve the reliability of channel estimate against the finite-sample estimation error (see Section V).

B. Computation of $\text{vec}(\tilde{\mathbf{G}})$

A crucial step for implementing the proposed channel estimation scheme is the computation of the product channel coefficient vector $\text{vec}(\tilde{\mathbf{G}})$ based on (3.14). In what follows we propose two methods for fulfilling this task.

- i) *Direct matrix inversion:* An immediate approach to solving (3.14) is through direct matrix inversion so that

$$\text{vec}(\tilde{\mathbf{G}}) = \mathbf{Q}^{-1} \text{vec}(\mathbf{R}_y(1)). \quad (4.3)$$

Observe from (3.14) that $\mathbf{Q}$ is BCCB and is characterized by the particular set of circulant matrices $\{p(0)^2 \mathbf{I}_N, p(1)^2 \mathbf{J}, \dots, p(N-1)^2 \mathbf{J}^{N-1}\}$. This appealing structure allows for a potentially low-complexity implementation via FFT operations. In Appendix B, we derive a simple closed-form expression of $\mathbf{Q}^{-1}$ based on which this figure of merit is justified.

- ii) *Solution via zero entry removal in (3.14):* It is noted from (2.6) that, for $1 \leq i \leq 2$, the vector $\mathbf{g}_i$ contains $L+1$ channel impulse response $h_i(n)$, $0 \leq n \leq L$, followed by $N-L-1$ trailing zeros. As a result, the $N^2 \times N^2$ matrix $\tilde{\mathbf{G}} (= \mathbf{g}_2 \mathbf{g}_1^T - \mathbf{g}_1 \mathbf{g}_2^T)$, and hence the associated vectorized representation $\text{vec}(\tilde{\mathbf{G}})$, has actually $(L+1)^2$ nonzero product unknowns. By removing the zero entries

in $\text{vec}(\tilde{\mathbf{G}})$, and the corresponding indexed columns of the matrix $\mathbf{Q}$, (3.14) can be simplified to a set of N^2 scalar equations in $(L+1)^2$ unknowns. Indeed, with $\mathbf{g}_i$ defined in (3.22), we have

$$\mathbf{g}_i \mathbf{g}_i^H = \begin{bmatrix} \mathbf{h}_i \mathbf{h}_i^H & \mathbf{0}_{(L+1) \times (N-L-1)} \\ \mathbf{0}_{(N-L-1) \times (L+1)} & \mathbf{0}_{N-L-1} \end{bmatrix}, \quad 1 \leq i, \bar{i} \leq 2 \quad (4.4)$$

and hence

$$\begin{aligned} \tilde{\mathbf{G}} &= \mathbf{g}_2 \mathbf{g}_1^T - \mathbf{g}_1 \mathbf{g}_2^T \\ &= \begin{bmatrix} \mathbf{H} & \mathbf{0}_{(L+1) \times (N-L-1)} \\ \mathbf{0}_{(N-L-1) \times (L+1)} & \mathbf{0}_{N-L-1} \end{bmatrix} \end{aligned} \quad (4.5)$$

where $\mathbf{H}$ is defined in (3.15). Based on (4.5) and by definition of the $\text{vec}(\cdot)$ operation, (3.14) can be shown (after some direct manipulations) to be reduced into

$$\underbrace{\mathbf{Q} \mathbf{J}_1 (\mathbf{I}_{L+1} \otimes \mathbf{J}_2)}_{:= \tilde{\mathbf{Q}}} \text{vec}(\mathbf{H}) = \text{vec}(\mathbf{R}_y(1)) \quad (4.6)$$

in which

$$\begin{aligned} \mathbf{J}_1 &= \begin{bmatrix} \mathbf{I}_{N(L+1)} \\ \mathbf{0}_{N(N-L-1) \times N(L+1)} \end{bmatrix} \in \mathbb{R}^{N^2 \times N(L+1)} \text{ and} \\ \mathbf{J}_2 &:= \begin{bmatrix} \mathbf{I}_{L+1} \\ \mathbf{0}_{(N-L-1) \times (L+1)} \end{bmatrix} \in \mathbb{R}^{N \times (L+1)}. \end{aligned} \quad (4.7)$$

The matrix $\tilde{\mathbf{Q}} \in \mathbb{R}^{N^2 \times (L+1)^2}$ in (4.7) is obtained by deleting $N^2 - (L+1)^2$ columns from $\mathbf{Q}$. It is thus of full column rank whenever $\mathbf{Q}$ is nonsingular and, if so, the product channel coefficients can be computed via

$$\text{vec}(\mathbf{H}) = (\tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}})^{-1} \tilde{\mathbf{Q}}^T \text{vec}(\mathbf{R}_y(1)). \quad (4.8)$$

Compared with the direct matrix inversion method (4.3), the solution (4.8) can yield better estimation accuracy at the expense of computational complexity (see Appendix B for complexity evaluation). Based on (4.3) and (4.8), the selection of precoder $p(n)$ for optimal numerical robustness is discussed next.

V. OPTIMAL PRECODER DESIGN

If the conjugate cross-correlation matrix $\mathbf{R}_y(1)$ is perfectly obtained, both solutions (4.3) and (4.8) are exact. In practice, however, only a finite number of data samples can be used for

estimating $\mathbf{R}_y(1)$; (3.14) and (4.6) should be accordingly modified as

$$\text{vec}(\hat{\mathbf{R}}_y(1)) = \mathbf{Q}\text{vec}(\hat{\mathbf{G}}) + \mathbf{w} \quad (5.1)$$

and

$$\text{vec}(\hat{\mathbf{R}}_y(1)) = \tilde{\mathbf{Q}}\text{vec}(\mathbf{H}) + \mathbf{w} \quad (5.2)$$

where $\hat{\mathbf{R}}_y(1)$ is an estimate of $\mathbf{R}_y(1)$ and $\mathbf{w}$ accounts for the data mismatch due to finite-sample estimation. Given the error-corrupted $\hat{\mathbf{R}}_y(1)$, it is impossible to recover the actual product channel coefficients. Instead, with (5.1) and (5.2), the estimated solutions are respectively

$$\text{vec}(\hat{\mathbf{G}}) := \mathbf{Q}^{-1}\text{vec}(\hat{\mathbf{R}}_y(1)) = \text{vec}(\tilde{\mathbf{G}}) + \mathbf{Q}^{-1}\mathbf{w} \quad (5.3)$$

and

$$\begin{aligned} \text{vec}(\hat{\mathbf{H}}) &:= (\tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}})^{-1} \tilde{\mathbf{Q}}^T \text{vec}(\hat{\mathbf{R}}_y(1)) \\ &= \text{vec}(\mathbf{H}) + (\tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}})^{-1} \tilde{\mathbf{Q}}^T \mathbf{w}. \end{aligned} \quad (5.4)$$

In what follows, we consider two different modeling schemes of $\mathbf{w}$ and propose the associated optimal criteria for designing $p(n)$ against imperfect data estimation.

A. Minimal Worst-Case Sensitivity to Error Perturbation

We will first treat $\mathbf{w}$ as an unknown “deterministic” perturbation since the statistical property of the data estimation error is in general difficult to characterize. From this standpoint, typical solution robustness measures for (5.3) and (5.4) are the *condition numbers* of the matrices $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$, respectively (see, e.g., [18] and [23]). Small $k(\mathbf{Q})$ and $k(\tilde{\mathbf{Q}})$, in particular, are known to ensure small worst-case sensitivity of the error-perturbed solution to data mismatch [18, p. 338]. Since both $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$ depend entirely on $p(n)$, a natural approach to improving the channel estimation accuracy is to choose $p(n)$ so that both $k(\mathbf{Q})$ and $k(\tilde{\mathbf{Q}})$ are kept as small as possible. This type of optimization problem would seem formidable to tackle since the condition number of a matrix is in general a highly nonlinear function in the entries. Toward a tractable design formulation, we note the crucial fact: since $\tilde{\mathbf{Q}}$ contains a subset of columns of $\mathbf{Q}$ (see (4.6)), it follows that [23, p. 27]

$$k(\tilde{\mathbf{Q}}) \leq k(\mathbf{Q}). \quad (5.5)$$

Inequality (5.5) suggests that, to jointly improve the accuracy of solutions (5.3) and (5.4), it is plausible to just minimize $k(\mathbf{Q})$ because a small $k(\mathbf{Q})$ will also guarantee $k(\tilde{\mathbf{Q}})$ to be small. Such a design strategy, on the one hand, can bypass direct minimization of $k(\tilde{\mathbf{Q}})$, which would appear rather intractable. More important, it will allow us to exploit the eigenvalue characteristics of the BCCB matrix $\mathbf{Q}$ (in Theorem 4.1) to analytically derive a solution, as is shown below. Hence, we specifically propose to minimize $k(\mathbf{Q})$, subject to the following two constraints:

$$\sum_{n=0}^{N-1} p(n)^2 = N \quad (5.6a)$$

and

$$\min p(n)^2 \geq \delta \text{ for some } 0 < \delta < 1. \quad (5.6b)$$

The constraint (5.6a) normalizes the average transmit power within one block to unity, and the constraint (5.6b) imposes a minimal threshold on the floor power. In the context of single channel blind identification based on modulation-induced-cyclostationarity, the two constraints have been used in [10], [24], and [37] for precoder design against the channel noise effect.

To derive the optimal solution, we shall first specify $k(\mathbf{Q})$ in terms of the eigenvalues of the matrix $\mathbf{Q}$. Since $\mathbf{Q}$ is BCCB, it can be factorized as $\mathbf{Q} = (\mathbf{F} \otimes \mathbf{F}) \Lambda (\mathbf{F}^H \otimes \mathbf{F}^H)$ for some diagonal Λ [12, p. 181]. This then implies that $\mathbf{Q}$ is a normal matrix [18, p. 100], as can be seen by

$$\begin{aligned} \mathbf{Q}^H \mathbf{Q} &= (\mathbf{F} \otimes \mathbf{F}) \Lambda^H (\mathbf{F}^H \otimes \mathbf{F}^H) (\mathbf{F} \otimes \mathbf{F}) \Lambda (\mathbf{F}^H \otimes \mathbf{F}^H) \\ &= (\mathbf{F} \otimes \mathbf{F}) \Lambda^H \Lambda (\mathbf{F}^H \otimes \mathbf{F}^H) \\ &= (\mathbf{F} \otimes \mathbf{F}) \Lambda \Lambda^H (\mathbf{F}^H \otimes \mathbf{F}^H) = \mathbf{Q} \mathbf{Q}^H. \end{aligned} \quad (5.7)$$

In deriving (5.7), we have used the identity $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ [19, p. 244]. As $\mathbf{Q}$ is normal, it is known that [18, p. 340]

$$k(\mathbf{Q}) = \rho(\mathbf{Q}) \rho(\mathbf{Q}^{-1}) \quad (5.8)$$

in which $\rho(\mathbf{M}) := \max \{|\lambda| : \lambda \text{ is an eigenvalue of the matrix } \mathbf{M}\}$. Equation (5.8) links the condition number of $\mathbf{Q}$ with the extreme magnitudes of the associated eigenvalues which, according to Theorem 4.1, are exactly the maximum and minimum among the N elements $\{|\mathbf{p}(1)|, |\mathbf{p}(w)|, \dots, |\mathbf{p}(w^{N-1})|\}$, where $\mathbf{p}(z)$ is the polynomial defined in (4.2). More precisely, we have

$$k(\mathbf{Q}) = \frac{\max |\mathbf{p}(w^k)|}{\min |\mathbf{p}(w^k)|}, \text{ for } 0 \leq k \leq N-1. \quad (5.9)$$

To find the minimal $k(\mathbf{Q})$ based on (5.9), we shall further characterize $\mathbf{p}(w^k)$'s under the two constraints in (5.6). With (5.6a), it is easy to see from (4.2) that, for $k = 0$

$$\mathbf{p}(w^0) = \mathbf{p}(1) = \sum_{n=0}^{N-1} p(n)^2 = N. \quad (5.10)$$

The following lemma provides an upper bound on $|\mathbf{p}(w^k)|$ for $1 \leq k \leq N-1$; the result is crucial for deriving the minimal $k(\mathbf{Q})$ (the proof of lemma is shown in Appendix C).

Lemma 5.1: For any $p(n)$ satisfying (5.6a) and (5.6b), we have

$$|\mathbf{p}(w^k)| \leq N(1 - \delta) \text{ for all } 1 \leq k \leq N-1. \quad (5.11)$$

□

With (5.9), (5.10), and (5.11), the minimal achievable $k(\mathbf{Q})$, and the corresponding optimal $p(n)$, are shown in the following theorem.

Theorem 5.2: Under the constraints (5.6a) and (5.6b), the minimal condition number associated with the matrix $\mathbf{Q}$ is given by

$$k_{\min}(\mathbf{Q}) = \frac{1}{1 - \delta}, \quad (5.12)$$

which is attained by the following two-level solution: for a fixed but arbitrary $0 \leq m \leq N - 1$

$$p(m)^2 = N - (N - 1)\delta, \text{ and } p(n)^2 = \delta \text{ for } n \neq m. \quad (5.13)$$

Proof: We claim that i) $\kappa(\mathbf{Q}) \geq 1/(1 - \delta)$ for any $p(n)$ satisfying (5.6a) and (5.6b), and ii) equality is attained by the solution (5.13); the theorem thus follows. To show claim i), it is noted from (5.9) and (5.10) that

$$k(\mathbf{Q}) = \frac{\max |\mathbf{p}(w^k)|}{\min |\mathbf{p}(w^k)|} \geq \frac{|\mathbf{p}(w^0)|}{\min |\mathbf{p}(w^k)|} = \frac{N}{\min |\mathbf{p}(w^k)|}. \quad (5.14)$$

Also, (5.10) and (5.11) imply

$$\min |\mathbf{p}(w^k)| \leq N(1 - \delta), \text{ i.e., } \frac{1}{\min |\mathbf{p}(w^k)|} \geq \frac{1}{N(1 - \delta)}. \quad (5.15)$$

Claim i) then follows immediately from (5.14) and (5.15). To prove claim ii), it is noted that solution (5.13) yields, for any $k \neq 0$

$$\begin{aligned} \mathbf{p}(w^k) &= \sum_{n=0}^{N-1} p(n)^2 w^{-kn} \\ &= \{N - (N - 1)\delta\} w^{-km} + \delta \sum_{n \neq m} w^{-kn} \\ &= \{N(1 - \delta)\} w^{-km} + \delta \sum_{n=0}^{N-1} w^{-kn} \\ &= \{N(1 - \delta)\} w^{-km} \end{aligned} \quad (5.16)$$

where the last equality follows since $\sum_{n=0}^{N-1} w^{-kn} = 0$ for any $k \neq 0$. Equations (5.10) and (5.16) show that, with solution (5.13), we have $\max |\mathbf{p}(w^k)| = |\mathbf{p}(w^0)| = N$ and $\min |\mathbf{p}(w^k)| = N(1 - \delta)$, and hence

$$k(\mathbf{Q}) = \frac{\max |\mathbf{p}(w^k)|}{\min |\mathbf{p}(w^k)|} = \frac{N}{N(1 - \delta)} = \frac{1}{1 - \delta}. \quad (5.17)$$

The proof is thus completed. $\square$

Theorem 5.2 shows that $k_{\min}(\mathbf{Q})$ depends entirely on the minimal power threshold δ , irrespective of the dimension of $\mathbf{Q}$ (and hence the symbol block length N). A small δ , in particular, is seen to yield small $k_{\min}(\mathbf{Q})$ and thus improves the channel estimation accuracy.

B. Minimization of Mean-Square Error

In this subsection, we alternatively formulate $\mathbf{w}$ as a zero-mean white noise vector with covariance matrix $\sigma_w^2 \mathbf{I}$, and resort to the well-known minimum mean-square error principle (see,

e.g., [6]), for constructing a solution. Although a theoretical justification of such statistical data error assumption is difficult to establish, our simulation study does confirm this tendency.

Since $\mathbf{w}$ is white, the mean-square errors incurred by solutions (5.3) and (5.4) are, respectively

$$E \left\| \text{vec}(\hat{\mathbf{G}}) - \text{vec}(\tilde{\mathbf{G}}) \right\|^2 = \sigma_w^2 \text{Tr} \left[(\mathbf{Q}^H \mathbf{Q})^{-1} \right] \quad (5.18)$$

and

$$E \left\| \text{vec}(\hat{\mathbf{H}}) - \text{vec}(\mathbf{H}) \right\|^2 = \sigma_w^2 \text{Tr} \left[(\tilde{\mathbf{Q}}^H \tilde{\mathbf{Q}})^{-1} \right]. \quad (5.19)$$

Toward an utmost reduction of the white noise effect, the precoder $p(n)$ should thus be chosen to jointly minimize

$$\text{Tr} \left[(\mathbf{Q}^H \mathbf{Q})^{-1} \right] \text{ and } \text{Tr} \left[(\tilde{\mathbf{Q}}^H \tilde{\mathbf{Q}})^{-1} \right] \quad (5.20)$$

subject to the constraints (5.6a) and (5.6b). Minimization of this type of cost function has been considered in least-squares-based channel estimation (e.g., [6] and [22, Ch. 9], among others). The reported solution approach therein is via the following inequality: Since both $\mathbf{Q}^H \mathbf{Q}$ and $\tilde{\mathbf{Q}}^H \tilde{\mathbf{Q}}$ are positive definite, it follows that

$$\begin{aligned} \text{Tr} \left[(\mathbf{Q}^H \mathbf{Q})^{-1} \right] &\geq \sum_i [\mathbf{Q}^H \mathbf{Q}]_{i,i}^{-1} \text{ and } \text{Tr} \left[(\tilde{\mathbf{Q}}^H \tilde{\mathbf{Q}})^{-1} \right] \\ &\geq \sum_i [\tilde{\mathbf{Q}}^H \tilde{\mathbf{Q}}]_{i,i}^{-1} \end{aligned} \quad (5.21)$$

and equalities in (5.21) hold whenever $\mathbf{Q}^H \mathbf{Q}$ and $\tilde{\mathbf{Q}}^H \tilde{\mathbf{Q}}$, respectively, are diagonal [28, p. 1041]. If the power normalization condition (5.6a) is the only design concern, it is easy to check that the impulse sequence

$$p(m)^2 = N, \text{ and } p(n)^2 = 0 \text{ for } n \neq m \quad (5.22)$$

where $0 \leq m \leq N - 1$ is fixed but arbitrary, simultaneously diagonalizes $\mathbf{Q}^H \mathbf{Q}$ and $\tilde{\mathbf{Q}}^H \tilde{\mathbf{Q}}$ and is thus the joint minimizer. However, given the additional threshold power requirement (5.6b), one cannot rely on this principle for finding a solution since, subject to the BCCB structure of $\mathbf{Q}$ and $p(n)^2 > 0$, it is impossible to choose $p(n)$ to render both $\mathbf{Q}^H \mathbf{Q}$ and $\tilde{\mathbf{Q}}^H \tilde{\mathbf{Q}}$ diagonal. In what follows we propose an alternative strategy to address the considered optimization problem. Our approach is grounded on a key fact shown in the next lemma, which directly establishes an inequality relation analogue to (5.5) regarding the two cost functions in (5.20) (the proof is given in Appendix D).

Lemma 5.3: Let $\mathbf{M}$ be a square nonsingular matrix and $\bar{\mathbf{M}}$ be constructed from $\mathbf{M}$ by deleting an arbitrary subset of its columns. Then

$$\text{Tr} \left[(\bar{\mathbf{M}}^H \bar{\mathbf{M}})^{-1} \right] \leq \text{Tr} \left[(\mathbf{M}^H \mathbf{M})^{-1} \right]. \quad (5.23)$$

$\square$

Lemma 5.3 asserts $\text{Tr} \left[(\tilde{\mathbf{Q}}^H \tilde{\mathbf{Q}})^{-1} \right]$ is upper bounded by $\text{Tr} \left[(\mathbf{Q}^H \mathbf{Q})^{-1} \right]$. This thus suggests a suboptimal, but would be a more simple and efficient, way of precoder design: we

can simply choose $p(n)$ to minimize $\text{Tr}[(\mathbf{Q}^H \mathbf{Q})^{-1}]$, since $\text{Tr}[(\tilde{\mathbf{Q}}^H \tilde{\mathbf{Q}})^{-1}]$ would in turn be kept small. The main advantage of the proposed design formulation, as expected, is that we can directly take profit of the BCCB property of $\mathbf{Q}$ to derive a closed-form solution. Indeed, since $\text{Tr}[(\mathbf{Q}^H \mathbf{Q})^{-1}]$ is the sum of the eigenvalues associated with $(\mathbf{Q}^H \mathbf{Q})^{-1}$ which, according to Theorem 4.1, are exactly the N replicas of the N -tuple $\{|\mathbf{p}(w^n)|^{-2}\}_{0 \leq n \leq N-1}$, we have

$$\text{Tr}[(\mathbf{Q}^H \mathbf{Q})^{-1}] = \sum_{k=0}^{N-1} \frac{N}{|\mathbf{p}(w^k)|^2}. \quad (5.24)$$

Equation (5.24) rewrites $\text{Tr}[(\mathbf{Q}^H \mathbf{Q})^{-1}]$ in terms of $\mathbf{p}(w^n)$'s; we can then further exploit (5.10) and Lemma 5.1 to construct an optimal solution, as is shown in the next theorem.

Theorem 5.4: The optimal $p(n)$ minimizing $\text{Tr}[(\mathbf{Q}^H \mathbf{Q})^{-1}]$, subject to constraints (5.6a) and (5.6b), is the two-level solution (5.13). The resultant minimal mean-square error is

$$MSE_{\min} = \frac{\sigma_w^2}{N} + \frac{\sigma_w^2(N-1)}{N(1-\delta)^2}. \quad (5.25)$$

Proof: From (5.10), we have

$$\begin{aligned} \text{Tr}[(\mathbf{Q}^H \mathbf{Q})^{-1}] &= \sum_{k=0}^{N-1} \frac{N}{|\mathbf{p}(w^k)|^2} \\ &= \frac{N}{|\mathbf{p}(1)|^2} + \sum_{k=1}^{N-1} \frac{N}{|\mathbf{p}(w^k)|^2} \\ &= \frac{1}{N} + \sum_{k=1}^{N-1} \frac{N}{|\mathbf{p}(w^k)|^2}. \end{aligned} \quad (5.26)$$

From Lemma 5.1, it follows that

$$\frac{1}{|\mathbf{p}(w^k)|^2} \geq \frac{1}{N^2(1-d)^2}, \quad 1 \leq k \leq N-1 \quad (5.27)$$

with equality attained by the two-level sequence (5.13) (this is easily seen from (5.16)). From (5.26) and (5.27), the minimal $\text{Tr}[(\mathbf{Q}^H \mathbf{Q})^{-1}]$ is thus

$$\begin{aligned} \text{Tr}[(\mathbf{Q}^H \mathbf{Q})^{-1}]_{\min} &= \frac{1}{N} + \sum_{k=1}^{N-1} \frac{1}{N(1-d)^2} \\ &= \frac{1}{N} + \frac{N-1}{N(1-d)^2}. \end{aligned} \quad (5.28)$$

Equation (5.25) follows directly from (5.28), and this thus proves the theorem. $\square$

Recall that the impulse sequence (5.22) is optimal with regard to the power normalization constraint (5.6a). When an additional power threshold is imposed, it turns out that the best choice is the “impulse-like” two-level solution (5.13). With (5.25), the resultant $MSE_{\min}$ is seen to decrease whenever δ is decreased. Hence, a small δ not only limits solution sensitivity to deterministic error perturbation (as we have shown in the previous subsection), but also improves the estimation accuracy

against white data estimation error. From the equalization point of view, it is, however, undesirable to keep δ unlimitedly small; this will be further discussed in the next section.

Remarks:

- From Theorems 5.2 and 5.4, it is somewhat surprising to see that, although the objective functions $k(\mathbf{Q})$ and $\text{Tr}[(\mathbf{Q}^H \mathbf{Q})^{-1}]$ are quite different in nature, the respective minimizing solutions, under constraints (5.6a) and (5.6b), are the same the two-level form choice (5.13); this is due to the BCCB property of the matrix $\mathbf{Q}$.
- The two-level solution (5.13) minimizes both $k(\mathbf{Q})$ and $\text{Tr}[(\mathbf{Q}^H \mathbf{Q})^{-1}]$ but its optimality with respect to $k(\tilde{\mathbf{Q}})$ and $\text{Tr}[(\tilde{\mathbf{Q}}^H \tilde{\mathbf{Q}})^{-1}]$ appears intractable to verify. Our simulation results seem to indicate that it is indeed the minimizing solution.
- Since $k(\tilde{\mathbf{Q}}) \leq k(\mathbf{Q})$ and

$$\text{Tr}[(\tilde{\mathbf{Q}}^H \tilde{\mathbf{Q}})^{-1}] \leq \text{Tr}[(\mathbf{Q}^H \mathbf{Q})^{-1}]$$

solution (5.4) can yield better estimation accuracy than (5.3); numerical simulations (see Simulation 2) also evidence this tendency.

- The optimal solution (5.13) does not depend on the index m at which the peak power occurs: any $0 \leq m \leq N-1$ allows for an utmost mitigation for data estimation error. However, since the trailing components in each symbol block will be duplicated as CP, the peak power in (5.13) should not be located within the corresponding index region so as to conserve the power resource.
- In the study of single channel blind identification via modulation-induced-cyclostationarity, the two-level sequence (5.13) is shown to be optimal for mitigating the channel noise effect for the serial transmission case [10], [24], and also for the FDE-based block transmission [37].

$\square$

VI. EQUALIZATION ASPECT

Toward symbol recovery in FDE-STBC systems, one commonly used approach is via frequency-domain per-tone equalization [2], [15] based on (2.8), commonly in conjunction with linear zero-forcing (ZF) or MMSE criterion. In this section, we resort to ZF-PEP analysis [35] for investigating the equalization performance regarding the optimal solution (5.13).

To proceed, based on (2.2), we shall first expand the linearly combined frequency-domain signal model (2.8) into

$$\begin{aligned} \underbrace{\mathbf{D}^H \begin{bmatrix} \mathbf{Y}_k \\ \mathbf{Y}_{k+1}^* \end{bmatrix}}_{:= \mathbf{Z}_{k,k+1}} &= \begin{bmatrix} \tilde{\mathbf{D}} & \mathbf{0}_N \\ \mathbf{0}_N & \tilde{\mathbf{D}} \end{bmatrix} \begin{bmatrix} \mathbf{X}_k \\ \mathbf{X}_{k+1} \end{bmatrix} + \mathbf{D}^H \begin{bmatrix} \mathbf{V}_k \\ \mathbf{V}_{k+1}^* \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} \tilde{\mathbf{D}} & \mathbf{0}_N \\ \mathbf{0}_N & \tilde{\mathbf{D}} \end{bmatrix} \begin{bmatrix} \mathbf{F}\mathbf{P} & \mathbf{0}_N \\ \mathbf{0}_N & \mathbf{F}\mathbf{P} \end{bmatrix}}_{:= \Phi} \underbrace{\begin{bmatrix} \mathbf{s}_k \\ \mathbf{s}_{k+1} \end{bmatrix}}_{:= \mathbf{s}_{k,k+1}} \\ &\quad + \underbrace{\mathbf{D}^H \begin{bmatrix} \mathbf{V}_k \\ \mathbf{V}_{k+1}^* \end{bmatrix}}_{:= \tilde{\mathbf{V}}_{k,k+1}} \end{aligned} \quad (6.1)$$

or by dropping the block index k and $k + 1$ for notational simplicity

$$\mathbf{Z} = \Phi \mathbf{s} + \tilde{\mathbf{V}}. \quad (6.2)$$

The PEP measures the probability that a symbol block $\mathbf{s}$ is transmitted but another $\tilde{\mathbf{s}} \neq \mathbf{s}$ is detected. Given the channel realizations $\mathbf{h}_1$ and $\mathbf{h}_2$, the conditional PEP is by definition given by

$$\Pr[\mathbf{s} \rightarrow \tilde{\mathbf{s}} | \mathbf{h}_1, \mathbf{h}_2] = \Pr[\|\tilde{\mathbf{s}} - \hat{\mathbf{s}}\| < \|\mathbf{s} - \hat{\mathbf{s}}\| | \mathbf{h}_1, \mathbf{h}_2] \quad (6.3)$$

where $\hat{\mathbf{s}}$ is the estimate of $\mathbf{s}$ under the ZF metric and, from (6.2), is given by

$$\hat{\mathbf{s}} := \Phi^{-1} \mathbf{Z} = \mathbf{s} + \Phi^{-1} \tilde{\mathbf{V}}. \quad (6.4)$$

By following the procedures as in [35] and defining $\tilde{d} := \|\mathbf{s} - \hat{\mathbf{s}}\|$, the conditional PEP in (6.3) can be upper bounded by

$$\begin{aligned} \Pr[\mathbf{s} \rightarrow \tilde{\mathbf{s}} | \mathbf{h}_1, \mathbf{h}_2] &\leq Q \left(\tilde{d} \left(\sqrt{4s_v^2 \|\mathbf{P}^{-1}\|_F^2 \|\tilde{\mathbf{D}}^{-1}\|_F^2} \right)^{-1} \right) \\ &= Q \left(\tilde{d} \left(\sqrt{4s_v^2 \left(\sum_{n=0}^{N-1} p(n)^{-2} \right) \|\tilde{\mathbf{D}}^{-1}\|_F^2} \right)^{-1} \right) \end{aligned} \quad (6.5)$$

where $Q(\cdot)$ denotes the Gaussian tail, and the equality in (6.5) follows directly from (2.1). For a given channel pair, and hence $\tilde{\mathbf{D}}$, the upper bound in (6.5) is minimized if the quantity $\sum_{n=0}^{N-1} p(n)^{-2}$ attains the minimum. Since, by the Cauchy's inequality

$$\begin{aligned} (p(0)^{-2} + \dots + p(N-1)^{-2}) (p(0)^2 + \dots + p(N-1)^2) \\ \geq (1 + \dots + 1)^2 = N^2 \end{aligned} \quad (6.6)$$

and $\sum_{n=0}^{N-1} p(n)^2 = N$ (cf. (5.6a)), we have $(p(0)^{-2} + \dots + p(N-1)^{-2}) \geq N$, with equality holds if and only if $p(n) = 1$ for $0 \leq n \leq N-1$. This shows the equal power scheme is optimal from an equalization point of view. Any form of precoding induced power variation, therefore, will incur a loss in the decision performance. The precoder (5.13), however, turns out to be the worst-case choice, as can be seen from the following theorem (see Appendix E for a proof).

Theorem 6.1: For all $p(n)$ satisfying (5.6a) and (5.6b), the solution (5.13) maximizes the quantity $\sum_{n=0}^{N-1} p(n)^{-2}$, leading to

$$\max \sum_{n=0}^{N-1} p(n)^{-2} = \frac{N-1}{\delta} + \frac{1}{N - (N-1)\delta}. \quad (6.7)$$

□

Based on (6.7), simple manipulation shows the maximum value, when viewed as a function of δ , will increase as δ is decreased. As a result, a small δ , although improving channel estimation accuracy, will enlarge the PEP upper bound in (6.5),

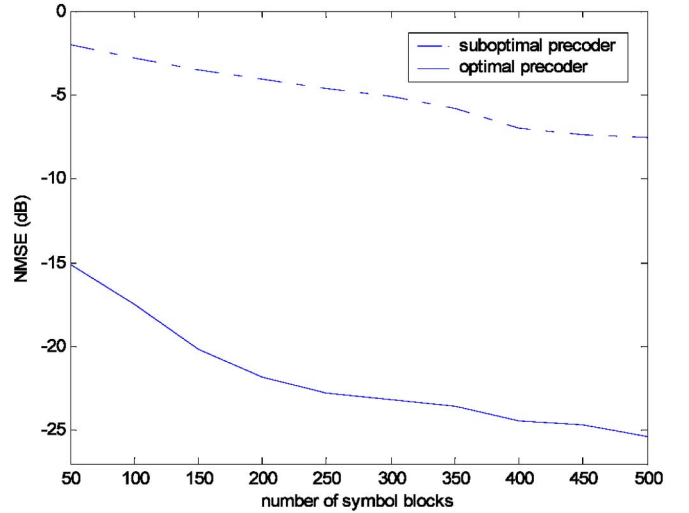


Fig. 2. Channel NMSE (optimal and suboptimal precoders).

and hence bring potentially poor equalization performance. This thus imposes a tradeoff in selecting δ ; our simulation study (see Simulation 5) indicates that $\delta = 0.7 \sim 0.8$ are the compromising choices.

VII. SIMULATION RESULTS

This section uses several numerical simulations to illustrate the performance of the proposed method. The symbol block length and the channel order, respectively, are set to be $N = 32$ and $L = 8$; the inserted CP spans eight symbol periods and the source constellation is quadrature phase-shift keying (QPSK). Unless otherwise stated, we will consider a block fading environment in which the channel taps, modeled as i.i.d. zero-mean unit-variance complex Gaussian random variables, remain constant over a burst of K symbol blocks and can vary independently between different bursts. The identification performance is measured by the normalized mean-square error (NMSE), namely $\text{NMSE} := (1/2I) \sum_{l=1}^2 \sum_{i=1}^I \|\hat{\mathbf{h}}_l^{(i)} - \mathbf{h}_l^{(i)}\|^2 \cdot \|\mathbf{h}_l^{(i)}\|^{-2}$, where $\mathbf{h}_l^{(i)}$ is the realization of the l th channel in the i th data packet, $\hat{\mathbf{h}}_l^{(i)}$ is the corresponding estimate, and I is the total number of trials. Throughout the simulations, the peak power index of the optimal precoder (5.13) is chosen to be $m = 0$; the signal-to-noise ratio (SNR) is defined as $\text{SNR} := (E \|\mathbf{h}_1\|^2 + E \|\mathbf{h}_2\|^2) / 2\sigma_v^2$. Simulations 1–4 investigate the intrinsic aspects pertaining to the proposed method, and we simply use the least-squares fit technique for matrix ambiguity removal, as is done in [4], [13], and [21]; in Simulations 1–4, we set $I = 200$.

1) Simulation 1—Effectiveness of the Optimal Precoder (5.13): This simulation illustrates the effectiveness of the proposed optimal precoder (5.13). For $\text{SNR} = 10$ dB and $\delta = 0.6$, we consider the optimal sequence (5.13) and another suboptimal choice given as $p(n)^2 = 0.6$ for $0 \leq n \leq 15$ and $p(n)^2 = 1.4$ for $16 \leq n \leq 31$. Fig. 2 shows the computed NMSE with various numbers of symbol blocks K (the product channel coefficients are computed via (5.4)). It can be seen

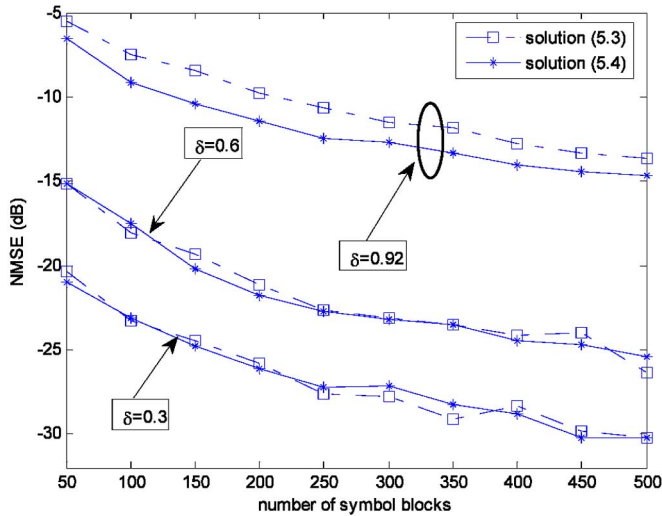


Fig. 3. Channels NMSE by two solutions.

that the optimal solution (5.13) significantly improves the performance.

2) *Simulation 2—Performance Comparison of Solutions (5.3) and (5.4)*: This simulation compares the estimation performance of solutions (5.3) and (5.4). Fig. 3 shows the respective NMSE versus number of symbol blocks for three power thresholds δ : 0.3, 0.6, and 0.92 (SNR is fixed at 10 dB). The result shows that the performances of the two methods are very close for $\delta = 0.3$ and 0.6; however, solution (5.4) seems to yield smaller NMSE when $\delta = 0.92$. This is because, for $\delta = 0.3$ and 0.6, the associated condition number pair $(k(\mathbf{Q}), k(\tilde{\mathbf{Q}}))$ are (1.4286, 1.1370) and (2.5, 1.5737): both the two matrices $\mathbf{Q}$ and $\tilde{\mathbf{Q}}$ remain well conditioned and can largely limit the error effect. However, for $\delta = 0.92$, we have $(k(\mathbf{Q}), k(\tilde{\mathbf{Q}})) = (12.5, 6.683)$: the matrix $\mathbf{Q}$ tends to be ill conditioned, and solution (5.3) becomes more susceptible to data errors (solution (5.4) will be adopted in subsequent simulations).

3) *Simulation 3—Robustness Against Channel Order Overestimation*: This simulation tests the proposed method when channel order is overestimated. We consider two different levels of SNR: 0 and 15 dB. For the overestimated channel order $8 \leq \hat{L} \leq 15$, Fig. 4 shows the respective computed NMSE ($K = 500$ and $\delta = 0.8$). It can be seen that the proposed method is quite robust with respect to channel order overestimation: the NMSE increment is only about 3 dB as $\hat{L}$ increases from 8 to 15.

4) *Simulation 4—Estimation Performance Against Blind Subspace Method With Transmit Redundancy [40]*: This simulation compares channel estimation performances of the proposed scheme with the identical-precoder subspace method [40, p. 1218], in which the FFT precoding matrix is adopted to convert the multicarrier scheme into single-carrier FDE-STBC systems considered in this paper. To implement the algorithm in [40], the last eight entries in each symbol block are set to be zero; this introduces the minimal amount of transmit redundancy for fulfilling the associated channel identifiability condition (cf. [40, p. 1218]). For fixed SNR = 10 dB, Fig. 5

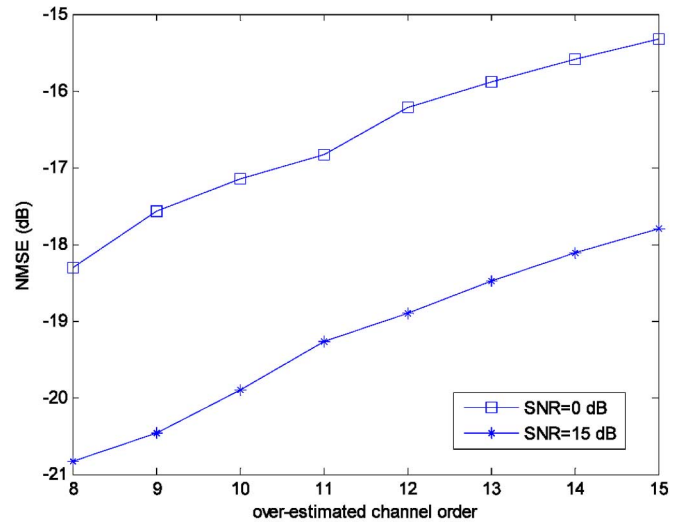


Fig. 4. NMSE in the presence of channel order overestimation.

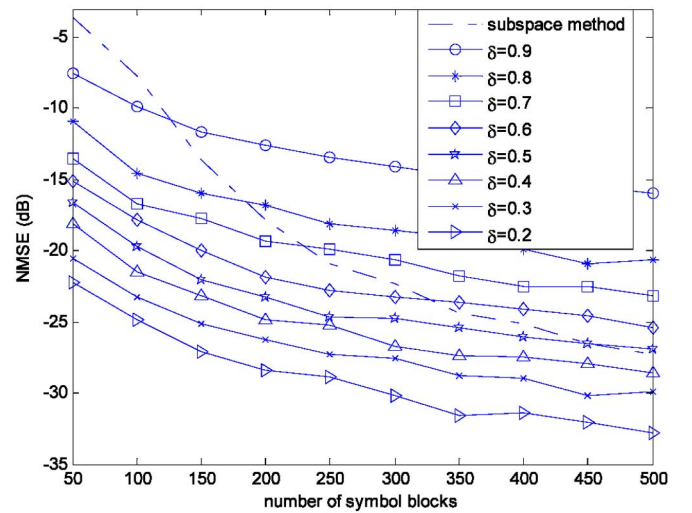


Fig. 5. NMSE of two methods at different numbers of symbol blocks.

shows the computed NMSE versus number of symbol blocks; the proposed method, depicted with solid lines, is implemented with various choices of δ . We can first see from the figure that the performance of the proposed method is improved as δ decreases: this is because small δ results in small $k(\mathbf{Q})$, and also reduces the mean square error incurred by white noise perturbation. Compared with the subspace method [40], the proposed approach can better track the channel with a small number of received data blocks. Fig. 6 shows the NMSE of the two methods at different SNR levels ($K = 500$). The result shows that, in the medium-to-low SNR region, our method performs better even with the moderate choice $\delta = 0.7$. When SNR increases, the output NMSE of [40] exhibits a fast decay. This is not unexpected since the method [40] is “deterministic” in nature: it benefits from the finite-sample-convergence property and can usually yield impressive estimation accuracy when SNR is high [31]. A similar tendency is also observed in [31, p. 1942] when nonredundant diagonal precoding based identification is compared with the (deterministic) multichannel subspace methods [29] and [38].

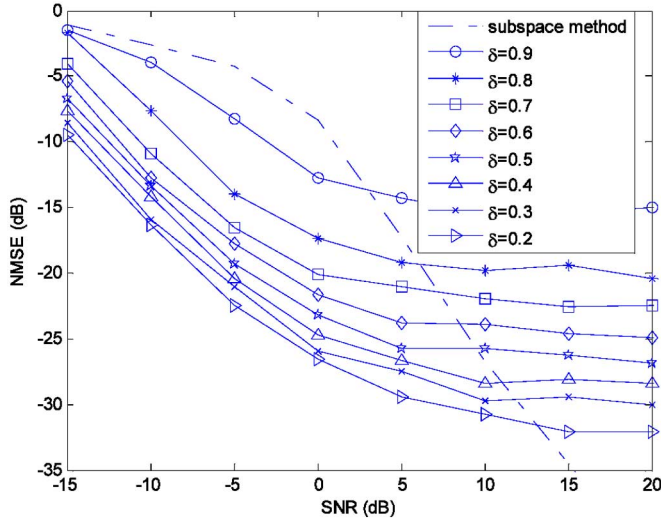
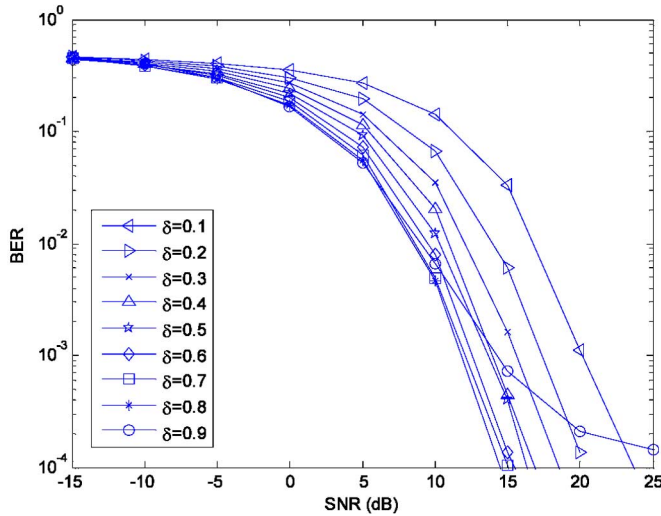


Fig. 6. NMSE of two methods at different SNR levels.

Fig. 7. BER performance of the proposed method at different minimal power threshold δ .

5) *Simulation 5—On Selection of Power Threshold δ* : This simulation considers the optimal precoder (5.13) and illustrates the impact of δ on equalization performance. Fig. 7 shows the bit-error-rate (BER) curves for $0.1 \leq \delta \leq 0.9$; we set $K = 500$, $I = 1000$, and use frequency-domain ZF equalizer [15] for symbol recovery. It can be seen that, although a large δ results in a less accurate channel estimate, the BER is however improved as δ increases from 0.1 to 0.8. This would reflect the ZF-PEP analysis given in Section VI: A large δ can, on the other hand, limit the power penalty and improve the equalization performance. However, if δ is too large ($\delta > 0.8$), the channels will be poorly estimated, resulting in large decision error floor. Hence, $\delta = 0.7 \sim 0.8$ seem to be the compromising choices, as far as equalization performance is concerned.

6) *Simulation 6—Equalization Performance Comparison*: In this simulation, we compare the proposed method (implemented using the optimal precoder (5.13) and $\delta = 0.8$) with the blind identical-precoder subspace algorithm in [40] and the training-based scheme [11] in terms of BER. To implement the method

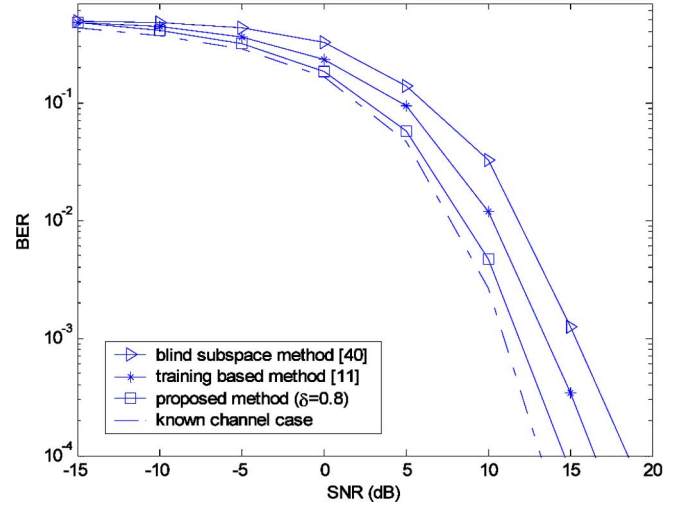


Fig. 8. BER performances of three methods (i.i.d. Gaussian channel model).

[11], pilot symbols (64 in total) are placed in the initial two blocks in each data burst and are optimally designed according to [11, p. 730]. To fairly compare the three methods under a fixed spectral efficiency, we will similarly use 64 training symbols, placed also in the initial block pair per data burst, for ambiguity removal in the two blind approaches; for simplicity, we just choose the optimal sequence reported in [11]. We note that, in the transmit-redundancy based blind scheme [40], eight entries in each symbol block are zero-padded for facilitating channel identification: Only 24 entries per symbol block can thus be used for carrying source data. To compensate for possible spectral efficiency loss, 20 elements out of which are loaded with 8-PSK symbols, whereas the remaining four are BPSK modulated: this maintains an overall data rate of 64 bits/block, as in the training scheme [11] and the proposed method (both with QPSK constellation). Fig. 8 shows the BER results of the three methods at different SNR levels ($K = 500$ and $I = 1000$); the known channel case is also included as the benchmark performance. The proposed method, as we can see, leads to the lowest BER. The performance advantage over the training method [11] would come from the reduction of number of unknowns from $2(L + 1)$ to three: This would further smooth the noise effect and thus improve performance. In contrast with [40], our method yields about a 3–4-dB SNR gain; this benefits from the nonredundant transmission of the proposed scheme: For a target data rate, one can otherwise use lower order constellations to buy more BER floor margin. Finally, we observe that, compared with the known channel case, our method seems to incur no more than 1-dB penalty. We repeat the above experiment with the exponential delay power profile channel model [30]. Fig. 9 shows the resultant BER curves, which are seen to exhibit a similar tendency as in the i.i.d. Gaussian channel case.

7) *Simulation 7—Equalization Performance in Slowly Time-Varying Channels*: This simulation compares the proposed method (with optimal precoder (5.13) and $\delta = 0.8$) against [11] and [40] in a slowly time-varying channel environment. We assume each channel tap varies according to the Jake's model with a maximal Doppler frequency of 52 Hz; this corresponds to a moving speed 3 m/s and a carrier frequency

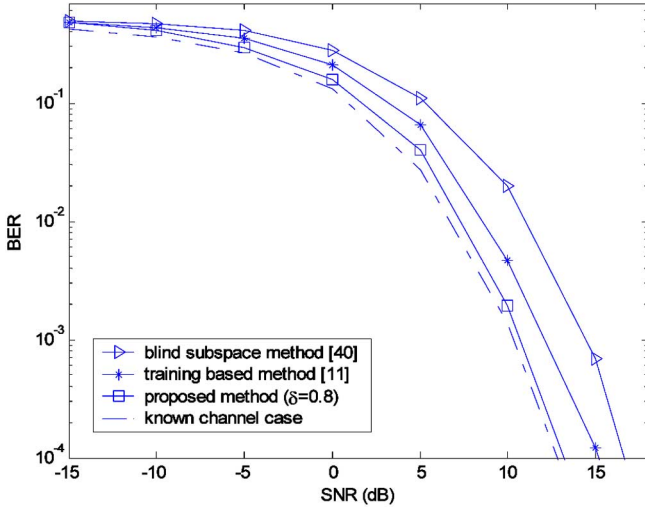


Fig. 9. BER performances of three methods (exponential delay-power profile channel model).

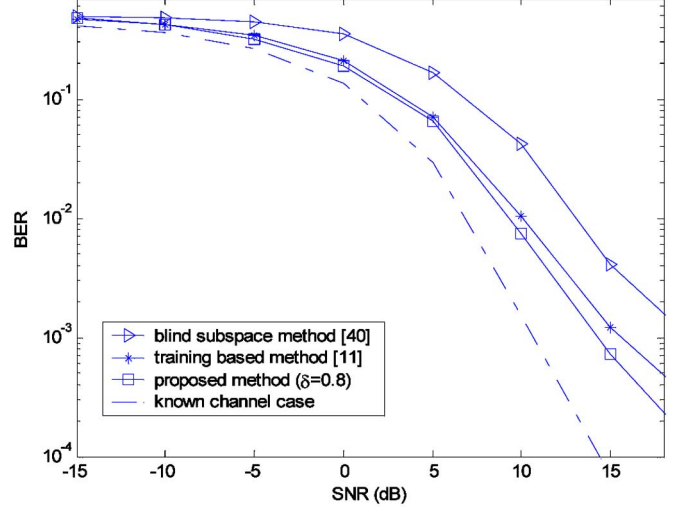


Fig. 10. BER performances of three methods (slowly time-varying channel with Jake's model).

of 5.2 GHz (the same simulation environment is considered in [40]). The number of symbol blocks in each data burst is set to be $K = 300$. To track the channel variation, in both blind methods, the receive data statistics are adaptively updated based on the rectangular sliding windowing scheme suggested in [40, p. 1218], with the window size set to be 150. Channel estimation is performed each time a new subburst of 50 symbol blocks are available. As in the previous simulation, the optimal pilot sequence designed in [11] is placed in the initial symbol block pair per subburst, for training-based estimation as well as for ambiguity removal in the two blind schemes; for the transmit-redundancy-based solution [40], symbol constellations are likewise loaded for comparison under a fixed data rate. Fig. 10 shows the respective BER curves (averaged over $I = 2000$ trials). Compared with the quasi-static case (Figs. 8 and 9), the performances of the three methods degrade due to the time-varying channel characteristic; the proposed method, still, leads to the lowest error rate.

8) *Simulation 8—On PAPR Performance:* This simulation investigates the PAPR results when the optimal precoder (5.13) is used. For the considered system parameters ($N = 32$, $L = 8$) and with Nyquist pulse shaping filter, the values of PAPR for various choices of δ with respect to different symbol constellations are tabulated in [37, p. 1124]. The results show that the two-level precoder (5.13) does increase PAPR over the equal-power case; in particular, the choice $\delta = 0.3$ raises the PAPR to a level comparable to that of multicarrier scenario. It is noted that the PAPR does not faithfully reflect the actual signal amplitude variation in all cases; this is because the probability that such a peak occurs could depend on the block length as well [36]. A more realistic performance metric for describing the actual power amplification portrait is the *instantaneous PAPR* [36, p. 383]. For 64-QAM and 256-QAM constellations, Fig. 11 shows the probability that the instantaneous PAPR exceeds a prescribed value γ with respect to five different choices of power thresholds: $\delta = 0.7, 0.76, 0.84, 0.93$ and 0.97 (the precoding induced power spike, i.e., $N - (N - 1)\delta$, are 10.3, 8.44, 6, 3, and

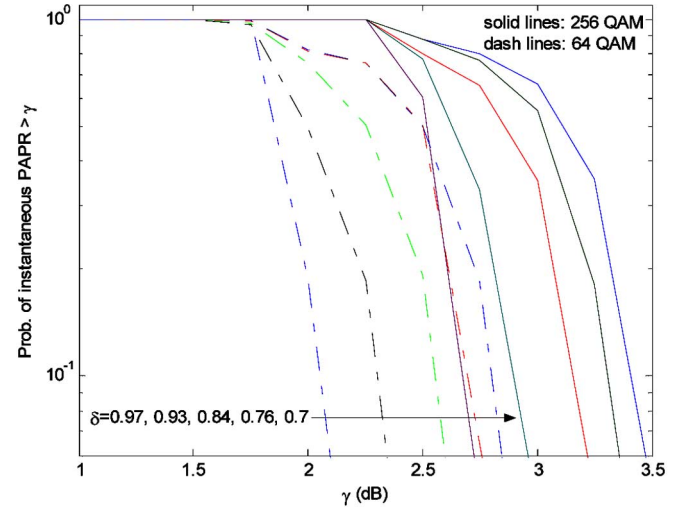


Fig. 11. Tendency of instantaneous PAPR for two QAM constellations.

1.8, respectively). As we can see from the figure, it is likely that no more than 1-dB power back-off is required as δ decreases from 0.97 to 0.7 (or, equivalently, the spike value increases from 1.8 to 10.3); a slight impact on the incurred instantaneous PAPR is also observed when different symbol constellations are used. As a result, with moderate choices of δ , the proposed optimal precoder (5.13) does not seem to induce a large power back-off in practice. It is noted that, although the instantaneous PAPR will increase when δ falls below 0.7, small δ should be precluded for maintaining the BER performance (see Fig. 7).

VIII. CONCLUSION

Blind channel estimation for STBC transmission in a MISO environment is a challenging research problem. This paper presents a solution for FDE-STBC systems. The proposed method relies on nonredundant diagonal precoding and i.i.d.

source assumption and exploits the resultant linear signal structure in the conjugate cross-correlation matrix of the received data. The circulant channel matrix property, which is unique to FDE-based block transmission, leads to an identification equation set with a BCCB nature. Such a distinctive system matrix structure simplifies the characterization of channel identifiability condition (in terms of precoder coefficients) and can also alleviate the underlying algorithm complexity. The proposed channel estimate has an appealing property: It is exact when perfect data statistic is available and channel noise is circularly Gaussian. In the presence of finite-sample estimation errors, our channel estimation framework easily incorporates the data mismatch effect, and allows for natural precoder design formulation and criteria for improving estimation accuracy. Through analysis, the optimization problems can be formulated to exploit the BCCB matrix property and are then analytically solved. The proposed solution tends to optimize the channel estimation robustness against deterministic error perturbation and also minimize the mean-square error when data mismatch is modeled as a white noise. The PEP analysis shows a tradeoff regarding the proposed optimal error-resistant precoder: It incurs the worst-case power penalty for symbol decision. Through numerical simulations compromising choices for precoder parameters are determined. Simulation results show that the proposed approach compares favorably with existing training and blind methods fitted for FDE-STBC systems.

APPENDIX A

PROOF OF THEOREM 4.1

We will denote by $\text{BCCB}_{N,N}$ the set of all $N^2 \times N^2$ block circulant matrices with circulant blocks [12, p. 184], each characterized by N circulant matrices of dimension $N \times N$. The proof of Theorem 4.1 is based on the following lemma.

Lemma A.1 [12, p. 185]: If $\mathbf{X} \in \text{BCCB}_{N,N}$, then $\mathbf{X}$ can be diagonalized by $\mathbf{F} \otimes \mathbf{F}$. More precisely, let $\{\mathbf{C}_0, \dots, \mathbf{C}_{N-1}\}$ be the set of $N \times N$ circulant matrices on the top row block of $\mathbf{X}$, and let Λ_n be the diagonal matrix containing the eigenvalues of $\mathbf{C}_n$. Then we have

$$\mathbf{X} = (\mathbf{F} \otimes \mathbf{F}) \left(\sum_{n=0}^{N-1} \Omega_N^n \otimes \Lambda_n \right) (\mathbf{F}^{-1} \otimes \mathbf{F}^{-1}), \quad (\text{A.1})$$

where $\Omega_N := \text{diag} \left\{ [1 w w^2 \dots w^{N-1}]^T \right\}$. Conversely, any matrix of the form $(\mathbf{F} \otimes \mathbf{F}) \bar{\Lambda} (\mathbf{F}^{-1} \otimes \mathbf{F}^{-1})$ for some diagonal $\bar{\Lambda}$ belongs to $\text{BCCB}_{N,N}$. $\square$

Sketch Proof of Theorem 4.1: It can be seen from (3.14) that the matrix $\mathbf{Q}$ is characterized by the N circulant matrices $\{p(0)^2 \mathbf{I}_N, p(1)^2 \mathbf{J}, \dots, p(N-1)^2 \mathbf{J}^{N-1}\}$ on its top row block.

If we stack the N eigenvalues of the circulant matrix $p(n)^2 \mathbf{J}^n$ into a vector, say $\mathbf{u}_n$, by definition of $\mathbf{J}$ (cf. (3.7)) it can be verified that

$$\begin{aligned} \mathbf{u}_0 &= \sqrt{N} \cdot p(0)^2 \mathbf{f}_0, \text{ and} \\ \mathbf{u}_n &= \sqrt{N} \cdot p(n)^2 \mathbf{f}_{N-n}, \text{ for } 1 \leq n \leq N-1 \end{aligned} \quad (\text{A.2})$$

where

$$\mathbf{f}_n := \frac{1}{\sqrt{N}} \cdot [1 \ w^n \ w^{2n} \ \dots \ w^{(N-2)n} \ w^{(N-1)n}]^T \in \mathbb{C}^N$$

is the $(n+1)$ th column of $\mathbf{F}^{-1}$, $0 \leq n \leq N-1$. Based on (A.2), Lemma A.1 and by going through essentially the same steps as in [37, App. A], it can be shown that the eigenvalues of $\mathbf{Q}$ are the N^2 entries of the vector (A.3), shown at the bottom of the page, and the assertion thus follows. $\square$

APPENDIX B

ON COMPUTATION OF PRODUCT UNKNOWNNS

In the following theorem, we will show that $\mathbf{Q}^{-1}$ also belongs to the BCCB category. Moreover, it possesses an identical structure as that of $\mathbf{Q}$ and is completely specified by N scalar parameters which are very easy to compute.

Theorem B.1: Let $\mathbf{p}(z)$ be the polynomial defined as in (4.2). Assume $p(n)$ is chosen so that $\mathbf{Q}$ is nonsingular. Then $\mathbf{Q}^{-1}$ is also a BCCB matrix with $[\alpha_0 \mathbf{I}_N, \alpha_1 \mathbf{J}, \dots, \alpha_{N-1} \mathbf{J}^{N-1}] \in \mathbb{R}^{N \times N^2}$ as the top row block, where

$$\begin{aligned} [\alpha_0 \ \dots \ \alpha_{N-1}]^T &= (\sqrt{N})^{-1} \mathbf{F}^{-1} \\ &\quad \times [\mathbf{p}(1)^{-1} \ \mathbf{p}(w)^{-1} \ \dots \ \mathbf{p}(w^{N-1})^{-1}]^T. \end{aligned} \quad (\text{B.1})$$

$\square$

Proof: From Lemma A.1, it is easy to see that $\mathbf{Q} \in \text{BCCB}_{N,N}$ implies $\mathbf{Q}^{-1} \in \text{BCCB}_{N,N}$, and hence we can write $\mathbf{Q}^{-1} = (\mathbf{F} \otimes \mathbf{F}) \bar{\Lambda} (\mathbf{F}^{-1} \otimes \mathbf{F}^{-1})$ for some diagonal $\bar{\Lambda}$. It suffices to check that, for the BCCB matrix with $\{\alpha_0 \mathbf{I}_N, \alpha_1 \mathbf{J}, \dots, \alpha_{N-1} \mathbf{J}^{N-1}\}$ on the top row block, the resultant $\bar{\Lambda}$ satisfies $\bar{\Lambda} \Lambda_{\mathbf{Q}} = \mathbf{I}_{N^2}$, where $\Lambda_{\mathbf{Q}} = \text{diag} \{\mathbf{q}_{\mathbf{Q}}\}$ and $\mathbf{q}_{\mathbf{Q}}$ is in (A.3). By following the procedures as in Appendix A, it can be shown $\bar{\Lambda} = \text{diag} \{\mathbf{q}\}$, where the vector $\mathbf{q}$ is defined in (B.2) shown at the bottom of the next page.

By definition of α in (B.1), it is easy to see that the l th entry of $\mathbf{q}$ is simply the reciprocal of the l th entry of $\mathbf{q}_{\mathbf{Q}}$ in (A.3). The proof is thus completed. $\square$

Theorem B.1 asserts that, to invert the $N^2 \times N^2$ matrix $\mathbf{Q}$, it merely requires to compute N scalars α_i 's via (B.1). This calls for two N -point FFT operations, one for computing

$$\mathbf{q}_{\mathbf{Q}} := \left[\left(\sqrt{N} \mathbf{F} \mathbf{p} \right)^T \ \left((\mathbf{J}^T) \sqrt{N} \mathbf{F} \mathbf{p} \right)^T \ \dots \ \left((\mathbf{J}^T)^{N-1} \sqrt{N} \mathbf{F} \mathbf{p} \right)^T \right]^T, \quad (\text{A.3})$$

$\mathbf{p}(w^n)$'s from $p(n)$ and the other for α_n 's based on $\mathbf{p}(w^n)^{-1}$'s; additional N multiplications are also needed for evaluating $\mathbf{p}(w^n)^{-1}$'s from $\mathbf{p}(w^n)$'s. When the symbol block length N is chosen to be a power of two, the number of flop counts is $4N \log_2 N - N$. The main computational cost for (4.8), on the other hand, lies in inverting an $(L+1)^2 \times (L+1)^2$ Hermitian Toeplitz matrix $\tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}}$; the complexity can be limited to $4(L+1)^4$ by using the Levinson algorithm [17, p. 196]. $\square$

APPENDIX C

PROOF OF LEMMA 5.1

The assertion relies on the following key observation: *any given $p(n)$ satisfying (5.6) can be constructed by “squeezing” the peak power of the two-level solution (5.13) so that the ground powers at other time instants are “raised” to the prescribed levels.* More precisely, let $p(n)$ be an admissible sequence such that $\delta < p(n)^2 < N - (N-1)\delta$ for $n \in \mathcal{I}$, where the index set $\mathcal{I}$ is a subset of $\{0, \dots, N-1\} \setminus \{m\}$. Then, $p(n)$ can be expressed as

$$p(m)^2 = N - (N-1)\delta - \sum_{n \in \mathcal{I}} \Delta_n \quad (\text{C.1})$$

and

$$p(n)^2 = \delta + \Delta_n \text{ for } n \in \mathcal{I}, \text{ and } p(n)^2 = \delta \text{ for } n \notin \mathcal{I} \quad (\text{C.2})$$

where $\Delta_n > 0$ models the excessive power over the ground level δ for $n \in \mathcal{I}$. The sequence of the form (C.1) and (C.2) satisfies the constraints (5.6a) and (5.6b); in particular, since $p(m)^2 \geq \delta$ is required, we can infer from (C.1) that

$$\sum_{n \in \mathcal{I}} \Delta_n \leq N(1 - \delta). \quad (\text{C.3})$$

We assume for the moment that $m = 0$; as one will see, the result for the $1 \leq m \leq N-1$ case easily follows. Associated with $p(n)$ in (C.1) and (C.2), we have, for $1 \leq k \leq N-1$

$$\begin{aligned} \mathbf{p}(w^k) &= \sum_{n=0}^{N-1} p(n)^2 w^{-kn} \\ &= \left[N - (N-1)\delta - \sum_{n \in \mathcal{I}} \Delta_n \right] \end{aligned}$$

$$\begin{aligned} &+ \sum_{n \in \mathcal{I}} (\delta + \Delta_n) w^{-kn} + \sum_{n \notin \mathcal{I}} \delta w^{-kn} \\ &= \left[N(1 - \delta) + \sum_{n \in \mathcal{I}} (w^{-kn} - 1) \Delta_n \right] \\ &+ \underbrace{\left[\delta + \sum_{n \in \mathcal{I}} \delta w^{-kn} + \sum_{n \notin \mathcal{I}} \delta w^{-kn} \right]}_{=\delta \sum_{n=0}^{N-1} w^{-kn} = 0} \\ &= \left[N(1 - \delta) + \sum_{n \in \mathcal{I}} (w^{-kn} - 1) \Delta_n \right]. \quad (\text{C.4}) \end{aligned}$$

Let us define the nonnegative number

$$d := N(1 - \delta) - \sum_{n \in \mathcal{I}} \Delta_n. \quad (\text{C.5})$$

Since $w^{-kn} = \cos n\theta_k - j \sin n\theta_k$, where $\theta_k := 2\pi k/N$, and with (C.5), it follows from (C.4) that

$$\begin{aligned} |\mathbf{p}(w^k)|^2 &= \left[d + \sum_{n \in \mathcal{I}} \Delta_n \cos n\theta_k \right]^2 + \left[\sum_{n \in \mathcal{I}} \Delta_n \sin n\theta_k \right]^2 \\ &= d^2 + 2d \left(\sum_{n \in \mathcal{I}} \Delta_n \cos n\theta_k \right) \\ &+ \left(\sum_{n \in \mathcal{I}} \Delta_n \cos n\theta_k \right)^2 + \left(\sum_{n \in \mathcal{I}} \Delta_n \sin n\theta_k \right)^2. \quad (\text{C.6}) \end{aligned}$$

Observe that

$$\begin{aligned} &\left(\sum_{n \in \mathcal{I}} \Delta_n \cos n\theta_k \right)^2 + \left(\sum_{n \in \mathcal{I}} \Delta_n \sin n\theta_k \right)^2 \\ &= \sum_{n \in \mathcal{I}} \Delta_n^2 + 2 \sum_{n_l, n_m \in \mathcal{I}} \Delta_{n_l} \Delta_{n_m} \\ &\quad \times (\cos n_l \theta_k \cos n_m \theta_k + \sin n_l \theta_k \sin n_m \theta_k) \\ &= \sum_{n \in \mathcal{I}} \Delta_n^2 + 2 \sum_{n_l, n_m \in \mathcal{I}} \Delta_{n_l} \Delta_{n_m} \cos(n_l - n_m) \theta_k \\ &\leq \sum_{n \in \mathcal{I}} \Delta_n^2 + 2 \sum_{n_l, n_m \in \mathcal{I}} \Delta_{n_l} \Delta_{n_m} = \left(\sum_{n \in \mathcal{I}} \Delta_n \right)^2 \quad (\text{C.7}) \end{aligned}$$

and that

$$2d \left(\sum_{n \in \mathcal{I}} \Delta_n \cos n\theta_k \right) \leq 2d \left(\sum_{n \in \mathcal{I}} \Delta_n \right). \quad (\text{C.8})$$

$$\mathbf{q} = \left[\left(\sqrt{N} \mathbf{F} \alpha \right)^T \quad \left((\mathbf{J}^T) \sqrt{N} \mathbf{F} \alpha \right)^T \quad \dots \quad \left((\mathbf{J}^T)^{N-1} \sqrt{N} \mathbf{F} \alpha \right)^T \right]^T \quad (\text{B.2})$$

From (C.7) and (C.8), $|\mathbf{p}(w^k)|^2$ in (C.6) can be upper bounded as

$$\begin{aligned} |\mathbf{p}(w^k)|^2 &\leq d^2 + 2d \left(\sum_{n \in \mathcal{I}} \Delta_n \right) + \left(\sum_{n \in \mathcal{I}} \Delta_n \right)^2 \\ &= \left(d + \sum_{n \in \mathcal{I}} \Delta_n \right)^2 = N^2(1 - \delta)^2 \end{aligned} \quad (\text{C.9})$$

in which the last equality follows from the definition of δ in (C.5). This thus proves the lemma, under the assumption $m = 0$ in (C.1). For $1 \leq m \leq N-1$, equation (C.4) is then accordingly modified as

$$\mathbf{p}(w^k) = \left[N(1 - \delta) + \sum_{n \in \mathcal{I}} (w^{-(k-m)n} - 1) \Delta_n \right] w^{-mn}. \quad (\text{C.10})$$

By going through the same procedures as in (C.5)–(C.8), the conclusion (C.9) will follow. $\square$

APPENDIX D PROOF OF LEMMA 5.3

Without loss of generality, we assume $\mathbf{M}$ is split as $\mathbf{M} = [\tilde{\mathbf{M}} \quad \mathbf{M}_d]$, in which $\mathbf{M}_d$ contains the columns to be deleted; otherwise, we can multiply $\mathbf{M}$ from the right by a permutation matrix to put it in this partition. It thus follows that

$$\mathbf{M}^H \mathbf{M} = \begin{bmatrix} \tilde{\mathbf{M}}^H \\ \mathbf{M}_d^H \end{bmatrix} [\tilde{\mathbf{M}} \quad \mathbf{M}_d] = \begin{bmatrix} \tilde{\mathbf{M}}^H \tilde{\mathbf{M}} & \tilde{\mathbf{M}}^H \mathbf{M}_d \\ \mathbf{M}_d^H \tilde{\mathbf{M}} & \mathbf{M}_d^H \mathbf{M}_d \end{bmatrix}. \quad (\text{D.1})$$

Since $\mathbf{M}$ is nonsingular, $\mathbf{M}^H \mathbf{M}$ is positive definite. By the inversion lemma for block matrix [20, p. 572], we have (D.2), shown at the bottom of the page, in which the notation “ $\times$ ” stands for the block off-diagonal submatrices irrelevant to the proof procedures. From (D.2), we have

$$\begin{aligned} &\text{Tr} \left[(\mathbf{M}^H \mathbf{M})^{-1} \right] \\ &= \text{Tr} \left[\left(\tilde{\mathbf{M}}^H \tilde{\mathbf{M}} - \tilde{\mathbf{M}}^H \mathbf{M}_d (\mathbf{M}_d^H \mathbf{M}_d)^{-1} \mathbf{M}_d^H \tilde{\mathbf{M}} \right)^{-1} \right] \\ &\quad + \text{Tr} \left[\left(\mathbf{M}_d^H \mathbf{M}_d - \mathbf{M}_d^H \tilde{\mathbf{M}} (\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}}^H \mathbf{M}_d \right)^{-1} \right]. \end{aligned} \quad (\text{D.3})$$

Since $(\mathbf{M}^H \mathbf{M})^{-1}$ is positive definite, so are its principle submatrices and (D.3) implies

$$\begin{aligned} &\text{Tr} \left[(\mathbf{M}^H \mathbf{M})^{-1} \right] \\ &\geq \text{Tr} \left[\left(\tilde{\mathbf{M}}^H \tilde{\mathbf{M}} - \tilde{\mathbf{M}}^H \mathbf{M}_d (\mathbf{M}_d^H \mathbf{M}_d)^{-1} \mathbf{M}_d^H \tilde{\mathbf{M}} \right)^{-1} \right]. \end{aligned} \quad (\text{D.4})$$

Using the matrix inversion lemma [20, p. 571], inequality (D.4) can be further expanded into

$$\begin{aligned} &\text{Tr} \left[(\mathbf{M}^H \mathbf{M})^{-1} \right] \\ &\geq \text{Tr} \left[\left(\tilde{\mathbf{M}}^H \tilde{\mathbf{M}} - \tilde{\mathbf{M}}^H \mathbf{M}_d (\mathbf{M}_d^H \mathbf{M}_d)^{-1} \mathbf{M}_d^H \tilde{\mathbf{M}} \right)^{-1} \right] \\ &= \text{Tr} \left[(\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} + (\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}}^H \mathbf{M}_d \right. \\ &\quad \times \left(\mathbf{M}_d^H \mathbf{M}_d - \mathbf{M}_d^H \tilde{\mathbf{M}} (\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}}^H \mathbf{M}_d \right)^{-1} \\ &\quad \left. \mathbf{M}_d^H \tilde{\mathbf{M}} (\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} \right] \\ &= \text{Tr} \left[(\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} \right] \\ &\quad + \text{Tr} \left[(\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}}^H \mathbf{M}_d \right. \\ &\quad \times \left(\mathbf{M}_d^H \mathbf{M}_d - \mathbf{M}_d^H \tilde{\mathbf{M}} (\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}}^H \mathbf{M}_d \right)^{-1} \\ &\quad \left. \mathbf{M}_d^H \tilde{\mathbf{M}} (\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} \right] \end{aligned} \quad (\text{D.5})$$

Since $(\mathbf{M}_d^H \mathbf{M}_d - \mathbf{M}_d^H \tilde{\mathbf{M}} (\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}}^H \mathbf{M}_d)^{-1}$ is a principle submatrix of $(\mathbf{M}^H \mathbf{M})^{-1}$ (cf. (D.1)), it is positive definite and so is

$$\begin{aligned} &(\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}}^H \mathbf{M}_d (\mathbf{M}_d^H \mathbf{M}_d \\ &\quad - \mathbf{M}_d^H \tilde{\mathbf{M}} (\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}}^H \mathbf{M}_d)^{-1} \mathbf{M}_d^H \tilde{\mathbf{M}} (\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1}. \end{aligned}$$

The result then follows from (D.5). $\square$

APPENDIX E PROOF OF THEOREM 6.1

We will prove the theorem by induction. We will first show that (6.7) holds for an arbitrary admissible three-level sequence which, according to (C.1) and (C.2), can be parameterized as

$$p(m)^2 = N - (N - 1)\delta - \Delta_{n_0} \quad (\text{E.1})$$

$$(\mathbf{M}^H \mathbf{M})^{-1} = \begin{bmatrix} \left(\tilde{\mathbf{M}}^H \tilde{\mathbf{M}} - \tilde{\mathbf{M}}^H \mathbf{M}_d (\mathbf{M}_d^H \mathbf{M}_d)^{-1} \mathbf{M}_d^H \tilde{\mathbf{M}} \right)^{-1} & \times \\ \times & \left(\mathbf{M}_d^H \mathbf{M}_d - \mathbf{M}_d^H \tilde{\mathbf{M}} (\tilde{\mathbf{M}}^H \tilde{\mathbf{M}})^{-1} \tilde{\mathbf{M}}^H \mathbf{M}_d \right)^{-1} \end{bmatrix}. \quad (\text{D.2})$$

and

$$p(n_0)^2 = \delta + \Delta_{n_0}, \text{ and } p(n)^2 = \delta \text{ for } n \notin \{m, n_0\} \quad (\text{E.2})$$

where $0 \leq n_0 \neq m \leq N-1$. From (C.3), we must have $\Delta_{n_0} \leq N(1-\delta)$. With (E.1) and (E.2), it follows that

$$\begin{aligned} \sum_{n=0}^{N-1} p(n)^{-2} &= \frac{N-2}{\delta} + \frac{1}{\delta + \Delta_{n_0}} + \frac{1}{N - (N-1)\delta - \Delta_{n_0}} \\ &= \frac{N-2}{\delta} + \frac{N(1-\delta)}{-\Delta_{n_0}^2 + N(1-\delta)\Delta_{n_0} + [N - (N-1)\delta]\delta}. \end{aligned} \quad (\text{E.3})$$

It thus suffices to show $J(\Delta_{n_0}) := -\Delta_{n_0}^2 + N(1-\delta)\Delta_{n_0} + [N - (N-1)\delta]\delta$ is minimized by either $\Delta_{n_0} = 0$ or $\Delta_{n_0} = N(1-\delta)$: This confirms that the maximizing $p(n)$ reduces to the two-level form (5.13). Indeed, simple manipulations show that the minimal value of $J(\Delta_{n_0})$ within the interval $\Delta_{n_0} \in [0, N(1-\delta)]$ is $[N - (N-1)\delta]\delta$, which is attained by the two boundary points. Now we assume that (6.7) holds for an arbitrary K -level sequence ($3 < K \leq N$), which is described as

$$p(m)^2 = N - (N-1)\delta - \sum_{k=1}^{K-2} \Delta_{n_k}, \quad (\text{E.4})$$

and

$$p(n)^2 = \delta + \Delta_{n_k} \text{ for } k \in \{1, \dots, K-2\}, \text{ and } p(n)^2 = \delta \text{ for } n \notin n_k. \quad (\text{E.5})$$

Hence, we have

$$\begin{aligned} \sum_{n=0}^{N-1} p(n)^{-2} &= \frac{N - (K-1)}{\delta} + \sum_{k=1}^{K-2} \frac{1}{\delta + \Delta_{n_k}} \\ &\quad + \left(N - (N-1)\delta - \sum_{k=1}^{K-2} \Delta_{n_k} \right)^{-1} \\ &\leq \frac{N-1}{\delta} + \frac{1}{N - (N-1)\delta}. \end{aligned} \quad (\text{E.6})$$

For any $(K+1)$ -level sequence given by

$$p(m)^2 = N - (N-1)\delta - \sum_{k=1}^{K-1} \Delta_{n_k}, \quad (\text{E.7})$$

$$p(n)^2 = \delta + \Delta_{n_k} \text{ for } k \in \{1, \dots, K-1\}, \text{ and } p(n)^2 = \delta \text{ for } n \notin n_k\text{'s} \quad (\text{E.8})$$

we have (E.9), shown at the bottom of the page. With (E.9), it thus remains to check $\Theta \leq 0$. It is easy to verify (E.10), shown at the bottom of the page.

$$\begin{aligned} \sum_{n=0}^{N-1} p(n)^{-2} &= \frac{N-K}{\delta} + \sum_{k=1}^{K-1} \frac{1}{\delta + \Delta_{n_k}} + \left(N - (N-1)\delta - \sum_{k=1}^{K-1} \Delta_{n_k} \right)^{-1} \\ &= \underbrace{\left\{ \frac{N - (K-1)}{\delta} + \sum_{k=1}^{K-2} \frac{1}{\delta + \Delta_{n_k}} + \left(N - (N-1)\delta - \sum_{k=1}^{K-2} \Delta_{n_k} \right)^{-1} \right\}}_{\leq (N-1/\delta) + (1/N - (N-1)\delta)} \\ &\quad + \underbrace{\left\{ \frac{1}{\delta + \Delta_{n_{K-1}}} - \frac{1}{\delta} + \left(N - (N-1)\delta - \sum_{k=1}^{K-1} \Delta_{n_k} \right)^{-1} - \left(N - (N-1)\delta - \sum_{k=1}^{K-2} \Delta_{n_k} \right)^{-1} \right\}}_{:=\Theta}. \end{aligned} \quad (\text{E.9})$$

$$\begin{aligned} \Theta &= \frac{-\Delta_{n_{K-1}}}{(\delta + \Delta_{n_{K-1}})\delta} \\ &\quad + \frac{\Delta_{n_{K-1}}}{\left(N - (N-1)\delta - \sum_{k=1}^{K-1} \Delta_{n_k} \right) \left(N - (N-1)\delta - \sum_{k=1}^{K-2} \Delta_{n_k} \right)}. \end{aligned} \quad (\text{E.10})$$

Subject to the power threshold constraint (5.6b), we must have

$$N - (N - 1)\delta - \sum_{k=1}^{K-1} \Delta_{n_k} \geq \delta,$$

and hence

$$N - (N - 1)\delta - \sum_{k=1}^{K-2} \Delta_{n_k} \geq \delta + \Delta_{n_{K-1}}. \quad (\text{E.11})$$

The two inequalities in (E.11) imply

$$\begin{aligned} & \left(N - (N - 1)\delta - \sum_{k=1}^{K-1} \Delta_{n_k} \right)^{-1} \\ & \times \left(N - (N - 1)\delta - \sum_{k=1}^{K-2} \Delta_{n_k} \right)^{-1} \leq \frac{1}{(\delta + \Delta_{n_{K-1}})\delta} \end{aligned} \quad (\text{E.12})$$

The assertion follows immediately from (E.10) and (E.12). $\square$

ACKNOWLEDGMENT

J.-Y. Wu would like to thank Prof. C.-A. Lin for his valuable suggestions about the organization of Section IV-B and the inclusion of footnote 3, and also thank S.-P. Fang for his help of providing the numerical results in Simulations 6 and 7.

REFERENCES

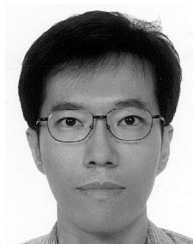
- [1] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.
- [2] N. Al-Dhahir, "Single-carrier frequency-domain equalization for space-time block-coded transmissions over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 5, no. 7, pp. 304–306, Jul. 2001.
- [3] —, "Overview and comparison of equalization schemes for space-time coded signals with applications to EDCG," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2477–2488, Oct. 2002.
- [4] N. Ammar and Z. Ding, "Channel estimation under space-time block coded transmission," in *Proc. Sensor Array Multi-Channel Signal Process. Workshop 2002*, pp. 422–426.
- [5] —, "Frequency selective channel estimation in time-reversed space-time coding," *Proc. Wireless Communications Networking Conf. (WCNC)*, pp. 1838–1843, 2004.
- [6] I. Barhumi, G. Leues, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.
- [7] E. Beres and R. Adve, "Blind channel estimation for orthogonal STBC in MISO systems," *Proc. IEEE Global Telecommunications (IEEE GLOBECOM) 2004*, pp. 2323–2328.
- [8] H. Bolcskei, R. W. Heath, and A. J. Paulraj, "Blind channel estimation in spatial multiplexing systems using nonredundant antenna precoding," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct. 1999, vol. 2, pp. 1127–1132.
- [9] —, "Blind channel identification and equalization in OFDM-based multiantenna systems," *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 96–109, Jan. 2002.
- [10] A. Chevreuil, E. Serpedin, P. Loubaton, and G. B. Giannakis, "Blind channel identification and equalization using periodic modulation precoders: Performance analysis," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1570–1586, Jun. 2000.
- [11] J. Coon, M. Beach, and J. McGeehan, "Optimal training sequence for channel estimation in cyclic-prefix-based single-carrier systems with transmit diversity," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 729–732, Sep. 2004.
- [12] P. J. Davis, *Circulant Matrices*. New York: Wiley, 1979.
- [13] Z. Ding, "Matrix outer-product decomposition method for blind multiple channel identification," *IEEE Trans. Signal Process.*, vol. 45, no. 12, pp. 3053–3061, Dec. 1997.
- [14] Z. Ding and Y. Li, *Blind Equalization and Identification*. New York: Marcel Dekker, 2001.
- [15] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single carrier broadband wireless systems," *IEEE Commun. Mag.*, vol. 40, no. 4, pp. 58–66, Apr. 2002.
- [16] G. B. Giannakis, Y. Hua, and P. Stoica, *Tong, Signal Processing Advances in Wireless and Mobil Communication Volume I: Trends in Channel Identification and Equalization*. Englewood Cliffs, NJ: Prentice Hall PTR, 2001.
- [17] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
- [18] —, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [19] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
- [20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.
- [21] T. P. Krauss and M. D. Zoltowski, "Bilinear approach to multiuser second-order statistics-based blind channel estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2473–2486, Sep. 2000.
- [22] E. G. Larsson and P. Stoica, *Space-Time Block Coding For Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [23] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice Hall, 1974.
- [24] C. A. Lin and J. Y. Wu, "Blind identification with periodic modulation: A time-domain approach," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2875–2888, Nov. 2002.
- [25] Y. P. Lin and S. M. Phoong, "BER minimized OFDM systems with channel independent precoders," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2369–2380, Sep. 2003.
- [26] E. Lindskog and A. Paulraj, "A transmit diversity scheme for delay spread channels," in *Proc. Int. Conf. Communications (ICC)*, New Orleans, LA, pp. 307–311.
- [27] 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. Sel. Areas Commun.*, vol. 19, no. 7, pp. 1352–1364, Jul. 2001.
- [28] Z. Q. Luo, T. N. Davidson, G. B. Giannakis, and K. M. Wong, "Transceiver optimization for block-based multiple access through ISI channels," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1037–1052, Apr. 2004.
- [29] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.
- [30] B. O'Hara and A. Petrick, *The IEEE 802.11 Handbook, A Designer's Companion*. Piscataway, NJ: IEEE Press, 1999.
- [31] E. Serpedin and G. B. Giannakis, "Blind channel identification and equalization with modulation-induced cyclostationarity," *IEEE Trans. Signal Process.*, vol. 46, no. 7, pp. 1930–1944, Jul. 1998.
- [32] S. Shahbazpanahi, A. B. Gershman, and J. H. Manton, "Closed-form blind decoding of orthogonal space-time block codes," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. IV, pp. 473–476, 2004.
- [33] A. L. Swindlehurst and G. Leus, "Blind and semi-blind equalization for generalized space-time block codes," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2489–2498, Oct. 2002.
- [34] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theor.*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.
- [35] C. Tepedelenlioglu, "Maximum multipath diversity with linear equalization in precoded OFDM systems," *IEEE Trans. Inf. Theor.*, vol. 50, no. 1, pp. 232–235, Jan. 2004.
- [36] Z. Wang, X. Ma, and G. B. Giannakis, "OFDM or single-carrier block transmissions?," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 380–394, Mar. 2004.
- [37] J. Y. Wu and T. S. Lee, "Periodic-modulation-based blind channel identification for single-carrier block transmission with frequency-domain equalization," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1114–1130, Mar. 2006.
- [38] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

- [39] S. Zhou and G. B. Giannakis, "Single-carrier space-time block-coded transmissions over frequency-selective fading channels," *IEEE Trans. Inf. Theor.*, vol. 49, no. 1, pp. 164–179, Jan. 2003.
- [40] S. Zhou, B. Muquet, and G. B. Giannakis, "Subspace-based (semi-) blind channel estimation for block precoded space-time OFDM," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1215–1228, May 2002.



Jwo-Yuh Wu (M'04) received the B.S., M.S., and Ph.D. degrees from the National Chiao Tung University, Taiwan, R.O.C., in 1996, 1998, and 2002, respectively, all in electrical and control engineering.

He is currently a Postdoctoral Research Fellow in the Department of Communication Engineering, National Chiao Tung University. His research interests are in signal processing and information theory, with current emphasis on communications and networking.



Ta-Sung Lee (M'94–SM'05) was born in Taipei, Taiwan, R.O.C., in 1960. He received the B.S. degree from National Taiwan University in 1983, the M.S. degree from the University of Wisconsin–Madison, in 1987, and the Ph.D. degree from Purdue University, West Lafayette, IN, in 1989, all in electrical engineering.

In 1990, he joined the Faculty of National Chiao Tung University (NCTU), Hsinchu, Taiwan, R.O.C., where he is a Professor and Chairman of Department of Communication Engineering. His other positions include Technical Advisor at Information and Communications Research Laboratories (CCL) of the Industrial Technology Research Institute (ITRI), Taiwan, R.O.C.; Managing Director of the MINDS Research Center, College of Electrical and Computer Engineering (ECE), NCTU; and Managing Director of Communications & Computer Training Program, NCTU. He is active in the research and development of advanced techniques for wireless communications, such as smart antenna and MIMO technologies, cross-layer design, and software designed radio (SDR) prototyping of advanced communication systems. He has been a coleader of several National Research Programs, such as the Program for Promoting Academic Excellence of Universities—Phases I & II and the 4G Mobile Communications Research Program sponsored by the Taiwan government.

Dr. Lee has won several awards for his research, engineering, and teaching contributions, including National Science Council (NSC) superior research award twice, the 1999 Young Electrical Engineer Award of the Chinese Institute of Electrical Engineers, and the 2001 NCTU Teaching Award.