Periodic-modulation-based blind channel identification for single-carrier block transmission with frequency-domain equalization — Source link ↗

Jwo-Yuh Wu, Ta-Sung Lee

Institutions: National Chiao Tung University

Topics: Circulant matrix, White noise, Pairwise error probability, System identification and Autocorrelation matrix

Related papers:

- Frequency domain equalization for single-carrier broadband wireless systems

- Blind channel identification and equalization with modulation-induced cyclostationarity

- Blind identification with periodic modulation: a time-domain approach

- Blind channel identification and equalization using periodic modulation precoders: performance analysis

- Matrix outer-product decomposition method for blind multiple channel identification

Periodic-Modulation-Based Blind Channel Identification for Single-Carrier Block Transmission With Frequency-Domain Equalization

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

Abstract—This paper proposes a periodic-modulation-based blind channel identification scheme for single-carrier (SC) block transmission with frequency-domain equalization (FDE). The proposed approach relies on the block system model and exploits the circulant channel matrix structure after the cyclic prefix is removed. It is shown that the set of linear equations relating the autocorrelation matrix of the block received signal and the product channel coefficients can be rearranged into one with a distinctive block circulant structure. The identification equations thus obtained lead to a very simple identifiability condition, as well as a natural formulation of the optimal modulating sequence design problem which, based on the block circulant signal structure, can be cast as a constrained quadratic problem that allows for a simple closed-form solution. The impact of the optimal modulating sequence on the peak-to-average power ratio (PAPR) is investigated. Also, it is shown that the optimal sequence results in a consistent channel estimate irrespective of white noise perturbation. Pairwise error probability analysis is used to examine the equalization performance and based on which several design tradeoffs are discussed. Simulation results are used for illustrating the performance of the proposed method.

Index Terms—Blind channel identification, circulant matrix, frequency-domain equalization (FDE), periodic modulation precoder, single-carrier (SC) block transmission, transmitter-induced cyclostationarity (TIC).

I. INTRODUCTION

BLIND channel identification can significantly reduce the training overhead and has been recognized as a bandwidth efficient means for acquiring the channel state information at the receiver [1]. Among the various blind techniques, the second-order-statistics-based methods have recently attracted significant interest (see [1], [7], [10] for a detailed literature review). The transmitter-induced-cyclostationarity (TIC) approach, originally introduced in [24], is known to be one of the major solution branches in this category. In contrast with the multichannel subspace methods [19], [26], which rely on channel diversity via over-sampling or multisensor at the receiver, the TIC approach resorts to signal precoding for facilitating channel identification. The transmitter precoding strategy can lead to identification algorithms free from any restriction on the channel zero location, which is recognized as a fundamental limit of the multichannel subspace methods [23, p. 1930]. Existing TIC precoders include periodic modulation [4], [16], [23], repetition coding [24], and the general filter bank precoder [9], [22]. The periodic modulation precoder, in particular, is attractive for it does not introduce redundancy into the symbol streams and thus potentially prevents the loss in the data rate [23]. Such a precoding technique has also been used for blind identification in spatial multiplexing systems [2], and in the associated multi-carrier scenario as well [3].

Single-carrier (SC) block transmission with cyclic prefix (CP) insertion recently has attracted much attention for its appealing low-complexity frequency-domain equalization (FDE) [8]. However, most of the existing blind identification schemes for SC modulation are devoted to the serial transmission case [7], [10]. It is noted that, with the CP-based block transmission, the signal model of the SC-FDE system shares the essential features as those of orthogonal frequency-division multiplexing (OFDM) [8]. As a result, a number of the blind identification algorithms tailored for OFDM, in particular those exploiting the transmit redundancy due to CP, can also be applied to the SC-FDE systems. Two typical such proposals are the deterministic subspace method [20] and the statistical subspace approach [28]. The former, although being attractive for its finite-sample-convergence property,[1] is still subject to the limitation of the channel zero locations and performance degradation in the low-SNR regime. The latter, on the other hand, is immune to the channel zero pattern but would usually need a large number of data to output a satisfactory identification performance. Instead of CP insertion, another form of guard interval adopted for SC block transmission is through zero-padding [22], [25]. Blind channel identification for such an alternative scheme is addressed in [22]; the proposed approach therein is nonetheless exclusively aimed for to the zero-padded transmit redundancy.

The authors are with the Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan, R.O.C. (e-mail: jywu@cc.nctu.edu.tw; tslee@mail.nctu.edu.tw).

[1]By this we mean the channel can be exactly estimated with a finite number of data samples in the absence of noise [27].
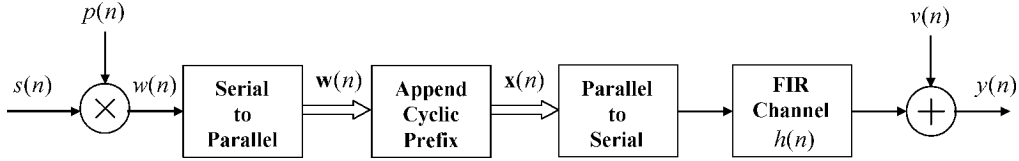
Fig. 1. Discrete-time baseband model of CP-based SC block transmission with periodic modulation.

Leveraging periodic modulation, this paper proposes a correlation-matching-based blind channel identification scheme for single-channel SC-FDE systems. By exploiting the circulant nature of the channel matrix after CP is discarded, it is shown that the set of linear equations relating the autocorrelation matrix of the block received signal and the products of the channel impulse response can be rearranged into one with a distinctive block circulant structure. The products of the channel taps are first computed by solving this set of linear equations, and the impulse response coefficients are then determined, up to a scalar ambiguity, by computing the dominant eigenvector of an associated Hermitian matrix. The proposed identification framework allows for a very simple identifiability condition depending on the modulating sequence alone. It also leads to a natural formulation of the modulating sequence design against the noise effect. By further exploiting the block circulant signal structure, the optimization problem thus formulated is shown to be a constrained quadratic problem, which can be analytically solved. The optimal noise-attenuation sequence would increase the peak-to-average power ratio (PAPR), and a suboptimal solution is provided for PAPR reduction. It is shown that both of the proposed modulating sequences result in consistent channel estimates whenever the channel noise is white. To examine the equalization performance, pairwise error probability (PEP) analysis is provided and based on which various design tradeoffs are discussed. Simulation results are used for illustrating the performance of the proposed method, showing that it is an attractive solution under severe SNR conditions or only a small number of data samples is available.

The rest of this paper is organized as follows. Section II describes the system model. Section III introduces the proposed method, and the associated key features are summarized in Section IV. Section V addresses the optimal design of modulating sequence. Section VI investigates the equalization performance. Section VII is the simulation results. Finally, Section VIII is the conclusion.

Notation: Let $\Re^{m \times n}$ and $C^{m \times n}$ be, respectively, the sets of $m \times n$ real and complex matrices. Denote by $(\cdot)^T, (\cdot)^*$, and $(\cdot)^H$, respectively, the transpose, the complex conjugate, and the complex conjugate transpose. 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 [13, p. 242]. For $\mathbf{X} \in C^{m \times n}$ with $\mathbf{x}_j \in C^m$ denoting the jth column, define $\text{vec}(\mathbf{X}) := [\mathbf{x}_1^T \cdots \mathbf{x}_n^T]^T \in C^{mn}$ to be the vector associated with $\mathbf{X}$ obtained by stacking its columns one on the top of another. The notation Ey stands for the expected value of the random variable y, and $j := \sqrt{-1}$. Throughout the paper, we denote by $\mathbf{F} \in C^{N \times N}$ the fast Fourier transform (FFT) ma-

trix with the klth entry $[\mathbf{F}]_{k,l} := 1/\sqrt{N} \cdot \omega^{-(k-1)(l-1)}$, where $\omega := \exp(j2\pi/N), 1 \le k, l \le N$. The notation $|| \cdot ||$ denotes the two-norm of a vector.

II. SYSTEM MODEL

We consider the discrete-time baseband model of an SC-FDE system as shown in Fig. 1. At the transmitter, the source symbol sequence $s(n)$ is modulated by a (real-valued) periodic sequence $p(n)$ with period N to obtain

$$w(n) = p(n)s(n). \tag{2.1}$$

The purpose of periodic modulation in (2.1) is to introduce some controlled variation in the signal power that can be exploited for blindly identifying the channel at the receiver. The modulated sequence $w(n)$ is then divided into consecutive N-dimensional blocks as

$$\mathbf{w}(n) := [w(nN) \cdots w(nN + N - 1)]^T, \quad n \ge 0. \tag{2.2}$$

Let $\mathbf{s}(n)$ be the source symbol block defined in an analogue way as in (2.2). From (2.1), $\mathbf{w}(n)$ is directly related to $\mathbf{s}(n)$ as

$$\mathbf{w}(n) = \mathbf{P}\mathbf{s}(n), \tag{2.3}$$

where $\mathbf{P}$ is an $N \times N$ diagonal matrix given by

$$\mathbf{P} := diag \{p(0) \cdots p(N-1)\} \in \Re^{N \times N}. \tag{2.4}$$

Before transmission, the block signal $\mathbf{w}(n)$ is appended with a cyclic prefix spanning L_{cp} symbol durations. The resultant data block, with dimension $N + L_{\text{cp}}$, is then transmitted in serial through a discrete-time (assuming symbol rate sampling) finite-impulse-response (FIR) channel $h(\cdot)$ with order L and is contaminated by an additive channel noise $v(n)$. At the receiver, the received samples $y(n)$ are collected into blocks of dimension $N + L_{\text{cp}}$. Assume that the length of CP is no less than the channel order, i.e., $L_{\text{cp}} \ge L$. By discarding the first L_{cp} samples in the received data block, the input–output channel characteristics, in terms of block signals, is described as (assuming that the receiver is synchronized with the transmitter)

$$\mathbf{y}(n) = \mathbf{GPs}(n) + \mathbf{v}(n) \tag{2.5}$$

where

$$\mathbf{y}(n) := [y(Nn) \, y(Nn + 1) \, \cdots \, y(nN + N - 1)]^T \in C^N. \tag{2.6}$$

$\mathbf{G} \in C^{N \times N}$ is circulant with

$$\mathbf{g} := [h(0) \quad h(1) \quad \cdots \quad h(L) \quad 0 \quad \cdots \quad 0]^T \in C^N \quad (2.7)$$

as the first column, and $\mathbf{v}(n)$ is the noise component. Since $\mathbf{G}$ is circulant, it can be diagonalized by the FFT matrix $\mathbf{F}$ so that $\mathbf{G} = \mathbf{F}^{-1}\mathbf{D}\mathbf{F}$, where $\mathbf{D}$ is an $N \times N$ diagonal matrix such that $[\mathbf{D}]_{ii} = \sum_{n=0}^{L} h(n)\omega^{-n(i-1)}$, $1 \le i \le N$. To recover the source symbol block based on (2.5), the FDE strategy [8] is to take an N-point FFT of the received data $\mathbf{y}(n)$ to obtain

$$\mathbf{Fy}(n) = \mathbf{DFPs}(n) + \mathbf{Fv}(n). \quad (2.8)$$

Since $\mathbf{D}$ is diagonal, one can design for each frequency tone a one-tap equalizer, e.g., by using the ZF or the MMSE criterion, to obtain an estimate of the frequency-domain source data $\mathbf{FPs}(n)$, and then transform it back to the time-domain for symbol decision. To accomplish this task, the channel gain matrix $\mathbf{D}$ must be known at the receiver. Based on the pre-FFT data model (2.5), this paper proposes a blind channel identification scheme by using the second-order statistics of the received signal and the knowledge of the modulating sequence $p(n)$. The following assumptions are made in the sequel.

a) The source sequence $s(n)$ is independent identically distributed (i.i.d.) with zero mean and $Es(k)s(l)^* = \delta(k-l)$, where $\delta(\cdot)$ is the Kronecker delta function.
b) The channel noise $v(n)$ is white circularly Gaussian with zero mean and variance σ_v^2 and is uncorrelated with $s(n)$.
c) An upper bound $\overline{L}$ on the channel order L is known and the length of CP satisfies $L_{\mathrm{cp}} \ge \overline{L}$.

III. Blind Channel Identification

A. Identification Equations

We will first consider the noiseless case. Also, we assume that the channel order L is exactly known and discuss the general case later. To identify the channel, one would focus on direct estimation of the N tones of the channel frequency response. Since the block length N could be large, such an identification strategy would be ineffective because the involved computational effort could be intensive. Moreover, for an L-order channel, the frequency-domain channel description in terms of the $N(> L+1)$ frequency response gains yields an excessive parameterization of the channel: direct search for the N unknown frequency tones would be more prone to data errors and in general leads to worse estimation accuracy than the time-domain-based strategies. As a result, we propose to instead estimate the channel impulse response coefficients $h(n), 0 \le n \le L$; the N associated frequency response gains can then be obtained by using FFT operations. The proposed approach lies in exploiting the circulant structure of the channel matrix $\mathbf{G}$. To see this, let us define the $N \times N$ permutation matrix

$$\mathbf{J} := \begin{bmatrix} 0_{1 \times (N-1)} & 1 \\ \mathbf{I}_{N-1} & 0_{(N-1) \times 1} \end{bmatrix} \in \Re^{N \times N}. \quad (3.1)$$

Since $\mathbf{G}$ is circulant, it can be expressed in terms of its first column [cf. (2.7)] as

$$\mathbf{G} = [\mathbf{g} \quad \mathbf{Jg} \quad \cdots \quad \mathbf{J}^{N-2}\mathbf{g} \quad \mathbf{J}^{N-1}\mathbf{g}]. \quad (3.2)$$

Based on (2.5), (3.2), and assumption a), the autocorrelation matrix of the block received signal $\mathbf{y}(n)$ is directly computed as

$$\mathbf{R_y}(0) := E\mathbf{y}(n)\mathbf{y}^H(n) = \sum_{n=0}^{N-1} p(n)^2 \mathbf{J}^n \mathbf{g}\mathbf{g}^H (\mathbf{J}^T)^n. \quad (3.3)$$

For a *given* $\mathbf{R_y}(0)$, the matrix equation (3.3) defines a set of N^2 scalar nonlinear equations in the unknowns $h(0), \cdots, h(L)$. However, if we consider the product channel coefficients of the form $h(k)h^*(l)$ as unknowns, we have a set of N^2 linear equations. As a result, in lieu of directly solving for $h(0), \cdots, h(L)$, we propose to first compute the product coefficients $h(k)h^*(l)$, $0 \le k, l \le L$. We also note from (3.3) that $\mathbf{R_y}(0)$ is a weighted sum of N rank-one matrices, each being an outer product of the circularly shifted zero-padding channel impulse response vector $\mathbf{g}$. In light of this observation, we can further rearrange the N^2 linear equations defined in (3.3) in a more tractable expression, based on which there will be a systematic way of solving for the desired product coefficients. Toward this end, we need the following lemma.

Lemma 3.1 [13, p. 255]: The matrix equation $\sum_{k=1}^{K} \mathbf{A}_k \mathbf{X} \mathbf{B}_k = \mathbf{C}$ can be equivalently expressed as $[\sum_{k=1}^{K} \mathbf{B}_k^T \otimes \mathbf{A}_k]\mathrm{vec}(\mathbf{X}) = \mathrm{vec}(\mathbf{C})$. $\square$

Based on Lemma 3.1, we can immediately rewrite (3.3) as

$$\left[\sum_{n=0}^{N-1} p(n)^2 \mathbf{J}^n \otimes \mathbf{J}^n \right] \mathrm{vec}(\mathbf{g}\mathbf{g}^H) = \mathrm{vec}(\mathbf{R_y}(0)). \quad (3.4)$$

By definition of the Kronecker product, (3.4) can be further rearranged into (3.5), shown at the bottom of the next page. The $N^2 \times N^2$ real-valued matrix $\mathbf{Q}$ defined in (3.5), which is characterized by the N circulant matrices $\{p(0)^2 \mathbf{I}_N, p(N-1)^2 \mathbf{J}^{N-1}, \cdots, p(1)^2 \mathbf{J}\}$ on the top row block, is a block circulant matrix with circulant blocks (BCCB) [5, p.184]. Equation (3.5) forms the basis for the proposed channel identification scheme, which is shown next.

B. Computation of the Product Channel Coefficients

The first step of the proposed approach is to determine the product channel coefficients $h(k)h^*(l)$ for $0 \le k, l \le L$, that is, the unknown vector $vec(\mathbf{g}\mathbf{g}^H)$, based on (3.5). To proceed, it is noted from (2.7) that the vector $\mathbf{g}$ contains $L+1$ channel impulse response taps $h(n)$ for $0 \le n \le L$ followed by $N-L-1$ trailing zeros. As a result, the $N^2 \times N^2$ outer-product matrix $\mathbf{g}\mathbf{g}^H$, and hence the associated vectorized representation $\mathrm{vec}(\mathbf{g}\mathbf{g}^H)$, has actually $(L+1)^2$ nonzero product unknowns. By removing the zero elements in $\mathrm{vec}(\mathbf{g}\mathbf{g}^H)$, and the corresponding indexed columns of the matrix $\mathbf{Q}$, (3.5) can be further simplified as a set

of N^2 scalar equations in $(L+1)^2$ unknowns. Specifically, let us define

$$\mathbf{h} := [h(0) \quad \cdots \quad h(L)]^T \in C^{L+1} \qquad (3.6)$$

to be the desired channel impulse response vector. Then, from (2.7), we immediately have

$$\mathbf{g} = \begin{bmatrix} \mathbf{h}^T & \mathbf{0}_{(N-L-1)\times 1}^T \end{bmatrix}^T \qquad (3.7)$$

and hence

$$\mathbf{g}\mathbf{g}^H = \begin{bmatrix} \mathbf{h}\mathbf{h}^H & \mathbf{0}_{(L+1)\times(N-L-1)} \\ \mathbf{0}_{(N-L-1)\times(L+1)} & \mathbf{0}_{N-L-1} \end{bmatrix}. \qquad (3.8)$$

From (3.8) and by definition of the $\mathrm{vec}(\cdot)$ operation, it can be seen that the $(L+1)^2$ nonzero product unknowns in $\mathrm{vec}(\mathbf{g}\mathbf{g}^H)$ are divided into $L+1$ groups of $L+1$ elements each, all clustering within the first $N(L+1)$ entries of the vectorized representation. Moreover, there are $N-L-1$ zero entries inserted between each pair of the unknown coefficient groups. Toward a more compact description of (3.5) by discarding the zero entries in $\mathrm{vec}(\mathbf{g}\mathbf{g}^H)$, let $\overline{\mathbf{Q}} \in \Re^{N^2 \times N(L+1)}$ be the matrix obtained from $\mathbf{Q}$ by deleting its last $N(N-L-1)$ columns. Also, define

$$\mathbf{J}_L := \begin{bmatrix} \mathbf{I}_{L+1} \\ \mathbf{0}_{(N-L-1)\times(L+1)} \end{bmatrix} \in \Re^{N\times(L+1)} \qquad (3.9)$$

which is simply the matrix containing the first $L+1$ columns of $\mathbf{I}_N$. Based on the above observations, (3.5) can be simplified as

$$\underbrace{\overline{\mathbf{Q}}(\mathbf{I}_{L+1} \otimes \mathbf{J}_L)}_{:=\widetilde{\mathbf{Q}}} \mathrm{vec}(\mathbf{h}\mathbf{h}^H) = \mathrm{vec}(\mathbf{R_y}(0)). \qquad (3.10)$$

Assume that the matrix $\widetilde{\mathbf{Q}} \in \Re^{N^2 \times (L+1)^2}$ defined in (3.10) is of full column rank. Then (3.10) defines a set of overdetermined and consistent linear equations, from which the product channel coefficient vector can be uniquely recovered as

$$\mathrm{vec}(\mathbf{h}\mathbf{h}^H) = (\widetilde{\mathbf{Q}}^T \widetilde{\mathbf{Q}})^{-1} \widetilde{\mathbf{Q}}^T \mathrm{vec}(\mathbf{R_y}(0)). \qquad (3.11)$$

It should be noted that, since $\mathbf{R_y}(0)$ is Hermitian, there are actually $N(N+1)/2$ independent equations in (3.10). Therefore, one can discard those redundant output measurements

from (3.10) for a further dimension reduction. This, however, will wipe off the imbedded structure in $\widetilde{\mathbf{Q}}$ that is inherited from the BCCB matrix $\mathbf{Q}$ (recall that $\widetilde{\mathbf{Q}}$ is obtained from $\mathbf{Q}$ simply by deleting its columns). As we will see in Section V, when channel noise is present, the signal structure specified by the matrix $\widetilde{\mathbf{Q}}$ can allow for a very effective way of modulating sequence design against noise effect. Hence, we propose to reserve the output redundancy and rely on (3.10) for computing the product channel coefficients. We note that the imbedded structure of the matrix $\widetilde{\mathbf{Q}}$ can be exploited for limiting the computational effort paid for this figure of merit. This will be discussed later on.

C. Identification of the Channel Impulse Response

Assume that the product unknowns $h(k)h^*(l)$, $0 \le k, l \le L$ are available from (3.11). To identify the channel, let us form the $(L+1) \times (L+1)$ Hermitain matrix defined as

$$\mathbf{H} = [\mathbf{H}_{k,l}]_{0 \le k,l \le L}, \quad \text{with} \quad \mathbf{H}_{k,l} = h(k)h(l)^*. \qquad (3.12)$$

Theoretically, the matrix $\mathbf{H}$ is of rank-one and can be factorized as $\mathbf{H} = \mathbf{h}\mathbf{h}^H$, where $\mathbf{h} \in C^{L+1}$ is defined in (3.6). The channel impulse response $h(0), \cdots, h(L)$, can thus be identified, up to a scalar ambiguity, by computing the eigenvector associated with the maximal eigenvalue of the matrix $\mathbf{H}$. We note that similar "bilinear" identification strategy is also used in the previous works [6], [14], and [16].

IV. CHARACTERISTICS OF THE PROPOSED APPROACH

This section shows several essential features regarding the proposed channel identification scheme. We characterize the identifiability condition and address the case when channel order overestimation occurs. We then summarize the proposed algorithm and conclude with some remarks.

A. Identifiability Condition

From the discussions in Section III-B, it is easy to see that the channel can be identified if the product channel coefficients $h(k)h^*(l)$, $0 \le k, l \le L$, are uniquely determined from (3.10), which is the case if the matrix $\widetilde{\mathbf{Q}}$ is of full column rank. Since $\widetilde{\mathbf{Q}}$ is a submatrix of $\mathbf{Q}$ obtained by deleting its columns, a sufficient condition for channel identifiability is that $\mathbf{Q}$ is nonsingular. Based on the block circulant structure of $\mathbf{Q}$, there is a rather elegant way of specifying this sufficient condition. More precisely, we have the following proposition, whose proof is given in Appendix A.

$$\underbrace{\begin{bmatrix} p(0)^2 \mathbf{I}_N & p(N-1)^2 \mathbf{J}^{N-1} & \cdots & p(2)^2 \mathbf{J}^2 & p(1)^2 \mathbf{J} \\ p(1)^2 \mathbf{J} & p(0)^2 \mathbf{I}_N & \cdots & p(3)^2 \mathbf{J}^3 & p(2)^2 \mathbf{J}^2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ p(N-2)^2 \mathbf{J}^{N-2} & p(N-3)^2 \mathbf{J}^{N-3} & \cdots & p(0)^2 \mathbf{I}_N & p(N-1)^2 \mathbf{J}^{N-1} \\ p(N-1)^2 \mathbf{J}^{N-1} & p(N-2)^2 \mathbf{J}^{N-2} & \cdots & p(1)^2 \mathbf{J} & p(0)^2 \mathbf{I}_N \end{bmatrix}}_{:=\mathbf{Q}} \mathrm{vec}(\mathbf{g}\mathbf{g}^H) = \mathrm{vec}(\mathbf{R_y}(0)). \qquad (3.5)$$

Proposition 4.1: Let $\mathbf{F}$ be the $N \times N$ FFT matrix and define

$$\mathbf{p} := \begin{bmatrix} p(0)^2 & p(1)^2 & \cdots & p(N-1)^2 \end{bmatrix}^T \in \Re^N. \quad (4.1)$$

Then the matrix $\mathbf{Q}$ defined in (3.5) is nonsingular if and only if the vector $\mathbf{Fp}$ contains no zero entries. $\quad\square$

Proposition 4.1 asserts that channel identifiability is guaranteed if there are no "spectral nulls" associated with the N-dimensional vector $\mathbf{p}$ defined in (4.1). This condition holds for almost all choices of $p(n)$. In particular, we can appropriately select $p(n)$ to guarantee channel identifiability and, moreover, to obtain a well-conditioned equation (3.10) against noise perturbation and finite-sample effect. This will be shown in Section V.

B. On Channel Order Overestimation

The previous discussions on channel identifiability assume that the channel order L is known. Now we consider the case that only an upper bound $\overline{L} \geq L$ is available. As long as the length of the CP is no less than $\overline{L}$, the circulant structure of the resultant data model (2.5) is still preserved; in particular, the channel matrix $\mathbf{G}$ is circulant with first column equal to $[h(0)\ h(1)\ \cdots\ h(\overline{L})\ 0\ \cdots\ 0]^T$. Assume that $p(n)$ is chosen to satisfy the sufficient condition for channel identifiability. The solution given in (3.11) will yield the product coefficients $h(k)h^*(l), 0 \leq k \leq l \leq \overline{L}$. The associated $(\overline{L}+1) \times (\overline{L}+1)$ rank-one matrix $\overline{\mathbf{H}}$ defined similarly as in (3.12) will accept the dominant eigenvector

$$\eta\begin{bmatrix} h(0) & h(1) & \cdots & h(L) & 0 & \cdots & 0 \end{bmatrix}^T \in C^{\overline{L}+1} \quad (4.2)$$

for some scalar η. From (4.2), the actual channel order, together with the channel impulse response, can be determined by the most significant entries of the unit-norm eigenvector associated with the maximal eigenvalue of $\overline{\mathbf{H}}$. The proposed method is thus applicable whenever an upper bound on the channel order is available. The price paid for channel order overestimation, however, is the increase in algorithm complexity (it requires the computation of $(\overline{L}+1)^2$ product unknowns followed by an eigen-decomposition of an $(\overline{L}+1) \times (\overline{L}+1)$ Hermitian matrix) as well as the decrease in the effective data rate (since a longer CP would be used).

C. Algorithm Outline and Discussions

We summarize the proposed channel identification procedures as an algorithm.

```
Proposed Blind Channel Identification
Algorithm
1)  Estimate the autocorrelation matrix
    R_y(0) via the time average
```

$$\hat{\mathbf{R}}_\mathbf{y}(0) = \frac{1}{K} \sum_{i=1}^{K} \mathbf{y}(i)\mathbf{y}^H(i) \quad (4.3)$$

```
    where K is the number of received
    data blocks.
```

```
2)  Form the matrix equation as in
    (3.10), and select a modulating se-
    quence p(n) with which the matrix Q
    is nonsingular.
3)  Compute the product channel coeffi-
    cients h(k)h*(l), 0 ≤ k ≤ l ≤ L, based
    on (3.11).
4)  Form the Hermitian matrix H as in
    (3.12). The channel impulse re-
    sponse vector is then computed as
    the eigenvector associated with the
    maximal eigenvalue of H.
```

We conclude this section with the following remarks.

a) *Circulant Channel Matrix Structure*: Based on periodic modulation, similar time-domain correlation matching approach is adopted in [16] for the SC serial transmission case. There is, however, a substantial difference in the resultant algorithm features between the previous work [16] and the present study for SC-FDE systems. In particular, the work [16] exploits the Toeplitz structure of the channel matrices in the associated block system model; the autocorrelation matrix therein is shown to define $L+1$ decoupled groups of linear equations with product channel taps as unknowns (cf. [16, p. 2878]). For the SC-FDE systems considered in this paper, the circulant property of the channel matrix, on the other hand, leads to a set of coupled equations with a distinctive BCCB structure as shown in (3.5).

b) *On Channel Identifiability*: Typical existing blind identification algorithms applicable to SC-FDE systems are the subspace-based methods [20], [28]. The deterministic approach [20] is particularly attractive for its good performance at high SNR region but is subject to the limitations on the locations of the channel zeros [20, p. 1702]. It would suffer a performance loss whenever the channel contains zeros exactly on, or nearby, the N frequency grids. The identifiability condition of the proposed method (cf. Proposition 4.1), on the contrary, is completely determined by the modulating sequence $p(n)$, which is at the designers' disposal (such an appealing feature is also seen in the serial transmission case [4], [16], [23]). In light of this observation, the proposed method can be a more robust solution against the channel zero locations. We note that the statistical-based method [28] also possesses this figure of merit but, as will be demonstrated via numerical simulations, would suffer from a slow decay of the estimation errors.

c) *Least-Squares Solution*: The solution (3.11) for product unknowns yields the exact solution provided that the autocorrelation matrix $\mathbf{R}_\mathbf{y}(0)$ is obtained perfectly and when noise is absent. When only a time average of $\mathbf{R}_\mathbf{y}(0)$ is available (as in practical situations), the solution (3.11) then gives the least squares solutions.

d) *On the Block Length N*: It is noted that the number of equations (N^2) in (3.10) increases with the block length N. In case that the equation errors, due to finite-sample estimation and noise effect, are present in (3.10) and can

be modeled as a zero-mean white noise with a fixed variance, the associated error covariance can be made arbitrary small if the number of equations is large enough [18, p. 178]. As a result, it seems preferable to choose a long block length not only to promote the spectral efficiency but also to improve the channel estimation accuracy. However, the unlimited increase in N would turn out to be problematic in practice. This is because, under a fixed number of received samples S, the available number of data blocks for estimating the time-average $\hat{\mathbf{R}}_{\mathbf{y}}(0)$ is about S/N. Hence, if N is too large and only a small number of samples is available, one would come up with a poorly estimated $\hat{\mathbf{R}}_{\mathbf{y}}(0)$, hence usually a less accurate channel estimate.

e) *Dependency Among the Product Unknowns*: We note that there are actually only $(L+2)(L+1)/2$ independent unknowns in the product channel coefficient vector $\text{vec}(\mathbf{h}\mathbf{h}^H)$, say, $h(k)h(l)^*$ for $0 \leq k \leq l \leq L$ (the complex-conjugated version of which will yield the rest). In the presence of noise and finite-sample estimation errors, the "joint" recovery of $h(k)h(l)^*$, for $0 \leq k$, $l \leq L$, via (3.11) does not necessarily restore the complex-conjugate dependency among these coefficients. If this inherent constraint is violated, one can first collect the estimate of $h(k)h(l)^*$'s for $0 \leq k \leq l \leq L$, and the respective complex conjugates, to form a Hermitian matrix $\mathbf{H}_1$, and likewise form $\mathbf{H}_2$ instead based on the estimated $h(l)h(k)^*$'s. The channel is then identified as the mean of the two dominant eigenvectors of $\mathbf{H}_1$ and $\mathbf{H}_2$. Another plausible technique to acquire the channel estimate is to bypass the inconsistency removal process, perform an SVD of the non-Hermitian $\mathbf{H}$, and then identify the channel as the associated dominant left singular vector. In our numerical tests the computed $\mathbf{H}$ matrix, however, is observed to remain Hermitian even when SNR is extremely low (< -30 dB), and an unreasonably small number of data blocks (such as $K = 5$) is used for estimating $\hat{\mathbf{R}}_{\mathbf{y}}(0)$ based on (4.3). The aforementioned inconsistency, therefore, does not seem to occur.

f) *Removal of the Scalar Ambiguity*: To recover the transmit signal, some training symbols must be inserted in the data blocks for removing the scalar ambiguity. Toward this end, let us express the true channel as $\mathbf{h} = \alpha\hat{\mathbf{h}}$, where $\hat{\mathbf{h}}$ denotes the dominant eigenvector of the rank-one matrix $\mathbf{H}$ defined in (3.12), and α is the scalar unknown. Since $\mathbf{F}\mathbf{h} = \alpha\mathbf{F}\hat{\mathbf{h}}$, the diagonal matrix $\mathbf{D}$ consisting of the true frequency response gains is thus $\mathbf{D} = \alpha\hat{\mathbf{D}}$, where $\hat{\mathbf{D}}$ is the estimate of $\mathbf{D}$ based on $\hat{\mathbf{h}}$. The frequency-domain system equation (2.8) then instead reads

$$\mathbf{F}\mathbf{y}(n) = \alpha\hat{\mathbf{D}}\mathbf{F}\mathbf{P}\mathbf{s}(n) + \mathbf{F}\mathbf{v}(n). \qquad (4.4)$$

To resolve α based on (4.4), we shall first "partly" equalize the channel to get

$$\hat{\mathbf{x}}(n) := \mathbf{F}^H\hat{\mathbf{D}}^{-1}\mathbf{F}\mathbf{y}(n) = \alpha\mathbf{P}\mathbf{s}(n) + \mathbf{F}^H\hat{\mathbf{D}}^{-1}\mathbf{F}\mathbf{v}(n). \qquad (4.5)$$

In the absence of noise, the unknown α can then be determined using only one pilot symbol θ_0, placed at the ith component of $\mathbf{s}(n)$, as $\alpha = [\hat{\mathbf{x}}(n)]_i/(p(i)\theta_0))$, where $[\hat{\mathbf{x}}(n)]_i$ is the ith entry of $\hat{\mathbf{x}}(n)$. When noise is present, more pilots should be incorporated so that the noise effect in the resultant estimate can be averaged out, thus improving the solution accuracy.

V. OPTIMAL DESIGN OF MODULATING SEQUENCE

This section considers the noisy case and addresses the problem of modulating sequence design for combating the noise effect. Based on the noise-perturbed equations, we will first introduce the design criterion. By further exploiting the BCCB structure of the matrix $\mathbf{Q}$, the optimization problem is then formulated as a constrained quadratic problem which yields an analytic solution. The optimal sequence, as one will see, may enlarge the PAPR and a suboptimal alternative for PAPR reduction is introduced. We then show that the proposed solutions yield consistent channel estimates irrespective of white noise perturbation. Finally, some computational aspects are discussed.

A. Optimality Criterion

In the following discussions, we assume without loss of generality that the channel order L is known. Assume that the channel noise is present. From (2.5) and (3.3), the autocorrelation matrix of the block received signal is thus

$$\mathbf{R}_{\mathbf{y}}(0) = \left(\sum_{n=0}^{N-1} p(n)^2 \mathbf{J}^n \mathbf{g}\mathbf{g}^H (\mathbf{J}^T)^n \right) + \sigma_v^2 \mathbf{I}_N. \qquad (5.1)$$

By following the same procedures as shown in Section III, the equations for computing the product channel coefficients, viz., the noisy version of (3.10), then becomes

$$\text{vec}\left(\mathbf{R}_{\mathbf{y}}(0)\right) = \widetilde{\mathbf{Q}}\text{vec}(\mathbf{h}\mathbf{h}^H) + \sigma_v^2\text{vec}(\mathbf{I}_N). \qquad (5.2)$$

Since the noise variance σ_v^2 is unknown, it is in general impossible to determine the exact solution $\text{vec}(\mathbf{h}\mathbf{h}^H)$ from the noisy data. For a given $\mathbf{R}_{\mathbf{y}}(0)$, we can only instead compute the least squares solution, given as

$$\begin{aligned} &\text{vec}(\mathbf{h}\mathbf{h}^H)_{LS} \\ &:= (\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}})^{-1}\widetilde{\mathbf{Q}}^T\text{vec}\left(\mathbf{R}_{\mathbf{y}}(0)\right) \\ &= \text{vec}(\mathbf{h}\mathbf{h}^H) + \sigma_v^2(\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}})^{-1}\widetilde{\mathbf{Q}}^T\text{vec}(\mathbf{I}_N). \end{aligned} \qquad (5.3)$$

From (5.3), it is easy to see that the least squares solution coincides with the exact solution if, and only if, the noise signature $\text{vec}(\mathbf{I}_N)$ is orthogonal to the range space of $\widetilde{\mathbf{Q}}$, that is

$$\widetilde{\mathbf{Q}}^T\text{vec}(\mathbf{I}_N) = \mathbf{0}_{(L+1)^2 \times 1}. \qquad (5.4)$$

In fact, if we think of the product channel coefficients $\text{vec}(\mathbf{h}\mathbf{h}^H)$ as the signal of interest in (5.2), the range space of $\widetilde{\mathbf{Q}}$ then defines the signal subspace, whereas the noise perturbation signature $\text{vec}(\mathbf{I}_N)$ spans the noise subspace. Accordingly, the condi-

tion (5.4) amounts to the requirement that the signal and noise subspaces are mutually orthogonal. Since the matrix $\widetilde{\mathbf{Q}}$ is completely determined by the modulating sequence $p(n)$, one natural design criterion, therefore, is to choose $p(n)$ to meet the orthogonality constraint (5.4), or to match it as exactly as we can if perfect fulfillment is impossible. This suggests the following performance measure:

$$\gamma := \max_i \frac{|\widetilde{\mathbf{q}}_i^T \text{vec}(\mathbf{I}_N)|}{\|\widetilde{\mathbf{q}}_i\| \cdot \|\text{vec}(\mathbf{I}_N)\|} \tag{5.5}$$

where $\widetilde{\mathbf{q}}_i \in \Re^{N^2}$ is the ith column of the matrix $\widetilde{\mathbf{Q}}$. We note that γ thus defined is the maximal correlation index among the pairs of vectors $\{\widetilde{\mathbf{q}}_i, \text{vec}(\mathbf{I}_N)\}$ for all $1 \le i \le (L+1)^2$; it gauges the worst-case tendency of noise contamination upon all the signal components, and can serve as a good measure of "closeness," and hence orthogonality, between the signal and noise subspaces. Small values of γ, in particular, imply small noise contribution on the desired signals and are expected to yield better channel estimation accuracy. To achieve an utmost noise reduction, we then propose to minimize the quantity γ in (5.5), subject to the following two constraints:

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

and

$$p(n)^2 \ge \delta > 0, \quad 0 \le n \le N-1. \tag{5.7}$$

The constraint (5.6) normalizes the average transmit power within one block to unity. The constraint (5.7), on the other hand, imposes a threshold on the minimal modulated power for equalization feasibility; this is because zero, or too small, transmit power will prevent symbol recovery since the source sequence is uncorrelated. For the serial transmission counterpart, these two constraints are used in [4] and [16] for modulating sequence design. We note that the proposed design formulation is similar to that reported in [16]. However, the underlying difference between the respective system models results in different nature of the cost functions: the performance metric in [16] is immediately specified by one vector pair, whereas the quantity γ in (5.5) calls for an exhaustive search over all the $(L+1)^2$ signal components.

Remark: A commonly adopted criterion for improving the accuracy of the least-squares estimate (5.3) is to minimize the squared norm of the noise perturbation, that is, $\sigma_v^2 \|(\tilde{\mathbf{Q}}^T \tilde{\mathbf{Q}})^{-1} \tilde{\mathbf{Q}}^T \text{vec}(\mathbf{I}_N)\|^2$. However, such a cost function is highly nonlinear in the design parameters $p(n)$'s; given the two constraints (5.6) and (5.7), there does not seem to exist tractable procedures for obtaining a solution. In the current formulation, another plausible design approach is to choose $p(n)$ toward the best match of the orthogonality condition (5.4), say, to minimize $\|\widetilde{\mathbf{Q}}^T \text{vec}(\mathbf{I}_N)\|^2$. In contrast with the proposed min-max correlation metric (5.5), this alternative cost function can be regarded as a measure of the sum correlation among the noise signature and all the columns of $\widetilde{\mathbf{Q}}$. Unfortunately, in our

case, the quantity $\|\widetilde{\mathbf{Q}}^T \text{vec}(\mathbf{I}_N)\|^2$ turns out to be a constant independent of $p(n)$ (see the proof of Proposition 5.1 given below). It is thus impossible to rely on this strategy for noise suppression. In the next subsection, we will see the advantages of the proposed design criterion: the associated cost function is quadratic in nature and allows for an analytic solution. $\quad\square$

B. Optimal Solution

The proposed optimization problem, which aims for minimizing the worst-case noise corruption, appears to be one of the min-max type. Toward a solution, one has to first determine the maximal correlation index among all the vector pairs $\{\widetilde{\mathbf{q}}_i, \text{vec}(\mathbf{I}_N)\}, 1 \le i \le (L+1)^2$. Since each $\widetilde{\mathbf{q}}_i$ is a column of the matrix $\mathbf{Q}$ in (3.5), by exploiting the BCCB structure of $\mathbf{Q}$ there is an elegant expression of γ in terms of $p(n)$, as is shown in the next proposition. This result will lead to an equivalent constrained quadratic problem formulation which can yield an analytic solution.

Proposition 5.1: Let γ be defined as in (5.5). Then we have

$$\gamma = \sqrt{\frac{N}{\sum_{n=0}^{N-1} p(n)^4}}. \tag{5.8}$$

$\quad\square$

Proof: For a fixed i, it is easy to check from (3.5) that $\|\widetilde{\mathbf{q}}_i\| = \sqrt{\sum_{n=0}^{N-1} p(n)^4}$, which is independent of the column index i. Since $\|\text{vec}(\mathbf{I}_N)\| = \sqrt{N}$, it thus remains to show

$$\max_i |\widetilde{\mathbf{q}}_i^T \text{vec}(\mathbf{I}_N)| = N. \tag{5.9}$$

Let $\mathbf{e}_n$ be the nth unit standard vector in $\Re^N$. Then, by definition, we have

$$\text{vec}(\mathbf{I}_N) = \begin{bmatrix} \mathbf{e}_1^T & \mathbf{e}_2^T & \cdots & \mathbf{e}_N^T \end{bmatrix}^T. \tag{5.10}$$

Denote by $\mathbf{q}_i$ the ith column of $\mathbf{Q}$, $1 \le i \le N^2$. For each $1 \le l \le N$, it follows from (3.5) that

$$\mathbf{q}_{(l-1)N+l} = \begin{bmatrix} c_{l,0}\mathbf{e}_1^T & c_{l,1}\mathbf{e}_2^T & \cdots & c_{l,N-2}\mathbf{e}_{N-1}^T & c_{l,N-1}\mathbf{e}_N^T \end{bmatrix}^T \tag{5.11}$$

where

$$\begin{bmatrix} c_{l,0} & c_{l,1} & \cdots & c_{l,N-1} \end{bmatrix}^T = \mathbf{J}^{l-1}\mathbf{p} \tag{5.12}$$

in which $\mathbf{J} \in \Re^{N \times N}$ and $\mathbf{p} \in \Re^N$ are defined in (3.1) and (4.1), respectively. From (5.10)–(5.12), we have, for each $1 \le l \le N$

$$\mathbf{q}_{(l-1)N+l}^T \text{vec}(\mathbf{I}_N) = \sum_{n=0}^{N-1} c_{l,n} = \sum_{n=0}^{N-1} p(n)^2 = N \tag{5.13}$$

where the last equality follows from (5.6). The maximal inner product between $\text{vec}(\mathbf{I}_N)$ and the columns of $\mathbf{Q}$ is indeed equal to N. This is because $\mathbf{q}_i^T \text{vec}(\mathbf{I}_N) = 0$ whenever $i \ne (l-1)N+l$, since there is no overlap between the locations of the respective nonzero entries. Since the matrix $\widetilde{\mathbf{Q}}$ is constructed

TABLE I
PAPR (IN DECIBELS) FOR VARIOUS CHOICES OF MINIMAL POWER THRESHOLD δ (BLOCK LENGTH $N = 32$)

constellations	$\delta = 1$	$\delta = 0.9$	$\delta = 0.8$	$\delta = 0.7$	$\delta = 0.6$	$\delta = 0.5$	$\delta = 0.4$	$\delta = 0.3$	$\delta = 0.2$	OFDM
PSK	0.97	7.10	9.54	11.10	12.24	13.14	13.89	14.53	15.09	15.05
16-QAM	3.52	9.65	12.10	13.65	14.79	15.70	16.44	17.08	17.64	17.60
64-QAM	4.65	10.78	13.22	14.78	15.92	16.82	17.57	18.21	18.77	18.73
256-QAM	5.20	11.32	13.77	15.33	16.47	17.37	18.12	18.76	19.31	19.28

from $\mathbf{Q}$ by first deleting its last $N - L - 1$ column blocks, and then in each remaining column block further deleting the last $N - L - 1$ columns [cf. (3.9) and (3.10)], we then have

$$
\widetilde{\mathbf{q}}_i^T \text{vec}(\mathbf{I}_N)
= \begin{cases} N, & \text{for } i = (l-1)(L+1)+l \text{ with } 1 \leq l \leq L+1 \\ 0, & \text{when otherwise.} \end{cases} \tag{5.14}
$$

Equation (5.9) then follows from (5.14). □

In contrast with equation (4.9) in [16, p. 2880], it is somewhat surprising to see from (5.8) that the cost function γ in (5.5) regarding the SC-FDE scenario turns out to be identical with the one reported for the serial transmission case, regardless of the fundamental difference in the respective signal models. Given the same design constraints (5.6) and (5.7), this coincidence could result from certain "inherent circularity" in the respective identification equations: the noise-corrupted equation group in [16] is described by a circulant matrix (cf. [16, p. 2879]), whereas the identification equations (5.2) inherit the BCCB structure from the matrix $\mathbf{Q}$ defined in (3.5). To find the solution, we can adopt the approach used in [16, p. 2880], which lies in a natural formulation of the original optimization problem as a constrained quadratic one.

Specifically, based on (5.8), the minimization of the worst-case orthogonality measure γ is equivalent to maximizing the quantity $\sum_{n=0}^{N-1} p(n)^4$, which is nonlinear in $p(n)$ but is quadratic if we think of $p(n)^2$'s as unknowns. In terms of the vector $\mathbf{p}$ defined in (4.1), the objective function then becomes

$$
\sum_{n=0}^{N-1} p(n)^4 = \|\mathbf{p}\|^2. \tag{5.15}
$$

The constraints (5.6) and (5.7), respectively, can be equivalently expressed via $\mathbf{p}$ as

$$
\|\mathbf{p}\|_1 = \sum_{n=0}^{N-1} p(n)^2 = N \tag{5.16}
$$

and

$$
\mathbf{p}_n \geq \delta > 0, \quad 0 \leq n \leq N - 1 \tag{5.17}
$$

where $\mathbf{p}_n (= p(n)^2)$ denotes the nth component of $\mathbf{p}$. With (5.15)–(5.17), the optimization problem is thus essentially a

constrained quadratic optimization problem: Maximize $\|\mathbf{p}\|^2$, subject to the constraints (5.16) and (5.17). The solution to this problem is detailed in [16, p. 2880], and the resulting optimal sequence is given in the next theorem.

Theorem 5.2 [16, p. 2880]: The optimal sequence that minimizes the objective function γ subject to the constraints (5.6) and (5.7) is given by, for any fixed $0 \leq m \leq N - 1$

$$
|p(m)| = \sqrt{N(1-\delta)+\delta}, \quad \text{and} \quad |p(n)| = \sqrt{\delta} \text{ for } n \neq m. \tag{5.18}
$$
□

The optimal sequence in (5.18) assumes two levels with one single power peak. It delivers the minimal threshold power, equal to δ, over $N - 1$ time periods, and then pins the remaining power on one single instant. It is noted that, in the serial transmission scenario, the same solution is also reported in [4] through maximizing a certain measure of the frequency-domain SNR.

C. Tradeoff Between Estimation Accuracy and PAPR

With the optimal sequence (5.18), the resultant optimal orthogonality measure can be found as

$$
\gamma_{\text{opt}} := \left(\sqrt{N - (N-1)\delta(2-\delta)} \right)^{-1}. \tag{5.19}
$$

It can be directly shown that, for $0 < \delta < 1$, $\delta(2 - \delta)$ is an increasing function in δ. As a result, for a fixed block length N, the optimal measure γ_{opt} will decrease as δ is decreased. Hence, a small power threshold δ implies small noise contribution on the signals, and hence attains better channel estimation accuracy. However, if δ is too small, the peak value $N(1-\delta)+\delta$ would be large, leading to high PAPR. Since one major advantage of the SC-FDE systems over OFDM is their low PAPR [8], it is thus desirable to examine the impact of the modulation-induced power peak on PAPR. For $N = 32$, $L_{\text{cp}} = 8$, and with the Nyquist pulse shaping filter, Table I lists the values of PAPR for various choices of δ with respect to different symbol constellations; the results are obtained based on [25, p. 383]. As benchmark performances, Table I also includes two sets of PAPR values for both the SC-FDE system without power modulation (designated by the $\delta = 1$ case) and OFDM. From the table, we can see that the optimal modulating sequence (5.18) does lead to an increased PAPR over the equal-power case. In particular, the choice $\delta = 0.3$ raises the PAPR to a level comparable to that of OFDM systems.

To limit the PAPR, one plausible approach, without resorting to essential modification and reformulation of the proposed optimality criterion, is to include a peak power bound in the constraint (5.7), leading to

$$\delta \leq p(n)^2 \leq \rho, \quad \text{for some } \rho \geq 1. \tag{5.20}$$

In (5.20), for a fixed δ, we must have $\rho \leq N(1-\delta)+\delta$, which is the maximal allowable power. Given the modified constraint (5.20), there appears to be no analytical approaches to finding a solution. A conjectured solution is instead provided in [16, p. 2881] as

$$|p(m)| = \sqrt{\rho}, \quad \text{and} \quad |p(n)| = \sqrt{\frac{(N-\rho)}{(N-1)}} \quad \text{for} \quad n \neq m \tag{5.21}$$

where $0 \leq m \leq N-1$ is fixed but arbitrary. If we form the Lagrangian associated with the modified optimization problem, the solution (5.21) does verify the first-order necessary condition for optimality (see [16, App. B]); the numerical tests, e.g., using the qp algorithm in [12, p. 3–38], tend to moreover confirm this postulation. We note that the sequence (5.21) is also in a two-level form with single power spike as the one given by (5.18), but it shows a different way of power allocation: the peak power is set to attain the upper bound ρ, and the remaining power is then evenly delivered over the other $N-1$ time instants. Since the floor power $(N-\rho)/(N-1)$ would often be larger than the minimal threshold δ, the solution (5.21), however, will enlarge the orthogonal measure (5.19): it reduces PAPR at the sacrifice of the channel estimation accuracy.

D. Consistency of the Channel Estimate

This subsection shows that the optimal modulating sequence (5.18), as well as the modified solution (5.21) for PAPR reduction, can result in a consistent channel estimate irrespective of white noise perturbation. This is a direct consequence of the following proposition (see Appendix B for the proof).

Proposition 5.3: Let $p(n)$ be a two-level modulating sequence of the form

$$p(m)^2 = \alpha, \quad \text{and} \quad p(n)^2 = \beta \text{ for } 0 \leq n \leq N-1 \text{ with } n \neq m \tag{5.22}$$

which satisfies the power normalization constrain (5.6). Let $\mathbf{H}$ defined in (3.12) be the rank-one matrix containing the product channel coefficients in the noise-free case. Assume that, in the presence of white channel noise, the matrix $\mathbf{H}$ is perturbed to $\mathbf{H}_w$, whose entries are computed based on (5.3). Then we have

$$\mathbf{H}_w = \mathbf{H} + \zeta \mathbf{I}_{L+1} \quad \text{for some scalar } \zeta. \tag{5.23}$$

$\square$

Proposition 5.3 asserts that, as long as a two-level sequence of the form (5.22) is used, the white-noise induced perturbation to the rank-one matrix $\mathbf{H}$ is simply a scalar multiple of $\mathbf{I}_{L+1}$. Such a perturbation preserves the eigenvector, in particular, the one associated with the maximal eigenvalue, the channel impulse response vector $\mathbf{h}$ can thus be identified, up to a scalar ambiguity,

by computing the dominant eigenvector of the matrix $\mathbf{H}_w$. Since the two modulating sequences given in (5.18) and (5.21) are of the form (5.22), they both yield consistent channel estimates; the optimal solution (5.18), in the mean while, achieves the largest noise reduction.

E. Computational Aspects

It is noted that the optimal orthogonality measure γ_{opt} in (5.19) is independent of the index m at which the peak power occurs: any $0 \leq m \leq N-1$ affords the best achievable noise reduction. However, from the computational point of view, the selection of the index m would be crucial to the computational accuracy of the least squares solution (5.3). This is because different m will result in different $\widetilde{\mathbf{Q}}$, and hence possibly different condition number associated with the resultant $\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}}$ matrix. In fact, as long as $p(n)$ is of the two-level form (5.22), it can be shown (after some manipulations) that all the choices $0 \leq m \leq N-1$ lead to an identical $\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}}$ matrix, which is symmetric Toeplitz with

$$\begin{bmatrix} a_0\mathbf{e}_1^T & a_1\mathbf{e}_2^T & a_1\mathbf{e}_3^T & \cdots & a_1\mathbf{e}_{L+1}^T \end{bmatrix} \tag{5.24}$$

as the first row, where, for $1 \leq l \leq L+1$, $\mathbf{e}_l$ in (5.24) instead denotes the lth unit standard vector in $\Re^{L+1}$, and

$$a_0 = \alpha^2+(N-1)\beta^2, \quad \text{and} \quad a_1 = 2\alpha\beta+(N-2)\beta^2. \tag{5.25}$$

As a result, the least-squares solution (5.3) is free from any additional risk of numerical sensitivity due to different location of m. Our simulations also indicate that, for moderate choices of the minimal power threshold δ, the matrix $\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}}$ remains well-conditioned, e.g., for $N = 32$ and $\delta = 0.7$, the condition number of $\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}}$ is about 3.84. This can be beneficial to relieving the effect of finite-sample estimation errors on the estimated channels.

Remarks:

1) Since the last L_{cp} symbols in each source data block will be duplicated in the guard period as the CP, the peak value should not be located within these trailing symbols so as to conserve the power resource.

2) One commonly used method for computing the least squares solution (5.3) is through the normal equation approach [11, p. 238]. Since $\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}} \in \Re^{(L+1)^2 \times (L+1)^2}$ is symmetric Toeplitz, one computationally efficient scheme for solving the normal equation is the Levinson algorithm [11, p. 196], which can limit the number of flops to $4(L+1)^4$ (counting both addition and multiplication). We note that there can be a further saving in the overall flop cost if we take into account the sparse nature of the $\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}}$ matrix (cf. (5.24)). $\square$

VI. ON EQUALIZATION PERFORMANCE

Our discussions thus far focus on exploiting the periodic transmit power variation for blind channel estimation. This section investigates the impact of periodic modulation on the equalization performance. To better predict the intrinsic system features due to power modulation, we assume that the channel is perfectly known at the receiver; the same assumption is also

made in [4] for performance analysis of the serial transmission strategy. To access the equalization performance, one candidate approach is through exact bit-error-rate (BER) analysis but this is in general quite involved. In what follows, we will instead resort to PEP analysis for the considered SC-FDE systems. This technique is widely used in the study of space–time communications [15], [21], and recently also in performance evaluation for multi-carrier block transmission [17]; it is believed to yield a good approximation to the exact BER when SNR is high [17].

A. ZF Pairwise Error Probability

Most of the PEP analysis in the literature is based on the maximum-likelihood (ML) decision metric [15], [17], [21]. In the considered block transmission scenario, in which the block length can be as large as 512 or 1024 [8], the computational load of the ML-based detection would be rather prohibitive. Since one major advantage of the SC-FDE systems is the frequency-domain per-tone equalization [8], commonly in conjunction with the linear ZF or MMSE criterion, it is thus desirable, and is what we will consider in the sequel, to conduct a PEP analysis within the linear equalization framework. In the following discussion we focus on the signal model (2.5) with ZF equalization; the time index n will be dropped for notational simplicity.

The PEP measures the probability that a symbol block $\mathbf{s}$ is transmitted but another $\widetilde{\mathbf{s}} \neq \mathbf{s}$ is detected. Given a channel realization $\mathbf{h}$, the PEP conditioned on $\mathbf{h}$ is by definition given by

$$\Pr[\mathbf{s} \to \widetilde{\mathbf{s}} | \mathbf{h}] = \Pr\left[\|\hat{\mathbf{s}} - \widetilde{\mathbf{s}}\| < \|\hat{\mathbf{s}} - \mathbf{s}\| \big| \mathbf{h}\right] \qquad (6.1)$$

where $\hat{\mathbf{s}}$ is the estimate of $\mathbf{s}$ under the ZF metric and, in the considered case, is equal to

$$\hat{\mathbf{s}} = \mathbf{s} + (\mathbf{GP})^{-1}\mathbf{v}. \qquad (6.2)$$

Let $d := \|\mathbf{s} - \widetilde{\mathbf{s}}\|$ be the distance between the two symbol blocks and define $\mathbf{e} := (\widetilde{\mathbf{s}} - \mathbf{s})/d$ to be the normalized error vector. With $\hat{\mathbf{s}}$ given in (6.2), it can be directly verified from (6.1) that

$$\Pr[\mathbf{s} \to \widetilde{\mathbf{s}} | \mathbf{h}] = \Pr\left[u > \left(\frac{d}{2}\right) \Big| \mathbf{h}\right]$$
$$\text{where} \quad u := \mathrm{Re}\left\{\mathbf{v}^H (\mathbf{GP})^{-H}\mathbf{e}\right\}. \quad (6.3)$$

Since the noise $v(\cdot)$ is zero mean Gaussian with variance σ_v^2, it is easy to check that, for a given $\mathbf{h}$, the random variable u defined in (6.3) is also Gaussian, with zero mean and variance

$$E\left|\mathrm{Re}\left\{\mathbf{v}^H (\mathbf{GP})^{-H}\mathbf{e}\right\}\right|^2 = \left(\frac{\sigma_v^2}{2}\right)\|(\mathbf{GP})^{-1}\mathbf{e}\|^2. \quad (6.4)$$

With (6.3) and (6.4), we immediately have

$$\Pr[\mathbf{s} \to \widetilde{\mathbf{s}} | \mathbf{h}] = Q\left(\frac{d}{\sqrt{2\sigma_v^2 \|(\mathbf{GP})^{-1}\mathbf{e}\|^2}}\right) \qquad (6.5)$$

where $Q(\cdot)$ denotes the error function. By the submultiplicative property of matrix norm [11, p. 55], we have

$$\|(\mathbf{GP})^{-1}\mathbf{e}\| = \|\mathbf{P}^{-1}\mathbf{G}^{-1}\mathbf{e}\| \leq \|\mathbf{P}^{-1}\|_F \cdot \|\mathbf{G}^{-1}\|_F \cdot \|\mathbf{e}\|$$
$$= \|\mathbf{P}^{-1}\|_F \cdot \|\mathbf{G}^{-1}\|_F \qquad (6.6)$$

where the last equality in (6.6) follows since $\mathbf{e}$ is by definition a unit-norm vector ($\|\cdot\|_F$ denotes the Frobenius norm). From (6.6) and since $\|\mathbf{P}^{-1}\|_F^2 = \sum_{n=0}^{N-1} p(n)^{-2}$ (cf. (2.4)), the conditional PEP in (6.5) is upper bounded by

$$\Pr[\mathbf{s} \to \widetilde{\mathbf{s}} | \mathbf{h}] \leq Q\left(\frac{d}{\sqrt{2\sigma_v^2 \left(\sum_{n=0}^{N-1} p(n)^{-2}\right) \|\mathbf{G}^{-1}\|_F^2}}\right). \qquad (6.7)$$

The upper bound in (6.7) allows us to investigate the effects of $p(n)$ on equalization performance and is discussed next.

B. Discussions

1) Design Tradeoff: From (6.7), it follows that the upper bound is minimized if the summation $\sum_{n=0}^{N-1} p(n)^{-2}$ attains the minimum. Since, by the Cauchy's inequality

$$\left(p(0)^{-2} + p(1)^{-2} + \cdots + p(N-1)^{-2}\right)$$
$$\cdot \left(p(0)^2 + p(1)^2 + \cdots + p(N-1)^2\right)$$
$$\geq (1 + \cdots + 1)^2 = N^2 \qquad (6.8)$$

and $\sum_{n=0}^{N-1} p(n)^2 = N$ (cf. (5.6)), we have

$$\left(p(0)^{-2} + p(1)^{-2} + \cdots + p(N-1)^{-2}\right) \geq N \qquad (6.9)$$

with equality holds in (6.9) if and only if $p(n) = 1$ for all $0 \leq n \leq N-1$, viz., the equal power case. As a result, leveraging periodic power modulation for channel identification can incur a BER performance loss. We note that, for the serial transmission case, this tendency is also shown in [4]. To sustain a target BER, an immediate remedy is to enlarge the average transmit power. Indeed, in case that an M-fold increase in the average power is allowed so that $\sum_{n=0}^{N-1} p(n)^2 = MN$, inequality (6.8) then leads to

$$\left(p(0)^{-2} + p(1)^{-2} + \cdots + p(N-1)^{-2}\right) \geq \frac{N}{M} \qquad (6.10)$$

which shows a resultant decrease in the minimal values of $\sum_{n=0}^{N-1} p(n)^{-2}$, and hence a decrease in the upper bound of PEP in (6.7). Hence, a large average power tends to buy an error floor margin for tolerating the performance drop due to power variation. If, however, the allowable transmit power is limited, one has to instead use a lower order constellation for maintaining the desired BER level: This would decrease the effective spectral efficiency. Such a design tradeoff pertaining to the periodic modulation-based identification is also pointed out in [3, p. 99]. Instead of increasing the average power, an advisable

TABLE II
TEST CHANNELS USED IN THE SIMULATIONS

	$h(0)$	$h(1)$	$h(2)$	$h(3)$	$h(4)$	$h(5)$	$h(6)$	$h(7)$	$h(8)$
Channel A	2.3823	-1.1217	-0.7250	0.5284	0.3561	-0.2979	0.2843	-0.1825	0.1082
	+1.8580j	-1.0801j	-0.8724j	-0.6103j	-0.6360j	-0.2570j	-0.3301j	+0.2194j	-0.0667j
Channel B	1.4590	0.5688	-0.3629	-1.8766	0.1717	-0.4309	-0.8961	-0.3983	0.1482
	+1.1379j	-1.4944j	-1.4702j	-0.9466j	+0.4604j	-0.0389j	+0.2675j	+0.5030j	+0.1900j

alternative to reducing the modulation-induced penalty is to "turn off" the periodic modulator, and hence restore the power efficiency for the subsequent source data, whenever a reliable channel estimate is available. Since the proposed method tends to yield a reasonably good identification performance with only a small number of data samples (see Simulation 2), the period of the power modulation phase could be kept small so as to relieve this penalty.

2) Equalization Aspect of the Optimal Sequence (5.18): We note that the optimal noise-attenuation sequence (5.18), as well as the PAPR-limiting solution (5.21), is of the two-level form (5.22). This motivates us to investigate the impact of the two-level modulation scheme, with single peak power (i.e., $\alpha > 1$) and satisfying (5.6) and (5.7), on the PEP. To see this, it is noted from (5.22) that

$$\sum_{n=0}^{N-1} p(n)^{-2} = \frac{1}{\alpha} + \frac{N-1}{\beta} = \frac{1}{\alpha} + \frac{(N-1)^2}{(N-\alpha)} \qquad (6.11)$$

where the last equality in (6.11) follows from the power normalization constraint, namely, $\alpha + (N-1)\beta = N$. From (6.11), it can be verified that $\sum_{n=0}^{N-1} p(n)^{-2}$ increases with α whenever $\alpha > 1$. As a result, the quantity $\sum_{n=0}^{N-1} p(n)^{-2}$ is maximized, and hence the upper bound on PEP in (6.7) assumes the maximum, if α grows to attain the largest allowable peak $\alpha = N(1-\delta) + \delta$: This is the case when $p(n)$ is chosen as in (5.18). As a result, the solution (5.18), albeit optimal in regard to the identification performance, appears to be the worst choice from the equalization point of view. In this sense, the modified solution (5.21) for PAPR reduction could avoid significant degradation in BER, even though it results in a less accurate channel estimate.

VII. SIMULATION RESULTS

In this section we use several numerical simulations to illustrate the performance of the proposed method. We consider two test channels, both with $L = 8$, whose coefficients are listed in Table II, in which channel B has three zeros very close to $\exp(j22\pi/32)$, $\exp(j42\pi/32)$, and $\exp(j56\pi/32)$. The lengths of the symbol block and CP, respectively, are $N = 32$ and $L_{\rm cp} = 8$; the source symbols are drawn from the QPSK constellation. The channel identification performance is gauged by the normalized mean square error defined by NMSE $:= (1/I)\sum_{i=1}^{I} \|\hat{\mathbf{h}}^{(i)} - \mathbf{h}\|^2 \cdot \|\mathbf{h}\|^{-2}$, where I is the total number of Monte Carlo runs (we set $I = 200$ in our simulations), and $\hat{\mathbf{h}}^{(i)}$ is the estimated channel in the ith trial. The SNR is defined as the ratio of the noiseless block
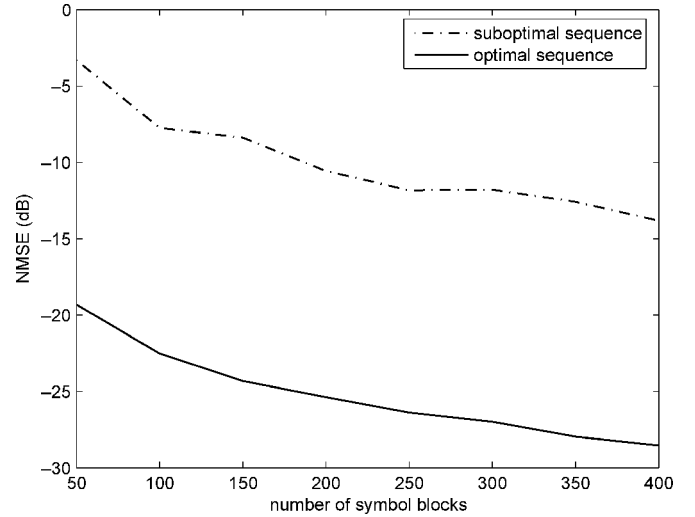

Fig. 2. NMSE obtained by two different modulating sequences (channel A).

received signal power to the corresponding noise power, that is, SNR $:= \|\mathbf{h}\|^2/\sigma_v^2$. Throughout the simulations, the index of peak power in the optimal sequence (5.18) is set to be $m = 0$.

Simulation 1—Effectiveness of the Optimal Modulating Sequence (5.18): This simulation demonstrates the effectiveness of the optimal sequence (5.18). For SNR = 10 dB and $\delta = 0.7$, we consider the optimal sequence (5.18) and another nonoptimal solution chosen as $p(n)^2 = 0.7$ for $0 \le n \le 15$ and $p(n)^2 = 1.3$ for $16 \le n \le 31$. We note that the comparative candidate also satisfies the two constraints (5.6) and (5.7), and restores the power efficiency of the last half symbol block as compared with the solution (5.18). Fig. 2 shows the respective computed NMSE (for channel A) versus number of symbol blocks. As we can see from the figure, the optimal solution significantly improves the performance.

Simulation 2—Comparison With Subspace Methods [20], [28]: This simulation compares the identification performances of the proposed method and the two subspace methods [20], [28] (abbreviated as d-subspace and s-subspace methods, respectively). In the first experiment, we use the three methods to estimate channel A; the proposed method (depicted by solid line in the figures) is implemented with various choices of the minimal power threshold δ. For fixed SNR = 10 dB, Fig. 3 shows the respective NMSE for different numbers of symbol blocks. From the figure, we can see that the performance of the proposed method is improved as δ decreases (for the moderate selection $\delta = 0.6$, the NMSE is kept below -20 dB when the number of symbol blocks is as small as 50). This is because a small δ leads to a small orthogonality measure (5.19) and
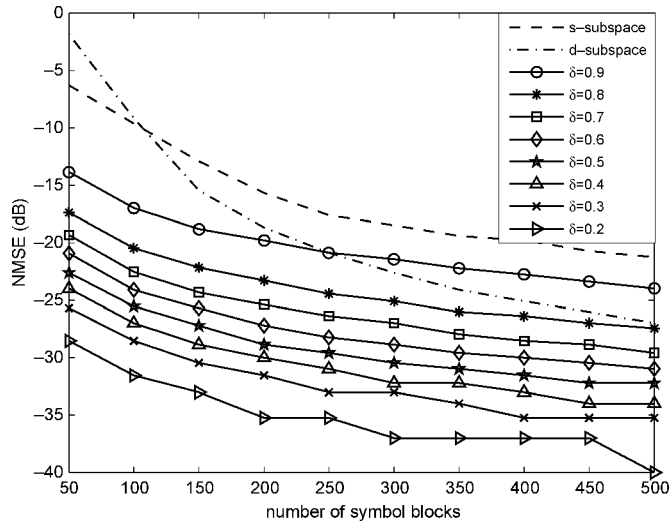
Fig. 3.   NMSE for different numbers of symbol blocks (channel A).
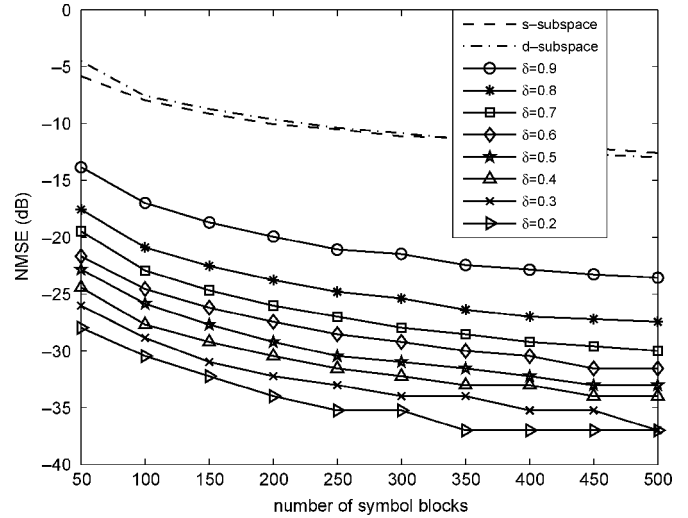

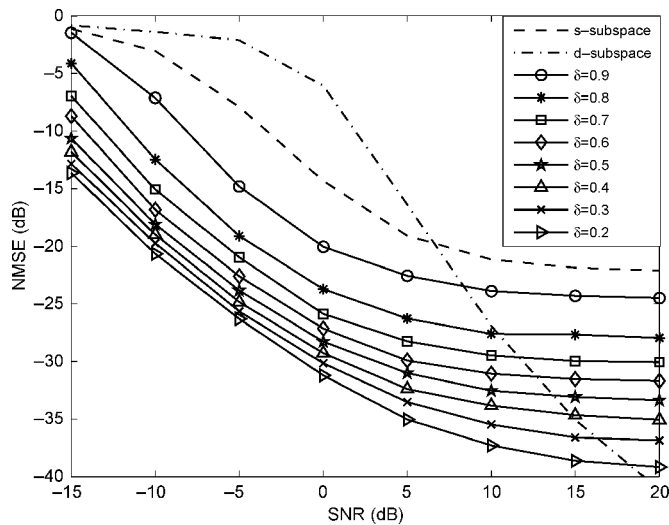Fig. 5.   NMSE for different numbers of symbol blocks (channel B).

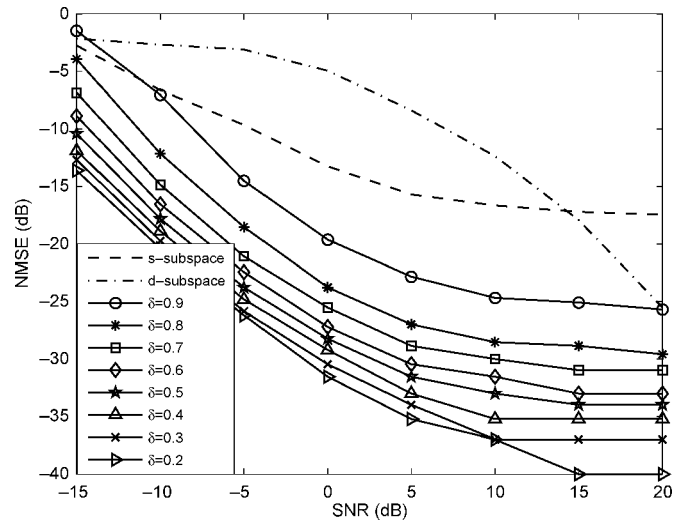Fig. 4.   NMSE at various SNR levels (channel A).

Fig. 6.   NMSE at various SNR levels (channel B).

hence good estimation accuracy. Also, compared with the subspace methods, the proposed method performs better as long as $\delta \leq 0.8$. Fig. 4 shows the NMSE of the three methods at various SNR levels (with 300 symbol blocks). The result shows that, in the medium-to-low SNR region (< 10 dB), the proposed method attains the best performance even with $\delta = 0.9$ for small power variation. For SNR larger than 15 dB, the output NMSE of the method [20] exhibits a fast decay. This is not unexpected since, owing to the finite-sample-convergence property, the deterministic-based approach can usually yield impressive estimation accuracy when SNR is high [24, p. 1789]. We note that, in the serial transmission scenario, similar tendency is also observed in [23, p. 1942] when modulation-based identification is compared with the (deterministic) multichannel subspace methods [19], [26]. In the second experiment, we repeat the above simulation procedures with respect to channel B, which contains zeros near the spectral nulls and tends to violate the identifiability condition of the deterministic approach [20, p. 1702]. Fig. 5 shows the NMSE for different numbers of data blocks (at SNR = 10 dB). In contrast with Fig. 3, there appears

to be around a 10-dB gap between the method [20] and our method with $\delta = 0.9$ irrespective the number of samples. Fig. 6 shows that NMSE at different SNR. We can see from the figure that, although the deterministic approach [20] still tracks the channel as SNR increases, the resultant NMSE significantly deteriorates for a wide range of SNR as compared with Fig. 4; the proposed method, on the other hand, appears to be more robust against the variation of the channel zero pattern. Based on simulations, the performance of the method [20] is observed to further degrade for channels with more zeros hitting (around) the FFT grids. We finally note that, although both the proposed approach and the subspace method [28] are statistical based in nature, the former can better track the channels. This is because the method [28] directly relies on the "raw" estimated autocorrelation matrix for channel identification, and could be more sensitive to the finite-sample estimation errors. For the proposed method, the identification equations (3.10) would usually remain well-conditioned whenever the optimal sequence (5.18) with a moderate (or small) δ is used: this could act as an in-built mechanism for combating the effect of imperfect
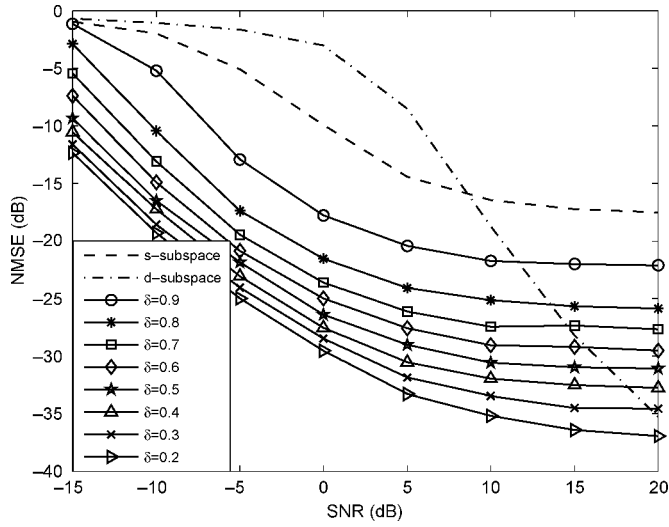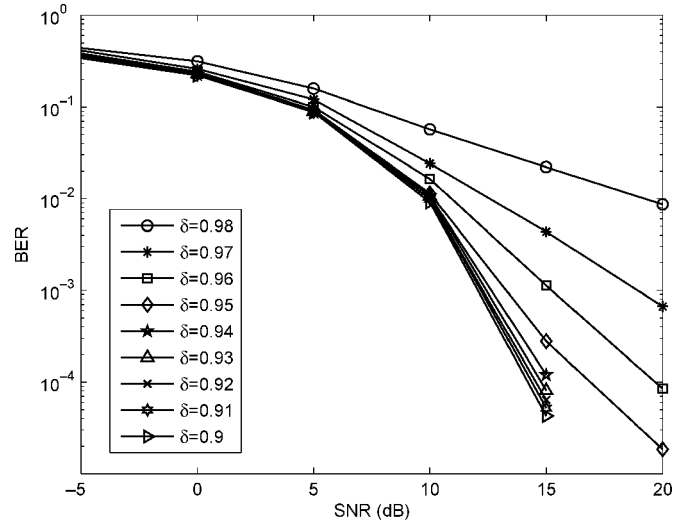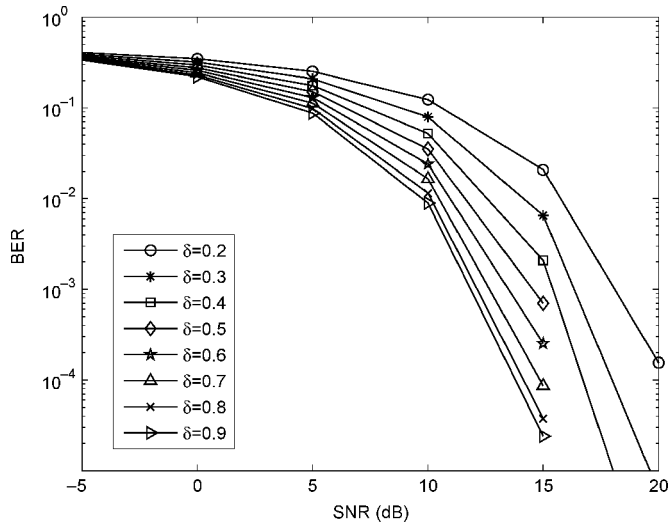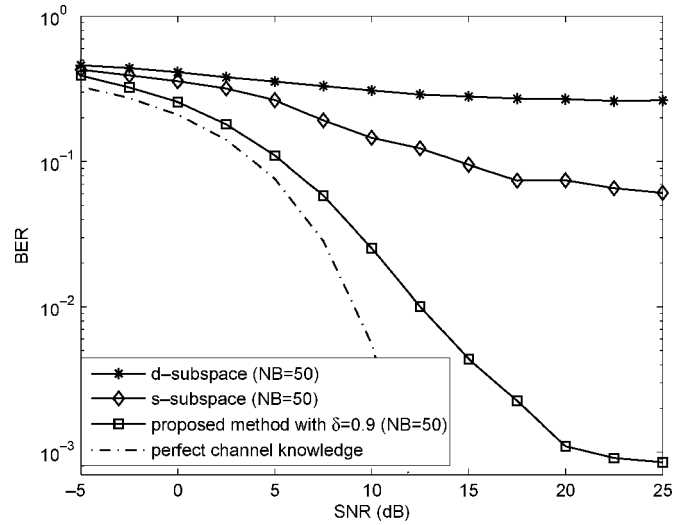
Fig. 7.   NMSE at various SNR levels (channel A; 16-QAM constellation).

Fig. 9.   BER performance for $0.9 \leq \delta \leq 0.98$.

Fig. 8.   BER performance for $0.2 \leq \delta \leq 0.9$.

Fig. 10.   BER performances of three methods (NB $= 50$).

data estimation. Finally, Fig. 7 depicts the computed NMSE at different SNR (for channel A) when the nonconstant modulus 16-QAM constellation is instead adopted. The curves are seen to show essentially the same tendencies as in the QPSK case (cf. Fig. 4). This is expected since the applicability of the three methods is not restricted to any particular constellation type: the proposed scheme and the method in [28] rely on the i.i.d. source assumption, whereas the deterministic approach [20] further relaxes the statistical requirement to certain "persistent excitation" condition [20, p. 1702].

Simulation 3—On Equalization Performances: This simulation investigates the equalization performance associated with the proposed scheme. We consider channel A and use the proposed method (with the optimal sequence (5.18)) for channel identification; the frequency-domain ZF equalizer is then implemented based on the estimated channel. Fig. 8 shows the BER versus SNR curves for $0.2 \leq \delta \leq 0.9$ (with 300 symbol blocks). From the figure we can see that, although a large δ results in a less accurate channel estimate, the BER is nonetheless improved as δ increases. This tendency reflects the result

of the ZF-PEP analysis in Section VI: A large minimal power threshold can limit the power penalty and improve the equalization performance. Another reason for this phenomenon is that, although small δ's bring good channel estimation accuracy, the resultant "effective" channel matrix **GP** [cf. (2.5)] is rendered ill-conditioned, and hence the decision process becomes more prone to the data errors. Fig. 9 shows the computed BER for $0.9 \leq \delta \leq 0.98$. It can be seen that $\delta = 0.9$ attains the lowest BER, and beyond which the performance degrades. This is because, if δ is too large, the channel estimation performance will be quite poor and then dominates the achievable BER floor ($\delta = 0.9$ thus seems to be a compromising selection). The last experiment compares the equalization performances achieved by the proposed method (with $\delta = 0.9$) and the two subspace methods [20], [28]. The number of symbol blocks is set to be 500. Fig. 10 and Fig. 11 show the BER versus SNR curves obtained, respectively, when NB $= 50$ and NB $= 80$ initial received data blocks are used for channel estimation (the perfect channel knowledge case is also shown as the benchmark performance). For the proposed method, only the first 50,
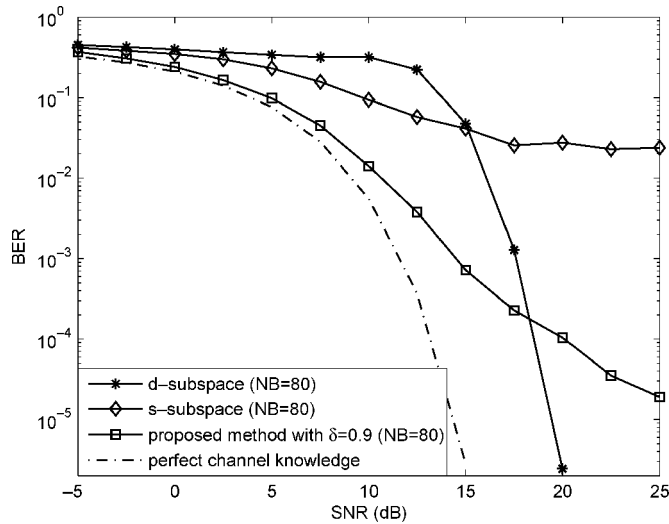
Fig. 11. BER performances of three methods (NB = 80).
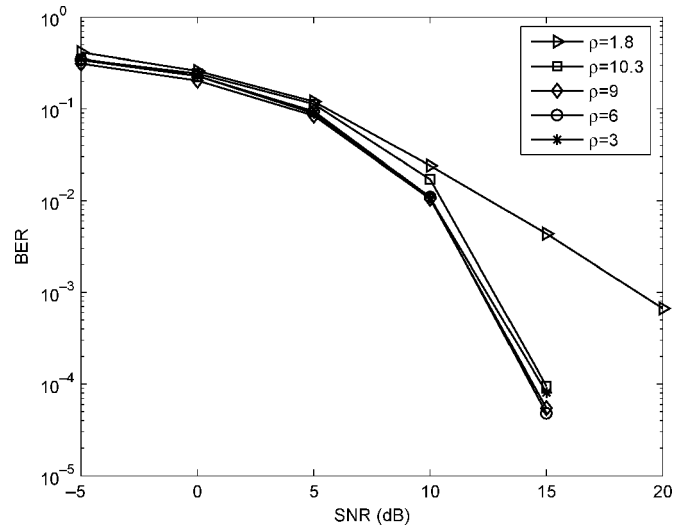

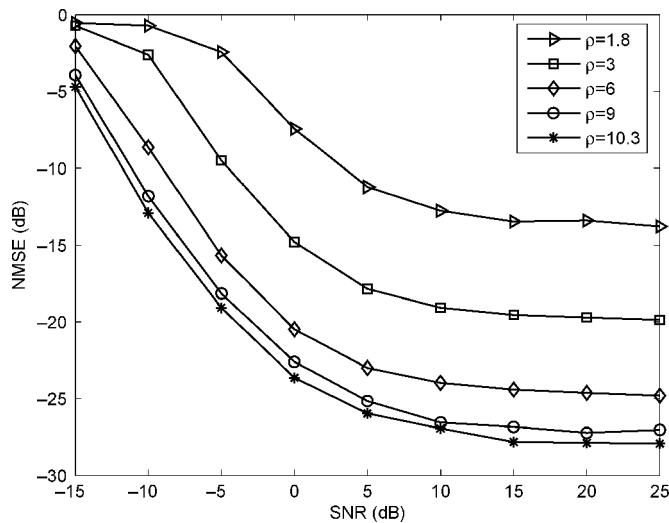Fig. 13. BER for different peak power upper bounds.

Fig. 12. NMSE for different peak power upper bounds.
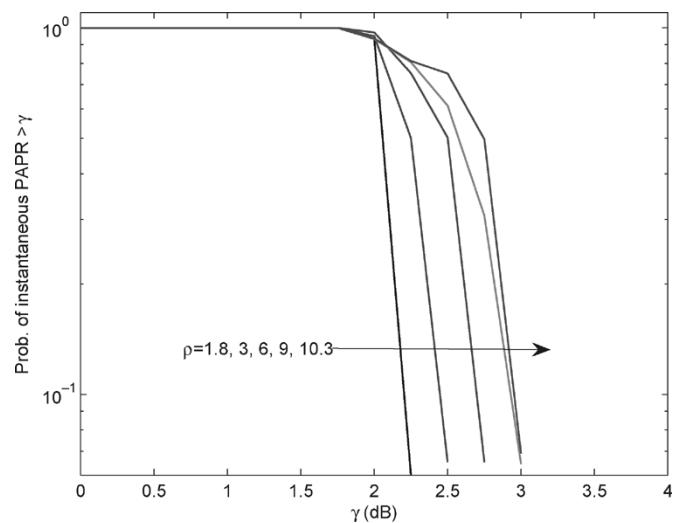

Fig. 14. Tendency of the instantaneous PAPR.

and 80, symbol blocks are modulated by the optimal sequence (5.18) and afterwards the symbols are transmitted with equal (unity) power. It is seen from the figures that, for NB = 50, the two subspace methods result in poor equalization outputs, whereas the proposed method significantly improves the performance. As NB increases, the BER curve of the deterministic approach [20] exhibits a fast roll-off for high SNR, but the proposed method still outperforms the two comparative solutions in the medium-to-low SNR region.

Simulation 4—Performance of the PAPR-Limiting Sequence (5.21): This simulation illustrates the performance of the PAPR-limiting solution (5.21) with respect to different upper bound ρ. We consider channel A and set the minimal power threshold to be $\delta = 0.7$, with which the maximal peak power is $32 - 31 \cdot 0.7 = 10.3$. Fig. 12 shows the computed NMSE for $\rho = 1.8, 3, 6, 9,$ and 10.3 (the corresponding floor powers are, respectively, 0.97, 0.93, 0.84, 0.74, 0.7). The result shows that the NMSE increases as ρ is decreased; this is because a small upper bound leads to a large floor power, and thus enlarges the optimal orthogonality measure (5.19). Fig. 13 shows the

respective computed BER, which is seen to slightly improve as ρ decreases from 10.3 to 3 (this tendency has also been seen in the previous simulations). It is thus seems plausible to choose a fairly small ρ for reducing a reasonable amount of PAPR without entailing BER degradation (in our case, there is about a 3-dB reduction in PAPR as ρ is decreased from 10.3 to 3). However, if ρ is too small, the incurred channel identification errors will then spoil the equalization performance. One should also note from Table I that, when the upper bound ρ increases from 1.8 to 10.3, the incurred PAPR could be large when high-order QAM constellations are used: for the 64-QAM constellation in particular, there is about a 10-dB penalty. However, 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 [25]. A more realistic performance metric for describing the actual power amplification portrait is the *instantaneous* PAPR [25, p. 383]. For each considered upper bound ρ, Fig. 14 shows the probability that the instantaneous PAPR exceeds a prescribed value γ (64-QAM constellation is used). As we
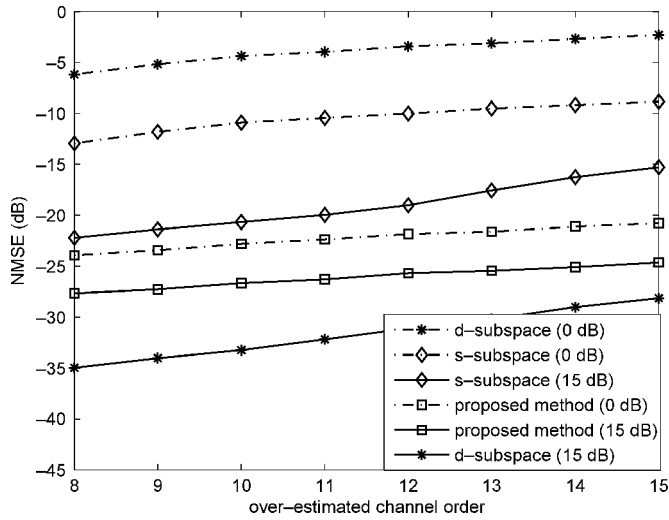
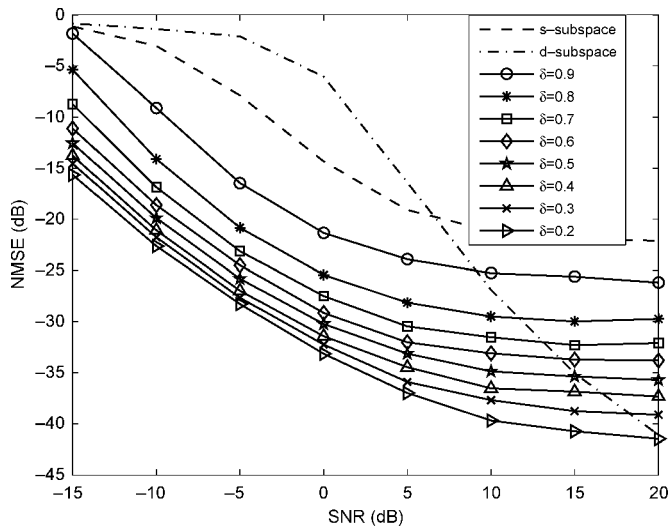Fig. 15.   NMSE at various overestimated channel orders.

Fig. 16.   NMSE at various SNR levels (random channel case).

can see from the figure, it is likely that no more than 1-dB power back-off is required as ρ increases from 1.8 to 10.3; a slight impact on the incurred instantaneous PAPR is also seen when different symbol constellations are used. As a result, the modulation-induced power spike does not seem to induce a substantial power back-off in practice.

Simulation 5—Robustness Against Channel Order Overestimation: This simulation compares the proposed method with the two subspace methods [20] and [28] when only an upper bound on the channel order is known. We consider channel A and 300 symbol blocks are generated. For the overestimated channel order $8 \leq \hat{L} \leq 15$, Fig. 15 shows the respective computed NMSE (the proposed method is implemented with $\delta = 0.8$) at two different levels of SNR (0 and 15 dB). The results show that, for a fixed SNR, all the three methods show good tolerance of channel order mismatch: the increase in NMSE is no more than 5 dB as $\hat{L}$ increases from 8 to 15. The proposed approach is seen to attain the best performance among the three when SNR is low and is quite insensitive to the level of noise contamination.
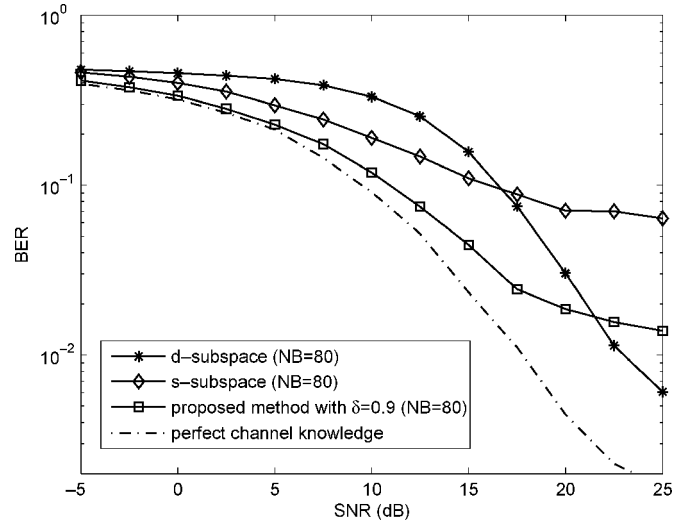

Fig. 17.   BER performances of three methods ($\mathrm{NB} = 80$; random channel case).

Simulation 6—Random Channel Case: The last simulation tests the proposed method in a random channel environment. The taps of the frequency-selective channel (with $L = 8$) are modeled as independent complex Gaussian random variables, each one being with zero mean and unit variance. In each Monte Carlo run (600 independent trails in total), a channel realization is drawn and is estimated using the proposed method and the two subspace methods [20] and [28] (500 symbol blocks are generated). Fig. 16 shows the computed NMSE; Fig. 17 shows the resultant equalization performance (the initial 80 received blocks are used for channel estimation). As we can see, the results exhibit essentially the same tendencies as in the fixed channel case (cf. Fig. 4 and Fig. 11); however, there is degradation in the BER performance as compared with Fig. 11.

VIII. CONCLUSION

In this paper, we propose a blind channel identification scheme for the SC-FDE systems based on periodic modulation. The presented study is the block transmission counterpart of the previous work [16]. The proposed method relies on the circulant channel matrix property unique to the SC-FDE systems. With the adopted correlation-matching-based formulation, the resulting set of identification equations is shown to exhibit a distinctive block circulant structure: this allows for a very simple identifiability condition as well as a natural approach to modulating sequence design against noise. The identifiability condition thus obtained inherits the appealing feature common to most TIC-based approaches: it entirely depends on the modulating sequence but not on the background channel characteristics. In fact, almost all sequences yield the channels identifiable. The problem of modulating sequence design against noise amounts to a constrained quadratic problem which yields a closed-form solution. The optimal solution is in a two-level form with single peak power; a suboptimal sequence for reducing the PAPR due to peak power modulation is introduced. Both solutions are shown to yield consistent channel estimates when the channel noise is white. There is a tradeoff between channel estimation accuracy and PAPR

limitation; in particular, a moderate reduction in PAPR can be achieved without incurring the loss in the resultant BER. Simulation results show that, in the low-to-medium SNR region or when the number of data samples is small, the proposed method outperforms existing subspace algorithms applicable to SC-FDE systems.

APPENDIX A
PROOF OF PROPOSITION 4.1

The following lemma characterizes the eigen-space of the matrix $\mathbf{Q}$ and will be used for proving Proposition 4.1. We define $\mathrm{BCCB}_{N,N}$ to be the set of all $N^2 \times N^2$ block circulant matrices with circulant blocks [5, p. 184], each characterized by N circulant matrices of dimension $N \times N$.

Lemma A.1 [5, p. 185]: If $\mathbf{X} \in \mathrm{BCCB}_{N,N}$, then $\mathbf{X}$ can be diagonalized by $\mathbf{F} \otimes \mathbf{F}$. More precisely, let $\{\mathbf{C}_0, \cdots, \mathbf{C}_{N-1}\}$ be the set of circulant matrices on the top row block characterizing $\mathbf{X}$, and let $\mathbf{\Lambda}_n$ be the $N \times N$ diagonal matrix containing the eigenvalues of $\mathbf{C}_n$. Then we have

$$\mathbf{X} = (\mathbf{F}^{-1} \otimes \mathbf{F}^{-1}) \left(\sum_{n=0}^{N-1} \mathbf{\Omega}_N^n \otimes \mathbf{\Lambda}_n \right) (\mathbf{F} \otimes \mathbf{F}) \qquad (A.1)$$

where $\mathbf{\Omega}_N := diag\{1, \omega, \omega^2, \cdots, \omega^{N-1}\}$. Conversely, any matrix of the form $(\mathbf{F}^{-1} \otimes \mathbf{F}^{-1})\mathbf{\Lambda}(\mathbf{F} \otimes \mathbf{F})$ for some diagonal $\mathbf{\Lambda}$ belongs to $\mathrm{BCCB}_{N,N}$. $\square$

Proof of Proposition 4.1: To show that $\mathbf{Q}$ is nonsingular, it is equivalent to verify that all its eigenvalues are nonzero. From (3.5), we can see that the matrix $\mathbf{Q}$ is characterized by the N circulant matrices $\{p(0)^2\mathbf{I}_N, p(N-1)^2\mathbf{J}^{N-1}, \cdots, p(1)^2\mathbf{J}\}$ in its top row block. For each $0 \le n \le N-1$, let $\mathbf{u}_n \in C^N$ be the vector containing the eigenvalues of the matrix $p(n)^2\mathbf{J}^n$. Then, from [5, p. 73], $\mathbf{u}_n$ can be expressed as $\mathbf{u}_n = \sqrt{N} \cdot \mathbf{F}^{-1}\mathbf{r}_n^T$, where $\mathbf{r}_n$ denotes the first row of $p(n)^2\mathbf{J}^n$. By definition of $\mathbf{J}$ (cf. (3.1)), we can see that all the entries of $\mathbf{r}_n$ are zero except that the $(n+1)$th entry is equal to $p(N-n)^2$ (note that, for $n=0$, $P(N)^2 = (0)^2$). This then implies $\mathbf{u}_n = \sqrt{N} \cdot p(N-n)^2 \mathbf{f}_n$, where

$$\mathbf{f}_n := \frac{1}{\sqrt{N}} \cdot \left[1 \ \omega^n \ \omega^{2n} \ \cdots \ \omega^{(N-2)n} \ \omega^{(N-1)n} \right]^T \in C^N \qquad (A.2)$$

denotes the $(n+1)$th column of $\mathbf{F}^{-1}$, $0 \le n \le N-1$. From (A.1), the eigenvalues of $\mathbf{Q}$ are given by the diagonal entries of the matrix $\sum_{n=0}^{N-1} \mathbf{\Omega}_N^n \otimes \mathbf{\Lambda}_n$, where $\mathbf{\Lambda}_n = diag\{\mathbf{u}_n\}$. Since $\mathbf{\Omega}_N^n = \sqrt{N} \cdot diag\{\mathbf{f}_n\}$ (this follows by definition of $\mathbf{\Omega}_N$ and from (A.2)), the eigenvalues of $\mathbf{Q}$ can simply be computed as entries of the vector $\sqrt{N} \cdot \sum_{n=0}^{N-1} \mathbf{f}_n \otimes \mathbf{u}_n$. From (A.2) and since $\mathbf{u}_n = \sqrt{N} \cdot p(N-n)^2 \mathbf{f}_n$, we have

$$\mathbf{f}_n \otimes \mathbf{u}_n = \left[p(N-n)^2\mathbf{f}_n^T \ \omega^n p(N-n)^2\mathbf{f}_n^T \right.$$
$$\left. \cdots \ \omega^{(N-1)n} p(N-n)^2\mathbf{f}_n^T \right]^T. \qquad (A.3)$$

From (A.3), it can be checked that

$$\sum_{n=0}^{N-1} \mathbf{f}_n \otimes \mathbf{u}_n = \begin{bmatrix} \sum_{n=0}^{N-1} p(N-n)^2\mathbf{f}_n \\ \sum_{n=0}^{N-1} \omega^n p(N-n)^2\mathbf{f}_n \\ \vdots \\ \sum_{n=0}^{N-1} \omega^{(N-1)n} p(N-n)^2\mathbf{f}_n \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{F}^{-1}\widetilde{\mathbf{J}}\mathbf{p} \\ \mathbf{F}^{-1}\mathbf{\Omega}_N\widetilde{\mathbf{J}}\mathbf{p} \\ \vdots \\ \mathbf{F}^{-1}\mathbf{\Omega}_N^{N-1}\widetilde{\mathbf{J}}\mathbf{p} \end{bmatrix} \qquad (A.4)$$

where $\mathbf{p}$ is defined in (3.8) and $\widetilde{\mathbf{J}}$ is a Hankel matrix with $[1 \ 0 \ \cdots \ 0]^T$ as the first column and $[0 \ 1 \ 0 \ \cdots \ 0]$ as the last row. It can be directly checked by definition that $\mathbf{F}^{-1}\mathbf{\Omega}_N^n = (\mathbf{J}^T)^n\mathbf{F}^{-1}$ and, as a result

$$\mathbf{F}^{-1}\mathbf{\Omega}_N^n\widetilde{\mathbf{J}}\mathbf{p} = (\mathbf{J}^T)^n\mathbf{F}^{-1}\widetilde{\mathbf{J}}\mathbf{p}, \quad 0 \le n \le N-1. \qquad (A.5)$$

From (A.4) and (A.5), we can see that the entries in the ith N-dimensional block of the vector $\sum_{n=0}^{N-1} \mathbf{f}_n \otimes \mathbf{u}_n$, $2 \le i \le N$, is simply a permutation of those in first one. As a result, the N^2 eigenvalues of $\mathbf{Q}$ thus assume N distinct values only. In particular, the matrix $\mathbf{Q}$ is nonsingular if and only if all the entries of $\mathbf{F}^{-1}\widetilde{\mathbf{J}}\mathbf{p}$ are nonzero. Since $\mathbf{J} = \mathbf{F}^2$ [5, p. 33], we have $\mathbf{F}^{-1}\widetilde{\mathbf{J}}\mathbf{p} = \mathbf{F}\mathbf{p}$ and the assertion follows. $\square$

APPENDIX B
PROOF OF PROPOSITION 5.3

We will fix $m = 0$; the results are exactly the same for $1 \le m \le N-1$. It suffices to show that, with the two-level solution (5.22), the noise perturbation term to the least-squares solution (5.3) becomes

$$\sigma_v^2(\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}})^{-1}\widetilde{\mathbf{Q}}^T\mathrm{vec}(\mathbf{I}_N) = \zeta\mathrm{vec}(\mathbf{I}_{L+1}), \text{ for some scalar } \zeta. \qquad (B.1)$$

With (5.3) and (B.1), we can see that the white noise contamination is merely to induce an "equal-strength" perturbation to $|h(n)|^2$ for $0 \le n \le L$, and hence (5.23) follows. From (5.14), it can be directly verified that

$$\widetilde{\mathbf{Q}}^T\mathrm{vec}(\mathbf{I}_N) = N \, \mathrm{vec}(\mathbf{I}_{L+1}). \qquad (B.2)$$

To prove (B.1), it thus remains to show that $\mathrm{vec}(\mathbf{I}_{L+1})$ is an eigenvector of the matrix $(\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}})^{-1}$. Since $(\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}})^{-1}$ is nonsingular, it suffices to show that $\mathrm{vec}(\mathbf{I}_{L+1})$ is an eigenvector of $\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}}$. This is indeed true by construction of the matrix $\widetilde{\mathbf{Q}}$. To see this, we first note that $\widetilde{\mathbf{Q}}\mathrm{vec}(\mathbf{I}_{L+1})$ is a linear combination of the $((l-1)(L+1)+l)$th columns, $1 \le l \le L+1$, of $\widetilde{\mathbf{Q}}$, each

of which being of the form given in (5.11) and (5.12). Based on this observation, it can be checked that

$$\widetilde{\mathbf{Q}}\text{vec}(\mathbf{I}_{L+1}) = \begin{bmatrix} f\mathbf{e}_1^T & \cdots & f\mathbf{e}_{L+1}^T & \hat{f}\mathbf{e}_{L+2}^T & \cdots & \hat{f}\mathbf{e}_N^T \end{bmatrix}^T \quad \text{(B.3)}$$

where $f = \alpha + L\beta$ and $\hat{f} = (L+1)\beta$. With (B.3) and by definition of $\widetilde{\mathbf{Q}}$, we then have

$$\widetilde{\mathbf{Q}}^T\widetilde{\mathbf{Q}}\text{vec}(\mathbf{I}_{L+1}) = \left\{(\alpha + L\beta)f + (N-L-1)\beta\hat{f}\right\}\text{vec}(\mathbf{I}_{L+1}) \tag{B.4}$$

and the result follows. □

REFERENCES

[1] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," in *Proc. IEEE*, vol. 85, Aug. 1997, pp. 1310–1332.

[2] H. Bolcskei, R. W. Heath, and A. J. Paulraj, "Blind channel estimation in spatial multiplexing systems using nonredundant antenna precoding," in *Proc. Asilomar Conf. Singlas, Systems, Computers*, vol. 2, Pacific Grove, CA, Oct. 1999, pp. 1127–1132.

[3] ——, "Blind channel identification and equalization in OFDM-based multiantenna systems," *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 96–109, Jan. 2000.

[4] 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.

[5] P. J. Davis, *Circulant Matrices*. New York: Wiley, 1979.

[6] 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.

[7] Z. Ding and Y. Li, *Blind Equalization and Identification*. New York: Marcel Dekker, 2001.

[8] D. Falconer, S. L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broad-band wireless systems," *IEEE Commun. Mag.*, pp. 58–66, Apr. 2002.

[9] G. B. Giannakis, "Filterbanks for blind channel identification and equalization," *IEEE Signal Process. Lett.*, vol. 4, no. 6, pp. 184–187, Jun. 1997.

[10] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, *Signal Processing Advances in Wireless and Mobil Communication Volume I: Trends in Channel Identification and Equalization*. Englewood Cliffs, NJ: Prentice-Hall PTR, 2001.

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

[12] A. Grace, *Optimization Toolbox, For Use With Matlab*. Natick, MA: The MathWorks, Inc., 1994.

[13] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[14] 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.

[15] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[16] 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.

[17] Z. Liu, Y. Xin, and G. B. Giannakis, "Linear constellation precoding for OFDM with maximal multipath diversity and coding gain," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 416–427, Mar. 2003.

[18] L. Ljung, *System Identification, Theory for the Users*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[19] 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.

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

[21] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[22] A. Scaglione, B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers. Part I: Unification and optimal design. Part II: Blind channel estimation, synchronization, and direct equalization," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1988–2022, Jul. 1999.

[23] 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.

[24] M. Tsatsanis and G. B. Giannakis, "Transmitter induced cyclostationarity for blind channel equalization," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1785–1794, Jul. 1997.

[25] Z. Wang, X. Ma, and G. B. Giannakis, "OFDM or single-carrier block transmission?," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 380–394, Mar. 2004.

[26] 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.

[27] Q. Zhao and T. Long, "Adaptive blind channel estimation by least squares smoothing," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 3000–3012, Nov. 1999.

[28] X. Zhuang, Z. Ding, and A. L. Swindlehurst, "A statistical subspace method for blind channel identification in OFDM communications," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 5, 2000, pp. 2493–2496.

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

He is currently a Postdoctoral Researcher in the Department of Communication Engineering, National Chiao Tung University. His current research interests are in signal processing and information theory.

Ta-Sung Lee (M'94–SM'05) was born in Taipei, Taiwan, R.O.C., in 1960. He received the B.S. degree form National Taiwan University, Taiwan, R.O.C., in 1983, the M.S. degree from the University of Wisconsin, Madison, in 1987, and the Ph.D. degree form 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 in the Department of Communication Engineering. His other positions include Technical Advisor at Computer 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 EECS, NCTU; Director of the Taiwan Internet Education Association; and Managing Director of Communications and Computer Training Program, NCTU. He is active in the research and development of advanced techniques for wireless communications, such as smart antenna and multiple-input multiple-output (MIMO) technologies, cross-layer design, and software-defined radio (SDR) prototyping of advanced communication systems. He has been a co-leader 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; these include the 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.