Redundant Filterbank Precoders and Equalizers Part II: Blind Channel Estimation, Synchronization, and Direct Equalization

Anna Scaglione, *Student Member, IEEE,* Georgios B. Giannakis, *Fellow, IEEE,* and Sergio Barbarossa, *Member, IEEE*

***Abstract*—Transmitter redundancy introduced using finite impulse response (FIR) filterbank precoders offers a unifying framework for single- and multiuser transmissions. With minimal rate reduction, FIR filterbank transmitters with trailing zeros allow for perfect (in the absence of noise) equalization of FIR channels with FIR zero-forcing equalizer filterbanks, *irrespective* of the input color and the channel zero locations. Exploiting this simple form of redundancy, blind channel estimators, block synchronizers, and direct self-recovering equalizing filterbanks are derived in this paper. The resulting algorithms are computationally simple, require small data sizes, can be implemented online, and remain consistent (after appropriate modifications), even at low SNR colored noise. Simulations illustrate applications to blind equalization of downlink CDMA transmissions, multicarrier modulations through channels with deep fades, and superior performance relative to CMA and existing *output diversity* techniques relying on multiple antennas and fractional sampling.**

***Index Terms*— Blind channel estimation, block transmissions, filterbanks, intersymbol interference, minimum mean-square error communication receivers, nondata aided synchronization, precoding, self-recovering equalization, zero forcing.**

I. Introduction

REDUNDANCY at the transmitter builds *input diversity* in digital communication systems and is well motivated for designing error correcting codes (e.g., [2]). Recently, however, input diversity has been exploited also for ISI suppression using precoders operating in the complex (as opposed to Galois) field, [3], [14], [19], [24], [26], [27], [32]. Different precoding schemes are possible. Multiplying the input by a known periodic sequence offers a precoding scheme that does not require any increase of the transmission rate, although the constellation's modulus is affected and equalization of FIR channels with FIR equalizers is impossible [3], [26]. On the other hand, repeating input symbols as in [27] leads to FIR equalizers but reduces information rate by half. Combining desired features, filterbank precoders and blind equalizers were proposed in [14] to minimize the rate reduction and obviate channel zero restrictions imposed by spatio-temporal *output diversity* methods that rely on fractional sampling (FS) and/or multiple-antenna reception [13], [15], [22], [28], [33]. However, identifiability in [14] was established only for white inputs and linear equation or subspace algorithms were developed for simple precoders (see also [19]). In a deterministic multirate framework, filterbanks for *nonblind* channel equalization were also proposed in [32] under restrictions on the channel zeros.

In the companion paper [24], we have shown that redundant filterbank precoders offer a unifying discrete-time model that encompasses a wide range of digital modulation and coding schemes [24]. Those include periodic and line codes, orthogonal frequency-division multiplexing (OFDM) and discrete multitone (DMT) [5], [30], fractional sampling [28], (de-)interleaving, as well as multiuser transmissions such as TDMA, FDMA, CDMA, and the most recent discrete wavelet multiple access (DWMA) schemes [23] (see also [1] and [31]). However, self-recovering (or *blind*) approaches to channel estimation, synchronization, and equalization were not addressed in [24]. Redundant precoding brings input diversity similar to that available with training sequences that have been exploited recently in a semi-blind channel estimation framework [18].

Motivated by the generality and importance of filterbank precoders, this paper builds on [14] and [24] and proposes novel *deterministic* methods for blind channel estimation, block synchronization, and direct equalization without imposing restrictions on channel zeros (Section III). The deterministic solution, as opposed to statistical methods, is particularly appealing for transmissions over slowly varying channels. The methods are also applicable to random inputs, white or colored (important for coded transmissions). General precoders equipped with trailing zeros (TZ) offer computationally simple algorithms and lead to zero-forcing (ZF) or minimum mean-square error (MMSE) equalizing filterbanks that accept also adaptive implementations. Exciting options become available for blind equalization in OFDM, which is a transmission scheme known to suffer from deep channel fades (see also [6]). The importance of increasing the robustness of OFDM against frequency selective fading is testified by its applications: OFDM is currently used in digital audio broadcasting (DAB) [8], and it has been selected for European digital terrestrial

Manuscript received January 6, 1998; revised January 18, 1999. This work was supported by the National Science Foundation under Grant CCR-9805350. It was presented in part at Allerton'97 and ICASSP'98. The associate editor coordinating the review of this paper and approving it for publication was Dr. Truong Q. Nguyen.

A. Scaglione and S. Barbarossa are with the Infocom Department, University of Rome "La Sapienza," Roma, Italy (e-mail: sergio@infocom.ing.uniroma1.it; annas@infocom.ing.uniroma1.it).

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

Publisher Item Identifier S 1053-587X(99)04656-5.

television broadcasting (DTTB) [9], and, in its DMT version [4], for the asymmetric digital subscriber loop (ASDL) in the United States. Our methods are also directly applicable in TDMA and CDMA systems for the blind equalization of the downlink channel. Consistency analysis and modifications needed at low SNR are given in Section IV. In Section V, the proposed methods are tested via simulations and compared with alternative methods that employ space diversity [13], [15], [22], [33], fractional sampling (FS) [28], or they exploit *a priori* knowledge of the symbol constellation, e.g., the constant modulus algorithm (CMA) [16].

II. Preliminaries

We will consider the same discrete-time multirate transmitter model presented in [24] for the baseband communication system reported in Fig. 1. We will also assume Nyquist signaling pulses. Downsamplers and upsamplers perform blocking (i.e., multiplexing) and unblocking (demultiplexing) operations. With $P > M$, the ratio $(P-M)/P$ represents the amount of redundancy introduced. The input to the upsampler of the mth branch is $s_m(n) := s(nM+m)$. It represents the mth symbol in the nth block of M symbols, whereas in the multiuser case, it stands for the mth user's bits. We will use the same notation as in [24], denoting the transmit data sequence [summation of all the sequences $u_m(n)$ at the output of each branch] as $u(n)$, the noise-free channel output sequence as $x(n)$, and the noisy received signal as $y(n) := x(n)+v(n)$, where $v(n)$ is additive noise (independent of the transmitted symbols) characterized by the covariance matrix $\boldsymbol{R}_{vv}$ and, finally, by $\hat{s}_m(n)$ the equalizing filterbank output. The block data model in [24] expresses the input–output relationship in matrix form. Let us define the $M \times 1$ vectors $\boldsymbol{s}(n)$ and $\hat{\boldsymbol{s}}(n)$ that are defined as $\boldsymbol{s}(n)$ in

$$\boldsymbol{s}(n) := (s_0(n), s_1(n), \cdots, s_{M-1}(n))^T. \tag{1}$$

The $P \times 1$ vectors $\boldsymbol{u}(n)$, $\boldsymbol{x}(n)$, $\boldsymbol{v}(n)$, and $\boldsymbol{y}(n)$, are all defined as $\boldsymbol{u}(n)$

$$\boldsymbol{u}(n) := (u(nP), u(nP+1), \cdots, u(nP+P-1))^T \tag{2}$$

where the channel vector $\boldsymbol{h} := (h(0), \cdots, h(L))^T$ and $P \times M$ precoder and $M \times P$ equalizing matrices $\boldsymbol{F}_i$ and $\boldsymbol{G}_j$, whose elements are, respectively, $\{\boldsymbol{F}_i\}_{p,m} := f_m(iP+p)$ $\{\boldsymbol{G}_j\}_{m,p} := g_p(jM+m)$, for $p = 0, \cdots, P-1$ and $m = 0, \cdots, M-1$, i.e., the columns of the ith (jth) matrix $\boldsymbol{F}_i(\boldsymbol{G}_j)$, contain the ith (jth) segment of length P (M) of the filters' impulse responses $\{f_m(n)\}_{m=0}^{M-1}(\{g_p(n)\}_{p=0}^{P-1})$. Introducing the channel matrices $\{\boldsymbol{H}_l\}_{k,p} = h(lP+k-p)$as well, for $k, p = 0, \cdots, P-1$, we can express (see [24] for details) the transmitted block-data sequence as $\boldsymbol{u}(n) = \sum_i \boldsymbol{F}_i \boldsymbol{s}(n-i)$, the noise-free blocked channel output as $\boldsymbol{x}(n) = \sum_{l=-\infty}^{\infty} \boldsymbol{H}_l \boldsymbol{u}(n-l)$, and the equalized block sequence as

$$\hat{\boldsymbol{s}}(n) = \sum_{j,l,i=-\infty}^{\infty} \boldsymbol{G}_j \boldsymbol{H}_l \boldsymbol{F}_i \boldsymbol{s}(n-l-i-j) + \sum_{j=-\infty}^{\infty} \boldsymbol{G}_j \boldsymbol{v}(n-j). \tag{3}$$

We assume, as in [24], the following:

a0) Channel $h(l)$ is Lth-order FIR with $h(0)$, $h(L) \neq 0$.

a1) For a given L, the pair (P, M) is chosen to satisfy $P > L$ and $P = M + L$.

In [24, Th. 1], extra conditions on the precoding strategy that guarantee the existence of FIR linear block equalizers that are *independent* of the channel zero locations are established. The condition $P > L$, according to the definition of $\boldsymbol{H}_l$, implies that $\boldsymbol{H}_l = \boldsymbol{H}_0\delta(l) + \boldsymbol{H}_1\delta(l-1)$ or, in words, that the interblock interference (IBI) is limited only to subsequent blocks. Existence of FIR linear equalizers requires, in any case, redundant precoding that is obtained setting $P > M$, while the complexity of the precoding strategy and equalization depends on the ratio $(P-M)/L$. In this paper, we consider the specific class of precoders that allows exact FIR-ZF equalization, fulfilling the requirements of [24, Th, 2]. Specifically, we assume the following.

a2) Precoder filters have L trailing zeros (TZ), i.e., $\{f_m(n)\}_{n=M}^{P} = 0$, $\forall m \in [0, M-1]$, and are linearly independent, i.e., rank$(\boldsymbol{F}) = M$, which guarantees one-to-one mapping and, thus, recovery of $s(n)$ from the coded symbols $u(n)$.

In force of a2), the precoder $\boldsymbol{F}_i = \boldsymbol{F}_0\delta(i)$ is of order zero, implying that $\boldsymbol{u}(n) = \boldsymbol{F}_0\boldsymbol{s}(n)$, and assumes the form

$$\boldsymbol{F}_0^T := (\boldsymbol{F}^T \boldsymbol{0}^T) \tag{4}$$

where $\boldsymbol{F}$ is a full-rank $M \times M$ matrix with elements $\{\boldsymbol{F}\}_{p,m} = f_m(p)$ for $m, p = 0, \cdots, M-1$, whereas $\boldsymbol{0}$ indicates[1] an $L \times M$ matrix of zeros that creates the guard interval of L trailing zeros. The insertion of a null guard time interval is not new in digital communications. DVB systems, for example, send a null guard time interval at the beginning of each frame, for synchronization purposes, whose duration varies from 168 to 1297 μs [9]. Exact knowledge of L is not required as long as M is chosen to satisfy $P > \overline{L}$, where $\overline{L} \geq L$ is a known upper bound on the channel order. Taking into account the FIR nature of both $\{h(l)\}_{l=0}^{L}$ and $\{f_m(n)\}_{n=0}^{M-1}$, it is not difficult to verify that TZ prevent IBI, i.e., $\boldsymbol{H}_1\boldsymbol{F}_0 = \boldsymbol{0}$ (see also [24]); thus, assumptions a0)–a2) imply that $\boldsymbol{x}(n) = \boldsymbol{H}_0\boldsymbol{F}_0\boldsymbol{s}(n)$, where, according to the definition of $\boldsymbol{H}_l$, matrix $\boldsymbol{H}_0$ is

$$\boldsymbol{H}_0 := \begin{pmatrix} h(0) & 0 & 0 & \cdots & 0 \\ \vdots & h(0) & 0 & \cdots & 0 \\ h(L) & \cdots & \ddots & \cdots & \vdots \\ \vdots & \ddots & \cdots & \ddots & 0 \\ 0 & \cdots & h(L) & \cdots & h(0) \end{pmatrix}_{P \times P} \tag{5}$$

and is $P \times P$ Toeplitz lower triangular matrix with first column $(h(0) \cdots h(L) 0 \cdots 0)^T$ and first row $(h(0) 0 \cdots 0)^T$.

[1] In general, we will use the symbol $\boldsymbol{0}$ to indicate null vectors and matrices without specifying their dimensions unless they are not obvious.

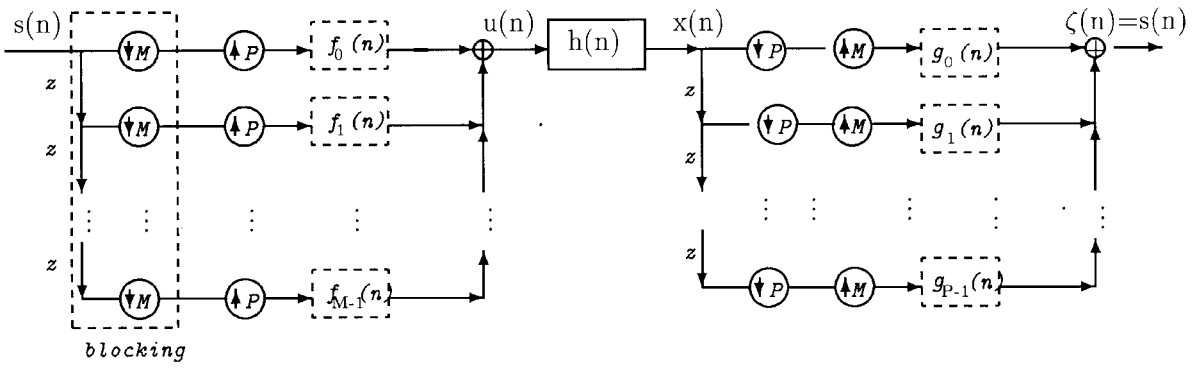


Fig. 1. Multirate discrete-time transmitter/channel/receiver model.

Denoting by $\boldsymbol{H}$ the $P \times M$ Toeplitz matrix formed by the first M columns of $\boldsymbol{H}_0$

$$\boldsymbol{H} := \begin{pmatrix} h(0) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ h(L) & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & h(0) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h(L) \end{pmatrix}_{P \times M} \tag{6}$$

we have that $\boldsymbol{H}_0\boldsymbol{F}_0 = \boldsymbol{H}\boldsymbol{F}$; therefore

$$\boldsymbol{x}(n) = \boldsymbol{H}\boldsymbol{F}\boldsymbol{s}(n) \tag{7}$$

and the received block data model is [24]

$$\boldsymbol{y}(n) = \boldsymbol{x}(n) + \boldsymbol{v}(n) = \boldsymbol{H}\boldsymbol{F}\boldsymbol{s}(n) + \boldsymbol{v}(n). \tag{8}$$

Henceforth, we will focus on the block-by-block transmission and reception framework described by the model in (8). Equation (8) shows that in the absence of noise, there is also a one-to-one mapping between input and output blocks. Thanks to a1) and a2), consecutive data blocks $\boldsymbol{s}(n)$ and $\boldsymbol{s}(n-1)$ do not interfere with each other. Hence, $\boldsymbol{y}(n)$ is a sufficient statistic with respect to $\boldsymbol{s}(n)$, and then, the linear equalizer can be without loss of generality (w.l.o.g.) of order zero, i.e., $\boldsymbol{G}_j = \boldsymbol{G}_0\delta(j)$. Indicating the $M \times P$ equalizing matrix as $\boldsymbol{G} := \boldsymbol{G}_0$, (3) becomes

$$\hat{\boldsymbol{s}}(n) = \boldsymbol{G}\boldsymbol{H}\boldsymbol{F}\boldsymbol{s}(n) + \boldsymbol{G}\boldsymbol{v}(n). \tag{9}$$

We report the statement of [24, Th. 2].

Theorem—TZ Precoders: Under a0)–a2), there exists a zero order $M \times P$ equalizing filterbank such that $\boldsymbol{G}\boldsymbol{x}(n) = \boldsymbol{s}(n)$. Moreover, the minimum-norm ZF filterbank is unique and is given by

$$\boldsymbol{G}_{zf} = \boldsymbol{F}^{-1}\boldsymbol{H}^{\dagger} = \boldsymbol{F}^{-1}(\boldsymbol{H}^H\boldsymbol{H})^{-1}\boldsymbol{H}^H \tag{10}$$

where $\dagger$ denotes pseudo inverse.

Later on, we will make use of the following persistence-of-excitation (p.e.) assumption on the transmitted symbols $s(n)$:

a3) There exists an $N \geq P$, such that the $M \times N$ matrix $\boldsymbol{S}_N := (\boldsymbol{s}(0) \cdots \boldsymbol{s}(N-1))$ has full rank M. Note that as $N \to \infty$, $(1/N)\boldsymbol{S}_N\boldsymbol{S}_N^H$ tends to the input correlation matrix $\boldsymbol{R}_{ss}$. In general, a3) is satisfied even for colored (e.g., coded) inputs provided that their spectra are nonzero for at least M frequencies (modes). Recall that p.e. assumptions like a3) are needed even with nonblind identification problems.

Given blocks of data $\{\boldsymbol{y}(n)\}_{n=0}^{N-1}$ only, and based on a0)–a3), the *objective* of this paper is threefold:

i) identify the channel $\{h(l)\}_{l=0}^{L}$;
ii) estimate directly the equalizer matrix $\boldsymbol{G}$ in (9);
iii) derive blind synchronizers to make block-coherent receiver processing possible.

Our goal for blind channel estimation, besides the evident purpose of avoiding the periodic transmission of bandwidth consuming training sequences, is also instrumental for i) sending channel status information (CSI) back to the transmitter, whenever a feedback channel is available, to optimize the transmission strategy, as in [24], or ii) for deriving zero-forcing (ZF), decision-feedback (DF), or minimum-mean square error (MMSE) equalizers that also rely on CSI. Even with moderate number of filters M in the precoder, the maximum likelihood (ML) receiver implemented with Viterbi's algorithm has prohibitively large complexity, which motivates the search for linear (and preferably low order FIR) equalizing filterbanks. ZF solutions offer (almost) perfect symbol recovery in (high SNR) noise-free environments, and their performance in terms of error probability is easily computable. At low SNR, a vector MMSE (or Wiener) equalizer is better motivated than the ZF solution (10) and can be derived by minimizing $J(\boldsymbol{G}) := E\{\text{tr}[\boldsymbol{G}\boldsymbol{y}(n) - \boldsymbol{s}(n)][\boldsymbol{G}\boldsymbol{y}(n) - \boldsymbol{s}(n)]^H\}$. The MMSE solution is obtained by equating to zero the gradient of $J(\boldsymbol{G})$ with respect to $\boldsymbol{G}$ (see [24]) and is

$$\boldsymbol{G}_{\text{mmse}} = \boldsymbol{R}_{ss}\boldsymbol{F}^H\boldsymbol{H}^H(\boldsymbol{R}_{vv} + \boldsymbol{H}\boldsymbol{F}\boldsymbol{R}_{ss}\boldsymbol{F}^H\boldsymbol{H}^H)^{-1}. \tag{11}$$

In the absence of noise, $\nabla_{\mathbf{G}} J(\boldsymbol{G}) = (\boldsymbol{G}\boldsymbol{H}\boldsymbol{F} - \boldsymbol{I})\boldsymbol{R}_{ss}\boldsymbol{F}^H\boldsymbol{H}^H = \boldsymbol{0}$, which is satisfied if and only if $\boldsymbol{G} = (\boldsymbol{H}\boldsymbol{F})^{\dagger}$ and therefore, [c.f. (10)] $\boldsymbol{G} \equiv \boldsymbol{F}^{-1}H^{\dagger} := \boldsymbol{G}_{zf}$. Observe, however, that $\boldsymbol{G}_{\text{mmse}}$ in (11) does not tend to $\boldsymbol{G}_{zf}$ in (10), as $\boldsymbol{R}_{vv} \to \boldsymbol{0}$, and matrix $(\boldsymbol{R}_{vv} + \boldsymbol{H}\boldsymbol{F}\boldsymbol{R}_{ss}\boldsymbol{F}^H\boldsymbol{H}^H)$ tends to become singular because the $P \times P$ matrix $\boldsymbol{H}\boldsymbol{F}\boldsymbol{R}_{ss}\boldsymbol{F}^H\boldsymbol{H}^H$ has rank M.

Note that [24, Th. 2] poses no constraints on the channel zeros. In contrast, FIR-ZF equalizers in [13], [15], [22], [27], [28], [32], [33] do not exist for certain configurations of channel zeros on the unit circle, and more important, their

performance degrades even when channels have zeros close to those noninvertible configurations. If an upper bound $\overline{L} \geq L$ is only available on the channel order, Theorem 2 holds true with $\overline{L}$ replacing L in a1). Because, for a fixed bandwidth, the information rate depends only on the ratio M/P, we underscore that under a1), the rate reduction can be made arbitrarily small by selecting M (and thus P) sufficiently large. Of course, this comes at the cost of increased decoding delay and greater complexity so that bandwidth efficiency is generally traded with computational complexity and maximum allowable delay.

We wish to emphasize that the computation of the equalizing matrices $\boldsymbol{G}_{zf}$ and $\boldsymbol{G}_{\text{mmse}}$ from (10) and (11) requires knowledge, or estimates, of the channel. In the ensuing sections, we will derive self-recovering channel estimation and direct equalization methods that take advantage of the redundancy introduced at the transmitter by the insertion of trailing zeros. The latter can be interpreted as a form of *distributed training*. Since both data and trailing zeros are exploited for channel identification and direct equalization, the proposed approaches can be classified as semi-blind (see also [18]). On the other hand, our methods rely on the received data only, and from this point of view, our receivers are self-recovering (or blind).

III. Blind Symbol Recovery

Blind channel estimation is well motivated for wireless environments where the multipath channel changes rapidly as the mobile communicators move. Self-recovering equalization schemes are important to avoid frequent retraining and thus increase bandwidth efficiency. In our block-transmission schemes, both channel estimation (Section III-A) and direct equalization (Section III-B) assume knowledge of the beginning of each block, which is a subject we also address later on (Section III-C).

A. Blind Channel Estimation

Under a2), $\boldsymbol{x}(n) = \boldsymbol{HFs}(n)$ and collecting N data vectors $\{\boldsymbol{x}(n)\}_{n=0}^{N-1}$ in a matrix, we arrive at

$$\boldsymbol{X}_N := (\boldsymbol{x}(0) \cdots \boldsymbol{x}(N-1)) = \boldsymbol{HFS}_N \tag{12}$$

where $\boldsymbol{S}_N$ is defined as in a3). From (6), it is sufficient that one channel coefficient is different from zero to ensure that $\text{rank}(\boldsymbol{H}) = M$, and this, along with a2) and a3), implies that $\text{rank}(\boldsymbol{X}_N) = M$. Therefore, the nullity of the matrix $\boldsymbol{X}_N\boldsymbol{X}_N^H$ is $\nu(\boldsymbol{X}_N\boldsymbol{X}_N^H) = P - M = L$, and the eigendecomposition

$$\boldsymbol{X}_N\boldsymbol{X}_N^H = (\overline{\boldsymbol{U}}\ \tilde{\boldsymbol{U}})\begin{pmatrix}\boldsymbol{\Sigma}_{M\times M} & \boldsymbol{0}_{M\times L}\\ \boldsymbol{0}_{L\times M} & \boldsymbol{0}_{L\times L}\end{pmatrix}\begin{pmatrix}\overline{\boldsymbol{U}}^H\\ \tilde{\boldsymbol{U}}^H\end{pmatrix} \tag{13}$$

yields the $P \times L$ matrix $\tilde{\boldsymbol{U}}$, whose columns span the null space $\mathcal{N}(\boldsymbol{X}_N)$. Because $\boldsymbol{FS}_N$ in (12) is full rank, $\mathcal{R}(\boldsymbol{X}_N) = \mathcal{R}(\boldsymbol{H}_0)$, where $\mathcal{R}$ stands for range space. However, since $\mathcal{R}(\boldsymbol{X}_N)$ is orthogonal to $\mathcal{N}(\boldsymbol{X}_N)$, it follows that

$$\tilde{\boldsymbol{U}}^H\boldsymbol{H} = \boldsymbol{0} \Rightarrow \tilde{\boldsymbol{u}}_l^H\boldsymbol{\mathcal{T}}(\boldsymbol{h}) = \boldsymbol{0}^H, \qquad l = 1, \cdots, L \tag{14}$$

where $\tilde{\boldsymbol{u}}_l$ denotes the lth column of $\tilde{\boldsymbol{U}}$, and $\boldsymbol{H} = \boldsymbol{\mathcal{T}}(\boldsymbol{h})$ is the Toeplitz matrix in (6) denoting convolution. Because convolution is commutative, (14) can be written as

$$\boldsymbol{h}^H\boldsymbol{\mathcal{U}} := \boldsymbol{h}^H(\boldsymbol{\mathcal{U}}_1 \cdots \boldsymbol{\mathcal{U}}_L) = \boldsymbol{0}^H \tag{15}$$

where each $\boldsymbol{\mathcal{U}}_l$ is an $(L+1) \times M$ Hankel matrix formed by $\tilde{\boldsymbol{u}}_l$ as

$$\boldsymbol{\mathcal{U}}_l = \begin{pmatrix}\tilde{u}_l(0) & \tilde{u}_l(1) & \cdots & \tilde{u}_l(P-L-1)\\ \tilde{u}_l(1) & \tilde{u}_l(2) & \cdots & \tilde{u}_l(P-L)\\ \vdots & \vdots & \vdots & \vdots\\ \tilde{u}_l(L) & \tilde{u}_l(L+1) & \cdots & \tilde{u}_l(P-1)\end{pmatrix}. \tag{16}$$

Our result and the corresponding algorithm rely on (15) and are summarized in the following.

Theorem 1: Let a0)–a3) hold true. Starting from the data matrix $\boldsymbol{X}_N$, we form the $(L+1) \times ML$ matrix $\boldsymbol{\mathcal{U}}$ as in (12)–(15). Channel vector $\boldsymbol{h}$ can then be obtained as the unique (within a scale factor) null eigenvector of $\boldsymbol{\mathcal{U}}$ in (15).

Proof: We will prove channel identifiability by contradiction. Let us assume that (15) admits two different solutions $\underline{\boldsymbol{h}} \neq \boldsymbol{0}$ and $\boldsymbol{h} \neq \boldsymbol{0}$, where $\boldsymbol{h}$ represents the true channel. Hence

$$\tilde{\boldsymbol{U}}^H\boldsymbol{H} = \tilde{\boldsymbol{U}}^H\boldsymbol{\mathcal{T}}(\underline{\boldsymbol{h}}) = \boldsymbol{0}. \tag{17}$$

It is easy to verify that the left null space of the convolution matrix $\boldsymbol{\mathcal{T}}(\boldsymbol{h})$, which is a Toeplitz matrix [see (6)], is spanned by L Vandermonde vectors of the form $\boldsymbol{\rho}_l = (1, \rho_l, \cdots, \rho_l^{P-1})^T$, where ρ_l, $l = 1, \cdots, L$ are the L distinct[2] roots of the channel transfer function $H(z) = \mathcal{Z}\{h(n)\}$. Therefore, there exists an $(L+1) \times (L+1)$ full-rank matrix $\boldsymbol{A}$ such that $\tilde{\boldsymbol{U}} = \boldsymbol{PA}$, where $\boldsymbol{P}$ is a $P \times L$ matrix formed with the L Vandermonde vectors $\boldsymbol{P} = (\boldsymbol{\rho}_1 \cdots \boldsymbol{\rho}_L)$. If the vectors $\underline{\boldsymbol{h}}$ and $\boldsymbol{h}$, with $\underline{\boldsymbol{h}} \neq \boldsymbol{h}$, must satisfy (17), then

$$\tilde{\boldsymbol{U}}^H\boldsymbol{\mathcal{T}}(\underline{\boldsymbol{h}}) = \tilde{\boldsymbol{U}}^H\boldsymbol{\mathcal{T}}(\boldsymbol{h}) = \boldsymbol{0}. \tag{18}$$

Using $\tilde{\boldsymbol{U}} = \boldsymbol{PA}$, we can rewrite (18) as $\boldsymbol{A}^H\boldsymbol{P}^H\boldsymbol{\mathcal{T}}(\underline{\boldsymbol{h}}) = \boldsymbol{A}^H\boldsymbol{P}^H\boldsymbol{\mathcal{T}}(\boldsymbol{h}) = \boldsymbol{0}$, or since $\boldsymbol{A}$ is full rank

$$\boldsymbol{P}^H\boldsymbol{\mathcal{T}}(\underline{\boldsymbol{h}}) = \boldsymbol{P}^H\boldsymbol{\mathcal{T}}(\boldsymbol{h}) = \boldsymbol{0}. \tag{19}$$

Equation (19) implies that $\underline{\boldsymbol{h}} = \alpha\boldsymbol{h}$ (where α is a complex constant) as a consequence of the fact that only impulse responses having the same roots (and with the same multiplicity) share the same Vandermonde (or generalized Vandermonde) basis for the null space of $\boldsymbol{\mathcal{T}}(\boldsymbol{h})$. In fact, if two polynomials having the same order share all the roots, their coefficients have to be proportional by a constant factor. This proves that the channel can be identified up to a scalar factor as in all blind channel identification methods. ■

The result and proof of Theorem 1 is similar in spirit with [22]. However, as we mentioned earlier, our method entails no assumptions on the FIR channel zeros, and generalizes the special TDMA-like result of [19] to arbitrary precoders satisfying a2) and with respect to [19] it is deterministic. Another distinct feature is our method's behavior under channel order mismatch. If the channel order is underestimated, similar to existing algorithms, our algorithm is not expected to work. However, in contrast to most output diversity methods,

[2]If $H(z)$ has multiple roots, it is possible to extend appropriately the definition using the so-called generalized Vandermonde vectors (see e.g., [21]).

we also verify in the simulations section that our method is interestingly *not affected by channel order overestimation*. Since in most applications the maximum delay is known, an upper bound on the maximum channel order is readily available. Furthermore, the method is robust with respect to Doppler shifts modeled as a multiplication of the transmit sequence by $\exp(j\omega_d n)$ in contrast with the standard channel estimation techniques of OFDM systems, based on pilot tones.

With the channel matrix $\boldsymbol{H}$ obtained from (15), we can proceed to determine either the ZF equalizer filterbank from (10) or the MMSE equalizer from (11). In fact, it is possible to derive MMSE equalizers involving a delayed decision (with or without feedback) or pursue the computationally complex but optimal ML receiver. Our channel estimation algorithm is summarized as follows.

1) With N blocks of data, form the matrix $\boldsymbol{X}_N := (\boldsymbol{x}(n) \cdots \boldsymbol{x}(n+N-1))$ as in (12).
2) Determine the L eigenvectors $\tilde{\boldsymbol{U}}_l$, $l = 1, \cdots, L$, corresponding to the null eigenvalues of the matrix $\boldsymbol{X}_N \boldsymbol{X}_N^H$ (to be replaced by the smallest eigenvalues for the case of noisy observations).
3) Estimate the channel vector $\boldsymbol{h}$ as the nontrivial solution of the system of linear equations in (15).
4) Use the channel estimate $\hat{\boldsymbol{h}}$ to form the matrix $\boldsymbol{\mathcal{T}}(\hat{\boldsymbol{h}})$ in (6).
5) Equalize the data with $\hat{\boldsymbol{G}}_{zf} = \boldsymbol{F}^{-1}(\boldsymbol{\mathcal{T}}(\hat{\boldsymbol{h}}))^{\dagger}$.

We focus next on direct blind equalizers that do not even require channel estimation as a first step, and being linear, they lend themselves naturally to online self-recovering algorithms. Recall, however, that channel estimation is indispensable for designing the optimal precoders [24].

B. Direct Blind Equalization

From (7), according to [24, Th. 2], the ZF-block equalizer is well defined because the system of linear equations is *always* invertible, and the solution that minimizes the error norm is given by $\boldsymbol{G} = \boldsymbol{H}^{\dagger}$, where $\dagger$ denotes matrix pseudoinverse. Intuitively, even deep fades can be equalized because the presence of guard bits allows us to equalize the channel by solving an *overdetermined* system of linear equations. In fact, each block of M input symbols is mapped to a block of $P = M + L$ data. Even if the *best* ZF equalizer, in terms of numerical stability of the result and noise enhancement, is given by $\boldsymbol{G} = \boldsymbol{H}^{\dagger}$, infinite many matrices solve the same overdetermined linear system of equations in (7), and they are all potential equalizers. In the following, we will show that some of these equalizers are self recovering, in the sense that they can be built directly form the received data without requiring channel estimation.

Collect $\{\boldsymbol{x}(n)\}_{n=0}^{N-1}$ blocks to form the data matrix $\boldsymbol{X}_N$ as in (12), and define $\boldsymbol{\Gamma} := \boldsymbol{H}_0^{-1}$ to arrive at

$$\boldsymbol{\Gamma}\boldsymbol{X}_N = \begin{pmatrix} \boldsymbol{F}\boldsymbol{S}_N \\ \boldsymbol{0}_{L\times N} \end{pmatrix}. \tag{20}$$

Hence, one possible equalizer $\boldsymbol{G} \neq \boldsymbol{G}_{zf}$ is $\boldsymbol{G} = (\boldsymbol{F}^{-1} \;\; \boldsymbol{0})\boldsymbol{\Gamma}$. Because the $P \times P$ matrix $\boldsymbol{H}_0$ is lower triangular Toeplitz, it follows easily (by forming $\boldsymbol{\Gamma}\boldsymbol{H}_0 = \boldsymbol{I}$) that the inverse $\boldsymbol{\Gamma}$ is also lower triangular Toeplitz. Thus, all the rows $\{\boldsymbol{\gamma}_i^T\}_{i=1}^P$ of $\boldsymbol{\Gamma}$ can be obtained from the last row $\boldsymbol{\gamma}_P^T$. Relying on (20), we will show how $\boldsymbol{\gamma}_P^T$ can be determined directly from the received data matrix $\boldsymbol{X}_N$ in (12).

Let $\boldsymbol{J}$ denote a square shift matrix

$$\boldsymbol{J} := \begin{pmatrix} 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \dots & \dots & 0 \end{pmatrix}. \tag{21}$$

Since $\boldsymbol{J}$ is not invertible, to simplify the derivations, we will use the symbol $\boldsymbol{J}^{-1}$ to denote

$$\boldsymbol{J}^{-1} := \boldsymbol{J}^{\dagger} \equiv \boldsymbol{J}^T \tag{22}$$

where superscript T is transpose. Matrix $\boldsymbol{J}^{-1}$, in analogy with the delay factor z^{-1} of the $\mathcal{Z}$-transforms, is a delay matrix, whereas $\boldsymbol{J}$ is the advance matrix similar to z.

Due to its lower triangular Toeplitz structure, using a $P \times P$ delay matrix $\boldsymbol{J}^{-1}$, the rows of $\boldsymbol{\Gamma}$ can be successively related as $\boldsymbol{\gamma}_i^T = \boldsymbol{\gamma}_{i+1}^T \boldsymbol{J}^{-1} = \boldsymbol{\gamma}_{i+2}^T \boldsymbol{J}^{-2} = \cdots = \boldsymbol{\gamma}_P^T \boldsymbol{J}^{-P+i}$, which implies that

$$\boldsymbol{\gamma}_P^T \boldsymbol{J}^{-P+i} \boldsymbol{X}_N = \boldsymbol{\gamma}_i^T \boldsymbol{X}_N, \qquad i = 1, \cdots, P. \tag{23}$$

However, focusing on the last L rows of (20), it follows that for $i = P,\ P-1, \cdots, P-L+1$, we have $\boldsymbol{\gamma}_i^T \boldsymbol{X}_N = \boldsymbol{0}^T$, which, after employing (23), leads to

$$\boldsymbol{\gamma}_P^T \boldsymbol{\mathcal{X}}_N = \boldsymbol{\gamma}_P^T \big(\boldsymbol{X}_N,\, \boldsymbol{J}^{-1}\boldsymbol{X}_N, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{X}_N\big) = \boldsymbol{0}^T. \tag{24}$$

We will show that the nullity $\nu(\boldsymbol{\mathcal{X}}_N) = 1$; thus, $\boldsymbol{\gamma}_P^T$ (and hence $\boldsymbol{\Gamma}$) can be determined from (24). In summary, we establish the following result.

Theorem 2: Let a0)–a3) hold true. Then, $\nu(\boldsymbol{\mathcal{X}}_N) = 1$, and $\boldsymbol{\gamma}_P^T$ can be identified from (24) as the unique (within a scalar ambiguity) null eigenvector of $\boldsymbol{\mathcal{X}}_N \boldsymbol{\mathcal{X}}_N^H$. With $\boldsymbol{\gamma}_P^T$ as the Pth row, the lower triangular Toeplitz matrix $\boldsymbol{\Gamma}$ can be built and used in (10) to obtain $\boldsymbol{G}_{zf} = (\boldsymbol{F}^{-1} \;\; \boldsymbol{0})\boldsymbol{\Gamma}$ directly.

Proof: To show that $\nu(\boldsymbol{\mathcal{X}}_N) = 1$, observe first that the matrix $\boldsymbol{\mathcal{X}}_N$ has dimensionality $P \times LN$ and that $N \geq P$. Thus, our goal is to prove that there are exactly $P-1$ linearly independent rows or columns that would imply that $\nu(\boldsymbol{\mathcal{X}}_N) = P - \text{rank}(\boldsymbol{\mathcal{X}}_N) = 1$. Note that by construction, (24) guarantees that $\boldsymbol{\gamma}_P^T \in \mathcal{N}(\boldsymbol{\mathcal{X}}_N)$ but not its uniqueness. Because $\boldsymbol{\gamma}_P^T \in \mathcal{N}(\boldsymbol{\mathcal{X}}_N)$, we know that $\nu(\boldsymbol{\mathcal{X}}) \geq 1$. Since $\boldsymbol{X}_N = \boldsymbol{H}\boldsymbol{F}\boldsymbol{S}_N$, matrix $\boldsymbol{\mathcal{X}}_N$ can be decomposed as

$$\boldsymbol{\mathcal{X}}_N = \big(\boldsymbol{H},\, \boldsymbol{J}^{-1}\boldsymbol{H}, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{H}\big)(\boldsymbol{I}_{L\times L} \otimes \boldsymbol{F}\boldsymbol{S}_N) \tag{25}$$

where $\otimes$ stands for the Kronecker product. In view of assumption a3), matrix $(\boldsymbol{I}_{L\times L} \otimes \boldsymbol{F}\boldsymbol{S}_N)$ is of rank LM; therefore, it does not affect the dimensionality of the range space of $\boldsymbol{\mathcal{X}}_N$, which depends only on the matrix $(\boldsymbol{H},\, \boldsymbol{J}^{-1}\boldsymbol{H}, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{H})$. The effect of multiplying successively $\boldsymbol{H}$ from the left-hand side by $\boldsymbol{J}^{-1}$ is to shift down the elements of each column,

adding every time one zero on top of it; thus, we have (26), shown at the bottom of the page. Observe that the last, and only the last, column of $\boldsymbol{J}^{-l}\boldsymbol{H}$ is linearly independent of all the M columns of $\boldsymbol{J}^{-l+1}\boldsymbol{H}$ for $l = 1, \cdots, L-1$, and each block $\boldsymbol{J}^{-l}\boldsymbol{H}$ adds to the matrix only one independent column. The matrix in (26) has thus exactly $M+L-1 = P-1$ linearly independent columns. Since all other columns are replicas of these $P-1$, $\text{rank}(\boldsymbol{\mathcal{X}}_N) = M+L-1 = P-1$ or, equivalently, $\nu(\boldsymbol{\mathcal{X}}_N) = 1$. ■

Our direct blind equalization algorithm consists of the following steps.

1) Form $\boldsymbol{X}_N$ with N blocks of data as $\boldsymbol{X}_N := (\boldsymbol{x}(n) \cdots \boldsymbol{x}(n+N-1))$ as in (12).
2) Form the matrix $\boldsymbol{\mathcal{X}}_N := (\boldsymbol{X}_N, \boldsymbol{J}^{-1}\boldsymbol{X}_N, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{X}_N)$ with $\boldsymbol{J}^{-1}$ transpose of the shift matrix defined in (21).
3) Determine the eigenvector $\boldsymbol{\gamma}_P$ corresponding to the null (smallest for noisy observations) eigenvalue of the matrix $\boldsymbol{\mathcal{X}}_N\boldsymbol{\mathcal{X}}_N^H$.
4) Equalize the data with $\boldsymbol{G} = (\boldsymbol{F}^{-1}\boldsymbol{0})\ \boldsymbol{\Gamma}$, where the ith row of $\boldsymbol{\Gamma}$ is given by $\boldsymbol{\gamma}_i = \boldsymbol{J}^{-P+i}\boldsymbol{\gamma}_P$ for $i = 1, \cdots, M$.

Neither the channel estimator in Theorem 1 nor the equalizer implied by Theorem 2 invoke any restrictions on the channel zeros or rely on any statistical input assumption (e.g., whiteness as in [14]). Thus, the input can be recovered exactly in the absence of noise. Contrary to CMA, all symbol constellations are allowed, and as long as the information rate is preserved, the symbol rate can be reduced in order to decrease the equivalent discrete-time channel order. The latter simplifies the channel estimation and equalization tasks considerably. Because only $N \geq P$ data blocks are required in $\boldsymbol{X}_N$, and each $\boldsymbol{x}(n)$ is $P \times 1$, the minimum number of symbols required is P^2.

Because the blind channel-equalizer $\boldsymbol{\Gamma}$ has a lower triangular Toeplitz form, it is easy to observe that the corresponding equalizer $\boldsymbol{G}$ can be written as

$$\boldsymbol{G} = \boldsymbol{F}^{-1}(\boldsymbol{\Gamma}_0 \quad \boldsymbol{0}_{M\times L}) \tag{27}$$

where $\boldsymbol{\Gamma}_0$ is $M \times M$ lower triangular Toeplitz, and it is built with first M rows and columns of $\boldsymbol{\Gamma}$. More specifically, $\boldsymbol{\Gamma}_0$ is the inverse of the matrix $\boldsymbol{H}_0$ built with the first M rows of $\boldsymbol{H}$. On the other hand, if we change the time reference so that the null guard interval is considered at the beginning of each block instead of at its end, each block of data can be written as

$$\begin{aligned}\boldsymbol{y}(n) &= \overline{\boldsymbol{H}}\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{F}\end{pmatrix}\boldsymbol{s}(n)\\ &:= \begin{pmatrix} h(L) & \cdots & h(0) & 0 & \cdots & 0\\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & \ddots & \ddots & 0\\ \vdots & \ddots & \ddots & \ddots & \ddots & h(0)\\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & \cdots & \cdots & 0 & h(L)\end{pmatrix}_{P\times P}\\ &\quad\cdot\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{F}\end{pmatrix}\boldsymbol{s}(n)\end{aligned} \tag{28}$$

where the $P \times P$ matrix $\overline{\boldsymbol{H}}$ is now Toeplitz upper triangular, and its inverse is, consequently, upper triangular and Toeplitz. In addition, in this case, it is possible to derive the equalizing matrix $\overline{\boldsymbol{\Gamma}}$ blindly, directly from the data, by forcing to zero the first L elements of $\overline{\boldsymbol{\Gamma}}\boldsymbol{y}(n)$ and arriving at the dual version of the algorithm described before for obtaining $\boldsymbol{\Gamma}_0$, where the last two steps as modified as follows.

3) Determine the eigenvector $\boldsymbol{\beta}_1$ associated with the null eigenvalue of $\overline{\boldsymbol{\mathcal{X}}}_N\overline{\boldsymbol{\mathcal{X}}}_N^H$, where
$$\overline{\boldsymbol{\mathcal{X}}}_N := (\boldsymbol{X}_N, \boldsymbol{J}\boldsymbol{X}_N, \cdots, \boldsymbol{J}^{L-1}\boldsymbol{X}_N).$$
4) Equalize the data with $\boldsymbol{G} = (\boldsymbol{0}, \boldsymbol{F}^{-1})\overline{\boldsymbol{\Gamma}}$, where the first row of $\overline{\boldsymbol{\Gamma}}$ is $\boldsymbol{\beta}_1$, and the generic ith row $\boldsymbol{\beta}_i$ is simply a right shifted replica of $\boldsymbol{\beta}_1$, i.e., $\boldsymbol{\beta}_i = \boldsymbol{\beta}_1\boldsymbol{J}^{i-1}$, with $i = 2, \cdots, P-1$.

Although the two solutions derived under the two different time references are in principle identical, they yield different performance in the presence of noise, as a function of the channel zero location. In fact, it is well known that in the equalization of SISO FIR channels, having transfer function $H(z)$, it is important to optimize the choice of the delay d leading to a stable equalizing filter with minimum norm. Specifically, denoting by $\gamma_d(n)$ the equalizing filter impulse response and by $\Gamma_d(z)$ its transfer function, the optimal delay results from the solution of the optimization problem

$$d_{opt} = \underset{d}{\text{argmin}} \sum_n |\gamma_d(n)|^2, \quad \text{with } \Gamma_{d_{opt}}(z)H(z) = z^{-d_{opt}}. \tag{29}$$

In the presence of white noise, this choice minimizes the noise variance at the equalizer output. In our transmission block

$$(\boldsymbol{H}, \boldsymbol{J}^{-1}\boldsymbol{H}, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{H}) = \begin{pmatrix} h(0) & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0\\ \vdots & h(0) & \ddots & \vdots & h(0) & \ddots & \vdots & \cdots & \vdots\\ h(L) & \cdots & \ddots & 0 & \vdots & \ddots & 0 & \cdots & \vdots\\ 0 & \ddots & \cdots & h(0) & \ddots & \cdots & 0 & \cdots & \vdots\\ \cdots & \cdots & \cdots & \cdots & \cdots & \ddots & h(0) & \cdots & \vdots\\ \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots & \cdots & 0\\ 0 & \cdots & 0 & h(L) & \cdots & h(L) & h(L-1) & \cdots & h(0)\end{pmatrix} \tag{26}$$

context, we face the same problem concerning the optimal delay, except that d is now confined between the values 0 and L. Introducing the notation

$$\boldsymbol{I}_d^T := \left(\boldsymbol{0}_{M\times d}\boldsymbol{I}_{M\times M}\boldsymbol{0}_{M\times(L-d)}\right) \tag{30}$$

the counterpart of (29) in our case is

$$d_{opt} = \underset{d}{\operatorname{argmin}} \|\boldsymbol{\Gamma}_d\| \quad \text{subject to} \quad \boldsymbol{\Gamma}_d\boldsymbol{I}_d^T\boldsymbol{H} = \boldsymbol{I}_{M\times M} \quad \text{and} \quad d_{opt}\in[0,\, L]. \tag{31}$$

Since $\boldsymbol{I}_d^T\boldsymbol{H}$ is a square matrix, the ZF constraint implies $\boldsymbol{\Gamma}_d = (\boldsymbol{I}_d^T\boldsymbol{H})^{-1}$, and therefore, $\min_d \|\boldsymbol{\Gamma}_d\| = \min_d \|(\boldsymbol{I}_d^T\boldsymbol{H})^{-1}\|$. Without loss of generality, we assume that the L channel roots ρ_l are ordered in the sense of decreasing amplitudes, i.e., $|\rho_1| \geq |\rho_2| \geq \cdots \geq |\rho_L|$. We are now able to solve the optimization problem stated in (31) via the following theorem.

Theorem 3: For $M \gg 1$, the optimal delay d_{opt} that guarantees $\boldsymbol{\Gamma}_{d_{opt}} = (\boldsymbol{I}_{d_{opt}}^T\boldsymbol{H})^{-1}$ and $\|\boldsymbol{\Gamma}_{d_{opt}}\| \leq \|\boldsymbol{\Gamma}_d\|$, $\forall\, d \neq d_{opt}$ is equal to the number of zeros ρ_l outside the unitary circle, i.e., $|\rho_l| > 1$. In particular

$$\lim_{M\to\infty} \frac{1}{M}\left\|(\boldsymbol{I}_d^T\boldsymbol{H})^{-1}\right\| = \begin{cases} \infty, & d \neq d_{opt} \\ 0, & d = d_{opt}. \end{cases} \tag{32}$$

Proof: The matrix $\boldsymbol{H}$, due to its Toeplitz structure, can always be decomposed as

$$\boldsymbol{H} = h(0)\boldsymbol{I}_0 + h(1)\boldsymbol{I}_1 + \cdots + h(L)\boldsymbol{I}_L. \tag{33}$$

It is straightforward to verify that $\boldsymbol{I}_j^T\boldsymbol{I}_i = \boldsymbol{J}^{j-i}$; hence, we can write the factorization of $\boldsymbol{I}_d^T\boldsymbol{H}$ as

$$\begin{aligned}\boldsymbol{I}_d^T\boldsymbol{H} &= h(0)\boldsymbol{J}^d + h(1)\boldsymbol{J}^{d-1} + \cdots + h(L)\boldsymbol{J}^{d-L} \\ &= h(0)\boldsymbol{J}^d \prod_{l=1}^{L}\left(\boldsymbol{I} - \rho_l\boldsymbol{J}^{-1}\right).\end{aligned} \tag{34}$$

Similar to $z^dH(z) = h(0)z^d\prod_{l=1}^{L}(1-\rho_l z^{-1})$, considering that approximately for $M \gg 1$, $\boldsymbol{I} \simeq \boldsymbol{J}\boldsymbol{J}^{-1}$ (because $\operatorname{diag}(\boldsymbol{J}\boldsymbol{J}^{-1}) = [\underbrace{1\cdots 1}_{M-1}, 0]$), we can write

$$\boldsymbol{I}_d^T\boldsymbol{H} \simeq h(0)\prod_{l=1}^{d}(-\rho_l)\left(\boldsymbol{I} - \rho_l^{-1}\boldsymbol{J}\right)\prod_{l=d+1}^{L}\left(\boldsymbol{I} - \rho_l\boldsymbol{J}^{-1}\right). \tag{35}$$

The approximation in (35) is valid in the sense that the ratio of the error norm over the matrix norm $\|\boldsymbol{I}_d^T\boldsymbol{H}\|$ tends to zero as M increases. From (35), we infer that $\boldsymbol{\Gamma}_d = (\boldsymbol{I}_d^T\boldsymbol{H})^{-1}$ can be factorized as[3]

$$\begin{aligned}\boldsymbol{\Gamma}_d &\simeq \frac{(-1)^d}{h(0)}\prod_{l=1}^{d}\frac{1}{\rho_l}\left(\boldsymbol{I} - \frac{1}{\rho_l}\boldsymbol{J}\right)^{-1}\prod_{l=d+1}^{L}\left(\boldsymbol{I} - \rho_l\boldsymbol{J}^{-1}\right)^{-1} \\ &= \frac{(-1)^d}{h(0)}\left[\prod_{l=1}^{d}\sum_{m=0}^{M-1}\frac{1}{\rho_l^{m+1}}\boldsymbol{J}^m\right] \cdot \left[\prod_{l=d+1}^{L}\sum_{m=0}^{M-1}\rho_l^m\boldsymbol{J}^{-m}\right].\end{aligned} \tag{36}$$

[3]In deriving this equation, we exploited the property $\boldsymbol{J}^m = \boldsymbol{0}$ for $m > M-1$ and the expansion $(I - r\boldsymbol{J})^{-1} = \sum_{m=0}^{M-1} r^m\boldsymbol{J}^m$, which is valid for *any* r.

From this decomposition, recalling that all matrix norms $\|.\|$ satisfy the submultiplicative property $\|\boldsymbol{A}\boldsymbol{B}\| \leq \|\boldsymbol{A}\|\,\|\boldsymbol{B}\|$, we have

$$\begin{aligned}\lim_{M\to\infty}\frac{1}{M}\|\boldsymbol{\Gamma}_d\| &= \lim_{M\to\infty}\frac{1}{M}\left\|\frac{(-1)^d}{h(0)}\prod_{l=1}^{d}\frac{1}{\rho_l}\left(\boldsymbol{I} - \frac{1}{\rho_l}\boldsymbol{J}\right)^{-1} \cdot \prod_{l=d+1}^{L}\left(\boldsymbol{I} - \rho_l\boldsymbol{J}^{-1}\right)^{-1}\right\| \\ &\leq \lim_{M\to\infty}\frac{1}{M}\frac{1}{|h(0)|^2}\prod_{l=1}^{d}\left\|\left(\boldsymbol{I} - \frac{1}{\rho_l}\boldsymbol{J}\right)^{-1}\right\| \cdot \prod_{l=d+1}^{L}\left\|\left(\boldsymbol{I} - \rho_l\boldsymbol{J}^{-1}\right)^{-1}\right\|.\end{aligned} \tag{37}$$

Therefore, to prove $\lim_{M\to\infty} 1/M\|\boldsymbol{\Gamma}_d\| = 0$, it is sufficient to focus on the behavior of the factors $(\boldsymbol{I} - (1/\rho_l)\boldsymbol{J})^{-1}$ or $(\boldsymbol{I} - \rho_l\boldsymbol{J}^{-1})^{-1}$, proving that $\lim_{M\to\infty} 1/M\|(\boldsymbol{I} - (1/\rho_l)\boldsymbol{J})^{-1}\| = 0$ or $\lim_{M\to\infty} 1/M\|(\boldsymbol{I} - \rho_l\boldsymbol{J}^{-1})^{-1}\| = 0$, respectively, only when d coincides with the number of zeros of $H(z)$ outside the unit circle. In fact, if we consider the Frobenious norm $\|.\|_F$, which is known to be always greater than or equal to the 2-norm $\|.\|_2$ (see [17, p. 57]), we can write the upper limit for the 2-norm of each factor in (37) as

$$\begin{aligned}\left\|\left(\boldsymbol{I} - \rho_l^{\pm 1}\boldsymbol{J}^{\mp 1}\right)^{-1}\right\|_2 &\leq \left\|\left(\boldsymbol{I} - \rho_l^{\pm 1}\boldsymbol{J}^{\mp 1}\right)^{-1}\right\|_F \\ &= \sqrt{M\sum_{m=0}^{M-1}(1 - m/M)\rho_l^{\pm 2m}}.\end{aligned} \tag{38}$$

Therefore, recalling that the roots are ordered in the sense of decreasing radii, we have

$$\begin{aligned}&\lim_{M\to\infty}\frac{1}{M}\left\|\left(\boldsymbol{I} - \rho_l^{\pm 1}\boldsymbol{J}^{\mp 1}\right)^{-1}\right\|_2 \\ &\quad\leq \lim_{M\to\infty}\frac{1}{\sqrt{M}}\sqrt{\sum_{m=0}^{M-1}(1 - m/M)\rho_l^{\pm 2m}} = 0\end{aligned} \tag{39}$$

if $d = d_{opt}$ because in such a case, all the roots outside the unit circle are inverted, whereas the other ones are not. Considering that the left-hand side of (37) is certainly non-negative, this proves that $\lim_{M\to\infty}(1/M)\|\boldsymbol{\Gamma}_d\| = 0$ when $d = d_{opt}$. Conversely, if $d \neq d_{opt}$, there exists at least one factor in (37) that is not invertible when M goes to ∞. In fact, since, in general, $1/\sqrt{M}\|.\|_F \leq \|.\|_2$ (e.g., [17, p. 57]), and $1/\sqrt{M}\|.\|_F = \sqrt{\sum_{m=0}^{M-1}(1 - m/M)\rho_l^{\pm 2m}}$, if $d \neq d_{opt}$, there certainly is a term with an infinite 2-norm. An unbounded 2-norm for one factor implies that this factor is not invertible, leading to $\lim_{M\to\infty} 1/M\|\boldsymbol{\Gamma}_d\| = \infty$, and this completes the proof of the theorem. ■

Interestingly, this theorem provides, as a by-product, a blind deterministic method for identifying the number of channel zeros outside the unit circle: a task that is impossible in the standard SISO setup with output second-order statistics. In fact

$$d_{opt} = \arg\min_d \left\|\left(\boldsymbol{I}_d^T\boldsymbol{X}_N\right)^{-1}\right\|. \tag{40}$$

Theorem 3 implies that $\mathbf{\Gamma}_L$ enhances the noise whenever there are channel zeros *inside* the unit circle. Hence, even if both $\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_L$ estimates are consistent, as will be proved in the following, the average (over several channels) performance in terms of bit error rate of both equalizers $\boldsymbol{G} = \boldsymbol{F}^{-1}\mathbf{\Gamma}_0\boldsymbol{I}_0^T$, $\boldsymbol{G} = \boldsymbol{F}^{-1}\mathbf{\Gamma}_L\boldsymbol{I}_L^T$ will not achieve performance similar to the MMSE ZF equalizer given by $\boldsymbol{G}_{\text{mmse-}zf} = \boldsymbol{F}^{-1}\boldsymbol{H}^{\dagger}$. Therefore, whenever the estimate of $d_{opt} \neq 0$ or L, we can estimate $\mathbf{\Gamma}_{d_{opt}}$ indirectly, by using the approach proposed in [25]. Alternatively, we can estimate the channel, up to a scale, using the equation $\boldsymbol{\gamma}^T\boldsymbol{H} = \boldsymbol{\beta}^T\boldsymbol{H} = \boldsymbol{0}^T$ as the null vectors $\boldsymbol{u}_l$ in Theorem 1 and then build, with this channel estimate, the optimal equalizer. Channel identification is guaranteed in this case, and even if we carry out the estimate using a single null vector instead of $\tilde{\boldsymbol{U}}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are unique.

Remark 1: We have shown in [24] that with $f_m(n) = \exp(j2\pi mn/M)$, the filterbank precoder of Fig. 1 reduces to the digital OFDM transmitter [5]. Under a2), the trailing zeros TZ-OFDM (which are detailed in [24]) can be equalized blindly even when $h(n)$ has unit circle zeros located at $2\pi m/M$; this is a case where deep fades deteriorate performance of conventional OFDM (see also [6] for a blind LMS adaptive solution).

C. Direct Blind Synchronization

Equation (8) assumes that block synchronization has been accomplished. Although techniques relying on training data are possible, it is of interest to achieve block synchronization blindly, which is a task complementing our blind equalizer nicely (see also [27] for a statistical method). A deterministic blind approach is proposed herein after observing that matrix $\boldsymbol{\mathcal{X}}_N$ in (24) becomes full rank when the receiver is not block synchronous. Specifically, we propose to retrieve a possible time-offset of d symbols between transmitter and receiver by checking the rank properties of the matrix $\boldsymbol{\mathcal{X}}_N^{(d)} := (\boldsymbol{X}_N^{(d)} \;\; \boldsymbol{J}^{-1}\boldsymbol{X}_N^{(d)} \cdots \boldsymbol{J}^{-L+1}\boldsymbol{X}_N^{(d)})$ as a function of the delay d. Each matrix $\boldsymbol{X}_N^{(d)}$ is obtained similar to (12) by collecting N vectors $\boldsymbol{x}^{(d)}(n) = (x(nP+d),\ x(nP+d+1),\ \cdots,\ x(nP+d+P-1))^T$ for $n = 0, \cdots, N-1$.

The following theorem summarizes our blind block synchronization result.

Theorem 4: If blocks at the receiver are off by $d = d_0$ samples and we form $\boldsymbol{\mathcal{X}}_N^{(d)}$ matrices for each possible shift $|d| < P/2$, then d_0 can be found as

$$d_0 = \arg\min_d \lambda_{\min}\left(\boldsymbol{\mathcal{X}}_N^{(d)}\boldsymbol{\mathcal{X}}_N^{(d)H}\right) \tag{41}$$

where $\lambda_{\min}(\boldsymbol{\mathcal{X}}_N^{(d)}\boldsymbol{\mathcal{X}}_N^{(d)H})$ denotes the minimum eigenvalue. In the noise-free case, we guarantee identifiability by showing that $\lambda_{\min}(\boldsymbol{\mathcal{X}}_N^{(d_0)}\boldsymbol{\mathcal{X}}_N^{(d_0)H}) = 0$, whereas $\lambda_{\min}(\boldsymbol{\mathcal{X}}_N^{(d)}\boldsymbol{\mathcal{X}}_N^{(d)H}) > 0$ for $d \neq d_0$.

Proof: We will assume, without loss of generality, that $d_0 = 0$ and that the offset d is positive since the extension of the proof for $d < 0$ is straightforward.

It is instrumental for the proof to represent the convolution in the transmit-filterbank through the matrix-vector multiplication between the $P \times (P+L)$ matrix

$$\boldsymbol{\mathcal{H}}_0 := \begin{pmatrix} h(L) & \cdots & h(0) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & h(L) & \cdots & h(0) \end{pmatrix} \tag{42}$$

and the input blocks $\boldsymbol{s}(n)$. For a shift d, the precoded block is given by the $(P+L)\times 1$ vector

$$\begin{aligned}\boldsymbol{u}^{(d)}(n) = (\boldsymbol{J}^{-d})^T \begin{pmatrix} \boldsymbol{0}_{L\times M} \\ \boldsymbol{F} \\ \boldsymbol{0}_{L\times M} \end{pmatrix} \boldsymbol{s}(n) \\ + \boldsymbol{J}^{-P+d} \begin{pmatrix} \boldsymbol{0}_{L\times M} \\ \boldsymbol{F} \\ \boldsymbol{0}_{L\times M} \end{pmatrix} \boldsymbol{s}(n+1)\end{aligned} \tag{43}$$

where $\boldsymbol{J}^{-1}$ is a $(P+L)\times(P+L)$ positive shift matrix similar to (21). There are two cases that have to be treated separately: i) $|d| \geq L$ and ii) $|d| < L$.

i) $d \geq L$: We will prove that in this case, the matrix $\boldsymbol{X}_N^{(d)}$ is full rank. Therefore, the matrix $\boldsymbol{\mathcal{X}}_N^{(d)}\boldsymbol{\mathcal{X}}_N^{(d)H}$ is necessarily full rank, and thus, its minimum eigenvalue is strictly positive. For $d \geq L$, the vector $\boldsymbol{u}^{(d)}(n)$ can be partitioned as

$$\boldsymbol{u}^{(d)}(n) = (\overbrace{\times\cdots\times}^{M-d+L}\ \overbrace{0\cdots 0}^{L}\ \overbrace{\times\cdots\times}^{d})^T \tag{44}$$

where $\times$ stands for a nonzero element. Collecting $N > P$ consecutive vectors, we form the matrix $\boldsymbol{U}_N^{(d)}$ as

$$\begin{aligned}\boldsymbol{U}_N^{(d)} &:= \left(\boldsymbol{u}^{(d)}(n), \cdots, \boldsymbol{u}^{(d)}(n+N-1)\right) \\ &= \begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{0}_{L\times N} \\ \boldsymbol{U}_2 \end{pmatrix}\end{aligned} \tag{45}$$

where $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ have dimensionalities $(M-d+L)\times N$ and $d \times N$, respectively. The matrix $\boldsymbol{X}_N$ can be written equivalently as

$$\boldsymbol{X}_N^{(d)} = \boldsymbol{\mathcal{H}}_0\boldsymbol{U}_N^{(d)} = \boldsymbol{\mathcal{H}}\begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{pmatrix} \tag{46}$$

where $\boldsymbol{\mathcal{H}}$ is a $P \times P$ matrix obtained by $\boldsymbol{\mathcal{H}}_0$, eliminating the L central columns [i.e., from the $(M-d+L+1)$th to the $(M-d+2L)$th column] that multiply the null elements of the input matrix, i.e., as shown in (47) at the bottom of the next page. The matrix in (47) is full rank since both $h(0)$ and $h(L)$ are nonzero. Relying on the persistence-of-excitation assumption a3), we infer that there exists a value of $N \geq P$ such that the matrix $(\boldsymbol{U}_1^T\boldsymbol{U}_2^T)^T$ is full row rank, implying that $\boldsymbol{X}_N^{(d)}$ [and thus $\boldsymbol{\mathcal{X}}_N^{(d)}$] is full rank.

ii) $d < L$: In this case, we will prove that the matrix $\boldsymbol{X}_N^{(d)}$ is not full rank for $d \neq 0$, but $\boldsymbol{\mathcal{X}}_N^{(d)}$ is still full rank, as in the previous case. In fact, the input vector $\boldsymbol{u}^{(d)}(n)$ now has more than L zeros

$$\boldsymbol{u}^{(d)}(n) = (\overbrace{0\cdots 0}^{L-d}\ \overbrace{\times\cdots\times}^{M}\ \overbrace{0\cdots 0}^{L}\ \overbrace{\times\cdots\times}^{d})^T \tag{48}$$

and correspondingly, the matrix $\boldsymbol{X}_N^{(d)}$ is

$$\boldsymbol{X}_N^{(d)} = \mathcal{H}_0 \begin{pmatrix} \boldsymbol{0}_{L-d\times N} \\ \boldsymbol{U}_1 \\ \boldsymbol{0}_{L\times N} \\ \boldsymbol{U}_2 \end{pmatrix}. \tag{49}$$

In addition, in this case, we can write $\boldsymbol{X}_N^{(d)}$ in terms of products between the matrices $\boldsymbol{\mathcal{H}}$ and $(\boldsymbol{U}_1^T\boldsymbol{U}_2^T)^T$ with their zero entries removed. However, since there are more zeros in the sequence, the resulting matrix will not be full row rank. In fact, in this case, $\boldsymbol{\mathcal{H}}$ is a $P\times(P-d)$ matrix, whose structure is similar to that obtained by eliminating the first $L-d$ columns from (47). Following the same steps used in Section III-B to show that $\nu(\boldsymbol{\mathcal{X}}_N^{(0)}) = \nu(\boldsymbol{\mathcal{X}}_N) = 1$, we can factorize the matrix $\boldsymbol{\mathcal{X}}_N^{(d)}$ as

$$\begin{aligned}\boldsymbol{\mathcal{X}}_N^{(d)} = &\left(\boldsymbol{\mathcal{H}}, \boldsymbol{J}^{-1}\boldsymbol{\mathcal{H}}, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{\mathcal{H}}\right) \\ &\cdot \left(\boldsymbol{I}_{L\times L}\otimes\left(\boldsymbol{U}_1^T\boldsymbol{U}_2^T\right)^T\right).\end{aligned} \tag{50}$$

By virtue of a3), $\boldsymbol{I}_{L\times L}\otimes(\boldsymbol{U}_1^T\boldsymbol{U}_2^T)^T$ is a square full rank matrix; hence, $\nu(\boldsymbol{\mathcal{X}}_N^{(d)}) = \nu((\boldsymbol{\mathcal{H}}, \boldsymbol{J}^{-1}\boldsymbol{\mathcal{H}}, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{\mathcal{H}}))$. Keeping in mind the structure of $\boldsymbol{\mathcal{H}}$, it follows by inspection (it is just more tedious to write it) that again, as observed about (26), every block $\boldsymbol{J}^{-l}\boldsymbol{\mathcal{H}}$ adds one, and only one, independent column to the previous one. Thus, since $\text{rank}(\boldsymbol{\mathcal{H}}) = P-d$, because there are $P-d$ independent columns for each block, and since $d<L$, the matrix $(\boldsymbol{\mathcal{H}}, \boldsymbol{J}^{-1}\boldsymbol{\mathcal{H}}, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{\mathcal{H}})$ is full row rank. Hence, $\nu(\boldsymbol{\mathcal{X}}_N^{(d)}) = 0$ for $d\neq 0$. ∎

Our blind block synchronization algorithm follows these steps.

1) With N blocks of data, form the matrix $\boldsymbol{\mathcal{X}}_N^{(d)} := (\boldsymbol{X}_N^{(d)}, \boldsymbol{J}^{-1}\boldsymbol{X}_N^{(d)}, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{X}_N^{(d)})$ as a function of the delay d for $|d| < P/2$. Each matrix $\boldsymbol{X}_N^{(d)}$ is obtained by collecting N vectors $\boldsymbol{x}^{(d)}(n) = (x(nP+d), x(nP+d+1), \cdots, x(nP+d+P-1))^T$, for $n=0, \cdots, N-1$.
2) For each d such that $|d| < P/2$, determine the smallest eigenvalue of the matrix $\boldsymbol{\mathcal{X}}_N^{(d)}\boldsymbol{\mathcal{X}}_N^{(d)H}$.
3) Determine the time offset d_0 in the block synchronization as in (41).

To augment the accuracy of the block time synchronization algorithm, we could exploit the dual solution and also collect matrices $\overline{\boldsymbol{\mathcal{X}}}^{(d)} = (\boldsymbol{X}_N, \boldsymbol{J}\boldsymbol{X}_N, \cdots, \boldsymbol{J}^{L-1}\boldsymbol{X}_N)$ and the corresponding $\overline{\lambda}_{\min}(\overline{\boldsymbol{\mathcal{X}}}^{(d)}\overline{\boldsymbol{\mathcal{X}}}^{(d)H})$. Similar steps prove that in the absence of noise, $d_0 = \arg\min_d[\overline{\lambda}_{\min}(\overline{\boldsymbol{\mathcal{X}}}^{(d)}\overline{\boldsymbol{\mathcal{X}}}^{(d)H})]$.

IV. Noisy Case—Consistency

In Section III-A, we established that, in the absence of noise, the channel can be identified exactly. In this section, we prove that also in the presence of noise, we can identify the channel uniquely, using the received data covariance matrix. When stationary additive noise $v(n)$ is present, the data correlation matrix $\boldsymbol{R}_{yy} := E\{\boldsymbol{y}(n)\boldsymbol{y}^H(n)\}$ is given by [cf. (8)]

$$\boldsymbol{R}_{yy} = \boldsymbol{R}_{xx} + \boldsymbol{R}_{vv} = \boldsymbol{H}\boldsymbol{F}\boldsymbol{R}_{ss}\boldsymbol{F}^H\boldsymbol{H}^H + \boldsymbol{R}_{vv}. \tag{51}$$

In practice, the ensemble correlation matrices are replaced by sample averages, which converge in the mean square sense to true correlation matrices since $\boldsymbol{y}(n)$ in (8) is mixing [input $s(n)$ has finite moments and $h(n)$ has finite memory]; thus

$$\hat{\boldsymbol{R}}_{yy}^{(N)} := \frac{1}{N}\sum_{n=0}^{N-1}\boldsymbol{y}(n)\boldsymbol{y}^H(n) \xrightarrow{\text{m.s.s.}} \boldsymbol{R}_{yy}, \qquad \text{as } N\to\infty \tag{52}$$

where m.s.s. stands for mean-square sense convergence. Assuming that the symbols and noise have zero mean, the correlation matrix coincides with the covariance matrix. We now establish consistency for both channel identification and direct equalization methods.

A. Channel Identification

Assuming that the noise covariance matrix $\boldsymbol{R}_{vv}$ is known, we prove that it is still possible to identify the channel, up to a constant factor, as the number of samples N goes to infinity. The proof is developed in two steps.

Step S1)

S1) In the absence of noise, we reformulate the blind identification algorithm presented in Section III-A in terms of the covariance matrix $\boldsymbol{R}_{xx}$ of the received data instead of the matrix $\boldsymbol{X}_N\boldsymbol{X}_N^H$ used in Theorem 2.
S2) In the presence of noise, we use the covariance matrix $\boldsymbol{R}_{yy}$ instead of $\boldsymbol{R}_{xx}$ if the noise is white; otherwise, we prewhiten the noise and then use the eigendecomposition of $\boldsymbol{\Phi}_v^{-1}\boldsymbol{R}_{yy}\boldsymbol{\Phi}_v^{-H}$, where $\boldsymbol{\Phi}_v$ is any matrix that factorizes $\boldsymbol{R}_{vv}$ as $\boldsymbol{R}_{vv} = \boldsymbol{\Phi}_v\boldsymbol{\Phi}_v^H$.

$$\boldsymbol{\mathcal{H}} := \begin{pmatrix} h(L) & \cdots & h(0) & \cdots & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \cdots & \ddots & \vdots & \vdots & \ddots & \cdots & \cdots & 0 \\ \vdots & \cdots & \ddots & \cdots & h(0) & \vdots & \ddots & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & \ddots & \vdots & \vdots & \ddots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & h(L) & 0 & \ddots & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \ddots & 0 & h(0) & 0 & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \vdots & \vdots & \ddots & \ddots & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & 0 & h(L) & \vdots & \ddots & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & \vdots & \ddots & \ddots & \vdots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \vdots & \vdots & \ddots & h(L) & \cdots & h(0) \end{pmatrix} \tag{47}$$

Step S1) Since the $M \times M$ covariance matrix $\boldsymbol{R}_{ss}$ is full rank, assumptions a2) and a3) imply that rank$(\boldsymbol{F}\boldsymbol{R}_{ss}\boldsymbol{F}^H) = M$. From (12), it follows that $\boldsymbol{R}_{xx} := E\{\boldsymbol{X}_N\boldsymbol{X}_N^H\} = \boldsymbol{H}\boldsymbol{F}\boldsymbol{R}_{ss}\boldsymbol{F}^H\boldsymbol{H}^H$; hence, we also have rank$(\boldsymbol{R}_{xx}) = M$. Therefore, the noise subspace of $\boldsymbol{R}_{xx}$ has dimension $P - M = L$. The noise subspace is thus spanned by the last L columns $\tilde{\boldsymbol{U}}_x$ of the matrix $\boldsymbol{U}_x := (\overline{\boldsymbol{U}}_x\tilde{\boldsymbol{U}}_x)$, which decomposes the covariance matrix as $\boldsymbol{R}_{xx} = \boldsymbol{U}_x\boldsymbol{\Sigma}_x\boldsymbol{U}_x^H$, assuming that the eigenvalues are written in a decreasing order. Since $\boldsymbol{H}$ is full column rank, $\boldsymbol{H}^H$ is full row rank, and hence, the matrix $\boldsymbol{F}\boldsymbol{R}_{ss}\boldsymbol{F}^H\tilde{\boldsymbol{H}}_0^H$ is full row rank. Therefore, the left null space of $\boldsymbol{R}_{xx} = \boldsymbol{H}\boldsymbol{F}\boldsymbol{R}_{ss}\boldsymbol{F}^H\boldsymbol{H}^H$ coincides with the left null space of $\boldsymbol{H}$. Once $\tilde{\boldsymbol{U}}_x$ has been obtained, we can proceed exactly as in (14). □

Step S2) Here, we show how to determine $\tilde{\boldsymbol{U}}_x$ from the covariance matrix of the noisy data $\boldsymbol{R}_{yy}$, assuming that the noise covariance $\boldsymbol{R}_{vv}$ is known.

If the noise is correlated, it is necessary to whiten it by multiplying the received data vector by any matrix $\boldsymbol{\Phi}_v^{-1}$ such that $\boldsymbol{R}_{vv} = \boldsymbol{\Phi}_v\boldsymbol{\Phi}_v^H$. The covariance matrix of the prewhitened data is

$$\boldsymbol{\Phi}_v^{-1}\boldsymbol{R}_{yy}\boldsymbol{\Phi}_v^{-H} = \boldsymbol{\Phi}_v^{-1}\boldsymbol{R}_{xx}\boldsymbol{\Phi}_v^{-H} + \boldsymbol{I}. \tag{53}$$

Defining the eigendecompositions

$$\boldsymbol{\Phi}_v^{-1}\boldsymbol{R}_{yy}\boldsymbol{\Phi}_v^{-H} := \boldsymbol{U}_y^{(\Phi)}\boldsymbol{\Sigma}_y^{(\Phi)}\boldsymbol{U}_y^{(\Phi)H} \tag{54}$$

and

$$\boldsymbol{\Phi}_v^{-1}\boldsymbol{R}_{xx}\boldsymbol{\Phi}_v^{-H} := \boldsymbol{U}_x^{(\Phi)}\boldsymbol{\Sigma}_x^{(\Phi)}\boldsymbol{U}_x^{(\Phi)H} \tag{55}$$

we obtain

$$\boldsymbol{U}_y^{(\Phi)}\boldsymbol{\Sigma}_y^{(\Phi)}\boldsymbol{U}_y^{(\Phi)H} = \boldsymbol{U}_x^{(\Phi)}(\boldsymbol{\Sigma}_x^{(\Phi)} + \boldsymbol{I})\boldsymbol{U}_x^{(\Phi)H} \tag{56}$$

implying that $\boldsymbol{U}_y^{(\Phi)}$ coincides with $\boldsymbol{U}_x^{(\Phi)}$. Let us partition $\boldsymbol{U}_y^{(\Phi)}$ as $(\overline{\boldsymbol{U}}_y^{(\Phi)}, \tilde{\boldsymbol{U}}_y^{(\Phi)})$, where $\tilde{\boldsymbol{U}}_y^{(\Phi)}$ contains the L eigenvectors associated with the smallest eigenvalues of $\boldsymbol{\Phi}_v^{-1}\boldsymbol{R}_{yy}\boldsymbol{\Phi}_v^{-H}$. Since $\nu(\boldsymbol{R}_{xx}) = \nu(\boldsymbol{\Phi}_v^{-1}\boldsymbol{R}_{xx}\boldsymbol{\Phi}_v^{-H}) = L$, the eigenvectors $\tilde{\boldsymbol{U}}_x^{(\Phi)} = \tilde{\boldsymbol{U}}_y^{(\Phi)}$ have to lie in the null space of $\boldsymbol{\Phi}_v^{-1}\boldsymbol{R}_{xx}\boldsymbol{\Phi}_v^{-H}$. Hence

$$\tilde{\boldsymbol{U}}_y^{(\Phi)H}\boldsymbol{\Phi}_v^{-1}\boldsymbol{R}_{xx}\boldsymbol{\Phi}_v^{-H} = \boldsymbol{0} \Rightarrow \boldsymbol{\Phi}_v^{-H}\tilde{\boldsymbol{U}}_y^{(\Phi)} \in \mathcal{N}(\boldsymbol{R}_{xx}) \tag{57}$$

which implies that $\boldsymbol{\Phi}_v^{-H}\tilde{\boldsymbol{U}}_y^{(\Phi)}$ is a basis for $\mathcal{N}(\boldsymbol{R}_{xx})$ or, equivalently, that $\tilde{\boldsymbol{U}}_x$ is related to $\boldsymbol{\Phi}^{-H}\tilde{\boldsymbol{U}}_y^{(\Phi)}$ through the product by an invertible matrix. Notice however, that, as far as channel estimation is concerned, nothing prevents us from using the matrix $\boldsymbol{\Phi}^{-H}\tilde{\boldsymbol{U}}_y^{(\Phi)}$ directly in (14) instead of $\tilde{\boldsymbol{U}}_x$ because the proof of identifiability holds for any basis of $\mathcal{N}(\boldsymbol{R}_{xx})$.

Finally, because $\boldsymbol{\Phi}^{-H}\tilde{\boldsymbol{U}}_y^{(\Phi)}$ is a continuous function of $\boldsymbol{R}_{yy}$ and the estimate (52) of $\boldsymbol{R}_{yy}$ is consistent, it follows that the proposed channel estimation algorithm is also consistent.

B. Direct Equalization

Similar to the previous subsection, we will also present our consistency results in two steps (noise-free and noisy cases).

Step S1) Using (12) and (20) and adopting the same approach as in Section III-B, we find that

$$E\{\boldsymbol{\Gamma}\boldsymbol{X}_N\boldsymbol{X}_N^H\} := \boldsymbol{\Gamma}\boldsymbol{R}_{xx} = \begin{pmatrix}\boldsymbol{F}\boldsymbol{R}_{ss}\boldsymbol{F}^H\boldsymbol{H}^H \\ \boldsymbol{0}_{L\times P}\end{pmatrix}$$
$$\Rightarrow \overline{\boldsymbol{\gamma}}_P^T\boldsymbol{R}_{xx} = \boldsymbol{0}^T. \tag{58}$$

Hence, arguing as in (23), we infer that $\boldsymbol{\gamma}_P^T\boldsymbol{J}^{-l}\boldsymbol{R}_{xx} = \boldsymbol{0}^T$, which implies that $\boldsymbol{\gamma}_P^T\boldsymbol{J}^{-l}\boldsymbol{R}_{xx}\boldsymbol{J}^l = \boldsymbol{0}^T$ for $l = 0, 1, \cdots, L-1$. Therefore,

$$\boldsymbol{\gamma}_P^T\boldsymbol{\mathcal{R}}_{xx} = \boldsymbol{0}^T \tag{59}$$

where

$$\boldsymbol{\mathcal{R}}_{xx} := \sum_{l=0}^{L-1}\boldsymbol{J}^{-l}\boldsymbol{R}_{xx}\boldsymbol{J}^l. \tag{60}$$

Matrix $\boldsymbol{\mathcal{R}}_{xx}$ will be shown to have nullity $\nu(\boldsymbol{\mathcal{R}}_{xx}) = 1$. In fact, matrix $\boldsymbol{\mathcal{R}}_{xx}$ is nothing but

$$\begin{aligned}\boldsymbol{\mathcal{R}}_{xx} &= E\{\boldsymbol{\mathcal{X}}_N\boldsymbol{\mathcal{X}}_N^H\} \\ &= \left(\boldsymbol{H}, \boldsymbol{J}^{-1}\boldsymbol{H}, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{H}\right)\left(\boldsymbol{I}_{L\times L}\otimes\boldsymbol{F}\boldsymbol{R}_{ss}\boldsymbol{F}^H\right) \\ &\quad\cdot\left(\boldsymbol{H}, \boldsymbol{J}^{-1}\boldsymbol{H}, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{H}\right)^H.\end{aligned} \tag{61}$$

It follows from (61) that $\mathcal{N}(\boldsymbol{\mathcal{R}}_{xx}) = \mathcal{N}((\boldsymbol{H}, \boldsymbol{J}^{-1}\boldsymbol{H}, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{H}))$ because $(\boldsymbol{I}_{L\times L}\otimes\boldsymbol{F}\boldsymbol{R}_{ss}\boldsymbol{F}^H)$ is full rank, and it has been already shown in Section III-A [see (26)] that $\nu((\boldsymbol{H}, \boldsymbol{J}^{-1}\boldsymbol{H}, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{H})) = 1$, which leads to our assertion that $\nu(\boldsymbol{\mathcal{R}}_{xx}) = 1$. Indeed, if the noise were not present, $\boldsymbol{\gamma}_P$ (and hence $\boldsymbol{\Gamma}$) could have been obtained (within a scale factor) as the null eigenvector of $\boldsymbol{\mathcal{R}}_{xx}$.

Step S2) In the noisy case, we define $\boldsymbol{\mathcal{R}}_{yy}$ ($\boldsymbol{\mathcal{R}}_{vv}$) as $\boldsymbol{\mathcal{R}}_{xx}$ in (60) with $\boldsymbol{R}_{yy}$ ($\boldsymbol{R}_{vv}$) replacing $\boldsymbol{R}_{xx}$. The procedure is now exactly equivalent to the one shown before for the channel estimation method, with the only difference that the null space of $\boldsymbol{\mathcal{R}}_{xx}$ has dimensionality one.

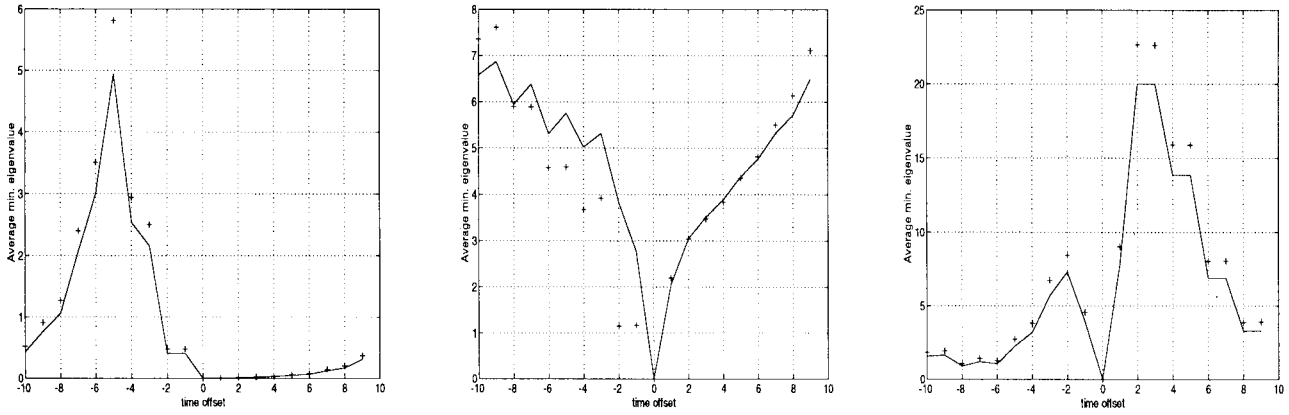
Factorizing $\boldsymbol{\mathcal{R}}_{vv} = \boldsymbol{\Phi}_v\boldsymbol{\Phi}_v^H$, we can prewhiten the noise by multiplying the data matrix by $\boldsymbol{\Phi}_v^{-1}$. Hence, using (51), we can write

$$\begin{aligned}\boldsymbol{\Phi}_v^{-1}\boldsymbol{\mathcal{R}}_{yy}\boldsymbol{\Phi}_v^{-H} &= \boldsymbol{\Phi}_v^{-1}\boldsymbol{\mathcal{R}}_{xx}\boldsymbol{\Phi}_v^{-H} + \boldsymbol{I}_{P\times P} \\ &= \boldsymbol{\mathcal{U}}_x^{(\Phi)}\left(\boldsymbol{\Lambda}_x^{(\Phi)} + \boldsymbol{I}\right)\boldsymbol{\mathcal{U}}_x^{(\Phi)H}\end{aligned} \tag{62}$$

where $\boldsymbol{\mathcal{U}}_x^{(\Phi)}$ ($\boldsymbol{\Lambda}_x^{(\Phi)}$) denotes the eigenvector (eigenvalue) matrix of $\boldsymbol{\Phi}_v^{-1}\boldsymbol{\mathcal{R}}_{xx}\boldsymbol{\Phi}_v^{-H}$. As in the previous case, we observe that $\nu(\boldsymbol{\mathcal{R}}_{xx}) = \nu(\boldsymbol{\Phi}_v^{-1}\boldsymbol{\mathcal{R}}_{xx}\boldsymbol{\Phi}_v^H) = 1$, and thus, the minimum eigenvector of $\boldsymbol{\mathcal{R}}_{yy}$ spans $\mathcal{N}(\boldsymbol{\Phi}_v^{-1}\boldsymbol{\mathcal{R}}_{xx}\boldsymbol{\Phi}_v^H)$. In fact, if we define $\overline{\boldsymbol{\gamma}}_P^T := \boldsymbol{\gamma}_P^T\boldsymbol{\Phi}_v$ and rewrite (60) as $\overline{\boldsymbol{\gamma}}_P^T\boldsymbol{\Phi}_v^{-1}\boldsymbol{\mathcal{R}}_{xx} = \boldsymbol{0}^T$, we obtain $\overline{\boldsymbol{\gamma}}_P^T\boldsymbol{\Phi}_v^{-1}\boldsymbol{\mathcal{R}}_{xx}\boldsymbol{\Phi}_v^{-H} = \boldsymbol{0}^T$. The latter, along with (62), yields

$$\overline{\boldsymbol{\gamma}}_P^T\boldsymbol{\Phi}_v^{-1}\boldsymbol{\mathcal{R}}_{yy}\boldsymbol{\Phi}_v^{-H} = \overline{\boldsymbol{\gamma}}_P^T \tag{63}$$

which shows that $\overline{\boldsymbol{\gamma}}_P^T$ is the eigenvector of $\boldsymbol{\Phi}_v^{-1}\boldsymbol{\mathcal{R}}_{yy}\boldsymbol{\Phi}_v^{-H}$ corresponding to its unitary eigenvalue, which, in force of (62), is also its minimum eigenvalue. Based on $\overline{\boldsymbol{\gamma}}_P^T$, we can find that $\boldsymbol{\gamma}_P^T = \overline{\boldsymbol{\gamma}}_P^T\boldsymbol{\Phi}_v^{-1}$, and thus, the direct equalizing matrix $\boldsymbol{\Gamma}$ relies only on the noisy covariance matrix $\boldsymbol{\mathcal{R}}_{yy}$ and knowledge of the noise covariance $\boldsymbol{\mathcal{R}}_{vv}$. □



(a) (b) (c)

Fig. 2. (a) Minimum-phase channel. (b) Channel with two zeros on the unit circle. (c) Nonminimum-phase channel (theoretical value: solid line, average value obtained by simulation: $+$, SNR $= \infty$).

Notice that even if the noise $v(n)$ is white, a prewhitening of the data is required because $\boldsymbol{\mathcal{R}}_{vv} := \sigma_v^2 \sum_{l=0}^{L-1} \boldsymbol{J}^{-l}\boldsymbol{J}^{l} = \sigma_v^2$ diag$(1\,2 \cdots L \cdots L)$, and its spectral factor in (63) is

$$\boldsymbol{\Phi}_v := \text{diag}(1\,\sqrt{2} \cdots \sqrt{L} \cdots \sqrt{L}). \tag{64}$$

With sample averages only available, consistency of our direct equalizer estimator follows easily from the definition $\boldsymbol{Y}_N := \boldsymbol{X}_N + \boldsymbol{V}_N$ and the fact that

$$\hat{\boldsymbol{\mathcal{R}}}_{yy}^{(N)} = \frac{1}{N}\boldsymbol{\mathcal{Y}}_N\boldsymbol{\mathcal{Y}}_N^H$$
$$\boldsymbol{\mathcal{Y}}_N := \left(\boldsymbol{Y}_N, \boldsymbol{J}^{-1}\boldsymbol{Y}_N, \cdots, \boldsymbol{J}^{-L+1}\boldsymbol{Y}_N\right). \tag{65}$$

As before, the sample average $\hat{\boldsymbol{\mathcal{R}}}_{yy}^{(N)}$ tends to the ensemble average $\boldsymbol{\mathcal{R}}_{yy}$ as N goes to infinity.

At this point, a natural question arises: Should we use in practice the "deterministic" approaches of Section III or their "statistical" counterparts in Section IV? Our recommendation depends on SNR versus complexity tradeoff's. When the SNR is high, the deterministic solution in Section III-C should be preferred, and the minimal number of blocks $N = P$ should be used for computational simplicity. However, at low SNR, N should be chosen large enough to obtain reliable estimates $\hat{\boldsymbol{\mathcal{R}}}_{yy}^{(N)}$ in (65).

Remark 2: Our blind channel estimation and direct equalization algorithms in Sections III and IV end up solving eigenvalue problems of the form $\boldsymbol{A\theta} = \boldsymbol{0}$, which can be turned into solving systems of linear equations like $\hat{\boldsymbol{A}}\hat{\boldsymbol{\theta}} = -\boldsymbol{a}_1$ after fixing the first entry of $\boldsymbol{\theta}$ to unity. Although it is beyond the size and scope of this paper, it is evident that adaptive variants of our batch algorithms are possible. In the context of OFDM, an adaptive LMS solution to direct equalization was reported recently in [6]. Notice, however, that with our multirate framework, both LMS and RLS channel estimators and equalizers are possible for general transmission schemes falling under the umbrella of redundant filterbank precoding.

V. Simulations—Comparisons

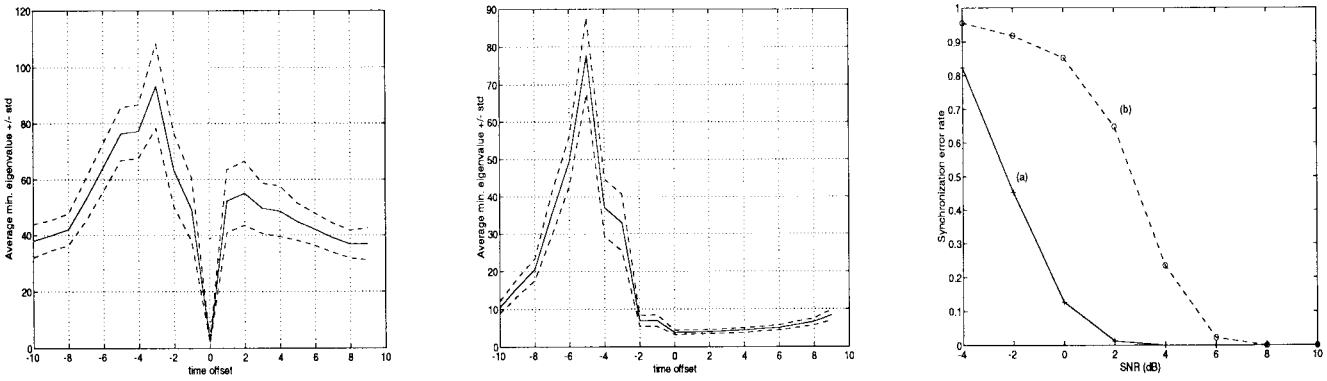
In this section, we test our proposed algorithms and compare their performance with existing input- and output-diversity methods. To make a fair comparison with methods not using any precoding, we have normalized the precoder matrix $\boldsymbol{F}$ to be unitary. In all examples, the filters of the precoder are $f_m(n) = (1/\sqrt{M})\exp(j2\pi mn/M)$; see also [24] for different precoder choices.

Example 1—Blind Synchronization: Using the transmission scheme shown in Fig. 1, with $(P, M, L) = (20, 16, 4)$ and $d_0 = 0$, we have estimated average value and standard deviation of $\lambda(d) := \lambda_{\min}(\boldsymbol{\mathcal{X}}_N^{(d)}\boldsymbol{\mathcal{X}}_N^{(d)H})$ as a function of d, using 100 independent realizations. In general, the behavior of $\lambda(d)$ depends on the channel. To analyze the impact of the channel transfer function on $\lambda(d)$, we have considered three typical classes corresponding to the following:

1) a minimum-phase channel transfer function with zeros $(0.8, -0.8, 0.5j, -0.5j)$;
2) a channel with unit circle zeros $(1, -1, 0.7j, -0.7j)$;
3) a nonminimum phase channel with zeros $(1.2, -1.2, 0.7j, -0.7j)$.

The results are shown in Fig. 2, where the theoretical delay value, obtained by substituting the matrices $\boldsymbol{\mathcal{X}}_N^{(d)}\boldsymbol{\mathcal{X}}_N^{(d)H}$ by their expected values (solid line), is reported together with the simulation results $(+)$. We observe a fairly good agreement between theory and simulation. Interestingly, peaks and valleys of $\lambda(d)$ for the minimum-phase channel appear reversed for the nonminimum-phase channel. It is important to observe that even if $\lambda(d)$ exhibits an almost flat behavior in the neighborhood of $d = 0$ for certain channels [see Fig. 2(a)], the absolute value of the gradient of $\lambda(d)$ around $d = 0$ is high in at least one direction. This consideration could be exploited to improve the performance of the method by incorporating the gradient into the estimation procedure, but this goes beyond the scope of this paper.

The noise effect on the synchronization procedure is illustrated in Fig. 3(a) and (b), where the average $\lambda(d)$ (solid line) is plotted together with its true value plus/minus the standard deviation (dashed-lines), as a function of d, for two different channels at SNR $= 6$ dB. It is important to notice that not only the minimum average value of $\lambda(d)$ is achieved for $d = 0$, as expected, but that the standard deviation assumes its minimum value at the same position as well.



(a) (b) (c)

Fig. 3. (a) Channel with zeros at (1.2, -0.9, $0.7j$, $-0.7j$). (b) Minimum-phase channel with zeros at (0.8, -0.8, $0.5j$, $-0.5j$) (average value: solid line, average value plus/minus the standard deviation: dashed line, SNR = 6 dB, $M = 16$, $N = P = 19$). (c) Synchronization error rate versus SNR (dB), $M = 8$, $L = 4$, and $N = P = 12$.

TABLE I
SET OF CHANNEL RESPONSES

$h_1(n)$	0	$-1.280 - j0.301$	$1.617 + j2.385$	$0.178 + j0.263$
$h_2(n)$	$-1.023 - j0.501$	$0.106 + j1.164$	$1.477 + j1.850$	$-0.482 - j0.523$
$h_3(n)$	0	$-0.282 + j0.562$	$0.371 - j1.001$	$0.041 - j0.110$
$h_4(n)$	$-0.227 + j0.487$	$0.031 - j0.211$	$0.336 - j0.866$	$-0.110 + j0.271$

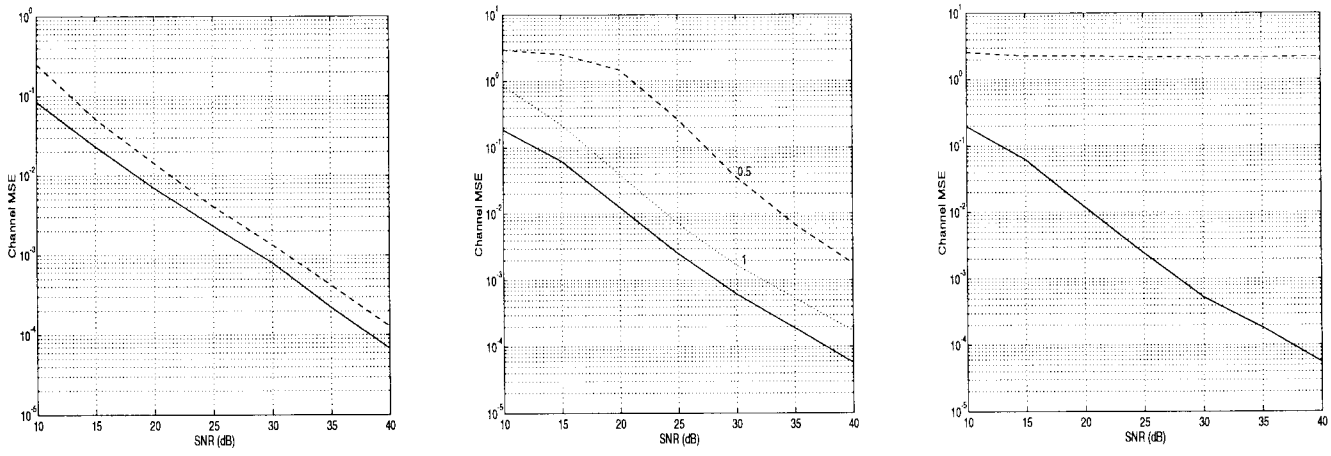
Performance of the proposed synchronization algorithm has been tested by simulating 400 independent trials. Fig. 3(c) shows the synchronization error rate as a function of the SNR for the two channels of Fig. 3(a) and (b). The system parameters were $(P = N, M, L) = (12, 8, 4)$. From Fig. 3(c), we observe that the synchronization procedure has better performance for channels characterized by more accentuated selectivity [case (a)]. This observation is also justified by looking at the behavior of $\lambda(d)$ shown in Fig. 3(a) and (b).

It is important to point out that the proposed deterministic blind synchronization method is able to provide satisfactory error rates also at low SNR with relatively short data records. Such a performance makes it an attractive choice for synchronizing block (e.g., multicarrier) transmissions through rapidly fading channels.

Example 2—Blind Channel Estimation versus [33]: We compare now our blind channel estimation method with the deterministic method proposed in [33], which is denoted hereafter as the XLTK method. The XLTK method has been implemented for the two basic configurations of single-input multiple-output (SIMO) systems obtained using either space diversity (antenna array) or fractional sampling at the receiver.

In the first case, we used four antennas, with the same four channels of [33, Table II], reported in Table I for convenience. For our algorithm, we used only one antenna, with channel impulse response given by the second column of Table I. We chose $M = 7$ and $P = N = 10$. Fig. 4(a) shows the mean square error (MSE) obtained in the blind channel estimation using our approach from (12) (solid line) and the XLTK method (dashed line) at various SNR levels. In both cases, 100 data samples were used for estimation. The advantage of the proposed approach is evident, although we used only one antenna instead of the four-sensor array used in [33]. This brings out that introducing redundancy at the transmitter improves the estimation performance, with respect to methods relying upon output redundancy, whereas at the same time, the receiver complexity is reduced. Moreover, although the algorithms use the same amount of data, the XLTK algorithm is heavier from a computational point of view. The main cost associated with our blind channel estimation procedure is, in fact, that related to the SVD of the $P \times P$ matrix $\boldsymbol{X}_N \boldsymbol{X}_N^H$, which is necessary to obtain the L vectors $\tilde{\boldsymbol{U}}_l$. Then, we have to solve an homogeneous linear system of $L + 1$ equations. It is important to remark that our method is robust against channel order mismatching, provided that we use an overestimate. Conversely, the XLTK method is particularly sensitive to channel order mismatching, and then, it needs a channel order estimate at the beginning (e.g., see [33, p. 298]). This requires the SVD of matrices, whose dimension is $(N - L_e + 1)Q(Q - 1)/2 \times (L_e + 1)Q$, where Q is the number of sensors in its array-based implementation, or the oversampling factor L_e is the channel order overestimate, and N is the number of information symbols. Therefore, assuming $P = N$, i.e., the same number of symbols for both systems, we see that for $Q > 2$, the XLTK method needs to compute the SVD of larger matrices, and this is indeed the case considered in our Example 2, where $Q = 4$. Furthermore, the number of unknowns with the XLTK method increases proportionally to Q.

Next, we compare our method with the XLTK algorithm using a single antenna at the receiver side and two channels obtained by fractional sampling. In Fig. 4(b), the two different techniques are analyzed by introducing the same amount of redundancy at the transmitter in the form of excess-bandwidth for the XLTK method, controlled by the roll-off factor of the raised cosine impulse response, and in the form of trailing zeros for the proposed strategy. With rolloff equal to 0.5, the information rate is $R_0 = (2/3)B$, where B indicates



(a) (b) (c)

Fig. 4. (a) Mean square error of the blind estimation (XLTK-4 antennas: dashed line; SGB: solid line). (b) Proposed method (solid line); XLTK-fractional sampling with roll-off factor equal to 0.5 (dashed line) and 1 (dotted line). (c) MSE of proposed method (solid line) and XLTK-fractional sampling (dashed line), with channel order overestimated by two samples.

the available bandwidth as well as the maximum theoretical information rate. Clearly, it is possible to have the same information rate in our scheme by setting $M/(M+L) = 2/3$ and assuming an ideal Nyquist characteristic for the pulse shaping filter (i.e., roll-off factor equal to 0). The channel impulse response at sampling rate $2B$ has been computed as the convolution of the raised-cosine impulse response with roll-off 0.5 with the channel impulse response obtained by interleaving the four channels reported in Table I. The resulting impulse response has been truncated by setting to zero the values below 10% of the maximum. Denoting by $h(t)$ the continuous time impulse response, the resulting samples at rate $2B$ are $h(n/2B) = (-0.1426 - j0.0833, -0.2094 - j0.1063, -0.1824 - j0.0334, -0.2254 + j0.0661, -0.2914 + j0.1942, -0.2463 + j0.2464, -0.0880 + j0.1998, 0.2284 + j0.4155, 0.5926 + j0.8055, 0.6436 + j0.6226, 0.3663 - j0.0710, 0.1099 - j0.3807, -0.0171 - j0.1990, -0.0884 - j0.0644)$, and the two-channel impulse responses of the equivalent SIMO model are thus of order $L = 6$. Therefore, the order of the channel at the symbol-rate is also $L = 6$. In the implementation of our method, we have used the impulse response corresponding to the even samples of $h(t)$, i.e., $h(n/B) = (-0.1426 - j0.0833, -0.1824 - j0.0334, -0.2914 + j0.1942, -0.0880 + j0.1998, 0.5926 + j0.8055, 0.3663 - j0.0710, -0.0171 - j0.1990)$. Imposing the constraint $M/(M+6) = 2/3$, we have $M = 12$, $P = 18$, and thus, our method requires the minimum amount of data samples equal to $PN = P^2 = 324$. We have used the same amount of data to test the performance of both methods, again, given in terms of mean square error of the channel impulse response estimates. Moreover, we have simulated the XLTK method using rolloff equal to 1, which simply means double bandwidth with respect to the information rate R_0. The channel samples are $h(n/2B) = (-0.3550 - j0.1739, -0.4442 - j0.1045, -0.0979 + j0.1950, 0.5126 + j0.6420, 0.1166 - j0.3005, -0.1673 - j0.1815, -0.0788 + j0.1690, 0.0368 + j0.4040, 0.5612 + j0.8277, 0.1288 - j0.3474, 0.0618 + j0.0913, -0.0382 + j0.0940)$, and the subchannel is of order $L = 5$.

From Fig. 4(b) we observe that the XLTK algorithm suffers from a threshold effect at low SNR and that the proposed method outperforms the XLTK algorithm as well, when a rolloff factor equal to 1 is used in the XLTK implementation, indicating that the structured redundancy introduced at the transmitter by the filterbank leads to better performance.

One more important remark about the comparison between our method and XLTK algorithm concerns the effect of channel order overestimation on the performance. More specifically, we plot, in Fig. 4(c), the MSE on the channel estimation obtained with our method (solid line) and XLTK method (dashed line) when the channel order has been overestimated by two samples. We observe that our method is robust to order overestimation, whereas the performance of XLTK depends strongly on knowing the exact channel order. The channel used for the XLTK method is the same as in Fig. 4(b), with rolloff factor equal to 1. In our case, the system parameters are $M = 12$, and the same channel of order $L = 6$ as in Fig. 4(b) was used, but in this case, $P = M + L + 2$, and $N = P$. This feature is particularly important for applications, where only an upper bound on the channel order is available.

Example 3—Direct Blind Equalization: Here, we compared our direct blind equalizer with the constant modulus algorithm (CMA) [16] and with the subspace method proposed in [15], which will be referred to as the GT method (see also [13] for an independent derivation of the same method).

First, we simulated $N = 19$ blocks of 8-PSK symbols, each block with $P = 19$ ($M = 16$ information symbols and $L = 3$ TZ's). The third-order channel had zeros at 0.9, j, $-j$, and the SNR at the receiver was fixed at 25 dB. With this relatively low SNR, we implemented the deterministic direct equalizer approach summarized in Theorem 3 using the minimal number of samples $P^2 = 361$. To compare with CMA, Fig. 5 shows scattering diagrams obtained in the following cases:

a) without equalization;
b) with the ZF equalizer of Theorem 1;
c) with the MMSE equalizer of (2);
d) with the CMA.

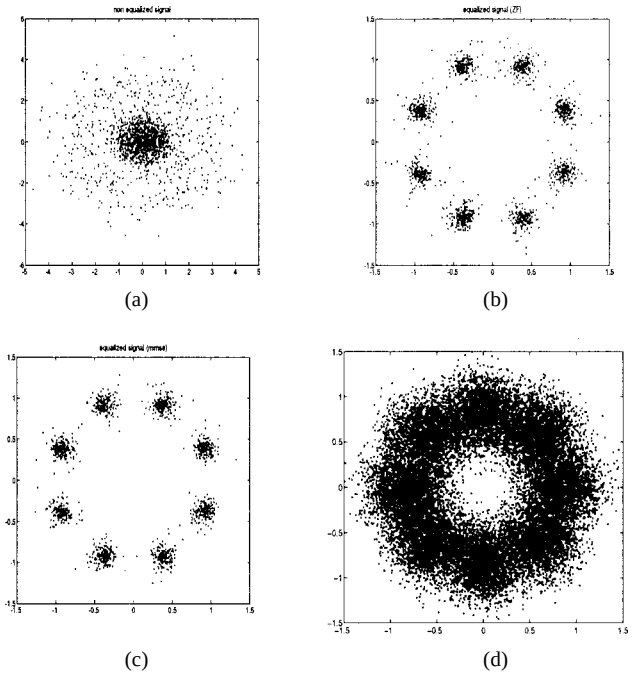


Fig. 5. Scattering diagrams. (a) Before equalization. (b) After direct ZF equalization. (c) After MMSE equalization (SNR = 25 dB). (d) After equalization based on the CMA (SNR = ∞, equalizer length 30 and step size 0.56×10^{-4}).

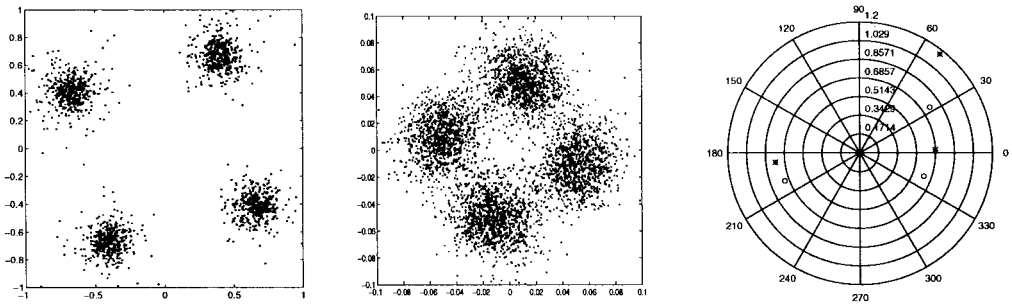
For the CMA, we used a 30-tap equalizer, and the adaptation rule was run over $N = 20\,000$ samples at SNR $= \infty$. In this situation, CMA suffers from the zeros on the unit circle so that, even in the absence of noise, there is no convergence toward the correct equalizer. Conversely, in our case, the convergence is guaranteed, and the performance is satisfactory, even with a relatively small amount of data. One more important remark is that CMA requires a prohibitively increasing number of samples for high order constellations, whereas the number of data required for our method to converge is *independent* of the constellation size.

Next, we compare our method with the method of [15] and [13] using space diversity built with a two-sensor array. The two impulse responses are $\boldsymbol{h}_1 = (0.5956,\ -0.3843 - j0.5020,\ -0.3145 - j0.1178,\ 0.1838 + j0.3148)$ and $\boldsymbol{h}_2 = (-0.7619 - j0.1887,\ 0.4166 + j0.0618,\ 0.1797 + j0.3201,\ -0.1815 - j0.1970)$. In our case, we have used only the second channel. The SNR was fixed at 20 dB in both cases. The system parameters were $(M, L, N) = (12, 3, 15)$. The scattering diagrams for a QPSK transmission obtained with both methods are reported in Fig. 6. Indeed, the performance of [13] and [15] depends strongly on the closeness of the zeros of the two channels. In the present case, the zeros corresponding to the two channels are located in the positions depicted in Fig. 6(c) ("$*$" refers to $\boldsymbol{h}_1$ and "o" to $\boldsymbol{h}_2$). From Fig. 6, we observe that our method provides better performance than [13] and [15], despite the use of only one sensor instead of two, even if the zeros of the two channels are not excessively close.

Example 4—Direct Blind Equalization versus Blind Channel Estimation: In Section III, we have proposed a blind channel estimation method and a direct equalization method. Clearly, in the first case, the channel estimate can also be used to derive the equalizer coefficients, according to (10), using the estimate $\hat{\boldsymbol{H}}$ instead of $\boldsymbol{H}$. In this section, we compare these two approaches. Since the performance of the blind direct equalizer strongly depends on the channel zero locations, we provide results averaged over several independent channels. More specifically, we considered the so-called "Vehicular A" model, which was adopted for the wideband CDMA in UMTS to test its performance [10]. The channel model is

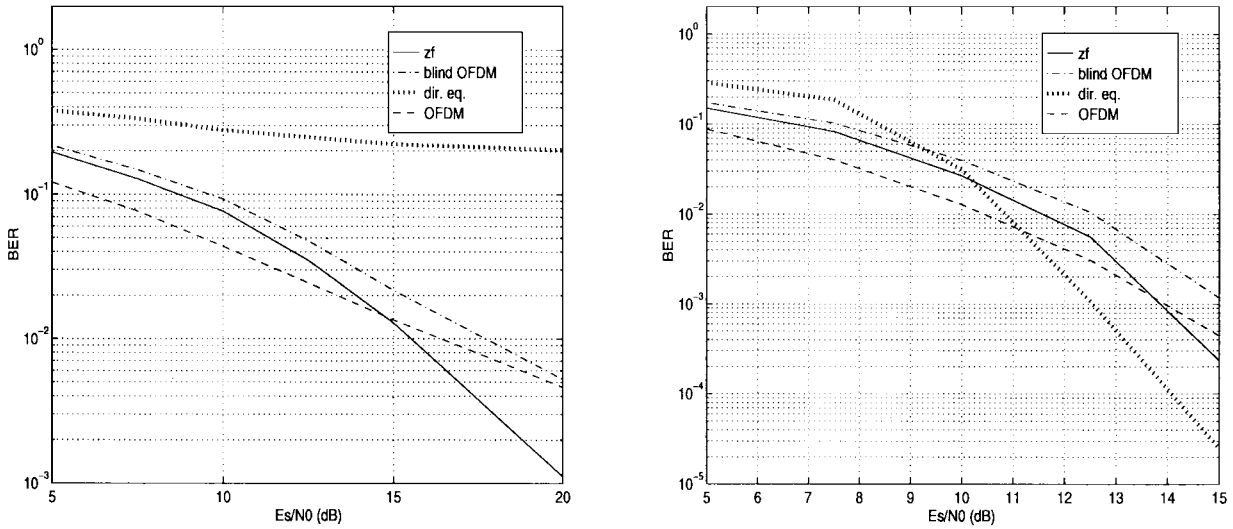
$$h(n) = \sum_{l=0}^{9} h_l g(n - \tau_l) \tag{66}$$

where the rays amplitudes h_l are independent complex Gaussian random variables with zero mean and variances σ_l^2 (expressed as 10 $\log_{10}\ \sigma_l^2$): [0, -2.5, -6.5, -9.5, -12.5, -13, -15.5, -25.5, -50, -21.5, -25.5]; $g(n)$ is a Nyquist pulse. The delays are [0, 0.25, 0.5, 0.75, 1, 1.2, 1.7, 1.9, 2.4, 2.7] μs, and the sampling rate is 4.096 Mcps. In Fig. 7(a), we report the BER versus the ratio E_s/N_0. The symbol constellation is QPSK, and the BER has been averaged over 100 independent channels. Specifically, in Fig. 7(a), the solid line refers to the ZF equalizer, where the data are equalized by $\hat{\boldsymbol{H}}^{\dagger}$, where $\hat{\boldsymbol{H}}$ is estimated by using our blind method; dashed and dotted lines refer to an OFDM scheme, where the FFT outputs are divided by the corresponding channel transfer function values estimated using our method; the dashed line refers again to OFDM, but in the ideal situation where the channel is supposed perfectly known at the receiver; finally, a dotted line refers to the blind direct equalizer. Both OFDM schemes insert cyclic prefixes to simplify the equalization. From Fig. 7(a) we notice that at high SNR, the blind method based on the channel identification outperforms all other methods, including the ideal OFDM scheme that assumes perfect knowledge of the channel. In fact, if some channels have zeros close to or on the unit circle, some of the OFDM symbols cannot be recovered reliably. This explains the smaller slope of the BER curve, at high SNR, with respect to our blind ZF method based on channel estimation, which, thanks to the trailing zeros, does not experience such a problem. Conversely, the blind method based on direct equalization presents the worst performance, and the reason for this is that some of the independent channel realizations are nonminimum phase, and thus, the delay d should be properly chosen, according to the theory developed in Section III-B, whereas we assumed $d = 0$ for all channel realizations. In Fig. 7(b), we tested the previous equalization algorithms against Rice fading channels, i.e., channels with a line-of-sight path. In particular, with reference to (66), we assumed $|h_0|^2 = 10 \sum_{l=1}^{9} |h_l|^2$, as in [9, p. 43] for the fixed channel, for each realization, whereas the statistics of the l path, with $l = 1, \cdots, 9$, are the same as before. As we can see, the performance of the direct equalizer improves considerably with respect to the Rayleigh fading channel case. Indeed, at high SNR, the direct equalizer is the method exhibiting the best performance. This behavior is due to the fact that with the Rice model and the same power profile as before, the probability of incurring into a nonminimum phase channel decreases, and thus, the choice $d = 0$ is optimal in most cases.



(a) (b) (c)

Fig. 6. Scattering diagrams obtained with (a) this paper's method and (b) the GT method.



(a) (b)

Fig. 7. Average BER versus E_s/N_0 using blind channel estimation (solid line), direct equalization (dotted line), OFDM with null guard time interval and channel perfectly known (dashed line) or estimated via our blind method (dashed and dotted line). (a) BER averaged over 100 independent Rayleigh fading channels. (b) BER averaged over 100 independent Rice fading channels.

Example 5—Downlink CDMA: We have simulated the downlink channel for a CDMA system, where the user codes are the Walsh–Hadamard codes, and we compared the performance of our blind methods with a RAKE receiver in Fig. 8. The curves report the BER averaged over 100 independent Rayleigh fading channels, assuming the same channel model as in Fig. 7. Specifically, the solid line refers to our blind estimation method, the dotted line refers to the direct equalizer, and the thick solid line refers to a RAKE receiver that assumes perfect knowledge of the channel impulse response. We can observe the superior performance of the method based on the channel estimation over the RAKE receiver at high SNR, i.e., when the multiuser interference is the dominating disturbance, whereas the poor performance of the direct equalizer is only due to the lack of optimization with respect to the delay.

VI. Conclusions

Building on the general precoding framework established in [24], we have proposed blind deterministic methods for channel identification, direct equalization, and synchronization. Comparing the performance with alternative procedures that introduce redundancy at the receiver in the form of space diversity or fractional sampling, we have shown that introducing redundancy at the transmitter offers distinct advantages.

Although the proposed methods have been derived for time-invariant channels, their deterministic structure also makes them suitable for application over time-varying channels, provided that the channel's coherence time is greater than the time interval necessary to transmit N blocks of data, where N is uniquely determined from the length of the precoding filters and the channel memory.

As far as *a priori* knowledge of the channel memory is concerned, it is important to remark that the proposed method is robust even when the channel order is overestimated. Since the maximum delay due to multipath propagation is generally known, depending on the carrier frequency, bandwidth, and application (e.g., indoor versus outdoor), the assumption about the upper bound on the order of the discrete-time equivalent channel is perfectly reasonable in practice.

Other important features of the proposed procedures are the lack of constraints on channel zero locations and the simplicity of the receiver, relative to existing blind approaches which

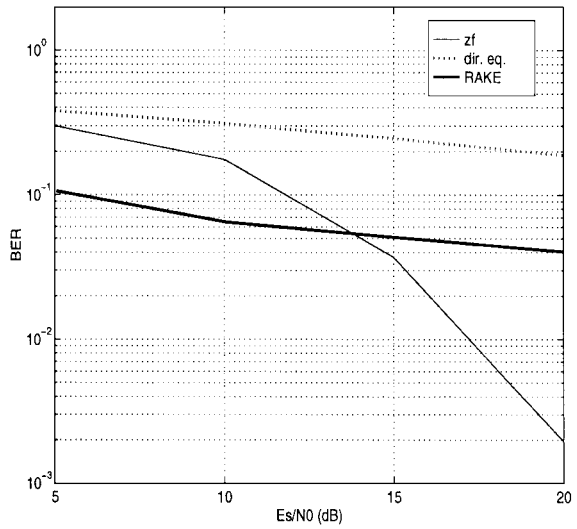


Fig. 8. Average BER versus E_s/N_0 for the downlink channel of a CDMA system employing Walsh–Hadamard codes using blind channel estimation (solid line), direct equalization (dotted line), and a RAKE receiver assuming perfect knowledge of the channel (thick solid line). The BER is averaged over 100 independent Rayleigh fading channels.

require multiple sensors or fractional sampling that assumes excess bandwidth.

REFERENCES

[1] A. Akansu, M. Tazebay, and R. Haddad, "A new look at digital orthogonal transmultiplexers for CDMA communications," *IEEE Trans. Signal Processing*, vol. 45, pp. 263–267, Jan. 1997.

[2] S. Benedetto, E. Biglieri, and V. Castellani, *Digital Transmission Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[3] A. Chevreuil and P. Loubaton, "Blind second-order identification of FIR channels: Forced cyclo-stationarity and structured subspace method," *IEEE Signal Processing Lett.*, vol. 4, pp. 204–206, July 1997.

[4] J. S. Chow, J. C. Tu, and J. M. Cioffi, "A discrete multitone transceiver system for ADSL applications," *IEEE J. Select. Areas Commun.*, vol. 9, pp. 895–908, 1991.

[5] L. J. Cimini, "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiple access," *IEEE Trans. Commun.*, vol. COMM-30, pp. 665–675, July 1985.

[6] M. de Courville, P. Duhamel, P. Madec, and J. Palicot, "A least mean-squares blind equalization technique for OFDM systems," *Annal. Telecom.*, pp. 4–11, 1997.

[7] B. W. Dickinson, "Efficient solution of linear equations with banded Toeplitz matrices," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP–27, pp. 421–423, Apr. 1979.

[8] Euro. Telecommun. Stand., "Radio broadcast systems: Digital audio broadcasting (DAB) to mobile, portable and fixed receivers," ETS 300 401, Mar. 1994.

[9] ———, "Radio broadcast systems for television, sound and data services: Framing structure, channel coding and modulation for digital terrestrial television," ETS 300 744, Mar. 1997.

[10] ———, "Wideband direct-sequence CDMA (WCDMA)—Evaluation document (3.0)," ETSI SMG, Meeting 24, Madrid, Spain, Dec. 15–19, 1997.

[11] K. Fazel and G. P. Fettweis, Eds., *Multi-Carrier Spread Spectrum.* Boston, MA: Kluwer, 1997.

[12] D. T. Gavel, "Solution to the problem of instability in banded Toeplitz solvers," *IEEE Trans. Signal Processing*, vol. 40, pp. 464–466, Feb. 1992.

[13] D. Gesbert, P. Duhamel, and S. Mayrargue, "On-line blind multichannel equalization based on mutually referenced filters," *IEEE Trans. Signal Processing*, vol. 45, pp. 2307–2317, Sept. 1997.

[14] G. B. Giannakis, "Filterbanks for blind channel identification and equalization," *IEEE Signal Processing Lett.*, vol. 3, pp. 184–187, June 1997.

[15] G. B. Giannakis and C. Tepedelenlioglu, "Direct blind equalizers of multiple FIR channels: A deterministic approach," *IEEE Trans. Signal Processing*, vol. 47, pp. 62–74, Jan. 1999.

[16] D. N. Godard, "Self-recovering equalization and carrier-tracking in two-dimensional data communication system," *IEEE Trans. Commun.*, vol. COMM-28, pp. 1867–1875, 1980.

[17] G. H. Golub and C. F. Van Loan, *Matrix Computation.* Baltimore, MD: Johns Hopkins Univ. Press, 1989, pp. 79–81.

[18] A. Gorokhov and P. Loubaton, "Semi-blind second-order identification of convolutive channels," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Munich, Germany, 1997, vol. V.

[19] S. D. Halford and G. B. Giannakis, "Blind equalization of TDMA wireless channels exploiting guard time induced cyclostationarity," in *Proc. First IEEE Signal Process. Workshop. Wireless Commun.*, Paris, France, Apr. 16–18, 1997, pp. 117–120.

[20] A. Erdogan, B. Hassibi, and T. Kailath "H-infinity equalization of communication channels," in *Proc. ICASSP*, Seattle, WA, May 1998, vol. VI, pp. 3433–3437.

[21] L. Lupas, "On the computation of generalized Vandermonde matrix inverse," *IEEE Trans. Automat. Contr.*, vol. 20, pp. 559–561, 1975.

[22] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

[23] S. D. Sandberg and M. A. Tzannes, "Overlapped discrete multitone modulation for high speed copper wire communications" *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1571–1585, 1995.

[24] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers Part I: Unification and optimal designs," *IEEE Trans. Signal Processing*, this issue, pp. 1988–2008; see also *Proc. 35th Annu. Allerton Conf.*, Monticello, IL, Sept. 1997, vol. II, pp. 1134–1137.

[25] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Inverting overdetermined Toeptlitz matrices with application to channel equalization in block transmission systems," in *Proc. 32nd Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1998.

[26] E. Serpedin and G. B. Giannakis, "Blind channel identification and equalization using modulation induced cyclostationarity," *IEEE Trans. Signal Processing*, vol. 46, pp. 1930–1944, July 1998.

[27] M. K. Tsatsanis and G. B. Giannakis, "Transmitter induced cyclostationarity for blind channel equalization," *IEEE Trans. Signal Processing*, vol. 45, pp. 1785–1794, July 1997.

[28] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, 1994.

[29] J. J. van de Beek, M. Sandell, and P. O. Börjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 1800–1805, July 1997.

[30] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency division multiplexing using the discrete Fourier transform," *IEEE Trans. Commun. Technol.*, vol. COMM-19, pp. 628–634, Oct. 1971.

[31] G. Wornell, "Emerging applications of multirate signal processing and wavelets in digital communications," *Proc. IEEE*, vol. 84, pp. 586–603, Apr. 1996.

[32] X.-G. Xia, "New precoding for ISI cancellation using nonmaximally decimated multirate filterbanks with ideal FIR equalizers," *IEEE Trans. Signal Processing*, vol. 45, pp. 2431–2441, Oct. 1997.

[33] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.

Anna Scaglione (S'97), for photograph and biography, see this issue, p. 2005.

Georgios B. Giannakis (F'96), for photograph and biography, see this issue, p. 2006.

Sergio Barbarossa (M'88), for photograph and biography, see this issue, p. 2006.