

# Performance Analysis of a Deterministic Channel Estimator for Block Transmission Systems With Null Guard Intervals

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

**Abstract**—A deterministic algorithm was recently proposed for channel identification in block communication systems. The method assumed that the channel is finite impulse response (FIR) and that null guard intervals of length greater than the channel order are inserted between successive blocks to prevent interblock interference and allow block synchronization. In the absence of noise, the algorithm provides error-free channel estimates, using a finite number of received data, without requiring training sequences and without imposing a restriction neither on the channel, except for finite order and time invariance, nor on the symbol constellation. Using small perturbation analysis, in this paper, we derive approximate expressions of the estimated channel covariance matrix, which are used to quantify the resilience of the estimation algorithm to additive noise and channel fluctuations. Specifically, we consider channel fluctuations induced by transmitter/receiver relative motion, asynchronism, and oscillators' phase noise. We also compare the channel estimation accuracy with the Cramér–Rao bound (CRB) and prove that our estimation method is statistically efficient at practical SNR values for any data block length. Finally, we validate our theoretical analysis with simulations and compare our transmission scheme with an alternative system using training sequences for channel estimation.

**Index Terms**—Channel estimation, equalization, theoretical bounds.

## I. INTRODUCTION

**B**LOCK transmission with the insertion of guard intervals between blocks is a commonly adopted strategy to prevent interblock interference, allow block synchronization, and facilitate channel equalization. Null guard intervals are inserted, for example, at the beginning of the frame in the digital audio broadcasting system for synchronization purposes [4], whereas guard intervals in the form of cyclic prefixes are commonly used in orthogonal frequency division multiplexing (OFDM) systems [4], [5] to convert linear convolution with the channel into cyclic convolution and then facilitate channel equalization. Since the performance of an OFDM system is highly penalized by deep null fades in the channel frequency response, which prevent zero forcing (ZF) equalization, OFDM is always equipped

with error correction coding before the fast Fourier transform (FFT), but this clearly comes at the expense of a transmission rate. However, in [12], we showed that if instead of using cyclic prefixes, we use null guard intervals of the same length as the cyclic prefix, perfect ZF equalization is possible, regardless of the channel nulls. This would thus alleviate the need for channel error correction coding. Furthermore, exploiting the only presence of the null guard intervals, in [13], we proposed a method for channel identification, synchronization, and direct equalization that is able to provide an error-free channel estimate, in the absence of noise, using a finite number of received symbols. Specifically, in [13], each block of  $M$  information symbols  $\mathbf{s}(n)$  was linearly mapped onto blocks of  $P$  coded data  $\mathbf{u}(n)$ , with  $P > M$ , by multiplying  $\mathbf{s}(n)$  with a  $P \times M$  precoding matrix  $\bar{\mathbf{F}}$ . The method of [13] was based on the following assumptions.

- a0) Channel  $h(l)$  is a linear time-invariant  $L$ th-order FIR filter.
- a1) For a given  $L$ , the pair  $(P, M)$  is chosen to satisfy  $P > M > L$  and  $P = M + L$ .
- a2) Precoder matrix  $\bar{\mathbf{F}}$  has  $L$  trailing zeros, i.e.,  $\bar{\mathbf{F}}^T = (\mathbf{F}^T \mathbf{0}^T)$ , where  $\mathbf{F}$  is a full-rank  $M \times M$  matrix.

The full rank of  $\mathbf{F}$  guarantees one-to-one mapping and, thus, recovery of  $\mathbf{s}(n)$  from the coded symbols  $\mathbf{u}(n) = \mathbf{F}\mathbf{s}(n)$ . Convolution with  $h(l)$  causes interference between successive symbols, but the insertion of the zero guard interval in a2) avoids interblock interference and guarantees channel identifiability and invertibility, irrespective of the channel nulls [12]. Compared with other deterministic methods [18], the algorithm of [13] does not impose any restrictive constraint on the channel nulls. Furthermore, as opposed to methods based on second- or higher order statistics, it is capable of providing error-free estimates, in the absence of noise, without requiring transmission of training sequences and using a finite number of data. This last feature makes it particularly appealing for mobile communications because the method can tolerate slow channel fluctuations provided that the duration of the block used for channel estimation is smaller than the channel coherence time [11].

To assess applicability of the algorithm proposed in [13] in a real mobile communication scenario, in this paper, we quantify its resilience to additive noise and to channel fluctuations induced by relative transmitter/receiver motion, carrier asynchronism, and oscillators' phase noise. Specifically, we use small perturbation analysis to derive the channel estimation covariance matrix in closed form. As a benchmark to measure the relative accuracy of the proposed estimation algorithm, we derive a closed-form expression for the CRB, assuming AWGN

Manuscript received August 19, 1999; revised October 9, 2001. This work was supported in part by the National Science Foundation Wireless Initiative Grant 9979443. The associate editor coordinating the review of this paper and approving it for publication was Prof. Michail K. Tsatsanis.

S. Barbarossa is with the INFOCOM Department, University of Rome "La Sapienza," Rome, Italy (e-mail: sergio@infocom.uniroma1.it).

A. Scaglione is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA.

G. B. Giannakis is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA.

Publisher Item Identifier S 1053-587X(02)01346-6.

and modeling the unknown transmitted symbols as deterministic (nuisance) parameters, and we prove that the channel estimation algorithm is *statistically efficient* at high SNR for any data block length. The theoretical analysis is finally validated with numerical results.

Even though null symbols cannot be assumed to be a training sequence because no channel estimate could be possible with a null training sequence, the introduction of guard zeros could be interpreted as a sort of distributed training. Thus, we compare the proposed transmission system with the conventional scheme using training preambles and the Wiener equalizer, assuming the same ratio between the number of information symbols and of extra symbols (zeros or training sequence).

The paper is organized as follows. In Section II, we briefly review the estimation algorithm of [13] and reveal two features of the algorithm that were not emphasized in [13], namely, its capability to cope with aperiodic precoding and with carrier offsets. In Section III, we derive the theoretical performance analysis using a small perturbation method. In Section IV, we compute the CRB and prove the statistical efficiency of the proposed method. In Section V, we compare the transmission strategy based on null guard intervals and our proposed channel estimation algorithm with an alternative system based on the transmission of training sequences and Wiener equalization. Finally, in Section VI, we provide several numerical results to validate our theoretical findings under a variety of channel conditions and coding strategies.

## II. CHANNEL ESTIMATION ALGORITHM

Based on a0)–a2), the received block data model is [13]

$$\mathbf{y}(n) = \mathbf{x}(n) + \mathbf{v}(n) = \mathbf{H}\mathbf{F}\mathbf{s}(n) + \mathbf{v}(n) \quad (1)$$

where  $\mathbf{v}(n)$  is AGWN vector with pdf  $\sim \mathcal{N}(\mathbf{0}, \sigma_v^2 \mathbf{I})$ ,  $\mathbf{F}$  is an  $M \times M$  full-rank precoding matrix, which could represent, for example, the IFFT matrix in an OFDM system or the matrix of spreading codes in the downlink of a code division multiple access (CDMA) system;  $\mathbf{H}$  is the  $P \times M$  channel Toeplitz matrix

$$\mathbf{H} := \mathcal{T}(\mathbf{h}) := \begin{pmatrix} h(0) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ h(L) & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & h(0) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ddots & 0 & h(L) \end{pmatrix} \quad (2)$$

where the symbol  $\mathcal{T}(\mathbf{h})$  is used to denote the Toeplitz structure of  $\mathbf{H}$ . Furthermore, we assume that the input symbol sequence satisfies a persistence of excitation condition, namely, the following.

**a3)** There exists an  $N \geq P$  such that the  $M \times N$  matrix  $\mathbf{S}_N := (\mathbf{s}(0), \dots, \mathbf{s}(N-1))$  is full row rank.

As  $N$  increases, matrix  $(1/N)\mathbf{S}_N\mathbf{S}_N^H$  tends to the input correlation matrix  $\mathbf{R}_{ss}$ , but a3) is satisfied even for colored (e.g., coded) inputs provided that their spectra are nonzero for at least  $M$  frequencies (modes).

Collecting  $N$  data vectors  $\{\mathbf{x}(n)\}_{n=0}^{N-1}$  from (1) in a matrix, we arrive at

$$\mathbf{X}_N := (\mathbf{x}(0) \cdots \mathbf{x}(N-1)) = \mathbf{H}\mathbf{F}\mathbf{S}_N \quad (3)$$

where  $\mathbf{S}_N$  is defined as in a3). Because of the structure (2), the matrix  $\mathbf{H}$  is always full column rank, i.e.,  $\text{rank}(\mathbf{H}) = M$ . This, along with a2) and a3), implies that  $\text{rank}(\mathbf{X}_N) = M$  and the nullity of  $\mathbf{X}_N$  is  $\nu(\mathbf{X}_N\mathbf{X}_N^H) = P - M = L$ . Hence, the SVD of  $\mathbf{X}_N$  assumes the form

$$\mathbf{X}_N = (\mathbf{U}\mathbf{\tilde{U}}) \begin{pmatrix} \Sigma_{M \times M} & \mathbf{0}_{M \times (N-M)} \\ \mathbf{0}_{L \times M} & \mathbf{0}_{L \times (N-M)} \end{pmatrix} \begin{pmatrix} \mathbf{\tilde{V}}^H \\ \mathbf{V}^H \end{pmatrix} \quad (4)$$

and the  $L$  columns of the  $P \times L$  matrix  $\mathbf{\tilde{U}}$  span the nullspace  $\mathcal{N}(\mathbf{X}_N)$ . Since  $\mathbf{F}\mathbf{S}_N$  in (3) is full rank,  $\mathcal{R}(\mathbf{X}_N) = \mathcal{R}(\mathbf{H})$ , where  $\mathcal{R}$  stands for range space, but since  $\mathcal{R}(\mathbf{X}_N)$  is orthogonal to  $\mathcal{N}(\mathbf{X}_N)$ , it follows that

$$\mathbf{\tilde{U}}^H \mathbf{X}_N = \mathbf{0} \Rightarrow \mathbf{\tilde{U}}^H \mathbf{H} = \mathbf{0} \Rightarrow \tilde{\mathbf{u}}_l^H \mathcal{T}(\mathbf{h}) = \mathbf{0}^H, \quad l = 1, \dots, L \quad (5)$$

where  $\tilde{\mathbf{u}}_l$  denotes the  $l$ th column of  $\mathbf{\tilde{U}}$ . Vector multiplication with a Toeplitz matrix denotes convolution that is commutative, and thus, (5) can be written as

$$\mathbf{h}^H \tilde{\mathbf{u}} := \mathbf{h}^H (\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_L) = \mathbf{0}^H \quad (6)$$

where  $\tilde{\mathbf{u}}_l$  denotes the  $(L+1) \times M$  Hankel matrix associated with the vector  $\tilde{\mathbf{u}}_l$

$$\tilde{\mathbf{u}}_l := \begin{pmatrix} \tilde{u}_l(1) & \tilde{u}_l(2) & \cdots & \tilde{u}_l(M) \\ \tilde{u}_l(2) & \tilde{u}_l(3) & \cdots & \tilde{u}_l(M+1) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{u}_l(L+1) & \tilde{u}_l(L+2) & \cdots & \tilde{u}_l(P) \end{pmatrix}. \quad (7)$$

Based on the previous assumptions and on (6), the channel identification method is summarized in the following (see [13] for the proof).

*Theorem 1:* Starting from the data matrix  $\mathbf{X}_N$ , let us form the  $(L+1) \times ML$  matrix  $\tilde{\mathbf{U}}$  as in (3)–(6). If a0)–a3) hold true, then the matrix  $\tilde{\mathbf{U}}$  has nullity one, and the channel vector  $\mathbf{h}^H$  can be obtained as the unique (within a scale) null eigen-vector of  $\tilde{\mathbf{U}}$  in (6).

Interestingly, even though the method was derived in [13] for time-invariant channels and precoders, its validity holds true even in the presence of i) nonconstant, or aperiodic, precoding or ii) carrier offsets. Indeed, since full rank is the only property of the matrix  $\mathbf{F}\mathbf{S}_N$  used in Theorem 1, the resulting channel estimation method also applies to systems using precoding matrices  $\mathbf{F}_i$  varying as a function of the block index  $i$ . This generalization is particularly important for the downlink channel of CDMA systems that rely on scrambling to increase their immunity against intercell interference [10] or for orthogonal frequency division multiple access (OFDMA) systems adopting frequency hopping from block to block.

Furthermore, the method provides reliable channel estimates even in the presence of carrier offsets due to lack of synchronism or Doppler effects (when all the paths exhibit the same Doppler shift). Denoting by  $u(n)$  the transmitted coded sequence [i.e.,

the entries of the coded vector  $\mathbf{u}(n) := \mathbf{F}\mathbf{s}(n)$ , the channel output sequence  $x(n)$  in the presence of a carrier offset  $f_D$  is multiplied by  $\exp(j2\pi f_D n)$ , and the observed signal is

$$\begin{aligned} \tilde{x}(n) &:= e^{j2\pi f_D n} x(n) = \sum_{l=0}^L e^{j2\pi f_D(n+l-l)} h(l) u(n-l) \\ &= \sum_{l=0}^L \tilde{h}(l) \tilde{u}(n-l) \end{aligned} \quad (8)$$

where  $\tilde{h}(l) := \exp(j2\pi f_D l) h(l)$ , and  $\tilde{u}(n) := \exp(j2\pi f_D n) \cdot u(n)$ . If  $u(n)$  satisfies a3), so does  $\tilde{u}(n)$ . Therefore, the method described in Theorem 1 is still able to provide, in the absence of noise, an error-free estimate of  $\tilde{h}(l)$  that differs from the LTI channel response  $h(l)$  only by a frequency shift. In matrix form, the  $n$ th received block in the presence of a frequency offset  $f_D$  is

$$\tilde{\mathbf{x}}(n) = \mathbf{\Omega}_P(n) \mathbf{H} \mathbf{F} \mathbf{s}(n) = \tilde{\mathbf{H}} \mathbf{\Omega}_M(n) \mathbf{F} \mathbf{s}(n) \quad (9)$$

where  $\mathbf{\Omega}_K$  is the  $K \times K$  diagonal matrix  $\mathbf{\Omega}_K = \text{diag}(1, \exp(j2\pi f_D), \dots, \exp(j2\pi f_D(K-1)))$ , with  $K = P$  or  $K = M$ , and  $\tilde{\mathbf{H}}$  is the Toeplitz matrix associated with the channel response  $\tilde{h}(l)$ . In the absence of noise, the method is able to yield an error-free estimate of  $\tilde{\mathbf{H}}$ . Depending on whether training sequences are transmitted or not,  $f_D$  and  $\mathbf{\Omega}_P$  can thus be estimated, using, for example, maximum likelihood or blind methods that exploit *a priori* information about the symbol source, as in [2]. The symbols are then recovered as

$$\hat{\mathbf{s}}(n) = \mathbf{F}^{-1} \hat{\mathbf{\Omega}}_P^H \tilde{\mathbf{H}}^\dagger \tilde{\mathbf{x}}(n) \quad (10)$$

where  $\dagger$  stands for pseudo-inverse, and  $\hat{\mathbf{\Omega}}_P$  is the phase rotating matrix built with the estimated frequency  $\hat{f}_D$ .

### III. PERFORMANCE ANALYSIS

In this section, we provide theoretical performance analysis of the channel estimation method described above in the presence of additive noise and of slow channel fluctuations. The theoretical results of this section will be validated by simulation results in Section VI. Although we will not detail the analysis here, the same guidelines of the approach herein can be extended with minor changes to all methods proposed in [13] and to their generalizations in [14].

#### A. Small Perturbation Analysis

Based on the assumptions underlying the channel estimation method described in Section II, we observe blocks of data of the form  $\mathbf{X}_N = \mathbf{H} \mathbf{F} \mathbf{S}_N$ , where the channel matrix  $\mathbf{H}$  is Toeplitz. However, in practice, the data are perturbed by the presence of noise and by possible channel fluctuations that render the matrix  $\mathbf{H}$  non-Toeplitz. In both cases, the observed data matrix  $\mathbf{Y}_N$  can be decomposed as

$$\mathbf{Y}_N := \mathbf{X}_N + \delta \mathbf{X}_N. \quad (11)$$

We assume the following.

- a4) Perturbation  $\delta \mathbf{X}_N$  is small (in its e.g., Frobenius norm).

We will evaluate first the effect of  $\delta \mathbf{X}_N$  on the channel estimation. In Sections III-B and C, we will specialize the analysis to two kinds of perturbations, namely, additive noise and channel time variations. In the presence of a data perturbation  $\delta \mathbf{X}_N$ , we estimate the basis spanning the null space of  $\mathbf{X}_N$  as the set composed of the left singular vectors associated with the  $L$  smallest singular values of  $\mathbf{X}_N + \delta \mathbf{X}_N$ . Let us denote by  $\hat{\mathbf{U}} + \delta \hat{\mathbf{U}}$  the  $P \times L$  matrix formed with such vectors. Under a4), the perturbation of the null space is small, and thus

$$(\hat{\mathbf{U}} + \delta \hat{\mathbf{U}})^H (\mathbf{X}_N + \delta \mathbf{X}_N) \approx \mathbf{0}. \quad (12)$$

Furthermore, the small perturbation assumption implies that the null space of  $\mathbf{X}_N$  and the space spanned by  $\hat{\mathbf{U}} + \delta \hat{\mathbf{U}}$  are close to each other. This allows us to express  $\hat{\mathbf{U}} + \delta \hat{\mathbf{U}}$  as a linear combination of the columns of  $\tilde{\mathbf{U}}$  plus a small perturbation

$$\hat{\mathbf{U}} + \delta \hat{\mathbf{U}} := \tilde{\mathbf{U}} \mathbf{Q} + \delta \hat{\mathbf{U}} \quad (13)$$

where  $\mathbf{Q}$  is a unitary matrix,<sup>1</sup> and  $\delta \hat{\mathbf{U}}$  is a small perturbation. Since  $\mathbf{Q}$  is unitary, we can always introduce a matrix  $\delta \tilde{\mathbf{U}}$  such that  $\delta \hat{\mathbf{U}} = \delta \tilde{\mathbf{U}} \mathbf{Q}$  so that (12) is equivalent to

$$\begin{aligned} (\tilde{\mathbf{U}} \mathbf{Q} + \delta \tilde{\mathbf{U}} \mathbf{Q})^H (\mathbf{X}_N + \delta \mathbf{X}_N) \\ = \mathbf{Q}^H (\tilde{\mathbf{U}} + \delta \tilde{\mathbf{U}})^H (\mathbf{X}_N + \delta \mathbf{X}_N) \approx \mathbf{0}. \end{aligned} \quad (14)$$

The factor  $\mathbf{Q}$  can be eliminated from this equality, and using (5), we can write the first order approximation of (14) as

$$\begin{aligned} \tilde{\mathbf{U}}^H \mathbf{X}_N + \delta \tilde{\mathbf{U}}^H \mathbf{X}_N + \tilde{\mathbf{U}}^H \delta \mathbf{X}_N \\ = \delta \tilde{\mathbf{U}}^H \mathbf{X}_N + \tilde{\mathbf{U}}^H \delta \mathbf{X}_N \approx \mathbf{0} \end{aligned} \quad (15)$$

which is valid for small perturbations  $\delta \mathbf{X}_N$  and  $\delta \tilde{\mathbf{U}}$  (and thus,  $\delta \hat{\mathbf{U}}$ ).

We assume, without loss of generality, that  $\delta \tilde{\mathbf{U}}$  is orthogonal to  $\tilde{\mathbf{U}}$  because any component of  $\delta \tilde{\mathbf{U}}$  lying in the subspace spanned by  $\tilde{\mathbf{U}}$  does not alter the identification of the null space of  $\mathbf{X}_N$  and, hence, channel estimation. From (4), it follows that

$$\tilde{\mathbf{U}} \tilde{\mathbf{U}}^H + \tilde{\mathbf{U}} \tilde{\mathbf{U}}^H = \mathbf{X}_N \mathbf{X}_N^\dagger + \tilde{\mathbf{U}} \tilde{\mathbf{U}}^H = \mathbf{I}. \quad (16)$$

Hence, exploiting the orthogonality between  $\delta \tilde{\mathbf{U}}$  and  $\tilde{\mathbf{U}}$ , we have

$$\delta \tilde{\mathbf{U}}^H = \delta \tilde{\mathbf{U}}^H (\mathbf{X}_N \mathbf{X}_N^\dagger + \tilde{\mathbf{U}} \tilde{\mathbf{U}}^H) = \delta \tilde{\mathbf{U}}^H \mathbf{X}_N \mathbf{X}_N^\dagger. \quad (17)$$

Multiplying (15) from the right-hand side by  $\mathbf{X}_N^\dagger$ , we get  $\delta \tilde{\mathbf{U}}^H \mathbf{X}_N \mathbf{X}_N^\dagger = -\tilde{\mathbf{U}}^H \delta \mathbf{X}_N \mathbf{X}_N^\dagger$ , and thus, using (17)

$$\delta \tilde{\mathbf{U}}^H \approx -\tilde{\mathbf{U}}^H \delta \mathbf{X}_N \mathbf{X}_N^\dagger. \quad (18)$$

Note that this equation, along with a4), implies that small  $\|\delta \mathbf{X}_N\|$  gives rise to a small  $\|\delta \tilde{\mathbf{U}}\|$ .

We note also that  $\delta \tilde{\mathbf{U}}$  does not depend on the unitary factor  $\mathbf{Q}$ . Interestingly, we will prove next that also, the perturbation

<sup>1</sup>The matrix  $\mathbf{Q}$  describes the property that the basis of the null space of  $\mathbf{X}_N$  is determined only within a unitary matrix. Indeed, only the incorporation of the factor  $\mathbf{Q}$  allows us to state that a *small* perturbation  $\delta \mathbf{X}_N$  causes a *small* perturbation  $\delta \tilde{\mathbf{U}}$ , even though, in general, the difference  $\hat{\mathbf{U}} + \delta \hat{\mathbf{U}} - \tilde{\mathbf{U}}$  might not have a small norm.

on the channel estimate  $\hat{\mathbf{h}}$  does not depend on  $\mathbf{Q}$ , but it is expressible in terms of  $\delta\tilde{\mathbf{U}}$  only.

Because the estimation method in (6) proceeds by building the  $L$  Hankel matrices  $\{\tilde{\mathbf{U}}_l\}_{l=1}^L$ , as in (7), we analyze next the perturbation of these matrices induced by the perturbation on the observed data. Let  $\hat{\mathbf{U}}_l$  and  $\delta\hat{\mathbf{U}}_l$  denote the Hankel matrices in (7) built with the columns  $\hat{\mathbf{u}}_l$  and  $\delta\hat{\mathbf{u}}_l$  of the matrices  $\hat{\mathbf{U}}$  and  $\delta\hat{\mathbf{U}}$ . In Appendix A, we prove that the matrix  $\hat{\mathbf{U}} := [\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_L]$  is related to the matrix  $\tilde{\mathbf{U}}$  by

$$\hat{\mathbf{U}} = \tilde{\mathbf{U}}\mathbf{P}\mathbf{Q}\mathbf{P}^T := \tilde{\mathbf{U}}\mathbf{B} \quad (19)$$

where  $\mathbf{P}$  is an  $(ML) \times (ML)$  permutation matrix,  $\mathbf{Q} := \mathbf{I}_{M \times M} \otimes \mathbf{Q}$ , and  $\mathbf{B} := \mathbf{P}\mathbf{Q}\mathbf{P}^T$ . Proceeding similarly with the matrix  $\delta\hat{\mathbf{U}} := [\delta\hat{\mathbf{U}}_1, \dots, \delta\hat{\mathbf{U}}_L]$ , we also end up with

$$\delta\hat{\mathbf{U}} = \delta\tilde{\mathbf{U}}\mathbf{B}. \quad (20)$$

Replacing  $\tilde{\mathbf{U}}$  in (6) with  $\hat{\mathbf{U}} + \delta\hat{\mathbf{U}}$  gives rise to a perturbed channel estimate  $\mathbf{h} + \epsilon_h$ , which comes out as the solution of

$$(\mathbf{h} + \epsilon_h)^H (\hat{\mathbf{U}} + \delta\hat{\mathbf{U}}) = \mathbf{0}^H. \quad (21)$$

From (19) and (20) and considering that  $\mathbf{B}$  is full rank, we infer that (21) is equivalent to

$$(\mathbf{h} + \epsilon_h)^H (\tilde{\mathbf{U}} + \delta\tilde{\mathbf{U}}) = \mathbf{0}^H. \quad (22)$$

Equation (22) implies that the error  $\epsilon_h$  does not depend on the unitary matrix  $\mathbf{Q}$  but only on  $\delta\tilde{\mathbf{U}}$ , which, in turn, depends only on  $\delta\tilde{\mathbf{U}}$ . This observation shows that the performance of the channel estimator is not affected by the unitary ambiguity matrix  $\mathbf{Q}$ .

Relying on the small perturbations assumption, we retain only the first-order perturbation terms in (22). Hence, using (6), the channel estimation error  $\epsilon_h$  can be expressed in terms of  $\delta\tilde{\mathbf{U}}$  as

$$\epsilon_h^H \tilde{\mathbf{U}} \approx -\mathbf{h}^H \delta\tilde{\mathbf{U}} \quad (23)$$

where  $\delta\tilde{\mathbf{U}} = (\delta\tilde{\mathbf{U}}_1, \delta\tilde{\mathbf{U}}_2, \dots, \delta\tilde{\mathbf{U}}_L)$ , and the matrices  $\delta\tilde{\mathbf{U}}_l$ ,  $l = 1, \dots, L$  are the Hankel matrices built with columns  $\delta\hat{\mathbf{u}}_l$  of  $\delta\hat{\mathbf{U}}$ , as in (7).

The commutativity of convolution implies that each product  $\mathbf{h}^H \delta\tilde{\mathbf{U}}_l$  can be written as  $\delta\tilde{\mathbf{u}}_l^T \mathbf{H}^*$ , where  $\mathbf{H}$  is the Toeplitz channel matrix given by (2). As a consequence, we can write the right-hand side of (23) as

$$\begin{aligned} \mathbf{h}^H \delta\tilde{\mathbf{U}} &= (\mathbf{h}^H \delta\tilde{\mathbf{U}}_1, \dots, \mathbf{h}^H \delta\tilde{\mathbf{U}}_L) \\ &= (\delta\tilde{\mathbf{u}}_1^T \mathbf{H}^*, \dots, \delta\tilde{\mathbf{u}}_L^T \mathbf{H}^*) \\ &= \text{vec}^T(\delta\tilde{\mathbf{U}}) (\mathbf{I}_{L \times L} \otimes \mathbf{H}^*) \end{aligned} \quad (24)$$

where  $\otimes$  denotes Kronecker product, and the operator  $\text{vec}(\mathbf{A})$  stacks all the columns of a matrix  $\mathbf{A}$  into one vector.

Since the observed data depend on the product between the unknown symbols and the channel parameters, the estimation method can provide the channel impulse response only up to a scalar factor. Therefore, to compare the estimated channel  $\hat{\mathbf{h}} := \mathbf{h} + \epsilon_h$  with the true channel  $\mathbf{h}$ , we need to normalize  $\hat{\mathbf{h}}$ . Several options may be pursued. We can, for example i) compare  $\mathbf{h}$  with  $\rho\hat{\mathbf{h}}$ , where  $\rho := \hat{\mathbf{h}}^H \mathbf{h} / (\hat{\mathbf{h}}^H \hat{\mathbf{h}})$  or ii) divide  $\hat{\mathbf{h}}$  by

one of its entries, say  $\hat{h}(d)$ , and compare  $\hat{\mathbf{h}}/\hat{h}(d)$  with  $\mathbf{h}/h(d)$ . Choice i) yields the minimum mean square error between  $\hat{\mathbf{h}}$  and  $\mathbf{h}$ , whereas choice ii) simplifies the comparison with the Cramér–Rao bound. For the sake of simplicity, we adopt choice ii) in this section. Nonetheless, in Section VI, we will pursue further the constrained CRB achieved by enforcing the unit norm of the estimated channel and its relationship with the normalization i).

We will assume in the following that both true and estimated channel vectors are normalized so that  $h(d) = \hat{h}(d) = 1$ , where  $d$  is such that  $|h(d)| \geq |h(l)|$ ,  $\forall l \neq d$ ,  $d \in [0, L]$ . As a consequence, the error vector after normalization has a null element in its  $d$ th position by construction and we can write  $\epsilon_h^H \tilde{\mathbf{U}} = \delta\mathbf{h}^H \tilde{\mathbf{V}}$ , where  $\delta\mathbf{h}$  and  $\tilde{\mathbf{V}}$  are obtained by  $\epsilon_h$  and  $\tilde{\mathbf{U}}$  by removing their  $d$ th entry and  $d$ th row, respectively. From (23), we can thus express the error vector as

$$\delta\mathbf{h}^H \approx -\mathbf{h}^H \delta\tilde{\mathbf{U}} \tilde{\mathbf{V}}^\dagger. \quad (25)$$

Substituting (24) into (25), the channel estimation error is

$$\delta\mathbf{h}^H \approx -\text{vec}^T(\delta\tilde{\mathbf{U}}) (\mathbf{I}_{L \times L} \otimes \mathbf{H}^*) \tilde{\mathbf{V}}^\dagger \quad (26)$$

where the perturbation  $\delta\tilde{\mathbf{U}}$  can be expressed directly in terms of the perturbation  $\delta\mathbf{X}_N$  through (18). Furthermore, recalling the  $\text{vec}$  operator property  $\text{vec}(\mathbf{ACB}) = (\mathbf{B}^T \otimes \mathbf{A})\text{vec}(\mathbf{C})$  [8], we have from (18)

$$\text{vec}(\delta\tilde{\mathbf{U}}) = -(\tilde{\mathbf{U}}^T \otimes \mathbf{X}_N^{\dagger H}) \text{vec}(\delta\mathbf{X}_N^H). \quad (27)$$

Thus, the error on the channel estimate in (26) becomes

$$\begin{aligned} \delta\mathbf{h}^H &= \text{vec}^T(\delta\mathbf{X}_N^H) (\tilde{\mathbf{U}}^T \otimes \mathbf{X}_N^{\dagger H})^T (\mathbf{I}_{L \times L} \otimes \mathbf{H}^*) \tilde{\mathbf{V}}^\dagger \\ &= \text{vec}^T(\delta\mathbf{X}_N^H) (\tilde{\mathbf{U}} \otimes \mathbf{X}_N^{\dagger*}) (\mathbf{I}_{L \times L} \otimes \mathbf{H}^*) \tilde{\mathbf{V}}^\dagger \\ &= \text{vec}^T(\delta\mathbf{X}_N^H) (\tilde{\mathbf{U}} \otimes (\mathbf{X}_N^{\dagger*} \mathbf{H}^*)) \tilde{\mathbf{V}}^\dagger \end{aligned} \quad (28)$$

where  $*$  indicates conjugation, and the last equality has been derived by exploiting the property  $(\mathbf{A}_1 \otimes \mathbf{A}_2)(\mathbf{B}_1 \otimes \mathbf{B}_2) = (\mathbf{A}_1 \mathbf{B}_1 \otimes \mathbf{A}_2 \mathbf{B}_2)$  [8]. We will now specialize this general expression to some important perturbation sources.

### B. Performance in Additive Noise

In the presence of additive white Gaussian noise (AWGN) only, the perturbation  $\delta\mathbf{X}_N$  coincides with the additive noise, i.e.,

$$\mathbf{X}_N + \delta\mathbf{X}_N = \mathbf{H}\mathbf{F}\mathbf{S}_N + \mathbf{W}_N \quad (29)$$

where  $\mathbf{W}_N \equiv \delta\mathbf{X}_N$  is the  $P \times N$  noise matrix. According to (28), the channel estimation error vector  $\delta\mathbf{h}$  is asymptotically (for high SNR) a Gaussian random vector with zero mean and covariance matrix

$$\begin{aligned} \mathbf{C}_{hh} &\approx \mathbf{R}_{hh} = E\{\delta\mathbf{h}\delta\mathbf{h}^H\} \\ &= \sigma_v^2 \tilde{\mathbf{V}}^{\dagger H} (\tilde{\mathbf{U}}^H \otimes \mathbf{H}^T \mathbf{X}_N^{\dagger T}) (\tilde{\mathbf{U}} \otimes \mathbf{X}_N^{\dagger*} \mathbf{H}^*) \tilde{\mathbf{V}}^\dagger \\ &= \sigma_v^2 \tilde{\mathbf{V}}^{\dagger H} \left[ (\tilde{\mathbf{U}}^H \tilde{\mathbf{U}}) \otimes (\mathbf{H}^T \mathbf{X}_N^{\dagger T} \mathbf{X}_N^{\dagger*} \mathbf{H}^*) \right] \tilde{\mathbf{V}}^\dagger \\ &= \sigma_v^2 \tilde{\mathbf{V}}^{\dagger H} \left[ \mathbf{I}_{L \times L} \otimes (\mathbf{H}^T \mathbf{X}_N^{\dagger T} \mathbf{X}_N^{\dagger*} \mathbf{H}^*) \right] \tilde{\mathbf{V}}^\dagger. \end{aligned} \quad (30)$$

However, since  $\mathbf{H}$  is full column rank, we also have

$$\begin{aligned} \mathbf{X}_N^\dagger \mathbf{H} &= (\mathbf{H} \mathbf{F} \mathbf{S}_N)^\dagger \mathbf{H} \\ &= (\mathbf{F} \mathbf{S}_N)^\dagger \mathbf{H}^\dagger \mathbf{H} = (\mathbf{F} \mathbf{S}_N)^\dagger = \mathbf{S}_N^\dagger \mathbf{F}^{-1} \end{aligned} \quad (31)$$

so that we can rewrite (30) as

$$\mathbf{C}_{hh} \approx \sigma_v^2 \tilde{\mathbf{V}}^{\dagger H} \left[ \mathbf{I}_{L \times L} \otimes \left( \mathbf{F}^{-T} \mathbf{S}_N^{\dagger T} \mathbf{S}_N^{\dagger*} \mathbf{F}^{-*} \right) \right] \tilde{\mathbf{V}}^\dagger. \quad (32)$$

Because, under a3)  $\mathbf{S}_N$  is full row rank  $\mathbf{S}_N^{\dagger H} \mathbf{S}_N^\dagger = (\mathbf{S}_N \mathbf{S}_N^H)^{-1}$ , and thus, (32) may be rewritten as

$$\mathbf{C}_{hh} \approx \sigma_v^2 \tilde{\mathbf{V}}^{\dagger H} \left[ \mathbf{I}_{L \times L} \otimes \left( \mathbf{F}^{-T} (\mathbf{S}_N^* \mathbf{S}_N^T)^{-1} \mathbf{F}^{-*} \right) \right] \tilde{\mathbf{V}}^\dagger. \quad (33)$$

This expression can be further simplified when we have the following.

- i)  $\mathbf{F}$  is unitary.
- ii) The symbols are zero mean, uncorrelated, with the same variance  $\sigma_s^2$ .
- iii)  $N$  is large enough so that  $\mathbf{S}_N \mathbf{S}_N^H \approx N \sigma_s^2 \mathbf{I}$ .

When i)–iii) are satisfied, we can approximate (33) as

$$\mathbf{C}_{hh} \approx \frac{1}{NSNR} \tilde{\mathbf{V}}^{\dagger H} \tilde{\mathbf{V}}^\dagger \quad (34)$$

where the signal-to-noise ratio (SNR) is defined as  $SNR = \sigma_s^2 / \sigma_v^2$ . Since  $\tilde{\mathbf{V}}^\dagger = \tilde{\mathbf{V}}^H (\tilde{\mathbf{V}} \tilde{\mathbf{V}}^H)^{-1}$ , we may simplify (34) further to obtain

$$\mathbf{C}_{hh} \approx \frac{1}{NSNR} \left( \tilde{\mathbf{V}} \tilde{\mathbf{V}}^H \right)^{-1}. \quad (35)$$

### C. Performance Under Slow Channel Fluctuations

The estimation algorithm reviewed in Section II requires the channel to be linear time invariant. However, the method was seen in simulations to provide reliable estimates even under slow channel variations. The purpose of this section is to quantify the resilience of that method to practically important sources of fluctuation, such as carrier asynchronism, Doppler effects, and oscillators phase noise, which are often present in mobile communication systems. Specifically, channel fluctuations are incorporated by letting the channel matrix  $\mathbf{H}$  in (1) be a function of the block index  $n$

$$\mathbf{x}(n) + \delta \mathbf{x}(n) = \mathbf{H}(n) \mathbf{F} \mathbf{s}(n) \quad (36)$$

where  $\mathbf{H}(n)$  is given by the sum of a Toeplitz matrix  $\mathbf{H} \equiv \mathcal{T}(\mathbf{h})$  plus a non-Toeplitz perturbation matrix  $\delta \mathbf{H}(n)$

$$\mathbf{H}(n) = \mathcal{T}(\mathbf{h}) + \delta \mathbf{H}(n) = \mathbf{H} + \delta \mathbf{H}(n). \quad (37)$$

Correspondingly, the observation  $\mathbf{X}_N$  is perturbed by a matrix

$$\begin{aligned} \delta \mathbf{X}_N &= [\delta \mathbf{H}(0) \mathbf{F} \mathbf{s}(0), \dots, \delta \mathbf{H}(N-1) \mathbf{F} \mathbf{s}(N-1)] \\ &= [\delta \mathbf{H}(0), \dots, \delta \mathbf{H}(N-1)] (\mathbf{I}_{N \times N} \otimes \mathbf{F}) \mathbf{S}_N \\ &:= \delta \mathcal{H}_N \Phi \mathcal{S}_N \end{aligned} \quad (38)$$

where we defined the matrices  $\delta \mathcal{H}_N := [\delta \mathbf{H}(0), \dots, \delta \mathbf{H}(N-1)]$ ,  $\Phi := (\mathbf{I}_{N \times N} \otimes \mathbf{F})$ , and

$$\mathbf{S}_N := \begin{pmatrix} \mathbf{s}(0) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{s}(N-1) \end{pmatrix} \quad (39)$$

with corresponding dimensions  $P \times MN$ ,  $MN \times MN$ , and  $MN \times N$ .

Using (38), the perturbation (18) can be expressed as

$$\delta \tilde{\mathbf{U}} = -\mathbf{X}_N^{\dagger H} (\delta \mathcal{H}_N \Phi \mathcal{S}_N)^H \tilde{\mathbf{U}}. \quad (40)$$

Substituting this expression in (26), we find the corresponding error on the channel estimate

$$\begin{aligned} \delta \mathbf{h}^T &\approx \text{vec}^T \left[ \mathbf{X}_N^{\dagger H} (\delta \mathcal{H}_N \Phi \mathcal{S}_N)^H \tilde{\mathbf{U}} \right] (\mathbf{I}_{L \times L} \otimes \mathbf{H}^*) \tilde{\mathbf{V}}^\dagger \\ &= \text{vec}^T \left[ \left( \mathbf{X}_N^{\dagger H} \mathcal{S}_N^H \Phi^H \right) \delta \mathcal{H}_N^H \tilde{\mathbf{U}} \right] (\mathbf{I}_{L \times L} \otimes \mathbf{H}^*) \tilde{\mathbf{V}}^\dagger \end{aligned} \quad (41)$$

and, after using the *vec* operator properties, we obtain

$$\begin{aligned} \delta \mathbf{h}^T &\approx \left[ \tilde{\mathbf{U}}^T \otimes \left( \mathbf{X}_N^{\dagger H} \mathcal{S}_N^H \Phi^H \right) \text{vec} \left( \delta \mathcal{H}_N^H \right) \right]^T (\mathbf{I}_{L \times L} \otimes \mathbf{H}^*) \tilde{\mathbf{V}}^\dagger \\ &= \text{vec}^T \left( \delta \mathcal{H}_N^H \right) \left[ \tilde{\mathbf{U}} \otimes \left( \Phi \mathcal{S}_N \mathbf{X}_N^\dagger \right)^* \right] (\mathbf{I}_{L \times L} \otimes \mathbf{H}^*) \tilde{\mathbf{V}}^\dagger \\ &= \text{vec}^T \left( \delta \mathcal{H}_N^H \right) \left[ \tilde{\mathbf{U}} \otimes \left( \Phi \mathcal{S}_N \mathbf{X}_N^\dagger \mathbf{H} \right)^* \right] \tilde{\mathbf{V}}^\dagger. \end{aligned} \quad (42)$$

We can use expression (42) for channel fluctuations that are modeled either as deterministic or as random. When Doppler shifts are present,  $\delta \mathbf{X}_N$  may be viewed as deterministic, and (42) provides an approximation for the channel estimation error. Conversely, performance under random channel fluctuations can be evaluated using the statistics of  $\delta \mathbf{h}$  in (42). If the covariance matrix of the channel fluctuations is known, we can use (42) to compute the correlation matrix of  $\delta \mathbf{h}$  as

$$\begin{aligned} \mathbf{R}_{hh} &= \tilde{\mathbf{V}}^{\dagger H} \left[ \tilde{\mathbf{U}}^H \otimes \left( \Phi \mathcal{S}_N \mathbf{X}_N^\dagger \mathbf{H} \right)^T \right] \\ &E \left\{ \text{vec} \left( \delta \mathcal{H}_N^T \right) \text{vec}^T \left( \delta \mathcal{H}_N^H \right) \right\} \cdot \left[ \tilde{\mathbf{U}} \otimes \left( \Phi \mathcal{S}_N \mathbf{X}_N^\dagger \mathbf{H} \right)^* \right] \tilde{\mathbf{V}}^\dagger. \end{aligned}$$

In Section VI, we will verify, using simulated examples, the validity of the previous expressions in predicting the performance of systems affected by different sources of fluctuations, including oscillators' phase noise and multipath propagation with different Doppler shift on different paths.

Before testing our derivations via numerical simulations, in the next section, we will prove that the channel estimation method of [13] is high-SNR efficient in the presence of AWGN.

## IV. CRAMÉR–RAO BOUND

In this section, we derive the Cramér–Rao bound (CRB) of the channel impulse response estimate, assuming that the transmitted symbols are nuisance parameters. Given the observation vector in (1), we assume that the noise  $\mathbf{v}(n)$  is composed of i.i.d. zero mean, complex Gaussian random variables uncorrelated from the symbol sequence  $\mathbf{s}(n)$ . The CRB for multiple FIR channels driven by a common unknown input was studied in [1], [3], [7], and [9]. Specifically, the CRB was used in [3] to compare blind, semi-blind, and training sequence based channel estimation techniques. We derive here the CRB for the transmission system using null guard intervals, using the general complex CRB form of [17]. We exploit the whiteness and circular symmetry of the noise  $\mathbf{v}(n)$  to arrive at a simple closed-form expression. Interestingly, the CRB coincides with (35), thus proving that the channel estimator of [13] is *high-SNR efficient* for any data block length  $N$ .

The probability density function (pdf) of each block of the received data  $\mathbf{y}(n)$ , conditioned on the channel parameter vector  $\mathbf{h}$  and the symbol vector  $\mathbf{s}(n)$ , is

$$p_{\mathbf{y}/\mathbf{h}, \mathbf{s}}(\mathbf{y}(n)/\mathbf{h}, \mathbf{s}(n)) = \frac{1}{(2\pi\sigma_v)^P} e^{-|\mathbf{y}(n) - \mathbf{H}\mathbf{F}\mathbf{s}(n)|^2/\sigma_v^2} \quad (43)$$

where  $\sigma_v^2$  is the variance of the complex AWGN, and  $|\cdot|$  is now indicating vector norm. Because the noise samples are uncorrelated, the joint pdf of a block  $\mathbf{Y}_N := [\mathbf{y}(1), \dots, \mathbf{y}(N)]$  is

$$p_{\mathbf{Y}_N/\mathbf{h}, \mathbf{S}_N}(\mathbf{Y}_N/\mathbf{h}, \mathbf{S}_N) = \frac{1}{(2\pi\sigma_v)^{PN}} \exp - \sum_{n=1}^N |\mathbf{y}(n) - \mathbf{H}\mathbf{F}\mathbf{s}(n)|^2/\sigma_v^2. \quad (44)$$

To simplify the CRB derivation, it is useful to introduce the equivalent  $1 \times 2(L+1)NM$  complex parameter vector

$$\boldsymbol{\theta} := [\mathbf{h}^T, \mathbf{s}^T(1), \dots, \mathbf{s}^T(N), \mathbf{h}^H, \mathbf{s}^H(1), \dots, \mathbf{s}^H(N)] \quad (45)$$

and write the log-likelihood function as

$$f(\boldsymbol{\theta}) := \log p_{\mathbf{Y}_N/\mathbf{h}, \mathbf{S}_N}(\mathbf{Y}_N/\mathbf{h}, \mathbf{S}_N) = -\log((2\pi\sigma_v)^{PN}) - \sum_{n=1}^N \frac{|\mathbf{y}(n) - \mathbf{H}\mathbf{F}\mathbf{s}(n)|^2}{\sigma_v^2}. \quad (46)$$

Similar to [17], we assume that the SNR is high enough to neglect the estimator's bias and introduce the  $2(L+1)NM \times 2(L+1)NM$  equivalent complex Fisher's information matrix (FIM)

$$\mathcal{J}(\boldsymbol{\theta}) := E_{\mathbf{Y}_N/\mathbf{h}, \mathbf{S}_N} \left\{ \left( \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right)^H \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right\} \quad (47)$$

where  $\partial f(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^T$  is an  $1 \times 2(L+1)NM$  row vector. Although differentiation of  $f(\boldsymbol{\theta})$  in (46) may look cumbersome, a series of simplifications are possible. First, we prove in Appendix B that the CRB of  $\boldsymbol{\theta}$  can be found by considering the reduced  $(L+1)NM \times (L+1)NM$  matrix

$$\begin{aligned} \bar{\mathcal{J}} &:= E_{\mathbf{Y}_N/\mathbf{h}, \mathbf{S}_N} \left\{ \left( \frac{\partial f(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^T} \right)^H \frac{\partial f(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^T} \right\} \\ &:= \begin{pmatrix} \mathcal{J}_{1,1} & \mathcal{J}_{1,2} \\ \mathcal{J}_{1,2}^H & \mathcal{J}_{2,2} \end{pmatrix} \end{aligned} \quad (48)$$

where  $\boldsymbol{\eta} := [\mathbf{h}^T, \mathbf{s}^T(1), \dots, \mathbf{s}^T(N)]$ . The expression of the submatrices  $\mathcal{J}_{1,1}$ ,  $\mathcal{J}_{1,2}$ ,  $\mathcal{J}_{2,2}$  [having dimensionalities  $(L+1) \times (L+1)$ ,  $(L+1) \times NM$ ,  $NM \times NM$ , respectively] are given explicitly in (67) in Appendix B.

With the symbols treated as nuisance parameters, the bound on the estimated channel covariance matrix is given by the  $(L+1) \times (L+1)$  upper left block of  $\bar{\mathcal{J}}^{-1}$ . However,  $\bar{\mathcal{J}}^{-1}$  does not exist because the channel coefficients can be identified only within a scalar factor. Hence, in accordance with the results derived in Section III, we normalize the  $L$ th-order channel impulse response by its  $d$ th tap, where  $d \in [0, L]$ :  $|h(d)| \geq |h(l)|, \forall l$ . This implies that the  $d$ th entry of the channel vector  $\mathbf{h}$  is no longer an unknown; therefore, we can remove the  $d$ th row and  $d$ th column from the FIM. The corresponding compacted FIM is

invertible, and the upper left  $L \times L$  portion  $\mathbf{C}_{CR}$  of its inverse is the CRB of the channel parameters, which is derived by treating the unknown symbols as nuisance parameters. Specifically, we prove in Appendix B that the CRB is

$$\mathbf{C}_{CR} = \sigma_v^2 \tilde{\mathbf{V}}^{\dagger H} \left[ \mathbf{I}_{L \times L} \otimes \left( \mathbf{F}^{-T} (\mathbf{S}_N^* \mathbf{S}_N^T)^{-1} \mathbf{F}^{-*} \right) \right] \tilde{\mathbf{V}}^{\dagger}. \quad (49)$$

Comparing (49) with (33), which was derived under high SNR assumptions, we can thus state that the deterministic method of [13] is *statistically efficient at high SNR*. Furthermore, it is worth pointing out that since (33) was derived for any data block length, the method is efficient for *any* data block length.

Even though the normalization of the channel estimate used above has been useful to simplify the comparison between the CRB and the estimated channel covariance matrix, in practice, the channel estimate  $\hat{\mathbf{h}}$  in (6) is computed as the eigenvector corresponding to the smallest eigenvalue of  $\tilde{\mathbf{U}}\tilde{\mathbf{U}}^H$ . Therefore,  $\hat{\mathbf{h}}$  has unit norm by construction. Thus, it is interesting to derive the CRB under the constraint that  $\hat{\mathbf{h}}$  has unit norm. Generalizing the procedure of [16] to the complex case, the constrained CRB  $\mathbf{C}_C$  is

$$\mathbf{C}_C = \mathbf{W} (\mathbf{W}^H \mathbf{D} \mathbf{W})^{-1} \mathbf{W}^H \quad (50)$$

where  $\mathbf{D}$  is given in (79), and  $\mathbf{W}$  is any  $(L+1) \times L$  matrix whose columns form an orthonormal basis for the null space of  $\mathbf{h}$ , or, in formulas

$$\mathbf{h}^H \mathbf{W} = \mathbf{0}^H. \quad (51)$$

Since  $\mathbf{h}$  satisfies (6), based on (51), we can express the SVD of  $\tilde{\mathbf{U}}$  as

$$\begin{aligned} \tilde{\mathbf{U}} &= (\mathbf{W}, \mathbf{h}) \begin{pmatrix} \boldsymbol{\Sigma}_{L \times L} & \mathbf{0}_{L \times (L-1)M} \\ \mathbf{0}_{1 \times L} & \mathbf{0}_{1 \times (L-1)M} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{V}}_u^H \\ \tilde{\mathbf{V}}_u^H \end{pmatrix} \\ &= \mathbf{W} \boldsymbol{\Sigma}_{L \times L} \hat{\mathbf{V}}_u^H. \end{aligned} \quad (52)$$

Considering that  $\tilde{\mathbf{U}}^\dagger = \mathbf{W} \boldsymbol{\Sigma}_{L \times L}^{-1} \hat{\mathbf{V}}_u^H$ , from (50), (52), and (79) we have

$$\begin{aligned} \mathbf{C}_C &= \frac{\mathbf{W}}{\sigma_v^2} \left\{ \mathbf{W}^H \tilde{\mathbf{U}} \left[ \mathbf{I}_{L \times L} \otimes (\mathbf{F}^* \mathbf{S}_N^* \mathbf{S}_N^T \mathbf{F}^T) \right] \tilde{\mathbf{U}}^H \mathbf{W} \right\}^{-1} \mathbf{W}^H \\ &= \frac{\mathbf{W}}{\sigma_v^2} \left\{ \boldsymbol{\Sigma}_{L \times L} \hat{\mathbf{V}}_u^H \left[ \mathbf{I}_{L \times L} \otimes (\mathbf{F}^* \mathbf{S}_N^* \mathbf{S}_N^T \mathbf{F}^T) \right] \right. \\ &\quad \left. \cdot \hat{\mathbf{V}}_u \boldsymbol{\Sigma}_{L \times L} \right\}^{-1} \mathbf{W}^H \\ &= \frac{\mathbf{W}}{\sigma_v^2} \boldsymbol{\Sigma}_{L \times L}^{-1} \hat{\mathbf{V}}_u^\dagger \left[ \mathbf{I}_{L \times L} \otimes (\mathbf{F}^* \mathbf{S}_N^* \mathbf{S}_N^T \mathbf{F}^T) \right]^{-1} \\ &\quad \cdot \hat{\mathbf{V}}_u^H \boldsymbol{\Sigma}_{L \times L}^{-1} \mathbf{W}^H \\ &= \frac{1}{\sigma_v^2} \tilde{\mathbf{U}}^\dagger \left[ \mathbf{I}_{L \times L} \otimes (\mathbf{F}^* \mathbf{S}_N^* \mathbf{S}_N^T \mathbf{F}^T)^{-1} \right]^{-1} \tilde{\mathbf{U}}^{\dagger H}. \end{aligned} \quad (53)$$

Hence, interestingly enough, the CRB constrained by the unit norm of the parameter vector is equal to the pseudo-inverse of the conditional FIM  $\mathbf{D}$  reported in (79). The use of the pseudo-inverse to restore invertibility of the FIM and derive the CRB was also suggested in [3] for SIMO-based channel estimators but without proving that the pseudo-inverse constitutes the CRB

subject to the unit norm constraint. In Section VI, we will compare this constrained CRB with the minimum mean square error achieved with the estimation method of Section II.

## V. SYMBOL DETECTION AND COMPARISON WITH TRAINING-BASED SYSTEMS

With the channel estimate  $\hat{\mathbf{h}}$  available, the symbols are detected by multiplying each block  $\mathbf{y}(n)$  in (1) by the matrix  $\mathbf{G} := (\hat{\mathbf{H}}\mathbf{F})^\dagger = \mathbf{F}^{-1}\hat{\mathbf{H}}^\dagger$ , where  $\hat{\mathbf{H}}$  is built from  $\hat{\mathbf{h}}$  similar to (2). The resulting decision vector is thus [cf. (1)]

$$\hat{\mathbf{s}}(n) = \mathbf{F}^{-1}\hat{\mathbf{H}}^\dagger\mathbf{H}\mathbf{F}\mathbf{s}(n) + \mathbf{F}^{-1}\mathbf{H}^\dagger\mathbf{w}(n). \quad (54)$$

Ideally,  $\hat{\mathbf{h}} = \mathbf{h}$ , and we obtain

$$\hat{\mathbf{s}}(n) = \mathbf{s}(n) + \mathbf{F}^{-1}\mathbf{H}^\dagger\mathbf{w}(n). \quad (55)$$

It is worth noticing that the  $P \times M$  matrix  $\hat{\mathbf{H}}$  is *always* full column rank so that it always admits a pseudo-inverse matrix  $\hat{\mathbf{H}}^\dagger$  such that  $\hat{\mathbf{H}}^\dagger\hat{\mathbf{H}} = \mathbf{I}_{M \times M}$  [12].

Since channel estimation is based on data received from  $N$  transmitted blocks, each containing  $M$  unknown information symbols followed by  $L$  guard zeros, we may interpret the zeros as distributed training bits. Therefore, it is of interest to compare the performance of such precoder-based channel estimator with an alternative scheme that transmits a sequence of  $NL$  consecutive training bits followed by  $NM$  information symbols. The two transmitters represent two alternative ways to distribute redundant data: Our system distributes the  $NL$  training (zero) bits over  $N$  blocks to eliminate interblock interference and guarantee FIR zero-forcing equalization [13]; the equivalent training-based system inserts all the  $NL$  known symbols at the beginning to minimize the channel estimation error variance. To compare the two approaches, we simulated a training-based system composed of the following blocks

- i) The transmitter sends blocks of  $N(M + L)$  data, composed of  $NL$  initial training symbols that are perfectly known at the receiver, followed by  $NM$  information symbols.
- ii) The receiver exploits the *a priori* knowledge of the training sequence to provide an MSE estimate  $\hat{\mathbf{h}}$  of the channel impulse response.
- iii) The received data are processed through a finite order discrete-time Wiener equalizer  $\hat{\mathbf{h}}$ .

The Wiener equalizer has impulse response  $g_W(n) = \tilde{g}^*(L_W - n)$ ,  $n = 0, \dots, L_W$ , where  $\tilde{g}(n)$  are the entries of the vector  $\tilde{\mathbf{g}} = [\tilde{g}(0), \dots, \tilde{g}(L_W)]$  obtained from the estimated channel  $\hat{\mathbf{H}}$  using

$$\tilde{\mathbf{g}} = \left( \sigma_s^2 \hat{\mathbf{H}}\hat{\mathbf{H}}^H + \sigma_n^2 \mathbf{I} \right)^{-1} \hat{\mathbf{H}}\mathbf{e}_d \quad (56)$$

where vector  $\mathbf{e}_d$  has all entries equal to 0, except the  $d$ th one, which is equal to 1, and  $d$  is a delay parameter. In our simulations, we used the delay  $d$  that minimizes the MSE.

The two systems exhibit the following tradeoffs: i) As far as channel estimation accuracy is concerned, the training-based system provides better performance because the estimation relies upon a long consecutive sequence of *known* symbols; however, ii) as far as channel equalization is concerned, the

scheme with null guard intervals guarantees perfect ZF equalization using a finite length block equalizer, provided that the channel estimate is sufficiently accurate. Conversely, the FIR Wiener equalizer cannot cancel ISI completely. To eliminate ISI exactly, the Wiener filter should be IIR, but then, the filter could become unstable.

Since the ultimate performance parameter in a digital communication system is bit error rate (BER), in Section VI, we will compare the two schemes in terms of BER.

## VI. SIMULATED PERFORMANCE

To validate the perturbation analysis of Section III, we compare now our theoretical expressions with simulation results and with the Cramér–Rao bound found in Section IV.

We generate sequences of i.i.d. symbols  $s(n)$  drawn from a QPSK constellation and simulate the Rayleigh fading channel, which is an  $L$ th-order FIR filter whose  $L + 1$  taps are i.i.d., zero-mean, unit variance complex Gaussian random variables. Since the performance depends on the symbol matrix  $\mathbf{S}_N$  and on the channel vector  $\mathbf{h}$ , we average the MSE over up to 1000 independent realizations of  $\mathbf{S}_N$  and  $\mathbf{h}$ . Furthermore, unless explicitly specified, to account for the scalar ambiguity present in all blind channel estimators, we compare the true channel  $\mathbf{h}/h(l_{\max})$ , where  $l_{\max}: |h(l_{\max})| \geq |h(l)|, \forall l \neq l_{\max}$ , with its estimate  $\hat{\mathbf{h}}/\hat{h}(l_{\max})$ .

*Example #1: [LTI Channel With AWGN]:* In this first example, we considered the effect of additive noise on the channel estimate. We tested our theoretical results by comparing the MSE computed as the trace of (32) with the corresponding simulation results averaged over a set of 100 independent Rayleigh fading channels of order  $L = 4$  and symbol blocks of size  $M = 12$ . We considered three precoding matrices.

- i) IFFT precoder with  $\{\mathbf{F}\}_{k,l} = e^{j2\pi kl/M}$ ;
- ii) Walsh–Hadamard (WH) precoder where the columns of  $\mathbf{F}$  are the Hadamard vectors;
- iii) complex pseudo-noise (PN) precoder where the columns of  $\mathbf{F}$  are i.i.d. complex Gaussian random vectors.

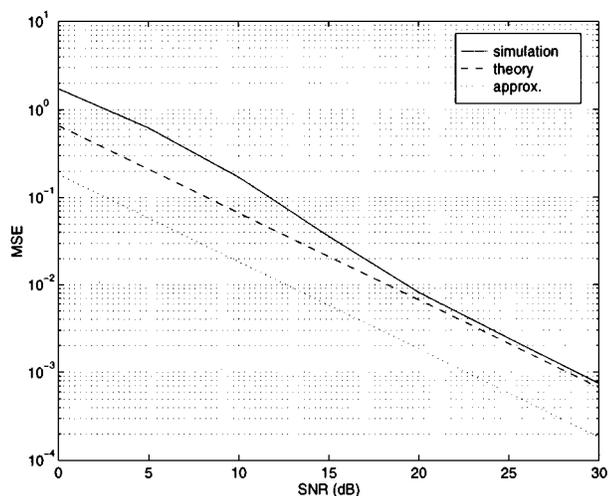
In all cases, each column of  $\mathbf{F}$  is normalized to have unit norm so that the energy per transmitted symbol equals one. We chose the IFFT matrix as a representative precoder having narrowband codes and the WH matrix for wideband codes. The PN code has been included as an example of nonunitary precoding matrix.

With WH precoding, Fig. 1 reports the channel MSE as a function of the SNR obtained using the following.

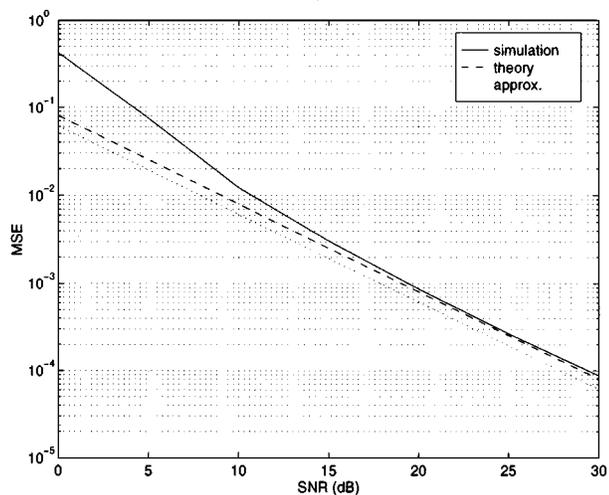
- i) simulations (solid line);
- ii) the trace of the theoretical covariance matrix  $\mathbf{C}_{hh}$  given in (32) (dashed line);
- iii) the trace of the approximate covariance matrix given by (34) (dotted line).

Specifically, Fig. 1(a) and (b) refer to  $N = M + L = 16$  and  $N = 3(M + L)$ , respectively. We observe a very good agreement between theory and simulation for SNR values greater than 10 dB, which validates the small perturbation assumption. Furthermore, from Fig. 1(b), we notice that for  $N = 3(M + L)$ , the approximate formula (34) is very close to the simulation results.

To quantify the effect of the encoder on the channel estimate, we compare in Fig. 2 the channel MSE obtained theoretically



(a)



(b)

Fig. 1. Channel MSE versus SNR (in decibels) obtained by simulations (solid line), trace of the covariance matrix  $C_{hh}$  in (32) (dashed line), and of the approximate covariance matrix in (34) (dotted line). (a)  $N = M + L$ . (b)  $N = 3(M + L)$ .

and by simulation using the same parameters as in Fig. 1(b), except for the precoders, which are now built with the IFFT, WH, and PN codes. Specifically, Fig. 2 shows the theoretical values (dashed line) for the IFFT, WH, and PN codes as well as the corresponding simulation results (solid line). From this figure, we see that, as predicted by the theory, unitary precoders yield the same channel MSE; therefore, the IFFT and WH codes perform similarly, whereas the nonunitary precoding induced by the PN code performs worse.

To quantify the effect of channel normalization implicit in the estimation of the channel vector as in Theorem 1, in Fig. 3, we compare the minimum mean square error (MMSE) obtained by multiplying the channel estimate by the correlation coefficient between the estimated channel and the true channel, as suggested in Section III-A, with the CRB subject to the unit norm constraint, as in (53). We can see that for SNR values greater than 10 dB, the MMSE is very close to the constrained CRB.

*Example #2: [LTI Channel With Doppler Shifts or Phase Noise]:* In Fig. 4(a), we compare the channel MSE obtained

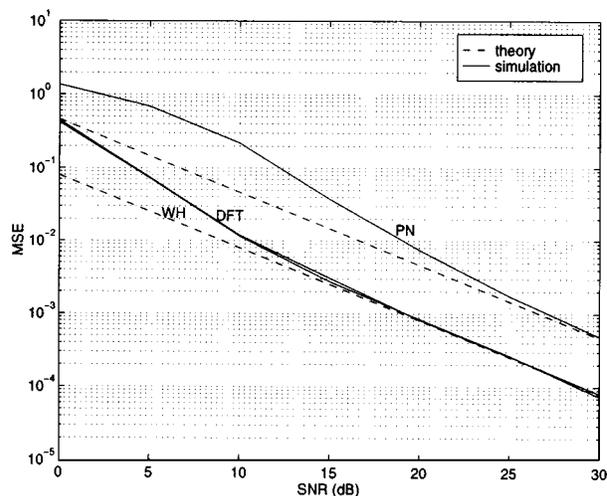


Fig. 2. Comparison between IFFT, WH, and PN precoders. Simulation (solid line) and theory (dashed line).

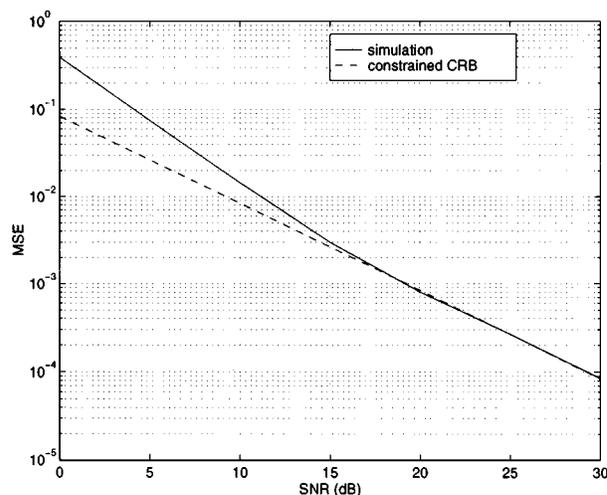
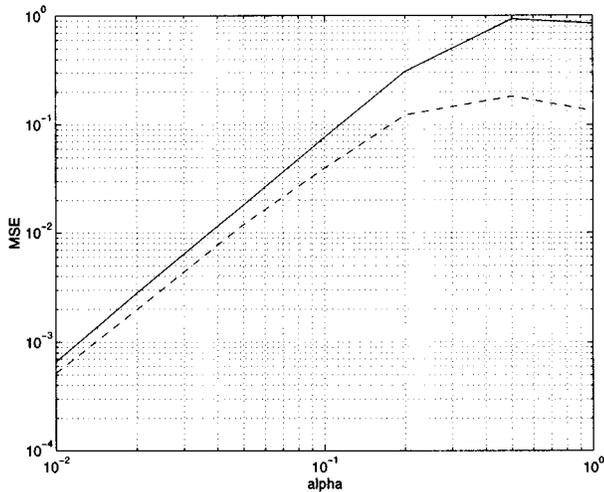


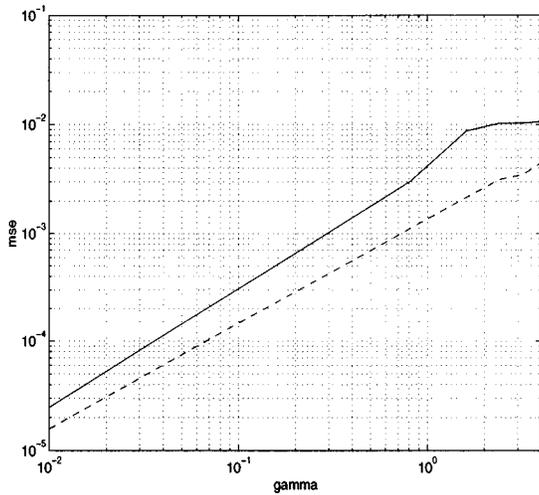
Fig. 3. Channel MSE versus SNR (in decibels). Comparison between MMSE and CRB subject to unit norm constraint.

theoretically (dashed line) using (42) and by simulation (solid line) on a time-varying multipath channel composed of three rays with amplitudes  $[1, 0.9, 0.9]$ , delays  $[0, T, 2T]$ , and corresponding Doppler shifts  $\alpha[0, 1, 2]/(NPT)$  as a function of the Doppler spread parameter  $\alpha$  ( $1/T$  is the symbol rate). We note that for Doppler drifts smaller than roughly 20% of the rate  $1/(NPT)$  (inverse of the observation time needed for channel estimation), we observe a good agreement between theory and simulations.

Fig. 4(b) shows the performance achieved in the presence of oscillators' phase noise modeled as a multiplicative noise  $\exp(j\phi(n))$ , where  $\phi(n)$  is a Wiener process obtained by integrating a white Gaussian noise with variance  $\sigma_n^2 = \gamma/(NPT)$ . The curves are obtained by averaging over 40 independent realizations of phase noise, and the channel is modeled here as a Rayleigh fading channel with three independent paths. Once more, the results of Section III-C are confirmed for a wide range of the phase noise bandwidth controlling parameter  $\gamma$ . Most important, Fig. 4 testifies a good resilience of the estimation error to slow fluctuations.



(a)



(b)

Fig. 4. Channel MSE obtained by simulation (solid line) and analytically (dashed line). (a) MSE versus Doppler spread. (b) MSE versus phase noise bandwidth.

*Example #3: [Comparison With Training-Based Systems]:* Finally, we compared the scheme with interblocks null guard intervals with the more conventional approach using training sequences for channel estimation and Wiener equalization for symbol recovery described in Section V. Since the ultimate figure of merit for the physical layer of a digital communication system is BER, in Fig. 5, we show the BER of the two alternative systems. We assumed the same ratio between the number of known (training or zero) data and the number of information symbols in both cases. To provide performance as independent as possible from the specific channel realization, we averaged the BER over 1000 independent realizations of Rayleigh fading channels modeled as FIR filters of order  $L = 6$ . We used Hadamard precoding vectors of length  $M = 8$  in our system and a Wiener equalizer of length 30 in the alternative scheme. In both cases, we considered transmission of  $N = 4P = 56$  blocks of  $P = M + L$  symbols. From Fig. 5, we notice that the scheme using precoding and guard zeros outperforms the conventional scheme for SNR greater than roughly 4 dB, i.e., when the ISI becomes the dominant source

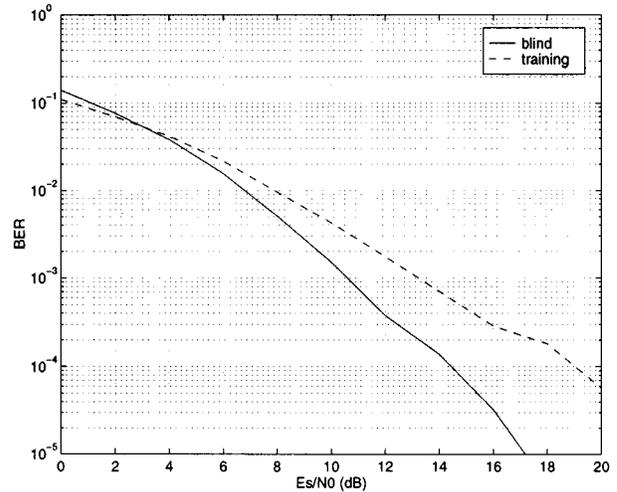


Fig. 5. Comparison between transmission strategies using null guard intervals and block equalization or training sequences and Wiener equalizer using the same number of redundant symbols for channel estimation.

of performance degradation, even though we used a relatively long Wiener filter to limit the ISI.

In fact, the scheme with null guard intervals prevents interblock interference and is capable, in principle, of eliminating ISI perfectly when the channel estimate is sufficiently good. Conversely, the scheme with training sequence is able to provide a better channel estimate, but it cannot offer zero ISI, unless an IIR equalizer is employed. Therefore, it is not surprising that at high SNR, i.e., when ISI is the dominant source of performance degradation, the scheme with null guard intervals outperforms the FIR Wiener equalizer. Clearly, the scheme based on training can be improved by using a decision feedback (DF) equalizer instead of the finite length Wiener filter. However, the precoder-based system can be improved as well by adding a block DF equalizer [15]. Interestingly, the block DF equalizer coupled with the presence of null guard intervals is capable of providing another important advantage. In fact, as shown in [15], the block DF is able to prevent the catastrophic error propagation phenomenon of scalar DF equalizers.

## VII. CONCLUDING REMARKS

The most common form of redundancy introduced for channel estimation and synchronization in digital communication systems relies on periodic transmission of training sequences. As an alternative to this standard approach, in [12] and [13], we proposed transmission of data blocks interlaced with null guard intervals. Although simple, this approach was shown in [13] to guarantee i) FIR zero-forcing equalization, irrespective of the underlying FIR channel zeros for a host of single-user transmission schemes and downlink multiple access systems and ii) the existence of deterministic and statistical channel estimation or direct equalization algorithms that require small data sizes and exhibit robustness to channel order overestimation.

In this paper, we proved that the channel estimation algorithm of [13] is statistically efficient at high SNR for any data block

length and quantified the resilience of the method to additive noise and to slow channel fluctuations induced by Doppler effects and oscillators' phase noise. We showed that the method can be applied to systems adopting aperiodic precoding and in the presence of a frequency asynchronism between transmitter and receiver. Finally, since the null guard intervals could be interpreted as a form of distributed training, we compared the scheme with an alternative scheme allocating all the known (nonzero) data at the beginning, in the form of a known preamble for channel acquisition, and then using a Wiener equalizer. We showed that even though the system with the long training sequence is capable of providing a better channel estimate, the scheme with null guard interval yields lower BER, at least at high SNR when ISI dominates, because it is capable of removing ISI completely. The other important competing alternative approach is OFDM with pilot tones for channel estimation and cyclic prefix. In such a case, a block of  $P = 2M + 2L$  would be sufficient to estimate the coefficients of an  $L$ th-order channel response, with the same efficiency  $M/(M + L)$  as our method using null guard intervals. However, OFDM with cyclic prefix does not guarantee perfect equalization when the channel impulse response has zeros on the unit circle, as opposed to our system, which also guarantees ZF equalization in such a case. The price paid for this advantage is essentially increased complexity because of the SVD required for the channel estimate.

#### APPENDIX A HANKEL MATRICES OF $\hat{\mathbf{U}} = \tilde{\mathbf{U}}\mathbf{Q}$

Let the  $P \times L$  matrices  $\hat{\mathbf{U}}$  and  $\tilde{\mathbf{U}}$  be related through the  $L \times L$  matrix  $\mathbf{Q}$  as  $\hat{\mathbf{U}} = \tilde{\mathbf{U}}\mathbf{Q}$ . Define the block matrix  $\hat{\mathbf{U}} := [\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_L]$ , where each  $\hat{\mathbf{U}}_l$  is the  $(L + 1) \times M$  Hankel matrix built with the  $L$  columns  $\hat{\mathbf{u}}_l$  of  $\hat{\mathbf{U}}$  and, similarly, the block matrix  $\tilde{\mathbf{U}} := [\tilde{\mathbf{U}}_1, \dots, \tilde{\mathbf{U}}_L]$  built from the columns  $\tilde{\mathbf{u}}_l$  of  $\tilde{\mathbf{U}}$ . We wish to prove that

$$\hat{\mathbf{U}} = \tilde{\mathbf{U}}\mathbf{B} := \tilde{\mathbf{U}}\mathbf{P}\mathbf{Q}\mathbf{P}^H \quad (57)$$

where  $\mathbf{P}$  is an  $(ML) \times (ML)$  permutation matrix, and  $\mathbf{Q} = \mathbf{I}_{M \times M} \otimes \mathbf{Q}$ .

Starting with the Hankel matrix  $\hat{\mathbf{U}}_l$  built from  $\hat{\mathbf{u}}_l$  as in (7), we can rearrange the columns of  $\hat{\mathbf{U}} = [\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_L]$  in order to form  $M$  blocks of dimension  $(L + 1) \times L$  according to the following rule. The  $k$ th block contains the  $k$ th column of the  $L$  matrices  $\hat{\mathbf{U}}_l$ ,  $l = 1, \dots, L$ . This operation is performed by multiplying  $\hat{\mathbf{U}}$  from the right-hand side by an  $(ML) \times (ML)$  permutation matrix  $\mathbf{P}$  obtained by permuting the columns of the identity matrix  $\mathbf{I}_{ML \times ML}$ , according to the aforementioned rule. The  $M$  blocks of the matrix  $\hat{\mathbf{U}}\mathbf{P}$  are composed of rows of the matrix  $\hat{\mathbf{U}}$ . Specifically

$$\hat{\mathbf{U}}\mathbf{P} = \begin{pmatrix} \hat{\mathbf{u}}(1) & \hat{\mathbf{u}}(2) & \dots & \hat{\mathbf{u}}(M) \\ \hat{\mathbf{u}}(2) & \hat{\mathbf{u}}(3) & \dots & \hat{\mathbf{u}}(M+1) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\mathbf{u}}(L+1) & \hat{\mathbf{u}}(L+2) & \dots & \hat{\mathbf{u}}(P) \end{pmatrix} \quad (58)$$

where  $\hat{\mathbf{u}}(k)$  indicates the  $k$ th row of  $\hat{\mathbf{U}}$ . Because  $\hat{\mathbf{U}} = \tilde{\mathbf{U}}\mathbf{Q}$ , we have  $\hat{\mathbf{u}}(k) = \tilde{\mathbf{u}}(k)\mathbf{Q}$ , where  $\tilde{\mathbf{u}}(k)$  is the  $k$ th row of  $\tilde{\mathbf{U}}$  so that we can write

$$\hat{\mathbf{U}}\mathbf{P} = \begin{pmatrix} \tilde{\mathbf{u}}(1) & \tilde{\mathbf{u}}(2) & \dots & \tilde{\mathbf{u}}(M) \\ \tilde{\mathbf{u}}(2) & \tilde{\mathbf{u}}(3) & \dots & \tilde{\mathbf{u}}(M+1) \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{\mathbf{u}}(L+1) & \tilde{\mathbf{u}}(L+2) & \dots & \tilde{\mathbf{u}}(P) \end{pmatrix} \mathbf{Q} \quad (59)$$

where  $\mathbf{Q} := \mathbf{I}_{M \times M} \otimes \mathbf{Q}$ . The first matrix in (59) has the same form as the first matrix in (58); hence

$$\hat{\mathbf{U}}\mathbf{P} = \tilde{\mathbf{U}}\mathbf{P}\mathbf{Q}, \quad \text{or,} \quad \hat{\mathbf{U}} = \tilde{\mathbf{U}}\mathbf{P}\mathbf{Q}\mathbf{P}^T. \quad (60)$$

It is easy to verify that  $\mathbf{P}\mathbf{Q}\mathbf{P}^T$  is also a unitary matrix.

#### APPENDIX B CRAMÉR–RAO BOUND

In this appendix, we compute the entries of the matrix  $\mathbf{J}$  in (47). For simplicity, we express the channel matrix  $\mathbf{H}$  in (3) as

$$\mathbf{H} = \sum_{l=0}^L h(l)\mathbf{J}_l \quad (61)$$

where  $\mathbf{J}_l$  is a  $P \times M$  diagonal matrix with entries  $J_l(i, k) := \delta[i - k - l]$ , and  $\delta[i]$  denotes Kronecker's delta. The log-likelihood function in (46) can then be written as

$$f(\boldsymbol{\theta}) = -\log((2\pi\sigma_v^2)^{PN}) - \frac{\left\| \mathbf{y}(n) - \sum_{l=0}^L h(l)\mathbf{J}_l\mathbf{F}\mathbf{s}(n) \right\|^2}{\sigma_v^2}. \quad (62)$$

Taking partial derivatives with respect to the unknown parameters, we obtain

$$\begin{aligned} \frac{\partial f(\boldsymbol{\theta})}{\partial h^*(k)} &= -\frac{1}{\sigma_v^2} \sum_{n=1}^N \mathbf{s}^H(n)\mathbf{F}^H\mathbf{J}_k^T(\mathbf{y}(n) - \mathbf{H}\mathbf{F}\mathbf{s}(n)) \\ \frac{\partial f(\boldsymbol{\theta})}{\partial s_m(k)} &= -\frac{1}{\sigma_v^2} (\mathbf{y}^H(m) - \mathbf{s}^H(m)\mathbf{F}^H\mathbf{H}^H)\mathbf{H}\mathbf{F}\mathbf{e}_m \\ \frac{\partial f(\boldsymbol{\theta})}{\partial h(k)} &= \left( \frac{\partial f(\boldsymbol{\theta})}{\partial h^*(k)} \right)^* \\ \frac{\partial f(\boldsymbol{\theta})}{\partial s_m^*(k)} &= \left( \frac{\partial f(\boldsymbol{\theta})}{\partial s_m(k)} \right)^* \end{aligned} \quad (63)$$

where  $\mathbf{e}_m$  is the  $m$ th canonical basis vector having all entries equal to zero except the  $m$ th one. Because the noise is white and circularly symmetric, we have

$$\begin{aligned} E \left\{ \frac{\partial f(\boldsymbol{\theta})}{\partial h^*(k)} \frac{\partial f(\boldsymbol{\theta})}{\partial h(m)} \right\} &= \frac{1}{\sigma_v^2} \sum_{n=1}^N \mathbf{s}^H(n)\mathbf{F}^H\mathbf{J}_k^T\mathbf{J}_m\mathbf{F}\mathbf{s}(n) \\ E \left\{ \frac{\partial f(\boldsymbol{\theta})}{\partial h^*(k)} \frac{\partial f(\boldsymbol{\theta})}{\partial s_l(m)} \right\} &= \frac{1}{\sigma_v^2} \mathbf{s}^H(m)\mathbf{F}^H\mathbf{J}_k^T\mathbf{H}\mathbf{F}\mathbf{e}_l \\ E \left\{ \frac{\partial f(\boldsymbol{\theta})}{\partial s_j^*(k)} \frac{\partial f(\boldsymbol{\theta})}{\partial s_l(m)} \right\} &= \frac{1}{\sigma_v^2} \mathbf{e}_j^T\mathbf{F}^H\mathbf{H}^H\mathbf{H}\mathbf{F}\mathbf{e}_l\delta[k - m] \\ E \left\{ \frac{\partial f(\boldsymbol{\theta})}{\partial h^*(k)} \frac{\partial f(\boldsymbol{\theta})}{\partial h^*(j)} \right\} &= E \left\{ \frac{\partial f(\boldsymbol{\theta})}{\partial h(k)} \frac{\partial f(\boldsymbol{\theta})}{\partial h(j)} \right\} = 0, \quad \forall k, j \\ E \left\{ \frac{\partial f(\boldsymbol{\theta})}{\partial s_m(k)} \frac{\partial f(\boldsymbol{\theta})}{\partial s_n(j)} \right\} &= E \left\{ \frac{\partial f(\boldsymbol{\theta})}{\partial s_m^*(k)} \frac{\partial f(\boldsymbol{\theta})}{\partial s_n^*(j)} \right\} = 0 \\ &\quad \forall m, n, k, j. \end{aligned} \quad (64)$$

The equivalent Fisher's information matrix defined in (47) then assumes the form

$$\mathcal{J} = \begin{pmatrix} \mathcal{J}_{1,1} & \mathcal{J}_{1,2} & \mathbf{0} & \mathbf{0} \\ \mathcal{J}_{1,2}^H & \mathcal{J}_{2,2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{J}_{1,1} & \mathcal{J}_{1,2} \\ \mathbf{0} & \mathbf{0} & \mathcal{J}_{1,2}^H & \mathcal{J}_{2,2} \end{pmatrix} \quad (65)$$

where the blocks have the following dimensions:  $\mathcal{J}_{1,1}$  is  $L \times L$ ,  $\mathcal{J}_{1,2}$  is  $L \times NP$ , and  $\mathcal{J}_{2,2}$  is  $NP \times NP$ . Given this block diagonal structure, we can simply invert the submatrix

$$\bar{\mathcal{J}} = \begin{pmatrix} \mathcal{J}_{1,1} & \mathcal{J}_{1,2} \\ \mathcal{J}_{1,2}^H & \mathcal{J}_{2,2} \end{pmatrix}. \quad (66)$$

In particular, using (64), we have

$$\begin{aligned} \mathcal{J}_{1,1} &= \frac{1}{\sigma_v^2} \sum_{n=1}^N \begin{pmatrix} \mathbf{s}^H(n) \mathbf{F}^H \mathbf{J}_0^T \\ \vdots \\ \mathbf{s}^H(n) \mathbf{F}^H \mathbf{J}_{L-1}^T \end{pmatrix} \\ &\quad \cdot (\mathbf{J}_0 \mathbf{F} \mathbf{s}(n), \dots, \mathbf{J}_{L-1} \mathbf{F} \mathbf{s}(n)) \\ \mathcal{J}_{1,2} &= \frac{1}{\sigma_v^2} \begin{pmatrix} \text{vec}^T (S_N^H \mathbf{F}^H \mathbf{J}_0^T \mathbf{H} \mathbf{F}) \\ \vdots \\ \text{vec}^T (S_N^H \mathbf{F}^H \mathbf{J}_{L-1}^T \mathbf{H} \mathbf{F}) \end{pmatrix} \\ \mathcal{J}_{2,2} &= \frac{1}{\sigma_v^2} \mathbf{I}_{N \times N} \otimes (\mathbf{F}^H \mathbf{H} \mathbf{H} \mathbf{F}). \end{aligned} \quad (67)$$

As discussed in Section IV,  $\bar{\mathcal{J}}$  is not invertible, but to recover invertibility, it is sufficient to remove one row and the corresponding column from  $\bar{\mathcal{J}}$ . Specifically, if we normalize the channel impulse response with respect to its  $d$ th entry, we may remove the  $d$ th column and row from  $\bar{\mathcal{J}}$ . In formulas, this contraction yields the compacted Fisher's matrix

$$\check{\mathcal{J}} = \begin{pmatrix} \mathbf{E}_d \mathcal{J}_{1,1} \mathbf{E}_d^T & \mathbf{E}_d \mathcal{J}_{1,2} \\ \mathcal{J}_{1,2}^H \mathbf{E}_d^T & \mathcal{J}_{2,2} \end{pmatrix} \quad (68)$$

where  $\mathbf{E}_d$  is the  $L \times (L+1)$  matrix obtained by removing the  $d$ th row from the  $(L+1) \times (L+1)$  identity matrix. Since we are interested in the covariance matrix of the channel estimate, we can also compute directly the upper left portion of the inverse of  $\check{\mathcal{J}}$ . In force of **a2)**,  $\mathbf{F}$  is always invertible, and thus,  $\mathcal{J}_{2,2}$  in (67) is also *always* invertible. Hence, using the block inversion formula [8], we derive the CRB  $\mathbf{C}_{CR}$  for the channel estimate as

$$\mathbf{C}_{CR} = (\mathbf{E}_d \mathbf{D} \mathbf{E}_d^T)^{-1}, \quad \mathbf{D} := \mathcal{J}_{1,1} - \mathcal{J}_{1,2} \mathcal{J}_{2,2}^{-1} \mathcal{J}_{1,2}^H. \quad (69)$$

Substituting (67) in (69), we find that the  $(k, l)$  entry of the matrix  $\mathbf{A} := \mathcal{J}_{1,2} \mathcal{J}_{2,2}^{-1} \mathcal{J}_{1,2}^H$  is given by

$$A(k, l) = \frac{1}{\sigma_v^2} \sum_{n=1}^N \mathbf{s}^H(n) \mathbf{F}^H \mathbf{J}_k^T \mathbf{H} \mathbf{H}^\dagger \mathbf{J}_l \mathbf{F} \mathbf{s}(n). \quad (70)$$

Therefore, the  $(k, l)$  element of the matrix  $\mathbf{D}$  is

$$D(k, l) = \frac{1}{\sigma_v^2} \sum_{n=1}^N \mathbf{s}^H(n) \mathbf{F}^H \mathbf{J}_k^T (\mathbf{I} - \mathbf{H} \mathbf{H}^\dagger) \mathbf{J}_l \mathbf{F} \mathbf{s}(n). \quad (71)$$

Introducing the SVD of the channel matrix  $\mathbf{H}$ , with a notation consistent with the one used in (4)

$$\mathbf{H} = (\bar{\mathbf{U}}, \tilde{\mathbf{U}}) \begin{pmatrix} \Sigma \\ \mathbf{0} \end{pmatrix} \mathbf{V}^H \quad (72)$$

we have  $\mathbf{H} \mathbf{H}^\dagger = \bar{\mathbf{U}} \bar{\mathbf{U}}^H$  or, equivalently,  $\mathbf{I} - \mathbf{H} \mathbf{H}^\dagger = \tilde{\mathbf{U}} \tilde{\mathbf{U}}^H$ . Hence, (71) simplifies to

$$D(k, l) = \frac{1}{\sigma_v^2} \sum_{n=1}^N \mathbf{s}^H(n) \mathbf{F}^H \mathbf{J}_k^T \tilde{\mathbf{U}} \tilde{\mathbf{U}}^H \mathbf{J}_l \mathbf{F} \mathbf{s}(n). \quad (73)$$

For  $k = 0, \dots, L$  and  $l = 0, \dots, L$ , the summands in (73) can be organized into an  $(L+1) \times (L+1)$  matrix  $\Xi(n)$ , which is defined as

$$\begin{aligned} \Xi(n) &:= (\mathbf{I}_{(L+1) \times (L+1)} \otimes \mathbf{F} \mathbf{s}(n))^H \begin{pmatrix} \mathbf{J}_0^T \tilde{\mathbf{U}} \\ \vdots \\ \mathbf{J}_L^T \tilde{\mathbf{U}} \end{pmatrix} \\ &\quad \cdot (\tilde{\mathbf{U}}^H \mathbf{J}_0, \dots, \tilde{\mathbf{U}}^H \mathbf{J}_L) (\mathbf{I}_{(L+1) \times (L+1)} \otimes \mathbf{F} \mathbf{s}(n)) \end{aligned} \quad (74)$$

and such that

$$\mathbf{D} = \frac{1}{\sigma_v^2} \sum_{n=1}^N \Xi(n). \quad (75)$$

Using (7), after some tedious but otherwise straightforward manipulations, (74) can be rewritten as

$$\begin{aligned} \Xi(n) &:= (\mathbf{I}_{(L+1) \times (L+1)} \otimes \mathbf{F} \mathbf{s}(n))^H \\ &\quad \cdot (\text{vec}(\tilde{\mathbf{u}}_1^T), \dots, \text{vec}(\tilde{\mathbf{u}}_L^T)) \\ &\quad \cdot \begin{pmatrix} \text{vec}^H(\tilde{\mathbf{u}}_1^T) \\ \vdots \\ \text{vec}^H(\tilde{\mathbf{u}}_L^T) \end{pmatrix} (\mathbf{I}_{(L+1) \times (L+1)} \otimes \mathbf{F} \mathbf{s}(n)). \end{aligned} \quad (76)$$

Again, using the property  $\text{vec}(\mathbf{A} \mathbf{C} \mathbf{B}) = (\mathbf{B}^T \otimes \mathbf{A}) \text{vec}(\mathbf{C})$ , every product  $(\mathbf{I}_{(L+1) \times (L+1)} \otimes \mathbf{F} \mathbf{s}(n))^H \text{vec}(\tilde{\mathbf{u}}_l^T)$  can be written as

$$\begin{aligned} &(\mathbf{I}_{(L+1) \times (L+1)} \otimes (\mathbf{F} \mathbf{s}(n))^H) \text{vec}(\tilde{\mathbf{u}}_l^T) \\ &= \text{vec}(\mathbf{s}^H(n) \mathbf{F}^H \tilde{\mathbf{u}}_l^T) = \tilde{\mathbf{u}}_l \mathbf{F}^* \mathbf{s}^*(n) \end{aligned} \quad (77)$$

and using the matrix  $\tilde{\mathbf{u}} := (\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_L)$ , (76) becomes

$$\begin{aligned} \Xi(n) &:= [\tilde{\mathbf{u}}_1 \mathbf{F}^* \mathbf{s}^*(n), \dots, \tilde{\mathbf{u}}_L \mathbf{F}^* \mathbf{s}^*(n)] \\ &\quad \cdot [\tilde{\mathbf{u}}_1 \mathbf{F}^* \mathbf{s}^*(n), \dots, \tilde{\mathbf{u}}_L \mathbf{F}^* \mathbf{s}^*(n)]^H \\ &= \tilde{\mathbf{u}} (\mathbf{I}_{L \times L} \otimes \mathbf{F}^* \mathbf{s}^*(n)) (\mathbf{I}_{L \times L} \otimes \mathbf{F} \mathbf{s}(n)) \tilde{\mathbf{u}}^H \\ &= \tilde{\mathbf{u}} (\mathbf{I}_{L \times L} \otimes \mathbf{F}^* \mathbf{s}^*(n) \mathbf{s}^T(n) \mathbf{F}^T) \tilde{\mathbf{u}}^H. \end{aligned} \quad (78)$$

Using (75), we then have

$$\begin{aligned} \mathbf{D} &= \frac{1}{\sigma_v^2} \tilde{\mathbf{u}} \left[ \mathbf{I}_{L \times L} \otimes \mathbf{F}^* \left( \sum_{n=1}^N \mathbf{s}^*(n) \mathbf{s}^T(n) \right) \mathbf{F}^T \right] \tilde{\mathbf{u}}^H \\ &= \tilde{\mathbf{u}} [\mathbf{I}_{L \times L} \otimes (\mathbf{F}^* \mathbf{S}_N^* \mathbf{S}_N^T \mathbf{F}^T)] \tilde{\mathbf{u}}^H. \end{aligned} \quad (79)$$

Therefore, plugging (79) in (69) and recalling that  $\tilde{\mathbf{V}} := \mathbf{E}_d \tilde{\mathbf{U}}$ , we end up with the final formula for the CRB

$$\begin{aligned} \mathbf{C}_{CR} &= \frac{1}{\sigma_v^2} \left[ \tilde{\mathbf{V}} [\mathbf{I}_{L \times L} \otimes (\mathbf{F}^* \mathbf{S}_N^* \mathbf{S}_N^T \mathbf{F}^T)] \tilde{\mathbf{V}}^H \right]^{-1} \\ &= \frac{1}{\sigma_v^2} \tilde{\mathbf{V}}^{\dagger H} \left[ \mathbf{I}_{L \times L} \otimes (\mathbf{F}^* \mathbf{S}_N^* \mathbf{S}_N^T \mathbf{F}^T)^{-1} \right] \tilde{\mathbf{V}}^{\dagger} \\ &= \frac{1}{\sigma_v^2} \tilde{\mathbf{V}}^{\dagger H} \left[ \mathbf{I}_{L \times L} \otimes (\mathbf{F}^{-T} (\mathbf{S}_N^* \mathbf{S}_N^T)^{-1} \mathbf{F}^{-*}) \right] \tilde{\mathbf{V}}^{\dagger}. \quad (80) \end{aligned}$$

## REFERENCES

- [1] C. Becchetti, G. Scarano, and G. Jacovitti, "Fisher information analysis for blind channel estimation based on second order statistics," in *Proc. Ninth IEEE Statist. Signal Array Process.*, Portland, OR, Sept. 14–16, 1998, pp. 312–315.
- [2] A. Belouchrani and W. Ren, "Blind carrier phase tracking with guaranteed global convergence," *IEEE Trans. Signal Processing*, vol. 45, pp. 1889–1894, July 1997.
- [3] E. De Carvalho and D. T. M. Slock, "Cramer–Rao bounds for semi-blind, blind and training sequence based channel estimation," in *Proc. First IEEE Signal Process. Workshop Signal Process. Adv. Wireless Commun.*, Paris, France, Apr. 16–18, 1997, pp. 129–132.
- [4] Eur. Telecommun. Stand., "Radio broadcast systems: Digital audio broadcasting (DAB) to mobile, portable, and fixed receivers," ETSI, pr ETS 300 401, Mar. 1994.
- [5] —, "Radio broadcast system for television, sound and data services; framing structures, channel coding and modulation for digital terrestrial television," ETSI, pr ETS 300 744, Mar. 1997.
- [6] G. D. Forney, Jr. and M. V. Eyuboğlu, "Combined equalization and coding using precoding," *IEEE Commun. Mag.*, pp. 25–34, Dec. 1991.
- [7] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing*, vol. 44, pp. 661–672, Mar. 1996.
- [8] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, second ed. San Diego, CA: Academic, 1985, p. 410.
- [9] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communications," *Signal Process.*, vol. 43, pp. 2714–2723, Nov. 1995.
- [10] E. Nikula, A. Toskala, E. Dahlman, L. Girard, and A. Klein, "FRAMES multiple access for UMTS and IMT-2000," *IEEE Pers. Commun.*, pp. 16–24, Apr. 1998.
- [11] J. G. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1989.
- [12] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers Part I: Unification and optimal designs," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988–2006, July 1999.
- [13] —, "Redundant filterbank precoders and equalizers—Part II: Blind channel estimation synchronization and direct equalization," *IEEE Trans. Signal Processing*, vol. 47, pp. 2007–2022, July 1999.
- [14] —, "Minimum redundancy filterbank precoder for blind channel identification irrespective of channel nulls," in *Proc. Wireless Commun. Netw. Conf.*, New Orleans, LA, Sept. 21–24, 1999.
- [15] A. Stamoulis, G. B. Giannakis, and A. Scaglione, "Self-recovering transceivers for block transmissions: Filterbank precoders and decision-feedback equalizers," in *Proc. 2nd IEEE Workshop Signal Process. Commun.*, Annapolis, MD, May 9–12, 1999, pp. 243–246.
- [16] P. Stoica and B. C. Ng, "On the Cramér–Rao bound under parametric constraints," *IEEE Signal Processing Lett.*, vol. 5, pp. 177–179, July 1998.
- [17] A. van den Bos, "A Cramér–Rao lower bound for complex parameters," *IEEE Trans. Signal Processing*, vol. 42, p. 2859, Oct. 1994.
- [18] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, pp. 2982–2993, Dec. 1995.

**Sergio Barbarossa** (M'88) received the B.S. degree (cum laude) in electrical engineering in 1984 and the Ph.D. degree in information and communication engineering in 1989, both from the University of Rome "La Sapienza," Rome, Italy.

He joined the Radar System Division of Selenia in 1985 as a radar system engineer. From November 1987 to August 1988, he was a Research Engineer at the Environmental Research Institute of Michigan (ERIM), Ann Arbor, where he was involved in research activities on synthetic aperture radars. After three years at the University of Perugia, Perugia, Italy, as Adjunct Professor, in 1991, he joined the University of Rome "La Sapienza," where he is now a Full Professor. He has held positions as Visiting Scientist and Visiting Professor at the University of Virginia, Charlottesville, and at the University of Minnesota, Minneapolis. His current research interests lie in the areas of signal processing for digital communications, multiple access methods, space–time coding, and communications over time-varying channels.

Dr. Barbarossa has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and, since 1997, he has been a member of the IEEE Technical Committee for Signal Processing in Communications. He is co-recipient of the 2000 IEEE Best Paper Award in the Signal Processing for Communications area.

**Anna Scaglione** (M'99) received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Rome, "La Sapienza," Rome, Italy, in 1995 and 1999, respectively.

In August 2000, she joined the Electrical and Computer Engineering Department, University of New Mexico, Albuquerque, as Assistant Professor. In 1999, she was with the the University of Minnesota, Minneapolis, as Postdoctoral Researcher. Her background is in statistical signal processing and digital communications. Her research field is in signal processing for communications, and her main publications are on the optimization of digital modulation and multiplexing techniques for frequency-selective fading channels and broadband transmission.

Dr. Scaglione, with co-authors, received the 2000 IEEE Signal Processing Best Paper Award.

**Georgios B. Giannakis** (F'97) received the Diploma degree in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1981. From September 1982 to July 1986, he was with the University of Southern California (USC), Los Angeles, where he received the M.Sc. degree in electrical engineering in 1983, M.Sc. degree in mathematics in 1986, and the Ph.D. degree in electrical engineering in 1986.

After lecturing for one year at USC, he joined the University of Virginia, Charlottesville, in 1987, where he became a Professor of electrical engineering in 1997. Since 1999, he has been with the University of Minnesota, Minneapolis, as a Professor of electrical and computer engineering. His general interests span the areas of communications and signal processing, estimation and detection theory, time-series analysis, and system identification—subjects on which he has published more than 120 journal papers, 250 conference papers, and two edited books. Current research topics focus on transmitter and receiver diversity techniques for single- and multiuser fading communication channels, redundant precoding, and space–time coding for block transmissions, multicarrier, and wideband wireless communication systems. He is a frequent consultant for the telecommunications industry.

Dr. Giannakis is the (co-) recipient of Best Paper Awards from the IEEE Signal Processing (SP) Society in 1992, 1998, and 2000. He also received the Society's Technical Achievement Award in 2000. He co-organized three IEEE-SP Workshops (HOS in 1993, SSAP in 1996, and SPAWC in 1997) and guest (co-) edited four special issues. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, a secretary of the SP Conference Board, a member of the SP Publications Board, and a member and vice chair of the Statistical Signal and Array Processing Committee. He is a member of the Editorial Board for the PROCEEDINGS OF THE IEEE, he chairs the SP for Communications Technical Committee, and he serves as the Editor in Chief for the IEEE SIGNAL PROCESSING LETTERS. He is a member of the IEEE Fellows Election Committee and the IEEE-SP Society's Board of Governors.