Blind Channel Identification and Equalization Using Periodic Modulation Precoders: Performance Analysis

Antoine Chevreuil, Erchin Serpedin, Philippe Loubaton, *Member, IEEE,* and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—The present paper deals with blind identification and equalization of communication channels within the so-called modulation induced cyclostationarity (MIC) framework, where the input symbol stream is modulated by a P periodic precoder with the purpose of inducing cyclostationarity in the transmit sequence. By exploiting the cyclostationarity induced by the periodic precoder, a subspace-based channel identification algorithm that is resilient to the location of channel zeros, channel order overestimation errors, and color of additive stationary noise, is developed. The asymptotic performance of the subspace-based identification approach is analyzed and compared with the asymptotic lower bound provided by the nonlinear cyclic correlation matching approach. Criteria for optimally designing the periodic precoder are also presented. The performance of MMSE-FIR and MMSE-DFE equalizers is quantified for the proposed cyclostationarity-induced framework in terms of the MMSE. Although cyclostationarity-inducing transmitters present several advantages relative to their stationary counterparts from a channel estimation viewpoint, it is shown that from an equalization viewpoint, MIC-based systems exhibit a slightly increased MMSE/BER when compared with the stationary case.**

***Index Terms*—Channels, equalizers, estimation, multipath, signal processing.**

I. Introduction

RECENT results [7], [11], [28] have established that blind identification of finite impulse response (FIR) single-input single-output (SISO) communication channels is possible only from the output second-order statistics, without using any restrictive assumption on the channel zeros, color of additive noise, channel order overestimation errors, and without increasing the transmission rate of the data stream. The common feature of this class of approaches [7], [11], [28] is that they induce cyclostationary (CS) statistics at the transmitter by means of a periodically time-varying precoder (e.g., a filterbank [11], [28] or a simple periodic modulator [7]) and exploit the resulting second-order cyclostationary statistics of the received samples. These algorithms have been termed transmitter-induced cyclostationarity (TIC) approaches.

In the present paper, we analyze the performance of a special class of TIC-precoders, namely, those using a periodic modulator in order to induce CS-statistics in the transmitted sequence. Henceforth, we will refer to this class of algorithms as modulation-induced cyclostationarity (MIC) approaches because the CS statistics are induced by modulating the input symbol stream with a deterministic strictly periodic sequence. We study the large sample behavior of two channel estimators, propose several criteria for optimally designing the periodic modulation precoder, and evaluate the performance of MIC-based approaches from a bit error rate (BER)—minimum-mean square error (MMSE) equalization viewpoint.

MIC-based approaches allow estimation of FIR channels from the output second-order cyclic statistics without imposing any restrictive assumptions on the channel. Although MIC approaches offer a net advantage over fractionally spaced (FS)-based approaches [16], [23], [26], [27] in terms of channel identifiability, it turns out that MIC approaches entail a slight deterioration relative to the stationary (no MIC precoder) case in terms of mean-square error (MSE) at the output of a Wiener equalizer. We show that for MIC-based transmissions, the MMSE at the output of a Wiener equalizer is slightly larger than the MMSE value corresponding to the stationary case. Equivalently, the MIC precoder implies a slight loss in power efficiency relative to the stationary framework. In the case of a memoryless channel, we also show that the MIC-based transmissions entail an increase of the symbol error rate when compared with the stationary transmission. However, the ability of MIC approaches to blindly estimate any FIR channel under very general conditions and without decreasing the transmission rate motivates a thorough analysis of the MIC framework.

The paper is organized as follows. Section II introduces the communication channel setup and the MIC precoder. Section III presents two channel identification algorithms: the subspace-based and the nonlinear correlation matching approaches. It is shown that the subspace approach is robust to channel order overestimation errors, location of channel zeros, and color of additive stationary noise. The large sample performance analysis of the proposed channel estimators is also derived. Criteria for optimally designing the periodic precoder are presented in Section IV. In Section V, the performance of MIC-based approaches

Manuscript received November 3, 1998; revised November 17, 1999. The work of A. Chevreuil was supported by INRIA, The French National Institute for Research in Computer Science and Control. The work of E. Serpedin and G. B. Giannakis was supported in part by NSF Grant NSF-MIP 9424305. Parts of this work have been published in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Seattle, WA, March 1998. The associate editor coordinating the review of this paper and approving it for publication was Dr. Jean Jacques Fuchs.

A. Chevreuil and P. Loubaton are with the Laboratoire "Systèmes de Communication," Université de Marne-la-Vallée, Marne-la-Vallée, France.

E. Serpedin is with the Department of Electrical Engineering, Texas A&M University, College Station, TX 77843 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(00)04055-1.

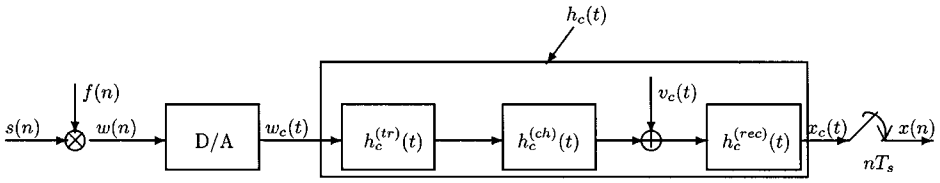


Fig. 1. Baseband transmission channel.

is analyzed from an equalization viewpoint. Comparisons with stationary transmissions are also presented. Computer simulations are presented in Section VI in order to corroborate the theoretical developments, and last, conclusions are outlined in Section VII.

II. Preliminaries

We consider the general communication channel of Fig. 1, where the independently and identically distributed (i.i.d.) symbol stream $s(n)$ is modulated by the periodic sequence $f(n)$ ($f(n) = f(n+P)$, $\forall n$). Multiplication of the symbol stream with the periodic sequence $f(n)$ forms the periodic modulation precoder, which is the basis of the MIC framework. Its role is to induce cyclostationary statistics in the transmit baseband sequence $w(n) := f(n)s(n)$. By defining the time-varying correlation of $w(n)$ at time n and lag τ via $c_{ww}(n;\tau) := \mathbb{E}w^*(n)w(n+\tau)$, it follows that $c_{ww}(n;\tau) = c_{ww}(n+lP;\tau) = \sigma_s^2|f(n)|^2\delta(\tau)$, $\forall n, l$ ($\sigma_s^2 := \mathbb{E}|s(n)|^2$, and superscript $*$ stands for conjugation). The sequence $w(n)$ is converted by the digital-to-analog (D/A) converter into the continuous time waveform $w_c(t) := \sum_n w(n)\delta(t-nT_s)$. The resulting $w_c(t)$ is pulse shaped by the transmit filter $h_c^{(tr)}(t)$ and propagates through the unknown channel $h_c^{(ch)}(t)$ and the receive filter $h_c^{(rec)}(t)$. At the receiver, the continuous-time waveform $x_c(t)$ is sampled at the symbol-rate $1/T_s$. Defining the composite channel response $h_c(t) := (h_c^{(rec)} \star h_c^{(ch)} \star h_c^{(tr)})(t)$, the continuous-time waveform $x_c(t)$ at the channel output can be expressed as

$$x_c(t) = \sum_n w(n)h_c(t-nT_s) + v_c(t) \tag{1}$$

where $v_c(t)$ stands for additive stationary noise, independently distributed from $s(n)$. After sampling at the symbol rate $1/T_s$ and supposing $h_c(t)$ has finite support LT_s, we obtain from (1) the discrete-time channel model shown in Fig. 2 and described by the input-output (I/O) relation

$$x(n) = \sum_{l=0}^{L} h(l)w(n-l) + v(n) \tag{2}$$

with $h(n) := h_c(nT_s)$, and $v(n) := v_c(nT_s)$, $\forall n$. Since $w(n)$ is CS, it follows that $x(n) := x_1(n) + v(n)$ is also CS; its time-varying correlation satisfies $c_{xx}(n;\tau) := \mathbb{E}x^*(n)x(n+\tau) = c_{xx}(n+P;\tau)$, $\forall n, \tau$. Being periodic, $c_{xx}(n;\tau)$ accepts a Fourier series expansion over the set of complex exponentials with harmonic cycles $2\pi k/P$, $k = 0, \cdots, P-1$. The coefficients of the Fourier series expansion are called cyclic correlation coefficients and are given by

$$C_{xx}(k;\tau) := \frac{1}{P}\sum_{n=0}^{P-1} c_{xx}(n;\tau)e^{-j2\pi kn/P}. \tag{3}$$

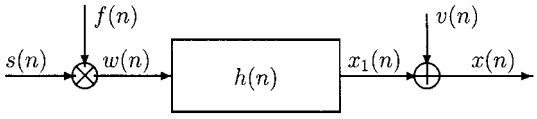


Fig. 2. Baseband discrete-time channel.

Henceforth, we consider working only with nonzero cycles $k \neq 0$ so that the contribution of the additive stationary noise $v(n)$ is cancelled out in the cyclic domain. It is easy to check that

$$C_{xx}(k;\tau) := \sigma_s^2 F_2(k) \sum_{m=-\infty}^{\infty} h^*(m)h(m+\tau)e^{-j2\pi km/P} \tag{4}$$

with

$$F_2(k) := \frac{1}{P}\sum_{n=0}^{P-1} |f(n)|^2 e^{-j2\pi kn/P}. \tag{5}$$

The Fourier transform of the cyclic correlation sequence $\{C_{xx}(k;\tau)\}_{\tau=-\infty}^{\infty}$, for a fixed cycle k, is called cyclic spectrum at cycle k and, for $k \neq 0$, it is given [c.f. (4)] by

$$S_{xx}(k;z) = \sigma_s^2 F_2(k)H(z)H^*(z^{-1}e^{-j2\pi k/P}) \tag{6}$$

with $H(z) := \sum_{l=0}^{L} h(l)z^{-l}$ denoting the channel transfer function, and $H^*(z) := \sum_{l=0}^{L} h^*(l)z^{-l}$. After these preliminaries, we now consider developing blind channel identification algorithms based only on the information provided by the output second-order cyclic statistics.

III. Blind Second-Order Channel Identification

In this section, a subspace-based channel identification approach will be developed. We will first introduce some definitions and notations. Then, we will establish the identifiability condition of the subspace approach, and we will close with the subspace and correlation matching algorithms.

Define for any vector $\boldsymbol{a} := [a_1 \cdots a_r]^T$ the $r(r-1)/2 \times r$ matrix $\mathcal{F}(\boldsymbol{a})$ via

$$\mathcal{F}(\boldsymbol{a}) := \left[\begin{array}{cccccc} -a_2 & a_1 & 0 & \cdots & 0 & 0 \\ -a_3 & 0 & a_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ -a_r & 0 & 0 & \cdots & 0 & a_1 \\ \hline 0 & -a_3 & a_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & -a_r & 0 & \cdots & 0 & a_2 \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline 0 & \cdots & \cdots & \cdots & -a_r & a_{r-1} \end{array}\right]. \tag{7}$$

Denote by $\mathbb{C}[z]$ the set of polynomials in z^{-1} over $\mathbb{C}$ and by $\mathbb{C}^r[z]$ the set of r-dimensional polynomial vectors: $\boldsymbol{p}(z) := [p_1(z)\cdots p_r(z)]^T$, with $p_k(z) \in \mathbb{C}[z]$, $1 \le k \le r$. Let $d(z)$ be the greatest common divisor (gcd) of the set of polynomials $p_k(z)$, $1 \le k \le r$; then, the irreducible part of $\boldsymbol{p}(z)$ is defined by $\overline{\boldsymbol{p}}(z) := [p_1(z)/d(z)\cdots p_r(z)/d(z)]^T$ [1]. Using the definitions of $\boldsymbol{p}(z)$ and its irreducible part $\overline{\boldsymbol{p}}(z)$, from (7), it follows easily that

$$\mathcal{F}(\overline{\boldsymbol{p}}(z))\boldsymbol{p}(z) = \boldsymbol{0} \tag{8}$$

holds for any z. This will be of particular importance in the following section.

A. Channel Identifiability

Let us collect all the cyclic spectra (6) corresponding to $k \neq 0$ in the $(P-1)\times 1$ vector

$$\boldsymbol{s}_{xx}(z) = \left[S_{xx}^*\left(1; z^{-1}\right), S_{xx}^*\left(2; z^{-1}\right) \cdots, S_{xx}^*\left(P-1; z^{-1}\right)\right]^T \tag{9}$$

and use (6) to express (9) as

$$\boldsymbol{s}_{xx}(z) = \sigma_s^2 \boldsymbol{h}_f(z) H^*\left(z^{-1}\right) \tag{10}$$

where $\boldsymbol{h}_f(z)$ is a $(P-1)\times 1$ vector given by

$$\boldsymbol{h}_f(z) := \left[F_2^*(1)H\left(ze^{-j2\pi/P}\right), F_2^*(2)H\left(ze^{-j4\pi/P}\right), \cdots F_2^*(P-1)H\left(ze^{-j2\pi(P-1)/P}\right)\right]^T. \tag{11}$$

A simple identifiability condition is obtained by noting that if the period P is chosen such that $P > L+1$, then the polynomials $\{H(ze^{-j2\pi k/P})\}_{k=1}^{P-1}$ are coprime, and hence, $H(z)$ can be extracted as the gcd of cyclic spectra $S_{xx}(k;z)$, $k = 1, \cdots, P-1$ [11]. Alternatively, it is possible to estimate each component of $\boldsymbol{h}_f(z)$ by using the noise subspace method of [16] or the least-squares approach of [29] (which is called the signal subspace approach in [1]) and then to estimate $H(z)$ by fitting the structure of $\boldsymbol{h}_f(z)$ to its "noise or signal" subspace. However, the order L of the channel $h(n)$ is usually unknown in practice, and in the presence of channel order mismatch errors, all these approaches are likely to fail. Therefore, we propose estimating the unknown channel $h(n)$ from $\boldsymbol{s}_{xx}(z)$ by using more specifically the structure of $\boldsymbol{h}_f(z)$. We will show that the resulting approach is quite robust to channel order overestimation errors. Henceforth, the condition $P > L+1$ is assumed, which implies that $\boldsymbol{h}_f(z)$ is the irreducible part of $\boldsymbol{s}_{xx}(z)$. Due to (8), we have

$$\mathcal{F}(\boldsymbol{h}_f(z))\boldsymbol{s}_{xx}(z) = \boldsymbol{0}. \tag{12}$$

Equation (12) suggests identifying the unknown channel $H(z)$ as the polynomial $Y(z)$ that satisfies the subspace equation

$$\mathcal{F}(\boldsymbol{y}_f(z))\boldsymbol{s}_{xx}(z) = \boldsymbol{0} \tag{13}$$

where $\boldsymbol{y}_f(z)$ is a $(P-1)$-dimensional rational vector structured similar to $\boldsymbol{h}_f(z)$ [see (11)]

$$\boldsymbol{y}_f(z) = \left[F_2^*(1)Y\left(ze^{-j2\pi/P}\right), F_2^*(2)Y\left(ze^{-j4\pi/P}\right), \cdots F_2^*(P-1)Y\left(ze^{-j2\pi(P-1)/P}\right)\right]^T \tag{14}$$

with $Y(z) := \sum_{l=0}^{\hat{L}} y(l)z^{-l}$, and $\hat{L}$ denoting an upper bound (or estimate) for the unknown channel order L. As in [1], (13) can be interpreted this way: The solution is chosen so that the row span of $\mathcal{F}(\boldsymbol{y}_f(z))$ belongs to the left kernel of $\boldsymbol{s}_{xx}(z)$. Hence, (13) is called signal subspace equation. Due to the particular structure of the solution—recall that $\boldsymbol{y}_f(z)$ is parameterized by (14)—(13) is called structured signal subspace equation. In the next subsection, we will rewrite (13) in the equivalent form $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$, where $\boldsymbol{y} := [y(0)\cdots y(\hat{L})]^T$ stands for the vector of unknown coefficients, and $\boldsymbol{A}$ is a matrix with entries that depend only on the available $\boldsymbol{s}_{xx}(z)$. This equivalence can be easily understood if we note that all the entries of $\mathcal{F}(\boldsymbol{y}_f^T(z))\boldsymbol{s}_{xx}(z)$ depend linearly on the coefficients $y(l)$, $l = 0, \cdots, \hat{L}$ of the unknown vector $\boldsymbol{y}$. Thus, finding $\boldsymbol{y}$ from (15) is equivalent to solving the system of equations

$$\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}. \tag{15}$$

Estimation of the vector of channel coefficients from (15) will be referred to as a subspace estimation approach. Before presenting the expression for $\boldsymbol{A}$, we show next that the channel $h(n)$ can be uniquely identified from (13) [or equivalently (15)]. Due to (12), $\boldsymbol{y} = \boldsymbol{h}$ is a solution of (13). Therefore, it suffices to show that $\boldsymbol{h}$ is the unique (within a scale factor) solution of (13). Denote by $\boldsymbol{y}$ an arbitrary solution of (13). Because all entries of $\mathcal{F}(\boldsymbol{y}_f(z))\boldsymbol{s}_{xx}(z)$ are of the form $F_2^*(k)F_2^*(l)H^*(z^{-1})[Y(ze^{-j2\pi k/P})H(ze^{-j2\pi l/P}) - Y(ze^{-j2\pi l/P})H(ze^{-j2\pi k/P})]$, with $1 \le k, l \le P-1$, and assuming[1] $F_2(k) \neq 0$, for $\forall\ k$, $1 \le k \le P-1$, we have

$$Y\left(ze^{-j2\pi k/P}\right)H\left(ze^{-j2\pi l/P}\right) - Y\left(ze^{-j2\pi l/P}\right)H\left(ze^{-j2\pi k/P}\right) = 0$$

which can be written as

$$\frac{Y\left(ze^{-j2\pi k/P}\right)}{H\left(ze^{-j2\pi k/P}\right)} = \frac{Y\left(ze^{-j2\pi l/P}\right)}{H\left(ze^{-j2\pi l/P}\right)} \quad \forall\, k, l = 1, \cdots, P-1. \tag{16}$$

It is easy to establish that (16) is equivalent to

$$\frac{Y(z)}{H(z)} = \frac{Y\left(ze^{-j2\pi/P}\right)}{H\left(ze^{-j2\pi/P}\right)} \qquad \forall\, z. \tag{17}$$

Indeed, note that by making the change of variable $z \leftrightarrow z\exp(j2\pi k/P)$ in (16) and assuming $k-l = -1$, (17) is obtained. In addition, by repeatedly making the change of variables $z \leftrightarrow z\exp(-j2\pi/P)$ in (17), it follows that (16) holds true for any $1 \le k, l \le P-1$. Define $B(z) := Y(z)/H(z)$. From (17), it follows that $B(z) = B(z\exp(-j2\pi/P))$. As $h(0) \neq 0$, function $B(z)$ admits a power series expansion around $z = \infty$, i.e., $B(z) = \sum_{k\ge 0} b(k)z^{-k}$. Since $B(z) = B(z\exp(-j2\pi/P))$, it follows that $b(k) = 0$ if $k \neq 0$

[1]The reader may check that the identifiability analysis still holds as soon as there exist $1 \le k_0 \le l_0 \le P-1$ such that 1) $F_2(k_0)$ and $F_2(l_0)$ are both nonzero and 2) that $l_0 - k_0$ and P are coprime. The more constraining condition that $F_2(k) \neq 0$ for all k is used here for the sake of simplicity.

modulo P, i.e., $B(z) = \sum_{k=0}^{\infty} b(Pk)z^{-Pk}$. Thus, the identity $Y(z) = B(z)H(z)$ takes the form

$$Y(z) = b(0)H(z) + b(P)z^{-P}H(z) + \cdots + b(nP)z^{-nP}H(z) + \cdots. \quad (18)$$

As $P > L+1$, no cancellation can occur between the various polynomials $b(0)H(z)$, $b(P)z^{-P}H(z), \cdots, b(nP)z^{-nP}H(z), \cdots$ in the summation (18). Thus, $Y(z)$ is a polynomial if and only if $B(z)$ is of finite degree. In this case, all the solutions of (17) are of the form $Y(z) = b(0)H(z) + b(P)z^{-P}H(z) + \cdots + b(nP)z^{-nP}H(z)$, and the degree of $Y(z)$ is $\hat{L} = L + nP$. It follows that the channel $H(z)$ is uniquely recoverable from (13) up to a scale factor if $L \leq \deg Y(z) = \hat{L} < L+P$. Thus, if the overestimated order $\hat{L}$ lies in the interval $[L, L+P)$, the channel can be identified uniquely up to a scale factor. Suppose now that $\hat{L} = L + nP \geq L + P$. Because $P > L+1$, it is interesting to note that the structure of $\boldsymbol{y}$ consists of scaled periodic replicas of the channel vector $\boldsymbol{h}$. Since the period of the repetition is known (P), it follows that the vector channel $\boldsymbol{h}$ can be obtained by selecting only the first P components of $\boldsymbol{y}$. Of course, any subvector of the form $[y(kP+1)\cdots y(kP+L+1)]^T$, $k = 0, \cdots, n-1$ also yields $\boldsymbol{h}$. The relation $P > L+1$ will be referred to as the subspace channel identifiability condition. Summarizing, we have established the following.

Theorem 1—Subspace-Based Channel Identifiability Result: Under the assumption that the period of modulating sequence is chosen such that $P > L+1$, any FIR channel can be identified from its output second-order cyclic spectra at nonnull cyclic frequencies, irrespective of channel order over-estimation errors.

Contrary to FS approaches, Theorem 1 does not constrain the location of channel zeros, and only an upper bound on the channel order needs to be available. We also remark that a similar channel identifiability condition has been derived before in [7] and [22], using a different approach. The main motivation for introducing the present subspace framework is that it leads to an "interpretable" solution for the asymptotic covariance of the channel estimate, as we will see in Section III-C. Having established the channel identifiability condition for the subspace approach, we next elaborate on the corresponding channel estimation algorithm.

B. Subspace Algorithm

Let $\overline{L}$ be an integer such that $\overline{L} \geq \hat{L}$. The polynomial vector $\boldsymbol{s}_{xx}(z)$ given by (9) admits the power series expansion

$$\boldsymbol{s}_{xx}(z) = \sum_{\tau=-\overline{L}}^{\overline{L}} \boldsymbol{c}(\tau)z^{-\tau}$$

where we set

$$\boldsymbol{c}(\tau) := [C_{xx}^*(1;-\tau)\ \ C_{xx}^*(2;-\tau)\cdots C_{xx}^*(P-1;-\tau)]^T. \quad (19)$$

Let $\boldsymbol{e}(z)$ be the $(P-1)(P-2)/2 \times 1$ vector defined as

$$\boldsymbol{e}(z) := \mathcal{F}(\boldsymbol{y}_f(z))\boldsymbol{s}_{xx}(z) = \sum_{\tau=-\overline{L}}^{\overline{L}+\hat{L}} \boldsymbol{e}(\tau)z^{-\tau}. \quad (20)$$

The unknown polynomial vector $\boldsymbol{y}_f(z)$ can be written as $\boldsymbol{y}_f(z) = \sum_{\tau=0}^{\hat{L}} \boldsymbol{y}_f(\tau)z^{-\tau}$, where

$$\boldsymbol{y}_f(\tau) := [F_2^*(1)y(\tau)e^{j2\pi\tau/P}\cdots F_2^*(P-1)y(\tau)exp^{j2\pi(P-1)\tau/P}]^T.$$

In addition, denote by $\boldsymbol{c}$ the $(P-1)(2\overline{L}+1) \times 1$ vector $\boldsymbol{c} := [\boldsymbol{c}^T(\overline{L}), \cdots, \boldsymbol{c}^T(-\overline{L})]^T$. In the time domain, (20) can be rewritten in a matrix form as in (21), shown at the bottom of the page.

One of this paper's objectives is to optimize the design of the periodic sequence $f(n)$ according to well-defined criteria. In the upcoming section, we propose selecting $f(n)$ to minimize the asymptotic covariance matrix of the subspace channel estimate. It is thus desirable to deal with a tractable expression of this asymptotic covariance matrix. As we will see below, this will be the case if the central part of (21) is considered. It is clear that if we choose $\overline{L} \geq 2\hat{L}$, then $\boldsymbol{c}(\tau) = \boldsymbol{0}$ for $|\tau| > \hat{L}$. In this case, the upper and lower parts of (21) are automatically satisfied. Henceforth, the condition $\overline{L} \geq 2\hat{L}$ is supposed. We set $\boldsymbol{e} := [\boldsymbol{e}^T(\overline{L}), \cdots, \boldsymbol{e}^T(-\overline{L}+\hat{L})]^T$ and define the $(2\overline{L}+1-\hat{L}) \times (P-1)(2\overline{L}+1)$ block Hankel matrix $\mathcal{D}(\boldsymbol{c})$ and the $(2\overline{L}+1-\hat{L}) \times (P-1)(\hat{L}+1)$ block Sylvester matrix $\mathcal{T}(\boldsymbol{y}_f)$ as

$$\mathcal{D}(\boldsymbol{c}) := \begin{bmatrix} \mathcal{F}(\boldsymbol{c}(\overline{L})) & \cdots & \cdots & \mathcal{F}(\boldsymbol{c}(\overline{L}-\hat{L})) \\ \vdots & \cdots & \cdots & \vdots \\ \mathcal{F}(\boldsymbol{c}(-\overline{L}+\hat{L})) & \cdots & \cdots & \mathcal{F}(\boldsymbol{c}(-\overline{L})) \end{bmatrix} \quad (22)$$

$$\left[\begin{array}{c} \boldsymbol{e}(\overline{L}+\hat{L}) \\ \vdots \\ \boldsymbol{e}(\overline{L}+1) \\ \hline \boldsymbol{e}(\overline{L}) \\ \vdots \\ \boldsymbol{e}(-\overline{L}+\hat{L}) \\ \hline \boldsymbol{e}(-\overline{L}+\hat{L}-1) \\ \vdots \\ \boldsymbol{e}(-\overline{L}) \end{array}\right] = \left[\begin{array}{cccccc} \mathcal{F}(\boldsymbol{y}_f(\hat{L})) & \boldsymbol{0} & \cdots & \cdots & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \cdots & \cdots & \vdots \\ \mathcal{F}(\boldsymbol{y}_f(1)) & \cdots & \mathcal{F}(\boldsymbol{y}_f(\hat{L})) & \cdots & \cdots & \boldsymbol{0} \\ \hline \mathcal{F}(\boldsymbol{y}_f(0)) & \cdots & \cdots & \mathcal{F}(\boldsymbol{y}_f(\hat{L})) & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \mathcal{F}(\boldsymbol{y}_f(0)) & \cdots & \cdots & \mathcal{F}(\boldsymbol{y}_f(\hat{L})) \\ \hline \boldsymbol{0} & \cdots & \boldsymbol{0} & \mathcal{F}(\boldsymbol{y}_f(0)) & \cdots & \mathcal{F}(\boldsymbol{y}_f(\hat{L}-1)) \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \cdots & \cdots & \cdots & \mathcal{F}(\boldsymbol{y}_f(0)) \end{array}\right] \boldsymbol{c} \quad (21)$$

and (23), shown at the bottom of the page, where we set $\boldsymbol{y}_f := [\boldsymbol{y}_f^T(0), \cdots, \boldsymbol{y}_f^T(\hat{L})]^T$. Simple algebraic manipulations show that

$$\boldsymbol{e} = \mathcal{T}(\boldsymbol{y}_f)\boldsymbol{c} = -\mathcal{D}(\boldsymbol{c})\boldsymbol{y}_f. \tag{24}$$

The parameterization of $\boldsymbol{y}_f(z)$ given by (14) provides in the time domain the relation

$$\boldsymbol{y}_f = \mathcal{P}\boldsymbol{y} \tag{25}$$

where $\boldsymbol{y}$ is the $(\hat{L}+1)\times 1$ vector $\boldsymbol{y} = [y(0)\cdots y(\hat{L})]^T$, and matrix $\mathcal{P}$ is block diagonal, with its (q, q) block entry given by

$$[\mathcal{P}]_{q,q} = \begin{bmatrix} F_2^*(1)e^{j2\pi q/P} \\ \vdots \\ F_2^*(P-1)e^{j2\pi q(P-1)/P} \end{bmatrix}. \tag{26}$$

Due to (21), (24), and (25), (13) can be rewritten in the form $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$, with $\boldsymbol{A} := \mathcal{D}(\boldsymbol{c})\mathcal{P}$. Let the superscript H stand for transposition and conjugation. From Theorem 1, (21), (24), and (25), we infer the following.

Proposition 1—Structured Subspace Identifiability Result: Time Domain: If $P > L+1$ and $L \le \hat{L} < L+P$, the quadratic form

$$\boldsymbol{Q} = \mathcal{P}^H\mathcal{D}^H(\boldsymbol{c})\mathcal{D}(\boldsymbol{c})\mathcal{P} \tag{27}$$

admits a kernel of dimension 1 spanned by the $(\hat{L}+1)\times 1$ vector $[h(0), \cdots, h(L), 0\cdots 0]^T$.

In practice, the cyclic correlations involved in $\boldsymbol{c}$ have to be estimated from a finite number of samples. Denote by T the number of available samples. A consistent estimate for the cyclic correlation at cycle k and lag τ is given by [8]

$$\hat{C}_{xx}(k;\tau) := \frac{1}{T}\sum_{n=0}^{T-1} x^*(n)x(n+\tau)e^{-j2\pi kn/P}. \tag{28}$$

Denote by $\hat{\boldsymbol{c}}$ an estimate for the vector of cyclic correlations $\boldsymbol{c}$, and define the matrix $\hat{\boldsymbol{Q}} := \mathcal{P}^H\mathcal{D}^H(\hat{\boldsymbol{c}})\mathcal{D}(\hat{\boldsymbol{c}})\mathcal{P}$. An estimate $\hat{\boldsymbol{h}}$ (up to a scale factor) of $\boldsymbol{h}$ can be found as a unit-norm eigenvector corresponding to the smallest eigenvalue of $\hat{\boldsymbol{Q}}$. The subspace-based channel estimation algorithm, which is outlined in Proposition 1, appears attractive from a computational viewpoint since it requires only one SVD of the reduced-dimension $(\hat{L}+1)\times(\hat{L}+1)$ matrix $\hat{\boldsymbol{Q}}$, as opposed to most FS-based subspace methods (e.g., [1]), which require two SVD's. This subspace algorithm also allows a tractable expression for its asymptotic covariance matrix, which is a result that will be shown later on to be useful for selecting the periodic precoder $f(n)$.

C. Asymptotic Performance Analysis

Because most of blind identification approaches allow estimation of the unknown channel under a scale factor ambiguity, the nonuniqueness of the channel estimate prevents us from computing the asymptotic performance. It is common to artificially remove the scale factor ambiguity of the subspace channel estimate by imposing a supplementary constraint. Denote by $\hat{\Pi}$ the orthogonal projector onto the one-dimensional subspace spanned by an eigenvector associated with the smallest eigenvalue. Define

$$\hat{\boldsymbol{h}} := \hat{\boldsymbol{\Pi}}\boldsymbol{h} \tag{29}$$

and $\boldsymbol{d}_{\overline{L}}(e^{j2\pi f}) := [1, \exp(-j2\pi f), \cdots, \exp(-j2\pi(\overline{L}-1)f)]^T$. We have the following asymptotic result.

Theorem 2—Asymptotic Performance of Subspace Approach: Let γ_{4s} be the kurtosis of $s(n)$, and

$$F_4(k) := \frac{1}{P}\sum_{n=0}^{P-1} |f(n)|^4 e^{-j2\pi kn/P}. \tag{30}$$

Define the $(P-1)\times(P-1)$ matrices $\boldsymbol{M}(\ell_1; e^{j2\pi f})$ and $\boldsymbol{\Theta}(e^{j2\pi\nu_1}, e^{j2\pi\nu_2})$, with (p, q) entries

$$\left[\boldsymbol{M}\left(\ell_1; e^{j2\pi f}\right)\right]_{p,q} = S_{xx}\left(\ell_1 + q - p; e^{j2\pi(f-(p/P))}\right) \tag{31}$$

$$\begin{aligned}&\boldsymbol{\Theta}(e^{j2\pi\nu_1}, e^{j2\pi\nu_2})_{p,q} \\ &= \theta(p, q)H\left(e^{j2\pi(\nu_1-(p/P))}\right)H^*\left(e^{j2\pi(\nu_2-(q/P))}\right) \\ &\quad\times H\left(e^{j2\pi\nu_2}\right)H^*\left(e^{j2\pi\nu_1}\right)\end{aligned} \tag{32}$$

and $\theta(p, q) := \gamma_{4s}F_4(p-q)$. In addition, define the matrices

$$\begin{aligned}&\boldsymbol{U}_1\left(\ell_1; e^{j2\pi f}\right) \\ &\quad:= \mathcal{F}\left(\boldsymbol{h}_f\left(e^{j2\pi f}\right)\right)\boldsymbol{M}\left(\ell_1; e^{j2\pi f}\right) \\ &\qquad\times S_{xx}^*\left(\ell_1; e^{j2\pi f}\right)\mathcal{F}^H\left(\boldsymbol{h}_f\left(e^{j2\pi(f-(\ell_1/P))}\right)\right) \\ &\boldsymbol{U}_2\left(e^{j2\pi\nu_1}, e^{j2\pi\nu_2}\right) \\ &\quad:= \mathcal{F}\left(\boldsymbol{h}_f\left(e^{j2\pi\nu_1}\right)\right)\boldsymbol{\Theta}\left(e^{j2\pi\nu_1}, e^{j2\pi\nu_2}\right) \\ &\qquad\times\mathcal{F}^H\left(\boldsymbol{h}_f\left(e^{j2\pi\nu_2}\right)\right)\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Sigma}_1 := &\sum_{\ell_1=0}^{P-1} e^{j2\pi(\ell_1\overline{L}/P)}\int_0^1 \boldsymbol{d}_{2\overline{L}-\hat{L}}\left(e^{j2\pi f}\right) \\ &\times\boldsymbol{d}_{2\overline{L}-\hat{L}}^H\left(e^{j2\pi(f-(\ell_1/P))}\right)\otimes\boldsymbol{U}_1\left(\ell_1; e^{j2\pi f}\right)df\end{aligned}$$

$$\begin{aligned}\boldsymbol{\Sigma}_2 := &\gamma_{4s}\int_0^1\int_0^1 \boldsymbol{d}_{2\overline{L}-\hat{L}}(e^{j2\pi\nu_1})\boldsymbol{d}_{2\overline{L}-\hat{L}}^H(e^{j2\pi\nu_2}) \\ &\otimes\boldsymbol{U}_2\left(e^{j2\pi\nu_1}, e^{j2\pi\nu_2}\right)d\nu_1\,d\nu_2\end{aligned}$$

where $\otimes$ stands for Kronecker product. Suppose that the symbols $s(n)$ are circularly distributed and $v(n)$ is normally distributed. Then, the asymptotic covariance of $\mathcal{T}(\boldsymbol{h})\hat{\boldsymbol{c}}$ is given by

$$\mathcal{T}(\boldsymbol{y}_f) := \begin{bmatrix} \mathcal{F}(\boldsymbol{y}_f(0)) & \cdots & \cdots & \mathcal{F}(\boldsymbol{y}_f(\hat{L})) & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \mathcal{F}(\boldsymbol{y}_f(0)) & \cdots & \cdots & \mathcal{F}(\boldsymbol{y}_f(\hat{L})) \end{bmatrix} \tag{23}$$

$\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2$. Moreover, the estimate $\hat{\boldsymbol{h}}$ given by (29) is a consistent estimate of $\boldsymbol{h}$ with asymptotic covariance

$$\lim_{T\to\infty} T\mathbb{E}\left[(\hat{\boldsymbol{h}}-\boldsymbol{h})(\hat{\boldsymbol{h}}-\boldsymbol{h})^H\right]=\boldsymbol{\Gamma}_1+\boldsymbol{\Gamma}_2 \tag{33}$$

where

$$\boldsymbol{\Gamma}_k=\boldsymbol{Q}^{\dagger}\mathcal{P}^H\mathcal{D}^H(\boldsymbol{c})\boldsymbol{\Sigma}_k\mathcal{D}(\boldsymbol{c})\mathcal{P}\boldsymbol{Q}^{\dagger},\ \text{for } k=1,\,2$$

where $\dagger$ stands for the Moore–Penrose pseudo-inverse, and T denotes the number of available samples.

Proof: See the Appendix. □

Of course, $\hat{\boldsymbol{h}}$ in (29) coincides with the estimate provided by the subspace method *up to a complex scale factor*. However, the evaluation of the asymptotic covariance matrix of $\hat{\boldsymbol{h}}$ provides useful information on the behavior of the subspace approach. We note that the above trick is implicit in most of the previous works [1], [15] dealing with the asymptotic performance of subspace-like methods. Numerical simulations have also shown that almost the same asymptotic variance for the channel estimates is achieved when different scaling operations are performed on the channel coefficients (e.g., unit norm channel vector or first channel tap fixed).

D. Nonlinear Correlation Matching

In the present subsection, we investigate an alternative channel estimation approach, namely, the nonlinear cyclic correlation matching approach introduced in [12] in the fractional sampling framework. For cyclic correlation matching, it suffices to use only the nonredundant information provided by the output second order cyclic correlations. From (4), it follows that

$$C_{xx}(k;m)=e^{j2\pi km/P}C^*_{xx}(P-k;-m) \tag{34}$$

$$C_{xx}(k;m)=0,\ \forall\ |m|>L. \tag{35}$$

The conjugate symmetry in the lags (34) and memory constraint (35) imply that the entire nonredundant second-order statistical information about the complex-valued channel is contained in the set $\mathcal{K}:=\{C_{xx}(k;m);k=1,\,2,\,\cdots,\,[P/2],\,m=-L,\,\cdots,\,L\}$. Therefore, it is sufficient that the cyclic correlations associated with the set $\mathcal{K}$ be matched to their sample estimates. The cyclic correlations associated with the set $\mathcal{K}$ and their sample estimates are collected in the vectors $\boldsymbol{c}_{xx}:=[C_{xx}(1;-L),\,\cdots,\,C_{xx}([P/2];L)]^T$, and $\hat{\boldsymbol{c}}_{xx}:=[\hat{C}_{xx}(1;-L),\,\cdots,\,\hat{C}_{xx}([P/2];L)]^T$, respectively. It is also useful to represent the complex-valued channel $\boldsymbol{h}$ and the vectors of cyclic correlations $\boldsymbol{c}_{xx}$, $\hat{\boldsymbol{c}}_{xx}$ in terms of their real and imaginary parts $\tilde{\boldsymbol{h}}:=[\mathrm{Re}(\boldsymbol{h})^T\,\mathrm{Im}(\boldsymbol{h})^T]^T$, $\tilde{\boldsymbol{c}}_{xx}:=[\mathrm{Re}(\boldsymbol{c}_{xx})^T\,\mathrm{Im}(\boldsymbol{c}_{xx})^T]^T$, $\hat{\tilde{\boldsymbol{c}}}_{xx}:=[\mathrm{Re}(\hat{\boldsymbol{c}}_{xx})^T\mathrm{Im}(\hat{\boldsymbol{c}}_{xx})^T]^T$. The nonlinear correlation matching estimator finds the estimate $\hat{\tilde{\boldsymbol{h}}}$ such that the cyclic correlations $\tilde{\boldsymbol{c}}_{xx}$ are closest to the observed correlations $\hat{\tilde{\boldsymbol{c}}}_{xx}$ in the weighted least-squares sense

$$\hat{\tilde{\boldsymbol{h}}}:=\arg\min_{\tilde{\boldsymbol{h}}} J\left[\hat{\tilde{\boldsymbol{c}}}_{xx};\tilde{\boldsymbol{h}}\right]$$

$$J\left[\hat{\tilde{\boldsymbol{c}}}_{xx};\tilde{\boldsymbol{h}}\right]:=\left[\hat{\tilde{\boldsymbol{c}}}_{xx}-\tilde{\boldsymbol{c}}_{xx}\right]^T\boldsymbol{R}\left[\hat{\tilde{\boldsymbol{c}}}_{xx}-\tilde{\boldsymbol{c}}_{xx}\right] \tag{36}$$

where the weighting matrix $\boldsymbol{R}$ is a positive-definite Hermitian matrix.

Although the nonlinear correlation matching approach has high computational complexity and local convergence problems, our motivation for studying it is due to the fact that the lowest asymptotic covariance matrix among all second-order statistics-based approaches is achieved by an optimally weighted nonlinear cyclic correlation matching approach. Thus, the performance of the weighted nonlinear correlation matching approach could serve as a benchmark for the subspace approach.

Before deriving the asymptotic covariance of the nonlinear correlation matching estimator, we next compute the asymptotic covariance matrix of the cyclic correlation estimate $\hat{\tilde{\boldsymbol{c}}}_{xx}$ defined as

$$\boldsymbol{\Sigma}:=\lim_{T\to\infty} T\mathbb{E}\left\{\left[\hat{\tilde{\boldsymbol{c}}}_{xx}-\tilde{\boldsymbol{c}}_{xx}\right]\left[\hat{\tilde{\boldsymbol{c}}}_{xx}-\tilde{\boldsymbol{c}}_{xx}\right]^T\right\}.$$

There are at least two ways to evaluate the normalized asymptotic covariance matrix $\boldsymbol{\Sigma}$. The first approach relies on the representation

$$\boldsymbol{\Sigma}=\begin{bmatrix}\mathrm{Re}\left\{\dfrac{\boldsymbol{\Gamma}_1+\boldsymbol{\Gamma}_2}{2}\right\} & \mathrm{Im}\left\{\dfrac{\boldsymbol{\Gamma}_1-\boldsymbol{\Gamma}_2}{2}\right\}\\ \mathrm{Im}\left\{\dfrac{\boldsymbol{\Gamma}_1+\boldsymbol{\Gamma}_2}{2}\right\} & \mathrm{Re}\left\{\dfrac{\boldsymbol{\Gamma}_2-\boldsymbol{\Gamma}_1}{2}\right\}\end{bmatrix}$$

with the matrices $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ representing the "unconjugated" asymptotic covariance matrix

$$\boldsymbol{\Gamma}_1:=\lim_{T\to\infty} T\mathbb{E}\left\{[\hat{\boldsymbol{c}}_{xx}-\boldsymbol{c}_{xx}][\hat{\boldsymbol{c}}_{xx}-\boldsymbol{c}_{xx}]^T\right\}$$

and the "conjugated" asymptotic covariance matrix

$$\boldsymbol{\Gamma}_2:=\lim_{T\to\infty} T\mathbb{E}\left\{[\hat{\boldsymbol{c}}_{xx}-\boldsymbol{c}_{xx}][\hat{\boldsymbol{c}}_{xx}-\boldsymbol{c}_{xx}]^H\right\}$$

respectively. By using several results in [8] and [12], the asymptotic covariance matrices $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ can be expressed in closed-form expressions under quite general conditions without requiring the circularity of the input $s(n)$ or Gaussianity of $v(n)$, as is the case with Lemma 1 (see the Appendix). However, it turns out that these asymptotic expressions are hard to implement numerically due to their high computational complexity. In order to avoid presenting intricate formulas, we next present a second approach for evaluating the asymptotic covariance matrix $\boldsymbol{\Sigma}$. Consider that all the cyclic correlations corresponding to all nonzero cycles and their sample estimates are collected in the vectors $C_{xx}:=[C_{xx}(1;-L)\cdots C_{xx}(P-1;L)]^T$ and $\hat{C}_{xx}$. In addition, consider the asymptotic covariance matrix $\boldsymbol{\Upsilon}:=\lim_{T\to\infty} T(\hat{C}_{xx}-C_{xx})(\hat{C}_{xx}-C_{xx})^H$, whose entries can be easily calculated using the closed-form formulas (62)–(64) from the Appendix. It is interesting to remark that once the matrix $\boldsymbol{\Upsilon}$ is computed, calculation of $\boldsymbol{\Sigma}$ can be greatly simplified by noting the dependence relation $\boldsymbol{\Sigma}=\boldsymbol{B}\boldsymbol{\Upsilon}\boldsymbol{B}^H$, where $\boldsymbol{B}$ is a matrix whose elements can be easily determined. Indeed, since $\mathrm{Re}C_{xx}(k;m)=(1/2)[C_{xx}(k;m)+C^*_{xx}(k;m)]$, according to (34), we have $\mathrm{Re}C_{xx}(k;m)=(1/2)[C_{xx}(k;m)+\exp(-j2\pi km/P)C_{xx}(P-k;-m)]$.

Similarly, we obtain $\mathrm{Im}C_{xx}(k;m) = (1/2j)[C_{xx}(k;m) - \exp(-j2\pi km/P)C_{xx}(P-k;-m)]$. We further infer that

$$\begin{aligned}
&\mathrm{Re}\hat{C}_{xx}(k;m) - \mathrm{Re}C_{xx}(k;m) \\
&\quad = \tfrac{1}{2}\Big[(\hat{C}_{xx}(k;m) - C_{xx}(k;m)) + e^{-(j2\pi km/P)} \\
&\qquad \cdot(\hat{C}_{xx}(P-k;-m) - C_{xx}(P-k;-m))\Big] \\
&\mathrm{Im}\hat{C}_{xx}(k;m) - \mathrm{Im}C_{xx}(k;m) \\
&\quad = \frac{1}{2j}\Big[(\hat{C}_{xx}(k;m) - C_{xx}(k;m)) - e^{-(j2\pi km/P)} \\
&\qquad \cdot(\hat{C}_{xx}(P-k;-m) - C_{xx}(P-k;-m))\Big]
\end{aligned}$$

which imply that there should exist a matrix $\boldsymbol{B}$ such that $\hat{\tilde{\boldsymbol{c}}}_{xx} - \tilde{\boldsymbol{c}}_{xx} = \boldsymbol{B}(\hat{C}_{xx} - C_{xx})$, and hence, $\boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Upsilon}\boldsymbol{B}^H$.

For establishing the asymptotic performance of the nonlinear correlation matching approach, the modulus one scale ambiguity inherent to the channel estimate is removed by considering a phase constraint on the channel vector. As a direct application of the standard results presented in [24, pp. 88–95] and [18, pp. 78–88], the following holds.

Theorem 3—Asymptotic Performance of the Nonlinear Correlation Matching Approach: The estimate $\hat{\tilde{\boldsymbol{h}}}$ obtained by minimizing $J[\hat{\tilde{\boldsymbol{c}}}_{xx};\ \boldsymbol{h}]$, under the phase constraint, converges in the mean-square sense to $\tilde{\boldsymbol{h}}$, provided that $s(n)$ has finite moments. The asymptotic covariance $\boldsymbol{\Sigma}(\hat{\tilde{\boldsymbol{h}}})$ of the correlation matching estimate (36) is given by

$$\boldsymbol{\Sigma}\left(\hat{\tilde{\boldsymbol{h}}}\right) = \lim_{T\to\infty} T\mathbb{E}\left(\hat{\tilde{\boldsymbol{h}}} - \tilde{\boldsymbol{h}}\right)\left(\hat{\tilde{\boldsymbol{h}}} - \tilde{\boldsymbol{h}}\right)^T = \boldsymbol{G}(\tilde{\boldsymbol{h}})\boldsymbol{\Sigma}\boldsymbol{G}^T(\tilde{\boldsymbol{h}}) \tag{37}$$

where

$$\boldsymbol{G}(\tilde{\boldsymbol{h}}) := \left[\boldsymbol{D}^T(\tilde{\boldsymbol{h}})\boldsymbol{R}\boldsymbol{D}(\tilde{\boldsymbol{h}})\right]^{-1}\boldsymbol{D}^T(\tilde{\boldsymbol{h}})\boldsymbol{R},$$

$$\boldsymbol{D}(\tilde{\boldsymbol{h}}) := [\nabla_{\tilde{h}(0)}\tilde{\boldsymbol{c}}_{xx}\cdots\nabla_{\tilde{h}(L)}\tilde{\boldsymbol{c}}_{xx}\nabla_{\tilde{h}(L+1)}\tilde{\boldsymbol{c}}_{xx}\cdots\nabla_{\tilde{h}(2L+1)}\tilde{\boldsymbol{c}}_{xx}]$$

provided that the Jacobian matrix $\boldsymbol{D}(\tilde{\boldsymbol{h}})$ is full column rank, and $\boldsymbol{R}$ is positive-definite.

When the normalized asymptotic covariance matrix $\boldsymbol{\Sigma}$ is nonsingular, it turns out [24, pp. 91–95] that the minimization of the asymptotic covariance matrix $\boldsymbol{\Sigma}(\hat{\tilde{\boldsymbol{h}}})$ is obtained by choosing the weighting matrix as $\boldsymbol{R} = \boldsymbol{\Sigma}^{-1}$. Thus, for the optimal weighting matrix $\boldsymbol{R}_{opt} = \boldsymbol{\Sigma}^{-1}$, the nonlinear correlation matching estimator is the asymptotically minimum variance (or asymptotically best consistent) estimator among the class of all unbiased estimators based on the output second-order cyclic statistics. When the asymptotic covariance matrix $\boldsymbol{\Sigma}$ is rank deficient, the "regularized" weighting matrix $\boldsymbol{R} = (\boldsymbol{\Sigma} + \boldsymbol{D}(\tilde{\boldsymbol{h}})\boldsymbol{D}(\tilde{\boldsymbol{h}})^T)^\dagger$ is usually adopted [24, pp. 88–90]. Next, we consider the problem of optimally designing the periodic modulating precoder $f(n)$.

IV. On the Precoder Design

The design of the optimal periodic precoder $f(n)$ will be performed under two conditions:

$$F_2(0) = \frac{1}{P}\sum_{k=0}^{P-1}|f(k)|^2 = 1 \tag{38}$$

$$\min_{0\le k\le P-1}|f(k)|^2 \ge \alpha, \qquad 0 < \alpha \le 1. \tag{39}$$

Condition (38) normalizes power and is natural, given the fact that all communication transmitters have power limitations. Condition (39) is necessary from an equalization viewpoint because modulating an input symbol with a zero value prevents us from recovering it at the receiver. We study the problem of optimally selecting the modulating sequence $f(n)$ from different perspectives. The first criterion we investigate consists of finding the periodic sequences for which the performance of the subspace channel estimator is optimal.

A. Minimum Asymptotic Covariance Criterion

Dependence of the asymptotic covariance matrix (33) on the modulating sequence $f(n)$ is quite involved, and the analytical minimization of trace$(\boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2)$, i.e., the asymptotic mean square error of the estimate, under (38) and (39) seems intractable. Interestingly, however, the next result shows that at least for high (ideally infinite) SNR, there exist periodic sequences $f(n)$ for which the asymptotic covariance matrix (33) is zero.

Proposition 2—Minimization of Asymptotic Covariance Matrix: In the noise-free case, the asymptotic covariance matrix of the subspace estimate $\hat{\boldsymbol{h}}$ in (29) vanishes if $f(n) = \delta(n)$ (mod P).

Proof: We show first that $\boldsymbol{\Sigma}_1$ is zero for this particular sequence. A similar result also holds for $\boldsymbol{\Sigma}_2$. It suffices to prove that the product $\mathcal{F}(\boldsymbol{h}_f(z))\boldsymbol{M}(\ell_1; e^{j2\pi f})$ is zero $\forall\ f,\ \ell_1$. The qth column of $\boldsymbol{M}(\ell_1; e^{j2\pi f})$ is

$$\begin{aligned}
&\boldsymbol{M}(\ell_1; e^{j2\pi f})\big|_{(:,\,q)} \\
&\quad = \begin{bmatrix} S_{xx}(\ell_1+q-1; e^{j2\pi(f-(1/P))}) \\ S_{xx}(\ell_1+q-2; e^{j2\pi(f-(2/P))}) \\ \vdots \\ S_{xx}(\ell_1+q-P+1; e^{j2\pi(f-(P-1/P))}) \end{bmatrix}
\end{aligned}$$

which, according to (6), can be rewritten as shown at the bottom of the next page.

An arbitrary entry of the column vector $\mathcal{F}(\boldsymbol{h}_f(z))\boldsymbol{M}(\ell_1;\ \exp(j2\pi f))|_{(:,\,q)}$ can be expressed as

$$\begin{aligned}
&[-F_2(k)F_2(\ell_1+q-l)+F_2(l)F_2(\ell_1+q-k)] \\
&\times H\left(e^{j2\pi(f-(l/P))}\right)H\left(e^{j2\pi(f-(k/P))}\right)H^*\left(e^{j2\pi(f-(\ell_1+q/P))}\right)
\end{aligned} \tag{40}$$

with $0 \le k,\ l \le P-1$. Since $f(n) = \delta(n)$ (mod P), it follows that $F_2(k) = F_2(0)$ for $\forall\, k = 0, \cdots, P-1$. Thus, we have

$$\mathcal{F}(\boldsymbol{h}_f(z))\boldsymbol{M}(\ell_1; e^{j2\pi f})\big|_{(:,\,q)} = \boldsymbol{0}.$$

Similar derivations prove that the other columns of $\mathcal{F}(\boldsymbol{h}_f(z))\boldsymbol{M}(\ell_1; e^{j2\pi f})$ are also null. □

Proposition 2 does not offer a good choice for the modulating sequence $f(n)$ since nulling $P-1$ out of P consecutive input symbols drops the bandwidth efficiency by a factor of P (in fact, $P-1$ of P symbols cannot be recovered). The result of Proposition 2 is not unexpected since it is known that for such scenarios, there exist deterministic algorithms [1]. The present subspace approach being statistical, Proposition 2 only states that the asymptotic covariance of $\hat{\boldsymbol{h}}$ is $O(1/T^{1+\epsilon})$, where $\epsilon > 0$. We have seen in the proof of Proposition 2 that zeroing the entries

of the asymptotic covariance matrix $\mathbf{\Gamma}_1 + \mathbf{\Gamma}_2$ is accomplished whenever all the values $F_2(k)$, $k = 0, \cdots, P-1$ are equal. However, the condition $F_2(k) = F_2(0)$ for $k = 1, \cdots, P-1$ holds iff $f(n)$ vanishes in $P-1$ values. By considering the relaxed condition $F_2(k) = F_2(1) \neq F_2(0)$ for $k = 2, \cdots, P-1$, we show that (39) can be fulfilled without nulling $f(n)$. We will see that the class of all periodic sequences $f(n)$ for which $F_2(k) = F_2(1) \neq F_2(0)$, $k = 2, \cdots, P-1$ can be parameterized as

$$\begin{aligned} |f(0)| &= \sqrt{P(1-\rho)+\rho} \\ |f(k)| &= \sqrt{\rho} \text{ for } k = 1, \cdots, P-1 \end{aligned} \tag{41}$$

where $0 \leq \rho \leq 1$ is an arbitrary scalar. By selecting $f(n)$ as in (41), we have $F_2(k) = 1-\rho, \forall\ k = 1, \cdots, P-1$. Due to (40), it follows that most of the entries in the qth column of $\mathcal{F}(\boldsymbol{h}_f(z))\boldsymbol{M}(\ell_1; e^{j2\pi f})$ are equal to zero. Numerical computations have shown that for such periodic sequences, the entries of the asymptotic covariance matrix assume small values that are close to zero (in absolute value). In addition, simulations with a large number of channels and periodic precoders satisfying (38) and (39) have pointed out that the minimum covariance of the subspace channel estimates is achieved for the periodic sequences (41), with $\rho = \alpha$, where α is the transmit power lower bound in (39). In the next subsection, we will elaborate more on the problem of selecting optimally $f(n)$ from a different viewpoint.

B. Maximum CS-Degree Criterion

Since the magnitude of the cyclic correlation $C_{xx}(k;\tau)$ for a given cycle $k \neq 0$ is proportional to $|F_2(k)|$ (4), it is natural to introduce the constants $|F_2(k)|$ as a measure of the cyclostationarity induced by the modulating sequence $f(n)$. We will refer to the value $|F_2(k)|$ as the degree of CS induced at cycle k in the transmitted baseband sequence $s(n)$ by the periodic precoder $f(n)$. A high value for $|F_2(k)|$ implies high values for the magnitudes of cyclic correlations $|C_{xx}(k;\tau)|$, $\forall\ \tau = -L, \cdots, L$. To justify the CS-degree criterion, we represent the estimated cyclic spectrum (6) as

$$\begin{aligned} \hat{S}_{xx}(k;z) &= \sigma_s^2 F_2(k) H(z) H^* \left(z^{-1} e^{-j2\pi k/P}\right) + S_{\nu\nu}(k;z) \\ &= S_{xx}(k;z) + S_{\nu\nu}(k;z) \end{aligned} \tag{42}$$

where the polynomial with random coefficients $S_{\nu\nu}(k;z)$ captures the finite sample/additive noise effects. We note that a small value of $|F_2(k)|(\approx 0)$ implies that $S_{xx}(k;z)$ assumes small values, and the additive component $S_{\nu\nu}(k;z)$ is expected to highly affect the factorization (6) of $\hat{S}_{xx}(k;z)$, and thus, performance of the estimation algorithm is expected to degrade.

The optimization problem that we pose is as follows: Find periodic sequences $\{f(n)\}_{n=1}^{P}$ that satisfy (38) and (39), and for a fixed k, $1 \leq k \leq P-1$, $|F_2(k)|$ is maximized. We refer to this optimization problem as the maximization of the CS-degree at the cycle k. We show the rather surprising result that whenever we maximize $|F_2(k)|$ at a fixed cycle $k \neq 0$, then $|F_2(l)|$ is also maximized for any other nonzero cycle $\forall\ l = 1, \cdots, P-1$, and moreover, the resulting maxima are all equal $|F_2(l)| = |F_2(k)|$, $\forall\ l = 1, \cdots, P-1$.

Proposition 3—Maximization of the CS-Degree Induced at a Nonzero Cycle: The optimal periodic precoders that induce the highest degree of CS under the constraints (38) and (39), are of the form (modulo a permutation and a phase-shift)

$$\{|f(n)|\}_{n=0}^{P-1} = \left\{\sqrt{P(1-\rho)+\rho}, \sqrt{\rho}, \cdots, \sqrt{\rho}\right\} \tag{43}$$

where $\rho = \alpha$, i.e., such sequences follow the parameterization already encountered in (41).

Proof: Consider an arbitrary periodic sequence $\{f(n)\}_{n=0}^{P-1}$ that satisfies the constraints (38) and (39), and define $\gamma := \min\{|f(n)|^2\}_{n=0}^{P-1} \geq \alpha$. For a fixed but arbitrary cycle $k \neq 0$, we have

$$\begin{aligned} F_2(k) &= \frac{1}{P} \sum_{n=0}^{P-1} |f(n)|^2 e^{-(j2\pi kn/P)} \\ &= \frac{1}{P} \sum_{n=0}^{P-1} \left(|f(n)|^2 - \gamma\right) e^{-(j2\pi kn/P)} \end{aligned} \tag{44}$$

where in deriving (44), we used $(1/P)\sum_{n=0}^{P-1} \exp(j2\pi kn/P) = \delta(k)$, $\forall\ k$. Since $|\sum_n z_n| \leq \sum_n |z_n|$, for any complex scalars z_n, we deduce from (44) that

$$|F_2(k)| \leq \frac{1}{P} \sum_{n=0}^{P-1} (|f(n)|^2 - \gamma) = 1 - \gamma \leq 1 - \alpha. \tag{45}$$

The last inequality in (45) is attained with equality provided that $\gamma = \alpha$ and the complex numbers $(|f(n)|^2 - \gamma^2)\exp(j2\pi kn/P)$, $\forall\ n = 0, \cdots, P-1$ and fixed k, $1 \leq k \leq P-1$ have the same phase or are, all but one, equal to zero. Because $F_2(k)$'s are not dependent on the phase of the modulating sequence, we obtain that all the optimal

$$\begin{aligned} &\begin{bmatrix} F_2(\ell_1+q-1)H\left(e^{j2\pi(f-(1/P))}\right) H^*\left(e^{j2\pi(f-(1/P)-(\ell_1+q-1/P))}\right) \\ F_2(\ell_1+q-2)H\left(e^{j2\pi(f-(2/P))}\right) H^*\left(e^{j2\pi(f-(2/P)-(\ell_1+q-2/P))}\right) \\ \vdots \\ F_2(\ell_1+q-P+1)H\left(e^{j2\pi(f-(P-1/P))}\right) H^*\left(e^{j2\pi(f-(P-1/P)-(\ell_1+q-P+1/P))}\right) \end{bmatrix} \\ &= \begin{bmatrix} F_2(\ell_1+q-1)H\left(e^{j2\pi(f-(1/P))}\right) \\ F_2(\ell_1+q-2)H\left(e^{j2\pi(f-(2/P))}\right) \\ \vdots \\ F_2(\ell_1+q-P+1)H\left(e^{j2\pi(f-(P-1/P))}\right) \end{bmatrix} \times H^*\left(e^{j2\pi(f-(\ell_1+q/P))}\right) \end{aligned}$$

sequences are given by (43) modulo a permutation and a phase factor. Since the maximization of $|F_2(k)|$, for a fixed $k \neq 0$, is independent of k, it follows that the maximization of $|F_2(k)|$, for a fixed $k \neq 0$, entails the maximization of $|F_2(l)|$ for $\forall\, l = 1, \cdots, P-1$. Note also that all these sequences satisfy $|F_2(l)| = 1-\alpha, \forall\, l = 1, \cdots, P-1$. □

As a conclusion, the choice of such sequences as the ones in (41) or (43) seems reasonable since, on the one hand, it corresponds to a maximum CS degree, and on the other hand, it makes most of the terms in the expression of the asymptotic covariance matrix vanish. In the following, we restrict the precoder $f(n)$ to belong to this class of functions.

V. Symbol Detection

In the previous sections, we have emphasized the benefit to expect from a MIC precoder on the blind channel estimation approach. In this respect, we have investigated the "best" modulating sequences. We have set forth a class of modulating functions described by (41) and characterized them by the single parameter ρ. Even though these functions are associated with an accurate channel estimation performance, a crucial point remains unclear: Independent of the channel estimation, what is the impact of the modulation precoder on the symbol detection? That is to say, assuming the unknown channel *perfectly estimated*, how does the BER behave as compared with a standard system (without precoding)?

A general answer is impossible because BER cannot be expressed analytically in closed form. In order to obtain partial results, we consider four particular problems, the solution of which gives an idea of how the BER is affected by MIC precoding. Specifically, we have the following.

- Evaluate the BER for a memoryless channel.
- Evaluate the SNR at the output of an infinite length MIMO MMSE (ISI channel).
- Evaluate the SNR at the output of an infinite length SISO MMSE (ISI channel).
- Evaluate the geometrical SNR at the output of a MIMO MMSE DFE (ISI channel).

We solve these problems and show that under these four viewpoints, the modulation of the input symbol stream with the periodic sequence $f(n)$ leads to a slight degradation in performance when compared with the stationary case.

A. BER for a Memoryless Channel

We compare analytically the expressions of BER in the presence and absence of an MIC precoder for a memoryless channel with an additive Gaussian noise component. Suppose first that the input is a binary PAM signal, and denote by E_0 the signal-to-noise ratio per bit. In the absence of a MIC precoder, it turns out that the average probability of error is given by $P_2 = Q(\sqrt{2E_0})$, where $Q(x) := 1/\sqrt{2\pi}\int_x^\infty \exp(-x^2/2)\,dx$ [19, p. 258]. In the presence of a MIC precoder (41), the average error probability can be expressed as

$$P_2^{\mathrm{MIC}} = \frac{P-1}{P}\,Q(\sqrt{\rho 2E_0}) + \frac{1}{P}\,Q(\sqrt{[P(1-\rho)+\rho]2E_0}). \tag{46}$$

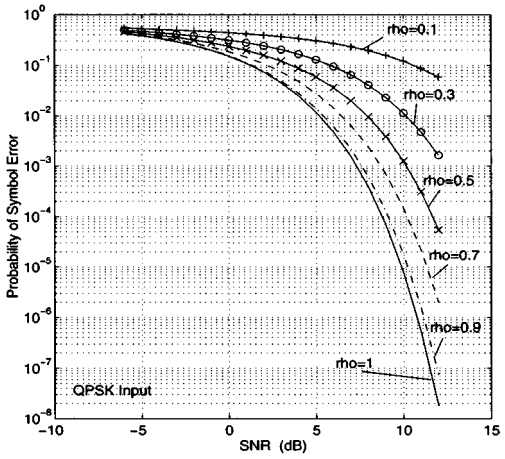


Fig. 3. Probability of symbol error. Comparisons.

Since the function $Q(\sqrt{x})$ is strictly convex, using Jensen's inequality and (46), we obtain

$$P_2^{\mathrm{MIC}} \geq Q\left(\sqrt{\frac{(P-1)}{P}\cdot\rho 2E_0 + \frac{1}{P}\cdot[P(1-\rho)+\rho]2E_0}\right) = Q\left(\sqrt{2E_0}\right) = P_2 \tag{47}$$

with equality iff $\rho 2E_0 = [P(1-\rho)+\rho]2E_0$, that is, iff $\rho = 1$. Thus, for a memoryless channel and 2-PAM constellations, the average error probability is higher for the MIC framework. This analysis can be easily extended to arbitrary M-ary PAM/QAM constellations. The extension to M-ary QAM constellations relies essentially on the convexity of the function $1-(1-Q(\sqrt{x}))^2$, Jensen's inequality, and the error probability formula [19, (5-2-79), p. 280]. Quantitatively, the symbol error rate loss induced by the MIC precoder for a QPSK input constellation is depicted in Fig. 3 for various values of parameter ρ. The extension of these average probability error results to the case of an unknown channel with memory appears more involved and not possible to be expressed in a closed form.

We next focus on ISI channels and evaluate the performance of well-known equalizers as a function of the periodic modulating sequence $f(n)$.

B. MIMO MMSE-Linear Equalization

Following, e.g., [9], the MMSE linear and DF equalizers will be derived for the equivalent time-invariant multi-input multi-output (MIMO) representation of the periodically time-varying SISO channel (2). It will be shown that the resulting matrix-valued equalizers can be expressed in a compact matrix form and are time-invariant, whereas the corresponding SISO equalizers are periodically time-varying. In order to introduce the time-invariant MIMO framework, it is helpful to define the following polyphase vectors

$$\boldsymbol{s}_n := [s(nP)s(nP+1)\ \cdots\ s(nP+P-1)]^T \tag{48}$$

and, similarly, the polyphase vectors $\boldsymbol{w}_n$, $\boldsymbol{v}_n$, and $\boldsymbol{x}_n$. Note that $\boldsymbol{w}_n = \boldsymbol{F}\boldsymbol{s}_n$, with the $P \times P$ diagonal matrix $\boldsymbol{F} := \mathrm{diag}\{f(0), f(1), \cdots, f(P-1)\}$. The I/O relation (2)

can be expressed equivalently in the MIMO operator-form notation as

$$\boldsymbol{x}_n = [\mathcal{H}(z)\boldsymbol{F}]\boldsymbol{s}_n + \boldsymbol{v}_n \tag{49}$$

with the $P \times P$ Toeplitz and polyphase channel matrix $\mathcal{H}(z)$ given by

$$\begin{aligned}\mathcal{H}(z) &:= \begin{bmatrix} h(0) & 0 & \cdots & 0 & h(L)z^{-1} & \cdots & h(1)z^{-1} \\ h(1) & h(0) & \ddots & \ddots & 0 & \ddots & h(2)z^{-1} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & h(L) & \cdots & h(0) \end{bmatrix} \\ &= \mathcal{H}(0) + \mathcal{H}(1)z^{-1} \end{aligned} \tag{50}$$

where the $P \times P$ Toeplitz matrices $\mathcal{H}(0)$ and $\mathcal{H}(1)$ have first column and row vectors given by the vectors $[h(0)\cdots h(L)0\cdots 0]^T$, $[h(0) \quad 0\cdots 0]$, and $[0\cdots 0]^T$, $[0\cdots 0\ h(L)\cdots h(1)]$, respectively.

For two arbitrary vector-valued stationary processes $\mu(n)$ and $\nu(n)$, define the cross-correlation matrix $\boldsymbol{R}_{\mu\nu}(k) := \mathbb{E}(\mu(n+k)\nu^H(n))$ and the corresponding cross-spectral density $\boldsymbol{S}_{\mu\nu}(z) := \sum_k \boldsymbol{R}_{\mu\nu}(k)z^{-k}$. Since the input symbol stream $s(n)$ is i.i.d., the spectrum of polyphase vector $\boldsymbol{s}_n$ is given by $\boldsymbol{S}_{\boldsymbol{ss}}(z) = \sigma_s^2\boldsymbol{I}$. Assuming that the additive noise $v(n)$ is white, it follows from (49) that the spectrum of $\boldsymbol{x}_n$ is equal to $\boldsymbol{S}_{\boldsymbol{xx}}(z) := \sigma_s^2\mathcal{H}(z)\mathcal{H}^H(z^{-1}) + \sigma_v^2\boldsymbol{I}$. The MMSE linear equalizer for the MIMO vector channel (49) is represented by the linear matrix filter $\boldsymbol{G}(z) := \sum_k \boldsymbol{G}(k)z^{-k}$ chosen to minimize the MSE $\mathbb{E}\|\boldsymbol{e}_n\|_2^2$, where $\boldsymbol{e}_n := \hat{\boldsymbol{s}}_n - \boldsymbol{s}_n$, $\hat{\boldsymbol{s}}_n := [\boldsymbol{G}(z)]\boldsymbol{x}_n$, and $\|.\|_2$ stands for the Euclidean norm. The MMSE linear equalizer $\boldsymbol{G}(z)$ can be easily derived using the orthogonality principle $\mathbb{E}[(\hat{\boldsymbol{s}}_{n+k} - \boldsymbol{s}_{n+k})\boldsymbol{x}_n^H] = \boldsymbol{0}$, $\forall\ k$, which implies $\boldsymbol{R}_{\hat{\boldsymbol{s}}\boldsymbol{x}}(k) = \boldsymbol{R}_{\boldsymbol{sx}}(k)$, $\forall\ k$, i.e., $\boldsymbol{S}_{\hat{\boldsymbol{s}}\boldsymbol{x}}(z) = \boldsymbol{S}_{\boldsymbol{sx}}(z)$. Since $\boldsymbol{S}_{\hat{\boldsymbol{s}}\boldsymbol{x}}(z) = \boldsymbol{G}(z)\boldsymbol{S}_{xx}(z)$ and $\boldsymbol{S}_{\boldsymbol{sx}}(z) = \sigma_s^2[\boldsymbol{F}^H\mathcal{H}^H(z^{-1})]$, it follows that

$$\begin{aligned}\boldsymbol{G}(z) = \boldsymbol{S}_{\boldsymbol{sx}}(z)\boldsymbol{S}_{\boldsymbol{xx}}^{-1}(z) &= \sigma_s^2\left[\boldsymbol{F}^H\mathcal{H}^H(z^{-1})\right] \\ &\cdot \left[\sigma_s^2\mathcal{H}(z)\boldsymbol{F}\boldsymbol{F}^H\mathcal{H}^H(z^{-1}) + \sigma_v^2\boldsymbol{I}\right]^{-1}\end{aligned} \tag{51}$$

and the corresponding MSE

$$\mathbb{E}\|\boldsymbol{e}_n\|_2^2 = \sigma_s^2\|\boldsymbol{G}(z)\boldsymbol{H}(z)\boldsymbol{F} - \boldsymbol{I}\|_F^2 + \sigma_v^2\|\boldsymbol{G}(z)\|_F^2 \tag{52}$$

where the Frobenius norm of the polynomial matrix $\boldsymbol{G}(z) := \sum_k \boldsymbol{G}(k)z^{-k}$ is defined as $\|\boldsymbol{G}(z)\|_F^2 := \text{trace}(\sum_k \boldsymbol{G}_k\boldsymbol{G}_k^H)$.

The expression for a finite-length MMSE linear equalizer can be obtained using a similar procedure; the result is immediate, although its requires some extra notations. As it does not provide additional insight as far as the impact of the modulating sequence is concerned, we will not show the solution in detail.

It is interesting to consider now the problem of selecting optimally the modulating sequence $f(n)$ such that the MMSE (52) is minimized. Due to the nonlinear dependence on $f(n)$, minimization of (52) does not seem to be possible in closed-form expression. Extensive numerical simulations using nonlinear optimization algorithms have shown that the minimum of (52) occurs whenever $|f(n)| = 1$, $n = 0, \cdots, P-1$ (that is, $\rho = 1$), i.e., the optimal MMSE occurs for the stationary case. A justification for this result is possible if we argue that for high SNR's ($\sigma_v^2 \to 0$), the IIR MMSE equalizer $\boldsymbol{G}(z)$ (51) converges to $[\mathcal{H}(z)\boldsymbol{F}]^{-1}$, which implies that the corresponding MSE (52) converges to

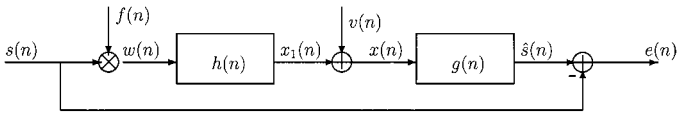


Fig. 4. Baseband MMSE-FIR channel equalization.

$$\mathbb{E}\|\hat{\boldsymbol{s}}(n) - \boldsymbol{s}(n)\|_2^2 \to \sigma_v^2\|\boldsymbol{G}(z)\|_F^2 = \sigma_v^2\|[\mathcal{H}(z)\boldsymbol{F}]^{-1}\|_F^2. \tag{53}$$

Note that for the stationary case the MSE is obtained by considering $\boldsymbol{F} = \boldsymbol{I}$ in (53). Consider that $\mathcal{H}^{-1}(z) = \sum_k \mathcal{L}(k)z^{-k}$. Since for the stationary case the polyphase MSE values $\mathbb{E}|\hat{s}_n(k) - s_n(k)|^2$ have the same value for $k = 0, \cdots, P-1$ and are equal to the diagonal elements of $\sum_k \mathcal{L}(k)\mathcal{L}^H(k)$, it follows that the diagonal elements of the matrix $\sum_k \mathcal{L}(k)\mathcal{L}^H(k)$ are all equal. From (53), we obtain that the MMSE value, corresponding to the case when the precoder $f(n)$ is present, is equal to the MSE corresponding to the stationary case times $\sum_{n=0}^{P-1}(1/|f(n)|^2)$. Appealing to the Cauchy–Schwarz inequality $[(\sum_{n=0}^{P-1}(1/|f(n)|^2))(\sum_{k=0}^{P-1}|f(k)|^2) \geq P^2]$, we obtain that the MMSE is achieved when $|f(n)| = 1$, $\forall\ n = 0, \cdots, P-1$.

C. SISO MMSE-Linear Equalization

The problem of finding the optimal precoder to minimize the MMSE under the constraints (38) and (39) can be carried out in a closed-form solution when the MMSE equalizer is a SISO FIR filter. Let $g(z)$ be an (IIR) filter driven by the noisy observations $x(n)$ (see Fig. 4). The output of $g(z)$ is, of course, CS, and $g(z)$ can be designed to minimize $\int_{-1/2}^{1/2} S_{ee}(0; \exp(i2\pi f))\,df$, where $e(n) := \hat{s}(n) - s(n)$, that is, to minimize

$$\min_{f(n)} \frac{1}{P}\sum_{n=0}^{P-1}\mathbb{E}|e(n)|^2. \tag{54}$$

Such a minimization can be easily coped with since it is a Wiener-like problem. The function $g(\exp(i2\pi f))$ achieving the minimization can be shown to satisfy

$$g(e^{i2\pi f})S_{xx}(0; e^{i2\pi f}) = S_{xs}(0; e^{i2\pi f})$$

where $S_{xs}(0; \exp(i2\pi f))$ denotes the cross-cyclospectrum between $x(n)$ and $s(n)$ at the null frequency. It is straightforward to prove that $S_{xs}(0; \exp(i2\pi f)) = \sigma_s^2 h(\exp(i2\pi f))(1/P)\sum_{k=0}^{P-1} f(k)$. In the noisy case, $S_{xx}(0; \exp(i2\pi f))$ is nonzero for all f, and the error corresponding to the optimal $g(z)$ has the average power

$$\begin{aligned}\frac{1}{P}\sum_{n=0}^{P-1}\mathbb{E}|e(n)|^2 &= \sigma_s^2 - S_{xx}(0; e^{i2\pi f})^{-1}|h(e^{i2\pi f})|^2 \\ &\times \frac{\sigma_s^2}{P^2}|f(0) + \cdots + f(P-1)|^2.\end{aligned} \tag{55}$$

Since the only factor dependent on the modulating sequence $f(n)$ on the right-hand side of (55) is $|f(0) + \cdots + f(P-1)|$,

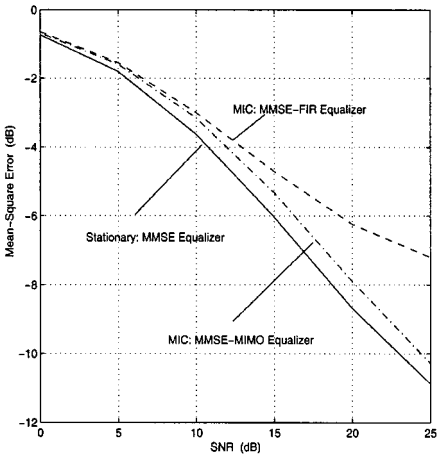


Fig. 5. MMSE linear equalization. MSE comparisons.

it follows that the optimization problem reduces to finding the sequences $\{f(n)\}$ that maximize

$$\max_{f(n)} |f(0)+\cdots+f(P-1)|. \tag{56}$$

Under the constraint (38), it follows that the optimal sequences that maximize (56) are of the form $f(k)=e^{j\phi}$, $k=1,\cdots,P$, where ϕ denotes an arbitrary phase. Indeed, using the inequality $|f(0)+\cdots+f(P-1)|\le|f(0)|+\cdots+|f(P-1)|$ and the Cauchy–Schwarz inequality, it follows that

$$\begin{aligned}&|f(0)|+\cdots+|f(P-1)|\\&\quad\le\sqrt{P|f(0)|^2+\cdots+|f(P-1)|^2}=P.\end{aligned}$$

Thus, the maximum of (56) is attained whenever $f(k)=e^{j\phi}$, $k=0,\cdots,P-1$ with arbitrary phase ϕ. However, such precoders imply a stationary framework and not a CS framework. Thus, the unknown channel cannot be estimated. Of course, the derivation of the SISO MMSE FIR equalizer leading to (55) can be viewed as a special case of the MMSE MIMO equalizer $G(z)$ in (52) with all its rows forced to be equal to $g(z)$.

Because the MIC framework entails increased MMSE w.r.t. the stationary case $f(n)=1$, it is of interest to quantify this increase. In Fig. 5, we plot the MMSE values at the output of the MMSE-MIMO (dash-dotted line) and MMSE-SISO (dashed line) equalizers versus SNR (SNR $:=10\log(\sigma_{x_1}^2/\sigma_v^2)$). As infinite length equalizers cannot be implemented, we considered finite length equalizers of length 7. The MMSE value corresponding to the stationary case ($f(n)=1,\ \forall\ n$) is depicted by the solid line. The channel and the optimal precoder $f(n)$ are given by $\boldsymbol{h}:=[0.4590+0.2650j,-0.2078-0.1200j,-0.4677-0.2700j,0.0953+0.0550j,-0.0312-0.0180j]^T$, $f(0)=\sqrt{P(1-\rho)+\rho}$, and $\{f(n)\}_{n=1}^{P-1}=\sqrt{\rho}$, respectively, with $P=5$, $\rho=0.5878$. We note that the MIC precoder entails, when the MMSE-MIMO equalizer is used, a slight loss (≈ 1 dB) in SNR relative to the stationary framework.

D. MMSE Decision-Feedback Equalization

Following, e.g., [9], the DF equalizer optimized according to the MMSE criterion is derived for the equivalent time-invariant MIMO representation. The stationary MIMO model described by (49) allows us to use the standard multivariable MMSE-DFE structure depicted in Fig. 6, where the filters involved are IIR, and the feedback filter $\boldsymbol{B}(z)$ is strictly causal, i.e., $\boldsymbol{B}(z)=\sum_{k>0}\boldsymbol{B}(k)z^{-k}$, where $\boldsymbol{B}(0)$ is lower triangular with ones on the diagonal. Under these conditions, the optimal choice (in the MMSE sense) of $\boldsymbol{A}(z)$ and $\boldsymbol{B}(z)$ can be calculated as in [9], [30]. The feedforward and feedback filters $\boldsymbol{A}(z)$ and $\boldsymbol{B}(z)$ are given by [30, Eqs. (9), (21), and (37)]

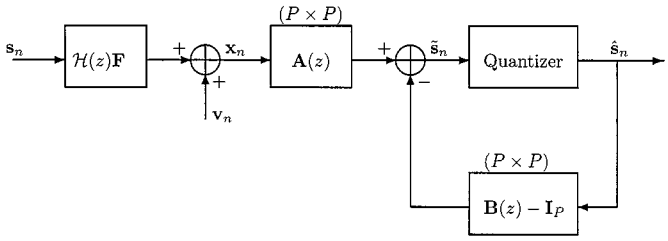


Fig. 6. MMSE-DFE.

$$\begin{aligned}&\boldsymbol{B}(z)=\boldsymbol{\Psi}(z)\\&\boldsymbol{A}(z)=\boldsymbol{\Psi}(z)\mathcal{H}^H(z^{-1})\boldsymbol{F}[\mathcal{H}(z)\boldsymbol{F}\boldsymbol{F}^H\mathcal{H}^H(z^{-1})+(\sigma_v^2/\sigma_s^2)\boldsymbol{I}]^{-1}\end{aligned} \tag{57}$$

where $\boldsymbol{\Psi}(z)$ is a lower triangular matrix, with only ones on the main diagonal, and is given uniquely by the spectral factorization

$$\begin{aligned}&\sigma_s^2\left\{\boldsymbol{I}-\mathcal{H}^H\left(z^{-1}\right)\left[\mathcal{H}(z)\mathcal{H}^H\left(z^{-1}\right)+\left(\sigma_v^2/\sigma_s^2\right)\boldsymbol{I}\right]^{-1}\mathcal{H}(z)\right\}\\&\quad=\boldsymbol{\Psi}(z)\boldsymbol{\Delta}\boldsymbol{\Psi}^H(z^{-1})\end{aligned} \tag{58}$$

where $\boldsymbol{\Delta}$ is a diagonal matrix with all its diagonal entries equal.

Performance of the MMSE-DFE is quantified in terms of the *geometric MSE*, which is defined as the log-determinant of the error covariance matrix $\mathbb{E}(\hat{\boldsymbol{s}}_n-\boldsymbol{s}_n)(\hat{\boldsymbol{s}}_n-\boldsymbol{s}_n)^H$ [30, p. 1336]. The *geometric SNR* at the output of the MMSE-DFE, which is defined as the ratio of the geometric MSE and signal power, is given in our case by (see also [9, p. 635], [30, p. 1338])

$$\begin{aligned}\mathrm{SNR}_{\mathrm{DFE}}=&-\frac{10}{2\pi P}\int_{-\pi}^{\pi}\log_{10}\\&\times\det\left(\frac{\sigma_s^2}{\sigma_v^2}\boldsymbol{F}^H\mathcal{H}^H(e^{i\lambda})\mathcal{H}(e^{i\lambda})\boldsymbol{F}+\boldsymbol{I}_P\right)d\lambda.\end{aligned} \tag{59}$$

If we assume that $\sigma_s^2\gg\sigma_v^2$, then (59) can be approximated as

$$\begin{aligned}\mathrm{SNR}_{\mathrm{DFE}}\simeq&-\frac{10}{2\pi P}\int_{-\pi}^{\pi}\log_{10}\det\left(\frac{\sigma_s^2}{\sigma_v^2}\mathcal{H}^H(e^{i\lambda})\mathcal{H}(e^{i\lambda})\right)\\&\times d\lambda-\frac{10}{P}\sum_{k=0}^{P-1}\log_{10}|f(k)|^2.\end{aligned} \tag{60}$$

The first term on the right-hand side of (60) is actually the SNR obtained in the case of a constant modulus periodic sequence, with $|f(k)|=1$ for all k; therefore, the second term of the right-hand side of (60) is merely the loss in decibels induced by nonconstant modulus periodic sequences. Due to the power constraint (38), and the inequality between the arithmetic and geometric means: $1=(1/P)\sum_{k=0}^{P-1}|f(k)|^2\ge\sqrt[P]{\sum_{k=0}^{P-1}|f(k)|^2}$, we obtain that the second term on the right-hand side of (60)

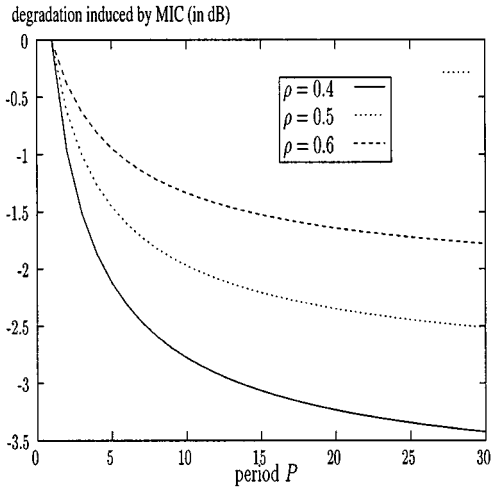


Fig. 7. Loss induced by nonconstant modulus modulating sequences.

is, indeed, a loss $(10/P)[\sum_{k=0}^{P-1} \log_{10} |f(k)|^2] \leq 0$, and the minimum geometric SNR occurs whenever $|f(k)| = 1, \forall\ k = 0, \cdots, P-1$. If we consider the sequences (41), this second term depends on ρ and is given by

$$\phi_1(\rho) = \frac{10}{P} \log_{10}[(P(1-\rho)+\rho)\rho^{P-1}]. \tag{61}$$

The criterion $\phi_1(\rho)$ is plotted versus ρ in Fig. 7 in order to quantify the loss in decibels induced by the periodic symbol modulation. Asymptotically, as $P \rightarrow \infty$, $\phi_1(\rho) \rightarrow 10 \log_{10}(\rho)$, which means that the degradation is the same as the one obtained if all the symbols were modulated by $\sqrt{\rho}$. In practice, the degradation is less than this worst case because the periods considered are small ($P = 4$ or 5).

VI. Simulation Results

In order to illustrate the performance of the proposed approaches, comparative simulation experiments are described in the present section. Throughout the simulations, with the exception of Experiment 3, the channel $\boldsymbol{h} := [0.4590 + 0.2650j, -0.2078 - 0.1200j, -0.4677 - 0.2700j,\ 0.0953 + 0.0550j, -0.0312 - 0.0180j]^T$ is considered. The input symbol stream $s(n)$ is drawn from an equiprobable QPSK constellation, and the additive noise $v(n)$ is white and normally distributed with zero mean. As channel estimation performance measures, we have used the normalized root-mean-square error (RMSE) and average bias (Avg. Bias) of channel estimates, which are defined via

$$\text{RMSE} := \frac{1}{||\boldsymbol{h}||}\sqrt{\frac{1}{R}\sum_{r=1}^{R}\left\|\hat{\boldsymbol{h}}^{(r)} - \boldsymbol{h}\right\|^2}$$

$$\text{Avg. Bias} := \frac{1}{R\cdot(L+1)||\boldsymbol{h}||}\sum_{l=0}^{L}\left|\sum_{r=1}^{R}\left[\hat{h}^{(r)}(l) - h(l)\right]\right|$$

where

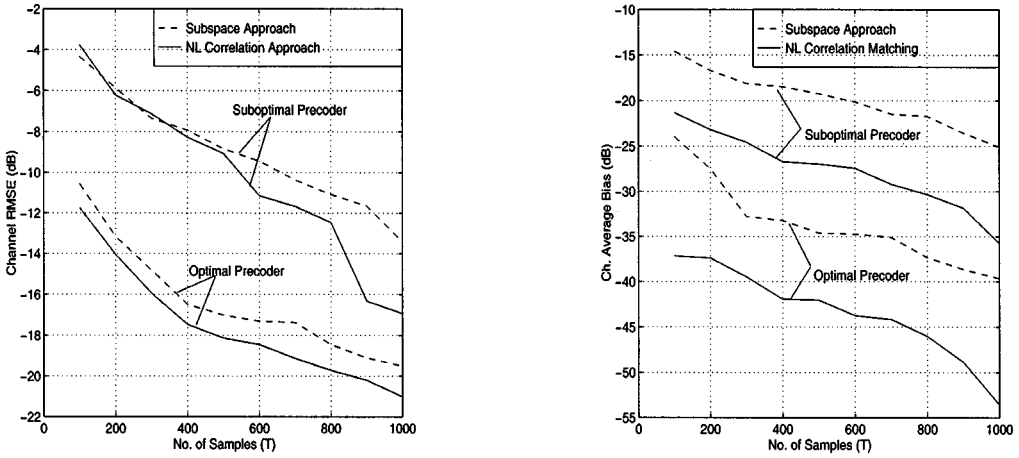
$\hat{\boldsymbol{h}}^{(r)}$	channel estimate in the rth trial;
R	number of Monte-Carlo runs;
$\|\|\boldsymbol{h}\|\|$	Euclidean norm of the channel vector $\boldsymbol{h}$.

We have also considered the SNR defined at the cyclostationary input of the equalizer as SNR $:= \sqrt{(\sum_{n=0}^{P-1} \mathbb{E}\{|x_1(n)|^2\}/P)/\mathbb{E}\{|v(n)|^2\}}$. The performance of the channel equalization approaches has been measured by means of the probability of symbol error.

Experiment 1–Optimal Precoder Selection: The performance of the subspace and nonlinear cyclic correlation matching approaches has been analyzed for two different precoders. In the first scenario, the precoder is chosen optimally according to (41), whereas in the second one, the precoder is suboptimally chosen as $f(0) = \sqrt{\rho}$, and $\{f(n)\}_{n=1}^{P-1} = \sqrt{P(1-\rho)+\rho}$. The parameters $P = 5$ and $\rho = 0.5878$ are used, and both modulating sequences are normalized to have unit power. In both scenarios, the nonlinear cyclic correlation matching approach is implemented without using a weighting matrix ($\boldsymbol{R} = \boldsymbol{I}$). In Fig. 8(a) and (b), the channel root-mean square error (RMSE) and average bias are plotted versus number of samples (T), assuming the fixed value SNR $= 10$ dB, and $R = 100$ Monte Carlo runs. The selection of an optimal precoder leads to a significant improvement in performance for both the subspace and the nonlinear correlation matching approaches. In the case of the optimal precoder, the performance of the subspace approach (dashed line) is quite close to the performance of the nonlinear cyclic correlation approach (solid line). In Fig. 9(a) and (b), the channel RMSE-average bias are plotted versus SNR. In all simulations, the number of samples is fixed to $T = 1000$, and $R = 100$ Monte Carlo runs are performed. We note again that the choice of the optimal precoder leads to a significantly better performance throughout the entire range of SNR's considered, and the proposed channel estimation approaches are quite robust to low SNR's. In Fig. 10(a) and (b), we plotted the channel RMSE-average bias values corresponding to the subspace channel estimator for different values of ρ, assuming the optimal MIC precoder (41). The performance of subspace estimator improves for small values of ρ, which confirms the conclusions of Propositions 2 and 3.

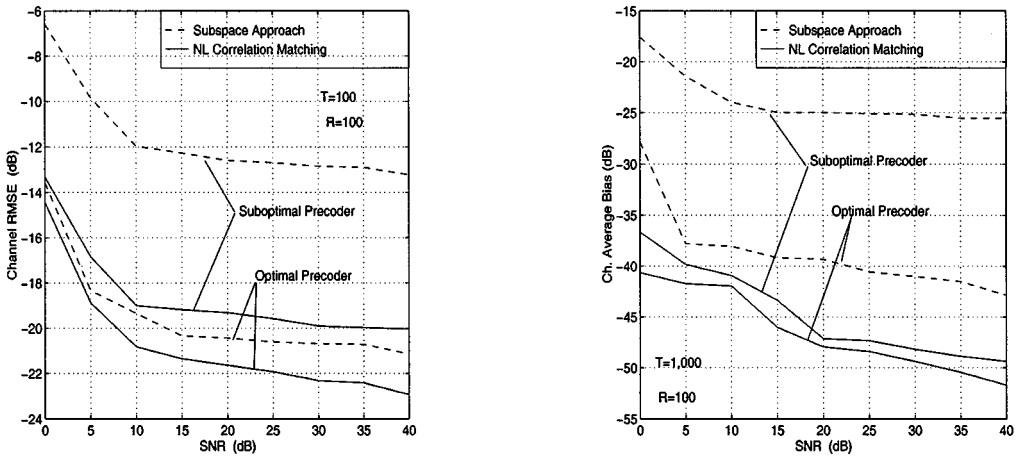
Experiment 2–Robustness to Channel Order Overestimation Errors: In Fig. 11(a) and (b), the channel RMSE-average bias are plotted versus the overestimated channel order $\hat{L}$. The channel order is overestimated with values in the range $4 \leq \hat{L} \leq 10$, and for each channel order, $R = 100$ Monte Carlo runs are used to compute the RMSE and the average bias of the channel estimates. The SNR is fixed to 10 dB, $\rho = 0.5878$, and $T = 1000$ samples are used per Monte Carlo simulation. Quite reliable estimates are obtained for small order overestimation errors. The errors increase as larger channel orders $\hat{L}$ are used. This latter characteristic can be explained since for a fixed period P, consistency of the subspace approach is guaranteed only for the range $L = 4 \leq \hat{L} < L + P - 1 = 8$. By increasing the period (P) of precoder $f(n)$, a larger range for the overestimated order $\hat{L}$ can be allowed.

Experiment 3–Robustness to the Location of Channel Zeros: In order to test the robustness of the proposed subspace approach to the location of channel zeros, we have considered a channel with zeros located at the identifiability limit of



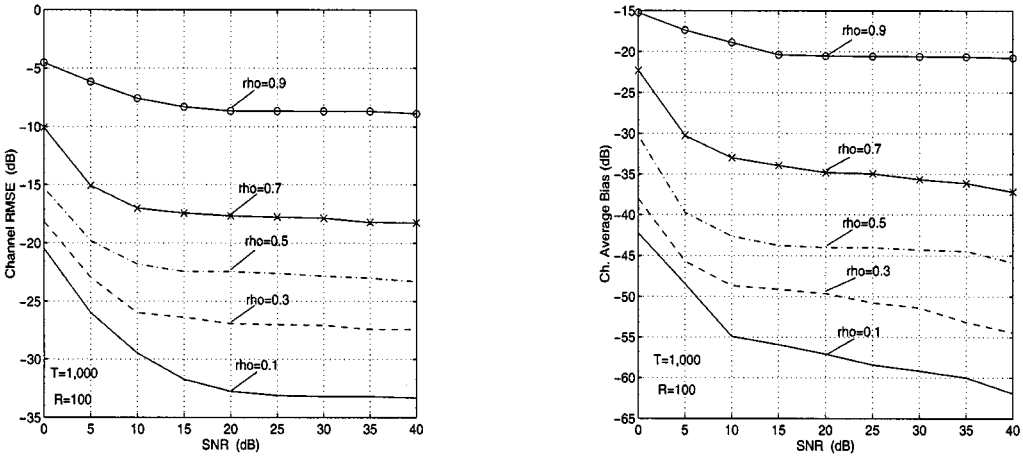
(a) (b)

Fig. 8. Channel average RMSE/Bias versus number of samples.



(a) (b)

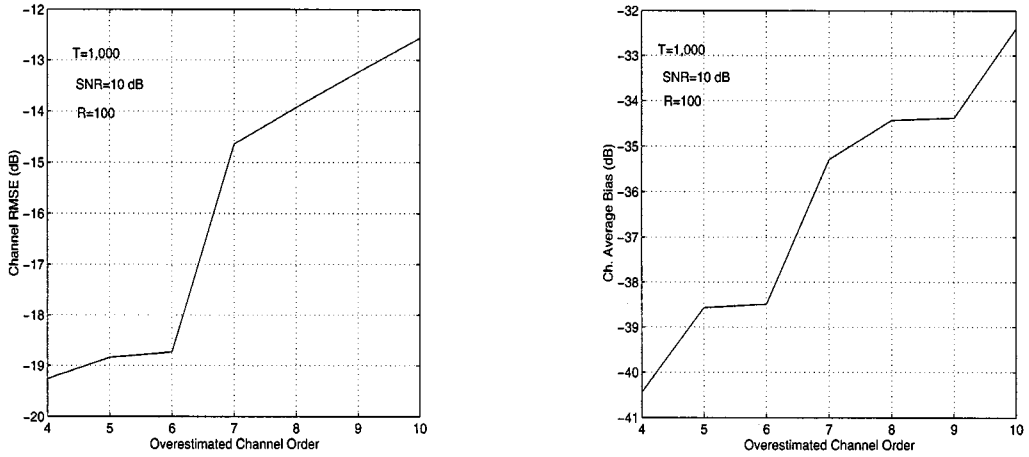
Fig. 9. Channel average RMSE/Bias versus SNR.



(a) (b)

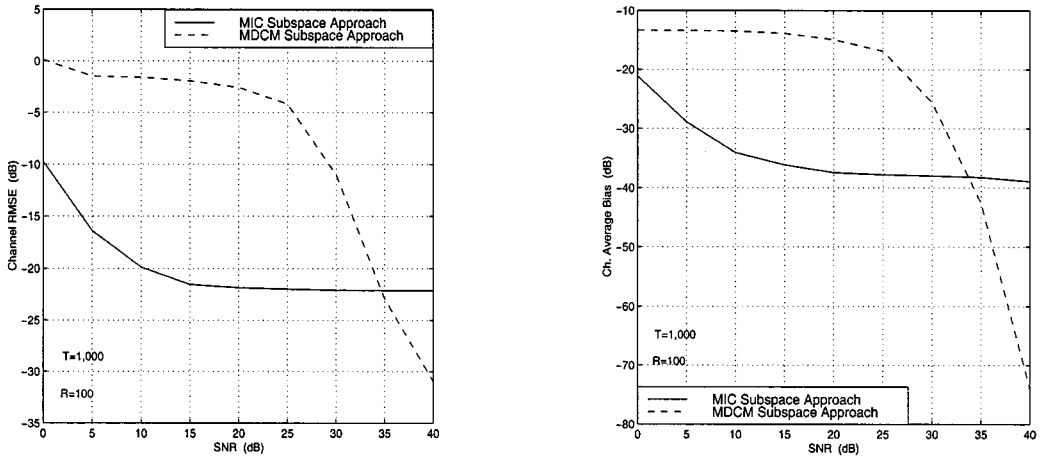
Fig. 10. Channel average RMSE/Bias versus SNR. Comparisons.

the channel subspace approach (MDCM) from [16]. The impulse response of the channel is given by $\tilde{\boldsymbol{h}} = [0.4215 + 0.2433j, 0.1264 + 0.0730j, -0.6955 - 0.4015j, 0.1138 + 0.0657j, \quad 0.2276 + 0.1314j, -0.0885 - 0.0511j]^T$. The MDCM subspace approach has been applied by considering the two subchannels resulting from downsampling the impulse response $\tilde{\boldsymbol{h}}$ by a factor 2. In Fig. 12(a) and (b), the channel RMSE and Avg. Bias are plotted versus SNR for the present MIC-based subspace approach (solid line) and the MDCM subspace approach (dashed line). We note that the MIC-based



(a) (b)

Fig. 11. Channel average RMSE/Bias versus channel order.



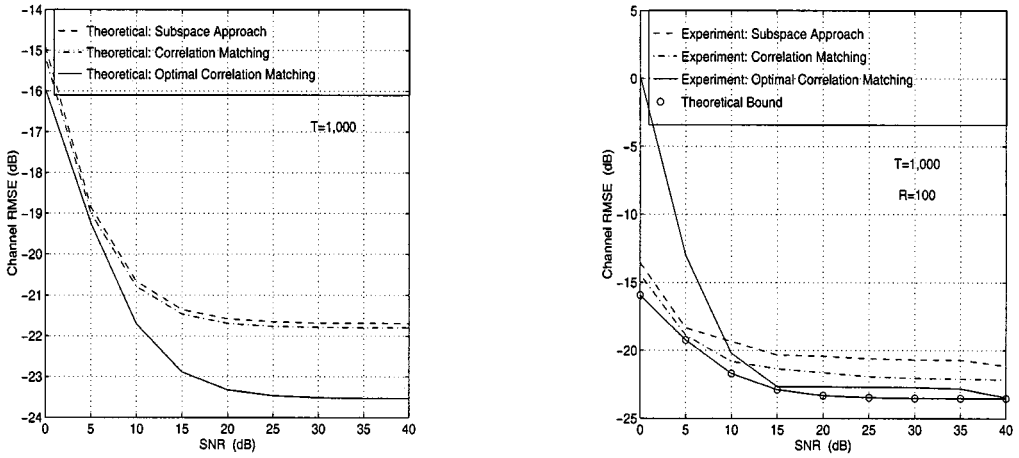
(a) (b)

Fig. 12. Identifiability limit. Channel average RMSE/Bias versus SNR.

approach presents a better performance than the MDCM approach for low to medium SNR's, and for high SNR's, MDCM outperforms the MIC approach. However, if the channel identifiability condition is satisfied, the MDCM approach provides an improved performance relative to the MIC approach in terms of the number of samples that are required to provide reliable channel estimates. This feature is due to the *deterministic* nature of the MDCM approach and the *stochastic* nature of the MIC approach.

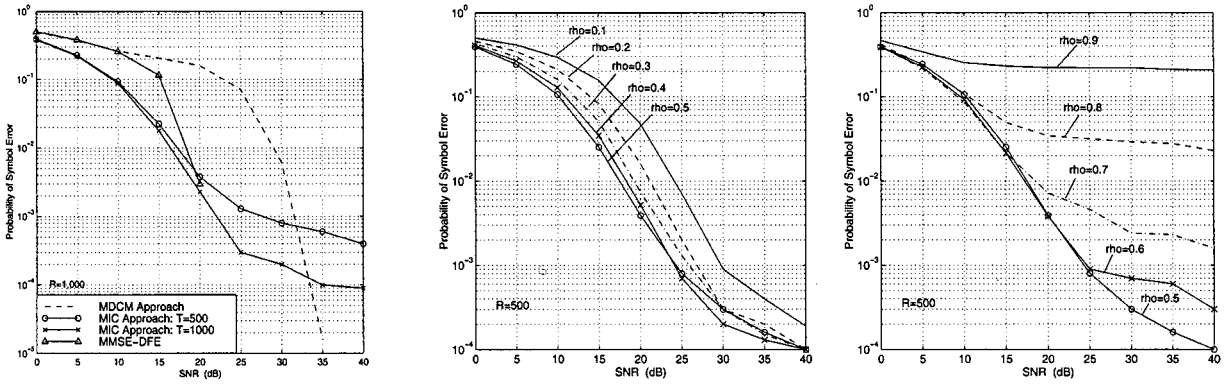
Experiment 4–Asymptotic Performance: Comparisons: In Fig. 13(a) and (b), the theoretical and experimental mean-square errors of the subspace, nonlinear cyclic correlation matching, and optimally weighted nonlinear cyclic correlation matching estimates are plotted versus SNR. The experimental values are averaged over a number of $R = 100$ Monte Carlo runs, assuming a number of $T = 1000$ samples per Monte Carlo trial. In all simulations, the nonlinear cyclic correlation matching approach has been initialized with the channel estimate provided by the subspace approach. Although the theoretical bound in Fig. 13(a) suggests an improvement of the optimally weighted matching approach relative to the subspace approach, the experimental results show that this improvement is hard to achieve due to the local minima. Extensive simulation experiments have shown that the nonlinear cyclic correlation matching approach works well, provided that the initial estimate is sufficiently close to the global minimum. For complex channels with an increased number of taps (6–8 taps), the initialization provided by the subspace approach does not appear to be useful since the nonlinear cyclic correlation matching approach cannot bypass the local convergence problems. For these reasons, only the linear subspace approach appears to be a good candidate for practical applications. In Fig. 13(b), the experimental mean-square errors of the optimally weighted (solid line) and suboptimally weighted ($\boldsymbol{R} = \boldsymbol{I}$, dash-dotted line) cyclic correlation matching approaches are compared with the experimental values corresponding to the subspace approach (dashed line). Fig. 13(b) indicates that the optimally and suboptimally weighted nonlinear approaches perform almost identically. Due to the high computational complexity associated with the optimal correlation matching approach and its lack of significant improvement relative to the suboptimal correlation matching approach, computation of the optimal weighting matrix may not be recommendable for real-time applications.

Experiment 5–Channel Equalization: Comparisons: Performance of the MDCM approach [16] and the proposed MIC



(a) (b)

Fig. 13. Asymptotic performance.



(a) (b) (c)

Fig. 14. Symbol error rate. Comparisons.

approach are compared with respect to the probability of symbol error for different SNR levels. In all experiments, symbol-spaced MMSE equalizers with 17 taps are used. For the MIC approach, the MMSE-MIMO equalizer is implemented as described in Section V-B by making use of the subspace channel estimate in the expression of the MMSE-MIMO equalizer. For MDCM, the channel estimates are obtained first within the FS framework, assuming a stationary input (no precoder), and then, the MMSE equalizer is designed based on the channel estimates. The symbol error curves are plotted in Fig. 14(a), where $R = 1000$ Monte Carlo runs are averaged per SNR point. The number of sample data used per Monte Carlo trial is $T = 500$. The equalization of the MIC framework has been also tested for an increased number of sample data $T = 1000$. It is remarkable that for low and medium SNR's, the MIC framework outperforms the subspace approach MDCM significantly. This good behavior is due to the robustness of the MIC framework to additive stationary noise. However, at high SNR's, MDCM outperforms MIC.

Implementation of MMSE-DFE has been performed, assuming that the channel is known. From the symbol error rate curves presented in Fig. 14(a), it follows that the MMSE-DFE is outperformed by the MMSE linear equalizer only for low SNR's. For high SNR's, the MMSE-DFE outperforms the MMSE linear equalizer. Simulation experiments for the MMSE-DFE have been also performed, assuming that the channel is unknown. At each Monte Carlo run, the unknown channel has been estimated using the MIC approach. It has been observed that the MMSE-DFE does not perform well for such scenarios, which can be explained by the sensitivity of the MMSE-DFE to variations in the location of channel zeros (due to the estimation errors in the channel estimates). The problem of designing MMSE-DFE's robust to variations of channel estimates is left for future research.

In Fig. 14(b) and (c), we plotted the symbol error rate curves for the MIC precoder (41), assuming different values of parameter ρ. For each Monte Carlo run, the equalization was performed by first estimating the unknown channel using the subspace approach and then recovering the symbols by means of the MMSE linear equalizer, as described above. Fig. 14(b) and (c) reveal the unexpected phenomenon that the detection performance improves as ρ increases in the interval $[0.1, 0.5]$ and degrades as ρ increases in the interval $[0.5, 0.9]$. The plots of Fig. 14(c) can be explained by the fact that the subspace channel estimates degrade as $\rho \to 1$, i.e., when the MIC framework approaches the stationary framework.

VII. Conclusions

The present paper has established a general identifiability condition for the MIC framework, which allows blind identification of any SISO FIR channel without imposing restrictions on the channel zeros, color of additive stationary noise, channel order over-estimation errors, and without increasing the transmission rate. However, this benefit is accompanied by a relative decrease in power efficiency. Criteria for optimally selecting the modulating precoder have been presented. The paper has also studied the asymptotic performance of two channel estimation algorithms. The asymptotic performance analysis has been motivated by the interest in quantifying the performance of the subspace approach with respect to the performance of the nonlinear cyclic correlation matching approach, as well as by the problem of selecting the modulating precoder optimally.

It appears that *redundant* input diversity techniques along the lines of [11] and [21] are very promising because they not only possess extra benefits over the MIC framework (e.g., possibility of ZF-FIR equalization), but they also obviate MIC limitations with high peak-to-average transmissions and the corresponding decrease in power efficiency.

Appendix
Proof of Theorem 2

We first provide a result [5] concerning the asymptotic covariance of the cyclic correlation estimates defined by (28).

Lemma 1–Asymptotic Covariance of Cyclic Correlation Coefficients: Suppose that the noise is white Gaussian and the symbols circular. For given cycles k_1, k_2, $1 \le k_1, k_2 \le P-1$, the following asymptotic result holds:

$$\lim_{T\to\infty} T \operatorname{cov}\left(\hat{C}_{xx}(k_1;\tau_1), \hat{C}_{xx}(k_2;\tau_2)\right) = \xi_1 + \xi_2 \quad (62)$$

where

$$\xi_1 = \sum_{\substack{\ell_1,\,\ell_2=0 \\ \ell_1-\ell_2\equiv k_1-k_2}}^{P-1} e^{j2\pi(\ell_1/P)\tau_2} \int_0^1 S_{xx}\left(\ell_1; e^{j2\pi f}\right) \times S^*_{xx}\left(\ell_2; e^{j2\pi(f-(k_1/P))}\right) e^{j2\pi f(\tau_1-\tau_2)}\, df \quad (63)$$

and

$$\xi_2 = \theta(k_1,k_2)\int_0^1\int_0^1 H\left(e^{j2\pi\nu_1}\right) H^*\left(e^{j2\pi(\nu_1-(k_1/P))}\right) \times H^*\left(e^{j2\pi\nu_2}\right) H\left(e^{j2\pi(\nu_2-(k_2/P))}\right) \cdot e^{j2\pi(\nu_1\tau_1-\nu_2\tau_2)}\, d\nu_1\, d\nu_2. \quad (64)$$

The proof of Theorem 2 follows now by using several standard techniques (see, e.g., [1] and [2]). A perturbation result from [14] states that there is an open neighborhood of $\boldsymbol{c}$ in which the mapping $\hat{\boldsymbol{c}} \hookrightarrow \hat{\boldsymbol{h}}$ is infinitely differentiable. We can write

$$\hat{\boldsymbol{h}} \cong \boldsymbol{h} - \boldsymbol{Q}^{\dagger}\mathcal{P}^H\mathcal{D}^H(\boldsymbol{c})\mathcal{D}(\Delta\boldsymbol{c})\mathcal{P}\boldsymbol{h}$$

which yields, according to (24) $\hat{\boldsymbol{h}} \cong \boldsymbol{h} + \boldsymbol{Q}^{\dagger}\mathcal{P}^H\mathcal{D}^H(\boldsymbol{c})\mathcal{T}(\boldsymbol{h})\Delta\boldsymbol{c}$. Theorem 2 is obtained after some straightforward calculations since computation of the asymptotic covariance of the correlation vector $\Delta\boldsymbol{c} = [\Delta\boldsymbol{c}(\overline{L}), \cdots, \Delta\boldsymbol{c}(-\overline{L})]^T$ with

$$\Delta\boldsymbol{c}(\tau) = [\Delta C^*_{xx}(1;-\tau), \Delta C^*_{xx}(2;-\tau), \cdots \Delta C^*_{xx}(P-1;-\tau)]^T$$

can be done by applying the result of Lemma 1. □

References

[1] K. Abed-Meraim, J. F. Cardoso, A. Gorokhov, P. Loubaton, and E. Moulines, "On subspace methods for blind identification of SIMO FIR systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 42–56, Jan. 1997.

[2] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. Inform. Theory*, vol. 43, pp. 499–511, Mar. 1997.

[3] N. Al-Dhahir and J. M. Cioffi, "MMSE decision feedback equalizers: Finite-length results," *IEEE Trans. Inform. Theory*, vol. 41, pp. 961–976, July 1995.

[4] D. R. Brillinger, *Time Series: Data Analysis and Theory*. San Francisco, CA: Holden Day, 1981, pp. 8–9.

[5] A. Chevreuil, "Blind equalization in a cyclostationary context," Ph.D. dissertation, ENST-Telecom, Dept. Signal Processing, Paris, France, 1997.

[6] A. Chevreuil, F. Desbouvries, A. Gorokhov, Ph. Loubaton, and C. Vignat, "Blind equalization in the presence of jammers and unknown noise: Solutions based on second-order cyclostationary statistics," *IEEE Trans. Signal Processing*, vol. 46, pp. 259–263, Jan. 1998.

[7] 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.

[8] A. V. Dandawaté and G. B. Giannakis, "Asymptotic theory of mixed time averages and k-th order cyclic-moment and cumulant statistics," *IEEE Trans. Inform. Theory*, vol. 41, pp. 216–239, Jan. 1995.

[9] A. Duel-Hallen, "Equalizers for multiple input/multiple output channels and PAM systems with cyclostationary input sequences," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 630–639, Apr. 1992.

[10] G. B. Giannakis, "Linear cyclic correlation approaches for blind channel identification of FIR channels," in *Proc. Asilomar Conf.*, Pacific Grove, CA, Nov. 1995, pp. 420–424.

[11] ——, "Filterbanks for blind channel identification and equalization," *IEEE Signal Processing Lett.*, vol. 4, pp. 184–187, June 1997.

[12] G. B. Giannakis and S. D. Halford, "Asymptotically optimal blind fractionally-spaced channel estimation and performance analysis," *IEEE Trans. Signal Processing*, vol. 45, pp. 1815–1830, July 1997.

[13] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[14] T. Kato, *Perturbation Theory for Linear Operators*. New York: Springer, 1980.

[15] M. Kristensson and B. Ottersten, "A statistical approach to subspace based blind identification," *IEEE Trans. Signal Processing*, vol. 46, pp. 1612–1623, June 1998.

[16] 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, Febr. 1995.

[17] E. E. Newhall, S. U. H. Qureshi, and C. F. Simone, "A technique for finding approximate inverse systems and its application to equalization," *IEEE Trans. Commun. Technol.*, vol. COMM-19, pp. 1116–1127, Dec. 1971.

[18] B. Porat, *Digital Processing of Random Signals: Theory and Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[19] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[20] S. U. H. Qureshi, "Adjustment of the position of the reference tap of an adaptive equalizer," *IEEE Trans. Commun.*, vol. COMM-21, pp. 1046–1052, Sept. 1973.

[21] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant filterbanks and equalizers—Part I: Unification and optimal designs. Part II: Blind channel estimation, synchronization and direct estimation," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988–2022, July 1999.

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

[23] D. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," in *Proc. ICASSP*, vol. 4, Adelaide, Australia, 1994, pp. 585–588.

[24] T. Söderström and P. Stoica, *System Identification*. London, U.K.: Prentice-Hall, 1989.
[25] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[26] 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, Mar. 1994.
[27] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind channel identification based on second-order statistics: A frequency-domain approach," *IEEE Trans. Inform. Theory*, vol. 41, pp. 329–334, Jan. 1995.
[28] 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.
[29] 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.
[30] J. Yang and S. Roy, "Joint transmitter–receiver optimization for multi-input multi-output systems with decision feedback," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1334–1347, Sept. 1994.

Antoine Chevreuil was born in Caen, France, in 1971. He received the M.Sc. and Ph.D. degrees from Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1994 and 1997, respectively.

In 1998, he was associated with the Telecommunications Department, University of Louvain-la-Neuve, Louvain-la-Neuve, Belgium. He is currently an Assistant Professor with the Communication Systems Laboratory, University of Marne-la-Vallée, Noisy le Grand, France. His research interests are statistical signal processing for communications, system identification, and equalization.

Erchin Serpedin received the Diploma of electrical engineering (with highest distinction) from the Polytechnic Institute of Bucharest, Bucharest, Romania, in 1991. He received the specialization degree in signal processing and transmission of information from Supélec, Paris, France, in 1992, the M.Sc. degree from Georgia Institute of Technology, Atlanta, in 1992, and the Ph.D. degree from the University of Virginia, Charlottesville, in December 1998.

From 1993 to 1995, he was an Instructor at the Polytehnic Institute of Bucharest. In 1999, he joined the Wireless Communications Laboratory, Texas A&M University, College Station, as an Assistant Professor. His research interests lie in the areas of statistical signal processing and wireless communications.

Philippe Loubaton (M'91) was born in 1958 in Villers Semeuse, France. He received the M.Sc. and Ph.D. degrees from Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France, in 1981 and 1988, respectively.

From 1982 to 1986, he was a member of the technical staff of Thomson CSF/RGS, where he worked in digital communications. From 1986 to 1988, he worked with the Institut National des Télécommunications as an Assistant Professor of electrical engineering. In 1988, he joined ENST, working in the Signal Processing Department. Since 1995, he has been a Professor of electrical engineering at the University of Marne-la-Vallée, Noisy le Grand, France. His present research interests are in statistical signal processing and linear system theory, including connections with interpolation theory for matrix-valued holomorphic functions and system identification.

Dr. Loubaton is a member of the Board of the GDR/PRC ISIS (the CNRS research group on signal and image processing), where he is in charge of the working group on blind identification.

Georgios B. Giannakis (F'96) received the Diploma 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, the 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 (UVA), Charlottesville, in 1987, where he became a Professor of electrical engineering in 1997, Graduate Committee Chair, and Director of the Communications, Controls, and Signal Processing Laboratory in 1998. Since January 1999, he has been with the University of Minnesota, Minneapolis, as a Professor of electrical and computer engineering. His general interests lie in the areas of signal processing and communications, estimation and detection theory, time-series analysis, and system identification—subjects on which he published more than 100 journal papers. Specific areas of expertise have included (poly)spectral analysis, wavelets, cyclostationary, and non-Gaussian signal processing with applications to sensor array and image processing. Current research focuses on transmitter and receiver diversity techniques for equalization of single- and multiuser communication channels, mitigation of rapidly fading channels, compensation of nonlinear amplifier effects, redundant filterbank transceivers for block transmissions, multicarrier, and wideband communication systems.

Dr. Giannakis was awarded the School of Engineering and Applied Science Junior Faculty Research Award from UVA in 1988 and the UVA-EE Outstanding Faculty Teaching Award in 1992. He received the IEEE Signal Processing Society's 1992 Paper Award in the Statistical Signal and Array Processing (SSAP) area and the 1999 Young Author Best Paper Award (with M. K. Tsatsanis). He co-organized the 1993 IEEE Signal Processing Workshop on Higher-Order Statistics, the 1996 IEEE Workshop on Statistical Signal and Array Processing, and the first IEEE Signal Processing Workshop on Wireless Communications in 1997. He guest (co-)edited two special issues on high-order statistics (*International Journal of Adaptive Control and Signal Processing* and the EURASIP journal *Signal Processing*) and the January 1997 Special Issue on Signal Processing for Advanced Communications of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, a Secretary of the Signal Processing Conference Board, a member of the SP Publications Board, and a member and vice-chair of SSAP Technical Committee. He now chairs the Signal Processing for Communications Technical Committee and serves as Editor-in-Chief of the IEEE SIGNAL PROCESSING LETTERS. He is also a member of the IEEE Fellow Election Committee, the IMS, and the European Association for Signal Processing.