

ML Approaches to Channel Estimation for Pilot-Aided Multirate DS/CDMA Systems

Xavier Mestre, *Student Member, IEEE*, and Javier R. Fonollosa, *Senior Member, IEEE*

Abstract—This paper analyzes the asymptotic performance of maximum likelihood (ML) channel estimation algorithms in wideband code division multiple access (WCDMA) scenarios. We concentrate on systems with periodic spreading sequences (period larger than or equal to the symbol span) where the transmitted signal contains a code division multiplexed pilot for channel estimation purposes. First, the asymptotic covariances of the training-only, semi-blind conditional maximum likelihood (CML) and semi-blind Gaussian maximum likelihood (GML) channel estimators are derived. Then, these formulas are further simplified assuming randomized spreading and training sequences under the approximation of high spreading factors and high number of codes. The results provide a useful tool to describe the performance of the channel estimators as a function of basic system parameters such as number of codes, spreading factors, or traffic to training power ratio.

Index Terms—Conditional maximum likelihood (CML), Gaussian maximum likelihood (GML), semi-blind channel estimation, wideband CDMA (WCDMA).

NOMENCLATURE

a	Scalar.
\mathbf{a}	Vector.
\mathbf{A}, \mathcal{A}	Matrix.
\mathbf{A}^T	Transpose of \mathbf{A} .
\mathbf{A}^H	Conjugate transpose of \mathbf{A} .
\mathbf{A}^*	Conjugate of \mathbf{A} .
\mathbf{I}_L	$L \times L$ identity matrix.
$\mathbf{0}_{M \times N}$	$M \times N$ matrix of zeros.
$\{\mathbf{A}\}_{i,j}$	$\{i, j\}$ th element of \mathbf{A} .
$\text{vec}(\mathbf{A})$	Vector reshaping \mathbf{A} column-wise.
$\text{tr}(\mathbf{A})$	Trace of \mathbf{A} .
δ_k	Kronecker delta.

I. INTRODUCTION

THE UNPRECEDENTED growth of the demand for wireless services has recently called for the introduction of third-generation mobile communication systems. These systems, which are specifically designed to accommodate multiple and higher data rates, are largely based on multirate

direct-sequence code division multiple access (DS/CDMA) technology. This is because classical DS/CDMA modulation schemes can be easily generalized to incorporate and homogenize the simultaneous transmission of several data streams at distinct bit rates through the same physical interface. Traditionally, there have been three distinct approaches to generalizing the classical DS-CDMA modulation scheme to a multirate framework, namely

- 1) multicode modulation, where users transmitting at high bit rates make use of more than one spreading code transmitted simultaneously;
- 2) variable spreading factor modulation, where the span of the symbol in terms of number of chips is variable and depends on the transmitting bit rate;
- 3) variable chip rate modulation, where users transmitting at a higher bit rate occupy a larger portion of the available bandwidth.

Currently, standardized systems such as the frequency division duplex (FDD) mode of UMTS terrestrial radio access (UTRA) transmit according to a modulation format usually referred to as wideband code division multiple access (WCDMA), which is basically designed as a combination of the multicode and variable spreading factor strategies. This means that each user is allowed to make use of more than one waveform or code sequence, while at the same time, the span of the bit interval in number of chips is variable, depending on the transmission bit rate. Fig. 1 depicts an example for real-valued signals where three different data streams are mapped to three codes with different spreading factor but equal chip and repetition periods.

One of the most interesting features of these newly deployed systems is the introduction of a pilot signal that, serving channel estimation purposes, is code division multiplexed and transmitted at the same time as the traffic information (one example of such a structure can be found in the uplink of the FDD mode of UTRA; see [1]). While the performance of channel estimation algorithms for systems in which a training sequence is time division multiplexed with the traffic data has been widely studied, little attention has been paid to systems where it is code division multiplexed. Given the importance that such systems may have in the near future, it seems crucial to investigate some specific aspects such as the dependence of the channel estimation performance on, e.g., the spreading codes repetition period, the signal-to-noise ratio (SNR), the quotient between the pilot and the traffic signal powers, or the spreading factor utilized. This paper tries to provide answers to these questions in a theoretical yet simple manner. It will be shown that under some asymptotic conditions, very simple expressions describing the behavior of the analyzed channel estimation algorithms can be obtained.

Manuscript received November 1, 2000; revised November 9, 2001. This work was supported in part by the Spanish and Catalan Governments under Grants TIC2000-1025, TIC99-0849, CIRIT 2000SGR-00083, and FIT-070000-2000-649 and by the European Commission under IST Project I-METRA. The associate editor coordinating the review of this paper and approving it for publication was Dr. Kristine Bell.

The authors are with the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, Barcelona, Spain (e-mail: mestre@gps.tsc.upc.es; fono@gps.tsc.upc.es).

Publisher Item Identifier S 1053-587X(02)01330-2.

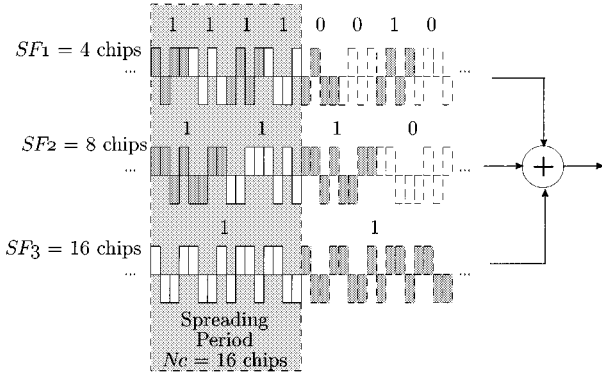


Fig. 1. Example of multirate modulation as a combination of multicode and variable spreading factor schemes. A single user simultaneously transmits three data streams mapped to three different codes, each having a different spreading factor but equal repetition period.

These simple expressions will prove most useful from a system designer point of view in that they will provide a precise description of the expected performance as a function of physical system parameters.

From the whole range of channel estimation algorithms available in the literature, we will focus on maximum likelihood (ML) estimation procedures since under certain regularity conditions, they provide asymptotically efficient estimates (see, e.g., [2]). In particular, we will study both classical training-only techniques (based on the knowledge of the pilot or training sequence exclusively) and semi-blind techniques [3], [4] (which make use of the training sequence and improve this estimation taking into account the signal structure). We will see that the performance gain obtained from the use of semi-blind techniques, which could be described as moderate in classical systems with time division multiplexed training sequences, turns out to be dramatic in systems with a code division multiplexed pilot.

The rest of the paper is organized as follows. Section II presents the multirate signal model and introduces the system parameters and assumptions. Section III describes the three different channel estimation methods under consideration and their asymptotic performance. In Section IV, the covariance of these estimators is particularized for some asymptotic cases assuming random training and code sequences. Finally, Section V presents a numerical validation of the results, and Section VI outlines some conclusions regarding the techniques under analysis.

II. MULTIRATE SIGNAL MODEL

The transmitting station is assumed to map the underlying data sequence to Q distinct and synchronized spreading sequences with period N_c chips. Both the period of the spreading sequences and the chipping rate are assumed constant for all the sequences. Let $s_q(m) \in \mathbb{C}$ represent the underlying complex symbol stream associated with the q th code sequence, and assume that $N_s(q)$, $q = 1 \dots Q$ consecutive symbols are mapped to each interval of N_c chips. The spreading factor associated with a particular spreading sequence is denoted as

$SF_q = N_c/N_s(q)$ chips/bit. According to these definitions, the total number of symbols mapped to a code sequence period (N_c chips) is $N_s = \sum_{q=1}^Q N_s(q)$.

We assume that at the basestation, the signal is synchronously sampled at the chip rate (modulation with no excess bandwidth) and that symbol detection is made in observation windows of MN_c chips, where M can be regarded as the number of spreading periods in the observation interval. Stacking N_c samples of the received signal corresponding to the m th code period into a column vector $\mathbf{x}(m) \in \mathbb{C}^{N_c \times 1}$, we can define the total received signal vector as $\mathbf{x} \in \mathbb{C}^{MN_c \times 1}$

$$\mathbf{x} = [\mathbf{x}^T(1) \quad \dots \quad \mathbf{x}^T(M)]^T = \mathbf{x}^k + \mathbf{G}\mathbf{s} + \mathbf{n} \quad (1)$$

where $\mathbf{x}^k \in \mathbb{C}^{MN_c \times 1}$ is the channel-filtered *known* training sequence, $\mathbf{G} \in \mathbb{C}^{MN_c \times M_s}$ is a matrix of received signatures, $\mathbf{s} \in \mathbb{C}^{M_s \times 1}$ is a vector containing the complex-valued transmitted symbols, and \mathbf{n} is the noise component. Here, M_s is the number of symbols received in an observation interval, which is equal to the sum $M_s = \sum_{q=1}^Q M_s(q)$, where $M_s(q)$ is the number of symbols corresponding to the q th code

$$M_s(q) = \underbrace{\left\lfloor \frac{L-1}{SF_q} \right\rfloor}_{\text{Past symbols}} + MN_s(q).$$

Note from (1) that the training sequence is transmitted code division multiplexed—rather than time division multiplexed—with the traffic information. We denote by $\mathbf{h} \in \mathbb{C}^{L \times 1}$ the column vector containing the channel impulse response, which is assumed stationary on the observation interval and of length L , where for simplicity, we only consider $L \leq N_c$. Accordingly, the known part of the received signal, \mathbf{x}^k can be expressed as

$$\mathbf{x}^k = \mathcal{T}\mathbf{h} = [\mathcal{T}(1)^T \quad \dots \quad \mathcal{T}(M)^T]^T \mathbf{h} \quad (2)$$

$$\mathcal{T}(m) = \begin{bmatrix} t((m-1)N_c + 1) & \dots & t((m-1)N_c - L + 2) \\ t((m-1)N_c + 2) & \dots & t((m-1)N_c - L + 3) \\ \vdots & \ddots & \vdots \\ t(mN_c) & \dots & t(mN_c - L + 1) \end{bmatrix} \in \mathbb{C}^{N_c \times L}, \quad m = 1 \dots M \quad (3)$$

where $t(n)$ denotes the chip-level training sequence.

Let us now concentrate on the matrix of received signatures \mathbf{G} . This matrix is formed by stacking side by side the signature matrices corresponding to each of the code sequences $\mathbf{G} = [\mathbf{G}_1 \quad \dots \quad \mathbf{G}_Q]$. These \mathbf{G}_q , $q = 1 \dots Q$ can in turn be expressed as

$$\mathbf{G}_q = C_q (\mathbf{I}_{M_s(q)} \otimes \mathbf{h})$$

$$C_q = \begin{bmatrix} \mathcal{D}_q(2) & C_q(1) & \dots & \mathbf{0} \\ \mathbf{0} & C_q(2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & C_q(2) & C_q(1) \end{bmatrix}_{M \times M_s(q) \text{ blocks}} \quad (4)$$

where \otimes is the Kronecker product, and $C_q \in \mathbb{C}^{MN_c \times M_s(q)L}$. Matrices $C_q(1)$ and $C_q(2)$ are the upper and lower parts of the next convolution matrix

$$\begin{bmatrix} C_q(1) \\ C_q(2) \end{bmatrix} = [C_{q,1} \ C_{q,2} \ \cdots \ C_{q,N_s(q)}] \in \mathbb{C}^{2N_c \times N_s(q)L}$$

$$C_{q,r} = \begin{bmatrix} c_{q,r}(1) & 0 & \mathbf{0}_{L-1 \times 1} \\ \vdots & \ddots & c_{q,r}(1) \\ c_{q,r}(N_c) & \ddots & \vdots \\ 0 & \ddots & c_{q,r}(N_c) \\ \vdots & \cdots & \mathbf{0}_{N_c-L+1 \times 1} \end{bmatrix} \in \mathbb{C}^{2N_c \times L} \quad (5)$$

with $c_{q,r}(n)$ defined from the original code sequences $c_q(n)$, $q = 1 \cdots Q$ setting to zero all the samples outside the r th symbol interval, i.e.,

$$c_{q,r}(n) = \begin{cases} c_q(n), & (r-1)SF_q < n \leq rSF_q \\ 0, & \text{otherwise.} \end{cases}$$

Matrix $\mathcal{D}_q(2)$ is obtained as the $\lceil L - 1/SF_q \rceil$ columns on the right of $C_q(2)$ and contains the contribution from symbols transmitted prior to the observation interval.

All this formulation is useful to understand the signal model in (1) but becomes a bit cumbersome when performing matrix manipulations. For this reason, we prefer to use the following partitioning throughout the paper.

$$\begin{aligned} \mathcal{C} &= [\mathcal{C}_1 \ \mathcal{C}_2 \ \cdots \ \mathcal{C}_Q] \in \mathbb{C}^{MN_c \times M_s L} \\ &= [\mathcal{S}_1 \ \mathcal{S}_2 \ \cdots \ \mathcal{S}_{M_s}] \end{aligned} \quad (6)$$

where now, each matrix $\mathcal{S}_i \in \mathbb{C}^{MN_c \times L}$ contains the convolution matrix of the code sequence associated with the i th symbol interval.

Returning to the signal model in (1), it is finally assumed that the components of the noise vector are circularly symmetric Gaussian distributed with zero mean and covariance $E[\mathbf{nn}^H] = \sigma^2 \mathbf{I}_{MN_c}$. This seems a reasonable approximation since we are concentrating on a single-user scenario. The results obtained here could, in principle, be generalized to the multichannel estimation case, modifying the signal model in (1) to include the contribution from several users in the signal term. This is, however, out of the scope of this paper.

III. ML CHANNEL ESTIMATION METHODS

In the following, we present some channel estimation methods based on the principle of ML. We will see that different ways of modeling the presence of the unknown symbols, transmitted at the same time as the training sequence, will lead to distinct types of estimators with different asymptotic performance. In particular, we will concentrate on three different types of estimators.

- 1) classical training-only (TO) estimator, which results from ignoring the presence of the unknown symbols;
- 2) CML estimator, based on a model in which unknown data are regarded as deterministic parameters;

- 3) GML estimator, which arises from modeling them as Gaussian-distributed random variables.

In order to ensure channel identifiability, it is assumed that a) the channel length is lower than or equal to the observation window ($L \leq MN_c$) and that b) the sequence $t(n)$ is persistently exciting¹ of order L or higher. These are sufficient conditions to ensure identifiability in both training-only and semi-blind GML methods. As for the semi-blind CML estimator, it will also be assumed that c) the Q codes are linearly independent and that d) the number of symbols per code period is lower than the period itself in number of chips ($N_s < N_c$).

In any case, it is important to distinguish between the model for which each channel estimator is derived and the model assumed taken when evaluating their asymptotic performance. In this paper, and as far as the performance evaluation is concerned, we will assume the following

As1) The unknown symbols are circularly symmetric i.i.d. random variables with zero mean, unit variance $E[s_q(i)s_p^*(j)] = \delta_{i-j}\delta_{p-q}$, $i = 1 \cdots M_s(q)$, $j = 1 \cdots M_s(p)$, $p, q = 1 \cdots Q$, zero third-order moment and a finite fourth-order moment corresponding to $E[|s_q(i)|^4] \stackrel{\text{def}}{=} \zeta$.

Note that no particular distribution is imposed. We will see that under the appropriate set of assumptions, results are independent of their particular statistical distribution.

Notation: In what follows, for a certain parameter vector, e.g., \mathbf{h} , we will distinguish between the minimization variables \mathbf{h}_{to} (training-only approach) \mathbf{h}_c (conditional approach) \mathbf{h}_g (Gaussian approach) the actual estimators $\hat{\mathbf{h}}_{to}$ $\hat{\mathbf{h}}_c$ $\hat{\mathbf{h}}_g$, and the vector of true values \mathbf{h} . The same notation will be used for all the other parameters.

A. Training-Only Approach

The training only estimator disregards the presence of the traffic channels, which is equivalent to setting $\mathbf{s} = \mathbf{0}$ in (1). The ML estimator for this signal model coincides with the conventional least squares solution and can be expressed as

$$\hat{\mathbf{h}}_{to} = (\mathcal{T}^H \mathcal{T})^{-1} \mathcal{T}^H \mathbf{x}. \quad (7)$$

Examining the form of the training-only estimator in (7) and making straightforward calculations, one can state the following.

Proposition 1: Under the statistical assumption **As1)**, the random vector $\sqrt{M}(\hat{\mathbf{h}}_{to} - \mathbf{h})$ is asymptotically (in M) circularly-symmetric Gaussian-distributed with zero mean and covariance equal to $\mathbf{C}_{to} = \lim_{M \rightarrow \infty} \mathbf{C}_{to}^M$, where

$$\mathbf{C}_{to}^M = M \mathbf{B}_{to}^M + M (\mathcal{T}^H \mathcal{T})^{-1} \mathcal{T}^H \mathbf{G} \mathbf{G}^H \mathcal{T} (\mathcal{T}^H \mathcal{T})^{-1} \quad (8)$$

and

$$\mathbf{B}_{to}^M = \sigma^2 (\mathcal{T}^H \mathcal{T})^{-1}$$

¹See [5, p. 177] for a definition of persistent excitation. Note that for training sequences belonging to a finite alphabet, this condition implies that they cannot be periodical with period lower than L [6].

²All these statistical assumptions are verified by most of the modulation constellations used in digital communications, cf. [7].

is the Cramér–Rao bound for the training-only scenario (cf. [8]).

Proof: Convergence in law to a Gaussian distribution follows from the central limit theorem for m -dependent processes [9, Th. 6.4.2] as $M \rightarrow \infty$ (in our case, $m = 2$ in the sense that

$$\mathcal{T}^H \mathbf{x} = \sum_{m=1}^M \mathcal{T}^H(m) \mathbf{x}(m)$$

and $E[\mathbf{x}(m)\mathbf{x}(m+t)^H] = \mathbf{0} \forall t > m = 2$). The covariance matrix can readily be obtained from (7). ■

Note that due to the presence of the traffic signal (second term in \mathbf{C}_{to}^M), the estimator becomes inefficient even at high SNRs [that is, the second term in (8) remains fixed when $\sigma^2 \rightarrow 0$]. This poor performance of the estimator can be improved with an explicit modeling of the traffic channels in the signal model of (1). In particular, one can model the unknown data either as deterministic parameters (conditional approach) or as random variables (unconditional approach); see [10] and references therein. These two approaches will lead to two distinct estimators, that, as already shown in [8], [11], and [12], do not perform equivalently.

B. Conditional (Deterministic) ML Approach

If we model the unknown data as deterministic parameters, the ML estimator for both data and channel impulse response can be obtained minimizing the following negative log-likelihood function

$$MN_c \log(\pi \sigma_c^2) + \frac{1}{\sigma_c^2} \|\mathbf{x} - \mathcal{T} \mathbf{h}_c - \mathbf{G}_c \mathbf{s}_c\|^2 \quad (9)$$

where $\mathbf{G}_c = \mathcal{C}(\mathbf{I}_{M_s} \otimes \mathbf{h}_c)$ and subindexes c in \mathbf{s}_c , \mathbf{h}_c and σ_c^2 denote conditional model. The unknown symbols can be estimated as

$$\hat{\mathbf{s}}_c = \left[\hat{\mathbf{G}}_c^H \hat{\mathbf{G}}_c \right]^{-1} \hat{\mathbf{G}}_c^H (\mathbf{x} - \mathcal{T} \hat{\mathbf{h}}_c) \quad (10)$$

where $\hat{\mathbf{h}}_c$ is the CML estimate of the channel impulse response, and $\hat{\mathbf{G}}_c$ is equal to \mathbf{G}_c replacing \mathbf{h}_c with $\hat{\mathbf{h}}_c$. Plugging (10) for the general case $\hat{\mathbf{h}}_c = \mathbf{h}_c$ into (9), we can reduce the dimensionality of the parameter space of the cost function

$$\eta_c(\mathbf{h}_c, \sigma_c^2) = MN_c \log(\pi \sigma_c^2) + \frac{1}{\sigma_c^2} (\mathbf{x} - \mathcal{T} \mathbf{h}_c)^H \mathbf{P}_{\mathbf{G}_c}^\perp (\mathbf{x} - \mathcal{T} \mathbf{h}_c) \quad (11)$$

where $\mathbf{P}_{\mathbf{G}_c}^\perp = \mathbf{I}_{MN_c} - \mathbf{G}_c [\mathbf{G}_c^H \mathbf{G}_c]^{-1} \mathbf{G}_c^H$ denotes the orthogonal projection matrix onto the null space of the columns of \mathbf{G}_c .

The semi-blind CML channel estimator has to be obtained after an exhaustive search for the global minimum of (11). However, if we have a good initialization for the estimator (this could be provided by a training-only estimation), we can get to the CML solution forcing the gradient of different sets of parameters to zero while assuming that the others are constant (this is the traditional alternating-maximizing approach proposed in [13]). Of course, depending on the initialization of the algorithm, the solution might converge to local minima rather than the global one. In addition, depending on the actual size of the matrices, solving the set of equations might be prohibitive from a computational point of view. In any case, it seems interesting

to analyze the estimator equations in the neighborhood of the actual channel response to gain insight into the inherent process of ML estimation.

In the CML semi-blind case, estimates of the channel impulse response and the noise power can be readily obtained minimizing either (9) or (11)

$$\hat{\sigma}_c^2 = \frac{1}{MN_c} (\mathbf{x} - \mathcal{T} \hat{\mathbf{h}}_c)^H \mathbf{P}_{\hat{\mathbf{G}}_c}^\perp (\mathbf{x} - \mathcal{T} \hat{\mathbf{h}}_c)$$

$$\hat{\mathbf{h}}_c = \left(\hat{\mathbf{R}}_c^H \hat{\mathbf{R}}_c \right)^{-1} \hat{\mathbf{R}}_c^H \mathbf{x}$$

being $\hat{\mathbf{R}}_c = \mathcal{T} + \hat{\mathcal{T}}_{u,c}$, where $\hat{\mathcal{T}}_{u,c} = \mathcal{C}(\hat{\mathbf{s}}_c \otimes \mathbf{I}_L)$ is a structured reconstruction of the convolution matrix associated with the unknown component of the received signal. Fig. 2(a) is a representation in a recursive block diagram of the equation that, when properly initialized, gives the CML semi-blind estimation. First, the known component of the signal is removed from the observation, and the result is passed through a zero-forcing (ZF) equalizer to obtain estimates of the unknown symbols. These estimates are then used to regenerate the unknown component of the received signal. The convolution matrix of the global regenerated signal $\hat{\mathbf{R}}_c$ is finally used to perform a least squares channel estimation. This is a well-known interpretation of the CML process of estimation; see, for instance, [14] and [15].

Let us now discuss the asymptotic performance of this estimator.

Proposition 2: Under **As1**, the normalized conditional channel estimator $\sqrt{M} \hat{\mathbf{h}}_c$ is asymptotically ($M \rightarrow \infty$) circularly symmetric Gaussian-distributed with covariance $\lim_{M \rightarrow \infty} \mathbf{C}_c$, where

$$\mathbf{C}_c^M = M \mathbf{B}_c^M + M \mathbf{B}_c^M \left[\frac{1}{M} \sum_{i,j=1}^{M_s} \chi_{i,j} \mathcal{S}_i^H \mathbf{P}_{\mathbf{G}_c}^\perp \mathcal{S}_j \right] M \mathbf{B}_c^M \quad (12)$$

$$\chi_{i,j} = \sigma^2 \left\{ (\mathbf{G}^H \mathbf{G})^{-1} \right\}_{j,i} \quad (13)$$

and \mathbf{B}_c^M is the conditional Cramér–Rao bound

$$M \mathbf{B}_c^M = \sigma^2 \left[\frac{1}{M} \left(\mathcal{T}^H \mathbf{P}_{\mathbf{G}_c}^\perp \mathcal{T} + \sum_{i=1}^{M_s} \mathcal{S}_i^H \mathbf{P}_{\mathbf{G}_c}^\perp \mathcal{S}_i \right) \right]^{-1}. \quad (14)$$

Furthermore, these results hold, regardless of the assumption about the distribution of the symbols vector \mathbf{s} .

Proof: See Appendix A. ■

The conclusion derived from the estimator covariance in (12) is clear. The conditional channel estimator is inefficient for finite values of the SNR. This is due to the finite number of chips per symbol SF_q , from which it is not possible to obtain a consistent estimation of the unknown data as M grows. This effect can be alleviated using a Gaussian statistical model for the transmitted symbols.

C. Gaussian ML Approach

According to the Gaussian ML approach, symbols are modeled as complex mutually independent circularly symmetric Gaussian random variables with zero mean and unit variance. A Gaussian ML approach is preferred to a strictly unconditional

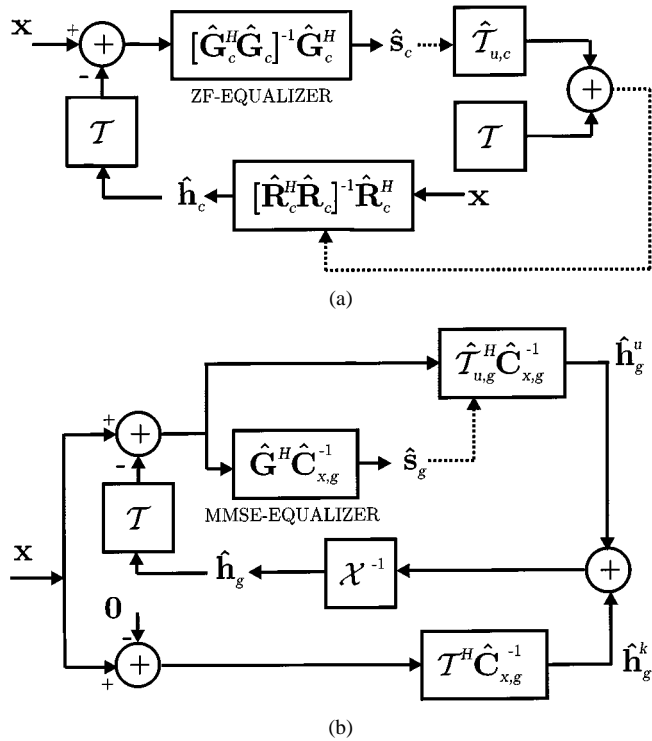


Fig. 2. Schematic representation of the (a) CML and (b) GML channel estimators.

ML estimator (where symbols would be modeled with a discrete uniform distribution) because we seek to obtain methods based on second-order statistics of the received signal, which are always more affordable from the computational point of view.

The negative log-likelihood function to be minimized can be expressed as

$$\eta_g(\sigma_g^2, \mathbf{h}_g) = \log \det(\pi \mathbf{C}_{x,g}) + \text{tr} \left(\mathbf{C}_{x,g}^{-1} \tilde{\mathbf{C}}_{x,g} \right) \quad (15)$$

where \mathbf{C}_x and $\tilde{\mathbf{C}}_x$ stand for the temporal covariance matrix associated with the random component of the signal and its rank-one sample estimate, respectively

$$\mathbf{C}_x = \mathbf{G}\mathbf{G}^H + \sigma^2 \mathbf{I}_{MN_c} \quad (16)$$

$$\tilde{\mathbf{C}}_x = (\mathbf{x} - \mathbf{T}\mathbf{h})(\mathbf{x} - \mathbf{T}\mathbf{h})^H \quad (17)$$

and $\mathbf{C}_{x,g}$ and $\tilde{\mathbf{C}}_{x,g}$ are defined replacing in the last equations the true parameters σ^2 , \mathbf{h} with the minimization variables σ_g^2 , \mathbf{h}_g . Note that unlike in the Conditional case, the estimation of channel impulse response and noise power are not decoupled.

The GML channel estimator can be obtained as one of the solutions to the equation obtained forcing the gradient of (15) with respect to the channel vector to zero

$$\hat{\mathbf{h}}_g = \left[\sum_{i=1}^{M_s} \mathcal{S}_i^H \hat{\mathbf{C}}_{x,g}^{-1} \mathcal{S}_i + \mathbf{T}^H \hat{\mathbf{C}}_{x,g}^{-1} \mathbf{T} \right]^{-1} \cdot \left[\hat{\mathbf{T}}_{u,g}^H \hat{\mathbf{C}}_{x,g}^{-1} (\mathbf{x} - \mathbf{T}\hat{\mathbf{h}}_g) + \mathbf{T}^H \hat{\mathbf{C}}_{x,g}^{-1} \mathbf{x} \right] \quad (18)$$

where $\hat{\mathbf{C}}_{x,g}$ is equal to $\mathbf{C}_{x,g}$, replacing the minimization variables with their GML estimates. Matrix $\hat{\mathbf{T}}_{u,g}$ can be interpreted

as a structured reconstruction of the unknown component of the signal

$$\hat{\mathbf{T}}_{u,g} = \sum_{i=1}^{M_s} \hat{\zeta}_i \mathcal{S}_i$$

where $\hat{\zeta}_i$ is the output of a minimum mean squared error (MMSE) equalizer matched to the global impulse response

$$\hat{\zeta}_i = \hat{\mathbf{h}}_g^H \mathcal{S}_i^H \hat{\mathbf{C}}_{x,g}^{-1} (\mathbf{x} - \mathbf{T}\hat{\mathbf{h}}_g). \quad (19)$$

The noise power estimator can be calculated as one of the solutions to

$$\text{tr} \left[\mathbf{C}_{x,g}^{-1} - \mathbf{C}_{x,g}^{-1} \tilde{\mathbf{C}}_{x,g} \mathbf{C}_{x,g}^{-1} \right] = 0$$

which is obtained forcing the derivative of η_g with respect to σ_g^2 to zero.

Fig. 2(b) is a schematic representation of the equation that gives the GML solution. We see from (18) that as it happened with the CML estimator, the semi-blind GML can be interpreted as a combination of linear estimators and signal regenerators. In this case, the observation is passed through two different branches. In the upper branch, the known component of the received signal is first removed; the result is then used to obtain an estimation of the unknown symbols through a MMSE equalizer (as opposed to the ZF equalizer in the conditional approach) in order to regenerate the unknown component of the received signal. This regenerated signal is used to obtain a first conditional mean estimate of the channel, which is denoted by $\hat{\mathbf{h}}_g^u$. In the lower branch, the estimator first removes the mean value of the unknown component of the received signal (assumed $\mathbf{0}$ here), and the result is used to obtain a second conditional mean estimate of the channel, which is represented as $\hat{\mathbf{h}}_g^k$ since it is obtained from the known component of the signal. Finally, these two estimations are combined, and the resulting bias is removed after projection onto the inverse of the matrix $\chi = \sum_{i=1}^{M_s} \mathcal{S}_i^H \hat{\mathbf{C}}_{x,g}^{-1} \mathcal{S}_i + \mathbf{T}^H \hat{\mathbf{C}}_{x,g}^{-1} \mathbf{T}$.

Let us now discuss the asymptotic performance of the estimator as $M \rightarrow \infty$.

Proposition 3: Under **As1**, the normalized real-valued estimator

$$\hat{\boldsymbol{\theta}}_g = \sqrt{M} \left[\hat{\sigma}_g^2 \quad \text{Re}^T(\hat{\mathbf{h}}_g) \quad \text{Im}^T(\hat{\mathbf{h}}_g) \right]^T \in \mathbb{R}^{2L+1 \times 1} \quad (20)$$

is asymptotically ($M \rightarrow \infty$) Gaussian-distributed with covariance depending on the statistical model for the unknown symbols through one of their fourth-order moments $\zeta = E[|s(n)|^4]$. The asymptotic (in M) covariance equals $\mathbf{C}_{\hat{\boldsymbol{\theta}}_g}^M = \lim_{M \rightarrow \infty} \mathbf{C}_{\hat{\boldsymbol{\theta}}_g}^M$, where

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}_g}^M = M \mathbf{B}_{\hat{\boldsymbol{\theta}}_g}^M + 2M^2(\zeta - 2) \mathbf{B}_{\hat{\boldsymbol{\theta}}_g}^M \cdot \begin{bmatrix} f_0^M & \text{Re}^T(\mathbf{f}_1^M) & \text{Im}^T(\mathbf{f}_1^M) \\ \text{Re}(\mathbf{f}_1^M) & \text{Re}[\mathbf{F}_2^M + \mathbf{F}_3^M] & -\text{Im}[\mathbf{F}_2^M - \mathbf{F}_3^M] \\ \text{Im}(\mathbf{f}_1^M) & \text{Im}[\mathbf{F}_2^M + \mathbf{F}_3^M] & \text{Re}[\mathbf{F}_2^M - \mathbf{F}_3^M] \end{bmatrix} \mathbf{B}_{\hat{\boldsymbol{\theta}}_g}^M \quad (21)$$

with $\mathbf{B}_{\theta_g}^M$ representing the Gaussian Cramér–Rao bound

$$M\mathbf{B}_{\theta_g}^M = \frac{1}{2} \begin{bmatrix} c_0^M & \text{Re}^T(\mathbf{e}_1^M) & \text{Im}^T(\mathbf{e}_1^M) \\ \text{Re}(\mathbf{e}_1^M) & \text{Re}[\mathbf{E}_2^M + \mathbf{E}_3^M] & -\text{Im}[\mathbf{E}_2^M - \mathbf{E}_3^M] \\ \text{Im}(\mathbf{e}_1^M) & \text{Im}[\mathbf{E}_2^M + \mathbf{E}_3^M] & \text{Re}[\mathbf{E}_2^M - \mathbf{E}_3^M] \end{bmatrix}^{-1}$$

and

$$\begin{aligned} c_0^M &= \frac{1}{2M} \text{tr} [\mathbf{C}_x^{-2}] \\ \mathbf{e}_1^M &= \frac{1}{M} \sum_{i=1}^{M_s} \mathcal{S}_i^H \mathbf{C}_x^{-2} \mathcal{S}_i \mathbf{h} \\ \mathbf{E}_2^M &= \frac{1}{M} \mathbf{T}^H \mathbf{C}_x^{-1} \mathbf{T} + \frac{1}{M} \sum_{i,j=1}^{M_s} (\mathbf{h}^H \mathcal{S}_i^H \mathbf{C}_x^{-1} \mathcal{S}_j \mathbf{h}) \mathcal{S}_i^H \mathbf{C}_x^{-1} \mathcal{S}_j \\ \mathbf{E}_3^M &= \frac{1}{M} \sum_{i,j=1}^{M_s} \mathcal{S}_i^H \mathbf{C}_x^{-1} \mathcal{S}_j \mathbf{h} \mathbf{h}^T \mathcal{S}_i^T \mathbf{C}_x^{-T} \mathcal{S}_j^* \\ f_0^M &= \frac{1}{2M} \text{tr} [(\mathbf{G}^H \mathbf{C}_x^{-2} \mathbf{G} \odot \mathbf{I}_{M_s})^2] \\ \mathbf{f}_1^M &= \frac{1}{M} \sum_{i=1}^{M_s} \mathcal{S}_i^H \mathbf{C}_x^{-1} \mathbf{G} [\mathbf{G}^H \mathbf{C}_x^{-2} \mathbf{G} \odot \mathbf{I}_{M_s}] \mathbf{G}^H \mathbf{C}_x^{-1} \mathcal{S}_i \mathbf{h} \\ \mathbf{F}_2^M &= \frac{1}{M} \sum_{i,j=1}^{M_s} \mathcal{S}_i^H \mathbf{C}_x^{-1} \mathbf{G} \\ &\quad \cdot [\mathbf{G}^H \mathbf{C}_x^{-1} \mathcal{S}_i \mathbf{h} \mathbf{h}^H \mathcal{S}_j^H \mathbf{C}_x^{-1} \mathbf{G} \odot \mathbf{I}_{M_s}] \mathbf{G}^H \mathbf{C}_x^{-1} \mathcal{S}_j \\ \mathbf{F}_3^M &= \frac{1}{M} \sum_{i,j=1}^{M_s} \mathcal{S}_i^H \mathbf{C}_x^{-1} \mathbf{G} \\ &\quad \cdot [\mathbf{G}^H \mathbf{C}_x^{-1} \mathcal{S}_j \mathbf{h} \mathbf{h}^T \mathcal{S}_i^T \mathbf{C}_x^{-T} \mathbf{G}^* \odot \mathbf{I}_{M_s}] \mathbf{G}^T \mathbf{C}_x^{-T} \mathcal{S}_j^* \end{aligned} \quad (22)$$

where \odot denotes the Schur–Hadamard product.

Proof: See Appendix B. \blacksquare

We see that in general terms, the GML semi-blind estimator is neither circularly symmetric nor asymptotically robust. Furthermore, given the complicated expressions in (22) and the intrinsic dependence on the training and code sequences, it is difficult to draw any conclusion about the behavior of the estimator. In the next section, we simplify these expressions, assuming that both training and codes are random variables.

IV. RANDOMIZED CODES AND TRAINING SEQUENCE

In this section, we investigate the asymptotic behavior of the covariance matrices for semi-blind and training-only estimation algorithms assuming the following.

As2) The spreading codes are circularly symmetric i.i.d. random variables with zero mean, variance $E[c_p(n)c_q^*(m)] = (\alpha_q/QN_c)\delta_{p-q}\delta_{n-m}$, $n, m = 1 \cdots N_c$, $q = 1 \cdots Q$, and finite fourth-order moments that are independent of the unknown symbols.

As3) The training sequences are circularly symmetric i.i.d. random variables with zero mean, variance $E[t(n)t^*(m)] = (\alpha^t/N_c)\delta_{n-m}$,

$n, m = -L + 2 \cdots MN_c$, and independent of the unknown symbols and the spreading sequences.

Note that under **As2** and **As3**, the sequences $c_q(n)$ $q = 1 \cdots Q$ are almost surely mutually linearly independent, and the training sequence $t(n)$ is persistently exciting of all orders so that identifiability is guaranteed with probability one whenever $MN_c \geq L$ (as well as $N_s < N_c$ in the semi-blind CML case).

Two different limits will be considered: a limit when the number of codes (Q) tends to infinity while their period (N_c) remains constant (high spectral efficiency case) and a limit as the period and spreading factors grow at the same rate for a fixed number of codes (low spectral efficiency case). Due to space limitations, we do not derive these asymptotic limits here. See [16] for a complete proof of these results.

A. High Spectral Efficiency Case

The limit as the number of codes Q grows without bound for a fixed code period can only be applied to the training-only and GML semi-blind estimators because the CML estimator is not defined for $N_s > N_c$.

1) *Training-Only Estimator:* Under **As2** and **As3** and as $Q \rightarrow \infty$, the covariance matrix of the training-only estimator tends in probability to³

$$\mathbf{C}_{to}^{heff} = \underbrace{\frac{\sigma^2}{\alpha^t} \mathbf{I}_L}_{MB_{to}^{heff}} + \frac{\bar{\alpha}}{\alpha^t N_c} \mathbf{R}_h \quad (23)$$

where $\bar{\alpha} = (1/Q) \sum_{q=1}^Q \alpha_q$

$$\mathbf{R}_h = \begin{bmatrix} r_h(0) & \cdots & r_h^*(L-1) \\ \vdots & \ddots & \vdots \\ r_h(L-1) & \cdots & r_h(0) \end{bmatrix} \in \mathbb{C}^{L \times L}$$

and

$$r_h(l) = \sum_{k=1}^{L-l} h^*(k)h(k+l), \quad l = 0 \cdots L-1.$$

Defining the global SNR as

$$\gamma_x = \frac{\|\mathbf{h}\|^2(\alpha^t + \bar{\alpha})}{\sigma^2} \quad (24)$$

the normalized asymptotic covariance matrix for the training-only case can be expressed as

$$\frac{\mathbf{C}_{to}^{heff}}{\|\mathbf{h}\|^2} = \frac{1}{\gamma_x} \left(1 + \frac{\alpha}{\alpha^t}\right) \mathbf{I}_L + \frac{1}{N_c} \left(\frac{\alpha}{\alpha^t}\right) \frac{\mathbf{R}_h}{\|\mathbf{h}\|^2}.$$

When the SNR increases without bound, the relative asymptotic covariance matrix tends to the constant value

$$\lim_{\gamma_x \rightarrow \infty} \frac{\mathbf{C}_{to}^{heff}}{\|\mathbf{h}\|^2} = \frac{1}{N_c} \left(\frac{\alpha}{\alpha^t}\right) \frac{\mathbf{R}_h}{\|\mathbf{h}\|^2}. \quad (25)$$

Due to the presence of the traffic channels ($\alpha \neq 0$) and the finite period of the code sequence ($N_c < \infty$), the estimator can never attain the Cramér–Rao bound, however high the SNR might be. Interestingly enough, the estimator seems to be inefficient even

³Invoking the Lebesgue dominated convergence theorem, it can also be seen that this is the mean value of the covariance matrix for any fixed Q .

when dealing with frequency nonselective channels ($L = 1$). This is a result that, due to the random code approximation, is too pessimistic. In practice, these codes are designed to be orthogonal to the training sequence to ensure that efficiency is obtained when $L = 1$. In any case, it is worth bearing in mind that the approximation of the channel estimator covariance matrix derived under random code sequences in (23) has to be handled carefully since it may be too pessimistic in certain scenarios.

2) *Semi-Blind GML Estimator*: Assuming that **As2** and **As3** hold and that the number of codes increases without bound, the quantities c_0^M , \mathbf{e}_1^M , \mathbf{E}_2^M and \mathbf{E}_3^M tend, respectively, in probability to

$$c_0^{heff} = \frac{N_c}{4\pi} \int_0^{2\pi} \frac{1}{\left(\frac{\bar{\alpha}}{N_c} S_h(w) + \sigma^2\right)^2} dw$$

$$\left\{ \mathbf{e}_1^{heff} \right\}_k = \frac{1}{2\pi} \int_0^{2\pi} \bar{\alpha} \frac{H(w)}{\left(\frac{\bar{\alpha}}{N_c} S_h(w) + \sigma^2\right)^2} \exp[jwk] dw$$

$$\left\{ \mathbf{E}_2^{heff} \right\}_{i,k} = \frac{1}{2\pi} \int_0^{2\pi} \left[\alpha^t \frac{1}{\frac{\bar{\alpha}}{N_c} S_h(w) + \sigma^2} + \bar{\alpha} \frac{\frac{\bar{\alpha}}{N_c} S_h(w)}{\left(\frac{\bar{\alpha}}{N_c} S_h(w) + \sigma^2\right)^2} \right] \cdot \exp[jw(k-i)] dw$$

$$\left\{ \mathbf{E}_3^{heff} \right\}_{i,k} = \frac{1}{2\pi} \int_0^{2\pi} \bar{\alpha} \frac{\frac{\bar{\alpha}}{N_c} H(-w)H(w)}{\left(\frac{\bar{\alpha}}{N_c} S_h(w) + \sigma^2\right)^2} \cdot \exp[jw(k-i)] dw$$

where

$$H(w) = \sum_{l=1}^L h(l) \exp[-jwl]$$

is the channel frequency response, and $S_h(w) = |H(w)|^2$ its spectral power density. The values of f_0^M , \mathbf{f}_1^M , \mathbf{F}_2^M , and \mathbf{F}_3^M tend to somewhat complicated quantities that depend on one of the fourth-order moments of the code sequences $c_q(n)$ and are thus omitted here for clarity. It suffices to say that their contribution to the actual covariance is relatively low for typical symbol kurtosis and that it vanishes at both high and low SNRs.

3) *Performance Comparison (High Spectral Efficiency)*: For comparison purposes, it is interesting to express the Gaussian Cramér–Rao bound associated with the reduced set of parameters $\hat{\boldsymbol{\varphi}}_g = \sqrt{M} [\text{Re}^T[\hat{\mathbf{h}}_g] \text{Im}^T[\hat{\mathbf{h}}_g]]^T$, namely

$$MB_{\hat{\boldsymbol{\varphi}}_g}^{heff} = \frac{1}{2} \left[\begin{array}{cc} \text{Re} \left[\mathbf{J}_2^{heff} + \mathbf{J}_3^{heff} \right] & -\text{Im} \left[\mathbf{J}_2^{heff} - \mathbf{J}_3^{heff} \right] \\ \text{Im} \left[\mathbf{J}_2^{heff} + \mathbf{J}_3^{heff} \right] & \text{Re} \left[\mathbf{J}_2^{heff} - \mathbf{J}_3^{heff} \right] \end{array} \right]^{-1}$$

where

$$\mathbf{J}_2^{heff} = \mathbf{E}_2^{heff} - \frac{\mathbf{e}_1^{heff} \left(\mathbf{e}_1^{heff} \right)^H}{2c_0^{heff}}$$

$$\mathbf{J}_3^{heff} = \mathbf{E}_3^{heff} - \frac{\mathbf{e}_1^{heff} \left(\mathbf{e}_1^{heff} \right)^T}{2c_0^{heff}}.$$

In [16], it is proven that

$$MB_{\hat{\boldsymbol{\varphi}}_g}^{heff} \leq \frac{1}{2} \left[\begin{array}{cc} \text{Re} \left[\mathbf{C}_{to}^{heff} \right] & -\text{Im} \left[\mathbf{C}_{to}^{heff} \right] \\ \text{Im} \left[\mathbf{C}_{to}^{heff} \right] & \text{Re} \left[\mathbf{C}_{to}^{heff} \right] \end{array} \right]$$

where for two square matrices, $\mathbf{A} \leq \mathbf{B}$ means that $\mathbf{B} - \mathbf{A}$ is positive semidefinite.

Fig. 3 represents the averaged trace of the channel estimation covariance under the approximation of high spectral efficiency for the two cases under consideration. In this example, the channel impulse response taps were randomly chosen with a uniformly distributed phase and an exponentially decaying power delay profile. The delay spread is given by the inverse of ϕ , as shown in the figure. The higher the angular spread, the higher the gain that can be obtained from using semiblind identification techniques.

B. Low Spectral Efficiency Case

We now analyze the asymptotic behavior of the different covariance matrices for a fixed number of codes (Q) and a code period N_c increasing without bound. The number of symbols per period $N_s(q)$ is assumed fixed for each code, which implicitly implies that the different spreading factors (SF_q) scale up at the same rate as N_c .

1) *Training-Only Estimator*: Under the statistical assumptions **As2** and **As3**, if the period of the spreading codes and the corresponding spreading factors increase without bound at the same rate ($N_c, SF_q \rightarrow \infty$) while their quotient remains constant, the covariance of the training-only estimator tends in probability to

$$\mathbf{C}_{to}^{eff} = MB_{to}^{eff} = \frac{\sigma^2}{\alpha^t} \mathbf{I}_L.$$

2) *Semi-Blind CML Estimator*: Assume that **As2** and **As3** hold. Then, as $N_c, SF_q \rightarrow \infty$, the covariance of the semi-blind CML channel estimator tends in probability to

$$\mathbf{C}_c^{eff} = MB_c^{eff} + \frac{\|\mathbf{h}\|^2 N_s}{\gamma_x^2} \mathbf{P}_h^\perp$$

$$MB_c^{eff} = \frac{\|\mathbf{h}\|^2}{\gamma_x} \left[\mathbf{P}_h^\perp + \left(1 + \frac{\bar{\alpha}}{\alpha^t} \right) \mathbf{P}_h \right]$$

where γ_x is defined in (24), and $\mathbf{P}_h = \mathbf{I}_L - \mathbf{P}_h^\perp = \mathbf{h}\mathbf{h}^H / \|\mathbf{h}\|^2$.

Two observations are in order. First, note that under these asymptotic conditions, the covariance of the CML channel estimator does not depend on how the symbols are distributed across the different codes or what the power associated with each code is. Instead, it depends on the total number of transmitted symbols per code period N_s and the global power associated with the codes α . Second, we see now more clearly that

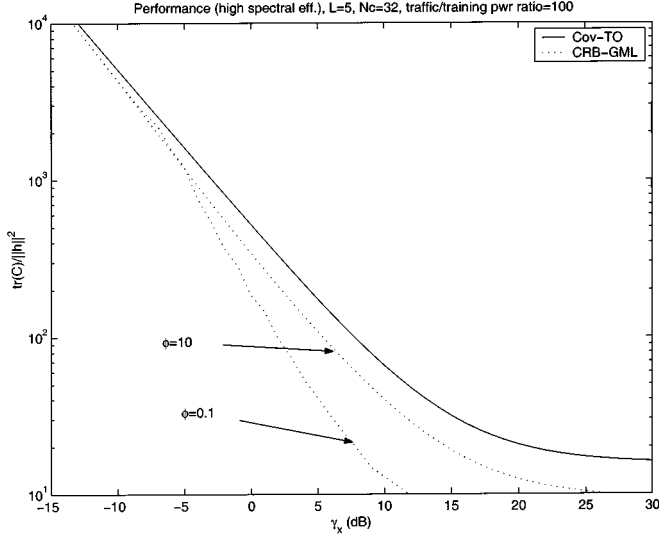


Fig. 3. Asymptotic performance of the training-only and semibind GML channel estimation algorithms for the high spectral efficiency case.

the second term of (12) depends quadratically with the inverse of γ_x , thus tending to zero faster than the Cramér–Rao bound as the SNR unit increases.

3) *Semi-Blind GML Estimator*: Assume once again that **As2** and **As3** hold. Then, provided that the period of the spreading codes and the corresponding spreading factors increase without bound at the same rate ($N_c, SF_q \rightarrow \infty$) while their quotient remains constant, the asymptotic covariance of the Gaussian channel estimator $\hat{\boldsymbol{\varphi}}_g = \sqrt{M}[\text{Re}^T[\hat{\mathbf{h}}_g] \quad \text{Im}^T[\hat{\mathbf{h}}_g]]^T$ tends in probability to⁴

$$\mathbf{C}_{\hat{\boldsymbol{\varphi}}_g}^{leff} = M\mathbf{B}_{\hat{\boldsymbol{\varphi}}_g}^{leff} + \frac{\zeta/2 - 1}{\|\mathbf{h}\|^2} \rho \cdot \begin{bmatrix} \text{Re}[\mathbf{h}\mathbf{h}^H + \mathbf{h}\mathbf{h}^T] & -\text{Im}[\mathbf{h}\mathbf{h}^H - \mathbf{h}\mathbf{h}^T] \\ \text{Im}[\mathbf{h}\mathbf{h}^H + \mathbf{h}\mathbf{h}^T] & \text{Re}[\mathbf{h}\mathbf{h}^H - \mathbf{h}\mathbf{h}^T] \end{bmatrix} \quad (26)$$

$$M\mathbf{B}_{\hat{\boldsymbol{\varphi}}_g}^{leff} = \frac{1}{2} \begin{bmatrix} \text{Re}[\mathbf{V}_2 + \mathbf{V}_3] & -\text{Im}[\mathbf{V}_2 - \mathbf{V}_3] \\ \text{Im}[\mathbf{V}_2 + \mathbf{V}_3] & \text{Re}[\mathbf{V}_2 - \mathbf{V}_3] \end{bmatrix} \quad (27)$$

where

$$\mathbf{V}_2 = \frac{1}{e_2^{leff}} \mathbf{P}_h^\perp + \frac{\sigma^2}{\alpha^t} \cdot \begin{bmatrix} \frac{\alpha^t}{\sigma^2} + \|\mathbf{h}\|^2 \left(e_3^{leff} - \frac{1}{e_0^{leff}} (e_1^{leff})^2 \right) \\ \frac{\alpha^t}{\sigma^2} + 2\|\mathbf{h}\|^2 \left(e_3^{leff} - \frac{1}{e_0^{leff}} (e_1^{leff})^2 \right) \end{bmatrix} \mathbf{P}_h \quad (28)$$

$$\mathbf{V}_3 = -\frac{\sigma^2}{\alpha^t} \begin{bmatrix} \|\mathbf{h}\|^2 \left(e_3^{leff} - \frac{1}{2e_0^{leff}} (e_1^{leff})^2 \right) \\ \frac{\alpha^t}{\sigma^2} + 2\|\mathbf{h}\|^2 \left(e_3^{leff} - \frac{1}{2e_0^{leff}} (e_1^{leff})^2 \right) \end{bmatrix} \frac{\mathbf{h}\mathbf{h}^T}{\|\mathbf{h}\|^2} \quad (29)$$

⁴We do not include the covariance of the noise power estimator for simplicity. However, it should be stressed that noise and channel estimator are not decoupled and that the influence of this coupling is included in (26).

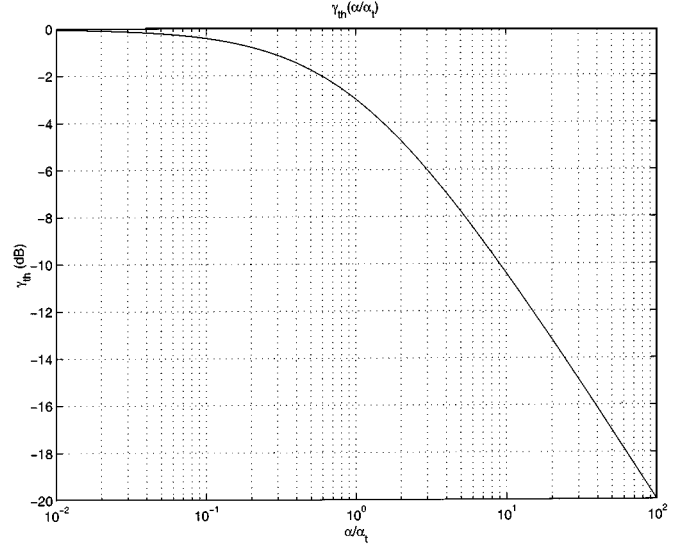


Fig. 4. Threshold symbol energy to noise power density γ_{th} as a function of the traffic to training power ratio α/α^t .

and where $\gamma_s(q)$ is the symbol energy to noise power spectral density associated with channel q

$$\gamma_s(q) = \frac{\alpha_q \|\mathbf{h}\|^2}{\sigma^2 Q N_s(q)} \quad (30)$$

and

$$e_0^{leff} = \frac{1}{2\sigma^4} \sum_{q=1}^Q N_s(q) \left[\frac{1}{1 + \gamma_s(q)} \right]^2$$

$$e_1^{leff} = \frac{1}{\sigma^2} \sum_{q=1}^Q N_s(q) \frac{\gamma_s(q)}{[1 + \gamma_s(q)]^2} \frac{1}{\|\mathbf{h}\|^2}$$

$$e_2^{leff} = \frac{1}{\sigma^2} \alpha^t + \frac{1}{\sigma^2} \sum_{q=1}^Q \frac{\alpha_q}{Q} \left[\frac{\gamma_s(q)}{1 + \gamma_s(q)} \right]$$

$$e_3^{leff} = \frac{1}{\sigma^2} \sum_{q=1}^Q \frac{\alpha_{q(i)}}{Q} \frac{\gamma_s(q)}{(1 + \gamma_s(q))^2} \frac{1}{\|\mathbf{h}\|^2}$$

$$\rho = \frac{f_2^{leff} - 2 \left(\frac{e_1^{leff}}{2e_0^{leff}} \right) f_1^{leff} + \left(\frac{e_1^{leff}}{2e_0^{leff}} \right)^2 2f_0^{leff}}{\left[\frac{\alpha^t}{\sigma^2} + 2 \left(e_3^{leff} - \frac{1}{e_0^{leff}} (e_1^{leff})^2 \right) \|\mathbf{h}\|^2 \right]^2}$$

$$f_0^{leff} = \frac{1}{2\sigma^4} \sum_{q=1}^Q N_s(q) \frac{\gamma_s^2(q)}{(1 + \gamma_s(q))^4}$$

$$f_1^{leff} = \frac{1}{\sigma^2} \sum_{q=1}^Q N_s(q) \left(\frac{\gamma_s(q)^3}{(1 + \gamma_s(q))^4} \right) \frac{1}{\|\mathbf{h}\|^2}$$

$$f_2^{leff} = \frac{1}{\sigma^2} \sum_{q=1}^Q \frac{\alpha_q}{Q} \frac{\gamma_s(q)^3}{(1 + \gamma_s(q))^4} \frac{1}{\|\mathbf{h}\|^2}.$$

The expression of the channel estimator covariances for the Gaussian scheme are quite cumbersome and difficult to handle. Great simplifications can be obtained assuming, in addition to **As2** and **As3**, that we have the following.

As4) The power of each code sequence is fixed according to its spreading factor so that the product $\alpha_q SF_q$ remains constant for $q = 1 \dots Q$.

This seems a logical working assumption since the larger the spreading factor, the smaller the received power must be in order to obtain a given quality. In other words, hypothesis **As4)** guarantees a constant reception quality $\gamma_s(q)$ for each and every one of the transmitted codes.

Simple algebra shows that under **As1–As4)**, the Gaussian semi-blind channel estimator becomes asymptotically robust (in the sense that the performance is independent of the model chosen for the unknown data) and circularly symmetric with covariance

$$\begin{aligned} \mathbf{C}_g^{leff} &= \mathbf{M}\mathbf{B}_g^{leff} \\ &= \frac{\|\mathbf{h}\|^2}{\gamma_x} \frac{\gamma_s(q) + 1}{\gamma_s(q) + \frac{\alpha^t}{\alpha^t + \alpha}} \mathbf{P}_h^\perp + \frac{\|\mathbf{h}\|^2}{\gamma_x} \left(1 + \frac{\alpha}{\alpha^t}\right) \mathbf{P}_h. \end{aligned} \quad (31)$$

Condition **As4)** is a structural restriction that guarantees the model symmetry needed for the asymptotic robustness of the estimator in the low efficiency case.

4) *Performance Comparison (Low Spectral Efficiency)*: It is easy to establish the following relationships between the limiting covariances for the conditional, Gaussian, and training-only ML methods under **As1–As4)** when SF_q and $N_c \rightarrow \infty$ at the same rate (for a constant Q)

$$\mathbf{M}\mathbf{B}_c^{leff} \leq \mathbf{C}_g^{leff} = \mathbf{M}\mathbf{B}_g^{leff} \leq \mathbf{C}_{to}^{leff} = \mathbf{M}\mathbf{B}_{to}^{leff} \quad (32)$$

$$\mathbf{M}\mathbf{B}_g^{leff} \leq \mathbf{C}_c^{leff} \quad (33)$$

$$\left. \begin{aligned} \mathbf{M}\mathbf{B}_{to}^{leff} &\geq \mathbf{C}_c^{leff} & \gamma_s(q) &\geq \gamma_{th} \\ \mathbf{M}\mathbf{B}_{to}^{leff} &< \mathbf{C}_c^{leff} & \gamma_s(q) &< \gamma_{th} \end{aligned} \right\} \quad (34)$$

with

$$\gamma_{th} = \frac{1}{1 + \frac{\alpha}{\alpha^t}}.$$

According to (32), under the asymptotic conditions specified, the semi-blind GML estimator performs better than the training-only method. On the other hand, the inequality in (33) indicates that the conditional method performs worse than the Gaussian method. As a conclusion, the GML covariance can be interpreted, under the present assumptions and asymptotic limits, as the performance bound for both conditional and Gaussian methods, whereas the asymptotic conditional Cramér–Rao bound can never be attained with second-order approaches. All these conclusions are in perfect agreement with the results presented in [10] in the context of direction-of-arrival estimation. It should be stressed, however, that here, these relationships are only valid for the low spectral efficiency case and that the relative position of conditional and Gaussian Cramér–Rao bounds will depend on system parameters such as the number of codes and their repetition period.

Finally, it is observed from (34) that, quite surprisingly, the conditional semi-blind method can perform worse than its training-only counterpart at low values of the SNR. The threshold symbol energy to noise power density γ_{th} establishes

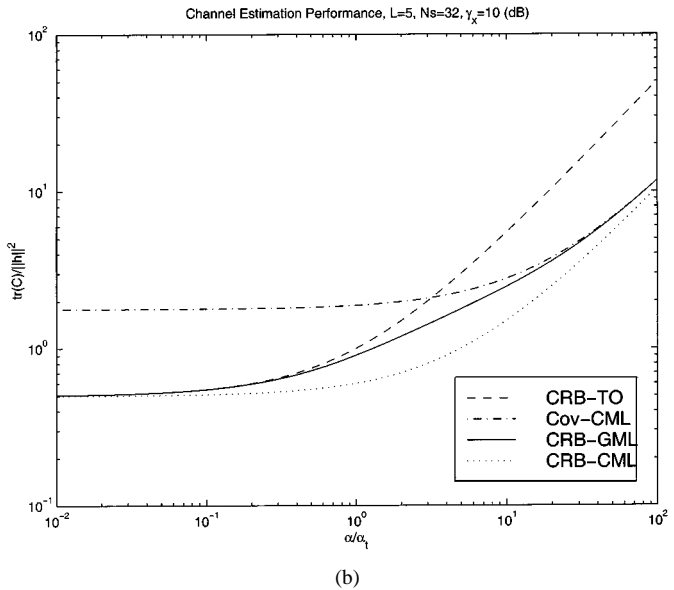
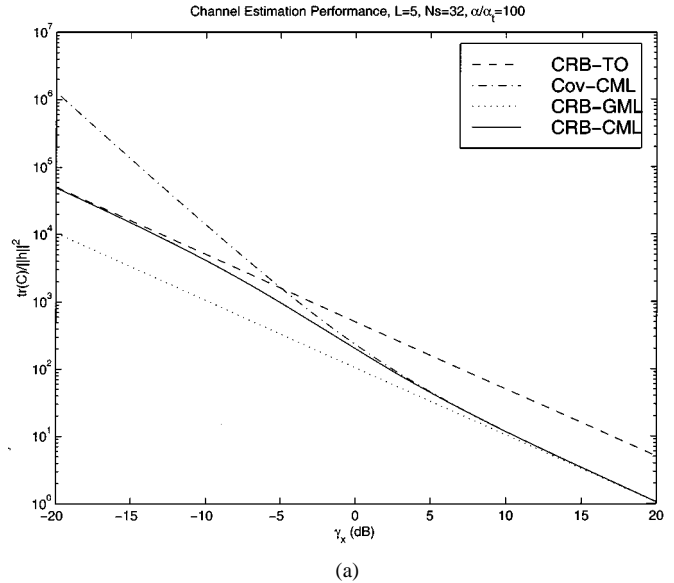
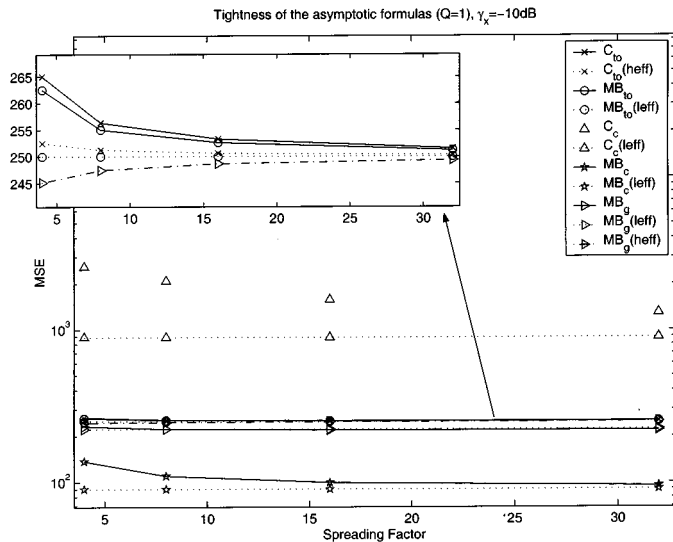


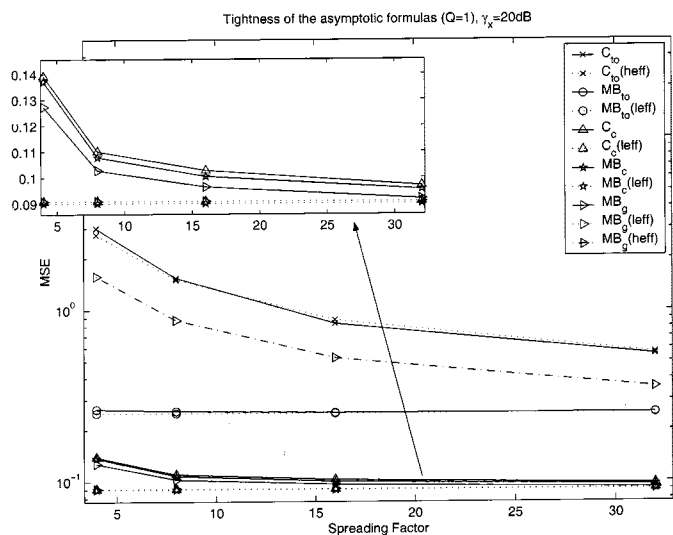
Fig. 5. Asymptotic performance of the channel estimation algorithms as a function of (a) the SNR and (b) the traffic to training power ratio.

a limit over which a performance gain can be expected from using semi-blind conditional estimation schemes under our asymptotic assumptions. For $\gamma_x < \gamma_{th}$, the introduction of new parameters (unknown symbols) in the system model will not eventually pay off in terms of asymptotic performance. In other words, the CML approach is using more information but only to estimate an increasing number of parameters, which in turn may jeopardize the overall asymptotic performance. In Fig. 4, γ_{th} is represented as a function of the traffic to training power ratio. As we see, the values of γ_{th} are low enough to guarantee that in practical situations, the semi-blind conditional scheme will perform better than the training-only estimator (values of the traffic to training power ratio are expected to be higher than one in actual WCDMA systems).

Fig. 5(a) represents the evolution of the trace of the different covariance matrices (\mathbf{C}_{to}^{leff} , \mathbf{C}_c^{leff} , \mathbf{C}_g^{leff} and $\mathbf{M}\mathbf{B}_c^{leff}$) normalized by the squared norm of the channel as a function of γ_x



(a)



(b)

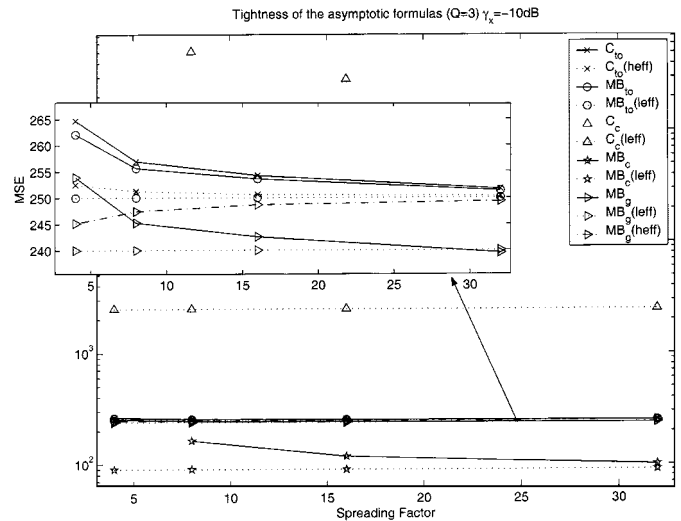
Fig. 6. Convergence of the deterministic formulas to the asymptotic ones as the spreading factor (SF) increases for a constant (a) $\gamma_x = -10$ dB and (b) $\gamma_x = 20$ dB and $Q = 1$.

for a constant $\alpha/\alpha_t = 100$. Fig. 5(b) represents the same quantity as a function of the quotient α/α_t for different values of γ_x . The theoretical relationships presented above can be verified in these two figures.

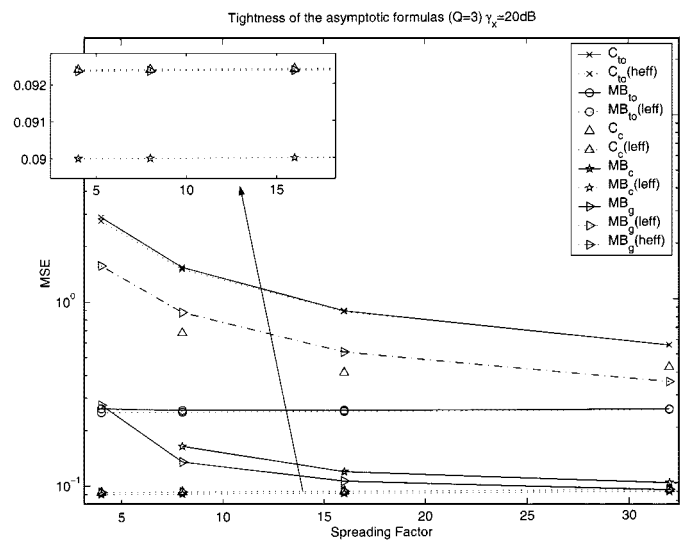
V. NUMERICAL VALIDATION

In this section, we evaluate, via simulations, the discrepancy between the asymptotic covariance matrices for deterministic training and spreading codes and those obtained under the randomized assumption. We compare the trace of the different covariances matrices normalized by the squared norm of the channel with the normalized trace of the covariance matrices under the assumption of deterministic codes and training sequence for finite N_c , SF_q , and Q .

Fig. 6 represents the convergence of the deterministic expressions to the ones obtained under the assumption of randomized training and spreading sequences as $N_c, SF_q \rightarrow \infty$ in a scenario with $N_s = 2, Q = 1, \alpha/\alpha^t = 6$ dB and $L = 5$.



(a)



(b)

Fig. 7. Convergence of the deterministic formulas to the asymptotic ones as the spreading factor (SF) increases for a constant (a) $\gamma_x = -10$ dB and (b) $\gamma_x = 20$ dB and $Q = 3$.

The channel impulse response was generated following an exponential power delay profile with time constant equal to the duration of 1 chip and random phases uniformly distributed within $(-\pi, \pi]$. The values corresponding to the covariances obtained under the deterministic codes assumption were calculated averaging 50 outcomes with $M = 10$ and using randomized codes. We represent the normalized trace of $C_{to}, C_{to}^{heff}, MB_{to}, MB_{to}^{heff}, C_c, C_c^{heff}, MB_c, MB_c^{heff}, MB_g, MB_g^{heff}$, and MB_g^{heff} (from top to bottom of the legend, respectively) with $\gamma_x = -10$ dB and $\gamma_x = 20$ dB. It is observed that the asymptotic expressions derived for the low-efficiency case are good approximations of the semi-blind variances for spreading factors of 16 or higher and become a bit too optimistic for lower values of the SF . On the other hand, we see that the covariance for the training-only scenario is, in general, much better approximated by the asymptotic covariance in the high-efficiency case

⁵In the examples considered here, $M = 10$ was sufficient to guarantee the convergence of these expressions.

(\mathbf{C}_{to}^{heff}) than the one obtained under the low spectral efficiency limit (note that \mathbf{C}_{to}^{heff} is also the mean variance with respect to training and code sequences for any fixed Q).

Fig. 7 illustrates the same evolution when the number of codes is fixed to $Q = 3$. As before, it is generally seen that in practical situations, the covariance of the training-only estimator is well approximated by the high spectral efficiency situation ($Q \rightarrow \infty$), whereas the formulas obtained for the low spectral efficiency situation ($N_c \rightarrow \infty$) are more representative of the behavior of the semi-blind estimators. This illustrates the potential benefits of semi-blind techniques applied to pilot-aided WCDMA contexts.

VI. CONCLUSIONS

We have derived asymptotic expressions describing the performance of three different channel estimation algorithms in a pilot-aided multi-rate framework. For the classic training-only channel estimation method, we have been able to provide a formula describing the mean asymptotic behavior as the number of spreading periods (M) tends to infinity. Investigating the resulting covariance matrix, we have shown that the performance of the algorithm tends to saturate as the SNR (γ_x) increases. The higher the channel length, the more inefficient training-only estimators become. Semi-blind techniques have been proposed as means to improve this bad performance. First, we have shown that the semi-blind channel estimation procedures can be described in terms of simple equalizers and signal regenerators. Then, their performance has been evaluated under asymptotic conditions in the number of spreading periods M when the spreading factor or the number of codes tend to infinity.

APPENDIX A PROOF OF PROPOSITION 2

Consistency

Let us first discuss the consistency of the estimator. The application of the strong law of large numbers (SLLN) guarantees almost-sure point-wise convergence as $M \rightarrow \infty$

$$\begin{aligned} \frac{1}{M} \eta_c(\mathbf{h}_c, \sigma_c^2) &\xrightarrow{\text{wpl}} \frac{1}{M} \bar{\eta}_c(\mathbf{h}_c, \sigma_c^2) \\ &\stackrel{\text{def}}{=} N_c \log(\pi \sigma_c^2) + \frac{1}{\sigma_c^2 M} \text{tr} [\mathbf{G}^H \mathbf{P}_{\mathbf{G}_c}^\perp \mathbf{G}] \\ &\quad + \frac{1}{\sigma_c^2 M} (\mathbf{h}_c - \mathbf{h})^H \mathcal{T}^H \mathbf{P}_{\mathbf{G}_c}^\perp \mathcal{T} (\mathbf{h}_c - \mathbf{h}). \end{aligned}$$

On the other hand, assuming that \mathbf{h}_c belongs to a dense subset of $\mathbb{C}^{L \times 1} \setminus \{\mathbf{0}\}$, it is easy to prove that the CML cost function is almost surely asymptotically uniformly equicontinuous, namely

$$\limsup_{M \rightarrow \infty} \sup_{\mathbf{h}'_c} \sup_{\mathbf{h}_c \in B(\mathbf{h}'_c, \delta)} \frac{1}{M} |\eta_c(\mathbf{h}_c) - \eta_c(\mathbf{h}'_c)| \xrightarrow{\text{wpl}} 0$$

as $\delta \rightarrow 0$, where $B(\mathbf{h}'_c, \delta)$ is a ball centered on \mathbf{h}'_c with radius δ and where the dependence on σ_c^2 has been dropped for ease of

notation. First, applying the triangular inequality and the mean value theorem, one can see that

$$\begin{aligned} |\eta_c(\mathbf{h}'_c) - \eta_c(\mathbf{h}_c)| &\leq \left| \frac{1}{\sigma^2} \Delta \mathbf{h}_c^H \mathcal{T}^H \mathbf{P}_{\mathbf{G}_{c1}}^\perp (\mathbf{x} - \mathcal{T} \mathbf{h}_{c1}) \right| \\ &\quad + \frac{1}{\sigma^4} \left| \sum_{i,j=1}^{M_s} \chi_{i,j}^{c1} \Delta \mathbf{h}_c^H \mathcal{S}_i^H \mathbf{P}_{\mathbf{G}_{c1}}^\perp \tilde{\mathbf{C}}_x^{c1} \mathcal{S}_j \mathbf{h}_{c1} \right| \end{aligned} \quad (35)$$

where $\Delta \mathbf{h}_c = \mathbf{h}_c - \mathbf{h}'_c$, \mathbf{h}_{c1} lies on the segment joining \mathbf{h}_c and \mathbf{h}'_c , and $\chi_{i,j}^{c1} = \sigma^2 \{(\mathbf{G}_{c1}^H \mathbf{G}_{c1})^{-1}\}_{i,j}$, $\tilde{\mathbf{C}}_x^{c1} = (\mathbf{x} - \mathcal{T} \mathbf{h}_{c1})(\mathbf{x} - \mathcal{T} \mathbf{h}_{c1})^H$. Note that the right-hand side of (35) is bounded, given a fixed ball radius δ . On the other hand, [17, Th. 4.2.3] implies that

$$\begin{aligned} \limsup_{M \rightarrow \infty} \frac{1}{\sqrt{M \log \log M}} \frac{1}{\sigma^2} \Delta \mathbf{h}_c^H \mathcal{T}^H \mathbf{P}_{\mathbf{G}_{c1}}^\perp (\mathbf{x} - \mathcal{T} \mathbf{h}_{c1}) \\ \leq \Delta \mathbf{h}_c^H \mathbf{k}_1 \end{aligned}$$

$$\begin{aligned} \limsup_{M \rightarrow \infty} \frac{1}{\sqrt{M \log \log M}} \frac{1}{\sigma^4} \sum_{i,j=1}^{M_s} \chi_{i,j}^{c1} \Delta \mathbf{h}_c^H \mathcal{S}_i^H \mathbf{P}_{\mathbf{G}_{c1}}^\perp \tilde{\mathbf{C}}_x^{c1} \mathcal{S}_j \mathbf{h}_{c1} \\ \leq \Delta \mathbf{h}_c^H \mathbf{k}_2 \end{aligned}$$

almost surely, where the vectors \mathbf{k}_1 and \mathbf{k}_2 are uniformly bounded for all M and $\mathbf{h}_{c1} \in B(\mathbf{h}'_c, \delta)$. Now, since $\sqrt{M \log \log M} = o(M)$, we conclude that for M sufficiently high, the right-hand side of (35) goes to zero as $\delta \rightarrow 0$. Asymptotic uniform equicontinuity follows. Now, applying the stochastic version of the Ascoli–Arzela’s theorem [18, Th. 3.1], we conclude that

$$\sup_{\mathbf{h}_c} \frac{1}{M} |\eta_c(\mathbf{h}_c) - \bar{\eta}_c(\mathbf{h}_c)| \rightarrow 0$$

almost surely as M increases without bound. This proves uniform convergence toward the limiting cost function

$$\begin{aligned} \bar{\eta}_c(\mathbf{h}_c, \sigma_c^2) &= N_c M \log(\pi \sigma_c^2) + \frac{1}{\sigma_c^2} \text{tr} [\mathbf{G}^H \mathbf{P}_{\mathbf{G}_c}^\perp \mathbf{G}] \\ &\quad + \frac{1}{\sigma_c^2} (\mathbf{h}_c - \mathbf{h})^H \mathcal{T}^H \mathbf{P}_{\mathbf{G}_c}^\perp \mathcal{T} (\mathbf{h}_c - \mathbf{h}). \end{aligned}$$

Note that the two last terms are always positive or zero. The minimum of the limiting cost function will therefore take place when the two terms vanish, and this can only happen for $\mathbf{h}_c = \mathbf{h}$ [the first term vanishes for $\mathbf{h}_c = e^{j\phi} \mathbf{h}$, where ϕ takes any value between $[0, 2\pi)$, but the second only vanishes for $\phi = 0$].

Asymptotic Normality and Covariance

Convergence in law to a Gaussian distribution follows from consistency and the asymptotic normality of the gradient vector, which can be expressed as

$$\begin{aligned} \frac{1}{\sqrt{M}} \eta'_c &= \frac{1}{\sqrt{M}} \left. \frac{\delta \eta_c}{\delta \mathbf{h}_c^*} \right|_{\mathbf{h}_c = \mathbf{h}} \\ &= -\frac{1}{\sigma^2 \sqrt{M}} \mathcal{T}^H \mathbf{P}_{\mathbf{G}}^\perp (\mathbf{x} - \mathcal{T} \mathbf{h}) \\ &\quad - \frac{1}{\sigma^4 \sqrt{M}} \sum_{i,j=1}^{M_s} \chi_{i,j} \mathcal{S}_i^H \mathbf{P}_{\mathbf{G}}^\perp \tilde{\mathbf{C}}_x \mathcal{S}_j \mathbf{h}. \end{aligned}$$

Thanks to the consistency of the conditional ML channel estimator and since channel and noise power estimators are decoupled, the asymptotic covariance matrix of the vector of the real parameters $\hat{\varphi}_c = \sqrt{M}[\text{Re}^T(\hat{\mathbf{h}}_c) \quad \text{Im}^T(\hat{\mathbf{h}}_c)]^T$ can be obtained as [19]

$$\mathbf{C}_{\hat{\varphi}_c} = E[(\hat{\varphi}_c - \varphi)(\hat{\varphi}_c - \varphi)^H] = \mathbf{H}_c^{-1} \mathbf{Q}_c \mathbf{H}_c^{-1} \quad (36)$$

where

$$\mathbf{H}_c = \lim_{M \rightarrow \infty} \frac{\delta^2 \eta_c}{\delta \varphi_c \delta \varphi_c^T} \Big|_{\varphi_c = \varphi}$$

$$\mathbf{Q}_c = \lim_{M \rightarrow \infty} E \left[\frac{\delta \eta_c}{\delta \varphi_c} \left(\frac{\delta \eta_c}{\delta \varphi_c} \right)^T \right]_{\varphi_c = \varphi}$$

and where φ_c and φ stand for the minimization variables and the true values corresponding to the real-valued channel estimator. The gradient and Hessian with respect to the real parameters can, in turn, be expressed as a function of the corresponding complex derivatives

$$\frac{\delta^2 \eta_c}{\delta \varphi_c \delta \varphi_c^T} \Big|_{\varphi_c = \varphi} = \frac{2}{M} \begin{bmatrix} \text{Re}[\eta_c'' + \bar{\eta}_c''] & -\text{Im}[\eta_c'' - \bar{\eta}_c''] \\ \text{Im}[\eta_c'' + \bar{\eta}_c''] & \text{Re}[\eta_c'' - \bar{\eta}_c''] \end{bmatrix}$$

$$\left[\frac{\delta \eta_c}{\delta \varphi_c} \left(\frac{\delta \eta_c}{\delta \varphi_c} \right)^T \right]_{\varphi_c = \varphi} = \frac{2}{M} \begin{bmatrix} \text{Re}[\eta_c'(\eta_c')^H + \eta_c'(\eta_c')^T] & -\text{Im}[\eta_c'(\eta_c')^H - \eta_c'(\eta_c')^T] \\ \text{Im}[\eta_c'(\eta_c')^H + \eta_c'(\eta_c')^T] & \text{Re}[\eta_c'(\eta_c')^H - \eta_c'(\eta_c')^T] \end{bmatrix} \quad (37)$$

where η_c' , η_c'' , and $\bar{\eta}_c''$ are defined, as shown later, following the description of complex gradient given in [20]. Using conventional formulas for matrix derivatives, it can be shown that (see [16] for details)

$$\frac{1}{M} \eta_c'' = \frac{1}{M} \frac{\delta^2 \eta_c}{\delta \mathbf{h}_c^* \delta \mathbf{h}_c^T} \Big|_{\mathbf{h}_c = \mathbf{h}}$$

$$= \frac{1}{M \sigma^2} \mathcal{T}^H \mathbf{P}_G^\perp \mathcal{T} + \frac{1}{M \sigma^2} \sum_{i=1}^{M_s} \mathcal{S}_i^H \mathbf{P}_G^\perp \mathcal{S}_i + O_p(M^{-1/2})$$

$$\frac{1}{M} \bar{\eta}_c'' = \frac{1}{M} \frac{\delta^2 \eta_c}{\delta \mathbf{h}_c^* \delta \mathbf{h}_c^H} \Big|_{\mathbf{h}_c = \mathbf{h}}$$

$$= O_p(M^{-1/2}) \quad (38)$$

where $O_p(\cdot)$ is the in-probability version of the corresponding deterministic notation. On the other hand, applying the classical formula for the expectation of four Gaussian-distributed random matrices [21]

$$E[\eta_c'(\eta_c')^H] = \frac{1}{\sigma^2} \mathcal{T}^H \mathbf{P}_G^\perp \mathcal{T} + \frac{1}{\sigma^2} \sum_{i,j=1}^{M_s} (\delta_{i-j} + \chi_{i,j}) \mathcal{S}_i^H \mathbf{P}_G^\perp \mathcal{S}_j$$

$$E[\eta_c'(\eta_c')^T] = \mathbf{0}_{L \times L}. \quad (39)$$

where $\chi_{i,j} = \sigma^2 \{(\mathbf{G}^H \mathbf{G})^{-1}\}_{i,j}$. Since the result of the expectation conditioned on \mathbf{s} depends quadratically on this vector, the result is valid, regardless of the unknown symbols statistical model assumption. Plugging (38) and (39) into (36), we obtain the expression for the asymptotic covariance matrix given in Proposition 2.

APPENDIX B PROOF OF PROPOSITION 3

Consistency

To prove strong consistency of the semiblind estimator, we proceed as in the conditional case. First, the SLLN allows us to prove almost sure point-wise convergence of the cost function to the following limiting value.

$$\frac{1}{M} \eta_g(\sigma_g^2, \mathbf{h}_g) \xrightarrow{\text{wp1}} \frac{1}{M} \bar{\eta}_g(\sigma_g^2, \mathbf{h}_g)$$

$$\stackrel{\text{def}}{=} \frac{1}{M} \log \det(\pi \mathbf{C}_{x,g}) + \frac{1}{M} \text{tr}(\mathbf{C}_{x,g}^{-1} \mathbf{C}_x)$$

$$+ \frac{1}{M} (\mathbf{h} - \mathbf{h}_g)^H \mathcal{T}^H \mathbf{C}_{x,g}^{-1} \mathcal{T} (\mathbf{h} - \mathbf{h}_g).$$

Now, to prove that this convergence is uniform in the parameters σ_g^2, \mathbf{h}_g , we must verify the a.s. asymptotic uniform equicontinuity condition, namely

$$\limsup_{M \rightarrow \infty} \sup_{\mathbf{h}'_g, \sigma_g'^2 \in B(\{\mathbf{h}'_g, \sigma_g'^2\}, \delta)} \sup_{\{\mathbf{h}_g, \sigma_g^2\} \in B(\{\mathbf{h}'_g, \sigma_g'^2\}, \delta)} \frac{1}{M} |\eta_g(\sigma_g^2, \mathbf{h}_g) - \eta_g(\sigma_g'^2, \mathbf{h}'_g)| = 0$$

with probability one, as $\delta \rightarrow 0$, where $B(\{\mathbf{h}'_g, \sigma_g'^2\}, \delta)$ is a ball defined in the space generated by the whole set of parameters centered on $\{\mathbf{h}'_g, \sigma_g'^2\}$ with radius δ . The proof is analogous to the one in the CML case and is therefore omitted here.

The limiting cost function $\bar{\eta}_g(\sigma_g^2, \mathbf{h}_g)$ can be expressed as

$$\bar{\eta}_g(\sigma_g^2, \mathbf{h}_g) = [\text{tr}(\mathbf{C}_{x,g}^{-1} \mathbf{C}_x) - \log \det(\mathbf{C}_{x,g}^{-1} \mathbf{C}_x) + MN_c] + (\mathbf{h} - \mathbf{h}_g)^H \mathcal{T}^H \mathbf{C}_{x,g}^{-1} \mathcal{T} (\mathbf{h} - \mathbf{h}_g) + k$$

where k is a constant independent of the parameters. We have implicitly grouped into two terms that are always higher than or equal to zero so that the global minimum is obtained when the two terms null out. The following lemma, which is proven in [22], is the key to understanding the behavior of the first term.

Lemma 4: For positive definite \mathbf{A} of dimension MN_c

$$\text{tr} \mathbf{A} - \log \det \mathbf{A} - MN_c \rightarrow 0$$

implies $\mathbf{A} \rightarrow \mathbf{I}_{MN_c}$.

The first term will thus be zero when

$$\mathbf{C}_x = \mathbf{G} \mathbf{G}^H + \sigma^2 \mathbf{I}_{MN_c} = \mathbf{G}_g \mathbf{G}_g^H + \sigma_g^2 \mathbf{I}_{MN_c} = \mathbf{C}_{x,g}$$

which necessarily implies $\sigma_g^2 = \sigma^2$ and $\mathbf{G}_g = \mathbf{G}$. If the identifiability conditions hold, $\mathbf{h} = e^{j\phi} \mathbf{h}_g$ is the minimum of the first term. When $\phi = 0$, this is also the minimizer of the second term, and we can conclude that this is the global minimum of the cost function.

$$\begin{aligned}
\frac{\delta\eta_g}{\delta\theta_g}\Big|_{\theta_g=\theta} &= \frac{1}{\sqrt{M}} \left[\frac{\delta\eta_g}{\delta\sigma_g^2}\Big|_{\theta_g=\theta} \quad \text{Re}^T[\nabla_g] \quad \text{Im}^T[\nabla_g] \right]^T \\
\frac{\delta^2\eta_g}{\delta\theta_g\delta\theta_g^T}\Big|_{\theta_g=\theta} &= \frac{2}{M} \begin{bmatrix} \frac{1}{2} \frac{\delta^2\eta_g}{(\delta\sigma_g^2)^2}\Big|_{\theta_g=\theta} & \text{Re}^T \left[\frac{\delta\nabla_g}{\delta\sigma^2} \right] & \text{Im}^T \left[\frac{\delta\nabla_g}{\delta\sigma^2} \right] \\ \text{Re} \left[\frac{\delta\nabla_g}{\delta\sigma^2} \right] & \text{Re}[\mathbf{H}_{g,c} + \mathbf{H}_{g,nc}] & -\text{Im}[\mathbf{H}_{g,c} - \mathbf{H}_{g,nc}] \\ \text{Im} \left[\frac{\delta\nabla_g}{\delta\sigma^2} \right] & \text{Im}[\mathbf{H}_{g,c} + \mathbf{H}_{g,nc}] & \text{Re}[\mathbf{H}_{g,c} - \mathbf{H}_{g,nc}] \end{bmatrix} \\
\frac{\delta\eta_g}{\delta\theta_g} \left(\frac{\delta\eta_g}{\delta\theta_g} \right)^T \Big|_{\theta_g=\theta} &= \frac{2}{M} \begin{bmatrix} \frac{1}{2} \left(\frac{\delta\eta_g}{\delta\sigma_g^2} \right)^2 & \text{Re} \left[\frac{\delta\eta_g}{\delta\sigma_g^2} \nabla_g \right]_{\theta_g=\theta}^T & \text{Im} \left[\frac{\delta\eta_g}{\delta\sigma_g^2} \nabla_g \right]_{\theta_g=\theta}^T \\ \text{Re} \left[\frac{\delta\eta_g}{\delta\sigma_g^2} \nabla_g \right]_{\theta_g=\theta} & \text{Re} [\nabla_g \nabla_g^H + \nabla_g \nabla_g^T] & -\text{Im} [\nabla_g \nabla_g^H - \nabla_g \nabla_g^T] \\ \text{Im} \left[\frac{\delta\eta_g}{\delta\sigma_g^2} \nabla_g \right]_{\theta_g=\theta} & \text{Im} [\nabla_g \nabla_g^H + \nabla_g \nabla_g^T] & \text{Re} [\nabla_g \nabla_g^H - \nabla_g \nabla_g^T] \end{bmatrix}
\end{aligned}$$

Asymptotic Normality and Covariance

Asymptotic normality follows from consistency and the asymptotic normality of the gradient vector, which can be expressed as

$$\begin{aligned}
\frac{1}{\sqrt{M}} \frac{\delta\eta_g}{\delta\sigma_g^2}\Big|_{\theta_g=\theta} &= \frac{1}{\sqrt{M}} \text{tr} \left[\mathbf{C}_x^{-1} - \mathbf{C}_x^{-1} \tilde{\mathbf{C}}_x \mathbf{C}_x^{-1} \right] \\
\frac{1}{\sqrt{M}} \nabla_g &\stackrel{\text{def}}{=} \frac{1}{\sqrt{M}} \frac{\delta\eta_g}{\delta\mathbf{h}_g^*}\Big|_{\mathbf{h}_g=\mathbf{h}} \\
&= \frac{1}{\sqrt{M}} \left[\sum_{i=1}^{M_s} \mathcal{S}_i^H (\mathbf{C}_x^{-1} - \mathbf{C}_x^{-1} \tilde{\mathbf{C}}_x \mathbf{C}_x^{-1}) \mathcal{S}_i \right] \mathbf{h} \\
&\quad - \frac{1}{\sqrt{M}} \mathbf{T}^H \mathbf{C}_x^{-1} (\mathbf{x} - \mathbf{T}\mathbf{h}).
\end{aligned}$$

The asymptotic covariance matrix of the vector of real parameters $\hat{\theta}_g$ defined in Proposition 3 can be obtained as

$$\begin{aligned}
\mathbf{C}_{\hat{\theta}_g} &= E \left[(\hat{\theta}_g - \theta) (\hat{\theta}_g - \theta)^H \right] \\
&= \mathbf{H}_g^{-1} \mathbf{Q}_g \mathbf{H}_g^{-1}
\end{aligned} \tag{40}$$

$$\begin{aligned}
\mathbf{H}_g &= \lim_{M \rightarrow \infty} \frac{\delta^2\eta_g}{\delta\theta_g\delta\theta_g^T}\Big|_{\theta_g=\theta} \\
\mathbf{Q}_g &= \lim_{M \rightarrow \infty} E \left[\frac{\delta\eta_g}{\delta\theta_g} \left(\frac{\delta\eta_g}{\delta\theta_g} \right)^T \right]_{\theta_g=\theta}
\end{aligned} \tag{41}$$

where θ stands for the vector of actual values. Depending on the statistical model chosen for the unknown symbols, the results of the expectation in (40) and (41) will be different. In any case, we can express these matrices in terms of complex derivatives as shown in the equations at the top of the page. It is shown in [16]

that the complex derivatives can asymptotically be expressed in compact form as

$$\begin{aligned}
\frac{1}{M} \frac{\delta^2\eta_g}{(\delta\sigma_g^2)^2}\Big|_{\theta_g=\theta} &= c_0^M + O_p(M^{-1/2}) \\
\frac{1}{M} \frac{\delta^2\eta_g}{\delta\mathbf{h}_g^*\delta\mathbf{h}_g^T}\Big|_{\mathbf{h}_g=\mathbf{h}} &\stackrel{\text{def}}{=} \frac{1}{M} \mathbf{H}_{g,c} = \mathbf{E}_2^M + O_p(M^{-1/2}) \\
\frac{1}{M} \frac{\delta^2\eta_g}{\delta\mathbf{h}_g^*\delta\sigma_g^2}\Big|_{\theta_g=\theta} &= \mathbf{e}_1^M + O_p(M^{-1/2}) \\
\frac{1}{M} \frac{\delta^2\eta_g}{\delta\mathbf{h}_g^*\delta\mathbf{h}_g^H}\Big|_{\mathbf{h}_g=\mathbf{h}} &\stackrel{\text{def}}{=} \frac{1}{M} \mathbf{H}_{g,nc} = \mathbf{E}_3^M + O_p(M^{-1/2})
\end{aligned} \tag{42}$$

where c_0^M , \mathbf{e}_1^M , \mathbf{E}_2^M , and \mathbf{E}_3^M are defined in Proposition 3. Before analyzing the covariance of the gradient, we introduce the following lemma.

Lemma 5: Given a random vector $\mathbf{x} = \mathbf{G}\mathbf{s} + \mathbf{n}$ such that \mathbf{s} and \mathbf{n} have the statistical properties established in Section II and a deterministic matrix \mathbf{K} of appropriate dimensions

$$\begin{aligned}
E[\mathbf{x}\mathbf{x}^H \mathbf{K}\mathbf{x}\mathbf{x}^H] &= \mathbf{C}_x \mathbf{K} \mathbf{C}_x + \mathbf{C}_x \text{tr}[\mathbf{K}\mathbf{C}_x] + (\zeta - 2) \mathbf{G} (\mathbf{G}^H \mathbf{K} \mathbf{G} \odot \mathbf{I}_{M_s}) \mathbf{G}^H \\
E[\mathbf{x}\mathbf{x}^H \mathbf{K}\mathbf{x}^* \mathbf{x}^T] &= \mathbf{C}_x \mathbf{K} \mathbf{C}_x^T + \mathbf{C}_x \mathbf{K}^T \mathbf{C}_x^T + (\zeta - 2) \mathbf{G} (\mathbf{G}^H \mathbf{K} \mathbf{G}^* \odot \mathbf{I}_{M_s}) \mathbf{G}^T
\end{aligned}$$

where \odot denotes element-wise product.

Proof: See [16]. ■

We just need to apply Lemma 5 to obtain

$$\begin{aligned}
\frac{1}{2M} E \left[\frac{\delta\eta_g}{\delta\sigma_g^2}\Big|_{\theta_g=\theta}^2 \right] &= c_0^M + (\zeta - 2) f_0^M \\
\frac{1}{M} E [\nabla_g \nabla_g^H] &= \mathbf{E}_2^M + (\zeta - 2) \mathbf{F}_2^M
\end{aligned}$$

$$\frac{1}{M} E \left[\frac{\delta \eta_g}{\delta \sigma_g^2} \bigg|_{\theta_g = \theta} \nabla_g \right] = \mathbf{e}_1^M + (\zeta - 2) \mathbf{f}_1^M$$

$$\frac{1}{M} E [\nabla_g \nabla_g^T] = \mathbf{E}_3^M + (\zeta - 2) \mathbf{F}_3^M$$

where we have used $\text{tr}[\mathbf{A}(\mathbf{A} \odot \mathbf{I})] = \text{tr}[(\mathbf{A} \odot \mathbf{I})^2]$ (which is valid for any square matrix \mathbf{A} of appropriate dimensions), where \mathbf{f}_0^M , \mathbf{f}_1^M , \mathbf{F}_2^M , and \mathbf{F}_3^M are defined in Proposition 3. Plugging this last result and (42) into (40), we obtain the final covariance matrix.

ACKNOWLEDGMENT

The authors would like to thank the associate editor and the anonymous reviewers for their helpful suggestions.

REFERENCES

- [1] 3GPP Techn. Spec. 25.211. (1999, Dec.) Physical channels and mapping of transport channels onto physical channels (FDD). [Online]. Available: www.3gpp.org
- [2] C. R. Rao, *Linear Statistical Inference and Its Applications*. New York: Wiley, 1973.
- [3] A. Gorokhov and Ph. Loubaton, "Semi-blind second order identification of convolutive channels," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Munich, Germany, Apr. 1997, pp. 3905–3908.
- [4] J. Ayadi, E. de Carvalho, and D. T. M. Slock, "Blind and semi-blind maximum likelihood methods for FIR multichannel identification," in *Proc. IEEE ICASSP*, Seattle, WA, May 1998.
- [5] T. Söderström and P. Stoica, *System Identification*. London, U.K.: Prentice-Hall, 1989.
- [6] F. Gustafsson, "Blind equalization by direct examination of the input sequences," *IEEE Trans. Signal Processing*, vol. 43, pp. 2213–2222, July 1995.
- [7] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, pp. 416–429, Mar. 2000.
- [8] E. de Carvalho and D. Slock, "Cramér–Rao bounds for semi-blind, blind a training sequence-based channel estimation," in *Proc. IEEE SP Workshop SPAWC*, Paris, France, Apr. 1997, pp. 129–132.
- [9] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. New York: Springer-Verlag, 1991.
- [10] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1783–1795, Oct. 1990.
- [11] E. de Carvalho and D. T. M. Slock, "Asymptotic performance of ML methods for semi-blind channel estimation," in *Proc. IEEE Asilomar Conf.*, 1998.
- [12] J. L. Bapat, "Partially blind estimation: ML-based approaches and Cramér–Rao bound," *Signal Process.*, vol. 71, pp. 265–277, 1998.
- [13] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1553–1560, Oct. 1988.

- [14] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing*, vol. 44, pp. 661–672, Mar. 1996.
- [15] E. Pité and P. Duhamel, "Bilinear methods for blind channel equalization: (No) local minimum issue," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Seattle, WA, 1998.
- [16] X. Mestre. (2001, June) Asymptotic covariance matrices for semi-blind ML channel estimation approaches in WCDMA contexts. [Online]. Available: <http://gps-tsc.upc.es/comm/Xavi/publications.html>
- [17] W. F. Stout, *Almost Sure Convergence*. New York: Academic, 1974.
- [18] B. M. Pötscher and I. R. Prucha, "Generic uniform convergence and equicontinuity concepts for random functions," *J. Econometr.*, vol. 60, pp. 23–63, 1994.
- [19] E. L. Lehmann, *Theory of Point Estimation*. New York: Chapman & Hall, 1991.
- [20] B. A. D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Proc. Inst. Elect. Eng. F*, vol. 130, pp. 11–16, Feb. 1983.
- [21] P. H. M. Janssen and P. Stoica, "On the expectation of the product of four matrix-valued Gaussian random variables," *IEEE Trans. Automat. Contr.*, vol. 33, pp. 867–870, Sept. 1988.
- [22] T. W. Anderson, "Multivariate linear relations," in *Proc. Second Int. Tampere Conf. Statist.*, Tampere, Finland, 1987, pp. 9–36.



Xavier Mestre (S'96) was born in Barcelona, Catalonia, Spain, in 1974. He received the degree in electrical engineering from the Universitat Politècnica de Catalunya (UPC), Barcelona, in 1998.

The same year, he joined the Department of Signal Theory and Communications at UPC, where he has since been working as a research assistant.



Javier R. Fonollosa (S'90–M'92–SM'98) received the Ph.D. degree in electrical and computer engineering from Northeastern University, Boston, MA, in 1992.

In 1989, he held a visiting appointment at the European Space Agency Centre in The Netherlands. From the Fall of 1989 until 1992, he was Research Assistant at Northeastern University and the University of Southern California, Los Angeles, where, upon completion of the Ph.D. degree, he became Research Associate. In 1993, he joined the Department of Signal

Theory and Communications, Universitat Politècnica de Catalunya, where he became Associate Professor in November 1996.

Dr. Fonollosa has been a member of the IEEE Signal Processing for Communications (SPCOM) Technical Committee of the IEEE Signal Processing Society since 1998.