

Channel Estimation for MIMO-OFDM Systems by Modal Analysis/Filtering

Marcello Cicerone, Osvaldo Simeone, *Member, IEEE*, and Umberto Spagnolini, *Senior Member, IEEE*

Abstract—In this paper, we investigate the benefits of exploiting the *a priori* information about the structure of the multipath channel on the performance of channel estimation for multiple-input multiple-output–orthogonal frequency-division multiplexing (MIMO-OFDM) systems. We first approach this problem from the point of view of estimation theory by computing a lower bound on the estimation error and studying its properties. Then, based on the insight obtained from the analysis, efficient channel estimators are designed that perform close to the derived limit. The proposed channel estimators compute the long-term features of the multipath channel model through a subspace tracking algorithm by identifying the invariant (over multiple OFDM symbols) space/time modes of the channel (modal analysis). On the other hand, the fast-varying fading amplitudes are tracked by using least-squares techniques that exploit temporal correlation of the fading process (modal filtering). The analytic treatment is complemented by thorough numerical investigation in order to validate the performance of the proposed techniques. MIMO-OFDM with bit-interleaved coded modulation and MIMO-turbo equalization is selected as a benchmark for performance evaluation in terms of bit-error rate.

Index Terms—Adaptive estimation, communication systems performance, fading channels, frequency-division multiplexing, multiple-input multiple-output (MIMO) systems.

I. INTRODUCTION

THE COMBINATION of orthogonal frequency-division multiplexing (OFDM) and multiple-input multiple-output (MIMO) technologies, referred to as MIMO-OFDM, is currently under study as one of the most promising candidate for next-generation communications systems, ranging from wireless LAN to broadband access. Recent works tackled the performance assessment (both through simulation and measurements) of MIMO-OFDM systems in the presence of practical impairments, such as synchronization and channel-estimation errors [1], [2]. As shown by these references, channel estimation is a critical issue for MIMO-OFDM systems, especially if multilevel modulation is employed in order to achieve high spectral efficiencies.

Paper approved by T. F. Wong, the Editor for Wideband and Multiple Access Wireless Systems of the IEEE Communications Society. Manuscript received May 17, 2004; revised April 11, 2005. This paper was presented in part at the ITG Workshop on Smart Antennas (WSA), Munich, Germany, March 2004.

M. Cicerone was with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, I-20133 Milano, Italy. He is now with Analysys, I-20122 Milano, Italy (e-mail: marcello.cicerone@analysys.com).

U. Spagnolini is with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, I-20133 Milano, Italy (e-mail: spagnoli@elet.polimi.it).

O. Simeone was with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, I-20133 Milano, Italy. He is now with the Center for Communications and Signal Processing Research, New Jersey Institute of Technology, Newark, NJ 07102 USA (e-mail: osvaldo.simeone@njit.edu).

Digital Object Identifier 10.1109/TCOMM.2006.884849

The scope of this paper is limited to pilot-based channel estimation, whereby channel estimation is performed by exploiting the transmission of training sequences. Semiblind approaches that also leverage on the received data symbols could be devised in order to improve the estimation performance [3]. So far, research on pilot-based channel estimation for MIMO-OFDM system has mainly focused on least squares (LS) channel estimation. The problem of designing an optimal training sequence for the LS estimator has been investigated in [4] and [5]. Conventional LS channel estimation does not assume any *a priori* information about the channel vector to be estimated, except for the finite length of the channel impulse response (CIR) in the time domain that is inherent in the choice of an OFDM modulation [4]–[6]. This assumption simplifies the channel-estimation process, but leaves room for improvement. For instance, modifications of the LS channel estimator that take into account additional information about the channel (i.e., equal power-delay profile for different transmitting antennas) have been proposed in [7], and proved to outperform the LS method.

The work presented in this paper is motivated by the following consideration. It is widely acknowledged that the propagation in a wireless scenario can be modeled by a multipath channel, where each path has some features related to the large-scale geometry of the environment (such as propagation delay, angles of departure and arrival, power-delay profile), and some others related to small-scale (i.e., on the order of a fraction of the wavelength) variations, namely the fading amplitude [8], [9]. Accordingly, the first class of features vary over a much larger time scale than the coherence time of the fading process. Our purpose is twofold. First, we try to assess analytically to what extent LS channel estimation can be improved if knowledge about the multipath channel structure is taken into account in deriving the channel estimator. Toward this goal, the lower bound on the channel-estimation error proposed in [10] for MIMO time-domain transmission is extended to multicarrier systems, and to a more realistic multipath scenario where each path may have a different Doppler spectrum (see, e.g., the standard in [32]). Second, based on the insight obtained from the aforementioned analysis, practical channel estimators are designed that perform close to the performance limit set by the lower bound with affordable computational complexity, even in presence of channel modeling mismatches.

As stated above, the derivation of the lower bound on the channel estimation in this paper extends the treatment presented in [10]. Both this paper and [10] consider asymptotic results for an observation window large enough. The theoretical foundation of these results is formally provided in [11], where it is proved, under a specific setting, that the lower bound considered herein coincides asymptotically with the hybrid Cramer–Rao

bound (CRB). The latter is a modification of the classical CRB for the case where the parameters to be estimated are a function of both deterministic (long-term channel features) and random variables (fading amplitudes).

The proposed channel estimators perform a separate estimation of the long-term and fast-varying features of the channel. In particular, the long-term features are computed through a subspace tracking algorithm by identifying the invariant (over multiple OFDM symbols) space/time modes of the channel (modal analysis). On the other hand, the fast-varying fading amplitudes are possibly tracked by using LS techniques that exploit temporal (i.e., over multiple OFDM symbols) correlation of the fading process. In case fading amplitudes are estimated separately on each OFDM symbol, the channel estimators reduce to a projection of the preliminary LS channel estimate over the space/time modes (modal filtering).

The analytic presentation is complemented by thorough numerical simulations. The proposed channel estimators are further tested for a MIMO-OFDM system with bit-interleaved coded modulation (BICM) [12] and MIMO-turbo equalization [13], in order to validate their performance for a system that has been recognized as a promising candidate for next-generation wireless systems [14].

The outline of the paper is as follows. Section II introduces the framework and state of the art (LS channel estimation) for training-based channel estimation in MIMO-OFDM systems. A simplified analytical derivation of the lower bound on the channel-estimation error is then derived in Section III. Based on the insight obtained from the analysis in the previous section, practical channel-estimation algorithms are proposed in Section IV, based on modal filtering/analysis. Finally, the proposed algorithms are validated for a promising MIMO-OFDM transceiver based on BICM and MIMO-turbo equalization in Section V.

Notation: In the paper, vectors or matrices are in bold; uppercase is used for frequency-domain quantities to distinguish them from the corresponding time-domain quantities, which are lowercase; $(\cdot)^T$ denotes transposition; $(\cdot)^H$ denotes Hermitian transposition; $(\cdot)^\dagger$ denotes matrix pseudoinverse; $\text{tr}[\cdot]$ represents trace; $E[\cdot]$ represents expectation; \otimes is the Kronecker matrix product; $\text{vec}\{\cdot\}$ is the stacking operator, and the following property is extensively used: $\text{vec}\{\mathbf{ABC}\} = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}\{\mathbf{B}\}$; $\text{vec}_M^{-1}\{\cdot\}$ is the inverse stacking operator defined so that $\text{vec}_M^{-1}\{\mathbf{V}\} = \mathbf{v}$ is equivalent to $\text{vec}\{\mathbf{V}\} = \mathbf{v}$, where \mathbf{v} is a $KM \times 1$ vector and \mathbf{V} is a $M \times K$ matrix. \mathbf{I}_P is the $P \times P$ identity matrix; $\mathbf{R}^{1/2}$ is a factorization of \mathbf{R} such that $\mathbf{R}^{H/2}\mathbf{R}^{1/2} = \mathbf{R}$; if \mathbf{X} is a given $M \times 1$ vector, $\text{diag}(\mathbf{X})$ is the $M \times M$ diagonal matrix with the entries of \mathbf{X} on the main diagonal; the i th entry of a given vector \mathbf{X} is denoted by $[X]_i$; $[\mathbf{X}]_{i,j}$ denotes the (i, j) th entry of the matrix \mathbf{X} .

II. MODELS AND PRELIMINARIES

A. System Model

The block diagram of a MIMO-OFDM system is depicted in Fig. 1. The $K \times 1$ vector $\mathbf{X}_\ell^{n_T} = [X_\ell^{n_T}[0] \dots X_\ell^{n_T}[K-1]]^T$ ($n_T = 1, \dots, N_T$) contains the complex symbols to be transmitted on the K subcarriers during the ℓ th OFDM symbol from

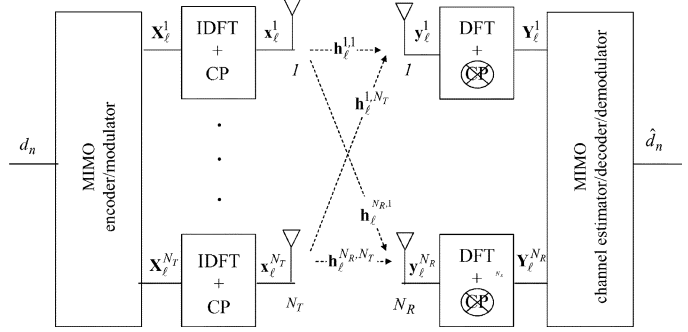


Fig. 1. Block diagram of a MIMO-OFDM system.

antenna n_T . The MIMO encoder/modulator (e.g., space-time block or trellis coding, spatial multiplexing, see [15] for an overview) determines how vectors $\mathbf{X}_\ell^{n_T}$ are generated from the input bit stream d_n and the pilot symbols (used for channel-estimation purposes). A K -point inverse discrete Fourier transform (IDFT) is performed on $\mathbf{X}_\ell^{n_T}$, and the cyclic prefix (CP) of length W_{CP} samples is appended in order to obtain the $(W_{\text{CP}} + K) \times 1$ vector $\mathbf{x}_\ell^{n_T}$ that, after parallel-to-serial conversion, is transmitted by the n_T th transmitting antenna.

The signal received by the n_R th receiving antenna ($n_R = 1, \dots, N_R$) $\mathbf{y}_\ell^{n_R}$ is the sum of the signals from the N_T transmitting antennas. The contribution of the n_T th transmitting antenna is given by the convolution of the transmitted signal $\mathbf{x}_\ell^{n_T}$, with the W -taps discrete-time CIR $\mathbf{h}_\ell^{n_R, n_T} = [h_\ell^{n_R, n_T}[0] \dots h_\ell^{n_R, n_T}[W-1]]^T$, modeling the propagation between the n_T th transmitting antenna and the n_R th receiving antenna (further discussion in the next section). Notice that the channel is assumed to be constant within one OFDM symbol, and varying over different OFDM symbols. After removal of the CP from the $(K + W_{\text{CP}}) \times 1$ vector $\mathbf{y}_\ell^{n_R}$, a K -point DFT is performed, and the $K \times 1$ frequency-domain vector $\mathbf{Y}_\ell^{n_R} = [Y_\ell^{n_R}[0] \dots Y_\ell^{n_R}[K-1]]^T$ is obtained. If the temporal support of $\mathbf{h}_\ell^{n_R, n_T}$ is $W \leq W_{\text{CP}} + 1$ for each n_R and n_T , the received signal by the n_R th receiving antenna at the k th subcarrier on the ℓ th OFDM symbol can be written as

$$Y_\ell^{n_R}[k] = \sum_{n_T=1}^{N_T} X_\ell^{n_T}[k] H_\ell^{n_R, n_T}[k] + N_\ell^{n_R}[k] \quad (1)$$

where the channel gains over the K subcarriers for the (n_R, n_T) link, gathered in the $K \times 1$ vector $\mathbf{H}_\ell^{n_R, n_T}$, are obtained as the K -point DFT of $\mathbf{h}_\ell^{n_R, n_T}$

$$\mathbf{H}_\ell^{n_R, n_T} = \mathbf{F} \mathbf{h}_\ell^{n_R, n_T}. \quad (2)$$

\mathbf{F} is the unitary $K \times W$ DFT matrix with elements $[\mathbf{F}]_{k,w} = 1/\sqrt{K} \exp(-j(2\pi/K)kw)$, for $k = 0, \dots, K-1$, $w = 0, \dots, W-1$. Furthermore, the $K \times 1$ vector $\mathbf{N}_\ell^{n_R}$ is the additive noise at the n_R th receiving antenna, with $E[\mathbf{N}_\ell^{n_R} \mathbf{N}_\ell^{n_R H}] = \sigma_n^2 \mathbf{I}_K$. The channel $\mathbf{H}_\ell^{n_R, n_T}$ and the noise $\mathbf{N}_\ell^{n_R}$ are assumed to be statistically independent.

The channel coefficients in the frequency domain $H_\ell^{n_R, n_T}[k]$ (or time domain $h_\ell^{n_R, n_T}[w]$) depend on four independent vari-

TABLE I
CHANNEL MATRICES AND VECTORS

Symbol	Size	Definition
$\mathbf{H}_\ell[k]$ $k = 0, \dots, K-1$	$N_R \times N_T$	$[\mathbf{H}_\ell[k]]_{n_R, n_T} = H_\ell^{n_R, n_T}[k]$
$\check{\mathbf{h}}_\ell$	$N_R N_T \times W$	$[\check{\mathbf{h}}_\ell]_{n_R + (n_T-1)N_T} = h_\ell^{n_R, n_T}[w]$
\mathbf{h}_ℓ	$N_R N_T W \times 1$	$[\mathbf{h}_\ell]_{n_R + (n_T-1)N_R + w N_R N_T} = h_\ell^{n_R, n_T}[w]$
$\mathbf{h}_\ell^{n_R}$ $n_R = 1, \dots, N_R$	$N_T W \times 1$	$[\mathbf{h}_\ell^{n_R}]_{n_T + w N_T} = h_\ell^{n_R, n_T}[w]$

ables, namely transmitting/receiving antennas (n_T and n_R), frequency bin k (or delay-time bin w) and OFDM symbol index ℓ . Throughout the paper, we will be elaborating on the signal model (1) in order to cast it in a suitable form for different computations. In doing so, the channel coefficients will be rearranged in matrices of different sizes, whose definitions are listed for reference in Table I.

For instance, it is useful to write the received signal so as to separate the measurements on different subcarriers $\mathbf{Y}_\ell[k] = [Y_\ell^1[k] \dots Y_\ell^{N_R}[k]]^T$:

$$\mathbf{Y}_\ell[k] = \mathbf{H}_\ell[k] \mathbf{X}_\ell[k] + \mathbf{N}_\ell[k] \quad (3)$$

where $\mathbf{X}_\ell[k] = [X_\ell^1[k] \dots X_\ell^{N_T}[k]]^T$ is an $N_T \times 1$ vector collecting the symbols transmitted simultaneously on the k th subcarrier by the N_T antennas, and $\mathbf{H}_\ell[k]$ is the channel matrix at the k th subcarrier, as defined in Table I. $\mathbf{N}_\ell[k]$ is the $N_R \times 1$ additive noise vector $\mathbf{N}_\ell[k] = [N_\ell^1[k] \dots N_\ell^{N_R}[k]]^T$ with spatial correlation \mathbf{R}_n , $E[\mathbf{N}_\ell[k] \mathbf{N}_\ell[k]^H] = \sigma_n^2 \mathbf{R}_n$. To simplify, noise will be assumed to be spatially white ($\mathbf{R}_n = \mathbf{I}_{N_R}$), unless stated otherwise.

B. Channel Model

The time-varying fading channel modeling the propagation on the (n_R, n_T) th link is characterized by a (continuous-time) CIR $h_\ell^{n_R, n_T}(\tau)$ given by the combination of D paths (let us consider the ℓ th OFDM symbol)

$$h_\ell^{n_R, n_T}(\tau) = \sum_{d=1}^D \sqrt{\Omega_{d,\ell}} [\mathbf{A}_{d,\ell}]_{n_R, n_T} g(\tau - \tau_{d,\ell}) \quad (4)$$

where for the d th path, $\Omega_{d,\ell}$ is the corresponding power, $\tau_{d,\ell}$ the delay, $g(\tau)$ is the equivalent transmission filter, $g(\tau) = g_T(\tau) * g_R(\tau)$, with $g_T(\tau)$ and $g_R(\tau)$ denoting, respectively, the transmitter's pulse-shaping function and the receiver's matched-filter impulse response. Moreover, $[\mathbf{A}_{d,\ell}]_{n_R, n_T}$ represents the complex gain relative to the (n_R, n_T) link for the d th path, incorporating fading effects and array responses at both sides, as will be specified below. According to our notation, $\mathbf{A}_{d,\ell}$ is an $N_R \times N_T$ matrix that contains the ensemble of gains for the d th path relative to the $N_R \times N_T$ links.

Let us now sample at a rate $1/T$ ($T = 1/B$, where B is the transmission bandwidth) the CIRs on the $N_R N_T$ links. The $W \times 1$ discrete-time CIR $\mathbf{h}_\ell^{n_R, n_T}$ introduced in the previous section is then obtained as $h_\ell^{n_R, n_T}[w] = h_\ell^{n_R, n_T}((w-1)T)$ for $w = 1, \dots, W$. From (4), it can be shown that the channel matrix as defined in Table I reads (see Appendix A)

$$\check{\mathbf{h}}_\ell = \mathbf{A}_\ell \mathbf{\Omega}_\ell^{1/2} \mathbf{G}(\boldsymbol{\tau}_\ell)^T \quad (5)$$

where $\mathbf{A}_\ell = [\text{vec}\{\mathbf{A}_{1,\ell}\} \dots \text{vec}\{\mathbf{A}_{D,\ell}\}]$ is $N_R N_T \times D$, $\mathbf{\Omega}_\ell^{1/2} = \text{diag}\{\sqrt{\Omega_{1,\ell}} \dots \sqrt{\Omega_{D,\ell}}\}$, and $\mathbf{G}(\boldsymbol{\tau}_\ell) = [\mathbf{g}(\tau_{1,\ell}) \dots \mathbf{g}(\tau_{D,\ell})]$ is a $W \times D$ matrix collecting the the T -spaced samples of the delayed waveform $g(\tau - \tau_{d,\ell})$ ($\boldsymbol{\tau}_\ell = [\tau_{1,\ell} \dots \tau_{D,\ell}]^T$).

In this paper, antennas at both the transmitting and receiving sides are assumed to be closely spaced to each other, so that we can apply the far-field approximation (i.e., assume locally plane wave). In this case, if the scattering environment is such that each path (composed of many subpaths) has an angular spread at the transmitting and receiving sides that is smaller than the resolution of the corresponding arrays, each path can be characterized by a direction of departure (DOD) $\alpha_d^{(T)}$ and a direction of arrival (DOA) $\alpha_d^{(R)}$. Furthermore, it has a propagation delay $\tau_{d,\ell}$ and a complex fading amplitude $\beta_{d,\ell}$. This model is widely used for outdoor scenarios, and has been adopted in standard models (see, e.g., [32]). Therefore, the set of paths is characterized within the ℓ th OFDM symbol by sets of DOAs $\boldsymbol{\alpha}_\ell^{(R)} = [\alpha_{1,\ell}^{(R)} \dots \alpha_{D,\ell}^{(R)}]$, DODs $\boldsymbol{\alpha}_\ell^{(T)} = [\alpha_{1,\ell}^{(T)} \dots \alpha_{D,\ell}^{(T)}]$, delays $\boldsymbol{\tau}_\ell$, and fading amplitudes $\boldsymbol{\beta}_\ell = [\beta_{1,\ell} \dots \beta_{D,\ell}]^T$. Defining the array response vector at the receiver (and transmitter side) as $\mathbf{a}^{(R)}(\alpha_{d,\ell}^{(R)})$ (and $\mathbf{a}^{(T)}(\alpha_{d,\ell}^{(T)})$) that depends on the DOA (and DOD), we get that the $N_R \times N_T$ matrices $\mathbf{A}_{d,\ell}$ in (28) can be expressed as $\mathbf{A}_{d,\ell} = \beta_{d,\ell} \mathbf{a}^{(R)}(\alpha_{d,\ell}^{(R)}) \mathbf{a}^{(T)}(\alpha_{d,\ell}^{(T)})^T$ [10]. The fading is uncorrelated on different paths ($E[\boldsymbol{\beta}_\ell \boldsymbol{\beta}_\ell^H] = \mathbf{I}_D$), but it is possibly temporally correlated, as detailed below.

1) *Long-Term Features of the Channel*: So far, we have denoted the time variability of the channel parameters by using the subscript ℓ . However, the quantities in (5) are known to vary over different time scales. In fact, the geometry and the characteristics of the scatterers are supposed to vary slowly, as compared with the coherence time of the fading processes. As a consequence, delays $\boldsymbol{\tau}_\ell$, DOAs $\boldsymbol{\alpha}_\ell^{(R)}$, DODs $\boldsymbol{\alpha}_\ell^{(T)}$, and powers $\mathbf{\Omega}_\ell$ can be considered as constant over multiple OFDM symbols, thus losing (as a first-order approximation) their dependence on the subscript ℓ over the interval of stationarity. On the other hand, the vector $\boldsymbol{\beta}_\ell$ accounts for the fast-fading variations of the path amplitudes and, as such, is variable on a much smaller time scale, typically on the order of the time it takes for a mobile to travel a fraction of the wavelength.

According to this assumption, the time-varying model in (5) becomes

$$\check{\mathbf{h}}_\ell = \mathbf{A}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)}) \cdot \text{diag}(\boldsymbol{\beta}_\ell) \cdot \mathbf{\Omega}^{1/2} \mathbf{G}(\boldsymbol{\tau})^T \quad (6)$$

where \mathbf{A}' has now been factorized into two parts, $\mathbf{A}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)}) = [\mathbf{a}^{(T)}(\alpha_1^{(T)}) \otimes \mathbf{a}^{(R)}(\alpha_1^{(R)}), \dots, \mathbf{a}^{(T)}(\alpha_D^{(T)}) \otimes \mathbf{a}^{(R)}(\alpha_D^{(R)})]$ and $\text{diag}(\boldsymbol{\beta}_\ell)$, separating the terms that depend on the (slow-varying) angles $\boldsymbol{\alpha}^{(T)}$, $\boldsymbol{\alpha}^{(R)}$ from the (fast-varying) fading coefficients $\boldsymbol{\beta}_\ell$.

To clearly isolate the slow-varying terms from the fast-varying fading amplitudes, we stack the elements of the MIMO channel impulse response into the $N_R N_T W \times 1$ vector $\mathbf{h}_\ell = \text{vec}\{\check{\mathbf{h}}_\ell\}$ (see Table I), obtaining

$$\mathbf{h}_\ell = \mathbf{T}(\boldsymbol{\tau}, \boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)}, \mathbf{\Omega}^{1/2}) \cdot \boldsymbol{\beta}_\ell. \quad (7)$$

The $N_R N_T W \times D$ matrix \mathbf{T} (we drop the functional dependence for simplicity of notation) collects the space–time signatures of the D paths, its d th column being given by $\Omega_d \mathbf{g}(\tau_d) \otimes \mathbf{a}^{(T)}(\alpha_d^{(T)}) \otimes \mathbf{a}^{(R)}(\alpha_d^{(R)})$.

Notice that the assumption of constant angles/delays/powers over multiple OFDM symbols will be relaxed when deriving the channel estimator (see Section IV-A).

2) *Fast-Varying Fading Amplitudes (Doppler Spectrum)*: The $D \times 1$ vector $\boldsymbol{\beta}_\ell$ in (7) models time-varying uncorrelated Rayleigh fading amplitudes. More specifically, $\boldsymbol{\beta}_\ell$ is a circularly symmetric Gaussian process with correlation matrix $\mathbf{R}_\beta(n) = E[\boldsymbol{\beta}_\ell \boldsymbol{\beta}_\ell^H] = \boldsymbol{\Phi}(n) = \text{diag}([\varphi_1(n) \cdots \varphi_D(n)])$, where $\varphi_d(n) = E[\beta_{d,\ell} \beta_{d,\ell}^*]$ is the temporal correlation of the fading coefficient relative to the d th path. The Doppler spectrum of each path, $S_{\varphi,d}(\omega)$, is obtained by taking the Fourier transform of the corresponding correlation function, $S_{\varphi,d}(\omega) = \mathcal{F}\{\varphi_d(n)\}$. Notice that different paths may have different Doppler spectra. For instance, in a typical scenario, a path with a smaller delay corresponds to a narrower Doppler spectrum [34]. Moreover, in the standard model [32], $S_{\varphi,d}(\omega)$ is a delta function centered on the Doppler frequency of the specific path.

3) *Channel Normalization*: Considering the powers $\{\Omega_d\}$ normalized to have $\sum_{d=1}^D \Omega_d = 1$, it can be easily shown that

$$E \left[|H_\ell^{n_R, n_T}[k]|^2 \right] = \frac{1}{K} \quad \forall k, \forall \ell. \quad (8)$$

As a consequence, it is convenient to define the signal-to-noise ratio (SNR) per frequency bin as

$$\text{SNR} = \frac{\frac{1}{K} \cdot E \left[\|\mathbf{X}_\ell[k]\|^2 \right]}{\sigma_n^2} = \frac{1}{K} \frac{P}{\sigma_n^2} \quad (9)$$

where $P = E[\|\mathbf{X}_\ell[k]\|^2]$ is the total power transmitted on a given subcarrier by all the N_T transmitting antennas, assumed to be independent on k (i.e., for nonadaptive power allocation).

C. Unconstrained (or LS) MIMO Channel Estimation

In training-based OFDM systems, channel estimation is based on the transmission of constellation points known at the receiver on predetermined subcarriers (referred to as pilot subcarriers), selected according to some periodic pattern in the time–frequency grid. OFDM symbols containing pilot subcarriers will be referred to here as training OFDM symbols. Denoting by K_p the number of pilot subcarriers per training OFDM symbol, divided by the number of transmitting antennas, condition $W N_T \leq K_p N_T (\leq K)$ is necessary to get an estimate immune to aliasing for an estimator¹ [4], [5].

Conventional unconstrained (or LS) channel estimation is carried out independently for each element (n_R) of the array at receiver side, and for each training OFDM symbol. Recalling (2), it is easy to show that the signal received from the n_R th

¹Notice that if the channel can be considered constant over several OFDM symbols, the K_p pilot subcarriers can be distributed in different OFDM symbols (we will not consider this case here).

receiving antenna over the pilot subcarriers, collected in the $K_p N_T \times 1$ vector $\tilde{\mathbf{Y}}_\ell^{n_R}$ can be stated as (see also [4])

$$\tilde{\mathbf{Y}}_\ell^{n_R} = \mathbf{B} \mathbf{h}_\ell^{n_R} + \tilde{\mathbf{N}}_\ell^{n_R} \quad (10)$$

where $\tilde{\mathbf{N}}_\ell^{n_R}$ is the $K_p N_T \times 1$ of additive noise; $\mathbf{B} = \text{diag}([\mathbf{X}[1]^T \cdots \mathbf{X}[K_p N_T]^T])(\tilde{\mathbf{F}} \otimes \mathbf{I}_{N_T})$ is a $K_p N_T \times W N_T$ matrix, that contains the $N_T \times 1$ vectors $\mathbf{X}[i]$ collecting the pilot symbols transmitted on the i th pilot subcarrier by the N_T transmitting antenna; $\tilde{\mathbf{F}}$ is composed by the $K_p N_T$ rows of \mathbf{F} corresponding to the pilot subcarriers; and $\mathbf{h}_\ell^{n_R}$ is the $N_T W \times 1$ vector defined in Table I.

Since we are interested in channel estimation, throughout the rest of the paper, index ℓ will be intended to run over the training OFDM symbols. The LS estimate of $\mathbf{h}_\ell^{n_R}$ is

$$\mathbf{h}_{LS,\ell}^{n_R} = \mathbf{B}^\dagger \tilde{\mathbf{Y}}_\ell^{n_R}. \quad (11)$$

The LS is known to be unbiased, and its variance can be minimized by appropriate design of the training sequences across the transmitting antennas and pilot subcarriers in order to guarantee the optimality condition $\mathbf{R}_x = \mathbf{B}^H \mathbf{B} = (K_p P / K) \mathbf{I}_{N_T W}$. This result can be obtained by letting the N_T transmitting antennas share the $K_p N_T$ pilot subcarriers with equipowered, equispaced, and phase-shift orthogonal sequences [4], [5].

In the following, we will consider optimally designed training sequences.

III. LOWER BOUND ON CHANNEL-ESTIMATION ERROR FOR RAYLEIGH MIMO CHANNELS

As discussed in Section II-B, the channel vector (7) can be parametrized so as to separate the slow-varying channel parameters (collected in matrix \mathbf{T}) from the fast-varying fading amplitudes $\boldsymbol{\beta}_\ell$. The conventional LS channel estimator, recalled in Section II-C, does not take into any account the structure of the channel vector (7). In this section, we try to assess to what extent channel-estimation accuracy can be improved if the knowledge of the structure of the channel vector is exploited in designing a channel-estimation algorithm.

Toward this goal, following the treatment in [10], here we derive a lower bound on the mean-square error (MSE) $\text{MSE} = E[\|\hat{\mathbf{h}} - \mathbf{h}\|^2]$ (recall that the average $E[\cdot]$ is taken with respect to both noise and fading amplitudes) of any unbiased estimator $\hat{\mathbf{h}}$ by considering the performance of a channel-estimation algorithm that fully exploits the knowledge about the structure of the multipath channel. In particular, we consider an estimator that:

- performs a consistent estimate of the long-term features of the channel;
- tracks the fading amplitudes according to the minimum MSE (MMSE) criterion.

Reference [11] formally proves that such an estimator achieves the hybrid CRB for an observation window large enough. This result can be easily justified, by noticing that once the long-term matrix \mathbf{T} is known (which is increasingly true for a consistent estimator as the number of observed OFDM symbols increases), the received signal (10) is linear in the unknown fading amplitudes.

In order to compute the performance of an estimator characterized as discussed above, and therefore, obtain the lower bound, a further observation has to be made on the model (7). The space–time signatures of different paths (i.e., the columns of \mathbf{T}) are generally linearly dependent, or equivalently, $r = \text{rank}(\mathbf{T}) \leq \min(D, N_R N_T W)$. Condition $r < \min(D, N_R N_T W)$ holds whenever different paths have at least one parameter among angles (DOA and DOD) and delay that is not resolvable at the receiver (given the resolution of the arrays and the bandwidth of the system). As a consequence, matrix \mathbf{T} is not directly identifiable. Appendix B shows how to overcome this problem in the computation of the lower bound, and the result is discussed below. As a final remark, notice that while parametric channel estimators (see, e.g., [8]) suffer from this lack of identifiability, the modal analysis/filtering estimators discussed in Section IV take advantage of this property by reducing the number of parameters to be estimated (see also [29] for previous work on the subject).

The lower bound reads (the frequency $f = (\omega/2\pi)$ is, by definition, normalized with respect to the time interval between two successive training OFDM symbols NT_{ofdm} , where $T_{\text{ofdm}} = T(K + W_{\text{CP}})$ is the duration of an OFDM symbol) (see Appendix B)

$$\begin{aligned} \text{MSE}_{\hat{h}} = \text{tr} & \left[\int_{-\pi}^{\pi} \mathbf{U} \left(\mathbf{S}_{dd}(\omega) - \mathbf{S}_{dd}(\omega) \mathbf{U}^H \check{\mathbf{X}}^H \right. \right. \\ & \times \left(\check{\mathbf{X}} \mathbf{U} \mathbf{S}_{dd}(\omega) \mathbf{U}^H \check{\mathbf{X}}^H + \mathbf{R}_n \otimes \mathbf{I}_{K_p N_T} \right)^{-1} \\ & \left. \left. \times \check{\mathbf{X}} \mathbf{U} \mathbf{S}_{dd}(\omega)^H \right) \mathbf{U}^H \frac{d\omega}{2\pi} \right] \end{aligned} \quad (12)$$

where \mathbf{U} is the $N_R N_T W \times r$ unitary basis of the subspace spanned by the columns of matrix \mathbf{T} that can be obtained from the r left eigenvectors of the singular value decomposition (SVD) of $\mathbf{T} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$ (notice that only nonzero singular values are considered in the SVD). Moreover, it is

$$\mathbf{S}_{dd}(\omega) = \mathbf{\Sigma} \mathbf{V}^H \mathbf{S}_{\Phi}(\omega) \mathbf{V} \mathbf{\Sigma} \quad (13)$$

the diagonal matrix $\mathbf{S}_{\Phi}(\omega) = \text{diag}[S_{\varphi,1}(\omega) \cdots S_{\varphi,D}(\omega)]$ collecting the power spectral densities of the fading variations for the D paths $S_{\varphi,d}(\omega)$. Finally, the $K_p N_T N_R \times W N_T N_R$ matrix $\check{\mathbf{X}}$ is defined as $\check{\mathbf{X}} = \mathbf{I}_{N_R} \otimes \mathbf{B}$.

The r columns of matrix \mathbf{U} define the space–time modes of the channel. This definition stems from the fact that the subspace spanned by the columns of \mathbf{T} is a stationary feature of the channel vector, varying at the same rate as \mathbf{T} .

A. Analytical Simplifications of the Bound (12)

The bound (12) needs to be evaluated numerically, given the Doppler spectra of different paths. However, further analytical insight of (12) can be obtained by dividing the integration range $\mathcal{I} = (-\pi, \pi]$ into J nonoverlapping subbands \mathcal{I}_j of support $0 < \Delta\omega_j \leq 2\pi$, so that $\mathcal{I} = \cup_{j=1}^J \mathcal{I}_j$ and D_j out of D paths have the related spectrum $S_{\varphi,d}(\omega)$ nonzero for ω ranging over \mathcal{I}_j (see Fig. 2). This operation allows the application of the matrix inversion lemma to (12). In fact, the latter cannot be directly

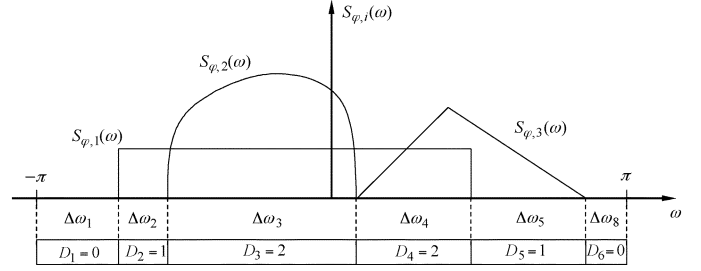


Fig. 2. Illustration of the procedure employed for evaluation of the lower bound (12).

applied to (12), since, in general, $\mathbf{S}_{dd}(\omega)$ is not invertible for every $\omega \in (-\pi, \pi]$. Accordingly, the bound (12) can be written as

$$\begin{aligned} \text{MSE}_{\hat{h}} = \sum_{j=1}^J \text{tr} & \left[\int_{\mathcal{I}_j} \mathbf{U}_j \left(\mathbf{S}_{dd,j}(\omega) \right)^{-1} \right. \\ & \left. + \mathbf{U}_j^H \left[\mathbf{R}_n^{-1} \otimes \mathbf{R}_x \right] \mathbf{U}_j \right)^{-1} \mathbf{U}_j^H \frac{d\omega}{2\pi} \right] \end{aligned} \quad (14)$$

where the $N_R N_T W \times r_j$ matrix \mathbf{U}_j is the unitary basis of the subspace spanned by the columns of the $N_R N_T W \times D_j$ matrix \mathbf{T}_j , that contains the columns of \mathbf{T} corresponding to the D_j paths of the j th subband (accordingly, $\mathbf{T}_j = \mathbf{U}_j \mathbf{\Sigma}_j \mathbf{V}_j^H$). Moreover, it is $\mathbf{S}_{dd,j}(\omega) = \mathbf{\Sigma}_j \mathbf{V}_j^H \mathbf{S}_{\Phi,j}(\omega) \mathbf{V}_j \mathbf{\Sigma}_j$ where the $D_j \times D_j$ diagonal matrix $\mathbf{S}_{\Phi,j}(\omega)$ contains the spectra of the D_j paths of the j th subband. Furthermore, since we restrict the analysis to optimally designed training sequences, the lower bound reduces to

$$\text{MSE}_{\hat{h}} = \sum_{j=1}^J \text{tr} \left[\int_{-\pi}^{\pi} \left(\mathbf{S}_{dd,j}(\omega) \right)^{-1} + \text{SNR} \cdot K_p \mathbf{I}_{r_j} \right)^{-1} \frac{d\omega}{2\pi} \right] \quad (15)$$

with the SNR defined as in (9).

To gain further insight on the bound (15), herein we consider the following simplifications.

Low SNR: For low SNR, the lower bound (15) becomes

$$\text{MSE}_{\hat{h}} \simeq \sum_{j=1}^J \left(\int_{-\pi}^{\pi} \text{tr} \left[\mathbf{S}_{\Phi,j}(\omega) \mathbf{\Sigma}_j^2 \right] \frac{d\omega}{2\pi} \right) = N_R N_T \quad (16)$$

that equals the energy of the channel (see Section II-B). This is consistent with the typical behavior of MMSE estimators.

High SNR: For high SNR, we have

$$\text{MSE}_{\hat{h}} \simeq \sum_{j=1}^J \frac{\Delta\omega_j \cdot r_j}{2\pi \cdot \text{SNR} \cdot K_p} \quad (17)$$

that is independent of the particular shape of the power spectral densities in \mathbf{S}_{Φ} . In fact, it only depends on the Doppler spectrum bandwidths of the different paths.

Uniform Doppler Spectra: Let us assume that all paths have equal uniform Doppler spectrum $S_{\varphi,d}(\omega) = S_{\varphi}(\omega) = 1/(2f_D)$ ($d = 1, \dots, D$) for $|\omega| \leq 2\pi f_D$ (f_D is usually referred to as Doppler spread, and $0 \leq f_D \leq 0.5$), in this case $\mathbf{S}_{dd,j}(\omega) =$

$1/(2f_D)\Sigma_j^2$, and we easily get from (15), by setting $J = 1$ in (17)

$$\text{MSE}_{\hat{h}} = \sum_{i=1}^r \frac{2f_D \lambda_i^2}{\lambda_i^2 \text{SNR} \cdot K_p + 2f_D} \quad (18)$$

that for high SNR becomes

$$\text{MSE}_{\hat{h}} \simeq \frac{2f_D r}{\text{SNR} \cdot K_p}. \quad (19)$$

B. Numerical Results

Let us first start by describing the MIMO-OFDM system that will be used throughout the paper for performance evaluations. $N_T = 4$ transmitting antennas and $N_R = 4$ receiving antennas (half-wavelength-spaced linear antenna arrays at both sides) communicate over a multicarrier link with $K = 32$ subcarriers (recall that we consider optimally designed training sequences, $\mathbf{R}_x = (K_p P/K)\mathbf{I}_{N_T W}$, and spatially white noise, $\mathbf{R}_n = \mathbf{I}_{N_R}$). The time-varying channel is generated according to the model presented in Section II-B with temporal support $W = 8$. The number of pilot tones for each transmitting antenna is $K_p = W = 8$ and, as a consequence, a total of $N_T K_p = K = 32$ subcarriers are used for training. It follows that in such a system, preliminary pilot OFDM symbols (i.e., training OFDM symbols containing only pilot subcarriers and no data) are sent before transmission of data OFDM symbols. This is the situation encountered in many wireless communication standards, such as IEEE 802.11 and IEEE 802.16. For simplicity, we will consider the case where a pilot OFDM symbol is inserted periodically every N data OFDM symbols. Furthermore, where not stated otherwise, the channel is characterized by $D = 4$ paths with uniform power delay profile (PDP), delays $\boldsymbol{\tau} = [1 \ 3 \ 4 \ 6]T$, DODs and DOAs equally spaced in the angular support $(-60, 60)^\circ$. With these choices, all the paths are resolvable in both temporal and spatial domains, and therefore, the number of modes equals the number of paths, i.e., $r = 4$. Notice that this simplified channel scenario is only to avoid the introduction of inessential complications in the analysis. We refer the reader to [33] for related simulation results with the standard channel model [32].

The lower bound (15) is evaluated numerically in order to show the impact of Doppler spectra and PDPs on the channel-estimation performance. The results in Fig. 3 show the lower bound $\text{MSE}_{\hat{h}}$ versus SNR for four cases: uniform Doppler spectra equal for all paths with $f_D = 0.5$ (see previous section) and uniform/nonuniform PDP [see boxes (a)–(b)]; uniform Doppler spectra with different shapes for different paths [see box (c)], and uniform/nonuniform PDP. The solid curves represent results for uniform PDP for the four paths [as in box (a)], while dashed curves are derived when 80% of the power is associated with the path with the smallest Doppler bandwidth [as in box (b)].

By comparing the lower bounds with the reference performance of the LS estimate, it is clear that exploiting the knowledge about the channel structure yields very relevant benefits (around 15 dB in this example). Moreover, we notice that smaller estimation error can be achieved when some paths

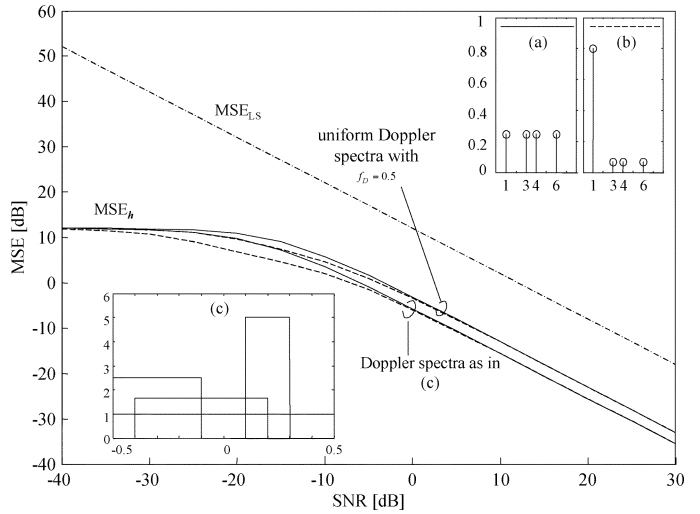


Fig. 3. MSE lower bound $\text{MSE}_{\hat{h}}$ versus SNR for paths with same/different Doppler spectra and uniform/nonuniform PDP. As a reference, the MSE of the LS channel estimate is shown.

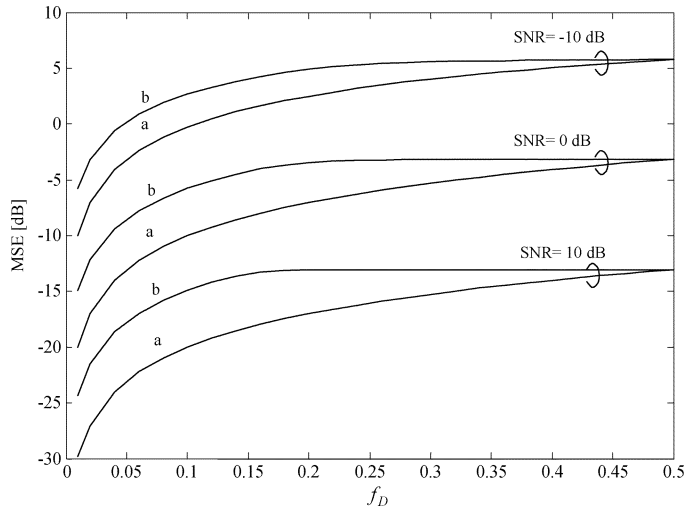


Fig. 4. Lower bound $\text{MSE}_{\hat{h}}$ versus Doppler spread f_D for uniform (a) and truncated Gaussian (b) Doppler profiles, and for different SNR.

have limited Doppler support, and that the improvement is more significant when a greater fraction of power is associated with these paths. This conclusion is consistent with the analytical result in (17) derived for high SNR, and here supported by simulation for $\text{SNR} > 0$ dB. Further discussion on the behavior of the bound versus SNR can be found in Section IV-D.

Fig. 4 shows the dependence of the lower bound $\text{MSE}_{\hat{h}}$ on the Doppler spread f_D in the case where the spectra of the four paths are identical. We remark that since f_D is normalized with respect to the time interval between two successive training OFDM symbols, the results shown here and in the following are independent of the number of data OFDM symbols in between. Uniform (a) and truncated Gaussian (b) spectra have been considered (in the latter case, f_D is the 3-dB cutoff frequency). The spectra are scaled so that they correspond to equal received power. We notice that the truncated Gaussian spectrum yields a larger error, as it is defined over a larger support. Moreover, as expected for $f_D = 0$, the lower bound is zero, showing

that in case the channel is static (no Doppler variations), it can be asymptotically estimated with accuracy.

IV. CHANNEL ESTIMATION BY MODAL ANALYSIS/FILTERING

As discussed in the previous section, the accuracy of channel estimation can be greatly improved if the structure of the channel vector (7) is taken into account when designing the channel-estimation algorithm. In particular, the derivation of the lower bound (12) suggested the following guidelines: 1) long- and short-term channel parameters have to be estimated separately; 2) long-term parameters have to be estimated consistently; 3) short-term parameters (i.e., fast-varying amplitudes) have to be tracked according to the MMSE criterion. Notice that many known estimators proposed in the literature under simplified settings have (at least one of) the aforementioned properties [8], [30].

A. Estimation of the Long-Term Features of the Channel

The long-term part of the MIMO channel can be obtained by following either a structured or an unstructured approach. The structured approaches are parametric methods that directly estimate angles and delays. Such techniques have been developed for single-input multiple-output (SIMO) systems to exploit the stationarity of angles/delays in \mathbf{T} [9]. The extension to MIMO systems is conceptually trivial, and it is just a matter of increased complexity. Even if these parametric methods guarantee a consistent estimate of the long-term features of the channel (but only for well-resolved paths), there are several drawbacks that prevent their practical use, such as high computational complexity and the requirement of regular spatial and temporal manifolds that imposes strict constraints on array calibration errors and modeling mismatches. Furthermore, angle and delay estimation suffers from threshold effects at low SNRs typical for nonlinear estimators.

1) *Multipath Modal Analysis (MMA)*: Instead of estimating angles and delays in \mathbf{T} , it is possible to directly evaluate the space-time modes \mathbf{U} (unstructured approach [22], [29]). This choice not only poses less stringent requirements on array calibration and modeling accuracy (the relationship between \mathbf{T} and angles/delays is not of concern), but also avoids the impairments of nonlinear estimation, since it reduces to a quadratic optimization problem.

Since $E[\mathbf{h}_\ell \mathbf{h}_\ell^H] = \mathbf{T}\mathbf{T}^H$ and $\text{span}\{\mathbf{T}\} = \text{span}\{\mathbf{U}\}$, the estimate $\hat{\mathbf{U}}$ of the modes \mathbf{U} [i.e., modal analysis (MA)] can be obtained from the r leading eigenvectors of the sample correlation matrix $(1/L) \sum_{\ell=1}^L \mathbf{h}_{\text{LS},\ell} \mathbf{h}_{\text{LS},\ell}^H$ computed from the set of L LS estimates $\mathbf{h}_{\text{LS},\ell}$ (method of moments estimator [27]). The estimate is consistent ($\text{cov}\{\hat{\mathbf{U}}\} \rightarrow 0$ if $L \rightarrow \infty$) if the fading is asymptotically uncorrelated (i.e., $\varphi_d(n) \rightarrow 0$ for $n \rightarrow \infty$) [22].

An adaptive computation of the space-time modes \mathbf{U} that alleviates the computational burden of the eigenvalue decomposition, and allows for continuous (but still slow) variations of the channel space-time modes, can be obtained through a subspace tracking algorithm. Here we selected the subspace tracker proposed by [23] for its good tradeoff between complexity and accuracy. Table II summarizes the subspace tracking algorithm with adaptive rank estimation (for this application, $\mathbf{v}_\ell = \mathbf{h}_{\text{LS},\ell}$,

TABLE II
SUBSPACE TRACKING ALGORITHM WITH ADAPTIVE RANK ESTIMATION

<p>Initialize: $r_{\max}; \mathbf{E}_0 = \begin{bmatrix} \mathbf{I}_{r_{\max}} \\ \mathbf{0} \end{bmatrix}; \Theta_0 = \mathbf{I}_{r_{\max}}; \mathbf{C}_0 = \mathbf{0}; p_0 = 0; 0 \leq \gamma \leq 1$</p> <p>For each symbol n:</p> <p>input: \mathbf{v}_ℓ, M</p> <p>1. Subspace tracking:</p> $\mathbf{Z}_\ell = \mathbf{E}_{\ell-1}^H \mathbf{v}_\ell$ $\mathbf{C}_\ell = \gamma \mathbf{C}_{\ell-1} \Theta_{\ell-1} + \mathbf{v}_\ell \mathbf{Z}_\ell^H$ $\mathbf{C}_\ell = \mathbf{E}_\ell \mathbf{R}_\ell \quad (\text{QR factorization})$ $\Theta_\ell = \mathbf{E}_{\ell-1}^H \mathbf{E}_\ell$ <p>2. Adaptive rank estimation:</p> $\hat{\lambda}_i = [\mathbf{R}_\ell]_{ii} \quad i = 1, 2, \dots, \bar{r}_{\max}$ $p_\ell = \gamma p_{\ell-1} + \frac{1}{M} \text{tr}\{\mathbf{v}_\ell \mathbf{v}_\ell^H\}$ $\hat{\sigma}^2 = \frac{M}{M-r_{\max}} p_\ell - \frac{1}{M-r_{\max}} \text{tr}\{\mathbf{R}_\ell\}$ $\hat{r}_\ell = \text{card}\{\hat{\lambda}_i : \hat{\lambda}_i > \beta \cdot \hat{\sigma}^2\}$ <p>updated basis: $\hat{\mathbf{U}}_\ell = [\mathbf{E}_\ell]_{:,1:\hat{r}_\ell}$</p>

$M = N_T N_R W$). Notice that an upper bound r_{\max} on the number of the space-time modes r has to be predetermined, which may be derived from *a priori* knowledge about the channel. The coefficient γ rules the memory of the algorithm, and can be adjusted to accommodate temporal variations of the channel modes. Moreover, the coefficient β is a threshold used for rank-estimation purposes. Its value can be adjusted as a function of system parameters and SNR, as thoroughly explained in [24]. The estimate of the modal matrix for the ℓ th OFDM training symbol $\hat{\mathbf{U}}_\ell$ is obtained by taking the first \hat{r}_ℓ columns of \mathbf{E}_ℓ .

We refer to the (adaptive) approach described in this section for computing the long-term modes of the channel as *M MA*, where *M* stands for multipath.

2) *Space-Time MA (ST MA)*: The estimation of the long-term modes of the channel \mathbf{U} as detailed above suffers from high computational complexity and slow convergence, due to the size of the space-time modal matrix \mathbf{U} . A suboptimal approach is preferred here, as in [22], which considers separately the stationarity of angles and delays. The idea is the following: decouple the space-time modes (\mathbf{U}) into spatial (\mathbf{U}_S) and temporal (\mathbf{U}_T) modes, to be estimated separately. The advantage is that the corresponding modal matrices (\mathbf{U}_S and \mathbf{U}_T) have reduced dimensions, as compared with the space-time modal \mathbf{U} . The key to obtain this result is designing the channel estimator from parametrization (6) instead of (7). In fact, the channel matrix $\hat{\mathbf{h}}_\ell$ in (6) can be written as

$$\hat{\mathbf{h}}_\ell = \mathbf{U}_S \Gamma_\ell \mathbf{U}_T^T \quad (20)$$

that can be derived from the SVD $\mathcal{A}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)}) = \mathbf{U}_S \boldsymbol{\Sigma}_S \mathbf{V}_S^H, \mathbf{G}(\boldsymbol{\tau}) \boldsymbol{\Omega}^{1/2} = \mathbf{U}_T \boldsymbol{\Sigma}_T \mathbf{V}_T^H$, and by definition of $\Gamma_\ell = \boldsymbol{\Sigma}_S \mathbf{V}_S^H \text{diag}(\boldsymbol{\beta}_\ell) \mathbf{V}_T^* \boldsymbol{\Sigma}_T$. Notice that \mathbf{U}_S is $N_T N_R \times r_S$ and \mathbf{U}_T $W \times r_T$, where the number of spatial and temporal modes are $r_S = \text{rank}[\mathcal{A}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)})]$ (upper bounded by $N_R N_T$) and $r_T = \text{rank}[\mathbf{G}(\boldsymbol{\tau}) \boldsymbol{\Omega}^{1/2}] = \text{rank}[\mathbf{G}(\boldsymbol{\tau})]$ (upper

bounded by W), that are determined by the number of resolvable paths in the separate angular and temporal domains.

A suboptimal MA can now be defined as the estimation of the spatial and temporal modes \mathbf{U}_S and \mathbf{U}_T through computation of the principal r_S (and r_T) eigenvectors of the spatial (and temporal) sample covariance matrices $(1/L) \sum_{\ell=1}^L \check{\mathbf{h}}_{LS,\ell} \check{\mathbf{h}}_{LS,\ell}^H$ (and $(1/L) \sum_{\ell=1}^L \check{\mathbf{h}}_{LS,\ell}^H \check{\mathbf{h}}_{LS,\ell}$). Again, the estimate is consistent, and coincides with the solution given by the method of moments [27].

As explained above, an adaptive and low-complexity alternative to the computation of eigenvalue decomposition is the implementation of a subspace tracking algorithm. With reference to the subspace tracker in Table II, the input of the algorithm to track the spatial (or temporal) modes are $\mathbf{v}_\ell = \check{\mathbf{h}}_{LS,\ell}$ (or $\mathbf{v}_\ell = \check{\mathbf{h}}_{LS,\ell}^H$) and $M = N_R N_T$ (or $M = W$), whereas the output is $\hat{\mathbf{U}}_{S,\ell}$, equal to the first $\hat{r}_\ell = \hat{r}_{S,\ell}$ columns of \mathbf{E}_ℓ (and similarly, for $\hat{\mathbf{U}}_{T,\ell}$).

We denote this (adaptive) technique for computing the long-terms modes of the channel as *S/T MA*, in order to recall that spatial and temporal modes are computed separately.

Remark: The proposed MA technique is applicable for LS estimates affected by uncorrelated estimation error. Thus, some modifications are to be introduced to the *S/T* technique in order to cope with correlated estimation error. This arises for spatially correlated noise at the receiver or for suboptimal training sequences (e.g., for decision-directed estimation). In this case, whitening of the LS estimate has to be performed before MA [22]. A detailed discussion on this topic is outside the scope of this paper.

3) *M MA versus S/T MA:* It is interesting to investigate the reasons why parametrization (20) will eventually lead to a (asymptotically with respect to L) suboptimum estimator. By stacking (20), we get $\mathbf{h}_\ell = \text{vec}\{\check{\mathbf{h}}_\ell\} = (\mathbf{U}_T \otimes \mathbf{U}_S) \boldsymbol{\gamma}_\ell$, where $\boldsymbol{\gamma}_\ell = \text{vec}\{\check{\boldsymbol{\Gamma}}_\ell\}$. Recalling the channel vector parametrization (7) [see also (29)], it is clear that we are here parametrizing the space-time modes \mathbf{U} as the Kronecker product of the spatial modal matrix \mathbf{U}_S and the temporal modal matrix \mathbf{U}_T . In other terms, we are assuming that the space-time modes are *separable* into spatial and temporal modes, or equivalently, that the number of space-time modes r can be written as $r = r_S r_T$. This condition only arises in very special situations, such as in the example sketched in Fig. 5. In this particular case (tailored for simplicity of representation of systems with $N_R = 1$ or $N_T = 1$, so that there is only one angle to be concerned with), there is a well-resolved path for each couple angle/delay. In other terms, we could say that the geometry of the paths in the angle/delay domain is separable over the two dimensions (and $r = 4 = r_S r_T$, where $r_S = r_T = 2$). In any other case, it is $r < r_S r_T$, and this condition implies a degraded performance of the estimator based on (20) for $L \rightarrow \infty$, as explained in Section IV-C.

B. Estimation of the Fast-Varying Fading Amplitudes

From the guidelines recalled at the beginning of this section, fading amplitudes β_ℓ should be tracked according to the MMSE criterion. In the context of MIMO systems, Kalman filtering has been recently proposed to track the fading amplitudes by as-

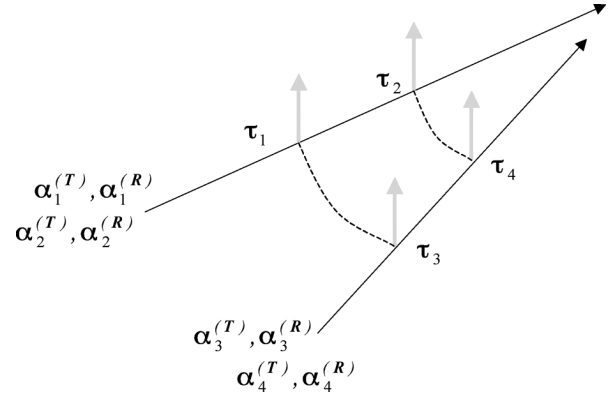


Fig. 5. Path geometry in the angle/delay domain that guarantees the optimality of the *S/T* MA approach.

suming an autoregressive model for Doppler spectra and known sample-spaced delays [30]. On the other hand, the estimator proposed in [28] for the case $N_T = 1$ estimates the basis \mathbf{U}_S or \mathbf{U}_T by following the suboptimum nonparametric approach described above, and tracks the fading variations by least mean square (LMS) or recursive least square (RLS) algorithms.

Tracking of the fading amplitudes vector has to be performed after estimation of the long-term features of the channel and, as such, its design depends on the approach selected for MA (see previous section). Since we are interested in ending up with a practical channel estimator, for presentation herein, we select the *S/T* MA method. Moreover, we propose two suboptimal ways (with respect to Kalman or MMSE) of tracking the fading amplitudes. The corresponding techniques related to *M MA* are conceptually identical, and can be straightforwardly derived.

1) *LMS Tracking:* Since for *S/T* MA, we have that $\mathbf{h}_\ell = \text{vec}\{\check{\mathbf{h}}_\ell\} = (\mathbf{U}_{T,\ell} \otimes \mathbf{U}_{S,\ell}) \boldsymbol{\gamma}_\ell$, the estimate of the amplitudes $\hat{\boldsymbol{\Gamma}}_\ell$ can be updated according to the LMS algorithm [31] as follows ($0 < \mu < 2$ in order to guarantee stability; \mathbf{b}_0 can be initialized as $\mathbf{b}_0 = \mathbf{0}$):

$$\boldsymbol{\epsilon}_\ell = \mathbf{h}_{LS,\ell} - (\hat{\mathbf{U}}_{T,\ell} \otimes \hat{\mathbf{U}}_{S,\ell}) \mathbf{b}_{\ell-1} \quad (21a)$$

$$\mathbf{b}_\ell = \mathbf{b}_{\ell-1} + \mu \left(\hat{\mathbf{U}}_{T,\ell}^H \otimes \hat{\mathbf{U}}_{S,\ell}^H \right) \boldsymbol{\epsilon}_\ell \quad (21b)$$

$$\hat{\boldsymbol{\Gamma}}_\ell = \text{vec}_{r_S, \max}^{-1} \{ \hat{\boldsymbol{\gamma}}_\ell \} = [\text{vec}^{-1} \{ \mathbf{b}_\ell \}]_{1:\hat{r}_{S,\ell}, 1:\hat{r}_{T,\ell}} \quad (21c)$$

where notation $[\cdot]_{1:a, 1:a}$ denotes the $a \times a$ minor matrix of its argument. The channel estimate is then obtained as

$$\hat{\mathbf{h}}_\ell = (\hat{\mathbf{U}}_{T,\ell} \otimes \hat{\mathbf{U}}_{S,\ell}) \hat{\boldsymbol{\gamma}}_\ell. \quad (22)$$

2) *No Tracking [Modal Filtering (MF)]:* In this case, no effort is made to estimate the fading amplitudes, and the estimator results in

$$\hat{\mathbf{h}}_\ell = \left(\hat{\mathbf{U}}_{T,\ell} \hat{\mathbf{U}}_{T,\ell}^H \otimes \hat{\mathbf{U}}_{S,\ell} \hat{\mathbf{U}}_{S,\ell}^H \right) \mathbf{h}_{LS,\ell} \quad (23)$$

which is simply the projection of the LS estimate onto the space and time modal subspaces. We define this operation as *MF*, so that the channel estimator complete with *S/T* MA will be referred to as *S/T* MA-MF. In [22], it is proved that this estimator is the maximum-likelihood solution for the case where

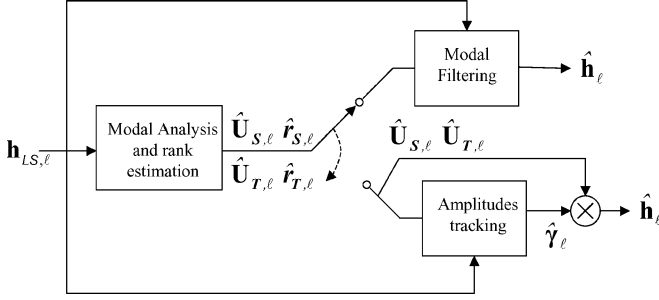


Fig. 6. Block diagram of the S/T MA channel estimator with MF or amplitude tracking.

the amplitudes γ_ℓ are modeled as deterministic variables. Moreover, in [21], a similar estimator that only performs spatial MA for a MIMO (not OFDM) system is proposed, whereas [20] presents an analogous technique based on temporal MA. The S/T MA-MF estimator is expected to perform (asymptotically) according to the bound derived in Section III, in case the fading amplitudes are uncorrelated across different OFDM symbols (so that tracking of the fading variations would be ineffective). This conjecture will be proved in Section IV-D.

In Fig. 6, the block diagram of a channel estimator based on S/T MA with or without tracking of the fading amplitudes is depicted.

C. Asymptotic MSE Performance Analysis

In this section, the asymptotic MSE of the S/T MA-MF channel estimator (23) is computed and compared with the MSE of the conventional LS estimate and with the lower bound derived in Section IV-C. The assessment of the performance of the S/T MA with LMS tracking is more complicated and will be omitted here. According to the model used for derivation of the lower bound, the long-term features of the channel are assumed to be constant over an infinite temporal horizon. Moreover, it will be assumed that the number of modes r_S and r_T are accurately estimated, which is only possible for sufficiently high SNR, as it will be shown by simulations in Section IV-D. The definition “asymptotic” in this section has thus to be interpreted both as a function of the number of training OFDM symbols ($L \rightarrow \infty$) and of SNR ($\text{SNR} \rightarrow \infty$).

The MSE for the LS channel estimate reads [4]

$$\text{MSE}_{\text{LS}} = E [\|\mathbf{h}_{\text{LS},\ell} - \mathbf{h}_\ell\|^2] = \frac{N_R N_T W}{\text{SNR} \cdot K_p}. \quad (24)$$

Recalling (23) and the consistency of the S/T MA estimate of the modal matrices (i.e., $\hat{\mathbf{U}}_{S,\ell} \rightarrow \mathbf{U}_S$ and $\hat{\mathbf{U}}_{T,\ell} \rightarrow \mathbf{U}_T$ for $L \rightarrow \infty$), it is easy to show that the MSE of the S/T MA-MF approaches

$$\begin{aligned} \text{MSE}_{\text{S/TMA-MF}}(L \rightarrow \infty) &= \frac{\text{tr} [\mathbf{U}_S \mathbf{U}_S^H] \text{tr} [\mathbf{U}_T \mathbf{U}_T^H]}{\text{SNR} \cdot K_p} \\ &= \frac{r_S r_T}{\text{SNR} \cdot K_p}. \end{aligned} \quad (25)$$

For reference, recall that the lower bound for high SNR and uncorrelated fading amplitudes over different OFDM symbols (i.e., (17) with $J = 1$, $\Delta f_j = 1$, $r_j = r$) reads

$$\text{MSE}_{\hat{\mathbf{h}}} \simeq \frac{r}{\text{SNR} \cdot K_p}. \quad (26)$$

Comparing (25) with (24), we can conclude that MA/filtering allows a reduction in the MSE equal with respect to the LS method equal to the ratio $(N_R N_T W / r_S r_T)$, that can be interpreted as the product of a spatial gain $(N_R N_T / r_S)$ and a temporal gain W / r_T . These gains quantify the reduction in the number of parameters to be estimated due to the exploitation of the structure of the channel vector. On the other hand, comparing (25) with (26), we get that suboptimality of S/T MA-MF for temporally uncorrelated fading amplitudes can be quantified as the ratio $r / (r_S r_T)$, that equals one only for special channel geometries, as explained in Section IV-A.2. As a final remark, we notice that the bound (26) could be attained (for high SNR) by M MA and corresponding filtering (M MA-MF, see Section IV-B). In fact, following the same computations that led to (25), it is easy to show that the MSE of the M MA-MF technique tends to

$$\text{MSE}_{\text{MMA-MF}}(L \rightarrow \infty) = \frac{r}{\text{SNR} \cdot K_p} \simeq \text{MSE}_{\hat{\mathbf{h}}}. \quad (27)$$

D. Numerical Results

The performance of the proposed channel-estimation method is evaluated through simulations for the MIMO-OFDM system described in Section III-B. Notice that the spatial and temporal modal dimensions read $r_S = 4$ and $r_T = 4$. Doppler spectra of different paths are equal and uniform with normalized Doppler spread $f_D = 0.03$. Validation of the proposed channel-estimation algorithms for the channel model standardized by the 3GPP/3GPP2 spatial channel modeling (SCM) adhoc group [32] can be found in [33].

Figs. 7 and 8 compare the performance of the S/T MA-MF channel estimator in terms of MSE versus SNR with the LS channel estimate, the M MA-MF channel estimate, and the lower bound $\text{MSE}_{\hat{\mathbf{h}}}$. For the S/T MA-MF, both the asymptotic results derived in Section IV-C and the MSE obtained after $L = 50$ training OFDM symbols are shown (performance as a function of L is considered next). On the other hand, the performance of M MA-MF is plotted only as the asymptotic results (27) for reference, given its prohibitive complexity and rate of convergence (see below). The lower bound $\text{MSE}_{\hat{\mathbf{h}}}$ is shown for the case $f_D = 0.5$, in order to confirm the analytical consideration presented in Section IV-C. Toward the goal of investigating the effect of an error in the estimation of the modal dimensions, Fig. 7 shows the performance of S/T MA-MF for a fixed and correctly estimated spatial modal dimension, $\hat{r}_S = r_S = 4$, and a varying temporal modal dimension, $\hat{r}_T = 1, \dots, 4 = r_T$. On the other hand, Fig. 8 considers the performance of S/T MA-MF with the adaptive rank estimation described in Table II. The upper part of the figure shows the estimated number of spatial and temporal modes after $L = 50$

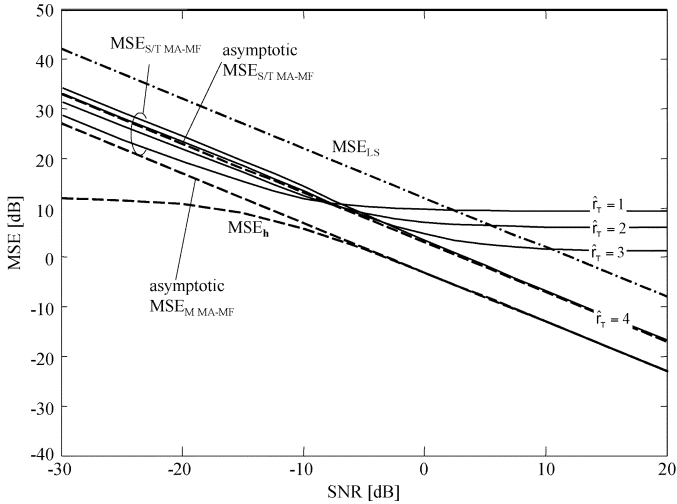


Fig. 7. MSE of the S/T MA-MF technique versus SNR for $L = 50$, compared with the limit performance given by the lower bound $MSE_{\hat{h}}$ for different fixed temporal modal dimensions $\hat{r}_T = 1, \dots, 4$ ($r_T = 4$, $\hat{r}_S = r_S = 4$).

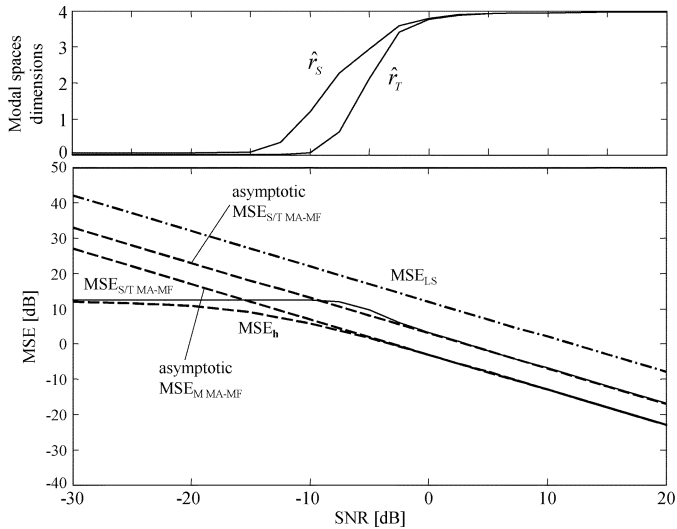


Fig. 8. MSE of the S/T MA-MF technique versus SNR for $L = 50$, compared with the limit performance given by the lower bound $MSE_{\hat{h}}$. The upper part of the figure shows the estimated number of spatial and temporal modes.

training symbols (with $\gamma = 0.999$, $\beta = 1.5$, $r_{S,\max} = 15$, and $r_{T,\max} = 7$).

Observing Figs. 7 and 8 from the upper curve (MSE_{LS}), we can make the following considerations: 1) the S/T MA-MF estimator outperforms (asymptotically) the LS channel estimate by a factor $N_R N_T W / (r_S r_T) \simeq 9$ dB, as expected from the analysis in Section IV-C. 2) For $L = 50$ training symbols, the S/T MA-MF essentially converges to the asymptotic performance if the modal dimensions are correctly estimated. Moreover, for small SNR (SNR < 0 dB), it is convenient to underestimate the number of modes, since this has the effect of lowering the MSE as compared with the asymptotic result (valid for large SNR). This phenomenon can be explained by noticing that reducing the number of modes implies trading a bias in the estimate (due to the underparametrization) for a reduced variance [22]. 3) The adaptive rank algorithm described in Table II is able to estimate correctly the modal dimensions for large SNR, whereas

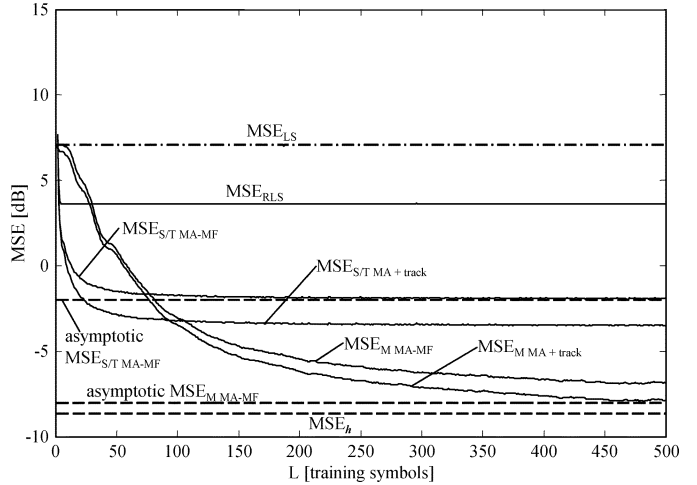


Fig. 9. MSE of the S/T MA-MF technique versus L for SNR = 5 dB, compared with the limit performance given by the lower bound. As a reference, performance of an RLS estimator is given.

for small SNR, it tends to underparametrize the model. As discussed above, this has a beneficial effect on the channel-estimation performance. 4) The asymptotic MSE of the M MA-MF technique coincides with the lower bound for large SNR, as anticipated in Section IV-C. 5) The S/T MA-MF technique has an asymptotic degradation of $r / (r_S r_T) = 6$ dB. 6) The qualitative behavior of the bound $MSE_{\hat{h}}$ for small SNR is analogous to that discussed above of S/T MA-MF. In this case, the underparametrization of the channel (and consequent tradeoff between bias and variance) is automatically performed by MMSE filtering. Accordingly, for very small SNR, the bound $MSE_{\hat{h}}$ tends to the channel norm (16).

The MSE as a function of the number L of processed symbols for SNR = 5 dB is shown in Fig. 9 for M MA and S/T MA with and without amplitude tracking. Moreover, as a further term of comparison, the performance of an RLS algorithm with optimized learning factor is considered [31]. It can be seen that LMS amplitude tracking (with $\mu = 0.8$) allows a gain of approximately 2 dB with respect to MF. Moreover, convergence of the performance of S/T MA-MF to the asymptotic value (25) is obtained (by a fraction of a decibel) for $L > 50$. On the other hand, M MA-MF has a very slow convergence, and becomes advantageous with respect to S/T MA-MF only for $L > 80$. MA clearly outperforms RLS for L large enough. Finally, suboptimality of S/T MA with LMS tracking can be quantified by approximately 5 dB through comparison with the lower bound.

V. MIMO-OFDM SYSTEM WITH BICM AND TURBO EQUALIZATION

In order to show the effectiveness of the proposed channel-estimation technique in terms of probability of error of a practical system, here we consider a MIMO-OFDM system based on BICM [12] and turbo equalization [13]. This combination has been recently recognized as a promising candidate for next-generation wireless LAN [14].

With reference to Fig. 1, the block diagram of the MIMO encoder/modulator for the considered system is shown in Fig. 10(a). The information bit stream d_n is passed through

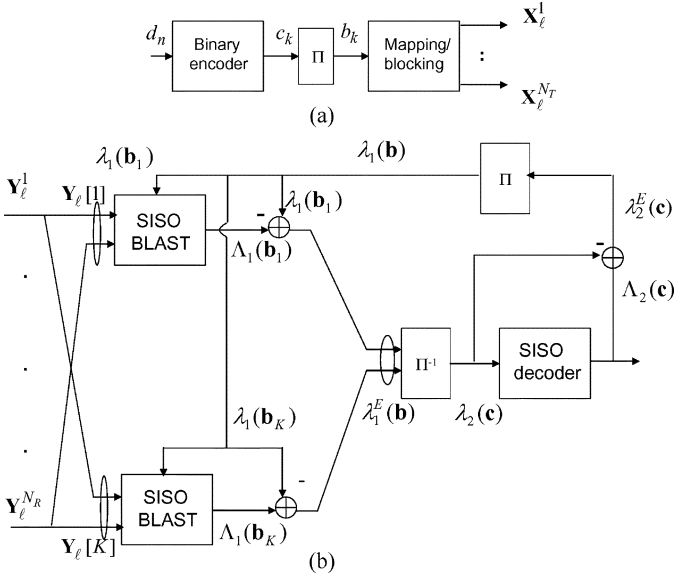


Fig. 10. Block diagram of the transmitter (a) and receiver (b) of a MIMO-OFDM system, based on BICM and turbo equalization.

a convolutional encoder producing the encoded bits c_k , then interleaved (interleaved bits are denoted as b_k), and finally modulated into a M -ary quadrature amplitude modulation (QAM) constellation and blocked into $K \times 1$ vectors $\mathbf{X}_\ell^{n_T}$ ($n_T = 1, \dots, N_T$), to be transmitted by different transmitting antennas (according to the model in Section II-A). No attempt of optimizing the interleaving operation over different transmitting antennas and frequencies is made here.

The receiver is depicted in Fig. 10(b). Channel estimation is not included in the block diagram for simplicity. The received signals on each frequency $\mathbf{Y}_\ell[k]$ ($k = 1, \dots, K$) [see (3)] are processed separately by a soft-input soft-output (SISO) BLAST decoder [13]. This performs MMSE linear filtering of the input signals and, based on the Gaussian approximation of the residual interference, computes the *a posteriori* log-likelihood ratios (LLRs) $\Lambda_1(\mathbf{b}_k)$ of the encoded and interleaved bits \mathbf{b}_k ($k = 1, \dots, K$). The $N_T \log_2 M \times 1$ vector \mathbf{b}_k represents the encoded and modulated bits that are mapped onto the symbols collected in the $N_T \times 1$ vector $\mathbf{X}_\ell[k]$ transmitted on the k th subcarrier. After subtraction of the *a priori* LLR $\lambda_1(\mathbf{b}_k)$ (obtained from the SISO decoder, as explained below, and initialized to zero for the first iteration), the so-obtained extrinsic LLR $\lambda_1^E(\mathbf{b})$ (where $\mathbf{b} = [\mathbf{b}_1^T, \dots, \mathbf{b}_K^T]^T$) are deinterleaved, producing the *a priori* LLR $\lambda_2(\mathbf{c})$ for the SISO decoder (\mathbf{c} is the $KN_T \log_2 M \times 1$ vector obtained by deinterleaving of \mathbf{b}). In a similar way, from the *a posteriori* LLR $\Lambda_2(\mathbf{c})$ produced by the decoder, the *a priori* LLRs $\lambda_2(\mathbf{c})$ are subtracted, yielding the extrinsic LLR $\lambda_2^E(\mathbf{c})$ that, interleaved, provide the *a priori* LLR for the SISO BLAST equalizer.

A. Simulation Results

The proposed MIMO-OFDM system with the following specific features is tested: the convolutional encoder has rate $R =$

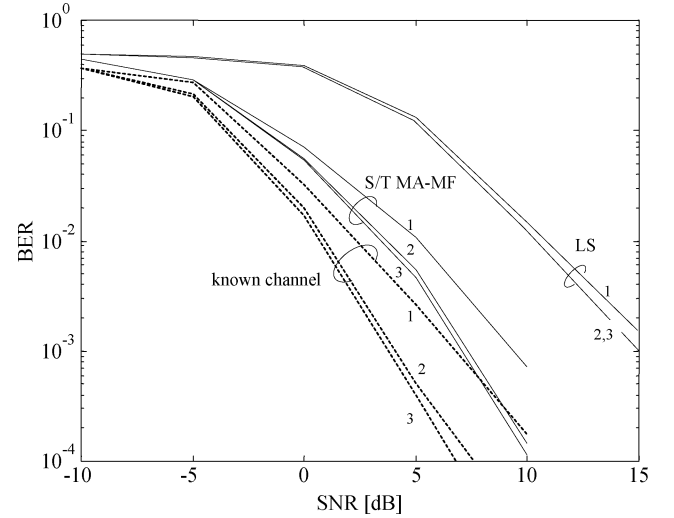


Fig. 11. BER for a MIMO-OFDM system based on BICM and turbo equalization with channel-estimation error.

$1/2$ and generators [7,5], the interleaver is random, the modulation is 4-QAM ($M = 4$), and the SISO decoder is log-maximum *a posteriori* (MAP) [26]. The bit-error rate (BER) of such a system is plotted in Fig. 11 for the ideal case of perfect knowledge of the channel, for an LS channel estimate, and for an S/T MA-MF estimator (with $L = 50$, as in Fig. 8) as a function of the equalization-decoding iteration (after the third, only minor improvements are obtained). For a BER equal to 10^{-3} , S/T MA-MF guarantees approximately 7 dB gain as compared with the LS channel estimate, and is only 3 dB away from the case of ideal channel knowledge.

VI. CONCLUSION

A lower bound for the channel-estimation error for MIMO-OFDM systems has been derived, in order to assess to what extent LS channel estimation can be improved if knowledge about the multipath channel structure is taken into account in deriving the channel estimator. Moreover, based on the insight obtained from the aforementioned analysis, practical channel estimators have been designed that perform close to the performance limit. The proposed channel estimators compute the long-term features through a subspace tracking algorithm by identifying the invariant (over multiple OFDM symbols) space-time modes of the channel. On the other hand, the fast-varying fading amplitudes are possibly tracked by using LS techniques that exploit temporal (i.e., over multiple OFDM symbols) correlation of the fading process.

Thorough numerical investigation has validated the performance of the proposed techniques. In particular, MIMO-OFDM with BICM and MIMO-turbo equalization has been selected as a benchmark for performance evaluation in terms of BER.

APPENDIX A

DERIVATION OF CHANNEL MODEL (5)

Let us define the $N_R \times N_T W$ MIMO CIR matrix $\bar{\mathbf{h}}_\ell$ as $\bar{\mathbf{h}}_\ell = [\bar{\mathbf{h}}_\ell[1] \cdots \bar{\mathbf{h}}_\ell[W]]$, where the $N_R \times N_T$ matrix $\bar{\mathbf{h}}_\ell[w]$ is the w th “MIMO tap” collecting the samples $h_\ell^{n_R, n_T}[w]$ over the $N_R N_T$

links, i.e., $[\bar{\mathbf{h}}_\ell[w]]_{n_R, n_T} = h_\ell^{n_R, n_T}[w]$. Using the model (4), we get

$$\bar{\mathbf{h}}_\ell = \sum_{d=1}^D \sqrt{\Omega_{d,\ell}} \mathbf{g}(\tau_{d,\ell})^T \otimes \mathbf{A}_{d,\ell} \quad (28)$$

where $\mathbf{g}(\tau_{d,\ell})$ is the $W \times 1$ vector containing the T -spaced samples of the delayed waveform $g(\tau - \tau_{d,\ell})$.

The channel model (28) can be recast in the $N_R N_T \times W$ matrix $\check{\mathbf{h}}_\ell$ by stacking the W MIMO taps $\{\bar{\mathbf{h}}_\ell[w]\}_{w=1}^W$ as $\check{\mathbf{h}}_\ell = [\text{vec}\{\bar{\mathbf{h}}_\ell[1]\} \cdots \text{vec}\{\bar{\mathbf{h}}_\ell[W]\}]$. Model (5) easily follows.

APPENDIX B DERIVATION OF THE ASYMPTOTIC CRB

Here we derive the lower bound on the channel-estimation error discussed in Section III. In order to obtain an identifiable parametrization of the channel vector, we consider the SVD of the stationary matrix $\mathbf{T} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ and rewrite the channel model (7) as

$$\mathbf{h}_\ell = \mathbf{T}\boldsymbol{\beta}_\ell = \mathbf{U}(\mathbf{\Sigma}\mathbf{V}^H\boldsymbol{\beta}_\ell) = \mathbf{U}\mathbf{d}_\ell \quad (29)$$

where the $N_R N_T W \times r$ matrix \mathbf{U} contains the (long-term) modes of the channel, i.e., the signatures of the r resolvable paths (out of D) of the multipath channel. The $r \times 1$ vector $\mathbf{d}_\ell = \mathbf{\Sigma}\mathbf{V}^H\boldsymbol{\beta}_\ell$ contains the amplitudes of the modes. Notice that

$$\mathbf{R}_d(n) = \mathbf{\Sigma}\mathbf{V}^H \mathbf{R}_\beta(n) \mathbf{V}\mathbf{\Sigma} = \mathbf{\Sigma}\mathbf{V}^H \boldsymbol{\Phi}(n) \mathbf{V}\mathbf{\Sigma}. \quad (30)$$

The signal model for the MIMO channel can be written from (10) as

$$\begin{aligned} \bar{\mathbf{Y}}_\ell &= (\mathbf{I}_{N_R} \otimes \mathbf{B}) \mathbf{h}_\ell + \bar{\mathbf{N}}_\ell \\ &= (\mathbf{I}_{N_R} \otimes \mathbf{B}) \mathbf{U}\mathbf{d}_\ell + \bar{\mathbf{N}}_\ell \\ &= \mathbf{P}\mathbf{d}_\ell + \bar{\mathbf{N}}_\ell \end{aligned} \quad (31)$$

where $\bar{\mathbf{Y}}_\ell = [\bar{\mathbf{Y}}_\ell^1 \cdots \bar{\mathbf{Y}}_\ell^{N_R}]^T$ and $\bar{\mathbf{N}}_\ell = [\bar{\mathbf{N}}_\ell^1 \cdots \bar{\mathbf{N}}_\ell^{N_R}]^T$ collect, respectively, the received signals and the additive noise at the N_R receiving antennas. Notice that matrix $\mathbf{P} = (\mathbf{I}_{N_R} \otimes \mathbf{B})\mathbf{U}$ is asymptotically known by the estimator, since this is assumed to perform a consistent estimate of the long-term features of the channel \mathbf{U} . MMSE estimation of the amplitudes can be carried out by the Wiener filter in the frequency domain

$$\hat{\mathbf{D}}(\omega) = \mathbf{S}_{dy}(\omega) \mathbf{S}_{yy}(\omega)^{-1} \bar{\mathbf{Y}}(\omega) \quad (32)$$

with $\hat{\mathbf{D}}(\omega) = \mathcal{F}\{\hat{\mathbf{d}}_\ell\}$ and $\bar{\mathbf{Y}}(\omega) = \mathcal{F}\{\bar{\mathbf{Y}}_\ell\}$, and $\mathbf{S}_{dy}(\omega) = \mathcal{F}\{E[\mathbf{d}_\ell \bar{\mathbf{Y}}_{\ell-n}^H]\}$ and $\mathbf{S}_{yy}(\omega) = \mathcal{F}\{E[\bar{\mathbf{Y}}_\ell \bar{\mathbf{Y}}_{\ell-n}^H]\}$, the operator $\mathcal{F}\{\cdot\}$ denoting the Fourier transformation. Estimation error depends only on the error in estimating the amplitudes $\hat{\mathbf{d}}_\ell$, as the modes are supposed to be (asymptotically) known. The error-

correlation matrix on the estimate of $\hat{\mathbf{h}}_\ell = \mathbf{U}\hat{\mathbf{d}}_\ell$ is $\mathbf{Q}_{\hat{h}} = E[(\mathbf{h}_\ell - \hat{\mathbf{h}}_\ell)(\mathbf{h}_\ell - \hat{\mathbf{h}}_\ell)^H]$ and

$$\text{MSE}_{\hat{h}} = \text{tr}[\mathbf{Q}_{\hat{h}}] = \text{tr}[\mathbf{U}\mathbf{Q}_{\hat{d}}\mathbf{U}^H] \quad (33)$$

that depends only on the correlation matrix on the estimate of the amplitudes $\hat{\mathbf{d}}_\ell$, $\mathbf{Q}_{\hat{d}} = E[(\mathbf{d}_\ell - \hat{\mathbf{d}}_\ell)(\mathbf{d}_\ell - \hat{\mathbf{d}}_\ell)^H]$. By using Parseval's theorem, we get

$$\text{MSE}_{\hat{h}} = \text{tr} \left[\mathbf{U} \left(\int_{-\pi}^{\pi} \mathbf{S}_{ee}(\omega) \frac{d\omega}{2\pi} \right) \mathbf{U}^H \right] \quad (34)$$

where $\mathbf{S}_{ee}(\omega) = \mathcal{F}\{E[(\mathbf{d}_\ell - \hat{\mathbf{d}}_\ell)(\mathbf{d}_{\ell-n} - \hat{\mathbf{d}}_{\ell-n})^H]\} = \mathbf{S}_{dd}(\omega) - \mathbf{S}_{dy}(\omega) \mathbf{S}_{yy}(\omega)^{-1} \mathbf{S}_{yd}(\omega)$. By using the equalities $\mathbf{S}_{yy}(\omega) = \mathbf{P}\mathbf{S}_{dd}(\omega)\mathbf{P}^H + \mathbf{R}_n$ and $\mathbf{S}_{dy}(\omega) = \mathbf{S}_{yd}(\omega)^H = \mathbf{S}_{dd}(\omega)\mathbf{P}^H$, the MSE of the estimate can be written as

$$\begin{aligned} \text{MSE}_{\hat{h}} &= \text{tr} \left[\int_{-\pi}^{\pi} \mathbf{U} (\mathbf{S}_{dd}(\omega) - \mathbf{S}_{dd}(\omega)\mathbf{P}^H \right. \\ &\quad \left. \times (\mathbf{P}\mathbf{S}_{dd}(\omega)\mathbf{P}^H + \mathbf{R}_n \otimes \mathbf{I}_{K_p N_T})^{-1} \mathbf{P}\mathbf{S}_{dd}(\omega)^H) \mathbf{U}^H \frac{d\omega}{2\pi} \right] \end{aligned}$$

that coincides with (12), and $\mathbf{S}_{dd}(\omega)$ can be expressed as (13).

REFERENCES

- [1] G. L. Stuber, J. R. Barry, S. W. McLaughlin, Y. G. Li, M. A. Ingram, and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," *Proc. IEEE*, vol. 92, no. 2, pp. 271–294, Feb. 2004.
- [2] A. van Zelst and T. C. W. Schenk, "Implementation of a MIMO OFDM-based wireless LAN system," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 483–494, Feb. 2004.
- [3] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1699–1712, Jul. 2002.
- [4] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.
- [5] Y. Li, "Optimum training sequences for OFDM systems with multiple transmit antennas," in *Proc. IEEE GLOBECOM*, 2000, vol. 3, pp. 1478–1482.
- [6] V. K. Jones and G. G. Raleigh, "Channel estimation for wireless OFDM systems," in *Proc. IEEE GLOBECOM*, pp. 980–985.
- [7] Y. G. Li, J. H. Winters, and N. R. Sollenberger, "MIMO-OFDM for wireless communications: Signal detection with enhanced channel estimation," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1471–1477, Sep. 2002.
- [8] A. Van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 405–418, Feb. 1998.
- [9] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 449–462, Feb. 1998.
- [10] O. Simeone and U. Spagnolini, "Lower bound on training-based channel estimation error for frequency-selective block-fading Rayleigh MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 11, pp. 3265–3277, Nov. 2004.
- [11] O. Simeone and U. Spagnolini, "Estimation of block-fading channels with reduced-rank correlation matrix," submitted for publication.
- [12] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
- [13] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: Theory and experiments," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2538–2546, Oct. 2002.
- [14] K.-B. Song and S. A. Mujtaba, "A low complexity space-frequency BICM MIMO-OFDM system for next-generation WLANs," in *Proc. GLOBECOM*, 2003, vol. 2, pp. 1059–1063.

- [15] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications—A key to gigabit wireless," *Proc. IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.
- [16] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.
- [17] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Pilot-based channel estimation for OFDM systems by tracking the delay subspace," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 315–325, Jan. 2004.
- [18] B. Yang, K. B. Letaief, R. S. Cheng, and Z. Cao, "Channel estimation for OFDM transmission in multipath fading channels based on parametric channel modeling," *IEEE Trans. Commun.*, vol. 49, no. 3, pp. 467–479, Mar. 2001.
- [19] T. A. Thomas and F. W. Vook, "Broadband MIMO-OFDM channel estimation via near maximum likelihood time of arrival estimation," in *Proc. IEEE ICASSP*, May 2002, vol. 3, pp. 2569–2572.
- [20] J. Du and Y. Li, "MIMO-OFDM channel estimation based on subspace tracking," in *Proc. IEEE VTC*, Apr. 2003, vol. 2, pp. 1084–1088.
- [21] M. Stege, P. Zillmann, and G. Fettweis, "MIMO channel estimation with dimension reduction," in *Proc. IEEE Int. Symp. WPMC*, Oct. 2002, vol. 2, pp. 417–421.
- [22] M. Nicoli, O. Simeone, and U. Spagnolini, "Multislot estimation of fast-varying space-time channels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1184–1195, May 2003.
- [23] P. Strobach, "Low-rank adaptive filters," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 2932–2947, Dec. 1996.
- [24] A. Kavcic and B. Yang, "Adaptive rank estimation for spherical subspace trackers," *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1573–1579, Jun. 1996.
- [25] H. L. Van Trees, *Optimum Array Processing*. New York: Wiley, 2002.
- [26] J. Tan and G. L. Stuber, "New SISO decoding algorithms," *IEEE Trans. Commun.*, vol. 51, no. 6, pp. 845–848, Jun. 2003.
- [27] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [28] M. Nicoli, M. Sternad, U. Spagnolini, and A. Ahlen, "Reduced-rank channel estimation and tracking in time-slotted CDMA systems," in *Proc. ICC*, May 2002, vol. 1, pp. 533–537.
- [29] F. Dietrich, M. T. Ivrlac, and J. A. Nossek, "On performance limits of optimum reduced rank channel estimation," in *Proc. GLOBECOM*, 2002, vol. 1, pp. 345–349.
- [30] C. Kominakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1065–1076, May 2002.
- [31] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
- [32] *Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations (Release 6)*, 3GPP TR 25.996, V.6.0, 3GPP, Jun. 2003, Tech. Spec. Group Radio Access.
- [33] M. Cicerone, N. Geng, O. Simeone, and U. Spagnolini, "Modal analysis/filtering to estimate time-varying MIMO-OFDM channels," in *Proc. ITG Workshop Smart Antennas*, Mar. 2004, pp. 35–40.

- [34] R. Janaswamy, *Radiowave Propagation and Smart Antennas for Wireless Communications*. Boston, MA: Kluwer, 2000.



Marcello Cicerone received the M.Sc. degree (with honors) in telecommunications engineering from Politecnico di Milano, Milano, Italy, in 2004.

During his studies, he was an R&D intern at Siemens, Munich, Germany, where he developed his Master thesis. Subsequently, he joined STMicroelectronics, Milan, Italy, working on MIMO signal processing techniques for 802.11n WLAN chips. Since June 2005, he has been with Analysis, Milan, Italy, providing consulting services to firms in the telecom field. His interests lie in the area of

broadband and wireless communications.



Osvaldo Simeone (M'02) received the M.Sc. degree (with honors) and the Ph.D. degree in information engineering from Politecnico di Milano, Milano, Italy, in 2001 and 2005, respectively.

From February to September 2002, he was a Visiting Researcher with the Center for Communications and Signal Processing Research, New Jersey Institute of Technology (NJIT), Newark, where he is currently a Postdoctoral Researcher and Adjunct Professor. His current research interests lie in the field of information theory and signal processing aspects of wireless systems, with emphasis on cooperative communications, MIMO systems, ad hoc wireless networks, and cognitive radio.



Umberto Spagnolini (SM'03) received the Dott.Ing. Elettronica degree (*cum laude*) from Politecnico di Milano, Milano, Italy, in 1988.

Since 1988, he has been with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, where he is a Full Professor in Telecommunications. His general interests are in the area of statistical signal processing. The specific areas of interest include channel estimation and space-time processing for wireless communication systems, parameter estimation and tracking, signal processing and wavefield interpolation applied to UWB radar, geophysics, and remote sensing.

Dr. Spagnolini serves as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.