

Lower Bound on Training-Based Channel Estimation Error for Frequency-Selective Block-Fading Rayleigh MIMO Channels

Oswaldo Simeone, *Student Member, IEEE*, and Umberto Spagnolini, *Senior Member, IEEE*

Abstract—A lower bound on the error correlation matrix of training-based channel estimators is derived for multiple-input multiple-output (MIMO) systems over block-fading frequency-selective channels with symbol-spaced receivers. The bound is obtained in a constructive way by evaluating the asymptotic performance of an estimator that fully exploits the algebraic structure of the multipath channel. In particular, the estimator is assumed to be able to estimate the long-term features of the channel consistently (e.g., second order statistics of fading, delays, angles) while tracking the fast-varying fading fluctuations by Wiener filtering. Known estimators that are able to attain the bound under simplified settings are referred to, and general guidelines for designing novel estimators are discussed.

Based on the simple analytical expression of the bound, the impact of channel estimation error on the link capacity is investigated for different system parameters and channel characteristics (e.g., Doppler shift, spatial correlation of fading). Numerical results are provided to corroborate the analysis.

Index Terms—Channel capacity, channel estimation, fading channels, MIMO systems.

I. INTRODUCTION

THEORETICAL analysis as well as algorithms devoted to multiple input multiple output (MIMO) systems have mostly relied on the assumption of perfect channel estimate (also referred to as perfect channel state information). However, a practical approach has to take into account the consequences of estimation errors. A setting of recognized effectiveness in the pursuit of Shannon's capacity sees the transmission organized in bursts, each divided into a training period and (one or more) payload section(s) (see, e.g., [1]–[3]). Training sequences are used to sound the radio channel that in a general framework can be modeled as frequency selective. The choice of system parameters (size of the burst, length of the training sequence) is usually aimed at guaranteeing that the channel is approximately constant within each burst and possibly varying from burst to burst (block-fading [3]–[5]). In this framework, it is of theoretical and practical relevance to derive a lower bound

on the error correlation matrix of the channel estimate. The bound is obtained by computing the asymptotic performance of an estimator that fully exploits the structure of the multipath channel.

In this paper, the lower bound will be derived for two different multipath MIMO channel models corresponding to different assumptions about the geometry of antenna arrays and scatterers (see Fig. 1).

- 1) *Beamforming model* [6]: The elements of both transmitting and receiving antenna arrays are co-located, and the scatterers can be considered as point sources. Each path of the multipath channel is characterized by a direction of departure (DOD) and a direction of arrival (DOA) that “steer” the array response at the transmitter and receiver, respectively, toward a delay and a complex amplitude (fading). The latter is in turn modeled as a temporally correlated Gaussian stationary process. As a general rule, this model appears to be well suited for outdoor channels [7] and has been used in a single input multiple output (SIMO) context in, e.g., [8] and [9].
- 2) *Diversity model* [4]: The elements of both the transmitting and receiving antenna arrays are not co-located and/or the different scatterers have to be modeled as distributed sources (see, e.g., [10]). These assumptions are generally appropriate for an indoor scenario [11]. For each delay of the multipath, the channel gains between different transmitting and receiving antennas can be modeled as spatially and temporally correlated jointly Gaussian random variable with zero mean (Rayleigh fading). This model has been used (with some simplifications) by [12] and [13] in the context of MIMO-OFDM transmission.

The time-varying multipath channel has some characteristics that are stationary (or varying over a long-term) and other features that are fast-varying (e.g., fading amplitudes). This property has been recently used in designing signal processing algorithms for wireless communications [8], [14], [15]. A simple example can illustrate the different varying rates: While the fading amplitudes can vary completely when either end of the communication link moves as little as $\lambda/4$ (λ is the carrier wavelength), the angles (DOAs and DODs) remain constant with changes in position of several wavelengths (i.e., 10 to 1000 λ). We show that entries of the frequency-selective MIMO channel matrix within the k th block can be arranged into a vector $\mathbf{h}_k = \mathbf{U}\mathbf{d}_k$ so that the slowly varying term, which is represented by the matrix \mathbf{U} , and the fast varying fading vector \mathbf{d}_k can be decoupled. We consider an estimator that a) is able to consistently estimate the long-term features of the channel

Manuscript received May 8, 2003; revised November 16, 2003. The material in this paper was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Hong Kong, April 2003. This work was supported by the Virtual Immersive Communication (VICOM) project of the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR). The associate editor coordinating the review of this paper and approving it for publication was Prof. Xiaodong Wang.

The authors with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, 32 I-20133 Milano, Italy (e-mail: simeone@elet.polimi.it; spagnolini@elet.polimi.it).

Digital Object Identifier 10.1109/TSP.2004.836534

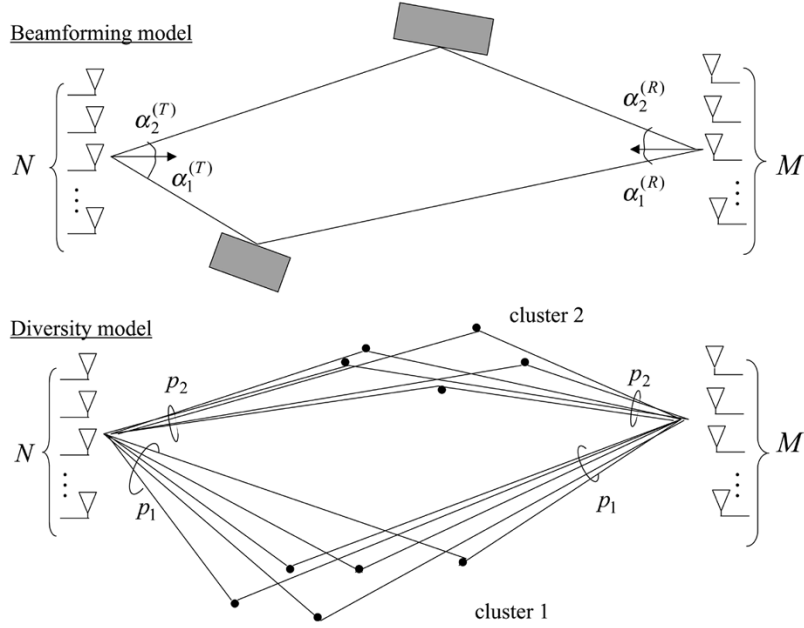


Fig. 1. Geometry of the beamforming (upper part) and diversity (lower part) models.

\mathbf{U} so that for a large number of bursts K (ideally $K \rightarrow \infty$) these can be assumed to be acquired with any accuracy and b) performs optimum [minimum mean square error (MMSE)] tracking of the variations of the fast-varying features \mathbf{d}_k . Notice that many known estimators proposed in the literature under simplified settings have (at least one of) the aforementioned properties [8], [12], [16]. By deriving the asymptotic ($K \rightarrow \infty$) error correlation matrix of the estimate for this method, we set a lower bound on the achievable performance of any channel estimation algorithm over the considered channel models.

The error correlation matrix provides a useful measure to compare the performance of different estimators. Nevertheless, in terms of system performance, it appears to be more sound to investigate the impact of channel estimation error on the link capacity. Toward this end, we extend the analysis carried out in [17] to a MIMO system over a block-fading frequency-selective channel, obtaining a lower bound on the information rate in presence of imperfect channel state information.

The outline of the paper is as follows. In Section II, the problem is formulated, and the notation is set. The beamforming and diversity channel models are introduced in Section III. It is shown that the channel can be decomposed into long-term/fast-varying features within a unifying mathematical framework for both models. Section IV elaborates on these models and derives the corresponding lower bound on the error correlation matrix of the channel estimate. Furthermore, it provides a thorough discussion on the bound, showing the connections with known results for a uniform Doppler spectrum. Practical channel estimators and their performance relative to the bound are discussed in Section V. The impact of channel estimation error on the information rate of the link is addressed in Section VI for a time-slotted MIMO system. Finally, in Section VII, numerical evaluations of the bound related to specific examples are proposed. Below are some essential notations useful to make the analysis of MIMO system more compact.

Notation: Lowercase (uppercase) bold denotes column vector (matrices), $(\cdot)^T$ is the matrix transpose, $(\cdot)^*$ is the

complex conjugate, $(\cdot)^H$ is the Hermitian transposition, and $(\cdot)^\dagger$ is the pseudoinverse. $\mu_i[\mathbf{A}]$ denotes the i th eigenvalue of the $n \times n$ matrix \mathbf{A} , and eigenvalues are arranged in decreasing order $\mu_1[\mathbf{A}] \geq \mu_2[\mathbf{A}] \geq \dots \geq \mu_n[\mathbf{A}]$. \otimes is the Kronecker matrix product, $\mathbf{v} = \text{vec}\{\mathbf{V}\}$ is the stacking operator, and the following property is extensively used: $\text{vec}\{\mathbf{ABC}\} = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}\{\mathbf{B}\}$. \mathbf{I}_P is the $P \times P$ identity matrix, and $\mathbf{R}^{1/2}$ is a factorization of \mathbf{R} such that $\mathbf{R}^{H/2}\mathbf{R}^{1/2} = \mathbf{R}$.

II. PROBLEM FORMULATION

A. Signal Model

A MIMO link with M receiving and N transmitting antennas over a frequency-selective fading channel constitutes the scenario of interest of this paper. Letting (m, n) be the link between the n th transmitting and the m th receiving antennas ($m = 1, \dots, M$ and $n = 1, \dots, N$), the baseband model of the transmitted data-sequences within the k th burst (denoted as subscript) $x_k^{(n)}[i]$ at symbol rate $1/T$ is

$$y_k^{(m,n)}(t) = \sum_i x_k^{(n)}[i] h_k^{(m,n)}(t - iT). \quad (1)$$

The multipath channel $h_k^{(m,n)}(t)$ is time-varying. According to the block-fading assumption, the time-variations are sufficiently slow to guarantee that the assumption of a static channel within each burst (and generally varying from burst to burst) is reasonably satisfied. The m th antenna receives the combination of the simultaneous transmissions from the N transmitting antennas so that $y_k^{(m)}(t) = \sum_{n=1}^N y_k^{(m,n)}(t) + n_k^{(m)}(t)$, where $n_k^{(m)}(t)$ is the additive Gaussian noise (AGN). The M received signals are arranged into the $M \times 1$ vector $\mathbf{y}_k(t) = [y_k^{(1)}(t) \dots y_k^{(M)}(t)]^T$, and the model (1) becomes

$$\mathbf{y}_k(t) = \sum_i \mathbf{H}_k(t - iT) \mathbf{x}_k[i] + \mathbf{n}_k(t) \quad (2)$$

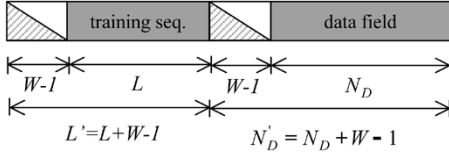


Fig. 2. Burst structure.

where $\mathbf{x}_k[i] = [x_k^{(1)}[i] \cdots x_k^{(N)}[i]]^T$ is $N \times 1$, and $\mathbf{H}_k(t)$ is the MIMO impulse response as $[\mathbf{H}_k(t)]_{m,n} = h_k^{(m,n)}(t)$.

After sampling at symbol-rate $1/T$, the discrete-time channel $h_k^{(m,n)}[i] = h_k^{(m,n)}(iT)$ is assumed causal and temporally supported over W samples: $h_k^{(m,n)}[i] = 0$ for $i \notin [0, W - 1]$. Each burst contains a data field and a training sequence (Fig. 2). Here, we tailor the model (2) for the estimation of the MIMO-FIR channel $\{\mathbf{H}_k[i]\}_{i=0}^{W-1}$ from periodic training sequences simultaneously transmitted by each of the N transmitting antennas during the training period.

Letting $\{x^{(n)}[\ell]\}_{\ell=0}^{L-1}$ be the L sample long training sequences ($n = 1, \dots, N$) that are independent on the burst, the discrete-time model is

$$\mathbf{y}_k[\ell] = \sum_{i=0}^{W-1} \mathbf{H}_k[i] \mathbf{x}[\ell - i] + \mathbf{n}_k[\ell]. \quad (3)$$

AGN $\mathbf{n}_k[\ell]$ is assumed temporally uncorrelated but spatially correlated with covariance matrix $\mathbf{R}_n : E[\mathbf{n}_k[\ell] \mathbf{n}_k[\ell - i]^H] = \mathbf{R}_n \delta[i]$. In order to avoid interference from the data-block, the training sequences contain a preamble of $W - 1$ samples (or, equivalently, are supported over $L' = W + L - 1$ samples) that is discarded at receiver to yield L received samples free from interference. Therefore, the received L samples of the training sequences can be arranged into the $M \times L$ matrix $\mathbf{Y}_k = [\mathbf{y}_k[0], \dots, \mathbf{y}_k[L - 1]]$, leading to

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{X} + \mathbf{N}_k \quad (4)$$

where $\mathbf{H}_k = [\mathbf{H}_k[0], \mathbf{H}_k[1], \dots, \mathbf{H}_k[W - 1]]$ is the $M \times NW$ MIMO-FIR channel matrix, \mathbf{X} is the $NW \times L$ convolution matrix obtained from the N training sequences so that the i th column of \mathbf{X} is $[\mathbf{x}[i]^T, \mathbf{x}[i - 1]^T, \dots, \mathbf{x}[i - W + 1]^T]^T$, \mathbf{N}_k has the same structure as \mathbf{Y}_k , and $1/L \cdot E[\mathbf{N}_k \mathbf{N}_k^H] = \mathbf{R}_n$. The spatial (among the transmitting antennas) and temporal correlation of the training sequences is given by the matrix $\mathbf{R}_x = \mathbf{X}^* \mathbf{X}^T$ (see the next section).

For channel estimation, it is convenient to arrange the L columns of the matrix \mathbf{Y}_k into a $ML \times 1$ vector $\mathbf{y}_k = \text{vec}\{\mathbf{Y}_k\}$, and according to the properties of $\text{vec}\{\cdot\}$, the model (4) becomes

$$\mathbf{y}_k = (\mathbf{X}^T \otimes \mathbf{I}_M) \mathbf{h}_k + \mathbf{n}_k = \tilde{\mathbf{X}} \mathbf{h}_k + \mathbf{n}_k. \quad (5)$$

The columns of \mathbf{H}_k are similarly stacked into a $MNW \times 1$ vector $\mathbf{h}_k = \text{vec}\{\mathbf{H}_k\}$. The spatial covariance matrix for the AGN $\mathbf{n}_k = \text{vec}\{\mathbf{N}_k\}$ is $E[\mathbf{n}_k \mathbf{n}_k^H] = \mathbf{I}_L \otimes \mathbf{R}_n$, and the training sequence matrix is now $\tilde{\mathbf{X}} = \mathbf{X}^T \otimes \mathbf{I}_M$.

B. Problem Definition

According to the propagation model described below (Section III), the channel vector \mathbf{h}_k can be parametrized as the

product of a stationary (or slowly varying) matrix \mathbf{T} and the fast-varying fading vector $\boldsymbol{\beta}_k$

$$\mathbf{h}_k = \mathbf{T} \boldsymbol{\beta}_k. \quad (6)$$

The d paths (or clusters) model \mathbf{T} is $MNW \times d$, and $\boldsymbol{\beta}_k$ is $d \times 1$. Matrix \mathbf{T} is stationary across a large number of bursts (say K) as it depends on the DODs, DOAs, delays, and power profile of fading. Instead of estimating the MNW entries of \mathbf{h}_k for each burst, methods based on the parameterization (6) estimate the matrix \mathbf{T} from $K \gg 1$ bursts, whereas the fading $\boldsymbol{\beta}_k$ is estimated on a burst-by-burst basis or tracked across bursts.

Different paths having the similar propagation parameters do not always contribute to the matrix \mathbf{T} with linearly independent columns (see [9] for a discussion in the context of SIMO systems). In fact, when the separation between delays (and/or angles) is below the temporal (and spatial resolution), the matrix \mathbf{T} is rank-deficient and, in practice, $r = \text{rank}\{\mathbf{T}\} \ll \min\{MNW, d\}$. Therefore, matrix \mathbf{T} is not identifiable. In order to write the model (6) in terms of an equivalent minimal and identifiable parametrization, it is thus necessary to introduce a $MNW \times r$ full rank matrix \mathbf{U} such that $\text{span}\{\mathbf{T}\} = \text{span}\{\mathbf{U}\}$. A convenient choice is to select \mathbf{U} as an orthonormal matrix, i.e., in terms of the singular value decomposition (SVD) of \mathbf{T} , $\mathbf{T} = \mathbf{U} \boldsymbol{\Lambda}^{1/2} \mathbf{V}^H$, so that the model (6) can be restated as

$$\mathbf{h}_k = \mathbf{U} \mathbf{d}_k \quad (7)$$

where the $r \times 1$ vector that contains the newly defined fading amplitudes is

$$\mathbf{d}_k = \boldsymbol{\Lambda}^{1/2} \mathbf{V}^H \boldsymbol{\beta}_k. \quad (8)$$

The equivalent parametrization (7) provides the rationale for efficient subspace-based channel estimators (Section V) and allows the computation of the lower bound that depends on the specific channel models (see Section III).

Let $\hat{\mathbf{h}}_k$ be any estimate of the MIMO channel. The purpose is to evaluate a bound $\mathbf{Q}_{\hat{\mathbf{h}}}$ on the error correlation matrix

$$\mathbf{Q}_{\hat{\mathbf{h}}} = E[(\hat{\mathbf{h}}_k - \mathbf{h}_k)(\hat{\mathbf{h}}_k - \mathbf{h}_k)^H] \geq \mathbf{Q}_{\hat{\mathbf{h}}} \quad (9)$$

for Rayleigh fading channels [expectation in (9) is w.r.t. noise and fading]. The corresponding MSE bound $\text{MSE}_{\hat{\mathbf{h}}}$ is

$$\text{MSE}_{\hat{\mathbf{h}}} = \text{tr}\{\mathbf{Q}_{\hat{\mathbf{h}}}\} \geq \text{MSE}_{\hat{\mathbf{h}}} = \text{tr}\{\mathbf{Q}_{\hat{\mathbf{h}}}\}. \quad (10)$$

A bound on the error correlation matrix will be obtained by computing the asymptotic performance of a channel estimator that fully exploits the structure (7). As a reference to conventional estimators, we recall that the unconstrained maximum likelihood (UML) estimate of the channel vector for $L \geq MNW$ is $\hat{\mathbf{h}}_{\text{UML},k} = \tilde{\mathbf{X}}^\dagger \mathbf{y}_k$ [5], [19]. The corresponding error correlation matrix

$$\mathbf{Q}_{\text{UML}} = E[(\hat{\mathbf{h}}_{\text{UML},k} - \mathbf{h}_{\text{UML},k})(\hat{\mathbf{h}}_{\text{UML},k} - \mathbf{h}_{\text{UML},k})^H] = \mathbf{R}_x^{-1} \otimes \mathbf{R}_n \quad (11)$$

depends on the covariance of AGN \mathbf{R}_n and on the correlation properties of the training sequences.

III. CHANNEL MODELS

In a multipath environment, the (m, n) link can be described by a combination of d delayed replica of the waveform $g(t)$ given by the convolution of the transmitted baseband pulse and the receiving filter

$$h_k^{(m,n)}(t) = \sum_{i=1}^d \sqrt{\Omega_{i,k}} \cdot a_{i,k}^{(m,n)} \cdot g(t - \tau_{i,k}) \quad (12)$$

where each path is characterized by delay $\tau_{i,k}$, power $\Omega_{i,k}$, and amplitude $a_{i,k}^{(m,n)}$. The MIMO channel impulse response obtained by arranging the $M \times N$ channels is

$$\mathbf{H}_k(t) = \sum_{i=1}^d \sqrt{\Omega_{i,k}} \cdot g(t - \tau_{i,k}) \mathbf{A}_{i,k} \quad (13)$$

where $[\mathbf{A}_{i,k}]_{m,n} = a_{i,k}^{(m,n)}$. After sampling and ordering the W matrices $\mathbf{H}_k[\ell] = \mathbf{H}_k(\ell T)$ into the $M \times NW$ MIMO-FIR channel matrix \mathbf{H}_k , we get

$$\mathbf{H}_k = \sum_{i=1}^d \sqrt{\Omega_{i,k}} \cdot \mathbf{g}(\tau_{i,k})^T \otimes \mathbf{A}_{i,k} \quad (14)$$

where $\mathbf{g}(\tau)$ is the $W \times 1$ vector that gathers the T -spaced samples of the delayed waveform $g(t - \tau)$. Appropriate definitions of the matrices $\mathbf{A}_{i,k}$ generalize the MIMO channel model (14) to beamforming and diversity models, as described below.

The channel matrix \mathbf{H}_k in (14) can be rearranged to separate the stationary and fast-varying terms by stacking the columns of the $M \times N$ channel taps $\mathbf{H}_k[\ell]$ (i.e., $\bar{\mathbf{H}}_k = [\mathbf{h}_k[0], \mathbf{h}_k[2], \dots, \mathbf{h}_k[W-1]]$, where $\mathbf{h}_k[\ell] = \text{vec}\{\mathbf{H}_k[\ell]\}$):

$$\begin{aligned} \bar{\mathbf{H}}_k &= \sum_{i=1}^d \sqrt{\Omega_{i,k}} \text{vec}\{\mathbf{A}_{i,k}\} \mathbf{g}(\tau_{i,k})^T \\ &= \mathbf{A}_k \cdot \Omega_k^{1/2} \cdot \mathbf{G}(\boldsymbol{\tau}_k)^T \end{aligned} \quad (15)$$

where $\mathbf{A}_k = [\text{vec}\{\mathbf{A}_{1,k}\}, \text{vec}\{\mathbf{A}_{2,k}\}, \dots, \text{vec}\{\mathbf{A}_{d,k}\}]$ is $MN \times d$, $\Omega_k^{1/2} = \text{diag}\{\sqrt{\Omega_{1,k}}, \sqrt{\Omega_{2,k}}, \dots, \sqrt{\Omega_{d,k}}\}$, and the $W \times d$ matrix $\mathbf{G}(\boldsymbol{\tau}_k) = [\mathbf{g}(\tau_{1,k}), \dots, \mathbf{g}(\tau_{d,k})]$ collects all the delayed waveforms. According to the model (5), the $MNW \times 1$ channel vector \mathbf{h}_k can now be obtained as $\mathbf{h}_k = \text{vec}\{\mathbf{H}_k\} = \text{vec}\{\bar{\mathbf{H}}_k\}$. Factorization (6) can be derived by considering that delays $\boldsymbol{\tau}_k = [\tau_{1,k}, \dots, \tau_{d,k}]^T$ and power Ω_k are independent on the bursts so that $\boldsymbol{\tau}_k = \boldsymbol{\tau}$ and $\Omega_k = \Omega$ for $k \in [1, K]$. Scattering corresponding to different delays can be assumed to be uncorrelated [20], and the amplitudes $\mathbf{A}_{i,k}$ are varying from burst to burst according to the mobility of the ends of the link.

A. Beamforming Model

According to the geometry depicted in Fig. 1 with antennas closely spaced apart ($\lambda/2$ or smaller), the beamforming model for narrowband signals describes the i th path ($i = 1, \dots, d$) as characterized by a DOD $\alpha_i^{(T)}$, a DOA $\alpha_i^{(R)}$, and a complex amplitude (fading) $\beta_{i,k}$ so that

$$\mathbf{A}_{i,k} = \beta_{i,k} \mathbf{a}_R(\alpha_i^{(R)}) \mathbf{a}_T(\alpha_i^{(T)})^T \quad (16)$$

where $\mathbf{a}_T(\alpha)$ (or $\mathbf{a}_R(\alpha)$) is the $N \times 1$ (or $M \times 1$) vector containing the array response to a plane wave transmitted (or received) with the angle α . Notice that the two arrays need not have the same arrangement of sensors.

The DODs and DOAs in (16) are considered to be slowly varying features of the channel in that even though the two ends of the link are moving (translating) during K bursts, the variations of DOAs and DODs are smaller than the angular resolution of the arrays [8], [14]. This implies that the vectors $\boldsymbol{\alpha}^{(T)} = [\alpha_1^{(T)}, \dots, \alpha_d^{(T)}]^T$ and $\boldsymbol{\alpha}^{(R)} = [\alpha_1^{(R)}, \dots, \alpha_d^{(R)}]^T$ are stationary. The fading amplitudes $\beta_{i,k}$ are uncorrelated zero mean circularly symmetric normal (Rayleigh fading) with unit power: $\boldsymbol{\beta}_k = [\beta_{1,k}, \dots, \beta_{d,k}]^T \sim \mathcal{CN}(0, \mathbf{I}_d)$. Their temporal correlation across different bursts is the same for all the paths, and it accounts for the mobility of the terminals $E[\beta_{i,k}^* \beta_{i,k+n}] = \varphi_n$ with $\varphi_o = 1$ for $\forall i = 1, \dots, d$. This assumption is mainly for mathematical convenience since a more realistic model would allow the Doppler spectrum to be dependent on the delay [21]. Recall that the power-delay profile is accounted for by the burst-independent matrix Ω . We emphasize that the stationary parameters (i.e., angles, delays, and power-delay profile) are assumed to be deterministic, whereas the fading amplitudes are random variables.

According to (16), the matrix \mathbf{A}_k in (15) can be further factorized into the burst-independent term $\mathcal{A}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)})$ that contains the spatial signatures of the d stationary angles $\boldsymbol{\alpha}^{(T)}$ and $\boldsymbol{\alpha}^{(R)}$ and the $d \times 1$ burst-varying random vector $\boldsymbol{\beta}_k$, where we have (17), shown at the bottom of the page. The $MNW \times 1$ channel vector \mathbf{h}_k reduces now to model (6), where each column of the stationary matrix \mathbf{T} contains the space-time signature of one path:

$$\begin{aligned} \mathbf{T} &= \left[\Omega_1^{1/2} \mathbf{g}(\tau_1) \otimes \mathbf{a}_T(\alpha_1^{(T)}) \otimes \mathbf{a}_R(\alpha_1^{(R)}), \dots \right. \\ &\quad \left. \Omega_d^{1/2} \mathbf{g}(\tau_d) \otimes \mathbf{a}_T(\alpha_d^{(T)}) \otimes \mathbf{a}_R(\alpha_d^{(R)}) \right] \\ &= \left(\mathbf{G}(\boldsymbol{\tau}) \diamond \mathcal{A}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)}) \right) \cdot \Omega^{1/2} \end{aligned} \quad (18)$$

where \diamond denotes the Khatri-Rao (or columnwise Kronecker product). The $d \times 1$ faded amplitudes $\boldsymbol{\beta}_k$ are correlated across time as $\mathbf{R}_{\boldsymbol{\beta}}(n) = E[\boldsymbol{\beta}_k \boldsymbol{\beta}_{k-n}^H] = \mathbf{R}_{\boldsymbol{\beta}}(0) \varphi_n = \varphi_n \mathbf{I}_d$. The

$$\mathbf{A}_k = \underbrace{\left[\mathbf{a}_T(\alpha_1^{(T)}) \otimes \mathbf{a}_R(\alpha_1^{(R)}), \dots, \mathbf{a}_T(\alpha_d^{(T)}) \otimes \mathbf{a}_R(\alpha_d^{(R)}) \right]}_{\mathcal{A}'(\boldsymbol{\alpha}^{(T)}, \boldsymbol{\alpha}^{(R)})} \cdot \text{diag}\{\boldsymbol{\beta}_k\}. \quad (17)$$

amplitudes of the channel model (7) are still randomly varying with correlation

$$\mathbf{R}_d(n) = E[\mathbf{d}_k \mathbf{d}_{k-n}^H] = \mathbf{\Lambda} \varphi_n \quad (19)$$

where $\mathbf{\Lambda} = \text{diag}\{\mu_1[\mathbf{T}\mathbf{T}^H], \mu_2[\mathbf{T}\mathbf{T}^H], \dots, \mu_r[\mathbf{T}\mathbf{T}^H]\}$. Since

$$\mathbf{T}\mathbf{T}^H = \sum_{i=1}^d \Omega_i \left(\mathbf{g}(\tau_i) \mathbf{g}(\tau_i)^H \otimes \mathbf{a}_T(\alpha_i^{(T)}) \mathbf{a}_T(\alpha_i^{(T)})^H \right. \\ \left. \otimes \mathbf{a}_R(\alpha_i^{(R)}) \mathbf{a}_R(\alpha_i^{(R)})^H \right) \quad (20)$$

these eigenvalues represent the power-delay/angles profile that accounts for the power on each vector of the basis \mathbf{U} .

B. Diversity Model

In case the elements of the antenna arrays are sufficiently far apart and/or the propagation occurs in a rich scattering environment with large angular spread, the model (16) is no longer appropriate. To simplify and referring to Fig. 1, let p paths be grouped into d subsets (or clusters) of p_1, p_2, \dots, p_d paths each such that $p = \sum_{i=1}^d p_i$. Each path within the i th cluster is characterized by the same delay τ_i but different amplitudes with overall power Ω_i . By assuming the stationarity of $\mathbf{\Omega}$ and $\boldsymbol{\tau}$, for d clusters, the model becomes

$$\mathbf{H}_k = \sum_{i=1}^d \sqrt{\Omega_i} \cdot \mathbf{g}(\tau_i)^T \otimes \sum_{\ell=1}^{p_i} \mathbf{a}_{i,k}[\ell] \mathbf{b}_{i,k}^T[\ell] \quad (21)$$

where $\mathbf{a}_{i,k}[\ell]$ and $\mathbf{b}_{i,k}[\ell]$ contain the fading gains between different receiving and transmitting antennas, respectively. The model (21) can be reduced to (14) once the $M \times N$ random matrix $\mathbf{A}_{i,k} = \sum_{\ell=1}^{p_i} \mathbf{a}_{i,k}[\ell] \mathbf{b}_{i,k}^T[\ell]$ is defined for each cluster. Notice that for the beamforming model, $p_i = 1$. Here, we consider p_i very large ($p_i \rightarrow \infty$) and assume that scattering from different paths is uncorrelated so that according to the central limit, $\text{vec}\{\mathbf{A}_{i,k}\} \sim \mathcal{CN}(0, \mathbf{R}_i)$, where the $MN \times MN$ matrix \mathbf{R}_i accounts for the correlation of fading among the transmitting and receiving antennas. This is normalized so that along the diagonal, $[\mathbf{R}_i]_{mm} = 1$. Even if this is not necessary for our analysis, the fading correlation can be considered to be separate as $\mathbf{R}_i = \mathbf{R}_{T,i} \otimes \mathbf{R}_{R,i}$, where $\mathbf{R}_{T,i}$ and $\mathbf{R}_{R,i}$ are the correlation matrix of the fading at the transmitting and receiving antennas, respectively. The separability of the spatial correlation matrix is an assumption that is widely accepted in the literature in order to simplify the analysis (see, e.g., [4]) or to have a simple model to match the measurements [22]. For a discussion on the limits of the separable model for the spatial correlation based on simple geometrical reasoning, see [23].

Vectorization of (21) reduces the channel vector \mathbf{h}_k to (6) with

$$\mathbf{T} = \mathbf{G}(\boldsymbol{\tau}) \mathbf{\Omega}^{1/2} \otimes \mathbf{I}_{MN} \quad (22)$$

where \mathbf{T} is the $MNW \times MNd$ matrix obtained from the $W \times d$ stationary matrix $\mathbf{G}(\boldsymbol{\tau}) \mathbf{\Omega}^{1/2}$. The burst-varying fading $MNd \times 1$ is $\boldsymbol{\beta}_k = \text{vec}\{\mathbf{A}_k\} = \text{vec}\{[\mathbf{A}_{1,k}, \mathbf{A}_{2,k}, \dots, \mathbf{A}_{d,k}]\}$. These

amplitudes are spatially (i.e., among sensors) correlated, and the correlation is

$$\mathbf{R}_\beta(n) = \mathbf{R}_\beta \varphi_n \quad (23)$$

with $\mathbf{R}_\beta = \text{diag}\{\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_d\}$ block-diagonal, and φ_n accounts for the temporal correlation. A simplification arises in (23) when spatial correlation is the same for each cluster (i.e., $\mathbf{R}_i = \mathbf{R}$), as

$$\mathbf{R}_\beta = \mathbf{I}_d \otimes \mathbf{R}. \quad (24)$$

If \mathbf{R} can be further assumed to be separable, it is $\mathbf{R}_\beta = \mathbf{I}_d \otimes \mathbf{R}_T \otimes \mathbf{R}_R$.

Temporal signatures $\Omega_i^{1/2} \mathbf{g}(\tau_i)$ in \mathbf{T} are not linearly independent when cluster delays are closely spaced, compared with the temporal resolution of the system [9]. Letting $r_g = \text{rank}\{\mathbf{G}(\boldsymbol{\tau}) \mathbf{\Omega}^{1/2}\} \leq \min\{W, d\}$ be the order of temporal resolution of the channel and taking the SVD of $\mathbf{G}(\boldsymbol{\tau}) \mathbf{\Omega}^{1/2} = \mathbf{U}_g \mathbf{\Lambda}_g^{1/2} \mathbf{V}_g^H$, we get the equivalent model (7) with

$$\mathbf{U} = \mathbf{U}_g \otimes \mathbf{I}_{MN}. \quad (25)$$

The $r = MNr_g$ amplitudes in (8) are $\mathbf{d}_k = (\mathbf{\Lambda}_g^{1/2} \mathbf{V}_g^H \otimes \mathbf{I}_{MN}) \boldsymbol{\beta}_k$, and their correlation depends on the eigenvalues of $\mathbf{G} \mathbf{\Omega} \mathbf{G}^H = \sum_{i=1}^d \Omega_i \mathbf{g}(\tau_i) \mathbf{g}(\tau_i)^H$ as

$$\sum_{i=1}^d \Omega_i \mathbf{g}(\tau_i) \mathbf{g}(\tau_i)^H = \mathbf{U}_g \mathbf{\Lambda}_g \mathbf{U}_g^H. \quad (26)$$

In general, the correlation of the amplitudes \mathbf{d}_k is

$$\mathbf{R}_d(n) = \left(\mathbf{\Lambda}_g^{1/2} \mathbf{V}_g^H \otimes \mathbf{I}_{MN} \right) \mathbf{R}_\beta \left(\mathbf{V}_g \mathbf{\Lambda}_g^{1/2} \otimes \mathbf{I}_{MN} \right) \varphi_n \quad (27)$$

that simplifies when the spatial correlation is independent on the path, as in (24):

$$\mathbf{R}_d(n) = (\mathbf{\Lambda}_g \otimes \mathbf{R}) \varphi_n. \quad (28)$$

Remark: Mixed beamforming/diversity models for MIMO links that present asymmetric geometry of arrays and/or scatters at their ends (see, e.g., [24]) can be derived by generalizing the discussion above. This aspect will not be covered here.

IV. LOWER BOUND ON THE ERROR CORRELATION MATRIX OF THE CHANNEL ESTIMATE

The previous section justified the parameterization of the MIMO channel vector \mathbf{h}_k into a stationary (or slowly varying) and a fast-varying term, here rewritten from (6) and (7):

$$\mathbf{h}_k = \mathbf{T} \boldsymbol{\beta}_k = \mathbf{U} \mathbf{d}_k. \quad (29)$$

The stationary features of the channel are given by the basis \mathbf{U} along with the statistics of noise (\mathbf{R}_n) and fading amplitudes ($\mathbf{R}_d(n)$). According to Section V, the matrix \mathbf{U} can be estimated in a structured (i.e., in term of angles, delays, and power-delay profile) or unstructured way (i.e., by estimating \mathbf{U} or \mathbf{U}_g for the diversity model). However, for the purpose of obtaining the MSE

bound, it is enough to consider a consistent (but not necessarily efficient) estimator of \mathbf{U} —be it structured or unstructured—so that \mathbf{U} can be considered to be known for $K \rightarrow \infty$. Moreover, the correlation of noise \mathbf{R}_n and fading amplitudes $\mathbf{R}_d(n)$ are assumed to be reliably estimated for $K \rightarrow \infty$. Similar assumptions have been exploited to evaluate the performance of a multichannel MLSE with channel estimation [25].

The signal model (5) can be restated in terms of the known matrix $\mathbf{F} = \bar{\mathbf{X}}\mathbf{U}$ as

$$\mathbf{y}_k = \mathbf{F}\mathbf{d}_k + \mathbf{n}_k. \quad (30)$$

The optimum (MMSE) channel estimator for $K \rightarrow \infty$ reduces to the MMSE estimation of the amplitudes \mathbf{d}_k . According to the (theoretically) infinite temporal horizon of our framework, we consider a Wiener filter that estimates the amplitudes in the frequency domain

$$\mathcal{F}\{\hat{\mathbf{d}}_k\} = \mathbf{S}_{dy}(\omega)\mathbf{S}_{yy}(\omega)^{-1}\mathcal{F}\{\mathbf{y}_k\} \quad (31)$$

where $\mathbf{S}_{dy}(\omega) = \mathcal{F}\{E[\mathbf{d}_k\mathbf{y}_{k-n}^H]\}$ denotes the discrete-time Fourier transform of the crosscorrelation matrix between $\{\mathbf{d}_k\}$ and $\{\mathbf{y}_k\}$, $\mathbf{S}_{yy}(\omega)$, is similarly defined. Since the spatio-temporal correlation of the fading is separable [i.e., $\mathbf{R}_d(n) = \mathbf{R}_d\varphi_n$ as in (19) or (28)], the MMSE estimate (31) depends on the power spectral density of the fading variations $S_\varphi(\omega) = \mathcal{F}\{\varphi_n\}$ as

$$\mathbf{S}_{dy}(\omega)\mathbf{S}_{yy}(\omega)^{-1} = S_\varphi(\omega)\mathbf{R}_d\mathbf{F}^H \times (S_\varphi(\omega)\mathbf{F}\mathbf{R}_d\mathbf{F}^H + (\mathbf{I}_L \otimes \mathbf{R}_n))^{-1} \quad (32)$$

where we used the equalities $\mathbf{S}_{dy}(\omega) = \mathbf{S}_{dd}(\omega)\mathbf{F}^H$, $\mathbf{S}_{yy}(\omega) = \mathbf{F}\mathbf{S}_{dd}(\omega)\mathbf{F}^H + (\mathbf{I}_L \otimes \mathbf{R}_n)$ and $\mathbf{S}_{dd}(\omega) = \mathbf{R}_d S_\varphi(\omega)$. The error correlation matrix depends on the estimate of the amplitudes \mathbf{d}_k , as $\mathbf{Q}_{\hat{\mathbf{h}}} = \mathbf{U}\mathbf{Q}_{\hat{\mathbf{d}}}\mathbf{U}^H$, where $\mathbf{Q}_{\hat{\mathbf{d}}} = E[(\hat{\mathbf{d}}_k - \mathbf{d}_k)(\hat{\mathbf{d}}_k - \mathbf{d}_k)^H]$. In the frequency domain, the error correlation matrix of the amplitudes reads

$$\begin{aligned} \mathbf{S}_{ee}(\omega) &= \mathcal{F}\{E[(\hat{\mathbf{d}}_k - \mathbf{d}_k)(\hat{\mathbf{d}}_{k-n} - \mathbf{d}_{k-n})^H]\} \\ &= \mathbf{S}_{dd}(\omega) - \mathbf{S}_{dy}(\omega)\mathbf{S}_{yy}(\omega)^{-1}\mathbf{S}_{yd}(\omega) \\ &= S_\varphi(\omega) (\mathbf{R}_d^{-1} + S_\varphi(\omega)\mathbf{F}^H(\mathbf{I}_L \otimes \mathbf{R}_n)\mathbf{F}) \\ &= S_\varphi(\omega) (\mathbf{R}_d^{-1} + S_\varphi(\omega)\mathbf{R}_w)^{-1} \end{aligned} \quad (33)$$

and for the second equality, we used the matrix inversion lemma,¹ and

$$\mathbf{R}_w = \mathbf{U}^H (\mathbf{R}_x \otimes \mathbf{R}_n^{-1}) \mathbf{U}. \quad (34)$$

Notice that \mathbf{R}_w can be written in terms of the error correlation matrix of the UML estimate (11) as $\mathbf{R}_w = \mathbf{U}^H\mathbf{Q}_{\text{UML}}^{-1}\mathbf{U}$ if $L \geq NW$.

By using the Parseval theorem, we have

$$\begin{aligned} \mathbf{Q}_{\hat{\mathbf{h}}} &= \mathbf{U} \int_{-\pi}^{\pi} \mathbf{S}_{ee}(\omega) \frac{d\omega}{2\pi} \mathbf{U}^H \\ &= \mathbf{U} \int_{-\pi}^{\pi} S_\varphi(\omega) (\mathbf{R}_d^{-1} + S_\varphi(\omega)\mathbf{R}_w)^{-1} \frac{d\omega}{2\pi} \mathbf{U}^H. \end{aligned} \quad (35)$$

¹ $(\mathbf{A}^{-1} + \mathbf{V}\mathbf{C}^{-1}\mathbf{V}^H)^{-1} = \mathbf{A} - \mathbf{A}\mathbf{V}(\mathbf{C} + \mathbf{V}^H\mathbf{A}\mathbf{V})^{-1}\mathbf{V}^H\mathbf{A}^H$

Computation of the bound (35) requires integration over the Doppler spectrum $S_\varphi(\omega)$. In order to ease the analysis and allow us to gain insight into the bound (35) in the following, we consider a uniform Doppler spectrum.

Remark: The bound (35) has been computed under the asymptotic condition $K \rightarrow \infty$. Whenever it is of interest to have the bound for any K , the analytical derivation can not be simplified as herein, but it has to be evaluated from the hybrid Cramér–Rao bound [10]. It can be shown [32] that the hybrid Cramér–Rao bound coincides with the lower bound (35) for $K \rightarrow \infty$.

A. Lower Bound for a Uniform Doppler Spectrum

The error correlation matrix bound $\mathbf{Q}_{\hat{\mathbf{h}}}$ can be easily evaluated in closed form for a uniform Doppler spectrum $S_\varphi(\omega) = 1/2f_D$ over the support $\omega \in [-2\pi f_D, +2\pi f_D]$ ($0 \leq f_D \leq 1/2$). In this case, the bound (35) simplifies as

$$\mathbf{Q}_{\hat{\mathbf{h}}} = 2f_D \cdot \mathbf{U} (2f_D\mathbf{R}_d^{-1} + \mathbf{R}_w)^{-1} \mathbf{U}^H \quad (36)$$

and the corresponding MSE (10)

$$\begin{aligned} \text{MSE}_{\hat{\mathbf{h}}} &= 2f_D \cdot \text{tr} \left\{ (2f_D\mathbf{R}_d^{-1} + \mathbf{R}_w)^{-1} \right\} \\ &= 2f_D \cdot \sum_{i=1}^r \frac{1}{\mu_i [2f_D\mathbf{R}_d^{-1} + \mathbf{R}_w]} \end{aligned} \quad (37)$$

depends on the r eigenvalues of the $r \times r$ matrix $2f_D\mathbf{R}_d^{-1} + \mathbf{R}_w$.

The bound of (36) and (37) generalizes some known results on the performance of MMSE or ML channel estimation, where some of these connections are discussed below under some simplifying assumptions (recall that $K \rightarrow \infty$).

1) *Static Channel* ($f_D = 0$): In a static channel, fading is not varying across blocks so that $f_D = 0$ and from (36), it is $\mathbf{Q}_{\hat{\mathbf{h}}} = \mathbf{0}$. Indeed, in this case, the channel vector is constant and can be consistently estimated with covariance $O(1/K)$ by just averaging the UML estimates $\hat{\mathbf{h}}_{\text{UML},k}$ obtained from the different slots.

2) *Ideal Training and Uncorrelated Noise* ($\mathbf{R}_x = L\sigma_x^2\mathbf{I}_{NW}$ and $\mathbf{R}_n = \sigma^2\mathbf{I}_M$): Let us consider spatially white noise ($\mathbf{R}_n = \sigma^2\mathbf{I}_M$) and ideal training sequences (i.e., orthogonal between any two transmitting antennas that are temporally uncorrelated: $\mathbf{R}_x = L\sigma_x^2\mathbf{I}_{NW}$); in this case, $\mathbf{R}_w = L\sigma_x^2/\sigma^2\mathbf{I}_r$, and the MSE can be evaluated for the two models from $\mu_i[2f_D\mathbf{R}_d^{-1} + \mathbf{R}_w]$.

Beamforming Model: The eigenvalues for the beamforming model can be evaluated as $\mu_i[2f_D\mathbf{R}_d^{-1} + \mathbf{R}_w] = L\sigma_x^2/\sigma^2 + 2f_D/\mu_i[\mathbf{T}\mathbf{T}^H]$, and the MSE (37) is

$$\text{MSE}_{\hat{\mathbf{h}}} = 2f_D \cdot \sum_{i=1}^r \frac{\mu_i[\mathbf{T}\mathbf{T}^H]}{2f_D + \mu_i[\mathbf{T}\mathbf{T}^H] \frac{L\sigma_x^2}{\sigma^2}}. \quad (38)$$

For low SNR (or $L\sigma_x^2/\sigma^2 \ll 2f_D/\mu_i[\mathbf{T}\mathbf{T}^H] \leq 1/\mu_i[\mathbf{T}\mathbf{T}^H]$), the MSE (38) is

$$\text{MSE}_{\hat{\mathbf{h}}} \simeq \sum_{i=1}^r \mu_i[\mathbf{T}\mathbf{T}^H] = MN \sum_{j=1}^d \Omega_j \|g(\tau_j)\|^2 \quad (39)$$

for high SNR or small Doppler frequency (i.e., $L\sigma_x^2/\sigma^2 \gg 2f_D/\mu_i[\mathbf{T}\mathbf{T}^H]$), the MSE

$$\text{MSE}_{\hat{\mathbf{h}}} \simeq 2f_D \frac{\sigma^2}{L\sigma_x^2} r \quad (40)$$

is proportional to $r = \text{rank}\{\mathbf{T}\}$. Notice that r is the number of parameters (amplitudes) to be varying on a burst-by-burst basis according to the minimal model (7). If terminals are moving fast enough to let the fading amplitudes be temporally uncorrelated across bursts ($f_D \simeq 1/2$) but the block-fading assumption and stationarity of \mathbf{T} hold true, the MSE (40) for $M = 1$ (SIMO system) coincides with the MSE bound on the channel estimation error derived in [26] for the multislot estimator (see Section V).

Diversity Model: According to the model (28), the $r = MNr_g$ eigenvalues $\{\mu_i[\mathbf{R}_d]\}$ depend on $\{\mu_i[\mathbf{G}\Omega\mathbf{G}^H]\}_{i=1}^{r_g}$ and $\{\mu_i[\mathbf{R}]\}_{i=1}^{MN}$, and if spatial correlation of MIMO channel is separable $\mathbf{R} = \mathbf{R}_T \otimes \mathbf{R}_R$, the set $\{\mu_i[\mathbf{R}]\}_{i=1}^{MN}$ depends on $\{\mu_i[\mathbf{R}_T]\}_{i=1}^N$ and $\{\mu_i[\mathbf{R}_R]\}_{i=1}^M$ according to the properties of Kronecker products. The MSE bound (37) becomes (41), shown at the bottom of the page. In spite of its complexity, bound (41) can be simplified in some useful cases. For low SNR (i.e., $L\sigma_x^2/\sigma^2 \ll 2f_D/\mu_i[\mathbf{G}\Omega\mathbf{G}^H]\mu_n[\mathbf{R}_T]\mu_m[\mathbf{R}_R]$), the MSE (41) coincides with (39), which has been derived for the beamforming model. For high SNR, it is

$$\text{MSE}_{\hat{\mathbf{h}}} \simeq 2f_D \frac{\sigma^2}{L\sigma_x^2} MNr_g \quad (42)$$

thus showing that for high SNR, the MSE bound is proportional to the minimal number of fading amplitudes \mathbf{d}_k of size $MNr_g \times 1$ that are varying on each burst. Furthermore, the MSE bound (42) is upper bounded by the worst case of uncorrelated fading ($f_D = 1/2$).

The diversity model for the frequency-flat fading channel is commonly used in the MIMO literature. This occurs if the delays are not temporally resolvable (compared with the system bandwidth) so that $d = 1$ and $W = 1$. In this case, $\mathbf{U} = \mathbf{I}_{MN}$, and the MSE bounds (36) and (37) depend on the spatial properties of fading as

$$\mathbf{Q}_{\hat{\mathbf{h}}} = 2f_D \cdot \left(\frac{2f_D}{\Omega} \mathbf{R}^{-1} + \frac{L\sigma_x^2}{\sigma^2} \mathbf{I}_{MN} \right)^{-1} \quad (43a)$$

$$\text{MSE}_{\hat{\mathbf{h}}} = 2f_D \sum_{i=1}^{MN} \frac{\Omega\mu_i[\mathbf{R}]}{2f_D + \frac{L\sigma_x^2}{\sigma^2} \Omega\mu_i[\mathbf{R}]} \quad (43b)$$

For spatially uncorrelated fading at both ends of the link ($\mathbf{R} = \mathbf{I}_{MN}$), the bounds (43a) and (43b) reduce to the results in [2]:

$$\mathbf{Q}_{\hat{\mathbf{h}}} = \frac{2f_D\Omega}{2f_D + \Omega \frac{L\sigma_x^2}{\sigma^2}} \mathbf{I}_{MN} \quad (44a)$$

$$\text{MSE}_{\hat{\mathbf{h}}} = \frac{2f_D\Omega MN}{2f_D + \Omega \frac{L\sigma_x^2}{\sigma^2}} \quad (44b)$$

Moreover, if the faded amplitudes are uncorrelated across bursts ($f_D = 1/2$) and for large SNR ($\Omega L\sigma_x^2/\sigma^2 \gg 1$), the MSE (44b) becomes

$$\text{MSE}_{\hat{\mathbf{h}}} = \frac{\sigma^2 MN}{L\sigma_x^2} \quad (45)$$

which coincides with the MSE for UML estimator (11), which is also derived as the Cramér–Rao bound in [5].

V. ESTIMATORS FOR BLOCK-FADING MIMO CHANNELS

The unconstrained ML method [5], [19] performs a burst-by-burst estimation of the MIMO channel matrix \mathbf{H}_k without the exploitation of the structural or statistical side information discussed in Section III. According to the model (6) or (7), different approaches can be devised to estimate the long-term (\mathbf{T} or \mathbf{U}) and the short-term (β_k or \mathbf{d}_k) parameters from a set of K bursts.

A. Estimators for Long/Short-Term Parameters

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 estimate angles and delays according to the models in Section III. Such techniques have been developed for SIMO systems to exploit the stationarity of angles/delays in \mathbf{T} for TDMA [8], [14], [9] or TD-CDMA systems [27]. The extension to MIMO system is conceptually trivial, and it is just a matter of increased complexity when compared with SIMO systems. Basically, the angle and delay estimation methods are based on the minimization of nonlinear objective functions to compute the d triplets DOD/DOA/delay (for the beamforming model) or the d delays (for the diversity model) according to the knowledge of the waveform $g(t)$ and the spatial manifolds. Even if these parametric methods can closely reach the MSE bounds derived in Section IV (see [27] for the analytic derivation of the MSE for SIMO systems), there are several drawbacks that prevent their practical use. Nonlinear optimizations are known to have local minima so that iterative methods can estimate and track the slowly-varying angles and delays, provided that the initialization has been very carefully chosen. The need for regular spatial and temporal manifolds imposes strict constraints on array calibration errors and modeling mismatches. Furthermore, angle and delay estimation suffers from threshold effects at low SNRs that are typical for nonlinear estimators.

Instead of estimating angles and delays in \mathbf{T} , it is possible to directly evaluate the basis \mathbf{U} (unstructured approach [26], [28]). 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.

$$\text{MSE}_{\hat{\mathbf{h}}} = 2f_D \sum_{i=1}^{r_g} \sum_{n=1}^N \sum_{m=1}^M \frac{\mu_i[\mathbf{G}\Omega\mathbf{G}^H]\mu_n[\mathbf{R}_T]\mu_m[\mathbf{R}_R]}{2f_D + \frac{L\sigma_x^2}{\sigma^2} \mu_i[\mathbf{G}\Omega\mathbf{G}^H]\mu_n[\mathbf{R}_T]\mu_m[\mathbf{R}_R]} \quad (41)$$

As an example, let us consider the ideal case of $\mathbf{R}_x = L\sigma_x^2\mathbf{I}_{NW}$ and $\mathbf{R}_n = \sigma^2\mathbf{I}_M$ (the more general case can be reduced to the one discussed here by prewhitening for spatial correlation of noise and temporal correlation of the training sequences: see [16] for a general discussion). Since $E[\mathbf{h}_k\mathbf{h}_k^H] = \mathbf{T}\mathbf{T}^H$ and $\text{span}\{\mathbf{T}\} = \text{span}\{\mathbf{U}\}$, the estimate $\hat{\mathbf{U}}$ of the basis \mathbf{U} can be obtained from the r leading eigenvectors of the sample correlation matrix $\hat{\mathbf{R}}_h$ computed from the set of K UML estimates $\hat{\mathbf{h}}_{\text{UML},k} = \bar{\mathbf{X}}^\dagger \mathbf{y}_k$:

$$\hat{\mathbf{R}}_h = \frac{1}{K} \sum_{k=1}^K \hat{\mathbf{h}}_{\text{UML},k} \hat{\mathbf{h}}_{\text{UML},k}^H. \quad (46)$$

The estimate is consistent ($\text{cov}\{\hat{\mathbf{U}}\} \rightarrow 0$ if $K \rightarrow \infty$) if the fading is asymptotically uncorrelated (i.e., $\varphi_n \rightarrow 0$ for $n \rightarrow \infty$ or equivalently $f_D \neq 0$) [26]. Once $\hat{\mathbf{U}}$ has been estimated from K bursts, the amplitudes \mathbf{d}_k can be evaluated burst-by-burst as $\hat{\mathbf{d}}_k = \hat{\mathbf{U}}^H \hat{\mathbf{h}}_{\text{UML},k}$ so that the estimate of the MIMO channel becomes

$$\hat{\mathbf{h}}_k = \hat{\mathbf{U}} \hat{\mathbf{U}}^H \hat{\mathbf{h}}_{\text{UML},k} \quad (47)$$

i.e., the projection onto $\text{span}\{\hat{\mathbf{U}}\}$ of the estimates $\hat{\mathbf{h}}_{\text{UML},k}$. It can be easily shown that this burst-by-burst estimate of the amplitudes corresponds to the optimum MMSE strategy for uncorrelated fading and high SNR so that the performance of the estimator (47) attains the bound (40) for $K \rightarrow \infty$, $f_D = 1/2$ and $\text{SNR} \rightarrow \infty$. Interestingly, [26] shows numerically that for a SIMO system (i.e., $N = 1$), the asymptotic MSE bound (40) can be reached within a reasonable small number of bursts ($K \simeq 20 \div 30$).

If $f_D < 1/2$, fading is correlated across bursts, and their variations can be tracked by Kalman filtering or any suboptimal techniques [29]. In the context of MIMO systems, Kalman filtering has been recently proposed to track the amplitudes of the diversity model by assuming an autoregressive model and known sample-spaced delays (or, equivalently, \mathbf{T} or \mathbf{U} is diagonal and known) [12]. On the other hand, the estimator proposed in [30] for a SIMO setting estimates the basis \mathbf{U} from the matrix (46) and tracks the fading variations by LMS or RLS approaches. The performance of both techniques are lower bounded by the MSE in (40).

B. SISO/SIMO/MISO versus MIMO Approach to Channel Estimation

Instead of performing channel estimation in a MIMO system by *jointly* considering all the (frequency-selective) SISO channels corresponding to each pair transmitting-receiving antenna, one could use suboptimum approaches that estimates separately the SISO or the MISO/SIMO links. These approaches are detailed below:

- *SISO approach*: Estimate separately the MN SISO channels corresponding to each pair transmitting-receiving antennas.
- *MISO approach*: Estimate jointly all the N SISO channels relative to the links between all the transmitting antennas and one receiving antenna (M separate channel estimates).

- *SIMO approach*: Estimate jointly all the M SISO channels relative to the links between one transmitting antennas and all the receiving antennas (N separate channel estimates).

Based on the discussion above, these approaches are just a special case of MIMO estimators. However, it is of interest here to evaluate the performance loss of these suboptimal approaches. To simplify the analysis, herein we assume that there is no interference between the training sequences at the transmitting antennas (i.e., orthogonal training sequences: $\mathbf{R}_x = L\sigma_x^2\mathbf{I}_{NW}$).

The MSE bound on the performance of the SISO approach can be easily obtained by evaluating the MSE (37) for $M = N = 1$ for each of the MN SISO channels $\mathbf{h}_k^{(m,n)} = \mathbf{G}(\tau)\boldsymbol{\Omega}^{1/2}\boldsymbol{\beta}_k^{(m,n)} = \mathbf{U}_{\text{SISO}}\mathbf{d}_k^{(m,n)}$ composing the MIMO link. The basis, which is referred to as \mathbf{U}_{SISO} , reduces to the temporal basis $\mathbf{U}_{\text{SISO}} = \mathbf{U}_g$ (26), as $\text{span}\{\mathbf{U}_{\text{SISO}}\} = \text{span}\{\mathbf{G}(\tau)\boldsymbol{\Omega}^{1/2}\}$, and $r_{\text{SISO}} = \text{rank}\{\mathbf{U}_{\text{SISO}}\} = r_g \leq W$. If all the SISO channels share the same characteristics (as for transmitting and receiving antennas spaced not too far apart), the MSE bound for the SISO approach can be derived by specializing the MSE bound (37) for $M = N = 1$ (with obvious notation):

$$\text{MSE}_{\text{SISO}} = MN \cdot \text{MSE}_{\text{MIMO}}|_{M=1,N=1}. \quad (48)$$

According to the previous section, for high SNR and $\mathbf{R}_n = \sigma^2\mathbf{I}_M$, the MSE (48) becomes

$$\text{MSE}_{\text{SISO}} = MN \cdot 2f_D \frac{\sigma^2}{L\sigma_x^2} r_g \quad (49)$$

either for beamforming or diversity models. For high SNR, the comparison between (49) and (40) shows that for beamforming model, the *joint* MIMO approach outperforms the SISO approach by

$$\frac{\text{MSE}_{\text{SISO}}}{\text{MSE}_{\text{MIMO}}} = \frac{MNr_g}{r_{\text{MIMO}}} \quad (50)$$

which is greater than one since $r_{\text{MIMO}} \ll MNW$. On the other hand, for the diversity model, the SISO approach shows no degradation compared with the MIMO approach in (42).

The MSE bound on the performance of the SIMO (or dually MISO) approach to estimate the MIMO channel can be similarly obtained by evaluating the MSE (37) for $N = 1$ for each of the N SISO channels $\mathbf{h}_k^{(m,n)} = \mathbf{U}_{\text{SIMO}}\mathbf{d}_k^{(m,n)}$ composing the MIMO link:

$$\text{MSE}_{\text{SIMO}} = N \cdot \text{MSE}_{\text{MIMO}}|_{N=1}. \quad (51)$$

Since the basis is now \mathbf{U}_{SIMO} with $r_{\text{SIMO}} = \text{rank}\{\mathbf{U}_{\text{SIMO}}\} \leq MW$, the performance comparison is

$$\frac{\text{MSE}_{\text{SIMO}}}{\text{MSE}_{\text{MIMO}}} = \frac{N \cdot r_{\text{SIMO}}}{r_{\text{MIMO}}} \quad (52)$$

which is larger than one as for the beamforming model (the MISO approach yields a similar conclusion). Under the said assumptions, the degree of improvement of the MIMO approach to channel estimation compared with the SISO or SIMO (or MISO) approaches depends on the dimension of the temporal and/or spatial subspaces, which are closely related to the specific

model. The rank order of these subspaces (r_{SISO} or r_{SIMO}) increases when the multipath environment is dense in time and/or space. For a number of (well-resolved in space and/or time) paths d large enough, we have $r_{\text{MIMO}} \simeq MNW$, $r_{\text{SISO}} \simeq W$ and $r_{\text{SIMO}} \simeq MW$ so that there is no practical advantage in estimating jointly the MIMO channel. These conclusions are validated numerically in Section VII-B.

VI. INFORMATION RATE WITH IMPERFECT CHANNEL KNOWLEDGE

Here we focus on a time-slotted system designed according to the principle of synchronized detection [1]. As discussed in Section II and shown in Fig. 2, the training sequence (of length L) is followed by a burst of data symbol of length N_D . The overhead for the transmission of one block of data with the guards required to avoid any block interference is $L' = L + W - 1$.

The impact of imperfect channel knowledge on the system performance is investigated in [17] for a SISO link in terms of capacity. In particular, an upper bound on the information rate assuming a Gaussian input distribution is derived. The choice of a Gaussian input distribution might not lead to the maximization of the information rate and, thus, to the capacity of the system, but it greatly simplifies the mathematical analysis. Extending the treatment of [17] to a frequency-selective block-fading MIMO system, it can be easily shown that for Gaussian uncorrelated (among transmitting antennas) signaling with power σ_D^2 and an estimate $\hat{\mathbf{H}}$, we have $I(\mathbf{y}, \mathbf{x} | \hat{\mathbf{H}}) \geq I_{lb}$, where the lower bound on the information rate is

$$I_{lb} = \frac{1}{N_D + L} \log_2 \left| \mathbf{I} + \sigma_D^2 \left[\sigma_n^2 (\mathbf{I} \otimes \mathbf{R}_n) + \sigma_D^2 \mathbf{Q}_{\hat{\mathbf{H}}} \right]^{-1} \hat{\mathbf{H}} \hat{\mathbf{H}}^H \right| \text{ [bit/s/Hz]} \quad (53)$$

where the temporal index k is dropped for convenience, and \mathbf{y} and \mathbf{x} denote the signal vectors, respectively, received and transmitted during the data burst. Herein, $\hat{\mathbf{H}}$ denotes the $M(N_D + W - 1) \times NN_D$ convolution matrix of the MIMO channel obtained from the estimate $\hat{\mathbf{H}}$ and $\mathbf{Q}_{\hat{\mathbf{H}}} = E[(\hat{\mathbf{H}} - \mathbf{H})(\hat{\mathbf{H}} - \mathbf{H})^H]$. Under the assumption $K \rightarrow \infty$ (which is consistent with the infinite temporal horizon of information theory), the signal model (30) is linear in the fading amplitudes, and similar to [2], [3], and [17], we consider the matrix $\hat{\mathbf{H}}$ as obtained from the MMSE estimator of Section IV so that $\mathbf{Q}_{\hat{\mathbf{H}}}$ can be derived from the error correlation matrix $\mathbf{Q}_{\hat{\mathbf{h}}}$ (see Appendix A). Notice that a different choice for the channel estimator would decrease the lower bound (53) so that the inequality $I(\mathbf{y}, \mathbf{x} | \hat{\mathbf{H}}) \geq I_{lb}$ holds for a channel estimator that reaches the performance (36).

According to [17], the lower bound (53) has a simple intuitive explanation that follows once compared with the capacity for perfect channel state information:

$$C = \frac{1}{N_D + L} \log_2 \left| \mathbf{I} + \sigma_D^2 / \sigma_n^2 (\mathbf{I} \otimes \mathbf{R}_n^{-1}) \mathbf{H} \mathbf{H}^H \right| \text{ [bit/s/Hz]} \quad (54)$$

Compared to the capacity (54), the worst possible effect of the channel estimation error is to behave as AGN since the lower bound (53) experiences a noise covariance matrix $\sigma_n^2 (\mathbf{I} \otimes \mathbf{R}_n)$ modified by the covariance matrix due to the estimation error

TABLE I
DEFAULT PARAMETER SETTING

M	4	\mathbf{Q}	$\sigma_n^2 \mathbf{Z}_M(\rho_n)$
N	4	\mathbf{R}_x	$\sigma_x^2 \mathbf{Z}_W(\rho_l) \otimes \mathbf{Z}_N(\rho_a)$
W	8	\mathbf{R}	$\mathbf{Z}_N(\rho_{tx}) \otimes \mathbf{Z}_M(\rho_{rx})$
d	4	ρ_n	0
τ_i/T	$i - 1$	ρ_l	0
$\alpha_i^{(T)}, \alpha_i^{(R)}$	$-60 + 120/d \cdot (i - 1)$ deg	ρ_a	0

$\sigma_D^2 \mathbf{Q}_{\hat{\mathbf{H}}}$. In other words, let the received signal be written (with appropriate definitions) as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} = \hat{\mathbf{H}}\mathbf{x} + ((\mathbf{H} - \hat{\mathbf{H}})\mathbf{x} + \mathbf{n}). \quad (55)$$

The worst case (which leads to the lower bound on the information rate) treats the term $(\mathbf{H} - \hat{\mathbf{H}})\mathbf{x}$ as additional Gaussian noise with covariance $\sigma_D^2 \mathbf{Q}_{\hat{\mathbf{H}}}$.

Similarly to the information rate (53), the capacity (54) is scaled by $1/(N_D + L')$ in order to take into account the loss due to the transmission of the training sequences. In other words, while computing C , we assume that training sequences of length L lead to perfect channel state information, whereas while computing I_{lb} , the channel is assumed to be estimated according to the method discussed in Section IV.

VII. NUMERICAL EXAMPLES

In the following, we present numerical evaluations of the bound (36) and (37) related to some specific problems. The simulation settings are in Table I. The power-delay profile Ω is scaled so that $E[|\mathbf{H}_k|^2] = M$, and the SNR is defined as $\text{SNR} = \sigma_x^2 / \sigma^2$ (see Section IV-A for definitions). To gain some insight into the effects of system parameters and channel characteristics on the estimation error, the correlation matrices involved in signal modeling are assumed to be obtained from an autoregressive model: $\mathbf{Z}_N(\rho)$ denotes a $N \times N$ Toeplitz matrix with first column $[1 \rho \dots \rho^{N-1}]^T$. Here, ρ_n accounts for the spatial correlation of noise (\mathbf{Q}), ρ_l and ρ_a for auto- and mutual-correlation of the training sequences of different antennas (training sequences in practical systems are usually designed to have $\rho_l = \rho_a \simeq 0$). For the beamforming model, the arrays at both ends are assumed to be uniform linear with half-wavelength interelement spacing. For the diversity model, the correlation matrix \mathbf{R} is separable; the correlation coefficient ρ_{tx} and ρ_{rx} characterize the spatial correlation of the fading at the transmitter and receiver side, respectively.

A. Performance of Known Estimators

Here, we corroborate the discussion of Sections IV-A and V with some numerical examples in order to show the relationship between the performance of known estimators, namely, the UML estimator and the parametric estimators with the lower bound derived in Section IV. By parametric estimators, we refer to methods (e.g., multislot [26] and JADE [8]) that perform a consistent estimate of the long-term features of the channel and a burst-by-burst estimate of the amplitudes (see also Section V). The parametric estimators thus exploit only structural information about the model (7). We set $L = NW = 32$ as necessary for the UML, JADE, and multislot estimators ($L \geq NW$).

The MSE of the UML (11) and parametric estimators (40) with $f_D = 1/2$ is evaluated for the beamforming (Fig. 3) and

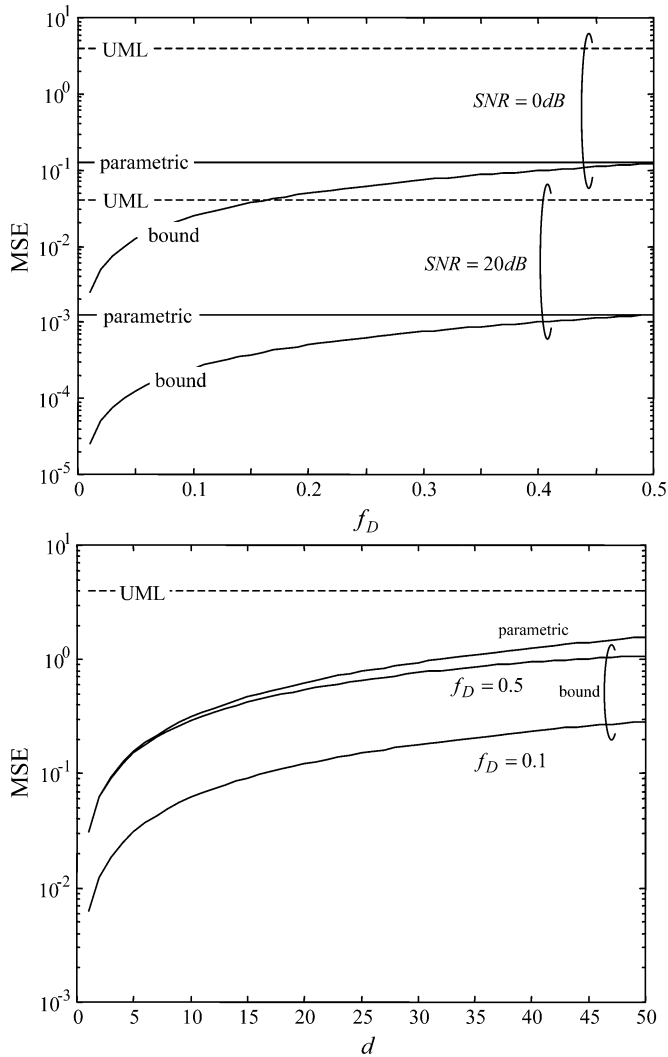


Fig. 3. Beamforming model: MSE of the UML and the parametric estimators compared with the MSE bound versus Doppler shift f_D . (Top) For SNR = 0, 20 dB and $d = 4$ and number of paths d . (Bottom) For SNR = 0 dB and $f_D = 0.1, 0.5$.

diversity model (Fig. 4) and compared with the MSE bound (37). Performance degradation of these techniques occurs as these estimators fail to fully exploit the information on the MIMO channel matrix. Indeed, as discussed in Section V, the UML estimator does not use any deterministic or statistical information, whereas the parametric methods only capitalize on structural modeling of the channel (29). However, when the information neglected by the estimator is not helpful in improving the performance, no degradation is expected.

Consider, for instance, the upper parts of Figs. 3 and 4, which show the MSE versus the Doppler frequency f_D . For increasing f_D , i.e., increasingly uncorrelated fading amplitudes across different bursts, no benefits can be obtained by tracking the amplitudes, and as a consequence, the degradation of the two estimators decreases. Similarly, for the diversity model, a decreasing spatial (ρ_{tx}) or temporal (ρ_{rx}) correlation renders the Wiener approach of the optimum estimator increasingly ineffective, and the degradation decreases, as shown in the lower part of Fig. 4. An increased SNR would make this effect less noticeable (not shown in the figure). The impact

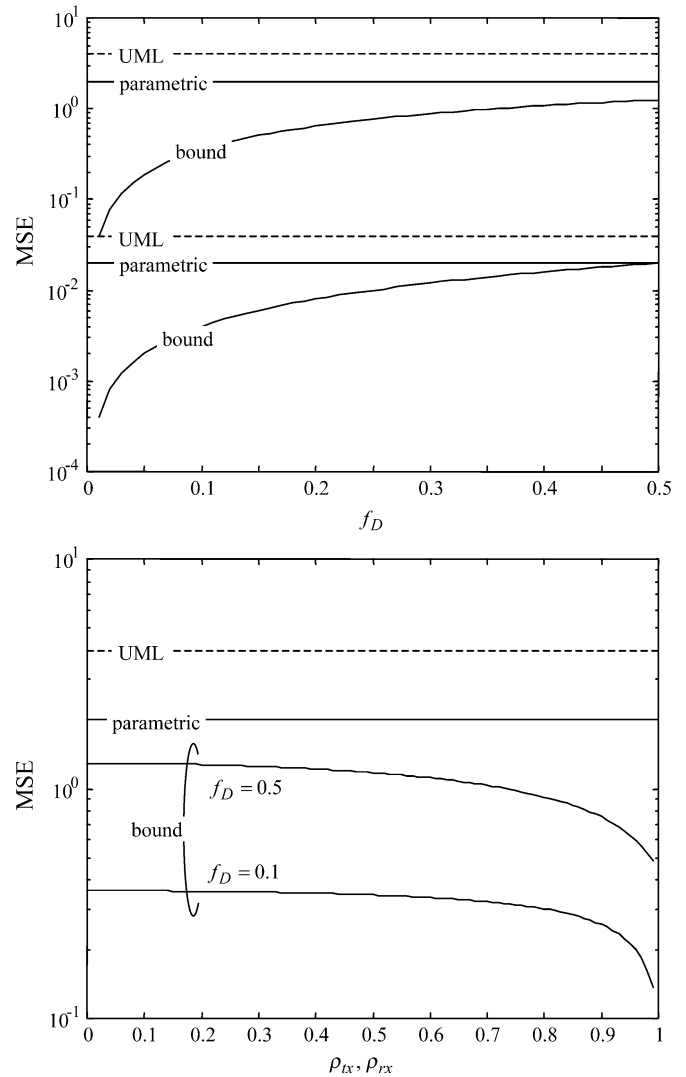


Fig. 4. Diversity model: MSE degradation of the UML and parametric estimators compared with the MSE bound versus Doppler shift f_D (Top) For SNR = 0, 20 dB and $\rho_{tx} = \rho_{rx} = 0.3$ and ρ_{tx} (or ρ_{rx}) with $\rho_{rx} = 0$ (or $\rho_{tx} = 0$). (Bottom) For SNR = 0 dB and $f_D = 0.1, 0.5$.

of the SNR, which is studied in the upper part of Figs. 3 and 4, can be interpreted in the same way: An increasing SNR makes the Wiener approach of the optimum estimator closer in performance to an ML approach, and as a consequence, the degradation decreases.

Let us focus on the performance comparison between the UML and the parametric estimators. Following the same logic of the discussion above, the parametric approach is expected to outperform the UML estimator as long as the structural information about the model (29) yields some benefits for the estimation process. In particular, this happens if the structural information allows us to reduce the set of parameters to be estimated. In fact, the UML approach entails the need to estimate MNW parameters for each block, whereas the parametric methods (asymptotically for $K \rightarrow \infty$) require the estimate of r amplitudes only. The rank r increases with the number of path d . Accordingly, the lower part of Fig. 3 shows that for increasing d , the performance gap of the two estimators shrinks and eventually vanishes for $r = MNW$ (not shown in the figure).

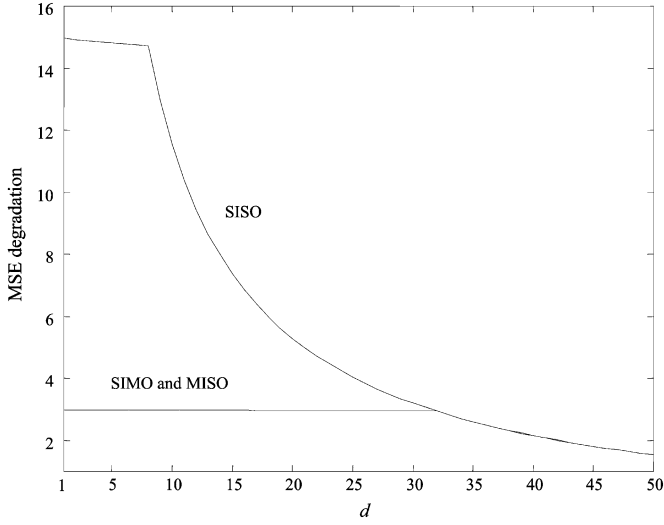


Fig. 5. MSE degradation of the SIMO, MISO, and SISO approaches compared with the MIMO estimation versus the number of paths d for the beamforming model (SNR = 10 dB, and $f_D = 0.1$, $L = 32$).

B. SISO/MISO/SIMO versus MIMO Channel Estimation

To compare different strategies, we study in Fig. 5 the MSE degradation $\text{MSE}_{\text{degradation}} = (\text{MSE} - \text{MSE}_{\hat{\mathbf{h}}})/\text{MSE}_{\hat{\mathbf{h}}}$ of SISO, MISO, or SIMO channel estimation approaches compared with the more general MIMO approach for the beamforming model and varying number of paths d . The Doppler shift is set to $f_D = 0.1$, SNR = 10 dB and $L = 32$. Notice that the choice $\rho_a = 0$ allows a simple decoupling of the channel estimation for different transmitting antennas (see Section V-B).

The benefits of using a MIMO channel estimator compared to SISO, MISO, or SIMO become smaller for the increasing number of paths d when $d \geq W = 8$ for the SISO approach, $d \geq NW = 32$ for the MISO, and $d \geq MW = 32$ for the SIMO approach. In addition, for $d \geq \max(MW, NW) = 32$, the SISO, MISO, and SIMO approaches have the same performance.

These results can be easily justified by recalling the discussion in Section V-B. For instance, from (50), we have that $\text{MSE}_{\text{SISO}}/\text{MSE}_{\hat{\mathbf{h}}} \simeq MN \cdot \min(d, W)/d$, where the last (approximate) equality stems from the direct proportionality of the rank of \mathbf{T} (or \mathbf{U}) and the number of paths d , as long as the signatures of the d paths are linearly independent. It follows that for $d \geq W = 8$, the degradation of the SISO approach decreases as $1/d$, as confirmed by Fig. 5. Similar considerations can be used to prove analogous conclusions for SIMO and MISO approaches.

C. Impact of Channel Estimation Error on System Performance

The influence of system parameters and channel characteristics on the lower bound on the information rate in the presence of imperfect channel state information (53) and on the capacity for a known channel (54) is investigated numerically. The computation of the bound (53) is carried out by averaging over the distribution of fading amplitudes (recall that angles, delays, and power-delay profile are deterministic) by using 10^3 runs of Monte Carlo simulations. Similarly to the experimental analysis carried out in [2], the estimate $\hat{\mathbf{d}}$ of the fading amplitudes has a circularly symmetric Gaussian

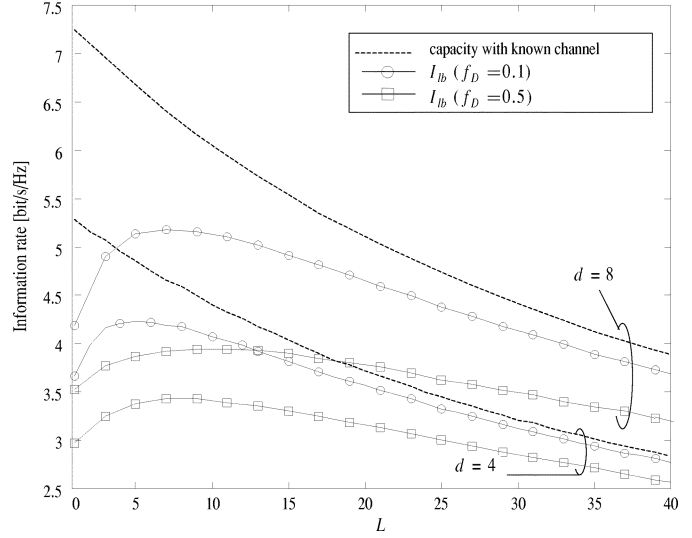


Fig. 6. Average capacity with known channel and lower bound on the information rate with imperfect channel state information I_{lb} versus the length of the training sequence L (diversity model $\rho_{tx} = \rho_{rx} = 0.3$).

distribution with covariance matrix $\mathbf{Q}_{\hat{\mathbf{d}}}$, i.e., $\hat{\mathbf{d}} \sim \mathcal{CN}(0, \mathbf{Q}_{\hat{\mathbf{d}}})$. The size of the block is $N_D = 30$, and the power of the data symbol is set equal to the power of the training sequence $\sigma_D^2 = \sigma_x^2$ (see [3] for a discussion on the benefits of exploiting the degree of freedom in the choice of the ratio σ_D^2/σ_x^2) and SNR = 10 dB.

Information Rate versus L : The optimal choice of the length L of training sequences is investigated first. We focus on the diversity model since the beamforming model leads to similar conclusions. Fig. 6 shows the average capacity (dashed lines) and the lower bound on the information rate (solid lines) versus L for $\rho_{rx} = \rho_{tx} = 0.3$ different Doppler shifts ($f_D = 0.1, 0.5$) and number of paths $d = 4, 8$. Increasing L has two opposite effects on the information rate: It increases the overhead L' (which reduces both the scaled capacity and the information rate), and it decreases the channel estimation error (which has a beneficial impact on the information rate). Therefore, the bound I_{lb} versus L shows an optimum tradeoff between transmission overhead and channel estimation error, as in Fig. 6. The optimum training sequence length L depends on both the number of paths d and the Doppler shift. For $f_D = 0.1$, the information rate I_{lb} peaks at approximately $L \simeq d$, whereas for larger f_D , the maximum moves toward slightly higher values of L . This can be simply interpreted by pointing out that in a block-fading channel, it is advisable to increase L in order to experience a constant channel for a longer interval if the channel itself varies rapidly over consecutive bursts. Interestingly, under the assumption of known long-term features of the channel, the optimum L is much smaller than the number of training symbols required to obtain the UML, i.e., $L \geq NW$, which is the choice of third-generation cellular standards [31].

I_{lb} versus $\rho_{tx} = \rho_{rx}$ and d : Fig. 7 shows the degradation of the average information rate when employing an estimator that reaches the MSE bound derived in this paper compared with the case of known channel as a function of Doppler spread f_D , spatial correlation $\rho_{tx} = \rho_{rx}$ (for the diversity model), and number of paths d (for the beamforming model). The degradation is measured as $(C - I_{lb})/C$ (notice that being $I_{lb} \leq I \leq C$,

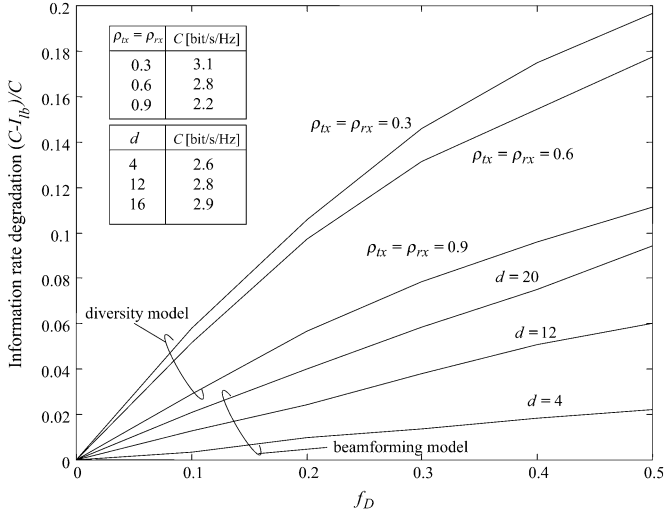


Fig. 7. Degradation of the average information rate compared with the capacity with perfect channel state information versus Doppler shift f_D for different values of spatial correlation $\rho_{tx} = \rho_{rx}$ (diversity model) and number of paths d (beamforming model).

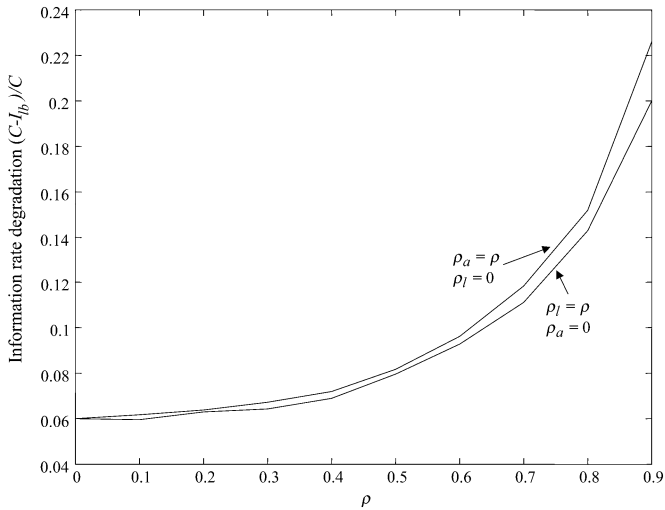


Fig. 8. Degradation of the average information rate compared with the capacity with perfect channel state information versus correlation ρ : i) $\rho_l = \rho$ and $\rho_a = 0$. ii) $\rho_a = \rho$ with $\rho_l = 0$.

the selected quantity is an upper bound on $(C - I)/C$ and set $L = 32$. The values of the capacity C are in a table in the upper left corner.

For decreasing Doppler shifts, the degradation decreases and vanishes for $f_D = 0$ since a static channel can be estimated with any accuracy as $K \rightarrow \infty$ (see Section IV-A). Moreover, the degradation decreases with increasing spatial correlation (for the diversity model) since higher correlations lead to a smaller channel estimation error. In order to keep the example consistent, as $d = 4$ for the diversity model, the delays are selected as $\tau_i/T = \lfloor (i-1)/4 \rfloor$ for $i = 1, \dots, d$ with $d = 4, 12, 20$ for the beamforming model (i.e., the delays span the first four samples of the channel). With this choice, increasing the number of paths in the beamforming model has qualitatively the same effect as decreasing the spatial correlation in the diversity model, as shown in Fig. 7. As a final remark, the degradation for the beamforming model is smaller than for the diversity model since

in the first case, the number of parameters (amplitudes) to be tracked is $r \simeq d$, whereas in the latter it is larger ($r \simeq MNd$).

I_{lb} versus Training Sequence Properties: The effect of the correlation properties of the training sequences (ρ_a and ρ_l) on the system performance is addressed in terms of the information rate degradation, as in the previous example. The results for the diversity model with $f_D = 0.1$ and $L = 32$ ($C = 3.1$ bit/sec/Hz, as shown in Figs. 6 and 7) are depicted in Fig. 8. The degradation of the average information rate versus ρ_l with $\rho_a = 0$ and versus ρ_a with $\rho_l = 0$ show that both ρ_a and ρ_l have a similar impact on the information rate. The correlations range from 0 to 0.9 since a complete correlation ($\rho_a = 1$ or $\rho_l = 1$) would not be compatible with the assumption of a consistent estimate of the basis.

VIII. CONCLUSION

A lower bound on the error correlation matrix for training-based channel estimators over a block-fading frequency-selective MIMO channel has been derived. The bound has been obtained in a constructive way by computing the asymptotic performance (performance for $K < \infty$ are in [32]) of an estimator that is able to estimate the long-term features of the channel (e.g., second-order statistics, delays, angles) consistently while tracking the fast-varying fading fluctuations by Wiener filtering. A connection of the bound with known results have been discussed both analytically and through simulation. Moreover, a lower bound on the information rate of a MIMO system over a block-fading frequency-selective MIMO channel with imperfect channel state information has been adapted from the SISO counterpart. The bound on the information rate has been thoroughly investigated through numerical results in order to show the impact of system parameters (e.g., length of the training sequence versus size of the data block and correlation properties of the training sequences) and channel characteristics (e.g., Doppler shift, spatial correlation) on the system performance.

APPENDIX A

COMPUTATION OF $\mathbf{Q}_{\hat{\mathcal{H}}}$ FROM $\mathbf{Q}_{\hat{\mathbf{h}}}$

The matrix $M(N_D + W - 1) \times NN_D \hat{\mathcal{H}}$ has a Toeplitz structure with first $MW \times N$ block-column given by $[\hat{\mathbf{H}}[0]^T \dots \hat{\mathbf{H}}[W-1]^T]^T$

$$\hat{\mathcal{H}} = \begin{bmatrix} \hat{\mathbf{H}}[0] & \mathbf{0} & & & \\ \vdots & \hat{\mathbf{H}}[0] & & & \\ \hat{\mathbf{H}}[W-1] & \vdots & \dots & \hat{\mathbf{H}}[0] & \\ \mathbf{0} & \hat{\mathbf{H}}[W-1] & & \vdots & \\ & & & & \hat{\mathbf{H}}[W-1] \end{bmatrix}. \quad (56)$$

The $M(N_D + W - 1) \times M(N_D + W - 1)$ covariance matrix $\mathbf{Q}_{\hat{\mathcal{H}}} = E[(\hat{\mathcal{H}} - \mathcal{H})(\hat{\mathcal{H}} - \mathcal{H})^H]$ can be written as a $(N_D + W - 1) \times (N_D + W - 1)$ block matrix

$$\mathbf{Q}_{\hat{\mathcal{H}}} = \begin{bmatrix} \mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{12} & \dots & \mathbf{\Gamma}_{1(N_D+W-1)} \\ \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_{22} & & \mathbf{\Gamma}_{2(N_D+W-1)} \\ \vdots & & \ddots & \\ \mathbf{\Gamma}_{(N_D+W-1)1} & & & \mathbf{\Gamma}_{(N_D+W-1)(N_D+W-1)} \end{bmatrix} \quad (57)$$

where the blocks Γ_{ij} are $M \times M$. If $N_D \gg W$, the matrix $\mathbf{Q}_{\hat{\mathbf{h}}}$ assumes a symmetric Toeplitz structure [33], i.e., $\Gamma_{ij} = \Gamma_{i-j}$ and $\Gamma_{-k} = \Gamma_k^H$, with $\Gamma_k = \sum_{i=0}^{W-k} E[(\hat{\mathbf{H}}[i] - \mathbf{H}[i])(\hat{\mathbf{H}}[i+k] - \mathbf{H}[i+k])^H]$.

The blocks Γ_k can be computed from the MSE matrix $\mathbf{Q}_{\hat{\mathbf{h}}}$ evaluated in (36). In fact, $\mathbf{Q}_{\hat{\mathbf{h}}}$ is a $W \times W$ block matrix such that the (i, j) th block, which is denoted as $\mathbf{Q}_{\hat{\mathbf{h}}}(i, j)$, is the $MN \times MN$ matrix $\mathbf{Q}_{\hat{\mathbf{h}}}(i, j) = E[\text{vec}\{\hat{\mathbf{H}}[i] - \mathbf{H}[i]\} \text{vec}\{\hat{\mathbf{H}}[j] - \mathbf{H}[j]\}^H]$ for $i, j = 0, \dots, W-1$. $\mathbf{Q}_{\hat{\mathbf{h}}}(i, j)$ can in turn be written as a $N \times N$ block matrix composed of $M \times M$ blocks such that $E[(\hat{\mathbf{H}}[i] - \mathbf{H}[i])(\hat{\mathbf{H}}[j] - \mathbf{H}[j])^H]$ is equal to the sum of the blocks on its main diagonal.

REFERENCES

- [1] H. Meyr, M. Moeneclaey, and S. Fechtel, *Digital Communication Receivers*. New York: Wiley, 1998.
- [2] J. Baltzer, G. Flock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 2358–2368, Dec. 2001.
- [3] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," Bell Labs, Lucent Technol., Apr. 2000. Tech. Memo. [Online] mars.bell-labs.com/cm/ms/what/mars/papers/training/.
- [4] D. Shiu, G. J. Foschini, and M. J. Gans, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000.
- [5] T. L. Marzetta, "Blast training: Estimating channel characteristics for high capacity space-time wireless," in *Proc. 37th Annual Allerton Conf. Commun., Contr., Comput.*, Sept. 1999, pp. 958–966.
- [6] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.
- [7] A. Richter, D. Hampicke, G. Sommerkorn, and R. Thoma, "MIMO measurement and joint parameter estimation of mobile radio channels," in *Proc. Veh. Technol. Conf.*, vol. 1, pp. 214–218.
- [8] A. Van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Processing*, vol. 46, pp. 405–418, Feb. 1998.
- [9] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Trans. Signal Processing*, vol. 46, pp. 449–462, Feb. 1998.
- [10] H. L. Van Trees, *Optimum Array Processing*. New York: Wiley, 2002.
- [11] P. Kyritsi and D. C. Cox, "Correlation properties of MIMO radio channels for indoor scenarios," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2001, pp. 994–998.
- [12] C. Kinninakis, C. Fragouli, A. H. Sayed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using Kalman estimation," *IEEE Trans. Signal Processing*, vol. 50, pp. 1065–1076, May 2002.
- [13] Z. Liu, Y. Xin, and G. B. Giannakis, "Space-time-frequency coded OFDM over frequency-selective fading channels," *IEEE Trans. Signal Processing*, vol. 50, pp. 2465–2476, Oct. 2002.
- [14] Y.-Y. Wang, J.-T. Chen, and W.-H. Fang, "TST-MUSIC for joint DOA-delay estimation," *IEEE Trans. Signal Processing*, vol. 49, pp. 721–729, Apr. 2001.
- [15] D. Gerlach and A. Paulraj, "Adaptive transmitting antenna arrays with feedback," *IEEE Signal Processing Lett.*, vol. 1, pp. 150–152, Oct. 1994.
- [16] O. Simeone and U. Spagnolini, "Multislot estimation of space-time channels," in *Proc. Int. Conf. Commun.*, Apr. 2002, pp. 802–806.
- [17] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, pp. 933–946, May 2000.
- [18] E. De Carvalho and D. T. M. Slock, "Cramer-Rao bounds for semiblind, blind and training sequence based channel estimation," in *Proc. IEEE Signal Processing Workshop Adv. Wireless Commun.*, 1997, pp. 129–132.
- [19] Q. Sun, D. C. Cox, H. C. Huang, and A. Lozano, "Estimation of continuous flat fading MIMO channels," *IEEE Trans. Wireless Comm.*, vol. 1, pp. 549–553, Oct. 2002.
- [20] P. A. Bello, "Characterization of randomly time-variant linear channel," *IEEE Trans. Circuits Syst.*, vol. CAS-11, pp. 360–393, Dec. 1963.
- [21] R. Janaswamy, *Radiowave Propagation and Smart Antennas for Wireless Communications*. Boston, MA: Kluwer, 2000.
- [22] K. I. Pedersen, J. B. Andersen, J. P. Kermaoal, and P. Mogensen, "A stochastic multiple-input-multiple-output radio channel model for evaluation of space-time coding algorithms," in *Proc. IEEE Veh. Technol. Conf.*, Boston, MA, 2000, pp. 893–897.
- [23] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Select. Areas Commun.*, vol. 20, pp. 550–560, Apr. 2002.
- [24] H. Bolckei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–234, Feb. 2002.
- [25] J.-T. Chen, Jr., J. Kim, and J.-W. Liang, "Multichannel MLSE equalizer with parametric FIR channel identification," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 1923–1935, Nov. 1999.
- [26] M. Nicoli, O. Simeone, and U. Spagnolini, "Multi-slot estimation of fast-varying space-time communication channels," *IEEE Trans. Signal Processing*, vol. 51, pp. 1184–1195, May 2003.
- [27] J. Picheral and U. Spagnolini, "Angles and delay estimation of space-time channels for multiuser detection," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 758–769, May 2004.
- [28] F. Dietrich, M. T. Ivrlac, and J. A. Nossek, "On performance limits of optimum reduced rank channel estimation," in *Proc. GLOBECOM*, vol. 1, 2002, pp. 345–349.
- [29] L. Lindbom, M. Sternad, and A. Ahlén, "Tracking of the time-varying mobile radio channels-Part I: The Wiener LMS algorithm," *IEEE Trans. Commun.*, vol. 49, pp. 2207–2217, Dec. 2001.
- [30] M. Nicoli, M. Sternad, U. Spagnolini, and A. Ahlen, "Reduced-rank channel estimation and tracking in time-slotted CDMA systems," in *Proc. Int. Conf. Commun.*, vol. 1, May 2002, pp. 533–537.
- [31] H. Holma and A. Toskala, *WCDMA for UMTS*. New York: Wiley, 2000.
- [32] O. Simeone and U. Spagnolini, "Hybrid CRB in linear regression for random parameters with unknown and rank deficient correlation matrix," submitted for publication.
- [33] J. Rissanen, "Algorithms for triangular decomposition of block Hankel and Toeplitz matrices with application to factoring positive matrix polynomials," *Math. Comput.*, vol. 27, no. 121, pp. 147–154, Jan. 1973.



Osvaldo Simeone (S'03) received the M.Sc. degree (with honors) from Politecnico di Milano, Milan, Italy in 2001, where he is currently working toward Ph.D. degree.

From February to September 2002, he was a visiting researcher at Center for Communications and Signal Processing Research, New Jersey Institute of Technology, Newark. He holds a patent on the work developed for his Master thesis. His current research interests lie in the field of signal processing for digital communications, with emphasis on multiple

input multiple output systems, multicarrier modulation, and channel estimation.



Umberto Spagnolini (SM'99) received the Dott. Ing. Elettronica degree (cum laude) from Politecnico di Milano, Milan, Italy, in 1988.

Since 1988, he has been with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, where he has held the position of Associate Professor of digital signal processing since 1998. His general interests are in the area of signal processing, estimation theory, and system identification. The specific areas of interest include channel estimation and array processing for communication systems, parameter

estimation and tracking, signal processing, and wavefield interpolation with applications to radar (SAR and UWB), geophysics, and remote sensing.

Dr. Spagnolini is a member of the SEG and EAGE and serves as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He received the AEI Award in 1991, the Van Weelden Award of EAGE in 1991, and the Best Paper Award from EAGE in 1998.