Reduced-Rank Channel Estimation for Time-Slotted Mobile Communication Systems

Monica Nicoli, *Member, IEEE,* and Umberto Spagnolini, *Senior Member, IEEE*

***Abstract*—In time-slotted mobile communication systems with antenna array at the receiver, the space-time channel matrix is conventionally estimated by transmitting pilot symbols within each data packet (or *block*). This paper is focused on reduced rank (RR) estimation methods that exploit the low-rank property of the space-time channel matrix to estimate single or multiple user channels from the observation of single or multiple training blocks. The proposed RR methods allow to improve the estimate accuracy by reducing the set of unknown parameters (rank reduction) and extending the training set (multiblock processing). The maximum likelihood RR estimate is obtained as the projection of the prewhitened full-rank (FR) estimate onto the spatial or temporal signal subspace. The paper shows that, even for time varying channels, these subspaces can be considered to be slowly varying, and therefore, they can be estimated with increased accuracy by properly exploiting training signals from several blocks. The analytical and numerical performance in terms of mean square error for the RR estimate shows that the main advantage of the proposed method with respect to the conventional FR one can be ascribed to the reduced complexity of the channel parameterization.**

***Index Terms*—Antenna arrays, array signal processing, code division multiaccess, fading channels, least mean square methods, maximum likelihood estimation, mobile communication, multipath channels, multiuser channels, parameter estimation, radio communication, reduced-rank processing, time division multiaccess, time-varying channels, training.**

I. Introduction

In mobile communication systems, the propagation channel is time-varying, and its estimation is mostly training based. In order to allow channel estimation, the moving terminal transmits periodically a sequence of known (pilot) symbols. Letting the N training signals be received by an antenna array of M elements through a frequency selective fading channel of length W (expressed in symbol intervals), the discrete-time model for the received signals is

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{n}(t), \quad t = 0, \ldots, N-1. \tag{1}$$

The $M \times 1$ signal vector $\mathbf{y}(t) = [y_1(t) \cdots y_M(t)]^{\mathrm{T}}$ is obtained by sampling at the symbol rate $1/T$ the matched filter output at each receiving antenna, while $\mathbf{x}(t) = [x(t)\ x(t-1) \cdots x(t-W+1)]^{\mathrm{T}}$ collects W symbols of the transmitted training sequence $\{x(t)\}_{t=-W+1}^{N-1}$ from a finite alphabet set. The $M \times W$ matrix $\mathbf{H}$ describes the discrete-time channel impulse response for the single-input-multiple-output (SIMO) link from the (single-antenna) transmitter to the (multiple-antenna) receiver. The impulse response includes the array gain/phase adjustments, the effects of path fading, the pulse waveform used in transmission, and the matched filter at the receiver. The channel matrix $\mathbf{H}$ is herein referred to as the space-time matrix, and it is considered to be quasi-static, at least within the training period (block-fading channel). The additive noise vector $\mathbf{n}(t)$ models both the co-channel interference and the background noise, and it is assumed to be a stationary normally distributed process with zero mean. To simplify, it is approximated as temporally uncorrelated but spatially correlated with unknown spatial covariance $\mathbf{Q}$: $\mathrm{E}[\mathbf{n}(t)\mathbf{n}^{\mathrm{H}}(t-m)] = \mathbf{Q}\delta(m)$. The matrix $\mathbf{Q}$ is positive definite, and its diagonal entries $\mathbf{Q}[m,m] = \sigma^2$ for $m = 1, \ldots, M$ represent the noise power at each antenna element.

The problem addressed in this paper is the estimation of the channel matrix $\mathbf{H}$ and the noise covariance $\mathbf{Q}$ under the assumption of Gaussian noise. A straightforward approach is the unconstrained maximum likelihood estimate (MLE) (see, e.g., [1]). However, a large number of channel unknowns (MW) or low training sequence length (N) may limit the estimate accuracy and increase the probability of error in symbol detection. In order to improve the accuracy of the estimate for block-fading channels, the space-time matrix $\mathbf{H}$ has to be reparameterized by using low-complexity models. Structured methods model the channel in terms of multipath parameters and perform a joint estimation of angles of arrival and times of delay [2], [3]. An alternative and practicable approach, which is considered in this paper, is the unstructured estimation based on rank reduction of $\mathbf{H}$. Low-rank channel matrices occur in several practical situations where the angle or delay spread of the propagation is small compared with the system resolution (see the analysis in [4] and references therein). In these situations, the MW entries of the space-time matrix $\mathbf{H}$ are more than those really necessary to describe the channel impulse response. Reduced-rank (RR) methods provide a parsimonious parameterization by introducing the low-rank constraint on the matrix $\mathbf{H}$. With respect to the aforementioned structured methods, the RR approach guarantees increased robustness against mismodeling (as the knowledge of the spatial/temporal manifold is not required) and lower computational complexity for channel estimation.

Several works are available in the literature about RR estimation and filtering (see [1], [5], and references therein). Maximum likelihood (ML) estimation for a RR linear regression in the presence of spatially correlated Gaussian noise with unknown covariance matrix is proposed by Viberg and Stoica [6]. The extension to both spatially and temporally correlated noise can be found in [7]. RR channel estimation and equalization in wireless

Manuscript received September 20, 2002; revised February 1, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiodong Wang.

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

Digital Object Identifier 10.1109/TSP.2004.842191

communication systems are in [8]–[10] for single-user systems and in [11] for multiuser systems (i.e., DS-CDMA systems). For rank-one channel matrices, RR methods have been proposed for joint space-time equalization [12], where the space-time filter can be equivalently obtained through the maximization of the signal-to-interference ratio [13], [14]. Field tests have shown that receivers based on the RR model outperform methods based on the structured approach in all practical cases [15].

In this paper, the RR approach is investigated for the estimation of the uplink (from mobile to base station) channel in time-slotted mobile communication systems, where the transmission is organized in data packets and each packet includes a training block of N symbols. The paper shows that the RR estimate proposed in [6] can be equivalently obtained as the projection of the whitened unconstrained MLE for the channel matrix $\mathbf{H}$ onto the spatial *or* temporal signal subspace. The rank order r_0 of the space-time matrix $\mathbf{H}$ is selected from the received signals as a tradeoff between distortion (due to under-parameterization) and variance (due to noise) or, equivalently, between simplicity of the model and reliability of the estimate [16]. For known rank order, the mean square error (MSE) of the estimate is derived in a new closed form, and it is shown to be uniformly lower than the MSE of the unconstrained MLE. In particular, for white noise and uncorrelated training sequence, the MSE depends on the ratio between the number of independent parameters that describe the low-rank channel matrix and the number of training symbols N. The MSE of the estimate is also evaluated for signals in temporally correlated noise showing the degradation due to noise mismodeling.

The new formulation proposed in this paper for the RR estimate allows the extension of the RR approach to the following cases: i) the estimation of the spatial or temporal structure of the channel matrix from multiple training blocks (see Section IV), and ii) the estimation of a multiuser channel through the redefinition of the whitening terms only (see Section V).

With regard to the first extension, for block-fading channels, the RR approach allows us to decrease the number of unknowns and to (virtually) extend the training length by constraining the temporal (or the spatial) subspace to be stationary across several blocks. Different approaches for multiblock channel estimation have been proposed in the literature. Structured methods exploit the invariance of angles and delays to explicitly estimate the path parameters from multiple blocks [17]. On the other hand, unstructured approaches translate the stationarity of angles/delays into the invariance of the spatial/temporal subspaces spanned by the corresponding channel signatures. In particular, subspace methods have been proposed to take advantage of the invariance of either the spatial [18], [19] or the temporal [19] subspace of the channel matrix. All these unstructured methods are derived in this paper in a common framework as an extension of the RR estimate to multiple training blocks. Depending on the specific multiblock method, the channel estimate is obtained by partitioning the low-rank channel matrix into a spatial (*or* temporal) time-varying component and a temporal (*or* spatial) stationary term. The MSE of both the multiblock estimators is calculated in closed form for *any* number of blocks showing that the convergence rate for the multiblock algorithms is related to the propagation and interference environments. For a large number of blocks, the analytical result coincides with the MSE bound derived in [19].

In the literature, RR channel estimation has been investigated mainly for single-user systems [such as the Global System for Mobile Communications (GSM)]. An extension to multiuser systems [such as the Universal Mobile Telecommunications System time-division duplex (UMTS-TDD) [20]] can be found in [4] and [21]. In this paper, we show how the channels of all users can be estimated *jointly* with the constraint that the space-time matrix for *each* user is low-rank. The method can effectively cope with multiple access interference (due to the non orthogonality of the training sequences) and co-channel interference from neighboring cells. As the rank of each channel depends on the interference level and the characteristics of the propagation environment (delay and angle spreads), the rank order needs to be estimated adaptively for each user.

In all the considered cases, the RR estimate is obtained through the singular value decomposition (SVD) of the whitened unconstrained [or full-rank (FR)] estimate. The computational complexity required by this approach is larger that the one of the conventional FR method. To lower the complexity, the SVD can be implemented by appropriate fast algorithms [22], or it can be approximated by subspace tracking (such as [23]–[25]) or alternating power [5] methods. These approximate methods are computationally more efficient than the exact evaluation of the eigenvectors, especially in adaptive multiblock estimation, where the subspaces are updated with additional data from successive data blocks (see [19]).

The coverage of all these topics is organized as follows. The essential theory for RR channel estimation is presented in the first part of the paper; the RR signal model is introduced in Section II; in Section III, RR channel estimation and order selection are investigated; the performance analysis is in Section III-C, the evaluation of the performance degradation for temporally correlated noise is in Section III-D. In the second part of the paper, the estimation method is extended to handle multiple block (Section IV) and multiple user (Section V). Two multiblock estimators are proposed, and their performance is evaluated analytically as a function of the number of blocks. Numerical results are provided in Section VI to validate the analytical results and the underlying theory, and the concluding remarks are in Section VII. The notational conventions used throughout the paper can be found in the Appendixes.

II. RR Signal Model

According to the multipath model for propagation, the channel response can be described as the superposition of P path contributions. The space-time matrix $\mathbf{H}$ is thus a function of $3P$ path parameters that will be considered to be stationary during the training period: the directions of arrival $\boldsymbol{\vartheta} = [\vartheta_1 \cdots \vartheta_P]$, the times of delay $\boldsymbol{\tau} = [\tau_1 \cdots \tau_P]$, and the complex-valued fading amplitudes $\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_P]$. The channel matrix can be expressed as

$$\mathbf{H} = \sum_{p=1}^{P} \alpha_p \mathbf{a}(\vartheta_p)\mathbf{g}^{\mathrm{T}}(\tau_p) = \mathbf{F}(\boldsymbol{\vartheta})\boldsymbol{\Lambda}(\boldsymbol{\alpha})\mathbf{G}^{\mathrm{T}}(\boldsymbol{\tau}). \qquad (2)$$

Here, the $W \times 1$ real-valued vector $\mathbf{g}(\tau_p) = [g(-\tau_p)g(T-\tau_p) \cdots g((W-1)T-\tau_p)]^{\mathrm{T}}$ contains samples of the delayed waveform $g(\tau)$ that represents the convolution between the transmitted pulse and the matched filter at the receiver. Similarly, the complex-valued vector $\mathbf{a}(\vartheta_p) = [a_1(\vartheta_p) \cdots a_M(\vartheta_p)]^{\mathrm{T}}$ denotes the $M \times 1$ array response to the narrowband signal impinging from the direction ϑ_p. For a linear array of half-wave-length spaced omnidirectional antennas, the entries of $\mathbf{a}(\vartheta_p)$ are $a_m(\vartheta_p) = \exp(-j\pi(m-1)\sin\vartheta_p)$. By collecting the set of P temporal/spatial vectors into the temporal/spatial matrices $\mathbf{G} = [\mathbf{g}(\tau_1) \ \ldots \ \mathbf{g}(\tau_P)]$ and $\mathbf{F}(\boldsymbol{\vartheta}) = [\mathbf{a}(\vartheta_1) \cdots \mathbf{a}(\vartheta_P)]$, the multipath formulation for the space-time channel matrix simplifies as indicated in the third member of (2), where the diagonal matrix $\boldsymbol{\Lambda}(\boldsymbol{\alpha}) = \mathrm{diag}[\alpha_1, \ldots, \alpha_P]$ embodies the fading amplitudes.

Let the spatial (r_{S}) and temporal (r_{T}) rank orders be defined as

$$r_{\mathrm{S}} = \mathrm{rank}\,[\mathbf{F}(\boldsymbol{\vartheta})] \le \min(P, M) \tag{3a}$$

$$r_{\mathrm{T}} = \mathrm{rank}\,[\mathbf{G}(\boldsymbol{\tau})] \le \min(P, W). \tag{3b}$$

From (2), it follows that the rank order of the channel matrix $r_0 = \mathrm{rank}[\mathbf{H}]$ is

$$r_0 = \min(r_{\mathrm{S}}, r_{\mathrm{T}}) \le \min(M, W) = r_{\max} \tag{4}$$

where $r_{\max}$ denotes the maximum rank order. Notice that r_{S} and r_{T} equal, respectively, the number of angles and delays that the system can resolve in the multipath pattern (namely, the spatial and temporal diversity orders). They depend on the angular and delay spread compared to the resolution of the antenna array (that is related to the array geometry and the number of antennas) and the resolution of the pulse waveform (that is related to the signal bandwidth).

Let us assume that the number of resolvable angles is $r_{\mathrm{S}} < r_{\max}$ (or, dually, that the number of resolvable delays is $r_{\mathrm{T}} < r_{\max}$), the channel matrix $\mathbf{H}$ is rank-deficient, and, thus, the number of independent entries in $\mathbf{H}$ is less than MW. This occurs when the paths in (2) can be grouped into few groups of scatterers (clusters) having angular (or delay) spread below the system resolution. In this case, the rank order r_0 can be approximated by the number of independent clusters that the system can resolve either in time or space. For instance, in the first example of Fig. 1(a), the rank order is determined by the spatial domain as the $P = 3$ paths are grouped into a single scattering cluster that can no longer be resolved in space as $\vartheta_1 = \vartheta_2 = \vartheta_3$: It is $r_0 = r_{\mathrm{S}} = 1, r_{\mathrm{T}} = 3$. In the second example (see the right side of Fig. 1), the $P = 3$ paths have all similar delays so that the rank order is given by $r_0 = r_{\mathrm{T}} = 1$, while the angular diversity order is $r_{\mathrm{S}} = 3$.

In the following, the channel matrix $\mathbf{H}$ is assumed to have rank

$$r_0 < r_{\max} \tag{5}$$

(as $r_0 = r_{\max}$ degenerates to known estimation methods), and it is rewritten as the product of two full column rank matrices $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_{r_0}]$ and $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_{r_0}]$:

$$\mathbf{H} = \mathbf{A}\mathbf{B}^{\mathrm{H}}. \tag{6}$$

The $M \times r_0$ spatial component $\mathbf{A}$ and the $W \times r_0$ temporal component $\mathbf{B}$ parameterize the spatial and the temporal structure of

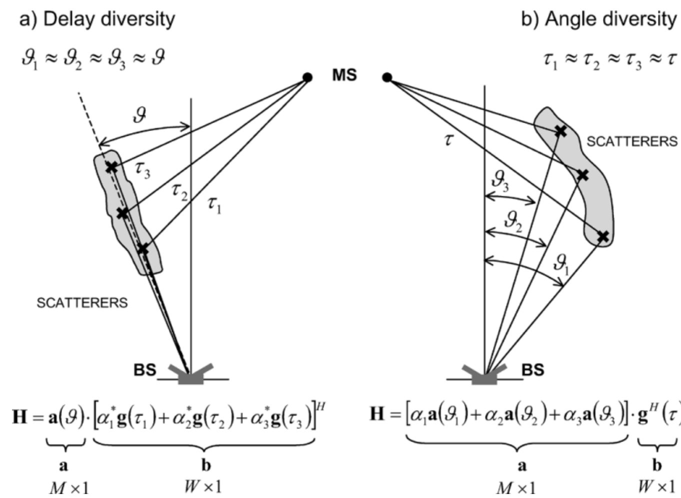

Fig. 1. Examples of rank-1 ($r_0 = 1$) channel for the link from mobile station (MS) to base station (BS). The channel is composed of $P = 3$ paths having similar angles or delays. The diversity orders are $r_{\mathrm{S}} = 1, r_{\mathrm{T}} = 3$ on the left and $r_{\mathrm{S}} = 3, r_{\mathrm{T}} = 1$ on the right. (a) Delay diversity. (b) Angle diversity.

the propagation channel, respectively. Each spatial (or temporal) vector $\mathbf{a}_\ell$ (or $\mathbf{b}_\ell$) can be obtained as a linear combination of the spatial (or temporal) signatures for all paths composing the ℓth cluster, and thus, it represents the equivalent signature for the cluster. As an example, a rank-1 channel $\mathbf{H} = \mathbf{a}\mathbf{b}^{\mathrm{H}}$ is obtained from (2) when the angular spread is null so that $\vartheta_p = \vartheta, \forall p$ (single cluster channel). The channel is described by the spatial signature $\mathbf{a} = \mathbf{a}(\vartheta)$ and the temporal filter $\mathbf{b} = \sum_{p=1}^{P} \alpha_p^* \mathbf{g}(\tau_p)$ (see the left side Fig. 1). The same holds if the channel has no delay spread, namely, $\tau_p = \tau, \forall p$. In this case, it is $\mathbf{a} = \sum_{p=1}^{P} \alpha_p \mathbf{a}(\vartheta_p)$ and $\mathbf{b} = \mathbf{g}(\tau)$ (see the right side of Fig. 1).

III. RR Channel Estimation

The signal model (1) can be rewritten into the conventional form

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N} \tag{7}$$

where the $M \times N$ matrix $\mathbf{Y} = [\mathbf{y}(0) \cdots \mathbf{y}(N-1)]$ collects N time samples of the received signal and $\mathbf{X} = [\mathbf{x}(0) \cdots \mathbf{x}(N-1)]$ is the $W \times N$ Toeplitz matrix for the convolution with the training sequence, while $\mathbf{N} = [\mathbf{n}(0) \cdots \mathbf{n}(N-1)]$ collects the noise samples. It is assumed that $N > M + W$.

The problem addressed in the following is the joint estimation of the channel matrix $\mathbf{H}$ and the noise covariance $\mathbf{Q}$ from the received signals $\mathbf{Y}$ and the known training sequence $\mathbf{X}$, under the assumption that the matrix $\mathbf{H}$ has rank $r_0 < r_{\max}$.

A. Method

According to the signal model (7), the normalized negative log-likelihood function is given by

$$\mathcal{L}(\mathbf{H}, \mathbf{Q}) = \log|\mathbf{Q}| + \frac{1}{N}\|\mathbf{Y} - \mathbf{H}\mathbf{X}\|^2_{\mathbf{Q}^{-1}} \tag{8}$$

where $\|\mathbf{Y} - \mathbf{H}\mathbf{X}\|^2_{\mathbf{Q}^{-1}} = \mathrm{tr}[\mathbf{Q}^{-1}(\mathbf{Y} - \mathbf{H}\mathbf{X})(\mathbf{Y} - \mathbf{H}\mathbf{X})^{\mathrm{H}}]$. The RR MLE $\{\hat{\mathbf{H}}, \hat{\mathbf{Q}}\}$ is given by the joint minimization of (8) with respect to $\mathbf{H}$ and $\mathbf{Q}$ under the RR constraint (5). The rank order r_0 is assumed to be known. Notice that for $r_0 = r_{\max}$, the solution reduces to the unconstrained FR MLE

$$\mathbf{H}_{\rm u} = \mathbf{R}_{yx}\mathbf{R}_{xx}^{-1} \tag{9a}$$

$$\mathbf{Q}_{\rm u} = \frac{1}{N}\left(\mathbf{R}_{yy} - \mathbf{R}_{yx}\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}\right) \tag{9b}$$

where $\mathbf{R}_{yx} = \mathbf{Y}\mathbf{X}^{\rm H}$, $\mathbf{R}_{xx} = \mathbf{X}\mathbf{X}^{\rm H}$, and $\mathbf{R}_{yy} = \mathbf{Y}\mathbf{Y}^{\rm H}$ are sample correlation matrices ($\mathbf{R}_{xx}$ is assumed to be positive definite). The estimate (9a) coincides with the least squares estimate (LSE) of the space-time channel matrix.

The RR estimate is reproposed below by two alternative expressions fully equivalent to [6].

Proposition 1: The joint MLE of $\{\mathbf{H}, \mathbf{Q}\}$ under the constraint (5) is

$$\hat{\mathbf{H}} = \mathbf{Q}_{\rm u}^{\frac{\rm H}{2}}\left(\hat{\mathbf{P}}_{{\rm S},r_0}\tilde{\mathbf{H}}_{\rm u}\right)\mathbf{R}_{xx}^{-\frac{\rm H}{2}} \tag{10a}$$

$$= \mathbf{Q}_{\rm u}^{\frac{\rm H}{2}}\left(\tilde{\mathbf{H}}_{\rm u}\hat{\mathbf{P}}_{{\rm T},r_0}\right)\mathbf{R}_{xx}^{-\frac{\rm H}{2}} \tag{10b}$$

$$\hat{\mathbf{Q}} = \frac{1}{N}(\mathbf{Y} - \hat{\mathbf{H}}\mathbf{X})(\mathbf{Y} - \hat{\mathbf{H}}\mathbf{X})^{\rm H} \tag{11}$$

where $\tilde{\mathbf{H}}_{\rm u}$ indicates the whitened FR channel estimate

$$\tilde{\mathbf{H}}_{\rm u} = \mathbf{Q}_{\rm u}^{-\frac{\rm H}{2}}\mathbf{H}_{\rm u}\mathbf{R}_{xx}^{\frac{\rm H}{2}} \tag{12}$$

and $\hat{\mathbf{P}}_{{\rm S},r_0}$ and $\hat{\mathbf{P}}_{{\rm T},r_0}$ denote, respectively, the projectors onto the subspaces spanned by the r_0 leading eigenvectors of the spatial ($\hat{\mathbf{R}}_{\rm S}$) and temporal ($\hat{\mathbf{R}}_{\rm T}$) correlation matrices:[1]

$$\hat{\mathbf{R}}_{\rm S} = \tilde{\mathbf{H}}_{\rm u}\tilde{\mathbf{H}}_{\rm u}^{\rm H} = \mathbf{Q}_{\rm u}^{-\frac{\rm H}{2}}\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}\mathbf{Q}_{\rm u}^{-\frac{1}{2}} \tag{13a}$$

$$\hat{\mathbf{R}}_{T} = \tilde{\mathbf{H}}_{\rm u}^{\rm H}\tilde{\mathbf{H}}_{\rm u} = \mathbf{R}_{xx}^{-\frac{\rm H}{2}}\mathbf{R}_{xy}\mathbf{Q}_{\rm u}^{-1}\mathbf{R}_{yx}\mathbf{R}_{xx}^{-\frac{1}{2}}. \tag{13b}$$

Proof: The RR MLE formulated as in (10a) is derived in Appendix A. ■

The RR solution depends on the matrix $\tilde{\mathbf{H}}_{\rm u}$ that is referred to as the whitened channel estimate, as it is $\text{Cov}[\text{vec}[\tilde{\mathbf{H}}_{\rm u}]] \to \mathbf{I}_{MW}$ for $N \to \infty$ [4]. Let the SVD of $\tilde{\mathbf{H}}_{\rm u}$ be

$$\tilde{\mathbf{H}}_{\rm u} = \mathbf{U}_{\rm S}\boldsymbol{\Sigma}\mathbf{U}_{\rm T}^{\rm H} = \left[\mathbf{U}_{{\rm S},r_0}\ \ \mathbf{U}_{{\rm S},r_0}^{\perp}\right]\begin{bmatrix}\boldsymbol{\Sigma}_{r_0} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Sigma}_{r_0}^{\perp}\end{bmatrix}\left[\mathbf{U}_{{\rm T},r_0}\ \ \mathbf{U}_{{\rm T},r_0}^{\perp}\right]^{\rm H} \tag{14}$$

where $\mathbf{U}_{{\rm S},r_0}$, and $\mathbf{U}_{{\rm T},r_0}$ denote, respectively, the left and right eigenvectors corresponding to the r_0 largest singular values $\boldsymbol{\Sigma}_{r_0} = \text{diag}[\sigma_1, \ldots, \sigma_{r_0}]$ arranged in nonincreasing order. The spatial and temporal projectors are

$$\hat{\mathbf{P}}_{{\rm S},r_0} = \mathbf{P}_{\hat{\mathbf{R}}_{\rm S},r_0} = \mathbf{U}_{{\rm S},r_0}\mathbf{U}_{{\rm S},r_0}^{\rm H} \tag{15a}$$

$$\hat{\mathbf{P}}_{{\rm T},r_0} = \mathbf{P}_{\hat{\mathbf{R}}_{\rm T},r_0} = \mathbf{U}_{{\rm T},r_0}\mathbf{U}_{{\rm T},r_0}^{\rm H}. \tag{15b}$$

From the definitions above and (14), it follows that $\hat{\mathbf{P}}_{{\rm S},r_0}\tilde{\mathbf{H}}_{\rm u} = \tilde{\mathbf{H}}_{\rm u}\hat{\mathbf{P}}_{{\rm T},r_0} = \mathbf{U}_{{\rm S},r_0}\boldsymbol{\Sigma}_{r_0}\mathbf{U}_{{\rm T},r_0}^{\rm H}$, which easily proves the equivalence between (10a) and (10b). The constrained channel estimation can thus be performed either in the spatial domain [with (10a)] or in the temporal domain [with (10b)] by evaluating the signal subspace from (13a) or (13b), respectively.

Furthermore, the terms within brackets in (10a), or in (10b), are fully equivalent to the SVD of $\tilde{\mathbf{H}}_{\rm u}$ truncated to the first r_0 singular values:

$$\hat{\mathbf{H}} = \mathbf{Q}_{\rm u}^{\frac{\rm H}{2}}\ \text{svd}_{r_0}[\tilde{\mathbf{H}}_{\rm u}]\mathbf{R}_{xx}^{-\frac{\rm H}{2}}. \tag{16}$$

[1]The sample estimate for $\{\hat{\tilde{\mathbf{R}}}_{\rm S,u}, \hat{\tilde{\mathbf{R}}}_{\rm T,u}\}$ is here briefly indicated as $\{\hat{\mathbf{R}}_{\rm S}, \hat{\mathbf{R}}_{\rm T}\}$, instead of $\{\hat{\tilde{\mathbf{R}}}_{\rm S,u}, \hat{\tilde{\mathbf{R}}}_{\rm T,u}\}$, just to simplify the notation.

This highlights that the RR MLE can be interpreted (and eventually implemented) as follows: i) pre-whitening of the full-rank channel estimate (left multiplication by $\mathbf{Q}_{\rm u}^{-{\rm H}/2}$ and right multiplication by $\mathbf{R}_{xx}^{{\rm H}/2}$ of $\mathbf{H}_{\rm u}$ to obtain $\tilde{\mathbf{H}}_{\rm u}$); ii) truncation of the SVD of $\tilde{\mathbf{H}}_{\rm u}$ to the r_0 largest singular values; iii) cancellation of the prewhitening (left multiplication by $\mathbf{Q}_{\rm u}^{{\rm H}/2}$ and right multiplication by $\mathbf{R}_{xx}^{-{\rm H}/2}$).

Remark 1: The RR problem was first solved by Stoica and Viberg [6] as an unconstrained estimation by replacing $\mathbf{H}$ with (6) in the expression (8) and then maximizing the likelihood function with respect to $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{Q}$. Similarly to (10b), the estimate therein obtained for $\mathbf{H}$ is a projection in the temporal domain where the spatial whitening $\mathbf{Q}_{\rm u}^{-{\rm H}/2}$ is replaced with the term $\mathbf{R}_{yy}^{-{\rm H}/2}$. Even if the whitening is different, the solution is fully equivalent to (10b) (see [4]), and it can be also rewritten into the dual formulation in the spatial domain, as in (10a) (see Appendix B). The solutions (10a) and (10b) have the advantage that they can be easily extended to multiuser systems (see Section V).

Remark 2: The estimate $\hat{\mathbf{H}}$ is also the minimizer of the norm

$$\mathcal{F} \triangleq \left\|\tilde{\mathbf{H}}_{\rm u} - \mathbf{Q}_{\rm u}^{-\frac{\rm H}{2}}\mathbf{H}\mathbf{R}_{xx}^{\frac{\rm H}{2}}\right\|^2 \tag{17}$$

(or, equivalently, the minimizer of $\|\mathbf{Y} - \mathbf{H}\mathbf{X}\|_{\mathbf{Q}_{\rm u}^{-1}}^2 = \mathcal{F} + M$). Indeed, as shown in Appendix A, by selecting the RR estimate $\hat{\mathbf{H}}$, each eigenvalue $\lambda_k[\mathbf{G}]$ of the matrix $\mathbf{G} = (\tilde{\mathbf{H}}_{\rm u} - \mathbf{Q}_{\rm u}^{-{\rm H}/2}\mathbf{H}\mathbf{R}_{xx}^{{\rm H}/2})^{\rm H}(\tilde{\mathbf{H}}_{\rm u} - \mathbf{Q}_{\rm u}^{-{\rm H}/2}\mathbf{H}\mathbf{R}_{xx}^{{\rm H}/2})$ reaches its minimum value. Thus, the optimizer of $\mathcal{L}$ in (82) is also the minimizer of $\mathcal{F} = \text{tr}[\mathbf{G}] = \sum_{k=1}^{r_{\max}}\lambda_k[\mathbf{G}]$. Notice that for a generic structured estimation, minimizing $\mathcal{L}$ is not equivalent to minimizing the trace $\mathcal{F}$, and the equivalence holds only asymptotically for $N \to \infty$ (see [26] for the proof).

Remark 3: From (16), it can be seen that the additional computational complexity required by the RR method with respect to the FR one is in the operations of whitening/dewhitening and the SVD (the overall complexity is comparable with the one of the RR method [6]). More precisely, the implementation of (16) requires the computation of the spatial/temporal whitening factors ($O(W^3) + O(M^3)$ flops), the evaluation of $\tilde{\mathbf{H}}_{\rm u}$ from $\mathbf{H}_{\rm u}$ ($O(W^2M) + O(M^2W)$ flops), the computation of the SVD of $\tilde{\mathbf{H}}_{\rm u}$ truncated to the first r_0 singular values, and finally, the cancellation of the whitening $(O(W^2M) + O(M^2W)$ flops$)$. Efficient algorithms can be used to compute the SVD, such as the alternated power (AP) method, that can be implemented with $O(MWr_0) + O(Mr_0^2) + O(Wr_0^2)$ flops for each iteration (the number of required operations is usually very small). Notice that for $r_0 \ll \min(M, W)$, the SVD cost becomes very small, and the computational burden reduces to the operations required before and after the SVD. Some of these operations can be avoided by precalculating the terms depending on $\mathbf{R}_{xx}$ (that are fixed). Still, for large r_0, the overall RR computational load is considerably larger than the FR one.

B. Selection of the Rank Order

In real systems, the rank order r_0 of the propagation channel is not known, and an estimate r has to be derived from the available measurements. The choice $r = r_0$ is the minimum order that yields an unbiased channel estimate, but in general, it does not provide the lowest MSE. As demonstrated in the following, for low signal-to-noise ratio (SNR), it might be more convenient to choose a biased estimator (i.e., $r < r_0$) and trade distortion for estimate error (see also [16]).

Assuming $N \to \infty$, define the "true" whitened channel and spatial correlation matrices as $\tilde{\mathbf{H}} = \mathbf{Q}^{-\mathrm{H}/2}\mathbf{H}\mathbf{R}_{xx}^{\mathrm{H}/2}$ and $\tilde{\mathbf{R}}_{\mathrm{S}} = \tilde{\mathbf{H}}\tilde{\mathbf{H}}^{\mathrm{H}}$, respectively, and let r denote the order adopted to calculate the RR channel estimator (10a). From (7) and (9a), the error for the FR estimate is

$$\Delta\mathbf{H}_{\mathrm{u}} = \mathbf{H}_{\mathrm{u}} - \mathbf{H} = \mathbf{N}\mathbf{X}^{\mathrm{H}}\mathbf{R}_{xx}^{-1}. \quad (18)$$

By using the latter result and recalling the definition for the whitened FR channel estimate (12), the RR channel estimator (10a) for rank order r can be rewritten as

$$\begin{aligned}\hat{\mathbf{H}}(r) &= \mathbf{Q}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}\hat{\mathbf{P}}_{\mathrm{S},r}\mathbf{Q}_{\mathrm{u}}^{-\frac{\mathrm{H}}{2}}\mathbf{H}_{\mathrm{u}}\\ &= \mathbf{Q}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}\hat{\mathbf{P}}_{\mathrm{S},r}\mathbf{Q}_{\mathrm{u}}^{-\frac{\mathrm{H}}{2}}\mathbf{H} + \mathbf{Q}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}\hat{\mathbf{P}}_{\mathrm{S},r}\mathbf{Q}_{\mathrm{u}}^{-\frac{\mathrm{H}}{2}}\Delta\mathbf{H}_{\mathrm{u}}. \quad (19)\end{aligned}$$

It follows that the asymptotic error $\Delta\mathbf{H}(r) = \hat{\mathbf{H}}(r) - \mathbf{H}$ is split into two terms:

$$\Delta\mathbf{H}(r) = \underbrace{\mathbf{Q}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}(\hat{\mathbf{P}}_{\mathrm{S},r} - \mathbf{P}_{\mathrm{S}})\mathbf{Q}_{\mathrm{u}}^{-\frac{\mathrm{H}}{2}}\mathbf{H}}_{\Delta\mathbf{H}_{\mathrm{d}}(r)} + \underbrace{\mathbf{Q}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}\hat{\mathbf{P}}_{\mathrm{S},r}\mathbf{Q}_{\mathrm{u}}^{-\frac{\mathrm{H}}{2}}\Delta\mathbf{H}_{\mathrm{u}}}_{\Delta\mathbf{H}_{\mathrm{n}}(r)} \quad (20)$$

where the spatial projector $\mathbf{P}_{\mathrm{S}}$ is the orthogonal projector onto the r_0-dimensional subspace $\mathcal{R}[\tilde{\mathbf{R}}_{\mathrm{S}}]$. Since $\mathbf{Q}_{\mathrm{u}} - \mathbf{Q} = O(1/N)$, in (20), it is assumed that $\mathbf{Q}_{\mathrm{u}}^{\mathrm{H}/2}\mathbf{P}_{\mathrm{S}}\mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}\mathbf{H} \to \mathbf{H}$ for $N \to \infty$.

The first term in (20) $\Delta\mathbf{H}_{\mathrm{d}}(r)$ accounts for the distortion of the RR estimate and depends on the difference $\mathbf{P}_{\mathrm{S}} - \hat{\mathbf{P}}_{\mathrm{S},r}$. Notice that for finite SNR, there is a mismatch (for any r) between the estimated projector $\hat{\mathbf{P}}_{\mathrm{S},r}$ and the true projector $\mathbf{P}_{\mathrm{S}}$, as the former is calculated from the matrix $\hat{\mathbf{R}}_{\mathrm{S}}$ that, in general, does not span the same subspace as $\tilde{\mathbf{R}}_{\mathrm{S}}$: $\mathcal{R}_r[\hat{\mathbf{R}}_{\mathrm{S}}] \neq \mathcal{R}[\tilde{\mathbf{R}}_{\mathrm{S}}]$. This bias decreases for increasing r. In particular, if $r = r_0$ the estimate for the signal subspace is $\mathcal{R}_r[\hat{\mathbf{R}}_{\mathrm{S}}] \to \mathcal{R}[\tilde{\mathbf{R}}_{\mathrm{S}}]$ for SNR $\to \infty$, which implies $\hat{\mathbf{P}}_{\mathrm{S},r} \to \mathbf{P}_{\mathrm{S}}$, and therefore, $\Delta\mathbf{H}_{\mathrm{d}}(r) = 0$. Similarly, for any $r > r_0$, it is $\mathcal{R}[\hat{\mathbf{R}}_{\mathrm{S}}] \supseteq \mathcal{R}[\tilde{\mathbf{R}}_{\mathrm{S}}]$, and again, $\Delta\mathbf{H}_{\mathrm{d}}(r) = 0$. Hence, for $r \geq r_0$, the distortion can be considered to be negligible, and the error is dominated by $\Delta\mathbf{H}_{\mathrm{n}}(r)$. This second term $\Delta\mathbf{H}_{\mathrm{n}}(r)$ depends mainly on the noise, and it increases with the noise power. Since it is obtained as a projection of the noise onto $\mathcal{R}_r[\hat{\mathbf{R}}_{\mathrm{S}}]$, the mean square value of the Frobenius norm for $\Delta\mathbf{H}_{\mathrm{n}}(r)$ increases with the subspace dimension r.

The reasoning above implies that there is a tradeoff between distortion (due to under-parameterization) and variance (due to over-parameterization) that minimizes the overall MSE of the estimate $\mathrm{E}[\|\Delta\mathbf{H}(r)\|^2]$. The choice $r = r_0$ provides low $\Delta\mathbf{H}_{\mathrm{d}}(r)$ but could lead to high MSE for low SNR, as the number of parameters to be estimated is large. Conversely, the choice $r = 1$, which corresponds to the simplest parameterization, minimizes the variance but it is distortion limited. The appropriate rank r, with $1 \leq r \leq r_0$, is the one that yields the best distortion/variance tradeoff (or parameterization simplicity/complexity) and minimizes the overall error. In particular, for low SNR, it is convenient to accept a little bias in order to gain a reduction of the overall error, whereas for large SNR, r should approach the effective value r_0.

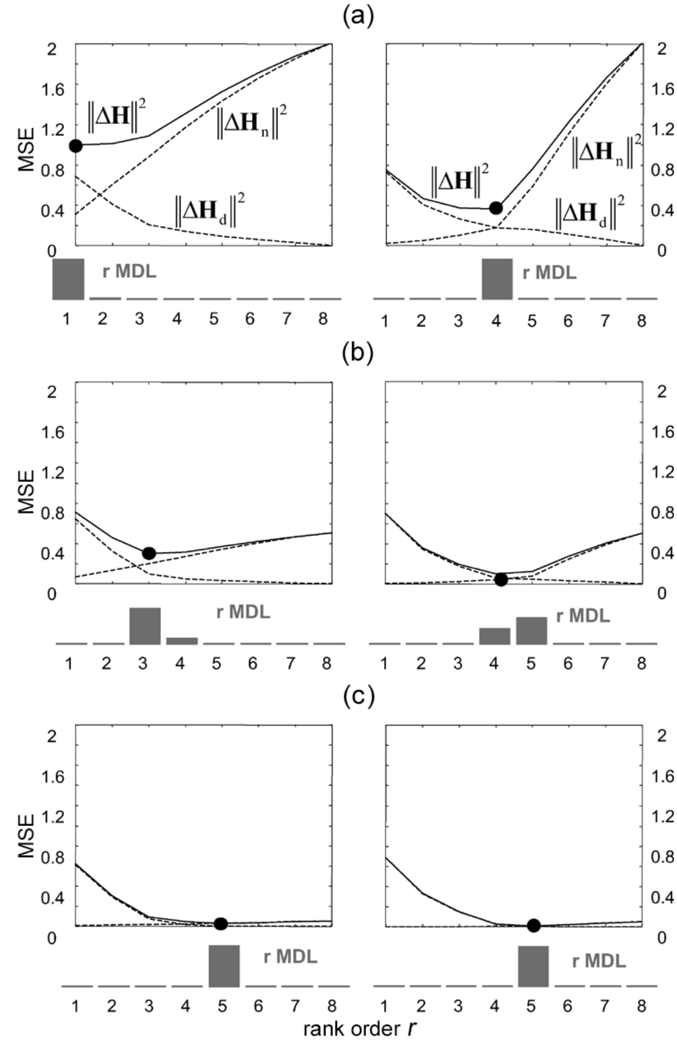

Fig. 2. MSE of the RR channel estimate versus the rank order r used for the estimation. (Left) White noise and (Right) Spatially correlated noise $M = 8$, $W = 57$, $r_0 = 5$. (a) SNR = −3 dB. (b) SNR = 3 dB. (c) SNR = 13 dB.

The model order r can be selected by using any method (e.g., Akaike [27] or MDL criterion [28]) aimed to estimate the number of uncorrelated sources from the analysis of the eigenvalues of $\hat{\mathbf{R}}_{\mathrm{S}}$. In the following example, the MDL estimate is adopted, and it is calculated as the minimizer of [28]

$$\mathrm{MDL}(r) = -W(M-r)\log\left(\frac{\prod_{i=r+1}^{M}\lambda_i^{\frac{1}{M-r}}}{\frac{1}{M-r}\sum_{i=r+1}^{M}\lambda_i}\right) + \frac{1}{2}r(2M-r)\log W \quad (21)$$

where $\lambda_i = \lambda_i[\hat{\mathbf{R}}_{\mathrm{S}}]$ are the eigenvalues of the matrix (13a). This criterion is shown to provide a reasonable tradeoff between distortion and variance.

Example 1: The channel matrix $\mathbf{H}$ is randomly generated with $M = 8$, $W = 57$, $r_0 = 5$, and $\mathrm{E}[\|\mathbf{H}\|^2] = 1$. A training sequence with $N = 456$ quadrature phase shift keying (QPSK) symbols is chosen from UMTS-TDD specifications [20] but for a single-user environment. The result of the comparison is shown in Fig. 2 for white noise (left) and spatially correlated (right) noise with covariance $[\mathbf{R}]_{m,\ell} =$

$\sigma^2 0.9^{|\ell-m|}\exp[-i\pi(\ell-m)/\sqrt{2}]$. Each simulated MSE value is the result of 500 independent runs of channel and noise. The plots (a)-(c) show the MSE of the RR estimate versus the order used for rank reduction, for SNR equal to −3, 3, and 13 dB, respectively. The SNR is defined as $\text{SNR} = \text{E}[||\mathbf{H}||^2]/\sigma^2$ (here, it is $\sigma_x^2 = 1$). Both $\text{E}[||\Delta\mathbf{H}_\text{n}||^2]$ and $\text{E}[||\Delta\mathbf{H}_\text{d}||^2]$ are represented, together with the total MSE $\text{E}[||\Delta\mathbf{H}||^2]$. As expected, for increasing r, the distortion decreases and the noise variance increases. For large SNR, the performance tends to the ideal case of known signal subspace. The distribution of the MDL estimate is shown below each graph as a histogram. Notice that when the noise is white, for low SNR, the optimum rank order is $r = 1$ [plot (a)], while for large SNR, it moves toward the correct rank order $r = 5$ [plot (c)]. For low SNR but correlated noise, this property does not hold.

C. Performance Analysis

The performance of the RR estimate is evaluated in terms of MSE of the estimate for large N and known rank order $r = r_0$. The proof of consistency and the derivation of the asymptotic covariance of the estimate are given in [6]; the result is recalled in the following proposition.

Proposition 2: Considering the constrained MLE (10a) and defining the vector $\hat{\mathbf{h}} = \text{vec}[\hat{\mathbf{H}}]$, the estimate is unbiased, and it is asymptotically (for $N \to \infty$) normally distributed with zero mean and covariance matrix

$$\text{Cov}[\hat{\mathbf{h}}] = \mathbf{D}\mathbf{J}^\dagger\mathbf{D}^\text{H} \tag{22}$$

where the $MW \times r(M+W)$ matrix $\mathbf{D}$ and the $r(M+W) \times r(M+W)$ matrix $\mathbf{J}$ are defined as

$$\mathbf{D} = [\mathbf{B}^* \otimes \mathbf{I}_M \quad \mathbf{I}_W \otimes \mathbf{A}] \tag{23}$$
$$\mathbf{J} = \mathbf{D}^\text{H}\left(\mathbf{R}_{xx}^* \otimes \mathbf{Q}^{-1}\right)\mathbf{D} \tag{24}$$

and, according to the rank-r_0 model (6), $\mathbf{A}$ and $\mathbf{B}$ are the spatial and temporal components of the channel matrix $\mathbf{H}$.

Proof: See [6]. ■

The asymptotic MSE calculated from (22)

$$\text{MSE}_\text{RR} = \text{E}\left[||\hat{\mathbf{H}} - \mathbf{H}||^2\right] = \text{tr}[\mathbf{D}\mathbf{J}^\dagger\mathbf{D}^\text{H}] \tag{25}$$

depends on the pseudoinverse of the matrix $\mathbf{J}$. The latter matrix is rank deficient as $\text{rank}[\mathbf{J}] = \text{rank}[\mathbf{D}] = r_0(M+W) - r_0^2$ (see the proof in [6]). For large samples, the right-hand side of (22) coincides with the Cramér–Rao lower bound, and $\mathbf{J}$ equals the Fisher information matrix of the parameter vector $\boldsymbol{\theta} = \left[\text{vec}[\mathbf{A}]^\text{T}\text{vec}[\mathbf{B}^\text{H}]^\text{T}\right]^\text{T}$.

For $r_0 = r_\text{max}$, the MLE of the channel reduces to the LSE: $\hat{\mathbf{H}} = \mathbf{H}_\text{u}$. In this case, the matrix $\mathbf{D}$ is full-row rank as $\text{rank}[\mathbf{D}] = r_\text{max}(M+W) - r_\text{max}^2 = MW$, and therefore, it is $\mathbf{D}\mathbf{J}^\dagger\mathbf{D}^\text{H} = \mathbf{D}\mathbf{D}^\dagger(\mathbf{R}_{xx}^* \otimes \mathbf{Q}^{-1})^{-1}\mathbf{D}^{\text{H}\dagger}\mathbf{D}^\text{H} = (\mathbf{R}_{xx}^{-1})^* \otimes \mathbf{Q}$. From (22), the covariance matrix $\mathbf{R}_\text{u} = \text{Cov}[\mathbf{h}_\text{u}]$ for the channel estimator $\mathbf{h}_\text{u} = \text{vec}[\mathbf{H}_\text{u}]$ simplifies as

$$\mathbf{R}_\text{u} = \mathbf{R}_{\text{u,T}}^* \otimes \mathbf{R}_{\text{u,S}} = \left(\mathbf{R}_{xx}^{-1}\right)^* \otimes \mathbf{Q}. \tag{26}$$

The corresponding MSE can be expressed as

$$\text{MSE}_\text{u} = \text{E}\left[||\mathbf{H}_\text{u} - \mathbf{H}||^2\right] = \Phi\left[\mathbf{I}_W, \mathbf{R}_{xx}^{-1}\right]\Phi[\mathbf{I}_M, \mathbf{Q}] \tag{27}$$

where $\Phi[\mathbf{P},\mathbf{T}] = \text{tr}[\mathbf{T}^{\text{H}/2}\mathbf{P}\mathbf{T}^{1/2}]$ (see [19] for the properties of the operator Φ).

The aim here is to further analyze the result (25) for $r_0 < r_\text{max}$ in order to derive a more explicit expression for the asymptotic MSE that could be easily compared to the MSE for the FR estimate (27). This new closed form is given in the following proposition.

Proposition 3: The MSE of the RR estimate is asymptotically given by one of the equivalent forms

$$\text{MSE}_\text{RR} = \Phi\left[\mathbf{I}_W, \mathbf{R}_{xx}^{-1}\right]\Phi[\mathbf{P}_\text{S}, \mathbf{Q}] + \Phi\left[\mathbf{P}_\text{T}, \mathbf{R}_{xx}^{-1}\right]\Phi\left[\mathbf{P}_\text{S}^\perp, \mathbf{Q}\right] \tag{28}$$
$$= \Phi\left[\mathbf{P}_\text{T}, \mathbf{R}_{xx}^{-1}\right]\Phi[\mathbf{I}_M, \mathbf{Q}] + \Phi\left[\mathbf{P}_\text{T}^\perp, \mathbf{R}_{xx}^{-1}\right]\Phi[\mathbf{P}_\text{S}, \mathbf{Q}] \tag{29}$$

where $\mathbf{P}_\text{S} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^\dagger$ and $\mathbf{P}_\text{T} = \tilde{\mathbf{B}}\tilde{\mathbf{B}}^\dagger$ denote, respectively, the projectors onto the column space of the whitened channel components $\tilde{\mathbf{A}} = \mathbf{Q}^{-\text{H}/2}\mathbf{A}$ and $\tilde{\mathbf{B}} = \mathbf{R}_{xx}^{1/2}\mathbf{B}$.

Proof: See Appendix C. ■

Let us compare (27) with (28) and (29). For $r_0 = r_\text{max}$, one of the two projectors in (28), (29) reduces to the identity matrix ($\mathbf{P}_\text{S} = \mathbf{I}_M$ if $r_\text{max} = M$, or $\mathbf{P}_\text{T} = \mathbf{I}_W$ if $r_\text{max} = W$) and, as expected, the MSE of the RR estimate equals the expression for the FR estimate (27).

For $r_0 \leq r_\text{max}$, consider the MSE (29), and notice that $\mathcal{R}[\mathbf{P}_\text{S}] \subseteq \mathcal{R}[\mathbf{I}_M]$. It is $\Phi[\mathbf{P}_\text{S}, \mathbf{Q}] \leq \Phi[\mathbf{I}_M, \mathbf{Q}]$, and therefore

$$\text{MSE}_\text{RR} \leq \Phi\left[\mathbf{P}_\text{T}, \mathbf{R}_{xx}^{-1}\right]\Phi[\mathbf{I}_M, \mathbf{Q}] + \Phi\left[\mathbf{P}_\text{T}^\perp, \mathbf{R}_{xx}^{-1}\right]\Phi[\mathbf{I}_M, \mathbf{Q}] = \text{MSE}_\text{u} \tag{30}$$

with equality for $r_0 = r_\text{max}$. This implies that the RR estimate outperforms the FR estimate for any $r_0 \leq r_\text{max}$:

$$\text{MSE}_\text{RR} \leq \text{MSE}_\text{u}. \tag{31}$$

This result can be explained by observing that in the RR approach (see Proposition 1), the FR estimate $\tilde{\mathbf{H}}_\text{u}$ is projected onto the spatial (by $\mathbf{P}_\text{S}$) or the temporal (by $\mathbf{P}_\text{T}$) subspace. The noise contained in $\tilde{\mathbf{H}}_\text{u}$ is reduced by the projection, as it can be seen from the bound (28) and (29). The MSE of the RR estimate is thus given by the residual noise that can no longer be eliminated as it belongs to the same subspace of the signal.

The relationship between the FR and the RR estimates can be easily derived for spatially white noise ($\mathbf{Q} = \sigma^2\mathbf{I}_M$) and uncorrelated training sequence ($\mathbf{R}_{xx} = N\sigma_x^2\mathbf{I}_W$). In this case, the MSE bounds (27) and (28) have the simple closed forms

$$\text{MSE}_\text{u} = \rho MW \tag{32}$$
$$\text{MSE}_\text{RR} = \rho r_0\left[(M+W) - r_0\right] \tag{33}$$

where $\rho = \sigma^2/N\sigma_x^2$ is the ratio between the noise power and the energy of the training sequence. The MSE (32) and (33) are linearly related to the SNR ρ and to the number of independent parameters that parameterize the $(M \times W)$ channel matrix $\mathbf{H}$. In the unconstrained model, $\mathbf{H}$ is full-rank, and therefore, the number of parameters is MW, while by using the rank-r_0 model, the number of degrees of freedom for the matrix $\mathbf{H}$ is reduced to $r_0(M+W-r_0)$. The matrix $\mathbf{H}$ is indeed univocally identified by its SVD that is composed of $r_0W - r_0(r_0+1)/2$ free parameters for the left eigenvector (recall that

$r_0(r_0+1)/2$ is the number of constrains of orthonormality), $r_0M - r_0(r_0+1)/2$ for the right eigenvectors and r_0 singular values. With respect to the FR estimate, the number of parameters and, therefore, the variance of the estimator, is reduced by a factor $MW/[(M+W)r_0 - r_0^2]$ that validates by simple arguments the inequality (31).

D. Performance Loss for Temporally Correlated Noise

The RR MLE (10a) and (10b) is suboptimal when applied to signals in temporally correlated noise. This is shown in the following by evaluating the performance of the estimate in temporally correlated noise for $N \to \infty$. The optimal estimate and its MSE are given in Appendix D for the simple case of known noise covariance.

The noise term in signal model (1) and (7) is assumed to be a separable vector noise process [29] with covariance $\mathrm{E}[\mathbf{n}(t)\mathbf{n}^{\mathrm{H}}(i)] = \mathbf{Q}_{\mathrm{T}}[i,t]\mathbf{Q}_{\mathrm{S}}$ for $i, t = 1,\ldots,N$. The matrices $\mathbf{Q}_{\mathrm{T}}$ $(N \times N)$ and $\mathbf{Q}_{\mathrm{S}}$ $(M \times M)$ denote, respectively, the temporal and spatial (Hermitian) covariance, with $\mathbf{Q}_{\mathrm{T}}[t,t] = 1$, $t = 1,\ldots,N$, and $\mathbf{Q}_{\mathrm{S}}[m,m] = \sigma^2$, $m = 1,\ldots,M$. The overall $MN \times 1$ vector $\mathbf{n} = \mathrm{vec}[\mathbf{N}]$ has covariance $\mathrm{E}[\mathbf{n}\mathbf{n}^{\mathrm{H}}] = \mathbf{Q}_{\mathrm{T}}^* \otimes \mathbf{Q}_{\mathrm{S}}$. This type of impairment occurs when the antenna array operates in a spectral homogeneous environment where all the sensors are exposed to noise having the same temporal correlation. Notice that for stationary noise, it is $\mathbf{Q}_{\mathrm{T}}[i,t] = q_{\mathrm{T}}(t-i)$ with $q_{\mathrm{T}}(t)$ denoting the scalar temporal autocorrelation.

From (18), the covariance for the unconstrained channel estimate $\mathbf{h}_{\mathrm{u}} = \mathrm{vec}[\mathbf{H}_{\mathrm{u}}]$ can be written as $\mathbf{R}_{\mathrm{u}} = \mathbf{R}_{\mathrm{u,T}}^* \otimes \mathbf{R}_{\mathrm{u,S}}$, where the spatial and temporal covariance are, respectively

$$\mathbf{R}_{\mathrm{u,T}} = \left[\mathbf{R}_{xx}^{-1}(\mathbf{X}\mathbf{Q}_{\mathrm{T}}\mathbf{X}^{\mathrm{H}})\mathbf{R}_{xx}^{-1}\right] \tag{34a}$$

$$\mathbf{R}_{\mathrm{u,S}} = \mathbf{Q}_{\mathrm{S}}. \tag{34b}$$

The corresponding $\mathrm{MSE}_{\mathrm{u}} = \mathrm{tr}[\mathbf{R}_{\mathrm{u,T}}]\mathrm{tr}[\mathbf{R}_{\mathrm{u,S}}]$ is given by

$$\mathrm{MSE}_{\mathrm{u}} = \Phi[\mathbf{I}_W, \mathbf{R}_{\mathrm{u,T}}]\Phi[\mathbf{I}_M, \mathbf{R}_{\mathrm{u,S}}] \tag{35}$$

that is equal to (27) for $\mathbf{Q}_{\mathrm{T}} = \mathbf{I}_N$, and it is larger for any $\mathbf{Q}_{\mathrm{T}} \neq \mathbf{I}_N$.

In the following, the MSE for the RR estimate is calculated from (34a) and (34b) by exploiting the result given in the Proposition below.

Proposition 4: Let $\hat{\mathbf{H}}$ be the minimizer, under the constraint $\mathrm{rank}(\mathbf{H}) = r_0$, of the weighted norm

$$\mathcal{F} = \left\|\mathbf{W}_{\mathrm{S}}(\mathbf{H}_{\mathrm{u}} - \mathbf{H})\mathbf{W}_{\mathrm{T}}^{\mathrm{H}}\right\|^2 \tag{36}$$

calculated from the unbiased unconstrained estimate $\mathbf{H}_{\mathrm{u}}$, having covariance $\mathbf{R}_{\mathrm{u}} = \mathbf{R}_{\mathrm{u,T}}^* \otimes \mathbf{R}_{\mathrm{u,S}}$ and for two given full-rank weighting matrices $\mathbf{W}_{\mathrm{S}}(M \times M)$ and $\mathbf{W}_{\mathrm{T}}(W \times W)$. The MSE for the estimate $\hat{\mathbf{H}}$ is

$$\begin{aligned}\mathrm{MSE}_{\mathrm{RR}} = &\,\mathrm{tr}\left[\mathbf{W}_{\mathrm{T}}^{-1}(\mathbf{P}_{\mathrm{T}}\tilde{\mathbf{R}}_{\mathrm{u,T}})\mathbf{W}_{\mathrm{T}}^{-\mathrm{H}}\right]\mathrm{tr}\left[\mathbf{W}_{\mathrm{S}}^{-1}\left(\mathbf{P}_{\mathrm{S}}^{\perp}\tilde{\mathbf{R}}_{\mathrm{u,S}}\right)\mathbf{W}_{\mathrm{S}}^{-\mathrm{H}}\right]\\ &+\mathrm{tr}[\mathbf{R}_{\mathrm{u,T}}]\mathrm{tr}\left[\mathbf{W}_{\mathrm{S}}^{-1}(\mathbf{P}_{\mathrm{S}}\tilde{\mathbf{R}}_{\mathrm{u,S}})\mathbf{W}_{\mathrm{S}}^{-\mathrm{H}}\right]\end{aligned} \tag{37}$$

where $\tilde{\mathbf{R}}_{\mathrm{u,S}} = \mathbf{W}_{\mathrm{S}}\mathbf{R}_{\mathrm{u,S}}\mathbf{W}_{\mathrm{S}}^{\mathrm{H}}$, and $\tilde{\mathbf{R}}_{\mathrm{u,T}} = \mathbf{W}_{\mathrm{T}}\mathbf{R}_{\mathrm{u,T}}\mathbf{W}_{\mathrm{T}}^{\mathrm{H}}$ are the spatial and temporal components of the covariance $\tilde{\mathbf{R}}_{\mathrm{u}} = \tilde{\mathbf{R}}_{\mathrm{u,T}}^* \otimes \tilde{\mathbf{R}}_{\mathrm{u,S}}$ for the weighted unconstrained estimate $\tilde{\mathbf{h}}_{\mathrm{u}} = \mathrm{vec}[\mathbf{W}_{\mathrm{S}}\mathbf{H}_{\mathrm{u}}\mathbf{W}_{\mathrm{T}}^{\mathrm{H}}]$, while $\mathbf{P}_{\mathrm{S}}$ and $\mathbf{P}_{\mathrm{T}}$ denote the projectors onto the column space of matrices $\mathbf{W}_{\mathrm{S}}\mathbf{H}$ and $\mathbf{W}_{\mathrm{T}}\mathbf{H}^{\mathrm{H}}$, respectively.

Proof: See Appendix E. ■

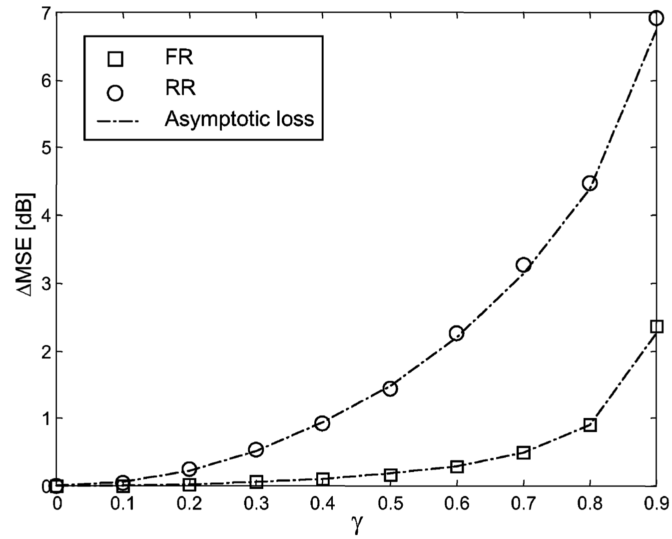

Fig. 3. Analytical (lines) and simulated (markers) performance loss for the FR and RR estimates due to noise temporal correlation.

Notice that (37) is minimized when $\tilde{\mathbf{R}}_{\mathrm{u,T}} = \mathbf{I}_W$ and $\tilde{\mathbf{R}}_{\mathrm{u,S}} = \mathbf{I}_M$, yielding the optimal performance:

$$\begin{aligned}\mathrm{MSE}_{\mathrm{RR}} = &\,\Phi[\mathbf{P}_{\mathrm{T}}, \mathbf{R}_{\mathrm{u,T}}]\Phi\left[\mathbf{P}_{\mathrm{S}}^{\perp}, \mathbf{R}_{\mathrm{u,S}}\right]\\ &+\Phi[\mathbf{I}_W, \mathbf{R}_{\mathrm{u,T}}]\Phi[\mathbf{P}_{\mathrm{S}}, \mathbf{R}_{\mathrm{u,S}}].\end{aligned} \tag{38}$$

This minimum is reached when the weights are chosen to whiten the error of the unconstrained estimate: $\mathbf{W}_{\mathrm{T}} = \mathbf{R}_{\mathrm{u,T}}^{-\mathrm{H}/2}$ and $\mathbf{W}_{\mathrm{S}} = \mathbf{R}_{\mathrm{u,S}}^{-\mathrm{H}/2}$ (see Appendix E).

As pointed out in Remark 2, the RR estimate (10a) and (10b) for $N \to \infty$ can be equivalently seen as the minimizer of (36) for $\mathbf{W}_{\mathrm{S}} = \mathbf{Q}_{\mathrm{S}}^{-\mathrm{H}/2}$ and $\mathbf{W}_{\mathrm{T}} = \mathbf{R}_{xx}^{1/2}$. For temporally correlated noise, this choice of weights is suboptimal as the covariance of the unconstrained estimate depends on

$$\tilde{\mathbf{R}}_{\mathrm{u,T}} = \mathbf{R}_{xx}^{-\frac{\mathrm{H}}{2}}(\mathbf{X}\mathbf{Q}_{\mathrm{T}}\mathbf{X}^{\mathrm{H}})\mathbf{R}_{xx}^{-\frac{1}{2}} \tag{39a}$$

$$\tilde{\mathbf{R}}_{\mathrm{u,S}} = \mathbf{I}_M \tag{39b}$$

where $\tilde{\mathbf{R}}_{\mathrm{u,T}}$ is nondiagonal, as the temporal whitening does not take into account the noise correlation $\mathbf{Q}_{\mathrm{T}}$. From (37), (39a), and (39b), the MSE of the estimate (for $N \to \infty$) is

$$\begin{aligned}\mathrm{MSE}_{\mathrm{RR}} = &\,\Phi\left[\mathbf{P}_{\mathrm{T}}\tilde{\mathbf{R}}_{\mathrm{u,T}}, \mathbf{R}_{xx}^{-1}\right]\Phi\left[\mathbf{P}_{\mathrm{S}}^{\perp}, \mathbf{Q}_{\mathrm{S}}\right]\\ &+\Phi[\mathbf{I}_W, \mathbf{R}_{\mathrm{u,T}}]\Phi[\mathbf{P}_{\mathrm{S}}, \mathbf{Q}_{\mathrm{S}}]\end{aligned} \tag{40}$$

that is larger than (28) for any $\mathbf{Q}_{\mathrm{T}} \neq \mathbf{I}_N$, and it is equal for $\mathbf{Q}_{\mathrm{T}} = \mathbf{I}_N$.

In Fig. 3, we evaluate the performance loss ΔMSE with respect to the minimum MSE (given in Appendix D or, equivalently, in (38)) for the FR estimate (9a) and the RR estimate (10a) and (10b) in temporally correlated noise. The channel matrix $\mathbf{H}$ is randomly generated with $M = 12$, $W = 20$, and $r_0 = 4$, where an m-sequence of length $N = 63$ is used as training sequence. The complex Gaussian noise is modeled as a stationary AR-1 process, with temporal covariance $[\mathbf{Q}_{\mathrm{T}}]_{m,\ell} = \gamma^{|\ell-m|}$ and spatial covariance $[\mathbf{Q}_{\mathrm{S}}]_{m,\ell} = \sigma^2 0.9^{|\ell-m|}\exp[-i\pi(\ell -$

$m)\sqrt{3}/2]$. The SNR is SNR $= 20$ dB (defined as in Example 1). Fig. 3 shows the performance loss for varying γ. The degradation, which is larger for the RR method, becomes relevant only for high correlation $\gamma \geq 0.5$.

IV. RR Estimation From Multiple Training Blocks

In time-slotted communication systems, the performance of channel estimation can be improved either by reducing the number of parameters that describe the channel matrix or by extending the training set. So far, we have investigated only the first solution (i.e., the RR parameterization). In this section, we propose to further improve the performance of the RR estimate by exploiting the training data from L successive data blocks to virtually extend the training set.

A. RR Model for Multiple Training Blocks

The discrete-time model (7) is adapted here to describe the signals over L disjointed training periods. The signal received within the ℓth training block, for any $\ell = 1, \ldots, L$, is expressed as

$$\mathbf{Y}(\ell) = \mathbf{H}(\ell)\mathbf{X} + \mathbf{N}(\ell) \tag{41}$$

where the training sequence $\mathbf{X}$ is assumed to be the same for all blocks, and the Gaussian noise $\mathbf{N}(\ell)$ is stationary with block-independent covariance matrix $\mathbf{Q}$ (the interference is considered to be stationary). Furthermore, the propagation channel $\mathbf{H}(\ell)$ is assumed to be constant within the ℓth training interval and to be generated by the same multipath structure (2), i.e.,

$$\mathbf{H}(\ell) = \mathbf{F}(\boldsymbol{\vartheta})\boldsymbol{\Lambda}(\boldsymbol{\alpha}(\ell))\,\mathbf{G}^{\mathrm{T}}(\boldsymbol{\tau}) \tag{42}$$

where $\boldsymbol{\alpha}(\ell) = [\alpha_1(\ell) \ \ldots \ \alpha_P(\ell)]$. The fading amplitudes $\boldsymbol{\Lambda}(\boldsymbol{\alpha}(\ell)) = \mathrm{diag}[\alpha_1(\ell) \ \ldots \ \alpha_P(\ell)]$ change rapidly due to the user movement, whereas the angle/delay pattern $\{\boldsymbol{\vartheta}, \boldsymbol{\tau}\}$ is considered to be stationary over the L blocks. This assumption is reasonable if the number of training periods L is chosen according to the mobile speed and multipath geometry [19]. Furthermore, the complex random variables $\boldsymbol{\alpha}(\ell)$ are assumed to be block-fading and uncorrelated from block to block (for the effect of correlated fading, see [19]). From (42), it follows that the rank order of the channel matrix $\mathbf{H}(\ell)$ is block-independent, and it is $\mathrm{rank}[\mathbf{H}(\ell)] = r_0$.

According to the signal model (41), the negative log-likelihood function is now given by

$$\mathcal{L}(\mathbf{H}(1), \ldots, \mathbf{H}(L), \mathbf{Q}) = \log|\mathbf{Q}| + \frac{1}{NL}\sum_{\ell=1}^{L} \|\mathbf{Y}(\ell) - \mathbf{H}(\ell)\mathbf{X}\|^2_{\mathbf{Q}^{-1}}. \tag{43}$$

The aim here is to derive the MLE of the ensemble $\{\mathbf{H}(\ell)\}_{\ell=1}^{L}$ and $\mathbf{Q}$ under the low-rank constraint $r_0 < r_{\max}$. For $r_0 = r_{\max}$, the (full-rank) solution is

$$\mathbf{H}_{\mathrm{u}}(\ell) = \mathbf{R}_{yx}(\ell)\mathbf{R}_{xx}^{-1} \tag{44a}$$

$$\mathbf{Q}_{\mathrm{u}} = \frac{1}{NL}\sum_{\ell=1}^{L}\left\{\mathbf{R}_{yy}(\ell) - \mathbf{R}_{yx}(\ell)\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}(\ell)\right\} \tag{44b}$$

where $\mathbf{R}_{yx}(\ell) = \mathbf{R}_{xy}^{\mathrm{H}}(\ell) = \mathbf{Y}(\ell)\mathbf{X}^{\mathrm{H}}$, $\mathbf{R}_{yy}(\ell) = \mathbf{Y}(\ell)\mathbf{Y}^{\mathrm{H}}(\ell)$, and $\mathbf{R}_{xx} = \mathbf{X}\mathbf{X}^{\mathrm{H}}$.

In the following, the multiblock constrained estimation is reformulated as an equivalent single-block RR problem distinguishing two different cases: spatial reduced-rank (S-RR) for $r_0 = r_{\mathrm{S}}$ and temporal reduced-rank (T-RR) for $r_0 = r_{\mathrm{T}}$. Both solutions are obtained by rewriting the multiblock signal model (41) in the form

$$\mathcal{Y} = \mathcal{H}\mathcal{X} + \mathcal{N} \tag{45}$$

where $\mathcal{Y}$, $\mathcal{H}$, and $\mathcal{N}$ collect the received signals $\{\mathbf{Y}(\ell)\}_{\ell=1}^{L}$, the overall channels $\{\mathbf{H}(\ell)\}_{\ell=1}^{L}$, and the noise $\{\mathbf{N}(\ell)\}_{\ell=1}^{L}$ for the whole set of L blocks, while $\mathcal{X}$ depends on the training symbols $\mathbf{X}$. The way each of these compound matrices is arranged depends on the specific RR method (S-RR or T-RR), as will be explained below. The noise is spatially correlated with covariance matrix $\mathcal{Q}$. Similar to the estimation problem discussed in Section II, the channel matrix is defined in such a way that $\mathrm{rank}[\mathcal{H}] = r_0$ and it is expressed as

$$\mathcal{H} = \mathcal{A}\mathcal{B}^{\mathrm{H}} \tag{46}$$

where $\mathcal{A}$ and $\mathcal{B}$ are full column rank matrices of rank r_0. The estimation of $\mathcal{H}$ and $\mathcal{Q}$ from (45) under the implicit constraint (46) is obtained by applying the RR method derived in Section III. Once again, we will show that the solution depends on the set of L FR estimates (44a) after a whitening operation: $\tilde{\mathbf{H}}_{\mathrm{u}}(\ell) = \mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}\mathbf{H}_{\mathrm{u}}(\ell)\mathbf{R}_{xx}^{\mathrm{H}/2}$, for $\ell = 1, \ldots, L$.

B. RR in the Space Domain (S-RR)

1) Signal Model: Let us assume that the rank order is determined by the spatial manifold: $r_0 = r_{\mathrm{S}} \leq r_{\mathrm{T}}$. The multipath channel model (42) can be expressed as

$$\mathbf{H}(\ell) = \mathbf{A}\mathbf{B}^{\mathrm{H}}(\ell), \quad \ell = 1, \ldots, L. \tag{47}$$

The $M \times r_0$ spatial component $\mathbf{A}$ is block-independent, whereas the $W \times r_0$ temporal component $\mathbf{B}(\ell)$ changes from block to block. This model is suitable for rich multipath environments where the delay diversity is large (e.g., $r_{\mathrm{T}} \simeq W$), but the angular spread is low compared to the receiver resolution [4]. An example is shown on the left side of Fig. 1, where the spatial/temporal diversity is given by $r_{\mathrm{S}} = 1$ and $r_{\mathrm{T}} \simeq 3$. The amplitudes $[\alpha_1(\ell)\alpha_2(\ell)\alpha_3(\ell)]$ are assumed to be block-dependent, and therefore, it is $\mathbf{H}(\ell) = \mathbf{a}\mathbf{b}^{\mathrm{H}}(\ell)$.

The L channels are arranged into the rank-r_0 matrix $\mathcal{H}$ in the following way:

$$\begin{aligned} &\mathcal{H} = [\mathbf{H}(1) \cdots \mathbf{H}(L)] && (M \times LW) \\ &\mathcal{A} = \mathbf{A} && (M \times r_0) \\ &\mathcal{B} = \left[\mathbf{B}^{\mathrm{T}}(1) \cdots \mathbf{B}^{\mathrm{T}}(L)\right]^{\mathrm{T}} && (LW \times r_0) \end{aligned} \tag{48}$$

where the components $\mathcal{A}$ and $\mathcal{B}$ are full column rank (as it is $WL \geq r_0$). The multiblock RR model is given by (45), with the following augmented matrices:

$$\begin{aligned} &\mathcal{Y} = [\mathbf{Y}(1) \cdots \mathbf{Y}(L)] && (M \times LN) \\ &\mathcal{X} = \mathbf{I}_L \otimes \mathbf{X} && (LW \times LN) \\ &\mathcal{N} = [\mathbf{N}(1) \cdots \mathbf{N}(L)] && (M \times LN). \end{aligned} \tag{49}$$

The training sequence has correlation $\mathcal{R}_{xx} = \mathcal{X}\mathcal{X}^{\mathrm{H}} = \mathbf{I}_L \otimes \mathbf{R}_{xx}$, whereas the noise covariance is $\mathrm{E}[\mathcal{N}\mathcal{N}^{\mathrm{H}}]/LN = \mathbf{Q}$. It is important to notice that in (48), the temporal length is virtually increased by a factor L, i.e., $W_{\mathrm{mb}} = LW$, as the temporal component varies from block to block. On the other hand, the spatial component is constant, and therefore, the spatial length is equal to M. The overall number of parameters in $\mathcal{H}$ is thus increased by L, and it is estimated by using a training sequence of length $N_{\mathrm{mb}} = LN$.

2) Method: The constrained estimation of $\mathcal{H}$ and $\mathcal{Q}$ can be obtained from the model (45)–(49) by using the RR single-block solution (10a) in the spatial domain:

$$\hat{\mathcal{H}} = \hat{\mathcal{A}}\hat{\mathcal{B}}^{\mathrm{H}} = \mathbf{Q}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}\hat{\mathbf{P}}_{\mathrm{S},r_0}\tilde{\mathcal{H}}_{\mathrm{u}}\mathcal{R}_{xx}^{-\frac{\mathrm{H}}{2}} \tag{50}$$

where $\tilde{\mathcal{H}}_{\mathrm{u}} = [\tilde{\mathbf{H}}_{\mathrm{u}}(1)\cdots\tilde{\mathbf{H}}_{\mathrm{u}}(L)]$ is the whitened multiblock estimate. According to the block partition of $\mathcal{H}$, the estimate for the ℓth block channel estimate is

$$\hat{\mathbf{H}}(\ell) = \hat{\mathbf{A}}\hat{\mathbf{B}}^{\mathrm{H}}(\ell) = \mathbf{Q}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}\hat{\mathbf{P}}_{\mathrm{S},r_0}\tilde{\mathbf{H}}_{\mathrm{u}}(\ell)\mathbf{R}_{xx}^{-\frac{\mathrm{H}}{2}}. \tag{51}$$

$\mathbf{Q}_{\mathrm{u}}$ is defined in (44b), and $\hat{\mathbf{P}}_{\mathrm{S},r_0} = \hat{\mathbf{U}}_{\mathrm{S},r_0}\hat{\mathbf{U}}_{\mathrm{S},r_0}^{\mathrm{H}}$ is the projector onto the subspace spanned by the r_0 leading eigenvectors $(\hat{\mathbf{U}}_{\mathrm{S},r_0})$ of the spatial correlation matrix

$$\hat{\mathbf{R}}_{\mathrm{S}}(L) = \frac{1}{L}\tilde{\mathcal{H}}_{\mathrm{u}}\tilde{\mathcal{H}}_{\mathrm{u}}^{\mathrm{H}} = \frac{1}{L}\sum_{\ell=1}^{L}\tilde{\mathbf{H}}_{\mathrm{u}}(\ell)\tilde{\mathbf{H}}_{\mathrm{u}}^{\mathrm{H}}(\ell). \tag{52}$$

Since the spatial basis is estimated by averaging the signals over L blocks, the accuracy is expected to increase with L when the fading is uncorrelated. The estimate for the covariance matrix of the noise is

$$\hat{\mathbf{Q}}(L) = \frac{1}{NL}\sum_{\ell=1}^{L}\hat{\mathbf{N}}(\ell)\hat{\mathbf{N}}^{\mathrm{H}}(\ell) \tag{53}$$

which is obtained from the residuals of the channel estimation: $\hat{\mathbf{N}}(\ell) = \mathbf{Y}(\ell) - \hat{\mathbf{H}}(\ell)\mathbf{X}$.

C. RR in the Time Domain (T-RR)

1) Signal Model: This variant assumes that the rank order depends on the temporal manifold: $r_0 = r_{\mathrm{T}} \le r_{\mathrm{S}}$. The multipath channel model (42) is rewritten as

$$\mathbf{H}(\ell) = \mathbf{A}(\ell)\mathbf{B}^{\mathrm{H}}, \quad \ell = 1,\ldots,L \tag{54}$$

where the $M \times r_0$ spatial component $\mathbf{A}(\ell)$ is varying, and the $W \times r_0$ temporal component $\mathbf{B}$ is constant. This model is suitable to describe environments with large angular spread but low delay spread [4]. An example of this scenario is given on the right side of Fig. 1, where the multipath is characterized by $r_{\mathrm{S}} = 3$, $r_{\mathrm{T}} \simeq 1$. The amplitudes $[\alpha_1(\ell)\alpha_2(\ell)\alpha_3(\ell)]$ change from block to block, and thus, it is $\mathbf{H}(\ell) = \mathbf{a}(\ell)\mathbf{b}^{\mathrm{H}}$.

The rank-r_0 matrix $\mathcal{H}$ collects the L channels in the following way:

$$\begin{aligned} \mathcal{H} &= \left[\mathbf{H}^{\mathrm{T}}(1)\cdots\mathbf{H}^{\mathrm{T}}(L)\right]^{\mathrm{T}} && (LM \times W) \\ \mathcal{A} &= \left[\mathbf{A}^{\mathrm{T}}(1)\cdots\mathbf{A}^{\mathrm{T}}(L)\right]^{\mathrm{T}} && (LM \times r_0) \\ \mathcal{B} &= \mathbf{B} && (W \times r_0). \end{aligned} \tag{55}$$

$\mathcal{A}$ and $\mathcal{B}$ have full column rank (it is $ML \ge r_0$). The equivalent RR model (45) is here obtained by defining

$$\begin{aligned} \mathcal{Y} &= \left[\mathbf{Y}^{\mathrm{T}}(1)\cdots\mathbf{Y}^{\mathrm{T}}(L)\right]^{\mathrm{T}} && (LM \times N) \\ \mathcal{X} &= \mathbf{X} && (W \times N) \\ \mathcal{N} &= \left[\mathbf{N}^{\mathrm{T}}(1)\cdots\mathbf{N}^{\mathrm{T}}(L)\right]^{\mathrm{T}} && (LM \times N) \end{aligned} \tag{56}$$

where the training sequence has correlation $\mathcal{R}_{xx} = \mathcal{X}\mathcal{X}^{\mathrm{H}} = \mathbf{R}_{xx}$, while the noise has spatial covariance $\mathcal{Q} = \mathbf{I}_L \otimes \mathbf{Q}$. It may be noticed that the spatial length in (55) is virtually increased by a factor L, i.e., $M_{\mathrm{mb}} = LM$.

2) Method: From the RR single-block solution (10b) derived in the time domain, it follows that

$$\hat{\mathcal{H}} = \hat{\mathcal{A}}\hat{\mathcal{B}}^{\mathrm{H}} = \mathcal{Q}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}\tilde{\mathcal{H}}_{\mathrm{u}}\hat{\mathbf{P}}_{\mathrm{T},r_0}\mathbf{R}_{xx}^{-\frac{\mathrm{H}}{2}} \tag{57}$$

where $\tilde{\mathcal{H}}_{\mathrm{u}} = [\tilde{\mathbf{H}}_{\mathrm{u}}^{\mathrm{T}}(1)\cdots\tilde{\mathbf{H}}_{\mathrm{u}}^{\mathrm{T}}(L)]^{\mathrm{T}}$ denotes the whitened FR channel estimate for the multiblock model and $\mathcal{Q}_{\mathrm{u}} = \mathbf{I}_L \otimes \mathbf{Q}_{\mathrm{u}}$. The RR estimate for the ℓth block can be expressed as

$$\hat{\mathbf{H}}(\ell) = \hat{\mathbf{A}}(\ell)\hat{\mathbf{B}}^{\mathrm{H}} = \mathbf{Q}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}\tilde{\mathbf{H}}_{\mathrm{u}}(\ell)\hat{\mathbf{P}}_{\mathrm{T},r_0}\mathbf{R}_{xx}^{-\frac{\mathrm{H}}{2}}. \tag{58}$$

$\hat{\mathbf{P}}_{\mathrm{T},r_0} = \hat{\mathbf{U}}_{\mathrm{T},r_0}\hat{\mathbf{U}}_{\mathrm{T},r_0}^{\mathrm{H}}$ is the projector onto the subspace spanned by the r_0 leading eigenvectors $(\hat{\mathbf{U}}_{\mathrm{T},r_0})$ of the temporal correlation matrix:

$$\hat{\mathbf{R}}_{\mathrm{T}}(L) = \frac{1}{L}\tilde{\mathcal{H}}_{\mathrm{u}}^{\mathrm{H}}\tilde{\mathcal{H}}_{\mathrm{u}} = \frac{1}{L}\sum_{\ell=1}^{L}\tilde{\mathbf{H}}_{\mathrm{u}}^{\mathrm{H}}(\ell)\tilde{\mathbf{H}}_{\mathrm{u}}(\ell). \tag{59}$$

The noise covariance matrix is estimated as in (53).

D. Performance Analysis

The MSE for the multiblock estimate (normalized by the number of blocks) is defined as

$$\mathrm{MSE} = \frac{1}{L}\mathrm{E}\left[\|\hat{\mathcal{H}} - \mathcal{H}\|^2\right] = \frac{1}{L}\sum_{\ell=1}^{L}\mathrm{E}\left[\left\|\hat{\mathbf{H}}(\ell) - \mathbf{H}(\ell)\right\|^2\right]. \tag{60}$$

This is calculated according to the results (28) and (29) by using the augmented matrices of the multiblock problem.

The asymptotic (for $N \to \infty$) MSE for the S-RR estimate is obtained from (28) by replacing $\Phi[\mathbf{P}_{\mathrm{T}}, \mathbf{R}_{xx}^{-1}]$ with $\mathrm{tr}[\mathcal{B}(\mathcal{B}^{\mathrm{H}}\mathcal{R}_{xx}\mathcal{B})^{-1}\mathcal{B}^{\mathrm{H}}]$ and recalling that $\mathcal{B} = [\mathbf{B}^{\mathrm{T}}(1)\cdots\mathbf{B}^{\mathrm{T}}(L)]^{\mathrm{T}}$ and $\mathcal{R}_{xx} = \mathbf{I}_L \otimes \mathbf{R}_{xx}$:

$$\begin{aligned} \mathrm{MSE}_{\mathrm{S}}(L) &= \frac{1}{L}\left[\Phi\left[\mathbf{P}_{\mathrm{S}}^{\perp}, \mathbf{Q}\right]\mathrm{tr}\left[\mathcal{B}(\mathcal{B}^{\mathrm{H}}\mathcal{R}_{xx}\mathcal{B})^{-1}\mathcal{B}^{\mathrm{H}}\right]\right. \\ &\qquad \left. +\Phi[\mathbf{P}_{\mathrm{S}}, \mathbf{Q}]\Phi\left[\mathbf{I}_{LW}, \mathcal{R}_{xx}^{-1}\right]\right] \\ &= \frac{1}{L}\Phi\left[\mathbf{P}_{\mathrm{S}}^{\perp}, \mathbf{Q}\right] \\ &\qquad \times \sum_{\ell=1}^{L}\mathrm{tr}\left[\mathbf{B}(\ell)\left(\sum_{\ell=1}^{L}\tilde{\mathbf{B}}^{\mathrm{H}}(\ell)\tilde{\mathbf{B}}(\ell)\right)^{-1}\mathbf{B}^{\mathrm{H}}(\ell)\right] \\ &\qquad + \Phi[\mathbf{P}_{\mathrm{S}}, \mathbf{Q}]\Phi\left[\mathbf{I}_W, \mathbf{R}_{xx}^{-1}\right] \end{aligned} \tag{61}$$

where $\tilde{\mathbf{B}}(\ell) = \mathbf{R}_{xx}^{1/2}\mathbf{B}(\ell)$, and $\mathbf{P}_{\mathrm{S}}$ is the projector onto the column space of $\tilde{\mathbf{A}} = \mathbf{Q}^{-\mathrm{H}/2}\mathbf{A}$.

For the multiblock estimate in the time domain (T-RR), the asymptotic MSE is obtained by using the result (29). By replacing $\Phi[\mathbf{P}_{\mathrm{S}}, \mathbf{Q}]$ with $\mathrm{tr}[\mathcal{A}(\mathcal{A}^{\mathrm{H}}\mathcal{Q}^{-1}\mathcal{A})^{-1}\mathcal{A}^{\mathrm{H}}]$ and recalling that the multiblock variables are $\mathcal{A} = [\mathbf{A}^{\mathrm{T}}(1)\cdots\mathbf{A}^{\mathrm{T}}(L)]^{\mathrm{T}}$ and $\mathcal{Q} = \mathbf{I}_L \otimes \mathbf{Q}$, it follows that

$$\begin{aligned}\mathrm{MSE}_{\mathrm{T}}(L) &= \frac{1}{L}\left[\Phi\left[\mathbf{P}_{\mathrm{T}}^{\perp}, \mathbf{R}_{xx}^{-1}\right] \mathrm{tr}\left[\mathcal{A}(\mathcal{A}^{\mathrm{H}}\mathcal{Q}^{-1}\mathcal{A})^{-1}\mathcal{A}^{\mathrm{H}}\right]\right.\\ &\qquad \left. +\Phi\left[\mathbf{P}_{\mathrm{T}}, \mathbf{R}_{xx}^{-1}\right]\Phi[\mathbf{I}_{LM}, \mathcal{Q}]\right]\\ &= \frac{1}{L}\Phi\left[\mathbf{P}_{\mathrm{T}}^{\perp}, \mathbf{R}_{xx}^{-1}\right]\\ &\quad \times \sum_{\ell=1}^{L} \mathrm{tr}\left[\mathbf{A}(\ell)\left(\sum_{\ell=1}^{L}\tilde{\mathbf{A}}^{\mathrm{H}}(\ell)\tilde{\mathbf{A}}(\ell)\right)^{-1}\mathbf{A}^{\mathrm{H}}(\ell)\right]\\ &\quad + \Phi\left[\mathbf{P}_{\mathrm{T}}, \mathbf{R}_{xx}^{-1}\right]\Phi[\mathbf{I}_M, \mathbf{Q}] \qquad (62)\end{aligned}$$

where $\tilde{\mathbf{A}}(\ell) = \mathbf{Q}^{-\mathrm{H}/2}\mathbf{A}(\ell)$, and $\mathbf{P}_{\mathrm{T}}$ is the projector onto the column space of $\tilde{\mathbf{B}} = \mathbf{R}_{xx}^{1/2}\mathbf{B}$.

For single-block ($L = 1$), the MSEs (61) and (62) for the multiblock estimates equal the MSEs (28), (29). On the other hand, for $L \to \infty$, the first term in both (61) and (62) is $O(1/L)$ and the MSE's reduce to

$$\mathrm{MSE}_{\mathrm{S}}(L\to\infty) = \Phi\left[\mathbf{I}_W, \mathbf{R}_{xx}^{-1}\right]\Phi[\mathbf{P}_{\mathrm{S}}, \mathbf{Q}] \qquad (63a)$$

$$\mathrm{MSE}_{\mathrm{T}}(L\to\infty) = \Phi[\mathbf{I}_M, \mathbf{Q}]\Phi\left[\mathbf{P}_{\mathrm{T}}, \mathbf{R}_{xx}^{-1}\right] \qquad (63b)$$

according to the bounds derived in [19]. In the trivial case of spatially white noise ($\mathbf{Q} = \sigma^2\mathbf{I}_M$) and ideal training sequences ($\mathbf{R}_{xx} = N\sigma_x^2\mathbf{I}_W$), it is

$$\begin{aligned}\mathrm{MSE}_{\mathrm{S}}(L) &= \frac{\rho}{L}r_{\mathrm{S}}(M + LW - r_{\mathrm{S}})\\ &= \rho_{\mathrm{mb}}r_{\mathrm{S}}(M + W_{\mathrm{mb}} - r_{\mathrm{S}}) \qquad (64a)\\ \mathrm{MSE}_{\mathrm{T}}(L) &= \frac{\rho}{L}r_{\mathrm{T}}(LM + W - r_{\mathrm{T}})\\ &= \rho_{\mathrm{mb}}r_{\mathrm{T}}(M_{\mathrm{mb}} + W - r_{\mathrm{T}}) \qquad (64b)\end{aligned}$$

where $\rho_{\mathrm{mb}} = \rho/L = \sigma^2/(N_{\mathrm{mb}}\sigma_x^2)$ is the noise to signal ratio for the multiblock. This result equals the MSE of the RR estimate for a system with a training sequence that is L times longer than the single-block one and a virtual channel length of LW samples (for $\mathrm{MSE}_{\mathrm{S}}$) or a virtual number of LM antennas (for $\mathrm{MSE}_{\mathrm{T}}$). By increasing L, the accuracy in estimating the spatial or temporal basis improves and the error reduces asymptotically to

$$\mathrm{MSE}_{\mathrm{S}}(L\to\infty) = \rho r_{\mathrm{S}}W \qquad (65a)$$

$$\mathrm{MSE}_{\mathrm{T}}(L\to\infty) = \rho r_{\mathrm{T}}M. \qquad (65b)$$

This means that for $L \to \infty$, the lower bound of the MSE depends only on the ratio between the number of parameters to be estimated on a block-by-block basis (the $r_{\mathrm{S}}W$ entries of $\mathbf{B}(\ell)$ or the $r_{\mathrm{T}}M$ entries of the space component $\mathbf{A}(\ell)$) and the training sequence length (N).

The performance comparison between the spatial/temporal multiblock methods and the RR/FR single-block methods is straightforward in the case of white noise and uncorrelated training sequences. Indeed, by comparing (65a) and (65b) with (32) and (33), it follows that

$$\mathrm{MSE}_{\mathrm{S}}(L\to\infty) \le \mathrm{MSE}_{\mathrm{RR}} \le \mathrm{MSE}_{\mathrm{u}}, \text{ for } r_0 = r_{\mathrm{S}} \qquad (66a)$$

$$\mathrm{MSE}_{\mathrm{T}}(L\to\infty) \le \mathrm{MSE}_{\mathrm{RR}} \le \mathrm{MSE}_{\mathrm{u}}, \text{ for } r_0 = r_{\mathrm{T}}. \qquad (66b)$$

Notice that the MSE reduction with respect to the FR single-block method depends on the spatial (M/r_{S}) and temporal (W/r_{T}) gains. For spatially correlated noise and/or temporally correlated training sequences, the relations (66a) and (66b) still hold (see also [19]), but the performance gains are modified, depending on the unconstrained estimate covariance projected onto the channel subspace.

V. RR Estimation for Multiuser Channels

The results obtained in Sections III and IV on the MSE of the RR estimate motivate the extension of the method to multiuser systems. In this section, we derive the MLE of a multiuser channel from a single training block ($L = 1$) under the constraint of reduced rank for each user channel. The extension to multiple training block follows the same steps described in Section IV (this topic is not covered here).

A. Signal Model

The $M \times N$ matrix $\mathbf{Y}$ collects the signals received by the antenna array in a K-user system

$$\mathbf{Y} = \sum_{k=1}^{K}\mathbf{H}_k\mathbf{X}_k + \mathbf{N} \qquad (67)$$

where the $M \times W$ channel matrix $\mathbf{H}_k$ and the $W \times N$ training matrix $\mathbf{X}_k$ refer to the kth user, for $k = 1, \ldots, K$. $\mathbf{N}$ accounts for the background noise and any other interfering signal (e.g., intercell interference in cellular systems), and it is temporally uncorrelated and spatially correlated with spatial covariance $\mathbf{Q}$. The channels are constrained to have rank

$$r_k = \mathrm{rank}(\mathbf{H}_k) < r_{\max}, \quad k = 1, \ldots, K. \qquad (68)$$

The estimation of the channel matrices $\{\mathbf{H}_k\}_{k=1}^{K}$ might be obtained in a single-user fashion by decomposing the multiuser problem into K RR single-user problems. This method will be referred as the single-user RR estimation. However, when the training sequences of different users are correlated, this single-user estimate turns out to be biased by multiaccess interference (MAI), as shown in Section VI. In order to properly account for MAI, the K channels have to be estimated jointly from (67), as described below.

B. Method

The joint MLE of the K channels from (67) is obtained by minimizing the negative log-likelihood function

$$\mathcal{L}(\mathbf{H}, \mathbf{Q}) = \log|\mathbf{Q}| + \frac{1}{N}\left\|\mathbf{Y} - \sum_{k=1}^{K}\mathbf{H}_k\mathbf{X}_k\right\|_{\mathbf{Q}^{-1}}^{2} \qquad (69)$$

with respect to $\{\mathbf{H}_k\}_{k=1}^{K}$ and $\mathbf{Q}$, under the constraints (68). For $r_k = r_{\max}$, the solution reduces to the conventional multiuser FR estimate

$$[\mathbf{H}_{\mathrm{u},1}\cdots\mathbf{H}_{\mathrm{u},K}] = \mathbf{R}_{yx}\mathbf{R}_{xx}^{-1} \qquad (70)$$

$$\mathbf{Q}_{\mathrm{u}} = \frac{1}{N}\left(\mathbf{R}_{yy} - \mathbf{R}_{yx}\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}\right) \qquad (71)$$

where it is $\mathbf{R}_{yy} = \mathbf{Y}\mathbf{Y}^{\mathrm{H}}$ and $\mathbf{R}_{yx} = [\mathbf{Y}\mathbf{X}_1^{\mathrm{H}}\cdots\mathbf{Y}\mathbf{X}_K^{\mathrm{H}}]$, and

$$\mathbf{R}_{xx} = \begin{bmatrix}\mathbf{X}_1\mathbf{X}_1^{\mathrm{H}} & \cdots & \mathbf{X}_1\mathbf{X}_K^{\mathrm{H}}\\ \vdots & \ddots & \vdots\\ \mathbf{X}_K\mathbf{X}_1^{\mathrm{H}} & \cdots & \mathbf{X}_K\mathbf{X}_K^{\mathrm{H}}\end{bmatrix} \qquad (72)$$

is a $KW \times KW$ block-partitioned matrix with blocks $[\mathbf{R}_{xx}]_{k,h} = \mathbf{X}_k\mathbf{X}_h^{\mathrm{H}}$ of dimension $W \times W$. Notice that, differently from Section III, $\mathbf{R}_{xx}$ account for the correlation between the training sequences of different users as well.

As pointed out in Remark 2, for large N, the RR MLE can be shown to be equivalent to the minimization of the loss function

$$\mathcal{F} = \left\| \mathbf{Q}_{\mathrm{u}}^{-\frac{\mathrm{H}}{2}}[\mathbf{H}_{\mathrm{u},1}\cdots\mathbf{H}_{\mathrm{u},K}]\mathbf{R}_{xx}^{\frac{\mathrm{H}}{2}} - \mathbf{Q}_{\mathrm{u}}^{-\frac{\mathrm{H}}{2}}[\mathbf{H}_1\cdots\mathbf{H}_K]\mathbf{R}_{xx}^{\frac{\mathrm{H}}{2}}\right\|^2. \tag{73}$$

However, unlike the single-user case, the whitening factor $\mathbf{R}_{xx}^{\mathrm{H}/2}$ in (73) introduces MAI on the channel matrices $\{\mathbf{H}_{\mathrm{u},k}\}_{k=1}^K$ due the correlation between the training sequences. This can be easily seen by partitioning the $KW \times KW$ Cholesky factor $\mathbf{R}_{xx}^{1/2}$ into blocks of dimension $K \times K$, with $[\mathbf{R}_{xx}^{1/2}]_{k,h} = \mathbf{0}$ for $k > h$ and $h, k = 1, \ldots, K$. The whitened FR estimate for the kth user is then given by

$$\tilde{\mathbf{H}}_{\mathrm{u},k} = \mathbf{Q}_{\mathrm{u}}^{-\frac{\mathrm{H}}{2}}\mathbf{H}_{\mathrm{u}}\left[\mathbf{R}_{xx}^{\frac{\mathrm{H}}{2}}\right]_{k,k} + \Delta\mathbf{H}_{\mathrm{MAI},k} \tag{74}$$

and

$$\Delta\mathbf{H}_{\mathrm{MAI},k} = \sum_{h=k+1}^{K}\mathbf{Q}_{\mathrm{u}}^{-\frac{\mathrm{H}}{2}}(\hat{\mathbf{H}}_{\mathrm{u},h} - \mathbf{H}_h)\left[\mathbf{R}_{xx}^{\frac{\mathrm{H}}{2}}\right]_{k,h} \tag{75}$$

is the interference from users $h = k+1, \ldots, K$. It is convenient to separate the likelihood function (73) as $\mathcal{F} = \sum_{k=1}^K \mathcal{F}_k$, where $\mathcal{F}_k = \|\tilde{\mathbf{H}}_{\mathrm{u},k} - \mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}\mathbf{H}_k[\mathbf{R}_{xx}^{\mathrm{H}/2}]_{k,k}\|^2$ represents the loss term for the kth users. The constrained MLE is given by the minimization of the sum of $\{\mathcal{F}_k\}_{k=1}^K$ with respect to $\{\mathbf{H}_k\}_{k=1}^K$ under the K constraints (68). Below, we propose an approximated solution based on a successive cancellation of MAI in (74) similar to the technique used for interference cancellation in multiuser detection (without any strategy for ordering the users). The optimization of $\mathcal{F}$ with respect to the successive users is performed starting from the Kth user down to the first one. At the kth step, the MLE $\hat{\mathbf{H}}_k$ is calculated on the base of $\mathcal{F}_k$ only, as this is the dominant term depending on $\mathbf{H}_k$. The estimates $\{\hat{\mathbf{H}}_h\}_{h=k+1}^K$, which are derived in the previous iterations, are substituted into $\mathcal{F}_k$, and the loss function is minimized with respect to $\mathbf{H}_k$. The result is summarized below.

The MLE for the multiuser channel $\{\hat{\mathbf{H}}_k\}_{k=1}^K$ and the noise covariance $\hat{\mathbf{Q}}$ under the constraints (68) is approximated by calculating the channel estimates iteratively for $k = K, K-1, \ldots, 1$ as

$$\hat{\mathbf{H}}_k = \mathbf{Q}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}\hat{\mathbf{P}}_{\mathrm{S},k}\tilde{\mathbf{H}}_{\mathrm{u},k}\left[\mathbf{R}_{xx}^{-\frac{\mathrm{H}}{2}}\right]_{k,k} \tag{76}$$

$$\hat{\mathbf{Q}} = \frac{1}{N}\left(\mathbf{Y} - \sum_{k=1}^K\hat{\mathbf{H}}_k\mathbf{X}_k\right)\left(\mathbf{Y} - \sum_{k=1}^K\hat{\mathbf{H}}_k\mathbf{X}_k\right)^{\mathrm{H}} \tag{77}$$

where $\tilde{\mathbf{H}}_{\mathrm{u},k}$ is obtained from (74) by setting $\mathbf{H}_h = \hat{\mathbf{H}}_h$ for $h = k+1, \ldots, K$, and $\hat{\mathbf{P}}_{\mathrm{S},k}$ denotes the projector onto the subspace spanned by the r_k leading eigenvectors of

$$\hat{\mathbf{R}}_{\mathrm{S},k} = \tilde{\mathbf{H}}_{\mathrm{u},k}\tilde{\mathbf{H}}_{\mathrm{u},k}^{\mathrm{H}}. \tag{78}$$

It is important to notice that in real mobile systems, the K training sequences are selected to have low correlation between different users. Therefore, the whitened unconstrained estimate (74) for the user k reduces to

$$\tilde{\mathbf{H}}_{\mathrm{u},k} = \mathbf{Q}_{\mathrm{u}}^{-\frac{\mathrm{H}}{2}}\hat{\mathbf{H}}_{\mathrm{u},k}[\mathbf{R}_{xx}]_{k,k}^{\frac{\mathrm{H}}{2}} \tag{79}$$

with temporal whitening $[\mathbf{R}_{xx}]_{k,k}^{\mathrm{H}/2} = (\mathbf{X}_k\mathbf{X}_k)^{\mathrm{H}/2}$ performed separately for each user. The RR estimate (76) is simplified as

$$\hat{\mathbf{H}}_k = \mathbf{Q}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}\hat{\mathbf{P}}_{\mathrm{S},k}\tilde{\mathbf{H}}_{\mathrm{u},k}[\mathbf{R}_{xx}]_{k,k}^{-\frac{\mathrm{H}}{2}} \tag{80}$$

where the projector $\hat{\mathbf{P}}_{\mathrm{S},k}$ is calculated from the correlation matrix $\hat{\mathbf{R}}_{\mathrm{S},k} = \tilde{\mathbf{H}}_{\mathrm{u},k}\tilde{\mathbf{H}}_{\mathrm{u},k}^{\mathrm{H}}$.

The simplified multiuser solution (80) can be seen as a straightforward extension of the RR estimate (10a) to multiuser systems. The correlation (and projection) matrices are calculated in a single-user fashion, while the whitening terms are properly redefined to take into account MAI. The temporal decorrelation $\mathbf{R}_{xx}^{-1}$ in the preliminary FR estimation (70) accounts for the interference between the training sequences of different users, while the spatial whitening factor $\mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}$ depends on the background noise only (and not on the users' signals). Notice that the solution (80) cannot be obtained as an extension of the RR estimate if the whitening is performed by the factor $\mathbf{R}_{yy}^{-\mathrm{H}/2}$, as in (86) (see Appendix B), as $\mathbf{R}_{yy}$ accounts for both MAI (intracell interference) and background noise (intercell interference).

Remark 4: Even if the method is based on a user-by-user temporal whitening, the multiuser RR estimate (80) outperforms the single-user RR estimate (see Section VI). The gain in performance becomes evident for large SNR, as the single-user approach is affected by an error floor due to MAI, while the multiuser RR estimate is unbiased. Indeed, the multiuser RR estimate is obtained as a rank reduction from the multiuser FR estimate $[\mathbf{H}_{\mathrm{u},1}\cdots\mathbf{H}_{\mathrm{u},K}]$, which is calculated jointly for the K users. As it is $\mathbf{H}_{\mathrm{u},k} \to \mathbf{H}_k$, from (80), it follows that $\hat{\mathbf{H}}_k \to \mathbf{H}_k$.

Remark 5: As shown in Section III for the single-user case, the multiuser RR estimate can be equivalently written as either a spatial projection, a temporal projection, or a SVD truncation of the whitened FR estimate. The computational complexity of the multiuser estimate is K times larger than the one required by the single-user estimate. The rank orders r_k for $k = 1, \ldots, K$ need to be evaluated user-by-user from the sample covariance matrices $\{\hat{\mathbf{R}}_{\mathrm{S},k}\}_{k=1}^K$, as shown in Section III.

VI. Numerical Results

A uniform linear array of $M = 8$ omnidirectional antennas with half wavelength spacing is used at the base station. Mobile users transmit a training sequence of $N = 456$ symbols chosen from the UMTS-TDD standard [20], as in Example 1. The modulation scheme is QPSK at rate $1/T = 3.84$ MHz and carrier frequency $f_c = 1.95$ GHz. The pulse shaping is a raised cosine pulse with roll-off 0.22. Both uncorrelated and spatially correlated Gaussian noise is considered. According to standard specifications [20], the channel length is $W \le W_{\max} = 57$, where $W_{\max}$ is the maximum length that can be estimated in the uplink of a UMTS-TDD system with the given training set. Notice that $W_{\max}$ is chosen rather large (14.8 μs) as the considered system is intended to work with a broad range of environments. This is

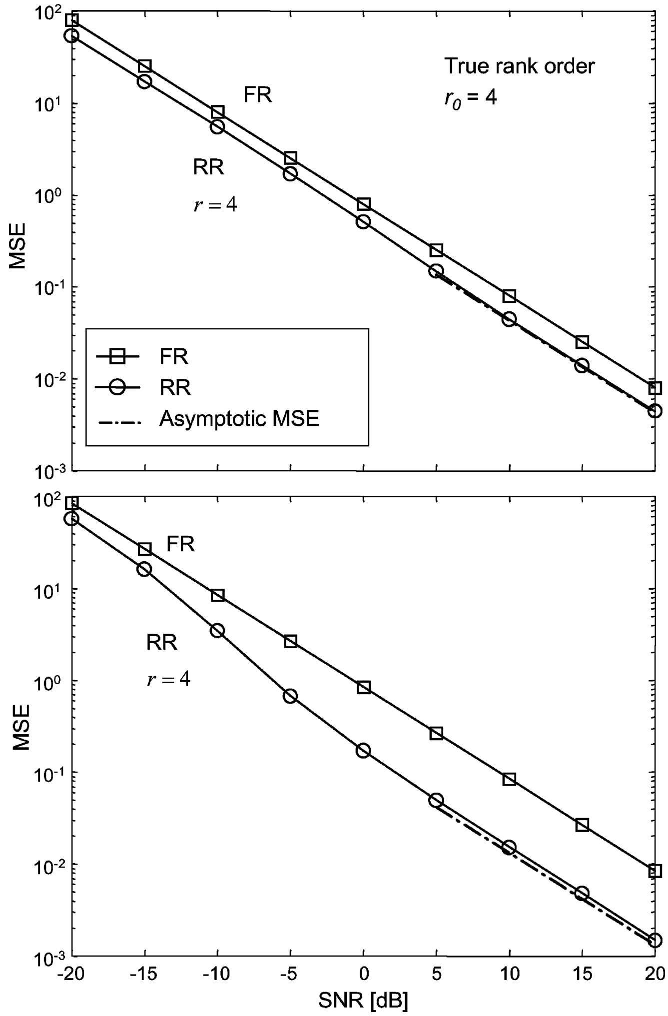

Fig. 4. Comparison of the performance of the FR estimate and the RR estimate with $r = r_0$ for (top) uncorrelated and (bottom) spatially correlated noise.

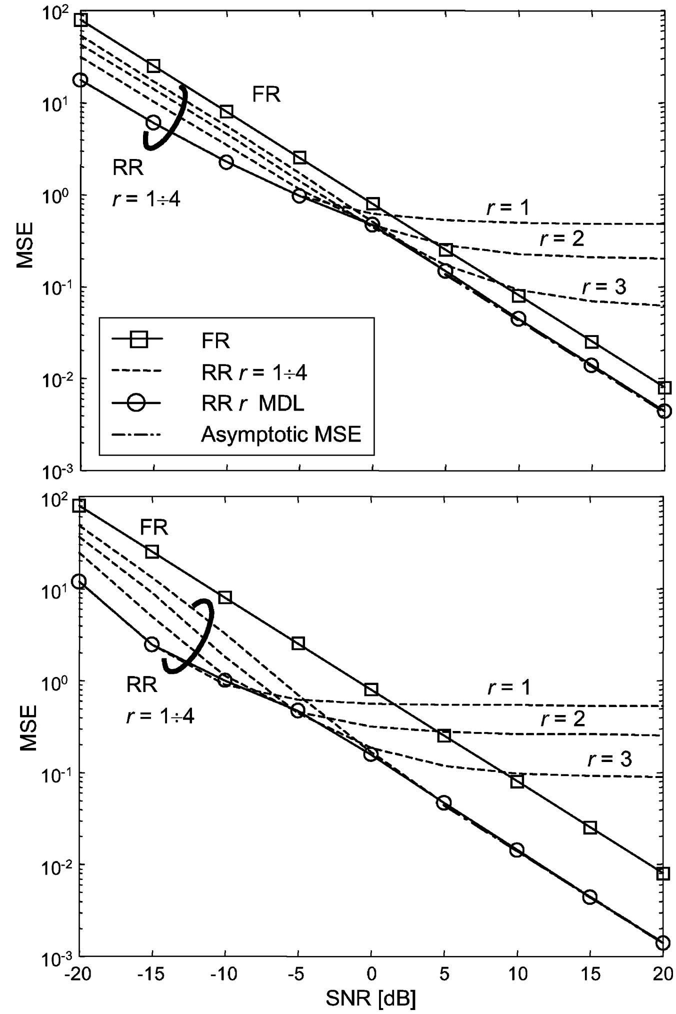

Fig. 5. Comparison of the performance of the FR estimate and the RR estimate for varying r. (Top) Uncorrelated and (bottom) spatially correlated noise.

the reason why in many situations, the full-rank parameterization results are too redundant.

The performances of RR estimation methods are evaluated in terms of normalized MSE $= \mathrm{E}[||\Delta\mathbf{H}||^2]/\mathrm{E}[||\mathbf{H}||^2]$ for different values of SNR $= \mathrm{E}[||\mathbf{H}||^2]/\sigma^2$ (here, it is $\sigma_x^2 = 1$). In order to highlight the MSE threshold effect (for low SNR) and to validate the asymptotic performance (for large SNR), the SNR is made varying from -20 dB up to 20 dB, even though some values might be not realistic. Simulations are performed at first for the single-user single-block scenario; then, the extensions to, respectively, multiblock and multiuser systems are considered; finally, simulation results on realistic multipath propagation channels are given.

A. Single-User Single-Block RR Estimate

Fig. 4 compares the MSE performance of the FR and the RR estimates for an unstructured channel modeled as a 8×45 random matrix with rank order $r_0 = 4$. The simulated MSE is obtained by averaging $||\Delta\mathbf{H}||^2$ over 200 independent runs of channel and noise. The rank order used for RR estimation is $r = r_0$. The additive Gaussian noise is spatially white, $\mathbf{Q} = \sigma^2\mathbf{I}_M$ (top figure), or spatially correlated with covariance $[\mathbf{Q}]_{m,\ell} = \sigma^2 0.9^{|\ell-m|}\exp[-i\pi(\ell-m)\sin\theta]$ and $\theta = \pi/3$ (bottom figure). In addition to the simulated MSE, the figure also shows the asymptotic MSE calculated from (28). It can be seen that the RR estimate outperforms the FR estimate by a factor equal to $r_0(M+W-r_0)/(MW) = 2.6$ dB for white noise, while it can be quantified from (28) in terms of approximately 7 dB for correlated noise.

Fig. 5 shows the performance of the RR estimate for the same example of Fig. 4 when the rank order r used to perform the reduction ranges from $r = 1$ to $r = 4$. As expected, for low SNR, the preferred choice is $r = 1$, as it is better to estimate the minimum number of channel dimensions. By underestimating the rank order, the variance of the channel estimate is indeed reduced by r/r_0, while the bias error introduced by the mismodeling remains negligible when compared with the noise error. For increasing SNR, the distortion becomes remarkable, and the noise error decreases; hence, the optimum rank order (that provides the minimum MSE) moves toward the true rank order ($r_0 = 4$). The minimum MSE for each SNR value can thus be obtained only by an adaptive selection of the rank order. This is confirmed by the numerical results in Fig. 5, where the MSE of the RR estimate with MDL rank-order selection (solid line with circle markers) envelopes the lower values of the fixed rank MSE, and it approaches the MSE bound for large SNR.

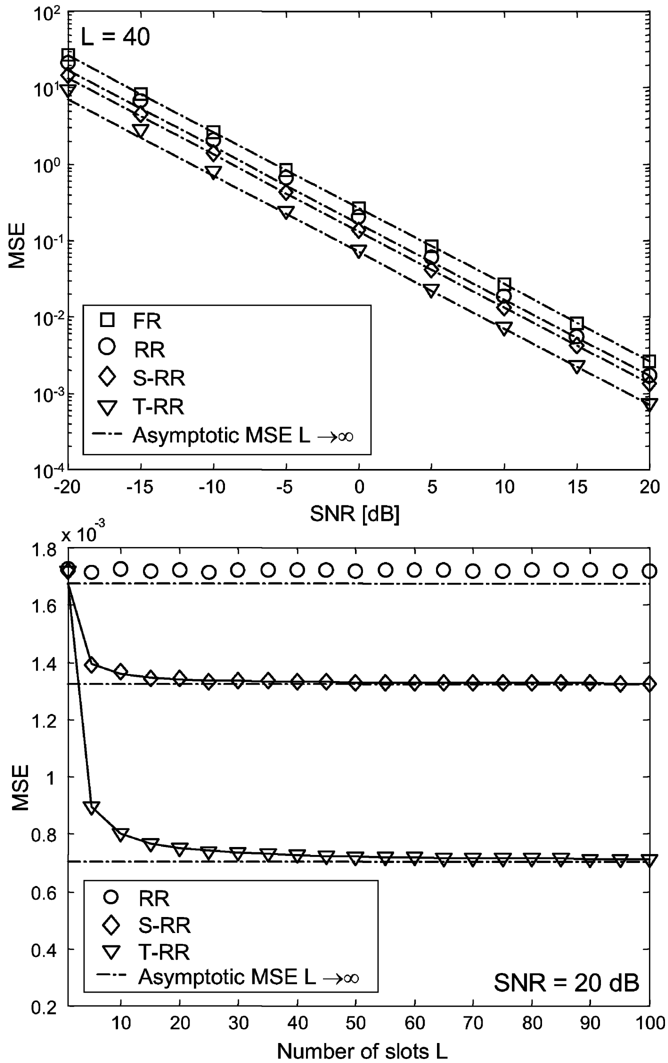

Fig. 6. MSE of the multiblock RR estimate in uncorrelated noise. (Top) For varying SNR and $L = 30$. (Bottom) For varying L and SNR $= 20$ dB.

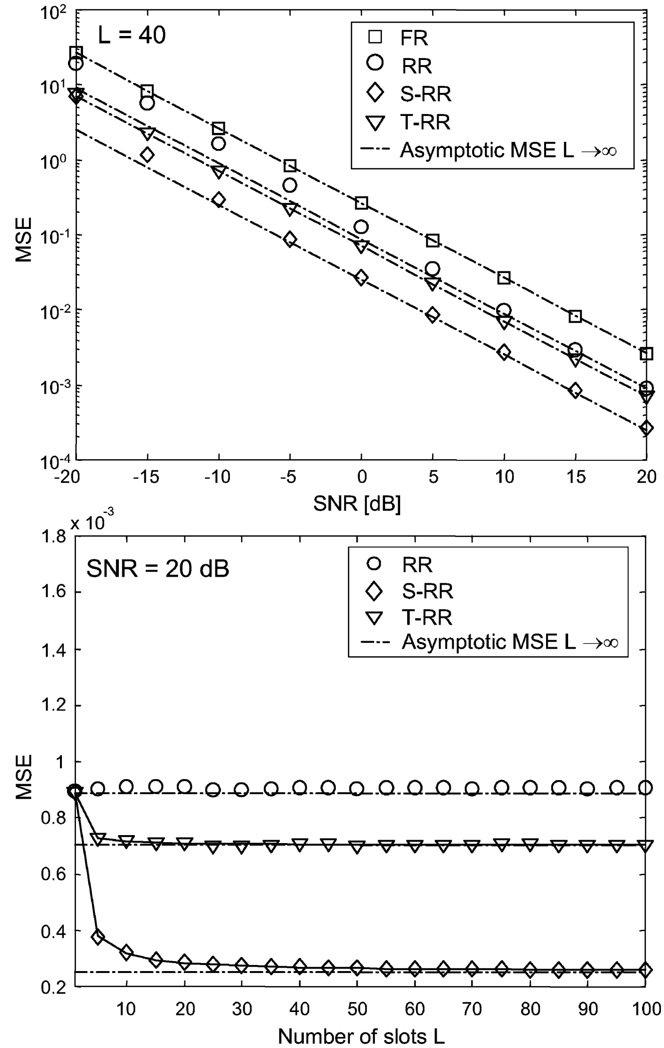

Fig. 7. MSE of the multiblock RR estimate in spatially correlated noise. (Top) For varying SNR and $L = 40$. (Bottom) For varying L and SNR $= 20$ dB.

B. Multiblock RR Estimate

Figs. 6 and 7 compare the simulations for the MSE of the multiblock estimates (markers) with the MSE bounds calculated in Section IV-D (lines). A single-user system is considered and the channel matrix is 8×15, and $r_{\mathrm{S}} = r_{\mathrm{T}} = 3$. These rank orders are assumed to be perfectly known. The MSE is evaluated for $L = 40$ and varying SNR on the top figure and for SNR $= 20$ dB and varying L on the bottom one. Fig. 6 refers to spatially uncorrelated noise, while in Fig. 7, the Gaussian noise is spatially correlated (as in Figs. 4 and 5). The results show that the multiblock RR methods approach the analytical MSE bound and outperform the FR estimate. Moreover, the bound for $L \to \infty$ can be easily reached with a reasonable number of blocks (in practice, $L \geq 20$). In particular, in Fig. 6, the simulation results show that for increasing L, the MSE of the multiblock estimates converge as $1/L$, according to (64a) and (64b): $\mathrm{MSE}_{\mathrm{S}}(L)/\mathrm{MSE}_{\mathrm{S}}(L \to \infty) = 1 + (M - r_{\mathrm{S}})/(WL)$, $\mathrm{MSE}_{\mathrm{T}}(L)/\mathrm{MSE}_{\mathrm{T}}(L \to \infty) = 1 + (W - r_{\mathrm{T}})/(ML)$. For $L \to \infty$, the performance gains of the RR methods with respect to the FR one are, from (32) and (33), (65a) and (65b): $\mathrm{MSE}_{\mathrm{u}}/\mathrm{MSE}_{\mathrm{RR}} = MW/[r_0(M + W - r_0)] = 3$ dB, $\mathrm{MSE}_{\mathrm{u}}/\mathrm{MSE}_{\mathrm{S}} = M/r_{\mathrm{S}} = 4.3$ dB, $\mathrm{MSE}_{\mathrm{u}}/\mathrm{MSE}_{\mathrm{T}} = W/r_{\mathrm{T}} = 7$ dB. Since the temporal gain in this case is larger than the spatial gain ($W/r_{\mathrm{T}} > M/r_{\mathrm{S}}$), it is $\mathrm{MSE}_{\mathrm{T}} < \mathrm{MSE}_{\mathrm{S}}$. Notice that in Fig. 7, the comparison between the two multiblock methods for the same example but with spatially correlated noise yields the opposite result, i.e., $\mathrm{MSE}_{\mathrm{T}} > \mathrm{MSE}_{\mathrm{S}}$. This is due to the spatial correlation of the noise that modifies the MSE of the S-RR method according to (63a), while the MSE of the T-RR method remains unchanged.

C. Multiuser RR Estimate

The performances of single-user and multiuser channel estimates are compared in Fig. 8 for $K = 8$ users and a $L = 1$ training block. The parameters are the same as in Fig. 4. The propagation channel $\mathbf{H}_k$ is simulated independently for each user as a 8×45 random matrix of rank $r_k = 4, \forall k$. The K complex training sequences are obtained according to standard specifications [20] from a single periodic basic code of length $N = 456$. Simulations confirm that single-user algorithms (dashed lines) reach an error floor for large SNR, as they are biased by MAI. On the other hand, multiuser estimates are

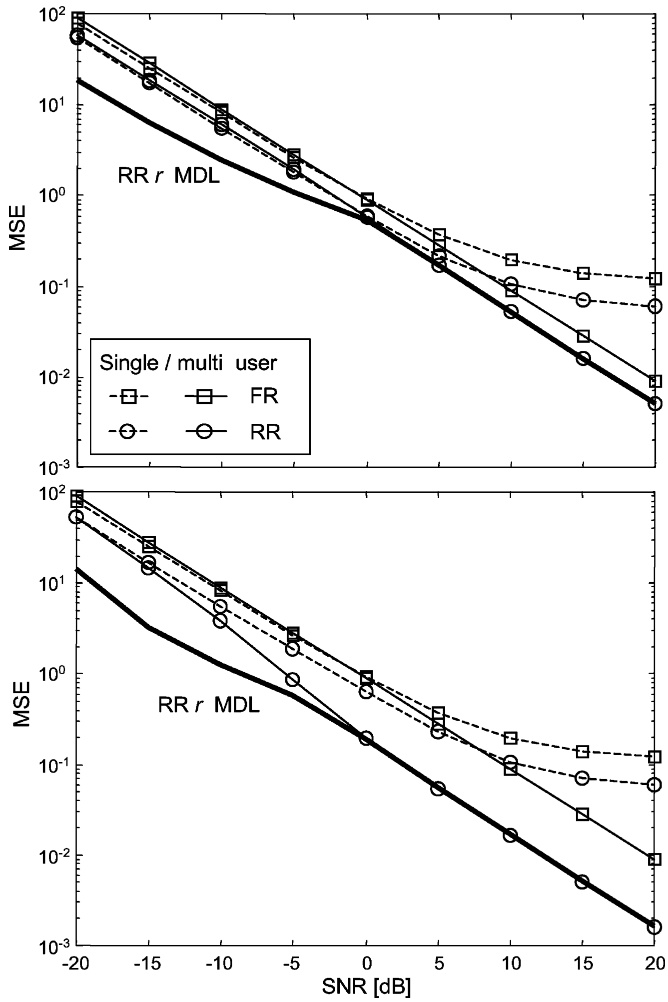

Fig. 8. MSE performance of the single-user and multiuser channel estimates for (top) uncorrelated and (bottom) spatially correlated noise.

shown to properly handle MAI and to outperform single-user methods.

The comparison between Figs. 4 and 8 shows that the MSE of the multiuser RR estimate for $K = 8$ (Fig. 8) has a small degradation compared with $K = 1$ (Fig. 4). Since the multiuser RR estimate is a re-estimation from the FR estimate, the performance for $K = 8$ can be obtained from the single-user case by considering the SNR degradation experienced at the output of the FR estimator due to the interference between the training sequences of different users. As shown in [4], the SNR degradation for UMTS-TDD training sequences is $D = 10\log_{10}[\mathrm{tr}[\mathbf{R}_{xx}^{-1}]/(N/KW)] = 0.65$ dB, which confirms that the training sequences are practically uncorrelated. Therefore, for RR estimation, the iterative approach (76) can be replaced by the solution (80), with a negligible performance degradation.

D. RR Estimate of Multipath Channels

The performance of the RR estimate is herein evaluated for multipath propagation channels with varying angular/delay spread. A stochastic model similar to COST-259 Directional Channel Model [30] is considered. The channel impulse response is modeled as in (2), and the P propagation paths are grouped into N_g disjoint clusters of N_p paths each, where N_g and N_p are deterministic parameters. Within each cluster, angles, delays, and amplitudes of the paths are described by random variables. To account for clustering, variables associated with distinct clusters depend on different cluster mean values. Furthermore, according to the wide sense stationary uncorrelated scattering (WSSUS) assumption, variables associated with different paths are independent from each other. A detailed description of the model is in Appendix F.

The rank analysis for this channel model can be found in [4], which shows that the dependence of the rank order on the number of clusters (i.e., on the overall angular/delay spread) is critical. For urban propagation environments, the number of clusters can be reasonably set to $N_g = 1$ (typical urban) or $N_g = 2$ (bad urban). In the following, we consider channels composed of $N_g \in \{1, 2\}$ clusters, each with delay spread $\sigma_{\tau\max} = 1$ μs, angular spread $\sigma_{\vartheta\max} = 10°$, and number of paths $N_p = 10$ (see Appendix F). This leads to a rank order $r_0 \simeq 2 \div 3$ for the typical urban and $r_0 \geq 4$ for the bad urban environment.

The performance of the RR estimate for single-clustered (top figure) and double-clustered (bottom figure) channels is shown in Figs. 9 and 10. The uplink of the synchronous TD-CDMA system [20] is considered with $K = 8$ users transmitting at fixed spreading factor $Q = 16$. The channel is simulated independently for each user, taking into account slow fading, fast fading, and path loss. Perfect power control is assumed, i.e., differences in attenuation between users due to shadowing and path loss (but not fast fading) are compensated so that $\mathrm{E}(\|\mathbf{H}_k\|^2) = 1$, $\forall k$ (but this power is not equally split between the N_g clusters). The background noise is generated spatially correlated, as in the example of Fig. 4, but with θ uniformly distributed in $[-60°, +60°]$.

Fig. 9 compares the MSE of the multiuser FR and RR estimates. Different RR estimates are obtained by using a fixed rank order (with $r = 1$, 2, 3, and 4, dashed lines) and adaptive selection of the rank order (marker ∘). The comparison between single and double-cluster performance shows that the error floor is larger for $N_g = 2$ as double-cluster channels call for higher rank orders.

The performance of the receiver, complete with channel estimation and MMSE multiuser detection (MUD [31]), is evaluated in Fig. 10 in terms of bit error rate (BER) versus $\mathrm{E}_b/N_0 = Q\mathrm{E}[\|\mathbf{H}_k\|^2]/(2M\sigma^2)$ (here, it is $\sigma_x^2 = 1$). If a fixed rank order is adopted for the RR estimate (dashed lines), the distortion becomes very marked for large SNR, and the RR estimate performs worse than the FR (thick line). On the other hand, the adaptive selection of rank order (marker ∘) guarantees a gain with respect to the FR estimate of approximately 2 to 3 dB for all SNR values. For a single cluster channel, the rank-1 estimate guarantees the lowest BER up to 12 dB, while for the double clustered channel, rank 2 is needed at SNR $= 0$ dB. A further improvement in performance can be obtained by a bootstrap approach [32], i.e., by refining the initial RR channel estimate through decided data symbols. In the simulation (dashed line with marker ∘), the symbols detected by the MMSE-MUD are used to extend the training set from the initial length $N = 456$ up to $N = 656$. As a result, the estimate accuracy is increased

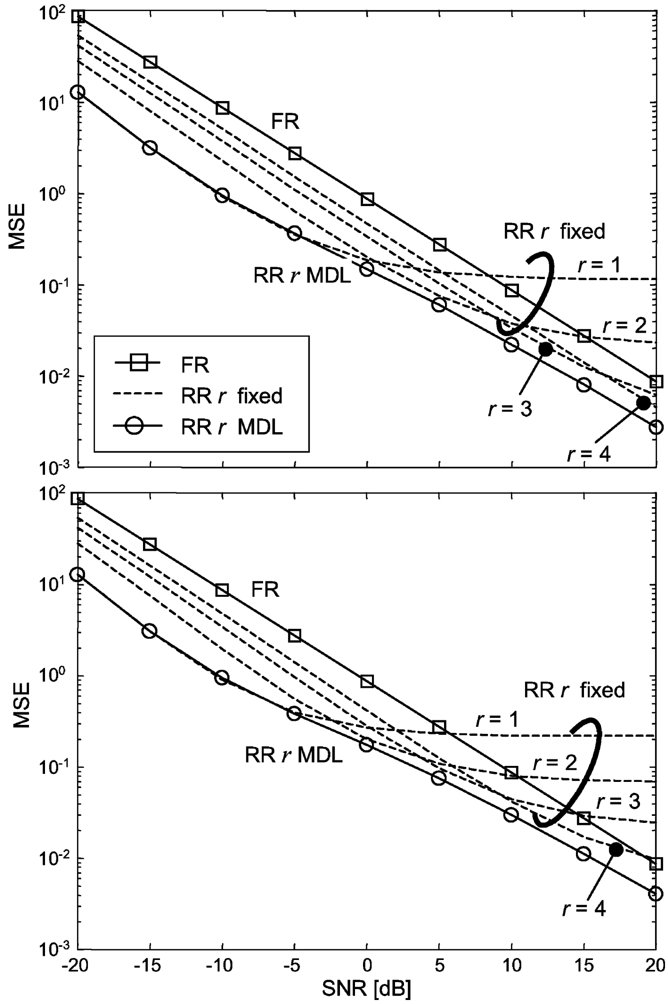

Fig. 9. MSE performance of the RR estimate for (top) single-cluster and (bottom) double-cluster multipath propagation channel in spatially correlated noise.

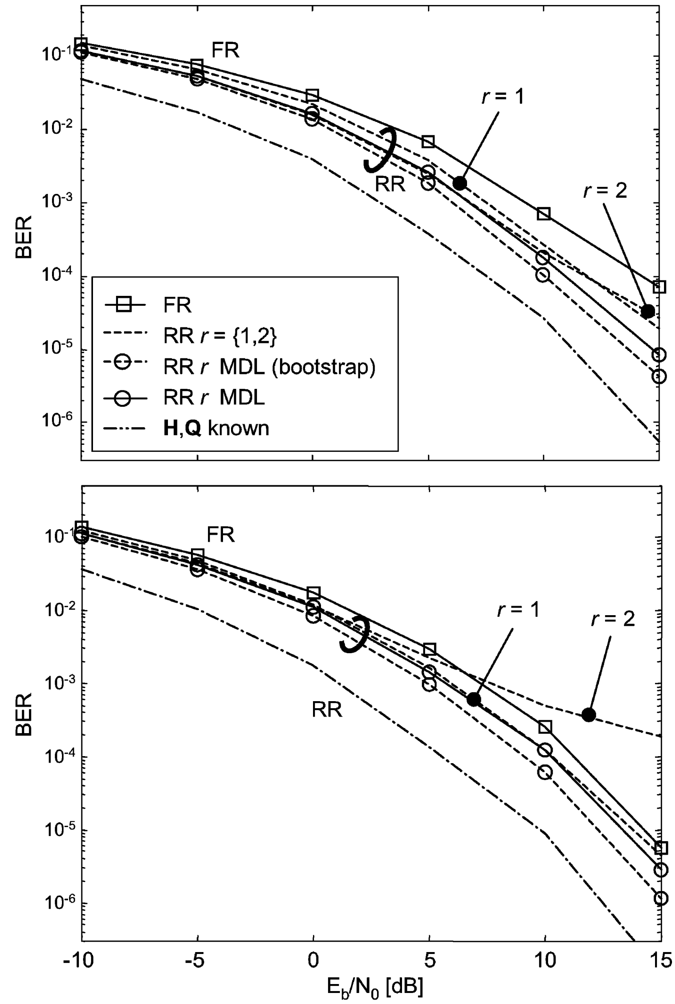

Fig. 10. BER performance of MMSE multiuser detection based on the RR estimate for a (top) single-cluster and a (bottom) double-cluster multipath propagation channel in spatially correlated noise.

by a factor approximately equal to 656/456, and the bit error probability is reduced.

VII. Conclusions

We have investigated reduced-rank estimation of the channel matrix and the covariance of the interference in time-slotted mobile systems. The method is based on an underparameterization of the channel model that constraints the space-time matrix to be low rank. The extension to multiple training periods is obtained by assuming as stationary the spatial or temporal structure of the fast varying channel. New closed-forms of the Cramér–Rao lower bound and the MSE highlight the dependence of the estimate error on the complexity of the channel model (number of unknowns), the overall training sequence length (number of pilot symbols within each training block and number of blocks), and the correlation of the noise and of the training sequence. The model order has been selected as tradeoff between distortion (due to the underparameterization) and variance (due to the limited training set). Simulations indicate that the RR estimate with adaptive selection of the rank order provides remarkable improvement of the performance with respect to the conventional unconstrained estimate.

Appendix A
Proof of (10a)

The derivation of RR MLE (10a) follows the same steps as the derivation of the equivalent solution [6], and only the main differences are highlighted in the following. The minimization of the negative log-likelihood function (8) with respect to $\mathbf{Q}$ yields $\hat{\mathbf{Q}} = (\mathbf{Y} - \hat{\mathbf{H}}\mathbf{X})(\mathbf{Y} - \hat{\mathbf{H}}\mathbf{X})^{\mathrm{H}}/N$. By insertion of $\hat{\mathbf{Q}}$ in (8), the cost function reduces to (see, e.g., [26])

$$\mathcal{L} \propto \log\left(|\mathbf{Q}_{\mathrm{u}}||\mathbf{I} + \mathbf{G}|\right) \tag{81}$$

where $\mathbf{G} = (\tilde{\mathbf{H}} - \tilde{\mathbf{H}}_{\mathrm{u}})^{\mathrm{H}}(\tilde{\mathbf{H}} - \tilde{\mathbf{H}}_{\mathrm{u}})$, $\tilde{\mathbf{H}} = \mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}\mathbf{H}\mathbf{R}_{xx}^{\mathrm{H}/2}$, and $\tilde{\mathbf{H}}_{\mathrm{u}}$ is defined as (12). Letting $\{\lambda_k[\mathbf{G}]\}_{k=1}^{r_{\max}}$ denote the eigenvalues of the matrix $\mathbf{G}$ arranged in nonincreasing order (see notational conventions in Appendix G), the MLE of $\mathbf{H}$ is equivalent to the minimizer of

$$\log|\mathbf{I} + \mathbf{G}| = \log\prod_{k=1}^{r_{\max}}\left(1 + \lambda_k[\mathbf{G}]\right). \tag{82}$$

The matrix $\mathbf{G}$ can be easily rewritten as $\mathbf{G} = \Delta\mathbf{G} + \mathbf{G}_0$, where $\Delta\mathbf{G} = (\tilde{\mathbf{H}} - \mathbf{P}_{\tilde{\mathbf{H}}}\tilde{\mathbf{H}}_{\mathrm{u}})^{\mathrm{H}}(\tilde{\mathbf{H}} - \mathbf{P}_{\tilde{\mathbf{H}}}\tilde{\mathbf{H}}_{\mathrm{u}}) \geq 0$, $\mathbf{G}_0 = \tilde{\mathbf{H}}_{\mathrm{u}}^{\mathrm{H}}(\mathbf{I} - \mathbf{P}_{\tilde{\mathbf{H}}})\tilde{\mathbf{H}}_{\mathrm{u}}$, and $\mathbf{P}_{\tilde{\mathbf{H}}}$ denotes the projector onto the column-space of $\tilde{\mathbf{H}}$. From the Poincaré separation problem (see, e.g., [33]), it can be shown that $\lambda_k[\mathbf{G}] \geq \lambda_k[\mathbf{G}_0] \geq \mu_k$, with

$$\mu_k = \lambda_{r_0+k}\left[\tilde{\mathbf{H}}_{\mathrm{u}}\tilde{\mathbf{H}}_{\mathrm{u}}^{\mathrm{H}}\right], \quad k = 1, \ldots, r_{\max} - r_0. \tag{83}$$

The minimum $\lambda_k[\mathbf{G}_0] = \mu_k$ is obtained if $\tilde{\mathbf{H}} = \hat{\mathbf{P}}_{\mathrm{S},r_0}\tilde{\mathbf{H}}_{\mathrm{u}}$, i.e., if $\mathbf{P}_{\tilde{\mathbf{H}}}$ equals the projector $\hat{\mathbf{P}}_{\mathrm{S},r_0}$ onto the subspace spanned by the r_0 leading eigenvectors of $\hat{\mathbf{R}}_{\mathrm{S}}$. Therefore, the MLE of the space-time matrix is given by $\mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}\hat{\mathbf{H}}\mathbf{R}_{xx}^{\mathrm{H}/2} = \hat{\mathbf{P}}_{\mathrm{S},r_0}\tilde{\mathbf{H}}_{\mathrm{u}}$ for the whitened channel, which yields (10a) after whitening cancellation.

Appendix B
Equivalent Formulations for the RR Estimate

Assuming $\mathbf{R}_{yy} > 0$, the RR MLE proposed in [6] is

$$\hat{\mathbf{H}}' = \mathbf{R}_{yx}\mathbf{R}_{xx}^{-\frac{1}{2}}\hat{\mathbf{P}}'_{\mathrm{T},r_0}\mathbf{R}_{xx}^{-\frac{\mathrm{H}}{2}} \tag{84}$$

where $\hat{\mathbf{P}}'_{\mathrm{T},r_0} = \mathbf{U}'_{\mathrm{T},r_0}\mathbf{U}'^{\mathrm{H}}_{\mathrm{T},r_0}$ is the projector onto the subspace spanned by the r_0 leading eigenvectors $(\mathbf{U}'_{\mathrm{T},r_0})$ of the temporal correlation matrix $\hat{\mathbf{R}}'_{\mathrm{T}} = \tilde{\mathbf{H}}'^{\mathrm{H}}_{\mathrm{u}}\tilde{\mathbf{H}}'_{\mathrm{u}}$, and $\tilde{\mathbf{H}}'_{\mathrm{u}} = \mathbf{R}_{yy}^{-\mathrm{H}/2}\mathbf{H}_{\mathrm{u}}\mathbf{R}_{xx}^{\mathrm{H}/2}$ is a weighted version of the FR estimate. With respect to (12), the spatial weighting in $\tilde{\mathbf{H}}'_{\mathrm{u}}$ is obtained by the term $\mathbf{R}_{yy}^{-\mathrm{H}/2}$. Similarly to the formulation given in Section III, the estimate (84) can be rewritten as the projection of $\tilde{\mathbf{H}}'_{\mathrm{u}}$ performed in the temporal or spatial domain

$$\hat{\mathbf{H}}' = \mathbf{R}_{yy}^{\frac{\mathrm{H}}{2}}\left[\tilde{\mathbf{H}}'_{\mathrm{u}}\hat{\mathbf{P}}'_{\mathrm{T},r_0}\right]\mathbf{R}_{xx}^{-\frac{\mathrm{H}}{2}} \tag{85}$$

$$= \mathbf{R}_{yy}^{\frac{\mathrm{H}}{2}}\left[\hat{\mathbf{P}}'_{\mathrm{S},r_0}\tilde{\mathbf{H}}'_{\mathrm{u}}\right]\mathbf{R}_{xx}^{-\frac{\mathrm{H}}{2}} \tag{86}$$

or, equivalently, as $\hat{\mathbf{H}}' = \mathbf{R}_{yy}^{\mathrm{H}/2} \cdot \mathrm{svd}_{r_0}[\tilde{\mathbf{H}}'_{\mathrm{u}}] \cdot \mathbf{R}_{xx}^{-\mathrm{H}/2}$ (see [5]). Here, $\hat{\mathbf{P}}'_{\mathrm{S},r_0} = \mathbf{U}'_{\mathrm{S},r_0}\mathbf{U}'^{\mathrm{H}}_{\mathrm{S},r_0}$ is the projector onto the subspace spanned by the r_0 leading eigenvectors $(\mathbf{U}'_{\mathrm{S},r_0})$ of the spatial correlation matrix $\hat{\mathbf{R}}'_{\mathrm{S}} = \tilde{\mathbf{H}}_{\mathrm{u}}\tilde{\mathbf{H}}'^{\mathrm{H}}_{\mathrm{u}}$. The estimate $\hat{\mathbf{H}}'$ can be shown to be fully equivalent to the RR solution (10a) and (10b) [4].

Appendix C
Proof of (28) and (29)

By recalling the definition (24) for $\mathbf{J}$ and using (26), the MSE (25) can be rewritten as

$$\begin{aligned}\mathrm{MSE}_{\mathrm{RR}} &= \mathrm{tr}\left[\mathbf{D}\left(\mathbf{D}^{\mathrm{H}}\mathbf{R}_{\mathrm{u}}^{-1}\mathbf{D}\right)^{\dagger}\mathbf{D}^{\mathrm{H}}\right] \\ &= \mathrm{tr}\left[\mathbf{R}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}\tilde{\mathbf{D}}(\tilde{\mathbf{D}}^{\mathrm{H}}\tilde{\mathbf{D}})^{\dagger}\tilde{\mathbf{D}}^{\mathrm{H}}\mathbf{R}_{\mathrm{u}}^{\frac{1}{2}}\right] \\ &= \mathrm{tr}\left[\mathbf{R}_{\mathrm{u}}^{\frac{\mathrm{H}}{2}}\mathbf{P}_{\tilde{\mathbf{D}}}\mathbf{R}_{\mathrm{u}}^{\frac{1}{2}}\right]\end{aligned} \tag{87}$$

where $\mathbf{P}_{\tilde{\mathbf{D}}} = \tilde{\mathbf{D}}(\tilde{\mathbf{D}}^{\mathrm{H}}\tilde{\mathbf{D}})^{\dagger}\tilde{\mathbf{D}}^{\mathrm{H}}$ is the projector onto the column space of the whitened matrix $\tilde{\mathbf{D}} = \mathbf{R}_{\mathrm{u}}^{-\mathrm{H}/2}\mathbf{D}$. Recalling the definition (23), it follows that the matrix $\tilde{\mathbf{D}} = [\tilde{\mathbf{D}}_1 \;\; \tilde{\mathbf{D}}_2]$ is partitioned into the submatrices

$$\tilde{\mathbf{D}}_1 = \tilde{\mathbf{B}}^* \otimes \mathbf{Q}^{-\frac{\mathrm{H}}{2}} \tag{88}$$

$$\tilde{\mathbf{D}}_2 = \mathbf{R}_{xx}^{\frac{\mathrm{T}}{2}} \otimes \tilde{\mathbf{A}}. \tag{89}$$

Therefore, $\mathbf{P}_{\tilde{\mathbf{D}}}$ is the projector onto the union $\mathcal{R}[\widetilde{\mathbf{D}}] = \mathcal{R}[\widetilde{\mathbf{D}}_1] \cup \mathcal{R}[\widetilde{\mathbf{D}}_2]$. The latter can be equivalently decomposed as

$$\mathcal{R}[\widetilde{\mathbf{D}}] = \mathcal{R}\left[\mathbf{P}^{\perp}_{\widetilde{\mathbf{D}}_1}\widetilde{\mathbf{D}}_2\right] \cup \mathcal{R}[\widetilde{\mathbf{D}}_1]$$

where the two subspaces are chosen to be orthogonal: $\mathcal{R}[\mathbf{P}^{\perp}_{\widetilde{\mathbf{D}}_1}\widetilde{\mathbf{D}}_2] \cap \mathcal{R}[\widetilde{\mathbf{D}}_1] = \varnothing$. The projection matrix simplifies then as

$$\mathbf{P}_{\widetilde{\mathbf{D}}} = \mathbf{P}_{\mathbf{P}^{\perp}_{\widetilde{\mathbf{D}}_1}\widetilde{\mathbf{D}}_2} + \mathbf{P}_{\widetilde{\mathbf{D}}_1} \tag{90}$$

where $\mathbf{P}_{\widetilde{\mathbf{D}}_1}$, $\mathbf{P}^{\perp}_{\widetilde{\mathbf{D}}_1}$, and $\mathbf{P}_{\mathbf{P}^{\perp}_{\widetilde{\mathbf{D}}_1}\widetilde{\mathbf{D}}_2}$ denote the orthogonal projections onto, respectively, $\mathcal{R}[\widetilde{\mathbf{D}}_1]$, the orthogonal complement $\mathcal{R}[\widetilde{\mathbf{D}}_1^{\perp}]$, and $\mathcal{R}[\mathbf{P}^{\perp}_{\widetilde{\mathbf{D}}_1}\widetilde{\mathbf{D}}_2]$.

The projectors in (90) can be easily calculated by noticing that for any given matrices $\mathbf{U}$ and $\mathbf{V}$, the projector onto $\mathcal{R}[\mathbf{U} \otimes \mathbf{V}]$ is $\mathbf{P}_{\mathbf{U}\otimes\mathbf{V}} = \mathbf{P}_{\mathbf{U}} \otimes \mathbf{P}_{\mathbf{V}}$. From (88) and (89), and by using the properties of the Kronecker product (see the list of operators in Appendix G), it follows that

$$\mathbf{P}_{\widetilde{\mathbf{D}}_1} = \mathbf{P}_{\mathrm{T}}^* \otimes \mathbf{I}_M \tag{91}$$

$$\mathbf{P}^{\perp}_{\widetilde{\mathbf{D}}_1} = \mathbf{P}_{\mathrm{T}}^{\perp *} \otimes \mathbf{I}_M \tag{92}$$

$$\mathbf{P}^{\perp}_{\widetilde{\mathbf{D}}_1}\widetilde{\mathbf{D}}_2 = \mathbf{P}_{\mathrm{T}}^{\perp *}\mathbf{R}_{xx}^{\frac{\mathrm{T}}{2}} \otimes \tilde{\mathbf{A}} \tag{93}$$

$$\mathbf{P}_{\mathbf{P}^{\perp}_{\widetilde{\mathbf{D}}_1}\widetilde{\mathbf{D}}_2} = \mathbf{P}_{\mathrm{T}}^{\perp *} \otimes \mathbf{P}_{\mathrm{S}} \tag{94}$$

where we used the definition for $\mathbf{P}_{\mathrm{S}}$ in Proposition 3. By substituting (91) and (94) into (90), we get

$$\mathbf{P}_{\tilde{\mathbf{D}}} = \mathbf{P}_{\mathrm{T}}^{\perp *} \otimes \mathbf{P}_{\mathrm{S}} + \mathbf{P}_{\mathrm{T}}^* \otimes \mathbf{I}_M = \mathbf{I}_W \otimes \mathbf{P}_{\mathrm{S}} + \mathbf{P}_{\mathrm{T}}^* \otimes \mathbf{I}_M - \mathbf{P}_{\mathrm{T}}^* \otimes \mathbf{P}_{\mathrm{S}}. \tag{95}$$

The latter result, along with (26) and (87), leads to

$$\mathrm{MSE}_{\mathrm{RR}} = \mathrm{tr}\left[\left(\mathbf{R}_{xx}^{-\frac{1}{2}} \otimes \mathbf{Q}^{\frac{\mathrm{H}}{2}}\right)\mathbf{P}_{\tilde{\mathbf{D}}}\left(\mathbf{R}_{xx}^{-\frac{\mathrm{H}}{2}} \otimes \mathbf{Q}^{\frac{1}{2}}\right)\right] \tag{96}$$

$$\begin{aligned}&= \mathrm{tr}\left[\mathbf{Q}^{\frac{\mathrm{H}}{2}}\mathbf{P}_{\mathrm{S}}\mathbf{Q}^{\frac{1}{2}}\right]\mathrm{tr}\left[\mathbf{R}_{xx}^{-1}\right] + \mathrm{tr}\left[\mathbf{R}_{xx}^{-\frac{1}{2}}\mathbf{P}_{\mathrm{T}}\mathbf{R}_{xx}^{-\frac{\mathrm{H}}{2}}\right]\mathrm{tr}[\mathbf{Q}] \\ &\quad - \mathrm{tr}\left[\mathbf{Q}^{\frac{\mathrm{H}}{2}}\mathbf{P}_{\mathrm{S}}\mathbf{Q}^{\frac{1}{2}}\right]\mathrm{tr}\left[\mathbf{R}_{xx}^{-\frac{1}{2}}\mathbf{P}_{\mathrm{T}}\mathbf{R}_{xx}^{-\frac{\mathrm{H}}{2}}\right]\end{aligned} \tag{97}$$

which can be rewritten into one of the two equivalent forms (28) or (29).

Appendix D
RR Estimate for Temporally and Spatially Correlated Noise

To simplify, both the temporal $(\mathbf{Q}_{\mathrm{T}})$ and the spatial $(\mathbf{Q}_{\mathrm{S}})$ covariance matrices are assumed to be known (e.g., estimated by a sample covariance over $N \to \infty$ observations). From the signal model (7), the negative log-likelihood function is [36]

$$\mathcal{L} = M\ln|\mathbf{Q}_{\mathrm{T}}| + N\ln|\mathbf{Q}_{\mathrm{S}}| + \mathrm{tr}\left[\mathbf{Q}_{\mathrm{S}}^{-1}(\mathbf{Y} - \mathbf{H}\mathbf{X})\mathbf{Q}_{\mathrm{T}}^{-1}(\mathbf{Y} - \mathbf{H}\mathbf{X})^{\mathrm{H}}\right]. \tag{98}$$

By redefining the received and training signals as $\tilde{\mathbf{Y}} = \mathbf{Y}\mathbf{Q}_{\mathrm{T}}^{-1/2}$ and $\tilde{\mathbf{X}} = \mathbf{X}\mathbf{Q}_{\mathrm{T}}^{-1/2}$, the RR MLE for $\mathbf{H}$ turns out to be the minimizer of $\|\tilde{\mathbf{Y}} - \mathbf{H}\tilde{X}\|^2_{\mathbf{Q}_{\mathrm{S}}^{-1}}$ under the constraint $\mathbf{H} = \mathbf{A}\mathbf{B}^{\mathrm{H}}$. It is therefore understood that the RR MLE can be calculated as in (16) by simply replacing $\mathbf{Q}_{\rm u}$ with the known spatial covariance $\mathbf{Q}_{\rm S}$ and $\mathbf{X}$ with $\tilde{\mathbf{X}}$. Similarly, the asymptotic performance for the RR MLE can be calculated from (28) by setting $\mathbf{Q} = \mathbf{Q}_{\rm S}$, replacing $\mathbf{R}_{xx}$ with $\mathbf{R}_{\tilde{x}\tilde{x}}$, and redefining $\mathbf{P}_{\rm S}$ and $\mathbf{P}_{\rm T}$ as the projectors onto the column space of $\tilde{\mathbf{A}} = \mathbf{Q}_{\rm S}^{-{\rm H}/2}\mathbf{A}$ and $\tilde{\mathbf{B}} = \mathbf{R}_{\tilde{x}\tilde{x}}^{1/2}\mathbf{B}$, respectively.

APPENDIX E
PROOF OF (37)

Let $\boldsymbol{\theta} = [\mathbf{a}^{\rm T}\mathbf{b}^{\rm T}]^{\rm T}$ be the vector containing the channel parameters $\mathbf{a} = {\rm vec}[\tilde{\mathbf{A}}]$ and $\mathbf{b} = {\rm vec}[\tilde{\mathbf{B}}^{\rm H}]$, where $\tilde{\mathbf{A}} = \mathbf{W}_{\rm S}\mathbf{A}$ and $\tilde{\mathbf{B}} = \mathbf{W}_{\rm T}\mathbf{B}$ are the spatial/temporal components of the weighted channel matrix $\tilde{\mathbf{H}}(\boldsymbol{\theta}) = \tilde{\mathbf{A}}\tilde{\mathbf{B}}^{\rm H}$. According to (36), the RR estimate $\tilde{\mathbf{H}}(\hat{\boldsymbol{\theta}})$ is obtained from the minimizer $\hat{\boldsymbol{\theta}} = [\hat{\mathbf{a}}^{\rm T}\hat{\mathbf{b}}^{\rm T}]^{\rm T}$ of $\mathcal{F}_N(\boldsymbol{\theta}) = \|\tilde{\mathbf{H}}_{\rm u} - \tilde{\mathbf{H}}(\boldsymbol{\theta})\|^2$. Since the RR estimate is consistent, for $N \to \infty$, the estimated parameters $\hat{\boldsymbol{\theta}}$ tend to the true channel parameters $\boldsymbol{\theta}_0 = [\mathbf{a}_0^{\rm T}\mathbf{b}_0^{\rm T}]^{\rm T}$. Therefore, the asymptotic Taylor expansion can be used for $\mathcal{F}'_N(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$, yielding

$$0=\mathcal{F}'_N(\boldsymbol{\theta}) \simeq \mathcal{F}'_N(\boldsymbol{\theta}_0) + \mathcal{F}''_N(\boldsymbol{\theta}_0)\Delta\boldsymbol{\theta} \simeq \mathcal{F}'_N(\boldsymbol{\theta}_0) + \mathcal{F}''_\infty(\boldsymbol{\theta}_0)\Delta\boldsymbol{\theta}. \quad (99)$$

Here, $\Delta\boldsymbol{\theta} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ denotes the parameter estimate error, $\mathcal{F}'_N(\boldsymbol{\theta}_0)$ is the gradient vector, $\mathcal{F}''_N(\boldsymbol{\theta}_0)$ is the Hessian matrix, and $\mathcal{F}''_\infty(\boldsymbol{\theta}_0) = \lim_{N\to\infty}\mathcal{F}''_N(\boldsymbol{\theta}_0)$ is the limiting Hessian. It is then

$$\Delta\boldsymbol{\theta} = -\left[\mathcal{F}''_\infty(\boldsymbol{\theta}_0)\right]^\dagger \cdot \mathcal{F}'_N(\boldsymbol{\theta}_0). \quad (100)$$

It is convenient to vectorize the channel matrices as $\tilde{\mathbf{h}}_{\rm u} = {\rm vec}[\tilde{\mathbf{H}}_{\rm u}]$, $\tilde{\mathbf{h}}_{\rm RR} = {\rm vec}[\tilde{\mathbf{H}}(\hat{\boldsymbol{\theta}})]$, and $\tilde{\mathbf{h}} = {\rm vec}[\tilde{\mathbf{H}}(\boldsymbol{\theta})]$, yielding

$$\tilde{\mathbf{h}}_{\rm u} = (\mathbf{W}_{\rm T}^* \otimes \mathbf{W}_{\rm S})\,\mathbf{h}_{\rm u} \quad (101)$$

$$\tilde{\mathbf{h}}_{\rm RR} = (\mathbf{W}_{\rm T}^* \otimes \mathbf{W}_{\rm S})\,\mathbf{h}(\hat{\theta}) \quad (102)$$

$$\tilde{\mathbf{h}} = (\tilde{\mathbf{B}}^* \otimes \mathbf{I}_M)\mathbf{a} = (\mathbf{I}_W \otimes \tilde{A})\mathbf{b}. \quad (103)$$

By defining $\Delta\tilde{\mathbf{h}}_{\rm u} = \tilde{\mathbf{h}}_{\rm u} - \tilde{\mathbf{h}}$, the norm to be minimized reduces to $\mathcal{F}_N = \Delta\tilde{\mathbf{h}}_{\rm u}^{\rm H}\Delta\tilde{\mathbf{h}}_{\rm u}$. It follows that $\mathcal{F}'_N(\boldsymbol{\theta}_0) = \tilde{\mathbf{D}}^{\rm H}\Delta\tilde{\mathbf{h}}_{\rm u}$ and $\mathcal{F}''_\infty(\boldsymbol{\theta}_0) = \tilde{\mathbf{D}}^{\rm H}\tilde{\mathbf{D}}$, where $\tilde{\mathbf{D}}$ denotes the derivative of (103) with respect to the parameter vector

$$\tilde{\mathbf{D}} \triangleq -\left[\frac{\partial\tilde{\mathbf{h}}^{\rm H}}{\partial\boldsymbol{\theta}}\right]^{\rm H}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = [\tilde{\mathbf{B}}^* \otimes \mathbf{I}_M \quad \mathbf{I}_W \otimes \tilde{\mathbf{A}}]. \quad (104)$$

From (100), we have $\Delta\boldsymbol{\theta} = -(\tilde{\mathbf{D}}^{\rm H}\tilde{\mathbf{D}})^\dagger\tilde{\mathbf{D}}^{\rm H}\Delta\tilde{\mathbf{h}}_{\rm u}$. The RR estimate error $\Delta\tilde{\mathbf{h}}_{\rm RR} = \tilde{\mathbf{h}}_{\rm RR} - \tilde{\mathbf{h}}$ is then

$$\Delta\tilde{\mathbf{h}}_{\rm RR} = -\tilde{\mathbf{D}}\boldsymbol{\Delta\theta} = \mathbf{P}[\tilde{\mathbf{D}}]\Delta\tilde{\mathbf{h}}_{\rm u} \quad (105)$$

where the projector $\mathbf{P}[\tilde{\mathbf{D}}]$ onto the column space of (104) can be calculated as in (95) from $\mathbf{P}_{\rm S} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^\dagger$ and $\mathbf{P}_{\rm T} = \tilde{\mathbf{B}}\tilde{\mathbf{B}}^\dagger$. From (105), the covariance for the weighted RR estimate is

$$\mathrm{Cov}(\tilde{\mathbf{h}}_{\rm RR}) = \mathbf{P}[\tilde{\mathbf{D}}]\mathrm{Cov}(\tilde{\mathbf{h}}_{\rm u}) \quad (106)$$

while the covariance for the RR estimate $\hat{\mathbf{h}} = \mathbf{h}(\hat{\boldsymbol{\theta}})$ is, from (102)

$$\mathrm{Cov}[\hat{\mathbf{h}}] = \left[\mathbf{W}_{\rm T}^{-1}(\mathbf{P}_{\rm T}\tilde{\mathbf{R}}_{\rm u,T})\mathbf{W}_{\rm T}^{-\rm H}\right]^* \otimes \left[\mathbf{W}_{\rm S}^{-1}\left(\mathbf{P}_{\rm S}^\perp\tilde{\mathbf{R}}_{\rm u,S}\right)\mathbf{W}_{\rm S}^{-\rm H}\right] + \mathbf{R}_{\rm u,T}^* \otimes \left[\mathbf{W}_{\rm S}^{-1}(\mathbf{P}_{\rm S}\tilde{\mathbf{R}}_{\rm u,S})\mathbf{W}_{\rm S}^{-\rm H}\right]. \quad (107)$$

The corresponding MSE is obtained as ${\rm MSE} = {\rm tr}[\mathrm{Cov}(\hat{\mathbf{h}})]$, yielding the results in (37).

APPENDIX F
CLUSTERED MULTIPATH CHANNEL MODEL

The paths of the propagation channel $\mathbf{H} = \sum_{i=1}^{N_g}\sum_{p=1}^{N_r}\alpha_{i,p}\mathbf{a}(\vartheta_{i,p})\mathbf{g}^{\rm T}(\tau_{i,p})$ are grouped into N_g independent clusters of N_r components, and the pth path within the ith cluster is characterized by the parameters $\alpha_{i,p}$, $\vartheta_{i,p}$, and $\tau_{i,p}$. According to the WSSUS assumption, within each cluster $\alpha_{i,p}$, $\vartheta_{i,p}$ and $\tau_{i,p}$ are i.i.d. random variables with probability density functions chosen as described below [30], while the number of clusters N_g and paths N_r are deterministic.

For the ith cluster, the mean value of the delays (τ_i), the angles (ϑ_i), and the powers (P_i) have the following distributions: $\tau_i \sim \tau_1 + \mathcal{U}[0, \tau_{\max}]$, for $i = 2, \ldots, N_g$, with $\tau_{\max} = 6.5\mu$ being the maximum delay; $\vartheta_i \sim \mathcal{U}[-\pi/3, \pi/3]$ and P_i [dB] $\sim \mathcal{N}(L_i\ [\mathrm{dB}], \sigma_p^2)$, for $i = 1, \ldots, N_g$. The delay for the line-of-sight cluster $(i = 1)$ is $\tau_1 = d/c$, where d denotes the distance between mobile and base station. The power P_i is subject to lognormal shadow fading with standard deviation $\sigma_p = 9$ dB. The path loss is $L_i \propto d_i^{-n}$, depending on the distance d_i between the ith cluster and the base station and on the path loss exponent $n = 3.8$. The clusters are uniformly distributed within a circular area of radius 2 km. Within the ith cluster, the delay $(\sigma_{\tau,i})$ and the angular $(\sigma_{\vartheta,i})$ spreads are uniformly distributed $\sigma_{\vartheta,i} \sim \mathcal{U}[0, \sigma_{\vartheta\max}]$, $\sigma_{\tau,i} \sim \mathcal{U}[0, \sigma_{\tau\max}]$, where $\sigma_{\vartheta\max}$ and $\sigma_{\tau\max}$ are the maximum spread values. According to [34], the pth path is characterized by the following parameters: delay $\tau_{i,p} = \tau_i + \Delta\tau_{i,p}$ with $\Delta\tau_{i,p} \sim \mathcal{E}(1.17\sigma_{\tau,i})$; angle $\vartheta_{i,p} = \vartheta_i + \Delta\vartheta_{i,p}$ with $\Delta\vartheta_{i,p} \sim \mathcal{N}(0, 1.9\sigma_{\vartheta,i}^2)$; amplitude $\alpha_{i,p} \sim \mathcal{CN}(0, P_i \cdot P(\Delta\vartheta_{i,p}, \Delta\tau_{i,p}))$ with power delay profile $P(\Delta\vartheta_{i,p}, \Delta\tau_{i,p})$ exponential both in $|\Delta\vartheta_{i,p}|/\sigma_{\vartheta,i}$ and $\Delta\tau_{i,p}/\sigma_{\tau,i}$.

APPENDIX G
NOTATION

Lowercase (uppercase) bold denotes a column vector (matrix). $\mathbf{A}[i,j]$ is the element (i,j) of the matrix $\mathbf{A}$, and $[\mathbf{A}]_{i,j}$ is the block (i,j) of the block-partitioned matrix $\mathbf{A}$. For a matrix $\mathbf{A}(N)$ such that $\mathbf{A}(N) - \mathbf{B} = O(1/N^\alpha)$ with $\alpha > 0$, the following notation is used: For $N \to \infty$ it is $\mathbf{A}(N) \to \mathbf{B}$ and $\mathcal{R}[\mathbf{A}(N)] \to \mathcal{R}[\mathbf{B}]$ (or, equivalently, $\mathbf{A}(N)\mathbf{A}^\dagger(N) \to \mathbf{B}\mathbf{B}^\dagger$); the same properties hold for random matrices with probability 1. For a random variable X, the normal, complex normal, or exponential distribution of mean μ and variance σ^2 is denoted by $X \sim \mathcal{N}(\mu, \sigma^2)$ or $X \sim \mathcal{CN}(0, \sigma^2)$, $X \sim \mathcal{E}(\mu)$, respectively.

In Table I, the top part of the table summarizes the definitions for the symbols used throughout the paper (the matrices

TABLE I

List of symbols

Symbol	**Meaning**
r_0	Rank order of the channel matrix - Sect. II
$r_\mathrm{S}, r_\mathrm{T}$	Spatial/temporal rank orders - Eq. (4)
$r_{\max}$	Maximum rank order - Eq. (4)
$\mathbf{H}\ (\mathcal{H})$	Channel matrix - Eq. (2), (42), (48), (55)
$\mathbf{A}\ (\mathcal{A})$	Spatial channel component - Eq. (6),(46),(48), (55)
$\mathbf{B}\ (\mathcal{B})$	Temporal channel component - Eq. (6),(46),(48), (55)
$\mathbf{Y}\ (\mathcal{Y})$	Received signal matrix - Eq. (7),(45),(49), (56)
$\mathbf{X}\ (\mathcal{X})$	Training sequence matrix - Eq. (7),(45),(49), (56)
$\mathbf{N}\ (\mathcal{N})$	Noise matrix - Eq. (7),(45),(49), (56)
$\mathbf{Q}\ (\mathcal{Q})$	Noise spatial covariance - Sect. I, IV-B.1, IV-C.1
$\mathbf{R}_{xx}\ (\mathcal{R}_{xx})$	Training sequence correlation - Sect. III, IV-B.1, C.1
$\mathbf{H}_\mathrm{u}\ (\mathcal{H}_\mathrm{u})$	Unconstrained MLE of $\mathbf{H}\ (\mathcal{H})$ - Eq. (9a), (44a), Sect. IV-B.2, IV-C.2
$\mathbf{Q}_\mathrm{u}\ (\mathcal{Q}_\mathrm{u})$	Reduced-rank MLE of $\mathbf{Q}\ (\mathcal{Q})$ - Eq. (9b), (44b), Sect. IV-C.2
$\hat{\mathbf{H}}\ (\hat{\mathcal{H}})$	Reduced-rank MLE of $\mathbf{H}\ (\mathcal{H})$ - Eq. (10a), (50), (57)
$\tilde{\mathbf{H}}_\mathrm{u}\ (\tilde{\mathcal{H}}_\mathrm{u})$	Whitened channel est. - Eq. (12), Sect. IV-B.2, C.2
$\mathbf{P}_\mathrm{S}, \mathbf{P}_\mathrm{T}$	Spatial/temporal projectors - Sect. III-C
$\hat{\mathbf{P}}_{\mathrm{S},r_0}, \hat{\mathbf{P}}_{\mathrm{T},r_0}$	Estimate of $\mathbf{P}_\mathrm{S}, \mathbf{P}_\mathrm{T}$ - Eq. (15)
$\tilde{\mathbf{R}}_{\mathrm{S,u}}, \tilde{\mathbf{R}}_{\mathrm{T,u}}$	Spatial/temporal correlations - Eq. (13)

List of operators

Operator	**Definition**
$\mathbf{X}^\mathrm{T}$	Transpose of the matrix $\mathbf{X}$
$\mathbf{X}^\mathrm{H}$	Conjugate transpose of $\mathbf{X}$
$\mathbf{X}^*$	Conjugate of $\mathbf{X}$
$\mathbf{X}^\dagger$	Pseudoinverse of $\mathbf{X}$
$\mathrm{tr}[\mathbf{X}]$	Trace of $\mathbf{X}$
$\lvert\mathbf{X}\rvert$	Determinant of $\mathbf{X}$
$\lVert\mathbf{X}\rVert_\mathbf{A}^2$	Frobenius norm weighted by $\mathbf{A}$: $\mathrm{tr}[\mathbf{A}\mathbf{X}\mathbf{X}^\mathrm{H}]$
$\mathrm{vec}[\mathbf{X}]$	Stack of columns of $\mathbf{X}$ into a vector
$\mathbf{X}^{1/2}$	Cholesky factor for $\mathbf{X}>0$: $\mathbf{X}=\mathbf{X}^{\mathrm{H}/2}\mathbf{X}^{1/2}$
$\mathbf{X}^\dagger$	Pseudoinverse of $\mathbf{X}$
$\lambda_k[\mathbf{X}]$	kth eigenvalue of $\mathbf{X}$ for non-increasing ordering
$\mathrm{svd}[\mathbf{X}]$	Singular value decomposition for $\mathbf{X}$
$\mathrm{svd}_r[\mathbf{X}]$	svd $[\mathbf{X}]$ truncated to the r largest singular values
$\mathcal{R}[\mathbf{X}]$	Subspace spanned by the columns of $\mathbf{X}$
$\mathcal{R}_r[\mathbf{X}]$	Subspace spanned by the columns of $\mathrm{svd}_r[\mathbf{X}]$
$\mathbf{P}_\mathbf{X}$	Orthogonal projector onto $\mathcal{R}[\mathbf{X}]$
$\mathbf{P}_\mathbf{X}^\perp$	Orthogonal projector onto $\mathcal{R}^\perp[\mathbf{X}]$
$\mathbf{P}_{\mathbf{X},r}$	Orthogonal projector onto $\mathcal{R}_r[\mathbf{X}]$
$\mathbf{I}_n$	$n \times n$ identity matrix
$\mathrm{diag}[x_1, ..., x_N]$	Diagonal matrix with elements $x_1, \ldots, x_N$
$\otimes$	Kronecker product, this property holds [35]: $\mathrm{vec}[\mathbf{ABC}]=(\mathbf{C}^\mathrm{T}\otimes\mathbf{A})\,\mathrm{vec}[\mathbf{B}]$,
$\Phi[\mathbf{P}, \mathbf{T}]$	$\mathrm{tr}[\mathbf{T}^{\mathrm{H}/2}\mathbf{P}\mathbf{T}^{1/2}]$, see properties in [19]

are in bold for single-block and in calligraphic within brackets for multiblock), while the bottom part of the table gives the matrix operations.

ACKNOWLEDGMENT

The authors acknowledge the anonymous reviewers for raising discussions that led us to complete the paper with Sections III-D and V.

REFERENCES

[1] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.

[2] M. Wax and A. Leshem, "Joint estimation of time delays and directions of arrival of multiple reflections of a known signal," *IEEE Trans. Signal Processing*, vol. 45, no. 10, pp. 2477–2484, Oct. 1997.

[3] A.-J. van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Processing*, vol. 46, no. 2, pp. 405–418, Feb. 1998.

[4] M. Nicoli, "Multiuser Reduced Rank Receivers for TD/CDMA Systems," Ph.D. dissertation, Politecnico di Milano, Milan, Italy, Dec. 2001.

[5] Y. Hua, M. Nikpour, and P. Stoica, "Optimal reduced rank estimation and filtering," *IEEE Trans. Signal Processing*, vol. 49, no. 3, pp. 457–469, Mar. 2001.

[6] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced rank multivariate linear regressions," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 3069–3078, Dec. 1996.

[7] T. Gustafsson and B. D. Rao, "Statistical analysis of subspace-based estimation of reduced-rank linear regressions," *IEEE Trans. Signal Processing*, vol. 50, no. 1, pp. 151–159, Jan. 2002.

[8] D. Giancola, A. Sanguanini, and U. Spagnolini, "Variable rank receiver structures for low-rank space-time channels," in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, May 1999, pp. 65–69.

[9] E. Lindskog and C. Tidestav, "Reduced rank channel estimation," in *Proc. IEEE Veh. Technol. Conf.*, vol. 2, May 1999, pp. 1126–1130.

[10] A. Dogandvžić and A. Nehorai, "Finite-length MIMO equalization using canonical correlation analysis," *IEEE Trans. Signal Processing*, vol. 50, no. 4, pp. 984–989, Apr. 2002.

[11] M. Nicoli and U. Spagnolini, "Reduced-rank channel estimation and rank order selection for CDMA systems," in *Proc. IEEE Int. Conf. Commun.*, vol. 1, Jun. 2001, pp. 2737–2741.

[12] U. Spagnolini, "Adaptive rank-one receiver for GSM-DCS systems," *IEEE Trans. Veh. Technol.*, vol. 51, no. 5, pp. 1264–1271, Sep. 2002.

[13] J. W. Liang, J. T. Chen, and A. J. Paulraj, "A two-stage hybrid approach for CCI/ISI reduction with space-time processing," *IEEE Commun. Lett.*, vol. 1, no. 6, pp. 163–165, Nov. 1997.

[14] M. A. Lagunas, A. I. Perez-Neia, and J. Vidal, "Joint beamforming and Viterbi equalizer in wireless communications," in *Proc. 31st Asilomar Conf. Signals, Syst., Comput.*, vol. 1, Nov. 1997, pp. 915–919.

[15] G. Bucci, A. Colamonico, M. Donati, A. Picciriello, and U. Spagnolini, "Smart antenna BTS based on software radio technique for GSM/DCS system," in *Proc. IEEE Veh. Technol. Conf.*, vol. 2, May 2000, pp. 1225–1229.

[16] L. L. Scharf, "The SVD and reduced rank signal processing," *Signal Process.*, vol. 25, no. 2, pp. 113–133, Nov. 1991.

[17] M. C. Vanderveen, A. Van der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Processing*, vol. 46, no. 3, pp. 682–690, Mar. 1998.

[18] P. Forster and T. Asté, "Maximum likelihood multichannel estimation under reduced rank constraint," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 6, May 1998, pp. 3317–3320.

[19] M. Nicoli, O. Simeone, and U. Spagnolini, "Multi-slot estimation of fast-varying space-time communication channels," *IEEE Trans. Signal Processing*, vol. 51, no. 5, pp. 1184–1195, May 2003.

[20] H. Holma and A. Toskala, *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*. New York: Wiley, 2000.

[21] M. Nicoli, O. Simeone, and U. Spagnolini, "Multi-slot estimation of fast-varying space-time channels in TD-CDMA systems," *IEEE Commun. Lett.*, vol. 6, no. 9, pp. 376–378, Sep. 2002.

[22] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1991.

[23] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[24] D. J. Rabideau, "Fast, rank adaptive subspace tracking and applications," *IEEE Trans. Signal Processing*, vol. 44, no. 9, pp. 2229–2244, Sep. 1996.

[25] P. Strobach, "Low-rank adaptive filters," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 2932–2947, Dec. 1996.

[26] J. Li, B. Halder, P. Stoica, and M. Viberg, "Computationally efficient angle estimation for signals with known waveforms," *IEEE Trans. Signal Processing*, vol. 43, no. 9, pp. 2154–2163, Sep. 1995.

[27] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. Inform. Theory*, B. N. Petrov and F. Caski, Eds., 1973, pp. 267–281.

[28] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 2, pp. 387–392, Apr. 1985.

[29] P. Strobach, "Low-rank detection of multichannel Gaussian signals using block matrix approximation," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 233–242, Jan. 1995.

[30] L. M. Correia, *Wireless Flexible Personalised Communications. COST 259: European Co-Operation in Mobile Radio Research*. New York: Wiley, 2001.

[31] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[32] E. Lindskog, "Space-Time Processing and Equalization for Wireless Communications," Ph.D. dissertation, Uppsala Univ., Uppsala, Sweden, May 1999.

[33] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[34] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "Dual-polarized model of outdoor propagation environments for adaptive antennas," in *Proc. IEEE Veh. Technol. Conf.*, vol. 2, Jul. 1999, pp. 990–995.

[35] A. Graham, *Kronecker Product and Matrix Calculus*. New York: Wiley, 1981.

[36] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Boca Raton, FL: Chapman and Hall/CRC, 1999.

Monica Nicoli (M'99) received the M.Sc. degree (with honors) and the Ph.D. degree in telecommunication engineering from Politecnico di Milano, Milan, Italy, in 1998 and 2002, respectively.

From March to August 2001, she was a visiting researcher with Signals and Systems, Uppsala University, Uppsala, Sweden. Currently, she is an Assistant Professor with D.E.I. Politecnico di Milano. Her research interests are in the area of signal processing for wireless communication systems, including antenna arrays, channel estimation, equalization, multiuser detection, turbo processing, Hidden Markov Model tracking for radiolocation, and remote sensing applications.

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

Since 1988, he has been with the Dipartimento di Elettronica e Informazione, Politenico 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.