Multislot Estimation of Fast-Varying Space-Time Communication Channels

Monica Nicoli, *Member, IEEE*, Osvaldo Simeone, *Member, IEEE*, and Umberto Spagnolini, *Member, IEEE*

***Abstract*—In mobile communications, coherent detection requires the estimate of an increasing number of channel parameters due to the promising performance of systems that deploy multiple antennas. However, the estimate of space-time channels requires a number of training symbols that grows with the number of unknowns. To overcome this problem, we propose a subspace-based estimation method that exploits the different varying rates in the structure of the space-time channel for moving terminals. Since the channel has some fast-varying (faded amplitudes of the paths) and slowly varying (delays and directions of arrival) features, the multislot (MS) estimate is composed of two terms: the slowly varying spatial and temporal bases estimated from L consecutive slots and the fast-varying amplitudes estimated on a slot-by-slot basis. Performance analysis and simulations confirm the expected benefits of the multislot approach and demonstrate that for large L, the mean square error (MSE) on the channel estimate depends only on the number of fast-varying parameters.**

I. Introduction

In MOST of existing and next-generation mobile communication systems, the multipath fading channel is estimated from known training symbols in order to allow a coherent detection. The accuracy of the estimate depends on the ratio between the number of channel parameters that have to be estimated and the length of the training sequence. Any increase of the latter comes at the price of a reduced transmission efficiency. In systems that deploy multiple antennas at the receiver and/or the transmitter, the augmented number of antennas leads to a proportionally larger number of channel parameters. Due to the prominent role of such systems in the actual and future generation of wireless communication networks, it is mandatory to design a parametric channel estimation technique that is efficient enough to overcome this problem.

In this paper, we propose novel techniques for the estimation of the propagation channel. These methods are here designed for single-user channels in the uplink of time slotted systems and are based on the stationarity of the spatial-temporal channel subspaces across several slots (see also [1]). The approach can be easily extended to multiuser systems (e.g., hybrid TDMA-CDMA systems), MIMO systems, or other transmission schemes such as multicarrier systems. All these applications are straightforward and appear very promising. The multislot (MS) method is based on the recognition that in a multipath channel the angles of arrival (or, in short, angles), and the delays of the paths are slowly varying, whereas the faded amplitudes of the paths are fast-varying. Even if the degree of fading decorrelation depends on the terminals' movement according to the Doppler effect, the amplitudes can be considered quasistatic within each slot (fading variations within the slot are not considered here), whereas the variations of angles and delays can be modeled as quasistatic in a larger time scale, say L slots. The multislot method estimates the slowly varying features (angles and delays) from the ensemble of L slots and the fast-varying amplitudes in a slot-by-slot fashion. The slowly varying terms are obtained without the explicit computation of the angles and delays but by estimating the corresponding spatial and temporal subspaces.

In a time-slotted system, the number of slots L can be chosen to have negligible variations of angles and delays compared with the angular and temporal resolution of the receiver. Notice that the angular resolution depends on the aperture of the antenna array, whereas the temporal resolution is approximately given by the inverse of the signal bandwidth. As an example, let us consider the uplink of a system with inter-slot interval of 5 ms, signal bandwidth of 1.28 MHz (e.g., TD-SCDMA system [2]), and an antenna array of $M = 8$ half wavelength spaced elements. If the mobile terminal is 250 m (or 500 m) from the base station (as for an outdoor environment macro cell), the angles and the delays can be considered stationary within $L = 40$ slots (or $L = 80$ slots), provided that the velocity of the mobile terminal is less than 500 km/h.

The problem of estimating the mobile-to-base station channel in time-slotted systems has been classically solved by using a slot-by-slot approach, such as the least squares estimate (LSE). The accuracy of the estimate can be improved through a parametric approach that reduces the number of the unknowns. This can be either a structured method based on angle-delay estimation [3], [4] or an unstructured method that exploits the low rank property of the space-time channel [5]–[7]. In principle, the estimate accuracy can be increased also by considering training sequences that are apparently longer, as obtained by appropriately merging information from multiple slots. The simplest way to implement such an approach is averaging over the channel estimates relative to consecutive slots [8]. This choice leads to good performance only under rather restrictive assumptions about the terminal mobility, and it cannot cope with time-varying channels, not even with moderately slow fading variations.

In order to make the multislot approach effective, the estimation should be based on the multipath structure of the channel to account for the fading amplitude variations over the slots. The joint estimation of angles and delays based on their invariance across multiple slots has been recently proposed in [9] as

Manuscript received April 17, 2002; revised November 5, 2002. The associate editor coordinating the review of this paper and approving it for publication was Dr. Rick S. Blum.

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

Digital Object Identifier 10.1109/TSP.2003.810302

an extension of the two-dimensional (2-D)-ESPRIT method [3]. When the receiver employs the MS angle-delay estimation, the system performance slightly degrades with respect to the case of perfectly known channel [10]. However, the MS angle-delay estimation is computationally prohibitive for communication systems and does not take into account the correlation properties of the noise. In this paper, we translate the slot-invariance property of angles and delays into the stationarity of the spatial and temporal channel subspaces since this permits us to avoid the expensive angle-delay estimation [9]. The exploitation of the stationarity of the spatial subspace in an MS approach was introduced in [11] (without proof of optimality). Here, we show that the maximum likelihood estimate of the space-time channel can be reduced to the projection of the LSE onto the spatial and temporal subspaces obtained from multiple slots. If the number of slots L is large enough (in practice $L = 10 \div 20$), these projectors can be estimated with enough accuracy so that the mean square error (MSE) for the channel estimate after the projection depends only on the (small) residual noise. Furthermore, differently from [10], here, both the (spatial) correlation properties of the noise and the (temporal) correlation of the training sequence can be taken into account in the estimation of the subspaces. Compared with the standard LSE, the benefits of the MS approach are relevant when the radio channel allows a significant reduction of complexity. It is proved that this occurs when the degree of spatial diversity is smaller than the number of antennas M and the degree of temporal diversity is smaller than the temporal support of the channel W.

The outline of the paper is as follows. Basic model and signals definitions, notations, and the unconstrained MLE are in Section II. Section III highlights the different varying rates in the multipath channel and introduces the corresponding spatial and temporal subspaces for a finite number of slots L. The MS method for channel estimation is proposed in Section IV for joint spatial and temporal subspaces (MS-ST method) and for spatial (MS-S method) or temporal (MS-T method) subspaces. Section V investigates the properties of the asymptotic (i.e., $L \to \infty$) MSE bound for all the proposed estimators and compares analytically the performance with other methods. Since the MS methods are based on the computations of the spatial and/or temporal subspaces from the ensemble of single-slot estimates, the fast subspace tracking method appears to be the mandatory for real-time implementations, which is discussed in Section VI. Section VII contains the numerical results to validate the advantage of the method and the rate of convergence to the analytic MSE bounds. As the interaction between spatial and temporal subspaces is difficult to visualize, some examples are used to improve the clarity.

II. Notations and Preliminaries

Let us introduce some notational conventions that will be used in this paper. Lowercase (uppercase) bold denotes column vector (matrix), $(\cdot)^{\mathrm{T}}$ is the matrix transpose, $(\cdot)^*$ is the complex conjugate, $(\cdot)^{\mathrm{H}}$ is the Hermitian transposition, and $[\cdot]_{m,n}$ is the element (m, n) of the matrix argument. $\otimes$ is the Kronecker matrix product, and $\odot$ is the element-wise multiplication for two matrices of the same size. $||\mathbf{X}||^2_{\mathbf{A}\otimes\mathbf{B}} = \mathrm{tr}(\mathbf{X}^{\mathrm{H}}\mathbf{B}\mathbf{X}\mathbf{A})$ is the matrix Frobenius norm weighted by positive definite matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{R}^{1/2}$ is the Cholesky factorization of a positive definite matrix $\mathbf{R}$: $\mathbf{R} = \mathbf{R}^{\mathrm{H}/2}\mathbf{R}^{1/2}$. $\mathbf{v} = \mathrm{vec}\{\mathbf{V}\}$ is the stacking operator, and the following property is extensively used: $\mathrm{vec}\{\mathbf{ABC}\} = (\mathbf{C}^{\mathrm{T}} \otimes \mathbf{A})\,\mathrm{vec}\{\mathbf{B}\}$ (see [12] for additional properties). $\mathbf{I}_P$ is the $P \times P$ unit matrix, $\mathbf{A}^\dagger$ is the pseudoinverse of $\mathbf{A}$, $\mathcal{R}\{\mathbf{A}\}$ is the subspace spanned by the columns of $\mathbf{A}$, and $\mathbf{\Pi}_{\mathbf{A}} = \mathbf{A}\mathbf{A}^\dagger$ is the projection matrix onto $\mathcal{R}\{\mathbf{A}\}$. For a matrix $\mathbf{A}(L)$ such that $\mathbf{A}(L) - \mathbf{B} = O(1/L^\alpha)$ for $\alpha > 0$, the following asymptotic (for $L \to \infty$) notation will be used: $\mathbf{A}(L) \to \mathbf{B}$ and $\mathcal{R}\{\mathbf{A}(L)\} \to \mathcal{R}\{\mathbf{B}\}$ (or, equivalently, $\mathbf{A}(L)\mathbf{A}^\dagger(L) \to \mathbf{B}\mathbf{B}^\dagger$); the same properties hold with probability 1 for random matrices.

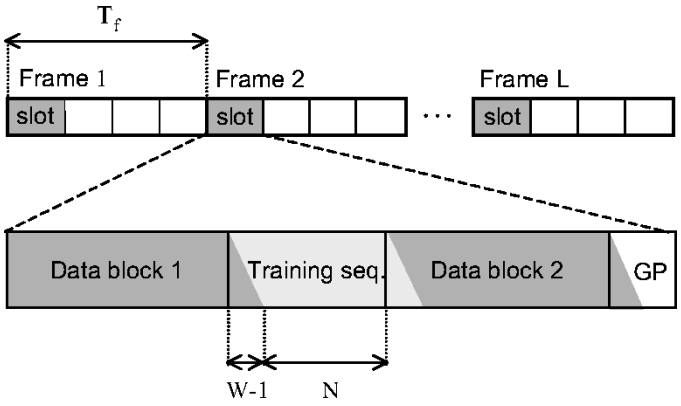


Fig. 1. Frame structure of time-slotted systems.

A. Signal Model

Let us consider a time-slotted wireless communication system in which a mobile user transmits bursts to a base station equipped with an antenna array of M elements. The frame structure is illustrated in Fig. 1; within each frame, a single slot is allocated to each user so that the time interval between two consecutive slots for the same user equals the frame duration T_f. During the time slot, the mobile terminal transmits a burst consisting of two data blocks, a training sequence (midamble), and a guard period. At each antenna, the baseband received signal is passed through a filter matched to the pulse waveform, and the signal at the output of the matched filter is modeled as the $M \times 1$ vector

$$\mathbf{y}(t;\,\ell) = \sum_i x(i;\,\ell)\mathbf{h}(t - iT;\,\ell) + \mathbf{n}(t;\,\ell) \qquad (1)$$

where t denotes the time variable within the time slot of the ℓth frame, whereas the complex value $x(i;\,\ell)$ denotes the ith symbol of the transmitted sequence (either information or training data) at rate $1/T$. The space-time channel from the mobile transmitter to the antenna array is described by the $M \times 1$ impulse vector $\mathbf{h}(t;\,\ell)$, which accounts for the array response, the fading channel, the symbol waveform, and the matched filter at the receiver. In general, the channel is time-varying; however, in this paper, the time slot is assumed to be short enough so that the channel can be simplified as invariant within the burst interval, whereas it can vary from frame to frame. The index ℓ denotes the dependence to the frame. Furthermore, the $M \times 1$ zero mean circularly symmetric Gaussian vector $\mathbf{n}(t;\,\ell)$

models both the co-channel interference and the background noise, and it is temporally uncorrelated but spatially correlated.

During the training period, the transmitted sequence $\{x(i;\,\ell)\}_{i=-W+1}^{N-1}$ is known at the receiver in order to allow the estimation of the propagation channel. The discrete-time model for the signal received within the midamble is obtained by sampling at the symbol rate $1/T$ the output of the matched filter. The temporal support of the channel is assumed to be $[0,\,WT)$. Notice that the first $W-1$ samples of the received signal cannot be used for channel estimation as they are affected by the interference from the preceding data symbols. Hence, by discarding the first $W-1$ samples of the signal $\mathbf{y}(t;\,\ell)$ and arranging the remaining samples into the $M\times N$ matrix $\mathbf{Y}(\ell)=[\mathbf{y}(0;\,\ell)\cdots\mathbf{y}((N-1)T;\,\ell)]$, the data model (1) can be rewritten by using the standard notation

$$\mathbf{Y}(\ell)=\mathbf{H}(\ell)\mathbf{X}(\ell)+\mathbf{N}(\ell). \tag{2}$$

Here, $\mathbf{H}(\ell)=[\mathbf{h}(0;\,\ell),\,\mathbf{h}(T;\,\ell),\,\ldots,\,\mathbf{h}((W-1)T;\,\ell)]^{\mathrm{T}}$ is the $M\times W$ space-time channel matrix, and $\mathbf{X}(\ell)$ is the $W\times N$ Toeplitz matrix that represents the convolution of the channel with the training sequence, i.e., $[\mathbf{X}(\ell)]_{m,\,n}=x(n-m;\,\ell)$. Moreover, the discrete-time noise $\mathbf{N}(\ell)=[\mathbf{n}(0;\,\ell)\cdots\mathbf{n}((N-1)T;\,\ell)]$ is Gaussian, still temporally uncorrelated (the transmitted symbol waveforms are orthogonal), but spatially correlated with covariance $\mathrm{E}[\mathbf{n}(iT;\,\ell)\mathbf{n}^{\mathrm{H}}((i+m)T;\,\ell)]=\mathbf{Q}\delta(m)$. The noise covariance matrix $\mathbf{Q}$ is full-rank, and it accounts for the thermal noise and the spatial arrangement of the interferers (see Section VII). Since the spatial features of the interference are slowly varying, $\mathbf{Q}$ is assumed to be independent of the slot. All the antennas have the same noise power $[\mathbf{Q}]_{m,\,m}=\sigma_n^2$.

B. Unconstrained Channel Estimate

From the model (2), the unconstrained maximum likelihood estimate (MLE) of the channel and the noise covariance matrix can be shown to be

$$\mathbf{H}_{\mathrm{u}}(\ell)=\mathbf{R}_{yx}(\ell)\mathbf{R}_{xx}^{-1} \tag{3a}$$

$$\mathbf{Q}_{\mathrm{u}}=\frac{1}{L}\sum_{\ell=1}^{L}\left(\mathbf{R}_{yy}(\ell)-\mathbf{R}_{yx}(\ell)\mathbf{R}_{xx}^{-1}\mathbf{R}_{yx}^{\mathrm{H}}(\ell)\right) \tag{3b}$$

where $\mathbf{R}_{yx}(\ell)=\mathbf{Y}(\ell)\mathbf{X}^{\mathrm{H}}(\ell)/N$, $\mathbf{R}_{yy}(\ell)=\mathbf{Y}(\ell)\mathbf{Y}^{\mathrm{H}}(\ell)/N$, and the training sequence $\mathbf{X}(\ell)$ is assumed to have correlation matrix $\mathbf{R}_{xx}=\mathbf{X}(\ell)\mathbf{X}^{\mathrm{H}}(\ell)/N$ independent on the slot. Notice that the estimate (3a) coincides with the LSE of the space-time channel matrix when this is carried out on a slot-by-slot basis.

The estimate (3a) is known to be unbiased, and $\mathbf{Q}_{\mathrm{u}}-\mathbf{Q}=O(1/NL)$; therefore, it is $\mathbf{Q}_{\mathrm{u}}\to\mathbf{Q}$ for $NL\to\infty$ [5]. The MSE of the LSE can be obtained from the vectorized error

$$\Delta\mathbf{h}_{\mathrm{u}}=\mathrm{vec}\{\mathbf{H}_{\mathrm{u}}(\ell)-\mathbf{H}(\ell)\}=\frac{1}{N}\left[\left(\mathbf{R}_{xx}^{-1}\mathbf{X}(\ell)\right)^{*}\otimes\mathbf{I}_M\right]\mathbf{n}(\ell)$$

where $\mathbf{n}(\ell)=\mathrm{vec}\{\mathbf{N}(\ell)\}$. Indeed, the covariance of the channel estimator is

$$\mathrm{Cov}\{\mathrm{vec}\{\mathbf{H}_{\mathrm{u}}(\ell)\}\}=\mathrm{E}\left[\Delta\mathbf{h}_{\mathrm{u}}\Delta\mathbf{h}_{\mathrm{u}}^{\mathrm{H}}\right]=\frac{1}{N}\,\mathbf{R}_{xx}^{-*}\otimes\mathbf{Q}$$

and therefore, the MSE is given by

$$\begin{aligned}\mathrm{MSE}_{\mathrm{u}}&=\mathrm{E}\left[\|\mathbf{H}_{\mathrm{u}}(\ell)-\mathbf{H}(\ell)\|^2\right]=\mathrm{tr}\{\mathrm{Cov}\{\mathbf{h}_{\mathrm{u}}\}\}\\&=\frac{1}{N}\,\mathrm{tr}\{\mathbf{Q}\}\,\mathrm{tr}\{\mathbf{R}_{xx}^{-1}\}=\frac{M\sigma_n^2}{N}\,\mathrm{tr}\{\mathbf{R}_{xx}^{-1}\}.\end{aligned} \tag{4}$$

For training sequences with ideal correlation properties ($\mathbf{R}_{xx}=\sigma_x^2\mathbf{I}_W$), the MSE (4) reduces to

$$\mathrm{MSE}_{\mathrm{u}}=\frac{\sigma_n^2}{\sigma_x^2}\cdot\frac{MW}{N}=\rho MW \tag{5}$$

with $\rho=\sigma_n^2/N\sigma_x^2$. As a result, the accuracy of the LSE depends on the ratio between the channel parameters (MW) and the training sequence length (N). The purpose of the paper is to reduce the MSE by a parsimonious parameterization of the space-time channel that takes advantage of the multislot redundancy.

III. Subspace Structure of the Propagation Channel

A. Subspace Channel Model

Within a set of L consecutive bursts, the channel can be modeled as a combination of P paths, each of them is characterized by a delay τ_p, an angle α_p, and an amplitude $\beta_p(\ell)$ that accounts for the fading variations:

$$\mathbf{h}(t;\,\ell)=\sum_{p=1}^{P}\beta_p(\ell)\mathbf{a}(\alpha_p)g(t-\tau_p). \tag{6}$$

The waveform $g(t)$ is the convolution of the transmitted pulse and the matched filter at the receiver, and the $M\times 1$ vector $\mathbf{a}(\alpha_p)$ is the array response to a plane wave with direction of arrival α_p. For instance, for a uniform linear array of half-wavelength spaced antennas, it is $\mathbf{a}(\alpha_p)=[1,\,\exp(-j\pi\sin\alpha_p),\,\ldots,\,\exp(-j\pi\sin\alpha_p(M-1))]^{\mathrm{T}}$.

The variations of the angles and delays over the L slots are assumed to be smaller than the angular-temporal resolution so that the parameters $\{\alpha_p,\,\tau_p\}_{p=1}^{P}$ in model (6) can be considered slot-independent. On the other hand, the slot-dependent amplitudes $\boldsymbol{\beta}(\ell)=[\beta_1(\ell),\,\ldots,\,\beta_P(\ell)]^{\mathrm{T}}$ are assumed to be uncorrelated with $\mathbf{R}_{\boldsymbol{\beta}}=\mathrm{E}[\boldsymbol{\beta}(\ell)\boldsymbol{\beta}^{\mathrm{H}}(\ell)]=\mathrm{diag}\{\sigma_1^2,\,\ldots,\,\sigma_P^2\}$ according to the WSSUS channel model [13]. Each amplitude $\beta_p(\ell)$ is a stationary and ergodic (up to the second order) process with $\mathrm{E}[\beta_p^{*}(\ell)\boldsymbol{\beta}_p(\ell+m)]=\sigma_p^2\cdot\varphi(m)$. According, for instance, to Clarke's isotropic scattering model [13], the normalized correlation function $\varphi(m)$ depends only on the time interval mT_f and on the terminal mobility. Except for the static channel $(\varphi(m)=1)$, for any moving terminal, the fading is uncorrelated over a large number of slots as $\varphi(m)\to 0$ for $m\to\infty$. It is worth anticipating that most of the reasonings proposed throughout the paper largely simplify when the fading is uncorrelated from slot-to-slot, i.e., $\varphi(m)=\delta(m)$. This occurs when the frame duration T_f is large compared with the channel coherence time or, as an extension of the model discussed here, when each slot is transmitted on a different frequency bandwidth (e.g., in frequency hopping or multicarrier system).

The space–time channel matrix is obtained by sampling at the symbol rate the response (6), and rearranging the terms as

$$\mathbf{H}(\ell) = \sum_{p=1}^{P} \beta_p(\ell)\mathbf{a}(\alpha_p)\mathbf{g}^{\mathrm{T}}(\tau_p) = \mathbf{A}\mathbf{D}(\ell)\mathbf{G}^{\mathrm{T}} \tag{7}$$

the $W \times 1$ vector $\mathbf{g}(\tau_p) = [g(-\tau_p),\, g(T-\tau_p),\, \ldots,\, g((W-1)T-\tau_p)]^{\mathrm{T}}$ is the sampled delayed waveform, and the set of P vectors for the delays $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_P]^{\mathrm{T}}$ are collected into the real-valued temporal matrix $\mathbf{G} = [\mathbf{g}(\tau_1), \ldots, \mathbf{g}(\tau_P)]$. Similarly, the spatial (or steering) matrix $\mathbf{A} = [\mathbf{a}(\alpha_1), \ldots, \mathbf{a}(\alpha_P)]$ depends on the P angles $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_P]^{\mathrm{T}}$; the diagonal matrix $\mathbf{D}(\ell) = \mathrm{diag}\{\boldsymbol{\beta}(\ell)\}$ contains the amplitudes.

The property of model (7) to separate the slot-dependent term $\mathbf{D}(\ell)$ from the slot-independent matrices $\mathbf{A}$ and $\mathbf{G}$ was exploited in [3] to jointly estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\tau}$. Here, in order to avoid the computationally expensive estimation of the angle-delay pairs, we propose to reparameterize the channel (7) in terms of unstructured slot-varying/unvarying matrices. This approach is derived from the multipath model (7) by exploiting the diversity of the propagation channel. The order of the spatial (r_{S}) and the temporal (r_{T}) diversity are

$$r_{\mathrm{S}} = \dim \mathcal{R}\{\mathbf{A}\} = \mathrm{rank}\{\mathbf{A}\} \le M \tag{8a}$$

$$r_{\mathrm{T}} = \dim \mathcal{R}\{\mathbf{G}\} = \mathrm{rank}\{\mathbf{G}\} \le W \tag{8b}$$

respectively. The first accounts for the number of angles that can be resolved in $\boldsymbol{\alpha}$ (given the array aperture), whereas the second equals the number of the resolvable delays in $\boldsymbol{\tau}$ (given the bandwidth of the transmitted signal). For P paths, it is $r_{\mathrm{S}} \le P$ and $r_{\mathrm{T}} \le P$. In many practical situations, P can be very large, but the order of diversity depends only on few clusters of scatterers with moderate angle-delay spread. This makes the diversity orders r_{S} and r_{T} smaller than P, and it allows the use of the parsimonious parameterization described below.

Let us define the spatial and temporal correlation of the multislot channel for a set of L slots as

$$\mathbf{R}_{\mathrm{S}}(L) = \frac{1}{L}\sum_{\ell=1}^{L} \mathbf{H}(\ell)\mathbf{H}^{\mathrm{H}}(\ell) \tag{9a}$$

$$\mathbf{R}_{\mathrm{T}}(L) = \frac{1}{L}\sum_{\ell=1}^{L} \mathbf{H}^{\mathrm{H}}(\ell)\mathbf{H}(\ell) \tag{9b}$$

so that $\mathrm{rank}\{\mathbf{R}_{\mathrm{S}}(L)\} = r_{\mathrm{S}}(L)$ and $\mathrm{rank}\{\mathbf{R}_{\mathrm{T}}(L)\} = r_{\mathrm{T}}(L)$. Given the eigenvector sets $\mathbf{U}_{\mathrm{S}}$ and $\mathbf{U}_{\mathrm{T}}$ for the correlation matrices (9a) and (9b), respectively, the channel matrix can be parameterized as

$$\mathbf{H}(\ell) = \mathbf{U}_{\mathrm{S}}\boldsymbol{\Gamma}(\ell)\mathbf{U}_{\mathrm{T}}^{\mathrm{H}} \tag{10}$$

where the $r_{\mathrm{S}}(L) \times r_{\mathrm{T}}(L)$ full-rank matrix $\boldsymbol{\Gamma}(\ell)$ is slot-dependent, and $\mathbf{U}_{\mathrm{S}}$ and $\mathbf{U}_{\mathrm{T}}$ are slot-independent matrices of dimension $M \times r_{\mathrm{S}}(L)$ and $W \times r_{\mathrm{T}}(L)$, respectively. Notice that, different from (7), here, $\mathbf{U}_{\mathrm{S}}$ and $\mathbf{U}_{\mathrm{T}}$ are nonstructured matrices. In the following, $\mathcal{R}\{\mathbf{R}_{\mathrm{S}}(L)\} = \mathcal{R}\{\mathbf{U}_{\mathrm{S}}\}$ will be referred to as the spatial subspace and $\mathcal{R}\{\mathbf{R}_{\mathrm{T}}(L)\} = \mathcal{R}\{\mathbf{U}_{\mathrm{T}}\}$ as the temporal subspace.

B. Spatial and Temporal Subspaces

For finite L, the structure of the sample correlation matrices (and therefore $\mathbf{U}_{\mathrm{S}}$ and $\mathbf{U}_{\mathrm{T}}$) cannot be ascribed only to the angle-delay pattern, as the interaction between the faded amplitudes can make the spatial and temporal subspaces interfere with each other. This can be easily proved by considering the correlation matrices (9a) and (9b) for the multipath model (7):

$$\mathbf{R}_{\mathrm{S}}(L) = \mathbf{A}\boldsymbol{\Lambda}_{\mathrm{S}}(L)\mathbf{A}^{\mathrm{H}} \tag{11a}$$

$$\mathbf{R}_{\mathrm{T}}(L) = \mathbf{G}\boldsymbol{\Lambda}_{\mathrm{T}}(L)\mathbf{G}^{\mathrm{T}} \tag{11b}$$

where the $P \times P$ matrices are

$$\boldsymbol{\Lambda}_{\mathrm{S}}(L) = \mathbf{R}_{\boldsymbol{\beta}}(L) \odot \left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right) \tag{12a}$$

$$\boldsymbol{\Lambda}_{\mathrm{T}}(L) = \mathbf{R}_{\boldsymbol{\beta}}(L) \odot \left(\mathbf{A}^{\mathrm{H}}\mathbf{A}\right) \tag{12b}$$

and $\mathbf{R}_{\boldsymbol{\beta}}(L) = (1/L)\sum_{\ell=1}^{L} \boldsymbol{\beta}(\ell)\boldsymbol{\beta}^{\mathrm{H}}(\ell)$ is the sample correlation matrix of fading. From (11a), it follows that $\mathcal{R}\{\mathbf{U}_{\mathrm{S}}\} \subseteq \mathcal{R}\{\mathbf{A}\}$ with equality if $\boldsymbol{\Lambda}_{\mathrm{S}}(L)$ is full rank. In fact, the following equality holds: $r_{\mathrm{S}}(L) = \min\{r_{\mathrm{S}}, \mathrm{rank}\{\boldsymbol{\Lambda}_{\mathrm{S}}(L)\}\}$, where the first term is the degree of spatial diversity, and the second term depends on the temporal characteristics of the channel and on the fading variations within the L slots. The rank of $\boldsymbol{\Lambda}_{\mathrm{S}}(L)$ is indeed a function of $\mathrm{rank}\{\mathbf{G}^{\mathrm{T}}\mathbf{G}\}$, which equals r_{T}, and of $\mathrm{rank}\{\mathbf{R}_{\boldsymbol{\beta}}(L)\}$, which depends on the degree of fading decorrelation in L slots.

Let us now consider two extreme cases. For a static channel $(v = 0)$ or $L = 1$, the fading variable $\boldsymbol{\beta}(\ell)$ is independent on the slot, and it is $\mathrm{rank}\{\boldsymbol{\Lambda}_{\mathrm{S}}(L)\} = r_{\mathrm{T}}$. Therefore, the spatial subspace dimension is $r_{\mathrm{S}}(L) = \min\{r_{\mathrm{S}}, r_{\mathrm{T}}\}$, which equals the rank-order of the single-slot channel matrix [7]. On the other hand, for $L \to \infty$, it is $\boldsymbol{\Lambda}_{\mathrm{S}}(L) \to \mathrm{diag}\,\{\sigma_1^2\|\mathbf{g}(\tau_1)\|^2, \ldots, \sigma_P^2\|\mathbf{g}(\tau_P)\|^2\}$, and thus , $\mathcal{R}\{\mathbf{U}_{\mathrm{S}}\} \to \mathcal{R}\{\mathbf{A}\}$ and $r_{\mathrm{S}}(\infty) = \min\{r_{\mathrm{S}}, P\} = r_{\mathrm{S}}$. The same reasoning can be dually considered for (11b) as $\mathcal{R}\{\mathbf{U}_{\mathrm{T}}\} \subseteq \mathcal{R}\{\mathbf{G}\}$.

The spatial subspace tends to the subspace spanned by the antenna array responses to the directions of arrival $\boldsymbol{\alpha}$, whereas the rank $r_{\mathrm{S}}(L)$ approaches the order of spatial diversity. Moreover, it is easy to show that for increasing L $\mathrm{rank}\{\mathbf{R}_{\boldsymbol{\beta}}(L)\}$ is nondecreasing, and so is $\mathrm{rank}\{\boldsymbol{\Lambda}_{\mathrm{S}}(L)\}$. A similar reasoning holds for the temporal domain. For moving terminals and L large enough, the spatial-temporal subspaces depend only on the slow varying features of the propagation channel (i.e., the angle-delay pattern). As a summary

$$\min\{r_{\mathrm{S}}, r_{\mathrm{T}}\} \le r_{\mathrm{S}}(L) \le r_{\mathrm{S}} \tag{13a}$$

$$\min\{r_{\mathrm{S}}, r_{\mathrm{T}}\} \le r_{\mathrm{T}}(L) \le r_{\mathrm{T}} \tag{13b}$$

and asymptotically, $r_{\mathrm{S}}(L) \to r_{\mathrm{S}}$, and $r_{\mathrm{T}}(L) \to r_{\mathrm{T}}$.

A simple example can clarify the preceding discussion. Let us consider a space-time channel with $r_{\mathrm{S}} = 2$ and $r_{\mathrm{T}} = 8$, $\varphi(m) = J_0(2\pi f_{\mathrm{D}} m T_{\mathrm{f}})$ (the model is exemplified with just these parameters, and the complete description of the simulation environment will be given in Section VII). Fig. 2 shows the average values of $r_{\mathrm{S}}(L)$ and $r_{\mathrm{T}}(L)$ obtained from 10^4 independent runs of fading versus the fading decorrelation within $L = \overline{L} = 10$ slots. The choice of $\overline{L} = 10$ is irrelevant as the rank order depends, namely, on the product $f_{\mathrm{D}} L T_{\mathrm{f}}$ for L large enough. For a static channel $(f_{\mathrm{D}} = 0)$, Fig. 2 shows that

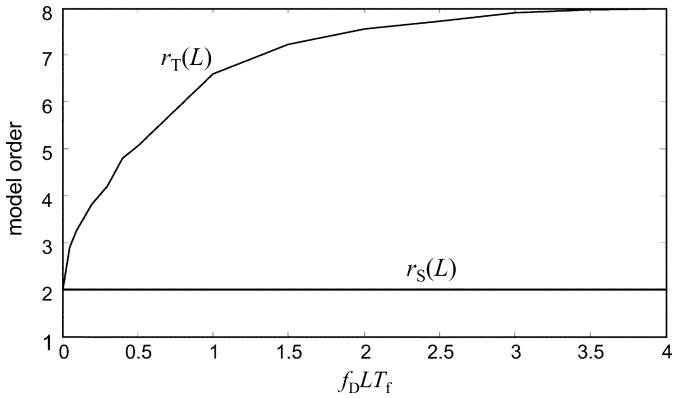


Fig. 2. Average model order $\{r_{\mathrm{S}}(L),\, r_{\mathrm{T}}(L)\}$ versus fading correlation measured in terms of $f_{\mathrm{D}}LT_{\mathrm{f}}$ ($\mathrm{rank}(\mathbf{A}(\boldsymbol{\alpha})) = 2$, $\mathrm{rank}(\mathbf{G}(\boldsymbol{\tau})) = 8$, $L = 10$, $M = 8$, $W = 15$, $N = 40$).

$r_{\mathrm{S}}(L) = r_{\mathrm{T}}(L) = \min(r_{\mathrm{S}}, r_{\mathrm{T}}) = 2$, which is the same result that would have been obtained for $L = 1$. Moreover, for $f_{\mathrm{D}}LT_{\mathrm{f}}$ large enough to guarantee uncorrelated fading within L slots, the model order tends to the limit imposed by the spatial and temporal resolution: $r_{\mathrm{S}}(L) \to r_{\mathrm{S}} = 2$ and $r_{\mathrm{T}}(L) \to r_{\mathrm{T}} = 8$. In conclusion, the interactions between the spatial and the temporal subspaces are very difficult to cope with when L is small or the fading is correlated across slots, whereas for large enough L or uncorrelated fading, each subspace depends only on the multipath pattern in the corresponding angular or temporal domain.

IV. Multislot Estimate of Spatial and Temporal Subspaces

In this section, we derive an MS estimation for the channel matrices $\mathbf{H}(\ell)$ ($\ell = 1, \ldots, L$) that accounts for both the slow and the fast-varying features of the multipath structure (10). By using the parameterization (10), the signal model (2) can be rewritten as

$$\mathbf{Y}(\ell) = \mathbf{U}_{\mathrm{S}}\boldsymbol{\Gamma}(\ell)\mathbf{U}_{\mathrm{T}}^{\mathrm{H}}\mathbf{X}(\ell) + \mathbf{N}(\ell), \qquad \ell = 1, \ldots, L. \quad (14)$$

The spatial covariance matrix $\mathbf{Q}$ (not known) and the correlation of the training sequence $\mathbf{R}_{xx}$ are both assumed to be constant over the L slots. The correlation of the fading amplitudes over the slots is not explicitly modeled (i.e., the knowledge of matrix $\mathbf{R}_{\boldsymbol{\beta}}(L)$ defined in Section III is not required).

In the following, we first discuss the MS maximum likelihood method that exploits the stationarity of the spatial *and* the temporal subspace (MS-ST algorithm). Then, we simplify the method by imposing only one constraint of stationarity, for either the spatial (MS-S) *or* the temporal (MS-T) subspace. An example will be used to illustrate the advantages in terms of noise reduction.

A. MS-ST

According to (14), the MLE under the constraint (10) is obtained by minimizing the negative log-likelihood function

$$\mathcal{L}(\boldsymbol{\theta}, \mathbf{Q}) = \ln|\mathbf{Q}| + \frac{1}{NL}\sum_{\ell=1}^{L} \cdot \left\|\mathbf{Y}(\ell) - \mathbf{U}_{\mathrm{S}}\boldsymbol{\Gamma}(\ell)\mathbf{U}_{\mathrm{T}}^{\mathrm{H}}\mathbf{X}(\ell)\right\|^2_{\mathbf{I}_N\otimes\mathbf{Q}^{-1}} \quad (15)$$

with respect to $\mathbf{Q}$ and the parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_{\mathbf{U}}^{\mathrm{T}}\boldsymbol{\theta}_{\boldsymbol{\Gamma}}^{\mathrm{T}}]^{\mathrm{T}}$, where $\boldsymbol{\theta}_{\mathbf{U}} = [\mathrm{vec}\{\mathbf{U}_{\mathrm{S}}\}^{\mathrm{T}}, \mathrm{vec}\{\mathbf{U}_{\mathrm{T}}\}^{\mathrm{T}}]^{\mathrm{T}}$ of size $(Mr_{\mathrm{S}}(L) + Wr_{\mathrm{T}}(L)) \times 1$ contains the slot-independent parameters and $\boldsymbol{\theta}_{\boldsymbol{\Gamma}} = \mathrm{vec}\{[\boldsymbol{\Gamma}(1)\cdots\boldsymbol{\Gamma}(L)]\}$ of size $(Lr_{\mathrm{S}}(L)r_{\mathrm{T}}(L)) \times 1$ the slot-dependent terms. The model orders $r_{\mathrm{S}}(L)$ and $r_{\mathrm{T}}(L)$ are assumed known within this section. It can be shown that the ML estimator reduces asymptotically (for $NL \to \infty$) to the minimizer of the loss function [14]

$$\Psi(\boldsymbol{\theta}) = \sum_{\ell=1}^{L}\left\|\mathbf{H}_{\mathrm{u}}(\ell) - \mathbf{U}_{\mathrm{S}}\boldsymbol{\Gamma}(\ell)\mathbf{U}_{\mathrm{T}}^{\mathrm{H}}\right\|^2_{\mathbf{R}_{xx}\otimes\mathbf{Q}_{\mathrm{u}}^{-1}} \quad (16)$$

which is analytically more convenient to handle. The asymptotic equivalence between the ML estimator and the minimizer of the loss function (16) can be intuitively justified by recalling that the estimator $\mathbf{Q}_{\mathrm{u}}$ is unbiased; hence, for $NL \to \infty$, the estimation of the channel matrix reduces to the case of known $\mathbf{Q}$, and the minimization of $\Psi(\boldsymbol{\theta})$ yields the exact ML estimator. We further observe that as $\mathbf{H}_{\mathrm{u}}(\ell)$ is the LSE of the channel matrix, according to (16), the constrained MLE is equivalent to a parametric re-estimation computed from the LSEs $\{\mathbf{H}_{\mathrm{u}}(\ell)\}_{\ell=1}^{L}$.

The loss function (16) can be rewritten in the following form:

$$\Psi(\boldsymbol{\theta}) = \sum_{\ell=1}^{L}\left\|\tilde{\mathbf{H}}_{\mathrm{u}}(\ell) - \tilde{\mathbf{U}}_{\mathrm{S}}\boldsymbol{\Gamma}(\ell)\tilde{\mathbf{U}}_{\mathrm{T}}^{\mathrm{H}}\right\|^2 \quad (17)$$

where $\tilde{\mathbf{H}}_{\mathrm{u}}(\ell) = \mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}\mathbf{H}_{\mathrm{u}}(\ell)\mathbf{R}_{xx}^{\mathrm{H}/2}$, $\tilde{\mathbf{U}}_{\mathrm{S}} = \mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}\mathbf{U}_{\mathrm{S}}$, and $\tilde{\mathbf{U}}_{\mathrm{T}} = \mathbf{R}_{xx}^{1/2}\mathbf{U}_{\mathrm{T}}$. Notice that $\tilde{\mathbf{H}}_{\mathrm{u}}(\ell)$ is equivalent to a whitening performed on $\mathbf{H}_{\mathrm{u}}(\ell)$ to take into account the noise correlation and the correlation properties of the training sequence. This is obtained by the spatial ($\mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}$) and the temporal ($\mathbf{R}_{xx}^{\mathrm{H}/2}$) whitening factors, as the covariance of $\tilde{\mathbf{H}}_{\mathrm{u}}(\ell)$ can be derived from the covariance of the LSE:

$$\begin{aligned}\mathrm{Cov}\left\{\mathrm{vec}\left\{\tilde{\mathbf{H}}_{\mathrm{u}}(\ell)\right\}\right\} &= \mathrm{Cov}\left\{\left(\mathbf{R}_{xx}^{*/2}\otimes\mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}\right)\mathrm{vec}\{\mathbf{H}_{\mathrm{u}}(\ell)\}\right\}\\ &= \mathbf{I}_W\otimes\left(\mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}\mathbf{Q}\mathbf{Q}_{\mathrm{u}}^{-1/2}\right)\end{aligned}$$

which tends to $\mathbf{I}_{MW}$ for $NL \to \infty$.

The optimization of $\Psi(\boldsymbol{\theta})$ in (17) can be carried out at first with respect to the slot dependent term $\boldsymbol{\theta}_{\boldsymbol{\Gamma}}$. The result is the slot-by-slot estimate

$$\hat{\boldsymbol{\Gamma}}(\ell) = \arg\min_{\boldsymbol{\Gamma}(\ell)}\Psi(\boldsymbol{\theta}) = \tilde{\mathbf{U}}_{\mathrm{S}}^{\dagger}\tilde{\mathbf{H}}_{\mathrm{u}}(\ell)\tilde{\mathbf{U}}_{\mathrm{T}}^{\dagger\mathrm{H}}. \quad (18)$$

The substitution of (18) into (17) yields

$$\Psi\left(\boldsymbol{\theta}_{\mathbf{U}}, \hat{\boldsymbol{\theta}}_{\boldsymbol{\Gamma}}\right) = \sum_{\ell=1}^{L}\left\|\tilde{\mathbf{H}}_{\mathrm{u}}(\ell) - \boldsymbol{\Pi}_{\mathrm{S}}\tilde{\mathbf{H}}_{\mathrm{u}}(\ell)\boldsymbol{\Pi}_{\mathrm{T}}\right\|^2 \quad (19)$$

where $\boldsymbol{\Pi}_{\mathrm{S}} = \tilde{\mathbf{U}}_{\mathrm{S}}\tilde{\mathbf{U}}_{\mathrm{S}}^{\dagger}$ is the projection matrix onto the subspace spanned by the columns of $\tilde{\mathbf{U}}_{\mathrm{S}}$ and, dually, $\boldsymbol{\Pi}_{\mathrm{T}} = \tilde{\mathbf{U}}_{\mathrm{T}}\tilde{\mathbf{U}}_{\mathrm{T}}^{\dagger}$ is the projector onto the subspace spanned by the columns of $\tilde{\mathbf{U}}_{\mathrm{T}}$. The estimation of $\tilde{\mathbf{U}}_{\mathrm{S}}$ and $\tilde{\mathbf{U}}_{\mathrm{T}}$ is turned into the minimization of (19) with respect to $\boldsymbol{\Pi}_{\mathrm{S}}$ and $\boldsymbol{\Pi}_{\mathrm{T}}$, constrained to be projection matrices of rank order $r_{\mathrm{S}}(L)$ and $r_{\mathrm{T}}(L)$, respectively. By making use of the trace operator properties [15], the minimization of (19) can be equivalently stated as

$$\left\{\hat{\boldsymbol{\Pi}}_{\mathrm{S}}, \hat{\boldsymbol{\Pi}}_{\mathrm{T}}\right\} = \arg\max_{\{\boldsymbol{\Pi}_{\mathrm{S}}, \boldsymbol{\Pi}_{\mathrm{T}}\}}\mathrm{tr}\left\{\boldsymbol{\Pi}_{\mathrm{S}}\sum_{\ell=1}^{L}\tilde{\mathbf{H}}_{\mathrm{u}}(\ell)\boldsymbol{\Pi}_{\mathrm{T}}\tilde{\mathbf{H}}_{\mathrm{u}}^{\mathrm{H}}(\ell)\right\} \quad (20)$$

$$= \arg \max_{\{\mathbf{\Pi}_\mathrm{S}, \mathbf{\Pi}_\mathrm{T}\}} \mathrm{tr}\left\{ \mathbf{\Pi}_\mathrm{T} \sum_{\ell=1}^{L} \tilde{\mathbf{H}}_\mathrm{u}^\mathrm{H}(\ell) \mathbf{\Pi}_\mathrm{S} \tilde{\mathbf{H}}_\mathrm{u}(\ell) \right\}. \tag{21}$$

The optimization (20) or (21) is nonlinear. The separable characteristics of the objective function suggests that the minimization can be carried out iteratively by alternating the search for $\mathbf{\Pi}_\mathrm{S}$ (given $\mathbf{\Pi}_\mathrm{T}$) and $\mathbf{\Pi}_\mathrm{T}$ (given $\mathbf{\Pi}_\mathrm{S}$). For $L = 1$ (or equivalently for a static channel), the minimization (20) or (21) yields the reduced-rank estimate [6], [7], whereas for $L \to \infty$, the closed-form solution can be based on a privileged choice of the initialization for the alternate search. Let us consider the optimization (20) with respect to $\mathbf{\Pi}_\mathrm{S}$ for any given initialization $\mathbf{\Pi}_\mathrm{T} = \mathbf{\Pi}_\mathrm{T}^{(0)}$ [the same reasoning applies dually for the optimization of (21)]. Any choice $\mathbf{\Pi}_\mathrm{T}^{(0)}$ such that $\mathcal{R}\{\mathbf{\Pi}_\mathrm{T}^{(0)} \tilde{\mathbf{H}}_\mathrm{u}^\mathrm{H}(\ell)\} \subset \mathcal{R}\{\tilde{\mathbf{H}}_\mathrm{u}^\mathrm{H}(\ell)\}$ affects the estimate of $\mathbf{\Pi}_\mathrm{S}$ since it reduces the set of the solutions to those that are compatible to the initial choice $\mathbf{\Pi}_\mathrm{T}^{(0)}$. In order not to bias the final solution, we prefer to decouple the optimizations by relaxing the constraint on the temporal structure and choosing $\mathbf{\Pi}_\mathrm{T}^{(0)} = \mathbf{I}_W$. With this choice, the estimate $\hat{\mathbf{\Pi}}_\mathrm{S}$ can be obtained very easily from the leading $r_\mathrm{S}(L)$ eigenvectors of the spatial correlation matrix

$$\tilde{\mathbf{R}}_\mathrm{S}(L) = \frac{1}{L} \sum_{\ell=1}^{L} \tilde{\mathbf{H}}_\mathrm{u}(\ell) \tilde{\mathbf{H}}_\mathrm{u}^\mathrm{H}(\ell). \tag{22}$$

For a similar reason, the solution $\hat{\mathbf{\Pi}}_\mathrm{T}$ is the projector onto the subspace spanned by the $r_\mathrm{T}(L)$ leading eigenvectors of the temporal correlation matrix

$$\tilde{\mathbf{R}}_\mathrm{T}(L) = \frac{1}{L} \sum_{\ell=1}^{L} \tilde{\mathbf{H}}_\mathrm{u}^\mathrm{H}(\ell) \tilde{\mathbf{H}}_\mathrm{u}(\ell). \tag{23}$$

The procedure could be iterated by alternating the search once initialized from the eigenvectors of $\tilde{\mathbf{R}}_\mathrm{S}(L)$ and $\tilde{\mathbf{R}}_\mathrm{T}(L)$, but in practice, there is no relevant improvement to justify the additional costs [6].

By substituting the projectors $\hat{\mathbf{\Pi}}_\mathrm{S}$ and $\hat{\mathbf{\Pi}}_\mathrm{T}$ estimated from $\tilde{\mathbf{R}}_\mathrm{S}(L)$ and $\tilde{\mathbf{R}}_\mathrm{T}(L)$, the *MS-ST estimator* of the channel matrix reduces to

$$\hat{\mathbf{H}}(\ell) = \mathbf{Q}_\mathrm{u}^{\mathrm{H}/2} \hat{\mathbf{\Pi}}_\mathrm{S} \tilde{\mathbf{H}}_\mathrm{u}(\ell) \hat{\mathbf{\Pi}}_\mathrm{T} \mathbf{R}_{xx}^{-\mathrm{H}/2}, \qquad \text{for } \ell = 1, 2, \ldots, L. \tag{24}$$

In summary, the MS method consists of the projection of the whitened LSE onto the spatial and temporal subspaces, which are obtained as the span of the first $r_\mathrm{S}(L)$ (and $r_\mathrm{T}(L)$) eigenvectors of the spatial (and temporal) correlation matrix (22) [and (23)]. We observe that for $L \to \infty$, the estimates $\hat{\mathbf{\Pi}}_\mathrm{S}$ and $\hat{\mathbf{\Pi}}_\mathrm{T}$ are the projection matrices onto the true subspaces of the whitened channel since $\mathcal{R}\{\tilde{\mathbf{R}}_\mathrm{S}(L)\} \to \mathcal{R}\{\mathbf{Q}^{-\mathrm{H}/2}\mathbf{A}\}$ and $\mathcal{R}\{\tilde{\mathbf{R}}_\mathrm{T}(L)\} \to \mathcal{R}\{\mathbf{R}_{xx}^{1/2}\mathbf{G}\}$, respectively. Appendix A proves analytically the optimality of solution (24) for $L \to \infty$.

B. MS-S and MS-T

In a dense multipath radio environment, the degree of temporal diversity could be as large as the support of the channel. In this case, the temporal order rises to $r_\mathrm{T}(L) \simeq W$, and $\mathcal{R}\{\mathbf{G}\}$ approaches $\mathbb{R}^W$. Under these conditions, it might be convenient to neglect the temporal projection and let $\mathbf{\Pi}_\mathrm{T} = \mathbf{I}_W$ in the optimization (20). The resulting channel estimate that exploits the stationarity of the spatial subspace only is referred to as *MS-S estimator:*

$$\hat{\mathbf{H}}(\ell) = \mathbf{Q}_\mathrm{u}^{\mathrm{H}/2} \hat{\mathbf{\Pi}}_\mathrm{S} \tilde{\mathbf{H}}_\mathrm{u}(\ell) \mathbf{R}_{xx}^{-\mathrm{H}/2}. \tag{25}$$

The projector $\hat{\mathbf{\Pi}}_\mathrm{S}$ is defined by the span of the first $r_\mathrm{S}(L)$ eigenvectors of the spatial correlation matrix (22). Following the same demonstration described in [5], it can be shown that the MS-S solution coincides with the MLE of the channel matrix for the model $\mathbf{H}(\ell) = \mathbf{U}_\mathrm{S}\mathbf{F}^\mathrm{H}(\ell)$, where $\mathbf{F}(\ell) = \mathbf{\Gamma}(\ell)\mathbf{U}_\mathrm{T}^\mathrm{H}$ is a $W \times r_\mathrm{S}(L)$ full-rank matrix. This algorithm was proposed (without proof of optimality) in [11].

Dually, for a large angle spread and/or a small number of antennas M ($r_\mathrm{S}(L) \simeq M$), it could be advisable not to use the spatial projection and choose $\mathbf{\Pi}_\mathrm{S} = \mathbf{I}_M$. The *MS-T estimator* exploits the stationarity of the temporal subspace:

$$\hat{\mathbf{H}}(\ell) = \mathbf{Q}_\mathrm{u}^{\mathrm{H}/2} \tilde{\mathbf{H}}_\mathrm{u}(\ell) \hat{\mathbf{\Pi}}_\mathrm{T} \mathbf{R}_{xx}^{-\mathrm{H}/2} \tag{26}$$

where $\hat{\mathbf{\Pi}}_\mathrm{T}$ is defined as usual. With the same considerations made above for the dual problem, it can be proved that the MS-T solution coincides with the MLE of the channel matrix given the model $\mathbf{H}(\ell) = \mathbf{C}(\ell)\mathbf{U}_\mathrm{T}^\mathrm{H}$, where $\mathbf{C}(\ell) = \mathbf{U}_\mathrm{S}\mathbf{\Gamma}(\ell)$ is a $M \times r_\mathrm{T}(L)$ full-rank matrix.

C. Remark on Model Order Selection

The model order $\{r_\mathrm{S}(L), r_\mathrm{T}(L)\}$ has to be estimated from data. The estimate $\{\hat{r}_\mathrm{S}(L), \hat{r}_\mathrm{T}(L)\}$ should be obtained as a tradeoff between distortion due to under parameterization ($\hat{r}_\mathrm{S}(L) < r_\mathrm{S}(L)$ and/or $\hat{r}_\mathrm{T}(L) < r_\mathrm{T}(L)$) and noise due to overparameterization ($\hat{r}_\mathrm{S}(L) > r_\mathrm{S}(L)$ and/or $\hat{r}_\mathrm{T}(L) > r_\mathrm{T}(L)$). A selection criterion could be the minimization of the MSE for the channel estimate with respect to $\{\hat{r}_\mathrm{S}(L), \hat{r}_\mathrm{T}(L)\}$ [7]. As proved by simulation later in this paper (Section VII), for low SNRs, the minimization of the MSE leads to a biased channel estimate, that is, $\hat{r}_\mathrm{S}(L) < r_\mathrm{S}(L)$ and/or $\hat{r}_\mathrm{T}(L) < r_\mathrm{T}(L)$. This property remains true also for $L \to \infty$.

D. Example

The underlying idea of the MS approach is to collect instantaneous estimates of the space-time channel as the user moves and then use these observations to estimate the slowly varying spatial-temporal basis by averaging over the fast-faded amplitudes. Let the LSEs be $\mathbf{H}_\mathrm{u}(\ell)$ for $\ell = 1, \ldots, L$. Then, the space-time observations can be used to form the sample correlation matrices $\hat{\mathbf{R}}_\mathrm{S}(L) = (1/L)\sum_{\ell=1}^{L} \mathbf{H}_\mathrm{u}(\ell)\mathbf{H}_\mathrm{u}^\mathrm{H}(\ell)$ and $\hat{\mathbf{R}}_\mathrm{T}(L) = (1/L)\sum_{\ell=1}^{L} \mathbf{H}_\mathrm{u}^\mathrm{H}(\ell)\mathbf{H}_\mathrm{u}(\ell)$ (here, we assume $\mathbf{Q} = \sigma_n^2\mathbf{I}_M$ and $\mathbf{R}_{xx} = \sigma_x^2\mathbf{I}_W$ so that $\hat{\mathbf{R}}_\mathrm{S}(L) \propto \tilde{\mathbf{R}}_\mathrm{S}(L)$, and $\hat{\mathbf{R}}_\mathrm{T}(L) \propto \tilde{\mathbf{R}}_\mathrm{T}(L)$). Next, the invariant spatial/temporal subspaces can be evaluated from the analysis of the eigenstructure of $\hat{\mathbf{R}}_\mathrm{S}(L)$ and $\hat{\mathbf{R}}_\mathrm{T}(L)$. This is illustrated by an example in Fig. 3. The multipath channel is composed of $P = 5$ paths having $\mathrm{E}[\boldsymbol{\beta}(\ell)\boldsymbol{\beta}^\mathrm{H}(\ell)] = \mathrm{diag}\{0.33, 0.25, 0.19, 0.14, 0.083\}$.

The path pattern is described in Fig. 3(a), whereas Fig. 3(b) shows the power-delay-angle diagram for a noisy estimate $\mathbf{H}_\mathrm{u}(\ell)$. This diagram can be obtained as $\mathcal{P}(i, \vartheta) = |[\mathbf{a}^\mathrm{H}(\vartheta)\mathbf{H}_\mathrm{u}(\ell)]_{1,i}|^2$, where $i = 1, \ldots, W$ and $\vartheta \in [-\pi/3, \pi/3]$ span the delay and angle axes, respectively. Since $\alpha_1 = \alpha_2$, $\alpha_3 = \alpha_4$ and $\tau_4 = \tau_5$, the spatial and temporal

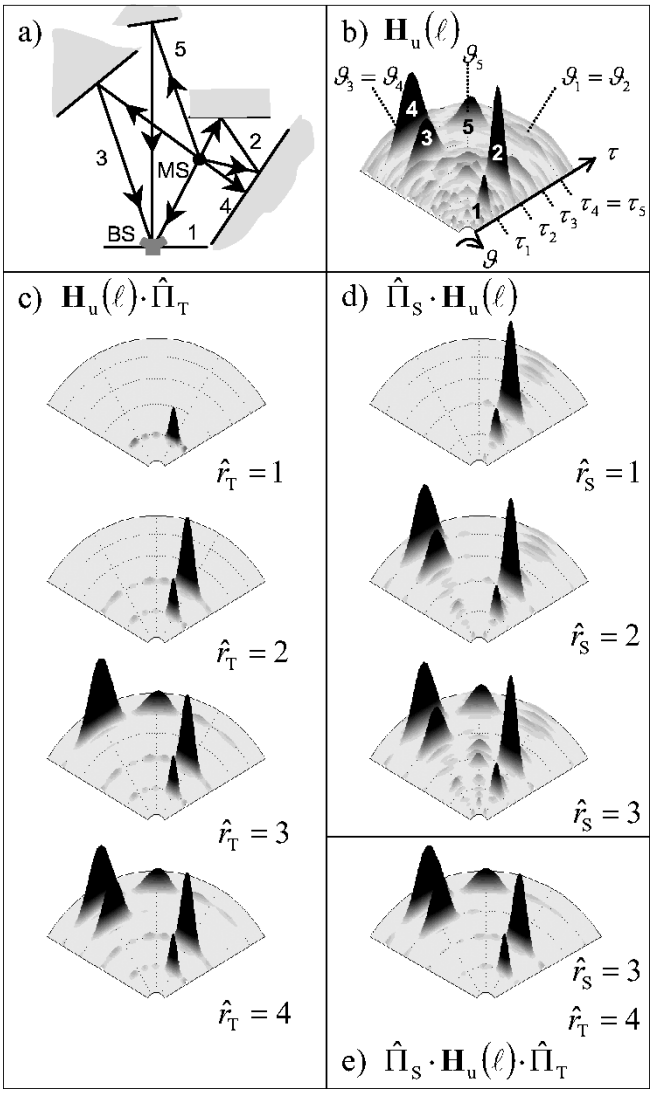


Fig. 3. Example of space and/or time projections for a channel with $P = 5$ paths and different degrees of space and time diversity: $r_{\mathrm{S}} = 3$ and $r_{\mathrm{T}} = 4$. (a) Multipath model. (b) Power-delay-angle diagram $\mathcal{P}(i, \vartheta)$ from the LSE. (c), (d) Projection of $\mathbf{H}_{\mathrm{u}}(\ell)$ onto temporal (c) or spatial (d) subspaces with increasing dimensions. (e) Projection onto spatial and temporal subspaces.

diversity orders are, respectively, $r_{\mathrm{S}} = 3$ and $r_{\mathrm{T}} = 4$. The projector $\hat{\mathbf{\Pi}}_{\mathrm{S}}$ onto the invariant spatial subspace is calculated from $\hat{\mathbf{R}}_{\mathrm{S}}(L)$ by using the diversity order $\hat{r}_{\mathrm{S}} = 1 \div 3$. The same holds for $\hat{\mathbf{\Pi}}_{\mathrm{T}}$ with $\hat{r}_{\mathrm{T}} = 1 \div 4$. As shown in Fig. 3(c)–(e), by projecting the space–time matrix $\mathbf{H}_{\mathrm{u}}(\ell)$, the background noise is reduced. Notice that in Fig. 3(c), the temporal projector $\hat{\mathbf{\Pi}}_{\mathrm{T}}$ with $\hat{r}_{\mathrm{T}} = 1$ selects the path of the channel that has the largest *mean* power σ_1^2, even though this is not the temporal component with the largest *local* power.

The residual noise after the projection is that component that can no longer be eliminated as it belongs to the same subspace of the channel. By comparing the power-delay-angle diagram for the initial space-time matrix $\mathbf{H}_{\mathrm{u}}(\ell)$ in Fig. 3(a) with the projected matrix $\hat{\mathbf{\Pi}}_{\mathrm{S}}\mathbf{H}_{\mathrm{u}}(\ell)\hat{\mathbf{\Pi}}_{\mathrm{T}}$ in Fig. 3(e), the visual inspection shows that the artifacts due to noise are reduced. This noise reduction depends on the ratio $r_{\mathrm{S}}r_{\mathrm{T}}/MW$. Indeed, it can be observed in Fig. 3(e) that the double projection selects $r_{\mathrm{S}}r_{\mathrm{T}} = 12$ "intersections" in the space–time domain, i.e., the multipath components that have angles in $\{\alpha_1, \alpha_3, \alpha_5\}$ and delays in $\{\tau_1, \tau_2, \tau_3, \tau_4\}$. Since the initial LSE in Fig. 3(a) contains all the MW channel samples, the noise-reduction after the double projection is $r_{\mathrm{S}}r_{\mathrm{T}}/MW$.

V. Performance Analysis

A. Asymptotic Performance ($L \to \infty$)

A lower bound on the MSE $= \mathrm{E}[\|\hat{\mathbf{H}}(\ell) - \mathbf{H}(\ell)\|^2]$ can be derived through the computation of the Cramér–Rao Bound (CRB) on the covariance for the channel estimate and then by taking the expectation with respect to the fading $\boldsymbol{\beta}(\ell)$. The CRB is largely simplified for $L \to \infty$ as the covariances of the temporal and spatial projectors are $O(1/L)$, and thus, the MSE is dominated by the slot-dependent terms. This is shown in Appendix B, where the CRB of the MS-ST estimator is derived for $L \to \infty$. The lower bound is MSE $\geq$ MSE$_{\text{MS-ST}}$ with

$$\mathrm{MSE}_{\text{MS-ST}} = \Phi\left(\mathbf{\Pi}_{\mathrm{S}}, \mathbf{Q}\right) \cdot \Phi\left(\mathbf{\Pi}_{\mathrm{T}}, \mathbf{R}_{xx}^{-1}\right) \tag{27}$$

where the operator $\Phi(\mathbf{\Pi}, \mathbf{T}) = \mathrm{tr}\{\mathbf{T}^{\mathrm{H}/2}\mathbf{\Pi}\mathbf{T}^{1/2}\}/\sqrt{N}$ has some useful properties that are listed in Appendix C. Since, from (4), the MSE for the LSE is known to be

$$\mathrm{MSE}_{\mathrm{u}} = \Phi(\mathbf{I}_W, \mathbf{R}_{xx}^{-1})\Phi(\mathbf{I}_M, \mathbf{Q}) \tag{28}$$

the bound (27) shows that the MS-ST method reduces the estimation error through the projection onto the spatial ($r_{\mathrm{S}} \leq M$) and temporal ($r_{\mathrm{T}} \leq W$) subspaces.

Recalling the derivation of the MS-S and the MS-T methods in Section IV-B, the MSE lower bound for the MS-S estimator can be inferred from (27) for $\mathbf{\Pi}_{\mathrm{T}} = \mathbf{I}_W$:

$$\mathrm{MSE}_{\text{MS-S}} = \Phi(\mathbf{\Pi}_{\mathrm{S}}, \mathbf{Q}) \cdot \Phi\left(\mathbf{I}_W, \mathbf{R}_{xx}^{-1}\right). \tag{29}$$

Dually for the MS-T method, we have $\mathbf{\Pi}_{\mathrm{S}} = \mathbf{I}_M$

$$\mathrm{MSE}_{\text{MS-T}} = \Phi(\mathbf{I}_M, \mathbf{Q}) \cdot \Phi\left(\mathbf{\Pi}_{\mathrm{T}}, \mathbf{R}_{xx}^{-1}\right). \tag{30}$$

The MSE bounds (27)–(30) are largely simplified for spatially uncorrelated noise ($\mathbf{Q} = \sigma_n^2\mathbf{I}_M$) and training sequence with ideal correlation properties ($\mathbf{R}_{xx} = \sigma_x^2\mathbf{I}_W$). Under these conditions, the bounds are summarized as MSE$_{\text{AWGN}}$ in Table I and have a simple interpretation: Asymptotically (or for a large L), the MSE depends on the number of parameters to be estimated on a slot-by-slot basis. For MS-ST, it depends on the $r_{\mathrm{S}}r_{\mathrm{T}}$ entries of the matrix $\mathbf{\Gamma}(\ell)$.

As a final remark, we note that a lower MSE bound can be obtained by parameterizing the channel matrix as a function of the joint spatial and temporal subspace (see [1] for details).

B. Relationship With Other Techniques

For single-slot processing ($L = 1$), all the MS methods proposed here coincide with the reduced-rank (RR) approach for the channel rank $r = \mathrm{rank}\{\mathbf{H}(1)\} \leq \min\{W, M\}$ [5], [7]. The RR method uses the same degree $r = \min\{r_{\mathrm{T}}, r_{\mathrm{S}}\}$ for the spatial and the temporal diversity as the rank of the channel matrix is dominated by the more restrictive of the two. On the other hand, a MS processing with uncorrelated fading gives the needed redundancy to allow the estimation of both the spatial and temporal subspaces by relaxing the constraint of the RR approach and differentiating between space and time rank orders.

TABLE I
MSE BOUNDS FOR $N \to \infty$ AND $L \to \infty$ ($\mathbf{\Pi}_{\mathrm{S}}(\ell) = \bar{\mathbf{H}}(\ell)\bar{\mathbf{H}}^{\dagger}(\ell)$, $\mathbf{\Pi}_{\mathrm{T}}(\ell) = \bar{\mathbf{H}}^{\dagger}(\ell)\bar{\mathbf{H}}(\ell)$)

Method	MSE	$\mathrm{MSE}_{\mathrm{AWGN}}$
LS	$\Phi(\mathbf{I}_W, \mathbf{R}_{xx}^{-1}) \cdot \Phi(\mathbf{I}_M, \mathbf{Q})$	$\rho \cdot MW$
RR	$\Phi(\mathbf{\Pi}_{\mathrm{S}}(\ell), \mathbf{Q}) \cdot \Phi(\mathbf{I}_W, \mathbf{R}_{xx}^{-1}) + \Phi(\mathbf{\Pi}_{\mathrm{S}}^{\perp}(\ell), \mathbf{Q}) \cdot \Phi(\mathbf{\Pi}_{\mathrm{T}}(\ell), \mathbf{R}_{xx}^{-1})$	$\rho \cdot [r(M+W) - r^2]$
MS-ST	$\Phi(\mathbf{\Pi}_{\mathrm{S}}, \mathbf{Q}) \cdot \Phi(\mathbf{\Pi}_{\mathrm{T}}, \mathbf{R}_{xx}^{-1})$	$\rho \cdot r_{\mathrm{S}} r_{\mathrm{T}}$
MS-S	$\Phi(\mathbf{I}_W, \mathbf{R}_{xx}^{-1}) \cdot \Phi(\mathbf{\Pi}_{\mathrm{S}}, \mathbf{Q})$	$\rho \cdot W r_{\mathrm{S}}$
MS-T	$\Phi(\mathbf{I}_M, \mathbf{Q}) \cdot \Phi(\mathbf{\Pi}_{\mathrm{T}}, \mathbf{R}_{xx}^{-1})$	$\rho \cdot M r_{\mathrm{T}}$

The MS channel estimators can thus be considered as the natural generalization of the RR algorithm to a MS framework.

To facilitate the comparison between the single-slot (LS, RR) and MS (MS-ST, MS-S, MS-T) methods in terms of performance, Table I summarizes the MSE bounds under asymptotic conditions: $N \to \infty$ and $L = 1$ for the RR method [7]; $\{N, L\} \to \infty$ for the MS techniques. Notice that $\mathbf{\Pi}^{\perp} = \mathbf{I} - \mathbf{\Pi}$ is the orthogonal projector onto the complement of $\mathcal{R}\{\mathbf{\Pi}\}$. The performance comparison between the MS and the RR method is carried out by simulation results in [16].

The following relation holds among the LS and the MS estimate performances (27)–(30):

$$\mathrm{MSE}_{\mathrm{u}} \geq \{\mathrm{MSE}_{\text{MS-T}}, \mathrm{MSE}_{\text{MS-S}}\} \geq \mathrm{MSE}_{\text{MS-ST}}. \tag{31}$$

See Appendix C for the proof.

For comparison with the single-slot RR method, we make use of the MSE lower bound given in Table I [7]. The error $\mathrm{MSE}_{\mathrm{RR}}$ depends on the fading amplitudes within the ℓth slot and, in particular, on the rank-r projectors $\mathbf{\Pi}_{\mathrm{S}}(\ell) = \tilde{\mathbf{H}}(\ell)\tilde{\mathbf{H}}^{\dagger}(\ell)$ and $\mathbf{\Pi}_{\mathrm{T}}(\ell) = \hat{\mathbf{H}}^{\dagger}(\ell)\hat{\mathbf{H}}(\ell)$, with $\hat{\mathbf{H}}(\ell) = \mathbf{Q}^{-\mathrm{H}/2}\mathbf{H}(\ell)\mathbf{R}_{xx}^{\mathrm{H}/2}$. In order to compare the performance of the RR estimate with those of the multislot methods for $L \to \infty$, here, the instantaneous fading $\mathrm{MSE}_{\mathrm{RR}}(\ell)$ needs to be averaged with respect to the fading amplitudes (or, equivalently, averaged over $L \to \infty$ slots): $\mathrm{MSE}_{\mathrm{RR}} = \mathrm{E}[\mathrm{MSE}_{\mathrm{RR}}(\ell)]$. The following inequalities hold:

$$\mathrm{MSE}_{\mathrm{u}} \geq \mathrm{MSE}_{\mathrm{RR}} \geq \begin{cases} \mathrm{MSE}_{\text{MS-T}}, & \text{for } r = r_{\mathrm{T}} \leq r_{\mathrm{S}} \\ \mathrm{MSE}_{\text{MS-S}}, & \text{for } r = r_{\mathrm{S}} \leq r_{\mathrm{T}} \end{cases} \tag{32}$$

which also imply $\mathrm{MSE}_{\mathrm{RR}} \geq \mathrm{MSE}_{\text{MS-ST}}$ for any r. The proofs are given in Appendix C.

The inequalities (31) and (32) can be easily justified in the case of $\mathbf{Q} = \sigma_n^2\mathbf{I}_M$ and $\mathbf{R}_{xx} = \sigma_x^2\mathbf{I}_W$ (column $\mathrm{MSE}_{\mathrm{AWGN}}$ in Table I). For all the methods, the bound $\mathrm{MSE}_{\mathrm{AWGN}}$ is proportional to $\rho = \sigma_n^2/N\sigma_x^2$ and to the number of independent unknowns. For instance, the RR approach reduces the number of unknowns with respect to LSE by a factor of $MW/(r(M+W)-r^2)$, which yields a MSE gain $\mathrm{MSE}_{\mathrm{u}}/\mathrm{MSE}_{\mathrm{RR}} \geq 1$ for any $r \leq \min\{W, M\}$. Furthermore, the gain of MS-ST with respect to LSE depends on the spatial (M/r_{S}) and temporal (W/r_{T}) gains as $\mathrm{MSE}_{\mathrm{u}}/\mathrm{MSE}_{\text{MS-ST}} = W/r_{\mathrm{T}} \cdot M/r_{\mathrm{S}}$. We conclude that the MS techniques show a definite advantage with respect to any single-slot method due to their capability to estimate the spatial and/or temporal projectors with any degree of accuracy. Indeed, for $L \to \infty$, the MSE of the MS methods depends only on the number of parameters to be estimated on each slot: $r_{\mathrm{S}}r_{\mathrm{T}}$ for MS-ST, $r_{\mathrm{S}}W$ for MS-S, and Mr_{T} for MS-T.

TABLE II
SUBSPACE TRACKING ALGORITHM

Initialize: $\mathbf{U}(0) = \begin{bmatrix} \mathbf{I}_r \\ \mathbf{0} \end{bmatrix}$; $\mathbf{\Theta}(0) = \mathbf{I}_r$; $\mathbf{A}(0) = \mathbf{0}$; $0 \leq \gamma \leq 1$; r

For each slot ℓ:

input: $\mathbf{B}(\ell)$

$\mathbf{Z}(\ell) = \mathbf{U}(\ell-1)^H\mathbf{B}(\ell)$

$\mathbf{A}(\ell) = \gamma\mathbf{A}(\ell-1)\mathbf{\Theta}(\ell-1) + \mathbf{B}(\ell)\,\mathbf{Z}(\ell)^H$

$\mathbf{A}(\ell) = \mathbf{U}(\ell)\mathbf{R}(\ell)$ (QR factorization)

$\mathbf{\Theta}(\ell) = \mathbf{U}(\ell-1)^H\,\mathbf{U}(\ell)$

It is worth noticing that when the characteristics of the radio environment are such that $r_{\mathrm{T}} \simeq W$ (or $r_{\mathrm{S}} \simeq M$), the additional computational load required to evaluate the temporal (or spatial) subspace in the MS-ST method provides no performance advantages. In this case, the MS-S (or the MS-T) estimator has to be preferred with respect to the MS-ST method.

VI. SUBSPACE TRACKING IMPLEMENTATION

The implementation of the MS techniques MS-ST, MS-S, and MS-T implies a certain latency in providing the channel estimate (approximately $L/2$ slots) and might turn out to be too demanding in terms of computational complexity. Moreover, for all the estimators, we have adopted the assumption that angles and delays have to be recomputed every L slots. An alternative implementation that allows the cancellation of the latency and alleviation of the computational burden of the eigenvalue decomposition (EVD) consists of updating the spatial and temporal subspace on a slot-by-slot basis through a subspace tracking technique [17], [18]. This way, angles and delays are allowed to vary continuously (but still slowly). It is advisable to select the most convenient algorithm in terms of performance and complexity. For the MS estimation, we have adopted the fast subspace tracker proposed in [19] and summarized in Table II with some minor modifications. The variable ℓ runs over the slots. In order to obtain the MS-ST estimate, the tracking algorithm needs to be performed both on the spatial and the temporal subspace for each slot ℓ, and the input of the algorithm is given by the channel matrix $\mathbf{B}(\ell) = \tilde{\mathbf{H}}_{\mathrm{u}}(\ell)$ for the spatial subspace tracking and by $\mathbf{B}(\ell) = \hat{\mathbf{H}}_{\mathrm{u}}^{\mathrm{H}}(\ell)$ for the temporal one. $\mathbf{U}$ is the orthonormal basis of the (spatial or temporal) subspace, and γ rules the effective memory (in terms of slots) of the algorithm. In Table II, the dimension r of the (spatial or temporal) subspace is assumed to be fixed. Nonetheless, the tracking algorithm can be modified in order to adaptively estimate r. The order of complexity for each slot is the combination of the two

tracking loops $O(Mr_{\mathrm{S}}^2(L))+O(Wr_{\mathrm{T}}^2(L))$. Notice that the algorithm as presented above can be made even more efficient while still retaining the same order of complexity [19]. Numerical results show that the performance of the fast subspace tracking algorithm is very close to the EVD implementation. A thorough study of the transient behavior of the subspace tracking implementation of the MS technique is under way, and it will be presented in future works.

VII. Simulation Results

In this section, the performances are evaluated for uncorrelated fading amplitudes ($\varphi(m) = \delta(m)$) since this condition leads to the largest model order and, as discussed in Section V, to the worst asymptotic performances. We consider $P = 8$ paths with exponentially decreasing power $\sigma_p^2 = \sigma^2 \times 0.5^{(p-1)}$ (σ^2 is scaled to have $\mathrm{E}[\|\mathbf{H}(\ell)\|^2] = 1$) clustered into two angles with four delays on each angle. In the first set $\alpha_p = \pi/3$ for $p = 1, \ldots, 4$, we have the delays $[\tau_1 \cdots \tau_4] = [3.2, 5.1, 6.2, 6.8]T$, and in the second $\alpha_p = \pi/6$ for $p = 5, \ldots, 8$, we have $[\tau_5 \cdots \tau_8] = [9.8, 11.1, 11.9, 12.8]T$. The transmitted pulse $g(t)$ is a raised cosine with roll-off factor 0.2. The receiver is equipped with a uniform linear antenna array of $M = 8$ elements with half-wavelength interelement spacing, and the temporal support of the channel is $W = 15$ symbols. The training sequence, of length $N = 40$, is randomly generated such that $\mathrm{E}[\mathbf{X}^{\mathrm{H}}(\ell)\mathbf{X}(\ell)] = N\sigma_x^2\mathbf{I}_W$. Five interferers have directions of arrival $[\overline{\alpha}_1 \cdots \overline{\alpha}_5] = [-\pi/3, -\pi/6, 0, \pi/6, \pi/3]$ equally spaced within the angular support $[-60\ 60]^\circ$. They are modeled as Gaussian disturbance with correlation matrix $\mathbf{R}_{\mathrm{i}} = (\sigma_{\mathrm{i}}^2/5)\sum_{k=1}^{5}\mathbf{a}(\overline{\alpha}_k)\mathbf{a}^{\mathrm{H}}(\overline{\alpha}_k)$ and $\mathrm{tr}\{\mathbf{R}_{\mathrm{i}}\} = M\sigma_{\mathrm{i}}^2$. The spatial correlation matrix is $\mathbf{Q} = \mathbf{R}_{\mathrm{i}} + \mathbf{R}_{\mathrm{b}}$, where $\mathbf{R}_{\mathrm{b}} = \sigma_{\mathrm{b}}^2\mathbf{I}_M$ accounts for the background AWGN. The signal-to-interference ratio is defined as $\mathrm{SIR} = \sigma_x^2\mathrm{E}[\|\mathbf{H}(\ell)\|^2]/(M\sigma_{\mathrm{i}}^2) = \sigma_x^2/(M\sigma_{\mathrm{i}}^2)$ and the signal to (background) noise ratio as $\mathrm{SNR} = \sigma_x^2\mathrm{E}[\|\mathbf{H}(\ell)\|^2]/(M\sigma_{\mathrm{b}}^2) = \sigma_x^2/(M\sigma_{\mathrm{b}}^2)$.

In Fig. 4, the simulations for the MSE on the channel estimate of MS-ST, MS-S, and MS-T (markers) are compared with the MSE bounds (27), (29), and (30) (dashed lines), respectively. At the top, the angles of arrival of the user are aligned with those of the interferers: $\alpha_p = \overline{\alpha}_5 = \pi/3$ for $p = 1, \ldots, 4$ and $\alpha_p = \overline{\alpha}_4 = \pi/6$ for $p = 5, \ldots, 8$. At the bottom, the angles of arrival are slightly misaligned: $\alpha_p = \pi/4$ for $p = 1, \ldots, 4$ and $\alpha_p = \pi/8$ for $p = 5, \ldots, 8$ (see the box on both figures). The simulations are carried out for $L \to \infty$ so that $\hat{\mathbf{\Pi}}_{\mathrm{S}} = \mathbf{\Pi}_{\mathrm{S}}$ and/or $\hat{\mathbf{\Pi}}_{\mathrm{T}} = \mathbf{\Pi}_{\mathrm{T}}$; in addition, $\hat{r}_{\mathrm{S}} = r_{\mathrm{S}} = 2$ and $\hat{r}_{\mathrm{T}} = r_{\mathrm{T}} = 8$. When the user angles are separated from those of the interferers (bottom), the spatial processing performed by the MS-ST and MS-S methods leads to a value of MSE that is independent on the interferer level, but it is ruled by the background noise (SNR $= 50$ dB). According to the analysis in Section V, here, the MS-S method outperforms the MS-T because in this radio environment, the spatial gain ($M/r_{\mathrm{S}} = 4$) exceeds the temporal gain ($W/r_{\mathrm{T}} = 1.875$). To be precise, this argument applies only to the ideal case ($\mathrm{MSE}_{\mathrm{AWGN}}$ in Table I), but it turns out to be a reliable rule of thumb to choose the MS estimators according to their expected performance. In the following simulations, we consider the angles of arrival for the user aligned with those of the interferers (see the top of Fig. 4) as this corresponds to the worst case.

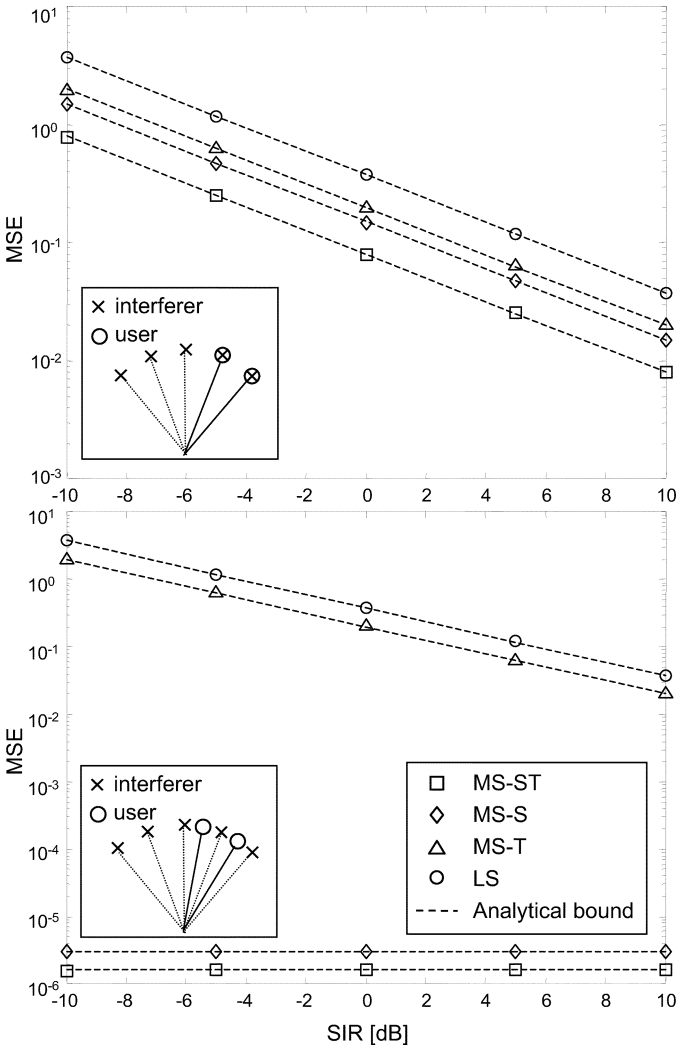


Fig. 4. MSE versus SIR for $L \to \infty$. Simulations (markers) and analytical bounds (27), (29), (30) (dashed lines). Angles of arrival of the user are aligned (top) or misaligned (bottom) with those of the interferers. ($M = 8$, $W = 15$, $N = 40$, SNR $= 50$ dB).

By now, we have validated the asymptotic ($L \to \infty$) performance of the MS estimators according to the theoretical bounds in Section V. In a practical situation, the number of the processed slot L is finite due to memory requirement and long-term angle-delay pattern variations. The former can be avoided by the subspace tracking method discussed in Section VI, and the latter makes the angle-delay invariance hold only within a finite number of slots, thus limiting the values of L (or, equivalently, the memory of the subspace tracking). In Fig. 5, numerical results for varying L show that for a number of slots $L \geq 20$, the performance of the MS estimators (solid lines) attains the asymptotic bounds (dashed line). As a simple rule of thumb, the convergence rate depends on the number of parameters to be estimated from the joint observation of the ensemble of slots, i.e., the number of entries of $\mathbf{U}_{\mathrm{S}}$ and/or $\mathbf{U}_{\mathrm{T}}$. This justifies the faster convergence of MS-S for this example. Moreover, recalling that the variance of any MLE approaches the CRB when the cardinality of the set of observations grows to infinity, we note that

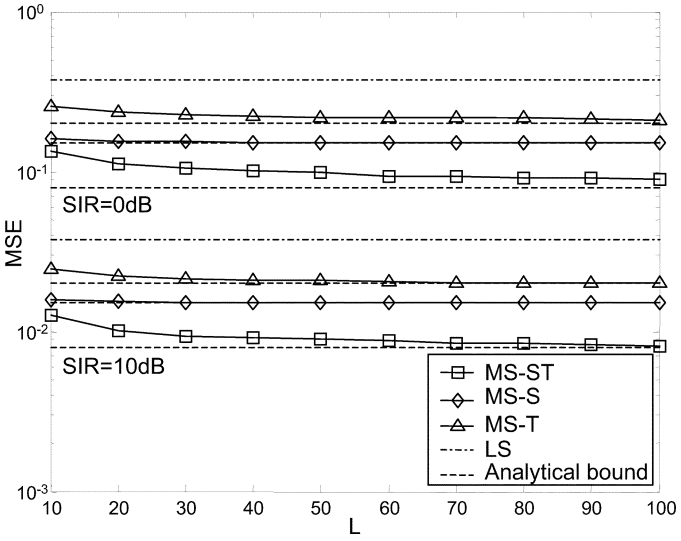


Fig. 5. Convergence properties of the MS techniques MS-ST, MS-S, and MS-T. For $L > 10$, the MSE attains the analytical bounds (27), (29), and (30) (dashed lines). As a reference, we show the MSE for LSE (dashed-dot lines) ($M = 8$, $W = 15$, $N = 40$, SNR $= 50$ dB).

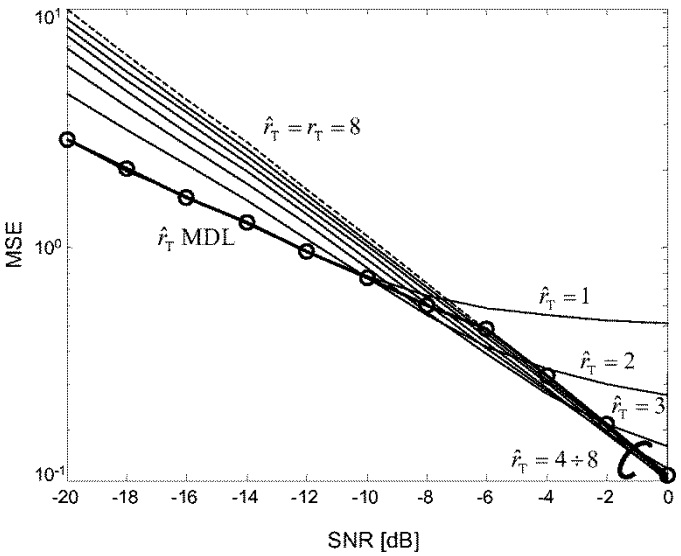


Fig. 6. MSE versus SIR for MS-ST and for varying values of $\hat{r}_T = 1, \ldots, r_T$. $\hat{r}_T = r_T = 8$ is the true model order (dashed lines); model order selection by the MDL criterion (markers). ($\hat{r}_S = r_S = 2$, $L = 10$, $M = 8$, $W = 15$, $N = 40$, SNR $= 50$ dB).

the convergence of the MSE of MS-ST to the bound computed from the CRB confirms numerically the optimality of MS-ST for large L, whereas MS-S and MS-T are close to the bound for any $L \geq 20$ (see Section IV).

All the simulations presented above were obtained for an unbiased estimate by choosing the degree of diversity $\hat{r}_S = r_S$ and $\hat{r}_T = r_T$. However, in Section IV-C, it was discussed that for low SNRs, it may be convenient to accept a biased channel estimate in order to reduce the MSE. This is shown in Fig. 6 for the MS-ST method with $L = 10$ (similar results can be shown for all the methods). For simplicity, the spatial model order is fixed to $\hat{r}_S = r_S = 2$, whereas $\hat{r}_T$ ranges from $\hat{r}_T = 1$ to $\hat{r}_T = r_T = 8$ (dashed line). The MSE versus SIR curves show

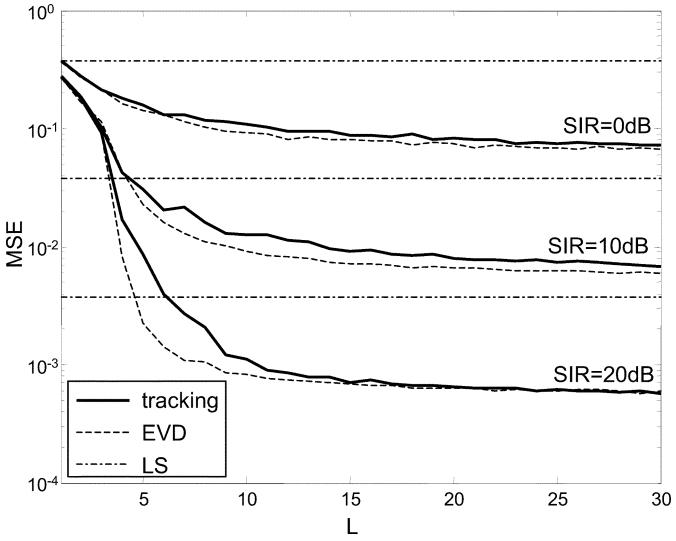


Fig. 7. MSE versus the number of slot L for the exact EVD-implementation (dashed line) and for the subspace tracking implementation (solid line) for varying SIRs ($\gamma = 1$). The MSE for the LSE (dash-dot line) is shown for reference.

that for SIR < -10 dB, the MSE is minimized by choosing $\hat{r}_T = 1$, whereas for larger SIR values, the MSE shows a floor due to the bias for underparameterization. Larger values of $\hat{r}_T$ have to be used for increasing SIR in order to select the degree of temporal diversity that minimizes the MSE. The minimum description length (MDL) criterion [20] has been used here to select $\hat{r}_T$ as it approximately provides the model order that minimizes the MSE [7].

The performances of the subspace tracker proposed in [19] are evaluated in Fig. 7 in terms of MSE versus the number of slots L. The exact EVD implementation (dashed lines) and the subspace tracking implementation (bold lines) are compared for varying SIRs (forgetting factor $\gamma = 1$). The MSE corresponding to the LSE is also shown as a reference (dash-dotted lines) and confirms the accuracy of the subspace tracking methods for real-time implementations.

VIII. Conclusion

The proposed MS estimation methods exploit the invariance over the slots of both the spatial and the temporal subspaces without estimating explicitly delays and angles of arrival. The paper has focused on the uplink of a single-user time-slotted system. Nonetheless, the approach (with straightforward modifications) appears to be a valid solution for a broader class of systems, such as TD-CDMA, FDMA, multicarrier, or generically MIMO. Notice that even when the receiving antennas are sufficiently spaced to cause the uncorrelation of the impinging wavefront, the MS-T method still retains its advantages. Analysis of the MSE shows that the asymptotic performance (i.e., for a large number of slots) depends on the number of parameters to be estimated on a slot-by-slot basis. Simulations proved the relevant advantage of the MS approach compared with single-slot techniques, as far as the quasistatic approximation of the space-time subspace holds true.

APPENDIX A
OPTIMALITY OF MS-ST

For $L \to \infty$, the channel estimate (24) selects $\hat{\mathbf{\Pi}}_{\mathrm{S}}$ as the projector onto the subspace spanned by the r_{S} first eigenvectors of $\hat{\mathbf{R}}_{ST} = \lim_{L\to\infty} \hat{\mathbf{R}}_{\mathrm{ST}}(L)$, where

$$\hat{\mathbf{R}}_{\mathrm{ST}}(L) = \frac{1}{L}\sum_{\ell=1}^{L} \hat{\mathbf{H}}_{\mathrm{u}}(\ell)\mathbf{\Pi}_{\mathrm{T}}\hat{\mathbf{H}}_{\mathrm{u}}^{\mathrm{H}}(\ell). \tag{33}$$

We prove here that $\hat{\mathbf{\Pi}}_{\mathrm{S}}$ can be equivalently obtained as the projector onto the subspace spanned by the r_{S} first eigenvectors of the correlation matrix $\tilde{\mathbf{R}}_{\mathrm{S}}$ defined in (22) for any choice of $\mathbf{\Pi}_{\mathrm{T}}$, provided that the diagonal matrix $(\tilde{\mathbf{G}}^{\mathrm{H}}\mathbf{\Pi}_{\mathrm{T}}\tilde{\mathbf{G}}) \odot \mathbf{I}_P = \mathrm{diag}\{[\tilde{\mathbf{G}}^{\mathrm{H}}\mathbf{\Pi}_{\mathrm{T}}\tilde{\mathbf{G}}]_{11}, \ldots, [\tilde{\mathbf{G}}^{\mathrm{H}}\mathbf{\Pi}_{\mathrm{T}}\tilde{\mathbf{G}}]_{PP}\}$, which was obtained from $\tilde{\mathbf{G}} = \mathbf{R}_{xx}^{1/2}\mathbf{G}$, is positive definite (or, equivalently, $[\tilde{\mathbf{G}}^{\mathrm{H}}\mathbf{\Pi}_{\mathrm{T}}\tilde{\mathbf{G}}]_{kk} \neq 0$ for $\forall k$). The proof below holds in the dual form for the estimation of $\mathbf{\Pi}_{\mathrm{T}}$ from (21).

Let us rewrite the prewhitened LSE as $\hat{\mathbf{H}}_{\mathrm{u}}(\ell) = \tilde{\mathbf{A}}\mathbf{D}(\ell)\tilde{\mathbf{G}}^{\mathrm{H}} + \Delta\hat{\mathbf{H}}_{\mathrm{u}}(\ell)$, with $\tilde{\mathbf{A}} = \mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}\mathbf{A}$. For $L \to \infty$, the entries of the error $\Delta\hat{\mathbf{H}}_{\mathrm{u}}(\ell) = \mathbf{Q}_{\mathrm{u}}^{-\mathrm{H}/2}\Delta\mathbf{H}_{\mathrm{u}}\mathbf{R}_{xx}^{\mathrm{H}/2}$ are independent normally distributed with zero mean and unitary variance ($\mathbf{Q}_{\mathrm{u}} \to \mathbf{Q}$ since $\mathbf{Q}_{\mathrm{u}} - \mathbf{Q} = O(1/L)$). Noise and faded amplitudes are independent, therefore, for the ergodicity:

$$\begin{aligned}
&\lim_{L\to\infty} \tilde{\mathbf{R}}_{\mathrm{S}}(L) \\
&\quad = \tilde{\mathbf{A}}\,\mathrm{E}\Big[\mathbf{D}(\ell)\tilde{\mathbf{G}}^{\mathrm{H}}\tilde{\mathbf{G}}\mathbf{D}^{\mathrm{H}}(\ell)\Big]\,\tilde{\mathbf{A}}^{\mathrm{H}} + \mathrm{E}\Big[\Delta\tilde{\mathbf{H}}_{\mathrm{u}}(\ell)\Delta\tilde{\mathbf{H}}_{\mathrm{u}}^{\mathrm{H}}(\ell)\Big] \\
&\quad = \tilde{\mathbf{A}}\tilde{\mathbf{\Lambda}}_{\mathrm{S}}\tilde{\mathbf{A}}^{\mathrm{H}} + W\mathbf{I}_M
\end{aligned} \tag{34}$$

$$\begin{aligned}
&\lim_{L\to\infty} \tilde{\mathbf{R}}_{\mathrm{ST}}(L) \\
&\quad = \tilde{\mathbf{A}}\,\mathrm{E}\Big[\mathbf{D}(\ell)\tilde{\mathbf{G}}^{\mathrm{H}}\mathbf{\Pi}_{\mathrm{T}}\tilde{\mathbf{G}}\mathbf{D}^{\mathrm{H}}(\ell)\Big]\,\tilde{\mathbf{A}}^{\mathrm{H}} \\
&\qquad + \mathrm{E}\Big[\Delta\tilde{\mathbf{H}}_{\mathrm{u}}(\ell)\mathbf{\Pi}_{\mathrm{T}}\Delta\tilde{\mathbf{H}}_{\mathrm{u}}^{\mathrm{H}}(\ell)\Big] \\
&\quad = \tilde{\mathbf{A}}\tilde{\mathbf{\Lambda}}_{\mathrm{ST}}\tilde{\mathbf{A}}^{\mathrm{H}} + r_{\mathrm{T}}\mathbf{I}_M
\end{aligned} \tag{35}$$

where, similarly to (12a), the diagonal matrices are $\tilde{\mathbf{\Lambda}}_{\mathrm{ST}} = (\tilde{\mathbf{G}}^{\mathrm{H}}\mathbf{\Pi}_{\mathrm{T}}\tilde{\mathbf{G}}) \odot \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_P^2\}$ and $\tilde{\mathbf{\Lambda}}_{\mathrm{S}} = (\tilde{\mathbf{G}}^{\mathrm{H}}\tilde{\mathbf{G}}) \odot \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_P^2\}$. The subspaces spanned by the r_{S} leading eigenvectors of the matrices (34) and (35) coincide with the column space of $\tilde{\mathbf{A}}$ if $[\tilde{\mathbf{G}}^{\mathrm{H}}\mathbf{\Pi}_{\mathrm{T}}\tilde{\mathbf{G}}]_{kk} \neq 0\ \forall k$. Since a sufficient condition is $\mathcal{R}\{\mathbf{\Pi}_{\mathrm{T}}\} \supseteq \mathcal{R}\{\tilde{\mathbf{G}}\}$, the choice $\mathbf{\Pi}_{\mathrm{T}} = \mathbf{I}_W$ guarantees that $\tilde{\mathbf{\Lambda}}_{\mathrm{ST}} = \tilde{\mathbf{\Lambda}}_{\mathrm{S}}$, which proves that the estimate $\hat{\mathbf{\Pi}}_{\mathrm{S}}$ can be obtained from the eigenvectors of $\hat{\mathbf{R}}_{\mathrm{S}}(L)$.

APPENDIX B
MSE LOWER BOUND FOR MS-ST

Let us assume $\mathbf{Q}$ known (or, equivalently, let $N \to \infty$), $\hat{r}_{\mathrm{S}} = r_{\mathrm{S}}$, and $\hat{r}_{\mathrm{T}} = r_{\mathrm{T}}$ (unbiased channel estimate). Recall that the unknown parameters are $\boldsymbol{\theta} = [\boldsymbol{\theta}_{\mathbf{U}}^{\mathrm{T}}, \boldsymbol{\theta}_{\mathbf{\Gamma}}^{\mathrm{T}}]^{\mathrm{T}}$, where $\boldsymbol{\theta}_{\mathbf{U}} = \mathrm{vec}\{[\mathbf{U}_{\mathrm{S}}\mathbf{U}_{\mathrm{T}}]\}$ contains the slot-independent parameters, and $\boldsymbol{\theta}_{\mathbf{\Gamma}} = \mathrm{vec}\{[\mathbf{\Gamma}(1)\cdots\mathbf{\Gamma}(L)]\}$ the slot-dependent terms. We can express the CRB for the MS estimator based on the parameterization (10) as a function of the matrix of sensitivities $\mathbf{D} = \partial\mathbf{h}/\partial\boldsymbol{\theta}^{\mathrm{T}}$, where $\mathbf{h} = \mathrm{vec}\{[\mathbf{U}_{\mathrm{S}}\mathbf{\Gamma}(1)\mathbf{U}_{\mathrm{T}}^{\mathrm{H}}\cdots\mathbf{U}_{\mathrm{S}}\mathbf{\Gamma}(L)\mathbf{U}_{\mathrm{T}}^{\mathrm{H}}]\}$:

$$\mathrm{Cov}\left\{\hat{\mathbf{h}}\right\} \geq \frac{1}{N}\mathbf{D}\mathbf{J}^{\dagger}\mathbf{D}^{\mathrm{H}}. \tag{36}$$

Since the Fisher information matrix (FIM) for the parameter vector $\boldsymbol{\theta}$

$$\mathbf{J} = \mathbf{D}^{\mathrm{H}}\left(\mathbf{I}_L \otimes \mathbf{R}_{xx}^{*} \otimes \mathbf{Q}^{-1}\right)\mathbf{D} \tag{37}$$

is singular, the CRB depends on the pseudoinverse $\mathbf{J}^{\dagger}$ [21]. The FIM matrix can be partitioned as

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\mathbf{UU}} & \mathbf{J}_{\mathbf{U\Gamma}} \\ \mathbf{J}_{\mathbf{U\Gamma}}^{\mathrm{H}} & \mathbf{J}_{\mathbf{\Gamma\Gamma}} \end{bmatrix} \tag{38}$$

where the $(Mr_{\mathrm{S}} + Wr_{\mathrm{T}}) \times (Mr_{\mathrm{S}} + Wr_{\mathrm{T}})$ submatrix $\mathbf{J}_{\mathbf{UU}}$ is related to the slot-independent channel parameters $\boldsymbol{\theta}_{\mathbf{U}}$, and the $(Lr_{\mathrm{S}}r_{\mathrm{T}}) \times (Lr_{\mathrm{S}}r_{\mathrm{T}})$ submatrix $\mathbf{J}_{\mathbf{\Gamma\Gamma}}$ corresponds to the slot-dependent parameters $\boldsymbol{\theta}_{\mathbf{\Gamma}}$. It can be shown that the FIM entries of $\mathbf{J}_{\mathbf{UU}}$ are $O(L)$, whereas the elements of $\mathbf{J}_{\mathbf{\Gamma\Gamma}}$ and $\mathbf{J}_{\mathbf{U\Gamma}}$ are limited for any L. Assuming for $\mathbf{J}^{\dagger}$ the same matrix partition as (38) and using the properties for the pseudoinverse of a partitioned matrix [22], it follows that for $L \to \infty$, all the blocks of $\mathbf{J}^{\dagger}$ are $O(1/L)$ apart from $[\mathbf{J}^{\dagger}]_{\mathbf{\Gamma\Gamma}}$. Therefore, asymptotically, the covariance of the estimate for $\boldsymbol{\theta}_{\mathbf{U}}$ approaches zero, and the parameters that affect the covariance of the estimate depend on $\boldsymbol{\theta}_{\mathbf{\Gamma}}$ only. It is easy to compute the matrix $\mathbf{D}$ for the parameter vector $\boldsymbol{\theta}_{\mathbf{\Gamma}}$ as

$$\mathbf{D} = \mathbf{I}_L \otimes \mathbf{U}_{\mathrm{T}}^{*} \otimes \mathbf{U}_{\mathrm{S}}. \tag{39}$$

Equation (36) then reduces to

$$\begin{aligned}
\mathrm{Cov}\left\{\hat{\mathbf{h}}\right\} \geq \frac{1}{N}\,\mathbf{I}_L \otimes &\Big(\mathbf{U}_{\mathrm{T}}^{*}\big(\mathbf{U}_{\mathrm{T}}^{\mathrm{T}}\mathbf{R}_{xx}^{*}\mathbf{U}_{\mathrm{T}}^{*}\big)^{-1}\mathbf{U}_{\mathrm{T}}^{\mathrm{T}}\Big) \\
&\otimes \Big(\mathbf{U}_{\mathrm{S}}\big(\mathbf{U}_{\mathrm{S}}^{\mathrm{H}}\mathbf{Q}^{-1}\mathbf{U}_{\mathrm{S}}\big)^{-1}\mathbf{U}_{\mathrm{S}}^{\mathrm{H}}\Big).
\end{aligned} \tag{40}$$

Since for $L \to \infty$ the MS-ST estimator coincides with the MLE (see Section V-B), the lower bound on the MSE (on a slot-by-slot basis) can be obtained by averaging with respect to the fading amplitudes

$$\begin{aligned}
\mathrm{MSE}_{\text{MS-ST}} &= \frac{1}{L}\,\mathrm{E}\Big[\mathrm{tr}\Big\{\mathrm{Cov}\Big\{\hat{\mathbf{h}}\Big\}\Big\}\Big] \\
&= \frac{1}{N}\,\mathrm{tr}\Big\{\mathbf{Q}^{\mathrm{H}/2}\mathbf{\Pi}_{\mathrm{S}}\mathbf{Q}^{1/2}\Big\}\,\mathrm{tr}\Big\{\mathbf{R}_{xx}^{-1/2}\mathbf{\Pi}_{\mathrm{T}}\mathbf{R}_{xx}^{-\mathrm{H}/2}\Big\}
\end{aligned} \tag{41}$$

where $\mathbf{\Pi}_{\mathrm{S}} = (\mathbf{Q}^{-\mathrm{H}/2}\mathbf{U}_{\mathrm{S}})(\mathbf{Q}^{-\mathrm{H}/2}\mathbf{U}_{\mathrm{S}})^{\dagger}$, and $\mathbf{\Pi}_{\mathrm{T}} = (\mathbf{R}_{xx}^{1/2}\mathbf{U}_{\mathrm{T}})(\mathbf{R}_{xx}^{1/2}\mathbf{U}_{\mathrm{T}})^{\dagger}$. Notice that the expectation has no effect in (41) since the covariance (40) is independent on $\{\mathbf{D}(\ell)\}_{\ell=1}^{L}$.

APPENDIX C
PROPERTIES OF $\Phi(\mathbf{\Pi}, \mathbf{T})$

Let $\mathbf{\Pi}_{\mathbf{X}}$ be the $N \times N$ projection matrix onto $\mathcal{R}\{\mathbf{X}\}$, and let $\mathbf{T}$ be an $N \times N$ positive definite matrix. The operator $\Phi(\mathbf{\Pi}_{\mathbf{X}}, \mathbf{T}) = \mathrm{tr}\{\mathbf{T}^{\mathrm{H}/2}\mathbf{\Pi}_{\mathbf{X}}\mathbf{T}^{1/2}\}/\sqrt{N}$ has the following properties that can be easily derived from the properties of the $\mathrm{tr}\{\cdot\}$ operator.

1) $\Phi(\mathbf{\Pi}_{\mathbf{X}}, \mathbf{T}) \geq 0$.
2) $\Phi(\mathbf{\Pi}_{\mathbf{X}}, \alpha\cdot\mathbf{T}) = \alpha\cdot\Phi(\mathbf{\Pi}_{\mathbf{X}}, \mathbf{T})$ for any real-valued $\alpha \geq 0$.
3) $\Phi(\mathbf{\Pi}_{\mathbf{X_1}} + \mathbf{\Pi}_{\mathbf{X_2}}, \mathbf{T}) = \Phi(\mathbf{\Pi}_{\mathbf{X_1}}, \mathbf{T}) + \Phi(\mathbf{\Pi}_{\mathbf{X_2}}, \mathbf{T})$ for any $N \times N$ matrices $\mathbf{X}_1$ and $\mathbf{X}_2$. In particular, it is $\Phi(\mathbf{I}_N, \mathbf{T}) = \Phi(\mathbf{\Pi}_{\mathbf{X}}, \mathbf{T}) + \Phi(\mathbf{\Pi}_{\mathbf{X}}^{\perp}, \mathbf{T})$.
4) $\Phi(\mathbf{I}_N, \mathbf{T}) = \mathrm{tr}\{\mathbf{T}\}/\sqrt{N}$.
5) $\Phi(\mathbf{\Pi}_{\mathbf{X}}, \mathbf{I}_N) = \mathrm{rank}\{\mathbf{\Pi}_{\mathbf{X}}\}/\sqrt{N}$.

Relation (31) between the MS performances can be easily derived from (29) and (30) by using Property 2, invoking Property 1, and recalling (27).

In order to prove (32), we need the use of the MSE bound $\mathrm{MSE}_{\mathrm{RR}}(\ell)$ given in Table I for the RR estimate [6]. We assume that the rank order for the channel matrix is the same in each slot, and it is determined by the spatial diversity $\mathrm{rank}\{\tilde{\mathbf{H}}(\ell)\} = r = r_{\mathrm{S}} \leq r_{\mathrm{T}}$ for any ℓ. This means that we are excluding the possibility of null fading amplitudes that can reduce the rank of $\tilde{\mathbf{H}}(\ell)$ below r_{S}. Under these conditions, it is $\mathrm{rank}\{\mathbf{D}(\ell)\mathbf{G}^{\mathrm{T}}\mathbf{R}_{xx}^{\mathrm{H}/2}\} = r_{\mathrm{T}}$, and $\mathbf{\Pi}_{\mathrm{S}}(\ell) = \mathbf{\Pi}_{\tilde{\mathbf{H}}}(\ell) = \mathbf{\Pi}_{\mathbf{Q}^{-\mathrm{H}/2}\mathbf{A}\mathbf{D}(\ell)\mathbf{G}^{\mathrm{T}}\mathbf{R}_{xx}^{\mathrm{H}/2}} = \mathbf{\Pi}_{\mathbf{Q}^{-\mathrm{H}/2}\mathbf{A}}$. On the other hand, for the temporal projection, it is $r \leq r_{\mathrm{T}}$, and therefore, $\mathcal{R}\{\mathbf{\Pi}_{\mathrm{T}}(\ell)\} \subseteq \mathcal{R}\{\mathbf{\Pi}_{\mathrm{T}}\}$. Since $\mathbf{\Pi}_{\mathrm{S}}(\ell) = \mathbf{\Pi}_{\mathrm{S}}$, from the MSE lower bound for the RR method (Table I) and Property 1, it follows that

$$\begin{aligned}\mathrm{MSE}_{\mathrm{RR}}(\ell) &= \Phi(\mathbf{\Pi}_{\mathrm{S}}, \mathbf{Q})\Phi\left(\mathbf{I}_W, \mathbf{R}_{xx}^{-1}\right) \\ &\quad + \Phi(\mathbf{\Pi}_{\mathrm{S}}^{\perp}, \mathbf{Q})\Phi(\mathbf{\Pi}_{\mathrm{T}}(\ell), \mathbf{R}_{xx}^{-1}) \\ &\geq \Phi(\mathbf{\Pi}_{\mathrm{S}}, \mathbf{Q})\Phi\left(\mathbf{I}, \mathbf{R}_{xx}^{-1}\right) \qquad (42)\end{aligned}$$

$$\mathrm{MSE}_{\mathrm{RR}} = \mathrm{E}[\mathrm{MSE}_{\mathrm{RR}}(\ell)] \geq \mathrm{MSE}_{\mathrm{MS\text{-}S}} \geq \mathrm{MSE}_{\mathrm{MS\text{-}ST}}. \qquad (43)$$

Dually, for $r = r_{\mathrm{T}} \leq r_{\mathrm{S}}$, it is $\mathcal{R}\{\mathbf{\Pi}_{\mathrm{S}}(\ell)\} \subseteq \mathcal{R}\{\mathbf{\Pi}_{\mathrm{S}}\}$ and $\mathbf{\Pi}_{\mathrm{T}}(\ell) = \mathbf{\Pi}_{\mathrm{T}}$, and therefore

$$\begin{aligned}\mathrm{MSE}_{\mathrm{RR}}(\ell) &= \Phi\left(\mathbf{\Pi}_{\mathrm{T}}, \mathbf{R}_{xx}^{-1}\right)\Phi(\mathbf{I}_M, \mathbf{Q}) \\ &\quad + \Phi\left(\mathbf{\Pi}_{\mathrm{T}}^{\perp}, \mathbf{R}_{xx}^{-1}\right)\Phi(\mathbf{\Pi}_{\mathrm{S}}(\ell), \mathbf{Q}) \\ &\geq \Phi\left(\mathbf{\Pi}_{\mathrm{T}}, \mathbf{R}_{xx}^{-1}\right)\Phi(\mathbf{I}_M, \mathbf{Q}) \qquad (44)\end{aligned}$$

$$\mathrm{MSE}_{\mathrm{RR}} = \mathrm{E}[\mathrm{MSE}_{\mathrm{RR}}(\ell)] \geq \mathrm{MSE}_{\mathrm{MS\text{-}T}} \geq \mathrm{MSE}_{\mathrm{MS\text{-}ST}}. \qquad (45)$$

From (43) and (45), we can finally conclude that for any rank $\{r_{\mathrm{S}}, r_{\mathrm{T}}\}$, it is $\mathrm{MSE}_{\mathrm{RR}} \geq \mathrm{MSE}_{\mathrm{MS\text{-}ST}}$.

REFERENCES

[1] O. Simeone and U. Spagnolini, "Multislot estimation of space-time channels," in *Proc. IEEE Int. Conf. Commun.*, Apr. 2002.

[2] CWTS, "China Wireless Telecommunication Standard (CWTS) Working Group 1 (WG1): Physical channels and mapping of transport channels onto physical channels," TS C102 v3.3.0, 2000–2009.

[3] A. Van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint angle and delay estimation using shift-invariance techniques," *IEEE Trans. Signal Processing*, vol. 46, pp. 405–418, Feb. 1998.

[4] A. L. Swindlehurst, "Time delay and spatial signature estimation using known asynchronous signals," *IEEE Trans. Signal Processing*, vol. 46, pp. 449–462, Feb. 1998.

[5] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions," *IEEE Trans. Signal Processing*, vol. 44, pp. 3069–3078, Dec. 1996.

[6] M. Nicoli, "Multi-user reduced rank receivers for TD/CDMA systems," Ph.D. dissertation, Politecnico di Milano, Milano, Italy, Dec. 2001.

[7] M. Nicoli and U. Spagnolini, "Reduced rank channel estimation for time-slotted mobile communication systems," *IEEE Trans. Signal Processing*, submitted for publication.

[8] H. Andoh, M. Sawahashi, and F. Adachi, "Channel estimation using time-multiplexed pilot symbols for coherent Rake combining for DS-CDMA mobile radio," in *PIRMC*, vol. 3, 1997, pp. 954–958.

[9] M. C. Vanderveen, A. Van der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Processing*, vol. 46, pp. 682–690, Mar. 1998.

[10] J. Picheral and U. Spagnolini, "Shift invariance algorithms for the delays/angles estimation of multipath space-time channel," in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, 2001, pp. 83–87.

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

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

[13] G. L. Stuber, *Principles of Mobile Communication*. Boston, MA: Kluwer, 1996.

[14] M. Viberg and P. Stoica, "Maximum likelihood array processing in spatially correlated noise fields using parametrized signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 996–1004, Apr. 1997.

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

[16] 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, pp. 376–378, Sept. 2002.

[17] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, pp. 95–107, Jan. 1995.

[18] D. J. Rabideau, "Fast, rank adaptive subspace tracking and applications," *IEEE Trans. Signal Processing*, vol. 44, pp. 2229–2244, Sept. 1996.

[19] P. Strobach, "Low-rank adaptive filters," *IEEE Trans. Signal Processing*, vol. 44, pp. 2932–2947, Dec. 1996.

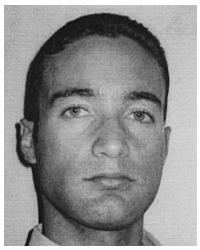
[20] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.

[21] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Processing*, vol. 49, pp. 87–90, Jan. 2001.

[22] W. E. Larimore, "Order-recursive factorization of the pseudoinverse of a covariance matrix," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 1299–1303, Jan. 2001.

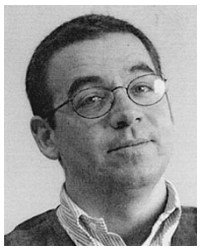
Monica Nicoli (M'99) received the M.Sc. degree (with honors) and the Ph.D. degree in telecommunication engineering from Politecnico di Milano, Milan, Italy, in 1998 and 2002, respectively.

From March to August 2001, she was a visiting researcher with the Signals and Systems Group, Uppsala University, Uppsala, Sweden. Currently, she is with the Dipartimento di Elettronica e Informazione, Politecnico di Milano. Her research interests include antenna array processing for wireless communication systems, channel estimation and equalization for multiuser systems, and multitarget detection and tracking for remote sensing applications.



Osvaldo Simeone (M'02) received the M.Sc. degree (with honors) from Politecnico di Milano, Milan, Italy, in 2001. He is currently pursuing the Ph.D. degree at Politecnico di Milano.

From February to September 2002, he was a visiting researcher with the Center for Communications and Signal Processing Research, New Jersey Institue of Technology, Newark. He holds a patent on the work developed for his Master's thesis. His current research interests lie in the field of signal processing for digital communications, with emphasis on MIMO systems, multicarrier modulation, and channel estimation.



Umberto Spagnolini (M'99) received the Dott. Ing. Elettronica (cum laude) degree 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. His 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 the EAGE in 1991, and the Best Paper Award from the EAGE in 1998.