Multislot Estimation of Frequency-Selective Fast-Varying Channels

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

***Abstract*—In mobile communications, the movement of terminals renders the multipath channel time varying. Even though the faded amplitudes are fast varying, the delays can be considered as stationary on a large temporal scale. In this paper, we propose a new subspace-based method that estimates the channel response from multiple slots by capitalizing on these different varying rates without explicitly computing the delays of the multipath. The temporal subspace is obtained from multiple single-slot training-based estimates of the (single-user or multiuser) channel response. Provided that the number of slots is large enough, the time basis can be calculated with any accuracy. As a consequence, the mean-square error on the channel response depends only on the number of fast-varying parameters that have to be estimated in a slot-by-slot fashion. Performance analysis and simulations confirm the expected benefits of the multislot approach in improving the efficiency of systems with short training sequences.**

***Index Terms*—Code-division multiple access (CDMA), delay estimation, fading channels, multipath channels, multiuser channels, reduced rank processing, subspace tracking, time-division multiple access (TDMA), time-varying channels, training.**

I. Introduction

IN MOBILE communication systems, the channel response is generally estimated by using training sequences. These are required to be long enough to reduce the estimation error at the expense of transmission efficiency. Even though blind or semiblind methods have been proposed to avoid the use of training sequences with moderate loss of performance (e.g., see [1] and [2]), it seems that this loss of efficiency is unavoidable in time-slotted systems [such as time-division multiple access (TDMA) or hybrid time-division code-division multiple access (TD-CDMA)] when the channel is time varying and/or a bootstrap estimate is needed. In these systems, the estimation accuracy can be increased by appropriately merging the information relative to channel estimates calculated in "neighboring" slots. Each single-slot estimate can have a poor accuracy, since it is obtained from a short training sequence. Nonetheless, the multislot estimate experiences a lower error, achieving the same performance of a virtually longer training sequence. In our framework, different slots can be distributed in time (e.g., asynchronous transmission) and frequency (e.g., frequency hopping). The multislot approach is trivial for a static channel, since the estimates can be simply averaged (provided that timing misalignments among different slots are negligible). For a slowly varying channel, weighted averaging of the channel estimates (WMSA method) shows reasonable improvements with respect to plain averaging [3]. Nonetheless, adaptation of the WMSA method to fast-varying fading channels in [4] and [5] showed unacceptable performance. In order to cope with fast-varying fading channels, a different approach has to be pursued that is based on the analysis of the stationarities of mobile communications channels.

Due to the movement of the terminal, the parameters that describe the channel response are time varying with different rates: the fading fluctuations are fast varying, while the delay pattern is slow varying, or even stationary. The fading fluctuations are related to the speed of mobiles, and the corresponding coherence time is roughly related to the inverse of the Doppler shift. On the other hand, the delay pattern can be considered as stationary within large time scales; for instance, since a terminal with radial speed 50 km/h moves approximately 1 m in 100 ms, in a symbol period $T = 1\mu$s, the shift of delays is much smaller than $T/100$. Methods have been proposed that explicitly estimate the delays of the multipath. In fact, high-resolution techniques (e.g., MUSIC, ESPRIT, or IQML) for frequency estimation can be applied to the delay-estimation problem by transforming time delays into linear phase shifts by using Fourier methods [6]. In this case, the fading amplitudes act as sources and their uncorrelation is a necessary prerequisite to exploit the resolution properties of these methods. Furthermore, the delays can be estimated from many slots with an accuracy that increases with the number of slots. These techniques are useful when the multipath parameters need to be estimated for an array of antennas [7], [8] and path selection can reduce the complexity of the detector [9]. However, such a high resolution is not mandatory for delay estimation in dense multipath [10]. Here, we capitalize on the delay-pattern stationarity, but instead of explicitly estimating the delays, we exploit the invariance of the subspace spanned by the corresponding delayed transmitted waveforms (temporal subspace) for channel estimation [26].

The proposed method estimates the temporal subspace from multiple-slots measurements. Then, it projects onto this subspace the estimate of the channel response obtained slot by slot according to the conventional least squares (LS) approach. For a large number of slots, the temporal subspace can be evaluated with high accuracy. This leads to an improvement with respect to the LS method (in terms of estimation error) that basically depends on the ratio between the channel support assumed in the estimation process (W) and the degree of time diversity (r_o). Since the channel support can be held longer than necessary to accommodate different propagation environments (such

Paper approved by G. M. Vitetta, the Editor for Equalization and Fading Channels of the IEEE Communications Society. Manuscript received December 20, 2001; revised October 7, 2002 and January 24, 2003.

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

Digital Object Identifier 10.1109/TCOMM.2003.815067

as urban areas with temporally dense paths or rural areas with few sparse paths [11]), the ratio is $W/r_o > 1$.

Differently from other subspace-based methods [12], the algorithm here proposed performs a multislot processing based on the (slot-by-slot) LS estimates of the channel response, and thus, it can be seen (and implemented) as a refinement of the traditional LS estimate.

For the sake of simplicity, at first a single-user system is considered. The extension to multiuser systems (e.g., TD-CDMA) is then treated by highlighting the main differences (and degradation) with respect to the single-user case. The overall organization of the paper is as follows. Multislot channel estimation for single-user systems is covered in Section II with emphasis on model definition, equivalence with the maximum-likelihood (ML) estimate, selection of the degree of temporal diversity, analytic evaluation of the performance in terms of mean-square error (MSE), and some remarks on computationally efficient implementations. The estimator is generalized to a multiuser system in Section III. Finally, computer simulations are presented in Section IV in order to validate the performance of the proposed algorithm in a realistic setting that includes fading correlation and timing error.

Basic notation: In this paper, lowercase (uppercase) bold denotes column vector (matrices), $(\cdot)^T$ is the matrix transpose, $(\cdot)^*$ is the complex conjugate, $(\cdot)^H$ is the Hermitian transposition, $\|\mathbf{X}\|^2_{\mathbf{A}} = \mathrm{tr}\{\mathbf{X}^H\mathbf{A}\mathbf{X}\}$ is the norm weighted by a positive definite matrix $\mathbf{A}$, $\mathbf{v} = \mathrm{vec}\{\mathbf{V}\}$ is the stacking operator and $\otimes$ is the Kronecker matrix product (for the properties, see [13]), $\mathbf{I}_P$ is the $P \times P$ unit matrix, and $\Pi_{\mathbf{A}}$ is the projection matrix onto range $\{\mathbf{A}\}$.

II. Multislot Processing

A. Model Definition

This section defines the signal model for a single-input (i.e., with a single-user) single-output time-slotted system. The user of interest transmits a sequence of bursts, each containing a set of known symbols (training sequence) used for channel estimation. The burst transmission can be either asynchronous (the time interval between two successive bursts depends on the burst index T_ℓ, as shown in Fig. 1), or synchronous with constant burst rate ($T_\ell = \Delta T$). Furthermore, consecutive slots can be transmitted on different carriers (i.e., frequency hopping). The baseband model for the signal received within the training period of the ℓth burst is (the receiver is assumed to be synchronized to the start of each burst)

$$y(t;\ell) = \sum_m x(m;\ell)h(t-mT;\ell) + n(t;\ell) \tag{1}$$

where $\{x(m;\ell)\}$ denotes the training sequence from a finite alphabet, $h(t;\ell)$ is the baseband channel assumed to be constant within the burst, and $n(t;\ell)$ is additive white Gaussian noise (AWGN). Henceforth, the terms "burst" and "slot" will be used interchangeably.

Within a set of L consecutive bursts (selected according to the terminal mobility), the channel can be modeled as a combination of d paths, each characterized by a burst-independent delay τ_i and a burst-dependent complex amplitude $\alpha_i(\ell)$

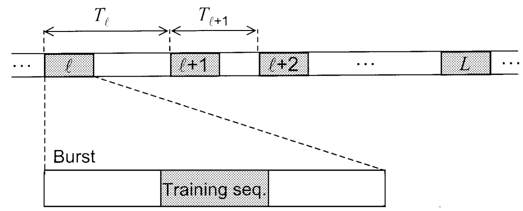


Fig. 1. Burst transmission in time-slotted systems.

$$h(t;\ell) = \sum_{i=1}^{d} \alpha_i(\ell)g(t-\tau_i), \qquad \ell = 1,\ldots,L \tag{2}$$

the waveform $g(t)$ is the convolution of the transmitted pulse and the matched filter at the receiver. For the model (2) to hold, the variation of the delays $\{\tau_i\}_{i=1}^d$ within L bursts should be smaller than the temporal resolution, i.e., approximately the inverse of the bandwidth of the transmitted signal. Furthermore, the variations of the fading amplitudes $\alpha_i(\ell)$ within the burst are assumed to be negligible.

After sampling at symbol rate $1/T$, the discrete-time system corresponding to (1) is obtained by collecting the N samples of the received signal into the $N \times 1$ vector $\mathbf{y}(\ell)$

$$\mathbf{y}(\ell) = \mathbf{X}(\ell)\mathbf{h}_o(\ell) + \mathbf{n}(\ell) \tag{3}$$

$\mathbf{X}(\ell) = [\mathbf{x}(0;\ell),\cdots,\mathbf{x}(-W+1;\ell)]$ is the $N \times W$ convolution matrix with $\mathbf{x}(n;\ell) = [x(n;\ell),\ldots,x(n+N-1;\ell)]^T$ and $\mathbf{h}_o(\ell) = [h(0;\ell),h(T;\ell),\ldots,h((W-1)T;\ell)]^T$ is the $W \times 1$ vector of the channel response. The AWGN is uncorrelated with respect to the slot, i.e., $\mathrm{E}[\mathbf{n}(\ell)\mathbf{n}^H(\ell+m)] = \sigma^2\mathbf{I}_N\delta_m$, and the power level σ^2 is not necessarily known.

According to the model (2), the channel can be rewritten as the linear combination of d vectors $\mathbf{g}(\tau_i) = [g(-\tau_i), g(T-\tau_i),\ldots,g((W-1)T-\tau_i)]^T$, each corresponding to a delayed waveform $g(t)$

$$\mathbf{h}_o(\ell) = \sum_{i=1}^{d} \alpha_i(\ell)\mathbf{g}(\tau_i) = \mathbf{G}(\boldsymbol{\tau})\boldsymbol{\alpha}(\ell). \tag{4}$$

The advantage of this model is to separate the burst-independent $(\mathbf{G}(\boldsymbol{\tau}))$ from the burst-dependent $(\boldsymbol{\alpha}(\ell))$ parameters. The temporal matrix $\mathbf{G}(\boldsymbol{\tau}) = [\mathbf{g}(\tau_1),\ldots,\mathbf{g}(\tau_d)]$ depends on the set of d delays $\boldsymbol{\tau} = [\tau_1,\ldots,\tau_d]^T$. No specific assumptions are made on the distribution of the fading amplitudes $\boldsymbol{\alpha}(\ell) = [\alpha_1(\ell),\ldots,\alpha_d(\ell)]^T$. In case of a single-carrier asynchronous transmission, the amplitudes can be characterized by the correlation matrix $\mathrm{E}[\boldsymbol{\alpha}(\ell+m)\boldsymbol{\alpha}^H(\ell)] = \mathrm{diag}\{A_1,\ldots,A_d\} \times \rho(\ell;m)$, where the normalized correlation function $\rho(\ell;m)$ generally depends on the time interval $\sum_{k=\ell}^{\ell+m-1} T_k$ (see Fig. 1) and on the terminal mobility. If the burst transmission is synchronous, the correlation $\rho(m)$ is a function of $m\Delta T$. Moreover, in case the carrier frequency is changed on a slot-by-slot basis, fading is uncorrelated and $\rho(m) = \delta(m)$.

According to the linear model (4), the channel vector $\mathbf{h}_o(\ell)$ can be equivalently parametrized as the product of a $W \times r_o$ burst-independent full column rank matrix $\mathbf{U}_o$ and a $r_o \times 1$ burst-dependent vector $\mathbf{b}_o(\ell)$

$$\mathbf{h}_o(\ell) = \mathbf{U}_o \mathbf{b}_o(\ell). \tag{5}$$

We define the r_o-dimensional subspace spanned by the columns of $\mathbf{U}_o$ as the temporal subspace. The factorization (5) implies that the channel $\mathbf{h}_o(\ell)$ can be obtained as a linear combination of the columns of $\mathbf{U}_o$. Since the following equality holds:

$$\mathbf{G}(\boldsymbol{\tau})\left[\boldsymbol{\alpha}(1), \ldots, \boldsymbol{\alpha}(L)\right] = \mathbf{U}_o\left[\mathbf{b}_o(1), \ldots, \mathbf{b}_o(L)\right] \tag{6}$$

it follows that the dimension of the temporal subspace is $r_o = \min\{\operatorname{rank}\{\mathbf{G}(\boldsymbol{\tau})\}, \operatorname{rank}\{\mathbf{R}_\alpha(L)\}\}$, where $\mathbf{R}_\alpha(L) = L^{-1}\sum_{\ell=1}^{L}\boldsymbol{\alpha}(\ell)\boldsymbol{\alpha}^H(\ell)$ is the sample correlation matrix of fading. In other words, the model order r_o depends both on the number of the delays that can be resolved according to the bandwidth (i.e., on $\operatorname{rank}\{\mathbf{G}(\boldsymbol{\tau})\} \leq d$), and on the correlation properties of the amplitudes (i.e., on $\operatorname{rank}\{\mathbf{R}_\alpha(L)\} \leq d$). Since each column of $\mathbf{G}(\boldsymbol{\tau})$ contains a single delayed waveform $\mathbf{g}(\tau_i)$, any column of $\mathbf{U}_o$ gathers the samples of a compound waveform which is obtained as a linear combination of the delayed waveforms in $\mathbf{G}(\boldsymbol{\tau})$. According to our discussion, this linear combination depends on the delay pattern $(\mathbf{G}(\boldsymbol{\tau}))$, but also on the fading amplitudes and their correlation over the slots $\mathbf{R}_\alpha(L)$. For instance, for a still terminal and a single-carrier transmission, the channel is static, $\mathbf{h}_o(\ell) = \bar{\mathbf{h}}_o$, and the temporal subspace is one-dimensional independently of the number of delays as $\operatorname{rank}\{\mathbf{R}_\alpha(L)\} = 1$. Thus it is $r_o = 1$ and $\mathbf{U}_o = \bar{\mathbf{h}}_o$ (apart from a scaling term).

The advantage of the channel model (5) is the redefinition of the burst-independent matrix $\mathbf{U}_o$ without making explicit use of the delays $\boldsymbol{\tau}$. The degree of temporal diversity r_o of $\mathbf{U}_o$ is not known and it needs to be estimated separately from the set $\{\mathbf{y}(\ell)\}_{\ell=1}^{L}$. The estimation of r_o will be discussed in Section II-C.

B. Multislot Channel Estimate

The maximum-likelihood estimate (MLE) of the parameters $\{\mathbf{U}, \mathbf{b}\}$ can be reduced to the minimization of the loss function

$$\mathcal{L}_N(\mathbf{U}, \mathbf{b}) = \frac{1}{L}\sum_{\ell=1}^{L}\|\mathbf{y}(\ell) - \mathbf{X}(\ell)\mathbf{U}\mathbf{b}(\ell)\|^2 \tag{7}$$

under the constraint that $\operatorname{rank}\{\mathbf{U}\} = r_o$. The vector $\mathbf{b} = [\mathbf{b}^T(1), \ldots, \mathbf{b}^T(L)]^T$ denotes the ensemble of Lr_o burst-dependent terms. Without any loss of generality, we can assume that the correlation of the training sequence remains the same across all the bursts so that $\mathbf{R}_{\mathbf{XX}} = \mathbf{X}^H(\ell)\mathbf{X}(\ell)$ is independent of ℓ. Optimization with respect to (w.r.t.) $\mathbf{b}(\ell)$ is quadratic (assuming $\mathbf{U}$ known)

$$\hat{\mathbf{b}}(\ell) = (\widetilde{\mathbf{U}}^H\widetilde{\mathbf{U}})^{-1}\widetilde{\mathbf{U}}^H\mathbf{R}_{\mathbf{XX}}^{-\frac{H}{2}}\mathbf{X}^H(\ell)\mathbf{y}(\ell) \tag{8}$$

label $\widetilde{(.)}$ denotes the premultiplication by $\mathbf{R}_{\mathbf{XX}}^{1/2}$ obtained from any factorization of the full-rank matrix $\mathbf{R}_{\mathbf{XX}}$ such that $\mathbf{R}_{\mathbf{XX}} = \mathbf{R}_{\mathbf{XX}}^{H/2}\mathbf{R}_{\mathbf{XX}}^{1/2}$ (e.g., the Cholesky factorization). By substituting (8) into (7), we can reduce the optimization to

$$\hat{\widetilde{\mathbf{U}}} = \arg\min_{\widetilde{\mathbf{U}}} \operatorname{tr}\left[(\mathbf{I}_W - \Pi_{\widetilde{\mathbf{U}}})\mathbf{R}(L)\right] \tag{9}$$

where $\Pi_{\widetilde{\mathbf{U}}} = \widetilde{\mathbf{U}}(\widetilde{\mathbf{U}}^H\widetilde{\mathbf{U}})^{-1}\widetilde{\mathbf{U}}^H$ is the projection matrix onto the subspace spanned by the columns of $\widetilde{\mathbf{U}} = \mathbf{R}_{\mathbf{XX}}^{1/2}\mathbf{U}$, and $\mathbf{R}(L)$ denotes the $W \times W$ matrix

$$\mathbf{R}(L) = \mathbf{R}_{\mathbf{XX}}^{-\frac{H}{2}}\left(\frac{1}{L}\sum_{\ell=1}^{L}\mathbf{X}^H(\ell)\mathbf{y}(\ell)\mathbf{y}^H(\ell)\mathbf{X}(\ell)\right)\mathbf{R}_{\mathbf{XX}}^{-\frac{1}{2}}. \tag{10}$$

Let $\{\mathbf{v}_n, \lambda_n\}$ be the nth eigenvector/eigenvalue pair of the positive semidefinite matrix $\mathbf{R}(L)$ for nonincreasing ordering of the eigenvalues $(\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_W)$, and assume $\lambda_{r_o} > \lambda_{r_o+1}$. The following inequality holds for the complementary objective function (9):

$$\operatorname{tr}\left[\Pi_{\widetilde{\mathbf{U}}}\mathbf{R}(L)\right] = \operatorname{tr}\left[\sum_{n=1}^{W}\lambda_n\Pi_{\widetilde{\mathbf{U}}}\mathbf{v}_n\mathbf{v}_n^H\right] \leq \sum_{n=1}^{r_o}\lambda_n. \tag{11}$$

The equality holds only if $\hat{\Pi}_{\widetilde{\mathbf{U}}} = \mathbf{V}_{r_o}\mathbf{V}_{r_o}^H$, where $\mathbf{V}_{r_o} = [\mathbf{v}_1, \ldots, \mathbf{v}_{r_o}]$. This is the minimizer in (9).

The channel estimate is then obtained as $\hat{\mathbf{h}}(\ell) = \mathbf{R}_{\mathbf{XX}}^{-1/2}\hat{\widetilde{\mathbf{U}}}\hat{\mathbf{b}}(\ell)$. Once the projector onto the time basis $\hat{\Pi}_{\widetilde{\mathbf{U}}} = \mathbf{V}_{r_o}\mathbf{V}_{r_o}^H$ is estimated from the leading eigenvectors of the multiburst matrix $\mathbf{R}(L)$ and the amplitudes $\hat{\mathbf{b}}(\ell)$ are computed slot by slot according to (8), the estimate is

$$\hat{\mathbf{h}}(\ell) = \mathbf{R}_{\mathbf{XX}}^{-\frac{1}{2}}\mathbf{V}_{r_o}\mathbf{V}_{r_o}^H\mathbf{R}_{\mathbf{XX}}^{-\frac{H}{2}}\mathbf{X}^H(\ell)\mathbf{y}(\ell). \tag{12}$$

Let us remark that the estimate of the slot-independent matrix $\widetilde{\mathbf{U}}$ is not really needed in (12) and it can be obtained as $\hat{\widetilde{\mathbf{U}}} = \mathbf{V}_{r_o}$ (or $\hat{\mathbf{U}} = \mathbf{R}_{\mathbf{XX}}^{-1/2}\mathbf{V}_{r_o}$) or by any convenient way that preserves the corresponding projector $\hat{\Pi}_{\widetilde{\mathbf{U}}} = \mathbf{V}_{r_o}\mathbf{V}_{r_o}^H$.

The MLE (12) has a simple interpretation that is useful when an efficient implementation based on the modification of the standard LS estimate is of interest. The LS estimate of the channel matrix for the ℓth burst is known to be $\mathbf{h}_{\mathrm{LS}}(\ell) = \mathbf{R}_{\mathbf{XX}}^{-1}\mathbf{X}^H(\ell)\mathbf{y}(\ell)$ and the covariance matrix $\operatorname{Cov}[\mathbf{h}_{\mathrm{LS}}(\ell)] = \sigma^2\mathbf{R}_{\mathbf{XX}}^{-1}$ depends on the correlation properties of the training sequence (as in general, $\mathbf{R}_{\mathbf{XX}}$ is not diagonal). Let

$$\widetilde{\mathbf{h}}_{\mathrm{LS}}(\ell) = \mathbf{R}_{\mathbf{XX}}^{\frac{1}{2}}\mathbf{h}_{\mathrm{LS}}(\ell) = \mathbf{R}_{\mathbf{XX}}^{-\frac{H}{2}}\mathbf{X}^H(\ell)\mathbf{y}(\ell) \tag{13}$$

be the LS estimate after whitening (i.e., $\operatorname{Cov}[\widetilde{\mathbf{h}}_{\mathrm{LS}}(\ell)] = \sigma^2\mathbf{I}_W$), it follows that

$$\mathbf{R}(L) = \frac{1}{L}\sum_{\ell=1}^{L}\widetilde{\mathbf{h}}_{\mathrm{LS}}(\ell)\widetilde{\mathbf{h}}_{\mathrm{LS}}^H(\ell) \tag{14}$$

and the multislot estimate (12) becomes

$$\hat{\mathbf{h}}(\ell) = \mathbf{R}_{\mathbf{XX}}^{-\frac{1}{2}}\hat{\Pi}_{\widetilde{\mathbf{U}}}\widetilde{\mathbf{h}}_{\mathrm{LS}}(\ell). \tag{15}$$

In conclusion, the multislot MLE (12) is the projection of the single-burst whitened estimate $\widetilde{\mathbf{h}}_{\mathrm{LS}}(\ell)$ onto the temporal subspace spanned by the r_o leading eigenvectors of the matrix

$\mathbf{R}(L)$ computed from the ensemble of L whitened LS estimates $\{\widetilde{\mathbf{h}}_{\mathrm{LS}}(\ell)\}_{\ell=1}^{L}$. In Appendix A, it is shown that the general property of the MLE of being a reestimate from the LS solution holds for every parametrization of the channel vector. This general result has found application in some system identification problems [15]. In addition, since the factorizations $\mathbf{R}_{\mathbf{XX}}^{-1/2}$ and $\mathbf{R}_{\mathbf{XX}}^{1/2}$ depend on the training sequence, they can be precalculated, stored, and accessed when necessary.

The multislot technique (15) introduces a latency of L bursts in providing the channel estimate. An adaptive slot-by-slot processing can be obtained by updating the projector $\hat{\Pi}_{\widetilde{\mathbf{U}}}$ (or equivalently, the eigenvectors $\mathbf{V}_{r_o}$) under the assumption of slowly varying temporal subspace. Subspace tracking algorithms such as Refinement Only-Fast Subspace Tracking (RO-FST) [18] have the advantage of updating both the temporal subspace and the model order $\hat{r}_o$ with a computational complexity $O(W\hat{r}_o)$ per slot. For a fixed (or known) model order, the loss function (7) can be minimized by solving sequentially two quadratic problems when optimizing w.r.t. $\mathbf{U}$ and $\mathbf{b}$. As shown in [17], this implementation can be reduced to an alternating-power (AP) algorithm [19].

C. Model Order Selection

The parameter r_o, introduced in Section II-A, is the minimum order that allows an unbiased channel estimate. It can be equivalently defined as the rank of the matrix $\mathbf{R}(L)$ in (10) for $\sigma^2 = 0$: $r_o = \mathrm{rank}\{\mathbf{R}_0(L)\}$, where

$$\mathbf{R}_0(L) = \mathbf{R}_{\mathbf{XX}}^{\frac{1}{2}}\mathbf{G}(\boldsymbol{\tau})\mathbf{R}_{\boldsymbol{\alpha}}(L)\mathbf{G}^H(\boldsymbol{\tau})\mathbf{R}_{\mathbf{XX}}^{\frac{H}{2}}. \qquad (16)$$

In practice, the model order to be used in (12) has to be estimated from the analysis of the eigenvalues of the matrix $\mathbf{R}(L)$. In the following, it will be shown by simple arguments that the estimated model order, here denoted as $\hat{r}_o$, is more conveniently chosen as $\hat{r}_o \le r_o$ for any $\sigma^2 > 0$.

According to the model (3) and the estimate (12), the error for the ℓth burst can be written as composed of two terms

$$\hat{\mathbf{h}}(\ell) - \mathbf{h}_o(\ell) = \mathbf{R}_{\mathbf{XX}}^{-\frac{1}{2}}\left(\hat{\Pi}_{\widetilde{\mathbf{U}}} - \Pi_{\widetilde{\mathbf{U}}_o}\right)\mathbf{R}_{\mathbf{XX}}^{\frac{1}{2}}\mathbf{h}_o(\ell) + \mathbf{R}_{\mathbf{XX}}^{-\frac{1}{2}}\hat{\Pi}_{\widetilde{\mathbf{U}}}\mathbf{R}_{\mathbf{XX}}^{-\frac{H}{2}}\mathbf{X}^H(\ell)\mathbf{n}(\ell) \qquad (17)$$

$\Pi_{\widetilde{\mathbf{U}}_o}$ is the projector onto the subspace spanned by the r_o leading eigenvectors of $\mathbf{R}_0(L)$, and the equality $\Pi_{\widetilde{\mathbf{U}}_o}\mathbf{R}_{\mathbf{XX}}^{1/2}\mathbf{h}_o(\ell) = \mathbf{R}_{\mathbf{XX}}^{1/2}\mathbf{h}_o(\ell)$ has been used. The first term in (17) accounts approximately for the distortion that depends on the mismatch between the estimate $\hat{\Pi}_{\widetilde{\mathbf{U}}}$ and the "true" projector $\Pi_{\widetilde{\mathbf{U}}_o}$, while the second term depends mainly on noise. The distortion decreases with increasing $\hat{r}_o$. For $L \to \infty$ (or for a number of bursts large enough) and $\hat{r}_o = r_o$, the distortion vanishes as $\hat{\Pi}_{\widetilde{\mathbf{U}}} \to \Pi_{\widetilde{\mathbf{U}}_o}$ (see Section II-D) so that the $\mathrm{MSE} = \mathrm{E}[\|\hat{\mathbf{h}}(\ell) - \mathbf{h}_o(\ell)\|^2]$ (here the averaging is carried out w.r.t. noise and fading amplitudes) depends only on the noise power. A similar conclusion can be drawn for $\hat{r}_o > r_o$. Moreover, with increasing $\hat{r}_o$, a larger noise subspace is retained in the estimate (12), leading to an increased noise contribution. This latter observation suggests that for low SNRs, it is convenient to accept a biased channel estimate $(\hat{r}_o < r_o)$ in order to reduce the noise contribution, trading some distortion for a lower MSE. The minimum description length (MDL) principle [16] is a suitable criterion for model order selection as it can be proved to approximately minimize the MSE on the estimate when evaluated as a function of $\hat{r}_o$ [17].

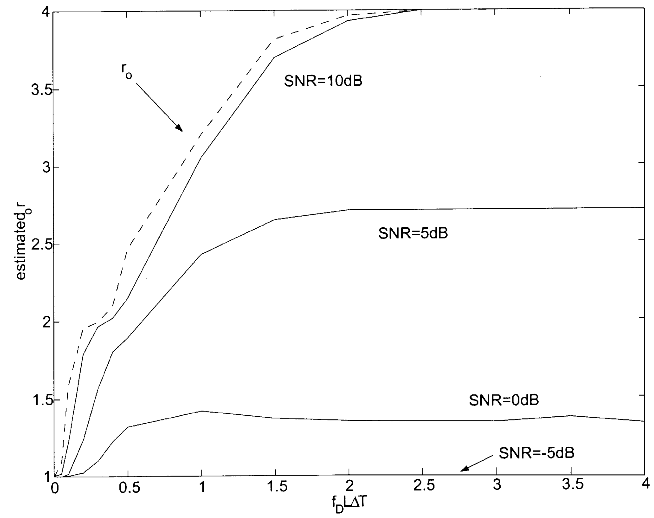


Fig. 2. Effect of fading correlation in the selection of the model order $\hat{r}_o$ for varying $\mathrm{SNR}'s$ (solid line) compared to $r_o = \min\{4, \mathrm{rank}(\mathbf{R}_\alpha(L))\}$ (dashed line) (rank $\{\mathbf{G}(\tau)\} = 4, W = 15, N = 20$).

A simple example can illustrate the preceding discussion. Let us consider a single-carrier synchronous transmission and let the fading variations be described by the Clarke's model of two–dimensional (2-D) isotropic scattering [14] so that $\rho(m) = J_0(2\pi f_D m \Delta T)$ where $f_D = v/\lambda$ is the Doppler shift and ΔT is the time interval between successive bursts. The channel length is $W = 15$ and the $d = 4$ paths have decreasing amplitudes $\{A_k\}_{k=1}^{4} = A \times [0, -3, -6, -9]$ dB (A is scaled to have $\mathrm{E}[\|\mathbf{h}_o(\ell)\|^2] = 1$) and delays $[5.1,\ 6.2,\ 6.8,\ 9.8]T$ so that $\mathrm{rank}\{\mathbf{G}(\boldsymbol{\tau})\} = 4$. The training sequence is randomly generated with $N = 20$, and the SNR is defined as $\mathrm{SNR} = \sigma_x^2\mathrm{E}[\|\mathbf{h}_o(\ell)\|^2]/\sigma^2 = \sigma_x^2/\sigma^2$. Fig. 2 shows the model order $r_o = \min\{4, \mathrm{rank}\{\mathbf{R}_\alpha(L)\}\}$ (evaluated with a threshold of 40 dB on the eigenvalues of $\mathbf{R}_\alpha(L)$[1]) and the model order $\hat{r}_o$ selected according to the MDL criterion averaged over 10^4 independent runs of simulation. The number of slots is irrelevant, since the behavior depends only on the term $f_D L\Delta T$ (here $L = 10$ is chosen). For $f_D L\Delta T > 2.5$, it is $\mathrm{rank}\{\mathbf{R}_{\boldsymbol{\alpha}}(L)\} = 4$ so that r_o approaches the number of resolvable paths (i.e., $\mathrm{rank}\{\mathbf{G}(\boldsymbol{\tau})\} = 4$). For a large SNR, the estimated model order $\hat{r}_o$ approaches $r_o = \min\{4, \mathrm{rank}\{\mathbf{R}_\alpha(L)\}\}$, the model order $\hat{r}_o$ is uniformly smaller for SNR < 10 dB, and it yields $\hat{r}_o = 1$ for SNR < 0 dB. Furthermore, for a static channel $(f_D L\Delta T = 0)$ or for $L = 1$, the dimension of the temporal subspace is independent of d and $r_o = 1$. In this case, the channel vector remains constant across the bursts and the spanned subspace is clearly 1-D.

D. Performance Analysis

Performance is evaluated in terms of MSE on the channel estimate. The Cramer–Rao bound (CRB) for an unbiased estimate

[1] $\mathrm{rank}(\mathbf{R}_\alpha(L)) = r_\alpha(L)$ for $\sum_{i=1}^{r_\alpha(L)}\mu_i(\mathbf{R}_\alpha(L))/\sum_{i=r_\alpha(L)+1}^{d}\mu_i(\mathbf{R}_\alpha(L)) > 10^4$, where $\mu_i(\cdot)$ denoted the ith eigenvalue of its argument.

($\hat{r}_o = r_o$) of the channel parameters $\{\mathbf{U}, \mathbf{b}\}$ is derived in Appendix B. The MSE bound can be evaluated by averaging w.r.t. the fading amplitudes

$$\mathrm{MSE}_{\mathrm{CRB}}[L] = \frac{\sigma^2}{L}\Big[(L-r_o)\mathrm{tr}\left\{\mathbf{R}_{\mathbf{XX}}^{-\frac{1}{2}}\Pi_{\widetilde{\mathbf{U}}_o}\mathbf{R}_{\mathbf{XX}}^{-\frac{H}{2}}\right\} + r_o\mathrm{tr}\left\{\mathbf{R}_{\mathbf{XX}}^{-1}\right\}\Big]. \quad (18)$$

It is of practical interest to evaluate the lower bound for a large number of slots by letting $L \to \infty$ in (18)

$$\begin{aligned}\mathrm{MSE}_{\mathrm{CRB}}[L\to\infty] &= \sigma^2\mathrm{tr}\left\{\mathbf{R}_{\mathbf{XX}}^{-\frac{1}{2}}\Pi_{\widetilde{\mathbf{U}}_o}\mathbf{R}_{\mathbf{XX}}^{-\frac{H}{2}}\right\}\\ &= \sigma^2\mathrm{tr}\left\{\mathbf{U}_o\left(\mathbf{U}_o^H\mathbf{R}_{\mathbf{XX}}\mathbf{U}_o\right)^{-1}\mathbf{U}_o^H\right\}. \quad (19)\end{aligned}$$

This shows that the structured reestimation of the LS solution (15) reduces the MSE of the original estimate, that is known to be (see Appendix C for single-user)

$$\mathrm{MSE}_{\mathrm{LS}} = \sigma^2\mathrm{tr}\left\{\mathbf{R}_{\mathbf{XX}}^{-1}\right\} \quad (20)$$

by projecting the estimate error onto the subspace spanned by $\widetilde{\mathbf{U}}_o$. The residual error after the projection depends only on that component of the noise $\mathbf{h}_{\mathrm{LS}} - \mathbf{h}_o$ that belongs to $\mathrm{range}(\widetilde{\mathbf{U}}_o)$ and therefore can no longer be suppressed.

For temporally uncorrelated training sequence ($\mathbf{R}_{\mathbf{XX}} = N\sigma_x^2\mathbf{I}_W$), the MSE (18) simplifies as

$$\mathrm{MSE}_{\mathrm{CRB}}[L] = \frac{1}{L}\frac{\sigma^2}{N\sigma_x^2}r_o(W+L-r_o) \quad (21)$$

showing that the MSE bound is proportional to the number of degrees of freedom in the parametrization (5) (see Appendix B). The corresponding asymptotic MSE bound

$$\mathrm{MSE}_{\mathrm{CRB}}[L\to\infty] = \frac{\sigma^2}{N\sigma_x^2}r_o \quad (22)$$

depends on the number of r_o parameters (i.e., the entries of vector $\mathbf{b}(\ell)$) to be estimated on a slot-by-slot basis.

The asymptotic bound (19) can be derived by following an alternative approach that, though not as rigorous as the one just presented, gives a remarkable insight into the performances attained by the algorithm for $L \to \infty$. As the number of processed bursts L grows, the slot-independent term $\mathbf{U}_o$ can be estimated with any accuracy, since the Fisher Information Matrix (FIM) for $\mathbf{U}$ is $O(L)$ [see (40a)], and thus, $\mathrm{Cov}[\hat{\mathbf{U}}] \to \mathbf{0}$. On the other hand, the accuracy of the estimate for the burst-varying term $\mathbf{b}_o(\ell)$ is independent of L, and from the corresponding FIM [see (40b)] we obtain $\mathrm{Cov}[\hat{\mathbf{b}}(\ell)] \to \sigma^2(\mathbf{U}_o^H\mathbf{R}_{\mathbf{XX}}\mathbf{U}_o)^{-1}$. From these remarks, $\mathrm{Cov}[\hat{\mathbf{h}}(\ell)] \to \sigma^2\mathbf{U}_o(\mathbf{U}_o^H\mathbf{R}_{\mathbf{XX}}\mathbf{U}_o)^{-1}\mathbf{U}_o^H$ and the MSE bound (19) can be easily derived. Furthermore, the preceding discussion justifies the MSE dependency in (22) on the slot-by-slot estimate.

From (20), the LS estimate for training sequence with $\mathbf{R}_{\mathbf{XX}} = N\sigma_x^2\mathbf{I}_W$ is

$$\mathrm{MSE}_{\mathrm{LS}} = \frac{\sigma^2}{N\sigma_x^2}W. \quad (23)$$

Thus, the advantage of the multislot estimate can be easily quantified in the ratio W/r_o between the different number of free parameters to be estimated on each slot. In general, it is $W/r_o \gg 1$. As a consequence, the multislot method allows reducing the length of the training sequence by W/r_o yielding asymptotically the same performance of the LS estimate with a longer training sequence (see Section IV).

III. Multiuser System

The signal model (3) is here extended to describe multiuser time-slotted systems (such as third-generation direct-sequence (DS)-CDMA systems, as the time-division duplex (TDD) mode of UMTS [11]). The signals received during the training period of the ℓth slot can be written as the sum of contributions from K active users

$$\mathbf{y}(\ell) = \sum_{k=1}^{K}\mathbf{X}^{(k)}(\ell)\mathbf{h}_o^{(k)}(\ell) + \mathbf{n}(\ell) = \mathcal{X}(\ell)\mathrm{h}_o(\ell) + \mathbf{n}(\ell) \quad (24)$$

where the superscript denotes the dependence on the kth user, $\mathcal{X}(\ell) = [\mathbf{X}^{(1)}(\ell)\cdots\mathbf{X}^{(K)}(\ell)]$, and $\mathrm{h}_o(\ell) = [\mathbf{h}_o^{(1)}(\ell)^T\cdots\mathbf{h}_o^{(K)}(\ell)^T]^T$ are the ensemble of training sequences and channels. The correlation properties of the training sequences are assumed to be independent of $\ell : \mathbf{R}_{\mathcal{X}\mathcal{X}} = \mathcal{X}^H(\ell)\mathcal{X}(\ell)$.

Appendix A shows that the MLE of $\mathrm{h}_o(\ell)$ from (24) can be equivalently obtained by reestimating the parametrized channel vectors from the LS estimate $\mathrm{h}_{\mathrm{LS}}(\ell) = \mathbf{R}_{\mathcal{X}\mathcal{X}}^{-1}\mathcal{X}^H(\ell)\mathbf{y}(\ell)$. The estimation reduces to the minimization of the loss function

$$\mathcal{L}_N(\mathbf{U},\mathbf{b}) = \frac{1}{L}\sum_{\ell=1}^{L}\|\mathrm{h}(\ell,\mathbf{U},\mathbf{b}) - \mathrm{h}_{\mathrm{LS}}(\ell)\|^2_{\mathbf{R}_{\mathcal{X}\mathcal{X}}}. \quad (25)$$

By partitioning the Cholesky factorization of the whitening matrix $\mathbf{R}_{\mathcal{X}\mathcal{X}}$ into blocks

$$\mathbf{R}_{\mathcal{X}\mathcal{X}}^{\frac{1}{2}} = \begin{bmatrix}\mathbf{Q}_{11} & \cdots & \mathbf{Q}_{K1}\\ \vdots & \ddots & \vdots\\ \mathbf{0} & \cdots & \mathbf{Q}_{KK}\end{bmatrix} \quad (26)$$

it is possible to highlight in (25) the contribution of each user as follows:

$$\begin{aligned}\mathcal{L}_N(\mathbf{U},\mathbf{b}) = \frac{1}{L}\sum_{k=1}^{K}\sum_{\ell=1}^{L}\Bigg\|&\mathbf{Q}_{kk}\mathbf{U}^{(k)}\mathbf{b}^{(k)}(\ell)\\ &-\left[\mathbf{Q}_{kk}\mathbf{h}_{\mathrm{LS}}^{(k)}(\ell) + \sum_{i=k+1}^{K}\left(\mathbf{Q}_{ki}\mathbf{h}_{\mathrm{LS}}^{(i)}(\ell) - \mathbf{U}^{(i)}\mathbf{b}^{(i)}(\ell)\right)\right]\Bigg\|^2. \quad (27)\end{aligned}$$

The optimization of $\mathcal{L}_N(\mathbf{U},\mathbf{b})$ in a closed form is not an easy task, but the analysis of (27) suggests an approximated solution similar to the successive interference cancellation in multiuser detection [20]. This solution is obtained by minimizing separately the terms corresponding to each user, starting with the Kth user down to the first. The iterative scheme is (for $k = K, K-1, \ldots, 1$)

$$\hat{\mathbf{h}}^{(k)}(\ell) = \mathbf{Q}_{kk}^{-1}\hat{\Pi}^{(k)}\tilde{\mathbf{h}}_{\mathrm{LS}}^{(k)}(\ell) \quad (28)$$

where the prewhitening accounts for the interference cancellation

$$\tilde{\mathbf{h}}_{\mathrm{LS}}^{(k)}(\ell) = \mathbf{Q}_{kk}\mathbf{h}_{\mathrm{LS}}^{(k)}(\ell) + \sum_{i=k+1}^{K} \mathbf{Q}_{ki}\left(\mathbf{h}_{\mathrm{LS}}^{(i)}(\ell) - \hat{\mathbf{h}}^{(k)}(\ell)\right) \quad (29)$$

at the cost of $W^2K(K-1)/2$ multiplications. $\hat{\Pi}^{(k)}$ is the projector onto the subspace spanned by the $\hat{r}_o^{(k)}$ principal eigenvectors of the correlation matrix $\mathbf{R}^{(k)}(L) = L^{-1}\sum_{\ell=1}^{L}\tilde{\mathbf{h}}_{\mathrm{LS}}^{(k)}(\ell)\tilde{\mathbf{h}}_{\mathrm{LS}}^{(k)H}(\ell)$. Let us remark that if the training sequences are mutually uncorrelated, the multiuser algorithm described by (28) and (29) reduces to the single-user case (14) and (15) to be carried out on a user-by-user basis. In this case (which is closely approximated by third-generation standards [11]), the multiuser approach does not increase the computational complexity of channel estimation per user. In addition, the near–far effect or the different correlation properties of the training sequences could be taken into account by sorting the users, and thus optimizing the performances of this iterative scheme.

Since the multislot estimate (28) and (29) is based on the LS estimate $\mathrm{h}_{\mathrm{LS}}(\ell)$, it is interesting to assess the performance of the latter in a multiuser environment. In the following, we consider the simple case of K training sequences that have the same temporal correlation properties and are reciprocally equicorrelated (with correlation coefficient $\rho_u \leq 1$), so that

$$\mathbf{X}^{(k)H}\mathbf{X}^{(k)} = N\sigma_x^2\mathbf{R}_t, \qquad k = 1, \ldots, K \quad (30a)$$

$$\mathbf{X}^{(n)H}\mathbf{X}^{(k)} = \rho_u N\sigma_x^2\mathbf{R}_t, \qquad n \neq k \quad (30b)$$

where $\mathbf{R}_t$ is the normalized correlation matrix (i.e., $[\mathbf{R}_t]_{i,i} = 1 \geq [\mathbf{R}_t]_{i,j}$) for the training sequence of each user. Therefore, the overall correlation matrix is

$$\mathbf{R}_{\mathcal{X}\mathcal{X}} = N\sigma_x^2\mathbf{R}_u \otimes \mathbf{R}_t \quad (31)$$

with $\mathbf{R}_u = \rho_u\mathbf{1}_K\mathbf{1}_K^T + (1-\rho_u)\mathbf{I}_K$. Notice that for a single-user system $(K = 1)$ the matrix (31) reduces to $\mathbf{R}_{\mathcal{X}\mathcal{X}} = \mathbf{R}_{\mathbf{XX}} = N\sigma_x^2\mathbf{R}_t$. The effect of multiaccess interference on the LS estimate can be quantified in closed form in terms of the degradation in MSE, or equivalently, SNR (due to the direct proportionality of the two measures), with respect to the single-user case (or $\rho_u = 0$). This degradation reads (see Appendix C)

$$D(\rho_u, K) = \frac{1}{1-\rho_u}\frac{1+\rho_u(K-2)}{1+\rho_u(K-1)}. \quad (32)$$

Notice that, as expected, $D(0,1) = 1$. In the following section, simulation results will show that the expression (32), although derived for the LS estimate, closely approximates the MSE degradation (with respect to the single-user case) measured on the parametric reestimate performed by the proposed algorithm.

IV. Simulation Results

In this section, the performance are evaluated for fading amplitudes uncorrelated from burst to burst $(\rho(m) = \delta(m))$, as this condition leads to the worst asymptotic performances, as discussed in Section II-C. It is worth pointing out that the assumption of uncorrelated fading across bursts implies that the estimate of the fading process cannot benefit from a multislot measurement. Nonetheless, the multislot approach is justified by the dependence of the received signal on the long-term parameters, namely the delays, or equivalently, the temporal basis (see Section II-A). The pulse $g(t)$ is a raised cosine with roll-off factor 0.2, the information symbols are independent and identically distributed (i.i.d.) binary phase-shift keying (BPSK) with the same power σ_x^2 of the training symbols. The system parameters and the delays are the same as in the example of Fig. 2. If not differently specified, the order of time diversity is selected according to the MDL criterion and the training sequence is randomly generated such that $\mathrm{E}[\mathbf{X}^H(\ell)\mathbf{X}(\ell)] = N\sigma_x^2\mathbf{I}_W, N = 20$. We consider first a single-user system, and then we evaluate the performance degradation for the multiuser case.

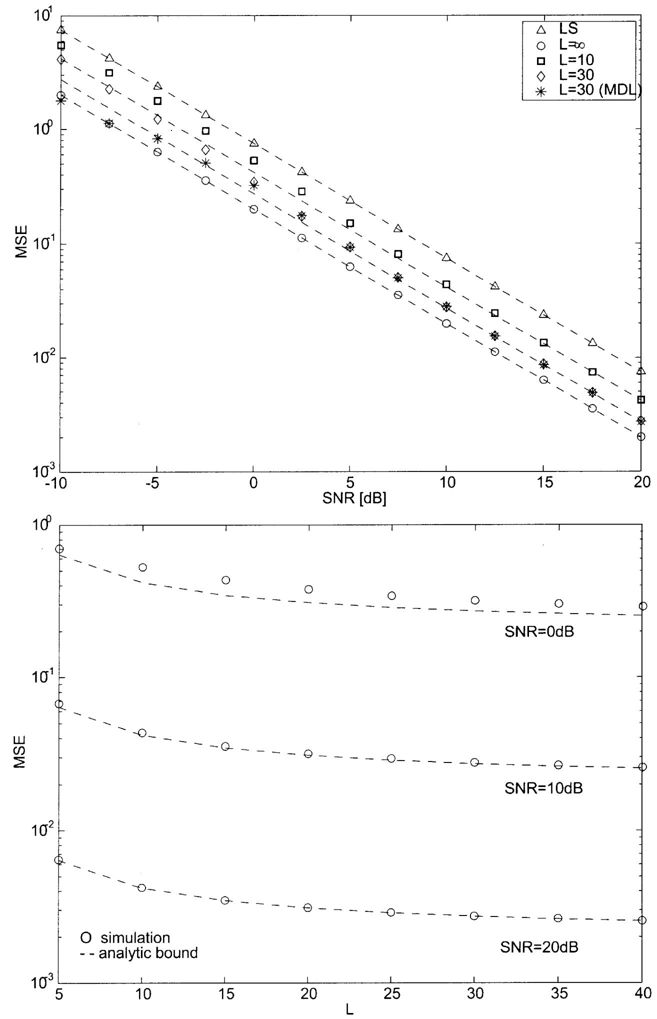


Fig. 3. MSE versus SNR (upper) and MSE versus L (lower) for LS and the multislot method (markers). MSE analytic bounds are in dashed line ($W = 15$, $N = 20$).

Fig. 3 compares the simulations for the MSE on the channel estimate (markers) with the MSE bounds (21) and (22) (dashed lines). The MSE versus SNR for L =10, 30, and $L \to \infty$ (or equivalently, $\hat{\Pi}_{\tilde{\mathbf{U}}} = \Pi_{\tilde{\mathbf{U}}_o}$) with the choice $\hat{r}_o = r_o = 4$ (upper figure) shows that the performance attains the MSE bound and

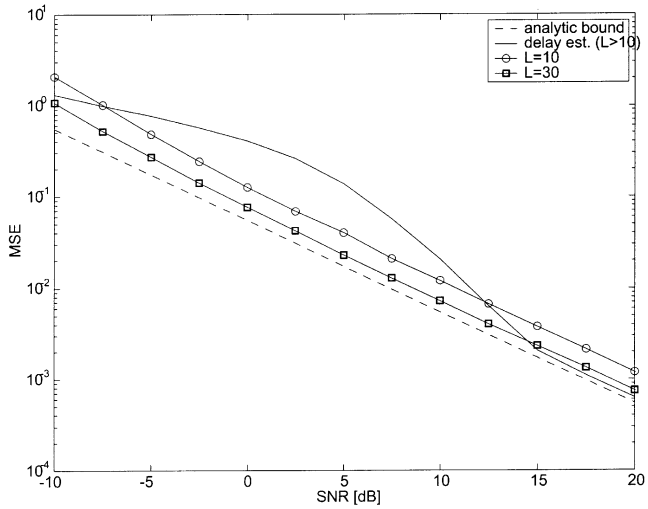


Fig. 4. MSE versus SNR for the structured channel estimator and the proposed unstructured method ($W = 15, N = 20$).

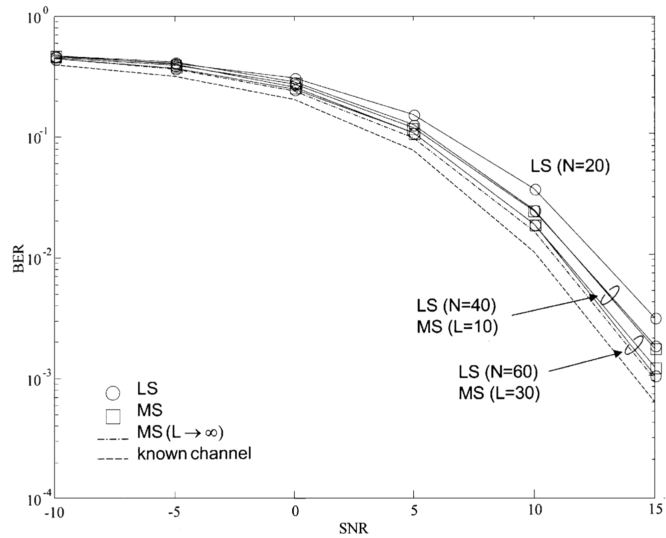


Fig. 5. BER versus SNR for the linear ZF with channel estimation ($W = 20$) performed by the multislot algorithm with different values of the number of slots $L (L = 10, 30)$ and $N = 20$, and the LS estimator with different training sequence lengths ($N = 20, 40, 60$).

that the multislot estimate outperforms the LS estimate. Furthermore, Fig. 3 demonstrates that for $\hat{r}_o$ selected according to the MDL criterion, the performance can be further improved for low SNRs as $\hat{r}_o < r_o$ (see Fig. 2). The MSE versus L (lower figure) for $\mathrm{SNR} =$0, 10, 20 dB shows again that the simulations closely approach the analytic MSE bounds (18), (19) and further proves that the asymptotic performance for $L \to \infty$ can be easily reached for a reasonable number of slots (in practice, $L \geq 20$).

Performance improvement with respect to single-slot techniques can be obtained when dealing with more realistic propagation environments in dense multipath, as far as the estimated channel impulse response length is larger than the delay diversity ($W > r_o$). Even though for $r_o = W$, it seems that there is no advantage in using the multislot method proposed here, it is still possible to improve the performance at small SNRs by selecting $\hat{r}_o < r_o$ according to the MDL criterion. In addition, the multislot method proposed here is not impaired by the threshold effects, typical of nonlinear estimation problems, that can be experienced when the delay estimation algorithms are used in low SNRs [6].

To emphasize the last statement, Fig. 4 compares the performance of the multislot technique presented in this paper with a structured technique consisting in the multislot ML estimate of the delays and the consequent LS estimate of the amplitudes [6], [8]. To simplify the problem at hand, we consider the same setting as before, except that now a single propagation path is considered ($d = 1$) with delay $5.1T$. The choice of a channel with a single delay implies that the ML delay estimator is $\hat{\tau} = \arg\max_\tau \mathbf{g}^T(\tau)(\sum_{\ell=1}^{L} \mathbf{h}_{\mathrm{LS}}(\ell)\mathbf{h}_{\mathrm{LS}}^H(\ell))\mathbf{g}(\tau)$ (recall that for ideal training sequence, the additive noise on $\mathbf{h}_{\mathrm{LS}}(\ell)$ is white) and the slot-by-slot amplitude estimate is $\hat{\alpha}(\ell) = (\mathbf{g}^T(\hat{\tau})/||\mathbf{g}(\hat{\tau})||^2) \cdot \mathbf{h}_{\mathrm{LS}}(\ell)$ [6]. Even though this constitutes a privileged scenario for delay estimation (as there are no resolution issues), the threshold effect causes the structured method to be outperformed by the proposed technique for $L > 10$ and sufficiently small SNRs, as shown in Fig. 4. For higher SNRs, the structured method attains its MSE bound derived from the CRB, which can be shown to coincide with (22) [8]. As expected, the unstructured technique has a slower convergence, owing to the larger number of parameters that have to be estimated from the multislot measurements.

The performance analysis in terms of error probability needs to take into account the channel estimation error at the detector. This study is rather complex, as the channel mismatch depends on the estimated model order $\hat{r}_o$ that cannot be easily related to the delays or fading correlation. However, for $L \to \infty$ and large SNR such that $\hat{r}_o = r_o$ the channel mismatch is known to be given exclusively by the error on amplitudes as $\hat{\Pi}_{\tilde{\mathbf{U}}} \to \Pi_{\tilde{\mathbf{U}}_o}$ (see Section II-C). In this case, the analysis of the probability of error can be reduced to the error probability bounds derived in [21] for maximum-likelihood sequence estimation (MLSE) by assuming the use of one antenna. Since the cost of the MLSE would be prohibitive for large W, here we evaluate the impact of channel estimation on the performance of linear equalizers, such as block equalizers [22]. In this case, it can be shown that the channel mismatch can be reduced to an equivalent additive noise at the receiver's input that yields a SNR degradation with respect to the performance for a perfect knowledge of the channel. This degradation depends on the covariance matrix of the channel estimate as $(1 + (\sigma_x^2/\sigma^2)\mathrm{tr}\{\mathrm{Cov}[\hat{\mathbf{h}}]\})$ [23]. For uncorrelated training sequence with $\mathbf{R}_{\mathbf{XX}} = N\sigma_x^2\mathbf{I}_W$, the SNR degradation can be evaluated in closed form for the LS estimate (i.e., $(\sigma_x^2/\sigma^2)\mathrm{tr}\{\mathrm{Cov}[\hat{\mathbf{h}}_{\mathrm{LS}}]\} = W/N$) and for the multislot method when $L \to \infty$ (i.e., $(\sigma_x^2/\sigma^2)\mathrm{tr}\{\mathrm{Cov}[\hat{\mathbf{h}}]\} = r_o/N$).

These analytical results are confirmed by the Monte–Carlo simulations shown in Fig. 5 for a ZF-equalizer (block length equal to 100 BPSK symbols) and $W = 20$. The performance of the LS estimate is shown for different training sequence lengths $N =$20, 40, and 60; the expected SNR degradations are $1 + W/N =$3, 1.76, and 1.25 dB, respectively. On the other hand, the performance of the multislot technique is evaluated for a fixed training sequence length $N = 20$ and varying number of slots $L =$10, 30, and $L \to \infty$; asymptotically ($L \to \infty$), the

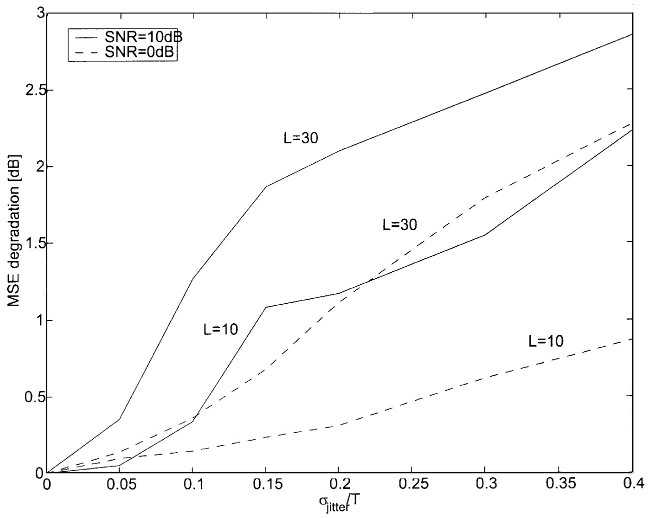


Fig. 6. Degradation of MSE due to random timing misalignements (SNR= 10 dB, $W = 15, N = 20$).

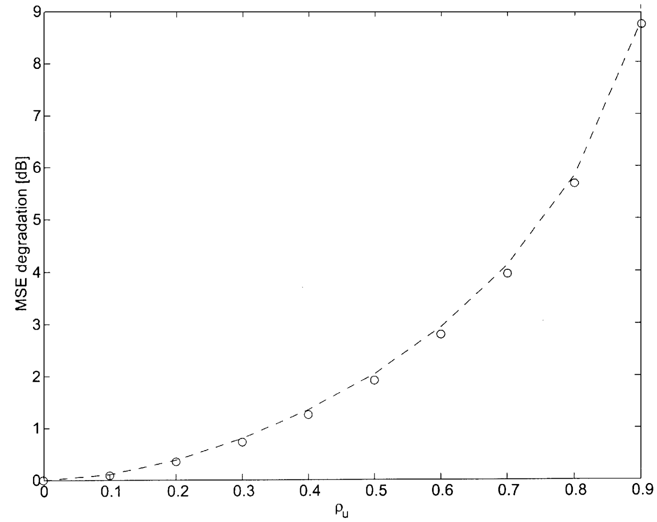


Fig. 7. The MSE degradation of the multislot estimate $\mathrm{MSE}(\rho_u)/\mathrm{MSE}(\rho_u = 0)$ for equicorrelated training sequences. Markers: simulation; dashed line: analytic model from (32) ($K = 4$ users, $L = 10$, $W = 15, N = 80$).

expected SNR degradation is $1 + r_o/N \simeq 0.8$ dB. Notice that, since $\mathrm{Cov}[\hat{\mathbf{h}}] \rightarrow (\sigma^2/\sigma_x^2)\cdot(r_o/N)\mathbf{I} \neq \mathbf{0}$ (see Appendix B), the performance of the multislot method for $L \rightarrow \infty$ does not attain the "known channel" bound.

The results in Fig. 5 allow us to discuss the impact of the multislot approach on the system throughput. It can be seen that the performance of a system employing the LS channel estimate with a certain training sequence length can be obtained by using a shorter training sequence and the multislot estimate. Theoretically, the multislot approach allows a reduction of the training sequence length (and a corresponding increase of the system efficiency) of a factor r_o/W, as confirmed by simulations.

To make the multislot approach valid in practice, the received signals $\{\mathbf{y}(\ell)\}_{\ell=1}^{L}$ should be aligned in time in order to have the same multipath delays in each slot. Here we investigate the effect of timing errors characterized by a random offset independently selected in each slot. According to the model (4), the matrix $\mathbf{G}(\boldsymbol{\tau})$ is modified by adding a time misalignment $\Delta\tau(\ell)$ that is independent from slot to slot: $\mathbf{G}(\boldsymbol{\tau}) = [\mathbf{g}(\tau_1 + \Delta\tau(\ell))\cdots\mathbf{g}(\tau_d + \Delta\tau(\ell))]$ with $\Delta\tau(\ell) \sim \mathcal{N}(0, \sigma_{jitter}^2)$. Even though it is reasonable to assume that the variables $\Delta\tau(\ell)$ are generally correlated (as it happens, for instance, due to the mismatch between the transmitter and receiver clocks), the simple model considered here constitutes an interesting worst-case scenario. Fig. 6 shows the MSE degradation compared to the MSE for $\sigma_{\mathrm{jitter}} = 0$ versus the timing error normalized to the symbol interval $(\sigma_{\mathrm{jitter}}/T)$ for SNR =0, 10 dB. The multislot processing is robust with respect to random timing error as the MSE degradation can be quantified to be less than 3 dB for $\sigma_{\mathrm{jitter}}/T < 0.4$, $L \leq 30$ bursts, and SNR < 10 dB. Analogous simulations in terms of bit-error rate (BER) can show that at probability of error 10^{-4}, no floor effects are encountered provided that $\sigma_{\mathrm{jitter}}/T$ is smaller than 5%.

Let us now consider a multiuser system with $K = 4$ users, $L = 10$ slots, and the same parameters as for the single-user simulations (the length of the training sequence is multiplied by the number of users, $N = 80$). To focus the attention on the difference with respect to single-user system, the performance is validated by considering the MSE degradation with respect to the case $\rho_u = 0$ (or equivalently, to the single-user case $K = 1$). Fig. 7 shows that for equicorrelated training sequences with varying ρ_u, the SNR degradation approximated by (32) for multiuser LS estimate closely describes the simulated values of MSE degradation for the iterative scheme (28) (here for $\mathrm{SNR} = 10$ dB). Since the performance depends on the correlation properties of the training sequences, i.e., on $\mathbf{R}_{\mathcal{X}\mathcal{X}}$, the entries of the matrix $\mathcal{X}(\ell)$ are generated here without the constraint of belonging to a finite alphabet. The performance in terms of error probability for linear multiuser detection schemes can be shown to be as consistent with the improvement in channel estimation accuracy as for the single-user case (Fig. 5).

V. Conclusion

The multislot channel estimation method capitalizes on the stationarity of the temporal subspace without explicitly estimating the delays of the multipath. As opposed to structured channel estimators, the approach does not suffer from the typical impairments of nonlinear estimators (i.e., threshold effect). The subspace is obtained by collecting multiple training-based channel responses evaluated slot by slot according to the standard LS approach. The model order can be selected adaptively using the MDL criterion in order to cope with different SNRs and fading correlations. Analysis and simulations show that fast variations of the radio channel do not represent an impairment for the proposed multislot technique. The advantages of the multislot approach are compared to single-slot channel response estimate in terms of MSE and error probability for linear equalizers. The performance analysis in terms of MSE on the channel estimate shows that asymptotically (i.e., for a large number of slots), the MSE depends only on the number of parameters that have to be estimated on a slot-by-slot basis (or equivalently, on the degree of temporal diversity r_o).

The extension to multiuser system is a simple generalization of the reestimation procedure starting from the multiuser LS channel estimate. The degradation due to the multiuser interference can be quantified in terms of the correlation properties of the training sequences.

APPENDIX A

REESTIMATION FROM LS ESTIMATE

Let $\mathrm{h}(\ell,\boldsymbol{\theta})$ be the channel vector and $\boldsymbol{\theta}$ the vector of parameters describing the channel response, the negative log-likelihood function (apart from uninteresting constants) can be written as

$$\begin{aligned}\mathcal{L}_N(\boldsymbol{\theta}) &= \frac{1}{L}\sum_{\ell=1}^{L}\|\mathbf{y}(\ell)-\mathcal{X}(\ell)\mathrm{h}(\ell,\boldsymbol{\theta})\|^2\\ &= \frac{1}{L}\sum_{\ell=1}^{L}\Big(\|\mathrm{h}(\ell,\boldsymbol{\theta})-\mathrm{h}_{\mathrm{LS}}(\ell)\|^2_{\mathbf{R}_{\mathcal{X}\mathcal{X}}}+\|\mathbf{y}(\ell)\|^2\\ &\qquad -\|\mathrm{h}_{\mathrm{LS}}(\ell)\|^2_{\mathbf{R}_{\mathcal{X}\mathcal{X}}}\Big) \end{aligned}\tag{33}$$

as it can be easily proved by recalling that $\mathrm{h}_{\mathrm{LS}}(\ell) = \mathbf{R}^{-1}_{\mathcal{X}\mathcal{X}}\mathcal{X}^H(\ell)\mathbf{y}(\ell)$. Since the last two terms in (33) do not depend on $\boldsymbol{\theta}$ they can be neglected in the minimization of $\mathcal{L}_N(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$. The equivalent loss function becomes

$$\begin{aligned}\mathcal{L}_N(\boldsymbol{\theta}) &= \frac{1}{L}\sum_{\ell=1}^{L}\|\mathrm{h}(\ell,\boldsymbol{\theta})-\mathrm{h}_{\mathrm{LS}}(\ell)\|^2_{\mathbf{R}_{\mathcal{X}\mathcal{X}}}\\ &= \frac{1}{L}\sum_{\ell=1}^{L}\left\|\widetilde{\mathrm{h}}(\ell,\boldsymbol{\theta})-\widetilde{\mathrm{h}}_{\mathrm{LS}}(\ell)\right\|^2. \end{aligned}\tag{34}$$

We remark that the equivalence between the minimizers of (33) and (34) holds only for AWGN.

APPENDIX B

CRB DERIVATION

As shown in Appendix A, the negative log-likelihood function can be rewritten as

$$\begin{aligned}\mathcal{L}_N(\mathbf{U},\mathbf{B}) &= \sum_{\ell=1}^{L}\|\mathbf{h}_{\mathrm{LS}}(\ell)-\mathbf{U}\mathbf{b}(\ell)\|^2_{\mathbf{R}_{\mathbf{xx}}}\\ &= \left\|\mathbf{H}_{\mathrm{LS}}-\mathbf{U}\mathbf{B}^H\right\|^2_{\mathbf{R}_{\mathbf{xx}}}\end{aligned}\tag{35}$$

where unimportant terms have been neglected. It is $\mathbf{H}_{\mathrm{LS}} = [\mathbf{h}_{\mathrm{LS}}(1),\ldots,\mathbf{h}_{\mathrm{LS}}(L)]$, and $\mathbf{B}=[\mathbf{b}(1),\ldots,\mathbf{b}(L)]^H$ is the $L\times r_o$ matrix containing the (deterministic) amplitudes for the L slots ($\mathrm{rank}\{\mathbf{B}\}=r_o$). Notice that we are assuming an unbiased channel estimate, i.e., $\hat{r}_o = r_o$. The overall parameter vector is $\boldsymbol{\theta}=[\mathbf{u}^T,\mathbf{b}^T]^T$, obtained from the $Wr_o\times 1$ vector $\mathbf{u}=\mathrm{vec}\{\mathbf{U}\}$ and the $Lr_o\times 1$ vector $\mathbf{b}=\mathrm{vec}\{\mathbf{B}^H\}$. Let us arrange the multislot channel into the vector $\mathbf{h}=\mathrm{vec}\{\mathbf{U}\mathbf{B}^H\}$ and recall the equality

$$\mathbf{h}=(\mathbf{I}_L\otimes\mathbf{U})\mathbf{b}=(\mathbf{B}^*\otimes\mathbf{I}_W)\,\mathbf{u}.\tag{36}$$

The derivative of the channel vector $\mathbf{h}$ taken in $\boldsymbol{\theta}=\boldsymbol{\theta}_o$ can be calculated as

$$\mathbf{D}=\left\{\frac{\partial\mathbf{h}}{\partial\boldsymbol{\theta}^T}\right\}_{\theta=\theta_o}=[\mathbf{B}_o^*\otimes\mathbf{I}_W\;\;\mathbf{I}_L\otimes\mathbf{U}_o].\tag{37}$$

The $r_o(W+L)\times r_o(W+L)$ FIM of $\boldsymbol{\theta}$ is thus given by [2]

$$\mathbf{J}=\frac{1}{\sigma^2}\mathbf{D}^H(\mathbf{I}_L\otimes\mathbf{R}_{\mathbf{XX}})\mathbf{D}\tag{38}$$

and can be partitioned as

$$\mathbf{J}=\begin{bmatrix}\mathbf{J}_{UU} & \mathbf{J}_{Ub}\\ \mathbf{J}_{Ub}^H & \mathbf{J}_{bb}\end{bmatrix}\tag{39}$$

where the $Wr_o\times Wr_o$ block $\mathbf{J}_{UU}$ and the $Lr_o\times Lr_o$ block $\mathbf{J}_{bb}$ depend on the burst-independent and on the burst-dependent terms, respectively. The diagonal blocks are obtained from (38)

$$\begin{aligned}\mathbf{J}_{UU} &= \frac{1}{\sigma^2}\sum_{\ell=1}^{L}(\mathbf{b}_o(\ell)\otimes\mathbf{I}_W)\,\mathbf{R}_{\mathbf{XX}}\left(\mathbf{b}_o^H(\ell)\otimes\mathbf{I}_W\right)\\ &= \frac{L}{\sigma^2}\mathbf{R}_{\mathbf{b}}(L)\otimes\mathbf{R}_{\mathbf{XX}} \end{aligned}\tag{40a}$$

$$\mathbf{J}_{bb}=\frac{1}{\sigma^2}\mathbf{I}_L\otimes\left(\mathbf{U}_o^H\mathbf{R}_{\mathbf{XX}}\mathbf{U}_o\right).\tag{40b}$$

Note that the matrix $\mathbf{I}_L\otimes\mathbf{R}_{\mathbf{XX}}$ in (38) is positive definite so that, by using the standard result on the rank of a partitioned matrix [25], it can be shown that $\mathrm{rank}\{\mathbf{J}\}=\mathrm{rank}\{\mathbf{D}\}=r_o(W+L)-r_o^2$. The rank order of $\mathbf{J}$ can be also derived as the number of degrees of freedom in the multislot channel model. Indeed, recalling that in L bursts, the parametrization (5) can be rewritten as $\mathbf{H}=\mathbf{U}\mathbf{B}^H$ with $\mathbf{U}$ and $\mathbf{B}$ full rank matrices, it follows the number of degrees of freedom from the singular value decomposition of matrix $\mathbf{H}$: $[r_oW-r_o(r_o+1)/2]$(left eigenvectors) $+[r_oL-r_o(r_o+1)/2]$ (right eigenvectors) $+[r_o]$ (eigenvalues) $=\mathrm{rank}\{\mathbf{J}\}$.

As $\mathrm{rank}\{\mathbf{J}\}<r_o(W+L)$ the FIM is singular. According to [24], the CRB on the channel estimate depends on the pseudoinverse $\mathbf{J}^\dagger$ through pre- and post-multiplication of the latter by the matrix $\mathbf{D}$ containing the "sensitivities" of the channel vector $\mathbf{h}$ to the parameters $\boldsymbol{\theta}$

$$\mathrm{Cov}[\hat{\mathbf{h}}]\geq\mathrm{CRB}[\hat{\mathbf{h}}]=\mathbf{D}\mathbf{J}^\dagger\mathbf{D}^H.\tag{41}$$

Next, by defining the matrices $\mathbf{D}_1=\mathbf{B}_o^*\otimes\mathbf{I}_W$, $\mathbf{D}_2=\mathbf{I}_L\otimes\mathbf{U}_o$, and $\mathbf{D}=[\mathbf{D}_1\;\mathbf{D}_2]$, the CRB (41) can be written as

$$\mathrm{CRB}[\hat{\mathbf{h}}]=\sigma^2\mathbf{D}(\widetilde{\mathbf{D}}^H\widetilde{\mathbf{D}})^\dagger\mathbf{D}^H\tag{42}$$

$$=\sigma^2\left(\mathbf{I}_L\otimes\mathbf{R}_{\mathbf{XX}}^{-\frac{1}{2}}\right)\Pi_{\tilde{\mathbf{D}}}\left(\mathbf{I}_L\otimes\mathbf{R}_{\mathbf{XX}}^{-\frac{H}{2}}\right)\tag{43}$$

$\widetilde{\mathbf{D}}=(\mathbf{I}_L\otimes\mathbf{R}_{\mathbf{XX}}^{1/2})\mathbf{D}$ and $\Pi_{\widetilde{D}}=\widetilde{\mathbf{D}}(\widetilde{\mathbf{D}}^H\widetilde{\mathbf{D}})^\dagger\widetilde{\mathbf{D}}^H$ is the corresponding projector. Since $\mathrm{range}\{\widetilde{\mathbf{D}}\}=\mathrm{range}\{\widetilde{\mathbf{D}}_1\}\cup\mathrm{range}\{\widetilde{\mathbf{D}}_2\}$, this can be equivalently decomposed into the orthogonal subspaces $\mathrm{range}\{\widetilde{\mathbf{D}}\}=\mathrm{range}\{\Pi^\perp_{\tilde{\mathbf{D}}_1}\widetilde{\mathbf{D}}_2\}\cup\mathrm{range}\{\widetilde{\mathbf{D}}_1\}$ such that $\mathrm{range}\{\Pi^\perp_{\tilde{\mathbf{D}}_1}\widetilde{\mathbf{D}}_2\}\cap\mathrm{range}\{\widetilde{\mathbf{D}}_1\}=\oslash$. The projection matrix reduces to

$$\Pi_{\tilde{\mathbf{D}}}=\Pi_{\Pi^\perp_{\tilde{\mathbf{D}}_1}\tilde{\mathbf{D}}_2}+\Pi_{\tilde{\mathbf{D}}_1}\tag{44}$$

where $\Pi_{\widetilde{\mathbf{D}}_1}$, $\Pi^{\perp}_{\widetilde{\mathbf{D}}_1}$, and $\Pi_{\Pi^{\perp}_{\widetilde{\mathbf{D}}_1}}\widetilde{\mathbf{D}}_2$ denote the orthogonal projections onto, respectively, range$\{\widetilde{\mathbf{D}}_1\}$, the orthogonal complement range$\{\widetilde{\mathbf{D}}_1^{\perp}\}$, and range$\{\Pi^{\perp}_{\widetilde{\mathbf{D}}_1}\widetilde{\mathbf{D}}_2\}$. According to the model exploited here, the projectors in (44) can be easily calculated as follows:

$$\Pi_{\tilde{\mathbf{D}}_1} = \Pi_{\mathbf{B}_o^*} \otimes \mathbf{I}_W \tag{45}$$

$$\Pi^{\perp}_{\tilde{\mathbf{D}}_1}\widetilde{\mathbf{D}}_2 = \left(\Pi^{\perp}_{\mathbf{B}_o^*} \otimes \mathbf{I}_W\right)(\mathbf{I}_L \otimes \widetilde{\mathbf{U}}_o) = \Pi^{\perp}_{\mathbf{B}_o^*} \otimes \widetilde{\mathbf{U}}_o \tag{46}$$

$$\Pi_{\Pi^{\perp}_{\tilde{\mathbf{D}}_1}\tilde{\mathbf{D}}_2} = \Pi^{\perp}_{\mathbf{B}_o^*} \otimes \Pi_{\tilde{\mathbf{U}}_o}. \tag{47}$$

The CRB (47) can be now evaluated as

$$\mathrm{CRB}[\hat{\mathbf{h}}] = \sigma^2\left(\mathbf{I}_L \otimes \mathbf{R}_{\mathbf{XX}}^{-\frac{1}{2}}\right) \times \left(\Pi^{\perp}_{\mathbf{B}_o^*} \otimes \Pi_{\tilde{\mathbf{U}}_o} + \Pi_{\mathbf{B}_o^*} \otimes \mathbf{I}_W\right)\left(\mathbf{I}_L \otimes \mathbf{R}_{\mathbf{XX}}^{-\frac{H}{2}}\right). \tag{48}$$

Since it is $\sum_{\ell=1}^{L}\mathrm{tr}\{\mathrm{CRB}[\hat{\mathbf{h}}(\ell)]\}/L = \mathrm{tr}\{\mathrm{CRB}[\hat{\mathbf{h}}]\}/L$, the average performance over L slots can be obtained as

$$\frac{1}{L}\mathrm{tr}\left\{\mathrm{CRB}[\hat{\mathbf{h}}]\right\} = \frac{\sigma^2}{L}\left\{(L-r_o)\mathrm{tr}\left\{\mathbf{R}_{\mathbf{XX}}^{-\frac{1}{2}}\Pi_{\tilde{\mathbf{U}}_o}\mathbf{R}_{\mathbf{XX}}^{-\frac{H}{2}}\right\} + r_o\mathrm{tr}\left\{\mathbf{R}_{\mathbf{XX}}^{-1}\right\}\right\}. \tag{49}$$

Notice that (49) does not depend on the amplitudes and it remains the same when averaging w.r.t. to the distribution of the fading amplitudes to obtain the MSE (18).

APPENDIX C

MSE FOR MULTIUSER LS CHANNEL ESTIMATE

The MSE on the LS channel estimate of each user [recall the multiuser model (24)] can be easily obtained as $\mathrm{MSE_{LS}} = 1/K \cdot \sigma^2 \mathrm{tr}\{\mathbf{R}_{\mathcal{XX}}^{-1}\}$. For the factorization $\mathbf{R}_{\mathcal{XX}} = N\sigma_x^2\mathbf{R}_u \otimes \mathbf{R}_t$ (31), it is

$$\mathrm{MSE_{LS}} = \frac{1}{K}\frac{\sigma^2}{N\sigma_x^2}\mathrm{tr}\left\{\mathbf{R}_u^{-1}\right\}\mathrm{tr}\left\{\mathbf{R}_t^{-1}\right\}. \tag{50}$$

Assuming reciprocally equicorrelated training sequences, $\mathbf{R}_u = \rho_u\mathbf{1}_K\mathbf{1}_K^T + (1-\rho_u)\mathbf{I}_K$, and using the Sherman–Morrison–Woodbury formula [25]

$$\mathbf{R}_u^{-1} = \frac{1}{1-\rho_u}\left\{\mathbf{I}_K - \frac{1}{\frac{1-\rho_u}{\rho_u}+K}\mathbf{1}_K\mathbf{1}_K^T\right\} \tag{51}$$

we get

$$\mathrm{tr}\left\{\mathbf{R}_u^{-1}\right\} = \frac{K}{1-\rho_u}\frac{1+\rho_u(K-2)}{1+\rho_u(K-1)}. \tag{52}$$

Notice that for $K \gg 1$ it is $\mathrm{tr}\{\mathbf{R}_u^{-1}\} = K/(1-\rho_u)$. Therefore, the MSE per user reads

$$\mathrm{MSE_{LS}} = \frac{\sigma^2}{N\sigma_x^2}\frac{1}{1-\rho_u}\frac{1+\rho_u(K-2)}{1+\rho_u(K-1)}\mathrm{tr}\left\{\mathbf{R}_t^{-1}\right\} \tag{53}$$

yielding a degradation of $\mathrm{MSE_{LS}}$ compared to the $\mathrm{MSE_{LS}}$ for $K=1$ given by (32). We remark that the expression (32) can be equivalently interpreted as a SNR degradation, given the direct proportionality between the two measures.

REFERENCES

[1] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channel," in *Proc. 25th Asilomar Conf.*, 1991, pp. 856–860.

[2] E. De Carvalho and D. T. M. Slock, "Cramer–Rao bounds for semiblind, blind and training sequence-based channel estimation," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, 1997, pp. 129–132.

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

[4] Z. Chunhui, M. Lin, L. Gang, W. Jing, and L. Weidong, "An order-adaptive filter for channel estimation in DS-CDMA mobile systems," in *Proc. Communication Technology Conf.*, vol. 1, 2000, pp. 355–358.

[5] H.-W. Wang and C.-H. Wei, "Pilot-symbol assisted channel estimation for coherent DS-CDMA communications," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 3, 2000, pp. 383–386.

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

[7] 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.

[8] J. Picheral and U. Spagnolini, "Angles and delay estimation of space-time channels for multiuser detection," *IEEE Trans. Wireless Commun.*, submitted for publication.

[9] Y. F. Chen and M. D. Zoltowski, "Joint angle and delay estimation for DS-CDMA with application to reduced dimension space-time RAKE receivers," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, Mar. 1999, pp. 2933–2936.

[10] M. Z. Win, G. Chrisikos, and N. R. Sollenberger, "Performance of RAKE reception in dense multipath channels: implications of spreading bandwidth and selection diversity order," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1516–1525, Aug. 2000.

[11] H. Holma and A. Toskala, *WCDMA for UMTS*. New York: Wiley, 2000.

[12] S. E. Bensley and B. Aazhang, "Subspace-based channel estimation for code-division multiple-access communication system," *IEEE Trans. Commun.*, vol. 44, pp. 1009–1020, Aug. 1996.

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

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

[15] T. Soderstrom and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[16] 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.

[17] O. Simeone, "Multi-slot space-time channel estimation in mobile communications," Master's thesis, Politec. di Milano, Milan, Italy, 2001.

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

[19] Y. Hua, M. Nikpour, and P. Stoica, "Optimal reduced-rank estimation and filtering," *IEEE Trans. Signal Processing*, vol. 49, pp. 457–468, Mar. 2001.

[20] S. Verdu, *Multiuser Detection*. Cambridge, U. K.: Cambridge Univ. Press, 1998.

[21] J.-T. Chen, J. Kim, and J.-W. Liang, "Multichannel MLSE equalizer with parametric FIR channel identification," *IEEE Trans. Veh. Technol.*, vol. 48, Nov. 1999.

[22] G. K. Kaleh, "Channel equalization for block transmission systems," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 110–121, Jan. 1995.

[23] B. Steiner and P. Jung, "Optimum and suboptimum channel estimation for the uplink of CDMA mobile radio system with joint detection," *Eur. Trans. Telecommun. Related Technol.*, vol. 5, pp. 39–50, Jan.–Feb. 1994.

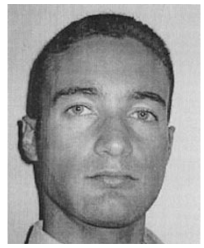
[24] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Processing*, vol. 49, pp. 87–90, Jan. 2001.

[25] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[26] M. Nicoli, O. Simeone, and U. Spagnolini, "Multislot estimation of fast-varying space–time channels in TD-CDMA systems," *IEEE Commun. Lett.*, vol. 6, pp. 376–378, Sept. 2002.

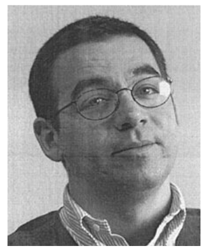
Monica Nicoli (M'99) received the M.Sc. degree (with honors) and the Ph.D. degree in telecommunication engineering from Politecnico di Milano, Milan, Italy, in 1998 and 2002, respectively.

From March to August 2001, she was a Visiting Researcher at the Signals and Systems Group, Uppsala University, Uppsala, Sweden. Currently, she is with the Dipartimento di Elettronica e Informazione, Politecnico di Milano, Milan, Italy. Her research interests include antenna array processing for wireless communication systems, channel estimation and equalization for multiuser systems, 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.

From February to September 2002, he was a Visiting Researcher at the Center for Communications and Signal Processing Research, New Jersey Institute of Technology, Newark. He is currently working toward the Ph.D. degree at Politecnico di Milano, Milan, Italy. He holds a patent on the work developed for his Master thesis. His current research interests lie in the field of signal processing for digital communications, with emphasis on MIMO systems, multicarrier modulation, and channel estimation.



Umberto Spagnolini (M'99–SM'03) 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 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 was awarded the AEI Award (1991), Van Weelden Award of EAGE (1991), and the Best Paper Award from EAGE (1998).