Two-stage Non-data Aided Adaptive Linear Receivers for DS/CDMA

Giuseppe Caire*

First submission: February 1999

Revised: January 2000

Abstract

Closed-form steady-state performance analysis of the signal-to-interference plus noise ratio (SINR) at the output of well-known adaptive implementations of the linear minimum mean-square error receiver for DS/CDMA show that non-data aided schemes may suffer from a considerable performance degradation with respect to their data-aided counterparts. Motivated by this fact, we propose a new two-stage non-data aided scheme where symbol-by-symbol pre-decisions at the output of a first adaptive stage are used to train a second stage. We derive closed-form steady-state performance analysis for both the two-stage and classical decision-directed schemes, taking into account detection errors in decision-directed adaptation. Our analysis shows that the SINR of the two-stage algorithm is close to optimal over a large range of values, while the SINR of the decision-directed scheme is far from optimal when the optimal SINR is small. Finally, we consider the case of time-varying fading channels. We derive modified RLS and LMS adaptation schemes by considering SINR maximization rather than mean-square error minimization (that is useless under the assumption of zero-mean random channels). The resulting two-stage receiver shows good tracking properties in heavy near-far conditions (at least for moderate normalized Doppler bandwidth), while the decision-directed receiver may easily loose tracking after deep fades.

Keywords: CDMA, adaptive algorithms, linear receivers.

1 Introduction

Multiuser detection has been a fruitful and rapidly growing research field for the last decade. Broadly speaking, this is motivated by the fact that the techniques developed for single-user communications,

*Giuseppe Caire is with the Mobile Communications Department, Institut EURECOM, Sophia-Antipolis, France. E-mail: giuseppe.caire@eurecom.fr.

mostly devoted to combat Gaussian white noise, fail to give near-optimal performance if used in the presence of multiple-access interference (MAI). Under the common name of *multiuser detection* we find a broad range of receivers differing in complexity and performance (see [1] for a complete survey and a comprehensive list of references). We can distinguish between *centralized* and *decentralized* receivers. Centralized receivers make use of side information about all interfering users (spreading sequences, timing and propagation channels). They are suited for base-station processing (uplink), where all this side information is either available or can be estimated consistently. Decentralized receivers exploit the knowledge of the spreading sequence, of the timing and of the propagation channel of the user of interest only. Remarkably, this is the same information necessary for a conventional single-user matched filter (SUMF) that ignores the presence of MAI. These receivers treat the superposition of MAI and background Gaussian noise as a random process, the statistics of which must be learned from the received signal, via some adaptive algorithm. In this paper, we are concerned with decentralized adaptive *linear receivers*, i.e., receivers formed by the concatenation of an adaptive linear filter with a suitable (non-linear) detection operation acting on the filter output.

In Section 2 we define a general discrete-time signal model for DS/CDMA and we review background results on adaptive linear receivers. We consider only simple algorithms of the LMS or RLS type [2] (recent algorithms based on subspace tracking [3] are not treated, even though they might be good alternatives). We distinguish between data-aided (DA) and non-data aided (NDA) adaptive algorithm, depending on whether the adaptation rule makes use of known data symbols (training sequence) or does not. [1]

In Section 3 we present closed form formulas for the steady-state signal-to-interference plus noise ratio (SINR) at the output of the adaptive receivers of Section 2. Our formulas generalize the results of [5]. Also, we derive a Gaussian approximation for the steady-state bit-error rate (BER) with 4PSK and Gray mapping. As confirmed by simulations (see Section 6), this approximation is very accurate and offers a simple tool to predict the performance of adaptive receivers. DA algorithm suffer from a SINR degradation of at most 3 dB with respect to optimum. On the contrary, NDA algorithms might be very far from optimum, especially when the optimal SINR is large, i.e., just in the case where the potential gain of linear multiuser receivers over the SUMF is large.

This observation motivates us to look for NDA algorithms that recover this performance loss,

[1] Algorithms that do not require a training sequence are referred to in the literature under different names, such as "blind" [4] or "code-aided" [5].

without renouncing to the simplicity of LMS or RLS. In Section 4, we consider a modified mean-square error (MSE) cost function that allows us to develop in a unified manner the steady-state analysis of decision-driven (DD) adaptive algorithms and of a new family of NDA algorithms referred to as "two-stage" algorithms. The proposed new algorithms have almost the same steady-state performance of DD algorithms, without requiring initial training. Moreover, their steady-state SINR is close to optimum over a range much wider than DD. Hence, it can be expected that two-stage schemes are better suited than DD schemes to track time-varying channels in deep fades.

This fact is confirmed by the results of Section 5, where we consider *moderately* time-varying frequency-selective fading channels. In this case, the standard minimum MSE criterion may be useless [6, 7]. Hence, we formulate an optimal filter design problem based on the maximization of the output SINR and we derive *modified* LMS and RLS algorithms approximating adaptively the SINR-maximizing filter. These algorithms can be seen as a non-trivial generalization of the scheme of [6], developed for a frequency-flat channel, to the frequency-selective case. They require channel estimation for the user of interest only, which is no more than what is required by a SUMF (usually approximated by a coherent rake receiver [8, 9]) and by the *pre-combining* adaptive receiver (referred to as "LMMSE-rake") of [10].

Numerical results show that the two-stage receiver with modified LMS/RLS adaptation is able to recover from deep fades even in heavy near-far conditions, while DD adaptation (also with modified LMS/RLS adaptation) is not. Also, our scheme requires only two adaptive algorithms to be run in parallel, while the LMMSE-rake requires one adaptive filter *per path*, and can be considerably more complex [10]. Then, the proposed scheme represents a good option from both the performance and the complexity point of views, at least for moderate (normalized) Doppler bandwidth of the channel.

2 Background

In this section we introduce a baseband equivalent signal model for DS/CDMA transmission over frequency-selective moderately time-varying channels and we recall some well-known facts about linear receivers and their adaptive implementation.

2.1 Discrete-time finite-memory signal model

Consider a system with K users. The k-th user signal is given by

$$u_k(t) = \sum_m b_k[m] s_k(t - mT) \tag{1}$$

where $s_k(t)$ and $b_k[m]$ are the signature waveform and the m-th information symbol of user k, respectively. Users transmit individually and mutually uncorrelated sequences of unit-variance zero-mean complex symbols. In DS/CDMA, the signature waveforms are given by $s_k(t) = \sum_{\ell=0}^{L-1} s_{k,\ell}\psi(t - \ell T_c)$, where $\mathbf{s}_k = (s_{k,0}, \ldots, s_{k,L-1})^T$ is the k-th user spreading sequence, $T_c = T/L$ is the chip interval, L is the processing gain and $\psi(t)$ is the chip pulse, common to all users, bandlimited in $[-W/2, W/2]$ and with normalized energy 1.

User k transmits with delay $\tau_k = q_k/W + \gamma_k$ (q_k integer and $0 \leq \gamma_k < 1/W$) through a channel with baseband equivalent time-varying impulse response $c_k(t;\tau)$ [9]. We assume that the channel Doppler bandwidth B_d and the signal bandwidth W satisfy $B_d T \ll 1 \ll WT$. Then, the signal bandwidth expansion (Doppler spread) can be safely neglected and the channel output can be sampled at rate W. From the sampling theorem, we can write the signal contribution of user k at the receiver in the form

$$v_k(t) = \sum_i \left[\sum_j c_k[i;j] u_k((i - j - q_k)/W) \right] \mathrm{sinc}(Wt - i) \tag{2}$$

where $\mathrm{sinc}(t) = \sin(\pi t)/(\pi t)$ and where the coefficients of the resulting time-varying discrete-time channel impulse response are given by

$$c_k[i;j] = \int c_k(i/W; j/W - \gamma_k - \tau)\mathrm{sinc}(W\tau)d\tau \tag{3}$$

The overall received signal, given by the superposition of all users' signals plus background noise, is given by $y(t) = \sum_{k=1}^K v_k(t) + \nu(t)$, where $\nu(t)$ is a white circularly-symmetric complex Gaussian process with power spectral density N_0.

The baseband receiver front-end is formed by an ideal lowpass filter with bandwidth $[-W/2, W/2]$ and gain $1/\sqrt{W}$ followed by sampling at rate W with arbitrary sampling epoch. For simplicity, we assume an integer number of samples per chip $N_c = WT_c$ and, without loss of generality, we restrict the integer part of the delays to satisfy $q_k \in [-LN_c/2, LN_c/2)$. In order to obtain an approximated finite-memory signal model, we assume that $c_k[i;j]$ and $\psi(j/W)$ are negligible for $j \notin [0, P-1]$ (for all i) and for $j \notin [-Q, Q]$, respectively, where P and Q are suitable integers. Moreover, we constrain the receiver to have a finite-length *processing window*, i.e., for each symbol time n it processes a window of samples with indexes $i \in [nLN_c - M_1, nLN_c + M_2]$ centered around the n-th symbol interval. The *processing window* size $M = M_1 + M_2 + 1$ is left as a design parameter and it may span more than one symbol interval. We let $y[i]$ denote the sample of $y(t)$ at instant i/W *after lowpass filtering* and form the n-th channel output vector $\mathbf{y}[n]$ as the content of the receiver processing window at symbol time n, i.e.,

$$\mathbf{y}[n] = (y[nLN_c + M_2], y[nLN_c + M_2 - 1], \dots, y[nLN_c - M_1])^T$$

Under the condition $B_d T \ll 1$, it is realistic to assume that the $c_k[i;j]$'s remain almost constant over the time interval spanned by the receiver processing window. Hence, we can consider $c_k[nLN_c + i; j] \approx c_k[nLN_c; j]$ for all $i = -M_1, \dots, M_2$ and represent the k-th channel impulse response during the n-th symbol interval by the vector

$$\mathbf{c}_k[n] = (c_k[nLN_c; 0], \dots, c_k[nLN_c; P-1])^T \tag{4}$$

From (1),(2) and (4), after some straightforward algebra, we can write the n-th channel output vector as

$$\mathbf{y}[n] = \sum_{k=1}^{K} \sum_{m=-B_1}^{B_2} \mathbf{S}_k[m]\mathbf{c}_k[n] b_k[n-m] + \boldsymbol{\nu}[n] \tag{5}$$

where $\boldsymbol{\nu}[n] \sim \mathcal{N}_{\mathcal{C}}(\mathbf{0}, N_0\mathbf{I})$ is the corresponding vector of noise samples [2] and where the matrices $\{\mathbf{S}_k[m] : m = -B_1, \dots, B_2\}$, of size $M \times P$, are uniquely defined by q_k and $s_k(t)$, and have (i,j)-th element given by

$$[\mathbf{S}_k[m]]_{i,j} = \frac{1}{\sqrt{W}} s_k((mLN_c - q_k + M_2 - i - j)/W) \tag{6}$$

for $i = 0, \dots, M-1$ and $j = 0, \dots, P-1$. The summation limits B_1 and B_2 are obtained by noticing that $\mathbf{S}_k[m]$ is not identically zero if and only if $-B_1 \le m \le B_2$, where

$$\begin{aligned} B_1 &= \lfloor (M_2 + Q + LN_c/2)/(LN_c) \rfloor \\ B_2 &= \lfloor (M_1 + Q + P + 3LN_c/2 - N_c - 1)/(LN_c) \rfloor \end{aligned} \tag{7}$$

Clearly, depending on the particular value of q_k, $\mathbf{S}_k[m]$ might be zero for some $m \in [-B_1, B_2]$. Then, each user contributes with *at most* $B = B_1 + B_2 + 1$ symbols to the vector $\mathbf{y}[n]$.

[2] $\mathcal{N}_{\mathcal{C}}(\boldsymbol{\mu}, \mathbf{R})$ denotes the circularly-symmetric complex multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{R}$.

It is useful to renumber the user symbols as $b_k[n-m] \to b_u[n]$ where $u = (k-1)B - m + 1$ for $m = -B_1, \ldots, 0$ and $u = (k-1)B + m + B_1 + 1$ for $m = 1, \ldots, B_2$, and to define the modified normalized spreading sequences $\mathbf{p}_u[n] = \mathbf{S}_k[m]\mathbf{c}_k[n]/\sqrt{\mathcal{E}_u[n]}$ where $\mathcal{E}_u[n] = |\mathbf{S}_k[m]\mathbf{c}_k[n]|^2$ is the energy contribution of symbol $b_u[n]$ to the total signal energy in the receiver processing window (if $\mathcal{E}_u[n] = 0$, we let $\sqrt{\mathcal{E}_u[n]}\mathbf{p}_u[n] = \mathbf{0}$ even though, strictly speaking, $\mathbf{p}_u[n]$ is not defined). Finally, we can rewrite (5) as

$$\mathbf{y}[n] = \sum_{u=1}^{U} \sqrt{\mathcal{E}_u[n]}\mathbf{p}_u[n]b_u[n] + \boldsymbol{\nu}[n] \tag{8}$$

where $U = BK$. Notice that, for the sake of symbol-by-symbol detection, (8) is equivalent to a synchronous system with U users.

2.2 Linear decentralized receivers

We focus on the *decentralized linear detection* of user 1. The receiver is formed by a linear (time-varying) FIR filter $\mathbf{h}[n]$ with output $z[n] = \mathbf{h}[n]^H\mathbf{y}[n]$ followed by a suitable (non-linear) processing of the filter output sequence $\{z[n]\}$. Receiver options are distinguished by different choices of the filter vector $\mathbf{h}[n]$ and of the non-linear processing. In the case of uncoded transmission, non-linear processing reduces to the simple symbol-by-symbol detection $\widehat{b}_1[n] = \mathrm{dec}(z[n])$, where $\mathrm{dec}(\cdot)$ denotes a decision rule based on the single filter output $z[n]$.

In the following, we consider the user channels as possibly unknown constant deterministic vectors and we drop the time index n for the sake of notation simplicity. The case of time-varying random channels will be examined in Section 5. Without loss of generality, we assume that $\mathcal{E}_1 \geq \mathcal{E}_u$ for all $u = 1, \ldots, B$ (if this is not the case, it is sufficient to renumber user 1 symbols), and we identify b_1 as the desired symbol. Thus, we can rewrite (8) by putting in evidence the useful signal component, as

$$\mathbf{y} = \sqrt{\mathcal{E}_1}\mathbf{p}_1 b_1 + \mathbf{w} \tag{9}$$

where $\mathbf{w}$ collects noise+ISI+MAI. A relevant measure of performance for the filter $\mathbf{h}$ is its output SINR. In our setting, this is defined by

$$\mathrm{SINR} \triangleq \frac{E[|\mathbf{h}^H(\mathbf{y}-\mathbf{w})|^2]}{E[|\mathbf{h}^H\mathbf{w}|^2]} \tag{10}$$

In the following, we review some well-known results on linear receivers.

Single-user matched filter (SUMF). The baseline linear receiver is the SUMF $\mathbf{h} = \mathbf{p}_1$, matched to the useful signal component as if $\mathbf{w}$ was a white noise vector. The SUMF requires the knowledge of user 1 signature waveform $s_1(t)$, coarse timing q_1 and channel vector $\mathbf{c}_1$. Remarkably, all decentralized receivers considered in this paper require no more side information than the SUMF (actually, some require less).

The output SINR achieved by the SUMF is given by $\text{SINR}_{\text{sumf}} = \mathcal{E}_1/(\mathbf{p}_1^H \mathbf{R}_w \mathbf{p}_1)$ where $\mathbf{R}_w = E[\mathbf{w}\mathbf{w}^H]$. In the absence of ISI and MAI, $\mathbf{R}_w = N_0 \mathbf{I}$ and $\text{SINR}_{\text{sumf}} = \mathcal{E}_1/N_0 \stackrel{\Delta}{=} \text{SNR}_1$. [3]

Linear minimum MSE receiver (LMMSER). A classical criterion for the design of the filter $\mathbf{h}$ is the minimization of the MSE [11]

$$J \stackrel{\Delta}{=} E[|b_1 - \mathbf{h}^H \mathbf{y}|^2] \tag{11}$$

The minimum MSE (MMSE) filter vector is the Wiener filter

$$\mathbf{h}_{\text{opt}} = \sqrt{\mathcal{E}_1} \mathbf{R}_y^{-1} \mathbf{p}_1 \tag{12}$$

where $\mathbf{R}_y = E[\mathbf{y}\mathbf{y}^H]$. The resulting output SINR is given by

$$\text{SINR}_{\text{opt}} = \mathcal{E}_1 \mathbf{p}_1^H \mathbf{R}_w^{-1} \mathbf{p}_1 \tag{13}$$

and it is the maximum SINR over all possible filters $\mathbf{h}$ [11] (this motivate the subscript "opt"). Notice that any two filter vectors $\mathbf{h}'$ and $\mathbf{h}''$ which differ by a scalar (non-zero) multiplicative term provide the same SINR (we shall write $\mathbf{h}' \propto \mathbf{h}''$). Then, any filter $\mathbf{h} \propto \mathbf{h}_{\text{opt}}$ is also optimal in terms of output SINR.

Adaptive implementations of the LMMSER are obtained from standard DA LMS and RLS algorithms [2]. Let $\mathbf{h}[n]$ be the filter vector at time n. The DA-LMS algorithm is given by:

$$\mathbf{h}[n] = \mathbf{h}[n-1] + \mu e[n]^* \mathbf{y}[n] \tag{14}$$

where $e[n] = b_1[n] - \mathbf{h}[n-1]^H \mathbf{y}[n]$ and $\mu > 0$ is the step-size. The DA-RLS algorithm [4] updates the estimate of the inverse covariance matrix by

$$\begin{aligned} \mathbf{k}[n] &= \left(\alpha + \mathbf{y}[n]^H \mathbf{M}[n-1]\mathbf{y}[n]\right)^{-1} \mathbf{M}[n-1]\mathbf{y}[n] \\ \mathbf{M}[n] &= \frac{1}{\alpha}(\mathbf{I} - \mathbf{k}[n]\mathbf{y}[n]^H)\mathbf{M}[n-1] \end{aligned} \tag{15}$$

[3] In general, we define SNR_k (the SNR of user k) as the maximum SINR achieved by the SUMF matched to user k in the absence of ISI and MAI, over all choices of the delay q_k, and for given N_0 channel vectors and spreading sequences.

[4] Here we consider only the straightforward transversal-filter RLS. See [2] for a more complete survey of RLS-type algorithms.

and the filter vector by

$$\mathbf{h}[h] = \mathbf{h}[n-1] + e[n]^* \mathbf{k}[n] \tag{16}$$

where $0 < \alpha < 1$ is the exponential forgetting factor and $\mathbf{M}[0] = \gamma \mathbf{I}$ for some $\gamma > 0$.

Constrained minimum MOE receiver (CMMOER). In [4], the receiver filter $\mathbf{h}$ is designed in order to minimize the mean output energy (MOE)

$$\xi \stackrel{\Delta}{=} E[|z|^2] = \mathbf{h}^H \mathbf{R}_y \mathbf{h} \tag{17}$$

subject to the constraint $\mathbf{h}^H \mathbf{p}_1 = 1$. The solution of this constrained minimization problem is readily obtained as

$$\mathbf{h}_{\text{moe}} = \xi_{\min} \mathbf{R}_y^{-1} \mathbf{p}_1 \tag{18}$$

where

$$\xi_{\min} = \left(\mathbf{p}_1^H \mathbf{R}_y^{-1} \mathbf{p} \right)^{-1} = \mathcal{E}_1 \left(1 + \frac{1}{\text{SINR}_{\text{opt}}} \right) \tag{19}$$

is the constrained minimum MOE. Since $\mathbf{h}_{\text{moe}} \propto \mathbf{h}_{\text{opt}}$, also the CMMOER attains SINR_{opt}.

Adaptive implementations of the CMMOER are obtained by using the NDA LMS and RLS algorithms described in [4, 5]. The NDA-LMS algorithm is based on the canonical decomposition $\mathbf{h}[n] = \mathbf{p}_1 + \mathbf{v}[n]$, where $\mathbf{v}[n]$ is constrained to be orthogonal to $\mathbf{p}_1$. In this way, the constraint $\mathbf{h}[n]^H \mathbf{p}_1 = 1$ is automatically satisfied. Then, $\mathbf{v}[n]$ is adapted according to [4]:

$$\mathbf{v}[n] = \mathbf{v}[n-1] - \mu z[n]^* \left(\mathbf{I} - \mathbf{p}_1 \mathbf{p}_1^H \right) \mathbf{y}[n] \tag{20}$$

where $z[n] = \mathbf{h}[n-1]^H \mathbf{y}[n]$.

The NDA-RLS algorithm is given by [5]:

$$\begin{aligned} \xi[n] &= \left(\mathbf{p}_1^H \mathbf{M}[n] \mathbf{p}_1 \right)^{-1} \\ \mathbf{h}[n] &= \xi[n] \mathbf{M}[n] \mathbf{p}_1 \end{aligned} \tag{21}$$

where $\mathbf{M}[n]$ is obtained by (15).

Generalized constrained minimum MOE receiver (GCMOER). An elegant generalization of the CMMOER which avoids explicit knowledge of $\mathbf{p}_1$ has been proposed in [12, 13]. This receiver, referred here as the GCMOER, is the result of the min-max problem: choose $(\mathbf{h}, \mathbf{g})$ such that the MOE (17) is minimized with respect to $\mathbf{h}$ subject to the constraint $\mathbf{S}_1[0]^H \mathbf{h} = \mathbf{g}$ and maximized with respect to $\mathbf{g}$ subject to the constraint $|\mathbf{g}|^2 = 1$. The resulting filter vector is given by [12]

$$\mathbf{h}_{\text{gmoe}} = \xi_1 \mathbf{R}_y^{-1} \mathbf{S}_1[0] \mathbf{u}_1 \tag{22}$$

where $1/\xi_1$ is the minimum eigenvalue of the matrix $\mathbf{S}_1[0]^H\mathbf{R}_y^{-1}\mathbf{S}_1[0]$ and $\mathbf{u}_1$ is a corresponding unit-norm eigenvector. By using (22) into (17), it is easy to show that the resulting MOE is given by $\xi = \xi_1$. The GCMOER is well-defined if the minimum eigenvalue of $\mathbf{S}_1[0]^H\mathbf{R}_y^{-1}\mathbf{S}_1[0]$ has multiplicity 1. This holds under the *unique identifiability condition* of [13, Prop. 1]. In the following we assume that this is always the case. The GCMOER needs only the prior knowledge of q_1 and $s_1(t)$ in order to calculate $\mathbf{S}_1[0]$. Therefore, it requires less side information than the SUMF.

The output SINR achieved by the GCMOER is easily obtained as

$$\text{SINR}_{\text{gmoe}} = \frac{1}{\xi_1/(\mathcal{E}_1|\mathbf{u}_1^H\mathbf{c}_1|^2) - 1} \tag{23}$$

In [13] it is shown that, if the GCMOER is well-defined, $\text{SINR}_{\text{gmoe}}$ is close to SINR_{opt} in most cases of interest, so that the GCMOER is only slightly suboptimal.

In this paper we consider the following straightforward NDA adaptive implementation of the GCMOER [14]: [5]

$$\begin{aligned}
\mathbf{u}_1[n] &= \arg\min_{\mathbf{g}} \frac{\mathbf{g}^H\mathbf{S}_1[0]^H\mathbf{M}[n]\mathbf{S}_1[0]\mathbf{g}}{\mathbf{g}^H\mathbf{g}} \\
\xi_1[n] &= \left(\mathbf{u}_1[n]^H\mathbf{S}_1[0]^H\mathbf{M}[n]\mathbf{S}_1[0]\mathbf{u}_1[n]\right)^{-1} \\
\mathbf{h}[n] &= \xi_1[n]\mathbf{M}[n]\mathbf{S}_1[0]\mathbf{u}_1[n]
\end{aligned} \tag{24}$$

where $\mathbf{M}[n]$ is obtained by (15). Because of the similarity between (24) and (21), the above algorithm will be referred to as "generalized" NDA-RLS (GNDA-RLS).

3 Steady-state performance analysis

For i.i.d. symbols and time-invariant channels $\mathbf{y}[n]$ is a wide-sense stationary (WSS) vector process. [6] We assume that, with WSS inputs, the adaptive algorithms described before satisfy the following convergence conditions [2]:

1. Convergence of the mean filter vector: $\lim_{n\to\infty} E[\mathbf{h}[n]] = \overline{\mathbf{h}}$.
2. Convergence of the MSE: $\lim_{n\to\infty} J[n] = J$ (where $J[n] = E[|b_1[n] - \mathbf{h}[n-1]^H\mathbf{y}[n]|^2]$).

[5] See [14, 15] for computationally-efficient stochastic-gradient adaptive GCMOER implementations.

[6] Strictly speaking, $\{\mathbf{y}[i] : i = 1, \ldots, n\}$ cannot be WSS since it starts at time 1. However, we assume that as $n \to \infty$ the transient effect due to *pre-windowing* can be neglected.

The constants $\overline{\mathbf{h}}$ and J depend on the specific algorithm and on the channel parameters (user channels, spreading sequences etc...).

Since we deal with steady-state, all the following expressions should be interpreted as limits for $n \to \infty$. Subject to the above assumptions, we can write $\mathbf{h} = \overline{\mathbf{h}} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is an asymptotically zero-mean WSS error vector [16], commonly assumed independent of b_1 and $\mathbf{y}$. With this *independence assumption* [2], the steady-state MSE can be written as $J = J_0 + J_{\rm ex}$ where $J_0 = E[|b_1 - \overline{\mathbf{h}}^H \mathbf{y}|^2]$ is the MSE achieved by the non-adaptive filter $\overline{\mathbf{h}}$ and $J_{\rm ex} = \mathrm{trace}(\mathbf{R}_y \mathbf{R}_\epsilon)$ is the steady-state excess MSE ($\mathbf{R}_\epsilon = E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^H]$). Closed-form expressions for $J_{\rm ex}$ are known for several adaptive algorithms [2, 17, 4, 5].

3.1 SINR analysis

We derive a general expression of the steady-state output SINR in terms of $\overline{\mathbf{h}}$, J_0 and $J_{\rm ex}$. Then, we evaluate it for the DA-LMS, DA-RLS, NDA-LMS, NDA-RLS and GNDA-RLS algorithms presented in Section 2. In order to account for the random component of the filter vector $\mathbf{h}$, we modify the definition of the SINR given in (10) as

$$\mathrm{SINR} = \frac{E[|\overline{\mathbf{h}}^H(\mathbf{y} - \mathbf{w})|^2]}{E[|\overline{\mathbf{h}}^H \mathbf{w} + \boldsymbol{\epsilon}^H \mathbf{y}|^2]} \tag{25}$$

where the filter error vector contributes to the noise energy. By using the independence assumption and expanding the above expression we can write

$$\begin{aligned} \mathrm{SINR} &= \frac{\mathcal{E}_1 |\overline{\mathbf{h}}^H \mathbf{p}_1|^2}{E[\boldsymbol{\epsilon}^H \mathbf{y}\mathbf{y}^H \boldsymbol{\epsilon}] + \overline{\mathbf{h}}^H \mathbf{R}_w \overline{\mathbf{h}}} \\ &= \frac{\mathcal{E}_1 |\overline{\mathbf{h}}^H \mathbf{p}_1|^2}{\mathrm{trace}(\mathbf{R}_y \mathbf{R}_\epsilon) + E[|b_1 - \overline{\mathbf{h}}^H \mathbf{y}|^2] - |1 - \sqrt{\mathcal{E}_1}(\overline{\mathbf{h}}^H \mathbf{p}_1)|^2} \\ &= \frac{\mathcal{E}_1 |\overline{\mathbf{h}}^H \mathbf{p}_1|^2}{J_0 + J_{\rm ex} - |1 - \sqrt{\mathcal{E}_1}(\overline{\mathbf{h}}^H \mathbf{p}_1)|^2} \end{aligned} \tag{26}$$

DA algorithms. Both the DA-LMS and the DA-RLS have the property that $\overline{\mathbf{h}} = \mathbf{h}_{\rm opt}$ and that $J_{\rm ex} = \eta J_0$, where η is referred to as *MSE misadjustment* [2]. By using this into (26), after some algebra we get

$$\mathrm{SINR}_{\rm DA} = \frac{\mathrm{SINR}_{\rm opt}}{1 + \eta + \eta / \mathrm{SINR}_{\rm opt}} \tag{27}$$

The MSE misadjustment is explicitly given by [2, 17]

$$\eta = \begin{cases} \frac{\sum_{\ell=1}^{M} \frac{\mu\lambda_\ell}{2-\mu\lambda_\ell}}{1-\sum_{\ell=1}^{M} \frac{\mu\lambda_\ell}{2-\mu\lambda_\ell}} & \text{DA-LMS} \\ \\ \frac{1-\alpha}{1+\alpha} M & \text{DA-RLS} \end{cases} \tag{28}$$

where $\{\lambda_\ell : \ell = 1, \ldots, M\}$ are the eigenvalues of $\mathbf{R}_y$.

NDA algorithms. Both the NDA-LMS and the NDA-RLS have the property that $\overline{\mathbf{h}} = \mathbf{h}_{\rm moe}$ and that $J_{\rm ex} = \eta\xi_{\rm min}$ [4, 5], where, in analogy with the terminology of DA algorithms, we refer to η as the MOE misadjustment. By using this into (26), after some algebra we get

$$\text{SINR}_{\rm NDA} = \frac{\text{SINR}_{\rm opt}}{1 + \eta + \eta\text{SINR}_{\rm opt}} \tag{29}$$

Approximated expressions of η are obtained in [5, 4],

$$\eta = \begin{cases} \frac{\sum_{\ell=1}^{M} \frac{\mu p_\ell\lambda_\ell}{2-\mu p_\ell\lambda_\ell}}{1-\sum_{\ell=1}^{M} \frac{\mu p_\ell\lambda_\ell}{2-\mu p_\ell\lambda_\ell}} & \text{NDA-LMS} \\ \\ \frac{1-\alpha}{2\alpha} (M-1) & \text{NDA-RLS} \end{cases} \tag{30}$$

where $\{p_\ell : \ell = 1, \ldots, M\}$ are the diagonal elements of the matrix $\mathbf{Q}^H(\mathbf{I} - \mathbf{p}_1\mathbf{p}_1^H)\mathbf{Q}$, and where $\mathbf{Q}$ is the unitary matrix such that $\mathbf{R}_y = \mathbf{Q}\text{diag}(\lambda_1, \ldots, \lambda_M)\mathbf{Q}^H$.

Next, we consider the more complicated GNDA-RLS algorithm. In [14] it is shown that, as long as the GCMOER is well-defined, the GNDA-RLS has the property that $\overline{\mathbf{h}} = \mathbf{h}_{\rm gmoe}$. The evaluation of $J_{\rm ex}$ for the GNDA-RLS algorithm is complicated by the presence of the eigenvector computation step in the recursion (24). Then, we approximate $J_{\rm ex}$ by the asymptotic excess MSE of the modified recursion obtained from (24) by eliminating the eigenvector computation step and by letting $\mathbf{u}_1[n] = \mathbf{u}_1$ for all n (recall that $\mathbf{u}_1$ is the unit-norm eigenvector corresponding to the minimal eigenvalue of the matrix $\mathbf{S}_1[0]^H\mathbf{R}_y^{-1}\mathbf{S}_1[0]$). This is motivated by the fact that, for large n, the inverse covariance matrix $\mathbf{M}[n]$ behaves like a *quasi-deterministic* quantity when $M(1-\alpha) \ll 1$ (see [5] and references therein). Therefore, $\lim_{n\to\infty} \mathbf{M}[n] \simeq E[\mathbf{M}[n]] = (1-\alpha)\mathbf{R}_y^{-1}$, which implies that, for large n, $\mathbf{u}_1[n] \simeq \mathbf{u}_1$.

The resulting modified recursion is formally equivalent to the NDA-RLS algorithm (21), provided that $\mathbf{p}_1$ is replaced by $\widetilde{\mathbf{p}}_1 = \mathbf{S}_1[0]\mathbf{u}_1$. Then, it is straightforward to duplicate the derivation of [5] with the change $\mathbf{p}_1 \to \widetilde{\mathbf{p}}_1$ and obtain $J_{\rm ex} \simeq \eta\xi_1$, where the MOE misadjustment η is the same of NDA-RLS. By using this into (26), after some algebra we can prove the following:

Proposition 1. *Under the convergence assumption, the steady-state SINR of the adaptive implementation of the GCMOER based on GNDA-RLS is given by*

$$\text{SINR}_{\text{GNDA}} = \frac{\text{SINR}_{\text{gmoe}}}{1 + \eta + \eta \text{SINR}_{\text{gmoe}}} \tag{31}$$

where $\eta = \frac{1-\alpha}{2\alpha}(M-1)$. □

Remark 1. Results (27) and (29) are proved in [5] directly for RLS. The general steady-state SINR formula (26) allows their immediate extension to LMS. Result (31) is new, up to the author's knowledge.

Remark 2. For $\text{SINR}_{\text{opt}} \gg 1$, we have that $\text{SINR}_{\text{DA}} \simeq \text{SINR}_{\text{opt}}/(1+\eta)$. In normal working conditions it is reasonable to expect that the excess MSE due to adaptation does not exceed the MMSE, therefore $0 < \eta < 1$ and DA algorithms at the steady-state are suboptimal by at most 3 dB. On the contrary, $\text{SINR}_{\text{NDA}} \leq 1/\eta$, and might be much less than SINR_{opt}. Notice that the use of LMMSER-type receivers makes sense precisely in the condition of large SINR_{opt}, since if SINR_{opt} is small, then the simpler SUMF would provide about the same performance. Thus, NDA algorithms prove to be poor just in the case where the potential gain of linear multiuser receivers is largest. This negative fact about NDA adaptive algorithms has been noticed in [5] in the case of RLS. The fact that similar expressions holds also for NDA-LMS and for GNDA-RLS induces us to conjecture that the poor steady-state performance of the adaptive CMMOER and GCMOER does not depend on the particular algorithm, but it is a consequence of the constrained MOE cost function defining the receivers.

3.2 Symbol-by-symbol error probability

It is well-known that blind equalization schemes based on second-order statistics are able to equalize the channel up to a phase rotation [18]. This means that, with CMMOER and GCMOER and their adaptive NDA implementations, the filter $\mathbf{h}$ is determined up to a factor $e^{j\phi}$. While this has no impact on the output SINR, it does have an impact on error probability, depending on the detector function dec(·).

By using (9) and $\mathbf{h} = \overline{\mathbf{h}} + \boldsymbol{\epsilon}$, we can write the filter output in the form

$$z = \sqrt{\mathcal{E}_1}(\overline{\mathbf{h}}^H \mathbf{p}_1) b_1 + \zeta \tag{32}$$

where ζ is the residual interference plus noise at the filter output, and takes into account also the effect of the random filter error $\boldsymbol{\epsilon}$. For simplicity, we assume that the phase of the deterministic useful signal component is perfectly known to the receiver and, without loss of generality, we let $\kappa \stackrel{\Delta}{=} \overline{\mathbf{h}}^H \mathbf{p}_1$ be real and positive. In this paper, we assume that the symbols b_u belong to a 4PSK signal set with Gray mapping [9], i.e., $b_u = (d_u^I + j d_u^Q)/\sqrt{2}$, where d_u^I and d_u^Q are i.i.d. antipodal random variables taking on values in $\{\pm 1\}$ with equal probability (the superscripts I and Q denote the in-phase and the quadrature rails), and we consider the following simple suboptimal symbol-by-symbol threshold detection rule:

$$\begin{aligned} \widehat{d_1^I} &= \text{sign}(\text{Re}\{z\}) \\ \widehat{d_1^Q} &= \text{sign}(\text{Im}\{z\}) \end{aligned} \tag{33}$$

We focus on the detection of the in-phase data symbol d_1^I (an analogous derivation applies to the quadrature symbols). Because of symmetry, we can assume $d_1^I = 1$. The decision variable in (33) can be written as

$$\text{Re}\{z\} = \kappa \sqrt{\frac{\mathcal{E}_1}{2}} \left(1 + \beta_0 + \sum_{i=1}^{2U-1} \beta_i d_i + \widetilde{\nu} \right) \tag{34}$$

where the d_i's are i.i.d., uniformly distributed over $\{\pm 1\}$, $\widetilde{\nu} \sim \mathcal{N}(0, N_0 |\mathbf{h}|^2/(\mathcal{E}_1 \kappa^2))$, $\beta_0 = \text{Re}\left\{ \boldsymbol{\epsilon}^H \mathbf{p}_1 \mathbf{p}_1^H \overline{\mathbf{h}} \right\} / \kappa^2$, $\beta_1 = \text{Im}\left\{ \boldsymbol{\epsilon}^H \mathbf{p}_1 \mathbf{p}_1^H \overline{\mathbf{h}} \right\} / \kappa^2$ and

$$\beta_i = \begin{cases} \sqrt{\mathcal{E}_u/\mathcal{E}_1} \text{Re}\left\{ \mathbf{h}^H \mathbf{p}_u \mathbf{p}_1^H \overline{\mathbf{h}} \right\} / \kappa^2 & i = 2u-2 \\ \sqrt{\mathcal{E}_u/\mathcal{E}_1} \text{Im}\left\{ \mathbf{h}^H \mathbf{p}_u \mathbf{p}_1^H \overline{\mathbf{h}} \right\} / \kappa^2 & i = 2u-1 \end{cases} \tag{35}$$

for $u = 2, \ldots, U$. The β_i's are random variables, since they are functions of $\boldsymbol{\epsilon}$. The BER conditioned on $\boldsymbol{\epsilon}$ is immediately obtained as [19]

$$P(e|\boldsymbol{\epsilon}) = \frac{1}{2^{2U-1}} \sum_{d_1,\ldots,d_{2U-1}} Q\left(\sqrt{\frac{\mathcal{E}_1 \kappa^2}{N_0 |\mathbf{h}|^2}} \left(1 + \beta_0 + \sum_{i=1}^{2U-1} \beta_i d_i \right) \right) \tag{36}$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ is the Gaussian distribution tail function.

Methods for the evaluation of (36) have been extensively studied in the framework of ISI channels, for which the channel output samples is formally analogous to (34) for fixed (i.e., non-random) coefficients β_i. Straightforward direct averaging with respect to the d_i has exponential complexity in U. A family of loose upper bounds are provided in [20] and a very efficient and numerically accurate method based on Discrete-Cosine Transform (DCT) is provided in [21].

In order to compute the steady-state $P(e)$ we should average (36) with respect to the steady-state distribution of the filter error vector $\boldsymbol{\epsilon}$. This appears to be prohibitively complex, since this distribution is not known exactly. The desired steady-state error probability can be evaluated by the semi-analytic Monte Carlo (MC) method:

$$P(e) \approx \frac{1}{N} \sum_{r=1}^{N} P(e|\boldsymbol{\epsilon}^{(r)}) \tag{37}$$

where $\boldsymbol{\epsilon}^{(r)}$ is the error vector resulting from the adaptive algorithm after a sufficiently large number of iterations and where the index r runs over N independent simulations of the algorithm. Each simulation run has the same channel parameters (SNRs, delays, spreading sequences and channel responses) but independently generated user data and noise sequences. The above approach is clearly very expensive in terms of computation.

As an efficient alternative to the MC method, we propose a steady-state Gaussian approximation (SSGA) consisting of modeling the residual noise variable ζ in (32) as a Gaussian zero-mean random variable. It is easy to check that $\kappa^2 \mathcal{E}_1 / E[|\zeta|^2]$ is equal to the steady-state SINR of the adaptive filter. Therefore we can write

$$P(e) \approx Q(\sqrt{\text{SINR}}) \tag{38}$$

where SINR is provided by (27), (29) and (31), for the algorithms considered here.

4 Decision-driven and two-stage NDA adaptive schemes

In this section we generalize of the classical minimum MSE problem by substituting to the desired symbol sequence $\{b_1[n]\}$ an auxiliary symbol sequence $\{\widetilde{b}_1[n]\}$ such that, for all n, $E[\widetilde{b}_1[n]] = 0$, $E[|\widetilde{b}_1[n]|^2] = 1$ and $E[b_1[n]\widetilde{b}_1[n]^*] = \rho$, where ρ is a given (complex) correlation coefficient. This allows us to derive and study decision-driven (DD) and a new class of improved NDA adaptive algorithms, referred to as *two-stage* NDA, in a unified manner.

We consider the minimization with respect to $\mathbf{h}$ of the modified MSE

$$\widetilde{J} = E[|\widetilde{b}_1 - \mathbf{h}^H \mathbf{y}|^2] \tag{39}$$

The resulting filter vector is given by

$$\widetilde{\mathbf{h}} = \rho \sqrt{\mathcal{E}_1} \mathbf{R}_y^{-1} \mathbf{p}_1 = \rho \mathbf{h}_{\text{opt}} \tag{40}$$

For all $\rho \neq 0$, we have that $\widetilde{\mathbf{h}} \propto \mathbf{h}_{\text{opt}}$. Then, $\widetilde{\mathbf{h}}$ achieves the same optimal output SINR of the LMMSER. LMS-type and RLS-type adaptive algorithms approximating $\widetilde{\mathbf{h}}$ are immediately obtained by using $\widetilde{b}_1[n]$ instead of $b_1[n]$ as training symbol in the standard DA LMS and RLS recursions (14) and (16). The resulting algorithms will be referred to as the *modified* LMS and RLS, respectively. By choosing appropriately the auxiliary sequence $\{\widetilde{b}_1[n]\}$, we can derive several adaptive schemes.

Decision-driven (DD) algorithms. DD adaptation can be interpreted in the general framework of modified LMS and RLS with the choice of the auxiliary sequence

$$\widetilde{b}_1[n] = \left\{ \begin{array}{ll} b_1[n] & \text{for} \quad n = 1, \ldots, T_r \\ \widehat{b}_1[n] & \text{for} \quad n > T_r \end{array} \right. \tag{41}$$

where T_r is the length of an initial training sequence and where $\widehat{b}_1[n] = \text{dec}(z[n])$ is the n-th symbol-by-symbol decision.

In order to improve the poor steady-state performance of NDA algorithms, several papers suggest dual-mode adaptive receivers starting with NDA adaptation and switching to DD adaptation as soon as the output SINR is sufficiently good for making reliable symbol-by-symbol decisions. Under standard convergence assumptions, the steady-state performance of the dual-mode receiver is the same of the DD receiver with initial training. Hence, we shall not distinguish between these two cases (the only difference being that the dual-mode receiver does not require the transmission of an initial training sequence).

Two-stage NDA algorithms. We propose a new class of NDA adaptive receivers formed by the serial concatenation of two adaptive stages. The first stage is based on any NDA adaptive algorithm (e.g., the NDA-LMS, NDA-RLS or GNDA-RLS described in Section 2). The second stage is based on modified LMS or RLS with auxiliary sequence

$$\widetilde{b}_1[n] = \text{dec}(z_1[n]) \tag{42}$$

where $z_1[n]$ is the output of the first stage. As for dual-mode receivers, no initial training sequence is required.

4.1 Steady-state SINR analysis

Subject to the convergence assumptions of Section 3, we let $\overline{\mathbf{h}}, J_0$ and J_{ex} denote the asymptotic (for large n) filter mean vector, the corresponding MSE and excess MSE of the modified LMS and RLS algorithms, respectively. The steady-state output SINR has the general form (26). In Appendix A, we show that for the modified LMS and RLS algorithms we have

$$\begin{aligned} \overline{\mathbf{h}} &= \widetilde{\mathbf{h}} \\ J_0 &= 1-(1-|1-\rho|^2)\chi \\ J_{\rm ex} &= \eta\left(1-|\rho|^2\chi\right) \end{aligned} \tag{43}$$

where $\widetilde{\mathbf{h}}$ is given in (40), where η is the MSE misadjustment of conventional DA LMS and RLS algorithms and where we let $\chi \triangleq \text{SINR}_{\rm opt}/(1+\text{SINR}_{\rm opt})$. By using (43) into (26), after some algebra we are able to write the steady-state SINR of the modified LMS and RLS algorithms as

$$\begin{aligned} \text{SINR} &= \frac{|\rho|^2\chi^2}{|\rho|^2\chi(1-\chi)+\eta(1-|\rho|^2\chi)} \\ &\triangleq F(\eta,\rho,\chi) \end{aligned} \tag{44}$$

The function $F(\eta,\rho,\chi)$ is increasing in $0<|\rho|\leq 1$, and strictly decreasing in $\eta \geq 0$, for all $0<\chi\leq 1$. Its maximum is attained for $\eta=0$ and any $\rho\neq 0$, and it is given by $F(0,\rho,\chi)=\text{SINR}_{\rm opt}$. This corresponds to the fact that, as noticed previously, $\widetilde{\mathbf{h}} \propto \mathbf{h}_{\rm opt}$ for any auxiliary variables $\widetilde{b}_1[n]$ having non-zero correlation with the true desired data variables $b_1[n]$, so that the deterministic (i.e., non-adaptive) filter $\widetilde{\mathbf{h}}$ achieves the optimal SINR. For $\rho=1$ we have $F(\eta,1,\chi)=\text{SINR}_{\rm DA}$, given in (27). In fact, for $\rho=1$ (that implies $\widetilde{b}_1[n]=b_1[n]$ in the mean-square sense), the modified and the conventional DA algorithms coincide.

Steady-state SINR of DD algorithms. We assume that users' symbols belong to a 4PSK signal set with Gray mapping and we consider the threshold detector (33) with ideal phase compensation. Then, we can use the SSGA and approximate the steady-state BER by $P(e)\approx Q(\sqrt{\text{SINR}})$. Implicitly, errors in the detection of the in-phase and quadrature bits are assumed to be statistically independent, since the SSGA treats the residual interference plus noise at the filter output as circularly-symmetric complex Gaussian noise. The resulting correlation coefficient is given by

$$\rho \approx 1-2Q(\sqrt{\text{SINR}}) \tag{45}$$

(45) and (44) form a system of two equations in the unknowns ρ and SINR. By eliminating ρ, we obtain the steady-state SINR of DD algorithms as the positive solution of the equation

$$x = F(\eta, 1-2Q(\sqrt{x}),\chi) \tag{46}$$

It is immediate to see that (46) is always verified for $x = 0$. Moreover, if

$$\left. \frac{d}{dx} F(\eta, 1 - 2Q(\sqrt{x}), \chi) \right|_{x=0} > 1$$

there exists a unique positive solution in the interval $(0, \text{SINR}_{\text{DA}}]$.

Steady-state SINR of two-stage algorithms. We make the same assumption of 4PSK with Gray mapping, and we consider the threshold detector (33) with ideal phase compensation at the output of the first stage. Then, we use the SSGA to approximate the steady-state BER of the first stage and we obtain ρ as

$$\rho \approx 1 - 2Q(\sqrt{\text{SINR}_{\text{1st}}})$$

where SINR_{1st} is the steady-state SINR of the NDA algorithm of stage 1. The steady-state SINR of the second stage is obtained from (44) as

$$\text{SINR}_{\text{2nd}} = F(\eta, 1 - 2Q(\sqrt{\text{SINR}_{\text{1st}}}), \chi) \tag{47}$$

For NDA-LMS, NDA-RLS and GNDA-RLS, we can use explicit expressions of SINR_{1st} given in (29) and (31).

Remark 3. The effect of detection errors in the DD loop and at the output of the first stage of the two-stage scheme is taken into account by the correlation coefficient ρ. For BER equal to 0.5 the desired sequence $b_1[n]$ and the auxiliary sequence $\widetilde{b}_1[n]$ are uncorrelated ($\rho = 0$) and the modified adaptive algorithms become useless. This effect is quite different from "error propagation" of *multistage* and *decision-feedback* multiuser detection [1]. In fact, errors in the sequence $\widetilde{b}_1[n]$ affect the detection of the data symbols only through a mismatch of the filter vector $\mathbf{h}[n]$. Simulations (see Section 6) show that the asymptotic BER of DD and two-stage algorithms can be accurately predicted by using the SSGA with the steady-state SINR expressions provided by (46) and by (47), respectively. Thus, the effect of errors in the auxiliary sequence is fully accounted for by our analysis.

Remark 4. The main advantage of the two-stage receiver is that the outputs of both stages are always available. Therefore, it is very easy to measure their SINR and select the best. This is not the case with a dual-mode receiver with a single adaptive filter switching between NDA and DD adaptation modes. In this case, the output SINR must be compared with an empirical threshold in order to determine which adaptation mode is to be used. In time-varying conditions, a bad choice of this threshold could make the dual-mode receiver to work in the wrong adaptation mode most of the time. Clearly, the advantage of the two-stage receiver is obtained at the cost of an increased computational complexity, since two adaptive algorithms must be run in parallel. The example below shows that a receiver selecting the best of the first and second stage SINRs is very close to the optimal LMMSER over a wide range of values of SINR_{opt}.

Example. Consider a two-stage NDA receiver where NDA-RLS and modified RLS are used in the first and second stage, respectively. By using (29) into (47) we obtain SINR_{2nd} as a function of $\text{SINR}_{\text{opt}}, \eta$ and η_1, where η_1 is the MOE misadjustment of the NDA-RLS at the first stage. Fig. 1 shows SINR_{1st} and SINR_{2nd} vs. SINR_{opt} for $\eta = \eta_1 = 0.1$. Both SINR_{1st} and SINR_{2nd} converge to finite values as $\text{SINR}_{\text{opt}} \to \infty$. These are given by $1/\eta_1$ and by

$$\frac{(1 + 2Q(\sqrt{1/\eta_1}))^2}{4\eta Q(\sqrt{1/\eta_1})(1 - Q(\sqrt{1/\eta_1}))}$$

respectively. For large SINR_{opt}, the second stage offers better steady-state performance. On the contrary, if SINR_{opt} is low, the second stage suffers from SINR degradation due to the large BER at the output of the first stage, and the first stage yields better performance. As anticipated in Remark 4, $\max\{\text{SINR}_{\text{1st}}, \text{SINR}_{\text{2nd}}\}$ is close to optimum over a wide range of SINR_{opt} values (from -15 to 15 dB). Fig. 1 shows also the steady-state SINR vs. SINR_{opt} of a DD algorithm (i.e., the non-negative solution of (46)), for $\eta = 0.1$. For $\text{SINR}_{\text{opt}} \geq 0$ dB, the DD algorithm yields almost optimal performance. If SINR_{opt} decreases below 0 dB, the SINR of the DD algorithm drops rapidly and goes to zero when (46) has no positive solution.

5 Time-varying channels

In this section, we propose modified LMS and RLS algorithms in order to cope with time-varying channels. The proposed approach is based on exploiting the estimation of user 1 channel in order to create an appropriate auxiliary sequence $\widetilde{b}_1[n]$ in the modified adaptive algorithm. This is a non-trivial extension to the case of general frequency-selective channels of the idea presented in [6], where a data-directed phase estimator is coupled with DD-LMS in order to work in frequency-flat Rayleigh channels. Our generalization applies both to DD and to two-stage receivers. However, as anticipated in the Introduction, when coupled with the two-stage scheme it yields a particularly attractive solution for tracking moderately time-varying channels, as it will be illustrated by the examples of Section 6.

Informally speaking, adaptive algorithms are based on the idea of exchanging ensemble averages with time averages. In order to track time-varying statistics, time averages are performed over a "sliding-window" of a given size, referred to as the algorithm *memory*. The components of the input signal $\mathbf{y}[n]$ that vary significantly over the algorithm memory are averaged over the sliding window and are to be treated as random [7]. In the steady-state analysis of Sections 3 and 4, the channel vectors were considered as deterministic constants. The underlying assumption is that the normalized Doppler bandwidth $B_d T$ is much smaller than the inverse algorithm memory. [7] In this section, we consider the case where $B_d T$ is comparable or larger than the inverse algorithm memory and we make the assumption that the $\mathbf{c}_k[n]$'s are WSS zero-mean vector processes, mutually independent and independent on the user data symbols.

The benchmark linear receiver is a centralized LMMSER having ideal instantaneous knowledge of all user channels. On the other hand, if the receiver has no knowledge of the channels, the standard MSE (averaged also with respect to the channels) is useless. In fact, we obtain $J = E[|b_1 - \mathbf{h}^H \mathbf{y}|^2] = 1 + \mathbf{h}^H \mathbf{R}_y \mathbf{h}$, whose minimization yields the trivial solution $\mathbf{h} = \mathbf{0}$ [6, 7]. Since we are interested in the decentralized detection of user 1, side information about users $k = 2, \ldots, K$ cannot be exploited, but we can assume that the receiver is provided with an estimate $\widehat{\mathbf{c}}_1[n]$ of $\mathbf{c}_1[n]$ by some external channel estimation device. From the general SINR formula (10), by averaging also with respect to the channel vectors, we obtain

$$\text{SINR} = \frac{\mathbf{h}^H \mathbf{S}_1[0] \mathbf{R}_c \mathbf{S}_1[0]^H \mathbf{h}}{\mathbf{h}^H \mathbf{R}_w \mathbf{h}} \tag{48}$$

where $\mathbf{R}_c = E[\mathbf{c}_1[n]\mathbf{c}_1[n]^H]$ is the covariance matrix of user 1 channel. [8] By using the maximization of the output SINR given by (48) as design criterion for $\mathbf{h}$, we obtained that the optimal filter vector is the generalized eigenvector corresponding to the maximal eigenvalue of the matrix pencil [22] $\{\mathbf{S}_1[0]\mathbf{R}_c\mathbf{S}_1[0]^H, \mathbf{R}_w\}$. In the following, this filter will be denoted by $\mathbf{h}_{\text{fast}}$.

We look for simple modified LMS and RLS algorithms approximating $\mathbf{h}_{\text{fast}}$. To this purpose, we consider an auxiliary sequence in the form $\widetilde{b}_1[n] = a[n]b_1[n]$, where $a[n]$ is a random variable uncorrelated with the user data symbols, such that $E[\mathbf{c}_1[n]a[n]^*] = \mathbf{r}$, where $\mathbf{r}$ is a non-zero vector of length P. Obviously, $a[n]$ is derived from the channel estimate $\widehat{\mathbf{c}}_1[n]$ of $\mathbf{c}_1[n]$, as we will see later. The

[7] For example, the memory of RLS with exponential forgetting factor α is about $1/(1-\alpha)$ symbols. Then, the channels can be considered as constants if $B_d T \ll 1 - \alpha$.

[8] The covariance matrix $\mathbf{R}_c$ is generally non-diagonal, even if the underlying continuous-time channel has uncorrelated scattering [9]. In fact, after low-pass filtering and sampling, the discrete-time channel coefficients given by (3) are correlated.

minimizer of the modified MSE cost function

$$\begin{aligned}\widetilde{J} &= E[|b_1[n]a[n] - \mathbf{h}^H \mathbf{y}[n]|^2] \\ &= 1 - 2\mathrm{Re}\{\mathbf{h}^H \mathbf{S}_1[0]\mathbf{r}\} + \mathbf{h}^H \mathbf{R}_y \mathbf{h} \end{aligned} \tag{49}$$

is given by $\widetilde{\mathbf{h}} = \mathbf{R}_y^{-1}\mathbf{S}_1[0]\mathbf{r}$. In order to approximate $\widetilde{\mathbf{h}}$ by DD and two-stage adaptive algorithms, we let $\widetilde{b}_1[n] = \widehat{b}_1[n]a[n]$, where $\widehat{b}_1[n]$ are symbol-by-symbol decisions.

Now, we have to find $a[n]$ as a function of $\widehat{\mathbf{c}}_1[n]$ in order to achieve $\widetilde{\mathbf{h}} = \mathbf{h}_{\text{fast}}$. The SINR attained by $\widetilde{\mathbf{h}}$ is given by

$$\text{SINR} = \frac{\mathbf{r}^H \mathbf{S}_1[0]^H \mathbf{R}_y^{-1} \mathbf{S}_1[0] \mathbf{R}_c \mathbf{S}_1[0]^H \mathbf{R}_y^{-1} \mathbf{S}_1[0]\mathbf{r}}{\mathbf{r}^H \mathbf{S}_1[0]^H \mathbf{R}_y^{-1} \mathbf{R}_w \mathbf{R}_y^{-1} \mathbf{S}_1[0]\mathbf{r}} \tag{50}$$

The following proposition, proved in Appendix B, provides the needed result:

Proposition 2. $\mathbf{h}_{\text{fast}}$ *maximizing (48) can be written as*

$$\mathbf{h}_{\text{fast}} = \mathbf{R}_y^{-1}\mathbf{S}_1[0]\mathbf{r}_{\text{fast}} \tag{51}$$

where $\mathbf{r}_{\text{fast}}$ *maximizes (50). Moreover, if the matrix* $\mathbf{A} \stackrel{\Delta}{=} \mathbf{S}_1[0]^H \mathbf{R}_y^{-1} \mathbf{S}_1[0]$ *is invertible,* $\mathbf{r}_{\text{fast}}$ *is the eigenvector corresponding to the maximal eigenvalue of the matrix* $\mathbf{R}_c\mathbf{A}$*.* □

The desired relation between $a[n]$ and $\widehat{\mathbf{c}}_1[n]$ is given by:

Corollary. *Assume that the channel estimator* $\widehat{\mathbf{c}}_1[n]$ *is uncorrelated with the user symbols and let* $\mathbf{\Sigma} = E[\mathbf{c}_1[n]\widehat{\mathbf{c}}_1[n]^*]$*. If det*$(\mathbf{\Sigma}) \neq 0$*, the sequence* $a[n]$ *achieving* $\widetilde{\mathbf{h}} = \mathbf{h}_{\text{fast}}$ *as the minimizer of (49) is given by*

$$a[n] = \mathbf{r}_{\text{fast}}^H \left(\mathbf{\Sigma}^{-1}\right)^H \widehat{\mathbf{c}}_1[n] \tag{52}$$

□

Remark 5. The condition $\det(\mathbf{\Sigma}) \neq 0$ is rather mild and should hold for any reasonably good channel estimator. It is interesting to notice that perfect channel knowledge is actually not needed for the deterministic (i.e., non-adaptive) filter $\mathbf{h}_{\text{fast}}$. Nevertheless, it is intuitively clear that good channel estimation can improve the performance of the corresponding *adaptive* receivers based on modified LMS or RLS. In practice, $\mathbf{\Sigma}$ is not known. However, for an unbiased channel estimator $\widehat{\mathbf{c}}_1[n] = \mathbf{c}_1[n] + \mathbf{e}[n]$, where $\mathbf{e}[n]$ is an error vector with mean zero, uncorrelated with $\mathbf{c}_1[n]$, we have that $\mathbf{\Sigma} = \mathbf{R}_c$. Moreover, $\mathbf{R}_c$ can be approximated by the exponentially weighted sample covariance matrix of $\widehat{\mathbf{c}}_1[n]$, $\widehat{\mathbf{R}}_c[n] = \sum_{i=1}^n \alpha^{n-i}\widehat{\mathbf{c}}_1[i]\widehat{\mathbf{c}}_1[i]^H$.

Remark 6. The approach proposed in this section for general frequency-selective time-varying channels is particularly simple when the channel of user 1 is frequency-flat. When the receiver has perfect timing for user 1, the channel $\mathbf{c}_1[n]$ reduces to the scalar $c_1[n]$. Then, if the receiver is provided with an estimator $\widehat{c}_1[n]$, such that $r = E[c_1[n]\widehat{c}_1[n]^*] \neq 0$, the simple choice $a[n] = \widehat{c}_1[n]$ achieves $\widetilde{\mathbf{h}} \propto \mathbf{h}_{\text{fast}}$.

With minor modifications, this approach is proposed in [6], where a classical DA-LMS working in DD mode is coupled with a decision-directed channel estimator tracking the phase of the (frequency-flat) complex channel gain $c_1[n]$. In [6], this approach is motivated from the observation that DD adaptive algorithms get into troubles during deep fades, when decisions become unreliable and the channel phase changes rapidly. Actually, we have shown that any channel estimator $a[n]$ yielding $r \neq 0$ provides an optimal filter. In particular, perfect knowledge of the phase of $c_1[n]$ is sufficient in the frequency-flat case. In fact, by letting $a[n] = c_1[n]/|c_1[n]|$ we get $r = E[|c_1[n]|] > 0$.

Remark 7. In [10] (see also references therein), the LMMSE-rake receiver is proposed to cope with frequency-selective time-varying channels. In brief, this consists of a bank of P adaptive filters followed by a "maximal-ratio combiner". The p-th filter is adapted by a modified DD LMS or RLS algorithm driven by the auxiliary sequence $\widetilde{b}_{1,p}[n] = a_p[n]\widehat{b}_1[n]$, where $a_p[n]$ is an estimate of the p-th channel coefficient $c_1[n;p]$. The p-th filter output, $z_p[n]$, is an estimate of the product $c_1[n;p]b_1[n]$ and the overall receiver output is obtained as $z[n] = \sum_{p=0}^{P-1} a_p[n]^* z_p[n]$.

Simulations and analysis show that the LMMSE-rake is effective even for fairly large B_dT [10]. On the other hand, its complexity is larger than the proposed two-stage NDA algorithm, if $P > 2$, since P adaptive algorithms must be run in parallel. Further comparisons between the LMMSE-rake and two-stage approaches are out of the scope of this paper and are left for future work.

6 Results

We consider a system with $K = 10$ users and processing gain $L = 31$. Each user is given a distinct sequence from a Gold set [23]. For simplicity, we assume ideal Nyquist chip pulses $\psi(t) = \frac{1}{\sqrt{T_c}}\text{sinc}(t/T_c)$ and we let $W = 1/T_c$, yielding $N_c = 1$ sample per chip. Without loss of generality, we let $q_1 = 0$ and we generate independently the delays q_k for $k = 2, \ldots, K$, uniformly distributed over the integers in $[-L/2, L/2)$, and γ_k for $k = 1, \ldots, K$, uniformly distributed over $[0, T_c)$.

The channel vectors $\mathbf{c}_k[n]$ are obtained from (3), where the underlying continuous-time channels have impulse responses

$$c_k(t;\tau) = \sum_p g_p(t)\delta(\tau - \tau_p) \tag{53}$$

where $g_p(t)$ are zero-mean mutually uncorrelated complex Gaussian WSS random processes with Jake's type Doppler spectrum $\sigma_p^2/(\pi\sqrt{B_d^2 - f^2})$ [24]. The *delay-intensity profile* [9], defined by the pairs (σ_p^2, τ_p), is given in Table 1. In our simulations, B_d and the delay-intensity profile are common to all users, and the channel vectors $\mathbf{c}_k[n]$ are independently generated for each user, with $P = 7$ discrete-time coefficients.

The receiver processing window is chosen to span two symbol intervals ($M = 62$). We let $M_1 = -15$ and $M_2 = 46$, so that the useful symbol falls approximately in the middle of the processing window.

6.1 Time-invariant channels

In order to validate the steady-state analysis of Sections 3 and 4, we let $B_d T = 0$ (time-invariant channels). The channel vectors are randomly generated and scaled in order to achieve the desired user SNRs. The assignment of the delays q_k, of the channel vectors $\mathbf{c}_k$ and of the spreading sequences $\mathbf{s}_k$ is fixed throughout the simulations. Therefore, *we are not averaging over these parameters.* We considered two SNR assignments: (a) all users have the same SNR= 13 dB (corresponding to $E_b/N_0 =$ 10 for uncoded 4PSK); (b) users $k = 1, \ldots, 5$ have SNR= 13 dB and users $k = 6, \ldots, 10$ have SNR= 28 dB. These situations are representative of perfect power-control and of uncompensated near-far effect.

Fig. 2 and 3 show BER vs. the number of symbol intervals (i.e., algorithm iterations) for DA, NDA, GNDA, DD and two-stage RLS algorithms in cases (a) and (b), respectively. The two-stage is based on GNDA-RLS in the first stage. The curves are obtained by the semi-analytic MC method (37) averaged over $N = 50$ independent simulation runs. At each iteration step, (36) is evaluated via the DCT method of [21], for the current value of the filter error vector. The horizontal lines indicate the steady-state BER obtained via the SSGA. The BER of the ideal (non-adaptive) SUMF, LMMSER and GCMOER are shown for comparison. Fig. 4 and 5 show analogous results for DA, NDA, DD and two-stage LMS algorithms, where the two-stage is based on NDA-LMS in the first stage.

DA and DD algorithms yields almost identical performance (the DD starts with an initial training sequence of length $T_r = 200$). The second stage of the two-stage algorithm has slightly inferior steady-state BER and slower convergence than DD and DA (this fact is less evident with LMS adaptation). The performance improvement of the second stage over NDA-GNDA adaptation (first stage) is remarkable. In all cases, the SSGA yields very accurate steady-state results.

It is interesting to notice that, in heavy near-far situations like case (b), the convergence of LMS-type algorithms is very slow. This can be intuitively explained by the fact that the convergence speed of LMS is affected by the eigenvalue spread [2] of $\mathbf{R}_y$, which increases with the users power imbalance. On the contrary, the convergence speed of RLS-type algorithms is insensitive to the eigenvalue spread of $\mathbf{R}_y$, and is almost independent of the near-far effect.

6.2 Time-varying channels

In order to demonstrate the superiority of the two-stage over the DD scheme in time-varying conditions, we consider moderately time-varying channels with normalized Doppler bandwidth $B_d T = 10^{-3}$. [9] For simplicity, we assume perfect knowledge of user 1 channel, i.e., $\widehat{\mathbf{c}}_1[n] = \mathbf{c}_1[n]$, and that $\boldsymbol{\Sigma} = \mathbf{R}_c$ is perfectly known. Because of space limitations, we show here only results for the modified DD-RLS and for the two-stage NDA based on GNDA-RLS and modified RLS. For all algorithms, the auxiliary sequence is given by $\widetilde{b}_1[n] = \widehat{b}_1[n]a[n]$, where $a[n]$ is given by (52).

Fig. 6 shows SINR vs. the number of symbols in SNR case (a). The curve labeled by "DD" shows the output SINR of standard DD-RLS. As expected, after the first deep fade, occurring between $n = 1000$ and $n = 1500$, the standard DD looses tracking. On the contrary, the modified DD ("Mod. DD") and the two-stage NDA keep on tracking and recover a good SINR after deep fades. Stage 2 of the two-stage receiver achieves about the same SINR of modified DD. Curves for the SUMF and for the ideal centralized LMMSER are shown for comparison. Fig. 7 shows analogous results for SNR case (b). Here, even the modified DD algorithm looses tracking after the first deep fade. This is probably due to the heavy near-far effect, generating severe error propagation in the DD feedback loop (as noticed in [6], in heavy near-far situations the DD may end up tracking another user during a deep fade of the desired user). On the contrary, the two-stage receiver is still able to recover after the deep fades. This shows that the proposed two-stage NDA is more robust than DD algorithms in near-far conditions.

7 Concluding remarks

We have derived closed form expressions for the steady-state SINR of some DA and NDA adaptive algorithms for linear decentralized multiuser detection. In order to fill the performance gap between DA and NDA algorithms, we considered DD and a new two-stage NDA algorithm, where the pre-

[9] This corresponds to a vehicle speed of 70 km/h, carrier frequency of 2 GHz, chip-rate of 4 Mchip/s and spreading gain 31, which can be regarded as fairly representative numbers for UMTS data transmission [25].

decisions obtained at the output of a NDA adaptive filter (first stage) are used for training a second stage. Closed-form steady-state SINR analysis, taking into account the effect of errors in the DD loop and at the output of the first stage, shows that both the DD and the two-stage schemes are able to recover the performance loss with respect to DA algorithms. In particular, a two-stage receiver selecting the output of the stage which yields the best SINR is very close to the optimal SINR over a wide range of values. We showed also that the steady-state BER at the output of adaptive filters (with 4PSK) can be accurately approximated by a straightforward Gaussian approximation based on the steady-state SINR formulas.

In order to motivate the use of the newly proposed two-stage scheme, we considered time-varying channels and we proposed a SINR-maximizing criterion for the filter design. LMS and RLS adaptive schemes can be coupled with a channel estimator for the desired user, in order to approximate adaptively the SINR-maximizing filter. The resulting modified LMS and RLS algorithms can work either in DD or two-stage mode. Simulations show that the two-stage is more robust than the DD scheme in heavy near-far conditions. Also, the two-stage scheme is simpler than the LMMSE-rake receiver of [10]. Therefore, it represents an attractive alternative for moderate values of the Doppler bandwidth.

Some issues for future research are: i) the characterization of the tracking performance of the two-stage receiver with time-varying channels; ii) the impact of actual channel estimation and phase recovery over the two-stage and DD receivers; iii) the comparison between the two-stage and the LMMSE-rake receivers, for different channel delay-spread and Doppler bandwidth.

APPENDIX

A Convergence analysis of modified LMS and RLS

From the definitions $J_0 = E[|b_1 - \overline{\mathbf{h}}^H \mathbf{y}|^2]$ and $\mathbf{h}_{\rm opt} = \sqrt{\mathcal{E}_1}\mathbf{R}_y^{-1}\mathbf{p}_1$, it is immediate to show that the second equality in (43) follows from the first. The first and the third equalities in (43) express the convergence of the mean filter vector and of the excess MSE. We shall consider separately the convergence of the modified LMS and RLS algorithms, which follows from rather standard approaches whose details can be found in [2, 17].

Convergence of the modified LMS algorithm. We let $\mathbf{h}[n] = \rho\mathbf{h}_{\rm opt} + \boldsymbol{\epsilon}[n]$ in the modified LMS recursion and we obtain

$$\boldsymbol{\epsilon}[n] = (\mathbf{I} - \mu\mathbf{y}[n]\mathbf{y}[n]^H)\boldsymbol{\epsilon}[n-1] + \mu\widetilde{e}_0[n]^*\mathbf{y}[n] \tag{54}$$

where $\widetilde{e}_0[n] = \widetilde{b}_1[n] - \rho^*\mathbf{h}_{\rm opt}^H\mathbf{y}[n]$. By taking the expectation of both sides, from the independence assumption [2] we obtain

$$E[\boldsymbol{\epsilon}[n]] = (\mathbf{I} - \mu\mathbf{R}_y)E[\boldsymbol{\epsilon}[n-1]] \tag{55}$$

where we used the fact that, because of the orthogonality principle, $E[\widetilde{e}_0[n]^*\mathbf{y}[n]] = \mathbf{0}$. Then, $\lim_{n\to\infty} E[\boldsymbol{\epsilon}[n]] = \mathbf{0}$ if all the eigenvalues of $\mathbf{I} - \mu\mathbf{R}_y$ have magnitude less than 1, implying that $\overline{\mathbf{h}} = \rho\mathbf{h}_{\rm opt}$.

In order to show convergence for the excess MSE, we let $\mathbf{R}_y = \mathbf{Q}{\rm diag}(\lambda_1,\ldots,\lambda_M)\mathbf{Q}^H$, with $\mathbf{Q}$ unitary, and we define $\boldsymbol{\theta}[n]$ as the vector of the diagonal elements of the transformed filter error covariance matrix $\widetilde{\mathbf{R}}_\epsilon[n] = \mathbf{Q}^H E[\boldsymbol{\epsilon}[n]\boldsymbol{\epsilon}[n]^H]\mathbf{Q}$. Then, by following the same steps of [2], we obtain the difference equation

$$\boldsymbol{\theta}[n] = \mathbf{B}\boldsymbol{\theta}[n-1] + \mu^2\widetilde{J}_{\rm min}\boldsymbol{\lambda} \tag{56}$$

where $\mathbf{B} = \mathbf{I} - 2\mu\boldsymbol{\Lambda} + \mu^2\boldsymbol{\Lambda}^2 + \mu^2\boldsymbol{\lambda}\boldsymbol{\lambda}^H$, where we define the vector $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_M)^T$ of the eigenvalues of $\mathbf{R}_y$, and where

$$\begin{aligned}\widetilde{J}_{\rm min} &= E[|\widetilde{b}_1[n] - \rho\mathbf{h}_{\rm opt}^H\mathbf{y}[n]|^2] \\ &= 1 - |\rho|^2\chi \end{aligned} \tag{57}$$

Since $\mathbf{B}$ has all-positive elements, from the Perron-Frobenius theorem [22] we get the stability condition [2]

$$\sum_{\ell=1}^{M} \frac{\mu\lambda_\ell}{2-\mu\lambda_\ell} < 1 \tag{58}$$

Under this condition, we have that

$$\begin{aligned} J_{\rm ex} &\stackrel{\Delta}{=} \lim_{n\to\infty} \boldsymbol{\lambda}^H\boldsymbol{\theta}[n] \\ &= \mu^2\widetilde{J}_{\rm min}\boldsymbol{\lambda}^H(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\lambda} \end{aligned} \tag{59}$$

By applying the matrix inversion lemma to $(\mathbf{I}-\mathbf{B})^{-1}$, after some algebra, we obtain the final desired result $J_{\rm ex} = \eta\widetilde{J}_{\rm min}$, where η is given in (28).

Convergence of the modified RLS algorithm. The solution $\mathbf{h}[n]$ of the exponentially-weighted least-squares problem

$$\min_{\mathbf{h}} \sum_{i=1}^{n} \alpha^{n-i} |\widetilde{b}_1[n] - \mathbf{h}^H \mathbf{y}[n]|^2$$

must satisfy

$$\widetilde{\mathbf{R}}_y[n]\mathbf{h}[n] = \alpha \widetilde{\mathbf{R}}_y[n-1]\mathbf{h}[n-1] + \mathbf{y}[n]\widetilde{b}_1[n]^* \tag{60}$$

where we define the sample covariance matrix $\widetilde{\mathbf{R}}_y[n] = \sum_{i=1}^{n} \alpha^{n-i}\mathbf{y}[i]\mathbf{y}[i]^H$. By substituting $\mathbf{h}[n] = \rho \mathbf{h}_{\text{opt}} + \boldsymbol{\epsilon}[n]$ into (60) we obtain

$$\widetilde{\mathbf{R}}_y[n]\boldsymbol{\epsilon}[n] = \alpha \widetilde{\mathbf{R}}_y[n-1]\boldsymbol{\epsilon}[n-1] + \mathbf{y}[n]\widetilde{e}_0[n]^* \tag{61}$$

where, again, $\widetilde{e}_0[n] = \widetilde{b}_1[n] - \rho^* \mathbf{h}_{\text{opt}}^H \mathbf{y}[n]$. For large n, we can multiply both sides by $\mathbf{M}[n] = \widetilde{\mathbf{R}}_y[n]^{-1}$ and use the fact that the inverse covariance matrix $\mathbf{M}[n]$ behaves like a *quasi-deterministic* quantity when $M(1-\alpha) \ll 1$ (see [5] and references therein). Therefore, $\lim_{n\to\infty} \mathbf{M}[n] \simeq E[\mathbf{M}[n]] = (1-\alpha)\mathbf{R}_y^{-1}$. We obtain

$$\boldsymbol{\epsilon}[n] = \alpha^n (1-\alpha)\mathbf{R}_y^{-1}\widetilde{\mathbf{R}}_y[0]\boldsymbol{\epsilon}[0] + (1-\alpha)\mathbf{R}_y^{-1}\sum_{i=1}^{n} \alpha^{n-i}\mathbf{y}[i]\widetilde{e}_0[i]^* \tag{62}$$

From the orthogonality principle, we get that $\sum_{i=1}^{n} \alpha^{n-i}\mathbf{y}[i]\widetilde{e}_0[i]^* = \mathbf{0}$. Then, by taking the expectation of both sides of (62) we obtain that $\lim_{n\to\infty} E[\boldsymbol{\epsilon}[n]] = \mathbf{0}$ for all $0 < \alpha < 1$, implying that $\overline{\mathbf{h}} = \rho \mathbf{h}_{\text{opt}}$.

In order to show convergence for the excess MSE, we let $J_{\text{ex}}[n] = E[\boldsymbol{\epsilon}[n]^H \mathbf{R}_y \boldsymbol{\epsilon}[n]]$. By using (61), the quasi-deterministic and the independence assumptions, after some algebra we obtain the difference equation

$$J_{\text{ex}}[n] = \alpha^2 J_{\text{ex}}[n-1] + (1-\alpha)^2 M \widetilde{J}_{\min} \tag{63}$$

where $\widetilde{J}_{\min}$ is again given by (57). Since $0 < \alpha < 1$, we have that

$$J_{\text{ex}} \stackrel{\Delta}{=} \lim_{n\to\infty} J_{\text{ex}}[n] = \eta \widetilde{J}_{\min} \tag{64}$$

where η is given in (28).

B Proof of Proposition 1.

The vector $\mathbf{r}$ maximizing (50) is the generalized eigenvector corresponding to the maximal eigenvalue of the matrix pencil $\{\mathbf{A}\mathbf{R}_c\mathbf{A}, \mathbf{S}_1[0]^H\mathbf{R}_y^{-1}\mathbf{R}_w\mathbf{R}_y^{-1}\mathbf{S}_1[0]\}$, where $\mathbf{A} = \mathbf{S}_1[0]^H\mathbf{R}_y^{-1}\mathbf{S}_1[0]$. From (9), we can write

$$\mathbf{R}_y = \mathbf{R}_w + \mathbf{S}_1[0]\mathbf{R}_c\mathbf{S}_1[0]^H$$

By using the above decomposition, we obtain that $\mathbf{h}$ and $\mathbf{r}$ maximizing (48) and (50), must satisfy

$$\begin{aligned} \mathbf{S}_1[0]\mathbf{R}_c\mathbf{S}_1[0]^H\mathbf{h} &= \phi_{\max}\mathbf{R}_y\mathbf{h} \\ \mathbf{A}\mathbf{R}_c\mathbf{A}\mathbf{r} &= \theta_{\max}\mathbf{A}\mathbf{r} \end{aligned} \quad (65)$$

where $\phi_{\max}$ and $\theta_{\max}$ are the maximum eigenvalues of the matrix pencils $\{\mathbf{S}_1[0]\mathbf{R}_c\mathbf{S}_1[0]^H, \mathbf{R}_y\}$ and $\{\mathbf{A}\mathbf{R}_c\mathbf{A}, \mathbf{A}\}$, respectively. Since $\mathbf{R}_y$ is invertible, from the first equation in (65) we obtain $\mathbf{h} = \frac{1}{\phi_{\max}}\mathbf{R}_y^{-1}\mathbf{S}_1[0]\mathbf{R}_c\mathbf{S}_1[0]^H\mathbf{h}$. Notice that $\mathbf{R}_c\mathbf{S}_1[0]^H\mathbf{h}$ is a vector of length P. Then, we have shown that the optimal $\mathbf{h}$ can always be written in the form

$$\mathbf{h}_{\text{fast}} = \frac{1}{\phi_{\max}}\mathbf{R}_y^{-1}\mathbf{S}_1[0]\mathbf{r}_{\text{fast}} \quad (66)$$

where $\mathbf{r}_{\text{fast}}$ is a suitable vector of length P. In order to show that $\mathbf{r}_{\text{fast}}$ satisfies the second equality of (65), we substitute (66) into the first equality of (65) and we obtain

$$\mathbf{S}_1[0]\mathbf{R}_c\mathbf{A}\mathbf{r}_{\text{fast}} = \phi_{\max}\mathbf{S}_1[0]\mathbf{r}_{\text{fast}} \quad (67)$$

Both the LHS and the RHS of the above equation belong to the range of $\mathbf{S}_1[0]$. Therefore, we can multiply both sides from the left by $\mathbf{S}_1[0]^H\mathbf{R}_y^{-1}$ without getting a trivial $0 = 0$ equation. This yields

$$\mathbf{A}\mathbf{R}_c\mathbf{A}\mathbf{r}_{\text{fast}} = \phi_{\max}\mathbf{A}\mathbf{r}_{\text{fast}} \quad (68)$$

Now, we compare (68) with the second line of (65) and we prove the statement by contradiction. If $\theta_{\max} > \phi_{\max}$, then $\mathbf{h} = \mathbf{R}_y^{-1}\mathbf{S}_1[0]\mathbf{r}$ with $\mathbf{r}$ being the eigenvector corresponding to $\theta_{\max}$ would achieve a better SINR than $\mathbf{h}_{\text{fast}}$ (contradiction). Then, it must be $\theta_{\max} = \phi_{\max}$ and $\mathbf{r}_{\text{fast}}$ be the eigenvector satisfying the second line of (65). Finally, if $\mathbf{A}$ is invertible, we can multiply both sides of (68) by $\mathbf{A}^{-1}$ and obtain the eigen-equation

$$\mathbf{R}_c\mathbf{A}\mathbf{r}_{\text{fast}} = \phi_{\max}\mathbf{r}_{\text{fast}} \quad (69)$$

This concludes the proof.

References

[1] S. Verdú, *Multiuser detection.* Cambridge, UK: Cambridge University Press, 1998.

[2] S. Haykin, *Adaptive filter theory*, 2nd ed., Englewood Cliffs, NJ: Prentice-Hall, 1994.

[3] X. Wang, V. Poor, "Blind multiuser detection: a subspace approach," *IEEE Trans. Inform. Theory,* Vol. 44, pp. 677 – 690, March 1998.

[4] M. Honig, U. Madhow and S. Verdú, "Blind multiuser detection," *IEEE Trans. Inform. Theory,* Vol. 41, pp. 944 – 960, July 1995.

[5] V. Poor, X. Wang, "Code-aided interference suppression for DS/CDMA communications – Part II: parallel blind adaptive implementations," *IEEE Trans. Commun.,* Vol. 45, pp. 1112 – 1122, Sept. 1997.

[6] A. Barbosa, S. Miller, "Adaptive detection of DS/CDMA signals in fading channels," *IEEE Trans. Commun.,* Vol. 46, pp. 115 – 124, Jan. 1998.

[7] S. Miller, M. Honig, L. Milstein, "Performance analysis of MMSE receivers for DS-CDMA in frequency selective fading channels," submitted to *IEEE Trans. on Commun.,* 1998.

[8] A. J. Viterbi, *CDMA – Principles of Spread Spectrum Communication.* Reading, MA: Addison-Wesley, 1995.

[9] J. Proakis, *Digital communications.* 3rd Edition. New York: McGraw-Hill, 1995.

[10] M. Latva-Aho, *Advanced receivers for wideband CDMA systems.* Ph.D. Thesis, Oulu University Library, Oulu, Finland, 1998.

[11] U. Madhow, M. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.,* Vol. 42, pp. 3178 – 3188, Dec. 1994.

[12] M. Tsatsanis, Z. Xu, "On minimum output energy CDMA receivers in the presence of multipath," *31st Conf. Inform. Sci. Systems (CISS),* Johns Hopkins Univ., Baltimore, MD, pp. 724 – 729, Mar. 19 – 21, 1997.

[13] M. Tsatsanis, Z. Xu, "Performance analysis of minimum variance CDMA receivers," *IEEE Trans. Sig. Proc.* Vol. 46, pp. 3014 – 3022, Nov. 1998.

[14] Z. Xu, M. Tsatsanis, "Adaptive minimum variance methods for direct blind multichannel equalization," Proc. *ICASSP '98*, Seattle, WA, May 1998.

[15] Z. Xu, M. Tsatsanis, "Blind stochastic gradient methods for optimal minimum variance CDMA receivers," Proc. *32nd Asilomar Conf. on Signals, Systems & Computers,* Pacific Grove, CA, Nov. 1998.

[16] J. Bucklew, T. Kurts, W. Sethares, "Weak convergence and local stability properties of fixed step size recursive algorithms," *IEEE Trans. Inform. Theory,* Vol. 30, pp. 966 – 978, May 1993.

[17] E. Eleftheriou, D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoust. Speech, Sig. Proc.* Vol. ASSP-34, pp. 1097 – 1109, Oct. 1986.

[18] L. Tong, G. Xu, T. Kailath, "Blind identification and equalization based on second-order statistics: a time-domain approach," *IEEE Trans. Inform. Theory,* Vol. 40, pp. 340 – 349, Mar. 1994.

[19] V. Poor, S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inform. Theory,* Vol. 43, pp. 858 – 871, May 1997.

[20] B. Saltzberg, "Intersymbol interference error bounds with applications to ideal bandlimited signals," *IEEE Trans. Inform. Theory,* Vol. IT-14, pp. 563 – 568, July 1968.

[21] A. Levy, "Fast error rate evaluation in the presence of intersymbol interference," *IEEE Trans. Commun.,* Vol. 33, pp. 479 – 481, May 1985.

[22] R. Horn, C. Johnson, *Matrix analysis.* Cambridge, UK: Cambridge University Press, 1985.

[23] D. Sarwate, M. Pursley, "Crosscorrelation properties of pseudorandom and related sequences," *Proc. IEEE,* Vol. 68, No. 5, pp. 563 – 619, May 1980.

[24] R. Clarke, "A statistical theory of mobile-radio reception," *Bell Syst. Tech. Journal,* pp. 957 – 1000, July-Aug. 1968.

[25] E. Dahlman, B. Gudmundson, M. Nilsson and J. Sköld, "UMTS/IMT-2000 based on wideband CDMA," *IEEE Comm. Magazine,* Vol. 36, No. 9, pp. 70 – 81, Sept. 1998.

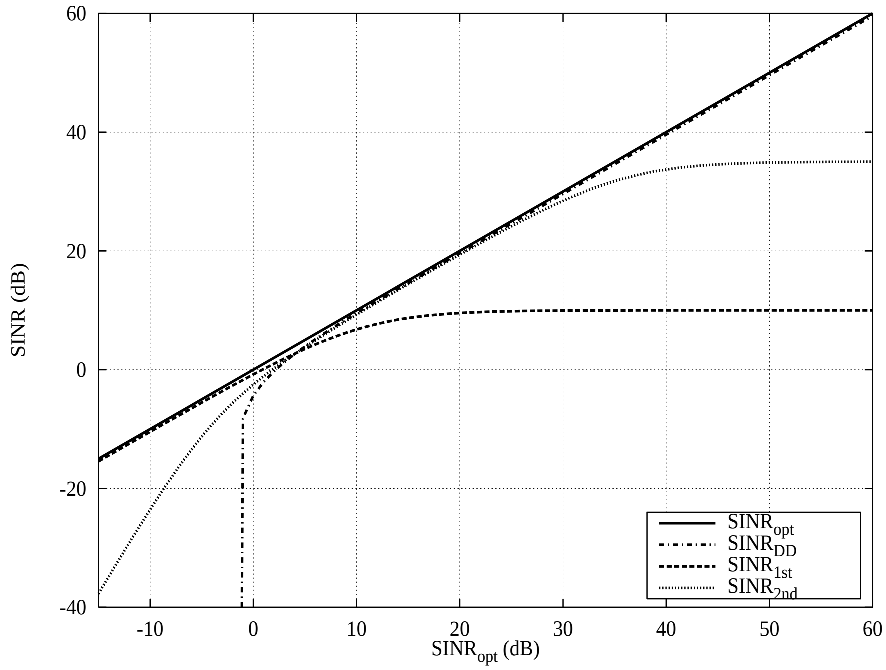

Figure 1: Steady-state output SINR of a DD algorithm and of stage 1 and stage 2 of a two-stage NDA algorithm as a function of $\mathrm{SINR}_{\mathrm{opt}}$.

p	σ_p^2	τ_p/T_c
0	1.0	0.0
1	0.5	1.2
2	0.2	3.4
3	0.1	5.6

Table 1: Delay-intensity profile of the Rayleigh channel used in the simulations.

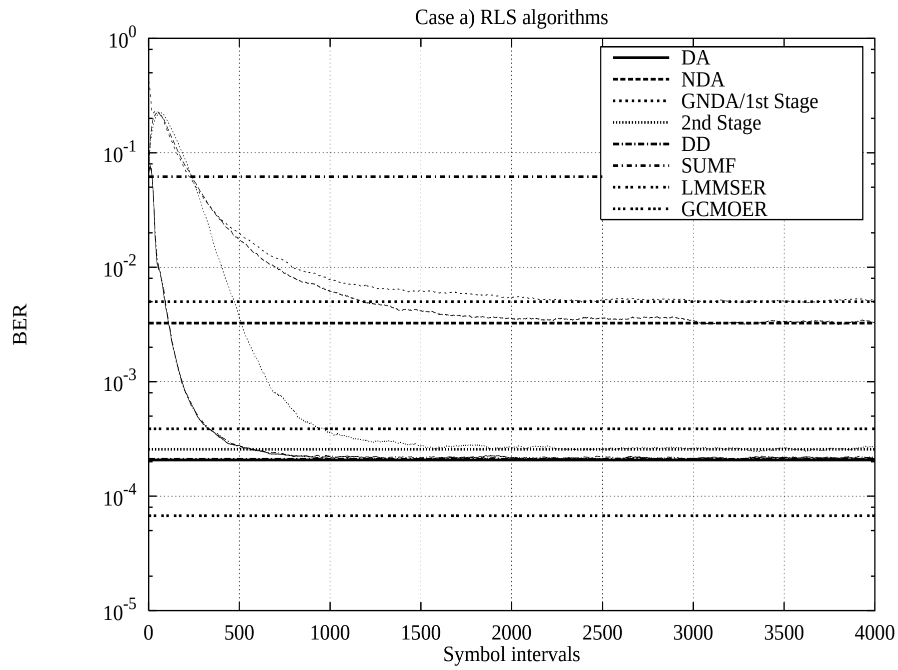

Figure 2: BER vs. number of symbols for the RLS algorithms in near-far case (a).

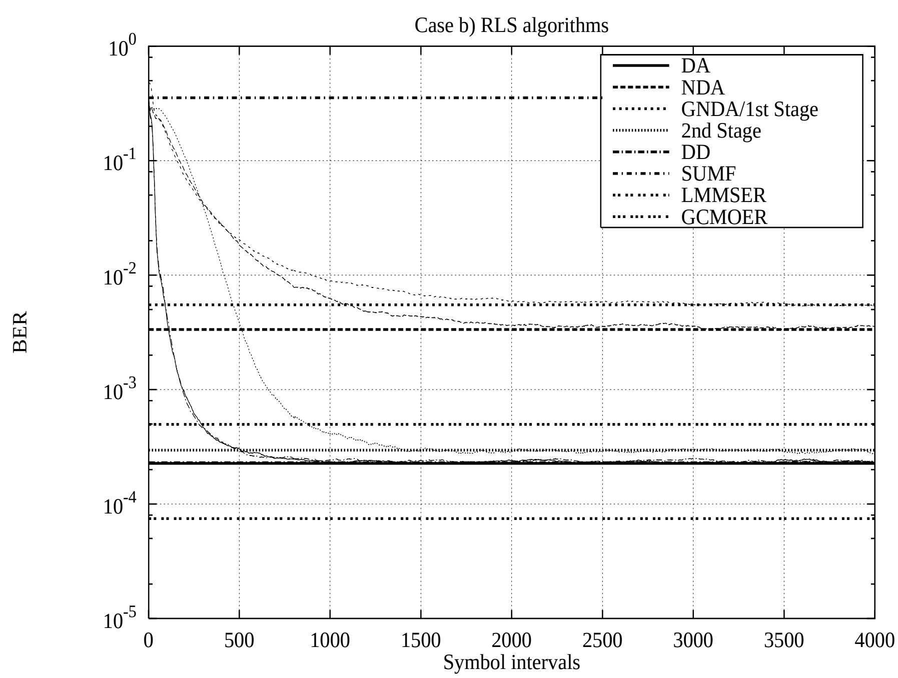

Figure 3: BER vs. number of symbols for the RLS algorithms in near-far case (b).

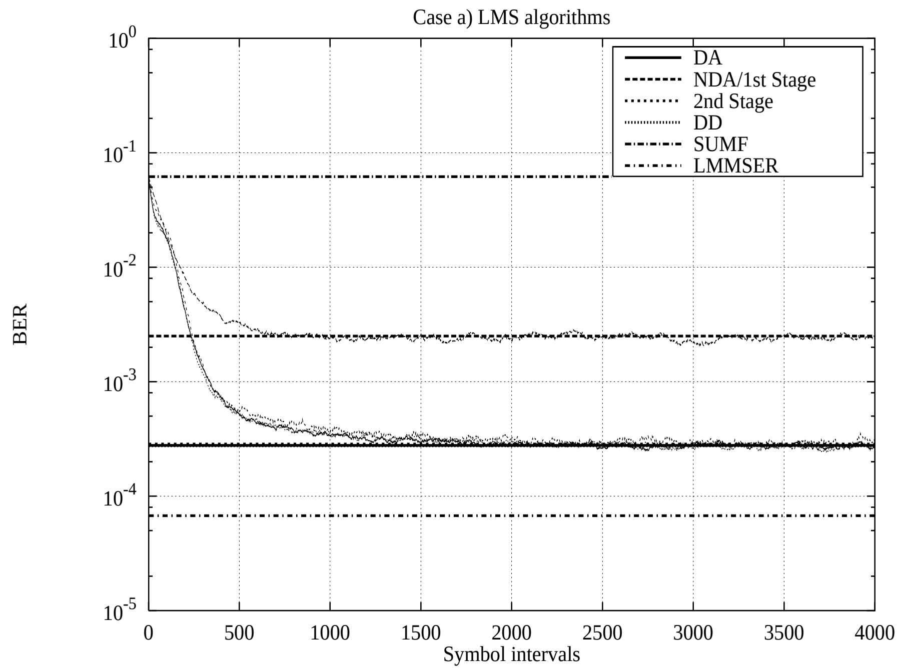

Figure 4: BER vs. number of symbols for the LMS algorithms in near-far case (a).

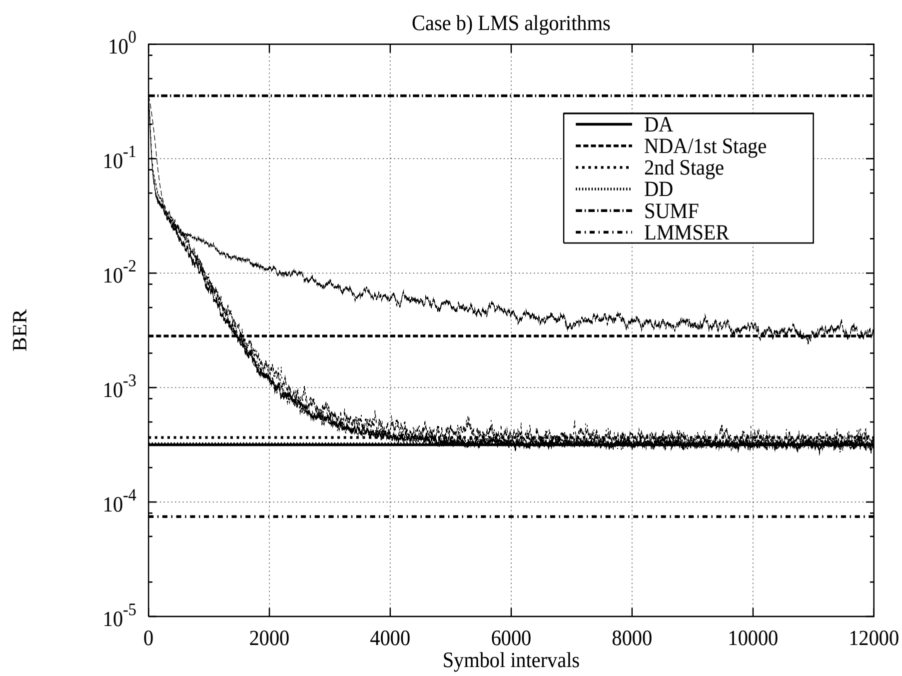

Figure 5: BER vs. number of symbols for the LMS algorithms in near-far case (b).

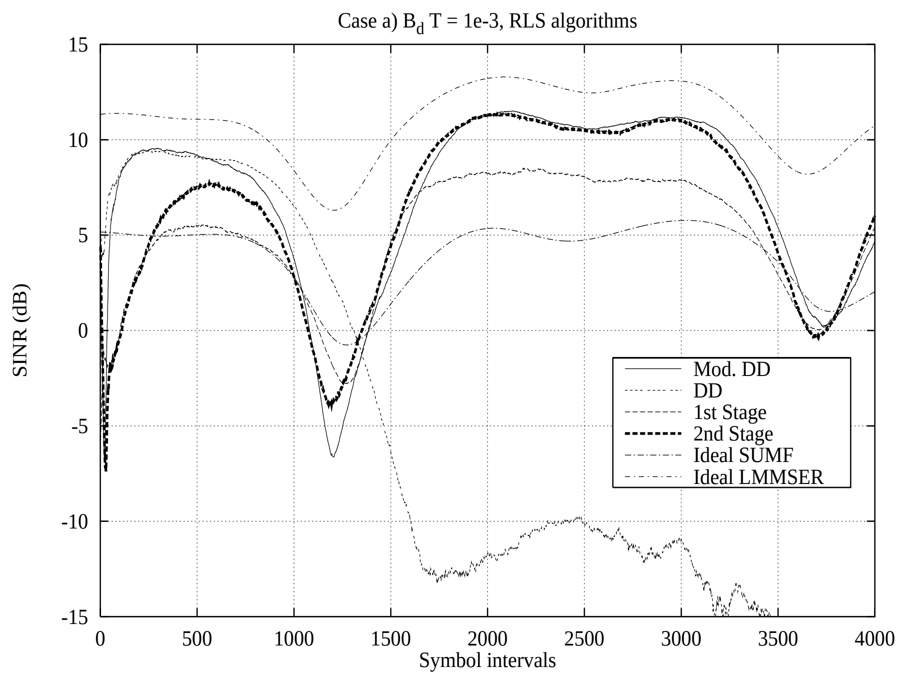

Figure 6: SINR vs. number of symbols for the RLS algorithms in near-far case (a), for time-varying channels with Doppler bandwidth $B_d T = 10^{-3}$.

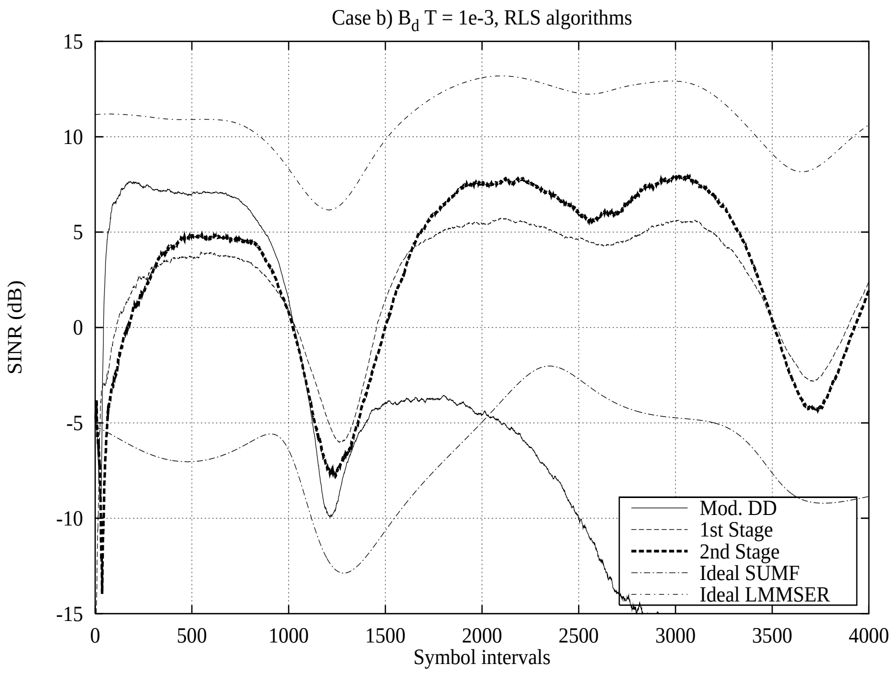

Figure 7: SINR vs. number of symbols for the RLS algorithms in near-far case (b), for time-varying channels with Doppler bandwidth $B_d T = 10^{-3}$.