

# Adaptive Multiuser CDMA Detector for Asynchronous AWGN Channels—Steady State and Transient Analysis

Lizhi Charlie Zhong, *Member, IEEE*, Zoran Siveski, *Member, IEEE*, Raafat E. Kamel, *Member, IEEE*, and Nirwan Ansari, *Senior Member, IEEE*

**Abstract**—A two-stage adaptive multiuser detector in an additive white Gaussian noise code-division multiple-access channel is proposed and analyzed. Its first stage is an asynchronous one-shot decorrelator which in terms of computational complexity only requires inversion of  $K$  symmetric  $K \times K$  matrices for all  $K$  users. In addition, the  $K$  inversions can be done in parallel, and the computed results for one user can be reused by all other users as well, resulting in a latency of only one bit, same as its synchronous counterpart. The decorrelated tentative decisions are utilized to estimate and subtract multiple-access interference in the second stage. Another novel feature of the detector is the adaptive manner in which the multiple-access interference estimates are formed, which renders prior estimation of the received signal amplitudes and the use of training sequences unnecessary. Adaptation algorithms considered include steepest descent (as well as its stochastic version), and a recursive least squares-type algorithm that offers a faster transient response and better error performance. Sufficient conditions for the receiver to achieve convergence are derived. The detector is near-far resistant, and is shown to provide substantial steady-state error performance improvement over the conventional and decorrelating detector, particularly in the presence of strong interfering signals.

**Index Terms**—Adaptive multiuser detector, CDMA, one-shot decorrelator, RLS.

## I. INTRODUCTION

Since the optimum receiver in a multiple-access, Gaussian noise channel was proposed ten years ago [1], a considerable amount of research in academia and industry has been devoted to what is usually referred to as multiuser detection. Its common viewpoint is a joint detection of all the users by exploiting the structure of the multiple-access interference, rather than focusing solely on the desired user and ignoring the rest. The complexity of the optimum receiver is exponential in the number of users, which makes it prohibitive in terms of practical implementation. The research efforts that followed were devoted to detectors that provided significant improvement over

the conventional one, but with a simpler structure in comparison to the optimum receiver allowing thus for their practical realization. An important example of such a receiver is a decorrelating detector that performs inversion of correlation matrix of users' signature sequences [2] so that its output is multiuser interference-free and its bit-error rate is therefore independent of the interfering amplitudes. In addition to this near-far resistance property, the decorrelating detector does not require an estimation of the received signal amplitudes. A somewhat inferior, but greatly simplified form of a decorrelator is the one with the processing window restricted to the bit of interest only (one-shot decorrelator). The one-shot decorrelator was introduced in [3] through a two-asynchronous-user example, while independently the same type of receiver was described in [4]. In the base station, with  $K$  active users in the system, all of them need to be demodulated, and in order to avoid the degradation of the decorrelated users' signal-to-noise ratios (SNRs), the processing window spans almost twice the length of the bit interval. The inclusion of the entire bit of every user means that a  $(3K - 2) \times (3K - 2)$  cross correlation matrix has to be inverted.

A minimum mean-square error (MMSE) detector which has both the conventional receiver and the decorrelator as its limiting cases was first proposed in [5]. It obtains a linear estimate of the transmitted bits from the matched-filter output by minimizing either a mean-squared error or a weighted-squared error performance criterion. As the SNRs go to infinity, the MMSE detector converges to the decorrelator. But as the SNRs go to zero, the MMSE detector approaches the conventional detector, which is better than the decorrelator for sufficiently low SNRs. The complexity of the MMSE detector is linear in the number of users, but it also requires knowledge of the received amplitudes. Adaptive linear MMSE detectors that do not require estimation of the received amplitudes were proposed in [6]–[8]. They, however, require the use of training sequences for the adaptive filter coefficients to settle to their proper values.

Another approach in multiuser detection is based on interference cancellation. It can be done in a single stage or in multiple stages. In the first case, the users are ordered according to their received amplitude levels and the strongest one, the one that provides the most reliable data bit estimate, is detected first. It is then removed (subtracted) from the other users. Because it is also the most damaging interferer, the overall multiuser interference level will be significantly reduced. The process is then repeated with the next strongest user until all of them are detected [9]. The technique requires very accurate estimates of the

Paper approved by B. Aazhang, the Editor for Spread Spectrum Networks of the IEEE Communications Society. Manuscript received September 25, 1995; revised July 17, 1997 and March 8, 2000.

L. C. Zhong was with the New Jersey Institute of Technology, Newark NJ 07102 USA. He is now with the Department of EECS, University of California at Berkeley, Berkeley, CA 94702 USA (e-mail: czhong@cory.eecs.berkeley.edu).

Z. Siveski and R. E. Kamel were with the New Jersey Institute of Technology, Newark NJ 07102 USA. They are now with Lucent Technologies, Whippany, NJ 07981 USA.

N. Ansari is with the ECE Department, New Jersey Institute of Technology, Newark, NJ 07102 USA.

Publisher Item Identifier S 0090-6778(00)07539-5.

received amplitudes. Unless the subtraction is done in the proper order of decreasing amplitude levels, the achieved performance may be worse when compared to that of the decorrelator, which uses no amplitude estimates.

Multistage detection was originally proposed in [10], where the detector's first stage is a bank of matched filters (conventional detector), and in [11], where it is a decorrelator. The cancellation occurs in the second stage, where the multiuser interference is estimated from the first-stage (tentative) decisions, and then subtracted from the received signal. When the first-stage decisions are error-free, the detector achieves the performance of the single-user channel. On the other hand, any erroneous tentative decision would double the interference from the corresponding user. Such a detector utilizes fixed weights that requires prior estimation of the received signals' amplitudes.

In searching for an adaptive structure that neither requires training sequences nor prior estimation of the received signals' amplitudes, an adaptive asynchronous version of a multistage receiver in [11] is proposed and analyzed in this paper. This nonlinear receiver uses the decorrelator as its first stage. It is near-far resistant and belongs to the category of blind multiuser detectors.<sup>1</sup> To see this property, there is no need to derive an expression for the near-far resistance of the receiver, which is defined as the worst-case asymptotic multiuser efficiency. We just need to show the asymptotic multiuser efficiency can never be zero. In the absence of noise, the estimates of interferences made by the decorrelating stage will be perfect and the interferences can be tuned out completely in the next stage, therefore the bit-error rate (BER) will vanish. As a result, the asymptotic multiuser efficiency will always be nonzero. The near-far resistance of the receiver is therefore nonzero and the receiver is near-far resistant.

The proposed receiver obtains tentative decisions from the outputs of a one-shot decorrelator. In this paper, the decorrelator is realized as a parallel structure which in terms of complexity requires inversion of only  $K$  symmetric  $K \times K$  matrices for all  $K$  users, and has latency of one bit only. Such decorrelator's architecture makes it suitable for parallel processing implementation in base stations.

For controlling the weights, the proposed adaptive part of the detector uses a steepest descent algorithm that minimizes the power at the output. The steady-state error performance is evaluated and comparisons are made with the conventional and the decorrelating detectors. It is followed by an investigation on the convergence and stability of the receiver. The conditions for the detector to achieve convergence are derived, and their properties analyzed. Since the steepest descent algorithm requires the knowledge of the channel statistics, a stochastic version of the algorithm is considered, with the average of the error function computed over a number of iterations. Averaging reduces the variance of the error function estimate, but entails storage requirement. Additionally, the choice of the step size of the algorithm enables a tradeoff between the speed of convergence and the average residual error.

<sup>1</sup>A linear blind multiuser detector for a single-user detection (decentralized detector) that is near-far resistant and requires no more knowledge than the conventional single-user detector was reported in [12].

As an alternative approach, the adaptation of the weights using the recursive least squares (RLS)-type algorithm based on the one described in [13] is also presented. The increased computational complexity of the algorithm provides a superior speed of convergence for a given amount of residual error.

The paper is organized as follows. The general description of the receiver is introduced in Section II. The parallel structure of the decorrelator is presented in Section III. The adaptation procedure is described in Section IV. Analytical evaluation of the steady-state error probability is given in Section V. The convergence of the adaptation procedure is proved in Section VI, and sufficient conditions for convergence are derived. The impact of parameters on convergence speed is discussed. The superiority of the RLS algorithm is also proved there. Section VII provides several numerical examples and a discussion.

## II. PRELIMINARIES

In a multiuser environment,  $K$  users share the same channel with the unit energy signature rectangular waveform  $s_k(t)$ ,  $k = 1, \dots, K$  of the duration  $T$  assigned to each of them. The information bits  $b_k(i) \in \{-1, +1\}$  for the symbol interval  $i$  have the same duration  $T$ . The waveform  $r(t)$  at the input of the receiver that has a bank of matched filters as its front end is expressed as

$$r(t) = \sum_{k=1}^K \sum_i b_k(i) \sqrt{a_k} s_k(t - iT - \tau_k) + n_w(t)$$

where  $n_w(t)$  is a zero-mean, white Gaussian noise with the two-sided power spectral density  $N_0/2$ , and  $a_k$  and  $\tau_k$  are the received energy and relative delay for user  $k$ , respectively. While it is assumed that precise relative delay estimates are available for all users, their amplitudes are considered to be unknown to the receiver. In reality, they are slowly time varying; however, here they are assumed to remain unchanged over the transmission horizon for each user.

Without loss of generality, the attention will be on detection of user 1, and it will be assumed that  $0 = \tau_1 \leq \tau_2 \leq \dots \leq \tau_K \leq \tau_{K+1} = T$ . The sampled output of the matched filter of bit  $i$  of user 1 is then

$$x_1(i) = \sqrt{a_1} b_1(i) + \sum_{k=2}^K \sqrt{a_k} \cdot [\rho_{k1} b_k(i-1) + \rho_{1k} b_k(i)] + n_1(i).$$

The partial cross correlations  $\rho_{k1}$  and  $\rho_{1k}$ , for  $k = 2, \dots, K$ , are

$$\rho_{k1} = \int_0^T s_1(t) s_k(t + T - \tau_k) dt,$$

$$\rho_{1k} = \int_0^T s_1(t) s_k(t - \tau_k) dt.$$

Also,  $n_1(i)$  is a zero-mean Gaussian random variable with variance  $N_0/2$ . Using the vector notations where

$$\boldsymbol{\rho}_1 = [\rho_{21}, \dots, \rho_{K1}, \rho_{12}, \dots, \rho_{1K}]^T$$

$$\mathbf{b}_1(i) = [b_2(i-1), \dots, b_K(i-1), b_2(i), \dots, b_K(i)]^T$$

$$\mathbf{A}_1 = \text{diag}[\sqrt{a_2}, \dots, \sqrt{a_K}]$$

and

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_1 \end{bmatrix}$$

the matched filter output is expressed as

$$x_1(i) = \sqrt{a_1}b_1(i) + \rho_1^T \mathbf{A}b_1(i) + n_1(i). \quad (1)$$

The multistage detector introduced in [10], forms an estimate of the multiuser interference as the weighted vector of tentative decisions on symbols that interfere with  $b_1(i)$  directly, and then subtracts this estimate from the matched filter output. The generic expressions for the final decision statistics  $y_1(i)$  and the corresponding final decision  $\hat{b}_1(i)$  for bit  $i$  of user 1 are

$$y_1(i) = x_1(i) - \mathbf{w}_1^T(i)\tilde{\mathbf{b}}_1(i), \quad \hat{b}_1(i) = \text{sgn}(y_1(i)), \quad (2)$$

where  $\tilde{\mathbf{b}}_1(i) = [\tilde{b}_2(i-1), \dots, \tilde{b}_K(i-1), \tilde{b}_2(i), \dots, \tilde{b}_K(i)]^T$  is the vector of tentative decisions affecting bit  $i$  of user 1, and  $\mathbf{w}_1(i) = [w_{21}^{(i-1)}, \dots, w_{K1}^{(i-1)}, w_{21}^{(i)}, \dots, w_{K1}^{(i)}]^T$  are the corresponding weights. In [10] and [11], the tentative decisions are obtained either from the decorrelator in the first stage or from the output of the bank of matched filters, while the weights are formed using the estimated received signals' amplitudes.

In this paper, the tentative decisions will be provided by the one-shot decorrelator. The vector of the one-shot decorrelator outputs  $\mathbf{z}'_1(i)$  and the corresponding tentative decisions  $\tilde{\mathbf{b}}_1(i)$  affecting bit  $i$  of user 1 are expressed as

$$\mathbf{z}'_1(i) = \mathbf{A}b_1(i) + \boldsymbol{\xi}'_1, \quad \tilde{\mathbf{b}}_1(i) = \text{sgn}(\mathbf{z}'_1(i)) \quad (3)$$

where  $\mathbf{z}'_1(i) = [z_2(i-1), \dots, z_K(i-1), z_2(i), \dots, z_K(i)]^T$ , and the noise at the output of the decorrelator  $\boldsymbol{\xi}'_1 = [\xi_2(i-1), \dots, \xi_K(i-1), \xi_2(i), \dots, \xi_K(i)]^T$  is a zero-mean Gaussian vector having the covariance matrix  $\boldsymbol{\Xi}_1$ , whose diagonal elements are evaluated in the following section.

The one-shot decorrelator was originally presented in [3]. The proposed parallel structure realization of such a decorrelator, which is particularly suitable for the multistage detection, is described next.

### III. PARALLEL DECORRELATOR STRUCTURE

To facilitate the description of the proposed parallel structure of the decorrelator, the signature waveform of each user in the  $i$ th bit interval of user 1,  $[(i-1)T, iT]$ , is shown in Fig. 1 as being partitioned into  $K$  blocks according to users' relative delays. Each block of time  $[\tau_j, \tau_{j+1}]$ ,  $j = 1, 2, \dots, K$ , can then be viewed as a  $K$ -user synchronous channel with unit-energy rectangular signature waveforms  $\tilde{s}_k^{(j)}(t) = s_k^{(j)}(t)/\sqrt{e_j}$ ,  $k = 1, 2, \dots, K$ , where

$$s_k^{(j)}(t) = \begin{cases} s_k(t + \beta_{kj}T - \tau_k), & \tau_j < t < \tau_{j+1} \\ 0, & \text{otherwise} \end{cases}$$

$$\beta_{kj} = \begin{cases} 1, & j < k \\ 0, & \text{otherwise.} \end{cases}$$

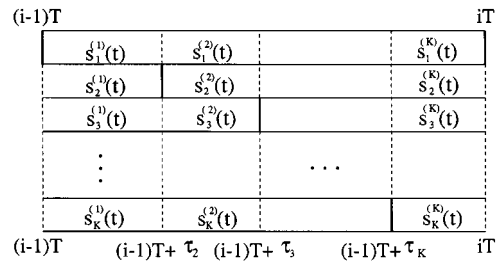


Fig. 1. Partition of the  $i$ th bit interval of user 1 into  $K$  blocks.

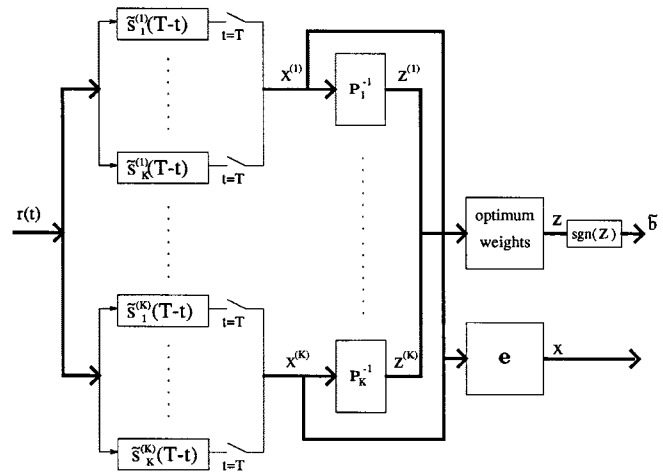


Fig. 2. Proposed decorrelator structure.

The partial energy  $e_j$  of  $s_k(t)$  within a particular block is given as

$$e_j = \int_{\tau_j}^{\tau_{j+1}} s_k^2(t + \beta_{kj}T - \tau_k) dt$$

$$= \int_0^T [s_k^{(j)}(t)]^2 dt, \quad k = 1, \dots, K$$

which, given the assumption that  $s_k(t)$  is rectangular<sup>2</sup> is equal to  $(\tau_{j+1} - \tau_j)/T$ .

The received signal  $r(t)$  in the block  $[\tau_j, \tau_{j+1}]$  can then be expressed as

$$r(t) = \sqrt{e_j} \left[ \sum_{k=1}^j \sqrt{a_k}b_k(i)\tilde{s}_k^{(j)}(t) + \sum_{k=j+1}^K \sqrt{a_k}b_k(i-1)\tilde{s}_k^{(j)}(t) \right] + n_w(t).$$

The essential structure of the decorrelator is shown in Fig. 2 in which the vector of sampled outputs of the bank of matched filters for the  $j$ th block is

$$\mathbf{x}^{(j)} = \sqrt{e_j}\mathcal{P}_j\mathbf{A}\mathbf{b}^{(j)} + \mathbf{n}^{(j)}$$

<sup>2</sup>The assumption implies that  $e_j$  is independent of  $k$ , although the analysis can be extended to a more general case in which the partial energy within each block depends on  $k$ .

where  $\mathcal{P}_j$  is the  $K \times K$  symmetric cross correlation matrix of signature waveforms  $\tilde{s}_k^{(j)}(t)$  in which the  $(l, m)$ th element  $\rho_{lm}^{(j)}$  is defined as

$$\rho_{lm}^{(j)} = \frac{1}{e_j} \int_0^T s_l^{(j)}(t) s_m^{(j)}(t) dt$$

and the diagonal elements are equal to one. Additionally,  $\mathbf{b}^{(j)} = [b_1(i), b_2(i), \dots, b_j(i), b_{j+1}(i-1), \dots, b_K(i-1)]^T$ , and  $\mathbf{n}^{(j)} = [n_1^{(j)}, n_2^{(j)}, \dots, n_K^{(j)}]^T$  is a Gaussian noise vector with a covariance matrix  $E\{\mathbf{n}^{(j)}\mathbf{n}^{(j)T}\} = (N_0/2)\mathcal{P}_j$ . (In all notations the superscript  $(j)$  denotes  $j$ th block, while the subscript usually refers to a particular user and the time index will be dropped whenever possible).

Assuming that  $\mathcal{P}_j$  is invertible and  $\mathbf{\Gamma}_j = \mathcal{P}_j^{-1}$ , the decorrelated vector output for the  $j$ th block is given as

$$\mathbf{z}^{(j)} = [z_1^{(j)}, z_2^{(j)}, \dots, z_K^{(j)}]^T = \mathcal{P}_j^{-1} \mathbf{x}^{(j)} = \sqrt{e_j} \mathbf{A} \mathbf{b}^{(j)} + \boldsymbol{\xi}^{(j)}$$

in which the  $k$ th element

$$z_k^{(j)} = \sqrt{e_j} \sqrt{a_k} b_k(i - \beta_{kj}) + \xi_k^{(j)} \quad (4)$$

and

$$\boldsymbol{\xi}^{(j)} = [\xi_1^{(j)}, \xi_2^{(j)}, \dots, \xi_K^{(j)}]^T = \mathcal{P}_j^{-1} \mathbf{n}^{(j)} = \mathbf{\Gamma}_j \mathbf{n}^{(j)}$$

is a Gaussian noise vector with a covariance matrix  $E\{\boldsymbol{\xi}^{(j)}\boldsymbol{\xi}^{(j)T}\} = (N_0/2)\mathbf{\Gamma}_j$ .

The next step in obtaining the decorrelator output for user  $k$  is to consider the vector

$$\begin{aligned} \mathbf{z}_k &= [z_k^{(1)}(i+1), z_k^{(2)}(i+1), \dots, z_k^{(k-1)}(i+1), \\ &\quad z_k^{(k)}(i), \dots, z_k^{(K)}(i)]^T \\ &= \sqrt{a_k} b_k \mathbf{e} + \boldsymbol{\xi}_k \end{aligned} \quad (5)$$

in which each element represents one of  $K$  nonoverlapping blocks corresponding to the bit  $i$  of user  $k$ ,<sup>3</sup> and where  $\mathbf{e} = [\sqrt{e_1}, \dots, \sqrt{e_K}]^T$  and

$$\boldsymbol{\xi}_k = [\xi_k^{(1)}(i+1), \xi_k^{(2)}(i+1), \dots, \xi_k^{(k-1)}(i), \xi_k^{(k)}(i), \dots, \xi_k^{(K)}(i)]^T.$$

Since  $E\{\boldsymbol{\xi}^{(j)}\boldsymbol{\xi}^{(j)T}\} = (N_0/2)\mathbf{\Gamma}_j$ , the covariance matrix of  $\boldsymbol{\xi}_k$  is

$$E\{\boldsymbol{\xi}_k \boldsymbol{\xi}_k^T\} = \frac{N_0}{2} \text{diag} [\gamma_k^{(1)}, \gamma_k^{(2)}, \dots, \gamma_k^{(K)}]$$

where  $\gamma_k^{(j)}$  is the  $k$ th diagonal element of the matrix  $\mathbf{\Gamma}_j$ .

<sup>3</sup>The argument  $(i+1)$  in some of the elements of  $\mathbf{z}_k$  (all belonging to user  $k$ ) refers to the  $(i+1)$ th interval along the time axis, not to the  $i$ th information bit; the "inconsistency" is due to the alignment of the  $i$ th interval ( $t \in [(i-1)T, iT]$ ) with the  $i$ th bit of user 1.

The decorrelator output for user  $k$ ,  $z_k$ , can then be constructed as the weighted sum of the elements of the vector

$$z_k = \mathbf{c}_k^T \mathbf{z}_k = \sum_{j=1}^K c_k^{(j)} \sqrt{e_j} \sqrt{a_k} b_k(i) + \xi_k \quad (6)$$

where  $\mathbf{c}_k = [c_k^{(1)}, \dots, c_k^{(K)}]^T$ , and  $\xi_k = \mathbf{c}_k^T \boldsymbol{\xi}_k$ .

The optimal selection of the weights  $\mathbf{c}_k$  minimizes the bit-error probability of the tentative decision<sup>4</sup>  $\tilde{b}_k(i) = \text{sgn}(z_k(i))$  at the decorrelator output. The evaluation of the optimal weights can be done in a straightforward, although somewhat tedious manner by minimizing the error probability at the decorrelator output

$$\Pr\{\tilde{b}_k(i) \text{ in error}\} = Q\left(\frac{\left[\sum_{j=1}^K c_k^{(j)} \sqrt{e_j}\right] \sqrt{a_k}}{\sqrt{\text{var}(\xi_k)}}\right)$$

where  $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$

which in this case is the same as maximizing the SNR. By recognizing, however, that the elements of  $\mathbf{z}_k$  represent  $K$  independent observations of  $b_k(i)$ , the expression for the optimal  $\mathbf{c}_k$  can be written directly, as shown in Appendix A, using the well-known solution for the MRC<sup>5</sup> [14]. One can also observe that premultiplying the optimum values of the weights  $\mathbf{c}_k$  from (A1) by any nonzero constant will not change the error probability. Here, the constant is chosen as

$$\Gamma_k = \frac{1}{\sum_{j=1}^K \frac{e_j}{\gamma_k^{(j)}}}$$

so that when the final results are substituted into (6), the  $k$ th element of the decorrelator output vector  $\mathbf{z}$ , shown in Fig. 2, where  $\mathbf{z} = [z_1, z_2, \dots, z_K]^T$ , becomes

$$z_k = \sqrt{a_k} b_k(i) + \xi_k. \quad (7)$$

The variance of the noise component is  $\text{var}(\xi_k^{(j)}) = \Gamma_k$  and  $\Pr\{\tilde{b}_k(i) \text{ in error}\} = Q(\sqrt{a_k}/\Gamma_k)$ . The vector  $\mathbf{z}'_k(i)$  in (3), required for the second stage of the proposed receiver, is then easily created. For processing one bit of all  $K$  users, the proposed realization of the decorrelator requires the inversion of  $K$  matrices with the dimensions  $K \times K$ . The procedure, however, can be done in parallel resulting in a complexity of  $O(K^3)$  per each inversion. Additionally, unlike [4], the latency is only one bit interval.

In Fig. 2, one also can observe that by using a linear combination of the matched filters' outputs, a sampled vector output of the conventional CDMA detector  $\mathbf{x} = [x_1, x_2, \dots, x_K]^T$  is obtained. For the  $k$ th user, we have  $x_k = \mathbf{x}_k^T \mathbf{e}$ , where

$$\mathbf{x}_k = [x_k^{(1)}(i+1), x_k^{(2)}(i+1), \dots, x_k^{(k-1)}(i+1), x_k^{(k)}(i), \dots, x_k^{(K)}(i)]^T.$$

<sup>4</sup>As mentioned toward the end of Section II, the tentative decisions are obtained from the decorrelator's outputs.

<sup>5</sup>This was pointed out by an anonymous reviewer.

The conventional output is needed to form the final decision statistics given by (2).

#### IV. ADAPTATION PROCEDURE

The proposed detector evaluates the weights  $\mathbf{w}_1(i)$  in (2) in an adaptive manner making received signal amplitudes estimations unnecessary. An algorithm which minimizes the output power  $E\{y_1^2(i)\}$  is employed for controlling the weights. A simple physical justification can be found by observing the output signal-to-multiuser-interference-plus-noise ratio (SINR) as a relevant performance measure, which is expressed as

$$\text{SINR} = \frac{a_1}{E\{y_1^2(i)\} - a_1}. \quad (8)$$

It is obvious that the output SINR will be maximized when the weights are selected in such a way that the output energy  $E\{y_1^2(i)\}$  is minimized. The weights that achieve this are obtained by an iterative search

$$\begin{aligned} \mathbf{w}_1(i+1) &= \mathbf{w}_1(i) - \frac{\mu}{2} \frac{\partial}{\partial \mathbf{w}_1(i)} E\{y_1^2(i)\} \\ &= \mathbf{w}_1(i) + \mu E\{y_1(i)\tilde{\mathbf{b}}_1(i)\} \end{aligned} \quad (9)$$

where  $\mu$  is a positive number. The algorithm, by minimizing the output power, forces the correlation between the desired output signal and vector of tentative decisions of interfering signals to zero. In a practical implementation, one would replace the statistical expectation in (9) by time averaging, or preferably by its instantaneous estimate  $y_1(i)\tilde{\mathbf{b}}_1(i)$ . Therefore, in the latter case, the weight update equation becomes

$$\begin{aligned} \mathbf{w}_1(i+1) &= \mathbf{w}_1(i) + \mu y_1(i)\tilde{\mathbf{b}}_1(i) \\ &= \mathbf{w}_1(i) + \mu [\tilde{\mathbf{b}}_1(i)x_1(i) - \tilde{\mathbf{b}}_1(i)\tilde{\mathbf{b}}_1^T(i)\mathbf{w}_1(i)]. \end{aligned} \quad (10)$$

As an alternative, an RLS-type adaptive algorithm is considered, in which at time  $i$  the weight vector  $\mathbf{w}_1^T(i)$  is chosen that minimizes  $\sum_{j=0}^i \lambda^{i-j} E\{[x_1(j) - \mathbf{w}_1^T(i)\tilde{\mathbf{b}}_1(j)]^2\}$  or in a practical implementation  $\sum_{j=0}^i \lambda^{i-j} [x_1(j) - \mathbf{w}_1^T(i)\tilde{\mathbf{b}}_1(j)]^2$ , where  $\lambda$  is called the forgetting factor and  $0 < \lambda < 1$ , i.e.,

$$\frac{\partial}{\partial \mathbf{w}_1^T(i)} \sum_{j=0}^i \lambda^{i-j} [x_1(j) - \mathbf{w}_1^T(i)\tilde{\mathbf{b}}_1(j)]^2 = 0, \quad (11)$$

The solution of the above equation is

$$\mathbf{w}_1(i) = \tilde{\mathbf{R}}^{-1}(i)\mathbf{d}(i), \quad (12)$$

where

$$\tilde{\mathbf{R}}(i) = \sum_{j=0}^i \lambda^{i-j} \tilde{\mathbf{b}}_1(j)\tilde{\mathbf{b}}_1^T(j)$$

and

$$\mathbf{d}(i) = \sum_{j=0}^i \lambda^{i-j} \tilde{\mathbf{b}}_1(j)x_1(j) \quad (13)$$

and  $\tilde{\mathbf{R}}(i)$  and  $\mathbf{d}(i)$  follow the recursive equations

$$\tilde{\mathbf{R}}(i+1) = \lambda\tilde{\mathbf{R}}(i) + \tilde{\mathbf{b}}_1(i+1)\tilde{\mathbf{b}}_1^T(i+1)$$

and

$$\mathbf{d}(i+1) = \lambda\mathbf{d}(i) + \tilde{\mathbf{b}}_1(i+1)x_1(i+1).$$

The inverse in (12) can be obtained using the matrix inversion lemma

$$\tilde{\mathbf{R}}^{-1}(i+1) = \frac{1}{\lambda} \left( \tilde{\mathbf{R}}^{-1}(i) - \mathbf{k}(i+1)\tilde{\mathbf{b}}_1^T(i+1)\tilde{\mathbf{R}}^{-1}(i) \right)$$

where

$$\mathbf{k}(i+1) = \frac{1}{\lambda + \chi_i} \tilde{\mathbf{R}}^{-1}(i)\tilde{\mathbf{b}}_1(i+1)$$

and

$$\chi_i = \tilde{\mathbf{b}}_1^T(i+1)\tilde{\mathbf{R}}^{-1}(i)\tilde{\mathbf{b}}_1(i+1).$$

The weight update equation is [14, pp. 658]

$$\begin{aligned} \mathbf{w}_1(i+1) &= \mathbf{w}_1(i) - \frac{\mu_i}{2} \tilde{\mathbf{R}}^{-1}(i) \frac{\partial}{\partial \mathbf{w}_1^T(i)} \\ &\quad \cdot \left( x_1(i+1) - \tilde{\mathbf{b}}_1^T(i+1)\mathbf{w}_1(i) \right)^2 \\ &= \mathbf{w}_1(i) + \mathbf{k}(i+1) \left[ x_1(i+1) - \tilde{\mathbf{b}}_1^T(i+1)\mathbf{w}_1(i) \right] \\ &= \mathbf{w}_1(i) + \mu_i \tilde{\mathbf{R}}^{-1}(i) \\ &\quad \cdot \left[ \tilde{\mathbf{b}}_1(i+1)x_1(i+1) - \tilde{\mathbf{b}}_1(i+1)\tilde{\mathbf{b}}_1^T(i+1)\mathbf{w}_1(i) \right] \end{aligned} \quad (14)$$

where

$$\mu_i = \frac{1}{\lambda + \chi_i}.$$

#### V. STEADY-STATE ERROR PERFORMANCE

To make an error performance comparison of the proposed multistage adaptive detector with that of the conventional receiver and the decorrelating detector, the steady-state behavior of the former is considered.

The steady-state values of the weights after completion of the iterative search in (9) are evaluated as follows:

$$\begin{aligned} &-\frac{1}{2} \frac{\partial}{\partial \mathbf{w}_1(i)} E\{y_1^2(i)\} \\ &= E\{y_1(i)\tilde{\mathbf{b}}_1(i)\} \\ &= E\{x_1(i)\tilde{\mathbf{b}}_1(i)\} - E\{\tilde{\mathbf{b}}_1(i)\tilde{\mathbf{b}}_1^T(i)\}\mathbf{w}_1(i) = 0. \end{aligned} \quad (15)$$

It is easy to show that  $E\{n_1(i)\xi_k(i+j)\} = 0$ , for  $k \neq 1$ ,  $j = -1, 0$ , so the first of the two expectations in (15) is

$$E\{x_1(i)\tilde{\mathbf{b}}_1(i)\} = \mathbf{A}E\{\mathbf{b}_1(i)\tilde{\mathbf{b}}_1^T(i)\}\boldsymbol{\rho}_1. \quad (16)$$

Matrix  $E\{\mathbf{b}_1(i)\tilde{\mathbf{b}}_1^T(i)\}$  is diagonal; therefore, the system of  $2K - 2$  linear equations (15), together with (16), gives the steady-state values of the weights affecting the first output as

$$\mathbf{w}_1^0 = \left[ E\{\tilde{\mathbf{b}}_1(i)\tilde{\mathbf{b}}_1^T(i)\} \right]^{-1} \mathbf{A}E\{\mathbf{b}_1(i)\tilde{\mathbf{b}}_1^T(i)\}\boldsymbol{\rho}_1. \quad (17)$$

The expressions for the expectations in (17) are shown in Appendix B.

By taking the expectation of (11), one can show that the steady-state values of the weights for the RLS-type of algorithm are the same as the one in (17).

The output error probability is evaluated as follows:

$$\begin{aligned}
P_{e_1} &= E_{b_1(i), \tilde{b}_1(i)} \Pr \left\{ \hat{b}_1 \text{ in error} \mid b_1(i), \mathbf{b}_1(i), \tilde{\mathbf{b}}_1(i) \right\} \\
&= \frac{1}{2} \sum_{b_1(i), \tilde{b}_1(i)} \left[ \Pr \left\{ n_1(i) > \sqrt{a_1} - \boldsymbol{\rho}_1^T \mathbf{A} \mathbf{b}_1(i) + \mathbf{w}_1^{0T} \tilde{\mathbf{b}}_1(i) \right\} \right. \\
&\quad \left. + \Pr \left\{ n_1(i) < -\sqrt{a_1} - \boldsymbol{\rho}_1^T \mathbf{A} \mathbf{b}_1(i) + \mathbf{w}_1^{0T} \tilde{\mathbf{b}}_1(i) \right\} \right] \\
&\quad \cdot \Pr \left\{ \tilde{\mathbf{b}}_1(i) \mid \mathbf{b}_1(i) \right\} \Pr \left\{ \mathbf{b}_1(i) \right\}.
\end{aligned}$$

Since  $\Pr \left\{ \mathbf{b}_1(i) \right\} = 2^{-(2K-2)}$ , and

$$\Pr \left\{ \tilde{\mathbf{b}}_1(i) \mid \mathbf{b}_1(i) \right\} = \Pr \left\{ -\tilde{\mathbf{b}}_1(i) \mid -\mathbf{b}_1(i) \right\}$$

the above error probability can be written as

$$\begin{aligned}
P_{e_1} &= 2^{-(2K-2)} \\
&\quad \cdot \sum_{b_1(i), \tilde{b}_1(i)} \left[ \Pr \left\{ n_1(i) > \sqrt{a_1} - \boldsymbol{\rho}_1^T \mathbf{A} \mathbf{b}_1(i) + \mathbf{w}_1^{0T} \tilde{\mathbf{b}}_1(i) \right\} \right] \\
&\quad \cdot \Pr \left\{ \tilde{\mathbf{b}}_1(i) \mid \mathbf{b}_1(i) \right\}.
\end{aligned}$$

Finally

$$\begin{aligned}
P_{e_1} &= 2^{-(2K-2)} \sum_{b_1(i), \tilde{b}_1(i)} \Pr \left\{ \tilde{\mathbf{b}}_1(i) \mid \mathbf{b}_1(i) \right\} \\
&\quad \cdot Q \left( \frac{\sqrt{a_1} - \boldsymbol{\rho}_1^T \mathbf{A} \mathbf{b}_1(i) + \mathbf{w}_1^{0T} \tilde{\mathbf{b}}_1(i)}{\sqrt{N_0/2}} \right) \quad (18)
\end{aligned}$$

where  $\Pr \left\{ \tilde{\mathbf{b}}_1(i) \mid \mathbf{b}_1(i) \right\}$  is the integral of the  $(2K-2)$ -variate Gaussian density function

$$\begin{aligned}
\Pr \left\{ \tilde{\mathbf{b}}_1(i) \mid \mathbf{b}_1(i) \right\} &= \frac{1}{\sqrt{(2\pi)^{2K-2} |\boldsymbol{\Xi}_1|}} \\
&\quad \cdot \int \cdots \int_{\substack{\tilde{b}_k(j) \xi_k(j) = -\sqrt{a_k} b_k(j) \tilde{b}_k(j) \\ k=2, \dots, K \\ j=i-1, i}} \exp \left( -\frac{1}{2} \boldsymbol{\xi}_1^T \boldsymbol{\Xi}_1^{-1} \boldsymbol{\xi}_1 \right) d\boldsymbol{\xi}_1.
\end{aligned}$$

## VI. CONVERGENCE AND STABILITY

As with any adaptive algorithm, we consider the two major issues: convergence and stability. To analyze the convergence and stability of the algorithms considered, we take a closer look at the power as a cost function. We examine the parameters affecting the convergence to the minimum cost, as well as the range of these parameters that would guarantee stability.

For the steepest descent algorithm, the cost function to be minimized is expressed as

$$\begin{aligned}
\mathcal{E}(\mathbf{w}_1(i)) &= E \left\{ y_1^2(i) \right\} \\
&= E \left\{ \left[ x_1(i) - \tilde{\mathbf{b}}_1^T(i) \mathbf{w}_1(i) \right]^2 \right\} \\
&= E \left\{ x_1^2(i) \right\} - 2\mathbf{w}_1^T(i) E \left\{ x_1(i) \tilde{\mathbf{b}}_1(i) \right\} \\
&\quad + \mathbf{w}_1^{0T}(i) E \left\{ \tilde{\mathbf{b}}_1(i) \tilde{\mathbf{b}}_1^T(i) \right\} \mathbf{w}_1(i). \quad (19)
\end{aligned}$$

Denoting the optimum<sup>6</sup> weight vector by  $\mathbf{w}_1^0$ , the corresponding (lowest) value of the cost function, is

$$\begin{aligned}
\mathcal{E}(\mathbf{w}_1^0) &= E \left\{ x_1^2(i) \right\} - 2\mathbf{w}_1^{0T} E \left\{ x_1(i) \tilde{\mathbf{b}}_1(i) \right\} \\
&\quad + \mathbf{w}_1^{0T} E \left\{ \tilde{\mathbf{b}}_1(i) \tilde{\mathbf{b}}_1^T(i) \right\} \mathbf{w}_1^0. \quad (20)
\end{aligned}$$

To assess the speed of convergence, we consider the speed with which  $\mathcal{E}(\mathbf{w}_1(i))$  approaches  $\mathcal{E}(\mathbf{w}_1^0)$ , or the excess cost approaches zero.

$$\begin{aligned}
\mathcal{E}(\mathbf{w}_1(i)) - \mathcal{E}(\mathbf{w}_1^0) &= 2(\mathbf{w}_1^{0T} - \mathbf{w}_1^T(i)) E \left\{ x_1(i) \tilde{\mathbf{b}}_1(i) \right\} \\
&\quad + \mathbf{w}_1^T(i) E \left\{ \tilde{\mathbf{b}}_1(i) \tilde{\mathbf{b}}_1^T(i) \right\} \mathbf{w}_1(i) \\
&\quad - \mathbf{w}_1^{0T} E \left\{ \tilde{\mathbf{b}}_1(i) \tilde{\mathbf{b}}_1^T(i) \right\} \mathbf{w}_1^0. \quad (21)
\end{aligned}$$

One can express the excess cost in terms of the difference of the weights. Let  $\boldsymbol{\delta}_1(i) = \mathbf{w}_1^0 - \mathbf{w}_1(i)$ . The above equation becomes

$$\begin{aligned}
\mathcal{E}(\mathbf{w}_1(i)) - \mathcal{E}(\mathbf{w}_1^0) &= \boldsymbol{\delta}_1^T(i) E \left\{ \tilde{\mathbf{b}}_1(i) \tilde{\mathbf{b}}_1^T(i) \right\} \boldsymbol{\delta}_1(i) \\
&= \boldsymbol{\delta}_1^T(i) \mathbf{R} \boldsymbol{\delta}_1(i). \quad (22)
\end{aligned}$$

To examine the parameters affecting the convergence of the algorithm, we will derive a recursive relation for the excess cost. We can rewrite (9) as

$$\begin{aligned}
\mathbf{w}_1(i+1) &= \mathbf{w}_1(i) + \mu E \left\{ \tilde{\mathbf{b}}_1(i) x_1(i) - \tilde{\mathbf{b}}_1(i) \tilde{\mathbf{b}}_1^T(i) \mathbf{w}_1(i) \right\} \\
&= \left( \mathbf{I} - \mu E \left\{ \tilde{\mathbf{b}}_1(i) \tilde{\mathbf{b}}_1^T(i) \right\} \right) \mathbf{w}_1(i) + \mu E \left\{ \tilde{\mathbf{b}}_1(i) x_1(i) \right\}
\end{aligned}$$

it follows that

$$\boldsymbol{\delta}_1(i+1) = \left( \mathbf{I} - \mu E \left\{ \tilde{\mathbf{b}}_1(i) \tilde{\mathbf{b}}_1^T(i) \right\} \right) \boldsymbol{\delta}_1(i) = (\mathbf{I} - \mu \mathbf{R}) \boldsymbol{\delta}_1(i). \quad (23)$$

Therefore, the excess cost may be obtained recursively as

$$\begin{aligned}
&\boldsymbol{\delta}_1^T(i+1) \mathbf{R} \boldsymbol{\delta}_1(i+1) \\
&= \boldsymbol{\delta}_1^T(i) (\mathbf{I} - \mu \mathbf{R}^T) \mathbf{R} (\mathbf{I} - \mu \mathbf{R}) \boldsymbol{\delta}_1(i) \\
&= \boldsymbol{\delta}_1^T(i) \mathbf{R} \boldsymbol{\delta}_1(i) - 2\mu \boldsymbol{\delta}_1^T(i) \mathbf{R}^2 \boldsymbol{\delta}_1(i) + \mu^2 \boldsymbol{\delta}_1^T(i) \mathbf{R}^3 \boldsymbol{\delta}_1(i). \quad (24)
\end{aligned}$$

Denote by  $\alpha$  and  $\beta$  the minimum and the maximum eigenvalue, respectively, of a symmetric positive definite matrix  $\mathbf{U}$ . Then

$$\alpha \mathbf{v}^T \mathbf{v} \leq \mathbf{v}^T \mathbf{U} \mathbf{v} \leq \beta \mathbf{v}^T \mathbf{v}. \quad (25)$$

Applying (25) repeatedly to (24)

$$\boldsymbol{\delta}_1^T(i+1) \mathbf{R} \boldsymbol{\delta}_1(i+1) \leq (1 - 2\alpha\mu + \mu^2\beta^2) \boldsymbol{\delta}_1^T(i) \mathbf{R} \boldsymbol{\delta}_1(i). \quad (26)$$

The excess cost

$$\boldsymbol{\delta}_1^T(i+1) \mathbf{R} \boldsymbol{\delta}_1(i+1) \rightarrow 0, \quad \text{if } (1 - 2\alpha\mu + \mu^2\beta^2) < 1. \quad (27)$$

Then, it follows that

$$\boldsymbol{\delta}_1^T(i+1) \mathbf{R} \boldsymbol{\delta}_1(i+1) \rightarrow 0, \quad \text{if } \mu^2\beta^2 - 2\alpha\mu < 0. \quad (28)$$

<sup>6</sup>Optimum in this context refers to the weight that will attain the minimum cost function.

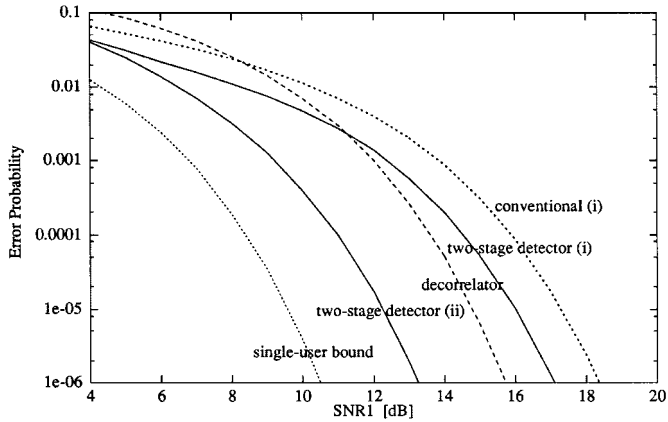


Fig. 3. Error probability of user 1 for  $K = 2$ ,  $\rho_{12} = 0.2$ ,  $\rho_{21} = 0.6$ ,  $e_1 = 0.4$ . (i)  $a_2/a_1 = 0.6$ . (ii)  $a_1/a_2 = 0.6$ .

The last inequality is satisfied if either  $\mu < 0$  and  $\mu > (2\alpha/\beta^2)$ , which is impossible, or  $\mu > 0$  and  $\mu < (2\alpha/\beta^2)$ .<sup>7</sup> From (22),  $0 < \mu < (2\alpha/\beta^2)$  are the sufficient conditions for convergence of the weights to their steady-state values. The smaller the eigenvalue spread, the larger  $\mu$  can be, resulting in higher convergence speed.

The analysis of the RLS algorithm is done by comparing the expression for its error function, given by (14), to that of the stochastic gradient algorithm, given by (10). The difference between the above expressions is in the occurrence of the term  $\tilde{\mathbf{R}}^{-1}$  in the former, such that in the steady state

$$\tilde{\mathbf{R}}^{-1}(i) \rightarrow \left[ E \left\{ \tilde{\mathbf{b}}_1(i) \tilde{\mathbf{b}}_1^T(i) \right\} \right]^{-1}.$$

So following similar steps leading to (23), we get

$$\boldsymbol{\delta}_1(i+1) = \left( \mathbf{I} - \mu_i \tilde{\mathbf{R}}^{-1}(i) \mathbf{R} \right) \boldsymbol{\delta}_1(i)$$

for the RLS algorithm, and hence the excess cost

$$\begin{aligned} & \boldsymbol{\delta}_1^T(i+1) \mathbf{R} \boldsymbol{\delta}_1(i+1) \\ &= \boldsymbol{\delta}_1^T(i) \tilde{\mathbf{R}}^{-1}(i) \mathbf{R} \boldsymbol{\delta}_1(i) - 2\mu_i \boldsymbol{\delta}_1^T(i) \left( \tilde{\mathbf{R}}^{-1}(i) \mathbf{R} \right)^2 \boldsymbol{\delta}_1(i) \\ & \quad + \mu_i^2 \boldsymbol{\delta}_1^T(i) \left( \tilde{\mathbf{R}}^{-1}(i) \mathbf{R} \right)^3 \boldsymbol{\delta}_1(i). \end{aligned} \quad (29)$$

At steady state  $\tilde{\mathbf{R}}^{-1}(i) \mathbf{R} \rightarrow \mathbf{I}$ . Therefore, as it is well known for the RLS algorithm [16], the eigenvalue spread will be smaller resulting in its faster convergence.

## VII. NUMERICAL EXAMPLES AND DISCUSSION

The steady-state error performance [with the optimal weights given by (17)] is illustrated in Figs. 3–5, while the transient error performance is shown in Figs. 6–8. In all the examples, the SNR for user  $k$  is defined as  $\text{SNR}_k = a_k/N_0$ . The relative energy  $e_1$  in the two-user examples is defined as  $e_1 = \int_{\tau_2}^T s_1^2(t) dt$ .

Fig. 3 shows the probability of error for user 1 versus  $\text{SNR}_1$ . The case labeled (i) in the same figure corresponds to a rela-

<sup>7</sup>An alternative derivation can be obtained in a similar fashion as done in [15]. If  $\mu$  is chosen to take any value within the range  $(0, 1)$ , the condition on  $\mu$  for the system to achieve convergence and stability is

$$\mu \leq (2/(1 + (K-2) d_{k\max}))$$

where  $d_{k\max} = \max_{i,j, i \neq j} |d_{ij}|$ , where  $d_{ij}$  is the  $(i, j)$ th element of matrix  $E\{\tilde{\mathbf{b}}_k \tilde{\mathbf{b}}_k^T\}$ .

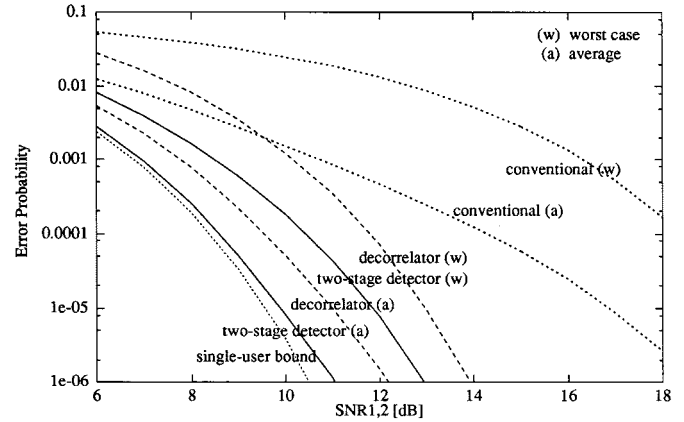


Fig. 4. Error probability of user 1 for  $K = 2$ .

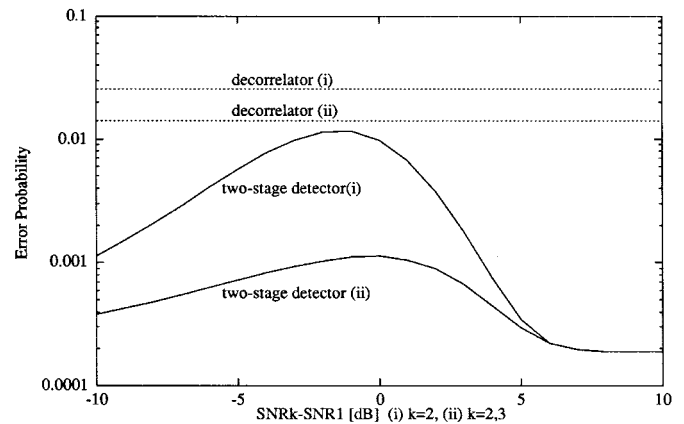
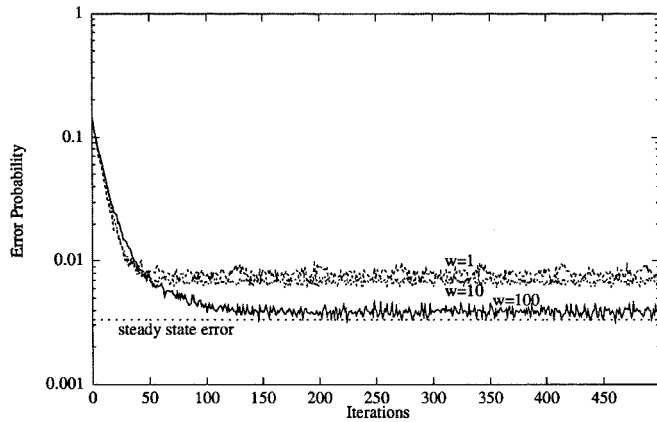


Fig. 5. Error probability of user 1 for  $\text{SNR}_1 = 8$  dB. (i)  $K = 2$ ,  $\rho_{12} = 0.2$ ,  $\rho_{21} = 0.6$ ,  $e_1 = 0.4$ . (ii)  $K = 3$ , Gold sequences (The three-user case outperforms the two-user case simply because of the parameters we are using. This is not true in general.).

tively weak level of interference. It describes a rather unfavorable scenario for this multistage detector due to tentative decisions which are unreliable. This results in the performance of the detector at the higher values of  $\text{SNR}_1$  to be somewhat inferior to the decorrelator, whose performance is insensitive to the level of interference. The multistage detector, however, outperforms the conventional detector. When the interference is stronger [case (ii)], due to the reliable tentative decisions, the multistage detector by far outperforms the other two, wherein the error probability (not shown in the figure) of the conventional receiver exceeds 0.1.

Fig. 4 shows the probability of error for user 1 when its energy and the energy of the interferer are the same, and Gold sequences of length 7 are used as the signature sequences. The worst case and the average error performance over the values of the relative delay  $\tau_2$  are shown. In this scenario, the multistage detector outperforms both the conventional and the decorrelating detector, and its average performance—due to the good cross correlation properties of the two signature sequences used—is very close to the single user-bound.

Fig. 5 shows the probability of error versus the relative energy of interferer(s) for the  $\text{SNR}_1$  fixed to 8 dB. For both the two-user scenario [case (i)] and the three-user scenario with Gold sequences of length 7 and  $\tau_2 = T/7$ ,  $\tau_3 = 5T/7$  [case

Fig. 6. Transient-error probability for  $\mu = 0.05$ .

(ii)], the multistage detector clearly outperforms the decorrelator, and approaches the single-user bound for the relative interference level above 5 dB (Here, the error probability of user 1 for the two-user scenario is lower than that of user 1 for the three-user case simply because of the parameters we are using. This is not the case in general).

In Figs. 6–8, the transient error probability curves were obtained by ensemble averaging over at least 50 000 independent trials of the simulation. The initial value of the weight vector  $\mathbf{w}_1$  was set to zero in each trial. Also  $\text{SNR}_1 = 8$  dB,  $a_1/a_2 = 0.6$ ,  $\rho_{12} = 0.2$ ,  $\rho_{21} = 0.6$ , and  $e_1 = 0.4$ . The updating rule in (9) used in the simulation was implemented as a sliding window of length  $W$ , i.e.,

$$E \left\{ y_1(i) \tilde{\mathbf{b}}_1(i) \right\} \approx \frac{1}{W} \sum_{j=i-W+1}^i y_1(i) \tilde{\mathbf{b}}_1(i).$$

Fig. 6 demonstrates the effect of the window length on the error, with the step size fixed at  $\mu = 0.05$ . When  $W = 100$  the steady-state error value is reached after about 50 iterations, while the stochastic gradient version ( $W = 1$ ) exhibits noticeable residual error. (For a smaller step size, e.g.,  $\mu = 0.01$ , no residual error was observed regardless of the value of  $W$ , but as expected the convergence was very slow—requiring about 150 iterations).

In Fig. 7, the effect of the value of the step size is examined, with  $W = 100$  which entails no residual error. The large step size  $\mu = 0.2$  when compared to  $\mu = 0.05$  does not offer any improvement in the convergence speed, because during the transient interval only a fraction of the length of the smoothing filter (window) is included.

Finally, in Fig. 8, it can be observed that the stochastic gradient algorithm, whose step size  $\mu$  is set to 0.01, provides roughly the same steady-state error performance as the RLS-type algorithm, whose  $\lambda = 0.999$ . The convergence speed superiority of the latter is evident. The stochastic gradient algorithm with  $\mu = 0.05$  provides approximately the same convergence rate as the RLS-type algorithm. The penalty that the former pays, however, is the substantially increased residual error.

## VIII. CONCLUSIONS

An analysis of a steady-state and transient error performance of a two stage multiuser detector is presented. The detector does

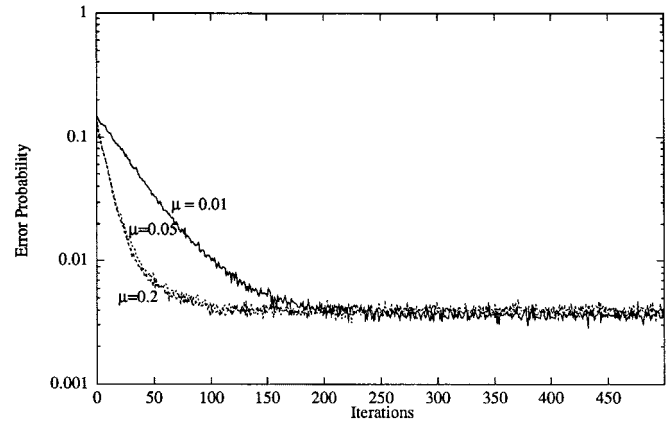
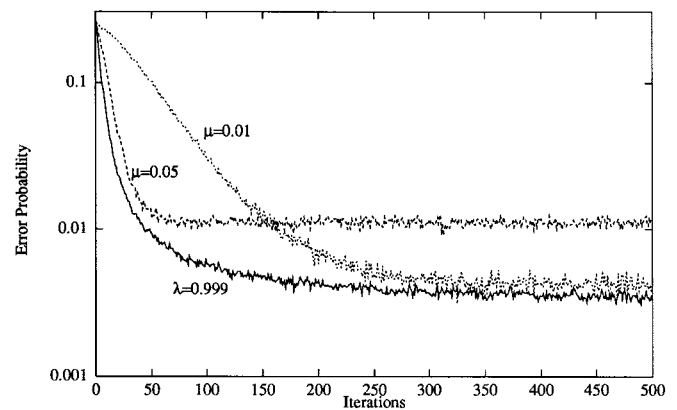
Fig. 7. Transient-error probability for  $W = 100$ .

Fig. 8. Transient-error probability.

not require prior estimation of the received signal amplitudes, or the use of training sequences. Its first stage is a one-shot decorrelator with a parallel structure that is particularly suitable for base station detection, and only requires an inversion of  $K \times K$  matrices in a  $K$  user system. The steady-state error performance of the detector is significantly better than that of the conventional receiver and generally outperforms the decorrelating detector as well. The stochastic gradient algorithm trades the speed of convergence with the residual error, while the RLS-type algorithm demonstrates a clear advantage over the former with regard to the speed of convergence for a given value of the residual error.

## APPENDIX A

Define  $\mathbf{T}_k = \text{diag}[\gamma_k^{(1)}, \gamma_k^{(2)}, \dots, \gamma_k^{(K)}]$ , after premultiplying (5) to obtain equal noise terms variances, we obtain

$$\mathbf{g}_k = (\mathbf{T}_k^{-1})^{1/2} \mathbf{z}_k = \sqrt{a_k} \mathbf{b}_k (\mathbf{T}_k^{-1})^{1/2} \mathbf{e} + (\mathbf{T}_k^{-1})^{1/2} \boldsymbol{\xi}_k.$$

Applying MRC on the elements of  $\mathbf{g}_k$

$$z_k = \mathbf{h}_k^T \mathbf{g}_k = \mathbf{h}_k^T (\mathbf{T}_k^{-1})^{1/2} \mathbf{z}_k$$

where  $\mathbf{h}_k = (\mathbf{T}_k^{-1})^{1/2} \mathbf{e}$ , we obtain the solution for the optimum  $\mathbf{c}_k$  in (6) as

$$\mathbf{c}_k = (\mathbf{T}_k^{-1}) \mathbf{e}. \quad (\text{A1})$$



APPENDIX B

The elements on the main diagonal of  $E\{\tilde{\mathbf{b}}_1(i)\tilde{\mathbf{b}}_1^T(i)\}$  are equal to unity, while the off-diagonal ones are

$$\begin{aligned} E\{\tilde{b}_k(m)\tilde{b}_l(j)\} &= E\{\text{sgn}(\sqrt{a_k}b_k(m) + \xi_k(m))\text{sgn}(\sqrt{a_l}b_l(j) + \xi_l(j))\} \\ &= \frac{1}{2} \sum_{b_k(m), b_l(j)} [\Pr\{\xi_k(m) < \sqrt{a_k}b_k(m), \xi_l(j) < \sqrt{a_l}b_l(j)\} \\ &\quad - \Pr\{\xi_k(m) < \sqrt{a_k}b_k(m), \xi_l(j) > \sqrt{a_l}b_l(j)\}], \\ &\quad m, j = i - 1, i, \quad k, l = 2, \dots, K, \quad k \neq l. \end{aligned}$$

Each probability term in the above summation defines four integrals. Therefore

$$E\{\tilde{b}_k(m)\tilde{b}_l(j)\} = \sum_{n=1}^8 \frac{(-1)^n}{2} \iint_{D_n} f_{\xi_k \xi_l} d\xi_k d\xi_l \quad (B1)$$

where  $f_{\xi_k \xi_l}$  denotes the bivariate Gaussian density function of random variables  $\xi_k(m)$  and  $\xi_l(j)$ , and  $D_n$  is an appropriate rectangular region of integration. □

The diagonal elements of  $E\{\tilde{\mathbf{b}}_1(i)\tilde{\mathbf{b}}_1^T(i)\}$  are

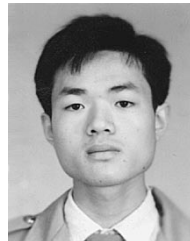
$$\begin{aligned} E\{b_k(j)\tilde{b}_k(j)\} &= 1 - 2\Pr\{\tilde{b}_k(j) \text{ in error}\} \\ &= 1 - 2Q\left(\sqrt{\frac{a_k}{\Gamma_k}}\right) \end{aligned} \quad (B2)$$

where  $k = 2, \dots, K; j = i - 1, i$ . □

REFERENCES

- [1] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 85–96, Jan. 1986.
- [2] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.
- [3] S. Verá, "Recent progress in multiuser detection," in *Advances in Communication and Signal Processing*, W. A. Porter and S. C. Kak, Eds. New York: Springer-Verlag, 1989, pp. 27–38.
- [4] A. Kajiwara and M. Nakagawa, "Crosscorrelation cancellation in SS/DS block demodulator," *IEICE Trans. Fundam. Electron., Commun., Comput. Sci.*, vol. E74, no. 9, pp. 2596–2602, Sept. 1991.
- [5] Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimum detectors for coherent multi-user communications," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 683–690, May 1990.
- [6] U. Madhow and M. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 3178–3188, Dec. 1994.
- [7] P. B. Rapajic and B. Vucetic, "Adaptive receiver structures for asynchronous CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 685–697, May 1994.
- [8] E. G. Strom and S. L. Miller, "A reduced complexity adaptive near-far resistant receiver for DS-CDMA," in *Proc. GLOBECOM'93*, Houston, TX, Nov. 1993, pp. 1734–1738.
- [9] P. Patel and J. Holtzman, "Analysis of a simple successive interference cancellation scheme in a DS/CDMA system," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 796–807, June 1994.
- [10] M. K. Varanasi and B. Aazhang, "Multistage detector in asynchronous code-division multiple access communications," *IEEE Trans. Commun.*, vol. 38, pp. 509–519, Apr. 1990.
- [11] —, "Near-optimum detector in synchronous code-division multiple-access system," *IEEE Trans. Commun.*, vol. 39, pp. 725–736, May 1991.
- [12] M. L. Honig, U. Madhow, and S. Verdú, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944–960, July 1995.
- [13] D. Godard, "Channel equalization using a Kalman filter for fast data transmission," *IBM J. Res. Develop.*, vol. 18, pp. 267–273, May 1974.

- [14] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [15] B. Zhu, N. Ansari, and Z. Siveski, "Convergence and stability analysis of a synchronous CDMA receiver," *IEEE Trans. Commun.*, vol. 43, pp. 3073–3079, Dec. 1995.
- [16] R. D. Gitlin and F. R. Magee, "Self-orthogonalizing adaptive equalization algorithms," *IEEE Trans. Commun.*, vol. COM-25, pp. 666–672, July 1977.



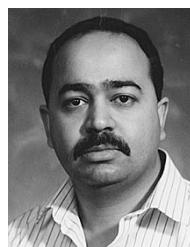
**Lizhi Charlie Zhong** was born in Xiangtan, Hunan, China, on July 13, 1970. He received the B.S. degree from Tsinghua University, Beijing, China, in 1993, and the M.S. degree from New Jersey Institute of Technology, Newark, in 1995, both in electrical engineering. He is currently working toward the Ph.D. degree in electrical engineering and computer science at the University of California at Berkeley.

From 1995 to 1998, he worked as a Consultant for the Network Wireless Group, AT&T Bell Laboratories, Whippany, NJ, which later became part of Lucent Technologies. His research interests include wireless communications with emphasis on CDMA. He has three patents in CDMA—one awarded in June 1999 and two others pending. Currently he is working in the Berkeley Wireless Research Center on Wireless Ad Hoc Networks, focusing on low-power MAC design for pico radios.

Mr. Zhong received two awards from the Bell Laboratories in 1996, including a CDMA Most Valuable Player (MVP) Award.

**Zoran Siveski** (S'83–M'89) received the Dipl. Ing. degree from the University of Belgrade and the Ph.D. degree from the City University of New York, both in electrical engineering.

He is currently a Member of Technical Staff in the Wireless Networks Group, Lucent Technologies, Whippany, NJ, and is involved in performance analysis and evaluation of CDMA wireless access technologies.



**Raafat E. Kamel** (S'91–M'94) received the B.Sc. degree from the University of Khartoum, and the M.S. and Ph.D. degrees from the New Jersey Institute of Technology (NJIT), Newark, all in electrical engineering. He was a Summer Intern at Hughes Network Systems in Germantown, MD, in the summer of 1992 and Research Associate at NJIT in 1994.

He is currently a Member of Technical Staff at Lucent Technologies. His research interest includes performance analysis of wireless systems.



**Nirwan Ansari** (S'78–M'88–SM'94) received the B.S.E.E. degree (*summa cum laude*) from the New Jersey Institute of Technology (NJIT) in 1982, the M.S.E.E. degree from the University of Michigan, Ann Arbor, in 1983, and the Ph.D. degree from Purdue University, West Lafayette, IN, in 1988.

In 1988, he joined NJIT where he has been Professor of Electrical and Computer Engineering since 1997. He authored (with E.S.H. Hou) *Computational Intelligence for Optimization* (1997), and edited (with B. Yuhua) *Neural Networks in Telecommunications* (1994), both by Kluwer Academic Publishers. His current research interests include various aspects of broad-band networks.

Dr. Ansari is a Technical Editor of the *IEEE Communications Magazine*. He was instrumental, while serving as its Chapter Chair, in rejuvenating the North Jersey Chapter of COMSOC which received the 1996 Chapter of the Year Award. He served as a Member-at-Large of the IEEE North Jersey Section (currently as the First Vice-Chair), and also serves in various IEEE committees. He was the 1998 recipient of the NJIT Excellence Teaching Award in Graduate Instruction, and a 1999 IEEE Region 1 Award.