Blind Multiuser Detection Using Linear Prediction

Richard T. Causey, *Student Member, IEEE*, and John R. Barry, *Member, IEEE*

***Abstract*—We propose a blind multiuser detection technique for array processing and code division multiple access (CDMA) systems that does not require knowledge of the array geometry or transmitter signature sequences. The technique has two key elements: an adaptive algorithm for separating the signal subspace from the noise subspace and an adaptive whitener based on linear prediction. The proposed algorithm offers low complexity, fast convergence, compatibility with shaped signal constellations, near-Wiener steady-state performance, and optimal near–far resistance.**

***Index Terms*—Adaptive subspace separation, array processing, blind source separation, cochannel demodulation, subspace tracking.**

I. Introduction

Blind multiuser detection is the process of recovering data from multiple simultaneously transmitting users without access to a training sequence. A partially blind multiuser detector exploits some known property of the channel. For example, the generalized sidelobe canceler [1], MUSIC [2], and ESPRIT [3] algorithms exploit knowledge of the array geometry, and the code division multiple access (CDMA) detectors of [4]–[6] exploit knowledge of the signature sequence of the desired user.

In contrast to the partially blind problem, this paper concerns the problem of blind multiuser detection without *a priori* knowledge of the channel. Thus, the receiver can exploit only its knowledge of the statistics of the channel input. For example, multiuser extensions of the constant-modulus algorithm (CMA) [7]–[11] attempt to restore the kurtosis of the channel input, but such detectors can exhibit slow convergence or misconvergence. Property-restoring methods based on fourth-order cumulants have been proposed in [12] and [13], but they disallow identically distributed sources. Methods based on explicit estimation of higher order statistics [14], [15] can be effective but are computationally complex.

Following [16]–[19], we decompose the blind multiuser detection problem into two steps by first whitening, then rotating. The whitening step exploits only second-order statistics and is well suited for blind implementation. The rotation step can be implemented by a unitary matrix chosen to restore some higher order statistical property of the channel input. The whiten-rotate (WR) structure is known to perfectly equalize a noiseless channel [20]. Batch techniques based on the WR structure were proposed in [16]–[18], and an adaptive technique for noiseless channels was presented in [19].

We introduce a blind WR detector based on linear prediction and a novel technique for subspace separation. We show that the steady-state performance of the WR detector closely approximates that of the minimum mean-square-error (MMSE) or Wiener detector and that it is optimally near–far resistant. The adaptive implementations we present offer a good compromise between complexity and convergence speed.

The paper is organized as follows. In the following section, we describe the problem statement. In Section III we define the structure of the WR detector. In Section IV we describe an alternative implementation of the WR detector based on subspace projection. In Section V we present low-complexity adaptive implementations of the proposed blind multiuser detector. Finally, in Section VI, we present simulation results for a linear-antenna-array and a synchronous-CDMA system.

II. Problem Statement

Let $\boldsymbol{x}_k$ denote a vector of symbols transmitted by n independent finite-alphabet transmitters at time k. Let $\boldsymbol{r}_k$ denote the corresponding receiver observation of dimension m, described by

$$\boldsymbol{r}_k = \mathbf{H}\boldsymbol{x}_k + \boldsymbol{n}_k \tag{1}$$

where $\mathbf{H}$ is an $m \times n$ memoryless channel matrix, and $\boldsymbol{n}_k$ is noise. In a narrowband m-sensor linear-array application, the columns of $\mathbf{H}$ represent the steering vectors for the n users, and in a synchronous-CDMA application with m chips per baud, the columns of $\mathbf{H}$ represent the signature sequences of the n users. The blind multiuser detection problem is to recover $\boldsymbol{x}_k$ from $\boldsymbol{r}_k$ without knowledge of $\mathbf{H}$. We assume that $\mathbf{H}$ has rank n, which implies that the channel is either square or tall $(m \geq n)$. We further assume that the signal and noise are independent and zero mean and satisfy[1] $E[\boldsymbol{x}_k\boldsymbol{x}_k^*] = \mathbf{I}$ and $E[\boldsymbol{n}_k\boldsymbol{n}_k^*] = \sigma^2\mathbf{I}$, with $\sigma > 0$.

A linear multiuser detector processes $\boldsymbol{r}_k$ with an $n \times m$ matrix $\mathbf{C}$, producing $\boldsymbol{z}_k = \mathbf{C}\boldsymbol{r}_k$. Let $\mathrm{MSE}_i = E[|z_k^{(i)} - x_k^{(i)}|^2]$ denote the resulting MSE for user i. The MMSE detector is the unique $\mathbf{C}$ that minimizes the MSE for each user. It can be expressed in two equivalent ways [4], [21], [22]

$$\mathbf{C}_{\mathrm{MMSE}} = \mathbf{H}^*\mathbf{R}_{\boldsymbol{rr}}^{-1} \tag{2}$$

$$= (\mathbf{H}^*\mathbf{H} + \sigma^2\mathbf{I})^{-1}\mathbf{H}^* \tag{3}$$

where the covariance matrix $\mathbf{R}_{\boldsymbol{rr}} = E[\boldsymbol{r}_k\boldsymbol{r}_k^*] = \mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I}$ is full rank, and hence invertible, when the noise is nonzero.

[1] The superscripts T, *, and $\dagger$ are used to denote transpose, Hermitian transpose, and the Moore–Penrose pseudoinverse, respectively.

Manuscript received October 20, 1997; revised September 24, 1998. This work was funded by the Center for Research in Applied Signal Processing (CRASP).

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0250 USA.

Publisher Item Identifier S 0733-8716(98)08659-4.

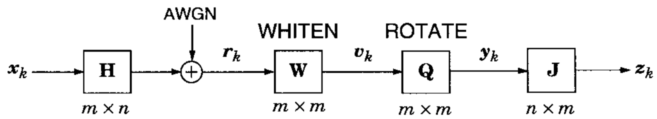


Fig. 1. The structure of the WR receiver.

We use the MMSE detector as a benchmark for assessing the performance of other detectors.

III. WR Detection

In the context of the system (1), an $n \times m$ matrix $\mathbf{C}$ is said to be a *whitener* if the covariance of $\boldsymbol{z}_k = \mathbf{C}\boldsymbol{r}_k$ is the identity matrix, $\mathbf{C}\mathbf{R}_{\boldsymbol{rr}}\mathbf{C}^* = \mathbf{I}$. We define the WR detector as the whitener with minimal MSE.

Definition 1: The WR $\mathbf{C}_{WR}$ for (1) is the $n \times m$ whitener that minimizes the MSE sum $E[||\mathbf{C}\boldsymbol{r}_k - \boldsymbol{x}_k||^2]$.

Any short whitener $\mathbf{C}$ of dimension $n \times m$ can be expressed as the first n rows of a larger whitener $\mathbf{B}$ of dimension $m \times m$. In particular, we can write $\mathbf{C} = \mathbf{J}\mathbf{B}$ where $\mathbf{J} = [\mathbf{I}, \mathbf{0}]$ is $n \times m$ and where $\mathbf{B}\mathbf{R}_{\boldsymbol{rr}}\mathbf{B}^* = \mathbf{I}$. Recall that, for any given $m \times m$ whitening matrix $\mathbf{W}$, every other whitening matrix $\mathbf{B}$ can be expressed in the form $\mathbf{B} = \mathbf{Q}\mathbf{W}$ for some $m \times m$ unitary matrix $\mathbf{Q}$ [23]. Thus, given any particular $m \times m$ whitening matrix $\mathbf{W}$, we can express every $n \times m$ whitener as $\mathbf{C} = \mathbf{J}\mathbf{Q}\mathbf{W}$ for some unitary matrix $\mathbf{Q}$. This suggests a three-stage implementation of the WR detector, as depicted in Fig. 1.

Observe from (2) that the MMSE detector can be expressed as $\mathbf{C}_{\text{MMSE}} = \mathbf{H}^*\mathbf{W}^*\mathbf{W} = (\mathbf{W}\mathbf{H})^*\mathbf{W}$ where we use the identity $\mathbf{W}^*\mathbf{W} = \mathbf{R}_{\boldsymbol{rr}}^{-1}$ for any whitener $\mathbf{W}$. Thus, the MMSE detector could be implemented by following a whitener $\mathbf{W}$ by the $n \times m$ filter $(\mathbf{W}\mathbf{H})^*$. By its definition, however, the WR detector must follow a whitener by a matrix of the form $\mathbf{J}\mathbf{Q}$. It can be shown that, rather than $(\mathbf{W}\mathbf{H})^*$, the best such filter (minimizing total MSE) is the unique so-called polar factor of $(\mathbf{W}\mathbf{H})^*$ [24], which is simply $(\mathbf{W}\mathbf{H})^*$ with its singular values replaced by unity

$$\mathbf{J}\mathbf{Q} = \left(\hat{\mathbf{U}}\mathbf{J}^T\hat{\mathbf{V}}^*\right)^* \tag{4}$$

where $\hat{\mathbf{U}}_{m \times m}$ and $\hat{\mathbf{V}}_{n \times n}$ are factors in a singular-value decomposition (SVD) of $\mathbf{W}\mathbf{H} = \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^*$. Note that $\mathbf{Q}$ satisfies (4) if and only if it is of the form

$$\mathbf{Q} = \begin{bmatrix} \hat{\mathbf{V}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_N \end{bmatrix} \hat{\mathbf{U}}^* \tag{5}$$

where $\mathbf{V}_N$ is an arbitrary unitary matrix of dimension $m - n$.

The following lemma summarizes the form of the WR detector.

Lemma 1: The WR detector of Definition 1 is unique, and it can be expressed in the following three equivalent ways:

$$\begin{aligned} \mathbf{C}_{WR} &= \mathbf{J}\mathbf{Q}\mathbf{W} && (6) \\ &= \mathbf{V}\mathbf{J}(\mathbf{S}\mathbf{S}^* + \sigma^2\mathbf{I})^{-1/2}\mathbf{U}^* && (7) \\ &= \mathbf{V}(\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I})^{-1/2}\mathbf{J}\mathbf{U}^*. && (8) \end{aligned}$$

In (6), $\mathbf{W}$ is any $m \times m$ whitener (satisfying $\mathbf{W}\mathbf{R}_{\boldsymbol{rr}}\mathbf{W}^* = \mathbf{I}$) and $\mathbf{J}\mathbf{Q}$ satisfies (4). In (7) and (8), $\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^*$ is an SVD, and $\tilde{\mathbf{S}} = \mathbf{J}\mathbf{S}$ with $\mathbf{J} = [\mathbf{I}, \mathbf{0}]$.

Proof: See Appendix I.

Using this lemma, we observe several properties of the WR detector.

Property 1: The WR detector is information lossless. This follows from (8) by observing that $\mathbf{J}\mathbf{U}^*$ discards no signal energy and that both $\mathbf{V}$ and $(\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I})^{-1/2}$ are invertible.

Property 2: The WR detector approaches the decorrelating or zero-forcing detector in the limit as the noise energy goes to zero

$$\lim_{\sigma \to 0^+} \mathbf{C}_{WR} = \mathbf{V}\tilde{\mathbf{S}}^{-1}\mathbf{J}\mathbf{U}^* = \mathbf{V}\mathbf{S}^{\dagger}\mathbf{U}^* = \mathbf{H}^{\dagger}. \tag{9}$$

Property 3: The WR detector is optimally near–far resistant [25]. Optimal near–far resistance is inherited from the zero-forcing detector.

Lemma 2: The MSE for the ith user of the WR and MMSE detectors, respectively, can be expressed as

$$\text{MSE}_i^{WR} = 2\boldsymbol{v}_i^*\left[\mathbf{I} - \left(\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I}\right)^{-1/2}\tilde{\mathbf{S}}\right]\boldsymbol{v}_i \tag{10}$$

$$\text{MSE}_i^{\text{MMSE}} = \sigma^2\boldsymbol{v}_i^*\left(\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I}\right)^{-1}\boldsymbol{v}_i \tag{11}$$

where $\boldsymbol{v}_i$ is the ith column of $\mathbf{V}^*$.

Proof: See Appendix II. Using this lemma, we arrive at the following property of the WR detector.

Property 4: The MSE of the WR detector approaches that of the MMSE detector in the limit as the noise energy goes to zero

$$\lim_{\sigma \to 0^+} \frac{\text{MSE}_i^{WR}}{\text{MSE}_i^{\text{MMSE}}} = 1. \tag{12}$$

The proof follows from (10) and (11) and a straightforward application of l'Hôpital's rule. In Fig. 2 we use (10) and (11) to compare the theoretical performance of the WR detector to that of the MMSE detector for a receiver with $m = 10$ sensors. We consider two cases, $n = 2$ users and $n = 10$, for which we plot MSE_1 versus $\text{SNR}_1 = \Sigma_{j=1}^m |h_{j,1}|^2/\sigma^2$, averaged over 1000 channels of dimension $10 \times n$. The coefficients of each channel were selected independently from a zero-mean unit-variance complex Gaussian distribution, and then the channel columns were scaled so that all odd-numbered users have energy 10 dB below that of even-numbered users. The curves for the two-user case show that even for a severe SNR_1 of -10 dB, user one suffers only a modest 2-dB MSE penalty. Moreover, for $\text{SNR}_1 > 10$ dB, the performance difference is negligible. The curves for the ten-user case show that as the number of users approaches the number of sensors, the performance of both detectors degrades. However, the performance difference between the two detectors widens only slightly.

IV. Subspace Separation Before Whitening

In the previous section, the rotator $\mathbf{Q}$ of (5) performs two tasks. First, it removes all signal energy from the last $m - n$ components of its output, thereby separating the signal and noise subspaces. Second, it also provides the best unitary

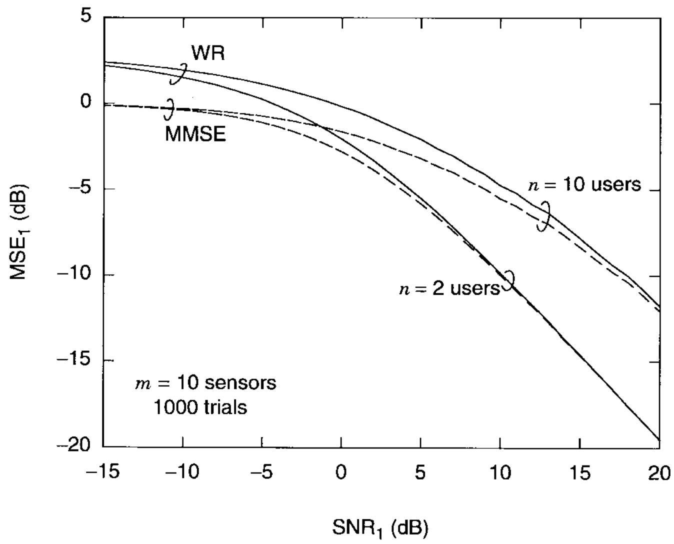


Fig. 2. A comparison of the WR detector with the minimum-MSE detector.

separation of users within the signal space. In other words, $\mathbf{Q}$ separates both signal from noise, and signal from signal.

It is often desirable to separate subspaces at the receiver front end, before whitening, by immediately projecting the m-dimensional receiver signal onto the n-dimensional signal subspace using a unitary matrix $\mathbf{\Theta}$. The advantages of such a front-end projector are two. First, it allows all subsequent signal processing to operate in n dimensions rather than m, which reduces the receiver complexity. Second, it reduces the number of receiver parameters, which often leads to faster receiver convergence as well.

We now precisely define a unitary matrix $\mathbf{\Theta}$ that separates the signal and noise subspaces.

Definition 2: Given the $m \times n$ channel $\mathbf{H}$ of (1), an $m \times m$ unitary matrix $\mathbf{\Theta}$ is a *subspace-separation matrix* if and only if the last $m-n$ rows of $\mathbf{\Theta H}$ are identically zero. Constraining the last $m-n$ rows of $\mathbf{\Theta H}$ to zero guarantees that the last $m-n$ components of $\mathbf{\Theta}\boldsymbol{r}_\mathbf{k}$ represent a projection of $\boldsymbol{r}_k$ onto the noise subspace. Furthermore, because a unitary $\mathbf{\Theta}$ always has full rank, it follows that the first n components of $\mathbf{\Theta}\boldsymbol{r}_\mathbf{k}$ represent a projection of $\boldsymbol{r}_k$ onto the signal subspace.

Lemma 3: Given the channel of (1), a unitary subspace-separating matrix must be of the form

$$\mathbf{\Theta} = \begin{bmatrix} \mathbf{U_S} & \mathbf{0} \\ \mathbf{0} & \mathbf{U_N} \end{bmatrix} \mathbf{U}^* \tag{13}$$

where $\mathbf{U}_S$ and $\mathbf{U}_N$ are arbitrary unitary matrices of dimension n and $m-n$, respectively, and where $\mathbf{U}$ is again the left factor of a channel SVD $\mathbf{H} = \mathbf{USV}^*$.

Proof: In terms of an SVD $\mathbf{H} = \mathbf{USV}^*$, we can write $\mathbf{\Theta H} = \mathbf{\Theta USV}^*$. Because $\mathbf{H}$ has rank n, the last $m-n$ rows of $\mathbf{SV}^*$ are already zero. Thus, the last $m-n$ rows of $\mathbf{\Theta H}$ remain zero as long as $\mathbf{\Theta U}$ passes none of the energy from the first n inputs to the last $m-n$ outputs, or equivalently, because $\mathbf{\Theta U}$ is also unitary, $\mathbf{\Theta}$ satisfies (13). □

The first n rows of $\mathbf{\Theta}$ in (13) form an orthonormal basis for the signal subspace, and the last $m-n$ rows form a basis for the noise subspace. From (13) and an SVD $\mathbf{H} = \mathbf{USV}^*$,

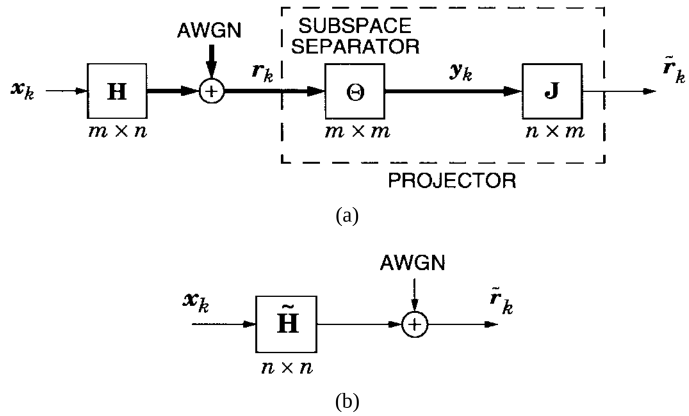


Fig. 3. Equivalent models. (a) Tall channel using a signal–space projector as the receiver front end. (b) Equivalent square channel.

the output $\boldsymbol{y}_k = \mathbf{\Theta}\boldsymbol{r}_\mathbf{k}$ of a subspace-separating matrix can be written as

$$\boldsymbol{y}_k = \begin{bmatrix} \mathbf{U}_S\tilde{\mathbf{S}}\mathbf{V}^* \\ \mathbf{0} \end{bmatrix} \boldsymbol{x}_k + \mathbf{\Theta}\boldsymbol{n}_\mathbf{k} \tag{14}$$

where $\tilde{\mathbf{S}} = \mathbf{JS}$. Because the last $m-n$ components of $\boldsymbol{y}_k$ contain only noise energy, they can be used to estimate the noise variance, if desired, or they can be discarded without any loss of signal information, thereby effectively producing a square channel, as shown in Fig. 3(b)

$$\tilde{\boldsymbol{r}}_k = \tilde{\mathbf{H}}\boldsymbol{x}_k + \tilde{\boldsymbol{n}}_k \tag{15}$$

where $\tilde{\boldsymbol{r}}_k = \mathbf{J\Theta}\boldsymbol{r}_\mathbf{k}$, $\tilde{\mathbf{H}} = \mathbf{U}_S\tilde{\mathbf{S}}\mathbf{V}^*$, and $E[\tilde{\boldsymbol{n}}_k\tilde{\boldsymbol{n}}_k^*] = \sigma^2\mathbf{I}$. The n components of the new observation vector $\tilde{\boldsymbol{r}}_k$ form a set of sufficient statistics for estimating $\boldsymbol{x}_k$. Thus, the WR detector of Section III (or any other blind multiuser detector) can be applied to this new channel without compromising performance.

Theorem 1: The cascade of a signal-subspace projector $\mathbf{J\Theta}$ and a WR detector $\tilde{\mathbf{C}}_{WR}$ designed for the reduced channel $\tilde{\mathbf{H}}$ precisely implements the WR detector $\mathbf{C}_{WR}$ designed for the original channel $\mathbf{H}$.

Proof: The proof applies to the MMSE detector (2) as well as the WR detector (8), since both can be expressed in terms of an SVD $\mathbf{H} = \mathbf{USV}^*$ as $\mathbf{C} = \mathbf{VDJU}^*$ where the diagonal matrix $\mathbf{D}$ is $\mathbf{D} = (\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I})^{-1/2}$ for the WR detector and $\mathbf{D} = \tilde{\mathbf{S}}(\tilde{\mathbf{S}}^2 + \sigma^2\mathbf{I})^{-1}$ for the MMSE detector. We need to show that the cascade of $\mathbf{J\Theta}$ and an $n \times n$ detector $\tilde{\mathbf{C}}$, designed for the reduced channel $\tilde{\mathbf{H}}$, is equivalent to the same type of detector designed for the original channel $\mathbf{H}$. In other words, we need to show that $\tilde{\mathbf{C}}\mathbf{J\Theta} = \mathbf{C}$. But based on the SVD $\tilde{\mathbf{H}} = \mathbf{U}_S\tilde{\mathbf{S}}\mathbf{V}^*$ of the reduced channel, we have $\tilde{\mathbf{C}} = \mathbf{VDU}_S^*$, so that $\tilde{\mathbf{C}}\mathbf{J}\Theta = \mathbf{VDU}_S^*\mathbf{J}\left[\begin{smallmatrix}\mathbf{U}_S & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_N\end{smallmatrix}\right]\mathbf{U}^* = \mathbf{VDJU}^* = \mathbf{C}$. □

V. Blind Adaptive Implementations

In this section we describe a blind adaptive implementation of the WR detector of Section III, as illustrated in Fig. 1, and of the subspace separator $\mathbf{\Theta}$ of Section IV, as illustrated in Fig. 3(a). With the exception of the algorithm for estimating the final unitary matrix $\mathbf{Q}$ in Fig. 1, they are all strictly second-order algorithms. They impose no constraints on the channel input distribution and are thus inherently compatible with shaped, or even Gaussian, channel inputs.

A. An Adaptive Whitener

A simple way to perform whitening is to use adaptive linear prediction. Suppose we wish to predict the ith component $r_k^{(i)}$ of $\boldsymbol{r}_k$ using a linear combination of the preceding components $r_k^{(1)} \cdots r_k^{(i-1)}$, yielding an estimate

$$\hat{\boldsymbol{r}}_k = \mathbf{P}\boldsymbol{r}_k \tag{16}$$

where $\mathbf{P}$ is a strictly lower triangular matrix of prediction coefficients. The prediction error is $\boldsymbol{e}_k = (\mathbf{I} - \mathbf{P})\boldsymbol{r}_k$. The best predictor in the least-mean-square sense, i.e., minimizing $E[||\boldsymbol{e}_k||^2]$, is closely linked to the Cholesky factorization of the covariance matrix $\mathbf{R}_{\boldsymbol{rr}}$ of $\boldsymbol{r}_k$.

Lemma 4—Generalized Cholesky Factorization: An Hermitian matrix $\mathbf{R}$ of dimension $m \times m$ and rank $n \times m$ can be factored in either of two ways

$$\mathbf{R} = \mathbf{G}\mathbf{G}^* \tag{17}$$

$$= \mathbf{M}\mathbf{D}^2\mathbf{M}^* \tag{18}$$

where $\mathbf{G} = \mathbf{MD}$ is a unique $m \times m$ lower triangular matrix with real, nonnegative diagonal elements, where $\mathbf{D} = \mathrm{diag}\,(\mathbf{G})$, and where $\mathbf{M}$ is lower triangular with ones on the main diagonal (monic). The matrix $\mathbf{M}$ is unique if and only if the first $m-1$ rows of $\mathbf{R}$ are linearly independent.

Proof: See Appendix III.

Theorem 2—Linear Prediction: Let $\boldsymbol{r}$ be a random $m \times 1$ vector with covariance matrix $\mathbf{R} = E[\boldsymbol{r}\boldsymbol{r}^*]$ and let $\boldsymbol{e} = (\mathbf{I} - \mathbf{P})\boldsymbol{r}$ denote the error of a linear predictor where $\mathbf{P}$ is strictly lower triangular. The $\mathbf{P}$ that minimizes $E[||\boldsymbol{e}||^2]$ is

$$\mathbf{P} = \mathbf{I} - \mathbf{M}^{-1} \tag{19}$$

where $\mathbf{M}$ is any valid monic factor in the generalized Cholesky factorization (18) of $\mathbf{R}$. The predictor is unique if and only if $\mathbf{M}$ is unique, or equivalently, if and only if the first $m-1$ rows of $\mathbf{R}$ are linearly independent.

Proof: See Appendix IV.

As long as the noise variance is nonzero, the covariance $\mathbf{R}_{\boldsymbol{rr}} = \mathbf{H}\mathbf{H}^* + \sigma^2\mathbf{I}$ is full rank and the Cholesky factor (18) and corresponding predictor (19) are both unique. (See Appendix V for a discussion of the noiseless case.) In practice, the coefficients of $\mathbf{P}$ can be adapted according to the least-mean-square algorithm

$$\mathbf{P}_{k+1} = (\mathbf{P}_k + \mu_p \boldsymbol{e}_k \boldsymbol{r}_k^*) \otimes \mathbf{L} \tag{20}$$

where $\boldsymbol{e}_k = \boldsymbol{r}_k - \mathbf{P}_k\boldsymbol{r}_k$ is the prediction error, where μ_p is a step size, where $\otimes$ denotes a component-wise (Schur) product, and where $\mathbf{L}$ is a mask, with ones below the main diagonal and zeros elsewhere, that constrains $\mathbf{P}$ to be strictly lower triangular. We remark that, because (20) is derived from a quadratic cost function, convergence to (19) is guaranteed for a sufficiently small step size μ_p.

After $\mathbf{P}$ converges to (19), the covariance of the resulting error $\boldsymbol{e}_k = (\mathbf{I} - \mathbf{P})\boldsymbol{r}_k$ is diagonal $\mathbf{R}_{\boldsymbol{ee}} = \mathbf{D}^2$. Therefore, a diagonal gain matrix $\mathbf{A} = \mathbf{D}^{-1}$ converts the prediction error $\boldsymbol{e}_k$ into the white signal $\boldsymbol{v}_k = \mathbf{A}\boldsymbol{e}_k$ with covariance matrix $\mathbf{R}_{\boldsymbol{vv}} = \mathbf{I}$. This gain matrix can be implemented adaptively by a bank of independent scalar automatic gain-control loops, designed to force the energy at each output to unity. We propose a simple first-order loop for adapting each diagonal component of $\mathbf{A} = \mathrm{diag}\,(A^{(1)}, \ldots, A^{(m)})$

$$A_{k+1}(i) = \left| A_k^{(i)} - \mu_a(|v_k^{(i)}|^2 - 1) \right|. \tag{21}$$

In summary, the proposed adaptive whitener is $\mathbf{W} = \mathbf{A}(\mathbf{I} - \mathbf{P})$, where $\mathbf{P}$ and $\mathbf{A}$ are adapted according to (20) and (21).

B. An Adaptive Rotator

Recall the structure of the WR detector of Fig. 1, $\mathbf{C} = \mathbf{JQW}$. Let $\boldsymbol{v}_k = \mathbf{W}\boldsymbol{r}_k$ denote the whitener output. With $\mathbf{W}$ adapted according to the previous section, it remains to specify an adaptive algorithm for the rotator $\mathbf{Q}$. Let $\boldsymbol{y}_k = \mathbf{Q}\boldsymbol{v}_k$ denote the rotator output. By Definition 1, the $\mathbf{Q}$ of the WR detector minimizes the MSE-sum cost function $E[||\mathbf{J}\boldsymbol{y}_k - \boldsymbol{x}_k||^2]$. If we define a new vector $\overline{\boldsymbol{x}}_k = [\boldsymbol{x}_k^T, \mathbf{0}]^T$ by stacking $\boldsymbol{x}_k$ above $m-n$ zeros, then the whiten-rotate $\mathbf{Q}$ also minimizes the following cost function:

$$E\Big[||\boldsymbol{y}_k - \overline{\boldsymbol{x}}_k||^2\Big] = E\Big[||\mathbf{J}\boldsymbol{y}_k - \boldsymbol{x}_k||^2\Big] + \sigma^2 \,\mathrm{tr}\,[\mathbf{W}\mathbf{W}^*]. \tag{22}$$

The two cost functions are simultaneously minimized because the last term $\sigma^2 \,\mathrm{tr}\,[\mathbf{W}\mathbf{W}^*]$ is independent of $\mathbf{Q}$. Thus, the whiten-rotate $\mathbf{Q}$ is the rotator that best maps $\boldsymbol{y}_k$ to $\overline{\boldsymbol{x}}_k$.

Let $\hat{\boldsymbol{x}}_k$ denote the receiver's estimate of $\overline{\boldsymbol{x}}_k$. Clearly the last $m-n$ components of this estimate should be zero. The first n components can be estimated in a decision-directed and blind manner by exploiting knowledge of the signal alphabets [26]

$$\hat{\boldsymbol{x}}_k = \begin{bmatrix} \boldsymbol{q}(\cdot) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{y}_k. \tag{23}$$

For $i \le n, \hat{x}_k^{(i)} = q_i(y_k^{(i)})$ is the point in the constellation of user i closest to $y_k^{(i)}$, but for $i > n, \hat{x}_k^{(i)}$ is set to zero.

We modify the multidimensional phase-locked loop (MPLL) of [27] to find a rotation that best maps $\boldsymbol{y}_k$ to $\hat{\boldsymbol{x}}_k$. Following [27], we define a partial rotation from $\boldsymbol{x}$ to $\boldsymbol{y}$ as

$$\mathcal{R}^\lambda(\boldsymbol{x} \to \boldsymbol{y}) = \mathbf{I} + [\boldsymbol{u}, \boldsymbol{v}] \begin{bmatrix} p-1 & \dfrac{-p}{|p}, \sqrt{1-|p|^2} \\ \sqrt{1-|p|^2} & |p|-1 \end{bmatrix} \begin{bmatrix} \boldsymbol{u}^* \\ \boldsymbol{v}^* \end{bmatrix} \tag{24}$$

where p is a normalized inner product, $p = \boldsymbol{x}^*\boldsymbol{z}/(||\boldsymbol{x}|| \cdot ||\boldsymbol{z}||)$ with $\boldsymbol{z} = \lambda\boldsymbol{y} + (1-\lambda)\boldsymbol{x}$, and where $\{\boldsymbol{u}, \boldsymbol{v}\}$ is a basis for the two-dimensional subspace spanned by $\boldsymbol{x}$ and $\boldsymbol{y}$: $\boldsymbol{u} = \boldsymbol{x}/||\boldsymbol{x}||$ and $\boldsymbol{v} = (\boldsymbol{z}/||\boldsymbol{z}|| - p\boldsymbol{u})/\sqrt{1-|p|^2}$. (For the singular case of $|p| = 1$, we take $\boldsymbol{v} = \mathbf{0}$.) As in [27], the estimate of $\mathbf{Q}$ is iteratively updated by accumulating a matrix that partially rotates $\hat{\boldsymbol{x}}_k$ to $\boldsymbol{y}_k$

$$\mathbf{Q}_{k+1} = \mathcal{R}^\lambda(\hat{\boldsymbol{x}}_k \to \boldsymbol{y}_k)^*\mathbf{Q}_k \tag{25}$$

$$= \mathbf{Q}_k + [\mathcal{R}^\lambda(\hat{\boldsymbol{x}}_k \to \boldsymbol{y}_k)^* - \mathbf{I}]\mathbf{Q}_k. \tag{26}$$

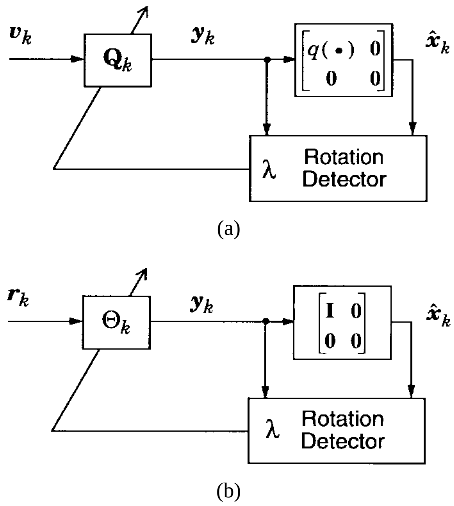


Fig. 4. Adaptive rotator. (a) Used to implement the MMSE rotator of (29). (b) Used to implement the subspace separator of (13).

Observe that the matrix within the square brackets of (26) has rank two, implying that (26) has lower complexity than (25) when $m > 2$. The modified MPLL is illustrated in Fig. 4(a).

Because of the ambiguities inherent in any fully blind detection algorithm, the MPLL (26) may not converge to (5) exactly. Because identically distributed users are statistically indistinguishable, they are arbitrarily labeled at the output of any fully blind detector. Moreover, the constellation of each user has rotational symmetries that cannot be blindly resolved. Rotating any square quadrature amplitude modulation (QAM) constellation by an integer multiple of 90,° for example, does not change its statistics. In practice, these ambiguities are of little consequence because they can be resolved by other means. Therefore, it is generally satisfactory if $\mathbf{J}\hat{\boldsymbol{x}}_k = \mathbf{K}\boldsymbol{x}_k$, where $\mathbf{K} = \mathbf{K}_P\mathbf{K}_R$ is the $n \times n$ product of a permutation matrix $\mathbf{K}_P$ and a diagonal unitary matrix $\mathbf{K}_R = \operatorname{diag}(\exp(j\theta_i))$, with the angles θ_i determined by the rotational symmetries of constellation i. If all users transmit 16-QAM, for example, then $\mathbf{K}$ is a *complex* permutation matrix, i.e., a matrix with exactly one nonzero element from $\{\pm 1, \pm j\}$ per row and per column. In Section VI, we present empirical evidence to support the conjecture that the MPLL converges to

$$\mathbf{Q} = \begin{bmatrix} \mathbf{K}\hat{\mathbf{V}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_N \end{bmatrix} \hat{\mathbf{U}}^* \tag{27}$$

where, as in (5), $\hat{\mathbf{U}}$ and $\hat{\mathbf{V}}$ are factors of an SVD of $\mathbf{WH} = \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^*$ and $\mathbf{V}_N$ is any arbitrary unitary matrix of dimension $m - n$.

C. An Adaptive Subspace Separator

We now describe an adaptive technique for implementing the subspace projector of Section IV, as shown in Fig. 3(a). The intent of the front-end filter $\boldsymbol{\Theta}$ is to implement the $m \times m$ unitary matrix (13), separating the signal subspace from the noise subspace. That is, $\boldsymbol{\Theta}$ minimizes the cost function $E[\|\boldsymbol{y}_k - \hat{\boldsymbol{x}}_k\|^2]$ that measures the energy in the last $m - n$ components of $\boldsymbol{y}_k = \boldsymbol{\Theta}\boldsymbol{r}_\mathbf{k}$ where we introduce

$$\hat{\boldsymbol{x}}_k = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{y}_k. \tag{28}$$

We can interpret $\hat{\boldsymbol{x}}_k$ as a projection of $\boldsymbol{r}_k$ onto the signal subspace, with the cost function measuring the projection error. Observe that this cost function is minimized by any $\boldsymbol{\Theta}$ of the form given in (13). Because of the similarity between (23) and (28), we can again use the modified MPLL update of (26) to update $\boldsymbol{\Theta}$, except that $\hat{\boldsymbol{x}}_k$ is defined by (28) instead of (23). This adaptive subspace separator is illustrated in Fig. 4(b). The subspace separator is thus adapted according to

$$\boldsymbol{\Theta}_\mathbf{k+1} = \boldsymbol{\Theta}_\mathbf{k} + \left[\mathcal{R}^\lambda(\hat{\boldsymbol{x}}_\mathbf{k} \to \boldsymbol{y}_\mathbf{k})^* - \mathbf{I}\right]\boldsymbol{\Theta}_\mathbf{k} \tag{29}$$

where $\boldsymbol{y}_k = \boldsymbol{\Theta}_\mathbf{k}\boldsymbol{r}_\mathbf{k}$ and $\hat{\boldsymbol{x}}_k$ is defined by (28). See Section VI-A for a demonstration that (29) with (28) converges to (13).

An adaptive WR detector that uses the subspace separator of (29) is shown in Fig. 5. With ideal subspace separation, the output $\boldsymbol{y}_k$ of the front-end rotator $\boldsymbol{\Theta}$ is given by (14), and in particular, the last $m - n$ components contain no signal energy. Thus, the subspace separator can be followed by an $n \times m$ nonadaptive matrix $\mathbf{J} = [\mathbf{I}, \mathbf{0}]$. The net effect is a square channel described by (15) so that the remainder of the algorithm (the predictor $\mathbf{I} - \mathbf{P}$, the automatic gain control (AGC) bank $\mathbf{A}$, and the MPLL $\mathbf{Q}$) can be implemented as before.

In practice, any signal-space estimation or tracking algorithm [28]–[31] can be used before the whitener. The proposed technique of (28) and (29) provides a precisely orthonormal estimate of both subspaces at each iteration. The cascade of the signal–space projector, predictor, and AGC in Fig. 5 amounts to an adaptive short whitener that, although higher in complexity than the adaptive prewhitener of [19], preserves all information about the signals.

VI. Numerical Results

A. Random Gaussian Channels

First, we demonstrate the convergence of the adaptive subspace separator of Section V-C. According to (14), we need only demonstrate that the last $m - n$ rows of the separator-channel cascade $\boldsymbol{\Theta}_\mathbf{k}\mathbf{H}$ converge to zero. We consider again the random Gaussian channels of Section III with $m = 10$, $n = 2$, and the signal-to-noise ratio (SNR) of each user fixed at 27 dB. Fig. 6 shows the energy in each row of $\boldsymbol{\Theta}_\mathbf{k}\mathbf{H}$ as a function of time k, averaged over 100 channels. We see that the energy of the last eight rows converges quickly to levels of -40 dB or less.

We now demonstrate convergence of the entire project-first algorithm of Fig. 5. We consider two users, each transmitting 16-QAM with 20-dB SNR. Fig. 7 shows $\mathrm{MSE}_1 = E[|z_k^{(1)} - x_k^{(1)}|^2]$ as a function of time, averaged over 1000 realizations of input, noise, and a 10 × 2 complex Gaussian channel. There are five curves in all. The bottom curve, labeled MMSE, is MSE_1 for the ideal MMSE equalizer. The initial subspace separator is adaptive for the curve above that, but all remaining functions are idealized. Similarly, the other curves are labeled

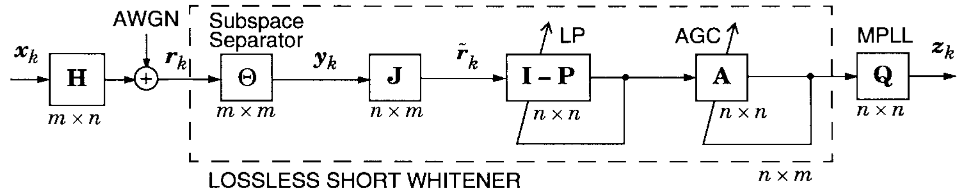


Fig. 5. A block diagram of an adaptive project-first WR receiver.

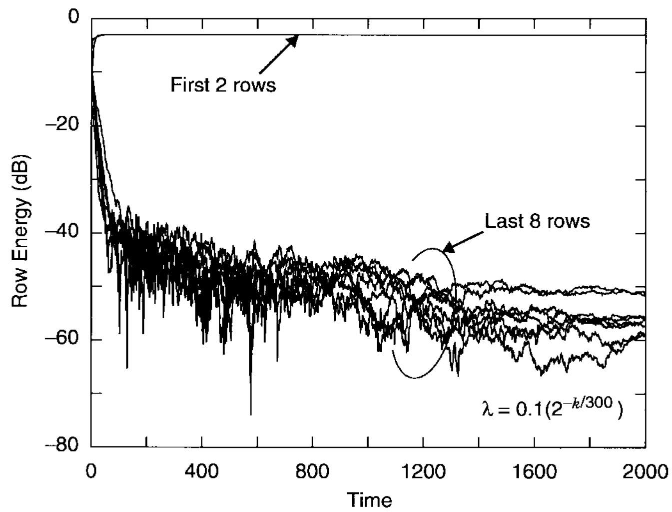


Fig. 6. Demonstration of the convergence of the subspace separator: energy in the rows of the separator-channel cascade $\mathbf{\Theta_k H}$ versus time k.

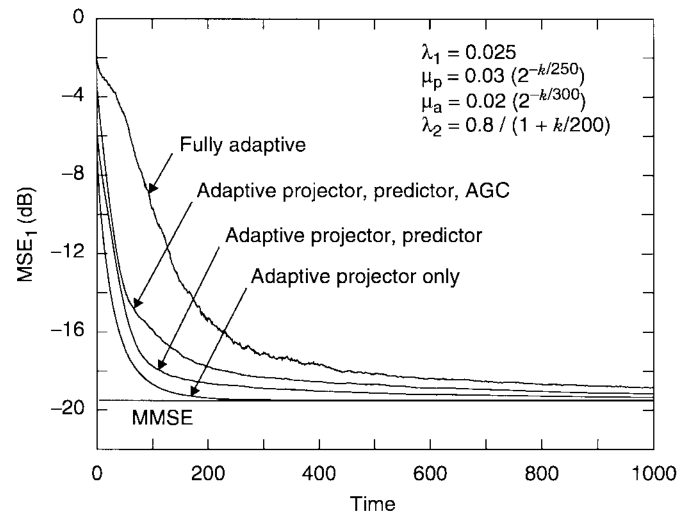


Fig. 7. Convergence of the project-first adaptive algorithm of Fig. 5, showing contributions to MSE from each stage.

to indicate which components of the algorithm are adaptive. Everything is adaptive for the top curve, with the effect of the ambiguous complex permutation matrix $\mathbf{K}$ removed for each trial. These curves illustrate the MSE contributed by each stage of the project-first algorithm. The subspace separator converges very quickly and has little impact on MSE. The linear predictor and AGC bank also converge quickly, and the receiver eventually closely approximates the MMSE solution.

B. A Linear Antenna Array Example

Consider a 20-sensor linear antenna array with half-wavelength spacing, and suppose that two signals are incident at angles $\theta_1 = 0°$ and $\theta_2 = 20°$ (measured from broadside). For this arrangement the channel model is (1) with $\mathbf{H}_{20\times 2} = (1/\sqrt{20})\mathbf{VB}$ where $V_{i,l} = \exp\{j(\pi/\lambda)(i-1)\sin(\theta_l)\}$ [32] and $\mathbf{B} = \mathrm{diag}(B_1, B_2)$ where B_i^2 is the received power of the ith user. Each user transmits 4-QAM with $B_2^2/B_1^2 = -20$ dB, $\mathrm{SNR}_1 = 35$ dB, and $\mathrm{SNR}_2 = 15$ dB. Fig. 8 shows a plot of MSE_2 versus time, averaged over 100 input and noise realizations, with the effect of the complex permutation removed. The inset shows constellations from time 4000 to time 5000 from the last trial. Once again, we see quick convergence to near-MMSE performance.

C. A Synchronous-CDMA Example

Consider now a synchronous direct-sequence-CDMA application with three interfering users, each transmitting 16-QAM. Let $\mathbf{c}_i \in \{\pm 1\}^{32}$ denote the binary signature sequence with length 32 of the ith user. If the transmitter pulse-shape filters

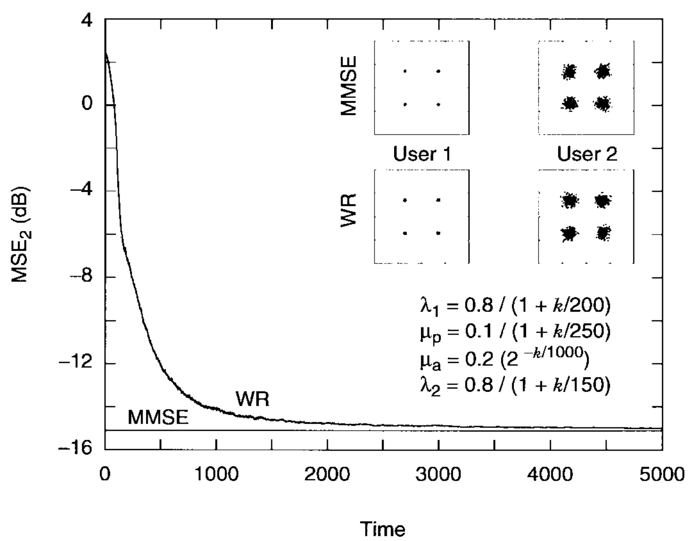


Fig. 8. Convergence of project-WR receiver for the linear-array application with inset constellations near convergence.

are Nyquist and the receiver uses a chip-rate-sampled matched filter followed by a serial-to-parallel converter, the resulting discrete-time channel is again given by (1) with $\mathbf{H}_{32\times 3} = (1/\sqrt{32})[\mathbf{c}_1 \quad \mathbf{c}_2 \quad \mathbf{c}_3]\mathbf{B}$ where $\mathbf{B}_{3\times 3} = \mathrm{diag}(B_1, B_2, B_3)$ is a matrix of signal amplitudes. The signature sequences have normalized correlations $\rho_{ij} = \frac{1}{32}\mathbf{c}_i^T\mathbf{c}_j$ of $\rho_{12} = -1/8, \rho_{13} = -1/4$, and $\rho_{23} = 1/4$. Fig. 9 shows an MSE learning curve, averaged over 100 input and noise realizations, with $\mathrm{SNR}_1 = 40$ dB and $\mathrm{SNR}_2 = \mathrm{SNR}_3 = 20$ dB. We see that the

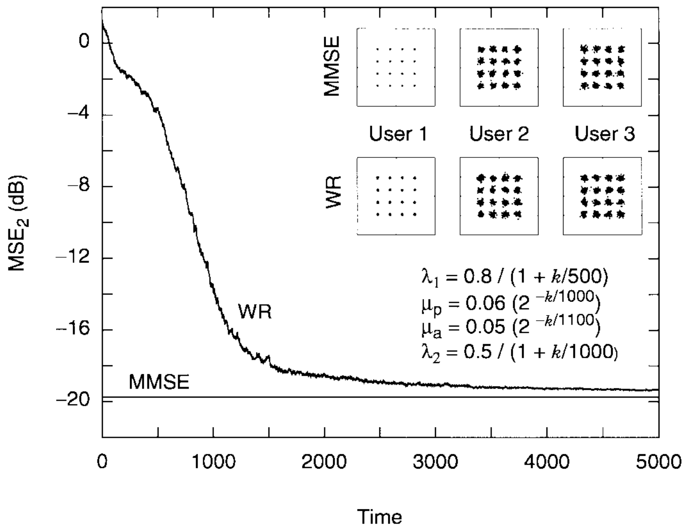


Fig. 9. Convergence of the project-WR receiver for the CDMA application with inset constellations near convergence.

algorithm converges quickly, eventually closely approximating the MMSE solution. The inset constellations show the last 1000 symbols from the last trial.

VII. Conclusions

We have presented a blind multiuser detector based on a lossless WR structure. The algorithm uses linear prediction and projection techniques to exploit second-order statistics first. The higher order statistics of the channel input are exploited only at the last step, by finding a unitary matrix that best restores the discrete nature of the channel inputs. We have also demonstrated a new technique for the adaptive separation of signal and noise subspaces. The proposed algorithms offer both low complexity and performance approaching that of the minimum-MSE detector. Future work should extend these techniques to channels with memory, perhaps by applying temporal prediction [33], [34].

Appendix I
Proof of Lemma 1

Substituting an SVD $\mathbf{H}=\mathbf{USV}^*$ into $\mathbf{R}_{\boldsymbol{rr}}=\mathbf{HH}^*+\sigma^2\mathbf{I}$ yields $\mathbf{R}_{\boldsymbol{rr}}=\mathbf{U\Lambda U}^*$ where $\mathbf{\Lambda}=\mathbf{SS}^*+\sigma^2\mathbf{I}$ is diagonal. It follows that $\mathbf{W}=\mathbf{\Lambda}^{-1/2}\mathbf{U}^*$ is a whitener, satisfying $\mathbf{WR}_{\boldsymbol{rr}}\mathbf{W}^*=\mathbf{I}$ and that $\mathbf{WH}=\mathbf{\Lambda}^{-1/2}\mathbf{SV}^*$. Replacing the singular values of $\mathbf{WH}$ by unity produces its polar factor $\mathbf{J}^T\mathbf{V}^*$. From Section III, the WR detector is then $\mathbf{C}_{WR}=\mathbf{JQW}$ where $\mathbf{JQ}$ is the polar factor of $(\mathbf{WH})^*$, or $\mathbf{JQ}=\mathbf{VJ}$. Thus, we have $\mathbf{C}_{WR}=\mathbf{VJW}=\mathbf{VJ\Lambda}^{-1/2}\mathbf{U}^*$. This proves (7). If we define the diagonal matrix $\tilde{\mathbf{\Lambda}}=\mathbf{J\Lambda J^T}=(\tilde{\boldsymbol{S}}^2+\sigma^2\mathbf{I})^{-1/2}$ then (8) follows from (7) and the identity $\mathbf{J\Lambda}=\tilde{\mathbf{\Lambda}}\mathbf{J}$.

We now establish by contradiction the uniqueness of the WR detector. Suppose $\mathbf{C}_1=\mathcal{P}\{\mathbf{W}_1\mathbf{H}\}^*\mathbf{W}_1$ and $\mathbf{C}_2=\mathcal{P}\{\mathbf{W}_2\mathbf{H}\}^*\mathbf{W}_2$ denote two distinct WR detectors derived from whiteners $\mathbf{W}_1$ and $\mathbf{W}_2$, respectively, where $\mathcal{P}\{\mathbf{A}\}=\mathbf{U}_A\mathbf{V}_A^*$ denotes the polar factor of $\mathbf{A}=\mathbf{U}_A\mathbf{S}_A\mathbf{V}_A^*$. Since $\mathbf{W}_1$ and $\mathbf{W}_2$ are both whiteners, there exists a unitary $\mathbf{Q}$ such that $\mathbf{W}_2=\mathbf{QW}_1$. It then follows that $\mathbf{C}_2=\mathcal{P}\{\mathbf{QW}_1\mathbf{H}\}^*\mathbf{QW}_1=(\mathbf{Q}\mathcal{P}\{\mathbf{W}_1\mathbf{H}\})^*\mathbf{QW}_1=\mathcal{P}\{\mathbf{W}_1\mathbf{H}\}^*\mathbf{W}_1=\mathbf{C}_1$, a contradiction. □

Appendix II
Proof of Lemma 2

Using (8), we can express the error of the WR detector as

$$\begin{aligned}\boldsymbol{e}_k&=(\mathbf{C}_{WR}\mathbf{H}-\mathbf{I})\boldsymbol{x}_k+\mathbf{C}_{WR}\boldsymbol{n}_k\\&=\mathbf{V}[\tilde{\mathbf{\Lambda}}\tilde{\mathbf{S}}-\mathbf{I}]\mathbf{V}^*\boldsymbol{x}_k+\mathbf{V}\tilde{\mathbf{\Lambda}}\mathbf{JU}^*\boldsymbol{n}_k\end{aligned}\tag{30}$$

where $\tilde{\mathbf{\Lambda}}=(\tilde{\mathbf{S}}^2+\sigma^2\mathbf{I})^{-1/2}$. The covariance $\mathbf{R}_{\boldsymbol{ee}}=E[\boldsymbol{e}_k\boldsymbol{e}_k^*]$ of this error is given by

$$\begin{aligned}\mathbf{R}_{\boldsymbol{ee}}&=\mathbf{V}\left[\tilde{\mathbf{\Lambda}}\tilde{\boldsymbol{S}}-\mathbf{I}\right]^2\mathbf{V}^*+\sigma^2\mathbf{V}\tilde{\mathbf{\Lambda}}^2\mathbf{V}^*\\&=\mathbf{V}\left[\tilde{\mathbf{\Lambda}}^2\tilde{\mathbf{S}}^2+\mathbf{I}-2\tilde{\mathbf{\Lambda}}\tilde{\mathbf{S}}+\sigma^2\tilde{\mathbf{\Lambda}}^2\right]\mathbf{V}^*\\&=2\mathbf{V}\left[\mathbf{I}-\tilde{\mathbf{\Lambda}}\tilde{\mathbf{S}}\right]\mathbf{V}^*.\end{aligned}\tag{31}$$

The MSE of the ith user $E[|e_k^{(i)}|^2]$ is then given by (10).

Similarly, using (3), we can express the error of the MMSE detector as

$$\begin{aligned}\boldsymbol{e}_k&=(\mathbf{C}_{\mathrm{MMSE}}\mathbf{H}-\mathbf{I})\boldsymbol{x}_k+\mathbf{C}_{\mathrm{MMSE}}\boldsymbol{n}_k\\&=[(\mathbf{H}^*\mathbf{H}+\sigma^2\mathbf{I})^{-1}\mathbf{H}^*\mathbf{H}-\mathbf{I}]\boldsymbol{x}_k+(\mathbf{H}^*\mathbf{H}+\sigma^2\mathbf{I})^{-1}\\&\quad\cdot\mathbf{H}^*\boldsymbol{n}_k\\&=(\mathbf{H}^*\mathbf{H}+\sigma^2\mathbf{I})^{-1}[\mathbf{H}^*\mathbf{H}-(\mathbf{H}^*\mathbf{H}+\sigma^2\mathbf{I})]\boldsymbol{x}_k\\&\quad+(\mathbf{H}^*\mathbf{H}+\sigma^2\mathbf{I})^{-1}\mathbf{H}^*\boldsymbol{n}_k\\&=-\sigma^2\mathbf{V}\tilde{\mathbf{\Lambda}}^2\mathbf{V}^*\boldsymbol{x}_\mathbf{k}+\mathbf{V}\tilde{\mathbf{\Lambda}}^2\mathbf{S}^*\mathbf{U}^*\boldsymbol{n}_\mathbf{k}.\end{aligned}\tag{32}$$

The covariance of this error is

$$\mathbf{R}_{\boldsymbol{ee}}=\boldsymbol{V}[\sigma^4\tilde{\mathbf{\Lambda}}^4+\sigma^2\tilde{\mathbf{\Lambda}}^4\tilde{\mathbf{S}}^2]\mathbf{V}^*=\sigma^2\mathbf{V}\tilde{\mathbf{\Lambda}}^2\mathbf{V}^*\tag{33}$$

and the corresponding MSE of the ith user is then given by (11). □

Appendix III
Proof of Lemma 4

Given an Hermitian matrix $\mathbf{R}$ of dimension $m\times m$ and rank $n\le m$, there exists a square-root matrix $\mathbf{S}$ such that $\mathbf{SS}^*=\mathbf{R}$. Since $\mathbf{R}$ has rank n, the rows of $\mathbf{S},\{\boldsymbol{s}_1,\boldsymbol{s}_2,\ldots,\boldsymbol{s}_m\}$, span an n-dimensional space $\mathcal{S}$. Performing the Gram–Schmidt orthonormalization procedure on the ordered rows of $\mathbf{S}$ produces a set of m row vectors $\{\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_m\}$, exactly $m-n$ of which are zero and n of which form an orthonormal basis of $\mathcal{S}$. Thus, we can write

$$\mathbf{S}=\mathbf{FV}\tag{34}$$

where the rows of $\mathbf{V}$ are the vectors $\boldsymbol{v}_i, 1\le i\le m$ and $\mathbf{F}$ is a Gram matrix

$$\mathbf{F}=\begin{bmatrix}F_{11} & & & \mathbf{0}\\ F_{21} & F_{22} & & \\ \vdots & & \ddots & \\ F_{m1} & F_{m2} & \cdots & F_{mm}\end{bmatrix}\tag{35}$$

with $F_{ij}\equiv\langle\boldsymbol{s}_i\boldsymbol{v}_j\rangle$. There exists a set of $m-n$ unit-norm vectors $\{\tilde{\boldsymbol{v}}_1,\tilde{\boldsymbol{v}}_2,\ldots,\tilde{\boldsymbol{v}}_{m-n}\}$ orthogonal to the n nonzero rows of $\mathbf{V}$. Let $\tilde{\mathbf{V}}$ be a unitary matrix formed by replacing the zero rows of $\mathbf{V}$ with this set of vectors. Because the columns of $\mathbf{F}$ multiplying the zero rows of $\mathbf{V}$ in (34) are also zero, it follows that $\mathbf{S} = \mathbf{F}\tilde{\mathbf{V}}$. Because $\mathbf{F}$ may have complex diagonal elements, we define $\mathbf{G} = \mathbf{F}\,\mathrm{diag}\,(|F_{ii}|/F_{ii})$ and $\mathbf{U} = \tilde{\mathbf{V}}\,\mathrm{diag}\,(F_{ii}/|F_{ii}|)$. It follows that $\mathbf{S} = \mathbf{GU}$ and that

$$\mathbf{R} = \mathbf{GG}^*. \tag{36}$$

The factorization in (36) is unique because for any other square-root $\tilde{\mathbf{S}} \neq \mathbf{S}$ there exists a unitary matrix $\mathbf{Q}$ such that $\tilde{\mathbf{S}} = \mathbf{SQ} = \mathbf{GUQ} = \mathbf{G}\tilde{\mathbf{U}}$ where $\tilde{\mathbf{U}}$ is also unitary.

The factor $\mathbf{G}$ can be decomposed as $\mathbf{G} = \mathbf{MD}$ where $\mathbf{M}$ is lower triangular and monic and where $\mathbf{D} = \mathrm{diag}\,(\mathbf{G})$ such that

$$\mathbf{R} = \mathbf{MD}^2\mathbf{M}^*. \tag{37}$$

If any of the first $m-1$ diagonal elements $D_{jj}, 1 \leq j \leq m-1$ of $\mathbf{D}$ are zero, then elements $M_{ij}, i > j$ of $\mathbf{M}$ are not unique. Therefore, $\mathbf{M}$ is unique if and only if the first $m-1$ rows of $\mathbf{S}$ are linearly independent, or equivalently, if and only if the first $m-1$ rows (or columns) of $\mathbf{R}$ are linearly independent. □

APPENDIX IV
PROOF OF THEOREM 2

Given a strictly lower triangular predictor $\mathbf{P}$ and prediction error $\boldsymbol{e}_k = (\mathbf{I} - \mathbf{P})\boldsymbol{r}_k$ the MSE $J = E[||\boldsymbol{e}_k||^2]$ can be expressed as

$$J = \mathrm{tr}\,[(\mathbf{I} - \mathbf{P})\mathbf{R}(\mathbf{I} - \mathbf{P})^*] \tag{38}$$

where $\mathbf{R} = E[\boldsymbol{r}_k\boldsymbol{r}_k^*]$. Applying the factorization in (18) yields

$$J = \mathrm{tr}\,[(\mathbf{I} - \mathbf{P})\mathbf{MD}^2\mathbf{M}^*(\mathbf{I} - \mathbf{P})^*]. \tag{39}$$

Because $(\mathbf{I} - \mathbf{P})\mathbf{M}$ is monic and lower triangular, it can be expressed as $\mathbf{I} + \mathbf{B}$ where $\mathbf{B}$ is strictly lower triangular. Thus, the cost function in terms of $\mathbf{B}$ is

$$\begin{aligned} J &= \mathrm{tr}\,[(\mathbf{I} + \mathbf{B})\mathbf{D}^2(\mathbf{I} + \mathbf{B})^*] \\ &= \mathrm{tr}\,[\mathbf{D}^2 + \mathbf{BD}^2\mathbf{B}^* + \mathbf{BD}^2 + \mathbf{D}^2\mathbf{B}^*]. \end{aligned} \tag{40}$$

Because $\mathbf{B}$ is strictly lower triangular, the traces of $\mathbf{BD}^2$ and $\mathbf{D}^2\mathbf{B}^*$ are both zero in (40). Furthermore, the trace of $\mathbf{D}^2$ is independent of $\mathbf{B}$. Thus, it suffices to minimize $\mathrm{tr}\,[\mathbf{BD}^2\mathbf{B}^*]$, which is clearly accomplished by any strictly lower triangular $\tilde{\mathbf{B}}$ in the left null space of $\mathbf{D}$. The best predictor can thus be expressed as

$$\mathbf{P} = \mathbf{I} - \left(\mathbf{I} + \tilde{\mathbf{B}}\right)\mathbf{M}^{-1} \tag{41}$$

with $\tilde{\mathbf{B}}$ so defined. Observe that $(\mathbf{I} + \tilde{\mathbf{B}})\mathbf{M}^{-1}$ is both monic and lower triangular, and that it diagonalizes $\mathbf{R}$:

$$\begin{aligned} &\left[\left(\mathbf{I} + \tilde{\mathbf{B}}\right)\mathbf{M}^{-1}\right]\mathbf{R}\left[(\mathbf{M}^{-1})^*\left(\mathbf{I} + \tilde{\mathbf{B}}\right)^*\right] \\ &\quad = \left(\mathbf{I} + \tilde{\mathbf{B}}\right)\mathbf{D}^2\left(\mathbf{I} + \tilde{\mathbf{B}}\right)^* = \mathbf{D}^2. \end{aligned} \tag{42}$$

Hence, for any inverse generalized Cholesky factor $\mathbf{M}^{-1}$, the product $(\mathbf{I} + \tilde{\mathbf{B}})\mathbf{M}^{-1} \equiv \tilde{\mathbf{M}}^{-1}$ is the inverse of some other Cholesky factor. Thus (41) reduces to (19). □

APPENDIX V
A WR DETECTOR FOR CHANNELS WITHOUT NOISE

Both the WR structure of Section III and its adaptive implementation of Section V are derived under the assumption of nonzero noise. Although a good assumption in practice, the special case of zero noise is also of interest because it provides insight into the behavior of the WR detector in the limit of high SNR. Therefore, this appendix contains a brief summary of a WR detector for channels without noise. We emphasize that the discussion that follows applies to the whiten-first strategy of Section III only. The project-first strategy of Section IV and its adaptive implementation of Section V-C can be applied to noiseless channels without modification.

Without noise, the covariance matrix of the observation vector $\boldsymbol{r}_k$ in (1) is $\mathbf{R}_{\boldsymbol{rr}} = \mathbf{HH}^*$ with rank n. Therefore, if the channel is tall $(m > n)$ there does not exist an $m \times m$ whitening matrix $\mathbf{W}$. Consequently, the whiten-first approach illustrated in Fig. 1 and described by (6) is not valid. Nevertheless, there does exist a *short* $n \times m$ whitener $\mathbf{W}$ satisfying $\mathbf{WHH}^*\mathbf{W}^* = \mathbf{I}$. And this identity implies that $\mathbf{WH} = \mathbf{T}$ for some unitary matrix $\mathbf{T}$. Hence, the WR filter $\mathbf{T}^*\mathbf{W}$ achieves zero MSE, which is certainly minimal. In fact, any zero-forcing detector of the form $\mathbf{C}_{ZF} = \mathbf{H}^\dagger + \mathbf{N}$ (where $\mathbf{N}$ is any $n \times m$ matrix in the left null space of $\mathbf{H}$) can be interpreted as a WR detector when it is factored according to the QR-factorization theorem [23], $\mathbf{C}_{ZF} = \mathbf{QW}$. Thus, in the absence of noise, the WR detector is not unique.

According to (18) and (19), the least-mean-square linear-prediction error $\boldsymbol{e}_k = (\mathbf{I} - \mathbf{P})\boldsymbol{r}_k$ has a diagonal covariance matrix $\mathbf{R}_{\boldsymbol{ee}} = \mathbf{D}^2$, where $\mathbf{HH}^* = \mathbf{MD}^2\mathbf{M}^*$ is a generalized Cholesky factorization. Because $\mathbf{H}$ has rank n, exactly $m - n$ of the diagonal elements of $\mathbf{D}$ are zero, which implies that the corresponding components of the error signal $\boldsymbol{e}_k$ are identically zero. Clearly we lose nothing by discarding these zeros, thereby producing a reduced vector $\tilde{\boldsymbol{e}}_k$ of dimension n. Let $\mathbf{J}$ denote the $n \times m$ matrix that extracts the nonzero components of $\boldsymbol{e}_k$, so that $\tilde{\boldsymbol{e}}_k = \mathbf{J}\boldsymbol{e}_k$ has covariance matrix $\mathbf{R}_{\tilde{\boldsymbol{e}}\tilde{\boldsymbol{e}}} = \mathbf{JD}^2\mathbf{J}^* = \tilde{\mathbf{D}}^2$ where $\tilde{\mathbf{D}} = \mathbf{JD}$ is a full-rank $n \times n$ diagonal matrix containing all of the nonzero components of $\mathbf{D}$. The diagonal filter $\tilde{\mathbf{D}}^{-1}$ then whitens the reduced error signal. An $n \times m$ whitening matrix can thus be expressed as $\mathbf{W} = \tilde{\mathbf{D}}^{-1}\mathbf{J}(\mathbf{I} - \mathbf{P})$.

The preceding development suggests a method for blind adaptive implementation of $\mathbf{W}$ for the case of low noise. The adaptive predictor of (20) is guaranteed to converge to a solution of the form (19). After convergence, the $m - n$ components of the prediction error that are nearly zero can be discarded. The n remaining error components can be adaptively scaled according to (21). (The MPLL can be applied without modification after the whitener.)

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful comments.

REFERENCES

[1] L. J. Griffiths and M. J. Rude, "The P-vector algorithm: A linearly constrained point of view," in *Proc. 20th Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, 1986, pp. 457–461.

[2] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 276–280, Mar. 1986.

[3] R. Roy and T. Kailath, "ESPRIT—Estimation of signal parameter via rotational invariance techniques," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 37, pp. 984–993, July 1989.

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

[5] X. Wang and H. V. Poor, "Blind adaptive interference suppression for CDMA communication based on eigenspace tracking," in *Proc. Conf. Information Sciences and Systems*, Baltimore, MD, 1997, pp. 468–473.

[6] M. L. Honig, "A comparison of subspace adaptive filtering techniques for DS-CDMA interference suppression," in *Proc. Military Conf. on Communications*, Monterey, CA, Nov. 1997, vol. 2, pp. 836–840.

[7] J. R. Treichler and M. G. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 33, pp. 420–431, Apr. 1985.

[8] A. Batra and J. R. Barry, "Blind cancellation of Co-channel interference," in *Proc. Globecom*, Singapore, China, 1995, vol. 1, pp. 157–162.

[9] H. Oda and Y. Sato, "A method of multi-dimensional blind equalization," in *Proc. Int. Symp. on Information Theory*, San Antonio, TX, 1993, p. 327.

[10] L. Castedo, C. J. Escudero, and A. Dapena, "A blind signal separation method for multiuser communications," *IEEE Trans. Signal Processing*, vol. 45, pp. 1343–1348, May 1997.

[11] C. B. Papadias and A. J. Paulraj, "A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays," *IEEE Signal Processing Lett.*, vol. 4, pp. 178–181, June 1997.

[12] J.-F. Cardoso, "Source separation using higher order moments," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, Glasgow, UK, May, 1989, vol. 4, pp. 2109–2112.

[13] L. Tong, *et al.*, "Indeterminacy and indentifiability of blind identification," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 499–509, May 1991.

[14] N. Delfosse and P. Loubaton, "Adaptive separation of independent source: A deflation approach," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, Adelaide, Australia, 1994, vol. 4, pp. 41–44.

[15] J. T. Tugnait, "On blind MIMO channel estimation and blind signal separation in unknown additive noise," in *Proc. IEEE Signal Processing Workshop Signal Processing Advances in Wireless Communications (SPAWC)*, Paris, France, 1997, pp. 53–60.

[16] M. Gaeta and J. L. Lacoume, "Source separation without *a priori* knowledge: The maximum likelihood solution," in *Proc. European Signal Processing Conf.*, 1990, pp. 621–624.

[17] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[18] L. Tong, Y. Inouye, and R. Liu, "Waveform preserving blind estimation of multiple independent sources," *IEEE Trans. Signal Processing*, vol. 41, pp. 2461–2470, July 1993.

[19] J. F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.

[20] K. Abed-Meraim, *et al.*, "Asymptotic performance of second-order blind source separation," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, Adelaide, Australia, Apr. 1994, vol. 4, pp. 277–280.

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

[22] Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Select. Areas Commun.*, vol. 8, pp. 683–690, May 1990.

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

[24] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Press, 1996.

[25] R. Lupas and S. Verdú, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Commun.*, vol. 38, no. 4, pp. 496–508, Apr. 1990.

[26] S. Bellini, "Bussgang techniques for blind deconvolution and equalization," in *Blind Deconvolution*. S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.

[27] J. R. Barry and A. Batra, "A multidimensional phase-locked loop for blind equalization of multi-input multi-output channels," in *Proc. Int. Conf. Communications*, Dallas, TX, 1996, vol. 3, pp. 1307–1312.

[28] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, no. 1, Jan. 1995.

[29] R. D. DeGroat, "Noniterative subspace tracking," *IEEE Trans. Signal Processing*, vol. 40, no. 3, pp. 571–577, Mar. 1992.

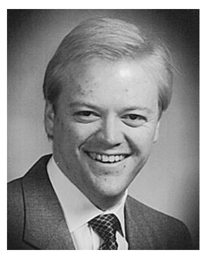
[30] G. W. Stewart, "An updating algorithm for subspace tracking," *IEEE Trans. Signal Processing*, vol. 40, no. 6, pp. 1535–1541, June 1992.

[31] C. H. Bischof, "On updating signal subspaces," *IEEE Trans. Signal Processing*, vol 40, no. 1, pp. 96–105, Jan. 1992.

[32] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

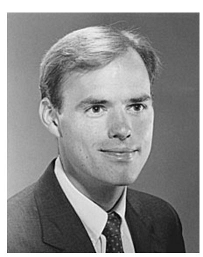
[33] A. Gorokhov, P. Loubaton, and E. Moulines, "Second order blind equalization in multiple input multiple output FIR systems: Part I: Identifiability and estimation," *IEEE Trans. Signal Processing*, submitted for publication.

[34] D. G. Gesbert, P. Duhamel, and S. Mayrargue, "A bias removal technique for the prediction-based blind adaptive multichannel deconvolution," in *Proc. 29th Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, 1995, pp. 275–279.



Richard T. Causey (S'92) received the B.S.E.E. degree (summa cum laude) from the University of Alabama in 1985 and the M.S.E.E. degree from the Georgia Institute of Technology, Atlanta, in 1993. He is currently a Ph.D. candidate at the Georgia Institute of Technology.

From 1985 to 1988 he was with Motorola, where he was involved in the design of high-speed, voiceband modems for the public telephone network. From 1988 to 1992 he was with ADTRAN, where he helped to develop ISDN and digital-subscriber-loop products. He is currently a Research Assistant at the Georgia Institute of Technology. His research interests include multiuser communication theory and broadband systems.



John R. Barry (S'85–M'86) received the B.S. degree in electrical engineering from the State University of New York at Buffalo in 1986 and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley in 1987 and 1992, respectively.

Since 1992, he has been with the Georgia Institute of Technology, Atlanta, where he is an Associate Professor at the School of Electrical and Computer Engineering. His current research interests include wireless communications, equalization, and multiuser communications. He is the author of *Wireless Infrared Communication*, (Norwell, MA: Kluwer, 1994.)