Blind and Semi-Blind Equalization for Generalized Space-Time Block Codes

A. Lee Swindlehurst, *Senior Member, IEEE*, and Geert Leus

***Abstract*—This paper presents a general framework for space-time codes (STCs) that encompasses a number of recently proposed STC schemes as special cases. The STCs considered are block codes that employ arbitrary redundant linear precoding of a given data sequence together with embedded training symbols, if any. The redundancy introduced by the linear precoding imposes structure on the received data that under certain conditions can be exploited for blind or semi-blind estimation of the transmitted sequence, a linear receiver that recovers the sequence, or both simultaneously. Algorithms based on this observation are developed for the single-user flat-fading case and then extended to handle multiple users, frequency-selective fading, as well as situations where the channel is rank deficient, or there are fewer receive than transmit antennas.**

***Index Terms*—Array signal processing, communication channels, diversity methods, equalizers, fading channels, MIMO systems, multipath channels.**

I. INTRODUCTION

THE advantages of using multiple antennas at both the transmit and receive ends of a wireless communications link have recently been noted [1], [2]. A number of space-time codes (STCs) have been proposed that exploit the potential for increased throughput and diversity that such systems offer. For most algorithms, these gains can only be realized when that the multiple-input multiple-output (MIMO) channel separating the transmitter and receiver has been identified. While training data can be used to estimate the channel, this approach consumes precious bandwidth and reduces throughput. One approach to overcoming this difficulty is the use of differential STCs [3], [4], although such techniques incur a 3-dB penalty in SNR.

The large body of previous research on blind multiuser, multichannel estimation and equalization is applicable to the MIMO problem since the data broadcast from different transmit antennas can be thought of as data from different users. However, only recently have techniques appeared that exploit the structure built into space-time encoded signals. Many of these techniques have focused on the special structure of the so-called *space-time block codes* (STBCs) described in [5]–[7], or generalizations thereof. Examples of methods that exploit STBCs for blind and semi-blind channel estimation include [8], [9] and also the work of [10]–[13] which combine the use of STBCs with redundant linear precoders. Algorithms have also been presented for "modulation-induced" block coding in [14]–[17] and for circulant codes in [18].

In this paper, two different approaches are presented in which the structure of a certain class of STCs is exploited either for blind (semi-blind) equalization of the channel or for direct estimation of the transmitted data sequence. The codes considered employ generalized redundant linear precoders; in other words, different linearly transformed versions of the same desired data sequence are broadcast from each of the transmit antennas. This framework is general enough to include the STBCs of [5]–[7] as well as the specific codes used in [10]–[23] as special cases. Codes that fit within the framework are referred to herein as *generalized* space-time block codes (GSTBCs). In the noiseless case, it is shown how the GSTBC structure can be exploited to construct a set of channel-independent linear equations whose solution simultaneously yields the transmitted data sequence and a vector containing all possible zero-forcing receivers. While the details of the algorithms are presented for the single-user flat-fading case, extensions to situations involving frequency-selective fading and multiple users are discussed, along with modifications of the algorithm required when there are more transmit than receive antennas or the channel is rank deficient.

The next section describes the general single-user flat-fading data model considered and presents the GSTBC framework. Specific examples of recently proposed codes that fit within the framework are included. The proposed blind and semi-blind algorithms are derived in Section III, and the various extensions mentioned above are discussed in Section IV. Section V describes the results from a number of simulation examples that illustrate the performance of the algorithms in various situations.

II. DATA MODEL

To begin, assume a single-user transmit array with $K > 1$ elements, a receive array with $M > 1$ elements, and a flat-fading channel. If the receive array is sampled once per symbol over N consecutive symbol periods, the following model results:

$$\hat{\mathbf{X}} = \mathbf{H}\mathbf{S}_K + \mathbf{N} \tag{1}$$

where $\hat{\mathbf{X}}$ is the $M \times N$ matrix of received data, $\mathbf{H}$ is the $M \times K$ channel matrix, $\mathbf{N}$ is the additive noise and interference, and $\mathbf{S}_K$ is the $K \times N$ matrix containing the transmitted symbols. The subscript on $\mathbf{S}$ is used to explicitly indicate the number of rows in the matrix. The $\hat{(\cdot)}$ symbol is used to differentiate $\hat{\mathbf{X}}$ from its

Manuscript received October 17, 2001; revised June 26, 2002. This work was supported by the National Science Foundation under Wireless Initiative Grant CCR-9979452 and Information Technology Research Grant CCR-0081476. The associate editor coordinating the review of this paper and approving it for publication was Prof. Georgios B. Giannakis.

A. L. Swindlehurst is with the Department of Electrical and Computer Engineering, Brigham Young University, Provo, UT 84602 USA (e-mail: swindle@ee.byu.edu).

G. Leus is with the Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven, Leuven, Belgium (e-mail: geert.leus@esat.kuleuven.ac.be).

Publisher Item Identifier 10.1109/TSP.2002.803319.

noise-free counterpart $\mathbf{X} = \mathbf{H}\mathbf{S}_K$. For now, it will be assumed that $M \geq K$ and that $\mathbf{H}$ is full rank, but these assumptions will be relaxed in Section IV, as will the restriction to the single-user and flat-fading cases.

Let the $N \times 1$ vector $\mathbf{s}_k = [s_k(0) \cdots s_k(N-1)]^T$ represent the symbol sequence transmitted from antenna k, and define

$$\mathcal{S} = \begin{bmatrix} \text{real}\left(\mathbf{S}_K^T\right) \\ \text{imag}\left(\mathbf{S}_K^T\right) \end{bmatrix}, \quad \tilde{\mathbf{s}}_k = \begin{bmatrix} \text{real}\left(\mathbf{s}_k\right) \\ \text{imag}\left(\mathbf{s}_k\right) \end{bmatrix} \tag{2}$$

so that the K vectors $\tilde{\mathbf{s}}_k$ form the columns of $\mathcal{S}$. We assume that embedded within each $\tilde{\mathbf{s}}_k$ are contributions from a set of N_u unknown, information-bearing symbols $\mathbf{u}$, together perhaps with contributions from some known training data $\mathbf{t}_k$ as well. In particular, we assume

$$\tilde{\mathbf{s}}_k = \mathcal{U}_k \tilde{\mathbf{u}} + \tilde{\mathbf{t}}_k, \qquad k = 1, \ldots, K \tag{3}$$

where

$$\tilde{\mathbf{u}} = \begin{bmatrix} \text{real}(\mathbf{u}) \\ \text{imag}(\mathbf{u}) \end{bmatrix} \quad \tilde{\mathbf{t}}_k = \begin{bmatrix} \text{real}\left(\mathbf{t}_k\right) \\ \text{imag}\left(\mathbf{t}_k\right) \end{bmatrix} \tag{4}$$

and $\mathcal{U}_k$ is a real-valued, $2N \times 2N_u$ linear precoder. The ratio $0 < N/N_u \leq K$ yields the rate of the code or the average number of symbols transmitted per symbol period. In most applications, the training and unknown data symbols are transmitted at different time instants, in which case, the rows of $\mathcal{U}_k$ will be zero when the elements of $\tilde{\mathbf{t}}_k$ are not, and vice versa. However, the formulation given in (3) is general enough to accommodate schemes where the training and unknown data symbols are mixed and transmitted simultaneously [24].

The data, precoders, and training are assumed to have been chosen so that $\mathbf{S}_K$ is full rank but, at this point, are otherwise arbitrary. The algorithms presented in the next section will require the following three additional conditions.

- $K(N-K) \geq N_u$.
- The following matrix is full rank N_u :

$$\mathcal{U} = \begin{bmatrix} \mathcal{U}_1 \\ \vdots \\ \mathcal{U}_K \end{bmatrix}. \tag{5}$$

- $\tilde{\mathbf{t}} \notin \text{span}(\mathcal{U})$ when $\tilde{\mathbf{t}} \neq 0$, where

$$\tilde{\mathbf{t}} = \begin{bmatrix} \tilde{\mathbf{t}}_1 \\ \vdots \\ \tilde{\mathbf{t}}_K \end{bmatrix}. \tag{6}$$

While splitting the data into real and imaginary parts as above provides the most general STC framework, it is often convenient to use a more compact notation involving complex quantities when the transformation matrix $\mathcal{U}_k$ has the following form:

$$\mathcal{U}_k = \begin{bmatrix} \text{real}\left(\mathbf{U}_k\right) & -\text{imag}\left(\mathbf{U}_k\right) \\ \text{imag}\left(\mathbf{U}_k\right) & \text{real}\left(\mathbf{U}_k\right) \end{bmatrix} \tag{7}$$

for some $N \times N_u$ complex matrix $\mathbf{U}_k$. In this case, (3) may be rewritten as

$$\mathbf{s}_k = \mathbf{U}_k \mathbf{u} + \mathbf{t}_k. \tag{8}$$

Both the models (3) and (8) allow each transmit antenna to use different training data and different transformations of the unknown data spread over different time instants. The use of linear (affine) precoders like (8) have been proposed for both single-channel block transmission schemes [25]–[27] and multichannel systems as well [11]–[13], [22], [28].

As mentioned above, codes that obey the model described by (3)–(8) will be referred to as *generalized* space-time block codes (GSTBCs). The GSTBC framework is very general and encompasses many types of popular STCs. Some of these include the following (all linear precoders are shown for the case where $\mathbf{t} = 0$).

1) Example 1: The $K = 2$ STBC of [5] satisfies (3) with $N = N_u$ and

$$\mathcal{U}_1 = \begin{bmatrix} \mathbf{I}_{N/2} \otimes \left(\mathbf{J}\tilde{\mathbf{I}}\right) & 0 \\ 0 & \mathbf{I}_N \end{bmatrix}$$
$$\mathcal{U}_2 = \begin{bmatrix} \mathbf{I}_{N/2} \otimes \tilde{\mathbf{I}} & 0 \\ 0 & \mathbf{I}_{N/2} \otimes \mathbf{J} \end{bmatrix} \tag{9}$$

where $\mathbf{I}_a$ indicates an $a \times a$ identity matrix, and

$$\tilde{\mathbf{I}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \mathbf{J} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{10}$$

Similar transformations exist for the STBCs with larger values of K described in [6], [7].

2) Example 2: The codes described in [10], [12], and [13] combine the structure of the $K = 2$ STBC together with the use of two $N/2 \times N_u/2$ "sub-precoders" $\mathbf{C}_1$ and $\mathbf{C}_2$, where $N > N_u$. When cast in the framework of (3), this method results in the following two linear precoders

$$\mathcal{U}_1 = \begin{bmatrix} \text{real}\left(\mathbf{C}_1\right) & 0 & -\text{imag}\left(\mathbf{C}_1\right) & 0 \\ 0 & -\text{real}\left(\mathbf{C}_2\right) & 0 & \text{imag}\left(\mathbf{C}_2\right) \\ \text{imag}\left(\mathbf{C}_1\right) & 0 & \text{real}\left(\mathbf{C}_1\right) & 0 \\ 0 & \text{imag}\left(\mathbf{C}_2\right) & 0 & \text{real}\left(\mathbf{C}_2\right) \end{bmatrix} \tag{11}$$

$$\mathcal{U}_2 = \begin{bmatrix} -\mathbf{J} \otimes \mathbf{I}_{N/2} & 0 \\ 0 & \mathbf{J} \otimes \mathbf{I}_{N/2} \end{bmatrix} \mathcal{U}_1 \tag{12}$$

where $\mathbf{J}$ is defined in (10).

3) Example 3: The so-called *linear dispersion* (LD) codes of [23] satisfy (3) as well. For these codes

$$\mathbf{S}_K^T = \sum_{q=1}^{N_u} \alpha_q \mathbf{A}_q + j\beta_q \mathbf{B}_q \tag{13}$$

where α_q and β_q are real scalars, and $\mathbf{A}_q, \mathbf{B}_q$ are $N \times K$ complex matrices. The LD coding scheme is equivalent to (3) with the following choice for $\mathcal{U}_k$:

$$\mathcal{U}_k = \begin{bmatrix} \text{real}\left(\mathcal{A}_k\right) & -\text{imag}\left(\mathcal{B}_k\right) \\ \text{imag}\left(\mathcal{A}_k\right) & \text{real}\left(\mathcal{B}_k\right) \end{bmatrix}$$

where

$$\mathcal{A}_k = [\mathbf{A}_1(:,k) \cdots \mathbf{A}_{N_u}(:,k)]$$
$$\mathcal{B}_k = [\mathbf{B}_1(:,k) \cdots \mathbf{B}_{N_u}(:,k)]$$

and $(:,k)$ denotes the kth column of the associated matrix (as in Matlab notation). A similar though slightly less general coding framework was also considered in [21].

4) Example 4: In [19] and [22], unitary constellation-rotating precoders are used that satisfy (8) with

$$\mathbf{U}_k = \mathbf{D}_k \mathbf{\Theta}_k$$

where $\mathbf{\Theta}_k$ is $N \times N$ and unitary, and $\mathbf{D}_k$ is a diagonal matrix whose N diagonal elements are formed from the kth row of another unitary matrix.

5) Example 5: For the code described in [18], $\mathbf{S}_K$ is circulant and Hankel; therefore, (8) applies with $N = N_u$ and

$$\mathbf{U}_k = \begin{bmatrix} 0 & \mathbf{I}_{N-1} \\ 1 & 0 \end{bmatrix}^{k-1}.$$

6) Example 6: For the modulation-induced code of [14] and [16], $\mathbf{U}_k$ is $N \times N$ and chosen to be diagonal. The method of [17] is similar, except it mixes several independent symbol streams together with different coding matrices for each signal and each transmit antenna (i.e., $\mathbf{S}_K$ is a sum of the type of single-user signal matrices employed by [14] and [16]).

7) Example 7: A method related to Example 6 is the full-rate ($N_u = KN$) code of [15], which employs a different periodic diagonal precoder for the unique data sequence broadcast from each transmit antenna. The corresponding linear precoders for this approach are given by the $N \times KN$ matrices

$$\mathbf{U}_k = \mathbf{I}_K(k,:) \otimes \mathbf{I}_P \otimes \mathbf{D}_k$$

where $\mathbf{D}_k$ is $N/P \times N/P$ and diagonal, and P is the period of the modulating code.

8) Example 8: A number of researchers have proposed beamforming-based MIMO systems in which one or more independent waveforms are broadcast using different (often orthogonal) transmit beamformers. In this approach, $\mathbf{S}_K = \mathbf{W}\mathbf{Y}$, where the columns of the $K \times K'$ matrix $\mathbf{W}$ represent the beamformer weights used to transmit the K' signals that make up the rows of the $K' \times N$ matrix $\mathbf{Y}$. If we let $\mathbf{y}_k^T$ represent the kth row of $\mathbf{Y}$, then for this method

$$\mathbf{U}_k = \mathbf{W}(k,:) \otimes \mathbf{I}_N \qquad \mathbf{u} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_{K'} \end{bmatrix}$$

which corresponds to a rate K' code (i.e., $\mathbf{U}_k$ is $N \times K'N$).

While the STCs mentioned above were not originally formulated with training data in mind, the presence of such data is easily accommodated by the GSTBC model for all of these codes.

As described in the next section, the structure induced in the data by (3)–(8) leads to a set of linear equations that can be solved for $\mathbf{u}$ and all possible zero-forcing receivers simultaneously. Algorithms for estimating $\mathbf{u}$ and these equalizers in the presence of noise are then presented. It is important to note that although the algorithms described below are very flexible and can be applied to a large number of different space-time codes, certain codes may admit simpler solutions that take advantage of additional structure that is not assumed here.

III. GSTBC Equalization and Sequence Estimation

Assume for the moment that $M \geq K$ and that the linear precoders and the transmitted symbols have been chosen so that $\mathbf{S}_K$ is full rank. To describe the proposed algorithms, consider an SVD of the rank K noiseless matrix $\mathbf{X}$

$$\mathbf{X} = \mathbf{H}\mathbf{S}_K = \mathbf{F}\mathbf{\Lambda}\mathbf{V}^* \tag{14}$$

where $\mathbf{F}$ is $M \times K$, $\mathbf{\Lambda}$ is $K \times K$, $\mathbf{V}$ is $N \times K$, and $(\cdot)^*$ denotes the complex conjugate transpose. If $\mathbf{S}_K$ is full rank, there exists a full-rank $K \times K$ matrix $\mathbf{H}_v$ that satisfies

$$\mathbf{V}^* = \mathbf{H}_v\mathbf{S}_K \qquad \Longrightarrow \qquad \bar{\mathbf{V}} = \mathbf{S}_K^T\mathbf{H}_v^T \tag{15}$$

where $\bar{(\cdot)}$ denotes conjugation.

Let $\mathbf{w}_k$ denote the zero-forcing receiver that recovers the kth transmitted signal from $\bar{\mathbf{V}}$

$$\bar{\mathbf{V}}\mathbf{w}_k = \mathbf{s}_k. \tag{16}$$

Rewriting this equation in terms of real and imaginary parts yields

$$\mathcal{V}\tilde{\mathbf{w}}_k = \tilde{\mathbf{s}}_k \tag{17}$$

where

$$\mathcal{V} = \begin{bmatrix} \text{real}(\mathbf{V}) & \text{imag}(\mathbf{V}) \\ -\text{imag}(\mathbf{V}) & \text{real}(\mathbf{V}) \end{bmatrix} \tag{18}$$

and $\tilde{\mathbf{w}}_k$ is defined similarly to $\tilde{\mathbf{s}}_k$. From (3), $\tilde{\mathbf{w}}_k$ will also satisfy

$$\mathcal{V}\tilde{\mathbf{w}}_k = \mathcal{U}_k\tilde{\mathbf{u}} + \tilde{\mathbf{t}}_k, \quad k = 1, \ldots, K. \tag{19}$$

Stacking all K of these equations together leads to the following general set of $2KN$ real-valued linear equations:

$$(\mathbf{I}_K \otimes \mathcal{V})\,\tilde{\mathbf{w}} = \mathcal{U}\tilde{\mathbf{u}} + \tilde{\mathbf{t}} \tag{20}$$

where

$$\tilde{\mathbf{w}} = \begin{bmatrix} \tilde{\mathbf{w}}_1 \\ \vdots \\ \tilde{\mathbf{w}}_K \end{bmatrix} \tag{21}$$

and $\mathcal{U}, \tilde{\mathbf{t}}$ are defined in (5) and (6). If (7) holds, then KN complex-valued linear equations result:

$$\left(\mathbf{I}_K \otimes \bar{\mathbf{V}}\right)\mathbf{w} = \mathbf{U}\mathbf{u} + \mathbf{t} \tag{22}$$

where $\mathbf{w}, \mathbf{U}$ and $\mathbf{t}$ result from stacking $\mathbf{w}_k, \mathbf{U}_k$ and $\mathbf{t}_k$, as in (21). While we will use the general formulation of (20) throughout the remainder of the algorithm derivations in this section, note that analogous solutions for the complex case of (22) can be derived in an identical fashion.

Equation (20) can be rewritten as

$$[(\mathbf{I}_K \otimes \mathcal{V}) \quad -\mathcal{U}] \begin{bmatrix} \tilde{\mathbf{w}} \\ \tilde{\mathbf{u}} \end{bmatrix} = \tilde{\mathbf{t}} \tag{23}$$

which, if solvable, would yield the unknown data sequence and all K zero-forcing receivers simultaneously. This observation forms the basis for the algorithms presented in what follows. Before presenting them, we discuss the identifiability of $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{u}}$ from (23). Note that in some instances, it may not be necessary to form (23) using all K of the equations from (19), due to the redundancy introduced by the space-time code. The dimension of (23) could be reduced by choosing a subset of $2 \leq K' \leq K$ equations from (19), provided that the identifiability conditions described later are met. While such an approach would, in general, be suboptimal, it does allow for an easy tradeoff between performance and computational load.

A. *Identifiability*

There are several conditions necessary for (23) to have a unique solution. The requirement that there be at least as many equations as unknowns leads to

$$\text{(C1)} \quad K(N-K) \geq N_u$$

which is typically not difficult to satisfy. As an example, for rate-one codes where $N = N_u$, $N \geq K+2$ is sufficient for all values of K. In the general case, (C1) is equivalent to

$$N \geq \frac{N_u}{K} + K$$

which implies that blind estimation via (23) is impossible for full-rate codes (although near-full-rate transmission is possible when $K \ll N$).

Since $(\mathbf{I}_K \otimes \mathcal{V})$ is full rank by construction, a necessary condition for eliminating trivial or ambiguous solutions to (23) is that

$$\text{(C2)} \quad \text{rank}(\mathcal{U}) = N_u$$

which can be guaranteed by proper code design. Additional requirements for identifiability depend on whether the blind ($\tilde{\mathbf{t}} = 0$) or semi-blind ($\tilde{\mathbf{t}} \neq 0$) case is considered.

1) Blind Case: When $\tilde{\mathbf{t}} = 0$, a "unique" solution to (23) can be obtained provided that

$$\text{(C3)} \quad \dim \text{null}\left[(\mathbf{I}_K \otimes \mathcal{V}) \quad \mathcal{U}\right] = 1.$$

The uniqueness of the solution is to within the scalar ambiguity common to all blind estimators. Condition (C3) is equivalent to

$$\dim\Big(\text{span}\,(\mathcal{Q}_1) \cap \cdots \cap \text{span}\,(\mathcal{Q}_K)\Big) = 1 \qquad (24)$$

where

$$\mathcal{Q}_i = \left(\mathcal{U}^\dagger\right)_i \mathcal{V}$$

and the pseudo-inverse $\mathcal{U}^\dagger$ is partitioned into K blocks of size $N_u \times N$:

$$\mathcal{U}^\dagger = \left(\mathcal{U}^T\mathcal{U}\right)^{-1}\mathcal{U}^T = \left[\left(\mathcal{U}^\dagger\right)_1 \cdots \left(\mathcal{U}^\dagger\right)_K\right].$$

From (2), (17), and (18), we have

$$\text{span}(\mathcal{V}) = \text{span}\left([\mathcal{S} \quad (\mathbf{J} \otimes \mathbf{I}_N)\,\mathcal{S}]\right) \qquad (25)$$

where $\mathbf{J}$ is defined in (10), so that

$$\text{span}\,(\mathcal{Q}_i) = \text{span}\Big(\Big[\left(\mathcal{U}^\dagger\right)_i \mathcal{U}_1\tilde{\mathbf{u}} \cdots \left(\mathcal{U}^\dagger\right)_i \mathcal{U}_K\tilde{\mathbf{u}} \quad \left(\mathcal{U}^\dagger\right)_i (\mathbf{J}\otimes\mathbf{I}_N)\mathcal{U}_1\tilde{\mathbf{u}} \cdots \left(\mathcal{U}^\dagger\right)_i (\mathbf{J}\otimes\mathbf{I}_N)\mathcal{U}_K\tilde{\mathbf{u}}\Big]\Big). \qquad (26)$$

Assuming that the elements of $\mathbf{u}$ are drawn from a finite-alphabet and that N_u is not too large, (C3) can be established offline by testing (24) using (26) and all possible realizations of the transmitted sequence.

For any choice of $\mathcal{U}$, there exist "bad" sequences $\tilde{\mathbf{u}}$ that will violate (C3). As an example, when $N = N_u$, choosing $\tilde{\mathbf{u}}$ to satisfy the generalized eigenvalue relationship

$$\left(\mathcal{U}^\dagger\right)_i \mathcal{U}_k\tilde{\mathbf{u}} = \lambda \left(\mathcal{U}^\dagger\right)_l \mathcal{U}_m\tilde{\mathbf{u}}$$

for any indices $i \neq l$ or $k \neq m$ will result in a nullspace of at least dimension two in (C3). Such a situation has not been observed in our extensive simulations for most of the example codes described in Section II; therefore, we postulate that the set of "bad" sequences for these codes is one of measure zero. However, one important situation where (C3) does not hold for any $\tilde{\mathbf{u}}$ is for the rate-one STBC schemes of [5]–[7]. The problem is due to a fundamental ambiguity associated with blind processing of these codes. In particular, note that for the $K = 2$ STBC, $\mathbf{AS}_2$ has exactly the same structure as $\mathbf{S}_2$ for any $\mathbf{A}$ of the form

$$\mathbf{A} = \begin{bmatrix} a_1 & -\bar{a}_2 \\ a_2 & \bar{a}_1 \end{bmatrix}.$$

Thus, it is impossible for any blind equalizer to distinguish between $\mathbf{HS}_2$ and $\mathbf{HA}^{-1}\mathbf{AS}_2$ using only the STBC structure. The ambiguity can be resolved by the insertion of pilot symbols in the data or by appropriately modifying the structure of the STBC. For example, in the approach of [11] and [13] (see Example 2 of Section II), a different linear precoder is used for even and odd symbols. Note that for STBCs with rates lower than one (e.g., such as the one considered in the simulation examples of [9]), there is no ambiguity other than the unknown scaling common to all blind algorithms.

2) Semi-Blind Case: The presence of training data can often establish the identifiability of the model without considering the influence of the unknown data $\tilde{\mathbf{u}}$. For example, if each transmit antenna broadcasts at least $T \geq K$ training symbols at sample times free from the influence of $\tilde{\mathbf{u}}$ and if $[\tilde{\mathbf{t}}_1 \cdots \tilde{\mathbf{t}}_K]$ is full rank, then $\tilde{\mathbf{w}}$ can be solved for explicitly, and $\tilde{\mathbf{u}}$ is clearly identifiable. While this is perhaps the most common situation, identifiability is possible under much weaker conditions. In the general case, the following condition similar to (C3) must hold when $\tilde{\mathbf{t}} \neq 0$ for the existence of a unique solution to (23):

$$\text{(C4)} \quad \dim \text{null}\left[(\mathbf{I}_K \otimes \mathcal{V}) \quad \mathcal{U} \quad \tilde{\mathbf{t}}\right] = 1.$$

Clearly, (C4) requires

$$\mathbf{P}_{\mathcal{U}}^{\perp}\tilde{\mathbf{t}} \neq 0 \qquad (27)$$

as well as

$$\mathbf{P}_{I_K\otimes\mathcal{V}}^{\perp}\tilde{\mathbf{t}} \neq 0 \qquad (28)$$

which, due to (25), is equivalent to stating that

$$\tilde{\mathbf{t}}_k \notin \text{span}\Big(\Big[\mathcal{U}_1\tilde{\mathbf{u}} + \tilde{\mathbf{t}}_1 \cdots \mathcal{U}_K\tilde{\mathbf{u}} + \tilde{\mathbf{t}}_K \quad (\mathbf{J}\otimes\mathbf{I}_N)\left(\mathcal{U}_1\tilde{\mathbf{u}} + \tilde{\mathbf{t}}_1\right) \cdots (\mathbf{J}\otimes\mathbf{I}_N)\left(\mathcal{U}_K\tilde{\mathbf{u}} + \tilde{\mathbf{t}}_K\right)\Big]\Big) \qquad (29)$$

for at least one $k \in \{1, \ldots, K\}$. In the previous expressions, $\mathbf{P}_A^{\perp} = \mathbf{I} - \mathbf{AA}^\dagger$.

Equation (27) will be satisfied when the influence of the training and the unknown data sequence are temporally disjoint. For example, if the first T samples transmitted from each antenna are training, then for each k

$$\tilde{\mathbf{t}}_k = \begin{bmatrix} \text{real}\,(t_{k,1}) \\ \vdots \\ \text{real}\,(t_{k,T}) \\ \mathbf{0}_{(N-T)\times 1} \\ \text{imag}\,(t_{k,1}) \\ \vdots \\ \text{imag}\,(t_{k,T}) \\ \mathbf{0}_{(N-T)\times 1} \end{bmatrix} \qquad \mathcal{U}_k = \begin{bmatrix} \mathbf{0}_{T\times 2N_u} \\ \mathcal{U}_{k,1} \\ \mathbf{0}_{T\times 2N_u} \\ \mathcal{U}_{k,2} \end{bmatrix} \qquad (30)$$

for some $(N-T)\times 2N_u$ matrices $\mathcal{U}_{k,1},\mathcal{U}_{k,2}$, where $t_{k,i}$ denotes the ith training symbol from antenna k, and $\mathbf{0}_{Y\times Z}$ indicates a $Y\times Z$ matrix of zeros. In such a case, $\tilde{\mathbf{t}}\perp$ span$(\mathcal{U})$, which implies (27).

B. Least-Squares Algorithms

Under the assumption of an identifiable model, the following least-squares problem can be solved for estimates of $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{u}}$ in the presence of noise:

$$\hat{\tilde{\mathbf{w}}},\hat{\tilde{\mathbf{u}}}=\arg\min_{\tilde{\mathbf{w}},\tilde{\mathbf{u}}}\left\|\left[(\mathbf{I}_K\otimes\mathcal{V})\quad -\mathcal{U}\right]\begin{bmatrix}\tilde{\mathbf{w}}\\ \tilde{\mathbf{u}}\end{bmatrix}-\tilde{\mathbf{t}}\right\|^2. \tag{31}$$

The solution to (31) can be found in several different, although equivalent ways.

1) Blind Case:

a) Solve for $\tilde{\mathbf{w}}$ and $\tilde{\mathbf{u}}$ simultaneously from the right singular vector of $[(\mathbf{I}_K\otimes\mathcal{V})\quad -\mathcal{U}]$ with smallest singular value.

b) Set

$$\hat{\tilde{\mathbf{w}}}=\begin{bmatrix}\mathcal{V}^T\mathcal{U}_1\\ \vdots\\ \mathcal{V}^T\mathcal{U}_K\end{bmatrix}\hat{\tilde{\mathbf{u}}} \tag{32}$$

where $\hat{\tilde{\mathbf{u}}}$ is found from the right singular vector of

$$\begin{bmatrix}\mathbf{P}_{\mathcal{V}}^{\perp}\mathcal{U}_1\\ \vdots\\ \mathbf{P}_{\mathcal{V}}^{\perp}\mathcal{U}_K\end{bmatrix} \tag{33}$$

with smallest singular value.

c) Set

$$\hat{\tilde{\mathbf{u}}}=\mathcal{U}^{\dagger}\left(\mathbf{I}_K\otimes\mathcal{V}\right)\hat{\tilde{\mathbf{w}}} \tag{34}$$

where $\hat{\tilde{\mathbf{w}}}$ is found from the right singular vector of

$$\mathbf{P}_{\mathcal{U}}^{\perp}\left(\mathbf{I}_K\otimes\mathcal{V}\right) \tag{35}$$

with smallest singular value.

2) Semi-Blind Case:

a) Solve for $\hat{\tilde{\mathbf{w}}}$ and $\hat{\tilde{\mathbf{u}}}$ simultaneously using

$$\begin{bmatrix}\hat{\tilde{\mathbf{w}}}\\ \hat{\tilde{\mathbf{u}}}\end{bmatrix}=\left[(\mathbf{I}_K\otimes\mathcal{V})\quad -\mathcal{U}\right]^{\dagger}\tilde{\mathbf{t}}. \tag{36}$$

b) Solve for $\hat{\tilde{\mathbf{w}}}$ in terms of $\hat{\tilde{\mathbf{u}}}$

$$\hat{\tilde{\mathbf{w}}}=\begin{bmatrix}\mathcal{V}^T\mathcal{U}_1\\ \vdots\\ \mathcal{V}^T\mathcal{U}_K\end{bmatrix}\hat{\tilde{\mathbf{u}}}+\begin{bmatrix}\mathcal{V}^T\tilde{\mathbf{t}}_1\\ \vdots\\ \mathcal{V}^T\tilde{\mathbf{t}}_K\end{bmatrix} \tag{37}$$

$$\begin{aligned}\hat{\tilde{\mathbf{u}}}&=-\left[\left(\mathbf{I}_K\otimes\mathbf{P}_{\mathcal{V}}^{\perp}\right)\mathcal{U}\right]^{\dagger}\left(\mathbf{I}_K\otimes\mathbf{P}_{\mathcal{V}}^{\perp}\right)\tilde{\mathbf{t}}\\ &=-\sum_{k=1}^{K}\left(\mathcal{U}_k^*\mathbf{P}_{\mathcal{V}}^{\perp}\mathcal{U}_k\right)^{-1}\mathcal{U}_k^*\mathbf{P}_{\mathcal{V}}^{\perp}\tilde{\mathbf{t}}_k.\end{aligned} \tag{38}$$

c) Solve for $\hat{\tilde{\mathbf{u}}}$ in terms of $\hat{\tilde{\mathbf{w}}}$

$$\hat{\tilde{\mathbf{u}}}=\mathcal{U}^{\dagger}\left(\left(\mathbf{I}_K\otimes\mathcal{V}\right)\hat{\tilde{\mathbf{w}}}-\tilde{\mathbf{t}}\right) \tag{39}$$

$$\hat{\tilde{\mathbf{w}}}=\left[\mathbf{P}_{\mathcal{U}}^{\perp}\left(\mathbf{I}_K\otimes\mathcal{V}\right)\right]^{\dagger}\mathbf{P}_{\mathcal{U}}^{\perp}\tilde{\mathbf{t}}. \tag{40}$$

While mathematically equivalent, one of the three approaches for each case may have a slight numerical or computational advantage over the others, depending on the values of K, N, and N_u.

C. Data Direct Methods

From (14) and (16), we see that

$$\bar{\mathbf{V}}\mathbf{w}_k=\mathbf{X}^T\bar{\mathbf{F}}\mathbf{\Lambda}^{-1}\mathbf{w}_k \tag{41}$$

which leads to an equation analogous to (17)

$$\mathcal{X}\tilde{\mathbf{w}}_{x,k}=\tilde{\mathbf{s}}_k=\mathcal{U}_k\tilde{\mathbf{u}}+\tilde{\mathbf{t}}_k \tag{42}$$

where

$$\mathcal{X}=\begin{bmatrix}\text{real}\left(\mathbf{X}^T\right) & -\text{imag}\left(\mathbf{X}^T\right)\\ \text{imag}\left(\mathbf{X}^T\right) & \text{real}\left(\mathbf{X}^T\right)\end{bmatrix} \tag{43}$$

$$\tilde{\mathbf{w}}_{x,k}=\begin{bmatrix}\text{real}\left(\mathbf{w}_{x,k}\right)\\ \text{imag}\left(\mathbf{w}_{x,k}\right)\end{bmatrix} \tag{44}$$

$$\mathbf{w}_{x,k}=\bar{\mathbf{F}}\mathbf{\Lambda}^{-1}\mathbf{w}_k. \tag{45}$$

Stacking all K equations like (43) yields

$$\left[(\mathbf{I}_K\otimes\mathcal{X})\quad -\mathcal{U}\right]\begin{bmatrix}\tilde{\mathbf{w}}_x\\ \tilde{\mathbf{u}}\end{bmatrix}=\tilde{\mathbf{t}} \tag{46}$$

where

$$\tilde{\mathbf{w}}_x=\begin{bmatrix}\tilde{\mathbf{w}}_{x,1}\\ \vdots\\ \tilde{\mathbf{w}}_{x,K}\end{bmatrix}.$$

Equation (46) is equivalent to (23), except that $\mathcal{V}$ is now replaced by $\mathcal{X}$. Instead of solving (23), we could therefore solve (46) directly for $\tilde{\mathbf{w}}_x$ and $\tilde{\mathbf{u}}$. This would alleviate the computation of an SVD of $\mathbf{X}$ and, hence, reduce the computational complexity. Note that in contrast to the zero-forcing equalizer $\tilde{\mathbf{w}}$, which has no degrees of freedom, the zero-forcing equalizer $\tilde{\mathbf{w}}_x$ has $K(M-K)$ degrees of freedom (the dimension of the right null space of $\mathbf{I}_K\otimes\mathcal{X}$). Hence, when solving (46), $\tilde{\mathbf{w}}_x$ should be prevented from lying in the right null space of $\mathbf{I}_K\otimes\mathcal{X}$.

In the blind case, we need to solve (46) under a specific constraint in order to avoid the trivial solution. This constraint should be chosen such that $\tilde{\mathbf{w}}_x$ cannot lie in the right null space of $\mathbf{I}_K\otimes\mathcal{X}$. Putting a unit constraint on $[\tilde{\mathbf{w}}_x^T\ \tilde{\mathbf{u}}^T]^T$ or $\tilde{\mathbf{w}}_x$ [cf. the blind methods (a) and (c)] clearly does not prevent this and generally leads to poor performance. On the other hand, a unit norm constraint on $\tilde{\mathbf{u}}$ [cf. the blind method (b)] does prevent $\tilde{\mathbf{w}}_x$ from lying in the right null space of $\mathbf{I}_K\otimes\mathcal{X}$. However, the blind method related to this constraint requires the computation of $\mathbf{X}^{\dagger}$, which has complexity comparable to the computation of the SVD of $\mathbf{X}$. Another appropriate constraint is the unit output energy constraint [29], i.e., $\|(\mathbf{I}_K\otimes\mathcal{X})\tilde{\mathbf{w}}_x\|^2=1$. The algorithm corresponding to this constraint can be shown to be equivalent to blind method (c), where $\bar{\mathbf{V}}$ is replaced by the $N\times M$ matrix $\mathbf{Q}$ obtained by computing the QR decomposition of $\mathbf{X}^T$: $\mathbf{X}^T=\mathbf{QR}$.

In the semi-blind case, there is no problem since the nonzero training sequence $\tilde{\mathbf{t}}\notin$ span$\{\mathcal{U}\}$ prevents $\tilde{\mathbf{w}}_x$ from lying in the right null space of $\mathbf{I}_K\otimes\mathcal{X}$ and methods similar to the semi-blind methods (a)–(c) can be applied.

D. Processing Subsequent Data Blocks

An important implementational issue is the redundancy of solving for both $\hat{\tilde{\mathbf{w}}}$ and $\hat{\tilde{\mathbf{u}}}$ rather than just $\hat{\tilde{\mathbf{u}}}$ directly. One advantage of estimating the receiver weights $\tilde{\mathbf{w}}$ or $\tilde{\mathbf{w}}_x$ along with $\tilde{\mathbf{u}}$ is that the weights can be used to process subsequent data blocks with minimal additional computation, provided that the channel is stationary. The first block of data would be used to find an estimate of $\tilde{\mathbf{w}}_x$ by either direct estimation using the results of Section III-C or by transforming the subspace-based estimate $\hat{\tilde{\mathbf{w}}}$ using (45). Estimates of the transmitted data $\tilde{\mathbf{u}}'$ for subsequent blocks of data $\mathcal{X}'$ are then found using the weights obtained from the first block

$$\hat{\tilde{\mathbf{u}}}' = \mathcal{U}^\dagger \left[(\mathbf{I}_K \otimes \mathcal{X}') \hat{\tilde{\mathbf{w}}}_x - \tilde{\mathbf{t}}' \right]. \tag{47}$$

E. Processing Real Symbols

When the transmitted symbols $\mathbf{S}_K$ are purely real, it is standard in multichannel problems to split the data into real and imaginary parts as

$$\hat{\hat{\mathbf{X}}} = \begin{bmatrix} \text{real}\left(\hat{\mathbf{X}}\right) \\ \text{imag}\left(\hat{\mathbf{X}}\right) \end{bmatrix} = \begin{bmatrix} \text{real}(\mathbf{H}) \\ \text{imag}(\mathbf{H}) \end{bmatrix} \mathbf{S}_K + \begin{bmatrix} \text{real}(\mathbf{N}) \\ \text{imag}(\mathbf{N}) \end{bmatrix}.$$

Working with $\hat{\hat{\mathbf{X}}}$ instead of $\hat{\mathbf{X}}$ effectively doubles the number of available receive channels and allows replacement of the condition $M \geq K$ with $2M \geq K$. Note that in this case, the simplified model of (8) should be used in describing the transmitted data.

IV. Extensions of the Algorithm

This section considers extensions of the above algorithms to cases involving multiple users, more transmit than receive antennas, a rank-deficient channel, or frequency-selective fading. In most cases, the approach taken is to modify the resulting data models so that they are isomorphic to the basic case considered in Section II. Once this is done, the algorithms of Section III-B can more or less be directly applied.

A. Multiple Users

If d symbol-synchronous users are present, then (1) becomes

$$\hat{\mathbf{X}} = \mathbf{\Pi}\mathbf{\Sigma}_K + \mathbf{N} \tag{48}$$

where

$$\mathbf{\Pi} = [\mathbf{H}_1 \cdots \mathbf{H}_d] \quad M \times K$$
$$\mathbf{\Sigma}_K = \begin{bmatrix} \mathbf{S}_{K_1,1} \\ \vdots \\ \mathbf{S}_{K_d,d} \end{bmatrix} \quad K \times N$$
$$K = \sum_{i=1}^{d} K_i$$

and where $\mathbf{S}_{K_i,i}$, $\mathbf{H}_i$, and K_i represent, respectively, the transmitted signals, channel, and number of transmit antennas for the ith user. Assume that each user employs space-time coding in the form of (3), where $\mathcal{U}^{(i)}$ represents the $2K_iN \times 2N_{u,i}$ matrix containing all K_i precoders for user i as in (5) and with $\tilde{\mathbf{u}}^{(i)}, \tilde{\mathbf{t}}^{(i)}$ representing, respectively, the $2N_{u,i} \times 1$ data sequence and $2K_iN \times 1$ training data vectors. Note that data obeying (48) could also be generated by a single user whose transmit antennas are divided into d groups, with each group transmitting a different data sequence. Such an approach could be used to trade off diversity for throughput.

When $M \geq K$, two separate situations must be considered for the case of blind estimation (if $M < K$, methods similar to those described in Section IV-B must be used).

- *Users with unique STCs*: Data from users that employ unique linear precoders $\mathcal{U}^{(i)}$ can typically be recovered with no modification to the algorithms described previously. The term "unique" here denotes that the K_i signals transmitted by user i should not be able to be generated using any K_i of the linear precoders from other users with K_i or more transmit antennas. A necessary condition for this to hold is that for all $l \neq i$, where $K_l \geq K_i$
$$[\mathcal{U}^{(i)} \quad (\mathcal{P} \otimes \mathbf{I}_{2N})\mathcal{U}^{(l)}]$$
must be full column rank for all $K_i \times K_l$ permutation matrices $\mathcal{P}$ (each row of $\mathcal{P}$ has a single one and zeroes elsewhere).
- *Users with nonunique STCs*. If, say, d' users share the same set of precoders $\mathcal{U}$, then (23) formed from $\mathcal{U}$ will have a nullspace of dimension d'. This nullspace will be spanned by the unknown information sequences of the d' users: $[\tilde{\mathbf{u}}^{(1)} \cdots \tilde{\mathbf{u}}^{(d')}]$. Additional information about the signals would be required to separate the individual user's symbol sequences in a second step, using, for example, the assumption of constant modulus [30] or finite alphabet signals [31].

If each user transmits linearly independent training data, then the semi-blind approach of Section III-B can be used directly, whether or not the users' codes are unique. This is done by simply rewriting (36) so that it is specific to user i:

$$\begin{bmatrix} \hat{\tilde{\mathbf{w}}}^{(i)} \\ \hat{\tilde{\mathbf{u}}}^{(i)} \end{bmatrix} = \left[(\mathbf{I}_{K_i} \otimes \mathcal{V}) \quad -\mathcal{U}^{(i)} \right]^\dagger \tilde{\mathbf{t}}^{(i)}. \tag{49}$$

B. More Transmit Than Receive Antennas

In certain cases, a slight modification to the data will allow the algorithms of Sections III-B and C to be applied in situations where $M < K$. In particular, assume that (7) holds and that $\mathbf{U}_k = \mathbf{U}^{k-1}$ for $k = 2, \ldots, K$ and some square ($N_u = N$) full-rank matrix $\mathbf{U}$. This constraint is satisfied by the circulant code of [18]. It can easily be implemented with the diagonal precoders of [14] and [16] and it trivially applies to all codes in the form of (7) with $K = 2$ and $\mathbf{U}_1 = \mathbf{I}_N$. It does not hold for orthogonal, full-diversity STBCs like the Alamouti code and its derivatives. Under these assumptions, $\mathbf{s}_k = \mathbf{U}^{k-1}\mathbf{s}_1$, and in the single-user case

$$\mathbf{X}\mathbf{U}^T = \mathbf{H}\mathbf{S}_K\mathbf{U}^T \tag{50}$$
$$= [\mathbf{0} \quad \mathbf{H}] \begin{bmatrix} \mathbf{S}_K \\ \mathbf{s}_1^T \left(\mathbf{U}^T\right)^K \end{bmatrix} \tag{51}$$
$$= [\mathbf{0} \quad \mathbf{H}] \mathbf{S}_{K+1} \tag{52}$$

where $\mathbf{0}$ is an $M \times 1$ vector of zeros. Since

$$\mathbf{H}\mathbf{S}_K = [\mathbf{H} \quad \mathbf{0}] \mathbf{S}_{K+1}$$

a stacking operation leads to

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{X}\mathbf{U}^T \end{bmatrix} = \mathbf{H}_2 \mathbf{S}_{K+1} \tag{53}$$

where $\mathbf{H}_2$ is a $2M \times (K+1)$ block Sylvester matrix identical in form to those obtained in single-input multiple-output blind equalization problems (e.g., see [32]). Thus, the same data model is obtained as before, except the row dimension (number of effective receive antennas) has been doubled, whereas the column dimension (number of effective transmit antennas) has increased by only one. Stacking $P-1$ times and adding the effects of noise leads to

$$\hat{\mathbf{X}}_P \overset{\text{def}}{=} \begin{bmatrix} \hat{\mathbf{X}} \\ \hat{\mathbf{X}}\mathbf{U}^T \\ \vdots \\ \hat{\mathbf{X}}\left(\mathbf{U}^T\right)^{P-1} \end{bmatrix} = \mathbf{H}_P \mathbf{S}_{K+P-1} + \mathbf{N}_P \tag{54}$$

where $\mathbf{N}_P$ is formed identically to $\hat{\mathbf{X}}_P$. The new "channel" matrix $\mathbf{H}_P$ is guaranteed to be full rank as long as $\mathbf{H}$ is and will have at least as many rows as columns, provided that

$$P \geq \frac{K-1}{M-1}.$$

The single-user algorithms of Section III-B can then be directly applied to $\hat{\mathbf{X}}_P$ rather than $\hat{\mathbf{X}}$. A larger value of P is typically required in cases involving multiple users. Note that in general, the stacking operation will lead to a noise term $\mathbf{N}_P$ that is neither temporally nor spatially white, even if $\mathbf{N}$ was. This can be accounted for, however, by prewhitening in both space and time.

C. Rank-Deficient Channels

If the channel is rank deficient, i.e., $\text{rank}(\mathbf{H}) = \rho < \min\{M, K\}$, then in the noiseless case, $\text{rank}(\mathbf{X}) = \rho$. If the SVD of $\mathbf{X}$ in (14) is partitioned so that $\mathbf{V}$ contains the first ρ right singular vectors, then (15) still holds, except that $\mathbf{H}_v$ will be $\rho \times K$. The transpose of (15) will then be equivalent to a noiseless version of the original model (1) for a case with more transmit than receive antennas (i.e., $\mathbf{H}_v$ is fat), and the approach of Section IV-B can be used. Instead of (54), the algorithms are applied to the matrix

$$\hat{\mathbf{X}}_P \overset{\text{def}}{=} \begin{bmatrix} \mathbf{V}^* \\ \mathbf{V}^*\mathbf{U}^T \\ \vdots \\ \mathbf{V}^*\left(\mathbf{U}^T\right)^{P-1} \end{bmatrix} \tag{55}$$

where P is chosen to satisfy

$$P \geq \frac{K-1}{\rho-1}.$$

D. Frequency-Selective Fading

In general, the technique presented here requires $M > K$. This requirement can be relaxed for certain types of codes (e.g., see [18] for an example) but not in the general case. Assume the MIMO channel can be represented as an L-tap FIR filter

$$\mathbf{H}(0) + \mathbf{H}(1)z^{-1} + \cdots + \mathbf{H}(L-1)z^{-(L-1)}$$

where each matrix tap $\mathbf{H}(i)$ is $M \times K$. Let $\mathbf{S}_K(t)$ represent the $K \times N'$ matrix containing the symbols broadcast from the K transmit antennas at symbol times $t, t+1, \ldots, t+N'-1$, where the value of N' will be specified below. Then, the $M \times N'$ matrix of received data at time t can be represented as

$$\hat{\mathbf{X}}(t) = [\mathbf{H}(0)\cdots\mathbf{H}(L-1)] \begin{bmatrix} \mathbf{S}_K(t) \\ \vdots \\ \mathbf{S}_K(t-L+1) \end{bmatrix} + \mathbf{N}(t) \tag{56}$$

$$= \mathbf{H}\mathbf{S}_{KL}(t) + \mathbf{N}(t). \tag{57}$$

As before, the subscript on $\mathbf{S}_{KL}(t)$ indicates that the matrix contains KL rows. Since it is likely that $M < KL$, several delayed versions of $\hat{\mathbf{X}}(t)$ can be stacked to create a low-rank model

$$\hat{\mathbf{X}}_P \overset{\text{def}}{=} \begin{bmatrix} \hat{\mathbf{X}}(t) \\ \vdots \\ \hat{\mathbf{X}}(t-P+1) \end{bmatrix} = \mathcal{H}\mathbf{S}_{K(L+P-1)} + \mathcal{N} \tag{58}$$

where $\mathcal{H}$ is an $MP \times K(L+P-1)$ block Sylvester matrix defined as in Section IV-B, and

$$\mathbf{S}_{K(L+P-1)} = \begin{bmatrix} \mathbf{S}_K(t) \\ \vdots \\ \mathbf{S}_K(t-L-P+2) \end{bmatrix}. \tag{59}$$

The stacking factor P is chosen so that $\mathcal{H}$ is tall, which requires

$$P \geq \frac{K(L-1)}{M-K}. \tag{60}$$

As with standard blind equalization problems, a low-rank model results because the intersection of the rowspace of successive delayed versions of the data $\hat{\mathbf{X}}(t-i)$ and $\hat{\mathbf{X}}(t-i-1)$ has dimension $K(L-1)$.

The key observation here is that (58) is essentially identical to (1), except that the temporal diversity of the channel and the data stacking have spread the STC structure over $L+P-1$ time shifts. To be more precise, let $N' = N+L+P-2$, and assume that the first encoded data sample in $\mathbf{S}_K(t)$ occurs at time t. In other words, $\mathbf{S}_K(t-i)$ contains i and $L+P-2-i$ samples from the previous and next N-sample blocks of transmitted data, respectively. Let $\tilde{\mathbf{w}}_{k,i}$ be the zero-forcing equalizer associated with the kth row of $\mathbf{S}_K(t-i)$, and let $\mathcal{I}_i$ be the set of indices shown in the equation at the bottom of the page. Then

$$\mathcal{V}(\mathcal{I}_i,:)\,\tilde{\mathbf{w}}_{k,i} = \mathcal{U}_k\tilde{\mathbf{u}} + \tilde{\mathbf{t}}_k \qquad k = 1,\ldots,K,\ i = 0,\ldots,L+P-2 \tag{61}$$

where $\mathcal{V}(\mathcal{I}_i,:)$ denotes the matrix formed from $\mathcal{V}$ using only the rows in $\mathcal{I}_i$. The algorithms of Section III-B can be directly applied to estimate $\tilde{\mathbf{w}}_{k,i}$ and $\tilde{\mathbf{u}}$, except that instead of having

$$\mathcal{I}_i = [\,i+1 \quad i+2 \quad \cdots \quad N+i \mid N'+i+1 \quad N'+i+2 \quad \cdots \quad N'+N+i\,].$$

$2KN$ equations with $2(K^2 + N_u)$ unknowns as in (23), there are now $2KN(L+P-1)$ equations with $2(K^2(L+P-1)+N_u)$ unknowns

$$\begin{bmatrix} \mathbf{I}_K \otimes \mathcal{V}(\mathcal{I}_0,:) & & \\ & \ddots & \\ & & \mathbf{I}_K \otimes \mathcal{V}(\mathcal{I}_{L+P-2},:) \end{bmatrix} \times \begin{bmatrix} \tilde{\mathbf{w}}_{1,0} \\ \vdots \\ \tilde{\mathbf{w}}_{K,0} \\ \tilde{\mathbf{w}}_{1,1} \\ \vdots \\ \tilde{\mathbf{w}}_{K,L+P-2} \end{bmatrix} = \begin{bmatrix} \mathcal{U} \\ \vdots \\ \mathcal{U} \end{bmatrix} \tilde{\mathbf{u}} + \begin{bmatrix} \tilde{\mathbf{t}} \\ \vdots \\ \tilde{\mathbf{t}} \end{bmatrix}. \quad (62)$$

The extra equations provided by the channel's temporal diversity relax identifiability condition (C1) to be $K(L+P-1)(N-K) \geq N_u$.

V. Simulation Examples

The first example considers a case with one user, $M = 3$ receive antennas, $K = 2$ transmit antennas, a block of $N_u = 40$ transmitted data symbols, and a variable number T of training symbols ($N = N_u + T$). Two different space-time coding strategies were implemented: 1) the circulant code of [18] and 2) the Alamouti code [5]. Unit-amplitude QPSK symbols were generated for both the training and unknown data, and the elements of the channel and noise matrices were zero-mean, circular complex Gaussian random variables, with variances chosen to achieve the desired SNR. In the plots shown for this and other examples, the SNR is defined by σ_h^2/σ_n^2, where σ_h^2 and σ_n^2 are the variances of the elements of $\mathbf{H}$ and $\mathbf{N}$, respectively. In each trial, a new random $\mathbf{H}$, $\mathbf{N}$, $\mathbf{t}$, and $\mathbf{u}$ are generated and used to create observations for both of the above codes side-by-side. The training data was always placed at the beginning of each block of data, as in (30). Both the subspace-based semi-blind algorithm of Section III-B and the direct-data semi-blind approach of Section III-C [corresponding to (46)] were implemented. In addition, when $T \geq 3$, the training data was also used by itself to estimate the channel $\hat{\mathbf{H}}$ and, in turn, a set of zero-forcing equalizers $\hat{\mathbf{H}}^\dagger$. Estimates of $\mathbf{u}$ were then obtained by substituting these equalizers into the following equation, which was derived using notation similar to that in (41)–(47):

$$\hat{\hat{\mathbf{u}}} = \mathcal{U}^\dagger \left[(\mathbf{I}_K \otimes \mathcal{X}) \hat{\hat{\mathbf{w}}}_t - \tilde{\mathbf{t}} \right]$$

where $\hat{\mathbf{w}}_t = \text{vec}[(\hat{\mathbf{H}}^\dagger)^T]$ and $\text{vec}(\cdot)$ is the column stacking operator.

Fig. 1 shows the symbol error rate (SER) achieved by the semi-blind algorithms presented in this paper, together with the performance obtained using training alone. The notation "$w/\mathbf{V}$" and "$w/\mathbf{X}$" in the legend indicates, respectively, whether the subspace or direct-data algorithm was used. The subspace algorithm achieves an SER that is about 10–50% lower than that of the direct-data approach for both codes with better relative performance at higher SNRs. The algorithms perform better for the circulant code than for the Alamouti code, especially when the number of training symbols is small, since in the limit where $T = 0$, the structure of the Alamouti code is not identifiable. However, the structure of the code still provides some information, as evidenced by the fact that an SER is achieved that is 1.5–4 times smaller than when using the training data alone.

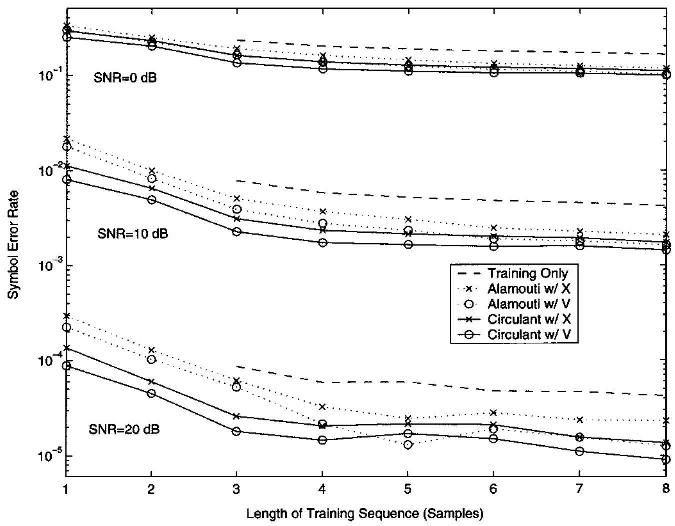


Fig. 1. Semi-blind symbol error rate versus training interval T for Alamouti and circulant space-time coding schemes.

The second example involves two users, each with two transmit antennas and a variable number of receive antennas. The noise, data, and training symbols were generated as above with $N_u = 40$ and $T = 5$ (hence, $N = 45$), and both users employed the same zero-padded diagonal linear precoders defined by

$$\mathbf{U}_1 = \begin{bmatrix} \mathbf{0}_{5\times 40} \\ \mathbf{I}_{40} \end{bmatrix}, \quad \mathbf{U}_2 = \begin{bmatrix} \mathbf{0}_{5\times 40} \\ \mathbf{D} \end{bmatrix}$$

where $\mathbf{D}$ is a 40×40 diagonal matrix with nonzero entries drawn at random from the unit circle. The signals from the two users can still be separated in this case since each transmitted linearly independent training data. Subspace-based stimates of $\mathbf{u}$ were obtained both with unstacked data and with data stacked once ($P = 2$), as described in Section IV-B. The SER results are plotted in Fig. 2 versus SNR. No result is shown for $M = 3$ and $P = 1$ since a low-rank model is not available in this case. While $P > 1$ is not required for $M > 3$, stacking provides a significant performance advantage; stacking once has roughly the same effect as adding a receive antenna, resulting in an order of magnitude improvement in SER.

In the final example, a single user with two antennas transmits over an $L = 2$ tap frequency-selective fading channel using a diagonal linear precoder with $\mathbf{U}_1 = \mathbf{I}_{40}$ and $\mathbf{U}_2 = \mathbf{D}$, where $\mathbf{D}$ is a 40×40 diagonal matrix with nonzero entries drawn at random from the unit circle. Data was collected by an $M = 3$ element receive array and stacked $P = 3$ times to create a 9×43 data matrix $\hat{\mathbf{X}}_3$, with a signal subspace of dimension eight. The 3×2 matrix taps had elements of equal average power and were generated as in the previous two examples. No training data was assumed to be present; therefore, $\mathbf{u}$ was estimated using the blind

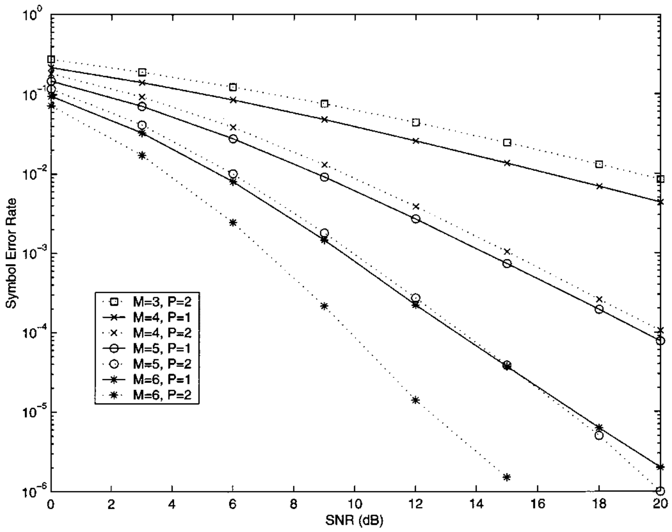


Fig. 2. Semi-blind symbol error rate versus SNR for diagonal linear precoding and various M and P.

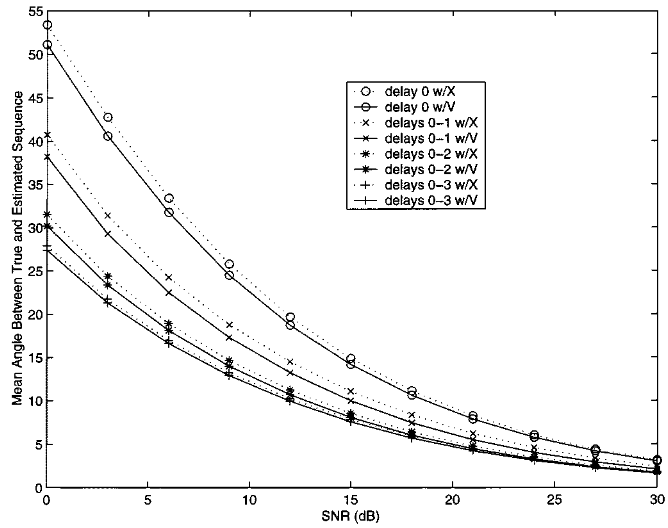


Fig. 3. Mean blind decoding performance versus SNR for diagonal linear precoding in a frequency selective fading channel.

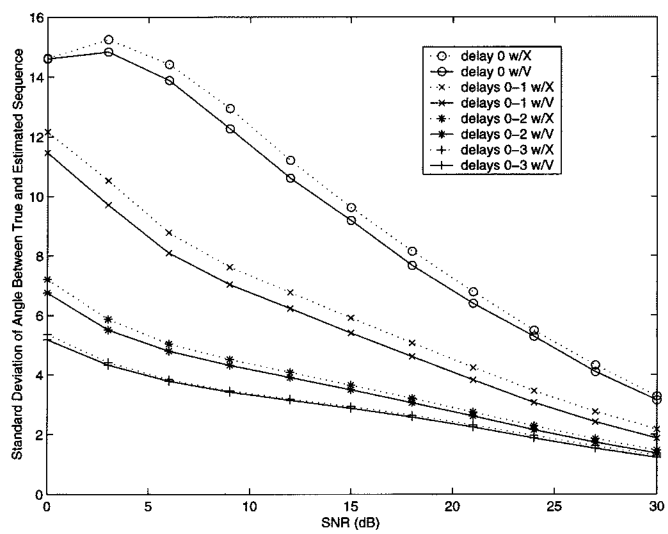


Fig. 4. Blind decoding standard deviations versus SNR for diagonal linear precoding in a frequency-selective fading channel.

subspace algorithm of Section III-B and the blind data-direct algorithm (with output normalization) of Section III-C applied to the model in (62). Equation (62) exploits all of the STC structure due to the temporal diversity added by stacking and the memory of the channel. However, $\mathbf{u}$ is identifiable using far fewer equations. To examine the tradeoff between computation and performance that results from taking only a subset of the equations represented by (62), four different estimates of $\mathbf{u}$ were obtained using both $\mathcal{V}$ and $\mathcal{X}$. The first estimate was calculated from the right singular vector of $[\mathbf{I}_K \otimes \mathcal{V}(\mathcal{I}_0,:) \quad \mathcal{U}]$ with the smallest singular value. This estimate only makes use of the structure present in the first $N_u = 40$ samples of the data block (those that correspond to delay zero) and ignores the information present due to the temporal diversity. The second estimate was calculated in the same way using the matrix

$$\begin{bmatrix} \mathbf{I}_K \otimes \mathcal{V}(\mathcal{I}_0,:) & & \mathcal{U} \\ & \mathbf{I}_K \otimes \mathcal{V}(\mathcal{I}_1,:) & \mathcal{U} \end{bmatrix}$$

and, thus, uses not only the zero-delay structure of samples 1–40 but that of the first delay present at samples 2–41 as well. The final two estimates build on these by exploiting delays 0–2 and 0–3, respectively. Figs. 3 and 4 plot, respectively, the mean and standard deviation of the angle (in degrees) between each of these four estimates and the true $\mathbf{u}$ as a function of SNR for both the subspace and direct-data algorithms. The angle between $\mathbf{u}$ and an estimate $\hat{\mathbf{u}}$ is defined to be

$$\theta(\mathbf{u}, \hat{\mathbf{u}}) = \cos^{-1}\left(\frac{|\mathbf{u}^*\hat{\mathbf{u}}|}{||\mathbf{u}||\,||\hat{\mathbf{u}}||}\right).$$

At low SNR, there is a large improvement associated with using all of the available temporal diversity, but this advantage decreases as the SNR increases. Exploiting the information from an additional delayed block of data provides about a 3-dB performance gain, except when going from delays 0–2 to 0–3, where the gain is only about 1 dB.

VI. CONCLUSIONS

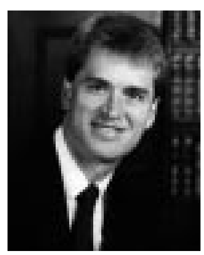
A general framework for space-time block coding has been presented and shown to include a number of recently proposed codes or code families as special cases. All codes within this framework employ linear precoders (or affine precoding when training data is present) and were referred to as generalized space-time block codes (GSTBCs). In the noiseless case, the redundant structure of GSTBCs allows for construction of a set of channel-independent linear equations whose solution, if it exists, simultaneously yields the transmitted data sequence and a vector containing all possible zero-forcing receivers. Conditions under which a unique solution exists were discussed, and least-squares blind and semi-blind algorithms were proposed for finding estimates with noisy data. While the algorithms were presented for the single-user flat-fading case, where the channel is full rank and there are more receive than transmit antennas, extensions to scenarios involving multiple users, more transmit than receive antennas, rank-deficient channels, and frequency-selective fading were presented. Several simulation studies were used to illustrate the performance of both the basic algorithm and some of its extensions.

ACKNOWLEDGMENT

The authors wish to thank S. Zhou of the University of Minnesota and the reviewers of the paper for their helpful suggestions.

REFERENCES

[1] I. Telatar, "Capacity of multiple antenna Gaussian channels," *AT&T Tech. Memo.*, June 1995.
[2] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
[3] B. L. Hughes, "Differential space-time modulation," in *Proc. IEEE Wireless Commun. Networking Conf.*, New Orleans, LA, Sept. 1999.
[4] B. Hochwald and W. Sweldens, "Differential unitary space-time modulation," Bell Labs., Lucent Technol., Tech. Rep., 1999.
[5] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Aug. 1998.
[6] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, Oct. 1999.
[7] G. Ganesan and P. Stoica, "Space-time diversity," in *Signal Processing Advances in Wireless and Mobile Communications*, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Englewood Cliffs, NJ: Prentice-Hall, 2001, vol. 2, ch. 2, pp. 59–88.
[8] C. Budianu and L. Tong, "Channel estimation for space-time orthogonal block codes," in *Proc. Int. Conf. Commun.*, Helsinki, Finland, June 2001.
[9] P. Stoica and G. Ganesan, "Space-time block codes: Trained, blind and semi-blind detection," *IEEE Trans. Signal Processing*, 2001, submitted for publication.
[10] Z. Liu, G. Giannakis, A. Scaglione, and S. Barbarossa, "Decoding and equalization of unknown multipath channels based on block precoding and transmit-antenna diversity," in *Proc. 33rd Asilomar Conf. Signals, Syst., Comput.*, vol. 2, 1999, pp. 1557–1561.
[11] Z. Liu, G. Giannakis, B. Muquet, S. Zhou, and M. Computing, "Space-time coding for broadband wireless communications," *Wireless Commun. Mobile Comput.*, vol. 1, no. 1, pp. 35–53, 2001.
[12] Z. Liu, G. Giannakis, S. Barbarossa, and A. Scaglione, "Transmit-antenna space-time block coding for generalized OFDM in the presence of unknown multipath," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 1352–1364, July 2001.
[13] S. Zhou, B. Muquet, and G. Giannakis, "Semi-blind channel estimation for block precoded space-time OFDM transmissions," in *Proc. IEEE Workshop Stat. Signal Process.*, Singapore, 2001, pp. 381–384.
[14] P. Vandaele, G. Leus, and M. Moonen, "A noniterative blind signal separation algorithm based on transmit diversity and coding," in *Proc. 33rd Asilomar Conf. Signals, Syst., Comput.*, Oct. 1999.
[15] H. Bölcskei, R. Heath, A. Paulraj, and Computers, "Blind channel estimation in spatial multiplexing systems using nonredundant antenna precoding," in *Proc. 33rd Asilomar Conf. Signals, Syst., Comput.*, 1999.
[16] G. Leus, P. Vandaele, and M. Moonen, "Deterministic blind modulation-induced source separation for digital wireless communications," *IEEE Trans. Signal Processing*, vol. 49, pp. 219–227, Jan. 2001.
[17] H. Artés and F. Hlawatsch, "Blind equalization of MIMO channels using deterministic precoding," in *Proc. ICASSP*, vol. VI, Salt Lake City, UT, 2001, pp. 2153–2156.
[18] A. Swindlehurst, "Simultaneous channel estimation and decoding for diagonal space-time codes," in *Proc. Sensor Array Multichan. Process. Workshop*, Cambridge, MA, 2000, pp. 173–177.
[19] Y. Xin, Z. Wang, and G. Giannakis, "Linear unitary precoders for maximum diversity gains with multiple transmit and receive antennas," in *Proc. 34th Asilomar Conf. Signals, Syst., Comput.*, 2000, pp. 1553–1557.
[20] Z. Liu and G. Giannakis, "Space-time block coded multiple access through frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 49, pp. 1033–1044, June 2001.
[21] R. Heath, H. Bölcskei, and A. Paulraj, "Space-time signaling and frame theory," in *Proc. ICASSP*, vol. IV, Salt Lake City, UT, 2001, pp. 2445–2448.
[22] Y. Xin, Z. Wang, and G. Giannakis, "Space-time diversity systems based on unitary constellation-rotating precoders," in *Proc. ICASSP*, vol. IV, Salt Lake City, UT, 2001, pp. 2429–2432.
[23] B. Hassibi and B. Hochwald, "High-rate linear space-time codes," in *Proc. ICASSP*, vol. IV, Salt Lake City, UT, 2001, pp. 2461–2464.
[24] C. Budianu and L. Tong, "Channel estimation for space-time orthogonal block codes," *IEEE Trans. Signal Processing*, vol. 50, pp. XXX–XXX, Oct. 2002.
[25] J. Manton, I. Mareels, and Y. Hua, "Affine precoders for reliable communications," in *Proc. ICASSP*, Istanbul, Turkey, 2000, pp. 2749–2752.
[26] A. Scaglione, G. Giannakis, and S. Barbarossa, "Redundant filterbank precoders and equalizers, Part I: Unification and optimal designs," *IEEE Trans. Signal Processing*, vol. 47, pp. 1988–2006, July 1999.
[27] ——, "Redundant filterbank precoders and equalizers, Part II: Blind channel estimation, synchronization and equalization," *IEEE Trans. Signal Processing*, vol. 47, pp. 2007–2022, July 1999.
[28] C. Martin and B. Ottersten, "Joint channel estimation and detection for interference cancellation in multi-channel systems," in *Proc. 10th Stat. Signal Array Process. Workshop*, 2000, pp. 186–190.
[29] D. Gesbert, P. Duhamel, and S. Mayrargue, "On-line blind multichannel equalization based on mutually referenced filters," *IEEE Trans. Signal Processing*, vol. 45, pp. 2307–2317, Sept. 1997.
[30] A. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 1136–1157, May 1996.
[31] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array. Part I. Algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1184–1197, May 1996.
[32] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.



A. Lee Swindlehurst (SM'99) received the B.S. (summa cum laude) and M.S. degrees in electrical engineering from Brigham Young University (BYU), Provo, UT, in 1985 and 1986, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1991.

From 1983 to 1984, he was with Eyring Research Institute, Provo, as a scientific programmer. From 1984 to 1986, he was a Research Assistant with the Department of Electrical Engineering, BYU, working on various problems in signal processing and estimation theory. He was awarded an Office of Naval Research Graduate Fellowship for 1985 to 1988, and during most of that time, he was with the Information Systems Laboratory, Stanford University. From 1986 to 1990, he was also with ESL, Inc., Sunnyvale, CA, where he was involved in the design of algorithms and architectures for several radar and sonar signal processing systems. He joined the faculty of the Department of Electrical and Computer Engineering, BYU, in 1990, where he holds the position of Full Professor. From 1996 to 1997, he held a joint appointment as a visiting scholar at both Uppsala University, Uppsala, Sweden, and at the Royal Institute of Technology, Stockholm, Sweden. His research interests include sensor array signal processing for radar and wireless communications, detection and estimation theory, and system identification.

Dr. Swindlehurst is currently serving as Secretary of the IEEE Signal Processing Society and is a past Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, a past member of the Statistical Signal and Array Processing Technical Committee of the IEEE Signal Processing Society, and past Vice-Chair of the Signal Processing for Communications Technical Committee within the same society. He has served as the Technical Program Chair for the 1998 IEEE Digital Signal Processing Workshop and for the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. He is also a recipient of the 2000 IEEE W. R. G. Baker Prize Paper Award and is co-author of a paper that received a Signal Processing Society Young Author Best Paper Award in 2001.

Geert Leus, photograph and biography not available at time of publicaiton.