Low Complexity Optimal Joint Detection for Oversaturated Multiple Access Communications

Rachel E. Learned, *Student Member, IEEE*, Alan S. Willsky, *Fellow, IEEE*, and Don M. Boroson, *Member, IEEE*

***Abstract*—Optimal joint detection for interfering (nonorthogonal) users in a multiple access communication system has, in general, a computational complexity that is exponential in the number of users. For this reason, optimal joint detection has been thought to be impractical for large numbers of users. A number of *suboptimal* low-complexity joint detectors have been proposed for direct sequence spread spectrum user waveforms that have properties suitable for mobile cellular and other systems. There are, however, other systems, such as satellite systems, for which other waveforms may be considered. This paper shows that there are user signature set selections that enable *optimal* joint detection that is *extremely low* in complexity. When a hierarchical cross-correlation structure is imposed on the user waveforms, optimal detection can be achieved with a tree-structured receiver having complexity that is, in typical cases, a *low-order-polynomial* in the number of users. This is a huge savings over the *exponential* complexity needed for the optimal detection of general signals.**

Work in recent literature has shown that a hierarchically structured signal set can achieve oversaturation (more users than dimensions) with no growth in required signal-to-noise ratio. The proposed tree detector achieves low-complexity optimal joint detection even in this oversaturated case.

I. Introduction

Multiple access (MA) communication represents an active area of current research since it is the only means of communication among users in wireless systems such as mobile and cellular terrestrial systems and satellite-based systems. In each of these applications, the possibility of many users sharing the available communication channel offers obvious advantages in terms of flexible and cost-efficient use of the channel. In addition, MA also poses a number of challenging research problems including many that fall within the domain of signal processing. This paper investigates one of those challenges, namely, the problem of optimal detection in uncoded MA communications.

The importance and difficulty of the problem of detection in an uncoded MA system has been recognized for some time [1]–[4]. In particular, consider a pulse-amplitude-modulated (PAM) communication system in which each user transmits a distinct waveform, the amplitude of which is modulated by a weight corresponding to the information to be communicated.[1] If there is only one user transmitting through an additive white Gaussian noise channel, optimal detection at the receiver is realized by a simple matched filter followed by a quantization to the closest weight used in transmission [5]. If, however, many users were to transmit through the channel, the situation can become far more complex.

A. The Problem

Time-bandwidth restrictions on any communication system limit the dimension N of the space of possible user waveforms. Adopting the commonly-used vector space framework, the N-dimensional signal space would correspond to $\mathbb{R}^N$, and the multiuser joint detection problem may be stated as follows: For a given set of user waveforms represented in signal space by the set of signal vectors, $\{\mathbf{s}_k\}_1^K$, $\mathbf{s}_k \in \mathbb{R}^N$, the general uncoded detection problem is to compute an estimate of weights $\mathbf{b}$ from an observation $\mathbf{r} \in \mathbb{R}^N$,

$$\begin{aligned}\mathbf{r} &= \sum_{k=1}^{K} b_k \mathbf{s}_k + \sigma \mathbf{n}\\ &= \mathbf{S}\mathbf{b} + \sigma\mathbf{n}\end{aligned} \tag{1}$$

where we have the following:

- K is the number of users.
- $\mathbf{b} \in \{[b_1 \cdots b_K]^T \mid b_i \in P_i\}$, where P_i is some finite set of amplitudes, and the b_i's are iid uniform. For P_i having M elements, this is M-ary PAM.
- $\mathbf{S} = [\mathbf{s}_1, \cdots, \mathbf{s}_K]$ is an $N \times K$ matrix whose columns are user signal vectors as seen at the receiver.
- $\mathbf{n}$ is a Gaussian vector of zero mean and identity covariance.
- σ is the noise standard deviation.

B. Background

One case in which detection is simple is where the user waveforms are orthogonal. In this case, as in the single-user case, a matched filter followed by a quantization to the closest

Manuscript received December 29, 1995; revised August 27, 1996. This work was supported in part by the Department of the Air Force under contract F19628-95-C-0002 and in part by the Air Force Office of Scientific Research under Grant F49620-95-1-0083 and the Army Research Office under Grant DAAL03-92-G-0115.

R. E. Learned and A. S. Willsky are with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

D. M. Boroson is with the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA USA.

Publisher Item Identifier S 1053-587X(97)00520-5.

[1] In binary communications, this weight takes on one of two values. Among the most popular methods for binary PAM is binary phase-shift-keying (BPSK) in which the weights are $\{+1, -1\}$. For general M-ary PAM, however, M possible values are allowed for these weights.

weight used in transmission is optimal for each user.[2] The restriction to orthogonal signal sets, however, is often not a satisfactory one.

The assumption of orthogonality among user signals must be abandoned if we are to offer service to more users than orthogonality would allow. In the absence of time-varying fading, time-varying multipath, or frequency dispersion, it is possible to constrain user signals to be orthogonal.[3] Of course, this choice limits the number of users to the dimension of the signal space available for transmission. "Oversaturating" the signal space with users can, in principle, be accomplished with minimal impact on system performance, assuming that *optimal* detection can be implemented.[4] It is, therefore, desirable to increase the number of users beyond the orthogonal limit in order to enhance both system utilization and throughput. The success of such a system requires that the problem of optimal detection for nonorthogonal signal sets be confronted.

The challenge, then, is to design optimal detectors for MA systems that employ nonorthogonal signal sets. As discussed in [2], the optimal joint detector for an arbitrary, nonorthogonal signal set has exponential complexity in the number of users K. This is a catastrophic increase over the linear complexity of a bank of matched filters: one for each user. Surprisingly, though, the convention currently used, even in the case of nonorthogonal users, is a bank of matched filters where for each user, the interference from all other users is assumed to be a second source of "noise." With this type of detection, however, it is understood that the error rate will be higher than that obtained by the computationally complex optimal detector.

Indeed, as argued by Lupas and Verdú in [1], the performance loss of the conventional approach, as compared with the optimal, can be significant.[5] This has motivated several researchers [1]–[4] to consider more complex, *suboptimal* detection algorithms that perform joint detection for all users; better performance than the simple matched filtering approach is achieved with complexity that is at most polynomial in the number of users. These methods, however, were intended for pseudo-noise user signals[6] that form a linearly independent set.[7]

In contrast, this paper addresses the problem of finding an *optimal* joint detection algorithm that can accommodate the case of $K > N$ users in N-dimensional signal space that, like the suboptimal detectors, has complexity that is a low-order polynomial in the number of users. The key to devising such a detection algorithm is to take advantage of the flexibility many MA systems have in choosing the set of user waveforms so that an advantageous geometric structure is present.[8] In particular, the class of signal sets we consider has a hierarchical tree structure that allows for a rich variety of possibilities. For example, this desired tree structure is present in signals of considerable current interest in the signal processing and communication literature such as wavelets and wavelet packets. Moreover, we find that Ross and Taylor [6], [8] have developed signal design guidelines that fit $K > N$ users in N dimensions while preserving the orthogonal minimum distance. The tree hierarchy is a byproduct of their design.

In the next section, the signal set structure of interest is described and illustrated. In Section III, an overview of the hierarchical tree joint detector is given via an example, a formal derivation of the low-complexity optimal detector is done, and a calculation of its computational complexity is derived. Section IV details the processing procedure of the tree detection algorithm and steps through a binary example. The paper is concluded in Section V.

II. The Signal Sets

A. Signal Vector Set Structure

The geometric structure imposed on the signal vector set[9] is best described by saying that the set of signatures has *tree-structured cross-correlations*. The columns of the matrix $\mathbf{S}$ can be assigned to the nodes of a tree like the one shown in Fig. 1. The tree pictorially conveys the following required relationships among user signal vectors:

- Each vector at a given level of the tree is orthogonal to all other vectors at that level.
- A signal vector is correlated only with its ancestor vectors and its descendant vectors.

Both linearly dependent and linearly independent sets of signature vectors may be created to have tree-structured cross correlations. The detector detailed in this paper finds the optimal solution for both cases.

The constraint of tree-structured cross-correlations, while very particular, actually allows a considerable amount of flexibility in designing user waveforms. Given a tree, a signal set may be constructed to possess the desired cross-correlation structure. The waveforms at the bottom level of the tree

[2] Forcing user transmissions to be orthogonal or nearly orthogonal, even at the expense of inserting wasteful buffer zones in which no user is permitted to transmit, is common practice in systems of present.

[3] For example, current MA processing-satellite systems employing narrow-beam terrestrial antennas assign each user a disjoint portion of the available frequency spectrum. This is frequency division multiple access (FDMA). The satellite uplink channel is well modeled by (1), where the received waveforms are, simply, translated and attenuated replicas of the transmitted waveforms.

[4] Moreover, as can be seen in the work of Ross and Taylor [6]–[8], it is, indeed, possible to design signal sets having more users than dimensions where the minimum interdecision-point distance resulting from use of this set is the same as that achieved by the orthogonal set. Their design constrains all users to have powers no higher than the users in the orthogonal set.

[5] In particular, for the "near-far" problem (large power variations among interfering users), the conventional detector fails consistently.

[6] The MA detection literature is heavily concentrated on the mobile cellular problem for which code division multiple access (CDMA)—through the use of pseudo-noise user signals—exhibits advantages over orthogonal signals for use with the conventional detector. By restricting the user waveforms to be pseudo-noise pulses, orthogonality among users is not possible.

[7] Although some of these detection algorithms may be applied in the linearly dependent case, they were not intended for the oversaturated problem and, therefore, give very poor performance.

[8] In contrast with previous work with suboptimal detectors, user signals are not constrained to be pseudo-noise pulses. As will be discussed in later sections, the structure is used as a guideline for the choice or design of the actual waveforms to be transmitted.

[9] For ease of notation, the abstract signal space representation of real signals is used, and hence, all properties imposed on the signal vectors will also be true for the real waveform counterparts. The signal vector set structure described in this section, therefore, can be viewed as design guidelines for the waveforms that would be used in practice.

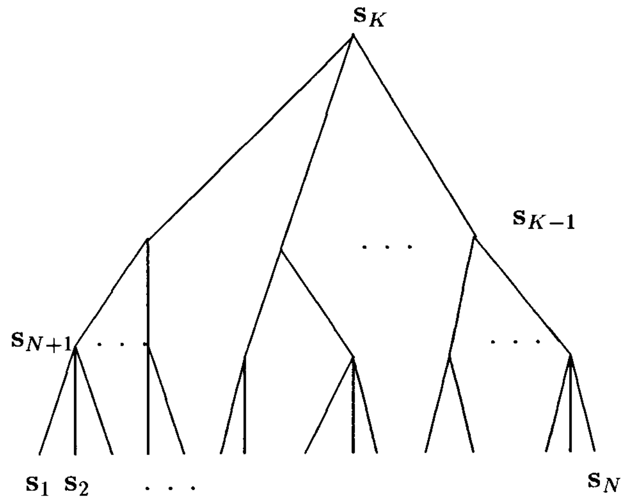


Fig. 1. Example of a general tree showing the correlation structure needed among signature vectors within the signature set.

comprise an orthogonal set. An orthogonal set is obtained at any level, i.e., the lth level, by constructing a signal at each node at this level as a linear combination of the signals at its bottom-most descendant nodes. Since orthogonal signals have been assigned to the lowest level nodes of the tree, the sets of bottom-level descendants for distinct nodes at the lth level are disjoint, and consequently, the signals created at level l are mutually orthogonal.

It follows that the general construction of a signal set with tree-structured cross-correlations requires the following:

a) the specification of a tree,
b) the specification of *any* orthogonal basis $\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N$ of $\mathbb{R}^N$ that is then assigned to the N nodes on the bottom of the tree,
c) the specification of the weights for each of the linear combinations used to construct signals from their bottom-level descendants,
d) possibly, the deletion of signals at any of the nodes.

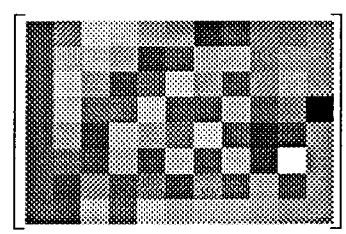
This formulation allows for considerable flexibility in designing the signal set since *any* choices that satisfy a)–d) will lead to the desired geometric structure on the signal set. Note also that d) provides us with the flexibility to capture linearly independent sets with the desired correlation.[10]

B. Some Examples of Signature Sets

In this section, we illustrate examples of two particular choices of signal vector sets, one of which involves signals of considerable current interest in the signal processing community, namely, wavelets and wavelet packets [9], [10], and one that was introduced in [6]–[8] directly in the context of designing signal vector sets for oversaturated MA systems.[11] Note that the following two examples are, simply, two different realizations of requirements a)–c), in the above discussion.

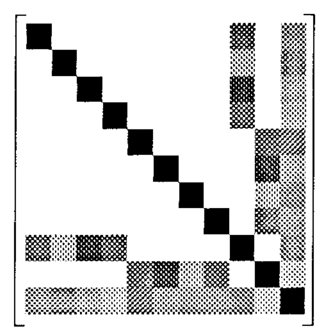
1) Wavelet Packet Sets: Wavelet and wavelet packet waveforms may be generated from a tree-structured procedure in which subspaces (generated by sets of orthogonal signals) are decomposed into Cartesian products of orthogonal lower dimensional subspaces.[12] The result is a wavelet or wavelet packet dictionary consisting of an overcomplete set of basis functions. A discrete wavelet packet dictionary offers a rich set of signal vectors from which to select many tree-structured sets. The use of wavelet packet waveforms has received much attention in recent literature for single user communications (i.e., bandwidth efficient and covert signal designs, mitigation of various types of jammers, transmission through non-Gaussian channels, etc.)[13] An example of a discrete wavelet packet signal set is shown below as an intensity matrix where each element of the matrix is shown as a pixel in the 8×11 image. The values are shown in gray scale where the smallest is denoted in white and the largest is denoted in black.

$\mathbf{S}$ =



Each column of $\mathbf{S}$ is a user signature vector. In order to reveal the tree-structured cross correlations among user signatures, the absolute values of the elements of $\mathbf{S}^T\mathbf{S}$ are displayed below, where 0 and 1 are denoted in white and black, respectively.

$\mathbf{S}^{\mathbf{T}}\mathbf{S}$ =



$\mathbf{S}^T\mathbf{S}$ is the matrix of cross-correlations between the received signals of the 11 different users. The wavelet packet signal vector set can be cast onto a tree with three levels as shown in Fig. 2.

2) Minimum Distance Sets: Another example is the minimum distance sets developed by Ross and Taylor in [7] and [8]. They begin with N orthogonal users in N dimensions. The set of possible received points based on an M-ary PAM MA system with an orthogonal set of signal vectors has associated with it a minimum distance. That is, if the vectors $\{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N\}$ are the orthogonal signal set, then there is

[10] For simplicity, however (and since we wish to emphasize the applicability of our methods to the oversaturated case), we will assume that our tree is *full*, i.e., that there is a user signature at each node on the tree. The extension of our low-complexity optimal detection scheme to the case in which there are fewer users is immediate.

[11] As we have indicated, however, there are many different signal sets that can be constructed to have the tree structure we have just described.

[12] For a tutorial treatment of wavelet packets, see the paper by Coifman and Wickerhauser [10].

[13] The reader is referred to [11] and [12] and the references therein.

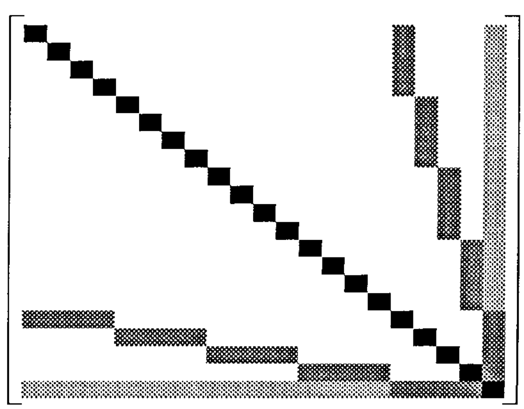
a specified minimum distance between any two points in the received constellation, $\{\sum_{k=1}^{N} b_k \mathbf{s}_k \mid b_k \in P_k\}$. Since the distance between the elements in this set is directly related to the probability that the optimal detector makes an error, maintaining a specified minimum distance is desirable. Ross and Taylor devise a method for adding additional, energy-constrained, linearly dependent users so that the minimum Euclidean distance between received points is preserved. The reader is referred to [8] for details of their construction.

Ross and Taylor, for antipodal binary modulation $P_k = \{+1, -1\}$, fit $\frac{4}{3}N - \frac{1}{3}$ unit energy signal vectors into N dimensions, where N must be a power of 4. A specific example detailed in [8] is repeated below.

$$\mathbf{S} = \left[\begin{array}{c|ccccc} & 1/2 & 0 & 0 & 0 & 1/4 \\ & 1/2 & 0 & 0 & 0 & 1/4 \\ & 1/2 & 0 & 0 & 0 & 1/4 \\ & 1/2 & 0 & 0 & 0 & 1/4 \\ & 0 & 1/2 & 0 & 0 & 1/4 \\ & 0 & 1/2 & 0 & 0 & 1/4 \\ & 0 & 1/2 & 0 & 0 & 1/4 \\ \mathbf{I}_{16} & 0 & 1/2 & 0 & 0 & 1/4 \\ & 0 & 0 & 1/2 & 0 & 1/4 \\ & 0 & 0 & 1/2 & 0 & 1/4 \\ & 0 & 0 & 1/2 & 0 & 1/4 \\ & 0 & 0 & 1/2 & 0 & 1/4 \\ & 0 & 0 & 0 & 1/2 & 1/4 \\ & 0 & 0 & 0 & 1/2 & 1/4 \\ & 0 & 0 & 0 & 1/2 & 1/4 \\ & 0 & 0 & 0 & 1/2 & 1/4 \end{array}\right] \tag{2}$$

Here, I_{16} is the 16-D identity matrix. The cross-correlation matrix $\mathbf{S}^T\mathbf{S}$ is given below.

$\mathbf{S}^T\mathbf{S} =$



The structure of $\mathbf{S}^T\mathbf{S}$ reveals that this minimum distance set of signature vectors may be cast onto a three level quad tree for which four children emanate from each parent node. These signature sets were designed for their minimum distance property. The tree hierarchy they possess is a byproduct that can be exploited.

3) Ideas on Future Work for Signal Design: From the above discussion, it is clear that oversaturation is achievable with minimal performance loss if optimal detection is used. Moreover, as is detailed in Section III, for MA systems employing tree-structured signatures, very low complexity optimal detection can be achieved. The existence of this low-complexity detector, in principle, enables an oversaturated MA system to operate with relatively the same detector complexity and probability of error as its corresponding orthogonal system. A challenging and interesting problem, then, is the study of user packing with tree-structured waveform sets. Any valuable work in this area will be achieved through careful understanding of the signal design problem and is beyond the scope of this paper. Our ideas for future work are discussed below.

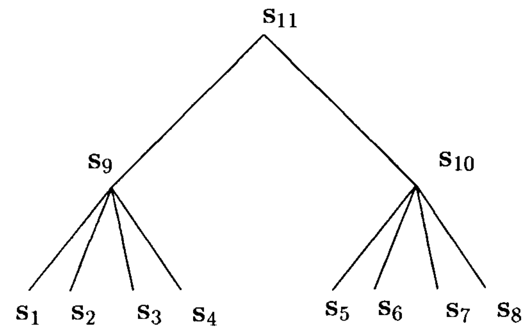


Fig. 2. Correlation tree for a wavelet packet signature set.

An interesting and challenging question that is more general than that answered by Ross and Taylor would ask how many users with specified power and performance constraints can be fit into a tree structure of a given dimension. Some rough, preliminary observations show promise in the increase in throughput using tree structured signatures.

The first scenario of this is, of course, the work of Ross and Taylor that is mentioned above and detailed in [6]. If users are constrained to be of equal energy, the Ross/Taylor sets realize, in typical cases, a 33% increase in the number of users relative to the orthogonal system while maintaining the minimum distance of the orthogonal system. This, by itself, is substantial for many applications, including military satellite systems.

A second, more realistic, scenario is where user powers are unequal (near-far problem). It is easy to come up with simple examples showing much greater than 33% increases in the number of users, also while maintaining minimum distance. For an example starting with two orthogonal users of identical energy, we realize a 50% increase in the number of users with no change in the minimum distance by squeezing in a third user of energy that is 18 dB higher than the first.

A third scenario leads us to user packing with criterion other than preservation of minimum distance. Specifically, an understanding of the relationship between the degradation of minimum distance (i.e., performance) and the number of additional users in a tree-structured signature set is not essential to the ultimate success of a tree-structured system but would offer Shannon-like limits against which to measure system throughput. The realization of a system that is allowed graceful degradation with the addition of each user might employ waveforms sets based on, for example, wavelet packets.[14]

[14]The minimum distance for the arbitrarily chosen 11 unit-energy users in the 8-D wavelet packet example in Section II-B is 1.13. If we were to have used a minimum distance construction, we would have been able to fit only 10 users but with a minimum distance of 2. The choice of particular wavelet packet signature vectors was not in any way optimized for this example.

III. The Tree Joint Detection Algorithm

A. Overview of the Detector

The optimal joint detector for the problem stated in (1) chooses the weight vector estimate $\hat{\mathbf{b}}$ according to the nearest neighbor or minimum distance rule.

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b} \in P^K} ||\mathbf{r} - \mathbf{S}\mathbf{b}||^2. \quad (3)$$

For ease of discussion, each user is assumed to employ the same M-ary PAM for the remainder of this paper; this assumption is not essential to the operation of the tree algorithm. An MA system employing an *arbitrary* set of signal vectors can achieve optimal detection through an exhaustive search, i.e., the detector needs to perform $M^K - 1$ comparisons to find the best estimate [2].

If the signal set has been chosen to have tree-structured cross correlations, the optimal detector of (3) can be achieved through a tree-structured algorithm that offers a huge reduction in the number of comparisons. In particular, because of this structure, a signature at a given node is *correlated* with all signatures at its ancestor and descendant nodes and is *orthogonal* to all other signatures on the tree. The weight estimate $\hat{b}_n$ at a given node n will effect the estimates at descendant and ancestor nodes but will not effect the other estimates on the tree.

Consider the tree structure in Fig. 2 and consider first the choice of the weight estimates for users 1–4 having signal vectors $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4$. These vectors are mutually orthogonal and are also orthogonal to $\mathbf{s}_5, \mathbf{s}_6, \mathbf{s}_7, \mathbf{s}_8$ and $\mathbf{s}_{10}$ but not to $\mathbf{s}_9$ and $\mathbf{s}_{11}$. Since $\mathbf{s}_5$–$\mathbf{s}_8$ and $\mathbf{s}_{10}$ are also correlated with $\mathbf{s}_{11}$, the decisions on weight estimates for $\mathbf{s}_5$–$\mathbf{s}_8$ and $\mathbf{s}_{10}$ are coupled with those for $\mathbf{s}_1$–$\mathbf{s}_4$. These estimates can, however, be decoupled by looking, instead, at the *conditional* estimates. Specifically, for each possible pair of weight estimates for $\mathbf{s}_9$ and $\mathbf{s}_{11}$, the optimal weight estimates for $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \mathbf{s}_4$ can be independently computed. The result of this calculation for $\mathbf{s}_1$–$\mathbf{s}_4$ can then be thought of as producing a *conditional weight estimate table*, i.e., for each possible pair of choices for the weight estimates for $\mathbf{s}_9$ and $\mathbf{s}_{11}$, the optimal weight estimates for $\mathbf{s}_1$–$\mathbf{s}_4$ are known. Similarly, for each pair of possible weight estimates for $\mathbf{s}_{10}$ and $\mathbf{s}_{11}$, the optimal estimates for $\mathbf{s}_5$–$\mathbf{s}_8$ can be computed. The estimation process may be iterated for $\mathbf{s}_9$: For each possible choice of weight value for its ancestor $\mathbf{s}_{11}$ and with knowledge of the just-constructed conditional estimate table for its descendants $\mathbf{s}_1$–$\mathbf{s}_4$, the optimal estimate for $\mathbf{s}_9$ may be computed in a manner decoupled from the analogous computation for $\mathbf{s}_{10}$. This gives conditional estimate tables for $\mathbf{s}_9$ and $\mathbf{s}_{10}$, which then can be used to determine the optimal estimate for $\mathbf{s}_{11}$ at the top of the tree. Conceptually, once this estimate is obtained, it is a simple matter of successive table lookups that propagate down the tree to determine the optimal estimates first for $\mathbf{s}_9$ and $\mathbf{s}_{10}$ and then for their descendants.

As this simple example illustrates, the tree detection algorithm takes advantage of the tree structure and sweeps through the tree from bottom to top, creating a conditional weight estimate table at each node. The table of decisions at a given node is conditioned on weight decisions of the ancestors and is a function of weight decisions of the descendants. Since each conditional estimate table requires an entry for each possible combination of weights at all ancestor nodes, the number of computations needed to create a table is exponential in the number of ancestors (since if there are l ancestors there are M^l possible sets of weight values for these ancestors). This complexity decreases exponentially as l decreases, i.e., as the algorithm moves from the bottom to the top of the tree, the number of decisions made at each level decreases exponentially until there is only one decision associated with the top node of the tree. The full weight vector estimate for all user weights is a byproduct of the last decision at the top of the tree.

While the complexity of the procedure as we have described it to this point is exponential in the number of levels in the tree (which bounds the number of ancestors of each node), the actual algorithm complexity is, in fact, extremely modest. If the tree were of uniform construction, i.e., if there are Q children emanating from each node, the number of levels of the tree is *logarithmic* in the number of users K. The overall complexity, then, is bounded by a very low-order polynomial in K. This is discussed more fully in Section III-C.

B. Derivation of Tree Detector

The global cost that must be minimized in (3) is

$$F(\mathbf{r}, \mathbf{b}) = ||\mathbf{r} - \mathbf{S}\mathbf{b}||^2. \quad (4)$$

In general, $F(\mathbf{r}, \mathbf{b})$ is not separable by weight variables b_i. Hence, the solution to (3) is found by the calculation and comparison of $F(\mathbf{r}, \mathbf{b}), \forall \mathbf{b} \in P^K$.

The introduction of tree-structured cross correlations transforms the structure of the cost function. The complexity of finding the smallest cost may be reduced by making decisions in stages. The independence of the conditional decisions discussed in the previous section can be seen mathematically. Specifically, the global cost may be separated into independent terms. First, (4) may be rewritten as

$$\begin{aligned} F(\mathbf{r}, \mathbf{b}) &= \sum_{i=1}^{N} f_i(\mathbf{r}, \mathbf{b}) \\ &= \sum_{i=1}^{N} (\mathbf{r}[i] - \mathbf{t}[i])^2 \end{aligned} \quad (5)$$

where $\mathbf{t} = \mathbf{S}\mathbf{b}$, and $\mathbf{t}[i]$ is the ith element of the vector $\mathbf{t}$. In general, each term $f_i(\mathbf{r}, \mathbf{b})$ is a function of all users' bits.

If the signature set were to exhibit tree-structured cross correlations, a rotation matrix $\mathbf{S}_R$ may be constructed from the orthogonal basis vectors that reside on the bottom of the tree.

$$\mathbf{S}_R = \left[\frac{\mathbf{s}_1}{||\mathbf{s}_1||} \; \frac{\mathbf{s}_2}{||\mathbf{s}_2||} \cdots \frac{\mathbf{s}_N}{||\mathbf{s}_N||} \right]. \quad (6)$$

Since the cost function of (4) is a squared Euclidean distance between a Gaussian random vector $\mathbf{r}$ and a deterministic vector $\mathbf{S}\mathbf{b}$, a rotation of this difference vector $\mathbf{r} - \mathbf{S}\mathbf{b}$ does not change its length or probability distribution. Hence,

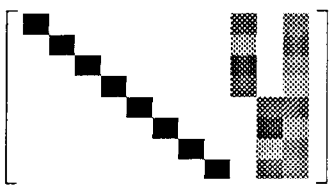
$$F(\mathbf{r}, \mathbf{b}) = \|\mathbf{S}_R^T\mathbf{r} - \mathbf{S}_R^T\mathbf{S}\mathbf{b}\|^2. \tag{7}$$

The tree structure is reflected in the position of the zero-valued elements of $\mathbf{S}_R^T\mathbf{S}$. In this form, the partition of $F(\mathbf{r}, \mathbf{b})$ into terms is

$$\begin{aligned} F(\mathbf{r}, \mathbf{b}) &= \sum_{i=1}^{N} \tilde{f}_i(\mathbf{r}, \mathbf{b}) \\ &= \sum_{i=1}^{N} (\tilde{\mathbf{r}}[i] - \tilde{\mathbf{t}}[i])^2 \end{aligned} \tag{8}$$

where $\tilde{\mathbf{r}} = \mathbf{S}_R^T\mathbf{r}$, $\tilde{\mathbf{t}} = \mathbf{S}_R^T\mathbf{S}\mathbf{b}$, and where the indices $i \in \{1, 2, \cdots, N\}$ correspond to the orthogonal users at the bottom nodes of the tree.

For example, the rotated version of the wavelet packet signal matrix from Section II-B is

$\mathbf{S}_R^T\mathbf{S} =$ 

and it is clear that $\tilde{\mathbf{t}}[i] = [\mathbf{S}_R^T\mathbf{S}\mathbf{b}][i]$ can be written as a linear combination of b_i and the elements of $\mathbf{b}_{ai}$. Here, $\mathbf{b}_{ai}$ denotes the vector of weights associated with the ancestors of node i.

In general, it follows that the "rotated" cost may be separated into additive terms, where each term is a function of only one of the weights b_i, $i \in \{1, 2, \cdots, N\}$ and all of the weights that correspond to its ancestors. Each term $\tilde{f}_i(\mathbf{r}, \mathbf{b})$ in (8) may be explicitly written as $\tilde{f}_i(\mathbf{r}, b_i, \mathbf{b}_{ai})$. Note that for $i, j \in \{1, 2, \cdots, N\}$ and $i \neq j$, $\tilde{f}_i(\mathbf{r}, b_i, \mathbf{b}_{ai})$ and $\tilde{f}_j(\mathbf{r}, b_j, \mathbf{b}_{aj})$ have no common unknown parameters, given the values for $\mathbf{b}_{ai}$ and $\mathbf{b}_{aj}$. It follows that the optimal solution may be determined through the optimization of each term conditioned on the values of the weights corresponding to the ancestors of the index i of that term. A dynamic program may be written to solve this minimization problem.

1) Some Practical Issues: Some practical issues of implementation arise from the assumptions made by the tree detector algorithm, namely, 1) the channel response $H(\cdot)$ is capable of being estimated, and the set of assigned users signals $\mathbf{x}_k$ leads to the set of received signals $\mathbf{s}_k = H(\mathbf{x}_k)$ that exhibit the needed tree-structure and 2) knowledge or accurate estimates of user timings and received powers are available. These assumptions are valid for many MA scenarios of present. For example, the processing satellite[15] MA uplink channel is free of both multipath and frequency dispersion, making assumption 1) practical using current technology. Assumption 2) is also reasonable within limitations since users request access from the satellite through a side channel. During this acquisition stage, the satellite can determine the power and symbol timing of the user. Another MA system having similar traits to the satellite is the wireless local loop [14].[16] This system may offer more of a challenge with assumptions 1) and 2), but since it is conceivable for channel probing to occur during off hours and periodically during normal operation, all multipaths and attenuations could be known for each transmitter.

Our ongoing work investigates the break down of assumption 2). In particular, although a user's power is often easily acquired during the acquisition process, and its envelope waveform and modulation frequency are assigned, the phase of each user cannot be assigned and may need to be estimated. A user's phase may be estimated with a tree-structured procedure as well. The effect phase uncertainty has on bit error rates throughout the tree is currently under investigation. Training sequences may be used to lower the variance of the phase estimate at a cost of higher computations, which may be impractical for some systems. Future work includes the study of suboptimal ways of obtaining good phase estimates along with low complexity ways of refining phase estimates for tree-structured MA. These issues are addressed as part of Learned's thesis [13].

C. Computational Complexity

The conceptual description of the tree detector in Section III-A and the dynamic programming description in Section III-B inevitably include many wasteful calculations and storage requirements. The resulting complexity of this "inefficient" version leads to an upper bound on tree detector complexity that is extremely low.

For simplicity of calculation of computational complexity, the tree is restricted to be of uniform composition in that there are exactly Q children emanating from each node. Recall that N is the number of signal space dimensions available (the number of nodes at the bottom of the tree), and M is the number of levels that can be modulated by each user. A measure of complexity that is in agreement with the MA joint detection literature is the number of compares c needed to perform the detection algorithm.[17]

The complexity is derived by counting the number of comparisons needed to execute the tree algorithm. Some facts used in the complexity calculation follow:

- Each node at level l has $l - 1$ ancestor nodes.
- There are Q^{l-1} nodes at level l of the tree.
- The tree has a total of L levels (counting the top as level 1).
- There are $N = Q^{L-1}$ nodes at the bottom of the tree, and thus, $L = \log_Q N + 1$.

The algorithm creates a conditional bit estimate table for each node. For a given $\hat{\mathbf{b}}_{an}$, the best of M possible values of $\hat{b}_n$ must be found. This requires $M - 1$ comparisons for

[15] At present, processing satellites, employing orthogonal MA, detect each user's transmission, correct errors caught by error correction coding, and then retransmit to close the link.

[16] The positions of all transmitters are known and fixed. This service would compete with existing wire-line service and might be delivered via an antenna on the top of buildings that are hooked up to the service.

[17] Counting the number of comparisons is equivalent to counting number of tentative decisions that must be made. Without computational optimization of the algorithm, each decision requires the computation of two metrics. Each metric requires several adds and subtracts. To find the order of the complexity of the tree algorithm, it is sufficient to count the number of compares.

a single configuration of $\hat{\mathbf{b}}_{an}$. Since there are $l-1$ ancestors of node n, there are M^{l-1} possible configurations of $\hat{\mathbf{b}}_{an}$. The tree detector, therefore, creates a single table at level l, node n, with $(M-1)M^{l-1}$ comparisons. There are Q^{l-1} tables needed for level l of the tree, and there are a total of $\log_Q N+1$ levels in the tree. It follows that the total number of comparisons needed for the tree algorithm is

$$\begin{aligned} c(N, Q, M) &= \sum_{l=1}^{\log_Q N+1} Q^{l-1}(M-1)M^{l-1} \\ &= (M-1)\frac{(QM)^{\log_Q N+1}-1}{QM-1} \\ &= \frac{(M-1)}{(QM-1)}(NQM^{\log_Q N+1}-1). \end{aligned}$$

For example, if a system were to employ antipodal modulation, $P=\{+1,-1\}$, $M=2$, and signal sets having quad-tree structure ($Q=4$) such as the minimum distance waveform sets, the number of comparisons needed for the tree detector estimate is

$$c(N, Q=4, M=2) = \frac{8N^{3/2}-1}{7}. \tag{9}$$

The computational complexity is polynomial in the number of dimensions. The number of users K in this special case is $K=\frac{4}{3}N-\frac{1}{3}$; hence, the tree detector is also polynomial in the number of users, resulting in a computational complexity of $O(K^{3/2})$.

IV. Signal Processing for the Optimal Tree Detector

This section examines the calculation and interpretation of key values used in the tree algorithm.

A. Calculation of the Estimate

Mathematically, the dependence between ancestors and descendants is revealed as the reduction of the general optimal estimator of (3) to the tree-structured optimal estimator below. For each node n of the tree, calculate the following estimate conditioned on the value of the set $\mathbf{b}_{an}$:

$$\begin{aligned} \hat{b}_n(\mathbf{r}|\mathbf{b}_{an}) = \arg\min_{b_n\in P} \|&\mathbf{r}-\mathbf{s}_n b_n - \mathbf{S}_{an}\mathbf{b}_{an} \\ &- \mathbf{S}_{dn}\hat{\mathbf{b}}_{dn}(\mathbf{r}|b_n, \mathbf{b}_{an})\|^2. \end{aligned} \tag{10}$$

The notation used throughout this section is shown in Table I. The set of estimates for all descendants of node n has already been calculated in the previous steps of the algorithm. Hence, $\hat{\mathbf{b}}_{dn}(\mathbf{r}|b_n, \mathbf{b}_{an})$ is best defined recursively. For a given set of values for $\{b_n, \mathbf{b}_{an}\}$,

$$\hat{\mathbf{b}}_{dn}(\mathbf{r}|b_n, \mathbf{b}_{an}) = \begin{bmatrix} \hat{\mathbf{b}}_{fcn_1}(\mathbf{r}|b_n, \mathbf{b}_{an}) \\ \hat{\mathbf{b}}_{fcn_2}(\mathbf{r}|b_n, \mathbf{b}_{an}) \\ \vdots \\ \hat{\mathbf{b}}_{fcn_{Kn}}(\mathbf{r}|b_n, \mathbf{b}_{an}) \end{bmatrix}. \tag{11}$$

For that same set of $\{b_n\ \mathbf{b}_{an}\}$, $\hat{b}_{cn_i}(\mathbf{r}|b_n\ \mathbf{b}_{an})$ has already been calculated; if the value found for $\hat{b}_{cn_i}(\mathbf{r}|b_n\ \mathbf{b}_{an})$ is ξ,

TABLE I
NOTATION

n	node index
pn	index of the parent to node n
$p^m n$	index of the ancestor to node n that is m levels above n
$an=\{pn, p^2n, p^3n, \cdots, p^{l-1}n\}$	set of indices corresponding to the ancestor nodes of node n
cn_i	node index for the i^{th} child of node n
Kn	number of children of node n
$cn=\{cn_1, cn_2, \ldots, cn_{Kn}\}$	set of indices corresponding to children of node n
$dn=\{cn_1, dcn_1, \cdots, cn_{Kn}, dcn_{Kn}\}$	set of indices corresponding to the descendant nodes of node n
$fn=\{n, dn\}$	the family of indices associated with node n
$\mathbf{S}_{an}$ ($\mathbf{S}_{dn}$)	signature vectors of all ancestors (descendants) of node n
$\hat{\mathbf{b}}_{an}$ ($\hat{\mathbf{b}}_{dn}$)	weight estimates for all ancestors (descendants) of node n
$\mathbf{y}_{i,an}=\mathbf{s}_i^T\mathbf{S}_{an}$	row vector of inner products
$\mathbf{y}_{i,dn}=\mathbf{s}_i^T\mathbf{S}_{dn}$	row vector of inner products
$\mathbf{Y}_{an,dn}=\mathbf{S}_{an}^T\mathbf{S}_{dn}$	matrix of inner products
$\mathbf{Y}_{dn,dn}=\mathbf{S}_{dn}^T\mathbf{S}_{dn}$	matrix of inner products

then the subvectors on the right-hand side of (11) are given by

$$\begin{aligned} &\hat{\mathbf{b}}_{fcn_i}(\mathbf{r}|b_n, \mathbf{b}_{an}) \\ &\quad= \begin{bmatrix} \hat{b}_{cn_i}(\mathbf{r}|b_n\ \mathbf{b}_{an}) = \xi \\ \hat{\mathbf{b}}_{dcn_i}(\mathbf{r}|b_{cn_i}=\xi, b_n, \mathbf{b}_{an}) \end{bmatrix}. \end{aligned} \tag{12}$$

We examine the argument of the minimization in (10) more closely. We may, of course, remove any terms that do not depend on b_n, and we can multiply by any positive constant.[18] As a result, some algebra shows that (10) is equivalent to

$$\hat{b}_n(\mathbf{r}|\mathbf{b}_{an}) = \arg\max_{b_n\in P} J(b_n|\mathbf{r}, \mathbf{b}_{an}) \tag{13}$$

where

$$\begin{aligned} J(b_n|\mathbf{r}, \mathbf{b}_{an}) = &[l_n b_n - \tfrac{1}{2}b_n^2 y_{n,n}] - b_n\mathbf{y}_{n,an}\mathbf{b}_{an} && (14)\\ &+ [\mathbf{l}_{dn}^T\hat{\mathbf{b}}_{dn}(\mathbf{r}|b_n, \mathbf{b}_{an}) \\ &- \tfrac{1}{2}\hat{\mathbf{b}}_{dn}^T(\mathbf{r}|b_n, \mathbf{b}_{an}) \\ &\quad \cdot \mathbf{Y}_{dn,dn}\hat{\mathbf{b}}_{dn}(\mathbf{r}|b_n, \mathbf{b}_{an})] && (15)\\ &- b_n\mathbf{y}_{n,dn}\hat{\mathbf{b}}_{dn}(\mathbf{r}|b_n, \mathbf{b}_{an}) && (16)\\ &- \mathbf{b}_{an}^T\mathbf{Y}_{an,dn}\hat{\mathbf{b}}_{dn}(\mathbf{r}|b_n, \mathbf{b}_{an}). && (17) \end{aligned}$$

Note that the only explicit processing of the data $\mathbf{r}$ is for the terms $l_n=\mathbf{s}_n^T\mathbf{r}$ and $\mathbf{l}_{dn}=\mathbf{S}_{dn}^T\mathbf{r}$ on lines (14) and (15).[19]

The first bracketed term $[l_n b_n - \frac{1}{2}b_n^2 y_{n,n}]$ on line (14) corresponds exactly to the decision statistic that would be used to choose b_n if there were no other users to consider or if all other users had orthogonal signals. The remaining terms, then, represent the *adjustments* of this decision statistic to reflect the impact of the nonorthogonality in the signals. The last term on line (14) $b_n y_{n,an}\mathbf{b}_{an}$ represents the interaction between

[18] or multiply by a negative constant and replace minimization by maximization

[19] The calculations of each l_i correspond to processing the data $\mathbf{r}$ through a filter matched to the signal $\mathbf{s}_i$. $\{l_i\}_1^K$ is the set of sufficient statistics needed for optimal detection. Reductions in the calculation of the set $\{l_i\}_1^K$ can be obtained by taking advantage of the exact relationships among user signatures on the tree. See Learned's thesis [13] for a detailed discussion.

TABLE II
TABLE CREATED AT NODE n FOR THE GENERAL BINARY CASE

$\mathbf{b}_{an}^T = [b_{pn} b_{p^2n} \cdots b_{p^{(l-1)}n}]$	$\hat{b}_{fn}(\mathbf{r}\vert\mathbf{b}_{an})$	$\eta_n(\mathbf{r}\vert\mathbf{b}_{apn}) = \eta_n(\mathbf{r}\vert b_{p^2n}, b_{p^3n}, \cdots, b_{p^{(l-1)}n})$
$[+1, +1, +1, \cdots, +1]$	$\hat{b}_{fn}(\mathbf{r}\vert +1, +1, +1, \cdots, +1)$	$\eta_n(\mathbf{r}\vert +1, +1, \cdots, +1)$
$[-1, +1, +1, \cdots, +1]$	$\hat{b}_{fn}(\mathbf{r}\vert -1, +1, +1, \cdots, +1)$	
$[+1, -1, +1, \cdots, +1]$	$\hat{b}_{fn}(\mathbf{r}\vert +1, -1, +1, \cdots, +1)$	$\eta_n(\mathbf{r}\vert -1, +1, \cdots, +1)$
$[-1, -1, +1, \cdots, +1]$	$\hat{b}_{fn}(\mathbf{r}\vert -1, -1, +1, \cdots, +1)$	
$\vdots$	$\vdots$	$\vdots$

the choice of b_n and the particular hypothesized choices for the ancestor weights. Note that since the values of $\mathbf{b}_{an}$ will be hypothesized, this term can be precomputed. Line (15) represents a counterpart to the bracketed term on line (14). Specifically, if all users other than the ones corresponding to the weights $\mathbf{b}_{dn}$ were not present, then

$$[\mathbf{1}_{dn}^T \mathbf{b}_{dn} - \tfrac{1}{2}\mathbf{b}_{dn}^T \mathbf{Y}_{dn,\,dn}\mathbf{b}_{dn}] \tag{18}$$

represents the decision statistic that would be used to determine the optimal choice of $\mathbf{b}_{dn}$. Since this is *not* the case, this term incorporates the decisions $\hat{\mathbf{b}}_{dn}(\mathbf{r}|b_n, \mathbf{b}_{an})$ conditioned on the value of $\{b_n, \mathbf{b}_{an}\}$. Line (16) accounts for the interactions between these descendant decisions and the possible decisions b_n. Likewise, line (17) accounts for the interactions between the descendant decisions and the hypothesized decisions for $\mathbf{b}_{an}$ at the ancestors of node n. Thus, all of the quantities needed in the last three lines (15)–(17) can be computed based on the value of b_n and the calculations that have already been performed at lower levels on the tree.

B. The Binary Conditional Decision Rule

This section focuses on the binary antipodal signaling case, i.e., when $P = \{+1, -1\}$. For each choice of $\mathbf{b}_{an}$, there is only one comparison to make for the minimization of (13).[20] The solution to (13) can, therefore, be expressed as[21]

$$\hat{b}_n(\mathbf{r}|\mathbf{b}_{an}) = \operatorname{sgn}\left[\frac{J(+1|\mathbf{r}, \mathbf{b}_{an}) - J(-1|\mathbf{r}, \mathbf{b}_{an})}{2}\right]. \tag{19}$$

Substituting the definition of $J(b_n|\mathbf{r}, \mathbf{b}_{an})$ from lines (14)–(17) into (19) and performing some algebra, (19) can be written as

$$\hat{b}_n(\mathbf{r}|\mathbf{b}_{an}) = \operatorname{sgn}[l_n - \delta_n(\mathbf{b}_{an}) - \varepsilon_n(\mathbf{r}|\mathbf{b}_{an})]. \tag{20}$$

The conditional decision rule at node n for each choice of ancestor bit vectors $\mathbf{b}_{an}$ corresponds to comparing the matched filter output l_n to a threshold.

$$l_n \underset{\hat{b}_n=-1}{\overset{\hat{b}_n=+1}{\gtrless}} \delta_n(\mathbf{b}_{an}) - \varepsilon_n(\mathbf{r}|\mathbf{b}_{an}). \tag{21}$$

The threshold on the right-hand side of (21) has both a deterministic component reflecting the influence of the hypothesized decisions at ancestor nodes

$$\delta_n(\mathbf{b}_{an}) = \mathbf{y}_{n,\,an}\mathbf{b}_{an} \tag{22}$$

[20] For the more general M-ary case, there would be $(M-1)$ comparisons.

[21] Dividing by 2 in (19) has no effect on the sign and is included to put the subsequent expressions into a form that can be compared with standard results.

and an adaptive component reflecting decision rules already constructed at descendant nodes

$$\begin{aligned}\varepsilon_n(\mathbf{r}|\mathbf{b}_{an}) = {} & \tfrac{1}{2}\mathbf{y}_{n,\,dn}[\hat{\mathbf{b}}_{dn}(\mathbf{r}|+1, \mathbf{b}_{an}) \\ & + \hat{\mathbf{b}}_{dn}(\mathbf{r}|-1, \mathbf{b}_{an})] \\ & + \tfrac{1}{4}[\hat{\mathbf{b}}_{dn}^T(\mathbf{r}|+1, \mathbf{b}_{an})\mathbf{Y}_{dn,\,dn}\hat{\mathbf{b}}_{dn}(\mathbf{r}|+1, \mathbf{b}_{an}) \\ & - \hat{\mathbf{b}}_{dn}^T(\mathbf{r}|-1, \mathbf{b}_{an})\mathbf{Y}_{dn,\,dn}\hat{\mathbf{b}}_{dn}(\mathbf{r}|-1, \mathbf{b}_{an})] \\ & + \tfrac{1}{2}[\mathbf{b}_{an}^T \mathbf{Y}_{an,\,dn} - \mathbf{1}_{dn}^T][\hat{\mathbf{b}}_{dn}(\mathbf{r}|+1, \mathbf{b}_{an}) \\ & - \hat{\mathbf{b}}_{dn}(\mathbf{r}|-1, \mathbf{b}_{an})].\end{aligned} \tag{23}$$

In particular, note that for nodes at the bottom level of the tree, there are *no* descendants and, consequently, $\varepsilon_n(\mathbf{r}|\mathbf{b}_{an}) = 0$. Hence, at the lowest level of the tree, the decision rules in (21) for each of the hypothesized set of values $\mathbf{b}_{an}$ correspond to comparing l_n to the fixed threshold given in (22). This nonzero threshold represents the adjustment of the test statistic to reflect the interference of users at ancestor nodes.

The calculation of $\varepsilon_n(\mathbf{r}|\mathbf{b}_{an})$, which is the adaptive portion of each threshold, has a child-separable structure.

$$\begin{aligned}\varepsilon_n(\mathbf{r}|\mathbf{b}_{an}) = {} & \eta_{cn_1}(\mathbf{r}|\mathbf{b}_{an}) + \eta_{cn_2}(\mathbf{r}|\mathbf{b}_{an}) \\ & + \cdots + \eta_{cn_{Kn}}(\mathbf{r}|\mathbf{b}_{an}).\end{aligned} \tag{24}$$

This is easy to see from the structure of $\mathbf{Y}_{dn,\,dn}$. Note the grouping structure of $\mathbf{S}_{dn} = [\mathbf{S}_{fcn_1}\ \mathbf{S}_{fcn_2} \cdots \mathbf{S}_{fcn_{Kn}}]$. That is, $\mathbf{S}_{dn}$ consists of orthogonal submatrices: one for each child and its descendants. Furthermore, for each child node cn_i, we have $\mathbf{S}_{fcn_i} = \{\mathbf{s}_{cn_i}\ \mathbf{S}_{dcn_i}\}$, where $\mathbf{S}_{dcn_i}$ consists of orthogonal submatrices. It is this nesting of orthogonal submatrices that gives $\mathbf{Y}_{dn,dn}$ a nested block diagonal structure that leads to the separation of $\varepsilon_n(\mathbf{r}|\mathbf{b}_{an})$ into terms.

In (24), the term $\eta_{cn_i}(\mathbf{r}|\mathbf{b}_{an})$ represents the contribution of the ith child of node n to the adaptive threshold at node n. Hence, the calculation of the adjustment ε_n may be done in parts: one for each child of node n. We use the family notation $fn = \{n, dn\}$ in showing the formula for the terms of (24).

$$\begin{aligned}\eta_n(\mathbf{r}|\mathbf{b}_{apn}) = {} & \tfrac{1}{2}\mathbf{y}_{pn,\,fn}[\hat{\mathbf{b}}_{fn}(\mathbf{r}|1, \mathbf{b}_{apn}) \\ & + \hat{\mathbf{b}}_{fn}(\mathbf{r}|-1, \mathbf{b}_{apn})] \\ & + \tfrac{1}{4}[\hat{\mathbf{b}}_{fn}^T(\mathbf{r}|1, \mathbf{b}_{apn})\mathbf{Y}_{fn,\,fn}\hat{\mathbf{b}}_{fn}(\mathbf{r}|1, \mathbf{b}_{apn})] \\ & - \hat{\mathbf{b}}_{fn}^T(\mathbf{r}|-1, \mathbf{b}_{apn})\mathbf{Y}_{fn,\,fn}\hat{\mathbf{b}}_{fn}(\mathbf{r}|-1, \mathbf{b}_{apn})] \\ & + \tfrac{1}{2}[\mathbf{b}_{apn}^T \mathbf{y}_{apn,\,fn} - \mathbf{1}_{fn}^T]\,[\hat{\mathbf{b}}_{fn}(\mathbf{r}|1, \mathbf{b}_{apn}) \\ & - \hat{\mathbf{b}}_{fn}^T(\mathbf{r}|-1, \mathbf{b}_{apn})].\end{aligned} \tag{25}$$

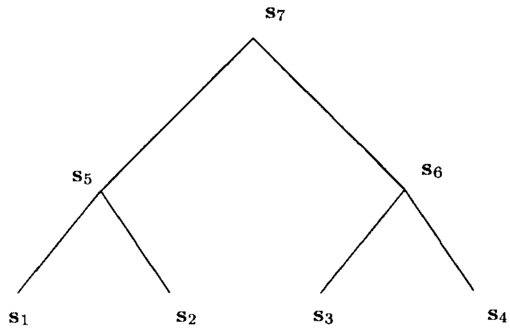


Fig. 3. Set of tree-structured signature vectors.

Implementation of the optimal decision rule may be organized as follows. Starting at the bottom of the tree and progressing to the top, construct augmented decision tables as illustrated in Table II.[22]

At each bottom node n, the conditional optimal decision $\hat{b}_n(\mathbf{r}|\mathbf{b}_{an})$ is computed by comparing l_n to the precomputed threshold $\delta_n(\mathbf{b}_{an})$. For each node at this level, calculate and store $\eta_n(\mathbf{r}|\mathbf{b}_{an})$ to be used at the next level. Move to the parent node pn, calculating the threshold for this node by starting with the precomputable portion $\delta_{pn}(\mathbf{b}_{apn})$ and adding to it the adjustments $\eta_n(\mathbf{r}|\mathbf{b}_{an})$ from each of the children of node pn. Compare l_{pn} to this threshold to make a decision.

For the root node n corresponding to the top level of the tree, $an = \emptyset$, and $\delta_n(\mathbf{b}_{an}) = 0$; there is a single threshold to be computed from the η's stored at the children of the root node.

C. A Binary Example

This procedure is illustrated for the simple signal set shown in Fig. 3. Consider node 1 at the lowest level. In this case, the table that is constructed for node 1 for user 1 is shown in Table III. Note that each value of $\eta_1(\mathbf{r}|b_7)$ depends on the two decisions $\hat{b}_1(\mathbf{r}|+1, +1)$ and $\hat{b}_1(\mathbf{r}|-1, +1)$.

$$\begin{aligned}\eta_1(\mathbf{r}|+1) = \tfrac{1}{2}y_{5,1}[&\hat{b}_1(\mathbf{r}|+1, +1) \\ &+ \hat{b}_1(\mathbf{r}|-1, +1)] \\ &+ \tfrac{1}{2}[y_{7,1} - l_1][\hat{b}_1(\mathbf{r}|+1, +1) \\ &- \hat{b}_1(\mathbf{r}|-1, +1)]\end{aligned} \tag{26}$$

and

$$\begin{aligned}\eta_1(\mathbf{r}|-1) = \tfrac{1}{2}y_{5,1}[&\hat{b}_1(\mathbf{r}|+1, -1) \\ &+ \hat{b}_1(\mathbf{r}|-1, -1)] \\ &+ \tfrac{1}{2}[-y_{7,1} - l_1][\hat{b}_1(\mathbf{r}|+1, -1) \\ &- \hat{b}_1(\mathbf{r}|-1, -1)].\end{aligned} \tag{27}$$

For example, suppose $y_{1,5} = y_{5,1} = 2$ and $y_{1,7} = y_{7,1} = 1$ and $l_1 = 5/2$. Table IV shows the values of the η_1's for this case. Similarly, tables are also constructed at the other bottom level nodes, 2–4.

[22] Notice that there are half as many values of η_n in the table as there are values of $\hat{b}_n$. Since there is one value of η_n for each value of $\mathbf{b}_{apn}$, we organize the values of $\mathbf{b}_{an}$ into pairs corresponding to $[\pm 1,\ \mathbf{b}_{apn}]$.

TABLE III
TABLE CREATED AT NODE 1 FOR THE EXAMPLE IN FIG. 3

$\mathbf{b}_{a1}^T = [b_5 b_7]$	$\hat{b}_1(\mathbf{r}\|\mathbf{b}_{a1})$	$\eta_1(\mathbf{r}\|b_7)$
$[+1\ \ +1]$	$\hat{b}_1(\mathbf{r}\|+1,\ +1) = sgn(l_1 - y_{1,5} - y_{1,7})$	$\eta_1(\mathbf{r}\|+1)$
$[-1\ \ +1]$	$\hat{b}_1(\mathbf{r}\|-1,\ +1) = sgn(l_1 + y_{1,5} - y_{1,7})$	
$[+1\ \ -1]$	$\hat{b}_1(\mathbf{r}\|+1,\ -1) = sgn(l_1 - y_{1,5} + y_{1,7})$	$\eta_1(\mathbf{r}\|-1)$
$[-1\ \ -1]$	$\hat{b}_1(\mathbf{r}\|-1,\ -1) = sgn(l_1 + y_{1,5} + y_{1,7})$	

TABLE IV
SPECIFIC INSTANCE OF THE TABLE AT NODE 1 OF OUR EXAMPLE

$\mathbf{b}_{a1}^T = [b_5 b_7]$	$\hat{b}_1(\mathbf{r}\|\mathbf{b}_{a1})$	$\eta_1(\mathbf{r}\|b_7)$
$[+1\ \ +1]$	-1	$3/2$
$[-1\ \ +1]$	1	
$[+1\ \ -1]$	1	2
$[-1\ \ -1]$	1	

TABLE V
TABLE CREATED AT NODE 5 IN OUR EXAMPLE

$\mathbf{b}_{a5}^T = b_7$	$\hat{b}_5(\mathbf{r}\|b_7)$	$\eta_5(\mathbf{r})$
$+1$	$\hat{\mathbf{b}}_{f5}(\mathbf{r}\|+1)$	$\eta_5(\mathbf{r})$
-1	$\hat{\mathbf{b}}_{f5}(\mathbf{r}\|-1)$	

Moving to the next level of the tree, consider node 5. The following conditional estimates are computed:

$$\begin{aligned}&\hat{b}_5(\mathbf{r}|+1) \\ &\quad = \text{sgn}\,[l_5 - y_{5,7} - \eta_1(\mathbf{r}|+1) - \eta_2(\mathbf{r}|+1)]\end{aligned} \tag{28}$$

and

$$\begin{aligned}&\hat{b}_5(\mathbf{r}|-1) \\ &\quad = \text{sgn}\,[l_5 + y_{5,7} - \eta_1(\mathbf{r}|-1) - \eta_2(\mathbf{r}|-1)]\end{aligned} \tag{29}$$

where $\eta_1(\mathbf{r}|+1)$ and $\eta_1(\mathbf{r}|-1)$ are the quantities in Table III for node 1. Similarly, $\eta_2(\mathbf{r}|+1)$ and $\eta_2(\mathbf{r}|-1)$ are the corresponding quantities that would be in the table for node 2. At this point, note that part of Table III and part of the corresponding table for node 2 may be discarded, and the remaining information may be consolidated into a single table for node 5. Specifically, suppose that $\hat{b}_5(\mathbf{r}|+1) = -1$. This implies that the best choice for b_5 is -1 if $b_7 = +1$. We may discard the first row of Table III since the first row corresponds to choosing $\hat{b}_5 = +1$ when $b_7 = +1$. Similarly, the analogous row of the table for node 2 may be discarded. That is, once the values in (28) and (29) have been computed, the following vectors may be assembled:

$$\hat{\mathbf{b}}_{f5}(\mathbf{r}|+1) = \begin{bmatrix}\hat{b}_5(\mathbf{r}|+1) \\ \hat{b}_1(\mathbf{r}|\hat{b}_5(\mathbf{r}|+1),\ +1) \\ \hat{b}_2(\mathbf{r}|\hat{b}_5(\mathbf{r}|+1),\ +1)\end{bmatrix},$$

$$\hat{\mathbf{b}}_{f5}(\mathbf{r}|-1) = \begin{bmatrix}\hat{b}_5(\mathbf{r}|-1) \\ \hat{b}_1(\mathbf{r}|\hat{b}_5(\mathbf{r}|-1),\ -1) \\ \hat{b}_2(\mathbf{r}|\hat{b}_5(\mathbf{r}|-1),\ -1)\end{bmatrix} \tag{30}$$

The table residing at node 5 (shown in Table V) may now be constructed.

Since node 7 has no ancestors, a single threshold correction $\eta_5(\mathbf{r})$ is calculated from (25) with $n = 5$ by dropping the last term since $ap5 = a7 = \emptyset$. The calculation of $\eta_5(\mathbf{r})$ from (25) would use the following substitutions:

$$\begin{aligned} \mathbf{y}_{pn,\,fn} &= \mathbf{y}_{7,\,f5} \\ &= [y_{7,\,5} \;\; y_{7,\,1} \;\; y_{7,\,2}], \\ \mathbf{Y}_{fn,\,fn} &= \mathbf{Y}_{f5,\,f5} \\ &= \begin{bmatrix} y_{5,\,5} & y_{5,\,1} & y_{5,\,2} \\ y_{5,\,1} & y_{1,\,1} & 0 \\ y_{5,\,2} & 0 & y_{2,\,2} \end{bmatrix}. \end{aligned}$$

Finally, at the top of the tree, since $an = \emptyset$, $\delta_7 = 0$, and the optimal decision rule at node 7 is

$$\hat{b}_7 = \operatorname{sgn}\left[l_7 - \eta_5(\mathbf{r}) - \eta_6(\mathbf{r})\right] \tag{31}$$

where $\eta_6(\mathbf{r})$ is computed in an analogous fashion to the computation of $\eta_5(\mathbf{r})$. Once we have $\hat{b}_7$, e.g., $\hat{b}_7 = +1$, the full optimal estimate may be read off the tables for nodes 5 and 6, e.g., $\hat{b}_{f5}(\mathbf{r}| + 1)$ and $\hat{b}_{f6}(\mathbf{r}| + 1)$.

V. Conclusion

This paper examines the problem of uncoded multiple access (MA) joint detection for the case in which user signatures are not orthogonal. The primary obstacle in a nonorthogonal MA system, however, is the complexity of detection; in general, the optimal detector has a complexity that is exponential in the number of users. The work presented in this paper proposes the use of signal design flexibility to allow a low complexity *optimal* algorithm for joint detection. Specifically, the user waveforms must have tree-structured interference. The tree structure is explained, and a low-complexity optimal tree detector is derived. The algorithm is pyramidal in that estimates and thresholds calculated at a given level of the tree are either discarded or used in the calculation of estimates at the next higher level of the tree. The tree detector gives the *optimal* estimate with an extremely *low* computational complexity. The complexity is derived and, for typical cases of interest, is bounded by a low-order polynomial in the number of users, e.g., $O(K^p)$, p small. This is an enormous savings in computations over the $O(M^K)$ computations needed if the signatures did not exhibit any structure.[23]

Since the tree detector is optimal for *any* set of tree-structured signatures, its use with the minimum distance sets proposed in [8] allows, in principle, for an oversaturated MA system that has comparable performance *and* computational complexity as its corresponding orthogonal MA system supporting less users. Our detection result has lifted a major computational obstacle and has opened up the area of oversaturated communications for more research.

[23] Here, it is assumed that M-ary PAM is used by the each of the K users.

Acknowledgment

The authors wish to thank Prof. S. Verdú for his most helpful discussion during the early stages of the hierarchical algorithm investigation. They are also thankful to Prof. M. Trott for his input on Section III-B. In addition, we thank the anonymous reviewers for their efforts and their helpful comments.

References

[1] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory,* vol. 35, Jan. 1989.
[2] S. Verdú, "Computational complexity of optimum multiuser detection," in *Algorithmica.* New York: Springer-Verlag, 1989.
[3] M. Varanasi and B. Aazhang, "Multistage detection in asynchronous code-division multiple-access communications," *IEEE Trans. Commun.,* vol. 38, Apr. 1990.
[4] A. Duel-Hallen, "Decision-feedback multiuser detector for synchronous code-division multiple access channel," *IEEE Trans. Commun.,* vol. 41, Feb. 1993.
[5] E. A. Lee and D. G. Messerschmitt, *Digital Signal Communication,* 2nd ed. Boston, MA: Kluwer, 1994.
[6] J. A. F. Ross and D. P. Taylor, "Vector assignment scheme for $M + N$ users in N-dimensional global additive channel," *Electron. Lett.,* vol. 28, Aug. 1992.
[7] J. A. F. Ross, "Multiple-user communications with nonorthogonal signaling," Ph.D. diessertation, McMaster Univ., Ont., Hamilton, Canada, Apr. 1994.
[8] J. A. F. Ross and D. P. Taylor, "Multiuser signaling in the symbol-synchronous AWGN channel," *IEEE Trans. Inform. Theory,* vol. 41, July 1995.
[9] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intel.,* vol. PAMI-11, July 1989.
[10] R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory,* vol. 38, Mar. 1992.
[11] G. W. Wornell, "Emerging application of multirate signal processing and wavelets in digital communications," *Proc. IEEE,* vol. 84, Apr. 1996.
[12] A. R. Lindsey and M. J. Medley, "Wavelet transforms and filter banks in digital communications," in *Proc. SPIE Mathematical Imaging: Wavelet Applications for Dual Use (2762-48),* Orlando, FL, Apr. 8–12, 1996, pp. 241–244.
[13] R. E. Learned, "Low complexity optimal joint detection for oversaturated multiple access communication," Ph.D. dissertation, Mass. Inst. Technol., Cambridge, Feb. 1997.
[14] A. Agarwal, "Wireless in local loop," *Telecommun.,* vol. 46, no. 3, June 1996.

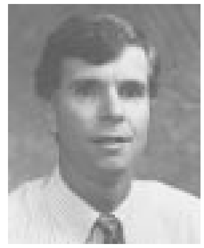
Rachel E. Learned (S'87) received the B.S. degree from Tufts University, Medford, MA, in 1990 and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in September 1992 and January 1997, respectively.

Since 1993, her research has been in the area of suboptimal and optimal multiple access joint detection for both CDMA and non-CDMA systems. Her Ph.D. work was funded by the MIT Lincoln Laboratory, Satellite Communications Technology Group. Between 1990 and 1992, she held a fellowship from the Charles Stark Draper Laboratory while working on her master's degree in the area of underwater acoustic transient signal classification. Both her Ph.D. and M.S. have been under the guidance of Prof. A. S. Willsky.

Alan S. Willsky (S'70–M'73–SM'82–F'86) received both the S.B. and the Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1969 and 1973, respectively.

He joined the MIT faculty in 1973, and his present position is Professor of Electrical Engineering. From 1974 to 1981, he served as assistant director of the MIT Laboratory for Information and Decision Systems. He is also a founder and member of the board of directors of Alphatech, Inc. His present research interests are in problems involving multidimensional and multiresolution estimation and imaging, discrete-event systems, and the asymptotic analysis of control and estimation systems.

In 1975, Dr. Willsky received the Donald P. Eckman Award from the American Automatic Control Council. He has held visiting positions at Imperial College, London, L'Universite de Paris-Sud, and the Institut de Recherche en Informatique et Systemes Aleatoires, Rennes, France. He was program chairman for the 17th IEEE Conference on Decision and Control, has been an associate editor of several journals including the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, has served as a member of the Board of Governors and vice president for Technical Affairs of the IEEE Control Systems Society, was program chairman for the 1981 Bilateral Seminar on Control Systems held in the People's Republic of China, and was Special Guest Editor of the 1992 Special Issue of the IEEE TRANSACTIONS ON INFORMATION THEORY on wavelet transforms and multiresolution signal analysis. In 1988, he was made a distinguished member of the IEEE Control Systems Society. In addition, he has given several plenary lectures at major scientific meetings including the 20th IEEE Conference on Decision and Control, the 1991 IEEE International Conference on Systems Engineering, the SIAM Conference on Applied Linear Algebra in 1991, and the 1992 Inaugural Workshop for the National Centre for Robust and Adaptive Systems, Canberra, Australia. He is the author of the research monograph *Digital Signal Processing and Control and Estimation Theory* and is coauthor of the widely used undergraduate text *Signals and Systems*. He was awarded the 1979 Alfred Noble Prize by the ASCE and the 1980 Browder J. Thompson Memorial Prize Award by the IEEE for a paper excerpted from his monograph.



Don M. Boroson (S'73–M'77) was born in Brooklyn, NY, on May 27, 1951. He received the B.S.E., M.S.E., and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 1973, 1974, and 1977, respectively.

From 1977 to 1986, he was a staff member in the Communications Division at the Massachusetts Institute of Technology Lincoln Laboratory, Lexington, where he worked in antenna nulling, narrowband modulation, analog signal processing, and spacecraft integration and testing. From 1986 to 1994, he was an assistant group leader of the Optical Communications Technology Group at Lincoln Laboratory, developing hardware and techniques related to intersatellite laser communications. Since 1994, he has been assistant group leader of the Satellite Communications Technology Group at Lincoln Laboratory, overseeing system engineering in military communications.