

A Novel Decoder for Unknown Diversity Channels Employing Space-Time Codes

Elona Erez

Department of Electrical Engineering-Systems, Tel Aviv University, Tel Aviv, 69978, Israel
Email: elona@eng.tau.ac.il

Meir Feder

Department of Electrical Engineering-Systems, Tel Aviv University, Tel Aviv, 69978, Israel
Email: meir@eng.tau.ac.il

Received 29 May 2001 and in revised form 11 January 2002

We suggest new decoding techniques for diversity channels employing space time codes (STC) when the channel coefficients are unknown to both transmitter and receiver. Most of the existing decoders for unknown diversity channels employ training sequence in order to estimate the channel. These decoders use the estimates of the channel coefficients in order to perform maximum likelihood (ML) decoding. We suggest an efficient implementation of the generalized likelihood ratio test (GLRT) algorithm that improves the performance with only slight increase in complexity. We also suggest an energy weighted decoder (EWD) that shows additional improvement without further increase in the computational complexity.

Keywords and phrases: maximum likelihood, generalized likelihood ratio test, Viterbi algorithm, recursive least square.

1. INTRODUCTION

Space time coding schemes have been shown to significantly improve the performance of communication over fading channels when multiple antennas are available at transmitter and receiver. Efficient space time codes (STC) for mobile cellular diversity channel are introduced by Tarokh et al. [1] where the channel coefficients are assumed to be known to the decoder. Many of the STC schemes assume full or partial receiver knowledge of the channel coefficients. When no knowledge of the channel is available, training sequences can be used in order to facilitate communication over an unknown channel. Since the sequence is known at the receiver, an estimate of the channel can be achieved at the receiver [2].

The use of training sequences has some drawbacks. First, there is a mismatch penalty. Because the training sequence is of limited length, the channel estimate formed at the receiver is imprecise which results in an increased error rate. Secondly, there is penalty in throughput, because the training sequence carries no information. This penalty is worse the longer the training sequence is, as compared to the length of the data sequence. When the channel changes rapidly over time, using training sequences might be inadequate. Since the channel can be assumed to be constant only over a very short period of time, training sequences would have to be

transmitted very often, which would severely increase the throughput penalty. In broadcast multipoint communication networks, if a channel from the master to one of the receiver stations goes down at any time following the initial training period and it is desired to retain only that receiver, the use of training sequences is unsuitable. In such situations, it is desirable for the receiver to decode without having a known training sequence available.

An efficient differential detection scheme which does not require training sequences and has a linear complexity was developed in [3]. The detection scheme was developed for a simple transmit encoding design, known as the Alamouti block coding, first introduced in [4]. The detection scheme in [3] requires equal-energy signal constellation such as PSK. A different approach which requires no pilot sequences is the unitary space-time modulation introduced in [5]. The decoder in [5] assumes a Rayleigh statistical model on the channel coefficients with a known covariance matrix. The encoding and decoding in [5] have exponential complexities.

From the error probability point of view the maximum-likelihood (ML) decoder is optimal for known channels. In the situation we are interested in this work, however, the channel coefficients are unknown and also do not have a known statistical model. A generalization of the ML decision rule for unknown channels is the generalized likelihood

ratio test (GLRT). Unlike the ML decoder for known channels, the GLRT is generally not optimal under the error probability criterion. In [6, 7] we develop a new decoder that improves the performance of the GLRT for unknown Gaussian linear channels. The improvement in [6, 7] is uniform and exponential, that is, for all the possible channel coefficients the error probability is either smaller by an exponential factor, or, only for some, remains the same. However, both the GLRT algorithm and the decoder we propose in [6, 7] have high computational complexity for practical implementation in space time systems. In this paper, we are interested in developing a computationally efficient decoder that could be applied to a large class of codes and on the other hand would assume no statistical model on the channel coefficients. To this end, we relax the complexity of the GLRT by using an appropriate approximation. We further suggest a new decoder that improves the performance of the approximated GLRT decoder for non equal-energy signal constellation while maintaining its low complexity.

The outline of the paper is as follows. In Section 2, the channel model is presented. In Section 3, an example of a space time system is given, later used for simulation. In Section 4, we introduce the GLRT for diversity channels and suggest an efficient approximation of the GLRT. We give motivation for a new decoder that improves the GLRT and propose the new energy weighted decoder (EWD), later in Section 4. In Section 5, a low complexity implementation of the approximated GLRT and the EWD is developed. In Section 6, simulation results for the performance of the new decoders will show that the GLRT decoder improves the existing decoder (that estimates the channel by the use of training sequences) and that the EWD shows further performance improvement. A summary and suggestions for further research will conclude the paper.

2. CHANNEL MODEL

The model of a complex diversity channel with L transmit antenna and J receive antenna is given by

$$\mathcal{H}_m : y_{n,j} = \sum_{i=1}^L \alpha_{i,j} x_{n,i}^{(m)} + z_{n,j}, \quad (1)$$

$$n = 0, 1, \dots, N-1, \quad m = 1, \dots, M,$$

where $\{y_{n,j}\}_{n=0}^{N-1}$ are the complex observed data samples at receive antenna j , $\{x_{n,i}^{(m)}\}_{n=0}^{N-1}$ are the complex symbols transmitted by the i th antenna for the m th codeword, $\alpha_{i,j}$ are the complex unknown fading coefficients from transmit antenna i to receive antenna j and $z_{n,j}$ are white noise samples modeled as independent zero-mean complex Gaussian random variables with variance σ^2 per dimension. We can write (1) as

$$\mathcal{H}_m : Y = X_m \alpha + Z, \quad (2)$$

where

$$X_m = \begin{bmatrix} x_{0,1}^{(m)} & x_{0,2}^{(m)} & \cdots & x_{0,L}^{(m)} \\ x_{1,1}^{(m)} & x_{1,2}^{(m)} & \cdots & x_{1,L}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-1,1}^{(m)} & x_{N-1,2}^{(m)} & \cdots & x_{N-1,L}^{(m)} \end{bmatrix},$$

$$\alpha = \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,J} \\ \alpha_{2,1} & \alpha_{2,2} & \cdots & \alpha_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{L,1} & \alpha_{L,2} & \cdots & \alpha_{L,J} \end{bmatrix}, \quad (3)$$

$$Y = \begin{bmatrix} y_{0,1} & y_{0,2} & \cdots & y_{0,J} \\ y_{1,1} & y_{1,2} & \cdots & y_{1,J} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-1,1} & y_{N-1,2} & \cdots & y_{N-1,J} \end{bmatrix},$$

$$Z = \begin{bmatrix} z_{0,1} & z_{0,2} & \cdots & z_{0,J} \\ z_{1,1} & z_{1,2} & \cdots & z_{1,J} \\ \vdots & \vdots & \ddots & \vdots \\ z_{N-1,1} & z_{N-1,2} & \cdots & z_{N-1,J} \end{bmatrix},$$

where matrices X_m , $m = 1, \dots, M$ are assumed to be full rank, as is the case for some of the coding methods we have encountered in the literature. For example, in [8] the matrix X_m has an orthogonal structure and in [5] the columns of X_m are designed to be (scaled) orthonormal.

3. AN EXAMPLE OF A SPACE-TIME CODE SYSTEM

3.1. The encoding algorithm

In this section we refer to the trellis based STC proposed by Tarokh et al. [1]. In [1], the channel coefficients were modeled as independent complex Gaussian random variables. This assumption is necessary for the criteria for optimal encoding design, but not for the decoding algorithm which is the main concern of our work. Therefore, we do not assume a statistical model of the channel coefficients until Section 6 for simulations.

In this example, a 16-QAM 16-state STC is used with two transmit antennas and two receive antennas ($L = 2, J = 2$). Figure 1 shows the 16-QAM constellation and the trellis description for this code. Each row in the matrix represents the edge labels for transitions from the corresponding state. At the beginning and end of each frame the encoder is required to be in zero state. At each time t , depending on the state of the encoder and the input bits a transition branch is chosen. If the label of this branch is $s_1 s_2$ then the s_1 is transmitted over the first transmit antenna and s_2 over the second one. In the general case, the label of each branch is $s_1 s_2 \cdots s_L$. We will illustrate the encoding with an example. Suppose the input stream to be encoded is

$$m = 0001101011010110. \quad (4)$$

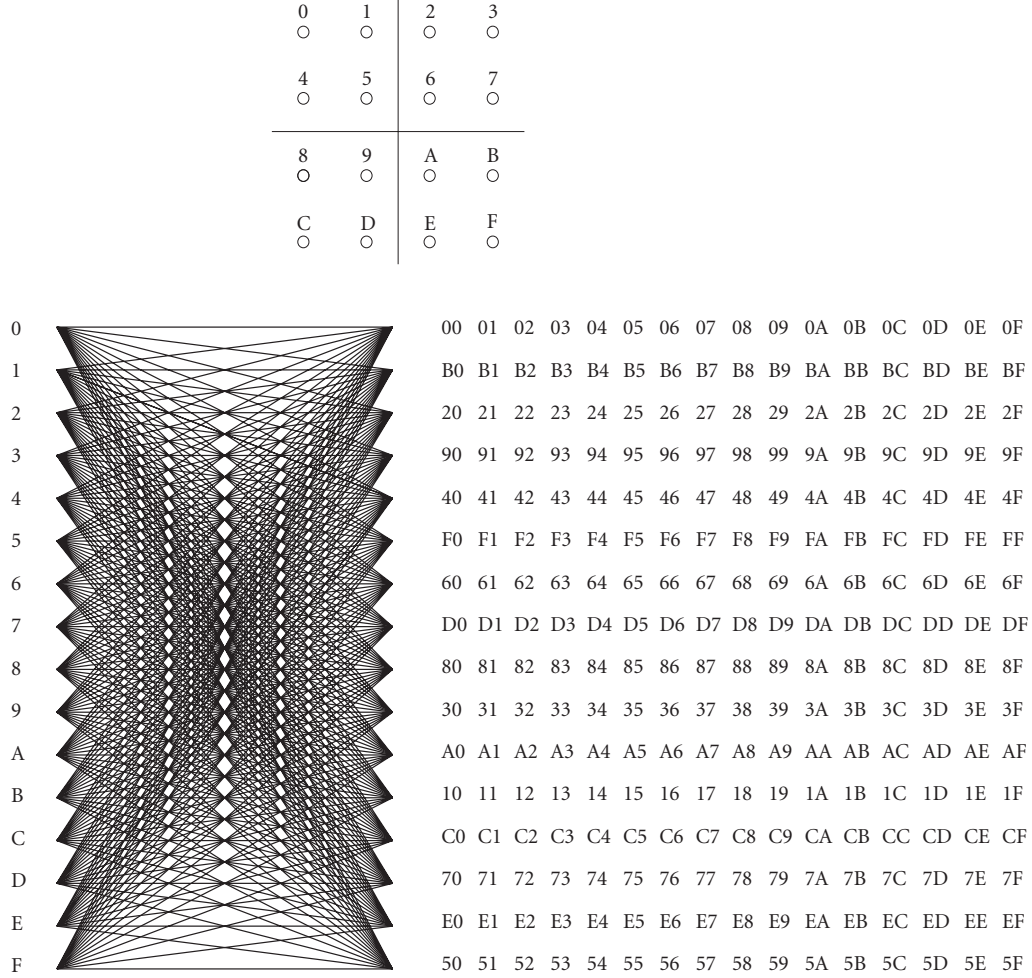


FIGURE 1: 16 QAM 16-state STC with 2 transmit antennas.

The bit stream is divided into groups of four bits and each group is mapped into one of sixteen constellation points. Thus the input to the encoder becomes

$$\mathcal{M} = \{1, A, D, 6\}. \quad (5)$$

The encoder is at the beginning in zero state. The first symbol to be encoded is 1. The encoder transits to state 1 in the trellis. The transmitted symbols $s_1 s_2$ are indicated in the zeroth row (current state) and first column (next state) of the matrix and therefore $s_1 s_2 = 01$. The first antenna transmits 0 and the second antenna transmits 1. Continuing the encoding process this way, the symbol sequences transmitted by the two antennas are

$$\mathcal{S}_1 = \{0, B, A, 7\}, \quad \mathcal{S}_2 = \{1, A, D, 6\}. \quad (6)$$

3.2. The decoding algorithm for known channels

In [1], the decoder is assumed to have full knowledge of the channel coefficients $\alpha_{i,j}$. When the channel is known, the optimal decoder uses the ML decision rule, which for Gaussian

channels selects the codewords according to the minimal Euclidean distance criterion. For the trellis STC described above, assuming that $y_{n,j}$ is the received signal at receive antenna j at time n , the branch metric for a transition labeled $s_1 s_2 \dots s_L$ is given by

$$\sum_{j=1}^J \left| y_{n,j} - \sum_{i=1}^L \alpha_{i,j} s_i \right|^2. \quad (7)$$

The Viterbi algorithm is then used to compute the path with the lowest accumulated metric.

3.3. Decoding algorithm for unknown channels

The more practical situation where the channel coefficients are unknown to the decoder is investigated by Naguib et al. [2]. The decoding method suggested in [2, 9] is to first estimate the channel using a training sequence. This is a mismatched decoder since the decoder operates according to the estimated channel and not according to the actual one. At the beginning of each frame, a sequence $\mathbf{W}_i$ of length k pilot

symbols is transmitted from transmit antenna i

$$\mathbf{W}_i = (W_{i,0}, W_{i,1}, \dots, W_{i,k-1}). \quad (8)$$

The sequences $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_L$ are designed to be orthogonal to each other

$$\mathbf{W}_p \cdot \mathbf{W}_q = \sum_{j=0}^{k-1} W_{p,j} \bar{W}_{q,j} = 0 \quad (9)$$

whenever $p \neq q$.

Let $\mathbf{y}_j = (y_{0,j}, y_{1,j}, \dots, y_{k-1,j})$ be the observed sequence of received signals at antenna j during the training period given by

$$y_{n,j} = \sum_{i=1}^L \alpha_{i,j} W_{i,n} + z_{n,j}, \quad 1 \leq j \leq L, \quad 0 \leq n \leq k-1. \quad (10)$$

The goal is to estimate $\alpha_{i,j}$, $i = 1, 2, \dots, L$, $j = 1, 2, \dots, J$ using the observed signals.

The least square (LS) estimate for $\hat{\alpha}_{i,j}$ is used, which is given by the inner product

$$\hat{\alpha}_{i,j} = \frac{\mathbf{y}_j \cdot \mathbf{W}_i}{\mathbf{W}_i \cdot \mathbf{W}_i}. \quad (11)$$

The estimated coefficient vector $\hat{\alpha}_{i,j}$ is then substituted in (7) instead of $\alpha_{i,j}$.

4. NOVEL DECODING METHODS

The classical decoding algorithm for unknown channels is the GLRT decision rule. This section begins by reviewing the GLRT decision rule in general and specifically for diversity channels. An approximation of the GLRT is suggested, which can be efficiently implemented. Later, we give the motivation for a new decoder that improves the GLRT and propose the new energy weighted decoder (EWD).

4.1. The generalized likelihood ratio test (GLRT)

Consider the family of channels specified by a set of channel laws,

$$\mathcal{F} = \{p_\theta(\mathbf{y}|\mathbf{x}), \theta \in \Theta\}, \quad (12)$$

where $p_\theta(\cdot|\cdot)$ denotes the conditional probability mass functions (PMF) governing the channel, which depends on the unknown and deterministic parameter $\theta \in \Theta$. For a known channel, the ML decoder is optimal in the sense of minimizing the probability of error (over messages). The decision rule of this decoder is given by

$$\phi(\mathbf{y}) = i \iff p_\theta(\mathbf{y}|\mathbf{x}^{(i)}) = \max_{1 \leq j \leq M} p_\theta(\mathbf{y}|\mathbf{x}^{(j)}), \quad (13)$$

where $\mathbf{x}^{(i)}$ is the i th codeword. Since the ML decision rule is in general different for different channels, it cannot be employed when the channel is unknown. A common approach in this case is the GLRT decoding rule. The GLRT decoder

can be defined as follows:

$$\phi_{\text{GLRT}}(\mathbf{y}) = i \iff \sup_{\theta \in \Theta} p_\theta(\mathbf{y}|\mathbf{x}^{(i)}) = \max_{1 \leq j \leq M} \sup_{\theta \in \Theta} p_\theta(\mathbf{y}|\mathbf{x}^{(j)}). \quad (14)$$

While the GLRT is intuitively appealing as joint channel and data estimation, it does not have optimal performance and lacks solid theoretical justification.

For Gaussian diversity channels the decoding criterion of the GLRT becomes

$$\begin{aligned} \hat{m} &= \arg \min_m \left\{ \min_\alpha \|Y - X_m \alpha\|^2 \right\} \\ &= \arg \min_m \left\{ \min_\alpha \sum_{n=0}^{N-1} \sum_{j=1}^J \left| y_{n,j} - \sum_{i=1}^L x_{n,i} \alpha_{i,j} \right|^2 \right\}, \end{aligned} \quad (15)$$

where Y , X_m , and α are defined in (2) and the norm of the matrix A is defined as

$$\|A\| = \left(\sum_i \sum_j |a_{i,j}|^2 \right)^{1/2}. \quad (16)$$

The LS solution for α yields

$$\hat{\alpha}_m = (X_m^H X_m)^{-1} X_m^H Y \quad (17)$$

and it can be easily derived that the closed form GLRT decision rule is

$$\hat{m} = \arg \max_m \left\{ Y^H X_m (X_m^H X_m)^{-1} X_m^H Y \right\}. \quad (18)$$

4.2. Approximated GLRT decoder

Since the decoding process in Section 3 involves the Viterbi algorithm, an approximated GLRT decoding that uses a sequential updated estimation of the channel coefficients can be incorporated into the existing decoder. The approximated GLRT decoder can be implemented, as we will show in Section 5 with only slight increase in complexity over the decoder for known channels. We define the matrices $X_m^0, X_m^1, \dots, X_m^{N-1}$ and the matrices $Y^0, Y^1, \dots, Y^{N-1}$ as

$$\begin{aligned} X_m^n &= \begin{bmatrix} x_{0,1}^{(m)} & x_{0,2}^{(m)} & \cdots & x_{0,L}^{(m)} \\ x_{1,1}^{(m)} & x_{1,2}^{(m)} & \cdots & x_{1,L}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1}^{(m)} & x_{n,2}^{(m)} & \cdots & x_{n,L}^{(m)} \end{bmatrix}, \\ Y^n &= \begin{bmatrix} y_{0,1} & y_{0,2} & \cdots & y_{0,L} \\ y_{1,1} & y_{1,2} & \cdots & y_{1,L} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,L} \end{bmatrix}. \end{aligned} \quad (19)$$

The LS solution of the following minimization problem:

$$\hat{\alpha}_m^n = \min_\alpha \|Y^n - X_m^n \alpha\| \quad (20)$$

is given by,

$$\hat{\alpha}_m^n = \left((X_m^n)^H X_m^n \right)^{-1} (X_m^n)^H Y^n, \quad (21)$$

where n is the update index. Note that $\hat{\alpha}_m^n$ can be computed at instant n without relying on future received signals.

The approximated GLRT criterion is

$$\hat{m} = \arg \min_m \left\{ \sum_{n=0}^{N-1} \sum_{j=1}^J \left| y_{n,j} - \sum_{i=1}^L x_{n,i} \hat{\alpha}_m^n \right|^2 \right\}. \quad (22)$$

Unlike the GLRT where the LS estimation of the channel coefficient matrix α is carried out over an entire codeword, here the estimated coefficients are updated at each time interval. The estimation of α at instant n depends only on present and past received signals.

4.3. Improving the GLRT: motivation

For a simple scalar case of fading (a single transmit antenna and a single receive antenna) a decoder that uniformly improves the GLRT is found in the appendix of [10] and is also suggested in [11]. That is, the new decoder improves the error probability performance for each choice of fading coefficients. The (real) fading channel model is

$$\mathbf{y} = \alpha \mathbf{x}^{(m)} + \mathbf{z}, \quad n = 0, 1, \dots, N-1, \quad m = 1, \dots, M, \quad (23)$$

where $\mathbf{y} = (y_0, y_1, \dots, y_{N-1})$ is the observed vector, $\mathbf{x}^{(m)}$ is the m th codeword, α is an unknown fading coefficient, and the components of the noise vector $\mathbf{z}$ are i.i.d. zero-mean, Gaussian random variables with variance σ^2 . Suppose that our codebook consists of two codewords ($M = 2$) of length N given by $\mathbf{x}^{(1)} = (a, 0, 0, \dots, 0)$ and $\mathbf{x}^{(2)} = (0, b, 0, \dots, 0)$. Any orthogonal code of two codewords can be transformed to this form. Since all of the coordinates of both codewords are zero for $n > 1$, the problem is actually two dimensional.

The decoding regions for the GLRT decoder appear in Figure 2. The GLRT projects the received signal (y_0, y_1) onto the directions of the two-dimensional vectors formed by the first two coordinates of $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, and decides according to the smaller between the distances of (y_0, y_1) to the vertical axis and to the horizontal axis of the coordinate system. The decoding rule decides $\mathbf{x}^{(1)}$ if $|y_0| \geq |y_1|$ and decides $\mathbf{x}^{(2)}$ if $|y_0| < |y_1|$. Thus, the boundaries between the two decision regions are straight lines through the origin at slopes of ± 45 degrees. Note that the decoding rule does not depend on the specific values of a and b . The distances of $\alpha \cdot (a, 0)$ and $\alpha \cdot (0, b)$ from the boundary lines dictate the error probability for the decoder. The distance d_1 of $\alpha \cdot (a, 0)$ from the boundary lines at slope ± 1 is $aa/\sqrt{2}$ and the distance d_2 of $\alpha \cdot (0, b)$ from the same lines is $ab/\sqrt{2}$. The probability of error is of the exponential order of $\exp\{-\alpha^2 \min\{a^2, b^2\}/(4\sigma^2)\}$.

The decoding regions of the new decoder suggested in [10] appear in Figure 3. This decoder projects the vector formed by the first two coordinates of each $\mathbf{x}^{(m)}$ in the direction of the first two coordinates of $\mathbf{y}$. The decoding rule decides $\mathbf{x}^{(1)}$ if $|y_0/a| \geq |y_1/b|$ and decides $\mathbf{x}^{(2)}$ if $|y_0/a| < |y_1/b|$.

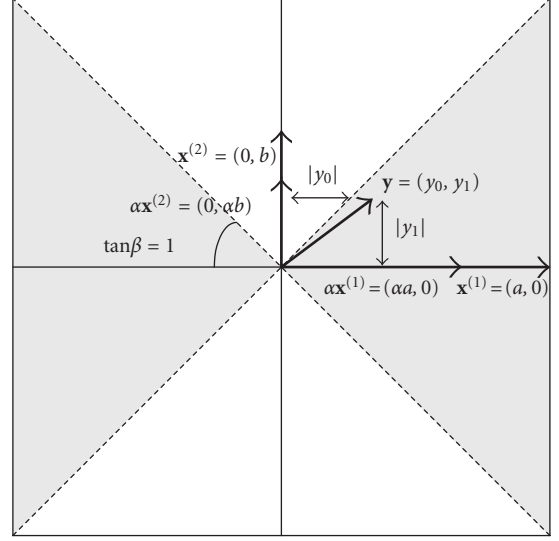


FIGURE 2: Signal space diagram of the GLRT decoder.

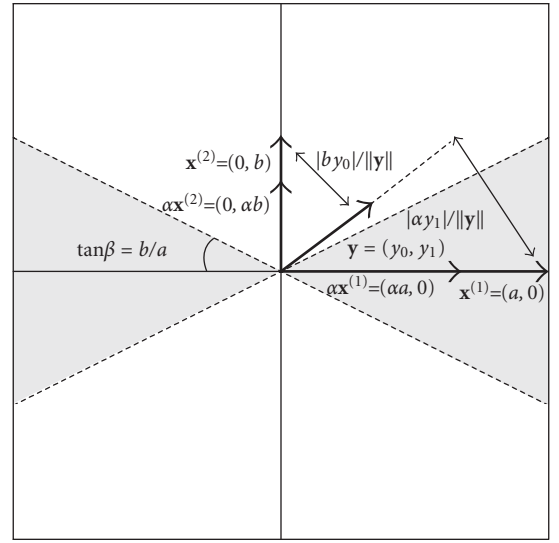


FIGURE 3: Signal space diagram of the new decoder.

In other words the new decoder decides $\mathbf{x}^{(1)}$ if $|b y_0| \geq |a y_1|$ and decides $\mathbf{x}^{(2)}$ if $|b y_0| < |a y_1|$. That is, the new decoder multiplies the metric of the GLRT for each codeword with the square root of the energy of that codeword. The boundary between the two decision regions is a pair of straight lines with slopes $\pm b/a$. For the new decoder the distances d_1, d_2 of $\alpha \cdot (a, 0)$ and $\alpha \cdot (0, b)$ from the boundary lines are both $aab/\sqrt{a^2 + b^2}$. The probability of error is of the exponential order of $\exp\{-\alpha^2 a^2 b^2/[2\sigma^2(a^2 + b^2)]\}$. Unless $a = b$, the exponential order of the new decoder is strictly better than that of the GLRT for any α . The improvement was achieved by maximizing the minimal of the distances d_1 and d_2 .

4.4. Energy weighted decoder

Motivated by the simple fading case with a single antenna in the transmitter and receiver we can derive a new decoder for STC over a diversity channel. Instead of (22) for the approximated GLRT decoder, the decision rule for the new decoder is

$$\hat{m} = \arg \min_m \left\{ \sum_{n=0}^{N-1} \left(\sum_{i=1}^L |x_{n,i}|^2 \right)^{1/2} \sum_{j=1}^J \left| y_{n,j} - \sum_{i=1}^L x_{n,i} \hat{\alpha}_m^n \right|^2 \right\}. \quad (24)$$

That is, the metric for each time interval is multiplied by the square root of the transmitted energy in that time interval, analogously to the new decoder suggested for the simple fading example.

Unlike the single fading coefficient case, for diversity channels this decoder does not generally improve the GLRT for all possible channel coefficient matrices α , as we have demonstrated in several examples of codebooks in [6]. Nevertheless, this decoder improves the GLRT in many cases. Indeed, our experimental results in Section 6 indicate the EWD outperforms the GLRT. An intuitive explanation for the EWD performance improvement is given as follows. For a certain decoder and a channel coefficient matrix, the error probability is given by

$$P(e|\alpha) = \frac{1}{M} \sum_{m=1}^M p(e|\alpha, m), \quad (25)$$

where α denotes the channel coefficients matrix and m denotes the index of transmitted codeword matrix $X^{(m)}$. The exponential order of the error probability is therefore

$$P(e|\alpha) \sim \max_m p(e|\alpha, m). \quad (26)$$

For the GLRT decoder the exponential order in (26) is determined, for most values of α , by the error probability of one of the less energetic codewords. Since the weighting of the EWD prefers selecting the less energetic codewords (in comparison to the GLRT), this compensation improves the exponential order in (26) for most values of α .

5. IMPLEMENTATION

When the encoder uses a trellis code, such as in the example in Section 2 the decoder uses the Viterbi algorithm. Both the approximated GLRT and the EWD can be incorporated with the Viterbi algorithms, using the recursive least square (RLS) algorithm.

At each step n of the algorithm we calculate the estimator $\hat{\alpha}_r^{(n-1) \rightarrow n}$ ($r = 1, \dots, T$, where T is the number of branches of the trellis at each interval) corresponding to each of the transitions from instant $n-1$ to instant n and the associated surviving sequences. The estimated channel coefficients can be substituted in the trellis metrics in (7). Each state at instant n has in addition to the cumulative metric and associated surviving sequence an associated estimated channel coefficient vector $\hat{\alpha}_r^n$.

In order to avoid increase in computational complexity we observe that the estimated coefficients $\hat{\alpha}_r^{(n-1) \rightarrow n}$, $n = 1, \dots, N-1$ do not have to be calculated separately for each n . The estimated coefficients can be updated via the recursive least square algorithm (RLS). We fully describe below the algorithm.

Step 0

At the beginning of the algorithm ($n = 0$) the decoder is at state 0. The decoder is supplied with an initial estimate of α , denoted as $\hat{\alpha}_0^0$. This estimate can be achieved, for example, by using a pilot sequence matrix with L (number of transmitters) rows. In this case we define

$$Q_0^0 = X_p^H X_p, \quad (27)$$

where X_p is the pilot sequence. When a pilot sequence is not available $\hat{\alpha}_0^0$ can be determined, for example, by using the most updated estimation of $\hat{\alpha}_0^0$ from the previous frame. When no pilot sequences are used we set $Q_0^0 = 0$ ($L \times L$ matrix).

Step n

In addition to the cumulative metric associated with each state in step $n-1$ and the surviving sequence, each state q is also associated with estimated coefficients $\hat{\alpha}_q^{n-1}$. Consider a transition r labeled $s_{n,1} s_{n,2} \dots s_{n,L}$. Define

$$\begin{aligned} \mathbf{s}_r^n &= [s_{n,1} \ s_{n,2} \ \dots \ s_{n,L}], \\ \mathbf{y}^n &= [y_{n,1} \ y_{n,2} \ \dots \ y_{n,J}]. \end{aligned} \quad (28)$$

(a) Updating the estimated channel coefficients

The decoder has to estimate the channel coefficients for each of the transitions from instant $n-1$ to instant n . For transition r emerging from state q to state p , $\hat{\alpha}_r^{n-1 \rightarrow n}$ is calculated by

$$\hat{\alpha}_r^{n-1 \rightarrow n} = \hat{\alpha}_q^{n-1} + K_r^n (\mathbf{y}^n - \mathbf{s}_r^n \hat{\alpha}_q^{n-1}), \quad (29)$$

where K_r^n is calculated according to

$$K_r^n = (Q_r^n)^{-1} \mathbf{s}_r^n \quad (30)$$

and where

$$Q_r^n = Q_r^{n-1} + (\mathbf{s}_r^n)^H \mathbf{s}_r^n. \quad (31)$$

Calculating (30) requires inversion of a $L \times L$ matrix. If L is large, the computational complexity can be decreased using the matrix inversion lemma

$$\begin{aligned} (Q_r^n)^{-1} &= (Q_r^{n-1})^{-1} - (Q_r^{n-1})^{-1} (\mathbf{s}_r^n)^H \\ &\quad \times [1 + \mathbf{s}_r^n (Q_r^{n-1})^{-1} (\mathbf{s}_r^n)^H]^{-1} \mathbf{s}_r^n (Q_r^{n-1})^{-1} \end{aligned} \quad (32)$$

which does not involve matrix inversion.

(b) Calculating the transition metric

Approximated GLRT:

$$\sum_{j=1}^J \left| y_{n,j} - \sum_{i=1}^L \hat{\alpha}_r^{(n-1) \rightarrow n} s_{n,i} \right|^2. \quad (33)$$

Energy Weighted Decoder:

$$\left(\sum_{i=1}^L |s_{n,i}|^2 \right)^{1/2} \sum_{j=1}^J \left| y_{n,j} - \sum_{i=1}^L \hat{\alpha}_r^{(n-1) \rightarrow n} s_{n,i} \right|^2. \quad (34)$$

The algorithm requires at each transition of the trellis only scalar inversions. The matrix multiplications in (29) is of order $L \times J$ which is the same order as the metric computation. The matrix multiplications in (30) and (31) or (32) are of order $L \times L$ (in practical systems $L \leq 16$ but most often $L = 2$), but do not depend on the received signal and therefore do not have to be computed in real-time. Therefore, the complexity of the approximated GLRT decoder is not significantly higher than that of the decoder designed for a known channel.

6. SIMULATION RESULTS

For the simulations, we used the 16-QAM 16-state STC described in Section 2 with two transmit antennas and two receive antennas. The channel coefficients were generated as samples of independent complex zero-mean Gaussian random variables with variance 0.5 per dimension. The frame of message to be encoded was 128 bits long. The channel coefficients were assumed to be constant during a frame period. The channel coefficients during different frame periods were assumed to be statistically independent (quasi-static conditions). The short pilot sequence was of two symbols only. The three decoding algorithms examined were the mismatched decoder described in Section 3.3, the approximated GLRT and the energy weighted decoder. The results of the simulations are shown in Figure 4. The ML decoder, employed when the channel is known to the receiver, is shown as a performance bound. The GLRT has about 2.5 dB advantage over the mismatched decoder and the EWD shows an additional advantage of 0.5 dB. Both algorithms allow communication at significantly lower SNR than the mismatched decoder for this QAM STC system.

7. CONCLUSIONS

In this paper, we have examined the problem of decoding for STC systems when the channel coefficients are unknown to the decoder. We have pointed out the difficulties which arise from employing pilot sequences, especially if those sequences are large in comparison to the frame length. These difficulties led us to suggest the GLRT decoder and its approximation for diversity channels. We later derived the EWD that improves the performance of the GLRT, motivated by the simple fading example in [10]. An efficient implementation of both

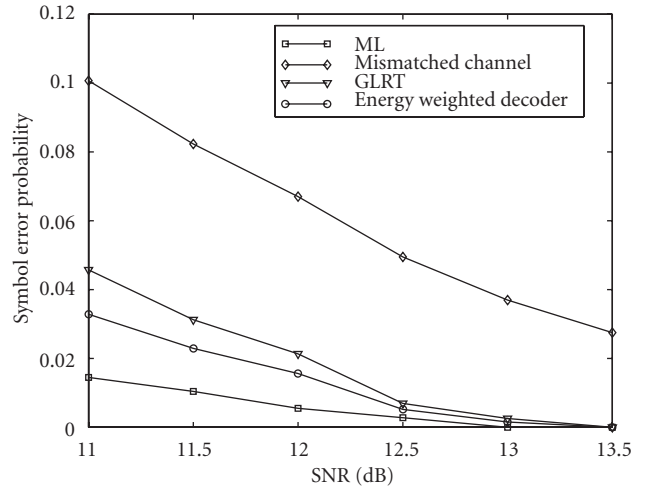


FIGURE 4: Symbol error probability for various decoders.

the approximated GLRT and the EWD was developed, with slight increase of complexity over the decoding scheme for known channels. Simulation results demonstrated the performance improvement of the approximated GLRT and the EWD. Since the EWD shows robust performance improvement for non equal-energy signal constellation, it may enable the usage of QAM even when the antenna fading gains are unknown.

One direction for future research is modifying our decoders to time variant channel model. A natural generalization of the GLRT and the EWD to time variant channels would modify the estimation of the channel coefficients involved in the algorithm. Instead of LS estimation it could involve the weighted least square (WLS) algorithm, where the weights are chosen to account for the changes in the channel.

The problem of the encoder design for unknown diversity channels can be investigated more closely in order to achieve a complete view of a universal communication system. A general discussion of robust communication for various classes of unknown channels can be found in [12], where both encoding and decoding methods were considered. The design criteria for good codes when the channel coefficients are assumed to have complex Gaussian distribution were investigated in [1]. Understanding the design criteria for good codes when the channel coefficients do not have a statistical model can be the subject of a future research.

REFERENCES

- [1] N. Tarokh, N. Seshardi, and A.R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.
- [2] A. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "A space-time coding modem for high-data-rate wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1459–1478, 1998.
- [3] V. Tarokh and H. Jafakhani, "A differential detection scheme for transmit diversity," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 7, pp. 1169–1174, 2000.

- [4] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.
- [5] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543–564, 2000.
- [6] E. Erez and M. Feder, "A universal decoder for unknown Gaussian linear channels," submitted to *IEEE Transactions on Information Theory*.
- [7] E. Erez, "A universal decoder for unknown Gaussian linear channels," M.S. thesis, Tel Aviv University, 2001.
- [8] V. Tarokh, H. Jafakhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.
- [9] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Trans. Communications*, vol. 47, no. 2, pp. 199–206, 1999.
- [10] M. Feder and N. Merhav, "Universal composite hypothesis testing—a competitive minimax approach," submitted to *IEEE Transactions on Information Theory*.
- [11] N. Merhav, "Universal decoding for memoryless Gaussian channels with deterministic interference," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1261–1269, 1993.
- [12] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2148–2177, 1998.

Elona Erez received the B.Sc. (summa cum laude) in electrical engineering in 2000 from Tel Aviv University, Israel. Currently, she is working on the M.Sc. thesis in the Department of Electrical Engineering-Systems, Tel Aviv University. Since 1998, she has been a teaching assistant in several courses in computer science, signal processing, and communications. Since 2000, she has been serving in the Israeli Air Force as a technical officer and an algorithm designer in the areas of electronic warfare, parameter estimation and signal processing. Her main research interests are universal encoding and decoding (communication in unknown channels), space-time codes and network information theory.



Meir Feder received the B.Sc. and M.Sc. degrees from Tel Aviv University, Israel and the Sc.D. degree from the Massachusetts Institute of Technology (MIT) Cambridge, and the Woods Hole Oceanographic Institution, Woods Hole, Mass, all in electrical engineering in 1980, 1984, and 1987, respectively. After being a research associate and lecturer in MIT he joined in 1989 the Department of Electrical Engineering-Systems, Tel Aviv University, where he is now a Professor. He also had visiting appointments at the Woods Hole Oceanographic Institution, Scripps Institute, and Bell laboratories. In the 1995/1996, he spent a sabbatical year as a visiting professor at the Electrical Engineering and Computer Science Department and the laboratory of information and decision systems in MIT. In addition to his academic activities, he is extensively involved in the high-tech industry. He was a co-founder of Peach Networks that developed a unique



server-based interactive TV solution. Peach Networks was acquired in March 2000 by Microsoft, and following that, he is currently a consultant to MicrosoftTV. Prof. Feder is a recipient of the 1993 IEEE Information Theory Best Paper Award. He also received the 1978 "creative thinking" award of the Israeli Defense Forces, the 1994 Tel Aviv University prize for Excellent Young Scientists, the 1995 Research Prize of the Israeli Electronic Industry, and the research prize in applied electronics of the Ex-Serviceman Association, London, awarded by Ben-Gurion University. He is a Fellow of IEEE. Between June 1993–June 1996 he served as an Associate Editor for Source Coding of the IEEE Transactions on Information Theory.