# Union Bound for Linear Space-Time Codes

Sumeet Sandhu
Packard 225
Stanford University
Stanford, CA 94305-9510
sumeet@stanford.edu

A. Paulraj
Packard 272
Stanford University
Stanford, CA 94305-9510
paulraj@rascals.stanford.edu

October 6, 2000

**Abstract**

Design of practical coding and modulation techniques for the multiple antenna fading wireless channel is a challenging problem. A number of interesting solutions have been proposed recently ranging from block codes to trellis codes for the MIMO (multiple input, multiple output) fading channel. We address the general problem of linear code design for the quasi-static, flat-fading, coherent MIMO channel. A *linear code* refers to an encoder that is linear with respect to scalar input symbols. The decoder performs ML (maximum likelihood) decoding of the received matrix symbols and is not assumed to be linear.

We provide a cohesive framework for analysis of linear codes in terms of a union bound on the *conditional* probability of symbol error. The error bound is conditioned on the channel realization and does not make any assumptions on channel statistics. Our analysis incorporates all existing linear spatial modulation techniques such as spatial multiplexing, space-time block codes, and truncated delay diversity. We show that space-time block codes achieve the lowest error bound among all orthogonal codes and are in fact optimal.

## 1 Introduction

A number of interesting coding and modulation techniques addressing the problem of transmit diversity design for the multiple-antenna fading channel have been proposed recently. They range from spatial multiplexing [1] to space-time block codes [2] to space-time trellis codes [3]. Among these schemes, the first two are linear and the third is nonlinear, where linearity is with respect to the scalar input symbols that are to be mapped to space-time matrix codewords.

In this paper we address the general problem of linear code design. We define a *linear code* as one that has a linear encoder, i.e., the encoder uses a set of modulation matrices to modulate the sequence of scalar input symbols and adds up the products to create the matrix output codeword. The output codeword is a matrix with dimensions representing transmit antennas and time. The decoder has perfect channel knowledge and performs ML decoding of the matrix symbols, and is not assumed to be linear or otherwise constrained.

We analyze all linear codes within the framework of probability of (matrix) symbol error. We obtain an upper bound on the error probability using the union bound on probabilities. The upper bound is *conditioned on the instantaneous channel realization.* It does not depend on channel statistics and is expected to be tight at high SNRs (signal to noise ratios). Focusing on the class of *orthogonal linear codes*, i.e., where the modulation matrices are orthogonal, we obtain *necessary and sufficient conditions for minimization* of the error bound. In the absence of any channel knowledge at the transmitter these conditions are met by the STBC (space-time block coding) solution. Our analysis does not presuppose decoupled detection as done in [4], leading to it instead in a natural fashion. Our error bound can be used to analyze any linear code, and work is in progress to obtain minimization conditions for all linear codes.

## 2 Data Model and Notation

Consider a system with $M_r$ receive antennas and $M_t > 1$ transmit antennas. The channel is flat-fading and quasi-static. It is unknown at the transmitter but is known at the receiver. At time $nL$, the channel output corresponding to the $n^{th}$ input block spanning $L$ symbol times is

$$\mathbf{Y}_{nL} \;=\; \mathbf{H}\mathbf{X}_{nL} + \mathbf{V}_{nL} \tag{1}$$

where the received signal $\mathbf{Y}_{nL}$ is $M_r \times L$, the fading channel $\mathbf{H}$ is $M_r \times M_t$, the encoded codeword $\mathbf{X}_{nL}$ is $M_t \times L$, and receiver noise $\mathbf{V}_{nL}$ is $M_r \times L$. The entries of $\mathbf{V}_{nL}$ are i.i.d.(independent, identically distributed), $v_{nL_{i,j}} \sim \aleph_c(0, N_0)$, and independent over $n$. The average power transmitted on $M_t$ antennas is $E_s$ per symbol time. Define $S = \frac{E_s}{N_0}$.

In our notation $\mathbf{A}^T$ denotes the transpose, $\mathbf{A}^*$ denotes the Hermitian, and $\mathrm{Tr}\mathbf{A}$ denotes the trace of matrix $\mathbf{A}$ respectively, and Re denotes the real part. The complementary error function integral is defined as $\mathrm{Q}(x) = \int_x^\infty \frac{dt}{\sqrt{2\pi}} \exp(-\frac{t^2}{2})$.

## 3 Encoder and Decoder

Following the linear modulation structure in [4], we consider codewords that consist of a set of modulation matrices modulated by scalar input symbols. The input to the encoder is a stream of symbols from a constellation such as PAM, QAM or PSK. Each complex symbol is expanded into two real symbols corresponding to its real and imaginary parts respectively. The encoder operates on the sequence of real symbols producing a matrix codeword whose rows correspond to antennas and columns correspond to symbol times. We define linear codes as follows.

**Definition 1 (Linear codes)** *A linear code is defined as a set of codewords that are linear in the scalar input symbols. Let $x^{(i)}$ and $x^{(j)}$ be two input symbol sequences with output codewords equal to $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$ respectively. Then the codeword corresponding to the sequence $\alpha x^{(i)} + \beta x^{(j)}$ is equal to $\alpha\mathbf{X}^{(i)} + \beta\mathbf{X}^{(j)}$, where $\alpha, \beta$ are any complex scalars. The decoder is the optimal decoder, i.e., it performs ML detection on the observations $\mathbf{Y}$.*

At time $nL$, the $m^{th}$ block of $K$ input symbols $\{x_{mK-k+1}\}_{k=1}^{K}$ is used to modulate a set of matrices $\{\mathbf{A}_k\}_{k=1}^{K}$ which may be real or complex. The block indices $n$ and $m$ may be different depending on the size $M_t \times L$ of matrices used to encode $K$ symbols. For e.g., $K \leq L$ for space-time block codes and $K > L$ for spatial multiplexing. The modulated matrices are summed to obtain the $M_t \times L$ codeword $\mathbf{X}_{nL} = \sum_{k=1}^{K} \mathbf{A}_k x_{mK-k+1}$, each row of which is then low-pass filtered, D/A sampled, upconverted and transmitted on the corresponding antenna. To simplify notation, we shall drop the time indices and consider the codebook $\mathcal{X} = \{\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(R)}\}$ where the $r^{th}$ codeword $\mathbf{X}^{(r)}$ is constructed from the $r^{th}$ sequence of input symbols $x^{(r)} = \{x_1^{(r)}, \ldots, x_K^{(r)}\}$ as follows

$$\mathbf{X}^{(r)} = \sum_{k=1}^{K} \mathbf{A}_k x_k^{(r)} \tag{2}$$

Depending on the input symbol constellation, each real symbol $x_k^{(r)}$ belongs to a PAM, one-dimensional "QAM" or one-dimensional "PSK" constellation of size M. The average energy of the PAM constellation is $E_s$ but the average energy of the real and imaginary symbols from the QAM and PSK constellations is $\frac{E_s}{2}$ each. The total number of codewords is $R = |\mathcal{X}| = \mathrm{M}^K$. We assume no additional outer coding on the input symbols in the subsequent analysis and comment on the effect of coded symbols in the conclusion. Since we consider the real and imaginary parts of a complex input symbol as two separate input symbols, the codeword in (2) is *linear in the input symbols* over the complex field.

We impose a power constraint on each modulation matrix by requiring $\|\mathbf{A}_k\|_F^2 \leq c$ for all $k$. This ensures that each codeword is bounded in power as follows

$$\begin{aligned} \|\mathbf{X}\|_F^2 &= \|\sum_{k=1}^{K} \mathbf{A}_k x_k\|_F^2 = \sum_{i=1}^{M_t}\sum_{j=1}^{L} |\sum_{k=1}^{K} \mathbf{A}_{k[i,j]} x_k|^2 \\ &\leq \sum_{i=1}^{M_t}\sum_{j=1}^{L}\sum_{k=1}^{K} |\mathbf{A}_{k[i,j]} x_k|^2 = \sum_{k=1}^{K} |x_k|^2 \sum_{i=1}^{M_t}\sum_{j=1}^{L} |\mathbf{A}_{k[i,j]}|^2 \qquad (3) \\ &= \sum_{k=1}^{K} |x_k|^2 \|\mathbf{A}_k\|_F^2 \leq c\sum_{k=1}^{K} |x_k|^2 \qquad (4) \end{aligned}$$

where (3) follows from the Cauchy-Schwartz inequality, and (4) follows from the power constraint $\|\mathbf{A}_k\|_F^2 \leq c$. From (4) we observe that the average power of the matrix codewords is at most $\frac{cK}{L} E_s$ per symbol time for PAM input constellations because

$$\frac{1}{L}\mathrm{E}\|\mathbf{X}\|_F^2 \leq \frac{c}{L}\mathrm{E}\sum_{k=1}^{K} |x_k|^2 = \frac{cK}{L}\mathrm{E}|x_k|^2 = \frac{cK}{L} E_s$$

In order to maintain the average power at $E_s$ per symbol time, we must normalize the modulation matrices for different input constellations as follows

$$\begin{aligned} &\text{PAM} && \|\mathbf{A}_k\|_F^2 \leq \frac{L}{K} \text{ for all } k \\ &\text{QAM , PSK} && \|\mathbf{A}_k\|_F^2 \leq \frac{2L}{K} \text{ for all } k \end{aligned} \tag{5}$$

We assume a coherent receiver, i.e., the channel is perfectly known. The received $M_r \times L$ signal $\mathbf{Y}$ is decoded using maximum likelihood decoding. Some practical examples of linear codes are presented in the next section to clarify concepts.

# 4 Examples of linear codes

In order to clarify the concept of linear codes, we provide a few examples in this section. Most of the current spatial modulation techniques can be considered as linear codes, the exceptions being space-time trellis codes.

**Example 1 (Spatial multiplexing)** *Spatial multiplexing [1], also called BLAST [5], is the simplest example of linear codes. Each incoming symbol is transmitted once on one antenna only. The modulation matrices are simply unit vectors, i.e., the $k^{th}$ matrix is an $M_t \times 1$ vector that is nonzero in the $k^{th}$ position. For e.g., for a PAM input constellation, $A_k = \frac{1}{\sqrt{M_t}}[0 \cdots 1 \cdots 0]^T$. For $M_t$ transmit antennas, $K = M_t$ and $L = 1$.*

**Example 2 (Space-time block coding)** *Space-time block coding is another example of linear codes. The modulation matrices are unitary matrices that are pairwise orthogonal with respect to the* Re Tr *matrix inner product. That is, for any weighting matrix $\mathbf{W}$,* $\mathrm{Re}\ \mathrm{Tr}(\mathbf{A}_k^* \mathbf{W} \mathbf{A}_l) = 0$ *for all $1 \leq k \neq l \leq K$. Such matrices exist for limited values of $K$, $M_t$ and $L$ [2]. For e.g., the complex, full rate Alamouti code [6] has $M_t = 2$ transmit antennas, $K = 4$ real symbols (from 2 complex input symbols), and $L = 2$. The two real, full rate, minimum-delay Tarokh codes have $M_t = K = L = 4$ and 8 respectively.*

**Example 3 (Delay diversity)** *Delay diversity [7, 8] can be considered as a linear code if the tail effects are ignored ("truncated" delay diversity). The $M_t \times L$ modulation matrices are proportional to $[0 I_{M_t} 0]$, where the order $M_t$ identity matrix is shifted right by $k-1$ columns in the $k^{th}$ modulation matrix. For any $L \geq M_t$ the total number of scalar symbols thus encoded is $K = L - M_t + 1$. In order that the loss in rate by ignoring the tail symbols be negligible, $L$ must be large.*

**Example 4 (Cyclic delay diversity)** *We propose a* new *class of linear codes called cyclic delay diversity. Instead of being dropped altogether, the tail symbols transmitted at the beginning of the block are repeated at the end of the block. This results in cyclic modulation matrices and an equivalent channel that is circulant. Similar to OFDM (Orthogonal Frequency Division Multiplexing) over Toeplitz channel matrices, encoding/decoding becomes much simpler because IFFT (Inverse Fast Fourier Transform) transmit matrices and FFT receive matrices diagonalize the equivalent circulant channel. Performance analysis of this scheme is currently in progress.*

# 5 Performance Analysis

We will analyze linear codes using the conditional probability of (matrix) symbol error, i.e., the error probability based on a given realization of the channel which is known at the receiver. First we compute an upper bound on the probability of symbol error in Lemma 1, then we set up conditions for its minimization in Lemmas 2 and 3. We state necessary and sufficient conditions for minimization over orthogonal codes in Lemma 4. Finally in Lemma 5 we show that the optimal linear code is the space-time block code, which is stated as the main result of the paper in Theorem 1.

**Lemma 1** *The probability of matrix symbol error can be bounded in the following manner*

$$P_e \;\leq\; \sum_{i=1}^{R} p_i \sum_{j\neq i}^{R} \mathrm{Q}\left(\sqrt{\Delta_1^{(ij)} + \Delta_2^{(ij)}}\right) \;=\; P_U \tag{6}$$

*where the argument of each Q-function is a function of two terms, one of which* $\Delta_1^{(ij)}$ *is determined by the norms of individual modulation matrices and the other* $\Delta_2^{(ij)}$ *is determined by their pairwise inner products.*

*Proof:* Lemma 1 is proved by construction. The probability of matrix symbol error is the average of the pairwise error probabilities over all matrix symbols as follows

$$P_e \;=\; \sum_{i=1}^{R} p_i P_{e|i} \tag{7}$$

where $p_i$ is the probability that codeword $i$ is transmitted, and $P_{e|i}$ is the probability that the receiver does not decode $i$ correctly. The probability of detecting $i$ incorrectly is equal to the probability that one of the other codewords $j$ is detected where $j \neq i$. Using the union bound on probabilities, $P_{e|i}$ can be bounded as follows

$$P_{e|i} \;\leq\; \sum_{j\neq i}^{R} \mathrm{Q}\left(\sqrt{D_{ij}S/2}\right). \tag{8}$$

For a given channel realization $\mathbf{H}$, $D_{ij}$ is the pairwise Euclidean distance at the receiver defined as

$$D_{ij} \;=\; \|\mathbf{H}\,(\mathbf{X}^{(i)} - \mathbf{X}^{(j)})\|_F^2 \;=\; \|\mathbf{H}\sum_{k=1}^{K}\mathbf{A}_k(x_k^{(i)} - x_k^{(j)})\|_F^2 \;=\; \|\mathbf{H}\sum_{k=1}^{K}\mathbf{A}_k\delta_k^{(ij)}\|_F^2 \tag{9}$$

where $\delta_k^{(ij)} = x_k^{(i)} - x_k^{(j)}$ is the difference between sequences $i$ and $j$ at the $k^{th}$ position. It can be further rewritten as

$$\begin{aligned}
D_{ij} \;&=\; \mathrm{Tr}\left(\left[\mathbf{H}\sum_{k=1}^{K}\mathbf{A}_k\delta_k^{(ij)}\right]\left[\sum_{l=1}^{K}\mathbf{A}_l^*\mathbf{H}^*(\delta_l^{(ij)})^*\right]\right) &(10)\\
&=\; \mathrm{Tr}\left(\sum_{k=1}^{K}\sum_{l=1}^{K}\mathbf{H}\mathbf{A}_k\mathbf{A}_l^*\mathbf{H}^*\delta_k^{(ij)}(\delta_l^{(ij)})^*\right) &(11)\\
&=\; \mathrm{Tr}\left(\sum_{k=1}^{K}\mathbf{H}\mathbf{A}_k\mathbf{A}_k^*\mathbf{H}^*\delta_k^{(ij)}(\delta_k^{(ij)})^*\right) + \mathrm{Tr}\left(\sum_{k\neq l}^{K}\mathbf{H}\mathbf{A}_k\mathbf{A}_l^*\mathbf{H}^*\delta_k^{(ij)}(\delta_l^{(ij)})^*\right) &(12)\\
&=\; \sum_{k=1}^{K}\|\mathbf{H}\mathbf{A}_k\|_F^2|\delta_k^{(ij)}|^2 + 2\sum_{k<l}^{K}\mathrm{Re}\;\mathrm{Tr}\left(\mathbf{H}\mathbf{A}_k\mathbf{A}_l^*\mathbf{H}^*\delta_k^{(ij)}(\delta_l^{(ij)})^*\right) &(13)
\end{aligned}$$

Here (10) follows from (9) because $\|\mathbf{H}\|_F^2 = \mathrm{Tr}(\mathbf{H}\mathbf{H}^*)$ [9], (11) and (12) are simple rearrangments of the terms in (10), and (13) uses the fact that $a + a^* = 2\mathrm{Re}(a)$. Invoking the fact that the input symbols and their differences $\delta_k^{(ij)}$ are real, we can write $D_{ij}$ as

$$D_{ij} = \sum_{k=1}^{K} \Omega_{kk} |\delta_k^{(ij)}|^2 + 2\sum_{k<l}^{K} \Omega_{kl} \delta_k^{(ij)} \delta_l^{(ij)} \tag{14}$$

Using the property that $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$ [9], $\Omega_{kl} = \mathrm{Re}\ \mathrm{Tr}(\mathbf{H}\mathbf{A}_k\mathbf{A}_l^*\mathbf{H}^*)$ is the symmetric, weighted, matrix inner product of $\mathbf{A}_k$ and $\mathbf{A}_l$ defined as

$$\Omega_{kl} = <\mathbf{A}_k, \mathbf{A}_l>_{\mathbf{H}^*\mathbf{H}} = \frac{1}{2}(\mathrm{Tr}(\mathbf{A}_k^*\mathbf{H}^*\mathbf{H}\mathbf{A}_l) + \mathrm{Tr}(\mathbf{A}_l^*\mathbf{H}^*\mathbf{H}\mathbf{A}_k)) = \mathrm{Re}\ \mathrm{Tr}(\mathbf{A}_k^*\mathbf{H}^*\mathbf{H}\mathbf{A}_l)$$

where the weighting matrix is $\mathbf{H}^*\mathbf{H}$. Note that $\Omega_{kk} = \|\mathbf{H}\mathbf{A}_k\|_F^2 \geq 0$. Define the following

$$\begin{aligned} \Delta_1^{(ij)} &\triangleq \frac{S}{2} \sum_{k=1}^{K} \Omega_{kk} |\delta_k^{(ij)}|^2 = \frac{S}{2} \sum_{k=1}^{K} \|\mathbf{H}\mathbf{A}_k\|_F^2 |\delta_k^{(ij)}|^2 , \\ \Delta_2^{(ij)} &\triangleq \frac{S}{2}\, 2\sum_{k<l}^{K} \Omega_{kl} \delta_k^{(ij)} \delta_l^{(ij)} = S \sum_{k<l}^{K} \mathrm{Re}\ \mathrm{Tr}(\mathbf{A}_k^*\mathbf{H}^*\mathbf{H}\mathbf{A}_l) \delta_k^{(ij)} \delta_l^{(ij)} \end{aligned} \tag{15}$$

Of these, $\Delta_1^{(ij)}$ is a function of the weighted norms of individual modulation matrices, and $\Delta_2^{(ij)}$ is a function of the weighted pairwise inner products of the modulation matrices.

Substituting (15) in (13), (8) and then in (7) we get

$$P_e \leq \sum_{i=1}^{R} p_i \sum_{j\neq i}^{R} \mathrm{Q}\left(\sqrt{D_{ij}S/2}\right) = \sum_{i=1}^{R} p_i \sum_{j\neq i}^{R} \mathrm{Q}\left(\sqrt{\Delta_1^{(ij)} + \Delta_2^{(ij)}}\right) = P_U \tag{16}$$

which finishes the proof of Lemma 1. ■

The goal of code design is to find modulation matrices $\{\mathbf{A}_k\}_{k=1}^K$ that minimize the upper bound on the conditional probability of symbol error $P_U$. The complementary error function (Q-function) in (16) is a monotonically decreasing function of its argument. In order to minimize $P_U$, we must maximize the argument of each of the Q-functions. From (16), the argument of each Q-function $\sqrt{D_{ij}S/2}$ is a function of two terms, one of which is determined by the norms of individual modulation matrices $\Omega_{kk}$ and the other is determined by their pairwise inner products $\Omega_{kl}$. If we make the assumption that all sequences are equally likely, i.e., $p_i = \frac{1}{R}$ for all $i$ in (16), then we have Lemma 2.

**Lemma 2** *: By carefully selecting terms over $i$ and $j$, we can always group terms in the expression for the upper bound on the probability of symbol error into pairs of the following form*

$$q = \frac{n}{R}\left[\mathrm{Q}\left(\sqrt{\Delta_1 + \Delta_2}\right) + \mathrm{Q}\left(\sqrt{\Delta_1 - \Delta_2}\right)\right] \tag{17}$$

*where $\Delta_1 \geq \Delta_2 \geq 0$ and $n$ is an integer denoting the number of such pairs.*

*Proof:* Recall that the $i^{th}$ input sequence consists of $K$ symbols, i.e., $x^{(i)}=\{x_1^{(i)},\ldots,x_K^{(i)}\}$. The difference between $x^{(i)}$ and $x^{(j)}$ is the sequence $\delta^{(ij)}=\{\delta_1^{(ij)},\ldots,\delta_K^{(ij)}\}$, where $\delta_k^{(ij)}=x_k^{(i)}-x_k^{(j)}$. When two symbol sequences differ only in $m$ positions, say $k_1,\ldots,k_m$, then $\delta_k^{(ij)}\neq 0$ for $k=k_1,\ldots,k_m$, and $\delta_k^{(ij)}=0$ for $k\neq k_1,\ldots,k_m$. Let $\delta_{k_i}^{(ij)}=\delta_i\neq 0$ for $1\leq i\leq m$. The range of values of $\delta_i$ is determined by the input constellation (PAM, QAM or PSK). The upper bound in (16) is a sum of contributions from difference sequences corresponding to all possible values of $\delta_i$, $k_i$ and $m$.

**Example 5** *As an example we illustrate the binary constellation with two modulation matrices in Table 1. The rows represent the $i^{th}$ sequence and the columns represent the*

| $x^{(i)}$ / $x^{(j)}$ | (-1 -1) | (-1 1) | ( 1 -1) | ( 1 1) |
|---|---|---|---|---|
| (-1 -1) | ( 0 0) | ( 0 -2) | (-2 0) | (-2 -2) |
| (-1 1) | ( 0 2) | ( 0 0) | (-2 2) | (-2 0) |
| ( 1 -1) | ( 2 0) | ( 2 -2) | ( 0 0) | ( 0 -2) |
| ( 1 1) | ( 2 2) | ( 2 0) | ( 0 2) | ( 0 0) |

Table 1: All difference sequences for M = 2, K = 2

*$j^{th}$ sequence. The $[ij]$ entry represents $\delta^{(ij)}$ which is a length two difference sequence. The total number of sequences is $M^K=4$. Computing (16) using Table 1, we have*

$$\begin{aligned} P_U &= \mathrm{Q}\left(\sqrt{\frac{S}{2}\Omega_{11}(2)^2}\right)+\mathrm{Q}\left(\sqrt{\frac{S}{2}\Omega_{22}(2)^2}\right)\ldots \\ &\quad +\frac{1}{2}\mathrm{Q}\left(\sqrt{\frac{S}{2}(\Omega_{11}(2)^2+\Omega_{22}(2)^2)+S\Omega_{12}(2)(2)}\right)\ldots \\ &\quad +\frac{1}{2}\mathrm{Q}\left(\sqrt{\frac{S}{2}(\Omega_{11}(2)^2+\Omega_{22}(2)^2)-S\Omega_{12}(2)(2)}\right) \end{aligned} \tag{18}$$

*where the first two terms satisfy Lemma 2 trivially with $\Delta_1=\frac{S}{2}\Omega_{11}(2)^2$ and $\Delta_2=0$, and $\Delta_1=\frac{S}{2}\Omega_{22}(2)^2$ and $\Delta_2=0$ respectively. The last two terms satisfy Lemma 2 with $\Delta_1=\frac{S}{2}(\Omega_{11}(2)^2+\Omega_{22}(2)^2)$ and $\Delta_2=|S\Omega_{12}(2)(2)|$.*

The key concept for the general proof of Lemma 2 follows from the simple case $m=1$, i.e., those pairs of sequences that differ in only one position. Consider all difference sequences such that $|\delta_k^{(ij)}|=\delta_1$ for $k=k_1$ and zero otherwise. There are two possible signs for the actual difference, either $\delta_{k_1}^{(ij)}=\delta_1$ or $\delta_{k_1}^{(ij)}=-\delta_1$. By symmetry, the number of difference sequences with $\delta_{k_1}^{(ij)}=\delta_1$ is the same as the number of difference sequences with $\delta_{k_1}^{(ij)}=-\delta_1$, say $n_1$ each [1]. Since only one difference position is nonzero in (15),

[1]If there is a sequence pair (i,j) with $x_{k_1}^{(i)}-x_{k_1}^{(j)}=\delta_1$, then by symmetry there is also the pair (j,i) such that $x_{k_1}^{(j)}-x_{k_1}^{(i)}=-\delta_1$.

$\Delta_2^{(ij)} = S \ \sum_{k<l}^{K} \Omega_{kl}\delta_k^{(ij)}\delta_l^{(ij)} = 0$. For both signs of $\delta_1$, $\Delta_1^{(ij)} = \frac{S}{2}\Omega_{k_1k_1}\delta_1^2$. Since *all sequences are equally likely*, the contribution to $P_U$ from these $2n_1$ difference sequences is

$$q \;=\; \frac{n_1}{R}\left[\mathrm{Q}\left(\sqrt{\Delta_1^{(ij)}+0}\right)+\mathrm{Q}\left(\sqrt{\Delta_1^{(ij)}+0}\right)\right] \;=\; \frac{2n_1}{R}\mathrm{Q}\left(\sqrt{\frac{S}{2}\Omega_{k_1k_1}\delta_1^2}\right)$$

This pair satisfies Lemma 2 trivially with $\Delta_1 = \Delta_1^{(ij)}$ and $\Delta_2 = 0$.

For the general proof, consider the Euclidean distance $D_m$ between two sequences when they differ in $m$ positions $k_1,\ldots,k_m$ by absolute values $|\delta_1|,\ldots,|\delta_m|$ respectively

$$D_m \;=\; (\Omega_{k_1k_1}\delta_1^2+\ldots+\Omega_{k_mk_m}\delta_m^2)+2(\Omega_{k_1k_2}\delta_1\delta_2+\ldots+\Omega_{k_{m-1}k_m}\delta_{m-1}\delta_m)$$

Now consider the distance when $m$ is increased to $m+1$.

$$\begin{aligned} D_{m+1} &= (\Omega_{k_1k_1}\delta_1^2+\ldots+\Omega_{k_{m+1}k_{m+1}}\delta_{m+1}^2)+2(\Omega_{k_1k_2}\delta_1\delta_2+\ldots+\Omega_{k_mk_{m+1}}\delta_m\delta_{m+1}) \\ &= D_m+\Omega_{k_{m+1}k_{m+1}}\delta_{m+1}^2+2(\Omega_{k_1k_{m+1}}\delta_1+\ldots+\Omega_{k_mk_{m+1}}\delta_m)\delta_{m+1} \qquad (19) \end{aligned}$$

Similar to the case $m = 1$, the actual difference at position $k_{m+1}$ is either $\delta_{m+1}$ or $-\delta_{m+1}$ with equal probability. For a fixed value of $D_m$, the argument of the Q function is either $\frac{S}{2}D_{m+1} = \Delta_1 + \Delta_2$ or $\frac{S}{2}D_{m+1} = \Delta_1 - \Delta_2$, where $\Delta_1 = \frac{S}{2}(D_m + \Omega_{k_{m+1}k_{m+1}}\delta_{m+1}^2) \geq 0$ and $\Delta_2 = S|(\Omega_{k_1k_{m+1}}\delta_1 + \ldots + \Omega_{k_mk_{m+1}}\delta_m)\delta_{m+1}| \geq 0$. For a given $D_m$ then, the contribution from $2n_{m+1}$ sequences with differences $\pm\delta_{m+1}$ at position $k_{m+1}$ can be written as

$$q \;=\; \frac{n_{m+1}}{R}\left[\mathrm{Q}\left(\sqrt{\Delta_1+\Delta_2}\right)+\mathrm{Q}\left(\sqrt{\Delta_1-\Delta_2}\right)\right]$$

Since this holds for all possible values of $\delta_i$, $k_i$, $1 \leq i \leq m$, and $m$, we conclude that all terms in (16) satisfy Lemma 2. ■

Lemma 2 sets the stage for minimization of $P_U$. We will prove Lemma 3 before stating sufficient conditions for minimization in Lemma 4.

**Lemma 3** *: For a given value of* $\Delta_1$*, the pair* $q = \mathrm{Q}\left(\sqrt{\Delta_1+\Delta_2}\right)+\mathrm{Q}\left(\sqrt{\Delta_1-\Delta_2}\right)$ *in (17) is minimized iff (if and only if)* $\Delta_2 = 0$*.*

*Proof:* This lemma follows from convexity of the Q-function. Since $\mathrm{Q}(\sqrt{x})$ is convex and nonincreasing in $\sqrt{x}$, *and* $\sqrt{x}$ is concave in $x$ for $x > 0$, it follows that $\mathrm{Q}(\sqrt{x})$ is convex in $x$ for $x > 0$ [10]. Applying Jensen's Inequality [11] to $q$ we get Lemma 3. ■

Applying the results of Lemmas 2 and 3 to Lemma 1, we state conditions for minimization of $P_U$ in Lemma 4, subject to structure on the modulation matrices which is described next.

Consider the case $K = 1$, i.e. the linear code consists of only one modulation matrix $\mathbf{A}_1$. This case provides the *MFB (matched filter bound) performance* of the code. The error sequence is at most length one. The error bound $P_U$ consists of terms of the form $\mathrm{Q}(\sqrt{\Delta_1})$ where $\Delta_1 = \frac{S}{2}\Omega_{11}\delta_1^2$ for all possible values of $\delta_1$. In order to minimize $P_U$ it is sufficient to maximize $\Omega_{11} = \|\mathbf{HA}_1\|_F^2$, which can be expanded as follows

$$\begin{aligned} \|\mathbf{HA}_1\|_F^2 &= \mathrm{Tr}(\mathbf{HA}_1\mathbf{A}_1^*\mathbf{H}^*) = \mathrm{Tr}(\mathbf{HU\Sigma\Sigma}^*\mathbf{U}^*\mathbf{H}^*) = \mathrm{Tr}(\tilde{\mathbf{H}}\mathbf{\Sigma\Sigma}^*\tilde{\mathbf{H}}^*) \qquad (20) \\ &= \mathrm{Tr}(\sum_{i=1}^{\min(M_t,L)}\sigma_i^2\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^*) = \sum_{i=1}^{\min(M_t,L)}\sigma_i^2\mathrm{Tr}(\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^*) = \sum_{i=1}^{\min(M_t,L)}\sigma_i^2\|\tilde{\mathbf{h}}_i\|^2 \qquad (21) \end{aligned}$$

Here (20) follows from the s.v.d.(singular value decomposition) of the $M_t \times L$ matrix $\mathbf{A}_1$ which is $\mathbf{A}_1 = \mathbf{U\Sigma V}^*$, and $\{\sigma_i\}_{i=1}^{\min(M_t,L)}$ are the singular values of $\mathbf{A}_1$. The outer product decomposition of the matrix product is used in (21) where $\tilde{\mathbf{h}}_i$ are the columns of $\mathbf{HU}$ . Since the channel is unknown at the transmitter, it is advisable to choose $\sigma_i^2 \neq 0$ for all $i$ in order to maximize the sum in (21). Since the values of $\|\tilde{\mathbf{h}}_i\|^2$ are also unknown, it advisable to choose all the singular values to be equal, i.e., $\sigma_i^2 = \frac{c}{\min(M_t,L)}$ where $c$ is the Frobenius norm of $\mathbf{A}_1$ from (5). This results in $\mathbf{A}_1$ being a unitary matrix.

This result was also obtained in [4] in a similar manner. The optimality of this solution may be also derived using game theory, and we believe that this is the minimax saddlepoint solution. From a stochastic viewpoint, we believe that this solution is optimal when $\mathbf{H}$ is spatially uniform, i.e., it has no preferred directions along its right singular vectors. If $\mathbf{H}$ has an arbitrary mean or covariance, then unitary matrices may not be optimal. Rigorous proof of this is in progress. In the sequel we will assume all matrices are orthogonal, i.e., $\mathbf{A}_k\mathbf{A}_k^* = I_{M_t}$, where $M_t \leq L$, such that

$$\Omega_{kk} = \|\mathbf{HA}_k\|_F^2 = \frac{c}{\min(M_t, L)}\|\mathbf{H}\|_F^2 \qquad \text{for } 1 \leq k \leq K \qquad (22)$$

where $c$ is the power normalization as in (5).

**Lemma 4** *A linear code consisting of orthogonal modulation matrices* $\{\mathbf{A}_k\}_{k=1}^K$ *achieves the minimum* $P_U$ *iff the matrices satisfy the following condition*

$$\Omega_{kl} = \frac{1}{2}(\mathrm{Tr}(\mathbf{A}_k^*\mathbf{H}^*\mathbf{HA}_l) + \mathrm{Tr}(\mathbf{A}_l^*\mathbf{H}^*\mathbf{HA}_k)) = 0 \qquad \text{for } 1 \leq k \neq l \leq K \qquad (23)$$

In other words, the modulation matrices must be *pairwise orthogonal* for any weight matrix $\mathbf{H}^*\mathbf{H}$.

*Proof:* We will prove this lemma by induction on $K$, the number of modulation matrices. To verify the initial case, consider $K = 1$, i.e. the linear code consists of only one modulation matrix $\mathbf{A}_1$. This trivially satisfies (23).

Now we make the inductive assumption, i.e., the optimal set of modulation matrices satisfies Lemma 4 when $K = K'$. Then $P_U$ consists of terms of the following form

$$q = \frac{n_m}{M^K}\mathrm{Q}(\sqrt{\Delta_1})$$

$$\text{where} \quad \Delta_1 = \frac{S}{2}\sum_{i=1}^{m}\Omega_{k_ik_i}\delta_i^2$$

where $1 \leq m \leq K'$ with $\Omega_{k_ik_i}$ as in (22).

When the $(K+1)^{th}$ modulation matrix is introduced, we know from (19) that it adds terms of the following form to $P_U$

$$q = \frac{n}{M^{K+1}}\left[\mathrm{Q}\left(\sqrt{\Delta_1 + \Delta_2}\right) + \mathrm{Q}\left(\sqrt{\Delta_1 - \Delta_2}\right)\right] \qquad (24)$$

$$\text{where} \quad \Delta_1 = \frac{S}{2}\sum_{i=1}^{m}\Omega_{k_ik_i}\delta_i^2 + \frac{S}{2}\Omega_{K'+1K'+1}\delta_{K'+1}^2 \ ,$$

$$\Delta_2 = S|\sum_{i=1}^{m}\Omega_{k_iK'+1}\delta_i||\delta_{K'+1}|$$

and there are $n$ sequences with a given absolute value of $\delta_{K'+1}$ when all other differences $\{\delta_1, \ldots, \delta_m\}$ are held constant. From Lemma 3 we know that for a fixed value of $\Delta_1$, (24) will be minimized iff $\Delta_2 = 0$. Since all modulation matrices are assumed to be orthogonal, $\Omega_{K'+1K'+1}$ is as in (22) and $\Delta_1$ is not a function of the actual $(K+1)^{th}$ matrix as long as $\mathbf{A}_{K'+1}$ is orthogonal.

The value of $\Delta_2$ is a function of $\Omega_{k_iK'+1}$, $1 \leq i \leq m$, the weighted inner products of matrices $\mathbf{A}_{k_i}$ with the new matrix $\mathbf{A}_{K'+1}$. The necessary and sufficient condition to ensure $\Delta_2 = 0$ for all values of $k_i$, $1 \leq i \leq m$ and $m$ is that $\Omega_{k_iK'+1} = 0$ for $1 \leq i \leq m$. This can be seen from the following system of linear equations that the matrix of all possible differences $\delta$ and the vector of pairwise inner products $\Omega$ must satisfy

$$\delta\Omega = \mathbf{0}$$

The matrix $\delta$ is $m^q \times m$, where $m^q$ is the number of all possible values that $m$ differences can take. The vector $\Omega$ is $m \times 1$ and consists of the inner products $\Omega_{k_iK'+1}$ for $1 \leq i \leq m$. This equation must be satisfied for all values of $1 \leq m \leq K'$. Note that $\delta$ is a full rank matrix for any $m^q$, therefore $\Omega = \mathbf{0}$ for all $m$. This proves that $K'+1$ matrices must also satisfy (23), thereby finishing the induction and the proof of Lemma 4. ∎

Lemma 4 takes us to the main result of this paper, which is Theorem 1.

**Theorem 1** *Among all orthogonal, linear space-time codes, space-time block codes minimize the union bound on the conditional probability of symbol error for equally likely input symbols.*

*Proof:* Space-time block codes [2] consist of orthogonal matrices that satisfy Lemma 4. Therefore they satisfy the sufficient and necessary condition for minimization of $P_U$. In the complete absence of channel knowledge at the transmitter, these are also the only orthogonal codes that minimize the union bound. ∎

## 6 Conclusions

We have provided a new criterion for performance evaluation of linear space-time codes, namely the union bound on symbol error probability. This bound is a function of the instantaneous channel realization and is independent of channel statistics. It is expected to be tight at high SNRs and can be used to analyze average performance over a stochastic channel.

We provided necessary and sufficient conditions for minimization of the union bound over all orthogonal, linear space-time codes. We showed that space-time block codes satisfy these conditions and are therefore optimal. Our analysis applies to equally likely input symbols and does not necessarily imply that space-time block codes are optimal for non equally likely input symbols. We considered modulation matrices for the case when the number of transmit antennas is smaller than the code block length. Extension to modulation matrices when block length is smaller than the number of transmit antennas is in progress, as is extension of our analysis to all linear codes.

# References


[1] A. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception (DTDR)," Tech. Rep. U.S. Patent #5345599, September 1994.

[2] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Information Theory*, vol. 45, pp. 1456–1467, July 1999.

[3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication I: performance criterion and code construction," *IEEE Trans. Information Theory*, vol. 44, pp. 744–65, March 1998.

[4] G. Ganesan and P. Stoica, "Space-time diversity using orthogonal and amicable orthogonal designs," in *Proc. ICASSP*, 2000.

[5] G. J.Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Comm.*, vol. 6, pp. 311–335, March 1998.

[6] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Select. Areas in Comm.*, vol. 16, pp. 1451–1458, October 1998.

[7] A. Wittneben, "Basestation modulation diversity for digital simulcast," in *Vehicular Technology Conference*, pp. 848–853, 1991.

[8] N. Seshadri and J. H. Winters, "Two signaling schemes for improving the error performance of frequency-division-duplex (FDD) transmission systems using transmitter antenna diversity," in *Vehicular Technology Conference*, pp. 508–11, 1993.

[9] H. Lutkepohl, *Handbook of Matrices.* Wiley, 1996.

[10] S. Boyd and L. Vandenberghe, "Convex Optimization." Course notes, 1996-97.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley, 1991.