UNION BOUND ON ERROR PROBABILITY OF LINEAR SPACE-TIME BLOCK CODES

S. Sandhu

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

A. Paulraj

Packard 272
Stanford University
Stanford, CA 94305-9510
apaulraj@stanford.edu

ABSTRACT

Design of practical coding techniques for the multiple antenna 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) channel. Here we consider linear block codes 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. We assume maximum likelihood decoding at the receiver.

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 a function of the instantaneous channel realization and does not make any assumptions on channel statistics. We show that the orthogonal block codes in [1] achieve the lowest error bound among all unitary codes and are in fact optimal.

1. INTRODUCTION

A number of interesting coding and modulation techniques for the multiple-antenna fading channel have been proposed recently. They range from spatial multiplexing [2] to space-time block codes [1] to space-time trellis codes [3]. The first two schemes 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 block 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 sums 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 (maximum likelihood) decoding of the matrix symbols.

We analyze all linear codes within the framework of probability of 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 *unitary linear codes*, i.e., where the square or wide modulation matrices are unitary, we obtain *necessary and sufficient conditions for minimization* of the error bound [4]. In the absence of any channel knowledge at the transmitter these conditions are met by the orthogonal block codes in [1]. Our analysis does not presuppose decoupled detection as done in [5], 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 over the class of 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) circular complex Gaussian variables, i.e., $v_{nL_{i,j}} \sim \aleph_c(0, N_0)$, and are independent over n. The average power transmitted on M_t antennas is E_s per symbol time. Define $S = \frac{E_s}{N_0}$.

The Hermitian of a matrix is denoted by $\mathbf{A}^*$, the trace by $\mathrm{Tr}\mathbf{A}$, and the real part by $\mathrm{Re}\mathbf{A}$, respectively. The error 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 [5], 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 $x^{(r)} = \{x_1^{(r)}, \ldots, x_K^{(r)}\}$ be the r^{th} input symbol sequence. Then the corresponding codeword is*

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

where $\mathbf{A}_k$ are $M_t \times L$ modulation matrices. Each real symbol $x_k^{(r)}$ belongs to a PAM, one-dimensional "QAM" or one-dimensional "PSK" constellation of size M. The total number of codewords is $R = \mathrm{M}^K$.

We assume no prior encoding of the input symbols. A power constraint is imposed on each modulation matrix such that $\|\mathbf{A}_k\|_F^2 \leq c$ for all k. To maintain average power at E_s per symbol time, c is as follows

$$c = \frac{L}{K} \quad (\mathrm{PAM}) \quad , \quad c = \frac{2L}{K} \quad (\mathrm{QAM}, \mathrm{PSK}) \tag{3}$$

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.

Most of the currently proposed spatial modulation techniques can be interpreted as linear codes. Spatial multiplexing [2], also called BLAST [6], is the simplest example of linear codes. The modulation matrices are simply unit vectors, 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 $L = 1$.

Orthogonal space-time block codes are another example of linear codes. The modulation matrices are unitary matrices that are pairwise orthogonal with respect to the $\mathrm{Re}\ \mathrm{Tr}$ 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 [1].

Delay diversity [7] 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 to obtain the k^{th} modulation matrix.

4. PERFORMANCE ANALYSIS

We will analyze linear codes using the conditional probability of symbol error at the receiver, i.e., the error probability conditioned on a given channel realization. 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 unitary codes in Lemma 4. Finally in Theorem 1 we show that the optimal unitary linear codes are the space-time block codes in [1], which is the main result of the paper.

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

$$P_e \leq \sum_{i=1}^{R} p_i \sum_{j \neq i}^{R} \mathrm{Q}\left(\sqrt{\Delta_1^{(ij)} + \Delta_2^{(ij)}}\right) = P_U \tag{4}$$

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

$$P_e = \sum_{i=1}^{R} p_i P_{e|i} \tag{5}$$

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 upper bounded as follows

$$P_{e|i} \leq \sum_{i=1}^{R} \sum_{j \neq i}^{R} \mathrm{Q}\left(\sqrt{D_{ij} S/2}\right) \tag{6}$$

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

$$D_{ij} = \left\| \mathbf{H}\left(\mathbf{X}^{(i)} - \mathbf{X}^{(j)}\right) \right\|_F^2 = \left\| \mathbf{H} \sum_{k=1}^{K} \mathbf{A}_k \epsilon_k^{(ij)} \right\|_F^2 \tag{7}$$

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

$$D_{ij} = \sum_{k=1}^{K} \Omega_{kk} |\epsilon_k^{(ij)}|^2 + 2 \sum_{k=1}^{K} \sum_{l=1}^{k-1} \Omega_{kl} \epsilon_k^{(ij)} \epsilon_l^{(ij)} \tag{8}$$

where $\Omega_{kl} = \mathrm{Re}\ \mathrm{Tr}(\mathbf{H}\mathbf{A}_k \mathbf{A}_l^* \mathbf{H}^*)$ is the symmetric, weighted, matrix inner product of $\mathbf{A}_k$ and $\mathbf{A}_l$ defined as

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

where the weighting matrix is $\mathbf{H}^*\mathbf{H}$. Define the following

$$\begin{aligned}\Delta_1^{(ij)} &\triangleq \frac{S}{2}\sum_{k=1}^{K}\Omega_{kk}|\epsilon_k^{(ij)}|^2\,,\\ \Delta_2^{(ij)} &\triangleq \frac{S}{2}\,2\sum_{k=1}^{K}\sum_{l=1}^{k-1}\Omega_{kl}\epsilon_k^{(ij)}\epsilon_l^{(ij)}\end{aligned} \tag{9}$$

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

Substituting (9) in (8), (6) and then in (5) we get

$$P_e \leq \sum_{i=1}^{R} p_i \sum_{j\neq i}^{R} \mathrm{Q}\left(\sqrt{\Delta_1^{(ij)}+\Delta_2^{(ij)}}\right) = P_U \tag{10}$$

which finishes the proof of Lemma 1. ■

Our code design criterion 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 in (10). 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 Ω_{kk} and the other is determined by their pairwise inner products Ω_{kl}. If we make the assumption that all sequences are equally likely, i.e., $p_i = \frac{1}{R}$ for all i in (10), then we have Lemma 2.

Lemma 2 *By carefully selecting terms over i and j, we can always pair up terms in the expression for P_U as follows*

$$q = \frac{n}{R}\left[\mathrm{Q}\left(\sqrt{\Delta_1+\Delta_2}\right)+\mathrm{Q}\left(\sqrt{\Delta_1-\Delta_2}\right)\right] \tag{11}$$

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 $\epsilon^{(ij)} = \{\epsilon_1^{(ij)},\ldots,\epsilon_K^{(ij)}\}$, where $\epsilon_k^{(ij)} = x_k^{(i)} - x_k^{(j)}$. When two symbol sequences differ only in m positions, say $k_1,\ldots,k_m$, then $\epsilon_k^{(ij)} \neq 0$ for $k = k_1,\ldots,k_m$, and $\epsilon_k^{(ij)} = 0$ for $k \neq k_1,\ldots,k_m$. Let $\epsilon_{k_i}^{(ij)} = \epsilon_i \neq 0$ for $1 \leq i \leq m$. The range of values of ϵ_i is determined by the input constellation. The expression in (10) is a sum of contributions from difference sequences corresponding to all possible values of ϵ_i, k_i and m.

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 $|\epsilon_k^{(ij)}| = \epsilon_1$ for $k = k_1$ and zero otherwise. There are two possible signs for the actual difference, either $\epsilon_{k_1}^{(ij)} = \epsilon_1$ or $\epsilon_{k_1}^{(ij)} = -\epsilon_1$. By symmetry, the number of difference sequences with $\epsilon_{k_1}^{(ij)} = \epsilon_1$ is the same as the number of difference sequences with $\epsilon_{k_1}^{(ij)} = -\epsilon_1$, say n_1 each [1]. Since only one difference position is nonzero in (9), $\Delta_2^{(ij)} = S\,\sum_{l<k}^{K}\Omega_{kl}\epsilon_k^{(ij)}\epsilon_l^{(ij)} = 0$. For both signs of ϵ_1, $\Delta_1^{(ij)} = \frac{S}{2}\Omega_{k_1k_1}\epsilon_1^2$. Since *all sequences are equally likely*, the contribution from these $2n_1$ difference sequences is

$$\begin{aligned} q &= \frac{n_1}{R}\left[\mathrm{Q}\left(\sqrt{\Delta_1^{(ij)}+0}\right)+\mathrm{Q}\left(\sqrt{\Delta_1^{(ij)}+0}\right)\right]\\ &= \frac{2n_1}{R}\mathrm{Q}\left(\sqrt{\frac{S}{2}\Omega_{k_1k_1}\epsilon_1^2}\right)\end{aligned}$$

This pair satisfies Lemma 2 trivially with $\Delta_1 = \Delta_1^{(ij)}$ and $\Delta_2 = 0$. The general proof follows from this and the details are provided in [4]. ■

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 Δ_1, the pair $q = \mathrm{Q}\left(\sqrt{\Delta_1+\Delta_2}\right) + \mathrm{Q}\left(\sqrt{\Delta_1-\Delta_2}\right)$ in (11) is minimized if and only if $\Delta_2 = 0$.*

Proof: This lemma follows from convexity of the Q-function. Since $\mathrm{Q}\left(\sqrt{x}\right)$ is convex and nonincreasing in $\sqrt{x}$, and $\sqrt{x}$ is concave in x for $x > 0$, it follows that $\mathrm{Q}\left(\sqrt{x}\right)$ is convex in x for $x > 0$ [8]. Applying Jensen's Inequality 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 the following structure on modulation matrices.

Assumption 1 *In the sequel we will assume all matrices are unitary, that is*

$$\mathbf{A}_k\mathbf{A}_k^* = I_{M_t} \tag{12}$$

for $1 \leq k \leq K$ where $M_t \leq L$. This ensures a constant value of Ω_{kk} over all k as follows

$$\Omega_{kk} = \|\mathbf{H}\mathbf{A}_k\|_F^2 = \frac{c}{M_t}\|\mathbf{H}\|_F^2 \tag{13}$$

where c is the power normalization in (3).

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

$$\Omega_{kl} = \mathrm{Re}\,\mathrm{Tr}(\mathbf{A}_k^*\mathbf{H}^*\mathbf{H}\mathbf{A}_l) = 0 \tag{14}$$

for $1 \leq k \neq l \leq K$. In other words, the modulation matrices must be *pairwise orthogonal* for any weight $\mathbf{H}^*\mathbf{H}$.

Proof: We will prove this lemma by induction on K, the number of modulation matrices. To verify the initial

[1] If there is a sequence pair (i,j) with $x_{k_1}^{(i)} - x_{k_1}^{(j)} = \epsilon_1$, then by symmetry there is also the pair (j,i) such that $x_{k_1}^{(j)} - x_{k_1}^{(i)} = -\epsilon_1$.

case, consider $K = 1$, i.e. the linear code consists of only one modulation matrix $\mathbf{A}_1$. This trivially satisfies (14).

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

$$q = \frac{n_m}{M^{K'}} \mathrm{Q}\left(\sqrt{\Delta_1}\right) \quad (15)$$

where $\Delta_1 = \frac{S}{2}\sum_{i=1}^{m} \Omega_{k_i k_i} \epsilon_i^2$. These are contributions from sequences differing in m positions, where $1 \le m \le K'$ with $\Omega_{k_i k_i}$ as in (13).

When the $(K'+1)^{th}$ modulation matrix is introduced, it will introduce difference sequences that are nonzero in the $(K'+1)^{th}$ position. Contributions from all difference sequences that are nonzero in the $(K'+1)^{th}$ position can be written as follows

$$q = \mathrm{Q}\left(\sqrt{\Delta_1 + \Delta_2}\right) + \mathrm{Q}\left(\sqrt{\Delta_1 - \Delta_2}\right) \quad (16)$$

where $\Delta_1 = \frac{S}{2}\sum_{i=1}^{m} \Omega_{k_i k_i} \epsilon_i^2 + \frac{S}{2}\Omega_{(K'+1)(K'+1)}\epsilon_{K'+1}^2$ and $\Delta_2 = S\left|\sum_{i=1}^{m} \Omega_{(k_i)(K'+1)}\epsilon_i\right| |\epsilon_{K'+1}|$. From Lemma 3 we know that for a fixed value of Δ_1, (16) will be minimized *if and only if* $\Delta_2 = 0$. Since all modulation matrices are assumed to be unitary, $\Omega_{(K'+1)(K'+1)}$ is as in (13) and Δ_1 is not a function of the actual $(K+1)^{th}$ matrix as long as $\mathbf{A}_{K'+1}$ is unitary.

The value of Δ_2 is a function of $\Omega_{(k_i)(K'+1)}$, $1 \le i \le m$, which are 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 is that $\Omega_{(k_i)(K'+1)} = 0$ for $1 \le i \le m$. See [4] for details. This proves that the set of $K'+1$ matrices must also satisfy (14), thus completing the induction to prove that (14) is a necessary and sufficient condition for minimization of P_U. ■

Lemma 4 takes us to Theorem 1, our main result.

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

Proof: Space-time block codes [1] consist of unitary 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 unitary codes that minimize the union bound. ■

5. CONCLUSIONS

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

We provided necessary and sufficient conditions for minimization of the union bound over all unitary, linear space-time codes. We showed that space-time block codes satisfy these conditions and are therefore optimal. Our analysis applies to uncoded, 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 of our analysis to modulation matrices when block length is smaller than the number of transmit antennas is in progress, as is extension to the class of all linear codes.

6. REFERENCES

[1] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, “Space-time block codes from orthogonal designs,” *IEEE Trans. Information Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[2] 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.

[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, no. 2, pp. 744–65, March 1998.

[4] S. Sandhu, R. Heath, and A. Paulraj, “Union bound for linear space-time codes,” in *Proc. 38th Annual Allerton Conference on Communication, Control and Computing*, 2000.

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

[6] 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, no. 3, pp. 311–335, March 1998.

[7] 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*, 1993, pp. 508–11.

[8] S. Boyd and L. Vandenberghe, “Convex Optimization,” Course notes, 1996-97.