On Outage and Capacity for Orthogonal Transmit Diversity

Tarik Muharemovic and Behnaam Aazhang

Authors are with the Department of Electrical and Computer Engineering (MS 366), Rice University, Houston, Texas 77251-1892. This research work was supported by Nokia.

Abstract

Assuming that only the receiver has perfect channel state information, we analyze the orthogonal transmit diversity (OTD). First, we consider the case of fast fading or *ergodic* channels. We demonstrate that for a single receive antenna, OTD does not result in a capacity loss. This is not the case when there are multiple receive antennas. Then, we consider block fading or *non-ergodic* channels. We propose a new, simple, and asymptotically tight upper bound for the outage probability in multiple transmit and single receive antenna systems. Assuming equal power distribution, we also demonstrate that for non-ergodic channels and a single receive antenna, OTD preserves the achievable outage probability. However, in the case of multiple receive antennas, OTD increases the achievable outage probability. In short, regardless of the nature of fading, OTD is optimal only in the case of a single receive antenna.

Keywords

Transmit diversity, space-time codes, orthogonal transmit diversity (OTD), outage probability.

I. Introduction

Deep fade in a wireless channel makes it extremely difficult for receiver to recover the transmitted signal. One way of combating this problem is diversity, i.e., multiple replicas of the transmitted signal present at the receiver. There are different types of diversity: time diversity, frequency diversity, and antenna diversity, to name a few. Here we consider the case of antenna diversity, where spatially separated antennas are available with a fixed total transmit power. This becomes the only source of diversity in slow, frequency non-selective fading channels, where time and frequency diversity are not feasible. Furthermore, a significant increase in transmission rate can be achieved in both fast [1] and slow fading channels using antenna diversity. Information theoretic capacity of this multiple-transmit multiple-receive antenna scheme was derived in [1]. Fading coefficients between antennas were assumed to be independent and identically Rayleigh distributed, and the coherence time of the channel is assumed to correspond to one symbol duration. This derivation was extended for coherence times greater than one symbol duration in [2]. It was shown that there is no change in the achievable information rate for larger coherence times. Following these discoveries, several researchers have designed some practical codes in order to approach these high rates [3], [4], [5]. Their guiding principle was code construction criterion, based on the pairwise error probability. In addition to these codes, orthogonal

[1]Here, by fast fading we mean fast as relative to codeword duration, *not* symbol duration.

transmit diversity was proposed as another simple and practical transmit scheme [6]. Several nice properties of this scheme were demonstrated in [6], [7]. It is somewhat similar to CDMA orthogonalisation; however, symbols are not spread only over time but over antennas as well, hence the expression: "orthogonal transmit diversity." It was further generalized to space-time block codes from orthogonal designs in [7].

In this paper, we present an information theoretic analysis of OTD. For fast fading channels we examine the capacity [2], while for block fading channels we examine the outage probability. For purposes of this discussion, fast fading or ergodic channels are defined as channels where we allow coding over multiple time-independent channel fades. Coherence time of the channel is assumed to be a lot less than codeword length. On the other hand, in block fading or non-ergodic channels codeword duration is restricted to a single channel state. In addition to this OTD analysis we propose a new, simple, and asymptotically tight upper bound for the outage probability. Throughout the discussion we assume a frequency non-selective fading channel.

II. Preliminaries and Previous Results

The channel model is sketched in Fig. 1, where fading coefficients to antenna j only are plotted for the sake of picture clarity. Receive antenna j receives noisy superposition of transmitted complex constellation symbols $(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_n)$, randomly attenuated by the channel $(\mathrm{h}_{j,1},\ \mathrm{h}_{j,2}, ...,\ \mathrm{h}_{j,n})$. Channel coefficients are independent, zero mean, complex Gaussian with varinace 0.5 per dimension. Noise, denoted as n_j, is assumed to be proper complex Gaussian with zero mean and variance $\frac{\sigma^2}{2}$ per dimension.

$$\mathrm{y}_j = \sum_{i=1}^{n} \mathrm{h}_{j,i}\mathrm{x}_i + \mathrm{n}_j \quad \text{for} \quad j \in \{1, 2, ..., m\} \qquad \Leftrightarrow \qquad \underline{\mathrm{y}} = \mathrm{H}\underline{\mathrm{x}} + \underline{\mathrm{n}}, \tag{1}$$

where we labeled $\underline{\mathrm{x}}^T = (\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_n)$ and stacked all y_j into a column vector. H is an $m \times n$ matrix of fading coefficients, row index cooresponding to receive and column index to transmit antenna.

[2] All units for capacity expressions are bits per transmission or bits/sec/Hz unless specified otherwise.

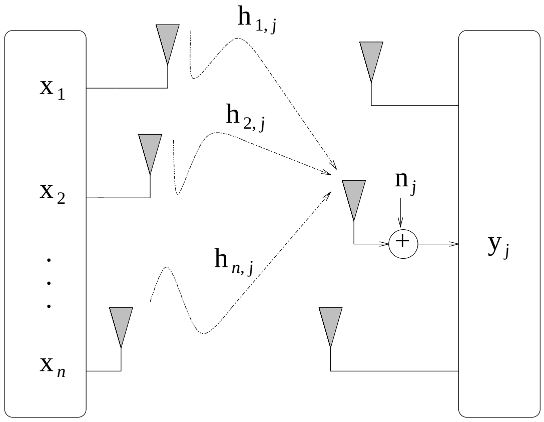

Fig. 1. Channel Model. Fading coefficients to antenna j only are plotted.

A. Ergodic Channel Capacity

The capacity of a multiple antenna communication system in fast Rayleigh fading channels was discussed in [1], [2]. Here we include only the main points as well as the result. We assume that consecutive channel realizations are independent. This independence allows averaging the capacity expression over all channel realizations, and that is why we also refer to fast fading channels as ergodic channels. Since capacity depends on the correlation among antennas, we label $Q = \mathbb{E}[\underline{\mathrm{x}}\underline{\mathrm{x}}^\star]$, where $\star$ denotes conjugate transposed. It is obvious that Q is non-negative definite, and that any non-negative definite matrix is a candidate for correlation among transmit antennas. The power constraint is given as $\mathrm{tr(Q)} \leq \mathrm{P}$. Therefore, we can write the capacity of this channel as

$$C \stackrel{(a)}{=} \sup_{\substack{\mathrm{tr(Q)}\leq \mathrm{P} \\ Q\geq 0}} I(\underline{\mathrm{x}};\underline{\mathrm{y}},\mathrm{H}) \stackrel{(b)}{=} \sup_{\substack{\mathrm{tr(Q)}\leq \mathrm{P} \\ Q\geq 0}} \{I(\underline{\mathrm{x}};\mathrm{H}) + I(\underline{\mathrm{x}};\underline{\mathrm{y}}|\mathrm{H})\}$$

$$\stackrel{(c)}{=} \sup_{\substack{\mathrm{tr(Q)}\leq \mathrm{P} \\ Q\geq 0}} I(\underline{\mathrm{x}};\underline{\mathrm{y}}|\mathrm{H}) \stackrel{(d)}{=} \sup_{\substack{\mathrm{tr(Q)}\leq \mathrm{P} \\ Q\geq 0}} \mathbb{E}_{\mathrm{H}}[I(\underline{\mathrm{x}};\underline{\mathrm{y}}|\mathrm{H}=H)]. \tag{2}$$

Here, the equalities carry since (a) is just the definition, (b) is by the chain rule for mutual information, and (c) is due to the fact that the transmitter has no knowledge of the channel and hence $\underline{\mathrm{x}}$ and H are independent. The last equality (d) follows straight from the definition where we distinguish between H as a random matrix and italics H as a realization.

Given the channel realization H and the covariance matrix $Q = \mathbb{E}[\underline{\mathrm{x}}\underline{\mathrm{x}}^\star]$, the mutual information $I(\underline{\mathrm{x}};\underline{\mathrm{y}}|\mathrm{H} = H)$ is maximized if $\underline{\mathrm{x}}$ is jointly Gaussian random vector with covariance Q. This is well established maximum entropy lemma [1], [8], and the mutual information becomes

$$I(\underline{\mathrm{x}};\underline{\mathrm{y}}|\mathrm{H}=H)=\log\det(I_m+\frac{HQH^\star}{\sigma^2}), \tag{3}$$

where I_m is an identity $m \times m$ matrix. When we make use of this, the capacity becomes

$$C=\sup_{\substack{\mathrm{tr(Q)}\leq \mathrm{P} \\ Q\geq 0}} \mathbb{E}_{\mathrm{H}}[\log\det(I_m+\frac{\mathrm{H}Q\mathrm{H}^\star}{\sigma^2})]. \tag{4}$$

The matrix $Q=UDU^\star$ is further diagonalized and since the distribution of $\mathrm{H}U$ is the same as the distribution of H we can consider the above optimization to be over diagonal matrices D. Alternatively, we restrict ourselves to diagonal Q. The property of the distribution on H, that it is invariant under orthogonal transformations U is called "spherically invariant property" [2].

It was shown in [1] that $\log\det(\cdot)$ is concave on the set of non-negative definite matrices. Also, $I_m+\frac{\mathrm{H}Q\mathrm{H}^\star}{\sigma^2}$ is a linear transformation of Q which results in a non-negative definite matrix if Q is non-negative definite. It follows that $\mathbb{E}_{\mathrm{H}}[\log\det(I_m+\frac{\mathrm{H}Q\mathrm{H}^\star}{\sigma^2})]$ is concave on the set of non-negative definite Q's. Furthermore, the above expression is invariant to permutations of Q. Averaging any diagonal Q over all possible permutations results in a matrix proportional to identity. By concavity, we conclude that Q proportional to the identity is optimal [1]. The capacity expression now simplifies to

$$C=\mathbb{E}_{\mathrm{H}}[\log\det(I_m+\frac{P}{n\sigma^2}\mathrm{H}\mathrm{H}^\star)], \tag{5}$$

which is one of the main results in [1]. The capacity was further evaluated in the case of Rayleigh fading; however, in order to illustrate our point the above result is sufficient. Note that in the case of one receive, antenna, H is a $1\times n$ vector and we can make further simplifications to obtain

$$C=\mathbb{E}_{\mathrm{H}}[\log(1+\frac{P\|\mathrm{H}\|^2}{n\sigma^2})]. \tag{6}$$

The above analysis is under assumption that the coherence time of the channel is one symbol period. In the case of larger coherence time $T_c > 1$, similar analysis carries through [2] and it follows that

$$C=T_c\mathbb{E}_{\mathrm{H}}[\log(1+\frac{P\|\mathrm{H}\|^2}{n\sigma^2})], \tag{7}$$

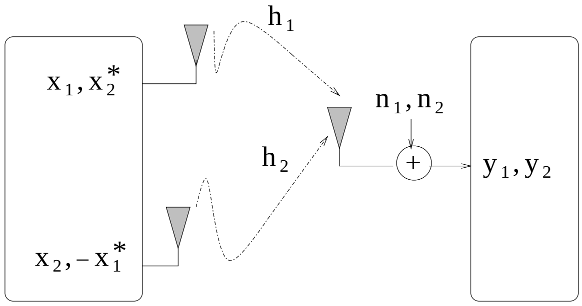

Fig. 2. Alamouti orthogonal transmit diversity technique

for T_c transmissions. This means that there is no alteration in the capacity expression for the case of greater coherence times.

B. Orthogonal Transmit Diversity

A simple two-branch transmit diversity scheme was proposed in [6] where using two transmit and one receive antenna one can achieve a diversity order of two [3]. The scheme is very appealing since it is simple and practical to implement. We first give a brief description and then evaluate the achievable information rate. We arrive to the interesting conclusion that *in the case of single receiver antenna there is no loss in the achievable information rate* as compared to a totally general two transmit antenna system. However, when there are multiple receivers, then there is a reduction in capacity if one constrains the transmitter to use OTD.

The OTD scheme can be described as follows: during the first time instant code symbol x_1 is transmitted on antenna 1 and x_2 on antenna 2. At the next time instant antenna 1 transmits $x_2^\star$ and antenna 2 transmits $-x_1^\star$, as described in Fig. 2. It was shown in [6] that coherence time of at least two symbols is necessary in order for receiver to decouple estimates for x_1 and x_2. This decoupling is the desired orthogonalisation which reduces computational complexity. Hence, we assume that the fading coefficients do not change between two close time instances.

[3] By diversity order we mean the number of independently faded observations of a single transmitted constellation simbol.

III. Rates Achievable with OTD

In this section, we evaluate the achievable information rate for the proposed OTD scheme. For the case of two transmit antennas, we assume that coherence time of the channel is two symbols, and if it happens to be larger, we perform a sufficient amount of interleaving to achieve independent coefficients beyond the two consecutive time instances. Here, we do allow coding over multiple channel realizations. Refering to Fig. 2, the receiver observes following statistics

$$\begin{aligned} \mathrm{y}_1 &= \mathrm{h}_1\mathrm{x}_1 + \mathrm{h}_2\mathrm{x}_2 + \mathrm{n}_1 \\ \mathrm{y}_2 &= \mathrm{h}_1\mathrm{x}_2^\star - \mathrm{h}_2\mathrm{x}_1^\star + \mathrm{n}_2, \end{aligned} \tag{8}$$

over two consecutive time instants. In order to determine the power constraint, note that if the vector $(\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, ..., \mathrm{x}_{2l-1}, \mathrm{x}_{2l})$ is concatenated with OTD then $(\mathrm{x}_1, \mathrm{x}_2^\star, \mathrm{x}_3, ..., \mathrm{x}_{2l-1}, \mathrm{x}_{2l}^\star)$ will be the output of the first branch of OTD. Similarily, $(\mathrm{x}_2, -\mathrm{x}_1^\star, \mathrm{x}_4, ..., \mathrm{x}_{2l}, -\mathrm{x}_{2l-1}^\star)$ will be the output of the second branch of OTD. Hence, the power transmitted on each antenna will be $\frac{1}{2l}\sum_{i=1}^{2l} \|\mathrm{x}_i\|^2$, meaning that the total transmitted power is given as

$$\frac{1}{l}\sum_{i=1}^{2l} \|\mathrm{x}_i\|^2 \xrightarrow{LLN} \mathbb{E}[\|\mathrm{x}_1\|^2] + \mathbb{E}[\|\mathrm{x}_2\|^2] \le P, \tag{9}$$

which is a two-antenna power constraint. Here, LLN refers to law of large numbers, and the convergence is in probability.

Theorem 1: Assume a fast Rayleigh fading channel, with a coherence time equaling two symbols. Then, using the OTD on the transmitter side does not result in a capacity reduction in the case of a single receive antenna.

Proof: Conjugating the statistic y_2 and stacking it into a matrix together with y_1 yields the following received signal model

$$\begin{bmatrix} \mathrm{y}_1 \\ \mathrm{y}_2^\star \end{bmatrix} = \begin{bmatrix} \mathrm{h}_1 & \mathrm{h}_2 \\ -\mathrm{h}_2^\star & \mathrm{h}_1^\star \end{bmatrix} \begin{bmatrix} \mathrm{x}_1 \\ \mathrm{x}_2 \end{bmatrix} + \begin{bmatrix} \mathrm{n}_1 \\ \mathrm{n}_2^\star \end{bmatrix} \Leftrightarrow \underline{\mathrm{y}}' = \mathrm{G}\underline{\mathrm{x}} + \underline{\mathrm{n}}, \tag{10}$$

where $\underline{\mathrm{y}}'$ is the obvious transformation of the vector $\underline{\mathrm{y}}$, and $\mathrm{G} = \begin{bmatrix} \mathrm{h}_1 & \mathrm{h}_2 \\ -\mathrm{h}_2^\star & \mathrm{h}_1^\star \end{bmatrix}$ is a Radon-Hurwitz [7] transform of the vector of fading coefficients. We evaluate the capacity of

OTD, again assuming that the receiver perfectly knows the fading coefficients. The mutual information is given as

$$I(\underline{\mathrm{x}}; \mathrm{G}, \underline{\mathrm{y}}) = I(\underline{\mathrm{x}}; \mathrm{G}, \underline{\mathrm{y}}') = I(\underline{\mathrm{x}}; \mathrm{G}) + I(\underline{\mathrm{x}}; \underline{\mathrm{y}}'|\mathrm{G}) = I(\underline{\mathrm{x}}; \underline{\mathrm{y}}'|\mathrm{G}). \tag{11}$$

We concluded that $I(\underline{\mathrm{x}}; \mathrm{G})$ is zero because transmitter has no knowledge of the channel. Noting that two time instances are used, it follows that

$$2C_{\mathrm{OTD}} \overset{(a)}{=} \sup_{\substack{\mathrm{tr(Q)} \leq \mathrm{P} \\ Q \geq 0}} I(\underline{\mathrm{x}}; \underline{\mathrm{y}}'|\mathrm{G}) \overset{(b)}{=} \sup_{\substack{\mathrm{tr(Q)} \leq \mathrm{P} \\ Q \geq 0}} \mathbb{E}_{\mathrm{G}}[I(\underline{\mathrm{x}}; \underline{\mathrm{y}}|\mathrm{G} = G)] \overset{(c)}{=} \sup_{\substack{\mathrm{tr(Q)} \leq \mathrm{P} \\ Q \geq 0}} \mathbb{E}_{\mathrm{G}}[\log\det(I_2 + \frac{\mathrm{G}Q\mathrm{G}^\star}{\sigma^2})]. \tag{12}$$

Here, (a) is a direct consequence of (11), (b) is just from the definition, and (c) is by the maximum entropy lemma, similar to (3). From here we can use the common matrix identity $\det(I + AB) = \det(I + BA)$, which yields

$$2C_{\mathrm{OTD}} = \sup_{\substack{\mathrm{tr(Q)} \leq \mathrm{P} \\ Q \geq 0}} \mathbb{E}_{\mathrm{G}}[\log\det(I_2 + \frac{\mathrm{G}^\star \mathrm{G} Q}{\sigma^2})] = \sup_{\substack{\mathrm{tr(Q)} \leq \mathrm{P} \\ Q \geq 0}} \mathbb{E}_{\mathrm{H}}[\log\det(I_2 + \frac{(\|\mathrm{h}_1\|^2 + \|\mathrm{h}_2\|^2)Q}{\sigma^2})]. \tag{13}$$

The last equation is because of the orthogonality of G, meaning that $\mathrm{G}^\star\mathrm{G} = (\|\mathrm{h}_1\|^2 + \|\mathrm{h}_2\|^2)I_2$. Since the last expression is a supremum over a set of Q's, when we consider Q proportional to identity, $Q = \frac{P}{2}I_2$, we obtain a lower bound as

$$C_{\mathrm{OTD}} \geq \frac{1}{2}\mathbb{E}_{\mathrm{H}}[\log\det(I_2 + \frac{P}{2\sigma^2}(\|\mathrm{h}_1\|^2 + \|\mathrm{h}_2\|^2)I_2)]$$

$$= \frac{1}{2}\mathbb{E}_{\mathrm{H}}[\log(1 + \frac{P}{2\sigma^2}(\|\mathrm{h}_1\|^2 + \|\mathrm{h}_2\|^2))^2] = \mathbb{E}_{\mathrm{H}}[\log(1 + \frac{P}{2\sigma^2}(\|\mathrm{h}_1\|^2 + \|\mathrm{h}_2\|^2)]. \tag{14}$$

Using (6) we have the reverse inequality because the rates achievable with OTD can not be greater than rates of a general multiple transmit antenna system.

$$C_{\mathrm{OTD}} \leq \mathbb{E}_{\mathrm{H}}[\log(1 + \frac{P}{2\sigma^2}(\|\mathrm{h}_1\|^2 + \|\mathrm{h}_2\|^2)]. \tag{15}$$

Combining (14) and (15) we have

$$C_{\mathrm{OTD}} = \mathbb{E}_{\mathrm{H}}[\log(1 + \frac{P}{2\sigma^2}(\|\mathrm{h}_1\|^2 + \|\mathrm{h}_2\|^2)]. \tag{16}$$

Hence, we have justified that $Q = \frac{P}{2}I_2$ is optimal and that there is no loss in capacity. ■

We conclude that for the case of a single receiver the achievable information rate is not reduced by using the OTD scheme. However, if there is more than one receive antenna using OTD results in a loss in capacity.

Proposition 1: If there are multiple receive antennas, using the OTD on the transmitter side results in an achievable information rate strictly less than the optimal given in (5).

Proof: We prove this for the case of two receive antennas. Extension to the case of more receivers is straightforward but tedious. We can again refer to Fig. 2; however, in order to distinguish between receive antennas, we add another subscript (first) to channel coefficients, noise, and received vector. We have four statistics available at the reciever over two consecutive time instants.

$$\mathrm{y}_{1,1} = h_{1,1}\mathrm{x}_1 + h_{1,2}\mathrm{x}_2 + \mathrm{n}_{1,1} \qquad \mathrm{y}_{2,1} = h_{2,1}\mathrm{x}_1 + h_{2,2}\mathrm{x}_2 + \mathrm{n}_{2,1}$$

$$\mathrm{y}_{1,2} = -h_{1,2}\mathrm{x}_1^\star + h_{1,1}\mathrm{x}_2^\star + \mathrm{n}_{1,2} \qquad \mathrm{y}_{2,2} = -h_{2,2}\mathrm{x}_1^\star + h_{2,1}\mathrm{x}_2^\star + \mathrm{n}_{2,2}.$$

We express this in the matrix notation as

$$\begin{bmatrix} \mathrm{y}_{1,1} \\ \mathrm{y}_{2,1} \\ \mathrm{y}_{1,2}^\star \\ \mathrm{y}_{2,2}^\star \end{bmatrix} = \begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \\ -h_{1,2}^\star & h_{1,1}^\star \\ -h_{2,2}^\star & h_{2,1}^\star \end{bmatrix} \begin{bmatrix} \mathrm{x}_1 \\ \mathrm{x}_2 \end{bmatrix} + \begin{bmatrix} \mathrm{n}_{1,1} \\ \mathrm{n}_{2,1} \\ \mathrm{n}_{1,2}^\star \\ \mathrm{n}_{2,2}^\star \end{bmatrix} \Leftrightarrow \underline{\mathrm{y}}' = \mathrm{G}\underline{\mathrm{x}} + \underline{\mathrm{n}}. \tag{17}$$

Once again the matrix transformation of channel coefficients has orthogonal columns, $\mathrm{G}^\star\mathrm{G} = (\|\underline{\mathrm{h}}_1\|^2 + \|\underline{\mathrm{h}}_2\|^2)I_2$. Here, $\underline{\mathrm{h}}_1 = [h_{1,1}, h_{1,2}]$ is the fading vector observed by the first receiver antenna, and $\underline{\mathrm{h}}_2 = [h_{2,1}, h_{2,2}]$ is the fading vector observed by the second receiver antenna. We now evaluate the capacity of this setup similar to before, noting that there are actually two time instances used

$$2C_{OTD} = \sup_{\substack{\mathrm{tr(Q)\leq P} \\ Q\geq 0}} \mathbb{E}_{\mathrm{G}}[\log\det(I_2 + \frac{\mathrm{G}Q\mathrm{G}^\star}{\sigma^2})] = \sup_{\substack{\mathrm{tr(Q)\leq P} \\ Q\geq 0}} \mathbb{E}_{\mathrm{G}}[\log\det(I_2 + \frac{\mathrm{G}^\star\mathrm{G}Q}{\sigma^2})]$$

$$\stackrel{(a)}{=} 2\mathbb{E}_{\mathrm{H}}[\log(1 + \frac{P}{2\sigma^2}(\|\underline{\mathrm{h}}_1\|^2 + \|\underline{\mathrm{h}}_2\|^2)] \Rightarrow C_{OTD} = \mathbb{E}_{\mathrm{H}}[\log(1 + \frac{P}{2\sigma^2}(\|\underline{\mathrm{h}}_1\|^2 + \|\underline{\mathrm{h}}_2\|^2)]. \tag{18}$$

Note that in (a) we have concluded that Q proportional to identity is optimal because of the concavity of $\log\det(\cdot)$ function. The last expression is exactly capacity of four transmit single receive antenna system. Now we evaluate the capacity in full generality, where the transmiter does not necessarily use OTD structure

$$C = \mathbb{E}_{\mathrm{H}}[\log\det(I_2+\frac{P}{2\sigma^2}\begin{bmatrix}\underline{\mathrm{h}}_1\\ \underline{\mathrm{h}}_2\end{bmatrix}[\underline{\mathrm{h}}_1^\star \;\; \underline{\mathrm{h}}_2^\star])] = \mathbb{E}_{\mathrm{H}}[\log\det(I_2+\frac{P}{2\sigma^2}\begin{bmatrix}\|\underline{\mathrm{h}}_1\|^2 & <\underline{\mathrm{h}}_1,\underline{\mathrm{h}}_2> \\ <\underline{\mathrm{h}}_2,\underline{\mathrm{h}}_1> & \|\underline{\mathrm{h}}_2\|^2\end{bmatrix}]$$

$$= \mathbb{E}_{\mathrm{H}}[\log(1+\frac{P}{2\sigma^2}(\|\underline{\mathrm{h}}_1\|^2+\|\underline{\mathrm{h}}_2\|^2)+(\frac{P}{2\sigma^2})^2(\|\underline{\mathrm{h}}_1\|^2\|\underline{\mathrm{h}}_2\|^2-|<\underline{\mathrm{h}}_1,\underline{\mathrm{h}}_2>|^2))]. \qquad (19)$$

Here, $<\cdot,\cdot>$ denotes Hermitian inner product. Since $\|\underline{\mathrm{h}}_1\|^2\|\underline{\mathrm{h}}_2\|^2 - |<\underline{\mathrm{h}}_1,\underline{\mathrm{h}}_2>|^2 \geq 0$ by the Cauchy-Schwarz inequality, we realize that for any channel realization, term inside the log of (19) is greater than or equal than term inside the log of (18). As long as $|<\underline{\mathrm{h}}_1,\underline{\mathrm{h}}_2>|^2 < \|\underline{\mathrm{h}}_1\|^2\|\underline{\mathrm{h}}_2\|^2$ with non-zero probability the strict inequality holds. This demonstrates that even though OTD is optimal in the case of a single receiver, it is not optimal if there is more of them. ∎

In order to find the exact amount of loss in this case, one needs to compare four transmit single receive antenna system with a two transmit two receive antenna system. Of course, it would depend on the distribution of fading coefficients. For Rayleigh distribution [1] gives a thorough treatment. The loss is about 3dB in the 20dB SNR range.

IV. Non-ergodic Channels and Outage Probability

In case that channel H is chosen randomly at the beginning of a codeword transmission frame and it is held constant during that particular frame, Shannon theoretic capacity is zero. No matter how small rate of transmission one chooses, there is always a finite probability that realized H and corresponding SNR will not be able to support that rate. This approximate model may be used for very slowly varying channels, which are also referred to as block fading channels. Hence, a single codeword is transmitted during a single channel state. However, we assume independent channel realizations between consecutive transmission frames.

Outage at a rate R is defined to be any event, dependent on the channel realization, for which the capacity of the resulting channel is less than R. For reasonable frame sizes, events of outage are necessarily followed by frame error rates. Of course, this is not exactly true for very small frame sizes. If one considers an extreme case of a frame being a single bit, then at least a frame error rate of a half can be obtained simply by guessing, even if the channel is in outage. However, by the converse to the Shannon coding theorem, if the transmission rate is above the channel capacity, then frame error rate goes exponentially to one as the frame length increases. For reasonable size frames we can safely assume this frame error rate to be exactly one, provided that the channel is in outage. Conversely, whenever the channel is not in outage, reliable transmission is possible by the Shannon capacity theorem. Therefore, provided that the number of symbols in a frame is large, there exist "good" codes for which the frame error rate is arbitrary close to the outage probability [10]. Clearly, for block fading channels, outage probability P_{out} is the ultimate benchmark for the frame error rate. It becomes very important to design systems which minimize this quantity. For the sake of simplicity, we consider multiple transmit and single receive antenna systems. To express P_{out}, we fix the available power to be P, and the desired rate of transmission to be R. In order to find the outage probability, we minimize the probability that channel capacity falls bellow R.

$$P_{\text{out}} = \inf_{\substack{\text{trQ} \leq \text{P} \\ Q \geq 0}} \mathbb{P}[\log(1 + \frac{\text{H}Q\text{H}^{\star}}{\sigma^2}) < R]. \tag{20}$$

Perhaps, an appropriate nomenclature for this quantity would be *achievable outage probability*, however, in consistency with [1], [9] we simply refer to this as the *outage probability.* Since P_{out} is an infimum of some expression taken over a particular set of Q's, we can easily evaluate an upper bound to P_{out} by considering Q to be diagonal with k entries being equal and $n-k$ zeros. The intuition behind this procedure is that we can turn antennas on and off based on the total power available. In addition to that it was conjectured in [1] that such power distribution would be optimal. Then we minimize this upper bound with respect to k. We obtain the following result:

Proposition 2: Assume we have n transmit and a single receive antenna available, together with the power constraint P. Also assume that the channel is Rayleigh distributed as well as block fading. An upper bound to outage probability (20) for the transmission rate R is given as

$$P_{\text{out}} \leq \min_{1 \leq k \leq n} \frac{(k(2^R-1))^k}{k!} (\frac{P}{\sigma^2})^{-k}. \tag{21}$$

This equation is very intuitively pleasing. Diversity level in fading channels corresponds

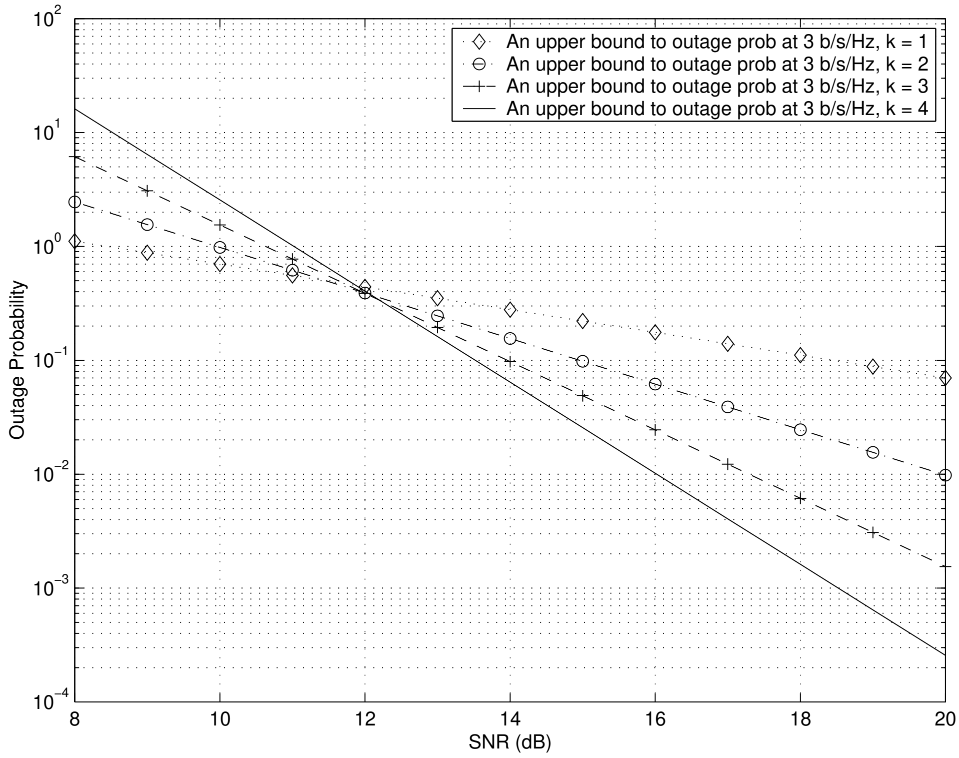

Fig. 3. Upper bound to the outage probability at 3 b/s/Hz is the lowest of the proposed curves at any SNR

to the inverse exponent of SNR in the expression for pairwise error probability. Here, we see all diversity levels up to n available in the outage probability expression, namely $1 \leq k \leq n$. SNR powered by these diversity levels gives us the diversity advantage $(\frac{P}{\sigma^2})^{-k}$. The first term is a constant independent of the SNR. We plot this upper bound in the Figure 3. It divides the plane in two regions. The actual outage probability curve lies in the region to the left of all the curves. Complement of this region containes achievable pairs of SNR's and frame error rates, but not all of them.

Proof: As mentioned before, P_{out} is the infimum of some quantity over all available Q as in (20). By considering Q to be diagonal with k entries beeing equal and $n-k$ zeros we can evaluate an upper bound. It follows from (20) that for every k we have an upper bound on outage probability as

$$P_{\text{out}} \leq \mathbb{P}[\log(1+\frac{P}{k\sigma^2}\sum_{i=1}^{k}\|\mathrm{h_i}\|^2) < R] \Rightarrow P_{\text{out}} \leq \mathbb{P}[\sum_{i=1}^{k}\|\mathrm{h_i}\|^2 < \frac{k(2^R-1)\sigma^2}{P}]$$

$$\Rightarrow P_{\text{out}} \leq \mathbb{P}[\sum_{i=1}^{k} ||\mathrm{h_i}||^2 < x]. \tag{22}$$

Note that we lumped some terms into one constant $x = \frac{k(2^R-1)\sigma^2}{P}$. Probability in (22) is essentially the cumulative distribution function of χ^2 (Chi squared) with $2k$ degrees of freedom [11].

$$P_{\text{out}} \leq 1 - e^{-x} \sum_{l=0}^{k-1} \frac{x^l}{l!}, \tag{23}$$

where the bound on the right is a cumulative distribution function of χ^2. Note that the summation is $k-1$-st order Taylor polynomial which approximates the exponential function, and with the help of the remainder we can reduce the expression to

$$P_{\text{out}} \leq 1 - e^{-x}(e^x - \sum_{l=k}^{\infty} \frac{x^l}{l!}) = e^{-x} \sum_{l=k}^{\infty} \frac{x^l}{l!} \leq \frac{x^k}{k!}, \tag{24}$$

where the last inequality is easily proven by multiplying through by e^x, expanding it, and noting that $\frac{1}{(k+i)!} \leq \frac{1}{k!i!}$. Now, using the definition for x we obtain

$$P_{\text{out}} \leq \frac{x^k}{k!} = \frac{(k(2^R-1))^k}{k!}(\frac{P}{\sigma^2})^{-k}. \tag{25}$$

Since this is true for every k , $1 \leq k \leq n$ then (21) follows. ■

Note that in Fig. 3 the slope of the outage increases with more available antennas, for sufficiently high SNR. We know that for ergodic channels the capacity cutoff curve has infinite slope on frame error rate versus SNR plot. Now we can see how the non ergodic outage curve approaches the ergodic curve as diversity level i.e, slope, increases. Also, it is apparent that, *for high SNR*, the diversity term $(\frac{P}{\sigma^2})^{-k}$ is the most important in the minimization and the above bound becomes

$$P_{\text{out}} \leq \frac{(n(2^R-1))^n}{n!}(\frac{P}{\sigma^2})^{-n}. \tag{26}$$

Furthermore, we will show that for equal power distribution among antennas, this bound becomes asymptotically tight in SNR. Note that for high SNR, equal power distribution is suggested by the symmetry of the problem and the law of large numbers. Following this reasoning all space-time codes proposed thus far [3], [4], [5], [12], [13] assume equal power distribution among antennas.

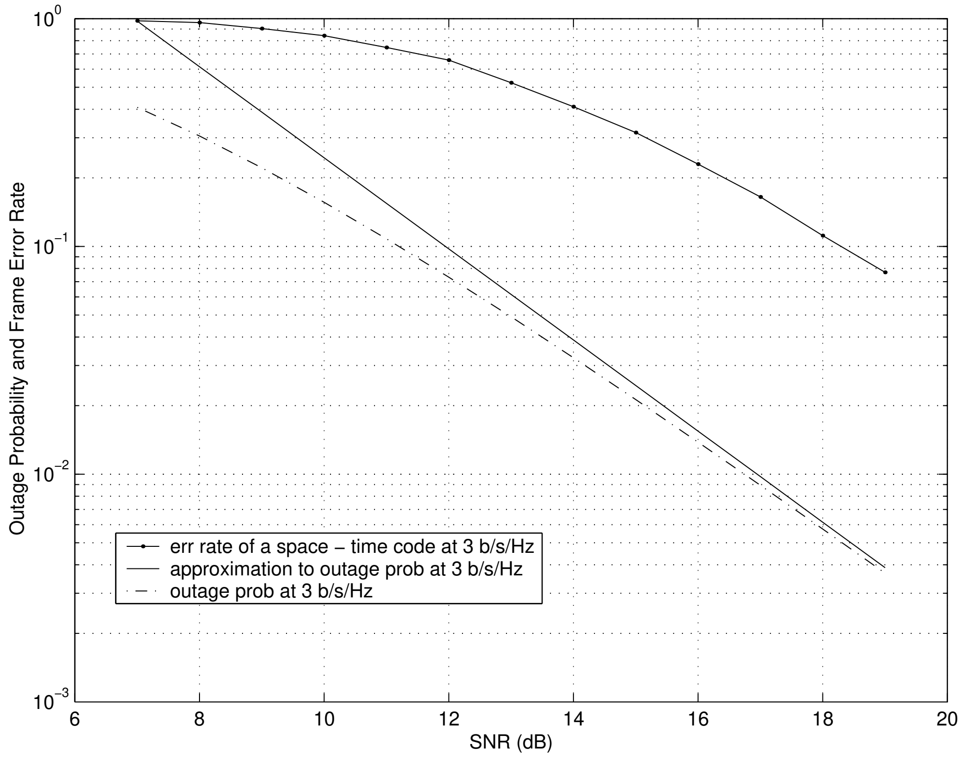

Fig. 4. Outage Probability and Space-Time Code Frame Error Rate at 3 b/s/Hz

Proposition 3: Assuming equal power distribution among transmit antennas the bound in (26) becomes asymptotically tight in SNR.

$$\lim_{P\to\infty} P_{\text{out}} P^n = \frac{(n(2^R-1)\sigma^2)^n}{n!}. \tag{27}$$

We write this as

$$P_{\text{out}} \sim \frac{(2^R-1)^n}{n!}(\frac{P}{n\sigma^2})^{-n}. \tag{28}$$

This result will become important later when we relate the outage probability with the space-time code frame error rate.

Proof: Starting from the definition in (20), we make use of the equal power assumption and the cumulative distribution function of χ^2 to obtain the following expression

$$P_{\text{out}} = \mathbb{P}[\sum_{i=1}^{n} \|\mathrm{h_i}\|^2 < \frac{n(2^R-1)\sigma^2}{P}] = 1 - e^{-x}\sum_{l=0}^{n-1}\frac{x^l}{l!}, \tag{29}$$

where $x = \frac{n(2^R-1)\sigma^2}{P}$. It follows that $P = \frac{n(2^R-1)\sigma^2}{x}$ and

$$\lim_{P\to\infty} P_{\text{out}} P^n = (n(2^R-1)\sigma^2)^n \lim_{x\searrow 0}[1 - e^{-x}\sum_{l=0}^{n-1}\frac{1}{l!}x^l]x^{-n}$$

$$
\begin{aligned}
&\overset{(a)}{=}(n(2^R-1)\sigma^2)^n \lim_{x\searrow 0} e^{-x}(\sum_{l=n}^{\infty}\frac{x^l}{l!})x^{-n} \\
&\overset{(b)}{=}(n(2^R-1)\sigma^2)^n \sum_{l=n}^{\infty}\lim_{x\searrow 0} e^{-x}\frac{x^{l-n}}{l!} \\
&\overset{(c)}{=}\frac{(n(2^R-1)\sigma^2)^n}{n!},
\end{aligned}
\tag{30}
$$

where (a) is obtained using the remainder of Taylor series for the exponential, and (b) is by interchanging the limit and the summation. Note that we are allowed to do so since the summation converges uniformly as a function of x in some neighborhood of 0. (c) follows because now there is only one nonzero term in the summation. ■

This result says that the outage probability behaves as inverse SNR to the full diversity power and we realize that the bound we derived is asymptotically tight. This becomes very apparent by looking at Fig. 4 which depicts asymptotic outage probability and the frame error rate of a space time code on the same plot. Note that asymptotically they have the same slope; however, there is a constant gap between the two curves. More complex and better space time codes attempt to close this gap with the "coding advantage" [3].

It is very interesting to algebraically establish the relationship between the result from Proposition 3 and the performance criterion for space time codes in block fading channels. Assume that the matrix of constellation symbols C was transmitted over n transmit antennas. The pairwise frame error probability $\mathbb{P}[C \to E]$ is defined as the probability that the maximum likelihood receiver will decide that E was transmitted. It was demonstrated in [3] that for Rayleigh faded channel

$$
\mathbb{P}[C \to E] \le \frac{4^n}{d^2(C,E)}(\frac{P}{n\sigma^2})^{-n}, \tag{31}
$$

for the unit energy constellation on each antenna. Here $d^2(C,E)$ is the product of all non zero eigenvalues of $(C-E)^\star(C-E)$. Minimum of this quantity over all (C,E) pairs is named d^2_{min}, or the *coding advantage.* For a good code we want this quantity to be as large as possible - in order to minimize the frame error probability. It is very intuitively pleasing to note the same SNR behavior of the outage probability in (28) and the pairwise frame error probability in (31).

V. Outage Probability for OTD

We have demonstrated the optimality of OTD in fast fading channels, however, many space-time codes were originally designed for block fading channels [7], [6], [12], [13]. In addition to that, many of them use OTD, together with some outer code. Since capacity analysis does not apply in block fading environment, it is instructive to examine the achievable outage probability when the transmitter uses OTD. In the case of a singe receive antenna, the result is again satisfactory.

Theorem 2: If we assume equal power distribution among antennas and a single receive antenna then OTD preserves the achievable outage probability described earlier.

Proof: If we assume equal power distribution then the outage probability becomes

$$P_{\text{out}} = \inf_{\substack{\mathrm{trQ} \leq \mathrm{P} \\ Q \geq 0}} \mathbb{P}[\log(1 + \frac{\mathrm{H}Q\mathrm{H}^\star}{\sigma^2}) < R] = \mathbb{P}[\log(1 + \frac{P}{2\sigma^2}(\|\mathrm{h}_1\|^2 + \|\mathrm{h}_2\|^2) < R]. \tag{32}$$

Referring to Fig. 2, and using the notations as in (10), if the transmitter uses OTD then the outage probability becomes

$$\begin{aligned} P_{\text{out,OTD}} &= \mathbb{P}[\log\det(I_2 + \frac{\mathrm{G}Q\mathrm{G}^\star}{\sigma^2}) < 2R] = \mathbb{P}[\log\det(I_2 + \frac{\mathrm{G}^\star\mathrm{G}Q}{\sigma^2}) < 2R] \\ &= \mathbb{P}[\log\det(I_2 + \frac{P}{2\sigma^2}(\|\mathrm{h}_1\|^2 + \|\mathrm{h}_2\|^2)I_2) < 2R] = \mathbb{P}[\log(1 + \frac{P}{2\sigma^2}(\|\mathrm{h}_1\|^2 + \|\mathrm{h}_2\|^2)^2) < 2R] \\ &= \mathbb{P}[2\log(1 + \frac{P}{2\sigma^2}(\|\mathrm{h}_1\|^2 + \|\mathrm{h}_2\|^2) < 2R]. \end{aligned} \tag{33}$$

Note that we have used rate of transmission of $2R$ on the right because the OTD transmission is made over two time instants. The last equation is equivalent to (32). ■

Indeed, the proof is similar as in fast fading channels. Following the same argument as in fast fading channels, we could discuss the multiple receive antenna case and conclude that the achievable outage probability would actually increase. As before, the argument boils down to Cauchy-Schwarz inequality. This implies that OTD is not optimal in the case of multiple receive antennas, even for block fading channels. However, some of the existing systems have two transmit and one receive antenna, and the fact that OTD preserves achievable outage probability in this setup gives further justification for its application in block fading channels. Some other properties of OTD are that it provides full diversity space-time codes and ensures a good coding advantage. More detailed work on this can be found in [12] and [13].

VI. Conclusion

At this point we should note the robustness of the OTD scheme to the nature of fading, for the case of a single receive antenna. In ergodic channels it does not decrease the capacity, while in non-ergodic channels it does not increase the achievable outage probability. However, it is somewhat disappointing that this is not the case for multiple receivers. This brings up another somewhat peculiar issue. The optimality of a *transmit scheme* depends on the number of *receivers*, which is very uncommon. Does this perhaps mean that some code construction criteria should depend on the number of receivers?

In addition to this OTD analysis, we have proposed a simple and asymptotically tight upper bound on the outage probability for non-ergodic channels. We believe that it is a good practice to compare frame error rates of space time codes to this bound in order to obtain a simple information theoretic benchmark.

VII. ACKNOWLEDGEMENT

Authors thank K.K Mukavilli, M.J. Borran and S.Rajagopal for extremly valuable comments and discussions.

References

[1] I.E.Telatar, "Capacity of Multi-Antenna Gaussian Channels,", AT&T-Bell Laboratories Internal Tech. Memo., June 1995.

[2] T.L.Marzetta and B.M. Hochwald "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading Environment," *IEEE Trans. Inform. Theory,* vol. 45, pp.139-158, Jan. 1999.

[3] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction," *IEEE Trans. Inform. Theory*, vol.44, pp.744–765, March 1998.

[4] A. R. Hammons and H. E. Gamal "On the Theory of Space-Time Codes for PSK Modulation," *IEEE Trans. Inform. Theory*, vol. 2, no. 2, pp. 524–542, March 2000.

[5] J. Grimm, M.P. Fitz, and J.V. Krogmeier, "Further Results on Space-Time Coding for Rayleigh Fading," Proc. of the 1998 Allerton Conference, Monticello, IL, 1998, pp. 391-400.

[6] S. M. Alamouti, "A Simple Diversity Scheme for Wireless Communications," *IEEE J. Select. Areas Commun.,* vol. 16, pp.1451-1458, Oct. 1998.

[7] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-Time Block Codes From Orthogonal Designs," *IEEE Trans. Commun.,* vol. 36, pp.1004-1012, Sep. 1988.

[8] T.M. Cover and J.A. Thomas, *Elements of Information Theory,* John Wiley and Sons, Inc., 1991.

[9] L.H.Ozarow and S.Shamai and A.D.Wyner, "Information Theoretic Considerations for Cellular Mobile Radio," *IEEE Transactions on Vehic. Tech.,* vol. 43, pp. 359-378, May 1994.

[10] R. Knopp and P.A.Humblet "On Coding for Block Fading Channels," *IEEE Trans. Inform. Theory,* vol. 46, pp.189-205, Jan. 2000.

[11] J. G. Proakis *Digital Communications,* McGraw-Hill, Inc., 1995.

[12] K. K. Mukkavilli, M. Ionescu, and B. Aazhang, "Design of Space-Time Codes With Optimum Coding Gain" *Proc. PIMRC2000*

[13] M. J. Borran, M. Memerzadeh and B. Aazhang, "Desig of Coded Modulation Schemes for Orthogonal Transmit Diversity" *Submitted to Intl. Symposium on Inform. Theory* June 20001.

[14] Wittenben "A new Bandwidth Efficient Transmit Antenna Modulation Diversity Scheme for Linear Digital Modulation," *in IEEE ICC,* 1993 pp. 1630-1634

[15] X. Li and J. A. Ritcey, "Bit-Interleaved Coded Modulation with Iterative Decoding," *in IEEE ICC*, vol. 2, pp. 858–863, June 1999.

[16] Y. Liu, M. P. Fitz, and O. Y. Takeshita, "Qpsk Space-Time Turbo Codes," *in IEEE ICC*, June 2000.

[17] X. Li and J. A. Ritcey, "Bit-Interleaved Coded Modulation with Iterative Decoding Using Soft Feedback," *Electronic Letters*, vol. 34, pp. 942–943, 4 March 1998.

[18] X. Li and J. A. Ritcey, "Trellis-Coded Modulation With Bit Interleaving and Iterative Decoding," *IEEE Journal on Select. Areas in Comm.*, vol. 17, pp. 715–724, April 1999.

[19] X. Li and J. A. Ritcey, "Bit-Interleaved Coded Modulation with Iterative Decoding," *IEEE Comm. Letters*, vol. 1, pp. 169–171, November 1997.

[20] D. Divsalar and M. K. Simon, "The Design of Trellis Coded MPSK for Fading Channels: Performance Criteria," *IEEE Trans. Inform. Theory,* vol. 45, pp.1456-1467, July 1999.

[21] A. Narula, M. D. Trott and G. W. Wornell "Performance Limits of Coded Diversity Methods for Transmitter Antenna Arrays," *IEEE Trans. Inform. Theory,* vol. 45, pp.2418-2433, Nov. 1999.

[22] E. Biglieri, J. Proakis, and S. Shamai "Fading Channel: Information-Theoretic and Communications Aspects" IEEE Transactions on Information Theory, vol. 44, pp.2619-2692, Oct. 1998.

[23] B.M. Hochwald and T.L.Marzetta "Unitary Space-time modulation for multiple-antenna communications in rayleigh flat fading," *IEEE Trans. Inform. Theory,* vol. 46, pp.543-564, March. 2000.

[24] T. Muharemovic and B. Aazhang, "Information Theoretic Optimality of Orthogonal Space Time Transmission Schemes and Concatenated Code Construction," Proc. 2000 Intl. Conf. Telecommunications, May 2000.