Information Theoretic Optimality of Orthogonal Space-Time Transmit Schemes and Concatenated Code Construction

Tarik Muharemovic and Behnaam Aazhang

Abstract— **We demonstrate that the information theoretic capacity of multiple-transmit antenna systems is not reduced by using space-time orthogonal transmit schemes as means of exploiting the transmit diversity. Then, we consider concatenation of a channel code with space-time orthogonal transmit scheme, where data stream is first encoded with the channel code, and then transmitted using orthogonal space-time transmit scheme. Using this orthogonal space time transmit scheme the elements in the product criterion for code construction are made more uniform and hence better gain is achieved. We derive code construction criterion for designing the channel code, taking into consideration coherence time of the channel.**

Keywords— **Transmit diversity, space-time codes, coherence time**

I. Summary

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. Here we consider case of antenna diversity, where spatially separated antennas at the transmitter side are available. Generalization to multiple antennas on both transmitter as well as receiver side is straightforward. Information theoretic capacity of this multiple-transmit multiple-receive antenna scheme was derived in [2], where fading coefficients between antennas were assumed to be iid Rayleigh distributed and the coherence time of the channel corresponded to one symbol. This derivation was extended for the coherence time larger than one symbol in [3]. It is easily shown that there is no improvement in the achievable information rate for larger coherence times. Several practical schemes were proposed to utilize this additional degree of freedom on the transmitter side [4]-[7]. Delay diversity transmit scheme was considered in [4], and subsequently generalized to space-time trellis codes [5]. Similarly, another simple diversity technique based on orthogonal matrices was proposed in [6], and generalized to space-time block codes from orthogonal designs in [7]. It is generally accepted that space-time trellis coder perform better than the space-time block codes, however here we demonstrate that if one uses orthogonal space time transmit scheme as transmit diversity technique there is no loss in the achievable information rate. In addition to that, space-time block codes (orthogonal space time transmit scheme is an example of a space time block code) have much lower decoding complexity than the space-time trellis codes [7]. Then, we consider concatenation of a channel code with the space-time orthogonal transmit scheme and derive the criterion for code design, taking into consideration coherence time of the channel. It is essential that the concatenation is done carefully so that channel code is designed optimally with respect to the orthogonal space time transmit scheme. Hence, we demonstrate that the orthogonal space time transmit scheme represent a simple low complexity transmit scheme with no impact on the channel capacity. Distance metric between two codeword symbols is derived to be the product of Euclidean distances taken within the coherence time, and based on that we propose a code design criterion.

All 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.

II. Information Theoretic Considerations

Consider the following system description: there are t transmit antennas available at transmitter side, and there is only one receiver antenna (taken only for simplicity of consideration and can be easily generalized to r receive antennas). The receiver antenna observes raw summation of randomly faded transmitted symbols, corrupted by AWGN. All fading coefficients are iid. In other words we deal with the channel model where the received symbol Y depends on the transmitted vector $\underline{X}$ via

$$Y = \underline{H}^T \underline{X} + N, \tag{1}$$

where $\underline{H} = [h_1 ... h_t]^T$ is a complex random vector which represents the channel, and N is complex AWGN. The power constraint is given as

$$E[\underline{X}^H \underline{X}] \leq P \tag{2}$$

Here, $\underline{X}^H$ denotes $\underline{X}$ conjugate transposed. We assume that the receiver observes fading coefficients, and therefore both Y and $\underline{H}$ are available at the output. Proceeding as in [2] we want to maximize the mutual information $I(\underline{X}; (Y, \underline{H}))$.

$$I(\underline{X}; (Y, \underline{H})) = I(\underline{X}; \underline{H})) + I(\underline{X}; Y|\underline{H})) = I(\underline{X}; Y|\underline{H})) \tag{3}$$

since the transmitter has no knowledge on the fading coefficients. Conditioned on the fade, channel is nothing more than attenuated AWGN channel, and averaging over all possible fades

$$I(\underline{X}; Y|\underline{H})) = E_{\underline{H}}(\log(1 + \frac{P}{t}\underline{H}^H \underline{H}) \tag{4}$$

ISBN 968-36-7763-0 / 22-25 May 2000 © ICT 2000

and therefore the achievable rate of the channel described above is given as

$$C = E_{\underline{H}}(\log(1 + \frac{P}{t}\underline{H}^H\underline{H}) \text{ bits per channel use} \tag{5}$$

which was derived in [2]. Note that the above analysis is under assumption that the coherence time of the channel corresponds to one transmission. In [3] this assumption we easily extended to the case when the coherence time is $T_c > 1$. It was derived that

$$C = T_c E_{\underline{H}}[\log(1 + \frac{P}{t}\underline{H}^H\underline{H})] \text{ per } \mathrm{T_c}\text{transmissions} \tag{6}$$

which means that per one transmission we essentially have (5). What we will do next is compare space time orthogonal block codes with (6). We will conclude that if one uses these codes as a transmit diversity scheme *there is no loss in the achievable information rate*. We will consider a simple transmission scheme [6], which uses two transmit antennas. This analysis is subsequently generalized to more complex orthogonal space time transmission schemes. For the case of two transmit antennas, we assume that the coherence time of the channel is two symbols, and if it happens to be larger we perform a sufficient ammount of interleaving.

The scheme is described as follows: at the first time instant symbol X_1 is transmitted on the antenna 1 and X_2 on the antenna 2. At the next time instant the antenna 1 transmits $X_2^\star$ and the antenna 2 transmits $-X_1^\star$. Hence, the receiver receives

$$Y_1 = h_1X_1 + h_2X_2 + N_1$$

at the first time instant, and

$$Y_2 = h_1X_2^\star - h_2X_1^\star + N_2.$$

at the second time instant. One can easily verify that the power constraint (2) is still pertinent to this transmission scheme, because if symbols $\{X_1^{(1)}, X_2^{(1)}, X_1^{(2)}, X_2^{(2)}, ..., X_1^{(n)}, X_2^{(n)}\}$ (superscript refers two time instants, total of 2n transmit time instants because of repetition) enter the block encoder the power transmitted by each antenna is given as

$$\frac{1}{2n}\sum_{i=1}^{n}|X_1^{(i)}|^2 + |X_2^{(i)}|^2 \tag{7}$$

which means that the total transmitted power is given as

$$\frac{1}{n}\sum_{i=1}^{n}|X_1^{(i)}|^2 + |X_2^{(i)}|^2 \xrightarrow{LLN} E[|X_1|^2 + E[|X_2|^2], \tag{8}$$

Where LLN refers to law of large numbers, and convergence is in probability. Hence the power constraint (2) is still a valid constraint, namely

$$E[|X_1|^2 + E[|X_2|^2] \leq P \tag{9}$$

We can compactly write the received vector $\underline{Y}$ as

$$\begin{bmatrix} Y_1 \\ Y_2^\star \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ -h_2^\star & h_1^\star \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} N_1 \\ N_2^\star \end{bmatrix} \tag{10}$$

or in matrix notation

$$\underline{Y}' = G\underline{X} + \underline{N} \tag{11}$$

Here, $\underline{Y}'$ is the vector $\underline{Y}$ with even components conjugated and

$$G = \begin{bmatrix} h_1 & h_2 \\ -h_2^\star & h_1^\star \end{bmatrix} \tag{12}$$

is essentially Radon-Hurwitz transform of the vector of fading coefficients $\underline{H}$ [7]. Another interesting observation is that using this transmit scheme, symbols X_1 and X_2 are spread across both time and antennas in an orthogonal manner. We evaluate the capacity of the above scheme, again assuming that the receiver perfectly knows the fading coefficients. The mutual information is given as

$$I(\underline{X}; (G, \underline{Y})) = I(\underline{X}; (G, \underline{Y'})) = I(\underline{X}; G) + I(\underline{X}; \underline{Y'}|G). \tag{13}$$

Again we conclude that the first term is zero (transmitter has no knowledge of the channel) and the second term is averaged over G, similar to (4). Hence the capacity of the scheme proposed in [6] is given as

$$E_G[\log\det(I + GRG^H)] \tag{14}$$

This is again a standard formula where the term inside the expectation is the capacity for the deterministic channel with the transfer function G, and can be found in any reference which considers capacity of fading channels, like [2]. The matrix R_{ij} is cross-correlation matrix between each two consecutive input symbols, $R_{ij} = E[X_iX_j]$. Power constraint is still given as (9). We expand (14) as

$$C = E_G[\log\det(I + \begin{bmatrix} h_1 & h_2 \\ -h_2^\star & h_1^\star \end{bmatrix} \begin{bmatrix} R_{11} & R_{21} \\ R_{12} & R_{22} \end{bmatrix} \begin{bmatrix} h_1^\star & -h_2 \\ h_2^\star & h_1^\star \end{bmatrix})] \tag{15}$$

Using the common matrix identity

$$\det(I + AB) = \det(I + BA) \tag{16}$$

We obtain

$$C = E_G[\log\det(I + \begin{bmatrix} h_1^\star & -h_2 \\ h_2^\star & h_1^\star \end{bmatrix} \begin{bmatrix} h_1 & h_2 \\ -h_2^\star & h_1^\star \end{bmatrix} \begin{bmatrix} R_{11} & R_{21} \\ R_{12} & R_{22} \end{bmatrix})] \tag{17}$$

Hence

$$C = E_G[\log\det(I + (|h_1|^2 + |h_2|^2) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_{11} & R_{21} \\ R_{12} & R_{22} \end{bmatrix})] \tag{18}$$

In order to maximize the above expression the matrix R should be diag$[P/2, P/2]$ (can be proven using Hadamard's inequality). Hence,

ISBN 968-36-7763-0 / 22-25 May 2000 © ICT 2000

$$C = E_G[\log\det(I + \frac{P}{2}(|h_1|^2 + |h_2|^2)I)]$$
$$= E_G[\log(1 + \frac{P}{2}(|h_1|^2 + |h_2|^2))^2)$$
$$= 2E_G[\log(1 + \frac{P}{2}(|h_1|^2 + |h_2|^2))] \quad (19)$$

which is exactly (6). Of course, this is per two transmit time instants, and if we normalize properly, we obtain (5). Hence, we conclude that *the achievable information rate is not reduced by using the described transmission scheme.* In addition to that, this scheme was proposed as low computational complexity at the receiver [6], [7] because it alows the receiver to decouple the estimate for symbols X_1 and X_2.

III. Generalizations

Here we generalize the above analysis for arbitrary 2^n transmit antennas. We again assume that the coherence time is at least 2^n. We first make a generalization of the transmission scheme proposed in [6]. Given a vector $\underline{X}$ of length 2^n to be transmitted, we decompose it into vectors $\underline{X}_1$ and $\underline{X}_2$, each of length 2^{n-1}. Each of these two vectors is sent similar to (6)

$$\underline{Y}_1 = G_1\underline{X}_1 + G_2\underline{X}_2 + \underline{N}_1$$
$$\underline{Y}_2 = G_1\underline{X}_2^\star - G_2\underline{X}_1^\star + \underline{N}_2$$

Here, G_1 and G_2 are again matrixes of fading coefficients, and $\underline{X}_2^\star$ denotes vector X conjugated (but not transposed). This can compactly we written in the form

$$\begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2^\star \end{bmatrix} = \begin{bmatrix} G_1 & G_2 \\ -G_2^\star & G_1^\star \end{bmatrix} \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix} + \begin{bmatrix} \underline{N}_1 \\ \underline{N}_2^\star \end{bmatrix} \quad (20)$$

or in the matrix notation

$$\underline{Y}' = G\underline{X} + \underline{N} \quad (21)$$

We continue partitioning, now vectors $\underline{X}_1$ and $\underline{X}_2$. Because of the iterative construction, the matrix G is orthogonal, meaning that

$$G^H G = \sum_{j=1}^{2^n} |h_j|^2 I \quad (22)$$

Where, $t = 2^n$ is the number of transmit antennas. The capacity of the generalized space time orthogonal transmission scheme is

$$E_G[\log\det(I + GRG^H)] \quad (23)$$

which, because of the orthogonal structure of G (and using the same argument as in the case of two antennas) is

$$tE_G[\log(1 + (P/t)\sum_{j=1}^{t} |h_j|^2)] \quad (24)$$

and, if normalized properly (per one transmission) is the same as (4). Hence, again the capacity stays the same.

IV. Comparison with multicode CDMA

One of the common questions is given that there are two transmit antennas available, how can one use this additional comodity to increase the data rate. Orthogonal space-time transmission scheme spreads symbols both across time and antennas, as opposed to multicode CDMA scheme which assigns different spreading sequences to each antenna [9]. Here we demonstrate that if we were to orthogonalize between antennas in a multicode CDMA manner we would decrease the achievable information rate. In order to have a fair comparison with the above transmit scheme, number of transmit antennas again has to be $t = 2^n$. We assume length of the spreading code would be less than the coherence time so the transmitted waveforms would stay orthogonal after passing through the channel. The spreading length is equal to the number of transmit antennas, so we don't have any loss in degrees of freedom. Hence we have a channel which can be described as

$$\underline{Y} = \sum_{j=1}^{t} h_j \underline{s_j} X_j + \underline{N} \quad (25)$$

where $\underline{s_j}$ is the spreading sequence of the j-th user (length t), X_j is the transmitted symbol of the user j, and the h_j is the fading coefficient of th j-th user. We can write the above equation as follows

$$\underline{Y} = SU\underline{X} + \underline{N} \quad (26)$$

where S is the matrix whose columns are spreading sequences of unit energy, and U is the diagonal matrix of fading coefficients. Considering again the power constraint, the average power transmitted on all t antennas is

$$\frac{1}{t}\sum_{j=1}^{t} E[|X_j|^2] \leq P \quad (27)$$

or

$$\sum_{j=1}^{t} R_{jj} \leq tP \quad (28)$$

The capacity is given as follows

$$C = E_U[\log\det(I + (SU)R(SU)^H))] \quad (29)$$

Now, using (16)

$$C = E_U[\log\det(I + S^H SURU^H)] \quad (30)$$

Assuming that the spreading sequences are orthogonal we have $S^H S = I$. Then,

$$C = E_U[\log\det(I + URU^H)] \quad (31)$$

which is maximized when R is diagonal because logdet is concave on positive definite matrixes [2]. The maximum value is

$$C = E_U[\log\det(I + PU^H U)] \quad (32)$$

ISBN 968-36-7763-0 / 22-25 May 2000 © ICT 2000

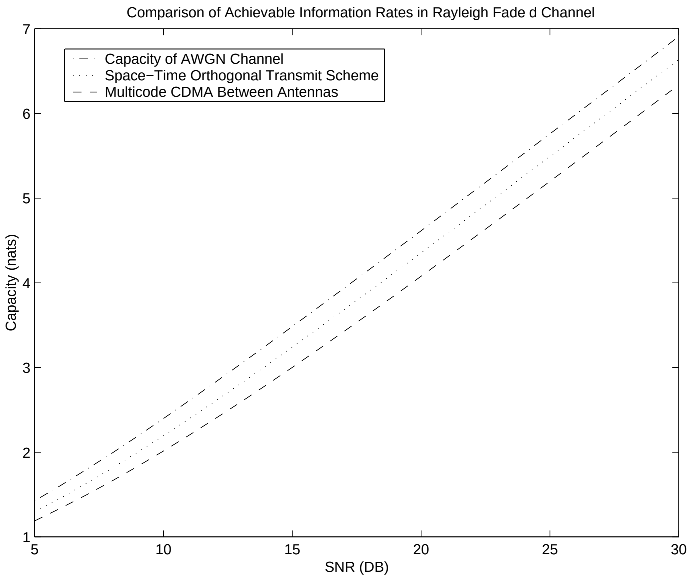

Fig. 1. Achievable Information Rates

However, since $U^H U = \text{diag}[|h_1|^2 \ldots |h_t|^2]$ then

$$C = \sum_{j=1}^{t} E_U[\log\det(1 + P|h_j|^2)] \tag{33}$$

Since, h_j are identically distributed, then

$$C = tE_h[\log(1 + P|h|^2)] \tag{34}$$

per t transmissions (spreading length t). If we compare this equation with (24) we see that the capacity with multicode orthogonalization is strictly less because $\log(1 + x)$ is a concave function. For the case of two transmit one receive antennas in iid Rayleigh faded channel, the above expressions are calculated in [2], and we plotted them in Figure 1.

V. Concatenated Code Construction Criterion

Inspired by the achievable rates in the previous section we now consider practical codes to try to reach the capacity. We will again do the analysis for the two transmit antennas, however it can be easily generalized for other powers of 2 (which allow the orthogonal transmit scheme [6]). We proceed similar to [8], [5], assuming that the symbol stream $\{x_1^{(1)}, x_2^{(1)}, ..., x_1^{(l)}, x_2^{(l)}\}$ is concatenated with the orthogonal block transmission scheme, we upper bound the probability that it is mistaken for $\{e_1^{(1)}, e_2^{(1)}, ..., e_1^{(l)}, e_2^{(l)}\}$. Here each two consecutive symbols are transmitted using the space time orthogonal transmission scheme i.e. symbols $\{x_1^{(1)}, (-x_2^{(1)})^\star, ..., x_1^{(l)}, (-x_2^{(l)})^\star\}$ are transmitted on the first antenna, while $\{x_2^{(1)}, (x_1^{(1)})^\star, ..., x_2^{(l)}, (x_1^{(l)})^\star\}$ is transmitted on the second antenna. We assume that coherence time of the channel corresponds to two symbols, and each consecutive fade is independent (for both of the antennas). Proceeding as in [8], [5] we calculate the Chernoff upper bound for

$$\text{Prob}\{X \to E|\alpha\} \le e^{-d^2(X;E|\alpha)(E_s/4N_0)} \tag{35}$$

where α are all fading coefficients (each proper complex gaussian, zero mean variance 0.5 per dimension), and $E_s/4N_0$ is properly normalized SNR. It is easy to verify that

$$\begin{aligned} d^2(x_1^{(t)}, x_2^{(t)}; e_1^{(t)}, e_2^{(t)}|\alpha) = \\ |\alpha_1^{(t)}(x_1^{(t)} - e_1^{(t)}) + \alpha_2^{(t)}(x_2^{(t)} - e_2^{(t)})|^2 + \\ +|\alpha_2^{(t)}(x_1^{(t)} - e_1^{(t)})^\star - \alpha_1^{(t)}(x_2^{(t)} - e_2^{(t)})^\star|^2 \end{aligned} \tag{36}$$

Since the cross terms cancel out, it reduces to

$$(|\alpha_1^{(t)}|^2 + |\alpha_2^{(t)}|^2)[|x_1^{(t)} - e_1^{(t)}|^2 + |x_2^{(t)} - e_2^{(t)}|^2] \tag{37}$$

And hence if we assume each two consecutive fades independent, because the coherence time is two symbols, we obtain

$$d^2(X;E|\alpha) = \sum_{t=1}^{l}(|\alpha_1^{(t)}|^2 + |\alpha_2^{(t)}|^2)[|x_1^{(t)} - e_1^{(t)}|^2 + |x_2^{(t)} - e_2^{(t)}|^2] \tag{38}$$

or in simplified notation

$$d^2(X;E|\alpha) = \sum_{t=1}^{l}|\beta(t)|^2 D(t) \tag{39}$$

$$\text{Prob}\{X \to E|\alpha\} \le \exp[-\sum_{t=1}^{l}|\beta(t)|^2 D(t)(E_s/4N_0)] \tag{40}$$

Since, $|\beta|^2$ is the sum of four Gaussian random variables with 0 mean and 0.5 variance, than by avearging the above equation with respect to α we obtain

$$\text{Prob}\{X \to E\} \le \prod_{t=1}^{l}[1 + (|x_1^{(t)} - e_1^{(t)}|^2 + |x_2^{(t)} - e_2^{(t)}|^2)(E_s/4N_0))]^{-2} \tag{41}$$

We can drop the 1 and get

$$\text{Prob}\{X \to E\} \le \prod_{x^{(t)} \ne e^{(t)}}[|x_1^{(t)} - e_1^{(t)}|^2 + |x_2^{(t)} - e_2^{(t)}|^2)(E_s/4N_0)]^{-2} \tag{42}$$

Hence, what we propose is to treat (x_1^t, x_2^t) as elements of multidimensional constalation with large Euclidean distance among them. On top of that we encode them with TCM which has large Hamming distance (because of the product).

The exponent (-2) reflects optimisation using the coherence time of the channel because it makes each two consecutive factors in the product distance equal, thereby increasing the product distance. This is the avantage of the space time orthogonal transmission scheme. This additional gain comes from considering the coherence time.

ISBN 968-36-7763-0 / 22-25 May 2000 © ICT 2000

VI. CONCLUSION

We demonstrated that the achievable information rate is not reduced by using orthogonal transmission schemes such as the one proposed by Alamouti [6], as a way of exploiting transmit diversity. In addition to that we show that orthogonalization among antennas in a CDMA fashion is not optimal because there is a decrease in the achievable information rate. Then we consider concatination of a channel code with the above orthogonal transmission scheme and derive the code construction criterion.

VII. ACKNOWLEDGEMENT

The authors thank Krishna K. Mukkavilli and Mohammad J. Borran for extremly valuable comments and discussions.

REFERENCES

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

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels,", AT&T-Bell Laboratories Internal Tech. Memo., June 1995.

[3] 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.

[4] Wittenben "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," *Proc IEEE'ICC,* 1993 pp. 1630-1634

[5] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: Performance analysis and code construction," *IEEE Trans. Inform. Theory,* vol. 44, pp.199-207, Feb. 1999.

[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] 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.

[9] J. Chen and U. Mitra, "Optimum Near-Far Resistance for Dual-Rate DS/CDMA Signals: Random Signature Sequence Analysis, *IEEE Transactions On Information Theory,* vol. 45, No. 7, November 1999.