DIVERSITY VS. ADAPTIVITY IN MULTIPLE ANTENNA SYSTEMS

Ronald Böhnke, Volker Kühn, and Karl-Dirk Kammeyer

Department of Communications Engineering
University of Bremen
Otto-Hahn-Allee, 28359 Bremen, Germany
Email: {boehnke, kuehn, kammeyer}@ant.uni-bremen.de

ABSTRACT

An information theoretic analysis of multiple antenna systems shows that the possible gain due to transmitter-sided channel knowledge is usually small for uncorrelated fading. However, it is not obvious how these results translate to real-world transmission strategies. Thus, in this paper the error performance of adaptive bit and power loading with perfect channel state information is compared to that of high-rate multistratum space-time codes relying only on transmit diversity. We demonstrate that the latter may even outperform the adaptive scheme if discrete rate constraints are taken into account.

1. INTRODUCTION

Multiple antenna or MIMO systems will be a key technology in future wireless communication systems as they enable very high data rates with limited bandwidth. This is especially true for indoor environments with rich scattering. With perfect channel knowledge at both ends of the link, the MIMO channel can be decomposed into parallel subchannels by appropriate precoding. This facilitates the application of adaptive modulation and power allocation techniques that were originally designed for discrete multitone transmission or orthogonal frequency division multiplexing (OFDM) [1, 2]. Without the rather optimistic assumption of ideal channel information at the transmitter, research has focused on two opposing directions: increasing data rate by spatial multiplexing [3], and improving link reliability by exploiting diversity [4].

In this paper, multistratum space-time codes [5] are compared to a channel aware adaptive transmit strategy in order to quantify the gain of channel knowledge with respect to the uncoded bit error rate for realistic systems. To this end, we present a unified view of some existing bit and power loading algorithms and derive a method that fits our needs. In order to allow for a fair comparison, maximum-likelihood detection is performed in both cases, though this comes at the price of higher receiver complexity for the non-orthogonal space-time codes.

After the definition of the channel model in Section 2, we will briefly review some motivating results from information theory in Section 3. The considered transmission schemes are described in Section 4 and 5, respectively, and simulation results are presented in Section 6. Concluding remarks can be found in Section 7.

2. CHANNEL MODEL

We consider a single-user multiple antenna system with N_T transmit and N_R receive antennas. For simplicity, we restrict to flat fading channels, and the $N_R \times N_T$ equivalent baseband channel matrix $\mathbf{H}$ is assumed to be constant during the transmission of a packet (block fading) and perfectly known to the receiver. Note that our analysis may easily be generalized to the frequency selective case if either OFDM is employed or $\mathbf{H}$ is replaced by a corresponding block Toeplitz convolution matrix for single-carrier systems. The receive signals at discrete time k are given by the $N_R \times 1$ vector

$$\mathbf{y}[k] = \mathbf{H}\,\mathbf{x}[k] + \mathbf{n}[k]\ , \tag{1}$$

where the $N_T \times 1$ vector $\mathbf{x}[k]$ contains all transmit symbols and $\mathbf{n}[k]$ represents complex white Gaussian noise with variance $\sigma_n^2 = 1$. Throughout this paper the following normalizations will be used:

$$\mathrm{E}\left\{\|\mathbf{H}\|_F^2\right\} = N_T\,N_R\ , \quad \mathrm{E}\left\{\|\mathbf{x}[k]\|^2\right\} = P\ . \tag{2}$$

For uncorrelated transmit signals with equal power, the received energy per information bit E_b is proportional to the number of receive antennas. Hence, we define

$$\frac{E_b}{N_0} = \frac{N_R\,P}{R} \tag{3}$$

where N_0 is the one-sided power spectral density of the noise and R denotes the data rate in bits per channel use.

3. RESULTS FROM INFORMATION THEORY

Assume that the transmitter also has perfect channel state information. Then, introducing the singular value decomposition $\mathbf{H} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^H$ and the relation $\mathbf{x}[k] = \mathbf{V}\,\mathbf{s}[k]$, the original channel model (1) can be transformed into an equivalent one consisting of decoupled subchannels

$$\mathbf{r}[k] = \mathbf{U}^H\,\mathbf{y}[k] = \boldsymbol{\Sigma}\,\mathbf{s}[k] + \mathbf{U}^H\,\mathbf{n}[k]\ . \tag{4}$$

Due to the circular symmetry of $\mathbf{n}[k]$, multiplication with the unitary matrix $\mathbf{U}^H$ does not change the statistical properties of the noise. Given a channel realization $\mathbf{H}$, the mutual information between transmit signals $\mathbf{x}[k]$ and receive signals $\mathbf{y}[k]$ under the power constraint (2) is maximized if $\mathbf{s}[k]$ contains independent elements taken from a Gaussian codebook. Denoting the squared singular values of $\mathbf{H}$ by λ_i, this yields

$$I(\mathbf{x}[k]\,;\,\mathbf{y}[k]\,|\,\mathbf{H}) = \sum_{i=1}^{N_T} \log_2(1+\lambda_i\,P_i)\ , \tag{5}$$

where, in order to achieve the capacity $C(\mathbf{H})$, the transmit powers P_i on the parallel subchannels must be chosen according to the waterfilling criterion [6]

$$P_i = \max\left\{\theta - \frac{1}{\lambda_i}\,,\,0\right\} \qquad \text{with} \quad \sum_{i=1}^{N_T} P_i = P\ . \tag{6}$$

For random $\mathbf{H}$, the mutual information is also a random variable. In this case, the ergodic capacity

$$C = \mathrm{E}\left\{I(\mathbf{x}[k]\,;\,\mathbf{y}[k]\,|\,\mathbf{H})\right\} \tag{7}$$

and the outage probability

$$P_{\text{out}} = \Pr\left\{I(\mathbf{x}[k]\,;\,\mathbf{y}[k]\,|\,\mathbf{H}) < R\right\} \tag{8}$$

are appropriate performance measures. As waterfilling maximizes the mutual information for each channel realization, it is also the optimum strategy with respect to (7) and (8), respectively.

Without channel knowledge at the transmitter, adaptation to the current channel conditions is not possible. In this situation, the ergodic capacity can be maximized by transmitting uncorrelated signals with equal power, i.e. using $P_i = P/N_T$ in (5). For the outage probability this power allocation is also optimal, at least in the interesting SNR range where $P_{\text{out}} \ll 1$.

Fig. 1 illustrates the achievable performance for uncorrelated Rayleigh fading channels. For $N_T = 4$ and $N_R = 1$, transmitter-sided channel state information provides a large capacity gain of 6 dB, because the total transmit power can be focused on the single subchannel with $\lambda_i > 0$ in this

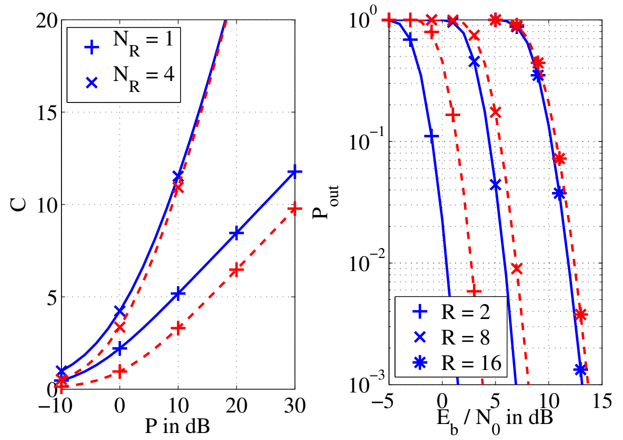

Fig. 1. Ergodic channel capacity for $N_T = 4$ (left) and outage probability for $N_T = N_R = 4$ (right) with waterfilling (solid) and equal power allocation (dashed).

case. However, for $N_T = N_R = 4$ the gap is significantly smaller and vanishes in the high SNR limit where the waterfilling solution also tends to distribute powers equally among the subchannels. The outage probabilities show the same trend: for low data rates the power can be concentrated on strong eigenmodes, while for high rates also weaker subchannels must be used. The question arises, how these information theoretic results that are valid only for ideal channel codes translate to more realistic system setups.

4. BIT AND POWER LOADING

As shown in (4), the channel may be decomposed into parallel subchannels by using $\mathbf{V}$ as a linear precoder. For Gray-labelled square QAM symbols $s_i[k]$ with $R_i = 2, 4, \ldots$ bits per symbol, the bit error probability on the i-th subchannel can be well approximated[1] by

$$P_{b,i} \approx \frac{4}{R_i}\left(1-\frac{1}{2^{R_i/2}}\right)\cdot \mathrm{Q}\left(\sqrt{\frac{3}{2^{R_i}-1}\,\lambda_i\,P_i}\right) \qquad \text{with} \quad \mathrm{Q}(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-t^2/2}dt\ . \tag{9}$$

Our goal is to minimize the average error probability for given data rate R and transmit power P by properly adapting R_i and P_i for $1 \le i \le N_T$, i.e.

$$\min_{\mathbf{R},\mathbf{P}} P_b = \frac{1}{R}\sum_{i=1}^{N_T} R_i\,P_{b,i} \qquad \text{s.t.} \quad \sum_{i=1}^{N_T} R_i = R\ , \quad \sum_{i=1}^{N_T} P_i = P\ . \tag{10}$$

[1] In (9) it has been assumed that one bit error occurs per symbol error, which is usually true for moderate to high SNR's.

For a fixed rate allocation, (10) can be solved by formulating the Lagrange dual function and applying the Karush-Kuhn-Tucker optimality conditions as shown in [7]. Then, the whole solution corresponds to the minimum over all possible rate distributions, where, without loss of generality, only those fulfilling $R_i \geq R_l$ for $\lambda_i \geq \lambda_l$ need to be checked. While this approach is feasible for a small number of subchannels, the computational complexity will be very high if, e.g., there are additionally multiple subcarriers like in MIMO-OFDM. Thus, we will subsequently introduce some simplifications.

For sufficiently large P, the minimization of P_b requires equal bit error rates (BER) on all subchannels, because otherwise weak channels will dominate the performance. This is most easily verified for a system with two subchannels if the Gaussian error function $\mathrm{Q}(x)$ is approximated by an exponential function

$$P_b \approx \beta_1 \, e^{-\alpha_1 \, P_1} + \beta_2 \, e^{-\alpha_2 \, (P - P_1)} \; , \tag{11}$$

where β_i depends only on the rate R_i, and α_i additionally on the channel gain λ_i. Setting $\partial P_b / \partial P_1 = 0$ leads to

$$P_1^{\mathrm{opt}} = \frac{\alpha_2 \, P + \ln\left(\frac{\alpha_1 \beta_1}{\alpha_2 \beta_2}\right)}{\alpha_1 + \alpha_2} \overset{P \gg 1}{\approx} \frac{\alpha_2 \, P}{\alpha_1 + \alpha_2} \; , \tag{12}$$

and hence $\alpha_1 P_1^{\mathrm{opt}} \approx \alpha_2 P_2^{\mathrm{opt}}$. Neglecting the factors β_1 and β_2 in (11) this implies $P_{b,1}^{\mathrm{opt}} \approx P_{b,2}^{\mathrm{opt}}$. By induction, this result can be generalized to any number of subchannels. Note that adaptive modulation limits the dynamics range of α_1/α_2, so the equal BER assumption is well justified in this case.

Replacing the factor in front of the Q function by an appropriate constant c, (9) can be solved for the rate

$$\begin{aligned} R_i &= \log_2\left(1 + \lambda_i \, \frac{P_i}{\Gamma}\right) \\ &\text{with} \quad \Gamma = \frac{1}{3} \, \left[\mathrm{Q}^{-1}\left(P_b/c\right)\right]^2 \; . \end{aligned} \tag{13}$$

Except for the scaling factor Γ which is a decreasing function of P_b, this is identical to the capacity of an AWGN channel with gain λ_i, hence (13) is usually referred to as gap approximation. For continuous rates, the minimization of P_b is equivalent to minimizing P and again solved by the waterfilling criterion (6), but now with a constraint on the total data rate. Afterwards, the transmit power resulting from the waterfilling procedure can be scaled by Γ to any desired value P.

Though derived in an alternative way, the widely used bit loading algorithm proposed in [2] is essentially based on a high SNR approximation of (13), namely

$$R_i \approx \log_2\left(\lambda_i \, \frac{P}{N_T \, \Gamma}\right) \; . \tag{14}$$

The continuous rates are quantized and, if necessary, bits are added or removed until the total rate constraint is fulfilled. However, this two-step procedure is suboptimal. Instead, we suggest to use an optimum algorithm for the discrete rate power minimization problem and scale the transmit powers as described above. The most prominent example is the Hughes-Hartogs algorithm [1], but there also exist far more efficient implementations that exploit the convexity of the rate-power region, e.g. [8]. Note that instead of simple scaling it is, of course, also possible to redistribute the transmit powers according to [7] once the bit allocation has been found. However, as argued before, the possible performance gain will be small compared to the additional computational effort.

5. MULTISTRATUM SPACE-TIME CODES

The adaptive transmission presented in the previous section requires perfect channel state information at the receiver and the transmitter. While for the receiver this assumption is justified for slowly fading channels, it is quite unrealistic for the transmitter as it implies either a high rate feedback link or a time division duplex (TDD) system with ideal reciprocity of the channel. Hence, we will now turn to the case without channel knowledge at the transmitter.

In the V-BLAST architecture [3], the information symbols $s_i[k]$ are directly transmitted, i.e. $\mathbf{x}[k] = \mathbf{s}[k]$ in (1). For $N_R \geq N_T$, it is possible to separate the signals at the receiver. To this end, many algorithms with different complexities have been developed. In combination with powerful channel coding and iterative a-posteriori probability (APP) detection and decoding, the capacity limits of the channel can be approached [9]. However, for uncoded transmission V-BLAST does not exploit transmit diversity.

On the other hand, orthogonal space-time block codes can significantly enhance the link reliability while having low decoding complexity. A simple example is the Alamouti scheme for $N_T = 2$ antennas [4]

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}[1] & \mathbf{x}[2] \end{pmatrix} = \begin{pmatrix} s_1 & -s_2^* \\ s_2 & s_1^* \end{pmatrix} \; . \tag{15}$$

Here, $K_{ST} = 2$ complex symbols are transmitted during $N_{ST} = 2$ channel uses. Thus, as opposed to V-BLAST, the Alamouti code (15) does not offer a multiplexing gain. Furthermore, for more than two transmit antennas there exists no linear orthogonal space-time code with rate one. Instead, we consider the following quasi-orthogonal code [10]

$$\mathbf{X} = \begin{pmatrix} s_1 & -s_2^* & -\tilde{s}_3^* & \tilde{s}_4 \\ s_2 & s_1^* & -\tilde{s}_4^* & -\tilde{s}_3 \\ \tilde{s}_3 & -\tilde{s}_4^* & s_1^* & -s_2 \\ \tilde{s}_4 & \tilde{s}_3^* & s_2^* & s_1 \end{pmatrix} \; , \tag{16}$$

where $\tilde{s}_i = e^{j\pi/4} s_i$ are rotated versions of the actual QAM information symbols. This constellation rotation is crucial for achieving the full diversity degree. Note that (16) follows from a recursive application of (15) and can easily be generalized to any number of transmit antennas.

Every linear space-time block code can be described by a generator matrix $\mathbf{G}$. In order to allow for complex conjugate information symbols, we define the real-valued vector

$$\mathbf{s} = \left(\mathrm{Re}\left\{s_1\right\}, \mathrm{Im}\left\{s_1\right\}, \ldots, \mathrm{Im}\left\{s_{K_{ST}}\right\}\right)^T . \qquad (17)$$

Then, a space-time codeword is given by the sequence of transmit vectors

$$\mathbf{x} = \left(\mathbf{x}^T[1], \ldots, \mathbf{x}^T[N_{ST}]\right)^T = \mathbf{G}\,\mathbf{s} . \qquad (18)$$

The channel matrix $\mathbf{H}$ and the generator matrix $\mathbf{G}$ can be merged to a system matrix $\mathbf{A} = (\mathbf{I}_{N_{ST}} \otimes \mathbf{H})\,\mathbf{G}$, which results in an overall system model having the same structure as (1). Thus, detection algorithms developed for V-BLAST may be used for general non-orthogonal codes. Comparing (18) and (15), it is easy to verify that the generator matrix of the Alamouti scheme is

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}[1] \\ \mathbf{G}[2] \end{pmatrix} = \left(\begin{array}{cccc} 1 & j & 0 & 0 \\ 0 & 0 & 1 & j \\ \hline 0 & 0 & -1 & j \\ 1 & -j & 0 & 0 \end{array}\right) , \qquad (19)$$

where the implicitly defined submatrices $\mathbf{G}[k]$ characterize the k-th transmit vector of one codeword. The 16×8 generator matrix of the quasi-orthogonal code (16) can be determined analogously and will be omitted here due to space limitations.

Multistratum space-time codes combine the benefits of both space-time block codes and multilayer transmission [5]. As in V-BLAST, the information symbols are demultiplexed into $N_S \leq \min\{N_T, N_R\}$ parallel data streams. However, instead of directly transmitting these so called strata, they are first encoded by a common space-time block code. Hence, each stratum experiences full transmit diversity. Afterwards, the strata are superimposed using an appropriate orthogonal transform that allows to separate them at the receiver. The Hadamard transform is one possible choice. For the Alamouti code with the maximum number of $N_S = 2$ strata, this results in the 4×8 matrix

$$\begin{aligned} \mathbf{G}^{(2)} &= \begin{pmatrix} \mathbf{G}[1] & \mathbf{G}[1] \\ \mathbf{G}[2] & -\mathbf{G}[2] \end{pmatrix} \\ &= \left(\begin{array}{cccc|cccc} 1 & j & 0 & 0 & 1 & j & 0 & 0 \\ 0 & 0 & 1 & j & 0 & 0 & 1 & j \\ \hline 0 & 0 & -1 & j & 0 & 0 & 1 & -j \\ 1 & -j & 0 & 0 & -1 & j & 0 & 0 \end{array}\right) . \end{aligned} \qquad (20)$$

Similarly, if $\mathbf{G}$ represents the quasi-orthogonal code from (16), a multistratum space-time code with rate four is given by the 16×32 generator matrix [11]

$$\mathbf{G}^{(4)} = \begin{pmatrix} \mathbf{G}[1] & \mathbf{G}[1] & \mathbf{G}[1] & \mathbf{G}[1] \\ \mathbf{G}[2] & -\mathbf{G}[2] & \mathbf{G}[2] & -\mathbf{G}[2] \\ \mathbf{G}[3] & \mathbf{G}[3] & -\mathbf{G}[3] & -\mathbf{G}[3] \\ \mathbf{G}[4] & -\mathbf{G}[4] & -\mathbf{G}[4] & \mathbf{G}[4] \end{pmatrix} . \qquad (21)$$

In contrast to V-BLAST, multistratum codes are also suited for systems with $N_R < N_T$ as strata may be switched off without transmission breaks. For $N_S = 1$ they degenerate to ordinary space-time block codes. This enables a flexible adaptation of the data rate and application in heterogenous networks.

6. SIMULATION RESULTS

In this section we compare multistratum space-time codes to the adaptive transmission scheme from Section 4. For the simulations, an uncorrelated Rayleigh fading channel was assumed, and the improved sphere decoder from [11] was used for maximum-likelihood detection of the multistratum codes. Taking practical implementation limitations into account, the maximum constellation size for bit loading was fixed to 256-QAM.

Fig. 2 shows the results for both codes discussed in this paper together with those using adaptive modulation and power allocation. As a reference, we also included the achievable performance for ”virtual” symbol alphabets without discrete and finite rate constraints.

For $R = 2$, the gain of pure beamforming over space-time coding is evident. This coincides with the theoretical analysis in Section 3. For higher rates, the maximum number of strata was used with varying modulation index. It can be observed that the code from (20) as proposed in [5] does not achieve full diversity, which may be prevented by proper a constellation rotation. Nevertheless, for $R = 12$ it outperforms the adaptive transmission mode at high SNR, because due to the rate restrictions, 16-QAM must be employed on the weak second subchannel. However, in general the diversity scheme is far from optimum in this case.

Things look different for the multistratum design (21) based on the quasi-orthogonal space-time block code from (16). The code exhibits the optimal diversity-multiplexing tradeoff and comes very close to the theoretical performance limit represented by continuous waterfilling. Hence, in line with the observations made from Fig. 1, channel state information at the transmitter yields no additional gain for uncorrelated fading and $N_R \geq N_T$ if high data rates are desired. Moreover, the practical bit loading scheme again suffers from discrete ($R = 8$) and finite ($R = 24$) rate constraints. As the results for the extreme cases of *ideal* and *no* channel coding show the same trends, it can be conjectured that the same conclusions hold true for realistic channel coding schemes.

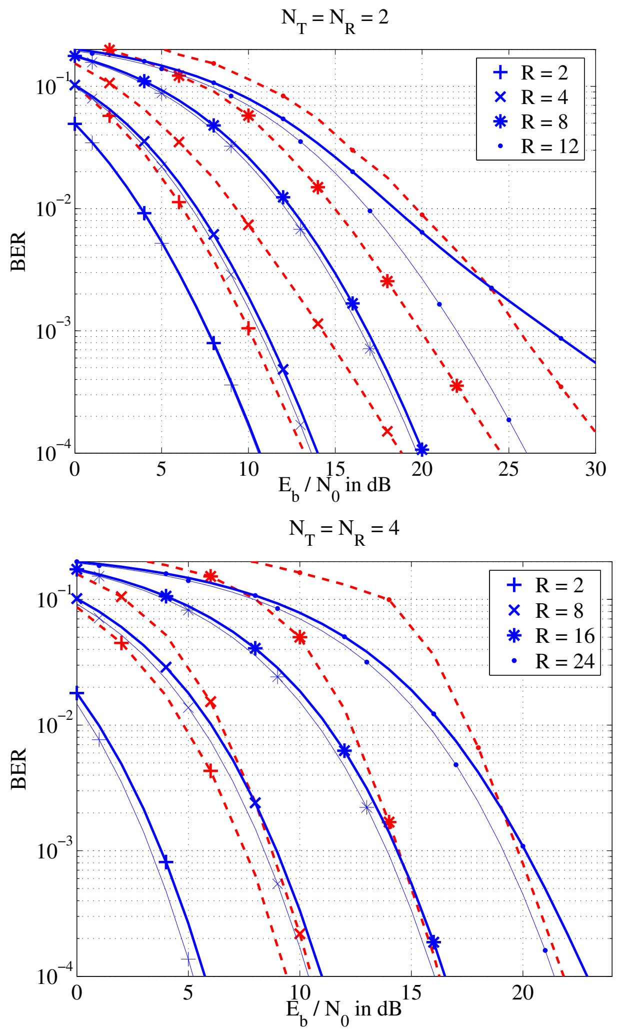

Fig. 2. Bit error rates for multistratum space-time codes (dashed) and adaptive transmission with $R_i \in \{2, 4, 6, 8\}$ (solid). The thin lines correspond to waterfilling without discrete rate constraints.

7. CONCLUSION

Inspired by results from information theory, we analyzed the uncoded bit error performance of adaptive bit and power loading with perfect channel state information at the transmitter, and compared it to that of multistratum space-time codes which simply rely on transmit diversity. While the code for $N_T = 2$ antennas taken from [5] is not able to exploit the full diversity of the channel, a recently proposed full-rate full-diversity multistratum design for $N_T = 4$ is close to optimal if maximum-likelihood detection is performed at the receiver. However, for low data rates channel knowledge may significantly improve the transmission quality.

8. REFERENCES

[1] D. Hughes-Hartogs, "Ensemble modem structure for imperfect transmission media," U.S. Patents Nos. 4,679,227 (July 1987) 4,731,816 (March 1988) 4,833,796 (May 1989).

[2] R. Fischer and J. Huber, "A new loading algorithm for discrete multitone transmission," in *Proc. IEEE Globecom*, London, November 1996, pp. 724–728.

[3] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An Architecture for Realizing Very High Data Rates Over the Rich-Scattering Wireless Channel," in *Proc. ISSE*, Pisa, Italy, September 1998.

[4] S. M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE Journal on Selected Areas in Communuications*, vol. 16, no. 8, pp. 1451–1458, October 1998.

[5] U. Wachsmann, J. Thielecke, and H. Schotten, "Exploiting the Data-Rate Potential of MIMO Channels: Multi-Stratum Space-Time Coding," in *Proc. IEEE VTC Spring 2001*, Rhodes, Greece, May 6-9 2001.

[6] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, November-December 2000.

[7] C. Mutti and D. Dahlhaus, "Adaptive power loading for multiple-input multiple-output ofdm systems with perfect channel state information," in *Joint COST Workshop on Antennas and Related System Aspects in Wireless Communications*, Goeteborg, Juli 2004.

[8] B. Krongold, K. Ramchandran, and D. Jones, "Computationally efficient optimal power allocation algorithms for multicarrier communication systems," *ITC*, vol. 48, no. 1, pp. 23–27, January 2000.

[9] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *ITC*, vol. 51, no. 3, pp. 23–27, March 2003.

[10] N. Sharma and C. Papadias, "Improved Quasiorthogonal Code through Constellation Rotation," *IEEE Transactions on Communications*, vol. 51, no. 3, pp. 332–335, March 2003.

[11] R. Böhnke, V. Kühn, and K. Kammeyer, "Efficient near maximum-likelihood decoding of multistratum space-time codes," in *Proc. IEEE Vehicular Technology Conference (VTC)*, Los Angeles, USA, September 2004.