MATRIX MODULATION AND ADAPTIVE RETRANSMISSION

Ari Hottinen and Olav Tirkkonen,

Nokia Research Center, P.O.Box 407 00045 Nokia Group, Finland

ABSTRACT

We consider different high rate space-time modulation methods and related (re)transmission strategies in a MIMO channel. The proposed matrix modulators exploit the MIMO channel in increasing the peak rate via power-efficient parallel transmission, whereas the self-orthogonalizing retransmission and signal combining strategies improve ultimate throughput.

1. INTRODUCTION

Modern high performance communication systems utilize various diversity resources, rate adaption concepts as Adaprive Modulation and Coding (AMC) and Automatic Repeat reQuest protocols (ARQ) in an attempt to maximize system throughput with minimal effort (transmit power) [1]. The most viable way to further increase throughput, at the cost of increased receiver complexity, is to exploit inherent “spatial signatures” in multi-antenna (or multiple-input multiple-output (MIMO)) channels [2, 3]. This may be done with well-designed non-orthogonal matrix modulation methods which allow to access both MIMO (rate) gains and diversity gains.

In time varying channel conditions, on-line rate adaptation allows to match channel peak data rates, and ARQ concepts provide a last resort, when decoding fails. ARQ and AMC can be seen as methods of providing low-rate side information to the transmitter. Both ARQ, AMC and transmit diversity solutions have been widely discussed in the literature, and found their way to 3G wireless systems. Typically they are considered as independent modules; AMC and (slow) ARQ concentrate on exploiting temporal diversity, in contrast to the spatial nature of transmit diversity. However, using non-orthogonal matrix modulation alphabets in a MIMO channel pave a way for new retransmission and rate adaptation concepts, as discussed in [4].

In this paper we consider high rate non-orthogonal matrix modulation, and space-time retransmission strategies, where the ARQ protocol and matrix modulation are jointly designed.

2. MATRIX MODULATION

2.1. Signal model

The signal model considered in this paper is given by

$$\underset{T\times N_\mathrm{r}}{\mathbf{Y}} = \underset{T\times N_\mathrm{t}}{\mathbf{X}}\ \underset{N_\mathrm{t}\times N_\mathrm{r}}{\mathbf{H}} + \underset{T\times N_\mathrm{r}}{\mathrm{noise}} \qquad (1)$$

Above, T designates the number of channel uses the matrix modulation extends over, N_t the number of transmit antennas, N_r the number of receive antennas, $\mathbf{Y}$ is the received signal matrix and $\mathbf{X}$ is the modulation matrix. The columns of the channel matrix $\mathbf{H}$ designate channel vectors from N_t transmit antennas to different receive antennas, with h_{mn} the channel from transmit antenna m to receive antenna n. In vector modulation $\mathbf{X}$ is an N_t-vector ($N_\mathrm{t} > 1$), while in matrix modulation $N_\mathrm{t} > 1, T > 1$.

2.2. Orthogonal modulation matrices

Symbol rate $R_\mathrm{s} = 1$ space–time codes with real symbol alphabets can be found for any number of Tx antennas. In dimensions 2, 4 and 8 these can be constructed from orthogonal designs [5]. For more than 8 Tx antennas, generalized real orthogonal designs were constructed in [5] that have $R_\mathrm{s} = 1$ for any number of Tx antennas. For these, the delay grows exponentially with N_t. The series of $R_\mathrm{s} = 1$ real modulation schemes can be converted in to $R_\mathrm{s} = 1/2$ complex modulation schemes, by transmitting the complex symbols and their complex conjugates separately using a $R_\mathrm{s} = 1$ real orthogonal design [5]. Complex modulation space–time codes with $R_\mathrm{s} = 1$ exist only for two transmit antennas(the Alamouti code). Similar square matrix designs in higher dimensions incur an exponential rate loss, see [4]. For this reason, orthogonal designs are not attractive when high throughput (or bandwidth efficiency) is the driving force in the design process.

2.3. Non-orthogonal matrix modulation

If the orthogonality requirement is relaxed a larger class of modulation matrices is allowed. They can be designed to capture both MIMO rate gains and transmit diversity gains. Often, especially with a few receive antennas, complex signal processing, analogous to multiuser detection or channel equalization is needed to recover the symbols or bits in the receiver. Non-orthogonal matrix modulation methods were first proposed in [6–8]. These first suggestions suffer from pathological error events leading to loss of diversity and suboptimal performance at high SNR. Applying symbol rotations to cure this was discussed in [9]. For an in-depth exposition on matrix modulation methods, see [4]. Here, we concentrate on particular examples for four transmit antennas.

2.3.1. *Symbol rate one: ABBA*

Take $\mathbf{X}_A$ and $\mathbf{X}_B$ to be Alamouti blocks [10].[1] The family of 4×4 matrices of ABBA type [6, 7], has correspondingly two

[1] For higher N_t, higher dimensional space-time block codes may be used.

quasi-orthogonal layers. In 2×2 block form, two representatives are

$$\mathbf{X}_{ABBA\pm} = \begin{bmatrix} \mathbf{X}_A & \pm\widetilde{\mathbf{X}}_B \\ \pm\widetilde{\mathbf{X}}_B & \mathbf{X}_A \end{bmatrix} \quad (2)$$

The ABBA's are optimal 4×4 matrix modulators from the self-interference [6], or mutual information [11] point of view, with optimal low SNR performance. With a scalar [9] or matrix [11] symbol rotation in the B-block, ABBA can be made full diversity, and high SNR performance is optimized as well. A tilde indicates a block with rotated symbols.

The family of modulators with ABBA structure is multidimensional. Here, for retransmission purposes, we concentrate on four special forms. In addition to (2) we consider

$$\mathbf{X}_{\mathrm{diag}\ ABBA\pm} = \begin{bmatrix} \mathbf{X}_A \pm \widetilde{\mathbf{X}}_B & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{X}_A \mp \widetilde{\mathbf{X}}_B \end{bmatrix} \quad (3)$$

Here $\mathbf{0}_2$ is the 2×2 null matrix. Note that (3) can be constructed by acting with a precoder consisting of a 2×2 Hadamard matrix and a symbol rotator on a supersymbol vector $[\mathbf{X}_A, \mathbf{X}_B]$ of two quaternionic (Alamouti) symbols. The resulting two quaternionic symbols are time-division multiplexed to the diagonal of a 2×2 matrix with quaternionic entries (any other multiplexing applies equally well). This is in contrast to [12, 13], where for a transmission from $N_\mathrm{t} = 4$ antennas, a 4×4 Hadamard transform and a symbol rotator is applied to a supersymbol vector of 4 complex symbols, and the result is multiplexed to the diagonal of a 4×4 matrix with complex entries.

The effect of different channel conditions on a non-orthogonal matrix modulation are most tranparently seen when writing the signal model in terms of the effective channel matrix as

$$\mathbf{y} = \mathcal{H}\mathbf{U}\mathbf{x} + \mathbf{n} \ ,$$

where $\mathbf{y}$ is obtained from $\mathbf{Y}$ by stacking and using complex conjugations. The vector of transmitted complex symbols is $\mathbf{x}$, and U is a complex precoding matrix realizing the symbol rotations. With two transmit and one receive antenna, the effective channel for the Alamouti code is $\mathcal{H}_{12} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}$. Similarly, for $N_\mathrm{r} = 1$ the effective channels for ABBA$\pm$ and diag-ABBA$\pm$ are, respectively

$$\begin{bmatrix} \mathcal{H}_{12} & \pm\mathcal{H}_{34} \\ \mathcal{H}_{34} & \pm\mathcal{H}_{12} \end{bmatrix}, \quad \begin{bmatrix} \mathcal{H}_{12} & \pm\mathcal{H}_{12} \\ \mathcal{H}_{34} & \pm\mathcal{H}_{34} \end{bmatrix} .$$

After matched filtering a model

$$\mathbf{z} = \mathbf{U}^\dagger \mathcal{H}^\dagger \widetilde{\mathbf{y}}$$

arises, giving the matched filter correlation matrix $\mathcal{H}^\dagger\mathcal{H} = \mathcal{D} + \mathcal{S}$. The entries of the diagonal part $\mathcal{D}$ measure the power of the channel an individual symbol is transmitted over. For any maximally symbolwise diverse scheme, these entries are all the same, and proportional to the total channel power $\sum_{mn} |h_{mn}|^2$. The off-diagonal part $\mathcal{S}$ measures the self-interference induced by the non-orthogonality of the modulation. For the four versions of ABBA, it is of the form

$$\mathcal{S} = s \begin{bmatrix} \mathbf{0}_2 & \mathbf{I}_2 \\ \mathbf{I}_2 & \mathbf{0}_2 \end{bmatrix} ; \quad (4)$$

$$s_{ABBA\pm} = \pm \sum_{n=1}^{N_\mathrm{r}} 2\,\mathrm{Re}\,[h_{1n}^* h_{3n} + h_{2n}^* h_{4n}] \quad (5)$$

$$s_{\mathrm{diag}\ ABBA\pm} = \pm \sum_{n=1}^{N_\mathrm{r}} |h_{1n}|^2 + |h_{2n}|^2 - |h_{3n}|^2 - |h_{4n}|^2 \quad (6)$$

For use in retransmission, it is important to note that by changing the transmission, the sign of the self-interference changes. More generically, if the relative sign of two quasi-orthogonal layers is changed, the corresponding self-interference changes sign.

2.3.2. *Symbol Rate Two: Double ABBA*

A number of ways to overlay two symbol rate one modulation matrices leads to a modulation method with four quasi-orthogonal layers, and symbol rate two [14]. One example is

$$\mathbf{X}_{DABBA} = \begin{bmatrix} \mathbf{X}_A & \widetilde{\mathbf{X}}_B \\ \widetilde{\mathbf{X}}_B & \mathbf{X}_A \end{bmatrix} + \begin{bmatrix} \widetilde{\mathbf{X}}_C & \widetilde{\mathbf{X}}_D \\ -\widetilde{\mathbf{X}}_D & -\widetilde{\mathbf{X}}_C \end{bmatrix} \quad (7)$$

Above $\mathbf{X}_C$ and $\mathbf{X}_D$ are again Alamouti blocks with independent symbols. Hence, eight symbols are transmitted in four time epochs. Choosing the symbol rotations to reach the full transmit diversity order of four is discussed in [11]. This requires using matrix valued constellation rotations within layers. Note that $\mathbf{X}_A$ and $\mathbf{X}_C$ have a diagonal ABBA structure, as have $\mathbf{X}_B$ and $\mathbf{X}_D$. The remaining cross-correlations are of worse-than-ABBA nature. Among rate two 4×4 modulation matrices, DABBA has minimal self-interference and maximal mutual information [4, 11]. Applying symbol rotations mixing symbols in two layers, DABBA can be related to Double STTD (DSTTD) [15] which transmits two Alamouti (STTD) blocks in parallel from four transmit antennas:

$$\mathbf{X}_{DSTTD} = \begin{bmatrix} \mathbf{X}_A & \mathbf{X}_B \end{bmatrix} \quad (8)$$

DSTTD only attains transmit diversity order two.

2.3.3. *Symbol Rate Four: Quad ABBA*

With $N_\mathrm{r} \geq 4$ receive antennas it becomes worthwhile to use a symbol rate four transmission. One possibility is Quad ABBA,

$$\begin{aligned} \mathbf{X}_{QABBA} &= \mathbf{X}_{DABBA}\Big(\mathbf{X}_A, \widetilde{\mathbf{X}}_B, \widetilde{\mathbf{X}}_C, \widetilde{\mathbf{X}}_D\Big) \\ &\quad + \mathrm{j}\ \mathbf{X}_{DABBA}\Big(\widetilde{\mathbf{X}}_E, \widetilde{\mathbf{X}}_F, \widetilde{\mathbf{X}}_G, \widetilde{\mathbf{X}}_H\Big) \ , \end{aligned} \quad (9)$$

written in terms of eight quasi-orthogonal layers based on Alamouti blocks. Quad ABBA achieves capacity and is symbol homogeneous in the sense that all symbols experience similar gain and self-interference. It uses the full Clifford basis for complex 4×4 matrices for transmission [4, 11].

3. RETRANSMISSION CONCEPTS

In retransmission concepts an erroneously decoded frame generates a NACK (Not-Acknowledged) feedback and a retransmission follows. With at most $\mathrm{L}_{\max}$ transmissions ($\mathrm{L}_{\max} - 1$

retransmissions), the throughput is

$$\text{Throughput} = R \sum_{l=1}^{\mathrm{L_{max}}} \pi_l / l \quad [\text{bps/Hz}] \tag{10}$$

where R is the peak data rate of the given encoding and modulation concept, π_l the probability that the lth transmission is successful and that all previous failed.

To obtain high throughput we would like to have a dominant π_1. Probabilities $\pi_l, l > 1$ can be affected by the chosen retransmission and signal combining concepts (incremental redundancy, chase combining, and the like). Here, we consider a related concept, where self-interference is mitigated by retransmissions [4].

In slowly varying channels, fast ARQ of the kind standardized by 3GPP [1] does not give temporal diversity benefits, it only acts as an adaptive coding concept. Here we propose space-time adaptive retransmission, where in addition to adaptive coding, the retransmission improves *spatial* diversity.

We consider a quasi-static fading channel that remains time-invariant during the time interval required by possible retransmissions. As a design criterion, we select the space-time modulation matrices so that when they are concatenated over retransmissions self-interference is mitigated. Ultimately, after a given number of retransmissions, all self-interference vanishes, and performance is defined by channel power (Frobenius norm of channel matrix), as opposed to channel rank.

Consider a symbol rate two initial transmission according to Double ABBA, as described in the previous section. We modify the signs in the modulation matrix for each retransmission. For initial transmission at t_1 we transmit (7) with an ABBA+ and CDDC+ block. If decoding fails, a retransmission (after NACK) at epoch t_2 uses (7) with the overall sign of the CDDC+ block changed. The self-interference between ABBA+ and CDDC+ thus changes sign between the two transmissions. Upon combining the matched filter outputs of the two transmissions, self-interference between ABBA+ and CDDC+ vanishes. The only self-interference remaining is within ABBA+ and CDDC+. If decoding still fails, the next two transmissions operate analogously, so that the third transmission applies DABBA with ABBA- and CDDC-, and the possible fourth transmission again changes the overall sign CDDC-. After three retransmission this cancels interference between all blocks, and a generalized complex orthogonal design with rate 1/2 and delay 16 arises. Similar process can be constructed for any non-orthogonal, or reduced diversity concept for any rate and number of antennas (such as STTD-OTD), as discussed in [4].

4. PERFORMANCE

First, in Figures 1 and 2, we compare different symbol rate two and four MIMO transmission methods. In particular, Double ABBA and Quad ABBA are compared to the threaded space-time architecture (TAST) of [13], where TAST uses real-valued rotation matrices. Here DABBA and Quad ABBA

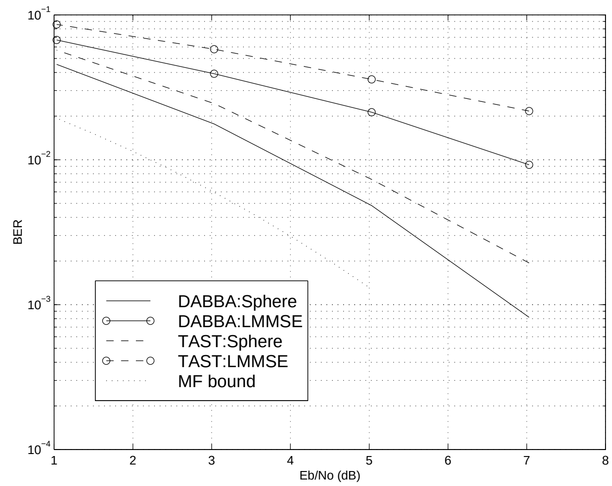

Fig. 1. Bit–error rate of DABBA and TAST. 4 transmit and 2 receive antennas in an iid Rayleigh channel, 4 bps/Hz

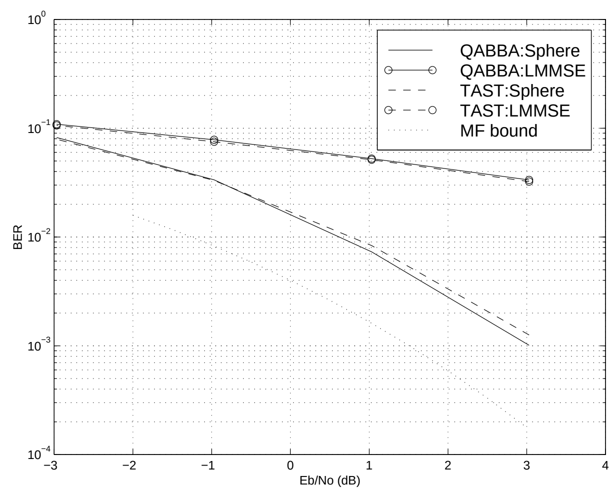

Fig. 2. Bit–error rate of DABBA and TAST. 4 transmit and 4 receive antennas in an iid Rayleigh channel, 8 bps/Hz.

apply only scalar symbol rotations, as in [14], so they do not reach full diversity. In the SNR regime considered, this does not affect the outcome. DABBA and Quad ABBA are within 1.5-2 dB from a matched filter bound, which assumes diversity order 8 or 16 with no self-interference. TAST provides essentially the same performance at symbol rate four, but is inferior at symbol rate two. This is due to loss in information of threaded concepts for rates $< N_\mathrm{t}$, and agrees perfectly with theory, see [4].

Next, consider an example of space-time retransmission for four Tx and two Rx antennas:

- Symbol rate two (DABBA and DSTTD) using two parallel QPSK streams (non-adaptive),
- DABBA and DSTTD using two parallel QPSK streams with adaptive retransmission, and

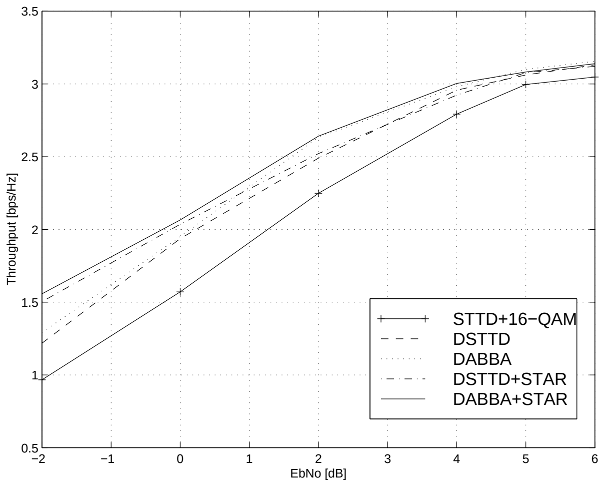

Fig. 3. Throughput for space–time retransmission with different space–time modulation schemes.

- Single stream 16-QAM using the Alamouti code.

In all cases the information bits are encoded with a rate 4/5 turbo code with frame size 378 bits, and randomly bit-interleaved. Thus, the peak rate for all transmission concepts is 3.2 bps/Hz. The maximum number of transmissions is set to four. The DABBA and DSTTD receivers use a sphere decoder.

Figure 3 reveals that the symbol rate two space–time modulation DABBA with space-time adaptive retransmission (STAR) (legend "DABBA+STAR") provides superior performance in a quasi-static Rayleigh fading channel. When adaptation is not used with DABBA (legend "DABBA") performance degrades especially for low E_b/N_0. Double STTD with adaptive retransmission performs well when E_b/N_0 is low, but performance is worse than in the corresponding DABBA case, especially when a very high throughput is desired. STTD transmission using 16-QAM and conventional ARQ with Chase combining is clearly inferior. For example, it requires about 2 dB more power than DABBA+STAR to reach 2 bps/Hz.

The throughput gains are apparent, and the gains are attained without the need to change the modulation order for retransmissions.

5. CONCLUSION

We have demonstrated the performance of efficient space–time modulation methods (DABBA and Quad ABBA) involving quasi-orthogonal layers. In addition, we showed how to couple such modulators with a retransmission scheme so that self-interference is mitigated after retransmissions.

6. REFERENCES

[1] 3GPP, "UTRA high speed downlink packet access (HSDPA)- overall description," Technical Specification TS 25.308, March 2002, available online from www.3gpp.org.

[2] G. Foschini, "Layered space–time architecture for wireless communication in a fading environment when using multi–element antennas," *Bell Labs Tech. J.*, pp. 41–59, 1996.

[3] E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, Nov/Dec 1999, based on AT& T Bell Laboratories, Internal Tech. Memo, June 1995.

[4] A. Hottinen, O. Tirkkonen, and R. Wichman, *Multiantenna Transceiver Techniques for 3G and Beyond*, Chichester: John Wiley and Sons, 2003.

[5] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space–time block codes from orthogonal designs," *IEEE Trans. Inf. Th.*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[6] O. Tirkkonen, A. Boariu, and A. Hottinen, "Minimal non-orthogonality rate one space–time block code for 3+ Tx antennas," in *Proc. IEEE ISSSTA*, Sept. 2000, vol. 2, pp. 429–432.

[7] H. Jafarkhani, "A quasi-orthogonal space–time block code," *IEEE Trans. Comm.*, vol. 49, no. 1, pp. 1–4, Jan. 2001.

[8] B. Hassibi and B. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inf. Th.*, vol. 48, no. 7, pp. 1804–1824, July 2002.

[9] O. Tirkkonen, "Optimizing space–time block codes by constellation rotations," in *Proc. Finnish Wireless Comm. Workshop'01*, Oct. 2001, pp. 59–60.

[10] S. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE J. Sel. Areas Comm.*, vol. 16, pp. 1451–1458, Oct. 1998.

[11] O. Tirkkonen and R. Kashaev, "Performance optimal and information maximal MIMO modulations," in *Proc. IEEE ISIT*, July 2002.

[12] M. Damen, K. Abed-Meraim, and J.-C. Belfiore, "Diagonal algebraic space–time block code," *IEEE Trans. Inf. Th.*, vol. 48, no. 3, pp. 628–636, Mar. 2002.

[13] H. El Gamal and M. Damen, "An algebraic number theoretic framework for space–time coding," in *Proc. IEEE ISIT*, June 2002, p. 132.

[14] A. Hottinen and O. Tirkkonen, "Space-time block code with symbol rate two," in *Proc. CISS*, March 2002.

[15] Texas Instruments, "Double-STTD scheme for HSDPA systems with four transmit antennas: link level simulation results," Tech. Rep. 21(01)-0701, Release 5 Ad hoc, 3GPP TSG RAN WG1, June 2001.