ANALYSIS OF LINEAR AND NON-LINEAR PRECODING TECHNIQUES FOR THE SPATIAL SEPARATION OF UNICAST AND MULTICAST USERS

Yuri C. B. Silva and Anja Klein

Darmstadt University of Technology, Communications Engineering Lab

Merckstr. 25, 64283 Darmstadt, Germany, {y.silva, a.klein}@nt.tu-darmstadt.de

ABSTRACT

The provision of multicast services is a relevant feature in the context of the further evolution of cellular communication systems. The scope of this paper lies on the analysis and comparison of linear and non-linear downlink precoding techniques for separating users of both unicast and multicast services in space. The paper investigates and derives a non-linear algorithm based on Tomlinson-Harashima precoding (THP) for this unicast/multicast scenario. It is seen that, differently from the linear case, the application of a non-linear multicast-aware algorithm is not capable of providing significant gains. Additionally, a hybrid linear/non-linear algorithm is proposed, which is shown to achieve a good trade-off between performance and complexity.

I. INTRODUCTION

The provision of multicast/broadcast services is an important feature in the context of the further evolution of cellular communication systems [1]. Services whose content is targeted at multiple users within the system, such as audio/video streaming, mobile TV, localized services, among others, may be implemented over point-to-multipoint connections. The use of such connections spares resources and is spectrally efficient. Nevertheless, they make it more difficult to adapt to the channel conditions of each specific user.

Adaptive antenna arrays may be employed at the base station in order to improve the quality perceived by the users within multicast groups, which has been investigated by previous work [2-4]. The implementation of space division multiple access (SDMA), so that multiple unicast and multicast users may share the same radio channel, is a further measure for increasing spectral efficiency.

Traditional unicast transmit processing techniques, such as zero-forcing (ZF), minimum mean square error (MMSE) precoding, and Tomlinson-Harashima precoding (THP), can be applied to this unicast/multicast scenario. However, since users of a multicast group expect the same information, it is not necessary to perform interference cancellation within the multicast group. For this reason it is expected that algorithms which take the multicast nature into account, i.e., multicast-aware algorithms, be capable of achieving better performance results.

Linear multicast-aware transmit processing techniques for the unicast/multicast scenario have been previously studied in [5, 6]. An SDMA algorithm based on semi-definite relaxation has been proposed in [5], which aims at providing a minimum target user quality, while allowing a certain degree of interference among the groups. In [6], an approach based on block diagonalization (BD) has been proposed for eliminating the interference among the groups and maximizing the worst-user signal-to-noise ratio (SNR).

The application of non-linear transmit processing techniques in unicast-only scenarios, such as Tomlinson-Harashima precoding (THP), has been shown to provide significant performance gains in comparison to linear processing algorithms [7-9]. The THP transmission chain introduces a feedback filter, which performs successive interference cancellation, and a modulo operator in order to keep the required transmit energy within feasible boundaries.

Motivated by the good results of linear multicast-aware techniques, and also by the fact that non-linear algorithms outperform linear strategies for the unicast scenario, this paper investigates and derives a non-linear multicast-aware algorithm based on THP. It is shown, however, that only slight gains can be obtained in comparison to traditional THP. Additionally, a hybrid approach which performs both linear and non-linear processing is investigated as well, and it is shown to achieve a good trade-off between performance and complexity.

This paper is organized as follows. In section II the signal model is presented. The proposed algorithms, which are referred to as the multicast aware THP (MA-THP) and the hybrid linear/non-linear precoding (HLNP), are described in sections III and IV, respectively. Section V presents the performance analysis, and, finally, conclusions are drawn in section VI.

II. SIGNAL MODEL

The system model corresponds to the downlink of a single cell in a cellular system containing both unicast and multicast users. The base station is equipped with an M-element antenna array, while the N mobile stations are single-antenna devices. It is assumed that $M \geq N$ and that the N users are divided into K multicast groups. Since the users of a multicast group expect the same symbol, K is also equivalent to the number of data streams. In this scenario the unicast users can be interpreted as multicast groups of unit size.

The number of users within each group is represented by vector $\mathbf{g}_{K \times 1}$, whose k^{th} element $g_k \in \{1, \ldots, N\}$ indicates the number of users within group k. Note that $\sum_{k=1}^{K} g_k = N$. In order to associate which users belong to which group, an index vector $\mathbf{b}_{N \times 1}$ is also introduced, whose n^{th} element $b_n \in \{1, \ldots, K\}$ indicates the group to which user n belongs. For example, in a scenario with two unicast users and one multicast group composed of two users, we would have: $N = 4$, $K = 3$, $\mathbf{g} = [1, 1, 2]^T$, and $\mathbf{b} = [1, 2, 3, 3]^T$, with $(\cdot)^T$ denoting the transpose operator.

The transmission chain considered by the proposed non-linear multicast-aware algorithms is the same as that of THP,
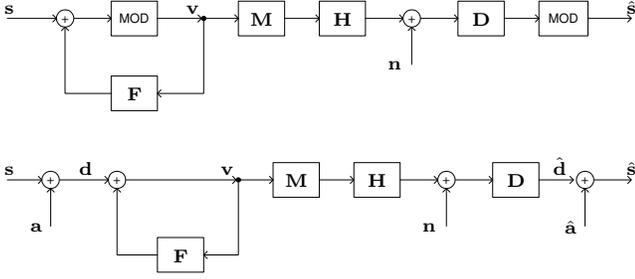
Figure 1: THP transmission chain and its linear representation.

which is depicted in Fig. 1. Its linear representation is also shown in the figure, which is obtained by expressing the modulo operator as the addition of auxiliary signals $\mathbf{a} \in \mathbb{C}^K$ and $\hat{\mathbf{a}} \in \mathbb{C}^N$ at both transmitter and receiver, respectively. Note that a multi-carrier system is considered which assumes flat-fading per sub-carrier and negligible inter-symbol interference, so that the data symbols can be treated individually.

The data symbol vector $\mathbf{s} \in \mathbb{C}^K$ is combined with the auxiliary signal $\mathbf{a}$ resulting in vector $\mathbf{d} \in \mathbb{C}^K$, which is fed into the feedback filter $\mathbf{F} \in \mathbb{C}^{K \times K}$. The output of the feedback block, vector $\mathbf{v} \in \mathbb{C}^K$, is modulated by a matrix $\mathbf{M} \in \mathbb{C}^{M \times K}$, transmitted over the radio channel $\mathbf{H} \in \mathbb{C}^{N \times M}$, subject to additive white Gaussian noise $\mathbf{n} \in \mathbb{C}^N$, and demodulated by matrix $\mathbf{D} \in \mathbb{C}^{N \times N}$. The downlink estimate $\hat{\mathbf{d}} \in \mathbb{C}^N$ of the equivalent symbol vector $\mathbf{d}$ may be written as

$$\hat{\mathbf{d}} = \mathbf{DHMv} + \mathbf{Dn}, \tag{1}$$

where $\mathbf{v} = (\mathbf{I} - \mathbf{F})^{-1}\mathbf{d}$, and $\mathbf{I}$ is an identity matrix of appropriate dimension. It should be mentioned that even though the multicast-aware algorithms employ the same transmission chain as THP, they differ essentially in how the system matrices are calculated, which is the subject of the next sessions.

III. MULTICAST-AWARE TOMLINSON-HARASHIMA PRECODING (MA-THP)

As described in the system model, the THP transmission chain comprises a linear transmission filter $\mathbf{M}$ as well as a non-linear feedback filter $\mathbf{F}$. The former is responsible for eliminating part of the interference, such that the equivalent channel matrix $\mathbf{HM}$ is lower triangular. In the case of unicast, this triangularization can be achieved by employing the Choleski factorization [7]. For the unicast/multicast scenario, however, this method does not apply, since $\mathbf{HM}$ is no longer a square matrix. Nevertheless, it is still possible to eliminate the interference by performing null space projections, such as in [6, 10, 11].

Let $\mathbf{h}_n$ denote the nth row of matrix $\mathbf{H}$ and $\mathbf{m}_k$ the kth column of matrix $\mathbf{M}$. The following constraints must be satisfied:

$$\mathbf{h}_n\mathbf{m}_k = 0 \quad \forall\, n, k \mid b_n < k\,, \tag{2}$$

where $n \in \{1, \dots, N\}$ and $k \in \{1, \dots, K\}$. This can be achieved by projecting each group beamformer $\mathbf{m}_k$ onto the null space associated to the channels of the users of previous groups.

The feedback matrix $\mathbf{F}$ is responsible for successively cancelling the remaining interference. In order to fulfill the causality constraint [8], $\mathbf{F}$ must be lower triangular with the main diagonal composed of zeros. Let matrix $\mathbf{P} \in \mathbb{C}^{K \times K}$ be defined as $\mathbf{P} = (\mathbf{I} - \mathbf{F})^{-1}$, then $\mathbf{P}$ is also lower triangular but with the main diagonal composed of ones.

The end effect that is expected from the transmission processing is that the interference among multicast groups be totally cancelled, therefore $\mathbf{P}$ must also satisfy

$$\mathbf{HMP} = \mathrm{diag_b}(\mathbf{HM})\,, \tag{3}$$

where the $\mathrm{diag_b}(\cdot)$ operator returns a matrix whose elements satisfy the following expression

$$(\mathbf{HMP})_{n,k} = \begin{cases} (\mathbf{HM})_{n,k} & , \quad \text{for} \quad b_n = k \\ 0 & , \quad \text{otherwise} \end{cases}. \tag{4}$$

One problem that arises from the fact that $\mathbf{P}$ has dimension K lower than N is that it is not always possible to find a lower triangular matrix $\mathbf{P}$ satisfying (3). In order to obtain a feasible solution it is necessary to impose additional constraints on matrix $\mathbf{M}$. These constraints can be obtained by writing (3) with $\mathbf{M}$ already satisfying (2) and $\mathbf{P}$ as lower triangular with unit diagonal. For example, considering a scenario with two unicast users and one multicast group composed of two users, (3) could be written as

$$\begin{bmatrix} \mathbf{h}_1\mathbf{m}_1 & 0 & 0 \\ \mathbf{h}_2\mathbf{m}_1 & \mathbf{h}_2\mathbf{m}_2 & 0 \\ \mathbf{h}_3\mathbf{m}_1 & \mathbf{h}_3\mathbf{m}_2 & \mathbf{h}_3\mathbf{m}_3 \\ \mathbf{h}_4\mathbf{m}_1 & \mathbf{h}_4\mathbf{m}_2 & \mathbf{h}_4\mathbf{m}_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ P_{2,1} & 1 & 0 \\ P_{3,1} & P_{3,2} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1\mathbf{m}_1 & 0 & 0 \\ 0 & \mathbf{h}_2\mathbf{m}_2 & 0 \\ 0 & 0 & \mathbf{h}_3\mathbf{m}_3 \\ 0 & 0 & \mathbf{h}_4\mathbf{m}_3 \end{bmatrix},$$

which only has a feasible solution if

$$\frac{\mathbf{h}_3\mathbf{m}_1}{\mathbf{h}_3\mathbf{m}_3} = \frac{\mathbf{h}_4\mathbf{m}_1}{\mathbf{h}_4\mathbf{m}_3} \quad \text{and} \quad \frac{\mathbf{h}_3\mathbf{m}_2}{\mathbf{h}_3\mathbf{m}_3} = \frac{\mathbf{h}_4\mathbf{m}_2}{\mathbf{h}_4\mathbf{m}_3}\,.$$

When generalizing the problem, the following set of constraints need to be taken into account for each group k:

$$\begin{aligned} \frac{\mathbf{h}_{f(k,1)}\mathbf{m}_1}{\mathbf{h}_{f(k,1)}\mathbf{m}_k} &= \frac{\mathbf{h}_{f(k,2)}\mathbf{m}_1}{\mathbf{h}_{f(k,2)}\mathbf{m}_k} = \cdots = \frac{\mathbf{h}_{f(k,g_k)}\mathbf{m}_1}{\mathbf{h}_{f(k,g_k)}\mathbf{m}_k}\,, \\ \frac{\mathbf{h}_{f(k,1)}\mathbf{m}_2}{\mathbf{h}_{f(k,1)}\mathbf{m}_k} &= \frac{\mathbf{h}_{f(k,2)}\mathbf{m}_2}{\mathbf{h}_{f(k,2)}\mathbf{m}_k} = \cdots = \frac{\mathbf{h}_{f(k,g_k)}\mathbf{m}_2}{\mathbf{h}_{f(k,g_k)}\mathbf{m}_k}\,, \\ &\quad\quad\quad \vdots \\ \frac{\mathbf{h}_{f(k,1)}\mathbf{m}_{k-1}}{\mathbf{h}_{f(k,1)}\mathbf{m}_k} &= \frac{\mathbf{h}_{f(k,2)}\mathbf{m}_{k-1}}{\mathbf{h}_{f(k,2)}\mathbf{m}_k} = \cdots = \frac{\mathbf{h}_{f(k,g_k)}\mathbf{m}_{k-1}}{\mathbf{h}_{f(k,g_k)}\mathbf{m}_k}\,, \end{aligned} \tag{5}$$

where the function $f(i, j)$ returns the index of the jth user of group i. This results in a total of $\sum_{k=1}^{K}(k-1)(g_k-1)$ additional constraints. Note that only groups with more than one user ($g_k > 1$) and which appear after the first position ($k > 1$) generate these constraints.

Due to this relationship among the beamforming vectors of the different groups, it is not possible to optimize them individually. The joint optimization problem corresponds to finding the matrix $\mathbf{M}$ which maximizes the minimum energy received by the users, i.e.,

$$\mathbf{M}_{opt} = \underset{\mathbf{M}}{\mathrm{argmax}} \ \min \ \mathrm{diag}(\mathbf{y}\mathbf{y}^H)\,,$$
$$\text{subject to (2), (5), and } \mathrm{trace}(\mathbf{M}^H\mathbf{M}\mathbf{R}_v) = E_{tr}\,, \tag{6}$$

where $\mathbf{y} \in \mathbb{C}^{N \times 1}$ is defined as $\mathbf{y} = \mathrm{diag}_\mathrm{b}(\mathbf{HM})\mathbf{1}$, in which $\mathbf{1}$ is a vector of ones of appropriate dimension, the $\mathrm{diag}(\cdot)$ operator applied to a matrix returns a vector containing the elements of the main diagonal and when applied to a vector it returns a diagonal matrix with the elements within the main diagonal, $\mathbf{R}_v \in \mathbb{C}^{K \times K}$ is the signal correlation matrix defined as $\mathbf{R}_v = \mathrm{E}\{\mathbf{v}\mathbf{v}^H\}$, $\mathrm{E}\{\cdot\}$ is the expectation operator, and E_{tr} is the maximum available transmit energy.

In order to avoid such a complex optimization procedure, a suboptimum methodology which performs the independent optimization of each beamforming vector is here proposed. The beamforming vector $\mathbf{m}_k$ of each group is assumed to lie in the null space of the following vectors:

$$
\begin{aligned}
&\mathbf{h}_n \,, \quad \forall\, n \mid b_n < k \,, \\
&\frac{\mathbf{h}_{f(i,g_i)}}{\mathbf{h}_{f(i,g_i)}\mathbf{m}_i} - \frac{\mathbf{h}_{f(i,j)}}{\mathbf{h}_{f(i,j)}\mathbf{m}_i} \,, \quad \forall\, i,j \mid i > k \text{ and } j < g_i \,,
\end{aligned}
\tag{7}
$$

where $n \in \{1, \ldots, N\}$, $i \in \{1, \ldots, K\}$, and $j \in \{1, \ldots, N\}$. After these null space projections, the remaining degrees of freedom for determining $\mathbf{m}_k$ can be exploited by performing multicast beamforming, such as the maximization of the minimum SNR in [2, 4].

Let $\mathbf{H}_k \in \mathbb{C}^{g_k \times M}$ denote the channel matrix of the users belonging to group k and $\mathbf{N}_k \in \mathbb{C}^{M \times L_k}$ the null space of (7), where the dimension L_k is given by

$$
L_k = M - \sum_{i=1}^{k-1} g_i - \sum_{i=k+1}^{K} (g_i - 1),
\tag{8}
$$

assuming that matrix $\mathbf{H}_k$ has full row rank. The equivalent channel matrix $\mathbf{H}'_k \in \mathbb{C}^{g_k \times L}$ after the projection is given by $\mathbf{H}'_k = \mathbf{H}_k \mathbf{N}_k$. The multicast beamforming procedure is done considering $\mathbf{H}'_k$ and results in the beamforming vector $\mathbf{w}_k \in \mathbb{C}^{L_k \times 1}$. The kth column of the modulation matrix $\mathbf{M}$ is then set to $\mathbf{m}_k = \mathbf{N}_k \mathbf{w}_k$.

It should be noted that the independent optimization of $\mathbf{m}_k$ balances the energy within each group, but not among different groups, due to the projections required by (7). For this reason it is required that the available energy be redistributed among the groups, so that the balancing effect can be achieved. Let $\mathbf{r}_k \in \mathbb{C}^{g_k \times 1}$ represent the channel gains within each group, i.e., $\mathbf{r}_k = \mathrm{diag}(\mathbf{H}_k \mathbf{m}_k \mathbf{m}_k^H \mathbf{H}_k^H)$, then the power redistribution matrix $\mathbf{\Gamma} \in \mathbb{R}^{K \times K}$ is defined as

$$
\mathbf{\Gamma} = \mathrm{diag}([\min(\mathbf{r}_1), \min(\mathbf{r}_2), \ldots, \min(\mathbf{r}_K)]^T)^{-1/2},
\tag{9}
$$

and the modulation matrix is reset to $\mathbf{M} = \mathbf{M}\mathbf{\Gamma}$. In order to satisfy the transmit energy constraint, the matrix $\mathbf{M}$ is additionally multiplied by a scalar variable $\beta \in \mathbb{R}$, which is defined as $\beta = \sqrt{E_{tr}/\mathrm{trace}(\mathbf{M}^H \mathbf{M} \mathbf{R}_v)}$, such that $\mathbf{M} = \beta \mathbf{M}$.

The matrix $\mathbf{M}$ lying on the null space of (7) allows for a feasible solution of (3). The filter $\mathbf{F}$ can then be calculated as

$$
\mathbf{F} = \mathbf{I} - [(\mathbf{HM})^+ \mathrm{diag}_\mathrm{b}(\mathbf{HM})]^{-1},
\tag{10}
$$

where $(\cdot)^+$ denotes the pseudo-inverse of a matrix.

Since independent single antenna users are considered, the demodulation matrix $\mathbf{D}$ is diagonal, and it is assumed that a matched filter is implemented at each receiver. The diagonal elements $D_{n,n}$ of matrix $\mathbf{D}$ can be expressed as

$$
D_{n,n} = \frac{(\mathbf{h}_n \mathbf{m}_{b_n})^*}{|\mathbf{h}_n \mathbf{m}_{b_n}|^2}, \quad \text{for} \quad n = 1, \ldots, N,
\tag{11}
$$

where $(\cdot)^*$ denotes the complex conjugate, and $|\cdot|$ is the absolute value operator.

The performance of the Tomlinson-Harashima precoding depends strongly on how the data streams are ordered prior to transmission. The best ordering is the one which minimizes the impact of the null space projections, such that the least amount of energy is lost. The optimum ordering can only be determined by exhaustively searching among all $N!$ possibilities. However, there exist more computationally efficient ordering methods which reasonably approximate the optimal performance [12].

In the case of the unicast/multicast scenario the number of possible orderings is reduced from $N!$ to $K!$, since the position of the users within each group does not impact the performance. One drawback of this procedure corresponds to the additional null space projections. This may lead to cases in which THP outperforms MA-THP.

In order to obtain the best possible performance, it is here proposed to employ both strategies. This means that of the $N!$ possibilities investigated by THP, for $K!$ of them the MA-THP algorithm is employed and its performance compared with that of THP. The one presenting the best performance is selected for transmission. This is possible since the receiver structure for both algorithms is the same. This strategy can also be adapted for a suboptimal lower complexity ordering algorithm, but such a study is out of the scope of this paper.

IV. Hybrid linear/non-linear precoding (HLNP)

Another possible approach for performing SDMA in a unicast/multicast scenario is to employ a mix of linear and non-linear precoding schemes. In order to avoid the additional null space projections of the unicast/multicast THP, this section considers that linear processing is applied to the multicast groups, while THP is applied to the unicast users.

It is here assumed that the block of unicast users comes first, and is then followed by the multicast blocks. Note that the ordering among blocks is not relevant, since the inter-block interference is assumed to be removed by the linear filter. However, it still plays an important role within the unicast block.

Let N_{uc} indicate the number of unicast users, then the constraints on matrix $\mathbf{M}$ can be written as

$$
\begin{aligned}
&\mathbf{h}_n \mathbf{m}_k = 0 \,, \\
&\forall n \in \{1, \ldots, N_{uc}\}, \, k \in \{1, \ldots, N_{uc}\} \mid n < k \,, \\
&\forall n \in \{N_{uc} + 1, \ldots, N\}, \, k \in \{N_{uc} + 1, \ldots, K\} \mid b_n \neq k \,.
\end{aligned}
\tag{12}
$$

The corresponding null space projections result in a block $\mathbf{HM}$ matrix with a triangular block corresponding to the unicast

Figure 2: Bit error rate results for scenario S1 ($N_{uc} = 2$).
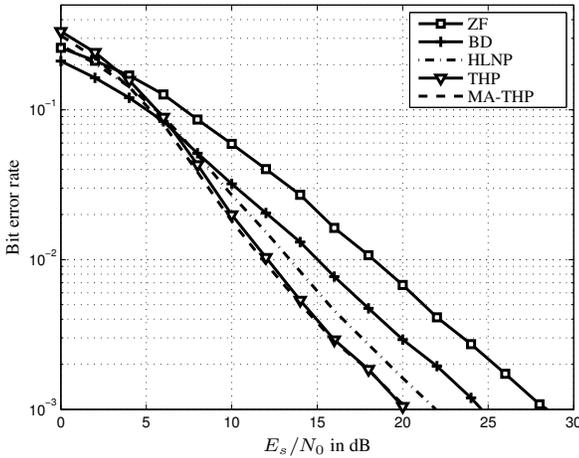

Figure 3: Bit error rate results for scenario S2 ($N_{uc} = 1$).

users. Multicast beamforming can be performed within each multicast block individually [2, 6]. Similar to the previous section, a power redistribution needs to be performed in order to balance the quality among the blocks and the transmit power constraint needs to be respected.

Let $\mathbf{H}_{uc} \in \mathbb{C}^{N_{uc} \times M}$, $\mathbf{M}_{uc} \in \mathbb{C}^{M \times N_{uc}}$, and $\mathbf{F}_{uc} \in \mathbb{C}^{N_{uc} \times N_{uc}}$ denote, respectively, the channel, modulation, and feedback matrices of the unicast block. The expression for the global feedback filter $\mathbf{F}$ can be expressed as

$$\mathbf{F}_{uc} = \mathbf{I} - \mathrm{diag}(\mathbf{H}_{uc}\mathbf{M}_{uc})^{-1}\mathbf{H}_{uc}\mathbf{M}_{uc}\,,$$
$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{uc} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\,, \tag{13}$$

where $\mathbf{0}$ are null matrices of appropriate dimension. The demodulation matrix $\mathbf{D}$ can be calculated as in (11).

The same transmitter structure of the previous section is also valid for this case, since the feedback matrix, and consequently the modulo operator, will not have any impact on the multicast blocks. The same applies to the receiver structure, which in the case of the multicast receivers can be additionally simplified, since they do not need to implement the modulo operator.

V. Performance Analysis

In this section, the performance of the algorithms is analyzed. The simulation scenario consists of a single cell equipped with a four-element uniform linear antenna array and single antenna mobile terminals. An uncorrelated channel matrix $\mathbf{H}$ is considered, which is composed of zero mean circularly symmetric complex Gaussian random variables with unit variance. A total of 10^5 channel realizations are simulated, and for each realization 100 QPSK symbols are transmitted. Note that the effects of path-loss and log-normal fading are assumed to be compensated by power control.

The symbols are assumed to have variance $\sigma_s^2 = 1$, but those affected by the modulo operator have a larger energy, which is given by $\sigma_v^2 = \tau^2/6$, where $\tau = 2\sqrt{2}$ for the QPSK constellation [8]. The symbols, as well as the noise, are assumed to be uncorrelated, so that their covariance matrices $\mathbf{R}_v$ and $\mathbf{R}_n$
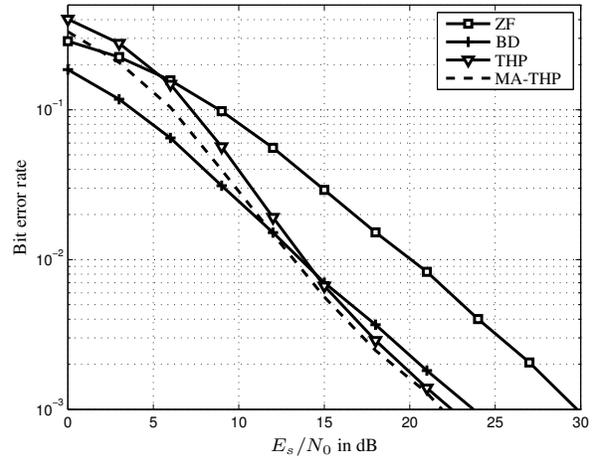
are diagonal. The total available transmit energy is set to be proportional to the number of streams, i.e., $E_{tr} = K\sigma_s^2$.

Two user scenarios are considered: scenario S1 assumes two unicast users and a two-user multicast group, while scenario S2 assumes one unicast user and a three-user multicast group.

The simulation results are obtained for the following algorithms: the multicast-aware algorithm of section III (MA-THP), the hybrid linear and non-linear algorithm of section IV (HLNP), the zero-forcing THP, the block diagonalization (BD), and the zero-forcing (ZF). Note that optimal stream ordering is considered within the simulations.

Figs. 2 and 3 present the uncoded bit error rate (BER) performance of the different algorithms for scenarios S1 and S2, respectively. The BER is depicted as a function of the E_s/N_0, which represents the ratio of the symbol energy to the spectral noise density.

In Fig. 2 it can be seen that, except for low E_s/N_0 values, the MA-THP algorithm presents the best results. However, its performance is closely followed by that of THP, which indicates that for this scenario the additional null space projections of the MA-THP algorithm are generally more energy-consuming than the channel triangularization process of THP. Fig 2 also shows that the HLNP algorithm provides significant gains with regard to the pure linear transmit processing of BD. The ZF algorithm, as expected, achieves the worst performance, which is due to the complete diagonalization of the channels.

From Fig. 3 it can be observed that the relative performance among the algorithms is maintained, with MA-THP presenting the best performance for a large potion of the E_s/N_0 range, and being followed by the THP, BD, and ZF algorithms. Note that HLNP is not depicted in the figure, since for the case in which there is only one unicast user its operation is the same as that of the BD algorithm. Still in Fig. 3 it can be seen that BD presents the best results for SNR values lower than approximately 12dB, and even for high SNR it maintains a performance close to that of the non-linear algorithms.

The increased number of multicast users in scenario S2 has a positive impact on the performance of both the BD and MA-THP, since the multicast beamforming can achieve higher gains
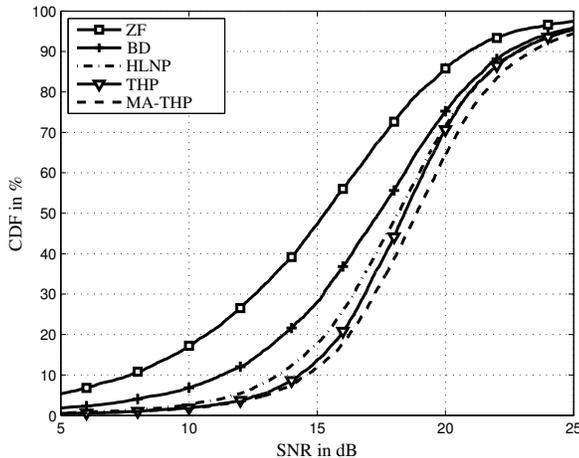
Figure 4: Average SNR distribution for scenario S1 and $E_s/N_0 = 10$dB.

and there is less interference to be suppressed when compared to scenario S1.

From the results it can be seen that the performance of MA-THP does not present much gains with regard to pure THP, especially for scenarios with lower number of multicast users. One reason for this can be said to be the additional null space projections that are required in order to achieve a feasible feedback matrix. Another reason is that a significant part of the gains of THP over linear processing techniques is due to the non-linearity of the modulo operator, which in the case of MA-THP is applied to a lower dimensional signal vector. Nevertheless, there is still room for improving MA-THP, which can be done by the determination of more efficient suboptimal solutions to the optimization problem in (6).

Additionally, for illustration purposes, the performance of the algorithms in terms of the SNR perceived by the users is shown in Fig. 4. In this figure the cumulative distribution function (CDF) of the average user SNR is presented for scenario S1 and an E_s/N_0 of 10dB. Note that in the case of the non-linear receivers this SNR is calculated prior to the application of the modulo operator.

VI. Conclusions

This paper analyzes and compares linear and non-linear downlink precoding techniques for separating users of both unicast and multicast services in space.

Motivated by the good results of linear multicast-aware techniques, this paper demonstrates that it is possible to derive a multicast-aware non-linear algorithm based on Tomlinson-Harashima precoding (MA-THP). This comes, however, at the cost of some undesirable constraints. Additionally, a hybrid linear/non-linear precoding algorithm (HLNP) is proposed, which consists of combining both the BD and THP strategies.

A simulation analysis has been conducted for two different user scenarios in terms of the bit error rate and the SNR. The results have shown that MA-THP achieves the best performance, but it is only slightly better than that of traditional THP, and that the HLNP algorithm outperforms both BD and ZF. It could

also be observed, as expected, that the increase of the multicast group size had a positive impact on the performance of the multicast-aware algorithms.

The formulation of the MA-THP algorithm has shown that it suffers from some limitations, which considerably hold back its performance and prevent that larger gains be achieved with regard to THP. It is therefore an interesting topic for further study the investigation of more efficient suboptimal solutions to the problem of determining the modulation matrix, such that the signal energy is maximized and at the same time the feedback filter feasibility constraints are satisfied.

Finally, it should be mentioned that the HLNP strategy presents good results, reasonably approaching the performance of the other non-linear techniques. It also presents the benefits of a lower stream ordering complexity than both THP and MA-THP, as well as lower complexity multicast receivers, since they are not required to implement the modulo operator.

REFERENCES

[1] 3GPP, "Multimedia Broadcast/Multicast service (MBMS); architecture and functional description," TS 23.246, 2005.

[2] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. on Signal Processing*, vol. 54, pp. 2239–2251, June 2006.

[3] J. Zhang *et al.*, "Optimal space-time transceiver design for selective wireless broadcast with channel state information," *IEEE Trans. on Wireless Communications*, vol. 3, pp. 2040–2050, Nov. 2004.

[4] Y. C. B. Silva and A. Klein, "Adaptive antenna techniques applied to multicast services in wireless networks," *Frequenz - Journal of RF-Engineering and Telecommunications*, vol. 60, pp. 199–202, Sep./Oct. 2006.

[5] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Transmit beamforming to multiple co-channel multicast groups," *in Proc. International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, pp. 109–112, Dec. 2005.

[6] Y. C. B. Silva and A. Klein, "Adaptive beamforming and spatial multiplexing of unicast and multicast services," *in Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2006.

[7] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. on Wireless Communications*, vol. 3, pp. 1305–1316, July 2004.

[8] M. Joham, "Optimization of linear and nonlinear transmit signal processing," Ph.D. dissertation, TU München, 2004.

[9] W. Qiu, M. Meurer, P. W. Baier, and T. Weber, "Power efficient CDMA broadcast system doing without any channel knowledge at the receivers, a nonobvious modification of THP," *in Proc. Asia-Pacific Conference on Communications*, vol. 2, pp. 793–799, Aug. 2004.

[10] J. Liu and W. A. Krzymień, "A null space constraint based block Tomlinson-Harashima precoding technique for the multi-user MIMO downlink," *in Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, pp. 61–64, Aug. 2005.

[11] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. on Signal Processing*, vol. 52, pp. 461–471, Feb. 2004.

[12] M. Joham, D. A. Schmidt, H. Brunner, and W. Utschick, "A symbol-wise order optimization for successive precoding," *in Proc. ITG/IEEE Workshop on Smart Antennas*, Feb. 2007.