Y. Silva, A. Klein, ”Linear transmit beamforming techniques for the multi-group multicast scenario,” in *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. XXX - XXX, Nov. 2009

Linear transmit beamforming techniques for the multi-group multicast scenario

Yuri C. B. Silva, *Member, IEEE,* Anja Klein, *Member, IEEE,*

***Abstract*—In the context of next-generation wireless systems, it is expected that services targeted at mass content distribution become widely popular. Multicast services, such as audio/video streaming and mobile TV, have the characteristic that the same information has to be transmitted to a group of recipients, which may include user terminals as well as relay stations, in the case of relaying networks. This paper deals with the problem of multi-group multicast beamforming for multi-antenna wireless cellular networks, which assumes that multiple multicast groups can share the same resource. The inter-group interference, which appears due to the resource sharing, needs to be suppressed by the beamforming algorithms. In this paper, new linear transmit beamforming techniques are proposed, which aim at providing a reasonable trade-off between performance and computational complexity. These techniques correspond to extensions of known algorithms, such as Zero-Forcing, MMSE, and SINR Balancing, to the multicast case. Furthermore, new "multicast-aware" techniques that take into account the peculiarities of multicast transmission are proposed for improving performance. In order to provide the necessary mathematical framework for the analysis, a novel general multi-group multicast system model is proposed as well.**

***Index Terms*—Multicast transmission, transmit beamforming, Spatial Division Multiple Access (SDMA).**

I. INTRODUCTION

IN the context of next-generation wireless systems, it is expected that services targeted at mass content distribution become widely popular. Examples of such services are audio/video streaming, mobile TV, messaging, news clips, localized services, download, among others. Multicast services have the characteristic that the same information has to be transmitted to a group of recipients. Broadcast services can be seen as a particular case of multicast services, in which there is not a specific target group, i.e., *all* users belong to the same group. Such services can be implemented through Point-to-Multipoint (P2M) connections, in which a single source transmits the data to all users belonging to the intended group.

The support of multicast services in cellular networks has been introduced by both the Global System for Mobile communications (GSM) and Universal Mobile Telecommunications System (UMTS) networks in the form of the Multimedia Broadcast/Multicast Service (MBMS) [1-3]. More recently, a multicast architecture based on MBMS has been proposed in [4] for Worlwide interoperability for Microwave Access (WiMAX) networks. Additionally, relaying strategies may also benefit from enhanced multicast transmission, since the same information is transmitted to one or more nodes.

The introduction of adaptive antenna arrays at the base station may contribute to the performance improvement of the multicast. The multicast beamforming problem consists of determining suitable antenna weight vectors, assuming that knowledge of the radio link of all multicast users is available at the transmitter. Let us define a radio resource in the time and frequency domains, e.g., a timeslot/frequency pair. In this regard, two different types of multicast beamforming techniques can be identified: single-group and multi-group beamforming. The former assigns a different radio resource to each multicast group [5-9], whereas the latter assumes that multiple multicast groups can share the same resource [10-15].

In this paper, the theme of multi-group multicast beamforming is approached. The motivation is to exploit the spatial dimension provided by the multiple antennas in order to provide an efficient utilization of the radio resources, also known as Spatial Division Multiple Access (SDMA). The challenge is to design efficient low-complexity algorithms capable of suppressing the inter-group interference, while at the same time providing the best possible quality to the users of the different multicast groups.

This paper is organized as follows. An overview of the recent research on multi-group multicast beamforming is presented in Section II. A novel general multi-group multicast system model is presented in Section III. It is shown in Section IV that linear algorithms originally designed for the multi-user unicast scenario can be extended to the multi-group multicast case. These are straightforward algorithms, which better situate and facilitate the comprehension of the more elaborate algorithms of the subsequent section. The main contribution of the paper is presented in Section V, where new Multicast-Aware (MA) algorithms are proposed in order to improve the performance of the multicast transmission. An analysis of the performance and complexity of the algorithms is presented in Section VI. Finally, the main conclusions are summarized in Section VII.

II. REVIEW OF PREVIOUS WORK

In a multi-group multicast scenario, differently from the single-group case, there are several data streams being trans-

Manuscript received May 14, 2008; revised November 17, 2008 and February 14, 2009. The associate editor coordinating the review of this paper and approving it for publication was Prof. Gerhard Bauch. This work was supported in part by CAPES-Brazil.

Y. C. B. Silva was with the Communications Engineering Lab, Technische Universität Darmstadt, Darmstadt, Germany. He is now with the Wireless Telecom Research Group - GTEL, Federal University of Ceará, Fortaleza, Brazil (e-mail: yuri@gtel.ufc.br).

A. Klein is with the Communications Engineering Lab, Technische Universität Darmstadt, Darmstadt, Germany (e-mail: a.klein@nt.tu-darmstadt.de).

mitted simultaneously on the same radio resource, each targeted at a different multicast group. Due to this simultaneous transmission of different streams at the same time and at the same resource, it can happen that each receiver sees the streams that are intended for other groups as interference. This inter-group interference has a significant impact on the solution of the optimization problem. Moreover, since within each group the users may be subject to different radio link qualities, a group measure must be defined, such as the worst-user Signal-to-Interference plus Noise Ratio (SINR).

Previous works have mainly dealt with the following two optimization problems for the multi-group multicast scenario:

- P1: maximization of the overall worst-user SINR, i.e., among all users of all groups, subject to transmit power constraints.
- P2: minimization of the total transmit power subject to minimum SINR requirements per user, i.e., each user of each group must achieve or exceed the demanded thresholds.

Both optimization problems correspond to quadratically constrained quadratic programming problems [16]. Problem P2 was claimed to be Nondeterministic Polynomial time hard (NP-hard) by Karipidis et al. in [10]. The equivalence of both problems for the single-group beamforming case indicates that the problem P1 might be NP-hard as well.

The multi-group multicast beamforming problem has first been discussed by Lopez in [17], where the use of null space projections has been suggested for eliminating the interference among the data streams of different groups. After the projections, an equivalent channel matrix is achieved, whose non-zero elements are grouped into "array processing subblocks". These subblocks determine the type of transmit processing technique to be employed, which can be: single-group multicast beamforming, single-user unicast beamforming, or non-linear precoding for a group of unicast users. Note that the author in [17] has suggested this approach, but the referred work has not presented any equations or performance results on this subject. This approach has served as inspiration for some of the algorithms presented in Section V.

A precoding strategy based on an extension of Dirty Paper Coding (DPC) [18] for the multi-group multicast scenario has been proposed by Khisti in [19]. The strategy in [19], however, is based on a sum rate maximization criterion that is not adequate to the optimization problems P1 or P2.

In [10], Karipidis et al. proposed a method based on Semi-Definite Relaxation (SDR) for the multi-group multicast optimization problem P2 of minimizing the transmit power subject to SINR constraints. It corresponds to an extension of the single-group multicast beamforming algorithm in [20, 7], and is based on the multi-user unicast case presented in [21]. Similar to the single-group multicast case, the rank 1 relaxation allows the problem to be solved efficiently through semi-definite optimization methods. If the obtained solution has rank 1, then it corresponds to the optimal solution, otherwise randomization techniques need to be employed in order to improve the solution [20, 10]. The problem is that, differently from the single-group case, it is no longer possible to simply scale the generated randomized beamforming vectors, due to the inter-group interference. In [10], the problem of converting each candidate vector into a feasible solution is called "multi-group power control" and is expressed as a linear programming problem, which can also be solved through semi-definite optimization. This additional optimization problem, however, increases the complexity of the algorithm. The specific case of multi-group multicast beamforming for Vandermonde channel matrices is approached in [22, 13], where it is shown that the relaxed problem always leads to rank 1 matrices, meaning that the optimal solution is always achieved for this case.

Gao and Schubert proposed in [11] another solution to problem P2 than that of [10]. The difference with regard to [10] is that DPC is initially employed and a block-triangular channel is taken into account. Such a channel structure allows a group-by-group algorithm, since the interference from previous groups is known. The beamforming vectors are successively determined for each group by employing single-group beamforming based on SDR [20]. The power allocated to each beamforming vector is also determined successively through a simple algorithm.

In [12], the same SDR approach of [10] is employed to solve problem P2, but instead of solving the "multi-group power control" through semi-definite programming, an iterative power allocation method based on worst-case interference functions is proposed.

With regard to the problem P1 of maximizing the worst-user SINR, differently from the single-group multicast case, its solution cannot be directly obtained by scaling the solution of P2. It has been shown in [12] and [13], however, that it can be solved through a bisection method. The idea is to specify an SINR interval within which the optimal solution must lie, and to determine the solution of problem P2 when considering the middle point of the interval as the target SINR. The interval is then successively bisected, based on whether the required amount of power P_{req} exceeds the transmit power constraint P or not. For each interval middle point, the corresponding problem P2 is solved. The bisection proceeds until a desired precision is reached with regard to $|P_{\mathrm{req}} - P|$.

Another method for determining a solution to P1, which is based on the alternating optimization procedure of [23], has been proposed in [12]. It employs an iterative power allocation algorithm, which determines the power allocation vector and the maximum achievable worst-user SINR, given a fixed set of beamforming vectors. Additionally, given a certain SINR target, the SDR approach of [11] is used to determine the beamforming vectors. The power allocation and beamforming algorithms are alternately executed, until the worst-user SINR stops increasing.

The extension of multi-user unicast beamforming techniques to the multi-group multicast case, however, has not been investigated by previous works. For this reason, formulations of new beamforming algorithms for the multi-group multicast scenario are proposed in Sections IV and V. Additionally, it is shown that these algorithms can be enhanced by introducing modifications that aim at improving the performance of multicast transmission. These new enhanced algorithms are further referred to as Multicast-Aware (MA) algorithms.

III. System model

In this section, a model for the multi-group multicast case is introduced, which is a generalization of the unicast-only models [24, 25], as well as of the single-group multicast models [6, 7]. It is also a further development of the multi-group multicast models presented in [10, 11]. The proposed model provides details on the transmission/reception chain and the system parameters are flexible enough, so that they can be adjusted to represent particular cases of the general model, such as single-group multicast, multi-user unicast, and single-user unicast.

A multi-user system is considered, which assumes flat-fading and negligible inter-symbol interference, so that the data symbols can be treated individually. The base station has an antenna array composed of L elements and serves N single-antenna users which are grouped into K multicast groups. Considering a vector $\mathbf{s} \in \mathbb{C}^N$ with N data symbols, which are modulated by a matrix $\mathbf{M} \in \mathbb{C}^{L\times N}$, transmitted over the radio channel $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_N]^{\mathrm{T}} \in \mathbb{C}^{N\times L}$, where $\mathbf{h}_n \in \mathbb{C}^L$ corresponds to the channel of the n^{th} user, subject to additive white Gaussian noise $\mathbf{z} \in \mathbb{C}^N$, and demodulated by a diagonal matrix $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_N) \in \mathbb{C}^{N\times N}$, the N downlink estimates $\hat{\mathbf{s}} \in \mathbb{C}^N$ of the N transmitted symbols $\mathbf{s}$ may be written as

$$\hat{\mathbf{s}} = \mathbf{DHMs} + \mathbf{Dz}. \tag{1}$$

Fig. 1 depicts the block diagram of the system. Note that the vectors $\mathbf{x} \in \mathbb{C}^L$ and $\mathbf{y} \in \mathbb{C}^N$ denote, respectively, the precoded transmit symbol vector and the received symbol vector prior to decoding, for which

$$\mathbf{x} = \mathbf{Ms}\,, \tag{2}$$

$$\mathbf{y} = \mathbf{HMs} + \mathbf{z}\,. \tag{3}$$

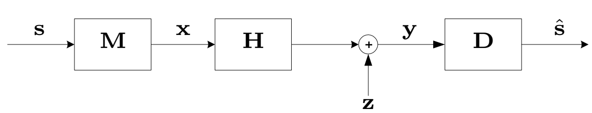

Fig. 1. Block diagram of the general system model in the frequency domain.

The system model is not yet complete at this point, since the multicast characterization is still missing. For this purpose it is necessary to introduce a pair of auxiliary vectors. Let the variables $l \in \{1, \ldots, L\}$, $n \in \{1, \ldots, N\}$, and $k \in \{1, \ldots, K\}$ denote the index of antenna elements, users, and multicast groups, respectively. The number of users within each group is represented by vector $\mathbf{g} \in \mathbb{Z}^K$, whose k^{th} element $g_k \in \{1, \ldots, N\}$ indicates the number of users within group k. Note that the unicast users can be interpreted as multicast groups of unit size and that $\sum_{k=1}^{K} g_k = N$. In order to associate which users belong to which group, an index vector $\mathbf{b} \in \mathbb{Z}^N$ 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 system with two unicast users and one multicast group composed of two users, we would have: $N = 4$, $K = 3$, $\mathbf{g} = [1, 1, 2]^{\mathrm{T}}$, and $\mathbf{b} = [1, 2, 3, 3]^{\mathrm{T}}$. In order to better illustrate some concepts, in the following, this particular system configuration will be referred to as the exemplary system.

An alternative representation for the system in (1), called reduced representation, is now presented. Since the users of a multicast group expect the same data stream, the number K of multicast groups is also equivalent to the number of different data streams. For this reason there are $N-K$ repeated entries within vector $\mathbf{s} \in \mathbb{C}^N$. The removal of such repeated entries results in vector $\mathbf{s}' \in \mathbb{C}^K$. This operation can be mathematically expressed as

$$\mathbf{s}' = \mathbf{Ts}\,, \tag{4}$$

where $\mathbf{T} \in \mathbb{R}^{K\times N}_{+,0}$ is a transformation matrix with the n^{th} column given by $\mathbf{t}_n = g_{b_n}^{-1}\mathbf{e}_{b_n}$, for which $\mathbf{e}_i$ corresponds to the i^{th} column of the identity matrix of dimension $K \times K$.

The reduced dimension of the data vector also leads to a reduced modulation matrix $\mathbf{M}' \in \mathbb{C}^{M\times K}$, i.e., instead of one beamforming vector per user there is now one beamforming vector per multicast group. Let $\mathbf{m}_i$ and $\mathbf{m}'_i$ represent the i^{th} column of matrices $\mathbf{M}$ and $\mathbf{M}'$, respectively. They are related by

$$\mathbf{m}'_k = \sum_{n=1,\, b_n=k}^{N} \mathbf{m}_n\,, \quad \text{for} \quad k = 1, \ldots, K\,. \tag{5}$$

Matrix $\mathbf{M}'$ can also be written as the following transformation of matrix $\mathbf{M}$:

$$\mathbf{M}' = \mathbf{MT}^{+}\,, \tag{6}$$

where $\mathbf{T}^{+} \in \mathbb{R}^{N\times K}$ is the right pseudoinverse of matrix $\mathbf{T}$ in (4). $\mathbf{T}^{+}$ has its n^{th} row given by $\mathbf{t}_n^{+} = \mathbf{e}_{b_n}^{\mathrm{T}}$. In the case of the exemplary system, matrices $\mathbf{T} \in \mathbb{R}^{3\times 4}$ and $\mathbf{T}^{+} \in \mathbb{R}^{4\times 3}$ are given by

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}, \quad \mathbf{T}^{+} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

The complete and reduced representations have different symbol vectors and modulation matrices, nevertheless they still represent the same system. This can be confirmed by the following equation:

$$\mathbf{M}'\mathbf{s}' = \mathbf{MT}^{+}\mathbf{Ts} = \mathbf{Ms}\,. \tag{8}$$

Note that, even though $\mathbf{T}^{+}\mathbf{T}$ is not an identity matrix, it can be shown that $\mathbf{T}^{+}\mathbf{Ts} = \mathbf{s}$, due to the repeated entries within $\mathbf{s}$. This same property can also be used to isolate $\mathbf{s}$ in (4), which leads to

$$\mathbf{s} = \mathbf{T}^{+}\mathbf{s}'\,. \tag{9}$$

After substituting $\mathbf{M}'$ and $\mathbf{s}'$ in (1), the system equation can be rewritten in reduced form as

$$\hat{\mathbf{s}} = \mathbf{DHM}'\mathbf{s}' + \mathbf{Dz}. \tag{10}$$

The reduced representation, however, does not completely eliminate the need for the complete representation. The $\mathbf{s}$ vector is still needed by some of the algorithms in upcoming sections, e.g., for the calculation of the mean square error. Other algorithms, such as the SINR balancing, require both reduced and complete representations.

From (10), the expression for the estimated data symbol of each user n can be written as

$$\hat{s}_n = d_n \mathbf{h}_n^{\mathrm{T}} \mathbf{m}'_{b_n} s'_{b_n} + \sum_{k=1,\, k\neq b_n}^{K} d_n \mathbf{h}_n^{\mathrm{T}} \mathbf{m}'_k s'_k + d_n z_n \,, \quad (11)$$

where $\mathbf{h}_n^{\mathrm{T}}$ corresponds to the n^{th} row of matrix $\mathbf{H}$. The three summands correspond, respectively, to the signal, interference, and noise parts of $\hat{s}_n$.

The design of the transmit filter is the topic of Sections IV and V. The receive filter, on the other hand, can already be determined at this point[1]. As it has been previously mentioned, there is an independent receive filter $d_n \in \mathbb{C}$ at each user terminal. It is assumed that each user terminal n knows the equivalent radio channel to the base station, which is given by $\mathbf{h}_n^{\mathrm{T}} \mathbf{m}'_{b_n} \in \mathbb{C}$. This information can be obtained, for example, if the base station transmits pilot symbols at the beginning of each Orthogonal Frequency Division Multiplexing (OFDM) frame, so that the user terminal can estimate the equivalent channel. It is here assumed that the receive filter satisfies the constraint that, in the absence of noise and interference, the estimated symbols are exactly the same as the original data symbols. This leads to

$$d_n \mathbf{h}_n^{\mathrm{T}} \mathbf{m}'_{b_n} s'_{b_n} = s'_{b_n} \Longrightarrow d_n = (\mathbf{h}_n^{\mathrm{T}} \mathbf{m}'_{b_n})^{-1} \,, \quad (12)$$

and the filter expression in matrix form is given by

$$\mathbf{D} = \operatorname{diag}(\mathbf{h}_1^{\mathrm{T}} \mathbf{m}'_{b_1}, \ldots, \mathbf{h}_N^{\mathrm{T}} \mathbf{m}'_{b_N})^{-1} \,, \quad (13)$$

where the $\operatorname{diag}(\cdot)$ operator returns a diagonal matrix when the argument is a vector or it returns a vector with the main diagonal elements when the argument is a matrix. Matrix $\mathbf{D}$ can also be alternatively expressed as

$$\mathbf{D} = \operatorname{diag}((\mathbf{H}\mathbf{M}' \odot \mathbf{T}^{+})\mathbf{1})^{-1} \,, \quad (14)$$

where the symbol $\odot$ denotes the entry-wise matrix product and $\mathbf{1}$ denotes a vector of ones. This expression is taken into account by all algorithms considered in this paper, except for the Zero-Forcing (ZF) and Minimum Mean Square Error (MMSE) algorithms, for which $\mathbf{D}$ is calculated as a result of the optimization procedure.

Next, the system is further characterized by defining the downlink SINR, the transmit power constraints, and the signal covariance matrices.

It is assumed that during a period of time T_{f}, which corresponds for example to an OFDM frame duration, the channel as well as the transmit and receive filters are time-invariant. The random variables correspond to the data symbols and noise. The instantaneous SINR of user n for a given channel realization is denoted by γ_n, which assumes a large enough number of symbols, such that the symbol and noise powers converge to their expectation. Taking (11) into account, the SINR γ_n, for $n = 1, \ldots, N$, is given by

$$\begin{aligned} \gamma_n &= \frac{\mathrm{E}\{|d_n \mathbf{h}_n^{\mathrm{T}} \mathbf{m}'_{b_n} s'_{b_n}|^2\}}{\mathrm{E}\left\{\sum\limits_{k=1,\, k\neq b_n}^{K} |d_n \mathbf{h}_n^{\mathrm{T}} \mathbf{m}'_k s'_k|^2\right\} + \mathrm{E}\{|d_n z_n|^2\}} \\ &= \frac{\sigma_s^2 \, |\mathbf{h}_n^{\mathrm{T}} \mathbf{m}'_{b_n}|^2}{\sum\limits_{k=1,\, k\neq b_n}^{K} \sigma_s^2 \, |\mathbf{h}_n^{\mathrm{T}} \mathbf{m}'_k|^2 + \sigma_z^2} \,, \end{aligned} \quad (15)$$

where σ_s^2 and σ_z^2 correspond, respectively, to the average symbol and noise power.

It is assumed that the maximum power available for transmission is denoted by P. As a consequence, the design of matrix $\mathbf{M}$ must satisfy the following power constraint:

$$\mathrm{E}\{||\mathbf{M}\mathbf{s}||^2\} = \operatorname{tr}(\mathbf{M}^{\mathrm{H}}\mathbf{M}\mathbf{R}_s) \leq P \,, \quad (16)$$

where $\operatorname{tr}(\cdot)$ denotes the trace of a matrix and $\mathbf{R}_s = \mathrm{E}\{\mathbf{s}\mathbf{s}^{\mathrm{H}}\} \in \mathbb{C}^{N\times N}$ is the signal covariance matrix. Note that, in the case of uncorrelated and equiprobable symbols, $\mathbf{R}_s$ corresponds to a block diagonal matrix, with each block k equal to $\sigma_s^2 \mathbf{J} \in \mathbb{R}^{g_k \times g_k}$, where $\mathbf{J}$ corresponds to a matrix of ones. Equivalently, with $\mathbf{M}'$ given by (5), the constraint may also be expressed as:

$$\operatorname{tr}(\mathbf{M}'^{\mathrm{H}}\mathbf{M}'\mathbf{R}'_s) \leq P \,, \quad (17)$$

for which $\mathbf{R}'_s = \mathrm{E}\{\mathbf{s}'\mathbf{s}'^{\mathrm{H}}\} \in \mathbb{C}^{K\times K}$. The assumption of uncorrelated and equiprobable symbols is considered throughout the paper, i.e., $\mathbf{R}'_s = \sigma_s^2 \mathbf{I} \in \mathbb{R}^{K\times K}$, where $\mathbf{I}$ is the identity matrix.

IV. Extension of Unicast Algorithms to the Multi-Group Multicast Case

A. Matched filter algorithm

The derivation of the Matched Filter (MF) for the multi-group multicast scenario is based on the optimization problem for the multi-user unicast case presented in [26]. It aims at maximizing the total SNR perceived at each terminal prior to the receive filter, without taking the inter-group interference into account. The unicast problem for determining the modulation matrix $\mathbf{M}$ can be expressed as

$$\begin{aligned} &\mathbf{M}_{\mathrm{MF}} = \underset{\mathbf{M}}{\operatorname{argmax}} \, \frac{|\mathrm{E}\{\mathbf{s}^{\mathrm{H}}\mathbf{y}\}|^2}{\mathrm{E}\{||\mathbf{s}||^2\}\mathrm{E}\{||\mathbf{z}||^2\}} \,, \\ &\text{subject to: } \mathrm{E}\{||\mathbf{M}\mathbf{s}||^2\} \leq P \,, \end{aligned} \quad (18)$$

where $\mathbf{y}$ is defined in (3). The multicast optimization can be obtained by substituting the modulation matrix $\mathbf{M}$ and symbol vector $\mathbf{s}$ by the reduced modulation matrix $\mathbf{M}'$ and reduced symbol vector $\mathbf{s}'$, respectively. With (9), the multicast problem is given by

$$\begin{aligned} &\mathbf{M}'_{\mathrm{MF}} = \underset{\mathbf{M}'}{\operatorname{argmax}} \, \frac{|\mathrm{E}\{(\mathbf{T}^{+}\mathbf{s}')^{\mathrm{H}}\mathbf{y}\}|^2}{\mathrm{E}\{||\mathbf{T}^{+}\mathbf{s}'||^2\}\mathrm{E}\{||\mathbf{z}||^2\}} \,, \\ &\text{subject to: } \mathrm{E}\{||\mathbf{M}'\mathbf{s}'||^2\} \leq P \,. \end{aligned} \quad (19)$$

The cost function of the optimization problem corresponds to an equivalent group SINR, denoted by γ_{eq}, which can be

[1] Note that the choice of the receive filter does not impact the SINR, thus any arbitrary scalar filters d_n are optimal in this sense. In terms of the bit error rate, however, a proper receive filter is essential for correctly decoding the received bits.

further expressed as

$$\gamma_{\mathrm{eq}} = \frac{|\mathrm{tr}(\mathbf{H}\mathbf{M}'\mathbf{R}'_s\mathbf{T}^{+,\mathrm{T}})|^2}{\mathrm{tr}(\mathbf{T}^{+}\mathbf{R}'_s\mathbf{T}^{+,\mathrm{T}})\mathrm{tr}(\mathbf{R}_z)}\,, \tag{20}$$

where $\mathbf{R}_z = \mathrm{E}\{\mathbf{z}\mathbf{z}^{\mathrm{H}}\}$ and $(\cdot)^{X,Y}$ corresponds to the sequential application of matrix operators X and Y. Note that, differently from the single-group multicast case, the optimization now involves the determination of a matrix, instead of a vector, which is due to the multiple data streams. The application of the same Lagrange optimization procedure as in [26] leads to the following solution:

$$\mathbf{M}'_{\mathrm{MF}} = \sqrt{\frac{P}{\sigma_s^2\,\mathrm{tr}(\mathbf{H}\mathbf{H}^{\mathrm{H}}\mathbf{T}^{+}\mathbf{T}^{+,\mathrm{T}})}}\mathbf{H}^{\mathrm{H}}\mathbf{T}^{+}\,. \tag{21}$$

B. Zero forcing algorithm

The Zero-Forcing (ZF) optimization problem for the multi-group scenario, utilizing (9), can be written as

$$\begin{aligned}\{\mathbf{M}'_{\mathrm{ZF}},\,\beta_{\mathrm{ZF}}\} = \ &\underset{\mathbf{M}}{\mathrm{argmin}}\,\mathrm{E}\{||\hat{\mathbf{s}} - \mathbf{T}^{+}\mathbf{s}'||^2\}\,,\\ \text{subject to: } &\begin{cases}\mathrm{E}\{||\mathbf{M}'\mathbf{s}'||^2\} \le P\\ \hat{\mathbf{s}}\big|_{\mathbf{z}=\mathbf{0}} = \mathbf{T}^{+}\mathbf{s}'\end{cases}\,,\end{aligned} \tag{22}$$

where the second constraint requires that, in the absence of noise, the estimated symbol vector $\hat{\mathbf{s}}$ must be equal to the complete symbol vector with repeated entries $\mathbf{s} = \mathbf{T}^{+}\mathbf{s}'$. Similarly to the single-group case, it is here assumed that the receive filter for each user is given by a scalar $\beta \in \mathbb{C}$, i.e., $\mathbf{D} = \beta\mathbf{I}$. The second constraint can be further expressed as

$$\hat{\mathbf{s}}\big|_{\mathbf{z}=\mathbf{0}} = \mathbf{T}^{+}\mathbf{s}' \Longrightarrow \beta\mathbf{H}\mathbf{M}' = \mathbf{T}^{+}\,. \tag{23}$$

The Mean Square Error (MSE) cost function, substituting $\hat{\mathbf{s}}$ and taking into account (23), is given by

$$\mathrm{E}\{||\beta\mathbf{H}\mathbf{M}'\mathbf{s}' + \beta\mathbf{z} - \mathbf{T}^{+}\mathbf{s}'||^2\} = \beta^2\mathrm{tr}(\mathbf{R}_z)\,. \tag{24}$$

The Lagrangian function can be expressed as

$$\begin{aligned}\mathrm{L}(\mathbf{M}',\beta,\mu,\mathbf{\Lambda}) = \beta^2\mathrm{tr}(\mathbf{R}_z) + \mu(\mathrm{tr}(\mathbf{M}'^{\mathrm{H}}\mathbf{M}'\mathbf{R}'_s) - P)\\ + \mathrm{tr}(\mathbf{\Lambda}(\mathbf{T}^{+} - \beta\mathbf{H}\mathbf{M}'))\,,\end{aligned} \tag{25}$$

where $\mu \in \mathbb{R}$ and $\mathbf{\Lambda} \in \mathbb{C}^{K\times N}$ are Lagrange multipliers. The optimization procedure employed in [26] for the multi-user unicast case can also be employed to obtain the solution to the multi-group multicast problem, which is given by

$$\mathbf{M}'_{\mathrm{ZF}} = \beta_{\mathrm{ZF}}\mathbf{H}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\mathbf{T}^{+}\,, \tag{26}$$

$$\mathbf{D}_{\mathrm{ZF}} = \beta_{\mathrm{ZF}}^{-1}\mathbf{I}\,, \tag{27}$$

with

$$\beta_{\mathrm{ZF}} = \sqrt{\frac{P}{\sigma_s^2\,\mathrm{tr}((\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\mathbf{T}^{+}\mathbf{T}^{+,\mathrm{T}})}}\,. \tag{28}$$

Note that the receive filter in (27), achieved as a result of the ZF optimization procedure, coincides with the expression obtained when plugging (26) into (14).

C. Minimum mean square error algorithm

The MMSE optimization, also referred to as the Wiener filter in [26], aims at minimizing the MSE subject to a transmit power constraint. For the multi-group scenario it is assumed that $\mathbf{s} = \mathbf{T}^{+}\mathbf{s}'$ and that each receiver implements a scalar filter $\beta \in \mathbb{C}$. The optimization problem can be written as

$$\begin{aligned}\{\mathbf{M}'_{\mathrm{MMSE}},\,\beta_{\mathrm{MMSE}}\} &= \underset{\{\mathbf{M}',\,\beta\}}{\mathrm{argmin}}\,\mathrm{E}\{||\hat{\mathbf{s}} - \mathbf{T}^{+}\mathbf{s}'||^2\}\,,\\ \text{subject to: } &\mathrm{E}\{||\mathbf{M}'\mathbf{s}'||^2\} \le P\,,\end{aligned} \tag{29}$$

The MSE cost function can be further expressed as

$$\begin{aligned}\mathrm{E}\{||\hat{\mathbf{s}} - \mathbf{T}^{+}\mathbf{s}'||^2\} = \mathrm{tr}\Big(&|\beta|^2\mathbf{M}'^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\mathbf{H}\mathbf{M}'\mathbf{R}'_s\\ &-2\mathrm{Re}(\beta\mathbf{T}^{+,\mathrm{T}}\mathbf{H}\mathbf{M}'\mathbf{R}'_s) + \mathbf{T}^{+,\mathrm{T}}\mathbf{T}^{+}\mathbf{R}'_s + |\beta|^2\mathbf{R}_z\Big)\,.\end{aligned} \tag{30}$$

The Lagrangian function is given by

$$\mathrm{L}(\mathbf{M},\beta,\mu) = \mathrm{E}\{||\hat{\mathbf{s}} - \mathbf{T}^{+}\mathbf{s}'||^2\} + \mu(\mathrm{tr}(\mathbf{M}'^{\mathrm{H}}\mathbf{M}'\mathbf{R}'_s) - P)\,, \tag{31}$$

where $\mu \in \mathbb{R}$ is a Lagrange multiplier. Similarly to the MF and ZF algorithms, the optimization procedure employed in [26] for the multi-user unicast case can be employed to obtain the solution to the multi-group multicast problem, which is given by

$$\mathbf{M}'_{\mathrm{MMSE}} = \beta_{\mathrm{MMSE}}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H} + \frac{N\sigma_z^2}{P}\mathbf{I}\right)^{-1}\mathbf{H}^{\mathrm{H}}\mathbf{T}^{+}\,, \tag{32}$$

$$\mathbf{D}_{\mathrm{MMSE}} = \beta_{\mathrm{MMSE}}^{-1}\mathbf{I}\,, \tag{33}$$

with

$$\beta_{\mathrm{MMSE}} = \sqrt{\frac{P}{\sigma_s^2\,\mathrm{tr}\Big(\mathbf{H}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H} + \frac{N\sigma_z^2}{P}\mathbf{I}\right)^{-2}\mathbf{H}^{\mathrm{H}}\mathbf{T}^{+}\mathbf{T}^{+,\mathrm{T}}\Big)}}\,. \tag{34}$$

D. SINR balancing algorithm

Different solutions to the SINR Balancing (SB) problem have been proposed in the literature for the multi-user unicast scenario, such as in [21, 23]. In [21], the problem of minimizing the transmit power subject to the condition that the users achieve a certain SINR target, is written as a semidefinite optimization problem, which can be solved through efficient semidefinite programming techniques. In [23], a different methodology for solving this problem, as well as the problem of maximizing the worst-user SINR subject to a transmit power constraint, is proposed. It takes advantage of the uplink/downlink duality [27, 28] and consists of an alternating optimization procedure, which adjusts both the unit-norm beamformers and the power allocation among the streams, converging to the optimal solution after only a few iterations.

In this section, the SINR balancing algorithm for unicast is reviewed and then it is shown that it can be extended to the multi-group multicast case by considering a suboptimal heuristic approach.

Let $\mathbf{U} \in \mathbb{C}^{L\times N}$ denote the unit-norm beamforming matrix, whose n^{th} column $\mathbf{u}_n \in \mathbb{C}^{L}$ is given by

$$\mathbf{u}_n = \frac{\mathbf{m}_n}{||\mathbf{m}_n||}\,, \tag{35}$$

and let $\mathbf{p} \in \mathbb{R}^N$ denote the power allocation vector, whose n^{th} element $p_n \in \mathbb{R}$ is

$$p_n = \sigma_s^2 \, ||\mathbf{m}_n||^2 \, . \tag{36}$$

From (35) and (36), it also follows that

$$\mathbf{m}_n = \sqrt{(p_n/\sigma_s^2)} \, \mathbf{u}_n \, . \tag{37}$$

The multi-user unicast SINR balancing optimization problem can be written as

$$\begin{aligned} \{\mathbf{p}_{\text{SB}}, \mathbf{U}_{\text{SB}}\} &= \underset{\{\mathbf{p},\mathbf{U}\}}{\operatorname{argmax}} \min_n \gamma_n \, , \quad \text{for} \quad n = 1, \ldots, N \, , \\ \text{subject to: } & \mathbf{1}^{\text{T}}\mathbf{p} \leq P, \;\; \mathbf{p} \geq \mathbf{0}, \;\text{ and }\; \mathbf{u}_n^{\text{H}}\mathbf{u}_n = 1 \; \forall n \, . \end{aligned} \tag{38}$$

where the unicast SINR γ_n is given by

$$\gamma_n = \frac{p_n \mathbf{u}_n^{\text{H}} \mathbf{G}_n \mathbf{u}_n}{\sum\limits_{i=1,\, i \neq n}^{N} p_i \mathbf{u}_i^{\text{H}} \mathbf{G}_n \mathbf{u}_i + 1} \, , \tag{39}$$

and $\mathbf{G}_n = (\mathbf{h}_n \mathbf{h}_n^{\text{H}})/\sigma_z^2 \in \mathbb{C}^{L \times L}$ denotes the normalized Gram matrix of the channel.

With the power and beamforming vectors being regarded separately, according to (35) and (36), the optimization problem is separated into two parts: power allocation and unit-norm beamforming determination. These two parts are explained in the following, and then the alternating optimization procedure is described.

The power allocation vector $\mathbf{p}$, given a fixed unit-norm beamforming matrix $\mathbf{U}$, can be determined by employing centralized power control [29]. Let $\mathbf{S} \in \mathbb{R}^{N \times N}$ denote a diagonal matrix corresponding to the signal part of the transmission, and $\mathbf{\Psi} \in \mathbb{R}^{N \times N}$ the interference part. The elements of $\mathbf{S}$ and $\mathbf{\Psi}$ are given by

$$S_{i,j} = \begin{cases} \mathbf{u}_i^{\text{H}} \mathbf{G}_i \mathbf{u}_i, & i = j \\ 0, & i \neq j \end{cases} , \quad \mathit{\Psi}_{i,j} = \begin{cases} 0, & i = j \\ \mathbf{u}_j^{\text{H}} \mathbf{G}_i \mathbf{u}_j, & i \neq j \end{cases} . \tag{40}$$

Assuming that all users achieve the same maximal SINR value γ_{max}, the following equation holds

$$\mathbf{S}\mathbf{p} = \gamma_{\text{max}}(\mathbf{\Psi}\mathbf{p} + \mathbf{1}) \implies \gamma_{\text{max}}^{-1}\mathbf{p} = \mathbf{S}^{-1}\mathbf{\Psi}\mathbf{p} + \mathbf{S}^{-1}\mathbf{1} \, . \tag{41}$$

In order to achieve the maximal SINR, the power vector needs to employ the total available power P, i.e., $\mathbf{1}^{\text{T}}\mathbf{p} = P$. When left-multiplying (41) by $\mathbf{1}^{\text{T}}$ it becomes

$$\gamma_{\text{max}}^{-1} = P^{-1}\mathbf{1}^{\text{T}}\mathbf{S}^{-1}\mathbf{\Psi}\mathbf{p} + P^{-1}\mathbf{1}^{\text{T}}\mathbf{S}^{-1}\mathbf{1} \, . \tag{42}$$

According to [23], an eigensystem can be formed based on (41) and (42):

$$\mathbf{\Upsilon}\mathbf{p}_{\text{ext}} = \gamma_{\text{max}}^{-1}\mathbf{p}_{\text{ext}} \, , \tag{43}$$

where $\mathbf{p}_{\text{ext}} = [\, \mathbf{p}^{\text{T}} \; 1]^{\text{T}} \in \mathbb{R}^{N+1}$ is an extended power vector, and $\mathbf{\Upsilon} \in \mathbb{R}^{(N+1)\times(N+1)}$ is an extended coupling matrix given by

$$\mathbf{\Upsilon} = \left[\begin{array}{cc} \mathbf{S}^{-1}\mathbf{\Psi} & \mathbf{S}^{-1}\mathbf{1} \\ P^{-1}\mathbf{1}^{\text{T}}\mathbf{S}^{-1}\mathbf{\Psi} & P^{-1}\mathbf{1}^{\text{T}}\mathbf{S}^{-1}\mathbf{1} \end{array} \right] . \tag{44}$$

The solution of the eigensystem leads to the optimal power vector, which is given by the first N components of the principal eigenvector of $\mathbf{\Upsilon}$ [23], such that the last entry is equal to one.

Next, the unit-norm beamforming optimization problem is discussed. Given a fixed power allocation, it has been shown in [23] for the unicast case that, due to the uplink/downlink duality, which we assume that also holds in this case, the optimal unit-norm beamformers can be obtained by performing maximization of the uplink SINR of each user independently. The optimization problem is expressed as

$$\begin{aligned} \mathbf{u}_{n,\text{opt}} &= \underset{\mathbf{u}_n}{\operatorname{argmax}} \frac{\mathbf{u}_n^{\text{H}}\mathbf{G}_n\mathbf{u}_n}{\mathbf{u}_n^{\text{H}}\mathbf{Q}_n\mathbf{u}_n} \, , \;\; \text{subject to: } ||\mathbf{u}_n||^2 = 1 \, , \\ \text{with } \mathbf{Q}_n &= \sum_{i=1,\, i \neq n}^{N} (q_i \mathbf{G}_i) + \mathbf{I} \, , \end{aligned} \tag{45}$$

where $\mathbf{q} \in \mathbb{R}^N$ represents the uplink power allocation vector, which may be obtained as the first N components of the principal eigenvector of the extended uplink coupling matrix $\mathbf{\Upsilon}^{(\text{ul})} = \mathbf{\Upsilon}(\mathbf{\Psi}^{\text{T}}) \in \mathbb{R}^{(N+1)\times(N+1)}$. The solution of (45) corresponds to the dominant generalized eigenvector of the pair $(\mathbf{G}_n, \mathbf{Q}_n)$.

The alternating optimization algorithm consists of the alternating execution of the power allocation and unit-norm beamforming procedures, such as described in [23]. The dominant eigenvalue λ_{max} of the power allocation problem monotonically decreases after each iteration, so that the stop criterion is defined based on λ_{max} reaching a certain precision ϵ, i.e., $\lambda_{\text{max}}^{(i-1)} - \lambda_{\text{max}}^{(i)} < \epsilon$, where $(\cdot)^{(i)}$ indicates the i^{th} algorithm iteration. Given an arbitrary initial uplink power vector $\mathbf{q}^{(0)}$, the following steps are repeated until the desired precision is reached:

- Calculate $\mathbf{U}^{(i)}$ given the previous vector $\mathbf{q}^{(i-1)}$,
- Calculate $\mathbf{q}^{(i)}$ given matrix $\mathbf{U}^{(i)}$.

Concluding the review of the SINR balancing algorithm for unicast, the downlink power allocation $\mathbf{p}$ is calculated for the final matrix $\mathbf{U}$. The resulting multi-user unicast beamforming matrix $\mathbf{M}_{\text{SB}} \in \mathbb{C}^{L \times N}$ is given by

$$\mathbf{M}_{\text{SB}} = \mathbf{U} \, \text{diag}(\mathbf{p})^{1/2} \, . \tag{46}$$

The extension of this method to the multi-group multicast case depairs itself with two problems. The first is that, when the SINR calculation in (39) is applied to the multicast case, pessimistic values are achieved, since even multicast users belonging to the same group are assumed to be interferers. The other issue is that the power constraint in (38) assumes that the symbols in $\mathbf{s}$ are uncorrelated. Since in the multicast case there can be repeated entries within $\mathbf{s}$, taking into account (6) and (46), the actual multicast power constraint is given by

$$\begin{aligned} &\text{tr}(\mathbf{M}_{\text{SB}}^{\text{H}}\mathbf{M}_{\text{SB}}\mathbf{T}^{+}\mathbf{T}^{+,\text{T}}) = \\ &\quad \mathbf{1}^{\text{T}}\mathbf{p} + 2\sum_{k=1}^{K} \sum_{\substack{(i,j)\in \mathcal{G}_k \\ i<j}} \sqrt{p_i p_j} \, \text{Re}\,(\mathbf{u}_i^{\text{H}}\mathbf{u}_j) \leq P \, , \end{aligned} \tag{47}$$

where $\mathcal{G}_k$ denotes the set of user indices of multicast group k. By taking only $\mathbf{1}^{\text{T}}\mathbf{p}$ into account within the optimization problem a suboptimal solution is obtained, as the transmit power constraint is usually not fulfilled with equality.

If these issues are taken into account by the optimization problem in (38), it is no longer possible to apply the SB algorithm. For these reasons, the most straightforward extension of SB to multicast, albeit suboptimal, is to directly apply the unicast optimization procedure and afterwards introduce a normalization factor to ensure that the power constraints are satisfied.

The reduced-form beamforming matrix $\mathbf{M}'_{\text{SB}} \in \mathbb{C}^{L\times K}$ is thus given by

$$\mathbf{M}'_{\text{SB}} = \beta\,\mathbf{U}\,\text{diag}(\mathbf{p})^{1/2}\,\mathbf{T}^{+}\,, \tag{48}$$

with $\beta \in \mathbb{R}$ defined as

$$\beta = \sqrt{\frac{P}{\text{tr}(\text{diag}(\mathbf{p})^{1/2}\,\mathbf{U}^{\text{H}}\mathbf{U}\,\text{diag}(\mathbf{p})^{1/2}\,\mathbf{T}^{+}\mathbf{T}^{+,\text{T}})}}\,. \tag{49}$$

V. Multicast aware beamforming algorithms

A. Basic idea of multicast awareness

The main objective of introducing the so-called multicast awareness is to better adjust the beamforming algorithms to the multicast case. The implementation of multicast awareness is algorithm-specific, but in general it relates to how the algorithms deal with the interference among data streams, such as avoiding the unnecessary suppression of interference among users of a same group.

Since the MF algorithm does not take interference into account, it has no multicast aware extension. In the next subsections new multicast aware algorithms based on ZF and MMSE criteria, as well as based on SINR balancing (SB), are proposed and discussed.

B. Multicast aware ZF algorithm

The ZF algorithm presented in Section IV-B corresponds to a direct method of implementing the zero-forcing filter based on channel inversion. Another possible method is to make use of null-space projections [30, 31] in order to eliminate the interference. In [30, 31], a null-space method called Block Diagonalization (BD) is proposed for the MIMO multi-user scenario. The idea of BD is to suppress only the interference among streams of different users, i.e., no energy is spent on mitigating the interference among the streams of a same user. The assumption is that this remaining intra-user interference can be suppressed by implementing receive processing techniques at each multi-antenna user terminal.

The MIMO multi-user scenario is to some extent analogous to the multi-group multicast scenario. For the latter, it is only necessary to suppress the interference among different groups. The users belonging to the same group do not require interference cancellation. Actually, since they expect the same data stream, no further interference suppression is required at the receiver side. In this section, an algorithm based on null-space projections is proposed for the multi-group multicast scenario. This algorithm will be referred to as multicast-aware zero-forcing (MA-ZF).

It is assumed that $\mathbf{M}' \in \mathbb{C}^{L\times K}$ and $\mathbf{m}'_k \in \mathbb{C}^{L}$ denote, respectively, the complete beamforming matrix and the beamforming vector of the k^{th} multicast group.

Let $\mathbf{H}_k \in \mathbb{C}^{g_k\times L}$ and $\tilde{\mathbf{H}}_k \in \mathbb{C}^{(N-g_k)\times L}$ denote, respectively, the channel matrix of all users belonging to group k and all users not belonging to group k. The latter can be written as

$$\tilde{\mathbf{H}}_k = [\,\mathbf{H}_1^{\text{T}}\,,\,\dots\,,\,\mathbf{H}_{k-1}^{\text{T}}\,,\,\mathbf{H}_{k+1}^{\text{T}}\,,\,\dots\,,\,\mathbf{H}_K^{\text{T}}\,]^{\text{T}}\,. \tag{50}$$

The channel $\tilde{\mathbf{H}}_k$ can be decomposed using Singular Value Decomposition (SVD) as follows:

$$\tilde{\mathbf{H}}_k = \tilde{\mathbf{U}}_k\tilde{\mathbf{S}}_k[\,\tilde{\mathbf{V}}_k^{(1)}\,,\,\tilde{\mathbf{V}}_k^{(0)}\,]^{\text{H}}\,, \tag{51}$$

where $\tilde{\mathbf{U}}_k \in \mathbb{C}^{(N-g_k)\times(N-g_k)}$ is a unitary matrix, $\tilde{\mathbf{S}}_k \in \mathbb{R}^{(N-g_k)\times L}$ is a diagonal matrix, $\tilde{\mathbf{V}}_k^{(1)} \in \mathbb{C}^{L\times\tilde{r}_k}$ and $\tilde{\mathbf{V}}_k^{(0)} \in \mathbb{C}^{L\times(L-\tilde{r}_k)}$ contain the right singular vectors of $\tilde{\mathbf{H}}_k$, with $\tilde{r}_k$ denoting the rank of matrix $\tilde{\mathbf{H}}_k$. Matrix $\tilde{\mathbf{V}}_k^{(0)}$ constitutes an orthogonal basis for the null space of $\tilde{\mathbf{H}}_k$. Due to this property, $\tilde{\mathbf{V}}_k^{(0)}$ can be used for specifying a beamforming vector that cancels the interference from the other groups of users. If $L-\tilde{r}_k = 1$, then $\tilde{\mathbf{V}}_k^{(0)}$ can be used directly as the beamforming vector, otherwise, if $L-\tilde{r}_k > 1$, then there are some degrees of freedom for determining a suitable beamforming vector. Note that $\tilde{r}_k = N - g_k$, when assuming that matrix $\mathbf{H}$ has full row rank.

The multiplication of the channel matrix $\mathbf{H}_j$ by $\tilde{\mathbf{V}}_k^{(0)}$, for all $j \neq k$, results in a matrix $\mathbf{0} \in \mathbb{R}^{g_j\times(L-\tilde{r}_k)}$ of zeros. The product $\mathbf{H}_k\tilde{\mathbf{V}}_k^{(0)}$, on the other hand, can be seen as an equivalent channel matrix $\mathbf{H}_k^{(\text{eq})} \in \mathbb{C}^{g_k\times(L-\tilde{r}_k)}$ after the null space projection. When multiplying $\mathbf{H}_k$ by $\tilde{\mathbf{V}}_k^{(0)}$ it is assured that the interference from the data streams of other users will be totally suppressed. For this reason, each multicast group can be processed individually, i.e., any single-group beamforming algorithm can be applied to the equivalent channel $\mathbf{H}_k^{(\text{eq})}$ of each group k. Let $\mathbf{m}_k^{(\text{eq})} \in \mathbb{C}^{(L-\tilde{r}_k)}$ denote the equivalent beamforming vector obtained after applying single-group beamforming to $\mathbf{H}_k^{(\text{eq})}$. For example, $\mathbf{m}_k^{(\text{eq})}$ can be set as the principal eigenvector of $\mathbf{H}_k^{(\text{eq}),\text{H}}\mathbf{H}_k^{(\text{eq})}$, if maximizing the average user SNR within the group is desired [17], or $\mathbf{m}_k^{(\text{eq})}$ can be obtained from the SDR methodology in [7], for maximizing the minimum user SNR within the group. The resulting normalized beamforming vector for group k is then given by

$$\mathbf{m}'_k = \tilde{\mathbf{V}}_k^{(0)}\mathbf{m}_k^{(\text{eq})}/||\tilde{\mathbf{V}}_k^{(0)}\mathbf{m}_k^{(\text{eq})}||\,, \tag{52}$$

and the beamforming matrix $\mathbf{M}'$ is

$$\mathbf{M}' = \left[\,\frac{\tilde{\mathbf{V}}_1^{(0)}\mathbf{m}_1^{(\text{eq})}}{||\tilde{\mathbf{V}}_1^{(0)}\mathbf{m}_1^{(\text{eq})}||}\,,\,\dots\,,\,\frac{\tilde{\mathbf{V}}_K^{(0)}\mathbf{m}_K^{(\text{eq})}}{||\tilde{\mathbf{V}}_K^{(0)}\mathbf{m}_K^{(\text{eq})}||}\,\right]. \tag{53}$$

Another aspect concerning the MA-ZF algorithm is that, differently from the ZF filter, the received power is not balanced among the users. In the case of ZF, all users receive the same power, which is due to the channel inversion step. The approach based on null-space projections, however, does not make any such guarantees regarding the received power. For this reason, it is necessary to perform power loading on matrix $\mathbf{M}'$. In [31], the power loading is done according to the waterfilling algorithm [32], which aims at maximizing the sum throughput. A more fair power loading, which balances the received power among the users, is considered here instead.

Let $\mathbf{r_k} = \mathbf{H}_k\mathbf{m}'_k \in \mathbb{C}^{g_k}$ denote the vector of complex coefficients of the received signal part of the users of multicast group k and $\mathbf{r} = [\,\mathbf{r}_1^{\mathbf{T}}, \ldots, \mathbf{r}_K^{\mathbf{T}}\,]^{\mathbf{T}} \in \mathbb{C}^N$ the vector relative to all users, whose n^{th} element is r_n. The optimization problem for determining the diagonal power loading matrix $\mathbf{\Gamma} \in \mathbb{R}^{K\times K}$ that maximizes the minimum SNR over all users can be expressed as

$$\begin{aligned}&\mathbf{\Gamma}_{\text{MA-ZF}} = \underset{\mathbf{\Gamma}}{\operatorname{argmax}} \min_{n\,\in\,\{1,\ldots,N\}} |r_n \Gamma_{b_n,b_n}|^2\,,\\ &\text{subject to: } \operatorname{tr}(\mathbf{\Gamma}) = P \text{ and } \Gamma_{i,j} = 0,\ \forall\, i \neq j\,,\end{aligned} \tag{54}$$

reminding that b_n indicates the group to which user n belongs. Since the power loading is applied group-wise, the factor Γ_k equally scales all users within group k, so that it suffices to take into account the worst user within the group. The solution of the problem is given by

$$\Gamma_{k,k} = P \frac{\min(\operatorname{diag}(\mathbf{r}_k \mathbf{r}_k^{\mathrm{H}}))^{-1/2}}{\sum\limits_{i=1}^{K} \min(\operatorname{diag}(\mathbf{r}_i \mathbf{r}_i^{\mathrm{H}}))^{-1/2}}\,, \tag{55}$$

which corresponds to normalizing each group by its worst user and then equally scaling all groups to satisfy the transmit power constraint. The solution for the MA-ZF algorithm can finally be summarized as

$$\mathbf{M}'_{\text{MA-ZF}} = \mathbf{M}'\mathbf{\Gamma}\,, \tag{56}$$

where $\mathbf{M}'$ and $\mathbf{\Gamma}$ are defined, respectively, in (53) and (55).

C. Multicast aware regularized ZF algorithm

In this section, the multicast aware regularized ZF (MA-RZF) algorithm is proposed. It is a heuristical algorithm based on the same method of null-space projections described in Section V-B. The main motivation of the algorithm is to improve the conditioning of the channel and consequently spend less energy on nulling out interference.

The difference with regard to MA-ZF is that, instead of making the projections with regard to the original channel, an equivalent regularized channel is taken into account. The null space projections totally suppress the inter-group interference of the equivalent channel. However, since it is not equal to the original channel there appears a residual inter-group interference.

Let $\mathbf{H}^{(\mathrm{R})} \in \mathbb{C}^{N\times N}$ denote the regularized channel. It is defined as

$$\mathbf{H}^{(\mathrm{R})} = \mathbf{H}\mathbf{H}^{\mathrm{H}} + \frac{N\sigma_z^2}{P}\mathbf{I}\,, \tag{57}$$

where the same regularization factor as that of the MMSE algorithm [26] is considered. This regularization factor is not necessarily optimal for MA-RZF, given the additional multicast beamforming functionality. The determination of the optimal factor possibly depends on the chosen multicast beamforming algorithm, such that a derivation similar to that of MMSE might not be possible. We therefore leave its determination as a topic for further investigation.

Proceeding with the algorithm, matrix $\tilde{\mathbf{H}}_k \in \mathbb{C}^{(N-g_k)\times N}$ and its SVD are given by

$$\tilde{\mathbf{H}}_k = [\,\mathbf{H}_1^{(\mathrm{R}),\mathrm{T}}, \ldots, \mathbf{H}_{k-1}^{(\mathrm{R}),\mathrm{T}}, \mathbf{H}_{k+1}^{(\mathrm{R}),\mathrm{T}}, \ldots, \mathbf{H}_K^{(\mathrm{R}),\mathrm{T}}\,]^{\mathrm{T}}, \tag{58a}$$

$$\tilde{\mathbf{H}}_k = \tilde{\mathbf{U}}_k \tilde{\mathbf{S}}_k [\,\tilde{\mathbf{V}}_k^{(1)}, \tilde{\mathbf{V}}_k^{(0)}\,]^{\mathrm{H}}\,, \tag{58b}$$

where $\mathbf{H}_k^{(\mathrm{R})} \in \mathbb{C}^{g_k\times N}$ results from extracting the corresponding columns of $\mathbf{H}^{(\mathrm{R})}$, $\tilde{\mathbf{U}}_k \in \mathbb{C}^{(N-g_k)\times(N-g_k)}$, $\tilde{\mathbf{S}}_k \in \mathbb{R}^{(N-g_k)\times N}$, $\tilde{\mathbf{V}}_k^{(1)} \in \mathbb{C}^{N\times \tilde{r}_k}$, $\tilde{\mathbf{V}}_k^{(0)} \in \mathbb{C}^{N\times(N-\tilde{r}_k)}$, and $\tilde{r}_k$ denotes the rank of matrix $\tilde{\mathbf{H}}_k$.

The multicast beamforming optimization, which can be implemented according to any single-group beamforming algorithm, is done for each group considering the equivalent channel after the null-space projection $\mathbf{H}_k^{(\mathrm{eq})} = \mathbf{H}_k^{(\mathrm{R})}\tilde{\mathbf{V}}_k^{(0)} \in \mathbb{C}^{g_k\times(N-\tilde{r}_k)}$, and resulting in vector $\mathbf{m}_k^{(\mathrm{eq})} \in \mathbb{C}^{(N-\tilde{r}_k)}$.

When considering the same power loading strategy described in the previous section, which maximizes the minimum SNR of all users, the beamforming matrix $\mathbf{M}'_{\text{MA-RZF}} \in \mathbb{C}^{L\times K}$ of the MA-RZF algorithm can be written as

$$\mathbf{M}'_{\text{MA-RZF}} = \mathbf{M}'\mathbf{\Gamma} \tag{59}$$

with

$$\mathbf{M}' = \left[\frac{\mathbf{H}^{\mathrm{H}}\tilde{\mathbf{V}}_1^{(0)}\mathbf{m}_1^{(\mathrm{eq})}}{||\mathbf{H}^{\mathrm{H}}\tilde{\mathbf{V}}_1^{(0)}\mathbf{m}_1^{(\mathrm{eq})}||}, \ldots, \frac{\mathbf{H}^{\mathrm{H}}\tilde{\mathbf{V}}_K^{(0)}\mathbf{m}_K^{(\mathrm{eq})}}{||\mathbf{H}^{\mathrm{H}}\tilde{\mathbf{V}}_K^{(0)}\mathbf{m}_K^{(\mathrm{eq})}||}\right], \tag{60a}$$

$$\Gamma_{k,k} = P \frac{\min(\operatorname{diag}(\mathbf{r}_k \mathbf{r}_k^{\mathrm{H}}))^{-1/2}}{\sum\limits_{i=1}^{K} \min(\operatorname{diag}(\mathbf{r}_i \mathbf{r}_i^{\mathrm{H}}))^{-1/2}}\,, \tag{60b}$$

where $\mathbf{M}' \in \mathbb{C}^{N\times K}$, $\mathbf{\Gamma} \in \mathbb{R}^{K\times K}$, and $\mathbf{r_k} = \mathbf{H}_k\mathbf{m}'_k \in \mathbb{C}^{g_k}$.

D. Multicast aware SINR balancing algorithm

In this section, the SINR balancing (SB) algorithm originally derived for the unicast case in [23] is enhanced with the purpose of improving the performance of the multicast users. The proposed multicast-aware SB (MA-SB) algorithm is based on alternating optimization and it aims at heuristically solving the problem of maximizing the minimum SINR with low computational complexity. As a comparison, the algorithms in [12, 13] provide closer-to-optimum results, but at the cost of a higher complexity introduced by the required bisection method for iteratively adjusting the SINR target until the maximum worst-user SINR is achieved.

The optimization problem of determining the optimal reduced-form beamforming matrix $\mathbf{M}'_{\text{MA-SB}} \in \mathbb{C}^{L\times K}$ that maximizes the minimum SINR is given by

$$\begin{aligned}&\mathbf{M}'_{\mathrm{MA-SB}} = \underset{\mathbf{M}'}{\operatorname{argmax}} \min_{n\,\in\,\{1,\ldots,N\}} \gamma_n\,,\\ &\text{subject to: } \sigma_s^2 \operatorname{tr}(\mathbf{M}'^{\mathrm{H}}\mathbf{M}') \leq P\,,\end{aligned} \tag{61}$$

with

$$\gamma_n = \frac{\sigma_s^2\, \mathbf{m}'^{\mathrm{H}}_{b_n}\mathbf{G}_n\mathbf{m}'_{b_n}}{\sum\limits_{k=1,\,k\neq b_n}^{K} \sigma_s^2\, \mathbf{m}'^{\mathrm{H}}_k\mathbf{G}_n\mathbf{m}'_k + 1}\,. \tag{62}$$

Even though the optimization problem is expressed in the reduced form, the proposed MA-SB algorithm is derived based on the complete form of the multi-group multicast scenario. This is necessary in order to make it possible to find a solution based on alternating optimization that is similar to the SB algorithm of the previous section. For this reason, the same notation is considered for the power vector $\mathbf{p} \in \mathbb{R}^N$ and unit-norm beamforming vectors $\mathbf{u}_n \in \mathbb{C}^L$ as defined in (36) and (35), respectively.

In order to express the set of equations that determines the downlink power assignment given a fixed matrix $\mathbf{U}$, it is initially assumed that all users are unicast and that they achieve the same maximum SINR γ_{max}. Let $\mathbf{S} \in \mathbb{C}^{N\times N}$ denote a diagonal matrix corresponding to the signal part of the transmission, and $\mathbf{\Psi} \in \mathbb{C}^{N\times N}$ the interference part. For the multi-user unicast case, the elements of $\mathbf{S}$ and $\mathbf{\Psi}$ are given in (40), and the power vector is determined based on the solution of the eigensystem expressed in (43).

For the multi-group multicast case, however, this procedure cannot be directly applied in the reduced form, since the power allocation would have to be done for each group, and not for each user. This results in a number of equations larger than the number of variables, i.e., there are still N SINR values to balance but only $K < N$ power elements to adjust. In this case it is not always possible to guarantee that all users achieve the same SINR and the problem cannot be solved as an eigenvalue problem.

In order to simplify the procedure and allow the multi-group multicast power allocation to be also expressed as an eigensystem, it is here assumed that the power allocation can be done user-wise, i.e., vector $\mathbf{p}$ contains N elements, and the elements of matrices $\mathbf{S}$ and $\mathbf{\Psi}$ are now defined as:

$$S_{i,j} = \begin{cases} \left(\sum\limits_{n=1,\, b_n=b_i}^{N} \mathbf{u}_n^{\mathrm{H}}\right)\mathbf{G}_i\left(\sum\limits_{n=1,\, b_n=b_i}^{N} \mathbf{u}_n\right), & i=j \\ 0, & i\neq j \end{cases}, \tag{63a}$$

$$\varPsi_{i,j} = \begin{cases} 0, & b_i = b_j \\ \mathbf{u}_j^{\mathrm{H}}\mathbf{G}_i\mathbf{u}_j, & b_i \neq b_j \end{cases}. \tag{63b}$$

Matrices $\mathbf{S}$ and $\mathbf{\Psi}$ are chosen so that when they are substituted in the SINR expression in (62), the actual SINR perceived by the users is approximated, while still allowing the system to be solved as an eigenvalue problem. The solution corresponds to the principal eigenvector of the extended coupling matrix, defined in (44), but considering the new $\mathbf{S}$ and $\mathbf{\Psi}$ matrices of (63). Assuming that

$$\mathbf{m}'_{b_n} = \sum_{i=1,\, b_i=b_n}^{N} \mathbf{m}_i = \sum_{i=1,\, b_i=b_n}^{N} \sqrt{p_i/\sigma_s^2}\,\mathbf{u}_i\,, \tag{64}$$

the actual and approximate complete-form SINR expressions are, respectively:

$$\gamma_n = \frac{\left(\sum\limits_{i=1,\, b_i=b_n}^{N} \sqrt{p_i}\mathbf{u}_i^{\mathrm{H}}\right)\mathbf{G}_n\left(\sum\limits_{i=1,\, b_i=b_n}^{N} \sqrt{p_i}\mathbf{u}_i\right)}{\sum\limits_{k=1,\, k\neq b_n}^{K}\left(\sum\limits_{i=1,\, b_i=k}^{N} \sqrt{p_i}\mathbf{u}_i^{\mathrm{H}}\right)\mathbf{G}_n\left(\sum\limits_{i=1,\, b_i=k}^{N} \sqrt{p_i}\mathbf{u}_i\right)+1}, \tag{65a}$$

$$\gamma_n \simeq \frac{p_n\left(\sum\limits_{i=1,\, b_i=b_n}^{N} \mathbf{u}_i^{\mathrm{H}}\right)\mathbf{G}_n\left(\sum\limits_{i=1,\, b_i=b_n}^{N} \mathbf{u}_i\right)}{\sum\limits_{k=1,\, k\neq b_n}^{K}\left(\sum\limits_{i=1,\, b_i=k}^{N} p_i\mathbf{u}_i^{\mathrm{H}}\mathbf{G}_n\mathbf{u}_i\right)+1}. \tag{65b}$$

The approximation of the signal part in (65b) corresponds to considering the power of only the n^{th} user and disregarding the power of the other users belonging to the same group. With regard to the interference part, it is an approximation that considers all interferers as unicast users, instead of taking into account the equivalent group beamforming vectors. This approximation is needed since when considering the actual SINR it is no longer possible to apply the SINR balancing algorithm, as the actual power allocated to a user depends on the power (in complete form) of other users of the same group (see (61) and (64)). The approximation consists essentially of decoupling these transmit powers, such that the SB algorithm can be applied.

Regarding the determination of the unit-norm beamformers, a similar approach to that of [23], which has been presented in the previous section, is considered. The optimization problem for the unit-norm beamformer of user n is written as

$$\mathbf{u}_{n,\text{opt}} = \underset{\mathbf{u}_n}{\operatorname{argmax}} \frac{\mathbf{u}_n^{\mathrm{H}}\mathbf{G}_n\mathbf{u}_n}{\mathbf{u}_n^{\mathrm{H}}\mathbf{Q}_n\mathbf{u}_n}\,, \quad \text{subject to: } ||\mathbf{u}_n||^2 = 1\,,$$
$$\text{with } \mathbf{Q}_n = \sum_{i=1,\, b_i\neq b_n}^{N} q_i\mathbf{G}_i + \mathbf{I}\,, \tag{66}$$

where $\mathbf{q} \in \mathbb{R}^N$ represents the uplink power allocation vector, which may be determined as the principal eigenvector of the previously defined extended uplink coupling matrix, with the $\mathbf{S}$ and $\mathbf{\Psi}$ matrices given in (63a) and (63b), respectively. The solution of (66) corresponds to the dominant generalized eigenvector of the pair $(\mathbf{G}_n, \mathbf{Q}_n)$. The difference with regard to the multi-user unicast case lies in the definition of matrix $\mathbf{Q}_n$, which has been modified in order to avoid interference within a same multicast group.

The MA-SB algorithm consists of the alternating optimization of the power allocation and unit-norm beamforming, such as described for the SB algorithm. At the end, the downlink power allocation $\mathbf{p}$ is calculated for the final matrix $\mathbf{U}$. The resulting complete-form modulation matrix is given by

$$\mathbf{M}_{\text{MA-SB}} = \mathbf{U}\,\mathrm{diag}(\mathbf{p})^{1/2}\,. \tag{67}$$

Due to the SINR approximation considered for the power allocation procedure, the SINR balancing is not achieved for all users. In fact, it is perceived that the SINR of the unicast users reaches a certain balanced level, and that the average SINR of the users of the multicast group also approaches this level, but not each individual multicast user.

In order to improve the worst-user performance, a power redistribution among the multicast and unicast users is proposed here. This procedure is a further refinement of the algorithm, and is performed only a single time after the iterative algorithm has stopped. Let $\mathbf{p}' \in \mathbb{R}^K$ represent the group power allocation vector and $\mathbf{u}'_k \in \mathbb{C}^L$ the unit-norm beamforming vector of group k, such that $p'_k = ||\mathbf{m}'_k||^2$, $\mathbf{u}'_k = \mathbf{m}'_k/||\mathbf{m}'_k||$, and $\mathbf{U}' = [\mathbf{u}'_1, \ldots, \mathbf{u}'_K] \in \mathbb{C}^{L\times K}$. The users with lowest SINR are selected to represent each group, such that $\mathbf{G}'_k = \mathbf{G}_{n\,|\,\gamma_n=\min\boldsymbol{\gamma}_k}$, where the vector $\boldsymbol{\gamma}_k \in \mathbb{R}^{g_k}$ contains the actual SINR of the users belonging to group k.

The unit-norm beamforming vectors $\mathbf{u}'_k$ calculated at the last iteration of the alternating optimization are maintained, and the power vector $\mathbf{p}'$ is recalculated by solving the system:

$$\begin{cases} \gamma_{\text{max}}^{-1}\,\mathbf{p}' = \mathbf{S}'^{-1}\mathbf{\Psi}'\mathbf{p}' + \mathbf{S}'^{-1}\mathbf{1} \\ \mathbf{1}^{\mathrm{T}}\mathbf{p}' = P \end{cases}, \tag{68}$$

with the elements of $\mathbf{S}' \in \mathbb{R}^{K\times K}$ and $\mathbf{\Psi}' \in \mathbb{R}^{K\times K}$ given by

$$S'_{i,i} = \begin{cases} \mathbf{u}_i'^{\mathrm{H}}\mathbf{G}_i'\mathbf{u}_i', & i = j \\ 0, & i \neq j \end{cases}, \quad \Psi'_{i,j} = \begin{cases} 0, & i = j \\ \mathbf{u}_j'^{\mathrm{H}}\mathbf{G}_i'\mathbf{u}_j', & i \neq j \end{cases}. \quad (69)$$

The solution corresponds to the principal eigenvector of the extended coupling matrix, defined in (44), but considering the new $\mathbf{S}'$ and $\mathbf{\Psi}'$ matrices. The obtained power re-allocation vector is denoted $\mathbf{p}'_{\text{PR}}$. It is applied to the unit-norm beamforming, without the need of further power normalization, such that the reduced-form modulation matrix is given by

$$\mathbf{M}'_{\text{MA-SB}} = \mathbf{U}' \,\mathrm{diag}(\mathbf{p}'_{\text{PR}})^{1/2} \,. \quad (70)$$

VI. Performance and complexity analysis

A. Analysis assumptions

The considered scenario consists of a single cell equipped with an L-element uniform linear antenna array and N single antenna mobile terminals, which belong to one of K multicast groups. The distribution of users among groups is characterized by the vectors $\mathbf{b} \in \mathbb{Z}^N$ and $\mathbf{g} \in \mathbb{Z}^K$ described in Section III. The group configurations considered by the analysis are summarized in Table I. The configurations specified by rows 1, 2, and 3, are further referred to as C_1, C_2, and C_3, respectively.

TABLE I
Group configurations of the performance analysis.

Config.	$\{L, N, K\}$	$\mathbf{b}$	$\mathbf{g}$
C_1	$\{4, 4, 3\}$	$[1, 2, 3, 3]^{\mathrm{T}}$	$[1, 1, 2]^{\mathrm{T}}$
C_2	$\{6, 6, 4\}$	$[1, 2, 3, 3, 4, 4]^{\mathrm{T}}$	$[1, 1, 2, 2]^{\mathrm{T}}$
C_3	$\{6, 6, 3\}$	$[1, 2, 2, 3, 3, 3]^{\mathrm{T}}$	$[1, 2, 3]^{\mathrm{T}}$

The cell area is assumed to be hexagonal and the base station is located at the cell corner, representing a sector cell. The radio link between base station and mobile stations takes into account a fast fading model [33, 24] which regards both Line-Of-Sight (LOS) and Non-Line-Of-Sight (NLOS) components and can be written as

$$\mathbf{H} = \sqrt{\kappa/(1+\kappa)}\,\overline{\mathbf{H}} + \sqrt{1/(1+\kappa)}\,\check{\mathbf{H}}, \quad (71)$$

where $\kappa \in \mathbb{R}$ is the Rician factor which determines the ratio of deterministic-to-scattered power, $\check{\mathbf{H}} \in \mathbb{C}^{N\times L}$ is composed of zero mean circularly symmetric complex Gaussian random variables with unit variance, and $\overline{\mathbf{H}} \in \mathbb{C}^{N\times L}$ models the LOS component, which has each row $\overline{\mathbf{h}}_n^{\mathrm{T}}$ given by

$$\overline{\mathbf{h}}_n^{\mathrm{T}} = [1, e^{j2\pi\delta\cos(\theta_n)}, \ldots, e^{j2\pi\delta(L-1)\cos(\theta_n)}], \quad (72)$$

where δ is the antenna spacing in wavelengths and θ_n is the angular direction of user n, which is assumed to be uniformly distributed within $[0,\, 2\pi/3]$ (base station at the corner).

The users are assumed to be at a same distance from the base station. This assumption is motivated by the fact that several other works found in the literature on multicast beamforming, such as [34, 7, 9], also disregard the path-loss in their evaluations. Additionally, preliminary simulation results indicate that the path-loss does not have a large impact on the relative performance of the algorithms.

The results are presented in terms of the uncoded Bit Error Rate (BER), which is averaged over all users and across several channel realizations, ranging from 10^4 to 10^6, depending on the SNR levels. The simulations take into account both QPSK and 16-QAM modulation schemes, and the constellation is normalized such that the average symbol power is $\sigma_s^2 = 1$. The total transmit power is equal to the summed power of all different symbols. Since a different symbol is transmitted to each multicast group, the total transmit power is given by the power of each symbol multiplied by the number K of multicast groups, i.e., $P = K\sigma_s^2 = K$.

Regarding the implementation of the algorithms, the following assumptions are taken into account:

- Regarding the Multicast-Aware algorithms, their single-group beamforming part must be specified. The SDR approach of [7] has been chosen, since for small group sizes it almost always converges to the optimal solution.
- A total of 5 iterations is assumed for the alternating optimization procedure of the SB and MA-SB algorithms. This number was determined by simulations and it was found to be enough for achieving in most cases a good convergence of the SINR balancing effect.
- The Bisec-SDR algorithm runs until a precision of $|P_{\text{req}} - P| \leq 10^{-3}$ is reached, and the solution to the power minimization problem is obtained through the SDR approach of [13].

B. Performance analysis

In this section, the performance of the linear multi-group multicast beamforming algorithms is analyzed. Besides the previously described algorithms – MF, ZF, MMSE, SB, MA-ZF, MA-RZF, and MA-SB – the bisection method based on SDR (Bisec-SDR) [12, 13] is analyzed as well.

The BER performance is shown in Figs. 2 and 3 for the QPSK and 16-QAM modulation schemes, respectively. The user configuration C_1 and an NLOS scenario are assumed. 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.

From Fig. 2 it can be seen that the MF is the algorithm presenting the worst performance by far, which is due to the fact that it does not implement any interference mitigation mechanism. The MF has a high error rate – above 10% – and not even high E_s/N_0 values are capable of improving its error floor. The ZF algorithm presents better performance than MF, as expected, since the channel inversion totally mitigates the interference among users. The MA-ZF algorithm, which is an enhanced version of ZF for the multi-group multicast scenario, clearly outperforms ZF. The subsequent ordering of the algorithms, in terms of their increasing performance, is given by: SB, MMSE, MA-RZF, MA-SB, and Bisec-SDR. Some explanations are given in the following.

The Multicast-Aware (MA) algorithms present significant performance gains with regard to their respective non-MA counterparts, which is due to the implemented multicast-aware enhancements. The most noticeable gain, of approximately 9dB, is the one achieved by MA-SB with regard to SB. When comparing the non-MA algorithms, it is seen that their order

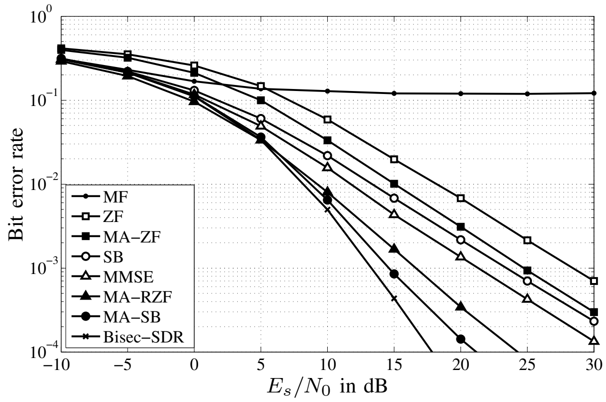

Fig. 2. BER performance considering QPSK, NLOS, C_1, cf. Table I.

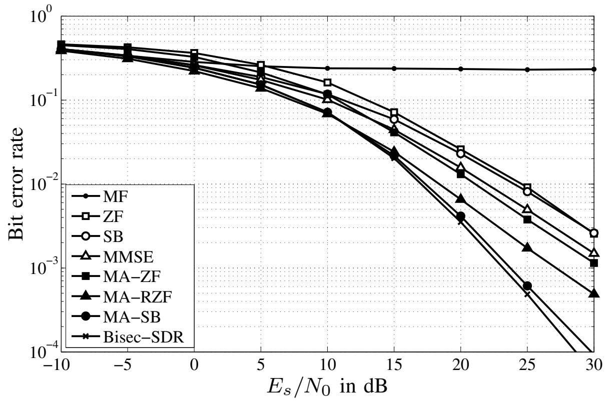

Fig. 3. BER performance considering 16-QAM, NLOS, C_1, cf. Table I.

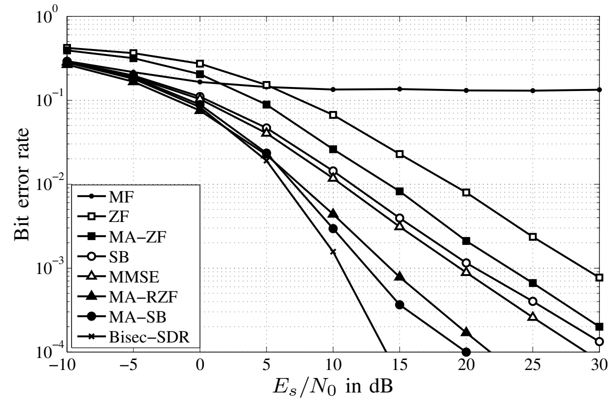

Fig. 4. BER performance considering QPSK, NLOS, C_2, cf. Table I.

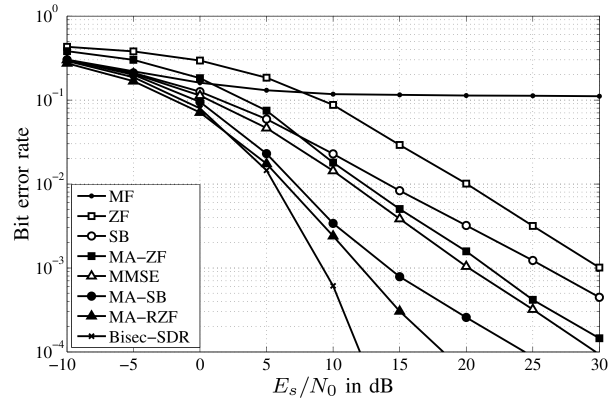

Fig. 5. BER performance considering QPSK, NLOS, C_3, cf. Table I.

of increasing performance is given by $\{\text{ZF} \rightarrow \text{SB} \rightarrow \text{MMSE}\}$. For the MA algorithms, the order is given by $\{\text{MA-ZF} \rightarrow \text{MA-RZF} \rightarrow \text{MA-SB}\}$. The advantage of MMSE over ZF, as well as the advantage of MA-RZF over MA-ZF, was expected and it is mainly due to the introduction of the regularization factor, which avoids the inversion of ill-conditioned matrices. With regard to the SB algorithm, if only unicast users were taken into account, then SB would achieve the best performance. For the multi-group multicast case, however, it turns out being an inadequate strategy, since its optimization is based on an SINR calculation that assumes that all users interfere with each other. The MA-SB algorithm provides, in general, a better approximation to the real SINR, thus approaching the optimal case and outperforming the other linear MA algorithms. The Bisec-SDR algorithm presents the best performance, but at the cost of a much higher complexity, as it will be discussed later in the complexity analysis section.

When changing the modulation scheme from QPSK to 16-QAM, the results are shown in Fig. 3. Besides the expected performance losses due to the higher order modulation, it can be seen that the MA-SB algorithm gets closer to the Bisec-SDR, with the difference between them dropping to less than 1dB. Furthermore, the relative performance among the MA algorithms and among the non-MA algorithms is still the same as in the previous case. What can be perceived is that MA-ZF outperforms both MMSE and SB for high E_s/N_0 values. This tendency could already be seen in Fig. 2 with QPSK, but in the case of 16-QAM it happens much sooner.

The results for configurations C_2 and C_3 are shown in Figs. 4 and 5, respectively, for QPSK modulation. When comparing the absolute results displayed in both these figures and in Fig. 2, it can be seen that, when considering the MA algorithms, C_3 presents better results than C_2, which has better results than C_1. The reason for this behavior lies in the number of available degrees of freedom of the antenna array for each configuration. This measure can be expressed as the ratio between the number of transmit antennas and the number of multicast groups, i.e., L/K. The calculation of this measure for each configuration leads to: $L_{C_3}/K_{C_3} > L_{C_2}/K_{C_2} > L_{C_1}/K_{C_1}$, which is in accordance with the achieved results. Another reason for the performance improvement of the MA algorithms when going from C_1 to C_3 is due to the increased number of multicast users, which leads to a more significant impact of the multicast enhancements on the results. Note that the MA-SB algorithm is an exception, which is discussed in the following.

The relative performance among the algorithms shown in Fig. 4 for configuration C_2 is similar to that obtained for C_1 in Fig. 2. The MA-SB algorithm has a performance close to that of Bisec-SDR, and MA-RZF is the third best algorithm. However, in Fig. 5, which depicts configuration C_3, it is seen that the performance of MA-SB becomes worse, being even surpassed by that of the MA-RZF algorithm. This occurs due to the fact that the MA-SB algorithm is based on an approximate SINR, and the accuracy of this approximation increases with the increasing number K of multicast groups. The closer K gets to the number N of users, the closer the

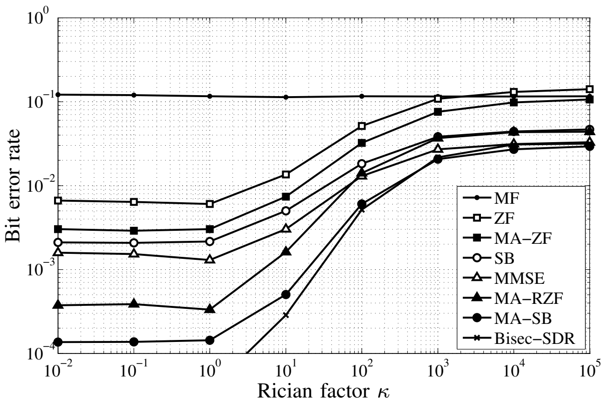

Fig. 6. Impact of Rician factor κ on the BER performance, considering QPSK, $E_s/N_0 = 20$dB, C_1, cf. Table I.

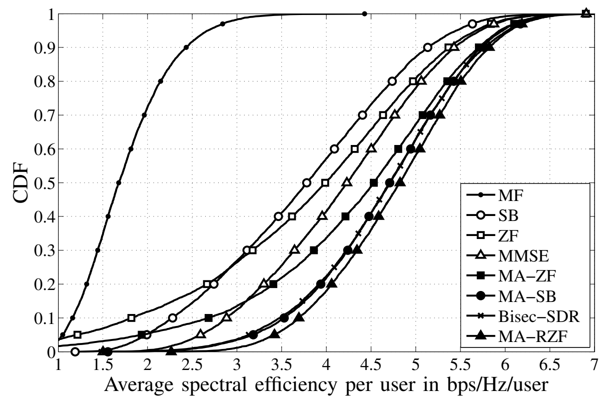

Fig. 7. Throughput performance of the algorithms considering Gaussian signaling, NLOS, $E_s/N_0 = 15$dB, C_1, cf. Table I.

SINR approximation gets to the actual SINR. The other way around, when K is reduced, the SINR approximation becomes more inaccurate, thus resulting in worse performance results. For example, $K/N = 0.75$ for C_1 and $K/N = 0.5$ for C_3, the latter presenting worse MA-SB results. The advantage of MA-RZF in this scenario is therefore mainly due to the degradation of MA-SB. A similar behavior is verified for the SB algorithm, which also takes into account an SINR that coincides with the real value only for the unicast case, i.e., when $K = N$.

In order to analyze the impact of the channel correlation on the performance of the algorithms, the Rician factor κ of (71) is gradually varied between NLOS ($\kappa \to 0$) and LOS ($\kappa \to \infty$) scenarios, given a fixed E_s/N_0. Fig. 6 shows the results when considering configuration C_1, QPSK modulation, and $E_s/N_0 = 20$dB. It can be seen that, with increasing κ, the BER of all algorithms increases. Up to $\kappa = 1$ the impact is not really relevant, but then it starts to significantly degrade the BER, leading to exceedingly high error rates as the pure LOS scenario is approached. Differently from the single-group case, for which the presence of LOS represented an improvement in terms of BER, the opposite behavior is observed for the multi-group scenario. Due to the increased channel correlation, it becomes more difficult to suppress the inter-group interference, thus resulting in a poor performance. Regarding the relative performance among the algorithms, it is similar to that of Fig. 2, but with the following two exceptions for large κ values: the ZF is outperformed by MF, since the channel matrix is mainly dominated by the LOS component, which leads to ill-conditioned matrices in the cases when the users have a small angular separation, and the MA-RZF gets worse than MMSE, due to the inefficiency of applying single-group beamforming on an equivalent regularized LOS channel. For $\kappa \geq 100$ the MA-SB algorithm achieves practically the same performance as Bisec-SDR.

The throughput performance of the algorithms is shown in Fig. 7. The results assume Gaussian signaling, an NLOS scenario, a fixed E_s/N_0 of 15dB and configuration C_1. The average spectral efficiency per user is defined as

$$S = \frac{1}{NF} \sum_{f=1}^{F} \sum_{n=1}^{N} \log_2(1 + \gamma_{n,f}) \,, \tag{73}$$

where F denotes the total number of channel realizations and $\gamma_{n,f}$ the SINR of user n at realization f. It can be seen that the MA algorithms present an overall better performance than the non-MA ones. In particular, the MA-SB and Bisec-SDR algorithms do not achieve the highest rates, which is not surprising, since they explicitly aim at maximizing the minimum SINR and not the average throughput. The MA-RZF algorithm stands out as the best choice with regard to this metric, and it could still be further improved if a power loading adequate to the throughput maximization purpose, e.g., waterfilling, were to be considered.

C. Complexity analysis

In this section, the complexity order of the algorithms is analyzed and compared. The complexity of an algorithm is here measured in terms of the required number of complex multiplications. Divisions and square roots have the same complexity as a multiplication, when they are efficiently implemented using Newton's method [16], and therefore are counted as such. Additions and subtractions are not taken into account and it is assumed that the algorithms are implemented as efficiently as possible. Repeated operations do not increase the complexity, i.e., when the same computation is employed at several points within the algorithm, its computational cost is computed only once, since its result can be stored in memory and reused when necessary. The following assumptions are considered when calculating the complexity order:

- The MA-ZF and MA-RZF algorithms have their complexity order determined essentially by the following two procedures: the null-space projections and the single-group beamforming. The null space projections are implemented through SVD. The term $\mathcal{O}(\mathrm{SGB}_{A \times B})$ expresses the complexity order of a Single-Group Beamforming (SGB) algorithm, taking into account an equivalent channel matrix with dimensions $A \times B$. This complexity depends on the chosen single-group beamforming algorithm. Note that the A and B dimensions are related to the size of each individual group g_k, being much smaller than the N and L dimensions. For example, the algorithm that maximizes the average SNR in [17] has complexity $\mathcal{O}(\frac{1}{2}AB^2 + B^3)$, the User-Selective Matched

TABLE II
COMPUTATIONAL COMPLEXITY OF THE BEAMFORMING ALGORITHMS.

Algorithm	Complexity order
MF	$\mathcal{O}(NLK)$
ZF	$\mathcal{O}(NLK + L^3 + L^2K + LK^2 + \frac{1}{2}NL^2)$
MMSE	$\mathcal{O}(NLK + L^3 + L^2K + LK^2 + \frac{1}{2}NL^2)$
MA-ZF	$\mathcal{O}(4L^3K + 2NL^2K) + \sum_{k=1}^{K} \mathcal{O}(\mathrm{SGB}_{g_k \times (L-N+g_k)})$
MA-RZF	$\mathcal{O}(6N^3K) + \sum_{k=1}^{K} \mathcal{O}(\mathrm{SGB}_{g_k \times g_k})$
SB	$\mathcal{O}(\frac{5}{3}\alpha NL^3 + \alpha N^2L^2)$
MA-SB	$\mathcal{O}(\frac{5}{3}\alpha NL^3 + \alpha N^2L^2)$
Bisec-SDR	$\mathcal{O}(3\sqrt{KL})\ \mathcal{O}(\frac{1}{2}(L^2K + N)^{3.5})$

Filter (USMF) algorithm in [8] has complexity $\mathcal{O}(2A^3B)$, and the SDR approach in [7] has $\mathcal{O}(\frac{1}{2}(A + B^2)^{3.5})$.

- The α parameter of the SB and MA-SB algorithms refers to the number of iterations considered by the alternating optimization procedure.
- According to [10, 13], the Bisec-SDR algorithm, which maximizes the minimum SINR for a given channel model, has its complexity divided into two parts: the number of iterations required for convergence, and the number of arithmetic operations required by each iteration. The first part takes into account a precision of 10^{-3}. With regard to the second part, the complexity order according to [10, 35] is expressed in terms of the number of arithmetic operations, i.e., both sums and multiplications are considered. Since in this section only the number of multiplications is taken into account, a factor of $1/2$ is introduced in order to roughly approximate the number of multiplications from the number of arithmetic operations.

The complexity of the algorithms is shown in Table II. It is expressed in terms of the complexity order and makes use of the big $\mathcal{O}$ notation. The algorithms are presented according to their increased order of complexity, which depends on the previously discussed assumptions concerning each algorithm.

In order to provide a better insight on the complexity of the different algorithms, Fig. 8 shows the complexity order as a function of the number N of users. Note that the y-axis is shown in logarithmic scale. The number L of transmit antennas is set to be equal to the number N of users, and the number K of groups is adjusted in such a way that half of the users are unicast users and the other half is roughly divided into multicast groups with 2 or 3 users per group. As in the results section, it is assumed that $\alpha = 5$ and that the SDR approach of [7] is selected as the single-group beamforming algorithm, which has complexity $\mathcal{O}(\frac{1}{2}(A + B^2)^{3.5})$.

The lowest complexity is presented by the MF algorithm, since it does not mitigate the interference. The drawback of its low complexity, as the previous performance analysis has shown, is that it achieves the worst BER results.

The ZF and MMSE algorithms introduce a channel inversion in order to mitigate the inter-group interference, for this reason they present a higher complexity than MF. Both ZF and MMSE have the same complexity order, since the only difference between them lies on the regularization factor, which does not increase the complexity order.

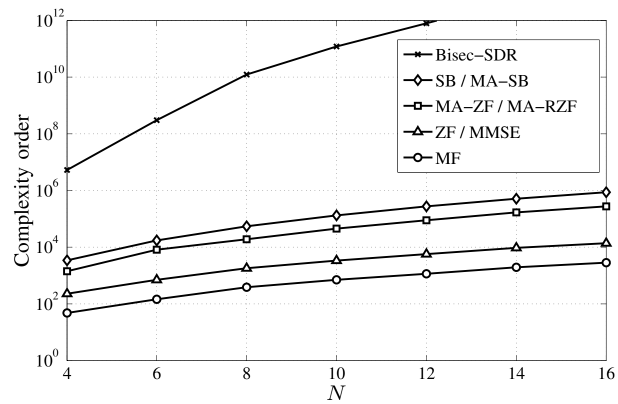

Fig. 8. Complexity order as a function of N, assuming that $L = N$.

Next, the MA-ZF and MA-RZF present the same complexity order. These algorithms have the null-space projections and single-group beamforming procedures in common, which are the preponderant factors for their complexity order, and for this reason they present an equivalent complexity order.

The alternating optimization employed by the SB and MA-SB algorithms is responsible for the increased complexity order with regard to the previous group of algorithms – MA-ZF and MA-RZF. The SB and MA-SB present practically the same complexity order. Both algorithms have a similar structure and the additional power redistribution of MA-SB is only performed once, thus not affecting the complexity order.

Finally, the Bisec-SDR algorithm has a much higher complexity than the other algorithms. This higher complexity of Bisec-SDR is due to the numerical optimization performed by the SDP solver. Even though the complexity order of Bisec-SDR may correspond to an upper complexity bound, as mentioned in [10], the actual complexity is still expected to be higher than that of the other algorithms.

VII. CONCLUSIONS

In this paper, the multi-group multicast problem has been investigated. Several linear algorithms have been formulated for the multi-group multicast case, multicast-aware enhancements have been proposed, and the performance and complexity of the algorithms have been analyzed throughout the paper.

In terms of performance, the best algorithm is the Bisec-SDR, which provides a tight approximation to the problem of maximizing the minimum SINR. It presents, however, the drawback of high computational complexity with regard to the other algorithms.

The proposed multicast-aware enhancements of the linear algorithms – MA-ZF, MA-RZF, and MA-SB – present significant gains with regard to the original algorithms – ZF, MMSE, and SB. In the case of MA-ZF and MA-RZF, the performance gain with regard to ZF and MMSE comes at the cost of a certain increase in complexity, due to the null space projections and single-group beamforming procedures. In the case of MA-SB, however, the proposed modifications do not significantly increase the complexity with regard to SB.

The best trade-off in terms of performance and complexity is achieved by the proposed MA-SB and MA-RZF algorithms. The choice among them depends on the ratio between the number of users and number of multicast groups, i.e., N/K. When regarding both performance and complexity aspects, the MA-RZF algorithm is more adequate for higher ratios ($K \rightarrow 1$), whereas the MA-SB is advised for lower ratios ($K \rightarrow N$).

References

[1] 3GPP, "Multimedia Broadcast/Multicast service (MBMS); architecture and functional description," Technical Specification Group Services and System Aspects, Tech. Rep. TS 23.246 v6.10.0, June 2006.

[2] M. Bakhuizen and U. Horn, "Mobile broadcast/multicast in mobile networks," *Ericsson Review*, issue no. 01/2005, May 2005.

[3] J. Ogunbekun and A. Mendjeli, "MBMS service provision and its challenges," in *Proc. IEE Int. Conf. on 3G Mobile Commun. Technologies*, June 2003, pp. 128–133.

[4] T. Jiang, W. Xiang, H.-H. Chen, and Q. Ni, "Multicast broadcast services support in OFDMA-based WiMAX systems," *IEEE Commun. Mag.*, vol. 45, no. 8, pp. 78–86, Aug. 2007.

[5] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1423–1436, Oct. 1998.

[6] Y. Sun and K. J. R. Liu, "Transmit diversity techniques for multicasting over wireless networks," in *Proc. IEEE WCNC*, vol. 1, Mar. 2004, pp. 593–598.

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

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

[9] R. Hunger, D. A. Schmidt, M. Joham, A. Schwing, and W. Utschick, "Design of single-group multicasting-beamformers," in *Proc. IEEE ICC*, June 2007.

[10] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Transmit beamforming to multiple co-channel multicast groups," in *Proc. IEEE CAMSAP*, Dec. 2005, pp. 109–112.

[11] Y. Gao and M. Schubert, "Group-oriented beamforming for multi-stream multicasting based on quality-of-service requirements," in *Proc. IEEE CAMSAP*, Dec. 2005, pp. 193–196.

[12] ——, "Power allocation for multi-group multicasting with beamforming," in *Proc. ITG/IEEE Workshop on Smart Antennas*, Mar. 2006.

[13] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Far-field multicast beamforming for uniform linear antenna arrays," *IEEE Trans. Signal Processing*, vol. 55, no. 10, pp. 4916–4927, Oct. 2007.

[14] Y. C. B. Silva and A. Klein, "Downlink beamforming and SINR balancing for the simultaneous provision of unicast/multicast services," in *Proc. IST Mobile and Wireless Communications Summit*, July 2007.

[15] ——, "Analysis of linear and non-linear precoding techniques for the spatial separation of unicast and multicast users," in *Proc. IEEE PIMRC*, Sep. 2007.

[16] S. Boyd and L. Vandenberghe, *Convex optimization*, 1st ed. Cambridge University Press, 2004.

[17] M. J. Lopez, "Multiplexing, scheduling, and multicasting strategies for antenna arrays in wireless networks," Ph.D. dissertation, Massachusetts Institute of Technology, 2002.

[18] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[19] A. Khisti, "Coding techniques for multicasting," Master's thesis, Massachusetts Institute of Technology, 2004.

[20] N. D. Sidiropoulos and T. N. Davidson, "Broadcasting with channel state information," in *Proc. IEEE SAM*, July 2004, pp. 489–493.

[21] M. Bengtsson and B. Ottersten, "Optimal downlink beamforming using semidefinite optimization," in *Proc. Annual Allerton Conf. on Communication, Control, and Computing*, Sep. 1999, pp. 987–996.

[22] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Convex transmit beamforming for downlink multicasting to multiple co-channel groups," in *Proc. IEEE ICASSP*, vol. 5, May 2006, pp. 973–976.

[23] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 18–28, Jan. 2004.

[24] A. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*, 1st ed. Cambridge University Press, 2003.

[25] A. Gershman and N. Sidiropoulos, Eds., *Space-Time Processing for MIMO Communications*. John Wiley & Sons, Apr. 2005.

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

[27] H. Boche and M. Schubert, "A general duality theory for uplink and downlink beamforming," in *Proc. IEEE VTC*, vol. 1, Sep. 2002, pp. 87–91.

[28] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 361–374, Feb. 2006.

[29] J. Zander, "Performance of optimum transmitter power control in cellular radio systems," *IEEE Trans. Veh. Technol.*, vol. 41, pp. 57–62, Feb. 1992.

[30] Q. H. Spencer and M. Haardt, "Capacity and downlink transmission algorithms for a multi-user MIMO channel," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, vol. 2, Nov. 2002, pp. 1384–1388.

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

[32] D. P. Palomar and J. R. Fonollosa, "Practical algorithms for a family of waterfilling solutions," *IEEE Trans. Signal Processing*, vol. 53, pp. 686–695, Feb. 2005.

[33] F. R. Farrokhi, A. Lozano, G. J. Foschini, and R. Valenzuela, "Spectral efficiency of wireless systems with multiple transmit and receive antennas," in *Proc. IEEE PIMRC*, vol. 1, Sep. 2000, pp. 18–21.

[34] J. Zhang, A. M. Sayeed, and B. D. V. Veen, "Optimal space-time transceiver design for selective wireless broadcast with channel state information," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 2040–2050, Nov. 2004.

[35] Y. Ye, *Interior point algorithms: theory and analysis*. John Wiley & Sons, 1997.

Yuri C. B. Silva (S'01-M'09) was born in Fortaleza, Brazil, in 1978. He received the B.Sc. and M.Sc. degrees from the Federal University of Ceará, Fortaleza, Brazil, in 2002 and 2004, respectively, and the Dr.-Ing. degree (with distinction) from the Technische Universität Darmstadt, Germany, in 2008, all in electrical engineering.

From 2001 to 2004 he was a researcher at the Wireless Telecom Research Group - GTEL, Fortaleza, Brazil. In 2003 he visited Ericsson Research, Stockholm, Sweden, for a three-month research internship. From 2005 to 2008 he was a research assistant at the Communications Engineering Lab of the Technische Universität Darmstadt, Germany. Since 2008 he has been with GTEL as a senior researcher and project leader. His research interests include multi-user and multi-antenna communications, radio resource management, and cooperative wireless networks.

Anja Klein (S'93-M'96) received the diploma and Dr.-Ing. (Ph.D.) degrees in electrical engineering from University of Kaiserslautern, Germany, in 1991 and 1996, respectively. From 1991 to 1996, she was a member of the staff of the Research Group for RF Communications at University of Kaiserslautern. In 1996, she joined Siemens AG, Mobile Networks Division, Munich and Berlin. She was active in the standardization of third generation mobile radio in ETSI and in 3GPP, for instance leading the TDD group in RAN1 of 3GPP. She was vice president, heading a development department and a systems engineering department.

In May 2004, she joined Darmstadt University of Technology, Germany, as full professor, heading the Communications Engineering Lab. Her main research interests are in mobile radio, including multiple access, single and multi carrier transmission schemes, multi antenna systems, radio resource management, network planning and dimensioning, relaying and cooperative communication.

Dr. Klein has published over 150 refereed papers and has contributed to five books. She is inventor and co-inventor of more than 45 patents in the field of mobile radio. In 1999, she was inventor of the year of Siemens AG. Dr. Klein is a member of IEEE and of Verband Deutscher Elektrotechniker - Informationstechnische Gesellschaft (VDE-ITG).