EFFICIENT RESOURCE ALLOCATION FOR ORTHOGONAL TRANSMISSION IN BROADCAST CHANNELS

Eduard Calvo * *and Javier R. Fonollosa*

SPCOM Group, Dept. of Signal Theory and Communications, Technical University of Catalonia (UPC)
Jordi Girona 1-3, Campus Nord, Ed. D5, 08034 Barcelona (SPAIN)
Email: {eduard,fono}@gps.tsc.upc.es

ABSTRACT

The allocation of network resources for spectral efficiency maximization in a broadcast channel under the practical restrictions of orthogonal transmission and the use of squared QAM constellations is addressed. Given a total transmit power constraint and (possibly) different per-user quality of service requirements in the form of target uncoded bit error rates, efficient power and bit loading algorithms are proposed to maximize a weighted sum of the users' rates that tightly match the performance of the optimal but computationally prohibitive allocation strategy.

1. INTRODUCTION

Efficient management of network resources has become a prominent issue for achieving high data rates in the downlink (broadcast channel) of upcoming wireless communication systems due to the limitations imposed by bandwidth scarcity and transmit power regulatory constraints. Such systems are likely to handle data flows of different service types sharing the same physical layer mechanism, hence giving rise to the need for Resource Allocation (RA) strategies complying with differentiated Quality of Service (QoS) constraints among users. To the end of further increasing the system spectral efficiency, dedicated feedback channels are used to gather Channel State Information (CSI) at the Base Station (BS) to adapt the RA policy to match the instantaneous network conditions.

We focus on broadcast channels where the transmit/receive strategies orthogonalize the users' signals [1][2][3]. In this way, the allocation of network resources (power and rate) is decoupled among users and only depends on their QoS requirements and the total transmit power constraint. More specifically, we consider QoS requirements in the form of target uncoded bit error rates and squared QAM modulation formats. We choose a weighted sum of the users rates as the figure of merit of the resource allocation.

In this work, an efficient algorithm for the computation of the optimal rate and power loading is derived in the case of continuously varying rates. When it comes to discrete-valued rates (feasible modulation formats), the optimization of the resource allocation becomes an integer optimization problem for which standard convex optimization techniques cannot be applied and whose complexity grows exponentially with the number of users. We tackle the integer optimization by proposing an efficient algorithm based on a modification of the allocation scheme in the case of unconstrained rates. The performance of the proposed algorithm is benchmarked against the optimal allocation obtained with a brute force exhaustive search for two different orthogonal transmission strategies. For zero-forcing beamforming the proposed allocation performs indistinguishably to the exhaustive search, while for zero-forcing Tomlinson-Harashima precoding it is able to capture most of the achievable system efficiency.

This paper is organized as follows. Section 2 introduces some preliminaries regarding the system model and notation, the implications of orthogonal transmissions, and the problem statement. In Section 3 we describe the optimal power and rate loading for continuous-valued rates, which allows us to address the proposed loading algorithm in Section 4. The numerical simulations for performance assessment are left to Section 5, and Section 6 concludes the paper.

2. PRELIMINARIES

2.1. System Model

We consider a BS equipped with n_T antennas and total transmit power P_T serving $\widetilde{K}$ geographically dispersed users, each of them equipped with n_R antennas. Although in practice $\widetilde{K} \gg n_T$, we assume that at any given time slot only a subset of $K \leq n_T$ users are allowed to communicate simultaneously with the BS as a result of some upper-layer Packet Scheduling

*This work was partially supported by the Spanish Government under the grant FPU-AP-2004-3549.

This work has been partially funded by the European Commission, the Spanish Ministry of Education and Science, the Catalan Government and FEDER funds under contracts TEC2006-06481, TEC2004-04526, 27187 SURFACE and 2005SGR-00639.

(PS) strategy. The received signal of the k-th user is

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{w}_k, \quad (1)$$

where $\mathbf{y}_k \in \mathbb{C}^{n_R \times 1}$, $\mathbf{H}_k \in \mathbb{C}^{n_R \times K}$ is the channel matrix, $\mathbf{x} \in \mathbb{C}^{K \times 1}$ is the transmitted vector, and $\mathbf{w}_k \in \mathbb{C}^{n_R \times 1}$ is the noise vector, with uncorrelated entries drawn $\mathcal{CN}(0, \sigma^2)$, $1 \leq k \leq K$. We further assume a Rayleigh model for $\mathbf{H}_k$, with each component drawn i.i.d. $\sim \mathcal{CN}(0, 1)$.

Vector $\mathbf{x}$ bears the information symbols[1] $\{s_1, s_2, \ldots, s_K\}$, each one uniformly drawn from a square M_k-QAM constellation yielding a rate $R_k = \log_2(M_k)$ [bit/ch.use] for user k. Let us denote by $\mathcal{R} = \{0, 2, 4, \ldots, R_{max}\}$ the set of available modulation formats, with R_{max} an even integer ($R = 0$ meaning no transmission). The transmission of s_k consumes a power p_k, such that $E\{\|\mathbf{x}\|^2\} = \sum_{i=1}^{K} p_k$.

The k-th user estimates the symbol s_k from its output $\mathbf{y}_k$, incurring in a bit error probability BER_k which is decreasing in p_k and increasing in R_k, and whose explicit expression depends on the transmit technique and on the receiver. We denote by BER_k^0 the target QoS for user k.

2.2. Orthogonal Transmission

Assuming that the BS has complete knowledge about the set of channel matrices $\{\mathbf{H}_k\}_{k=1}^K$, the downlink resource allocation problem can be simplified by orthogonalization of the users. In this way, the power and rate allocated to a specific user does not impact on the performance of the rest of them. Orthogonalization can be either implemented by presubtracting multi-user interference at the transmitter (using, for instance zero-forcing transmit beamforming, ZFBF, or zero-forcing Tomlinson-Harashima precoding, ZF-THP [4], see [2] for a thorough comparison of the aforementioned techniques) or by combined transmitter-receiver processing [5] [6]. In either case, the received instantaneous signal-to-noise ratio (SNR) of user k can be written as

$$\mathsf{snr}_k = c_k p_k, \quad (2)$$

where c_k is a channel dependent quantity that depends also on the transmit/receive technique. Additionally, we define the global system SNR as $\mathsf{snr} = P_T / \sigma^2$.

For illustrative purposes, we consider the case $n_R = 1$, for which the received signal of the k-th user (1) reduces to

$$y_k = \mathbf{h}_k^T \mathbf{x} + w_k, \quad (3)$$

and we can define the aggregated channel matrix $\mathbf{H} = [\mathbf{h}_1\ \mathbf{h}_2 \ldots \mathbf{h}_k]^T$. In this case,

$$c_k = |\mathbf{h}_k^T \mathbf{w}_k|^2 / \sigma^2 \quad (4)$$

[1]Note that we are implicitly assuming that if $n_R > 1$ all the received antennas are used for diversity purposes only.

for ZFBF, where $\mathbf{w}_k \in \mathbb{C}^{n_T \times 1}$ is the k-th normalized column of $\mathbf{H}^\dagger (\mathbf{H}\mathbf{H}^\dagger)^{-1}$, and

$$c_k = |[\mathbf{G}]_{k,k}|^2 / \sigma^2 \quad (5)$$

for ZF-THP, where $\mathbf{H} = \mathbf{G}\mathbf{Q}$ with $\mathbf{G}$ upper-triangular and $\mathbf{Q}$ unitary is the QR decomposition[2] of $\mathbf{H}$.

Thanks to user orthogonalization, the bit error rate of the k-th user can be expressed using the standard approximation for square M_k-QAM modulations [7]

$$\mathrm{BER}_k(p_k, R_k) = 4 \frac{1 - 2^{-R_k/2}}{R_k} \mathcal{Q}\Big(\sqrt{\frac{3c_k}{2^{R_k} - 1} p_k}\Big), \quad (6)$$

where $\mathcal{Q}(\cdot)$ is the Gaussian Q-function and we have used $M_k = 2^{R_k}$.

2.3. Problem Statement

We aim at finding the resource allocation $\{p_k, R_k\}_{k=1}^K$ that maximizes a weighted sum of the users rates (spectral efficiencies), for some non-negative weights[3] $\{\mu_k\}_{k=1}^K$, while satisfying the transmit power constraint and the QoS requirements of the users. The optimal allocation can be formulated as follows.

$$\underset{\{p_k, R_k\}}{\text{maximize}} \ \sum_{k=1}^{K} \mu_k R_k \quad (7)$$

$$\text{subject to } \ \mathrm{BER}_k(p_k, R_k) \leq \mathrm{BER}_k^0,\ 1 \leq k \leq K \quad (8)$$

$$p_k \geq 0, R_k \in \mathcal{R},\ 1 \leq k \leq K \quad (9)$$

$$\sum_{k=1}^{K} p_k \leq P_T \quad (10)$$

Due to (9) the optimization has to be carried over integer rates, and hence combinatorial search on $\mathcal{R}$ needs to be performed for each R_k. In order to be able to propose alternative simpler methods, we study first the optimal solution to (7)-(10) when (9) is relaxed so as to admit non-integer rates.

3. OPTIMAL LOADING FOR CONTINUOUS RATES

A relaxation of (7)-(10) is obtained by replacing the constraint $R_k \in \mathcal{R}$ in (9) by $0 \leq R_k \leq R_{max}$. If we rewrite (8) as

$$h(R_k; \mathrm{BER}_k^0) - 3c_k p_k \leq 0, \quad (11)$$

where

$$h(R; \mathrm{BER}_0) = (2^R - 1)\Psi^2\Big(\frac{1}{4}\mathrm{BER}_0 \frac{R}{1 - 2^{-R/2}}\Big) \quad (12)$$

[2]Note that due to the QR decomposition of $\mathbf{H}$, c_k depends on the user ordering (user encoding order).

[3]The weights can be determined by the PS according to other QoS parameters such as queue lengths, delays, and/or service type.

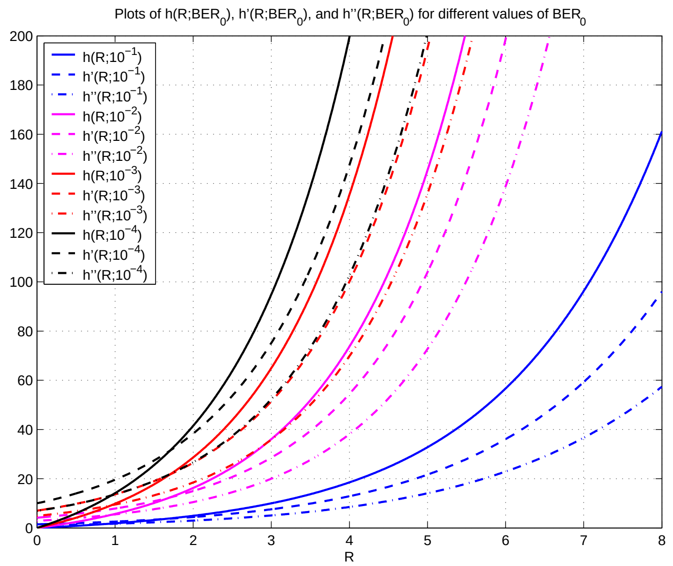

Fig. 1. Different plots of the function $h(R; \text{BER}_0)$ in the range of interest ($R \in [0, R_{max}]$), $R_{max} = 8$, for different possible values of the target BER_0.

and $\Psi(\cdot) = \mathcal{Q}^{-1}(\cdot)$ is the inverse Q-function. The relaxed problem can then be formulated as

$$\underset{\{p_k, R_k\}}{\text{maximize}} \quad \sum_{k=1}^{K} \mu_k R_k \tag{13}$$

$$\text{subject to} \quad h(R_k; \text{BER}_k^0) - 3c_k p_k \leq 0, \; 1 \leq k \leq K \tag{14}$$

$$p_k \geq 0, 0 \leq R_k \leq R_{max}, \; 1 \leq k \leq K \tag{15}$$

$$\sum_{k=1}^{K} p_k \leq P_T, \tag{16}$$

an optimization problem whose convexity depends exclusively on the convexity of the function $h(R; \text{BER}_0)$ with respect to R for some fixed BER_0. To that end, we should verify that

$$h''(R; \text{BER}_0) \equiv \frac{\partial^2 h(R; \text{BER}_0)}{\partial R^2} > 0 \;\text{ for }\; 0 \leq R \leq R_{max}. \tag{17}$$

Unfortunately, the analytical expression of $h''(R; \text{BER}_0)$ is not amenable to work with. For this reason, we have adopted a rather practical approach by plotting h, its first derivative, h', and h'' in Figure 1 from their lengthy analytical expressions not reported here for the sake of brevity.

We conclude from Figure 1 that h, h', and h'' are strictly increasing functions of R and that h is a convex function of R in the range $0 \leq R \leq R_{max}$ for practical values of R_{max} and BER_0. Therefore, (13)-(17) can be solved optimally using standard convex optimization tools. Alternatively, we can use Algorithm 1 (based on the KKT conditions of (13)-(17)) to obtain the optimal resource allocation.

Algorithm 1 Optimal Loading

1: Initializations: $\theta_{low} = 0$, $\theta_{up} = 3 \max_k \left\{ \frac{\mu_k c_k}{h'(0; \text{BER}_k^0)} \right\}$.
2: **while** $\theta_{up} - \theta_{low} > \epsilon$ **do**
3: $\quad \theta = (\theta_{low} + \theta_{up})/2$.
4: $\quad$ **for** $k = 1 \ldots K$ **do**
5: $\quad\quad R_k(\theta) = [R \,:\, h'(R; \text{BER}_k^0) = 3\mu_k/\theta]_0^{R_{max}}$.
6: $\quad\quad p_k(\theta) = h(R_k(\theta); \text{BER}_k^0)/3c_k$.
7: $\quad$ **end for**
8: $\quad$ **if** $\sum_k p_k(\theta) \leq P_T$ **then**
9: $\quad\quad \theta_{up} = \theta$.
10: $\quad$ **else**
11: $\quad\quad \theta_{down} = \theta$.
12: $\quad$ **end if**
13: **end while**
14: $R_k^\star = R_k(\theta_{up})$, $p_k^\star = p_k(\theta_{up})$ for $1 \leq k \leq K$.

Algorithm 1 is based on a bisection search for a water-level parameter θ up to a precision ruled by any arbitrarily small positive ϵ. The need for the bisection arises from the expression of $h'(R; \text{BER}_0)$, which is not analytically invertible. The larger θ, the lower the rates and the powers and vice versa. When $\theta = \theta_{up}$ in step 1 of Algorithm 1, all the rates and powers are zero. Therefore, bisection arises so as to approximate as close as desired the optimal θ for which the transmit power constraint (16) is satisfied with equality. Note that another bisection is needed in step 5 to determine $R_k(\theta)$.

4. PROPOSED LOADING ALGORITHM

In order to tackle the integer optimization of (7)-(10) we propose to use a simplified version of Algorithm 1, where step 5 is replaced by

$$R_k(\theta) = \max\{R \in \mathcal{R} : h'(R; \text{BER}_k^0) \leq 3\mu_k/\theta\}, \tag{18}$$

and $R_k(\theta) = 0$ if no rate in $\mathcal{R}$ satisfies the right hand side of (18). In other words, we avoid the bisection algorithm of step 5 of Algorithm 1 by selecting the best out of the available modulation formats in $\mathcal{R}$. The selection process can be efficiently performed if the values $\{h'(R; \text{BER}_k^0), h(R; \text{BER}_k^0)\}_{R \in \mathcal{R}}$ are computed offline $1 \leq k \leq K$.

One drawback of using (18) in step 5 of Algorithm 1 is that the available power is always underutilized: there exists a fraction of power which is not used because of the rounding of (18). In order to exploit the remaining transmit power we propose to increase (when possible) the rates of users with large priorities or small power demands to increase their rate until power shortage is achieved, as in [8]. The complete description of the proposed loading algorithm follows in Algorithm 2, where Δ_k is the extra power required to increase R_k in 2 bits per symbol.

Algorithm 2 Proposed Loading

1: Initializations: $\theta_{low} = 0$, $\theta_{up} = 3 \max_k \left\{ \frac{\mu_k c_k}{h'(0;\mathrm{BER}_k^0)} \right\}$.
2: **while** $\theta_{up} - \theta_{low} > \epsilon$ **do**
3: $\theta = (\theta_{low} + \theta_{up})/2$.
4: **for** $k = 1 \ldots K$ **do**
5: $R_k(\theta) = \max\{R \in \mathcal{R} : h'(R;\mathrm{BER}_k^0) \leq 3\mu_k/\theta\}$.
6: $p_k(\theta) = h(R_k(\theta);\mathrm{BER}_k^0)/3c_k$.
7: **end for**
8: **if** $\sum_k p_k(\theta) \leq P_T$ **then**
9: $\theta_{up} = \theta$.
10: **else**
11: $\theta_{down} = \theta$.
12: **end if**
13: **end while**
14: $R_k = R_k(\theta_{up})$, $p_k = p_k(\theta_{up})$ for $1 \leq k \leq K$.
15: **for** $k = 1 \ldots K$ **do**
16: **if** $R_k < R_{max}$ **then**
17: $\Delta_k = h(R_k + 2;\mathrm{BER}_k^0)/3c_k - p_k$.
18: **else**
19: $\Delta_k = +\infty$.
20: **end if**
21: **end for**
22: **repeat**
23: $k' = \arg \min_{k:(\sum_i p_i)+\Delta_k \leq P_T} \Delta_k/\mu_k$.
24: Set $R_{k'} = R_{k'} + 2$, $p_{k'} = p_{k'} + \Delta_{k'}$, and update $\Delta_{k'}$.
25: **until** $(\sum_i p_i) + \Delta_k > P_T \; \forall k$

The idea of Algorithm 2 is to use the step-by-step quantized allocation of Algorithm 1 as an starting allocation for launching the rate increase procedure comprised between steps 15 and 25. In this way, we guarantee a desirable closeness to the optimal real-valued rate quantization, from which we deviate so as not to leave any fraction of power unused. Additionally, and thanks to the quantization effect of step 5, the intermediate allocation of step 14 is not as sensitive to errors is θ, and the bisection method for can be alleviated with a larger ϵ than in Algorithm 1.

5. PERFORMANCE EVALUATION

The performance of the proposed rate and power loading algorithm (Algorithm 2) has been evaluated in terms of achievable spectral efficiency region when using ZFBF and ZF-THP with single-antenna receivers. A fair benchmarking of the proposed algorithm has been provided through the comparison with a brute force algorithm selecting the allocation maximizing the weighted sum rate (7) after an exhaustive search on all the possible choices for $\{R_1, R_2, \ldots, R_K\} \in \mathcal{R}^K$. When analyzing ZF-THP, the exhaustive search takes also into account all possible encoding orders.

For the simulations, 5000 Montecarlo runs of a $K = 2$

Algorithm 3 Allocation quantization

1: **if** $P < P_T$ **then**
2: **repeat**
3: $(k', n') = \arg \min_{k,n \,:\, P+\Delta_k^+[n] \leq P_T} \Delta_k^+[n]/\mu_k$
4: Increase $R_{k'}[n']$ one modulation format, update $P = P + \Delta_{k'}^+[n']$, and update $(p_{k'}^q[n'], \Delta_{k'}^+[n'], \Delta_{k'}^-[n'])$.
5: **until** $P + \Delta_k^+[n] > P_T \; \forall n, k$
6: **else**
7: **repeat**
8: $(k', n') = \arg \max_{k,n \,:\, \Delta_k^-[n] > 0} \Delta_k^-[n]/\mu_k$
9: Decrease $R_{k'}[n']$ one modulation format, update $P = P - \Delta_{k'}^-[n']$, and update $(p_{k'}^q[n'], \Delta_{k'}^+[n'], \Delta_{k'}^-[n'])$.
10: **until** $P \leq P_T$
11: **end if**

broadcast channel with $n_R = 1$, unit power noise ($\sigma^2 = 1$), and Rayleigh fading were averaged. For each simulation, the channel dependant gains were computed for ZFBF (4) and ZF-THP (5). For ZF-THP, and the proposed allocation of Algorithm 2, the user ordering was performed in decreasing order of weighted channel energies $\mu_k \|\mathbf{h}_k\|^2$. The rationale behind this ordering is the following: it is shown in [9] that $E\{|[\mathbf{G}]_{k,k}|^2\}$ (and hence, $E\{c_k\}$) is highly decreasing in k and users encoded first experience (on average) much better channel conditions. We therefore try to take advantage of this effect by encoding first those users with larger priorities or better channel conditions.

We considered two different scenarios: one with equal QoS requirements ($\mathrm{BER}_1^0 = \mathrm{BER}_2^0 = 10^{-3}$) and another with asymmetric requirements ($\mathrm{BER}_1^0 = 10^{-3}$ and $\mathrm{BER}_2^0 = 10^{-4}$). Figures 2 and 3 show the achievable spectral efficiency regions obtained for both scenarios at $\mathsf{snr} = 10$dB and $\mathsf{snr} = 20$dB, respectively. In order to study the impact of the arbitrary rate increase procedure of steps 15 to 25 of Algorithm 2 on the final performance, we distinguish the intermediate results obtained up to step 14 of Algorithm 2 (regions labeled as '1:') and the complete application of Algorithm 2 (regions labeled as '2:'); brute force regions are labeled 'bf:'.

Figures 1 and 2 show that the intermediate allocation up to step 15 of Algorithm 2 is able to capture a significant fraction of the achievable spectral efficiency, specially when user priorities are highly unbalanced or totally equal. The rate increase procedure of steps 15 to 25 is able to raise the achievable efficiencies so that for ZFBF the performance is indistinguishable from the brute force allocation.

For ZF-THP, the gap between the proposed allocation and the brute force search reduces with snr, and can be mostly put on the specific choice of the encoding order and not on the allocation algorithm. Attending to the results of Figure 1 and 2, we believe that the goodness of the results for ZFBF can be extrapolated to other techniques that are not sensitive to the user ordering. Finally, the achievable spectral efficiencies for

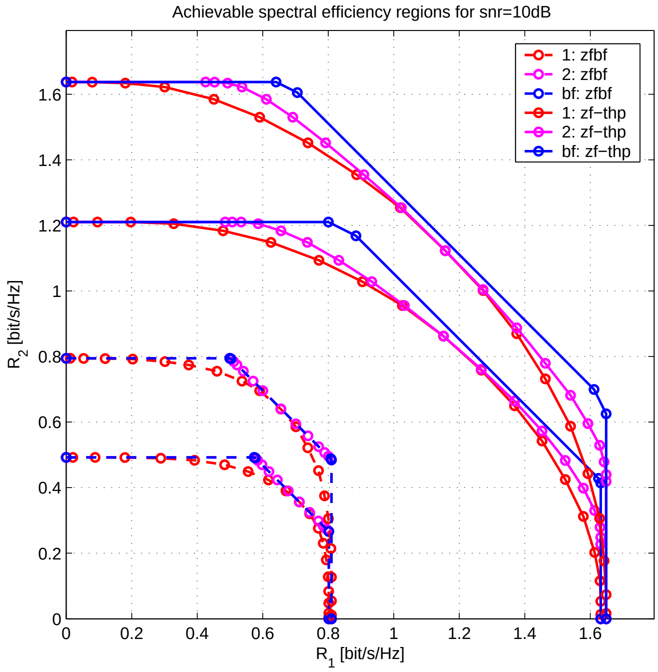

Fig. 2. Achievable spectral efficiency region [bit/s/Hz] for $\mathsf{snr} = 10$ dB. The symmetric regions correspond to the QoS constraints $\mathrm{BER}_1^0 = \mathrm{BER}_2^0 = 10^{-3}$, while the asymmetric regions correspond to $\mathrm{BER}_1^0 = 10^{-3}$ and $\mathrm{BER}_2^0 = 10^{-4}$.

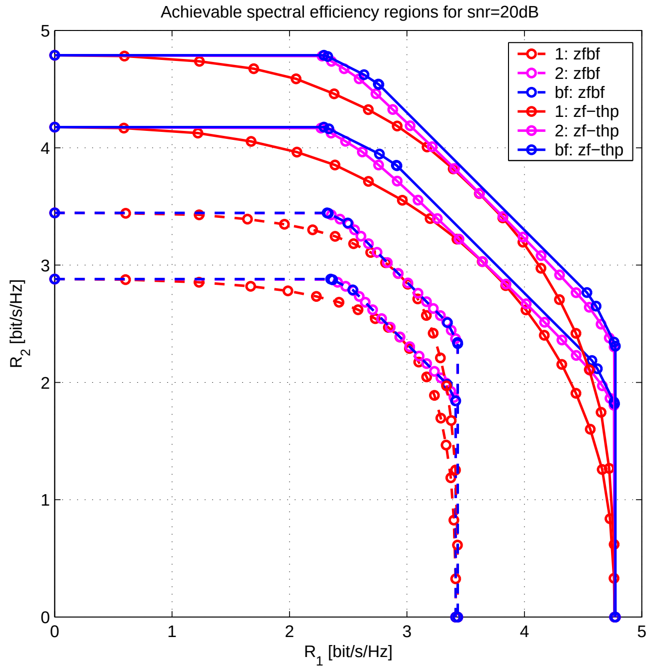

Fig. 3. Achievable spectral efficiency region [bit/s/Hz] for $\mathsf{snr} = 20$ dB. The symmetric regions correspond to the QoS constraints $\mathrm{BER}_1^0 = \mathrm{BER}_2^0 = 10^{-3}$, while the asymmetric regions correspond to $\mathrm{BER}_1^0 = 10^{-3}$ and $\mathrm{BER}_2^0 = 10^{-4}$.

ZF-THP are larger than for ZFBF, specially at low snr.

6. CONCLUSIONS

We have addressed joint power and rate loading in the broadcast channel with uncoded bit error rate constraints as QoS metric and square QAM modulation formats. By further constraining the transmit/receive processing to orthogonalize the users, the resource allocation optimization simplifies into decoupled problems linked by the total power constraint.

We found that the optimal allocation in the case of continuously varying rates is the solution to a convex optimization problem, and we proposed an efficient algorithm for its computation (Algorithm 1). With respect to the optimization of the allocation by considering only feasible values for the rates (even integers), we proposed an allocation scheme based on Algorithm 1 that tackled integer optimization. Performance comparison with a brute force allocation scheme showed excellent behavior at a complexity load several orders of magnitude below. Future work should address the extension of this work to the multicarrier setting (OFDMA).

7. REFERENCES

[1] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 528–541, March 2006.

[2] S. Thoen, L. Van der Perre, M. Engels, and H. De Man, "Adaptive loading for OFDM/SDMA-based wireless networks," *IEEE Trans. Commun.*, vol. 50, no. 11, pp. 1798–1810, Nov 2002.

[3] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication - Part ii: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, Mar 2005.

[4] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multi-antenna and multi-user communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305–1316, July 2004.

[5] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb 2004.

[6] G. Primolevo, O. Simeone, and U. Spagnolini, "Channel aware scheduling for broadcast MIMO systems with orthogonal linear precoding and fairness constraints," in *IEEE Intern. Conference on Communications*. Korea, May 2005.

[7] S. Benedetto and E. Biglieri, *Principles of digital transmission: with wireless applications*, New York: Kluwer Academic, 1999.

[8] E. Calvo and J. R. Fonollosa, "Near-optimal joint power and rate allocation for OFDMA broadcast channels," in *Proc. IEEE ICASSP*. Honolulu, USA, April 2007.

[9] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," in *Foundations and Trends in Communications and Information Theory*, 2002, vol. 1.