MIMO Broadcast Channel Rate Region with Linear Filtering at High SNR: Full Multiplexing

Paul de Kerret*, Raphael Hunger†, Michael Joham†, Wolfgang Utschick†, and Rudolf Mathar*
* Institute for Theoretical Information Technology, RWTH Aachen University,
52056 Aachen, Germany, email:dekerret@eurecom.fr,mathar@ti.rwth-aachen.de
† Associate Institute for Signal Processing, Technische Universität München,
80290 Munich, Germany, email:{hunger,joham,utschick}@tum.de

Abstract—**In this paper, the rate region of the two user MIMO broadcast channel (BC) is investigated in the high signal-to-noise ratio (SNR) regime, when linear precoding is used without time sharing. The transmitter is assumed to have more antennas than the sum of the receiving antennas, which is a meaningful assumption in cellular network after a scheduler has been used. To reach the rate region's boundary, the sum rate is maximized subject to a given ratio between the user's rates. The sum rate is first considered asymptotically when the SNR tends to infinity and the well known high SNR affine approximation in terms of the logarithm of the SNR is used. The multiplicative and the additive parameters are called respectively the multiplexing gain (MG) and the rate offset (RO) and the maximal values of these two parameters, which were not known in that case with proportional rate fairness considered, are derived at every point of the boundary of the rate region, as well as the optimal stream allocations associated with them. Analytical bounds for the boundary of the high SNR approximated rate region are then developed. Finally, the maximization of the rate subject to a rate ratio constraint is studied at finite SNR and algorithmic inner and outer bounds for the rate region boundary are derived. They are shown to be very close to each other and accurate even at intermediate SNR.**

I. Introduction

We consider a two user broadcast channel (BC) where both users have several antennas and the base station (BS) is assumed to have more antennas than the sum of the antennas of the two users. This is a common scenario in cellular network where the BS can have many antennas while the users have less antennas and the users to transmit to are selected by a scheduler. Moreover, the transmitter and the receivers have perfect channel state information and the power available at the BS is assumed to be very large. This could for example corresponds to a "hot spot" scenario where a large rate is provided to users in a favorable environment. On the opposite to the weighted sum-rate maximization approaches, we derive the true rate region and not the one obtained after the convex hull operation.

The capacity region is known to be achievable with dirty paper coding (DPC) [1], [2] and globally optimum algorithms are available to maximize the weighted sum rate [3], [4]. Still, DPC requires a demanding implementation [5], while linear precoding is a suboptimal alternative with good performance and low complexity. Thus, we assume in the following that the transmitter applies linear precoding. In that case, the computation of the rate region is complicated, and algorithms have only been derived to obtain a lower bound for the convex hull via weighted sum rate maximization [6], [7].

In the high SNR regime, analytical expressions could be derived for the optimal weighted sum rate of linear precoding [8], [9]. Still, only the convex hull of the rate region has obtained.

Algorithms to maximize a common rate achieved by all users simultaneously have been derived with DPC in [10], [11], for the linear case with single antenna receivers in [12] and with multi-antenna receivers in [13]. These works also consider the maximization of the sum rate with constraints on the minimal rates achieved by the users which can then be used with bisection to fulfill a proportional fairness constraint. Yet, no algorithm deals directly with arbitrary rate ratio constraint and more relevant to us, no analytical analysis is available, even at high SNR and the results for this case are completely different from those for the weighted sum rate maximization. The case when the BS has fewer antennas than the sum of the receiving antennas has been studied in [14], and both works complement each other to cover every antenna configuration and point of the rate region boundary.

The main contributions of the paper are as follows.

- A new model and new notations are developed to deal with the high SNR rate expressions.
- The multiplexing gain (MG) and the rate offset (RO) are derived at every point of the rate region boundary.
- The previous usual high SNR approximation is further improved to highly accurate bounds.

In this work, the calculations are based on the rate duality between the MIMO BC and a dual MIMO Multiple Access Channel (MAC) with the same sum power constraint [15].

In Section II, we introduce our system model and the optimization problem. In Section III, we derive the MG and the RO. Analytical bounds are then derived in Section IV and algorithmic ones in Section V. Finally, some simulations are presented in Section VI.

Notation: The operators $\|\cdot\|_{\mathrm{F}}$, $|\cdot|$, and $(\cdot)^{\mathrm{H}}$ denote the Frobenius norm, the determinant operator, and the Hermitian

This work was supported by Deutsche Forschungsgemeinschaft (DFG) under fund Ut 36/9-1.

transposition, respectively, and $\lceil\cdot\rceil$ denotes the ceiling operator. We write *streams* instead of *independent data streams*.

II. System Model

A. Rate Expressions

We consider a broadcast channel (BC) with two users, having r_1 and r_2 antennas, respectively, while the base station (BS) has t antennas with $t \geq r_1 + r_2$. We denote the antenna configuration of the users by $\boldsymbol{r} \triangleq (r_1, r_2)$ and the stream allocation by $\boldsymbol{b} \triangleq (b_1, b_2)$, where b_i is the number of streams allocated to user i. The total power available at the BS is P and is normalized such that the variance of the noise is one. P can then be associated to the SNR. We consider the transmission in the dual MAC in which user i transmits to the BS with the power P_i and $P_1 + P_2 = P$ [15]. The channel seen by user i is given by $\mathbf{H}_i \in \mathbb{C}^{t\times r_i}$ and is assumed to be full rank and perfectly known at the BS and at both users. Each element of the channel is generated randomly from an independent identically distributed standard Gaussian distribution, and the same holds for every element of the noise vector. If user i applies the full rank precoding matrix $\mathbf{T}_i \in \mathbb{C}^{r_i\times b_i}$, the rates of user one and two are given by [15]

$$\begin{aligned} R_1 &= \log\left|\mathbf{I}+\mathbf{T}_1^{\mathrm{H}}\mathbf{H}_1^{\mathrm{H}}(\mathbf{I}-\mathbf{H}_2\mathbf{T}_2(\mathbf{I}+\mathbf{T}_2^{\mathrm{H}}\mathbf{H}_2^{\mathrm{H}}\mathbf{H}_2\mathbf{T}_2)^{-1}\mathbf{T}_2^{\mathrm{H}}\mathbf{H}_2^{\mathrm{H}})\mathbf{H}_1\mathbf{T}_1\right|, \\ R_2 &= \log\left|\mathbf{I}+\mathbf{T}_2^{\mathrm{H}}\mathbf{H}_2^{\mathrm{H}}(\mathbf{I}-\mathbf{H}_1\mathbf{T}_1(\mathbf{I}+\mathbf{T}_1^{\mathrm{H}}\mathbf{H}_1^{\mathrm{H}}\mathbf{H}_1\mathbf{T}_1)^{-1}\mathbf{T}_1^{\mathrm{H}}\mathbf{H}_1^{\mathrm{H}})\mathbf{H}_2\mathbf{T}_2\right|. \end{aligned} \tag{1}$$

We now decompose the precoding matrices via $\mathbf{T}_i = \sqrt{P_i/b_i}\bar{\mathbf{T}}_i$, with the *normalized precoding matrix* (NP matrix) $\bar{\mathbf{T}}_i$, such that $\|\bar{\mathbf{T}}_i\|_{\mathrm{F}}^2 = b_i$ and $\|\mathbf{T}_i\|_{\mathrm{F}}^2 = P_i$. Since we consider the high SNR regime, every non-zero eigenvalue of the transmit covariance matrix can be assumed to be very large. Thus, the identity in the inverse term in (1) can be neglected, and we get the high SNR approximated rates

$$\begin{aligned} R_1' &= \log\left|\mathbf{I}+\mathbf{T}_1^{\mathrm{H}}\mathbf{H}_1^{\mathrm{H}}(\mathbf{I}-\mathbf{H}_2\bar{\mathbf{T}}_2(\bar{\mathbf{T}}_2^{\mathrm{H}}\mathbf{H}_2^{\mathrm{H}}\mathbf{H}_2\bar{\mathbf{T}}_2)^{-1}\bar{\mathbf{T}}_2^{\mathrm{H}}\mathbf{H}_2^{\mathrm{H}})\mathbf{H}_1\mathbf{T}_1\right|, \\ R_2' &= \log\left|\mathbf{I}+\mathbf{T}_2^{\mathrm{H}}\mathbf{H}_2^{\mathrm{H}}(\mathbf{I}-\mathbf{H}_1\bar{\mathbf{T}}_1(\bar{\mathbf{T}}_1^{\mathrm{H}}\mathbf{H}_1^{\mathrm{H}}\mathbf{H}_1\bar{\mathbf{T}}_1)^{-1}\bar{\mathbf{T}}_1^{\mathrm{H}}\mathbf{H}_1^{\mathrm{H}})\mathbf{H}_2\mathbf{T}_2\right|. \end{aligned}$$

The assumption to let both users be allocated with large power is now further used to neglect the offset identities inside the determinants. With this approximation, the high SNR approximated rates then read as

$$\begin{aligned} R_1'' &= \log\left|\mathbf{T}_1^{\mathrm{H}}\mathbf{H}_1^{\mathrm{H}}(\mathbf{I}-\mathbf{H}_2\bar{\mathbf{T}}_2(\bar{\mathbf{T}}_2^{\mathrm{H}}\mathbf{H}_2^{\mathrm{H}}\mathbf{H}_2\bar{\mathbf{T}}_2)^{-1}\bar{\mathbf{T}}_2^{\mathrm{H}}\mathbf{H}_2^{\mathrm{H}})\mathbf{H}_1\mathbf{T}_1\right|, \\ R_2'' &= \log\left|\mathbf{T}_2^{\mathrm{H}}\mathbf{H}_2^{\mathrm{H}}(\mathbf{I}-\mathbf{H}_1\bar{\mathbf{T}}_1(\bar{\mathbf{T}}_1^{\mathrm{H}}\mathbf{H}_1^{\mathrm{H}}\mathbf{H}_1\bar{\mathbf{T}}_1)^{-1}\bar{\mathbf{T}}_1^{\mathrm{H}}\mathbf{H}_1^{\mathrm{H}})\mathbf{H}_2\mathbf{T}_2\right|. \end{aligned} \tag{2}$$

In the following, we will study the rate region obtained with the approximated rates from (2). The error induced by using (2) instead of (1) will be evaluated qualitatively in the simulations in Section VI. A quantitative study of the error and the derivation of an upper bound is done in [16], [17].

If user two is allocated with as many streams as he has antennas, i.e., $b_2 = r_2$, he is said to be applying *full multiplexing* (FM). The NP matrix of the user applying FM cancels out in the interference term in (2) and the only remaining occurence of the NP matrix is in the determinant of the transmit covariance matrix. It is then optimal to choose the NP matrix unitary, so that the NP matrix also cancels out there. The rates (2) then read as

$$\begin{aligned} R_1'' &= b_1\left(\log\left(\frac{P_1}{b_1}\right)+\log\left|\bar{\mathbf{T}}_1^{\mathrm{H}}\bar{\mathbf{H}}_1^{\mathrm{H}}\bar{\mathbf{H}}_1\bar{\mathbf{T}}_1\right|^{\frac{1}{b_1}}\right), \\ R_2'' &= r_2\left(\log\left(\frac{P_2}{r_2}\right)+\log\left(\left(d_2\frac{\left|\bar{\mathbf{T}}_1^{\mathrm{H}}\bar{\mathbf{H}}_1^{\mathrm{H}}\bar{\mathbf{H}}_1\bar{\mathbf{T}}_1\right|}{\left|\bar{\mathbf{T}}_1^{\mathrm{H}}\mathbf{H}_1^{\mathrm{H}}\mathbf{H}_1\bar{\mathbf{T}}_1\right|}\right)^{\frac{1}{r_2}}\right), \end{aligned} \tag{3}$$

with the *projected channels* defined as

$$\begin{aligned} \bar{\mathbf{H}}_1 &\triangleq (\mathbf{I}_t-\mathbf{H}_2(\mathbf{H}_2^{\mathrm{H}}\mathbf{H}_2)^{-1}\mathbf{H}_2^{\mathrm{H}})\mathbf{H}_1, \\ \bar{\mathbf{H}}_2 &\triangleq (\mathbf{I}_t-\mathbf{H}_1(\mathbf{H}_1^{\mathrm{H}}\mathbf{H}_1)^{-1}\mathbf{H}_1^{\mathrm{H}})\mathbf{H}_2, \end{aligned} \tag{4}$$

and $d_2 \triangleq |\mathbf{H}_2^{\mathrm{H}}\mathbf{H}_2|$. The projected channel corresponds to the orthogonal projection of the channel into the null-space of the Hermitian channel of the interfering user. Thus, it arises when the interfering user applies FM and transmits in all the directions of the channels. We can observe in (3) that the influence of the power allocation and the NP matrix are divided into two different terms. We define for each user the *rate shift* containing the influence of the NP matrix, as

$$\begin{aligned} c_{b_1,\mathrm{FM2},1}(\bar{\mathbf{T}}_1) &\triangleq \left|\bar{\mathbf{T}}_1^{\mathrm{H}}\bar{\mathbf{H}}_1^{\mathrm{H}}\bar{\mathbf{H}}_1\bar{\mathbf{T}}_1\right|^{-\frac{1}{b_1}}, \\ c_{b_1,\mathrm{FM2},2}(\bar{\mathbf{T}}_1) &\triangleq \left(d_2\frac{\left|\bar{\mathbf{T}}_1^{\mathrm{H}}\bar{\mathbf{H}}_1^{\mathrm{H}}\bar{\mathbf{H}}_1\bar{\mathbf{T}}_1\right|}{\left|\bar{\mathbf{T}}_1^{\mathrm{H}}\mathbf{H}_1^{\mathrm{H}}\mathbf{H}_1\bar{\mathbf{T}}_1\right|}\right)^{-\frac{1}{r_2}}. \end{aligned} \tag{5}$$

When user two applies FM, using (5) in (3) gives

$$\begin{aligned} R_1'' &= b_1\left(\log(P_1)-\log(b_1)-\log(c_{b_1,\mathrm{FM2},1}(\bar{\mathbf{T}}_1))\right), \\ R_2'' &= r_2\left(\log(P_2)-\log(r_2)-\log(c_{b_1,\mathrm{FM2},2}(\bar{\mathbf{T}}_1))\right). \end{aligned} \tag{6}$$

We consider that user two applies FM because it will be proven in Section III that it is optimal at high SNR to let at least one user apply FM. The symmetric case in which it is user one who transmits $b_1 = r_1$ streams can then be obtained by exchanging the indices between the users.

B. Optimization Problem

We consider the maximization of the sum rate $R'' \triangleq R_1'' + R_2''$ subject to a *rate ratio constraint*, i.e., subject to a given ratio between the rates R_1'' and R_2''. This constraint is expressed by means of the non-negative *rate coefficients* $\boldsymbol{\gamma} \triangleq (\gamma_1, \gamma_2)$ defined such as $R_1''/\gamma_1 = R_2''/\gamma_2$. The rates coefficients are normalized as $\gamma_1 + \gamma_2 = 1$, which implies that $R_1'' = \gamma_1 R''$ and $R_2'' = \gamma_2 R''$. Note that knowing γ_1, γ_2 or $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$ is equivalent. Similarly, once the rate ratio constraint is fulfilled, knowing R_1'', R_2'', or R'' is equivalent.

With these definitions, the optimization problem is

$$\underset{\bar{\mathbf{T}}_1,\bar{\mathbf{T}}_2,P_1,P_2,b_1,b_2}{\text{maximize}}\; R'' \quad \text{subject to: } \frac{R_1''}{\gamma_1}=\frac{R_2''}{\gamma_2},\; P_1+P_2=P, \quad P_i \geq 0, \quad \|\bar{\mathbf{T}}_i\|_{\mathrm{F}}^2 = b_i, \quad i=1,2. \tag{7}$$

III. Asymptotic Analysis

In this section, we consider the optimization (7) asymptotically when P tends to infinity and use the approximation

$$R_\infty(\boldsymbol{\gamma}) \triangleq \mathrm{M_G}(\boldsymbol{\gamma})\left(\log(P)-\mathrm{R_O}(\boldsymbol{\gamma})\right) \approx R''(\boldsymbol{\gamma}) \tag{8}$$

where $\mathrm{M_G}(\boldsymbol{\gamma})$ and $\mathrm{R_O}(\boldsymbol{\gamma})$ are the multiplexing gain (MG) and the rate offset (RO), respectively. We see that the MG is the most important figure of merit when P tends to infinity. However, the convergence is slow and the RO has also a large influence at finite SNR. We start by computing the maximal MG and the stream allocation to achieve it, and use the remaining degrees of freedom to maximize the RO.

A. Multiplexing Gain

The MG of the i-th user, $\mathrm{M}_{\mathrm{G}i}(\boldsymbol{\gamma})$, is defined as in (8) with $R_i''(\boldsymbol{\gamma})$ instead of $R''(\boldsymbol{\gamma})$. For given $\boldsymbol{b}$, the MG region is

$$\mathcal{R}_{\mathrm{MG}}(\boldsymbol{b}) \triangleq \{(\mathrm{M}_{\mathrm{G}1}(\boldsymbol{\gamma}), \mathrm{M}_{\mathrm{G}2}(\boldsymbol{\gamma}))|\boldsymbol{b} = (b_1, b_2)\}, \quad (9)$$

and is equal to the rate region normalized by the $\log(P)$, when P tends to infinity. The MG region with the optimal stream allocation is denoted by $\mathcal{R}_{\mathrm{MG}}$ and reads as

$$\mathcal{R}_{\mathrm{MG}} \triangleq \bigcup_{\{\boldsymbol{b}|0\leq b_1\leq r_1, 0\leq b_2\leq r_2\}} \mathcal{R}_{\mathrm{MG}}(\boldsymbol{b}). \quad (10)$$

We first study $\mathcal{R}_{\mathrm{MG}}(\boldsymbol{b})$ for a given $\boldsymbol{b}$ and use the results to obtain $\mathcal{R}_{\mathrm{MG}}$ and the *asymptotically optimal* stream allocation.

1) Multiplexing Gain Region for a fixed $\boldsymbol{b}$:

Theorem 1. *For given rate coefficients $\boldsymbol{\gamma}$ and a stream allocation $\boldsymbol{b}$, only one user has a power allocation scaling linearly in P. He is called the* limiting user *and denoted as user ℓ, while the other one is denoted as user $n\ell$ and has a sub-linear power allocation, i.e., scaling with P raised to the exponent $b_\ell\gamma_{n\ell}/(b_{n\ell}\gamma_\ell) < 1$ The limiting user is the user with the largest quotient γ_i/b_i and is hence user two if $\gamma_1 < b_1/(b_1+b_2)$, and user one if $\gamma_1 > b_1/(b_1+b_2)$. The rate coefficients corresponding to the transition between the two cases are called the* transition coefficients of the stream allocation $\boldsymbol{b}$ *and are denoted by $\boldsymbol{\gamma}_{\mathrm{tr}}(\boldsymbol{b}) \triangleq (\gamma_{1,\mathrm{tr}}(\boldsymbol{b}), \gamma_{2,\mathrm{tr}}(\boldsymbol{b})) \triangleq (b_1/(b_1+b_2), b_2/(b_1+b_2))$. For given $\boldsymbol{\gamma}$, the MG reads as:*

$$\begin{aligned} &\textit{if } \gamma_1 < \gamma_{1,\mathrm{tr}}(\boldsymbol{b}): \quad M_G(\boldsymbol{\gamma}) = \frac{b_2}{\gamma_2}, \\ &\textit{if } \gamma_1 = \gamma_{1,\mathrm{tr}}(\boldsymbol{b}): \quad M_G(\boldsymbol{\gamma}) = b_1 + b_2, \\ &\textit{if } \gamma_1 > \gamma_{1,\mathrm{tr}}(\boldsymbol{b}): \quad M_G(\boldsymbol{\gamma}) = \frac{b_1}{\gamma_1}. \end{aligned}$$

Proof: The MG does not depend on the rate shifts, which we thus denote simply as c_1 and c_2. To be perfectly rigorous, the definition of the rate shifts should be extended to an arbitrary stream allocation $\boldsymbol{b}$ (see [14]). The rate expressions are then exactly the same as (6) only with b_2 instead of r_2, and with the rate shifts depending also on $\bar{\boldsymbol{T}}_2$. Using this generalization of (6), the rate ratio constraint reads as

$$\frac{R_1''}{\gamma_1} = \frac{R_2''}{\gamma_2} \quad \Leftrightarrow \quad P_2 = b_2 c_2 \left(\frac{P_1}{b_1 c_1}\right)^{\frac{b_1\gamma_2}{\gamma_1 b_2}}. \quad (11)$$

Inserting (11) into the sum power constraint gives

$$P_1 + P_2 = P_1 + b_2 c_2 \left(\frac{P_1}{b_1 c_1}\right)^{\frac{b_1\gamma_2}{\gamma_1 b_2}} = P. \quad (12)$$

Considering w.l.o.g. that $\gamma_1 > \gamma_{1,\mathrm{tr}}(\boldsymbol{b})$, it implies that $\gamma_2 < \gamma_{2,\mathrm{tr}}(\boldsymbol{b})$ and thus $\frac{b_1\gamma_2}{\gamma_1 b_2} < 1$. Letting P tend to infinity in (12), P_1 tends also to infinity such that the term with P_1 raised to the largest exponent is dominant and P_1 scales linearly with P. The scaling of P_2 and the MG then follow. ■

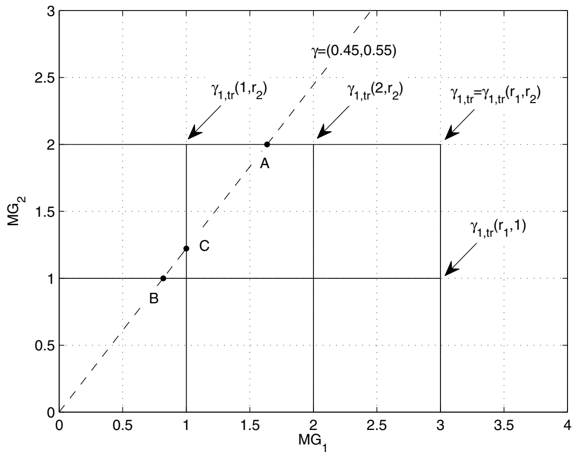

Fig. 1. MG regions for all the stream allocations with the antenna configuration $\boldsymbol{r} = (3, 2)$.

Geometrical interpretation: The MG of one user is fixed and the MG of the other user varies with $\boldsymbol{\gamma}$. The case $\gamma_1 = \gamma_{1,\mathrm{tr}}(\boldsymbol{b})$ corresponds to user one and two being both limiting users. Thus, the boundary of the MG region is a rectangle with b_1 as horizontal dimension and b_2 as vertical dimension, i.e., $\mathcal{R}_{\mathrm{MG}}(\boldsymbol{b}) = \{(\mathrm{M}_{\mathrm{G}1}, \mathrm{M}_{\mathrm{G}2})|0 \leq \mathrm{M}_{\mathrm{G}1} \leq b_1, 0 \leq \mathrm{M}_{\mathrm{G}2} \leq b_2\}$. The vertex corresponds to the transition coefficients $\boldsymbol{\gamma}_{\mathrm{tr}}(\boldsymbol{b})$ and is the only point at which both users have a power allocation scaling linearly in P. If the ray starting from the origin associated with a given $\boldsymbol{\gamma}$ intersects the rectangle on the vertical part, user one is the limiting user, and if it is on the horizontal part, user two is then the limiting user.

2) The Multiplexing Gain Region $\mathcal{R}_{MG}$:

Theorem 2. *For given $\boldsymbol{\gamma}$, the maximal $M_G(\boldsymbol{\gamma})$ and the stream allocation $\boldsymbol{b} = (b_1, b_2)$ to achieve it read as follows:*

$$\begin{aligned} &\textit{For } \gamma_1 \leq \gamma_{1,\mathrm{tr}}\textit{: } M_G(\boldsymbol{\gamma}) = \frac{r_2}{\gamma_2}, \textit{ if: } b_1 \geq \left\lceil \frac{\gamma_1 r_2}{\gamma_2} \right\rceil, b_2 = r_2 \\ &\textit{For } \gamma_1 \geq \gamma_{1,\mathrm{tr}}\textit{: } M_G(\boldsymbol{\gamma}) = \frac{r_1}{\gamma_1}, \textit{ if: } b_2 \geq \left\lceil \frac{\gamma_2 r_1}{\gamma_1} \right\rceil, b_1 = r_1 \end{aligned}$$

with $\boldsymbol{\gamma}_{\mathrm{tr}} \triangleq (\gamma_{1,\mathrm{tr}}, \gamma_{2,\mathrm{tr}}) \triangleq \boldsymbol{\gamma}_{\mathrm{tr}}(\boldsymbol{r})$.

Proof: Consider w.l.o.g. $\boldsymbol{b}$ and $\boldsymbol{\gamma}$ given so that user two is the limiting user if both users apply FM. From Theorem 1, the MG is b_2/γ_2 and is maximized with $b_2 = r_2$ if user two remains the limiting user, i.e., if $r_2/\gamma_2 < b_1/\gamma_1$. ■

Geometrical interpretation: We have plotted in Fig. 1 the MG regions for all the stream allocations when $\boldsymbol{r} = (3, 2)$. Using Theorem 1, the MG region $\mathcal{R}_{\mathrm{MG}}$ is the rectangle of dimensions $r_1 \times r_2$ and equal to $\mathcal{R}_{\mathrm{MG}}(\boldsymbol{r})$. For example, if we chose $\boldsymbol{\gamma} = (0.45, 0.55)$, the maximal MG is achieved at point A in Fig. 1. If user two transmits only one stream, the point B is reached and the maximal MG is not achieved, thus user two has to apply FM. Similarly, if user one transmits only one stream, point C is then reached.

B. The Rate Offset

The study of the MG has led to necessary conditions for the stream allocation $\boldsymbol{b}$ to achieve the maximum MG, but not defined $\boldsymbol{b}$ completely. Moreover, the coefficients of the NP matrix $\bar{\mathbf{T}}_1$ are also free to be optimized.

Theorem 3. *For $\gamma_1 \le \gamma_{1,tr}$, the asymptotically optimal stream allocation is $\boldsymbol{b} = (\lceil \gamma_1 r_2/\gamma_2 \rceil, r_2)$ and $R_\infty(\boldsymbol{\gamma})$ reads as*

$$R_\infty(\boldsymbol{\gamma}) = \frac{r_2}{\gamma_2}\left(\log\left(\frac{P}{r_2}\right) + \frac{1}{r_2}\left(\log(d_2) + \sum_{i=1}^{b_1}\log(\lambda_{1,\mathrm{inv},i})\right)\right),$$

where $d_2 \triangleq |\mathbf{H}_2^{\mathrm{H}}\mathbf{H}_2|$ and $\lambda_{1,\mathrm{inv},i}$ is the i-th largest eigenvalue of the matrix pencil $(\bar{\mathbf{H}}_1^{\mathrm{H}}\bar{\mathbf{H}}_1, \mathbf{H}_1^{\mathrm{H}}\mathbf{H}_1)$.

For $\gamma_1 \ge \gamma_{1,tr}$, the asymptotically optimal stream allocation is $\boldsymbol{b} = (r_1, \lceil \gamma_2 r_1/\gamma_1 \rceil)$ and $R_\infty(\boldsymbol{\gamma})$ reads as

$$R_\infty(\boldsymbol{\gamma}) = \frac{r_1}{\gamma_1}\left(\log\left(\frac{P}{r_1}\right) + \frac{1}{r_1}\left(\log(d_1) + \sum_{i=1}^{b_2}\log(\lambda_{2,\mathrm{inv},i})\right)\right),$$

where $d_1 \triangleq |\mathbf{H}_1^{\mathrm{H}}\mathbf{H}_1|$ and $\lambda_{2,\mathrm{inv},i}$ is the i-th largest eigenvalue of the matrix pencil $(\bar{\mathbf{H}}_2^{\mathrm{H}}\bar{\mathbf{H}}_2, \mathbf{H}_2^{\mathrm{H}}\mathbf{H}_2)$.

Proof: We consider w.l.o.g. that $\gamma_1 < \gamma_{1,\mathrm{tr}}$ such that user two is the limiting user. From Theorem 2, we know that user two has to apply FM and user one has to transmit $b_1 \ge \left\lceil \frac{\gamma_1 r_2}{\gamma_2} \right\rceil$ streams in order to achieve the maximal MG. Thus, we use the rate expressions (6) when user two is the limiting user with the rate ratio constraint fulfilled:

$$R'' = \frac{R_2''}{\gamma_2} = \frac{r_2}{\gamma_2}\left(\log\left(\frac{P}{r_2(c_{b_1,\mathrm{FM2},2}(\bar{\mathbf{T}}_1))}\right) + \log\left(1 - \frac{P_1}{P}\right)\right). \tag{13}$$

The power allocation P_1 scales sub-linearly with P such that P_1/P tends to zero when P tends to infinity. Using (5), the asymptotic optimal NP matrix, denoted as $\bar{\mathbf{T}}_{1,\mathrm{inv}}$, is then made of the b_1 principal eigenvectors of the matrix pencil $(\bar{\mathbf{H}}_1^{\mathrm{H}}\bar{\mathbf{H}}_1, \mathbf{H}_1^{\mathrm{H}}\mathbf{H}_1)$. We know $\forall i, \lambda_{2,\mathrm{inv},i} \le 1$ due to the orthogonal projection which ends the proof. ■

C. Interpretation

The parameters of the high SNR approximation (8) and the corresponding stream allocations have been derived in Theorems 2 and 3. We have shown that it is not necessarily optimal to let both users apply FM when maximizing the sum rate subject to a given rate ratio constraint, i.e., along a ray in the rate region. This is a new result and is in contrast to the weighted sum rate maximization in which case both users should apply FM to maximize the MG [9]. Both power allocations are then linear in P, and the points reached in the rate region converge to $\boldsymbol{\gamma}_{\mathrm{tr}}$ when P tends to infinity.

IV. Analytical Bounds

Assuming w.l.o.g. that user two applies FM, we will now derive analytical bounds for the high SNR approximated rate region for a given stream allocation $\boldsymbol{b}$, denoted as $\mathcal{R}''(\boldsymbol{b})$

$$\mathcal{R}''(\boldsymbol{b}) \triangleq \{(R_1'', R_2'')|P_1 + P_2 = P, \boldsymbol{b} = (b_1, r_2)\}. \tag{14}$$

Since $\boldsymbol{b}$ is fixed, we use in the following the shorthanded notation c_1 and c_2 for $c_{b_1,\mathrm{FM2},1}$ and $c_{b_1,\mathrm{FM2},2}$, respectively.

A. Outer Bound

We denote as $\mathcal{R}_{\mathrm{R_O}}(\boldsymbol{b})$ the rate region whose boundary is described by the rate pair $(R_{1,\mathrm{R_O}}(\boldsymbol{b}), R_{2,\mathrm{R_O}}(\boldsymbol{b}))$ when the rate ratio constraint $R_{1,\mathrm{R_O}}(\boldsymbol{b})/\gamma_1 = R_{2,\mathrm{R_O}}(\boldsymbol{b})/\gamma_2$ is fulfilled. One of the two following equations is valid depending on γ_1:

$$\gamma_1 \ge \gamma_{1,\mathrm{R_O},\mathrm{tr}}(\boldsymbol{b}): R_{1,\mathrm{R_O}}(\boldsymbol{b}) \triangleq b_1 \log\left(\frac{P}{b_1}\right) + \sum_{i=1}^{b_1}\log(\lambda_{1,\mathrm{proj},i}),$$

$$\gamma_1 \le \gamma_{1,\mathrm{R_O},\mathrm{tr}}(\boldsymbol{b}): R_{2,\mathrm{R_O}}(\boldsymbol{b}) \triangleq r_2 \log\left(\frac{P d_2^{\frac{1}{r_2}}}{r_2}\right) + \sum_{i=1}^{b_1}\log(\lambda_{1,\mathrm{inv},i}),$$

with $\gamma_{1,\mathrm{R_O},\mathrm{tr}}(\boldsymbol{b}) \triangleq R_{1,\mathrm{R_O}}(\boldsymbol{b})/(R_{1,\mathrm{R_O}}(\boldsymbol{b}) + R_{2,\mathrm{R_O}}(\boldsymbol{b}))$ and $\lambda_{1,\mathrm{proj},i}$ the i-th largest eigenvalue of $\bar{\mathbf{H}}_1^{\mathrm{H}}\bar{\mathbf{H}}_1$. Note that $P_1 = P$ in the first line and $P_2 = P$ in the second line.

Theorem 4. *The rate region $\mathcal{R}_{R_O}(\boldsymbol{b})$ is an outer bound for $\mathcal{R}''(\boldsymbol{b})$ at arbitrary SNR, and is asymptotically tight.*

Proof: We consider asymptotically (13) and write the rate when user one is the limiting user (but user two is still the one applying FM) to obtain:

$$R_{1,\mathrm{R_O}}(\boldsymbol{b}) = b_1\left(\log\left(\frac{P}{b_1}\right) - \log(c_1(\bar{\mathbf{T}}_1))\right), \text{if } \gamma_1 \ge \gamma_{1,\mathrm{tr}}(\boldsymbol{b}),$$

$$R_{2,\mathrm{R_O}}(\boldsymbol{b}) = r_2\left(\log\left(\frac{P}{r_2}\right) - \log(c_2(\bar{\mathbf{T}}_1))\right), \text{if } \gamma_1 \le \gamma_{1,\mathrm{tr}}(\boldsymbol{b}).$$

In both cases, the constraint $R_{1,\mathrm{R_O}}(\boldsymbol{b})/\gamma_1 = R_{2,\mathrm{R_O}}(\boldsymbol{b})/\gamma_2$ is fulfilled by the power allocation (11) so that only the rate expression of one user needs to be optimized and the other one follows from the rate ratio constraint. Using (5), the optimal NP matrices and the resulting expressions are: For $\gamma_1 \ge \gamma_{1,\mathrm{tr}}(\boldsymbol{b})$, the NP matrix maximizing $R_{1,\mathrm{R_O}}(\boldsymbol{b})$ made of the b_1 principal eigenvectors of $\bar{\mathbf{H}}_1^{\mathrm{H}}\bar{\mathbf{H}}_1$, and denoted as $\bar{\mathbf{T}}_{1,\mathrm{proj}}$, and for $\gamma_1 \le \gamma_{1,\mathrm{tr}}(\boldsymbol{b})$ the NP matrix maximizing $R_{2,\mathrm{R_O}}(\boldsymbol{b})$ denoted as $\bar{\mathbf{T}}_{1,\mathrm{inv}}$. We introduce $\gamma_{1,\mathrm{R_O},\mathrm{tr}}(\boldsymbol{b})$ to avoid a discontinuity at $\gamma_{1,\mathrm{tr}}(\boldsymbol{b})$ at finite SNR, but we still obtain an upper bound since the limiting user is allocated with P and the rate shift is optimal. The asymptotic tightness of the bounds follows from Theorem 3 after noticing that $\gamma_{1,\mathrm{R_O},\mathrm{tr}}(\boldsymbol{b})$ tends to $\gamma_{1,\mathrm{tr}}(\boldsymbol{b})$ when P tends to infinity. ■

For fixed P, the rate of the limiting user in $\mathcal{R}_{\mathrm{R_O}}(\boldsymbol{b})$ is a constant given by Theorem 4, while the rate of the other user is not constant and is given by $R_1''/\gamma_1 = R_2''/\gamma_2$ such that the boundary of $\mathcal{R}_{\mathrm{R_O}}(\boldsymbol{b})$ is a rectangle. $\mathcal{R}''(\boldsymbol{b})$ converges to $\mathcal{R}_{\mathrm{R_O}}(\boldsymbol{b})$ as P tends to infinity, but for finite P, it corresponds to $\mathcal{R}_{\mathrm{R_O}}(\boldsymbol{b})$ with a rounded down corner instead of the sharp edge. Indeed, the corner, located at $\gamma_{1,\mathrm{R_O},\mathrm{tr}}(\boldsymbol{b})$, is a point at which both users consume a power equal to P. Thus, the corner is never reached, but the domain of the rounded down corner has a measure tending to zero when P tends to infinity.

B. Inner Bound

The derivation of the inner bound is based on the study of the rates at the transition coefficient $\gamma_{1,\mathrm{tr}}(\boldsymbol{b})$. The power allocation solving the rate ratio constraint at $\gamma_{1,\text{tr}}(\boldsymbol{b})$ is

$$P_1=\frac{b_1c_1(\bar{\mathbf{T}}_1)}{b_1c_1(\bar{\mathbf{T}}_1)+b_2c_2(\bar{\mathbf{T}}_1)}P, P_2=\frac{b_2c_2(\bar{\mathbf{T}}_1)}{b_1c_1(\bar{\mathbf{T}}_1)+b_2c_2(\bar{\mathbf{T}}_1)}P \tag{15}$$

which yields a sum rate R'', denoted as $R''_{\text{tr}}(\boldsymbol{b})$, equal to

$$R''_{\text{tr}}(\boldsymbol{b})\triangleq(b_1+r_2)\left(\log(P)-\log\left(b_1c_1(\bar{\mathbf{T}}_1)+b_2c_2(\bar{\mathbf{T}}_1)\right)\right). \tag{16}$$

The optimization of $\bar{\mathbf{T}}_1$ to maximize $R''_{\text{tr}}(\boldsymbol{b})$ is intricate but we want to obtain a lower bound so that we can use any full rank NP matrix. We chose $\bar{\mathbf{T}}_{1,\text{inv}}$ and define $c_{\text{tr}}(\boldsymbol{b})\triangleq b_1c_1(\bar{\mathbf{T}}_{1,\text{inv}})+b_2c_2(\bar{\mathbf{T}}_{1,\text{inv}})$. $\mathcal{R}_{\text{R}_\text{O},\text{LB}}(\boldsymbol{b})$ is defined as the region whose boundary is made of the rate pair $(R_{1,\text{R}_\text{O},\text{LB}}(\boldsymbol{b}),R_{2,\text{R}_\text{O},\text{LB}}(\boldsymbol{b}))$ fulfilling $R_{1,\text{R}_\text{O},\text{LB}}(\boldsymbol{b})/\gamma_1=R_{2,\text{R}_\text{O},\text{LB}}(\boldsymbol{b})/\gamma_2$ and one of these equations, depending on γ_1:

$$\begin{aligned} R_{1,\text{R}_\text{O},\text{LB}}(\boldsymbol{b})&\triangleq\gamma_{1,\text{tr}}(\boldsymbol{b})R''_{\text{tr}}(\boldsymbol{b}) \quad ,\text{if } \gamma_1\geq\gamma_{1,\text{tr}}(\boldsymbol{b}),\\ R_{2,\text{R}_\text{O},\text{LB}}(\boldsymbol{b})&\triangleq\gamma_{2,\text{tr}}(\boldsymbol{b})R''_{\text{tr}}(\boldsymbol{b}) \quad ,\text{if } \gamma_1\leq\gamma_{1,\text{tr}}(\boldsymbol{b}).\end{aligned} \tag{17}$$

By construction and monotonicity arguments of R''_1 and R''_2, it is clear that $\mathcal{R}_{\text{R}_\text{O},\text{LB}}(\boldsymbol{b})$ is an inner bound for the rate region $\mathcal{R}''(\boldsymbol{b})$. The domain $\mathcal{R}_{\text{R}_\text{O},\text{LB}}(\boldsymbol{b})$ is also a rectangle with the dimensions given by (17) and the vertex at $\gamma_{1,\text{tr}}(\boldsymbol{b})$.

V. The Rate Region at Finite SNR

We now consider the approximated rate R'' at high but finite SNR. The region $\mathcal{R}''$ is defined as the union of the $\mathcal{R}''(\boldsymbol{b})$ over all $\boldsymbol{b}$. We assume w.l.o.g. user two to be the limiting user and to apply FM as a consequence of the asymptotic analysis in Section III. We can then use the rate expressions from (6). Optimization (7) is hard to handle directly, even with the simplified rate expressions (6) because the discrete stream allocation $\boldsymbol{b}$ has to be optimized. Therefore, we consider the stream allocation $\boldsymbol{b}$ to be fixed and study then (7). Comparing the sum rate R'' obtained for each stream allocation then leads to the rate region $\mathcal{R}''$.

A. Double Full Multiplexing case: $\boldsymbol{b}=(r_1,r_2)$

We assume now that both users apply FM and call it the *Double Full Multiplexing* case (DFM). The NP matrices are then square such that the rate shifts read as [cf. (5)]

$$\begin{aligned} c_{1,\text{DFM}}&\triangleq c_{r_1,\text{FM2},1}(\mathbf{I}_{r_1})=\left|\bar{\mathbf{H}}_1^{\text{H}}\bar{\mathbf{H}}_1\right|^{-\frac{1}{r_1}},\\ c_{2,\text{DFM}}&\triangleq c_{r_1,\text{FM2},2}(\mathbf{I}_{r_1})=\left|\bar{\mathbf{H}}_2^{\text{H}}\bar{\mathbf{H}}_2\right|^{-\frac{1}{r_2}}.\end{aligned} \tag{18}$$

Only the power allocation remains to be chosen, and since the power allocation satisfying the rate ratio constraint is unique, it means that we only need to solve the rate ratio constraint in P_1 (or P_2). No closed form solution can be found to solve the exact power allocation but it is very easy to obtain numerical solutions via convex optimization. An approximate closed form solution was also derived in [16].

B. The Multi Beamforming Stream Allocation

We now consider the *multi beamforming* stream allocation in which only the limiting user applies FM, while the other user transmits $0\leq b_1<r_1$ streams.

In the *beamforming case* (BF) case, when $b_1=1$, it is possible to apply a method presented in [16], [18] and called *Fixed Coordinate Rate Maximization*. This method allows for the derivation of more accurate lower bounds and of an algorithm converging almost surely to a local maximum of the sum rate. Furthermore, the high SNR approximation for the BF case can be avoided for the user applying BF, thanks to the simplified form of the rate. This reduces the high SNR approximation error considerably and is particularly interesting because the BF stream allocation has been shown in Section III to be asymptotically Pareto optimal for the region close to the axes, i.e., exactly on the part of the rate region where the rate of one user is small and the high SNR approximation error is significant. This is shown in the simulations in Section VI and studied in detail in [16], [17].

The NP matrix $\bar{\mathbf{T}}_1$ needs now to be optimized to maximize the sum rate R''. However, reaching the optimal NP matrix is intricate because of the interdependency between the NP matrix and the power allocation due to the requirement to fulfill the rate ratio constraint. Therefore, we focus on the derivation of bounds for R''. The idea is to optimize the rate shifts given in (5) first and then consider the power allocation. Indeed, once the NP matrix (from which the rate shifts follow) is given, the power allocation can be found exactly as in the DFM case discussed in Subsection V-A. From (5), it is easily shown that the projected NP matrix $\bar{\mathbf{T}}_{1,\text{proj}}$ is optimal for the rate shift of user one, and the inverse NP matrix $\bar{\mathbf{T}}_{1,\text{inv}}$ for the rate shift of user two.

Finally, we define the *virtual* rate shifts as the pair of rate shifts made of the maximal values of the rate shift of user one and user two, i.e., $(c_{1,\text{virt}}(b_1),c_{2,\text{virt}}(b_1))\triangleq(c_{b_1,\text{FM2},1}(\bar{\mathbf{T}}_{1,\text{proj}}),c_{b_1,\text{FM2},2}(\bar{\mathbf{T}}_{1,\text{inv}}))$. Using these rate shifts leads to an upper bound for the rate region boundary.

Theorem 5. *When P is large enough and γ_1 tends to zero or one, or when P tends to infinity, the NP matrix optimizing the rate of the limiting user is optimal. The upper bound obtained with the virtual rate shift is tight for all these cases.*

Proof: The scaling in P of the power allocations fulfilling the rate ratio constraint as given in Theorem 1 inserted in the rate expressions (6) leads to the results. ∎

VI. Simulations

In Fig. 2, we have plotted the boundary of the rate region obtained when using the two types of suboptimal rate shifts and the virtual rate shifts [cf. subsectionV-B] in the approximated rate expressions (2) first and then in the exact ones (1). At every point of the boundary, one of the two lower bounds is very close to the upper bound. As shown theoretically, it is the lower bound optimizing the rate shift of the limiting user which is tight when the rate coefficient tends to zero or. The comparison between the rate from (1) and (2) allows us to evaluate the accuracy of the high SNR approximation, which means to observe qualitatively the error induced by neglecting the identities in Section II. The approximation error is significant only close to the axes. A trust domain in which the high SNR approximation error

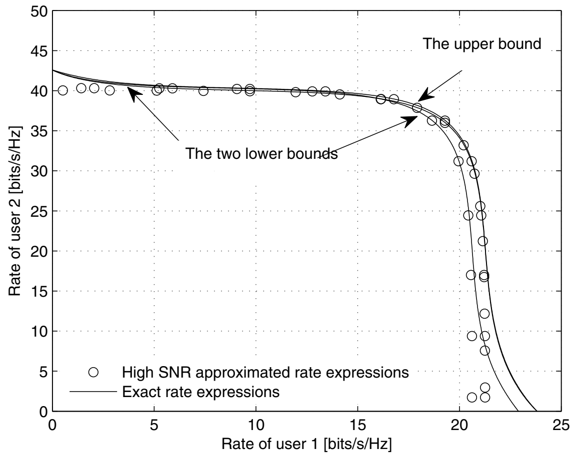

Fig. 2. Boundary of the rate region with $\boldsymbol{r} = (4, 4)$, $t = 8$, and $P = 30$ dB for the stream allocation $\boldsymbol{b} = (3, 4)$ with the exact/approximated power allocation in the exact/approximated rate expressions.

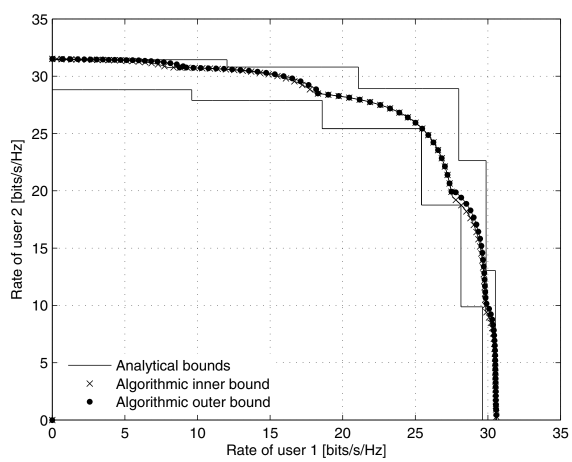

Fig. 3. Boundary of the algorithmic/analytical inner and outer bounds for the high SNR approximated rate region with $\boldsymbol{r} = (3, 3)$, $t = 6$, and $P = 30$ dB.

is smaller than a given threshold is derived in [17], and it is also shown that the problem of having the observed large deviation close to the axes can be solved thanks to the improved approximation for the BF case, such that the rate region is obtained in its totality with a very good accuracy.

In Fig. 3, we have plotted the boundary of the analytical [cf. Section IV] and algorithmic [cf. Section V] inner and outer bounds for the high SNR approximated rate region. The algorithmic bounds are once more very close to each other, and the algorithmic bounds fit inside the analytical ones. Furthermore, the analytical ones are not very accurate but still follow the main properties of the boundary and can be used to get a good idea of the general shape of the rate region, and this with extremely simple closed form expressions.

VII. Conclusion

In this work, the rate region of the two user BC has been studied in the high SNR regime when linear precoding is applied. We have characterized precisely the boundary of the high SNR approximated rate region by deriving first the asymptotic optimal parameters, and then by developing accurate bounds at finite SNR. The high SNR approximation error is shown in [16], [17] to be small for $P = 30$ dB and the results give a good insight into the rate region at intermediate SNR. We have developed new tools and a new method to improve the accuracy of the high SNR approximation which could be used for other high SNR analysis. It also appears possible to extend the general idea to the K user case. Finally, most of the results are given in terms of the eigenvalues of Wishart matrices, and can be easily evaluated in fading channels (e.g. Rayleigh or Rician channels).

References

[1] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates, and Sum-rate Capacity of Gaussian MIMO Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, Oct. 2003.

[2] H. Weingarten, S. Steinberg, and S. Shamai, "The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel," *IEEE Transactions on Information Theory*, vol. 52, no. 9, Sept. 2006.

[3] R. Hunger, D. Schmidt, M. Joham, and W. Utschick, "A General Covariance-Based Optimization Framework using Orthogonal Projections," in *Proc. SPAWC*, 2008.

[4] J. Liu, Y. T. Hou, and H. D. Sherali, "On the Maximum Weighted Sum-rate of MIMO Gaussian Broadcast Channels," in *Proc. ICC*, 2008.

[5] U. Erez and S. ten Brink, "A Close-to-capacity DirtyPaper Coding Scheme," *IEEE Transactions on Information Theory*, vol. 51, no. 10, Oct. 2005.

[6] R. Hunger, D. Schmidt, and M. Joham, "A Combinatorial Approach to Maximizing the Sum Rate in the MIMO BC with Linear Precoding," in *Proc. Asilomar*, 2008.

[7] S. Shi, M. Schubert, and H. Boche, "Rate Optimization for Multiuser MIMO systems with Linear Processing," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, Aug. 2008.

[8] J. Lee and N. Jindal, "High SNR Analysis for MIMO Broadcast Channels: Dirty Paper Coding versus Linear Precoding," *IEEE Transactions on Information Theory*, vol. 53, no. 12, Dec. 2007.

[9] R. Hunger and M. Joham, "An Asymptotic Analysis of the MIMO Broadcast Channel under Linear Filtering," in *Proc. CISS*, 2009.

[10] M. Mohseni, R. Zhang, and J. Cioffi, "Optimized Transmission for Fading Multiple-Access and Broadcast Channels with Multiple Antennas," *IEEE Journal on Selected Areas in Communications,*, vol. 24, no. 8, pp. 1627 –1639, 2006.

[11] J. Lee and N. Jindal, "Symmetric Capacity of MIMO downlink channels," in *Proc. ISIT*, 2006.

[12] M. Stojnic, H. Vikalo, and B. Hassibi, "Rate Maximization in Multi-antenna Broadcast Channels with Linear Preprocessing," in *Proc. GLOBECOM*, vol. 6, nov. 2004, pp. 3957 – 3961 Vol.6.

[13] S. Shi, M. Schubert, and H. Boche, "Capacity Balancing for Multiuser MIMO Systems," in *Proc. ICASSP*, 2007.

[14] P. de Kerret, M. Joham, W. Utschick, and R. Mathar, "MIMO Broadcast Channel Rate Region with Linear Precoding at High SNR without Full Multiplexing," in *Proc. ISITA 2010.*

[15] R. Hunger and M. Joham, "A General Rate Duality of the MIMO Multiple Access Channel and the MIMO Broadcast Channel," in *Proc. GLOBECOM*, 2008.

[16] P. de Kerret, "The Geometry of the High-power MIMO BC Rate Region under Linear Filtering," Master's thesis, Technische Universität München (TUM), 2009.

[17] P. de Kerret, R. Hunger, M. Joham, W. Utschick, and R. Mathar, "Analysis of the two user MIMO-BC rate region with linear precoding at high SNR," Draft.

[18] R. Hunger, P. de Kerret, and M. Joham, "The Geometry of the MIMO Broadcast Channel Rate Region under Linear Filtering at High SNR," in *Proc. Asilomar*, 2009.