An Asymptotic Analysis of the MIMO Broadcast Channel under Linear Filtering

Raphael Hunger and Michael Joham
Associate Institute for Signal Processing, Technische Universität München, 80290 Munich, Germany
Telephone: +49 89 289-28508, Fax: +49 89 289-28504, Email: hunger@tum.de

Abstract— **We investigate the MIMO broadcast channel with multi-antenna terminals in the high SNR regime when linear filtering is applied instead of dirty paper coding. Using a recent rate duality where the streams of every single user are not treated as self-interference as in the hitherto existing stream-wise rate dualities for linear filtering, we solve the weighted sum rate maximization problem of the broadcast channel in the dual multiple access channel. Thus, we can exactly quantify the asymptotic rate loss of linear filtering compared to dirty paper coding for any channel realization. We come up with the first rate loss expression that only depends on the channel matrices of all users and not on the precoders of the users as hitherto existing results for multi-antenna terminals do. Having converted the optimum covariance matrices to the broadcast channel by means of the duality, we observe that the optimal covariance matrices in the broadcast channel feature quite complicated but still closed form expressions although the respective transmit covariance matrices in the dual multiple access channel share a very simple structure. We immediately come to the conclusion that block-diagonalization is the asymptotically optimum transmit strategy in the broadcast channel. Out of the set of block-diagonalizing precoders, we present the one which achieves the largest sum rate and thus corresponds to the optimum solution found in the dual multiple access channel. Additionally, we quantify the ergodic rate loss of linear coding compared to dirty paper coding for Gaussian channels with correlations at the mobiles.**

I. INTRODUCTION

While the sum capacity of the single-user MIMO point-to-point link can be expressed semi-analytically in closed form [1], the simplest multi-user setup with single antenna terminals already allows for the presumption that this will remain infeasible in the *broadcast channel* (BC) and *multiple access channel* (MAC) irrespective of whether linear or nonlinear filtering is considered. Fortunately, the high signal-to-noise ratio regime is an exception to this deflating circumstance, since there, asymptotic results on the sum capacity have been discovered for dirty paper coding and partly for linear filtering.

A. Literature Overview

The single user point-to-point MIMO case was treated in [2], [3], where the *Grant-Gauthier* lower bound on the mutual information, that becomes asymptotically tight, was decomposed into a supremum capacity term, an instantaneous SNR effect term, and an instantaneous capacity degradation term due to the eigenvalue spread. Outage capacity and throughput of a fading point-to-point MIMO system are analyzed in [4]. The first high-SNR sum capacity analysis of the point-to-multipoint broadcast channel appeared in [5], where single-antenna receivers were considered. Therein, the affine approximation of the sum capacity introduced in [6] and elaborately discussed in [7] was utilized. First, [5] shows that the single-antenna broadcast channel has the same asymptotic sum-capacity as the corresponding point-to-point MIMO link with cooperating receive antennas, and second, how the power offset term in the broadcast channel looks like. Furthermore, the instantaneous and ergodic spectral efficiency loss of linear zero-forcing beamforming with respect to *dirty-paper-coding* (DPC) was derived in [5], again for single antenna receivers. The extension to multi-antenna receivers was presented in [8], [9], where the asymptotic equivalence of the nonlinear dirty paper coding sum capacity in the broadcast channel and the sum capacity of the equivalent point-to-point MIMO link with cooperating receivers was proven to hold in the multi-antenna case. Out of the class of linear precoding schemes, zero-forcing and block-diagonalization are considered. While the asymptotic rate loss of zero-forcing filtering in [5] features a closed-form solution that only depends on the channels of the users, the asymptotic rate loss expression of block-diagonalization with respect to DPC in [8, Eq. (21)] and [9, Eq. (13)] is given as a function of the channels *and* the precoders of the users. In this paper, we find the *optimum* precoders that minimize the rate loss such that the resulting rate loss expression *only* depends on the channel matrices of the users. Since the rate loss in [8], [9] still depends on the precoders, *ergodic* statements for the asymptotic sum rate and the asymptotic rate loss with respect to dirty paper coding require the assumption that block-diagonalization and equal power loading are asymptotically optimal and require a very special fading model as key prerequisite for the presented results. In contrast, our precoder free expressions allows for an arbitrary fading model in principle. Finally, it is neither known yet, whether block-diagonalization is the asymptotically optimum transmission strategy or not in the broadcast channel when linear filtering is considered, nor how the optimum block-diagonalizing precoder looks like.

B. Contributions

The main contributions of this paper are summarized in the following list:

1) The derivation of the maximum weighted sum rate asymptotically achievable with linear filtering.
2) A precoder-free closed form expression for the asymptotic rate loss of linear filtering with respect to dirty

paper coding for multi-antenna terminals that only depends on the channels of the users.

3) A closed form solution of the covariance matrices in the dual uplink achieving this maximum weighted sum rate.
4) We prove, that block diagonalization is asymptotically optimum in the broadcast channel.
5) Finally, we derive the optimum precoding and transmit covariance matrices in the broadcast channel by means of our recent rate duality in [10].

II. System Model

We consider the communication between an N antenna base station and K multi antenna terminals, where user k multiplexes B_k data streams over his r_k antennas. For a short notation, we define r as the sum of all antennas at the terminals, i.e., $r = \sum_{k=1}^{K} r_k$, and b as the sum of all transmitted streams, i.e., $b = \sum_{k=1}^{K} B_k$. In the MAC, user k applies a precoding matrix $\boldsymbol{T}_k \in \mathbb{C}^{r_k \times B_k}$ generating his $r_k \times r_k$ transmit covariance matrix $\boldsymbol{Q}_k = \boldsymbol{T}_k \boldsymbol{T}_k^{\mathrm{H}}$. The precoded symbol vector propagates over the channel described by the matrix $\boldsymbol{H}_k \in \mathbb{C}^{N \times r_k}$. At the receiver side, zero-mean noise $\boldsymbol{\eta} \in \mathbb{C}^N$ with identity covariance matrix is added and the receive filter for user k is denoted by $\boldsymbol{G}_k \in \mathbb{C}^{B_k \times N}$. Due to the reversed signal flow in the BC, we characterize the transmission from the base station to terminal k by the Hermitian channel $\boldsymbol{H}_k^{\mathrm{H}}$ in the BC, the precoder dedicated to the B_k streams of user k is denoted by $\boldsymbol{P}_k \in \mathbb{C}^{N \times B_k}$, and zero-mean noise $\boldsymbol{\eta}_k \in \mathbb{C}^{r_k}$ with identity covariance matrix is added at user k. Throughout this paper, we assume that the base station has at least as many antennas as the terminals have in sum, i.e., $N \geq r$.

III. Optimum Signaling in the Dual MAC

Introducing the composite channel matrix $\boldsymbol{H}$ and the composite block-diagonal precoder matrix $\boldsymbol{T}$ of all K users via

$$\begin{aligned} \boldsymbol{H} &= [\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K] \in \mathbb{C}^{N \times r}, \\ \boldsymbol{T} &= \mathbf{blockdiag}\{\boldsymbol{T}_k\}_{k=1}^{K} \in \mathbb{C}^{r \times b}, \end{aligned}$$

the rate of user k seeing interference from all other users can be expressed as (see [10])

$$\begin{aligned} R_k &= \log_2 \left| \mathbf{I}_N + \left(\mathbf{I}_N + \sum_{\ell \neq k} \boldsymbol{H}_\ell \boldsymbol{Q}_\ell \boldsymbol{H}_\ell^{\mathrm{H}} \right)^{-1} \boldsymbol{H}_k \boldsymbol{Q}_k \boldsymbol{H}_k^{\mathrm{H}} \right| \\ &= -\log_2 \left| \mathbf{I}_{B_k} - \boldsymbol{T}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{X}^{-1} \boldsymbol{H}_k \boldsymbol{T}_k \right|, \end{aligned} \tag{1}$$

where the substitution $\boldsymbol{X}$ reads as

$$\boldsymbol{X} = \mathbf{I}_N + \sum_{\ell=1}^{K} \boldsymbol{H}_\ell \boldsymbol{Q}_\ell \boldsymbol{H}_\ell^{\mathrm{H}} = \mathbf{I}_N + \boldsymbol{H} \boldsymbol{T} \boldsymbol{T}^{\mathrm{H}} \boldsymbol{H}^{\mathrm{H}}.$$

Reformulating the rate expression (1), we get

$$\begin{aligned} R_k &= -\log_2 \left| \boldsymbol{E}_k^{\mathrm{T}} \left(\mathbf{I}_b - \boldsymbol{T}^{\mathrm{H}} \boldsymbol{H}^{\mathrm{H}} \boldsymbol{X}^{-1} \boldsymbol{H} \boldsymbol{T} \right) \boldsymbol{E}_k \right| \\ &= -\log_2 \left| \boldsymbol{E}_k^{\mathrm{T}} \left(\mathbf{I}_b + \boldsymbol{T}^{\mathrm{H}} \boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \boldsymbol{T} \right)^{-1} \boldsymbol{E}_k \right|, \end{aligned} \tag{2}$$

where the kth block unit matrix is defined via

$$\boldsymbol{E}_k^{\mathrm{T}} = [\mathbf{0}, \ldots, \mathbf{0}, \mathbf{I}_{B_k}, \mathbf{0}, \ldots, \mathbf{0}] \in \{0; 1\}^{B_k \times b}$$

with the identity matrix at the kth block. Due to the assumption that the base station has more antennas than the terminals in sum, all r streams can be activated leading to square precoders $\boldsymbol{T}_k$ with $B_k = r_k \ \forall k$. Raising P_{Tx}, all r streams become active, $\boldsymbol{T}$ becomes full rank, and all eigenvalues of $\boldsymbol{T}^{\mathrm{H}} \boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \boldsymbol{T}$ become much larger than one. In the asymptotic limit, we obtain

$$\begin{aligned} R_k &\cong -\log_2 \left| \boldsymbol{T}_k^{-1} \boldsymbol{E}_k^{\mathrm{T}} \left(\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \right)^{-1} \boldsymbol{E}_k \boldsymbol{T}_k^{-\mathrm{H}} \right| \\ &= \ \log_2 \left| \boldsymbol{Q}_k \right| - \log_2 \left| \boldsymbol{E}_k^{\mathrm{T}} \left(\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \right)^{-1} \boldsymbol{E}_k \right|, \end{aligned} \tag{3}$$

since $\boldsymbol{E}_k^{\mathrm{T}} \boldsymbol{T}^{-1} = \boldsymbol{T}_k^{-1} \boldsymbol{E}_k^{\mathrm{T}}$. The notation $x \cong y$ means that the difference $x - y$ vanishes when the sum power P_{Tx} goes to infinity. Interestingly, the rate of user k depends only on the determinant of his own transmit covariance matrix $\boldsymbol{Q}_k$, and not on the covariance matrices of the other users! Consequently, the eigenbases of all transmit covariance matrices do not influence the rates of the users, only the powers of the eigenmodes are relevant. Let the eigenvalue decomposition of $\boldsymbol{Q}_k$ read as $\boldsymbol{Q}_k = \boldsymbol{V}_k \boldsymbol{\Lambda}_k \boldsymbol{V}_k^{\mathrm{H}}$ with unitary $\boldsymbol{V}_k$ and the diagonal nonnegative power allocation $\boldsymbol{\Lambda}_k$. Due to the determinant operator, $\boldsymbol{V}_k$ can be chosen arbitrarily and therefore, we set $\boldsymbol{V}_k = \mathbf{I}_{r_k} \ \forall k$ without loss of generality. Let the power allocation matrix be composed by the entries $\boldsymbol{\Lambda}_k = \mathbf{diag}\{\lambda_k^{(i)}\}_{i=1}^{r_k}$. Due to the sum-power constraint, the determinant $|\boldsymbol{Q}_k| = |\boldsymbol{\Lambda}_k|$ is then maximized by setting

$$\lambda_k^{(1)} = \ldots = \lambda_k^{(r_k)} =: \lambda_k, \tag{4}$$

i.e., by evenly distributing the power allocated to that user onto his individual modes, so $\boldsymbol{Q}_k = \lambda_k \mathbf{I}_{r_k} \ \forall k$ with the sum-power constraint $\sum_{k=1}^{K} r_k \lambda_k = P_{\mathrm{Tx}}$. Introducing nonnegative weight factors $w_1, \ldots, w_K$ for the rates of the users, the weighted sum rate asymptotically reads as

$$\sum_{k=1}^{K} w_k R_k \cong \sum_{k=1}^{K} w_k \left(r_k \log_2 \lambda_k - \log_2 \left| \boldsymbol{E}_k^{\mathrm{T}} \left(\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \right)^{-1} \boldsymbol{E}_k \right| \right). \tag{5}$$

Subject to the sum power constraint $\sum_{k=1}^{K} r_k \lambda_k = P_{\mathrm{Tx}}$, the weighted sum rate in (5) is maximized for

$$\lambda_k = \frac{w_k}{\sum_{\ell=1}^{K} w_\ell r_\ell} P_{\mathrm{Tx}}, \tag{6}$$

so the power is allocated to the users according to their weights (similar to the single-antenna case proven in [8]), and every user evenly distributes his fraction of power onto his modes. In the case of identical weights $w_k = 1 \ \forall k$, the conventional sum rate asymptotically reads as

$$\sum_{k=1}^{K} R_k \cong r \log_2 P_{\mathrm{Tx}} - r \log_2 r - \sum_{k=1}^{K} \log_2 \left| \boldsymbol{E}_k^{\mathrm{T}} \left(\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \right)^{-1} \boldsymbol{E}_k \right|, \tag{7}$$

and is achieved with $\boldsymbol{Q}_k = P_{\mathrm{Tx}}/r \cdot \mathbf{I}_{r_k} \ \forall k$. So, we are able to quantify the asymptotic sum rate that can be achieved by means of linear filtering for every single channel realization and antenna/user profile in terms of the transmit power P_{Tx} and the channel itself. Note that no precoders arise in (7) in contrast to [9, Eq. (10)] and [8, Eq. (21)]. In principle, the ergodic rate and the ergodic rate offset to dirty paper coding can be obtained by averaging corresponding to *any* distribution of the channel. In [9], results on the *ergodic* rate offset with respect to dirty paper coding were presented for the specific case of Rayleigh fading only, where the channel entries of $\boldsymbol{H}_1, \ldots, \boldsymbol{H}_K$ all have the same distribution. More complicated fading models cannot be captured due to this restricting assumption. Moreover, the instantaneous rate offset expression is given by means of bases representing null spaces of shortened channel matrices taken from [11] and not as a function of the channel purely as we do in (7).

Concerning the asymptotic rate expressions, we have now created a smooth transition from the r single-antenna-users system configuration in [5] where no cooperation exists between the antenna elements at the terminals, to the single-user point-to-point MIMO link where all r antennas fully cooperate, see [1] for example. In between, we can now specify any antenna/user profile we want and compute the feasible rate in the asymptotic limit. Using dirty paper coding, the asymptotic sum rate reads as

$$\sum_{k=1}^{K} R_k^{\mathrm{DPC}} \cong r \log_2 P_{\mathrm{Tx}} - r \log_2 r + \log_2 \left| \boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \right| \qquad (8)$$

and corresponds to the rate of the fully cooperating point-to-point link [9]. Combining (8) and (7), the rate loss $\Delta R = \sum_{k=1}^{K} (R_k^{\mathrm{DPC}} - R_k)$ of optimal linear filtering with respect to optimal dirty paper coding reads as

$$\Delta R \cong \sum_{k=1}^{K} \log_2 \left| \boldsymbol{E}_k^{\mathrm{T}} (\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H})^{-1} \boldsymbol{E}_k \right| - \log_2 \left| (\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H})^{-1} \right|, \quad (9)$$

which of course vanishes, if all channels are pairwise orthogonal, i.e., if $\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H}$ is block-diagonal. Of course, a block-type Hadamard inequality quickly leads to the inequality

$$-\log_2 \left| (\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H})^{-1} \right| \geq -\sum_{k=1}^{K} \log_2 \left| \boldsymbol{E}_k^{\mathrm{T}} (\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H})^{-1} \boldsymbol{E}_k \right|,$$

so linear filtering is obviously inferior to dirty paper coding.

IV. Optimum Signaling in the BC

Using our rate duality in [10], we can convert the simple solution for the covariance matrices in the dual MAC to covariance matrices in the BC, where the Hermitian channels are applied. Since this duality explicitly uses the receive filters in the MAC as scaled transmit matrices in the BC, we first compute the MMSE receivers in the dual MAC, as they are optimum and generate sufficient statistics. The receiver $\boldsymbol{G}_k$ for user k in the dual MAC reads as

$$\boldsymbol{G}_k = \boldsymbol{E}_k^{\mathrm{T}} \boldsymbol{T}^{\mathrm{H}} \boldsymbol{H}^{\mathrm{H}} \left(\mathbf{I}_N + \boldsymbol{H} \boldsymbol{T} \boldsymbol{T}^{\mathrm{H}} \boldsymbol{H}^{\mathrm{H}} \right)^{-1}.$$

With the asymptotically optimum precoders $\boldsymbol{T}_k = \sqrt{P_{\mathrm{Tx}}/r}\,\mathbf{I}_{r_k}$, above expression asymptotically converges to

$$\boldsymbol{G}_k \cong \sqrt{r/P_{\mathrm{Tx}}} \cdot \boldsymbol{E}_k^{\mathrm{T}} \left(\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \right)^{-1} \boldsymbol{H}^{\mathrm{H}}. \qquad (10)$$

Let $\boldsymbol{P}_k$ denote the precoder of user k in the BC, then the ith column $\boldsymbol{p}_{k,i}$ of $\boldsymbol{P}_k$ follows from the conjugate ith row $\boldsymbol{g}_{k,i}'^{\mathrm{T}}$ of the matrix $\boldsymbol{G}_k' = \boldsymbol{W}_k^{\mathrm{H}} \boldsymbol{G}_k$ via (see [10])

$$\boldsymbol{p}_{k,i} = \alpha_{k,i} \boldsymbol{g}_{k,i}'^{*} = \frac{\alpha_{k,i}}{\sqrt{P_{\mathrm{Tx}}/r}} \cdot \boldsymbol{H} \left(\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \right)^{-1} \boldsymbol{E}_k \boldsymbol{W}_k \boldsymbol{e}_i, \quad (11)$$

where the scaling factor $\alpha_{k,i}$ is obtained by the duality transformation and $\boldsymbol{W}_k$ is a unitary decorrelation matrix. Since we convert only the asymptotically optimum transmit precoders and receive filters, the duality transformation from the MAC to the BC in [10] drastically simplifies and can even be computed in closed form. In particular, the matrices $\boldsymbol{M}_{a,b}$ in [10, Eq. (23)] vanish for $a \neq b$ yielding a diagonal matrix $\boldsymbol{M}$ and therefore, the scaling factors read as

$$\alpha_{k,i} = \frac{\sqrt{P_{\mathrm{Tx}}/r}}{\| \boldsymbol{g}_{k,i}' \|_2}. \qquad (12)$$

In combination with (11), the ith column of the precoder associated to user k reads as

$$\boldsymbol{p}_{k,i} = \sqrt{P_{\mathrm{Tx}}/r} \cdot \frac{\boldsymbol{H} \left(\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \right)^{-1} \boldsymbol{E}_k \boldsymbol{W}_k \boldsymbol{e}_i}{\left\| \boldsymbol{H} \left(\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \right)^{-1} \boldsymbol{E}_k \boldsymbol{W}_k \boldsymbol{e}_i \right\|_2},$$

generating the precoder matrix

$$\boldsymbol{P}_k = \sqrt{P_{\mathrm{Tx}}/r} \cdot \boldsymbol{H} \left(\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \right)^{-1} \boldsymbol{E}_k \boldsymbol{W}_k \boldsymbol{D}_k^{-1}, \qquad (13)$$

where the ith diagonal element of the diagonal matrix $\boldsymbol{D}_k$ is

$$[\boldsymbol{D}_k]_{i,i} = \sqrt{\boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{W}_k^{\mathrm{H}} \boldsymbol{E}_k^{\mathrm{T}} \left(\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H} \right)^{-1} \boldsymbol{E}_k \boldsymbol{W}_k \boldsymbol{e}_i}. \qquad (14)$$

We can immediately see, that the precoding filters in (13) lead to a block diagonalization of the transmission, since $\boldsymbol{H}_\ell^{\mathrm{H}} \boldsymbol{P}_k = \mathbf{0}$ holds for $k \neq \ell$. Next, the decorrelation matrix $\boldsymbol{W}_k$ which enables the duality is usually chosen as the eigenbasis of $\boldsymbol{G}_k \boldsymbol{H}_k \boldsymbol{T}_k \cong \mathbf{I}_{r_k}$, which asymptotically coincides with the identity matrix due to (10). Since all eigenvalues are identical to one, the decorrelation matrices $\boldsymbol{W}_k$ are not given a priori, but can easily be computed such that the BC features the same sum rate as the dual MAC. By means of (13) and the block diagonalization property of the precoders, we obtain for user k's receive signal

$$\boldsymbol{y}_k = \boldsymbol{H}_k^{\mathrm{H}} \boldsymbol{P}_k \boldsymbol{s}_k + \boldsymbol{\eta}_k = \sqrt{P_{\mathrm{Tx}}/r} \cdot \boldsymbol{W}_k \boldsymbol{D}_k^{-1} \boldsymbol{s}_k + \boldsymbol{\eta}_k, \quad (15)$$

where $\boldsymbol{\eta}_k \in \mathbb{C}^{r_k}$ is the noise and $\boldsymbol{s}_k$ the symbol vector of user k both having an identity covariance matrix. From (15), the rate of user k achieved in the BC reads as

$$R_k = \log_2 \left| \mathbf{I}_{r_k} + P_{\mathrm{Tx}}/r \cdot \boldsymbol{W}_k \boldsymbol{D}_k^{-2} \boldsymbol{W}_k^{\mathrm{H}} \right|,$$

which asymptotically converges to

$$R_k \cong r_k \log_2 P_{\mathrm{Tx}} - r_k \log_2 r - \log_2 |\boldsymbol{D}_k^2|. \qquad (16)$$

Above expression is maximized, if we choose $\boldsymbol{W}_k$ as the unitary eigenbasis of $\boldsymbol{E}_k^{\mathrm{T}} (\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H})^{-1} \boldsymbol{E}_k$, see (14), such that $\boldsymbol{D}_k^2$ contains the eigenvalues. As a consequence, the elements of $\boldsymbol{D}_k^2$ are as different as possible since the eigenvalues of any positive definite matrix majorize its diagonal elements according to *Schur's* theorem [12]. Thus, the transmit covariance matrix $\boldsymbol{S}_k = \boldsymbol{P}_k \boldsymbol{P}_k^{\mathrm{H}}$ of user k reads as

$$\boldsymbol{S}_k = \frac{P_{\mathrm{Tx}}}{r} \cdot \boldsymbol{H}^{+\,\mathrm{H}} \boldsymbol{E}_k \left(\boldsymbol{E}_k^{\mathrm{T}} (\boldsymbol{H}^{\mathrm{H}} \boldsymbol{H})^{-1} \boldsymbol{E}_k \right)^{-1} \boldsymbol{E}_k^{\mathrm{T}} \boldsymbol{H}^{+} \quad (17)$$

with the channel pseudo-inverse $\boldsymbol{H}^{+} = (\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{H}}$. Note that r_k eigenvalues of $\boldsymbol{S}_k$ are P_{Tx}/r whereas the remaining $N - r_k$ ones are zero. Thus, $\boldsymbol{S}_k$ is a weighted orthogonal projector. Furthermore, $\mathrm{tr}(\boldsymbol{S}_k) = P_{\mathrm{Tx}} \cdot r_k/r = \mathrm{tr}(\boldsymbol{Q}_k)$, so the power is uniformly allocated to the individual users in the broadcast channel according to their numbers of antennas as well. Comparing (17) with the simple solution of the transmit covariance matrix $\boldsymbol{Q}_k = P_{\mathrm{Tx}}/r \cdot \mathbf{I}_{r_k}$ in the dual MAC, it becomes obvious that the optimum covariance matrices are much more difficult to find directly in the BC without using the rate duality, than in the dual MAC. Plugging the optimum $\boldsymbol{D}_k^2$ containing the eigenvalues of $\boldsymbol{E}_k^{\mathrm{T}}(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1}\boldsymbol{E}_k$ into (16) finally yields

$$R_k \cong r_k \log_2 P_{\mathrm{Tx}} - r_k \log_2 r - \log_2 \left| \boldsymbol{E}_k^{\mathrm{T}} (\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1} \boldsymbol{E}_k \right|.$$

Hence, the maximum sum rate (7) in the dual MAC is also achieved in the BC.

V. Ergodic Rate Expressions

In this section, we derive expressions for the asymptotic sum rate when averaging over the channel realizations. The simple channel model in [9], [5] is a prerequisite for the application of the ergodic analysis due to the fact that an instantaneous analysis is not possible there because of the precoder dependent expressions. As a simple example, we choose a more realistic channel where channel correlations at the terminals are modeled as well. Thanks to our closed form expression of the maximum asymptotic rate for an instantaneous channel realization that only depend on the channels of the users, the following ergodic analysis is basically feasible for *any* distribution of the channel coefficients. The channel matrices of the chosen channel model with transmit correlations (in the MAC) are defined by $\boldsymbol{H}_k = \bar{\boldsymbol{H}}_k \boldsymbol{C}_k^{\frac{1}{2}} \ \forall k$, where the elements of $\bar{\boldsymbol{H}}_k$ are uncorrelated and share a zero-mean i.i.d. Gaussian distribution with variance one, and the Hermitian matrix $\boldsymbol{C}_k^{\frac{1}{2}}$ contains the correlations. An uncorrelated channel purely modeling the near-far effect can be obtained by setting $\boldsymbol{C}_k = c_k \mathbf{I}_{r_k}$, where $c_k > 0$ is then the inverse path loss of user k. Let the $r \times r$ matrix $\boldsymbol{C}$ be defined via

$$\boldsymbol{C} = \mathbf{blockdiag}\{\boldsymbol{C}_k\}_{k=1}^{K},$$

then the frequently arising inverse of $\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}$ reads as

$$(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1} = \boldsymbol{C}^{-\frac{1}{2}} (\bar{\boldsymbol{H}}^{\mathrm{H}} \bar{\boldsymbol{H}})^{-1} \boldsymbol{C}^{-\frac{1}{2}},$$

where $\bar{\boldsymbol{H}}^{\mathrm{H}}\bar{\boldsymbol{H}} \sim \mathcal{W}_r(N, \mathbf{I}_r)$ has a *Wishart* distribution with N degrees of freedom and $(\bar{\boldsymbol{H}}^{\mathrm{H}}\bar{\boldsymbol{H}})^{-1} \sim \mathcal{W}_r^{-1}(N, \mathbf{I}_r)$ has an *inverse Wishart* distribution, see [13], [14]. Thus, the ergodic value for the channel dependent log-summand in the DPC sum rate expression (8) reads as [15]

$$\mathrm{E}\left[\log_2 |\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H}| \right] = \frac{1}{\ln 2} \sum_{\ell=0}^{r-1} \psi(N-\ell) + \sum_{k=1}^{K} \log_2 |\boldsymbol{C}_k|, \quad (19)$$

where the *Digamma*-function $\psi(\cdot)$ with integer arguments is defined via [15]

$$\psi(n+1) = \psi(n) + \frac{1}{n} \quad \text{if } n \in \mathbb{N}, \quad \psi(1) = -\gamma, \qquad (20)$$

and γ is the *Euler-Mascheroni* constant. Note from (19) that different path losses and correlations in the channel coefficients simply lead to a shift of the asymptotic rate curve. Concerning the rate expressions with linear filtering, we exploit the property that the kth main diagonal block of $(\bar{\boldsymbol{H}}^{\mathrm{H}}\bar{\boldsymbol{H}})^{-1}$ is also inverse Wishart [14]:

$$\boldsymbol{E}_k^{\mathrm{T}} (\bar{\boldsymbol{H}}^{\mathrm{H}}\bar{\boldsymbol{H}})^{-1} \boldsymbol{E}_k \sim \mathcal{W}_{r_k}^{-1}(N - r + r_k, \mathbf{I}_{r_k}),$$

In combination with $\boldsymbol{E}_k^{\mathrm{T}} \boldsymbol{C}^{-\frac{1}{2}} = \boldsymbol{C}_k^{-\frac{1}{2}} \boldsymbol{E}_k^{\mathrm{T}}$, this leads to the ergodic expression

$$\begin{aligned} &\mathrm{E}\left[\log_2 \left| \boldsymbol{C}_k^{-\frac{1}{2}} \boldsymbol{E}_k^{\mathrm{T}} (\bar{\boldsymbol{H}}^{\mathrm{H}}\bar{\boldsymbol{H}})^{-1} \boldsymbol{E}_k \boldsymbol{C}_k^{-\frac{1}{2}} \right| \right] = \\ &\quad - \log_2 |\boldsymbol{C}_k| - \frac{1}{\ln 2} \sum_{\ell=0}^{r_k-1} \psi(N - r + r_k - \ell). \end{aligned} \qquad (21)$$

By means of (19) and (21), averaging over the asymptotic rate difference $\mathit{\Delta} R$ in (9) between linear filtering and DPC yields

$$\mathrm{E}[\mathit{\Delta} R] \cong \frac{1}{\ln 2} \Big[\sum_{\ell=0}^{r-1} \psi(N{-}\ell) - \sum_{k=1}^{K} \sum_{\ell=0}^{r_k-1} \psi(N{-}r{+}r_k{-}\ell) \Big], \quad (22)$$

from which we can observe that different path losses and channel correlations do not influence the rate difference, since both DPC and linear filtering are affected in the same way.

The general expression (22) for the ergodic rate loss $\mathrm{E}[\mathit{\Delta} R]$ can be simplified by means of (20), when all users are equipped with the same number of antennas. For the first special case, assume that each user has $\bar{r} > 1$ antennas, i.e., $r_1 = \ldots = r_K = \bar{r}$, such that the total number of antennas therefore is $r = K\bar{r}$. After some manipulations, we obtain

$$\mathrm{E}[\mathit{\Delta} R] \cong \frac{1}{\ln 2} \left[\sum_{\ell=1}^{(K-1)\bar{r}} \frac{\ell}{N - \ell} + \sum_{\ell=1}^{\bar{r}-1} \frac{(K-1)\ell}{N - K\bar{r} + \ell} \right], \quad (23)$$

which coincides with the results in [8], [9], but is a different representation. For convenience, we assume that the summation vanishes if the upper limit of a sum is smaller than the lower one, which happens for $\bar{r} = 1$. In this second special case with single antenna receivers, i.e., $\bar{r} = 1 = r_k \ \forall k$ and $r = K$, the second sum in (23) consequently vanishes, and the ergodic rate loss simplifies to

$$\mathrm{E}[\mathit{\Delta} R] \cong \frac{1}{\ln 2} \sum_{\ell=1}^{K-1} \frac{\ell}{N - \ell}, \qquad (24)$$

which is also a result of [5].

VI. Numerical Examples

In Table I, we present the ergodic rate loss of linear filtering with respect to dirty paper coding for different parameters N, K, $\bar{r}$, r_1, and r_2, where we employed (23) and (24) for the case $\bar{r} = r_1 = \ldots = r_K$ (cf. [5], [8], [9]) and (22) for the case of different numbers of antennas r_1, r_2. It can be seen that a fully loaded single antenna system with $K = N$ and $\bar{r} = 1$ has to face a significant rate reduction when switching from nonlinear to linear filtering. Moreover, comparing the $K = 2$ and $\bar{r} = 3$ system with the one where $K = 3$ and $\bar{r} = 2$, we observe that the rate loss in the first system is only 65 percent of the one in the second system for $N = 6$. We can infer that fewer terminals with many antennas have to face smaller rate losses than many terminals with only few antennas.

Next, we plot the ergodic sum capacity with DPC and the ergodic sum rate when linear filtering is applied versus the transmit power P_{Tx} to see how large P_{Tx} must be to let the asymptotic affine approximations become tight. To this end, we choose a system configuration where $K = 2$ users each having $\bar{r} = 2$ antennas are served by an $N = 5$ antenna base station. Different path losses are modeled by setting $\boldsymbol{C}_1 = \mathbf{I}_2$ and $\boldsymbol{C}_2 = 2 \cdot \mathbf{I}_2$, i.e., user 2 has a stronger channel on average, and we averaged over 1000 channel realizations. While the DPC sum capacity can easily be computed via the algorithms in [16] or [17], an algorithm proven to reach the maximum sum rate under linear filtering does not seem to be available yet. Hence, we utilize our combinatorial approach in [18], which obtains the best sum rate hitherto known in the case of linear filtering. Fig. 1 shows that the asymptotic affine approximations become tight already for P_{Tx} smaller than 20dB and confirms the asymptotic ergodic rate loss $\mathrm{E}[\Delta R] \cong 2.04$ from Table I which is independent of the different average channel powers. For a multiplexing gain of $r = 4$ as in the chosen system configuration, this translates to an asymptotic power loss of 1.54dB of linear filtering with respect to DPC.

VII. Conclusion

In this paper, we derived the asymptotic sum capacity which is maximally achievable with *linear* filtering in the broadcast channel by means of our rate duality linking the rate region of the multiple access channel with the broadcast channel rate region. Due to the closed form expression of the asymptotic sum capacity for every single channel realization, the instantaneous rate loss with respect to dirty paper coding was presented, and the ergodic rate loss can quickly be computed or simulated for any distribution of the fading process. Analytic averaging for arbitrary fading models is not feasible with the state-of-the-art results in the literature. As an example, we presented the solution of the ergodic rate loss for a simple fading model incorporating correlations at the mobiles. Another key result proven is that block-diagonalization is the asymptotically optimum transmission strategy in the broadcast channel.

$K,\bar{r}$	$N=2$	$N=3$	$N=4$	$N=5$	$N=6$
2, 1	1.443	0.721	0.481	0.361	0.289
3, 1	-	3.607	1.924	1.322	1.010
4, 1	-	-	6.252	3.487	2.453
5, 1	-	-	-	9.257	5.338
6, 1	-	-	-	-	12.551
2, 2	-	-	3.366	2.044	1.491
2, 3	-	-	-	-	5.338
3, 2	-	-	-	-	8.223
r_1, r_2	$N=2$	$N=3$	$N=4$	$N=5$	$N=6$
1,2	-	2.164	1.202	0.842	0.649
1,3	-	-	2.645	1.563	1.130
1,4	-	-	-	3.006	1.851
2,3	-	-	-	4.208	2.693
2,4	-	-	-	-	4.857

TABLE I

Asymptotic ergodic rate loss $\mathrm{E}[\Delta R]$ in bits/s/Hz; N antennas at the base, K users, r_k antennas at user k.

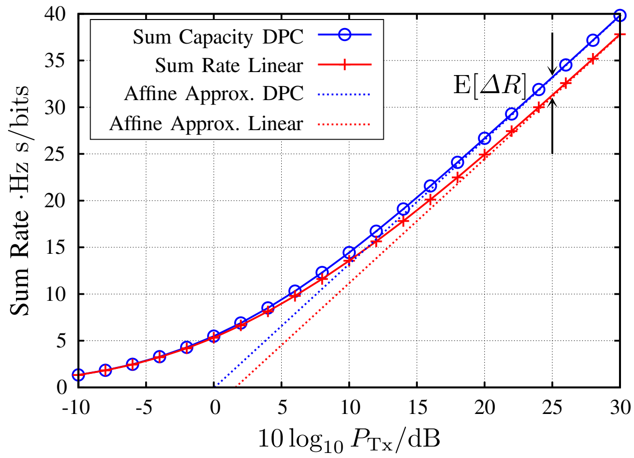

Fig. 1. Ergodic sum rate of linear filtering and DPC and their respective affine approximations with $K = 2$, $N = 5$, and $r_1 = r_2 = \bar{r} = 2$.

References

[1] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–596, November/December 1999.

[2] J. Salo, P. Suvikunnas, H. M. El-Sallabi, and P. Vainikainen, "Some results on MIMO mutual information: the high SNR case," in *Global Telecommunications Conference (Globecom '04)*, vol. 2, December 2004, pp. 943–947.

[3] J. Salo and P. Suvikunnas and H. M. El-Sallabi and P. Vainikainen, "Some Insights into MIMO Mutual Information: The High SNR Case," *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 2997–3001, November 2006.

[4] N. Prasad and M. K. Varanasi, "Throughput analysis for MIMO systems in the high SNR regime," in *International Symposium on Information Theory (ISIT)*, July 2006, pp. 1954–1958.

[5] N. Jindal, "High SNR Analysis of MIMO Broadcast Channels," in *International Symposium on Information Theory (ISIT 2005)*, September 2005, pp. 2310–2314.

[6] S. Shamai and S. Verdú, "The Impact of Frequency-Flat Fading on the Spectral Efficiency of CDMA," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.

[7] A. Lozano, A. M. Tulino, and S. Verdú, "High-SNR Power Offset in Multiantenna Communication," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4134–4151, December 2005.

[8] J. Lee and N. Jindal, "Dirty Paper Coding vs. Linear Precoding for MIMO Broadcast Channels," in *40th Asilomar Conference on Signals, Systems, and Computers (Asilomar 2006)*, October 2006, pp. 779–783.

[9] ——, "High SNR Analysis for MIMO Broadcast Channels: Dirty Paper Coding Versus Linear Precoding," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4787–4792, December 2007.

[10] R. Hunger and M. Joham, "A General Rate Duality of the MIMO Multiple Access Channel and the MIMO Broadcast Channel," in *Global Telecommunications Conference (Globecom '08)*, November 2008.

[11] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-Forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, February 2004.

[12] A.W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, ser. Mathematics in Science and Engineering, R. Bellman, Ed. Academic Press, 1979, vol. 143.

[13] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, 2nd ed. Wiley, 2005.

[14] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*. Chapman & Hall /Crc, 1999.

[15] A. M. Tulino and S. Verdu, *Random Matrix Theory and Wireless Communications*. Now Publishers Inc, 2004.

[16] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. J. Goldsmith, "Sum Power Iterative Water-Filling for Multi-Antenna Gaussian Broadcast Channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1570–1580, 2005.

[17] R. Hunger, D. A. Schmidt, and W. Utschick, "Sum-Capacity and MMSE for the MIMO Broadcast Channel without Eigenvalue Decompositions," in *IEEE International Symposium on Information Theory (ISIT)*, Nice, June 2007.

[18] R. Hunger, D. A. Schmidt, and M. Joham, "A Combinatorial Approach to Maximizing the Sum Rate in the MIMO BC with Linear Precoding," in *42nd Asilomar Conference on Signals, Systems, and Computers (Asilomar 2008)*, October 2008.