On the Inseparability of Parallel MIMO Broadcast Channels With Linear Transceivers

Christoph Hellings and Wolfgang Utschick

IEEE Transactions on Signal Processing

vol. 59, no. 12, Dec. 2011

Technische Universität München
Associate Institute for Signal Processing
http://www.msv.ei.tum.de

On the Inseparability of Parallel MIMO Broadcast Channels With Linear Transceivers

Christoph Hellings and Wolfgang Utschick

***Abstract*—Parallel multiple-input multiple-output (MIMO) broadcast channels have been shown to be separable from an information theoretic point of view, i.e., capacity can be achieved with a strategy that performs separate encoding and decoding on each subchannel subject to a power allocation across subchannels. In this paper, we show that separability does not necessarily hold if the transmit strategy is restricted to linear transceivers. The proof will be done by identifying a rate tuple that is achievable in a certain set of channels with joint treatment of the subchannels, but lies outside the achievable rate region of separate linear precoding. The implications of this result are that any algorithm to optimize transmit strategies in parallel MIMO broadcast channels that is based on both linear transceivers and separate encoding on each subchannel does in fact introduce two suboptimalities and not only one.**

***Index Terms*— Linear transceivers, multiple-input multiple-output (MIMO), multi-user multi-carrier systems, parallel broadcast channels, rate region, separable and inseparable channels.**

I. Introduction

The concept of parallel channels is a way to model a communication system that consists of a set of orthogonal resources, such as (groups of) subcarriers of a frequency selective channel or time intervals in a fading channel. Unlike parallel Gaussian interference channels, which are known to be inseparable in general [1], parallel Gaussian broadcast channels were shown to be separable for the single-antenna case in [2] and [3] and for the multiantenna case in [4] and [5]. Therefore, many algorithms that have been proposed to optimize the transmit strategy in parallel MIMO broadcast channels (e.g., [6]–[10]) are based on the assumption that each of the parallel channels can be treated separately. A mathematical interpretation of separate and joint treatment of parallel channels will be given together with the introduction of the system model in Section II.

However, recent results on multicarrier broadcast channels have suggested that separate treatment of the subchannels may be suboptimal if linear transceivers are used [11]. In particular, it was shown that quality of service requirements might become infeasible due to separation of the subchannels although they would be feasible for unseparated channels. Although this observation is rather specific in that it is restricted to feasibility considerations for quality of service problems, i.e., it only applies to systems without time-sharing that have more receivers than degrees of freedom, it motivates further research on the subject.

The main result of this paper is derived in Section III, where we will show that for a certain channel realization, there exists a rate tuple that is achievable with linear precoding if the subchannels are treated jointly, but is no longer part of the rate region if the subchannels are separated. This result is much more general than the abovementioned one since it holds for various objective functions that take their optimal values on the Pareto boundary of the rate region, and it even holds if time-sharing is allowed and if the system is not fully loaded. This generality is somewhat surprising: a potential reason for inseparability of parallel broadcast channels in the case of quality of service constrained problems in overloaded systems without time-sharing could be an unsuitable combination of the individual requirement sizes [11]. However, this is no longer a possible explanation for the general inseparability theorem presented in this paper. Instead, the examples constructed in Section III will demonstrate that inseparability for linear transceivers is a property inherent to certain channel realizations. More examples for channel realizations that lead to inseparability are given in Section IV, and the implications of the result will be discussed in detail in Section V.

Notation: Vectors are typeset in boldface lowercase letters and matrices in boldface uppercase letters. We write $\mathbf{0}$ for the zero matrix or vector and $\mathbf{I}_N$ for the identity matrix of size N, and we use $\boldsymbol{A}^{\mathrm{T}}$ and $\boldsymbol{A}^{\mathrm{H}}$ to denote the transpose and conjugate transpose of a matrix $\boldsymbol{A}$, respectively. The notation $\|\boldsymbol{a}\|_2$ is used for the Euclidean norm of a vector $\boldsymbol{a}$.

II. System Model and Mathematical Interpretation of Subchannel Separation

We consider a set of C parallel broadcast channels such as, e.g., carriers in a multicarrier system. Each subchannel $c \in \{1,\ldots,C\}$ is characterized by a set of channel matrices $(\boldsymbol{H}_k^{(c),\mathrm{H}} \in \mathbb{C}^{N_k \times M})_{k\in\{1,\ldots,K\}}$ and noise covariance matrices $(\boldsymbol{C}_{\boldsymbol{\eta}_k} \in \mathbb{C}^{N_k \times N_k})_{k\in\{1,\ldots,K\}}$. Here, M is the number of transmit antennas, K is the number of receivers, and N_k is the number of antennas at receiver k. To characterize the overall system, these matrices can be written in block-diagonal channel matrices

$$\boldsymbol{H}_k^{\mathrm{H}} = \begin{bmatrix} \boldsymbol{H}_k^{(1),\mathrm{H}} & & \\ & \ddots & \\ & & \boldsymbol{H}_k^{(C),\mathrm{H}} \end{bmatrix} \in \mathbb{C}^{N_k C \times MC} \tag{1}$$

and noise covariance matrices

$$\boldsymbol{C}_{\boldsymbol{\eta}_k} = \begin{bmatrix} \boldsymbol{C}_{\boldsymbol{\eta}_k}^{(1)} & & \\ & \ddots & \\ & & \boldsymbol{C}_{\boldsymbol{\eta}_k}^{(C)} \end{bmatrix} \in \mathbb{C}^{N_k C \times N_k C} \tag{2}$$

(cf., e.g., [12]). The data transmission with linear transceivers can now be described by the equation

$$\hat{\boldsymbol{x}}_k = \boldsymbol{V}_k^{\mathrm{H}} \boldsymbol{H}_k^{\mathrm{H}} \sum_{k'=1}^{K} \boldsymbol{B}_{k'} \boldsymbol{x}_{k'} + \boldsymbol{V}_k^{\mathrm{H}} \boldsymbol{\eta}_k \tag{3}$$

for all receivers k, where $\boldsymbol{x}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{S_k})$ is the vector of circularly symmetric complex Gaussian data symbols intended for receiver k with $S_k \leq C \min\{N_k, M\}$ being the number of independent data streams, and $\hat{\boldsymbol{x}}_k \in \mathbb{C}^{S_k}$ is the corresponding estimate. Moreover, the matrices $\boldsymbol{B}_k \in \mathbb{C}^{MC \times S_k}$ are the beamforming matrices, and $\boldsymbol{V}_k^{\mathrm{H}} \in \mathbb{C}^{S_k \times N_k C}$ are the receive filters. Note that $\boldsymbol{x}_k$ is a concatenation of all symbols intended for user k no matter across which subchannel(s) they are transmitted.

If we do not impose any constraints on the structure of the beamforming matrices $\boldsymbol{B}_k$, we allow that a transmit symbol is spread over several subchannels. This corresponds to the case of joint encoding

Manuscript received June 19, 2011; revised July 26, 2011; accepted August 12, 2011. Date of publication September 12, 2011; date of current version November 16, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Merouane Debbah.

The authors are with the Associate Institute for Signal Processing, Technische Universität München, 80290 München, Germany (e-mail: hellings@tum.de; utschick@tum.de).

Digital Object Identifier 10.1109/TSP.2011.2167619

over the parallel channels and was called carrier-cooperative transmission in [12]. However, we could also impose the constraint that the matrices $\boldsymbol{B}_k$ (and consequently, the matrices $\boldsymbol{V}_k^{\mathrm{H}}$) have to match the block-diagonal structure of the channel matrices and the noise covariance matrices. This case, which corresponds to separate treatment of the subchannels, was called carrier-noncooperative transmission in [12] and is equivalent to an independent data transmission

$$\hat{\boldsymbol{x}}_k^{(c)} = \boldsymbol{V}_k^{(c),\mathrm{H}} \boldsymbol{H}_k^{(c),\mathrm{H}} \sum_{k'=1}^{K} \boldsymbol{B}_{k'}^{(c)} \boldsymbol{x}_{k'}^{(c)} + \boldsymbol{V}_k^{(c),\mathrm{H}} \boldsymbol{\eta}_k^{(c)} \tag{4}$$

on all subchannels c. In this case, the only coupling between the subchannels is that each receiver achieves a data rate r_k that is the sum of its rates $r_k^{(c)}$ on each subchannel, i.e., $r_k = \sum_{c=1}^{C} r_k^{(c)}$, and the total transmit power P is distributed into subchannel powers $P^{(c)}$ according to a certain power allocation strategy.

III. Inseparability in the Case of Linear Transceivers

In order to precisely define the notion of separability, we first introduce the rate region $\mathcal{C}(P)$, i.e., the set of all rate vectors $\boldsymbol{r} = [r_1, \ldots, r_K]^{\mathrm{T}}$ that can be achieved with sum power P, and the rate region for separate coding on each subchannel $\mathcal{C}_{\mathrm{sep}}(P)$, i.e., the set of rate vectors that can be achieved using separate transmission with sum power P. Note that $\mathcal{C}(P)$ is a convex set since the applied strategy can involve time-sharing between different operation points, which is equivalent to allowing convex combinations between different transmit strategies.[1]

Definition 1: We call a set of parallel channels *separable* if for all transmit powers P and all rate vectors $\boldsymbol{r} \in \mathcal{C}(P)$, it holds that $\boldsymbol{r} \in \mathcal{C}_{\mathrm{sep}}(P)$, i.e., $\boldsymbol{r}$ can be achieved using separate transmission with transmit power P. On the other hand, if there exists a power P such that there are rate vectors $\boldsymbol{r} \in \mathcal{C}(P)$ with $\boldsymbol{r} \notin \mathcal{C}_{\mathrm{sep}}(P)$, the parallel channels are called *inseparable*.

The main result of this paper is stated in the following theorem. As the proof is performed by construction of an explicit example, it is not very technical, but rather insightful. In fact, the aim of this section is not only to prove the inseparability, but also to give some intuition about why this phenomenon occurs.

Theorem 1: Parallel MIMO broadcast channels with linear transceivers are not always separable.

Proof of Theorem 1: We provide a proof by construction. We first choose the system parameters of an example system as well as a rate vector $\boldsymbol{\rho}$, and we compute the sum transmit power P such that $\boldsymbol{\rho}$ lies on the Pareto boundary of $\mathcal{C}_{\mathrm{sep}}(P)$. To prove inseparability, it is not necessary to know $\mathcal{C}(P)$. Instead, we will find a rate vector $\boldsymbol{r} \in \mathcal{R}(P) \subseteq \mathcal{C}(P)$ with $r_k > \rho_k\ \forall k$, which implies $\boldsymbol{r} \notin \mathcal{C}_{\mathrm{sep}}(P)$. Here, $\mathcal{R}(P)$ is used to denote a subset of $\mathcal{C}(P)$, which contains the rate vectors that can be achieved with a certain type of suboptimal strategies. In our case, such a potential suboptimality will be introduced by refraining from the use of time-sharing and by the restriction to K data streams. As will be seen, it is possible to find a vector $\boldsymbol{r}$ that is a scaled version of the chosen vector $\boldsymbol{\rho}$.

Consider a set of parallel broadcast channels with $C = 2$, $M = 2$, $K = 3$, and $N_k = 1\ \forall k$. In this case, the noise covariance matrices are reduced to scalars, which we assume to be $\sigma_k^{(c)} = 1\ \forall k,\ \forall c$, and the channel matrices become row vectors $\boldsymbol{h}_k^{(c),\mathrm{H}}$. We choose the channel realization

$$\boldsymbol{H}_1^{\mathrm{H}} = \begin{bmatrix} \boldsymbol{h}_1^{(1),\mathrm{H}} & \\ & \boldsymbol{h}_1^{(2),\mathrm{H}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{5}$$

$$\boldsymbol{H}_2^{\mathrm{H}} = \begin{bmatrix} \boldsymbol{h}_2^{(1),\mathrm{H}} & \\ & \boldsymbol{h}_2^{(2),\mathrm{H}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{6}$$

$$\boldsymbol{H}_3^{\mathrm{H}} = \begin{bmatrix} \boldsymbol{h}_3^{(1),\mathrm{H}} & \\ & \boldsymbol{h}_3^{(2),\mathrm{H}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{\mathrm{j}}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{-\mathrm{j}}{\sqrt{2}} \end{bmatrix}. \tag{7}$$

We will now derive the minimal transmit power that is needed to achieve the rates $\rho_k = \rho = 1\ \forall k$ with separate treatment of the subchannels. Using the power minimization algorithm from [10], which can compute the globally optimal solution numerically up to an arbitrarily small error tolerance, it can be easily verified that the strategy derived in the following is the globally optimal one.

Note that all channel vectors have unit norm and that the setting on each subchannel is totally symmetric in that the absolute value of the inner product of two channel vectors is given by

$$\left| \boldsymbol{h}_k^{(c),\mathrm{H}} \boldsymbol{h}_j^{(c)} \right| = \begin{cases} 1, & \text{if } k = j, \\ \frac{1}{\sqrt{2}}, & \text{if } k \neq j, \end{cases} \tag{8}$$

for all $k, j \in \{1, 2, 3\}$ and $c \in \{1, 2\}$. As the vectors are normalized, this measure can be interpreted as the cosine of a generalized angle between two complex vectors, i.e., $\cos \theta_{k,j}^{(c)} = |\boldsymbol{h}_k^{(c),\mathrm{H}} \boldsymbol{h}_j^{(c)}|$. Following the nomenclature of [13], we call $\theta_{k,j}^{(c)} = 45^\circ$ the *Hermitian angle* between the vectors $\boldsymbol{h}_k^{(c)}$ and $\boldsymbol{h}_j^{(c)}$.

Since we have two transmit antennas, no more than two users should be scheduled on each subchannel in each time slot. For reasons of symmetry, one of the following two strategies must attain the minimal sum transmit power: either using three time slots with equal length, where each possible pair of users is scheduled on both subchannels during one of the time slots, or exclusively serving one user in each of the three time slots. In the latter case, the sum power can be easily calculated by equally dividing the rate and the power between the two subchannels:

$$P_1 = 2\left(2^{\frac{3}{2}\rho} - 1\right) \approx 3.6569. \tag{9}$$

The former possibility will be discussed in the following paragraphs.

As the three time slots, two subchannels, and two spatial dimensions correspond to twelve degrees of freedom, each of the three users is scheduled four times, always with a duration of one third of the total time. In each time slot, the same sum transmit power is used on each subchannel, so that the average power is $P = 2P^{(c)}$, where $P^{(c)}$ is the power needed to serve a user pair (k, j) at the rates $\rho_k^{(c)} = \rho_j^{(c)} = \frac{3}{4}\rho$ on one of the subchannels. For reasons of symmetry, this power has to be distributed equally among the users such that for each of the two users, the power $\frac{P}{4}$ is spent on that subchannel. The transmit strategy is visualized in Fig. 1.

[1]Note that time-sharing can be defined in two ways: either with a short-term average power constraint, i.e., the power constraint has to be fulfilled in each time slot, or with a long-term average power constraint, i.e., the power constraint only has to be fulfilled after averaging over all time slots. In both cases, the rate region becomes convex. Whenever time-sharing is applied in this paper, it turns out that the optimal time-sharing strategy for the case of a long-term average power constraint uses constant power so that it also fulfills the corresponding short-term average power constraint. Thus, the results presented in this paper hold for both types of time-sharing.

Time Slot 1

	Subchannel 1	Subchannel 2
User 1	$\rho_1^{(1)} = \frac{3}{4}\rho$ $p_1^{(1)} = \frac{1}{4}P$	$\rho_1^{(2)} = \frac{3}{4}\rho$ $p_1^{(2)} = \frac{1}{4}P$
User 2	$\rho_2^{(1)} = \frac{3}{4}\rho$ $p_2^{(1)} = \frac{1}{4}P$	$\rho_2^{(2)} = \frac{3}{4}\rho$ $p_2^{(2)} = \frac{1}{4}P$
User 3		

Time Slot 2

	Subchannel 1	Subchannel 2
User 1	$\rho_1^{(1)} = \frac{3}{4}\rho$ $p_1^{(1)} = \frac{1}{4}P$	$\rho_1^{(2)} = \frac{3}{4}\rho$ $p_1^{(2)} = \frac{1}{4}P$
User 2		
User 3	$\rho_3^{(1)} = \frac{3}{4}\rho$ $p_3^{(1)} = \frac{1}{4}P$	$\rho_3^{(2)} = \frac{3}{4}\rho$ $p_3^{(2)} = \frac{1}{4}P$

Time Slot 3

	Subchannel 1	Subchannel 2
User 1		
User 2	$\rho_2^{(1)} = \frac{3}{4}\rho$ $p_2^{(1)} = \frac{1}{4}P$	$\rho_2^{(2)} = \frac{3}{4}\rho$ $p_2^{(2)} = \frac{1}{4}P$
User 3	$\rho_3^{(1)} = \frac{3}{4}\rho$ $p_3^{(1)} = \frac{1}{4}P$	$\rho_3^{(2)} = \frac{3}{4}\rho$ $p_3^{(2)} = \frac{1}{4}P$

Fig. 1. Visualization of the symmetric time-sharing solution.

Using a dual uplink formulation [14], the rate of user k on subchannel c can be expressed as

$$\begin{aligned}\frac{3}{4} = \frac{3}{4}\rho = \rho_k^{(c)} &= \log_2\left(1 + \frac{P}{4}\boldsymbol{h}_k^{(c),\mathrm{H}}\left(\mathbf{I}_M + \frac{P}{4}\boldsymbol{h}_j^{(c)}\boldsymbol{h}_j^{(c),\mathrm{H}}\right)^{-1}\boldsymbol{h}_k^{(c)}\right)\\ &= \log_2\left(1 + \frac{P}{4}\boldsymbol{h}_k^{(c),\mathrm{H}}\boldsymbol{h}_k^{(c)} - \frac{\left(\frac{P}{4}\right)^2\left|\boldsymbol{h}_k^{(c),\mathrm{H}}\boldsymbol{h}_j^{(c)}\right|^2}{1+\frac{P}{4}\boldsymbol{h}_j^{(c),\mathrm{H}}\boldsymbol{h}_j^{(c)}}\right)\\ &= \log_2\left(1 + \frac{\frac{P}{4}+\frac{P^2}{32}}{1+\frac{P}{4}}\right)\end{aligned} \tag{10}$$

where j is the user that is scheduled together with user k in the respective time slot. The third line is due to the matrix inversion lemma,[2] and the last equality follows from (8). Solving this for P, we get $P \approx 3.5684 < 3.6569 \approx P_1$. Thus, the solution with two users per time slot on each subchannel is optimal for the given setting. Due to the optimality of P, the rate vector $\boldsymbol{\rho} = [1,1,1]^{\mathrm{T}}$ lies on the Pareto boundary of the rate region $\mathcal{C}_{\mathrm{sep}}(P)$ for separate coding on each subchannel with sum power P.

Let us now consider the unseparated case. To prove the theorem, it is sufficient to derive a suboptimal strategy without separation that achieves a strictly larger rate tuple with the same sum transmit power. As will be shown in the following, this can even be done without using time-sharing. As the overall system has 4 degrees of freedom, we decide that the number of data streams should not exceed 4. Thus, for reasons of symmetry, we serve each user with only one data stream. The receive filters are reduced to row vectors $\boldsymbol{v}_k^{\mathrm{H}}$, and, again for reasons of symmetry, they have to be

$$\boldsymbol{v}_k^{\mathrm{H}} = \frac{\left\|\boldsymbol{v}_k^{\mathrm{H}}\right\|_2}{\sqrt{2}}\left[\mathrm{e}^{\mathrm{j}\varphi_k^{(1)}} \quad \mathrm{e}^{\mathrm{j}\varphi_k^{(2)}}\right] = \frac{1}{\sqrt{2}}\left[1 \quad \mathrm{e}^{\mathrm{j}\varphi_k}\right] \tag{11}$$

for $k \in \{1,2,3\}$, where $\left\|\boldsymbol{v}_k^{\mathrm{H}}\right\|_2 = 1$ and $\varphi_k^{(1)} = 0$ were chosen without loss of generality. Applying these receive filters, we get an effective vector broadcast channel with channel vectors

$$\tilde{\boldsymbol{h}}_k^{\mathrm{H}} = \boldsymbol{v}_k^{\mathrm{H}}\boldsymbol{H}_k^{\mathrm{H}}. \tag{12}$$

Note that $\tilde{\boldsymbol{h}}_k^{\mathrm{H}}\tilde{\boldsymbol{h}}_k = 1$ due to $\left\|\boldsymbol{v}_k^{\mathrm{H}}\right\|_2 = 1$. For given values of φ_k and given per-user powers p_k, the rates of all users can be computed as

$$r_k = \log_2\left(1 + p_k\tilde{\boldsymbol{h}}_k^{\mathrm{H}}\left(\mathbf{I}_{MC} + \sum_{j\neq k} p_j\tilde{\boldsymbol{h}}_j\tilde{\boldsymbol{h}}_j^{\mathrm{H}}\right)^{-1}\tilde{\boldsymbol{h}}_k\right) \tag{13}$$

in the dual uplink [14] of the effective vector broadcast channel.

Applying the matrix inversion lemma, these equations can be rewritten such that they only depend on the powers p_k and on the inner products $\tilde{\boldsymbol{h}}_i^{\mathrm{H}}\tilde{\boldsymbol{h}}_j$ between the effective channels. As functions of the phases φ_k, the values of these inner products are given by

$$\tilde{\boldsymbol{h}}_1^{\mathrm{H}}\tilde{\boldsymbol{h}}_2 = \frac{1}{2\sqrt{2}}\left(1 + \mathrm{e}^{\mathrm{j}(\varphi_1-\varphi_2)}\right) \tag{14}$$

$$\tilde{\boldsymbol{h}}_2^{\mathrm{H}}\tilde{\boldsymbol{h}}_3 = \frac{\mathrm{e}^{-\mathrm{j}\frac{\pi}{4}}}{2\sqrt{2}}\left(1 + \mathrm{e}^{\mathrm{j}(\varphi_2-\varphi_3+\frac{\pi}{2})}\right) \tag{15}$$

$$\tilde{\boldsymbol{h}}_3^{\mathrm{H}}\tilde{\boldsymbol{h}}_1 = \frac{1}{2\sqrt{2}}\left(1 + \mathrm{e}^{\mathrm{j}(\varphi_3-\varphi_1)}\right). \tag{16}$$

To obtain high rates, the absolute values of these inner products should be as small as possible for $i \neq j$, and from symmetry considerations, it can be concluded that they should all be equal, which is fulfilled for $\varphi_1 - \varphi_2 = \varphi_2 - (\varphi_3 - \frac{\pi}{2}) = \varphi_3 - \varphi_1$. This leads to the choice $\varphi_1 = 0$, $\varphi_2 = -\frac{5\pi}{6}$, and $\varphi_3 = \frac{5\pi}{6}$, which is visualized in Fig. 2.[3] This particular choice yields high angular separations between the channel vectors, i.e., $\cos\theta_{k,j} = |\tilde{\boldsymbol{h}}_k^{\mathrm{H}}\tilde{\boldsymbol{h}}_j|$ is given by

$$\left|\tilde{\boldsymbol{h}}_k^{\mathrm{H}}\tilde{\boldsymbol{h}}_j\right| = \begin{cases}1, & \text{if } k = j,\\ \frac{1}{\sqrt{2}}\cos\left(\frac{5\pi}{12}\right) \approx 0.1830, & \text{if } k \neq j,\end{cases} \tag{17}$$

for all $k, j \in \{1,2,3\}$.

[2] $(\boldsymbol{A}+\boldsymbol{BCD})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{C}^{-1}+\boldsymbol{DA}^{-1}\boldsymbol{B})^{-1}\boldsymbol{DA}^{-1}$ (e.g., [15]).

[3] The choice is only symmetric with respect to user 2 and user 3 due to the phase shift $+\frac{\pi}{2}$ in (15).

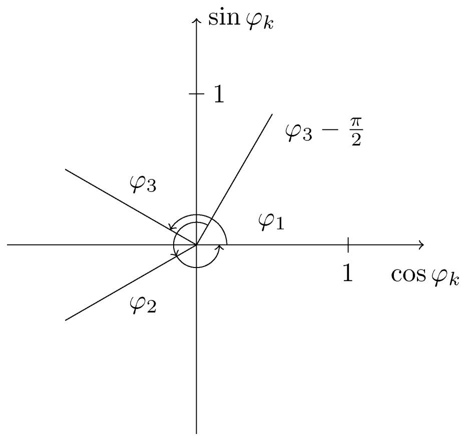

Fig. 2. Visualization of the choice for φ_k.

Choosing the symmetric power allocation $p_k = P/3\ \forall k$, which fulfills $\sum_{k=1}^{K} p_k = P$, we achieve the rates $r_k \approx 1.0986\ \forall k$. Clearly, the rate vector $\boldsymbol{r} \approx 1.0986\boldsymbol{\rho}$ lies outside of the rate region for separate coding on each subchannel with sum power P. ■

Note that instead of increasing the rates, the advantage of joint transmission over separate transmission can also be used to reduce the transmit power. After choosing the phases φ_k, the minimal transmit power needed to achieve the rate vector $\boldsymbol{\rho}$ in the effective vector broadcast channel can be computed numerically using one of the methods proposed in [16] and [17]. In the case discussed above, this would lead to a sum transmit power $P_{\text{joint}} \approx 3.1174 < P$.

To get some intuition about why the (potentially suboptimal) scheme with joint treatment of the subchannels achieves higher rates than the optimal scheme with subchannel separation, it helps to look at the Hermitian angles between the channel vectors. With the proposed choice for the phases φ_k in the case of joint encoding, the data can be transmitted across very dissimilar effective channels that enclose Hermitian angles $\theta_{k,j} \approx 79.5°$ [cf. (17)] while the Hermitian angles $\theta_{k,j}^{(c)} = 45°$ between the original channel vectors on each subchannel are significantly smaller [cf. (8)]. Apparently, the loss in performance resulting from the small angular separation of the channels in the separate case cannot be compensated by the fact that two subchannels are available and by the use of optimal time-sharing.

To show that (in)separability is a property inherent to the channel realization, we also state the following proposition, which can again be proven by construction.

Proposition 1: For certain channel realizations, parallel MIMO broadcast channels with linear transceivers are separable.

Proof of Proposition 1: Consider a channel realization where for all $c \neq 1$, the channel matrices are given by $\boldsymbol{H}_k^{(c),\mathrm{H}} = \boldsymbol{0} \forall k$. As transmission over the zero-channels is neither possible with separate nor with joint treatment of the subchannels, the parallel broadcast channels are obviously separable for this channel realization. ■

IV. Examples for Inseparable Channel Realizations

In order to get more intuition about when inseparability does occur, we keep considering the system introduced in the proof of Theorem 1, but we look at more channel realizations. If we ignore negligible rotations that change only the absolute, but not the relative spatial directions of the channel vectors, all possible channel realizations on subchannel c can be parametrized as follows:

$$\boldsymbol{h}_1^{(c),\mathrm{H}} = \ell_1^{(c)} \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{18}$$

$$\boldsymbol{h}_2^{(c),\mathrm{H}} = \ell_2^{(c)} \begin{bmatrix} \cos\alpha_1^{(c)} & \sin\alpha_1^{(c)} \end{bmatrix} \tag{19}$$

$$\boldsymbol{h}_3^{(c),\mathrm{H}} = \ell_3^{(c)} \begin{bmatrix} \cos\alpha_3^{(c)} & \mathrm{e}^{\mathrm{j}\psi^{(c)}} \sin\alpha_3^{(c)} \end{bmatrix}. \tag{20}$$

Restricting ourselves to channels with norm $\ell_k^{(c)} = 1\ \forall k,\ \forall c$, and choosing

$$\cos\psi^{(c)} = \frac{\cos^2\alpha_2^{(c)} - \cos^2\alpha_1^{(c)}\cos^2\alpha_3^{(c)} - \sin^2\alpha_1^{(c)}\sin^2\alpha_3^{(c)}}{2\cos\alpha_1^{(c)}\cos\alpha_3^{(c)}\sin\alpha_1^{(c)}\sin\alpha_3^{(c)}}, \tag{21}$$

the channel on subchannel c is parametrized by three Hermitian angles

$$\theta_{1,2}^{(c)} = \arccos\left|\boldsymbol{h}_1^{(c),\mathrm{H}}\boldsymbol{h}_2^{(c)}\right| = \alpha_1^{(c)} \in \left[0, \frac{\pi}{2}\right] \tag{22}$$

$$\theta_{2,3}^{(c)} = \arccos\left|\boldsymbol{h}_2^{(c),\mathrm{H}}\boldsymbol{h}_3^{(c)}\right| = \alpha_2^{(c)} \in \left[0, \frac{\pi}{2}\right] \tag{23}$$

$$\theta_{3,1}^{(c)} = \arccos\left|\boldsymbol{h}_3^{(c),\mathrm{H}}\boldsymbol{h}_1^{(c)}\right| = \alpha_3^{(c)} \in \left[0, \frac{\pi}{2}\right] \tag{24}$$

which have to fulfill

$$\sum_{i=1}^{3}\alpha_i^{(c)} \leq \pi \quad \text{and} \quad 2\alpha_k^{(c)} - \sum_{i=1}^{3}\alpha_i^{(c)} \leq 0\ \forall k \tag{25}$$

for (21) to be feasible. Note that $\psi^{(1)}$ can be assumed to be positive without loss of generality, but then, the positive as well as the negative solution of (21) have to be considered for $\psi^{(c)}$ with $c > 1$.

For a parametrization of the considered system with two parallel channels, we would need six angles and the choice for the sign of $\psi^{(2)}$. As the aim of this section is only to find examples for inseparable channel realizations, we will restrict ourselves to the special case of symmetric channel realizations, where the angles $\alpha_k^{(c)}$ have the same value α for all users and on both subchannels. In this case, the scenario can be parametrized by one angle[4] $\alpha \leq \frac{\pi}{3}$ and the sign of $\psi^{(2)}$. The channel realization considered in the proof of Theorem 1 is obtained for $\alpha = 45°$ and $\psi^{(2)} = -\psi^{(1)}$.

For other values of α, the optimal separate strategy can be computed by inserting the respective value of the inner product $|\boldsymbol{h}_k^{(c),\mathrm{H}}\boldsymbol{h}_j^{(c)}| = \cos\alpha$ into (10). If the resulting P is smaller than P_1 from (9), it is the minimal power for separate transmission. Otherwise, P_1 is the optimal power. In Fig. 3, the interval where the solid curve is constant corresponds to P_1 being the optimal power, i.e., in this interval, single-user transmission on each subchannel is optimal.

If the aim is to use the advantage of joint transmission to reduce the transmit power, a suboptimal power minimization algorithm for MIMO broadcast channels with linear transceivers as proposed in [18], [19], or [20] can be applied. If the aim is to increase the per-user rates, as was done in the proof of Theorem 1, a suboptimal rate balancing algorithm

[4]The inequality is due to (25).

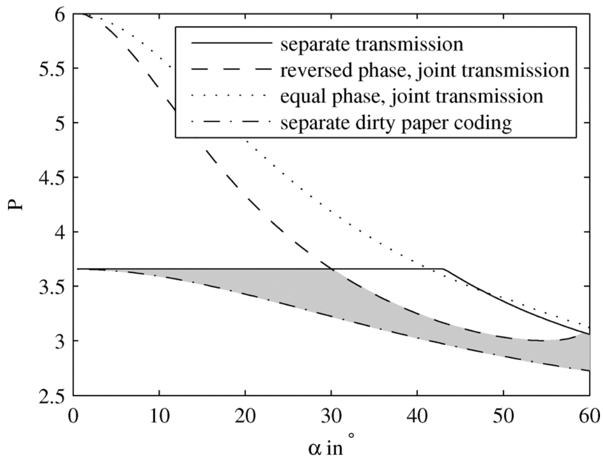

Fig. 3. Transmit power needed to achieve $\boldsymbol{\rho} = [1,1,1]^{\mathrm{T}}$ with optimal separate and with suboptimal joint transmission.

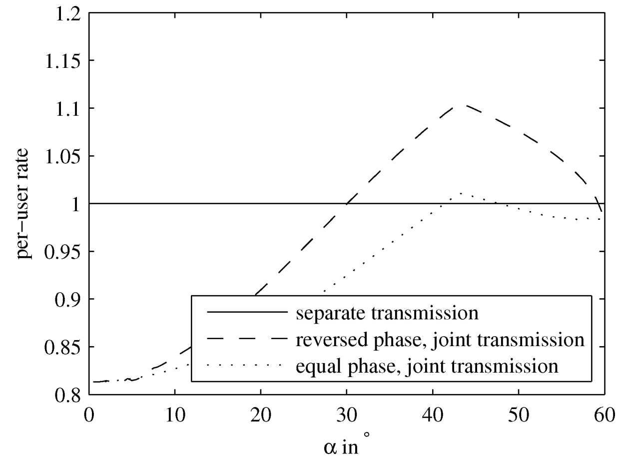

Fig. 4. Per-user rate achieved by suboptimal joint transmission with a transmit power sufficient to achieve $\boldsymbol{\rho} = [1,1,1]^{\mathrm{T}}$ with optimal separate transmission.

as proposed in [18], [19], or [21] can be used. All these algorithms refrain from using time-sharing.

For the results presented in Fig. 3, we have used the power minimization algorithm from [20]. The terms "reversed phase" and "equal phase" refer to the sign of $\psi^{(2)}$, i.e., the curves refer to different channel realizations with the same angle α. For separate transmission, the sign of $\psi^{(2)}$ does not play any role.

It can be seen that especially for the channel realizations with reversed phase on the second subchannel, there is a large angular range where the suboptimal joint transmission can outperform the optimal separate transmission. For the channel realizations with equal sign of $\psi^{(c)}$ on both subchannels, this observation applies at least for a small angular range. If α lies within these intervals, the parallel broadcast channels with linear transceivers are inseparable. The example discussed in the proof of Theorem 1 can be found in the plot by comparing the solid curve and the dashed curve at $\alpha = 45^\circ$.

Whenever the curves for suboptimal joint transmission lie above the curve for separate transmission, this is a result of the suboptimality of the algorithm applied in the joint case: the optimal separate strategy is also a valid joint strategy and could be applied instead of the joint strategy computed by the suboptimal power minimization algorithm. However, in these cases, there might be an even better joint strategy that could be found by some other algorithm. Therefore, if α lies in one of these intervals, it is not clear whether or not the parallel broadcast channels with linear transceivers are separable. To further study these intervals, we have also included the curve for separate transmission with nonlinear dirty paper coding (DPC), which is known to be the optimal strategy in parallel MIMO broadcast channels (e.g., [4]). In the case with reversed phase, the (unknown) optimum for joint linear transmission has to lie inside the gray area as it can never lie below the DPC curve. Apparently, for small angles α the height of this area diminishes, which implies that the distance to separability decreases for small α, i.e., the scenario with linear transceivers must at least be close to being separable if α is close to zero.

To produce the plot in Fig. 4, we have first computed the transmit power needed to achieve $\boldsymbol{\rho} = [1,1,1]^{\mathrm{T}}$ with separate transmission. Then, by means of the algorithm from [21], we have computed a suboptimal rate balancing solution using the same power. As expected, we get an increased rate in the intervals where joint transmission led to a power reduction in Fig. 3. The example discussed in the proof of Theorem 1 can again be found in the plot by comparing the solid curve and the dashed curve at $\alpha = 45^\circ$.

Even though we have restricted ourselves to symmetric channel realizations in this section, we have also been able to observe inseparability also for asymmetric ones. Just like in the symmetric case, inseparability has happened as long as the channels have been neither close to being orthogonal nor close to being linearly dependent, i.e., for moderate values of the angles α_k.

V. Discussion and Outlook

By constructing an adequate channel realization, we have proven that parallel MIMO broadcast channels are inseparable in systems that employ linear transceivers to transmit circularly symmetric complex Gaussian data symbols. The result might affect any such system since no further assumptions on the transmit strategy were imposed. In particular, the inseparability cannot be resolved by the application of time-sharing. Quite the contrary, for certain channel realizations, a joint strategy without time-sharing might outperform a strategy based on subchannel separation that uses time-sharing.

Allowing time-sharing in the joint strategy can only enlarge the rate region of the joint case, just as excluding time-sharing in the separate strategy can only reduce the size of the rate region of the separate case. From this, it can be easily concluded that inseparability still holds if both strategies are allowed to employ time-sharing as well as if both strategies are restricted to not make use of time-sharing.

As the theorem is based on a rate region formulation, it implies that subchannel separation might, depending on the actual channel conditions, result in suboptimal performance for any reasonable optimization criterion, including (weighted) sum rate maximization, quality of service constrained problems, rate balancing, and fairness based metrics. This can be easily concluded from the fact that the optimal transmit strategies for these criteria always correspond to points on the Pareto boundary of the rate region.

Interestingly, the underlying reasons for the inseparability of parallel MIMO broadcast channels with linear precoding seem to differ significantly from those for the inseparability of interference channels. As mentioned in [1], interference alignment [22], [23] is a way to find strategies with joint encoding over multiple frequencies that can outperform separate strategies. However, this technique is not applicable to the broadcast system considered in this paper, so that the reasons for inseparability must be different. With the following considerations, we will try to shed some light on them.

If the rate is to be maximized in a single-user MIMO channel with block-diagonal channel matrices and noise covariance matrices, the optimal strategy according to [24] obviously uses block-diagonal transmit covariance matrices and is, thus, a separate strategy. Intuitively, it is not surprising that in a broadcast channel with DPC, the same has to hold for the user that does not receive any interference. Therefore, the data transmitted for this user has a block-diagonal covariance matrix, and the next user, who receives just this one interfering signal, has a block-diagonal interference-plus-noise covariance matrix at the receiver, too. Consequently, for the next user, it also makes sense to use a block-diagonal transmit covariance matrix, i.e., a separate strategy. This reasoning can be successively applied to all users in order to intuitively understand the optimality of separate strategies in parallel broadcast channels with DPC.

If linear precoding is applied, each user receives interference caused by the signals of all other users, and a successive study as in the DPC case is no longer possible. Let us instead consider an arbitrary user k. If all other users apply a separate strategy, we have a block-diagonal interference-plus-noise covariance matrix at the kth receiver, and intuition tells us to use a separate strategy for user k as this maximizes the rate of user k for given strategies of the other users. However, as soon as one user transmits with a covariance matrix that does not match the block-diagonal structure of the channels, all other users get interference-plus-noise covariance matrices that are no longer block-diagonal, and it is no longer clear that the transmit covariance matrices of the signals intended for them should be block-diagonal.

Based on the result presented in this paper, it is clear that any approach that is claimed to be optimal for parallel MIMO broadcast channels with linear precoding has to include the possibility of joint coding across the subchannels, although this possibility is not required for optimality in systems with optimal nonlinear dirty paper precoding. The implications of this insight are far-reaching. First of all, the optimization of a transmit strategy without subchannel separation involves solving mathematical optimization problems in far more variables than in the case of separate treatment of the subchannels. The reason for this is that also the off-diagonal blocks of the beamforming matrices $\boldsymbol{B}_k$ and the receive filters $\boldsymbol{V}_k$ need to be optimized. Beyond that, allowing joint encoding across the subchannels might also change the qualitative properties of the optimization problems. In particular, in the case of parallel vector broadcast channels, i.e., in the case $N_k = 1 \forall k$, the receive filters on each subchannel are reduced to scalars, leading to pure power allocation problems in the dual uplink if the subchannels are treated separately. For instance, for the power minimization problem with per-user rate constraints, this simplified problem can be solved in a globally optimal manner as was shown in [10]. However, due to the structural difference between scalar receive filters and the receive filter matrices needed for the case without subchannel separation, these solutions cannot be transferred to the case where joint coding across subchannels is allowed.

These examples make clear that finding the optimal linear transmit strategy is far more involved when the subchannels are not separated. Therefore, a highly interesting topic for future research is to derive conditions under which a set of parallel MIMO broadcast channels is separable. Furthermore, it would be interesting to derive bounds on the maximal loss in performance resulting from separation of an inseparable setting.

References

[1] V. Cadambe and S. Jafar, "Parallel Gaussian interference channels are not always separable," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3983–3990, Sep. 2009.

[2] D. N. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," presented at the Int. Symp. Inf. Theory (ISIT), Ulm, Germany, Jun. 29–Jul. 4, 1997.

[3] D. N. C. Tse and S. V. Hanly, "Multi-access fading channels—Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.

[4] M. Mohseni, R. Zhang, and J. Cioffi, "Optimized transmission for fading multiple-access and broadcast channels with multiple antennas," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1627–1639, Aug. 2006.

[5] P. Tejera, W. Utschick, J. Nossek, and G. Bauch, "Rate balancing in multiuser MIMO OFDM systems," *IEEE Trans. Commun.*, vol. 57, no. 5, pp. 1370–1380, May 2009.

[6] N. Hassan and M. Assaad, "Margin adaptive resource allocation in downlink multiuser MIMO-OFDMA system with multiuser eigenmode transmission," in *Proc. Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2008, pp. 545–549.

[7] N. Hassan and M. Assaad, "Low complexity margin adaptive resource allocation in downlink MIMO-OFDMA system," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3365–3371, Jul. 2009.

[8] C. Guthy, W. Utschick, and G. Dietl, "Spatial resource allocation for the multiuser multicarrier MIMO broadcast channel—A QoS optimization perspective," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2010, pp. 3166–3169.

[9] F. She, W. Chen, H. Luo, T. Huang, and X. Wang, "Joint power allocation and scheduling of multi-antenna OFDM system in broadcast channel," in *Proc. Int. Conf. Commun. (ICC)*, Dresden, Germany, Jun. 14–18, 2009, pp. 1–5.

[10] C. Hellings, M. Joham, M. Riemensberger, and W. Utschick, "Minimal transmit power in parallel vector broadcast channels with linear precoding," *IEEE Trans. Signal Process.*, 2010, submitted for publication.

[11] M. Joham, C. Hellings, and R. Hunger, "QoS feasibility for the MIMO broadcast channel: Robust formulation and multi-carrier systems," in *Proc. WiOpt 2010*, May 2010, pp. 610–614.

[12] D. P. Palomar, M. A. Lagunas, and J. M. Cioffi, "Optimum linear joint transmit-receive processing for MIMO channels with QoS constraints," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1179–1197, May 2004.

[13] K. Scharnhorst, "Angles in complex vector spaces," *Acta Applicandae Mathematicae*, vol. 69, pp. 95–103, 2001.

[14] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[15] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[16] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.

[17] A. Mezghani, M. Joham, R. Hunger, and W. Utschick, "Transceiver design for multi-user MIMO systems," presented at the Int. ITG/IEEE Workshop on Smart Antennas (WSA), Ulm, Germany, Mar. 13–14, 2006.

[18] S. Shi, M. Schubert, and H. Boche, "Capacity balancing for multiuser MIMO systems," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2007, vol. 3, pp. III-397–III-400.

[19] S. Shi, M. Schubert, and H. Boche, "Rate optimization for multiuser MIMO systems with linear processing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4020–4030, Aug. 2008.

[20] C. Hellings, M. Joham, and W. Utschick, "Gradient-based power minimization in MIMO broadcast channels with linear precoding," *IEEE Trans. Signal Process.*, 2011, accepted for publication.

[21] C. Hellings, M. Joham, and W. Utschick, "Gradient-based rate balancing for MIMO broadcast channels with linear precoding," presented at the Int. ITG Workshop on Smart Antennas (WSA), Aachen, Germany, Feb. 24–25, 2011.

[22] S. Jafar and S. Shamai, "Degrees of freedom region of the MIMO X channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 151–170, Jan. 2008.

[23] V. Cadambe and S. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[24] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.