

# Spatial transmission mode switching in multi-user MIMO-OFDM systems with user fairness

Malte Schellmann, *Student Member, IEEE*, Lars Thiele, *Student Member, IEEE*, Thomas Haustein, *Member, IEEE* and Volker Jungnickel, *Member, IEEE*

**Abstract**—Multi-antenna radio systems allow to access the channel in diversity or spatial multiplexing mode. Adequate switching between these modes according to current channel conditions was shown to yield significant performance improvements while requiring little feedback from the receiving side. We present a transmission concept for the downlink of a multi-user MIMO-OFDM system aiming at high user rates with limited feedback demands. An extended score-based scheduling approach ensures a fair resource allocation to the users, while transmission mode switching is used to guarantee high user rates. The degree of fairness of the scheduler can be adapted by adequately configuring a weighting function for the scores. Comparison to single-mode schemes reveals substantial throughput gains of the adaptive switching concept. Furthermore, targeting maximum throughput, we show that a considerable proportion of the capacity of the MIMO broadcast channel can be achieved with a comparatively low amount of required feedback.

## I. INTRODUCTION

The spatial dimension of the multiple-input multiple-output (MIMO) radio channel can be accessed in spatial multiplexing (SMUX) or in spatial diversity transmission mode. While in SMUX mode multiple data streams are transmitted simultaneously via the radio channel, the diversity mode utilizes the independent propagation paths between different antenna pairs to improve the quality of the signal transmission. Early results from investigations on system-level [1] indicated that the mode achieving the highest spectral efficiency depends on the actual channel and signal conditions experienced by a user. Later, this relationship was substantiated by the fundamental tradeoff pointed out between diversity and SMUX mode in [2]. This crucial finding motivated the development of an adaptive transmission system which selects the transmission mode depending on the actual channel quality in order to improve the error rate performance for fixed data rate transmission [3], [4] or to increase the spectral efficiency [5]–[8]. In those work, the channel conditions are evaluated at the receiver's side, and the decision on the mode together with a channel quality identifier (CQI) is fed back to the transmitter, resulting in little feedback demands. While the aforementioned concepts benefit from adapting the transmission to the instantaneous channel conditions, it was shown in [9] that the capacity can

also be enhanced if the mode selection is based on the long-term channel statistics represented by the covariance matrix of the MIMO channel.

In recent years, the focus in MIMO communications has been shifted from point-to-point towards multi-user links [10]. The available multi-user diversity [11] in those systems enables further enhancements of the achievable spectral efficiency. The key idea is to assign the simultaneously transmitted streams of the SMUX mode to different users, thus enabling space-division multiple access (SDMA). The capacity of the MIMO broadcast channel (BC) exploiting SDMA transmission has been shown in [12]–[14] to be achievable with the dirty paper coding technique [15]. For individual user constraints like desired target signal to interference and noise ratios (SINRs), a framework for the optimum solution of the multi-user downlink beamforming problem has been presented in [16]. All those solutions are based on the assumption that full channel state information (CSI) is available at the base station (BS). In recent work [17], information theory analysis revealed that for the low SNR regime, it is optimum to transmit a single stream to a single user only, while with increasing SNR, support for additional SDMA users should successively be added. This result already suggests that mode switching can be applied beneficially also in multi-user MIMO systems.

For proper application of SDMA, information on the state of each user's channel is required at the BS. As availability of full CSI from all the users at the BS is hardly possible in frequency division duplex (FDD) systems, solutions based on limited feedback have been introduced that provide partial CSI to the BS [18]. A promising concept with comparatively low demands on feedback is opportunistic beamforming [19], where the BS generates random precoding vectors (beams). These are evaluated at the user terminals (UTs) in terms of their reception SNR, which is then reported to the BS. In case of a high SNR value, it is likely that the selected precoding vector points close into the direction of the actual channel, thus yielding the desired partial CSI. While the original concept was designed for the transmission of a single stream to a single user, a solution for SDMA has been introduced in [20], where the BS provides sets of unitary beams that are served simultaneously with equal power. In this latter work, it was proposed that the UTs determine the SINR for each received beam and convey the index of the best received beam together with the corresponding SINR value to the BS, which then assigns each beam to the user achieving highest rate. It was shown that for a large number of users, the achievable downlink capacity of this system scales equivalently to the

Part of the material in this paper has been presented at the Asilomar Conference, Pacific Grove, CA, October 2007.

M. Schellmann, L. Thiele and V. Jungnickel are with the Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut, Einsteinufer 37, 10587 Berlin, Germany

Thomas Haustein is with Nokia Siemens Networks, Sankt-Martin Str. 76, 81541 Munich, Germany

capacity of the MIMO BC.

To ease system implementation, fixed instead of random beams can be used, as suggested in proposals for next generation radio systems [21], [22]. As the fixed beams are available at any time, the user does not have to wait until a beam is provided that matches his channel conditions, and thus potential latency problems can be avoided.

While [19], [20] exclusively support a single transmission mode, initial performance results on mode switching in multi-user links based on a fixed set of unitary beams have been presented in [8], where the throughput performance of a  $2 \times 2$  MIMO configuration with two users was compared to a  $4 \times 2$  point-to-point link. Further results for single-antenna UTs have been presented in [23], [24], where the switching problem was formulated in terms of a balance between multi-user diversity and multiplexing gain. Feedback is given in terms of CQI for all provided beams, which is still a comparatively small amount. [23] further reports significant performance gains compared to the single mode schemes [19], [20], revealing that the mode switching concept has a high potential especially in cases where the number of available users is not exceptionally large (sparse network [10]).

In a multi-user system we have to care for fairness in the resource assignment process. In particular, it has to be ensured that resources will also be assigned to users experiencing relatively poor channel conditions. This can be achieved by applying scheduling policies at the BS tailored to meet predefined fairness constraints. One of these policies is the proportional fair scheduler [19], which enables each user to realize a constant fraction of his total achievable rate. The score-based scheduler proposed in [25] represents an efficient heuristic approach that aims at assigning all users the same amount of resources. In case all users have identical fading conditions, it asymptotically achieves proportional fairness. The set of resources to be scheduled can be defined arbitrarily over all dimensions (time/frequency/space) and can be flexibly adapted to any latency constraints. Due to these properties, the approach enables a simple and efficient resource allocation process, which makes it a favourable solution for the employment in practical systems.

The motivation for the work presented in this paper was the development of a practical concept for downlink transmission in a real-time multi-user MIMO-OFDM system, aiming at a high throughput while meeting desired fairness constraints within a predefined time frame. Therefore, we extend the score-based scheduling approach for the support of spatial mode switching, including SDMA access. Precoding based on fixed beams is adopted to obtain partial CSI at the BS. Opposed to existing work in the field [23], we assume UTs with multiple antennas, which enable interference suppression to improve the SINRs of the received beams. Assuming linear receivers, the UTs determine the achievable per-beam rates for the different spatial transmission modes and convey information on their preferred beams and the corresponding rates per mode to the BS via a low-rate feedback channel. This serves as the input for the extended score-based scheduler, which implicitly selects the spatial transmission mode per user. By introducing a weighting of the user-specific scores

generated by the scheduler, we can adapt the targeted degree of fairness, which is illustrated with the example of equal rates for all users. The capabilities of the scheduling concept are numerically analyzed for a  $2 \times 2$  MIMO configuration, showing that mode switching as well as enabling SDMA access lead to substantial performance gains. The downlink capacity achievable with the proposed system concept for the  $2 \times 2$  MIMO configuration is finally compared with the capacity of the MIMO BC based on full CSI. This comparison reveals that the proposed system concept is capable of realizing a large proportion of that BC capacity.

The paper is structured as follows: The system model and the fundamental ideas of the concept are sketched in section II. Section III provides a brief overview on the applied fixed beamforming concept, termed the grid of beams (GoB). In section IV, the concept of the adaptive downlink with fair multi-user scheduling is described in detail. The system is then investigated on link-level, i.e. for multiple users in a single (isolated) cell, for the exemplar case of a  $2 \times 2$  MIMO configuration in section V. Section VI draws some conclusions.

## II. SYSTEM MODEL

We consider the downlink of a broadband multi-user MIMO-OFDM system, where a BS with  $N_t$  antennas communicates with  $K$  UTs equipped with  $N_r$  antennas each. The BS provides  $B \geq N_t$  fixed beams  $\mathbf{b}_i$ , which are used for spatial precoding of the transmission signals (GoB concept, presented in [26]). We assume a uniform transmit power allocation over all subcarriers; hence the transmission equation for each subcarrier's signal is given by

$$\mathbf{r} = \mathbf{H}\mathbf{C}\mathbf{s} + \mathbf{n} \quad (1)$$

where  $\mathbf{H}$  is the  $N_r \times N_t$  dimensional MIMO channel and  $\mathbf{C}$  is the precoding matrix comprising  $N_t$  of the  $B$  beams  $\mathbf{b}_i$  with unitary property, which may be simultaneously active. The transmit vector  $\mathbf{s}$  contains up to  $N_t$  non-zero transmit symbols with constant transmit power  $E\{\mathbf{s}^H\mathbf{s}\} = P_s$ , i.e. the power  $P_s$  is uniformly distributed over all non-zero transmit symbols in  $\mathbf{s}$ . Finally,  $\mathbf{n}$  is the noise vector with  $N_r$  circularly symmetric complex Gaussian entries, and its covariance is given by  $E\{\mathbf{n}\mathbf{n}^H\} = N_0 \cdot \mathbf{I}$ .  $E\{\cdot\}$  is the expectation operator,  $(\cdot)^H$  the conjugate transpose operator and  $\mathbf{I}$  represents the identity matrix.

Transmission is based on a slotted time structure, where each slot is constituted of several consecutive OFDM symbols. The total transmission resources of a slot are given by the subcarriers available for signal transmission. By subdividing the signal bandwidth into single sub-bands confined to a fixed number of consecutive subcarriers, we partition these transmission resources into blocks, which are denoted as *chunks* in the following. The chunks form the basic scheduling resources that can be individually assigned to distinct users. Each chunk is processed separately, and thus the spatial transmission mode may be selected individually per chunk. The supported spatial modes are termed as single-stream (ss) and multi-stream (ms), which relate to the diversity mode and

SMUX mode, respectively. Thanks to the multiple antennas at the UTs, a single user can also receive multiple spatial streams in ms mode. To differentiate between the two cases in ms mode where the available spatial streams are assigned either to a single user or to multiple different users, the terms single-user MIMO (SU-MIMO) and multi-user MIMO (MU-MIMO) are used, respectively.

The channel is assumed to be perfectly known at the receiving UTs only. Based on the GoB, the UTs can evaluate the channel per chunk and determine the rates they are able to achieve in the different spatial modes with each beam. Beams achieving highest rate are selected for potential transmission. Each user conveys information on the beam selection per spatial mode and the achievable rate per beam to the BS, which then carries out the user scheduling process. To simplify the concept further, we assume linear equalization techniques are being applied at the multi-antenna receivers. Hence, to recover the  $i$ -th symbol  $s_i$  within vector  $\mathbf{s}$ , we multiply the received vector  $\mathbf{r}$  with the equalization vector  $\mathbf{w}_i$  according to

$$\hat{s}_i = \mathbf{w}_i^H \mathbf{r} \quad (2)$$

In this paper, we consider maximum ratio combining (MRC) for ss mode and minimum mean square error (MMSE) equalization for ms mode. For these techniques, closed form expressions for the post-detection SINRs<sup>1</sup> of the spatial streams exist, facilitating the evaluation and proper comparison of the rates that can be achieved with the different spatial modes.

The applied scheduling process is based on the score-based scheduling strategy [25], which is a simple heuristic process aiming to assign each user his best resources from a set of resources defined over arbitrary dimensions: The resources of each user within the set are ranked by their quality, and corresponding scores are assigned. The BS then assigns a resource to the user providing the best score. On average, this scheduling strategy assigns an equal amount of resources to each user. For identical fading statistics, each user thus can realize a constant fraction of his total achievable rate, yielding (asymptotically) a degree of fairness similar to that of the proportional fair scheduler [19]. In our research, we confine the set of resources to those comprised in a transmission slot. Hence, all users will be scheduled within the same time-slot, and the score-based scheduling translates to an even distribution of the chunks over all users. In this work, we consider two kinds of fairness constraints: First, fairness in terms of an equal number of resources being assigned to the users. Hereby, we account for the users' capabilities to support transmission of multiple spatial streams in ms mode, i.e. users supporting ms mode are assigned correspondingly more spatial streams than users supporting ss mode. On the other hand, we show in section IV-E that the fairness target of the score-based scheduler can be altered by introducing a weighting of the scores. As an example, we tune the scheduler to achieve fairness in terms of an equal rate for all users in the system.

<sup>1</sup>The post-detection SINR is the SINR of the useful signal achieved after equalization.

### III. PRECODING: THE GRID OF BEAMS

In this section, we will give some details on the proper choice of the vectors constituting the GoB: The GoB concept itself can be understood as a quantization scheme for the precoding vectors to be used for signal transmission. It is well known that the optimum choice of precoding vectors in single-user MIMO links are the eigenvectors of the matrix product  $\mathbf{H}^H \mathbf{H}$ . The optimum quantization of these vectors in uncorrelated MIMO Rayleigh-fading channels has been elaborated in [27] for diversity transmission and in [28] for SMUX transmission. For the Rayleigh-fading assumption, it follows an isotropic distribution of the eigenvectors in the  $N_t$ -dimensional vector space of  $\mathbf{H}^H \mathbf{H}$ . Under this condition, the authors in [27] formulated the problem of finding the optimum quantization vectors as a subspace packing problem in the manifold of all linear subspaces of this vector space, which is called the Grassmannian manifold. By using a suited distance metric, the optimum set of quantization vectors for diversity mode is found as the one where the minimum distance between any pair of these vectors is maximized. For SMUX mode, the optimum solution is a set of unitary  $N_t \times N_r$  matrices, which are found in a similar fashion, but based on different distance metrics.

However, due to per-antenna power constraints encountered in practice, it is desirable to use beams for precoding that distribute the power uniformly over the transmit antennas. In the literature, this is termed equal gain transmission (EGT). Quantization of precoding vectors for EGT has been studied in [29], where it was shown for uncorrelated Rayleigh-fading that a set of  $N_t$  unitary vectors is sufficient to guarantee achieving the full diversity gain of the channel  $\mathbf{H}$ . To obtain  $B = rN_t$  suitable precoding vectors for EGT, with  $r$  being an integer specifying the number unitary beam sets with  $N_t$  beams, the authors in [29] propose to take the first  $N_t$  rows of the  $rN_t$ -dimensional discrete Fourier transform (DFT) matrix. These precoding vectors are commonly referred to as DFT beams.

In practice, MIMO channels are often spatially correlated to some extent, so that the uncorrelated Rayleigh fading assumption, which was the basis for the investigations mentioned above, does not hold in general. One of the major differences is that in correlated channels, the distribution of the channel's eigenmodes is biased. This has a direct impact on the achievable spectral efficiency, which has been studied in [30]. In that work, it was shown that in case of an isotropic input (equal power per active beam,  $N_t$  beams simultaneously active), the channel correlation diminishes the achievable spectral efficiency compared to Rayleigh fading. Opposed to that, it was also shown that for a non-isotropic input (less than  $N_t$  beams are simultaneously active), the spectral efficiency can be improved.

If the channels are spatially correlated, the dominant eigenmodes are concentrated around the dominant eigenvector of the channel's covariance matrix  $\mathbf{R}_H = E\{\mathbf{H}^H \mathbf{H}\}$ . If the channels between transmit/receive antenna pairs have the same statistical properties – which is a common assumption also for realistic MIMO channels –  $\mathbf{R}_H$  is a Toeplitz matrix with hermitian property, i.e.  $\mathbf{R}_H^H = \mathbf{R}_H$ , with a real-valued

diagonal and, in general, with all entries being non-zero. The dominant eigenvector of a matrix with these properties is close to the  $N_t$ -dimensional vector with equal-weighted entries. This vector fulfils the properties of an EGT vector, suggesting that precoding based on DFT beams as described above may be a suitable choice especially for the case of correlated MIMO channels.

In a multi-user scenario, a solution achieving near-optimal performance is to serve multiple users simultaneously on their dominant channel eigenmodes [31]. This result suggests that the same quantization techniques from above can also be deemed suitable in the multi-user context.

Remarks from above suggest that DFT beams are a convenient choice for GoB-based precoding in practical systems, and hence we have also adopted them for our research. For SMUX transmission, we allow only sets of unitary beams to be active simultaneously, as precoding matrices of this kind were shown to optimize various performance measures in single user links (see [4] and references therein). Hence, the beams provided by the BS are constituted from  $r$  independent sets of  $N_t$  unitary DFT beams, i.e.  $B = rN_t$ . For example, for  $N_t = 2$ , the  $r = 2$  unitary DFT beam sets  $\mathbf{C}_1$  and  $\mathbf{C}_2$  that will be used as precoding matrices in (1) are given as:

$$\mathbf{C}_1 = \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}, \quad \mathbf{C}_2 = \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad (3)$$

where the  $B = 4$  beams  $\mathbf{b}_i, i \in \{1, \dots, B\}$  are given as the columns of the two matrices.

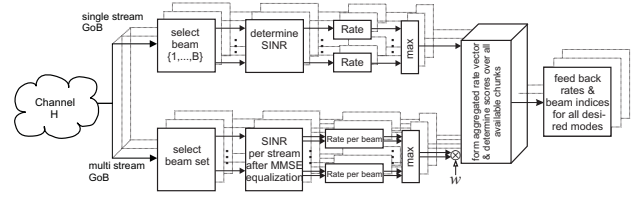
#### IV. SCHEDULER WITH SPATIAL MODE SELECTION

We will now describe the adaptive transmission concept, which is based on a 2-step procedure illustrated in Fig. 1. It consists of a channel evaluation unit at the side of the UT (step 1) and the resource scheduling and transmission mode selection unit at the BS (step 2). In step 1, a user carries out a chunk-wise evaluation of the different transmission modes and determines the achievable rates per beam. The single per-beam rates from all modes over all chunks are then ranked by their quality, and corresponding scores are assigned. This also yields a ranking of the single chunks of that user. The scores are used by the BS in step 2 to assign the beams in a chunk individually to the users and to make a final decision on the transmission mode per chunk, which is taken under the premise of guaranteeing a high throughput for each user. Note that the chunk-wise selection of the transmission mode allows to serve users in different modes simultaneously, enabling an improved link adaptation and thus a higher user throughput.

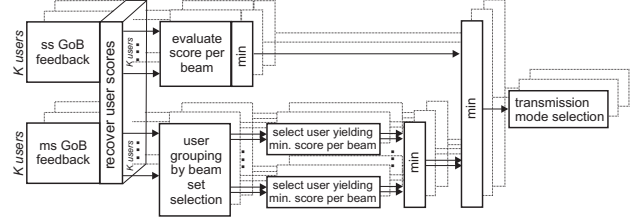
##### A. Channel evaluation at UT, step 1

Based on the actual channel  $\mathbf{H}$ , a UT determines for each transmission mode the beams it can achieve the highest data rate with. Evaluation is carried out for each chunk separately, which is represented by the different layers in Fig. 1(a).

In ss GoB mode (upper branch of Fig. 1(a)), a single beam is assumed to be powered with full transmit power  $P_s$ . At the receiver MRC is used, where the equalization vector for beam



(a) Step 1: Channel evaluation and determination of the scores at the user terminal.



(b) Step 2: Resource scheduling with transmission mode selection at the base station.

Fig. 1. Structure of the 2-step procedure constituting the adaptive transmission concept.

$\mathbf{b}_i$  is defined as  $\mathbf{w}_i = \mathbf{H}\mathbf{b}_i$ . Based on MRC, the post-detection SINR for each beam  $\mathbf{b}_i$  can be determined according to

$$\text{SINR}_{ss}(i) = \frac{P_s}{N_0} \|\mathbf{H}\mathbf{b}_i\|^2, \quad i \in \{1, \dots, B\} \quad (4)$$

where  $\|\mathbf{x}\|^2 = \mathbf{x}^H \mathbf{x}$  is the square of the Euclidean norm of vector  $\mathbf{x}$ . The equation yields the SINR for a single subcarrier signal. The SINR for the entire chunk can be obtained by determining an effective SINR from the per-subcarrier SINRs of that chunk, which can be based on the methods introduced in [32]. However, if the channel conditions do not vary considerably over the frequency width of a chunk, it may be sufficient to determine the SINR for the center subcarrier within the chunk only. Once the SINR values for all  $B$  beams are obtained, the achievable rates  $r_{ss}(i)$  per beam  $\mathbf{b}_i$  can be determined by using a suitable mapping function  $\mathcal{M}(\cdot)$ :

$$r_{ss}(i) = \mathcal{M}(\text{SINR}_{ss}(i))$$

The beam favoured for ss mode is the one achieving maximum rate, i.e.  $R_{ss} = \max_i r_{ss}(i)$ .

In ms GoB mode (lower branch of Fig. 1(a)),  $Q \leq N_t$  unitary beams are served in parallel with equal transmit power  $P_s/Q$  per beam. As  $r$  unitary beam sets of dimension  $N_t$  are provided, there exist a total of  $M = r \binom{N_t}{Q}$  sets of unitary beams of dimension  $Q$ . The following evaluation is performed for each set of beams  $m \in \{1, \dots, M\}$  separately: To recover the data stream transmitted on the  $i$ -th beam of set  $m$ , denoted as  $\mathbf{b}_{mi}, i \in \{1, \dots, Q\}$ , we use the MMSE equalizer. The corresponding MMSE equalization vector is defined as

$$\mathbf{w}_{mi} = \mathbf{Z}^{-1} \mathbf{H}\mathbf{b}_{mi}, \quad \mathbf{Z} = \frac{QN_0}{P_s} \cdot \mathbf{I} + \sum_{k=1}^Q \mathbf{H}\mathbf{b}_{mk} \mathbf{b}_{mk}^H \mathbf{H}^H \quad (5)$$

The post-detection SINR for each beam  $\mathbf{b}_{mi}$  is then given by

$$\text{SINR}_{ms}(m, i) = \frac{\|\mathbf{w}_{mi}^H \mathbf{H}\mathbf{b}_{mi}\|^2}{\mathbf{w}_{mi}^H \mathbf{Z} \mathbf{w}_{mi} - \|\mathbf{w}_{mi}^H \mathbf{H}\mathbf{b}_{mi}\|^2} \quad (6)$$

From the SINR values obtained for the chunks via from (6), one can determine the achievable rate per beam by applying the mapping function again:  $r_{ms}(m, i) = \mathcal{M}(\text{SINR}_{ms}(m, i))$ . The beam set  $\tilde{m}$  preferred for ms transmission is the one that comprises the beam achieving highest overall rate.<sup>2</sup>

$$\tilde{m} = \arg \max_m \left( \max_i r_{ms}(m, i) \right)$$

Once the beam set is selected, the  $Q$  per-beam rates belonging to that set are stored, yielding  $R_{ms,i} = r_{ms}(\tilde{m}, i)$ , while all others may be discarded.<sup>3</sup>

For  $N_t > 2$ , the number of active streams  $Q$  supported in ms mode may be in the range  $\{2, \dots, N_t\}$ . In this case, the evaluation branch for ms mode in Fig. 1(a) may be processed for each single value of  $Q$ , yielding a set of rates  $R_{ms,i}$  for each of the single ms mode options.

### B. Determination of the scores

In the following, the score concept is introduced. Therefore, let  $u$  be the chunk index and  $U$  be the total number of available chunks. The scores are used to rank the per-beam rates  $R_{ss}(u)$  and  $R_{ms,i}(u)$  of each user over all chunks  $U$  according to their quality. We use a single score set  $\mathcal{S}$  for the ranking of the user rates from all transmission modes to enable an implicit selection of the transmission mode within the score-based scheduling process following in step 2. However, this approach requires a direct comparison of the single per-beam rates from the different spatial modes, whereby it must be taken into account that each mode supports a different number of simultaneously active beams. A practical solution to enable the desired comparison with simple means is the introduction of a weighting factor  $w$ , which is used to weight the rates of the ms mode to account for its spatial multiplexing capability. For proper choice of  $w$ , we will take into account some basic considerations:

As we aim for a high user throughput, spatial mode selection should follow the rationale to favour ss mode whenever the user rate can be expected to be larger than the rate expected in ms mode. Consider that if a user decides globally for ms mode with  $Q$  active beams, the available spatial streams compared to ss mode are increased by factor  $Q$ . As a general result from that, we can assume that the user will be assigned also  $Q$  times the streams he would get if he globally selected ss mode. Hence, we can conclude that decision in favour of ss mode should be taken if the rates for the different modes in a single chunk  $u$  fulfil

$$R_{ss}(u) > Q \cdot \max_i R_{ms,i}(u), \quad (7)$$

suggesting  $w = Q$  as a suitable choice for the weighting factor of per-beam rates in ms mode.

We now return to the generation of the user scores: The per-beam rates from ss mode  $R_{ss}(u)$  as well as the weighted rates from ms mode  $R_{ms,i}(u)$  from all chunks  $U$  are aggregated

into one vector, which is sorted by magnitude in descending order.<sup>4</sup> The index within the sorted vector represents the score  $\varsigma$  of each beam.

Optionally, the user may use the scores to make a preselection of his best chunks as well as of his preferred mode per chunk  $u$ , which could be a suitable measure to further reduce the amount of required feedback.<sup>5</sup> For all selected chunks, the user finally feeds back the achievable rates for the beams supported by the desired transmission modes as well as the corresponding beam indices.

### C. Resource scheduling at BS, step 2

The second step of the process comprises the resource scheduling with implicit transmission mode selection and is carried out at the BS, which collects the feedback information from the  $K$  users, see Fig. 1(b). As a first step, it recovers the scores  $\varsigma^k$  for user  $k$  from the provided rates over all  $U$  chunks. Hereafter, the scores are partitioned according to the transmission mode they refer to, yielding  $\varsigma_{ss}^k(u)$  and  $\varsigma_{ms}^k(u, i)$ , respectively. Resource allocation with transmission mode selection is then carried out for all chunks successively. Hereby, an individual set of users is selected per chunk, and the number of beams assigned to each user is continuously tracked.

For each chunk  $u$ , the user selection process is carried out for each transmission mode separately: For ss mode (upper branch of Fig. 1(b)), the favoured user is the one providing the minimum score for that mode.<sup>6</sup>

$$k(u) = \arg \min_{k \in \{1, \dots, K\}} \varsigma_{ss}^k(u) \quad \forall u \in \{1, \dots, U\}$$

The lower branch of Fig. 1(b) illustrates the user selection for ms mode. Here, users that chose the same beam set  $m$  are possible candidates for MU-MIMO access and are thus put into one group, forming the user set  $\mathcal{K}$ . In each group  $m \in \{1, \dots, M\}$ , each of the  $Q$  available beams is assigned to the user providing the minimum score for that beam.

$$k(u, m, i) = \arg \min_{k \in \mathcal{K}} \varsigma_{ms}^k(u, i), \quad \forall i \in \{1, \dots, Q\}$$

Obviously, this user selection implicitly includes the SU-MIMO access mode, as all spatial streams will be assigned to the same user if he provides the minimum scores for all  $Q$  available beams. After user selection has been carried out for all groups, we pick the group  $\tilde{m}$  containing the user with minimum score

$$\tilde{m}(u) = \arg \min_m \left( \min_i \varsigma_{ms}^{k(u, m, i)}(u, i) \right)$$

Finally, we compare the scores of the users selected for the different transmission modes in chunk  $u$  and select the transmission mode embracing the user with minimum score. In essence, selection of the transmission mode and beam set for each chunk is thus dictated by the user providing minimum

<sup>2</sup>This criterion for beam set selection for MU-MIMO mode has also been suggested in [33].

<sup>3</sup>It is clear that this approach is suboptimum, however, it keeps the required feedback demand low and thus forms a suitable solution for practical application.

<sup>4</sup>In the case of multiple resources yielding identical rates, these are ordered in a random fashion.

<sup>5</sup>This kind of user-driven chunk selection has also been suggested in [34].

<sup>6</sup>In the case of multiple users providing identical scores, the stream is given to the one been assigned the least beams so far.

overall score. The decision on the mode and the user allocation per chunk is then signalled forward to the UTs, who configure their receivers accordingly.

Remark: Although the concept of the scheduling process has been developed for an isolated cell scenario, where only intra-cell interference from simultaneously active beams is taken into account in (4) and (6), the scheme can readily be employed in a multi-cell environment. The only modification required is to substitute the noise figure  $N_0\mathbf{I}$  in the MMSE equalizer by an interference plus noise correlation matrix that takes into account the inter-cell interference. Evaluation of the scheduling concept in a multi-cell environment has been carried out in [35]–[37].

#### D. Feedback

For the evaluations carried out in section V, we assume the UTs to feed back the information on the achievable rates per evaluated transmission mode for all available chunks (full feedback). For practical applications, however, the concept offers a high potential for further feedback reduction, as the score-based ranking allows each user to preselect his best chunks as well as his preferred mode to be served with. A score-based preselection of the transmission mode may result in a severe performance loss, though, which is illustrated by the following two examples: If there are not sufficient users in the system (sparse network), it may occur that no MU-MIMO partner can be found for a user who provided the best score for the ms mode. In this case, this user could be served either via multiple beams in SU-MIMO mode or, alternatively, in diversity mode. To achieve highest possible throughput for that user, the final mode selection should be based on his achievable rate – which requires also the availability of the ss rate at the BS. On the other hand, if a user prefers ss mode, but a ms user is selected, the former could still be assigned resources if he provided appropriate rates for ms mode.

As the scheduling process aims at assigning each user his best resources only, it is certainly not economical to let the users report on all available resources, but on their best chunks only. These can easily be selected after the scores have been determined by the UTs, which is related to the suggestion in [34] denoted as *Top-M feedback*. Furthermore, a similar selection can also be done for the rates referring to the transmission modes in a chunk, including the options for different  $Q$ . For example, for  $N_t = N_r = 2$ , a practical solution could be to let the users report two rates: The best ms rate enabling MU-MIMO access, and additionally the ss rate (enabling diversity mode) or the next ms rate (enabling dual-stream SMUX for that user). The adequate choice of the second rate to report could be based on the rate achievable with diversity mode and dual-stream SMUX mode, respectively.

Moreover, the frequency-selective feedback information for the utilized transmission band will be highly correlated, so that proper compression techniques can conveniently be applied, yielding a further reduction of the required feedback per user. Elaborating on the adequate amount of total required feedback is an interesting field for further studies, but lies out of the scope of this paper.

#### E. Fairness steering

As for every user the same score set is used to rank the available resources, the score-based scheduler will on average distribute the available resources evenly over all users (see also [25]). As a result, each user will be able to realize about the same fraction of his total achievable rate. However, one can think of applications where it might be desirable to distribute the resources with a different scheduling target. Forcing our proposed scheduler to assign the resources according to such an altered target can easily be achieved by applying a user-specific weighting of the scores determined at the BS (step 2 of the process). As this weighting will have a direct impact on the degree of fairness the scheduler is able to establish, we denote this process as *fairness steering*.

As an example, we will present an algorithm targeting at equal rates for all users in the time-slot. Assume that the achievable rates of all users are known. To go for equal user rates, the resources can be redistributed over all users by following the 'Robin-Hood principle', where resources are taken from high rate users and given to low rate users in order to improve their realizable rate. Specifically, this can be achieved by weighting the scores of the users by a factor proportional to the achievable rate of each user. This results in the fact that high rate users are considered less frequently in the resource scheduling process in favour of the low rate users, so that the latter will be assigned more resources than the high rate users. To develop a suitable algorithm for the fairness steering, we will firstly derive some basic relations based on the average behaviour of the score-based scheduler. These are then used as guideline for the development of an algorithm tailored to approach the equal rate target in an iterative fashion.

Let  $R_k$  be the achievable rate of user  $k \in \{1, \dots, K\}$ . If no weighting of the scores is applied, the scheduler assigns on average all users the same amount of resources  $u$ , so that the average rate per resource for user  $k$  can be given as  $R_k/u$ . Now let  $w_k$  and  $w_l$  be the weighting factor of  $k$ -th and  $l$ -th user, respectively, and  $x = w_k/w_l$  be their ratio. Then we can conclude that user  $l$  will be considered within the resource scheduling process by factor  $x$  more frequently than user  $k$ . As a result, the average amount of resources  $u_l$  assigned to user  $l$  will be by factor  $x$  larger than the amount of resources  $u_k$  assigned to user  $k$ . It follows

$$\frac{u_l}{u_k} = x = \frac{w_k}{w_l} \quad (8)$$

The scheduling based on the weighted scores yields as average result for the new achievable rate  $\bar{R}_k$  for user  $k$  approximately

$$\bar{R}_k \approx u_k \cdot R_k/u \quad (9)$$

The target rate  $\bar{R}_k$  should be identical for all users, so we have  $\bar{R}_k = \bar{R}_l \quad \forall k \neq l$ . With equations (8) and (9), we obtain

$$w_k = \frac{R_k}{R_l} \cdot w_l \quad (10)$$

If we define user  $l$  as the reference user, whose weighting factor we set to unity,  $w_l = 1$ , the weighting factors for all users  $k$  can be obtained simply from the ratio of their achievable rate  $R_k$  and the achievable rate of the reference user,  $R_l$ .

Evidently, the obtained result cannot be applied directly to our scheduler, as some relations have been simplified and hold for the average behaviour of the scheduler only. However, we can use it as a guideline for the construction of an iterative weighting and scheduling process, which enables us to approach the equal rate target in a stepwise manner. In each iteration, the closeness to the equal rate target is checked by determining the deviation of each user rate from the mean rate over all users. The stopping criterion is fulfilled if this deviation cannot be further decreased. The iterative algorithm is described as follows:

- 1) run scheduling process with unweighted scores, yielding an achievable rate  $R_k$  for each user
- 2) determine mean rate  $\bar{R}$  over all users,  $\bar{R} = \frac{1}{K} \sum_{k=1}^K R_k$
- 3) calculate sum of deviations between user rates and  $\bar{R}$ , i.e.  $D = \sum_{k=1}^K |R_k - \bar{R}|$
- 4) determine  $R_{min} = \min_k R_k$  as the reference rate
- 5) normalize all rates to the reference rate, i.e.  $\bar{R}_k = R_k / R_{min}$
- 6) determine weighting factors  $w_k = q \cdot \bar{R}_k + (1 - q)$ , where  $q \in [0, 1]$  is a tuning parameter used to control the strength of the weighting
- 7) multiply each user's score vector with corresponding  $w_k$
- 8) rerun scheduling process
- 9) determine  $D$  for new obtained rates  $R_k$
- 10) if  $D$  is smaller than previous value, goto step 4; else use resource assignment from previous iteration and end process

The tuning parameter  $q$  used for the weighting factors is intended to enable smooth transitions of the scheduling results from successive iterations, as this better accommodates the heuristic nature of the scheduling process. Certainly, the parameter  $q$  will also influence the convergence speed of the algorithm, as the stopping criterion will be achieved faster for large  $q$ . However, it can be expected that a smaller  $q$  in return will achieve a better final result in terms of a smaller deviation  $D$ , which will be confirmed by the simulation results presented later in section V. Eventually, there will be a tradeoff between convergence speed of the algorithm and quality of the result.

## V. INVESTIGATIONS ON LINK-LEVEL

The properties of the scheduler with transmission mode switching are investigated in a single-cell link-level simulation environment with multiple users. These kind of link-level investigations enable an isolated examination of the system behaviour depending on fixed SNR conditions, which are assumed to be common for all involved users. Thus, we are able to gain insights into the basic relationships that influence the performance of multi-user MIMO communication systems. As our main interest here is to illustrate the fundamental behaviour of the adaptive system concept, we focus on the simplest MIMO configuration with  $N_t = N_r = 2$  antennas. For performance results for MIMO configurations of higher order in realistic environments, the interested reader is referred to [35]–[37].

### A. Link-level assumptions

We assume  $K$  UTs being equipped with  $N_r = 2$  antennas each and a BS with  $N_t = 2$  antennas, which provides the two unitary DFT beam sets  $\mathbf{C}_1$  and  $\mathbf{C}_2$  introduced in (3). Possible multi-stream modes are thus 2-user MU-MIMO or dual-stream SU-MIMO. We use the channel model provided by the European WINNER project (WIM) in its configuration for a wide area urban macro scenario. This model assumes a uniform linear array of co-polarized antennas; antenna spacing is set to  $4\lambda$  at the BS and to  $0.5\lambda$  at the UTs, yielding a low degree of correlation between the paths emanating from the antennas. The mean channel power is normalized to unity; channels for different users are modeled independently. The mean reception SNR thus is  $P_s/N_0$  for any user. An OFDM system with 1024 subcarriers spanning a bandwidth of 40 MHz is assumed, accommodating 128 chunks of 8 subcarriers width. To obtain the per-stream SINR  $\gamma$  for the different transmission modes per chunk, (4) and (6) are calculated for the subcarrier in the center of the chunk<sup>7</sup> based on ideal knowledge of the channel  $\mathbf{H}$ .

For a given SINR value  $\gamma$ , the corresponding achievable rate can be determined via the Shannon information rate  $\log_2(1 + \gamma)$ . To obtain rates that are closer to those achievable in practice, we instead use a quantized rate mapping function, which was introduced as a component of the WINNER link to system interface, presented in [38]. This rate mapping function is based on a puncturable low density parity check (LDPC) code with constant block length of 1152 bits and supports the fixed symmetric modulation formats up to 64QAM. The discrete steps of the mapping function are derived from the SINR values required to meet a block error rate performance of  $10^{-2}$  in an equivalent additive white Gaussian noise (AWGN) channel. As both the block length of the code as well as the modulation formats are limited, the rates supported by the mapping function are confined to a minimum rate of 0.5 and a maximum rate of 5.538 bit/s/Hz. This former value corresponds to BPSK modulation with code rate 0.5, while the latter is achieved with 64QAM modulation with code rate 24/26. All simulation results are obtained from a total of 10,000 independent channel realizations.

### B. Performance of spatial mode switching

First we examine the system performance of the adaptive system when only the beam set  $\mathbf{C}_1$  is available. We focus on the low SNR regime,  $P_s/N_0 = 0$  dB, which is relevant for cell-edge users, where we expect the benefits from switching to single-stream mode to become prominent. First results are based on Shannon information rates. Fig. 2 presents cumulative distribution functions (CDFs) of the achievable user throughput divided by the signal bandwidth (left) and the spectral efficiency in the cell (right) for  $P_s/N_0 = 0$  dB for  $K = 10$  users. Focussing on the user throughput (left), we first point out that all users achieve non-zero rates, so indeed all users in the system are conveniently scheduled. We compare the adaptive system described above to a system supporting

<sup>7</sup>Note that with the chosen set of system parameters, the channel variations over the frequency width of a chunk can be considered negligible.

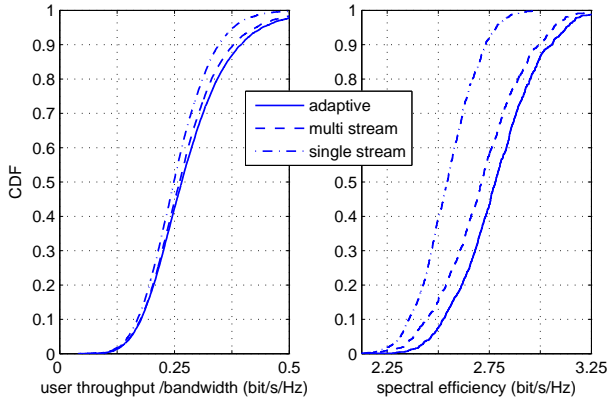


Fig. 2. CDFs of the achievable user throughput (left) and spectral efficiency in the cell (right) based on Shannon information rates. 1 beam set,  $K=10$  users,  $\text{SNR} = 0$  dB.

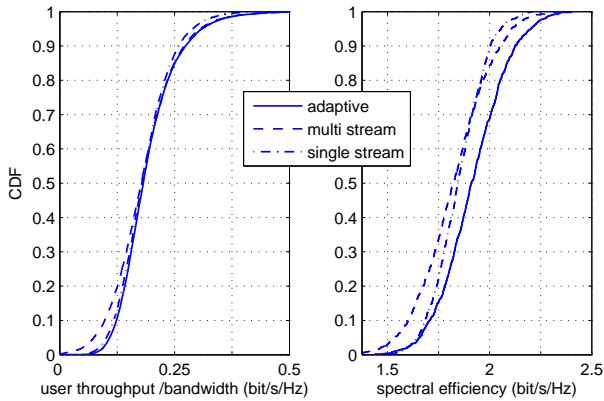


Fig. 3. System performance based on quantized rate mapping function. System setting as above.

either single-stream or multi-stream mode exclusively. For the user throughput (left), we observe that the performance of the adaptive system benefits slightly from switching in the region where the CDF is above 0.5. Further, for the CDF region below 0.2, the single-stream curve is nearly identical to the multi-stream curve, and hence no gains from switching can be realized here. This observation can be explained as follows: Recall that a beam in single-stream mode is served with double the power used for a multi-stream beam. In the low SNR regime, where the noise dominates the interference from simultaneously active beams, we thus can expect that the SINR of the selected beam for single-stream mode is about twice as large as the SINR  $\gamma$  for the corresponding beam in multi-stream mode. Moreover, in multi-stream mode, the amount of beams assigned to each user is twice as large as in single-stream mode. As

$$\log_2(1 + 2\gamma) \approx 2 \log_2(1 + \gamma) \quad \text{if } \gamma \ll 1$$

holds, the rates achievable with the two different modes are nearly identical. Considering the spectral efficiency within the cell (right subfigure in Fig. 2), we observe that the adaptive system benefits significantly from the mode switching over the entire CDF region.

Fig. 3 depicts performance curves for the same setting,

but this time the quantized rate mapping function is used. For the user throughput (left), we observe here that the CDF of the adaptive system represents a hull curve of the two single-mode schemes. As the minimum supported rate to be assigned is bound here to 0.5 bit/s/Hz, the adaptive system now significantly gains from switching to single-stream mode if the SINR conditions are low (left region of the CDF curve). The CDF of the adaptive system is quite close to the one supporting single-stream mode only, suggesting that this mode is predominantly chosen at low SNR. Considering the spectral efficiency in the cell (right subfigure in Fig. 3), we observe that only the left tail of the adaptive system's CDF approaches the curve of the pure single-stream mode. In the remaining region, substantial gains from mode switching become visible.

In the next step, we will examine the system behaviour for varying SNR. Therefore, we focus on the median spectral efficiency in the cell based on quantized rate mapping, i.e. the value determined from the CDF for a probability of 0.5. Furthermore, we draw our attention to the probabilities of mode selection, which reveal the dominantly chosen mode depending on the SNR conditions. Fig. 4 depicts the median spectral efficiency in the cell versus the SNR for different configurations of the adaptive mode switching system. The corresponding probabilities of mode selection are found in Fig. 5. The different configurations are as follows:

- 1) *MU-MIMO, 2 beam sets*: Similar as *MU-MIMO, chunk-adaptive* (see next entry), but here the two beam sets  $C_1$  and  $C_2$  from (3) are available.
- 2) *MU-MIMO, chunk-adaptive*: adaptive system as described in section IV with chunk-wise selection of the spatial mode, i.e. a user may be served in different modes simultaneously.
- 3) *MU-MIMO, fixed scheme*: For each user, a fixed mode and, in case of ss mode, a fixed beam is selected. Therefore, each user sums up the rates of his best beams in ms mode and the rates of all beams in ss mode over all chunks of the frequency band. By considering the weighting factor  $w$  for ms mode, decision is taken in favour of the mode (and beam) achieving highest sum rate.
- 4) *SU-MIMO, chunk-adaptive*: *MU-MIMO* option is switched off, i.e. ms mode reduces to *SMUX* to a single user. Now only one user is served per chunk either in ss or *SMUX* mode.
- 5) *SU-MIMO, fixed scheme*: Fixed mode and, in case of ss mode, fixed beam per user. Selection is based on the mode (and the beam) achieving highest sum rate over the entire frequency band.

Reading this list from bottom to top, the different configurations can be understood to successively add additional degrees of freedom in the spatial domain to the user selection and resource allocation process. Fig. 4 and 5 clearly show that these additional degrees of freedom do not only increase the achievable system performance successively, but also promote the selection of the ms mode. The crossing point of the probability curves for ss and ms mode in Fig. 5 highlights the point in the SNR region where the ms mode becomes

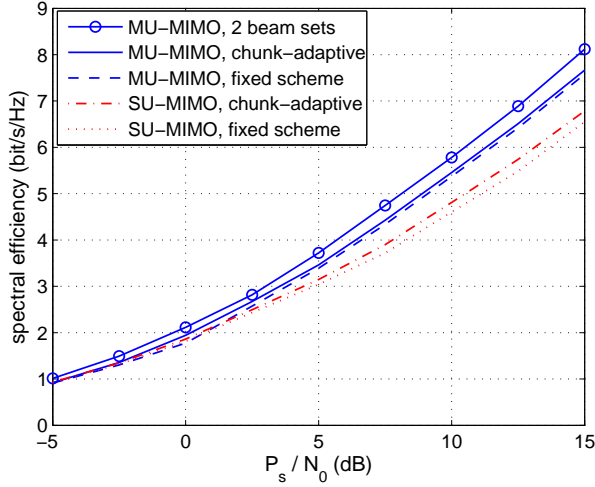


Fig. 4. Median spectral efficiency in the cell for different system configurations. Quantized rates,  $K=10$  users, beam set  $C_1$ .

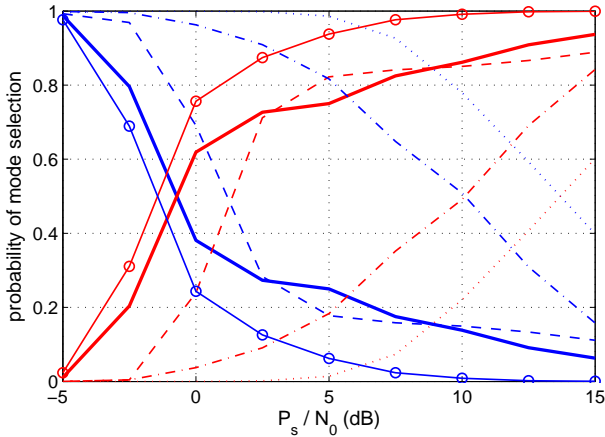


Fig. 5. Probability of mode selection vs. SNR. Blue: ss mode, red: ms mode. Legend and system setup as in Fig. 4

the dominantly selected one. Most interestingly, Fig. 5 reveals that by activating MU-MIMO (configuration 3), this crossing point is shifted by about 8 dB compared to configuration 4 to an SNR of about 1.5 dB. This result underlines that the MU-MIMO mode is the key to support the transmission of multiple data streams in the spatial domain already at low SNR. Together with the throughput gains shown in Fig. 4, the high potential of MU-MIMO to increase the overall system throughput is substantiated. Configuration 2 shifts the crossing point further to the left to an SNR below 0 dB. The throughput, however, is thereby increased only slightly. A significant additional gain in throughput can be achieved if an additional beam set is provided by the BS (configuration 1), which amounts to about 5% compared to configuration 2. The crossing point in Fig. 5 is also shifted further down to about -1.5 dB.

Close inspection of Fig. 5 further reveals that with increasing degrees of freedom, the probability curves exhibit a steeper slope and approach the upper and lower boundary areas more rapidly with increasing SNR. In particular, for the

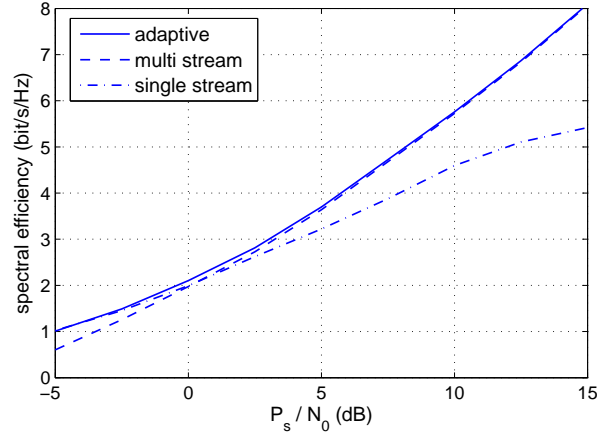


Fig. 6. Median spectral efficiency in the cell achievable with 2 beam sets and comparison to single mode systems. Quantized rates,  $K=10$  users.

leftmost ss curve (MU-MIMO, 2 beam sets), we observe a rapid decline that falls below a probability of 0.1 above 5 dB SNR, suggesting that the adaptive system tends to behave similar to a system that uses a fixed mode in the low and high SNR regime, respectively, with a switching point set at a fixed SNR level. This conjecture is confirmed by Fig. 6, where we compare the median cell throughput for configuration 1 (adaptive MU-MIMO, 2 beam sets) with a similar system supporting either ss or ms mode exclusively, as done at the beginning of this section: The figure reveals that the adaptive system represents a hull curve of the performance of the two single mode systems, which turns out to be very tight.

We conclude this subsection with the important observation that proper application of the MU-MIMO mode enables to conveniently serve even users in the ms mode who experience relatively poor SNR conditions<sup>8</sup>. Thus, the MU-MIMO mode establishes a win/win situation for low- and high-rate users competing for a frequency or time resource, as low-rate user can now be served without blocking this resource for any high-rate user, who can support a rate on any of the available beams.

### C. Steering the fairness to equal rates

In this subsection, we examine the behaviour of the scheduler if we apply the fairness steering option to achieve an equal rate scheduling target as described in section IV-E. We use different tuning factors  $q$  and compare the achievable distribution of the user rates as well as the required number of iterations. Investigations have been carried out for a mean SNR of 10 dB, which is equal for all  $K = 10$  users in the time-slot. The CDFs representing the rates of the successfully scheduled users are depicted in Fig. 7. We observe that for any choice of the tuning factor  $q$ , the fairness steering process seems to operate conveniently, as the CDF significantly gains in steepness and thus approaches the equal rate target, which would be represented by a vertical line. It is interesting to note that the median user throughput does not change due to the

<sup>8</sup>in a cellular system, these are the users at cell-edge

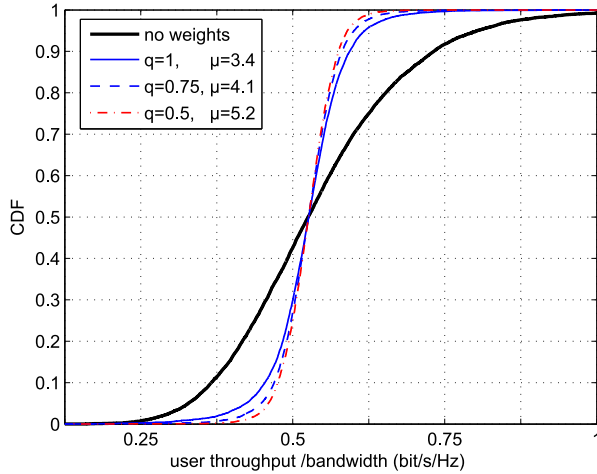


Fig. 7. Fairness steering towards an equal rate target based on iterative algorithm with tuning factor  $q$ .  $\mu$  - mean required number of iterations.  $K = 10$  users and SNR = 10 dB.

applied weighting, highlighting the convenient practicability of the proposed score weighting.

Taking a closer look at the curves for different  $q$  reveals that choosing a smaller value enables a better match of the equal rate target, as a steeper CDF of the user rates can be achieved. The price we pay for this is an increased number of iterations, whose expected value  $\mu$  grows from 3.4 for  $q = 1$  continuously to 5.2 for  $q = 0.5$ . For the application considered here, the choice of  $q = 0.75$  would probably be well suited, as it achieves a CDF which is quite close to the desired target while requiring a moderate number of iterations. The price we have to pay for the improved fairness is a loss of the achievable spectral efficiency in the cell. For the case considered here where all users have the same mean SNR, the loss is negligibly small, as it drops by 3% only. However, it should be noted that this loss will be substantially larger in real-world scenarios, where the mean SNRs of different users are likely to be significantly different.

#### D. Capacity scaling of the adaptive system

Finally we examine the downlink capacity achievable with the adaptive system and compare it to the upper bound, which is the capacity of the  $2 \times 2$  BC when full CSI is available at the receivers as well as at the transmitter. As mentioned in the introduction, the capacity of the BC was shown to be achievable with the dirty paper coding (DPC) technique. In [39] an algorithm was presented to compute it in an iterative manner for any given set of flat-fading user channels. While maintaining the equal power distribution over all chunks, we use this algorithm to compute the optimal user allocation and the corresponding precoding matrices per chunk to obtain the upper bound for the (flat-fading) capacity of the BC, which is depicted in Fig. 8 versus the SNR for  $K = 10$  users. The achievable downlink capacity of our adaptive system is obtained by applying Shannon's information rates and carrying out maximum throughput scheduling (MT) based on the reported rates at the BS, which selects for each chunk

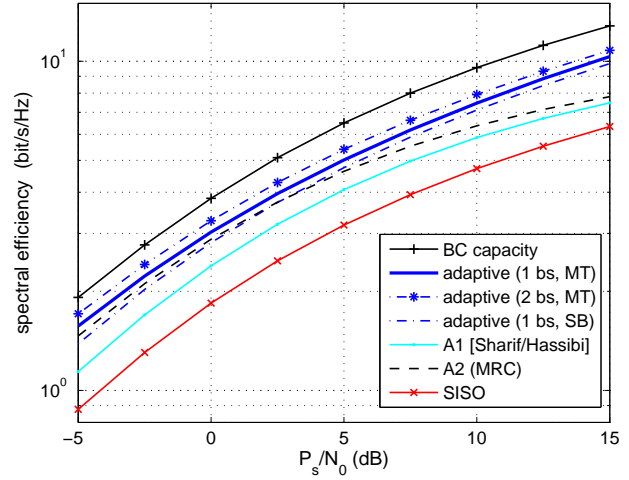


Fig. 8. Comparison of the downlink capacity for various systems vs. SNR.  $K = 10$  users, Shannon information rates. bs - beam sets; scheduling: MT - max. throughput, SB - score-based.

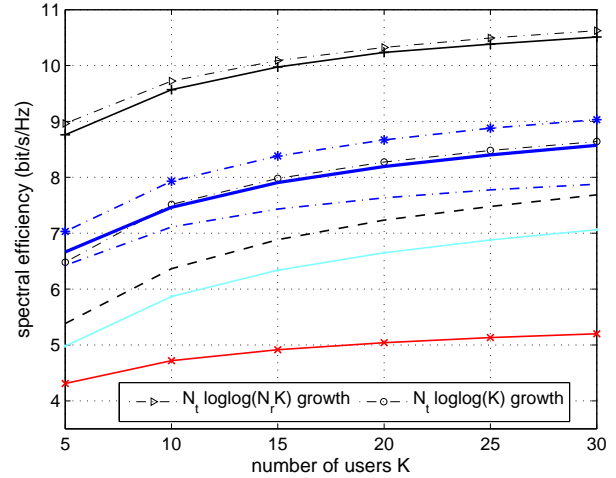


Fig. 9. Scaling of the downlink capacity with number of users at SNR = 10 dB. Same legend as in Fig. 8

the user (single-stream) or user constellation (multi-stream) that achieves the highest throughput. In Fig. 8 we observe that for an SNR above 0 dB, the capacity of our adaptive system utilizing partial CSI achieves a constant fraction of the capacity of the BC, which amounts to about 80% if one beam set is available. Utilization of two beam sets provides an extra gain in capacity of about 5%. Additionally, we included the capacity of the single-input single-output (SISO) channel achieved in an equivalent scenario. While we observe here that the capacity of the BC scales with factor 2 (corresponding to  $\min(N_r, N_t)$ ) compared to the capacity of the SISO channel in the high SNR range, the capacity of our adaptive system (with one beam set) achieves a factor of 1.6. For comparison, we also added the spectral efficiency achievable with the fair score-based scheduling (SB) technique. It can be seen that the price we have to pay to obtain user fairness within one time-slot is only marginal, as the loss in spectral efficiency is only

about 5%.<sup>9</sup>

To see potential gains in the downlink capacity compared to other well-known limited feedback schemes, we further compare the adaptive system to two multi-stream approaches with  $Q = 2$  simultaneously active beams, which we denote as A1 and A2. For both approaches, we assume beam set  $\mathbf{C}_1$  to be available only. A1 is the approach presented in [20] and is sketched as follows: Each receive antenna at the UT is treated as an independent receiver. Hence, the per-antenna reception SINR is calculated for each beam, assuming that the other beam interferes. For each antenna, the UT feeds back the best beam together with the corresponding SINR, and the BS assigns each beam to the user having provided the highest SINR value for it.

In A2, we consider a multi-stream system where the UTs simply carry out MRC for each received beam, i.e. nothing is done to actively combat the interference. The corresponding post-detection SINR can be determined by equation (6), with the MRC equalization vector  $\mathbf{w}_i = \mathbf{H}\mathbf{b}_i$ . Feedback and beam assignment is then carried out as in A1.

Note that for the  $2 \times 2$  MIMO setup considered here, the feedback required for the adaptive system is by 50% larger than that for the two systems A1 and A2, as additionally to the two per-stream rates in ms mode the rate for the ss mode has to be reported. Within these investigations, however, the amount of feedback has not been taken into account, as the main focus here lies in the achievable downlink capacity relative the capacity of the BC. A similar framework for this kind of performance evaluation has also been used in [40].

In Fig. 8, we observe that A1 achieves a significantly lower performance than the mode switching system – a result similar to the one found in [23] for single-antenna UTs. The relative difference in throughput increases with increasing SNR, as A1 suffers from the interference between active beams, which is actively suppressed in our adaptive system by the MMSE equalizer. The throughput of A2 is quite close to the performance of the adaptive system in the low SNR region. Obviously, this is due to the fact that in the low SNR region the noise dominates the interference from the other spatial stream active in ms mode. In this case, the MMSE solution approaches the MRC solution, yielding similar post-detection SINRs for both equalizers, which finally translates into a similar throughput performance. For increasing SNR, however, the performance of A2 degrades significantly and approaches the throughput of A1. As in the former case, A2 suffers here from the interference between the simultaneously active beams.

In Fig. 9 we examine the downlink capacity for a constant high SNR = 10 dB for a variable number of users. First we compare the capacity of the adaptive system (MT) with the achievable spectral efficiency of the fair scheduling approach (SB). Although the loss in throughput to provide the desired fairness is again not exceptionally high, we observe that the gap between MT and SB scheduling increases with increasing number of users. This is not very surprising, as with a growing

number of users, the probability of a user experiencing poor channel conditions increases, who imperatively has to be served by the system based on fairness. The support of these users by the fair scheduler thus costs a growing proportion of the maximum sum capacity.

Next we focus on the scaling of the downlink capacity versus the number of users and compare it to the other reference systems. In [20], it has been shown that the BC capacity for independent and identically distributed (i.i.d.) Rayleigh fading channels scales for a large number of users with  $N_t \log \log(N_r K)$ . This growth has also been plotted in Fig. 9, and we observe that the capacity of the BC for  $K \geq 10$  users and the correlated channels considered is scaling similarly. For our adaptive system, it has been shown in the previous investigations that at SNR = 10 dB, the ms mode dominates the selection (compare with Fig. 5). In this case, the use of the MMSE equalizer reduces the spatial degrees of freedom at the receiver from  $N_r = 2$  to 1 [41], resulting in a capacity scaling equivalent to  $N_t \log \log(K)$  [20] for the setup given here. The plot of this growth in the figure shows a convenient agreement with the scaling of the corresponding efficiency curve of the adaptive system for  $K \geq 10$  users.

We also plot the capacity of the SISO system, which grows roughly with a slope of  $\log \log(K)$  [20] and thus less steep than the adaptive system. We observe that the capacity scaling factor relative to the SISO system of 1.6 is achieved by the adaptive system for  $K = 10$  users, remaining about constant for further increasing  $K$ . Comparing finally the adaptive system to the reference systems A1 and A2, we observe that the adaptive system achieves a significantly higher downlink capacity, highlighting the gains that can be achieved by using the additional receive antennas to actively suppress the interference from other active beams. Note that A1 exhibits the same scaling over the number of users as the BC capacity, as has been shown in [20]. In the depicted range of users, the MRC-based system A2 exhibits the steepest slope over all curves, suggesting that for a large number of users its downlink capacity approaches the one of the adaptive system. This is reasonable, as we can expect that for a large number of users, we can find a user in the active set whose MRC equalization vector is close to the corresponding MMSE equalization vector, which is the receiver architecture used in the adaptive system.

## VI. CONCLUSION

We have presented a concept for the downlink of a multi-user MIMO-OFDM system combining spatial transmission mode switching with a fair scheduling approach. Based on unitary fixed beamforming, the users report information on the rates they can achieve for distinct beams in the supported transmission modes via a low-rate feedback channel. According to the feedback given, the BS then selects the transmission mode per frequency resource and assigns each user the beams yielding highest rates. As an example, performance evaluation has been carried out for a  $2 \times 2$  MIMO configuration, revealing significant gains in spectral efficiency as well as in user throughput due to mode switching. It turned out that MU-MIMO transmission, where the spatial beams are assigned to

<sup>9</sup>Note that this loss may increase substantially if users with different mean SNRs are considered. This scenario has been in the focus of the investigations presented in [35]–[37].

different users, is the key to enable simultaneous transmission of multiple spatial data streams even at low SNR conditions. If multiple beam sets are provided, the gains from instantaneous mode switching vanish in most parts of the SNR region, and the system turns into a pure spatial multiplexing system in the higher SNR range. We further introduced a fairness steering that can be configured to realize a desired fairness target the scheduler is supposed to provide; its potential has been shown with the example of an equal rate target. Finally, the achievable downlink capacity of the adaptive system has been compared to the capacity of the SISO and the MIMO broadcast channel as well as of two limited-feedback reference systems supporting MU-MIMO, but no interference suppression. This comparison illustrated that the system effectively achieves a high performance, excelling the one of the two reference systems significantly. The scheduling concept introduced here can readily be employed in a multi-cell environment, where users with different mean SNR are present, as investigated in [35], [36]. Note that the elementary functionality of the proposed system concept has been tested already in real-time experiments, which have been conducted in a real-world broadband mobile communication environment; for details refer to [37], [42].

Remark: The concept proposal presented here goes beyond the current standardization of 3G Long Term Evolution (LTE) [21]. In particular, we allow each user to select the spatial mode as well as the beam to be served on individually per chunk, while in LTE each user is supposed to make a selection for the entire band (similar to the *fixed scheme* configuration in the list given in section V-B). Further, feedback in LTE is supposed to be given for sets of contiguous chunks instead of per chunk individually. Both measures significantly reduce the required feedback, but clearly come at the cost of a decreased system performance.

#### ACKNOWLEDGEMENT

The authors are grateful to Elena Costa and Wolfgang Zirwas for their continuous interest in the progress of the work and for stimulating discussions. The first author would further like to thank Howard Huang for discussions, helpful comments and additional material provided, which have proven invaluable to this work. All authors wish to thank the German Ministry for Education and Research (BMBF) and Nokia Siemens Networks for financial support in the project 3GeT. Part of this work has been performed in the framework of the IST project IST-4-027756 WINNER II, which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues.

## REFERENCES

- [1] F. Farrokhi, A. Lozano, G. Foschini, and R. Valenzuela, "Spectral efficiency of FDMA/TDMA wireless systems with transmit and receive antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 591–599, Oct. 2002.
- [2] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.
- [3] R. Heath and A. J. Paulraj, "Switching between diversity and multiplexing in MIMO systems," *IEEE Trans. Commun.*, vol. 53, no. 6, June 2005.
- [4] D. Love and R. Heath, "Multi-mode precoding using linear receivers for limited feedback MIMO systems," *Communications, 2004 IEEE International Conference on*, pp. 448–452, 2004.
- [5] S. Catreux, V. Erceg, D. Gesbert, and J. Heath, R.W., "Adaptive modulation and MIMO coding for broadband wireless data networks," *Communications Magazine, IEEE*, vol. 40, no. 6, pp. 108–115, June 2002.
- [6] S. Chung, A. Lozano, H. Huang, A. Sutivong, and J. Cioffi, "Approaching the MIMO capacity with a low-rate feedback channel in V-BLAST," *EURASIP JASP*, no. 5, pp. 762–771, 2004.
- [7] J. Lopez-Vicario and C. Anton-Haro, "Adaptive switching between spatial diversity and multiplexing: A cross-layer approach," *IST Mobile Communications Summit*, June 2005.
- [8] M. Schellmann, V. Jungnickel, A. Sezgin, and E. Costa, "Rate-maximized switching between spatial transmission modes," *40th Asilomar Conference on Signals, Systems and Computers*, pp. 1635–1639, 2006.
- [9] A. Foreza, M. McKay, A. Pandharipande, R. Heath, and I. Collings, "Adaptive MIMO transmission for exploiting the capacity of spatially correlated channels," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 619–630, Mar. 2007.
- [10] D. Gesbert, M. Kountouris, R. Heath, C.-B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *Signal Processing Magazine, IEEE*, vol. 24, no. 5, pp. 36–46, Sept. 2007.
- [11] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," *Communications, IEEE International Conference on*, pp. 331–335, 1995.
- [12] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.
- [13] P. Viswanath and D. 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.
- [14] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.
- [15] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, 1983.
- [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] H. Boche and E. Jorswieck, "On the performance optimization in multiuser MIMO systems," *European Transactions on Telecommunications*, vol. 18, no. 3, pp. 287–304, Sept. 2007.
- [18] D. Love, J. Heath, R.W., W. Santipach, and M. Honig, "What is the value of limited feedback for MIMO channels?" *Communications Magazine, IEEE*, vol. 42, no. 10, pp. 54–59, Oct. 2004.
- [19] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.
- [20] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.
- [21] 3GPP TS 36.211V8.5.0, "E-UTRA – Physical channels and modulation (Release 8)," Dec. 2008.
- [22] IST-4-027756 WINNER II – D6.13.10, "Final CG 'wide area' description for integration into overall system concept and assessment of key technologies," Nov. 2007.
- [23] J. Wagner, Y.-C. Liang, and R. Zhang, "Random beamforming with systematic beam selection," *Personal, Indoor and Mobile Radio Comm., IEEE 17th Int. Symp. on*, Sept. 2006.
- [24] R. de Francisco, D. Slock, and Y.-C. Liang, "Balance of multiuser diversity and multiplexing gain in near-orthogonal MIMO systems with limited feedback," *Wireless Communications and Networking Conference, WCNC 2007. IEEE*, pp. 1269–1274, 2007.
- [25] T. Bonald, "A score-based opportunistic scheduler for fading radio channels," *Proc. of European Wireless*, 2004.
- [26] IST-2003-507581 WINNER – D2.7, "Assessment of advanced beam forming and MIMO technologies," Feb. 2005.
- [27] D. Love, J. Heath, R.W., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.
- [28] D. Love and J. Heath, R.W., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.
- [29] —, "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1102–1110, July 2003.
- [30] A. Tulino, A. Lozano, and S. Verdu, "Impact of antenna correlation on the capacity of multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, July 2005.
- [31] F. Boccardi and H. Huang, "A near-optimum technique using linear precoding for the MIMO broadcast channel," *Acoustics, Speech and Signal Processing, ICASSP 2007. IEEE International Conference on*, vol. 3, pp. III–17–III–20, Apr. 2007.
- [32] S. Nanda and K. Rege, "Frame error rates for convolutional codes on fading channels and the concept of effective  $E_b/N_0$ ," *IEEE Trans. Veh. Technol.*, vol. 47, no. 4, pp. 1245–1250, Nov 1998.
- [33] H. Kim, J. Kim, J. Li, and M. Kountouris, "On the performance of limited feedback multiuser MIMO transmission in 3GPP HSDPA," *Vehicular Technology Conference, VTC-2005-Fall. IEEE*, vol. 1, pp. 473–476, Sept. 2005.
- [34] J. van de Beek, "Channel quality feedback schemes for 3GPP's Evolved-UTRA downlink," *Global Telecommunications Conference, GLOBE-COM '06, IEEE*, pp. 1–5, Nov. 2006.
- [35] L. Thiele, M. Schellmann, V. Jungnickel, and W. Zirwas, "Capacity scaling of multi-user MIMO with limited feedback in a multi-cell environment," *41st Asilomar Conference on Signals, Systems and Computers*, Nov. 2007.
- [36] L. Thiele, M. Schellmann, T. Wirth, and V. Jungnickel, "Interference-Aware Scheduling in the Synchronous Cellular Multi-Antenna Downlink," *IEEE 69th Vehicular Technology Conference, VTC2009-Spring*, Apr. 2009, invited.
- [37] M. Schellmann, L. Thiele, T. Wirth, T. Haustein, and V. Jungnickel, "Resource Management in MIMO-OFDM systems," in *OFDMA: Fundamentals and Applications*, T. Jiang, L. Song, and Y. Zhang, Eds. CRC Press, Taylor&Francis Group, 2009.
- [38] IST-4-027756 WINNER II – D2.2.3, "Modulation and coding schemes for the WINNER II system," Nov. 2007.
- [39] N. Jindal, W. Rhee, S. Vishwanath, S. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.
- [40] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, 2006.
- [41] J. Winters, J. Salz, and R. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 1740–1751, Apr. 1994.
- [42] T. Wirth, V. Jungnickel, A. Forck, S. Wahls, T. Haustein, et al., "Realtime multi-user multi-antenna downlink measurements," *Wireless Communications and Networking Conference, WCNC 2008. IEEE*, pp. 1328–1333, Apr. 2008.