

Spatial Transmission Mode Switching in Multiuser 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—Multiantenna radio systems allow accessing the channel in diversity or spatial multiplexing (SMUX) 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 multiuser multiple-input–multiple-output orthogonal frequency-division-multiplexing (MU-MIMO-OFDM) system aiming at high user rates with limited feedback demands. An extended score-based scheduling (SB) approach ensures fair-resource allocation to the users, whereas 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 with 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 (BC) can be achieved with a comparatively low amount of required feedback.

Index Terms—Multiple-input multiple-output orthogonal frequency-division multiplexing (MIMO-OFDM), multiuser systems, resource allocation, score-based scheduling (SB), spatial mode switching.

I. INTRODUCTION

THE SPATIAL dimension of the multiple-input–multiple-output (MIMO) radio channel can be accessed in spatial multiplexing (SMUX) or spatial diversity transmission mode. While in SMUX mode multiple data streams are simultaneously transmitted 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

Manuscript received August 23, 2008; revised March 13, 2009 and June 16, 2009. First published August 11, 2009; current version published January 20, 2010. This work was supported in part by the German Ministry for Education and Research, by Nokia Siemens Networks under Project 3GeT, and by the European Union under the Information Society Technologies project IST-4-027756 WINNER II. This paper was presented in part at the Asilomar Conference, Pacific Grove, CA, October 2007. The review of this paper was coordinated by Prof. G. Bauch.

The authors are with the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, 10587 Berlin, Germany (e-mail: schellmann@hhi.fraunhofer.de; thiele@hhi.fraunhofer.de; thomas.haustein@hhi.fraunhofer.de; jungnickel@hhi.fraunhofer.de).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TVT.2009.2029551

of an adaptive transmission system that selects the transmission mode depending on the actual channel quality to improve the error rate performance for fixed data rate transmission [3], [4] or to increase the spectral efficiency [5]–[8]. In those works, 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 links to multiuser links [10]. The available multiuser 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 (DPC) technique [15]. For individual user constraints like desired target signal-to-interference-plus-noise ratios (SINRs), a framework for the optimum solution of the multiuser 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, whereas with increasing SNR, support for additional SDMA users should successively be added. This result already suggests that mode switching can also be beneficially applied in multiuser MIMO systems.

For proper application of SDMA, information on the state of each user's channel is required at the BS. As the availability of full CSI from all the users at the BS is hardly possible in frequency-division duplex systems, solutions based on limited feedback, which provide partial CSI to the BS, have been introduced [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 simultaneously served 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 the highest rate. It was shown that for a large number of users, the achievable downlink capacity of this system scales similarly as the capacity of the MIMO BC does.

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 that matches his channel conditions is provided, and thus, potential latency problems can be avoided.

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

In a multiuser 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 that are 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 to assign to all users the same amount of resources. In case all users are experiencing identical fading conditions, the scheduler asymptotically achieves proportional fairness. The set of resources to be scheduled can arbitrarily be defined over all dimensions (time/frequency/space) and can flexibly be adapted to any latency constraints. Due to these properties, the approach enables a simple and efficient resource allocation process, which makes it a favorable solution for 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 multiuser MIMO orthogonal frequency-division-multiplexing (OFDM) system, aiming at high throughput while meeting the desired fairness constraints within a predefined time frame. Therefore, we extend the score-based scheduling (SB) 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. As 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 2×2 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 2×2 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.

This 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 of the applied fixed beamforming concept, which is termed the grid of beams (GoB). In Section IV, the concept of adaptive downlink with fair multiuser 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 2×2 MIMO configuration in Section V. Section VI draws some conclusions.

II. SYSTEM MODEL

We consider the downlink of a broadband multiuser 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 simultaneously be active. The transmit vector \mathbf{s} contains up to N_t nonzero 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 nonzero 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, $(\cdot)^H$ is 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 subbands 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 individually be assigned to distinct users. Each chunk is separately processed, and thus, the spatial transmission mode may individually be selected per chunk. The supported spatial modes are termed as single stream (ss) and multistream (ms), which relate to the diversity mode and the SMUX mode, respectively. Due 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 multiuser 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 the 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 further simplify the concept, 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 postdetection SINRs¹ 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 SB 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 can thus 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 SB translates to an even distribution of the chunks over all users. In this paper, 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

correspondingly assigned 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.

III. PRECODING: THE GRID OF BEAMS

In this section, we will give some details on the proper choice of 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 choices of precoding vectors in SU-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 suitable 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 the per-antenna power constraints encountered in practice, it is desirable to use beams for precoding that uniformly distribute the power 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 of 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 previously mentioned, 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 with Rayleigh fading. As opposed to that, it was also shown that, for a nonisotropic input (less than N_t beams are simultaneously active), the spectral efficiency can be improved.

If the channels are spatially correlated, then the dominant eigenmodes are concentrated around the dominant eigenvector

¹The postdetection SINR is the SINR of the useful signal achieved after equalization.

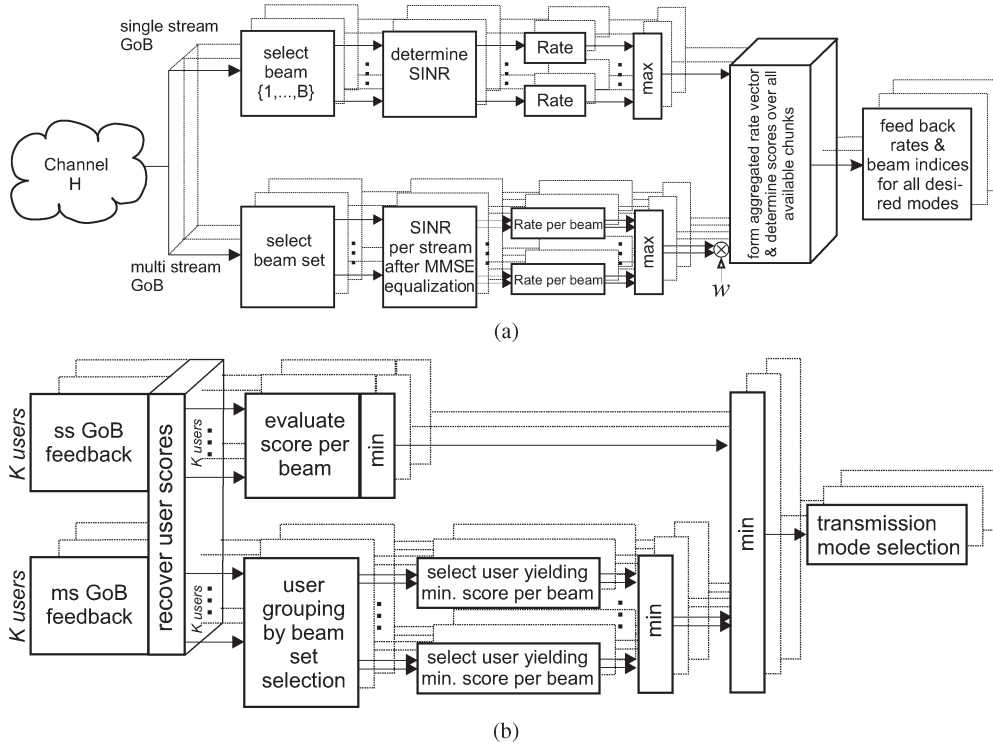


Fig. 1. Structure of the two-step procedure constituting the adaptive transmission concept. (a) Step 1: Channel evaluation and determination of the scores at the UT. (b) Step 2: Resource scheduling with transmission mode selection at the BS.

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—then \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 nonzero. 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 previously described may be a suitable choice for the case of correlated MIMO channels in particular.

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

Preceding remarks 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 only allow 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 two-step procedure illustrated in Fig. 1. It consists of a channel evaluation procedure 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 individually assign the beams in a chunk 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 simultaneously serving users in different modes, 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 separately carried out for each chunk, 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 \mathbf{b}_i is defined as $\mathbf{w}_i = \mathbf{H}\mathbf{b}_i$. Based on MRC, the postdetection SINR for each beam \mathbf{b}_i can be determined according to

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

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 considerably vary 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 rate $r_{\text{ss}}(i)$ per beam \mathbf{b}_i can be determined by using a suitable mapping function $\mathcal{M}(\cdot)$ as

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

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

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 separately performed for each set of beams $m \in \{1, \dots, M\}$: To recover the data stream transmitted on the i th beam of set m , which is denoted as \mathbf{b}_{mi} , $i \in \{1, \dots, Q\}$, we use the MMSE equalizer. The corresponding MMSE equalization vector is defined as

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

The postdetection SINR for each beam \mathbf{b}_{mi} is then given by

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

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

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

Once the beam set is selected, the Q per-beam rates belonging to that set are stored, yielding $R_{\text{ms},i} = r_{\text{ms}}(\bar{m}, i)$, whereas all other may be discarded.³

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

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

For $N_t > 2$, the number of active streams Q supported in ms mode may be in the range of $\{2, \dots, N_t\}$. 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 $R_{\text{ms},i}$ for each of these ms mode options.

B. Determination of the Scores

In the following, the score concept is introduced. Therefore, let u be the chunk indices, and let U be the total number of available chunks. The scores are used to rank the per-beam rates $R_{\text{ss}}(u)$ and $R_{\text{ms},i}(u)$ 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 SB 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 SMUX capability. For a 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 favor ss mode whenever the user rate can be expected to be larger than the rate expected in ms mode. Consider that if a user globally decides for ms mode with Q active beams, the available spatial streams compared with ss mode are increased by a factor Q . As a general result from that, we can assume that the user will also be assigned Q times the streams he would get if he globally selected ss mode. Hence, we can conclude that decision in favor of ss mode should be taken if the rates for the different modes in a single chunk u fulfil

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

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 $R_{\text{ss}}(u)$ and the weighted rates from ms mode $R_{\text{ms},i}(u)$ from all chunks U are aggregated into one vector, which is sorted by magnitude in descending order.⁴ The indices within the sorted vector represents the score ζ of each beam.

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

⁴In the case of multiple resources yielding identical rates, these are ordered in random fashion.

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

C. Resource Scheduling at BS, Step 2

The second step of the process consists of 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 ζ^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 $\zeta_{\text{ss}}^k(u)$ and $\zeta_{\text{ms}}^k(u, i)$, respectively. Resource allocation with transmission mode selection is then successively carried out for all chunks. 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 separately carried out for each transmission mode. For ss mode [upper branch of Fig. 1(b)], the favored user is the user providing the minimum score for that mode,⁶ e.g.,

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

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 as

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

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 as

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

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

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

⁶In the case of multiple users providing identical scores, the stream is given to the user assigned the least beams so far.

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, the concept offers a high potential for further feedback reduction, as the score-based ranking allows each user to preselect his best chunks and his preferred mode to be served with. However, a score-based preselection of the transmission mode may result in a severe performance loss, which is illustrated by the following two examples: If the number of users in the system is small (sparse network), then 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 the highest possible throughput for that user, the final mode selection should be based on his achievable rate—which also requires the availability of the ss rate at the BS. On the other hand, if a user prefers ss mode, but an ms user is selected, the former could still be assigned resources if he provided appropriate rates for the ms mode.

As the scheduling process only aims at assigning each user his best resources, 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: 1) the best ms rate enabling MU-MIMO access and, additionally, 2) 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 higher rate achievable with diversity mode or dual-stream SMUX mode.

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 the same score set is used to rank the available resources for every user, the score-based scheduler will, on average, evenly distribute the available resources over all users (see also [25]). For identical fading statistics, 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 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 favor of the low-rate users, and therefore, the latter will be assigned more resources than the high-rate users. To develop a suitable algorithm for fairness steering, we will firstly derive some basic relations based on the average behavior of the score-based scheduler. These are then used as a 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, then 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 the k th and l th user, respectively, and let $x = w_k/w_l$ 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 that

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

With the weighted scores, the scheduling yields the new achievable rate \bar{R}_k for user k , which can be approximated by

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

The target rate \bar{R}_k should be identical for all users, and therefore, we have $\bar{R}_k = \bar{R}_l \forall k \neq l$. With (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 as $w_l = 1$, then the weighting factors for all users k can simply be obtained from the ratio of their achievable rate R_k and the achievable rate of the reference user R_l .

Evidently, the obtained result cannot directly be applied to our scheduler, as some relations have been simplified and hold for the average behavior 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 fur-

ther be decreased. The iterative algorithm is described as follows.

- 1) Run the scheduling process with unweighted scores, yielding an achievable rate R_k for each user.
- 2) Determine the mean rate \bar{R} over all users, i.e., $\bar{R} = (1/K) \sum_{k=1}^K R_k$.
- 3) Calculate the 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 the 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 the scheduling process.
- 9) Determine D for the new obtained rates R_k .
- 10) If D is smaller than the previous value, then go to step 4; else, use the resource assignment from the previous iteration and end the 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 a 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 kinds of link-level investigations enable an isolated examination of the system behavior 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 multiuser MIMO communication systems. As our main interest here is to illustrate the fundamental behavior 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, see [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). The possible ms modes are thus the two-user MU-MIMO or the 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 copolarized antennas; the antenna spacing is set

to 4λ at the BS and to 0.5λ 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 independently modeled. The mean reception SNR is thus 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 eight subcarrier width. To obtain the per-stream SINR γ for the different transmission modes per chunk, (4) and (6) are calculated for the subcarrier in the center of the chunk⁷ based on ideal knowledge of the channel \mathbf{H} .

For a given SINR value γ , 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 code with constant block length of 1152 bits and supports the fixed symmetric modulation formats up to 64 quadrature-amplitude modulation (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 channel. As both the block length of the code and 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 binary phase-shift-keying modulation with a code rate of 0.5, whereas the latter is achieved with 64QAM modulation with a code rate of 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 ss mode to become prominent. First, the results are based on Shannon information rates. Fig. 2 presents cumulative distribution functions (cdf's) of the achievable user throughput divided by the signal bandwidth (left) and the spectral efficiency in the cell (right) for $K = 10$ users. Focussing on the user throughput (left), we first point out that all users achieve nonzero rates, so indeed all users in the system are conveniently scheduled. We compare the adaptive system described above to a system exclusively supporting either ss or ms mode. For the user throughput (left), we observe that the performance of the adaptive system slightly benefits 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 with the multistream curve, and hence, no gains from switching can be realized here. This observation can be explained as follows: Recall that a beam in ss mode is served with double the power used for a multistream beam. In the low-SNR regime, where the noise dominates the interference from simultaneously active beams, we can thus

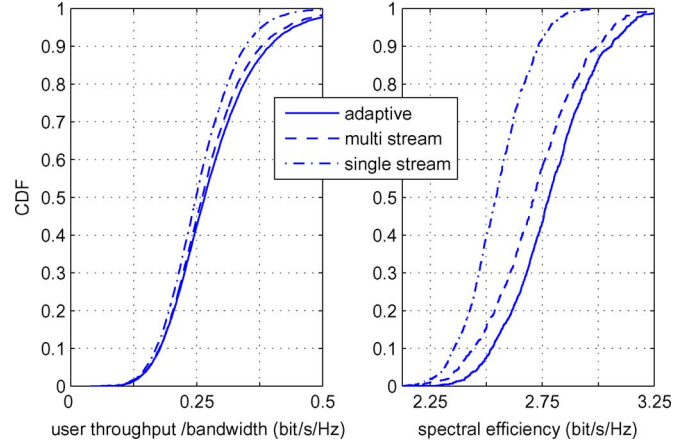


Fig. 2. CDFs of the achievable (left) user throughput and (right) spectral efficiency in the cell based on Shannon information rates. Beam set \mathbf{C}_1 , $K = 10$ users, and SNR = 0 dB.

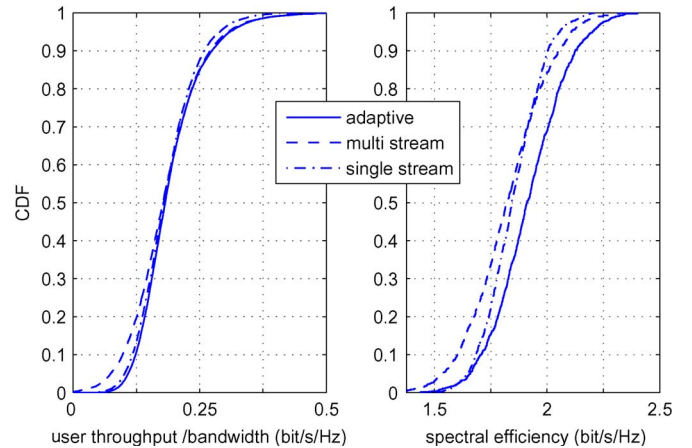


Fig. 3. System performance based on quantized rate mapping function. System setting as in Fig. 2.

expect that the SINR of the selected beam for ss mode is about twice as large as the SINR γ for the corresponding beam in ms mode. Moreover, in ms mode, the amount of beams assigned to each user is twice as large as in ss 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 significantly benefits from the mode switching over the entire cdf region.

Fig. 3 depicts the 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 ss mode if the SINR conditions are low (see the left region of the cdf curve). The cdf of the adaptive system is quite close to the cdf only supporting ss mode, suggesting that this mode is predominantly chosen at low SNR. Considering the spectral efficiency in the cell (see the

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

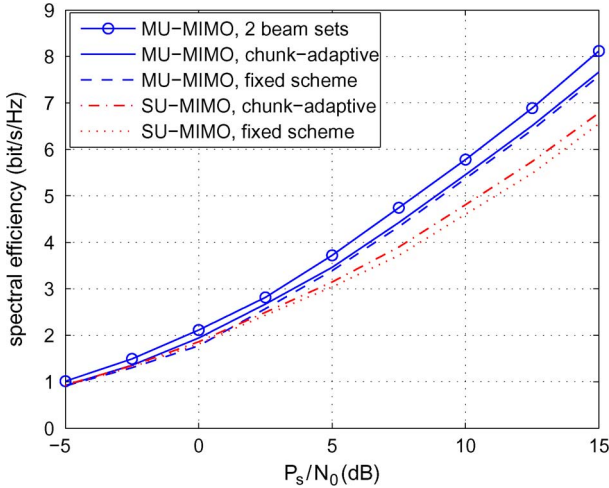


Fig. 4. Median spectral efficiency in the cell for different system configurations. Quantized rates, $K = 10$ users, and beam set C_1 .

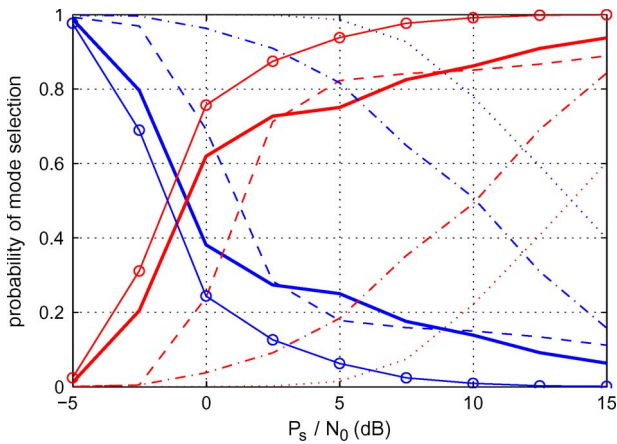


Fig. 5. Probability of mode selection versus SNR. Dark lines: ss mode; light lines: ms mode. Legend and system setup as in Fig. 4.

right subfigure in Fig. 3), we observe that only the left tail of the adaptive system's cdf approaches the curve of the pure ss mode. In the remaining region, substantial gains from mode switching become visible.

In the next step, we will examine the system behavior 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, two beam sets: Similar to 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 simultaneously be served in different modes.

- 3) MU-MIMO, fixed scheme: For each user, a fixed mode and, in case of ss mode, a fixed beam are 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 favor of the mode (and beam) achieving the 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 in either 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 the 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. Figs. 4 and 5 clearly show that these additional degrees of freedom do not only successively increase the achievable system performance but also promote the selection of the ms mode. The crossing point of the probability curves for the ss and ms modes in Fig. 5 highlights the point in the SNR region where the ms mode becomes the dominantly selected mode. Most interestingly, Fig. 5 reveals that, by activating MU-MIMO (configuration 3), this crossing point is shifted by about 8 dB compared with 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 only slightly increased. 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 with 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 more rapidly approach the upper and lower boundary areas with increasing SNR. In particular, for the leftmost ss curve (MU-MIMO, two 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 in Fig. 6, where we compare the median cell throughput for configuration 1 (adaptive MU-MIMO, two beam sets) with a similar system exclusively supporting either ss or ms mode, 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 section with the important observation that with proper application of the MU-MIMO mode, users can conveniently be served in the ms mode even if they experience

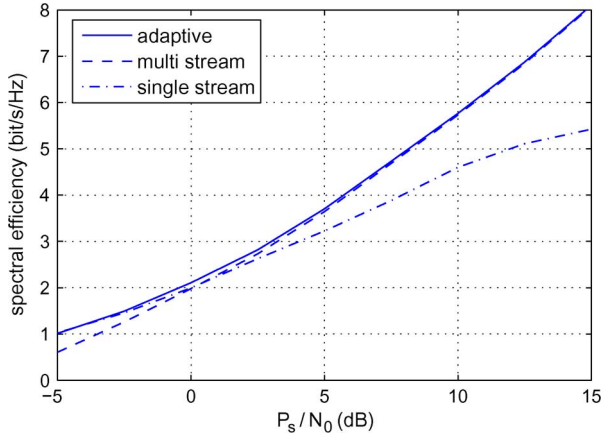


Fig. 6. Median spectral efficiency in the cell achievable with two beam sets and comparison to single-mode systems. Quantized rates and $K = 10$ users.

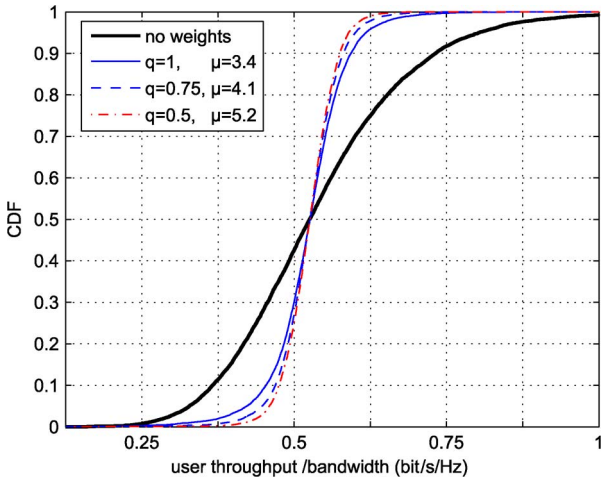


Fig. 7. Fairness steering toward an equal-rate target based on iterative algorithm with tuning factor q . μ —mean required number of iterations. $K = 10$ users, and SNR = 10 dB.

relatively poor SNR conditions.⁸ Thus, the MU-MIMO mode establishes a win/win situation for low- and high-rate users competing for a frequency or time resource: A 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 section, we examine the behavior of the scheduler when 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 and 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 cdf's 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 conveniently operate, as the cdf significantly gains in steepness and thus approaches the equal-rate target, which would be represented

⁸In a cellular system, these are the users at cell edge.

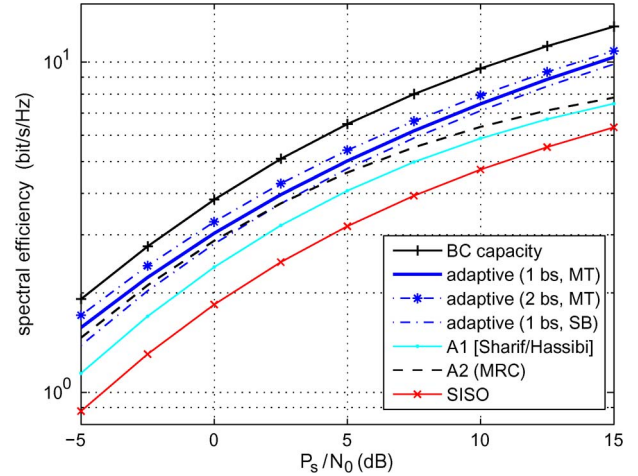


Fig. 8. Comparison of the downlink capacity for various systems versus SNR. $K = 10$ users, and Shannon information rates. bs—beam sets; scheduling: MT—max. throughput, SB—score based.

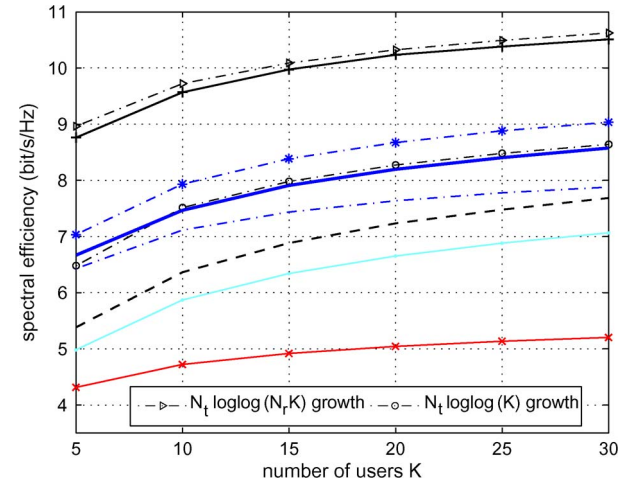


Fig. 9. Scaling of the downlink capacity with number of users at SNR = 10 dB. Same legend as in Fig. 8.

by a vertical line. It is interesting to note that the median user throughput does not change due to the 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 μ 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 that 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 only drops by 3%. However, it should be noted that this loss will substantially be 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 with the upper bound, which is the capacity of the 2×2 BC when full CSI is available at the receivers and the transmitter. As mentioned in Section I, the capacity of the BC was shown to be achievable with the 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 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, the user (single stream) or the user constellation (multistream) 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 with 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 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%.⁹

To see potential gains in the downlink capacity compared with other well-known limited-feedback schemes, we further compare the adaptive system to two multistream approaches with $Q = 2$ simultaneously active beams, which we denote as A1 and A2. For both approaches, we assume the 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 multistream system where the UTs simply carry out MRC for each received beam, i.e., nothing is done to actively combat the interference. The corresponding postdetection SINR can be determined by (6), with the MRC equalization vector $\mathbf{w}_i = \mathbf{H}\mathbf{b}_i$. Feedback and beam assignment are then carried out as in A1.

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

Remark: For the 2×2 MIMO setup considered here, the feedback required for the adaptive system is 50% larger than that for the two systems A1 and A2, as in addition 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 to 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 that found in [23] for single-antenna UTs. The relative difference in throughput increases with increasing SNR, as A1 suffers from 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 postdetection SINRs for both equalizers, which finally translate into a similar throughput performance. For increasing SNR, however, the performance of A2 significantly degrades 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 with the other reference systems. In [20], it has been shown that the BC capacity for independent and identically distributed 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 similarly scaling. 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 roughly grows with a slope of $\log \log(K)$ [20] and is, 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 . Finally, comparing 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 capacity 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 multiuser 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 given feedback, the BS then selects the transmission mode per frequency resource and assigns each user the beams yielding the highest rates. As an example, performance evaluation has been carried out for a 2×2 MIMO configuration, revealing significant gains in spectral efficiency and in user throughput due to mode switching. It turned out that MU-MIMO transmission, where the spatial beams are assigned to 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, then the gains from instantaneous mode switching vanish in most parts of the SNR region, and the system turns into a pure SMUX 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 with the capacity of the SISO and the MIMO BC and of two limited-feedback reference systems supporting MU-MIMO but no interference suppression. This comparison illustrated that the system effectively achieves high performance, significantly outperforming that of the two reference systems. The scheduling concept introduced here can readily be employed in a multicell environment, where users with different mean SNRs are present, as investigated in [35] and [36]. Note that the elementary functionality of the proposed system concept has already been tested in real-time experiments, which have been conducted in a real-world broadband mobile communication environment; for details, see [37] and [42].

Final 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 and the beam to be served on individually per chunk, whereas 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 individually per chunk. Both measures significantly reduce the required feedback but clearly come at the cost of decreased system performance.

ACKNOWLEDGMENT

The authors would like to thank E. Costa and W. Zirwas for their continuous interest in the progress of the work and for stimulating discussions, as well as H. Huang for discussions, helpful comments, and additional material provided, which have proven invaluable to this paper. The authors would further like to acknowledge the contributions of their colleagues in the 3GeT and the WINNER II project.

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, pp. 962–968, Jun. 2005.
- [4] D. Love and R. Heath, "Multi-mode precoding using linear receivers for limited feedback MIMO systems," in *Proc. IEEE Int. Conf. Commun.*, 2004, pp. 448–452.
- [5] S. Catreux, V. Erceg, D. Gesbert, and R. W. Heath, Jr., "Adaptive modulation and MIMO coding for broadband wireless data networks," *IEEE Commun. Mag.*, vol. 40, no. 6, pp. 108–115, Jun. 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 J. Appl. Signal Process.*, vol. 2004, no. 5, pp. 762–771, Jan. 2004.
- [7] J. Lopez-Vicario and C. Anton-Haro, "Adaptive switching between spatial diversity and multiplexing: A cross-layer approach," in *Proc. IST Mobile Wireless Commun. Summit*, Jun. 2005.
- [8] M. Schellmann, V. Jungnickel, A. Sezgin, and E. Costa, "Rate-maximized switching between spatial transmission modes," in *Proc. 40th Asilomar Conf. Signals, Syst., Comput.*, 2006, pp. 1635–1639.
- [9] A. Forenza, 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," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.
- [11] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. Commun.*, 1995, pp. 331–335.
- [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, Jul. 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. IT-29, no. 3, pp. 439–441, May 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 multi-user MIMO systems," *Eur. Trans. Telecommun.*, vol. 18, no. 3, pp. 287–304, Sep. 2007.
- [18] D. Love, R. W. Heath, Jr., W. Santipach, and M. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Commun. Mag.*, 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, Jun. 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 36.211 V8.5.0, *E-UTRA—Physical Channels and Modulation (Rel. 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," in *Proc. IEEE 17th Int. Symp. Pers., Indoor, Mobile Radio Commun.*, Sep. 2006, pp. 1–5.
- [24] R. de Francisco, D. Stock, and Y.-C. Liang, "Balance of multiuser diversity and multiplexing gain in near-orthogonal MIMO systems with limited feedback," in *Proc. IEEE WCNC*, 2007, pp. 1269–1274.
- [25] T. Bonald, "A score-based opportunistic scheduler for fading radio channels," in *Proc. Eur. Wireless*, 2004, pp. 283–292.
- [26] IST-2003-507581 WINNER—D2.7, *Assessment of Advanced Beamforming and MIMO Technologies*, Feb. 2005.
- [27] D. Love, R. W. Heath, Jr., 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 R. W. Heath, Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.
- [29] D. Love and R. W. Heath, Jr., "Equal gain transmission in multiple-input multiple-output wireless systems," *IEEE Trans. Commun.*, vol. 51, no. 7, pp. 1102–1110, Jul. 2003.
- [30] A. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, Jul. 2005.
- [31] F. Boccardi and H. Huang, "A near-optimum technique using linear precoding for the MIMO broadcast channel," in *Proc. IEEE ICASSP*, Apr. 2007, vol. 3, pp. III-17–III-20.
- [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," in *Proc. IEEE VTC—Fall*, Sep. 2005, vol. 1, pp. 473–476.
- [34] J. van de Beek, "Channel quality feedback schemes for 3GPP's evolved-UTRA downlink," in *Proc. IEEE GLOBECOM*, Nov. 2006, pp. 1–5.
- [35] L. Thiele, M. Schellmann, V. Jungnickel, and W. Zirwas, "Capacity scaling of multi-user MIMO with limited feedback in a multi-cell environment," in *Proc. 41st Asilomar Conf. Signals, Syst., Comput.*, Nov. 2007, pp. 93–100.
- [36] L. Thiele, M. Schellmann, T. Wirth, and V. Jungnickel, "Interference-aware scheduling in the synchronous cellular multi-antenna downlink," in *Proc. 69th IEEE VTC—Spring*, Apr. 2009, pp. 1–6.
- [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. Boca Raton, FL: CRC, 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, Nov. 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. 2–4, pp. 1740–1751, Feb.–Apr. 1994.
- [42] T. Wirth, V. Jungnickel, A. Forck, S. Wahls, H. Gaebler, T. Haustein, J. Eichinger, D. Monge, E. Schulz, C. Juchems, F. Luhn, and R. Zavrtak, "Realtime multi-user multi-antenna downlink measurements," in *Proc. IEEE WCNC*, Apr. 2008, pp. 1328–1333.



Malte Schellmann (S'05) received the Dipl.-Ing. (M.S.) degree in electrical engineering and information technology from the Technische Universität München, München, Germany, in 2003 and the Dr.-Ing. (Ph.D.) degree from the Technische Universität Berlin, Berlin, Germany, in 2009.

For his diploma thesis, he worked on equalization in multiple-input–multiple-output (MIMO) systems with the Advanced Micro Devices (AMD), Dresden, Germany. Since 2004, he has been with the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, where he pursued his Ph.D. thesis on the topic of "multiuser MIMO-orthogonal frequency division multiplexing (OFDM) transmission in practice." His current research interests concern practical aspects of multiuser MIMO-OFDM communications and particularly focus on transmission via time-varying channels.



Lars Thiele (S'05) received the Dipl.-Ing. (M.S.) degree in electrical engineering from the Technische Universität Berlin, Berlin, Germany, in 2005. He is currently working toward the Dr.-Ing. (Ph.D.) degree with the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin.

From October 2003 to April 2004, he was a Research Assistant with the vision research lab of Prof. Manjunath with the University of California, Santa Barbara. His main research interests are in the field of performance analysis for multiple-input–multiple-output transmissions in cellular orthogonal frequency division multiplexing systems. He is working on fair-resource allocation algorithms in combination with optimization at the receiver and/or transmitter under limited-feedback assumptions.



Thomas Haustein (M'09) received the Dr.-Ing. (Ph.D.) degree in mobile communications from the Technische Universität Berlin, Berlin, Germany, in 2006.

In 1997, he was with the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute (HHI), Berlin, where he worked on wireless infrared systems and radio communications with multiple antennas and orthogonal frequency division multiplexing. He focused on real-time algorithms for baseband processing and advanced multiuser resource allocation. In 2006, he was with Nokia Siemens Networks, where he conducted research for Long-Term Evolution (LTE) and LTE-advanced. He is currently the Head of the Broadband Mobile Communications Department, HHI.



Volker Jungnickel (M'00) received the Dipl.-Phys. (M.S.) and Dr. rer. nat. (Ph.D.) degrees in physics from Humboldt University of Berlin, Berlin, Germany, in 1992 and 1995, respectively.

He worked on the photoluminescence of semiconductor quantum dots and minimal-invasive laser surgery before joining the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, in 1997. He has been a Lecturer in the field of mobile communications with Technische Universität Berlin since 2002. He has worked on 155-Mb/s wireless indoor communication based on infrared, and on broadband multiple-input–multiple-output (MIMO) systems. He was in charge of the first 1-Gb/s real-time MIMO-orthogonal frequency division multiplexing radio link and its further evolution to the first real-time prototype of the Third Generation Partnership Project–Long Term Evolution standard used for the first adaptive multiuser MIMO field trials. Very recently, his team has realized the first coordinated multipoint downlink transmission in real time. He has contributed to several German and European research projects, including ATMmobil, HyEff, 3GeT, ScaleNet, Winner, WinnerII, Winner+, and EASY-C.