

# Optimal Resource Allocation for Multiuser MIMO-OFDM Systems with User Rate Constraints

Winston W. L. Ho, *Student Member, IEEE*, and Ying-Chang Liang, *Senior Member, IEEE*

**Abstract**—With the proliferation of wireless services, personal connectivity is fast becoming ubiquitous. As the user population demands greater multimedia interactivity, data rate requirements are set to soar. Future wireless systems such as multiple-input multiple-output orthogonal frequency division multiplexing (MIMO-OFDM) need to cater to not only a burgeoning subscriber pool, but also to a higher throughput per user. Furthermore, resource allocation for multiuser MIMO-OFDM systems is vital in optimizing the subcarrier and power allocations to improve the overall system performance. Using convex optimization techniques, this paper proposes an efficient solution to minimize the total transmit power subject to each user's data rate requirement. Through the use of a Lagrangian dual decomposition, the complexity is reduced from one that is exponential in the number of subcarriers  $M$  to one that is only linear in  $M$ . To keep the complexity low, linear beamforming is incorporated at both the transmitter and the receiver. Although frequency-flat fading has been known to plague OFDM resource allocation systems, a modification termed *dual proportional fairness* handles flat or partially frequency-selective fading seamlessly. Due to the non-convexity of the optimization problem, the proposed solution is not guaranteed to be optimal. However, for realistic number of subcarriers, the duality gap is practically zero, and the optimal resource allocation can be evaluated efficiently. Simulation results show large performance gains over a fixed subcarrier allocation.

**Index Terms**—MIMO-OFDM, multiuser, resource allocation, dual decomposition, dual proportional fairness, convex optimization, subcarrier selection.

## I. INTRODUCTION

A multiple-input multiple-output (MIMO) wireless link makes use of multiple antennas at both the transmitter and the receiver. Compared to a single input single output (SISO) random channel, a random MIMO channel has a capacity that grows linearly with the minimum of the number of transmit and receive antennas [1, 2], without requiring additional power or frequency spectrum. In orthogonal frequency division multiplexing (OFDM), a broadband frequency-selective channel is decoupled into multiple flat fading channels, through efficient fast Fourier transform (FFT) operations. The combination of these two technologies, termed MIMO-OFDM [3], is a strong candidate for next generation wireless systems, like 4th generation mobile communications. With the increase in the technology savvy population, there is now a huge demand for rich multimedia interactivity. Commercial cellular systems

have to cope with not only an increase in the number of users, but also with an increase in the data rate requirement per user. MIMO-OFDM addresses these two concerns aptly. Not only is there an increase in overall throughput, there are also more degrees of freedom to accommodate a larger number of users. This is because users can be separated in space as well as frequency.

In practical scenarios, users may be located at different distances from the base station (BS), resulting in different variances for each independent user's channel matrix. Furthermore, users may have subscribed to plans of different data rates. Therefore, practical resource allocation schemes have to take those into consideration. In a cellular system, users experience interference from the BSs of neighbouring cells. Consequently, an important question to answer is how to minimize the transmit power of each individual BS, while maintaining the rate requirements for the group of users currently served. This would help to reduce the interference that each BS produces to neighbouring cells, and as a result improve the whole cellular system's performance.

Convex optimization [4, 5] offers iterative methods to solve several nonlinear communications problems. [10–12] solve the flat-fading uplink/downlink power minimization given user target rates, with the help of convex optimization and the uplink-downlink duality [6–9]. [10] starts with a weighted sum rate maximization for an initial weight vector and an initial sum power. The iterations involve an inner loop, where the weight vector is updated, and an outer loop, where the power is updated by a one-dimensional bisection search. The oscillations near the end are used to derive the time-sharing rate points. [11] solves the sum power minimization problem for the fading broadcast channel by using a dual decomposition. For an initial vector of Lagrange multipliers, the Lagrangian is minimized. Following that, the Lagrange multipliers are updated iteratively by the ellipsoid method. [12] obtains the differentiated capacity for an initial sum power. A bisection search is then used to find the minimum sum power. For these papers, decision feedback equalization (DFE) is performed at the BS during the uplink. Equivalently, dirty paper coding (DPC) is assumed at the BS during the downlink. Time-sharing between the different decoding/encoding orders is required when the target rate-tuple lies on the convex hull of the respective vertices in the capacity region. The time-sharing scheme can be solved by a linear program [10].

The methods above are referred to as interference-balancing (IB) because they take noise into account and allow interference between users/subchannels. On the other hand, methods that cancel out all interference between users as well as

Manuscript received October 1, 2007; revised March 26, 2008. This work will be presented in part at the Vehicular Technology Conference, Calgary, Canada, September 2008.

W. W. L. Ho and Y.-C. Liang are with the Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613 (e-mail: stuwlh@i2r.a-star.edu.sg, ycliang@i2r.a-star.edu.sg)

between subchannels are referred to as zero-forcing (ZF) schemes. Generally, IB techniques have higher complexity than ZF ones, but may have better performance for the low SNR region.

For both the ZF and IB classifications, schemes can be further subdivided into linear and nonlinear schemes. The methods described earlier are known as nonlinear schemes because they involve nonlinear processing like DFE at the receiver or DPC at the transmitter. In MIMO-OFDM, each subcarrier represents a flat fading MIMO channel. Using the nonlinear solutions above, each subcarrier may require a different decoding/encoding order, leading to an undesirable increase in complexity. While this is optimal in terms of minimizing the total transmit power, the demands on the hardware processing capability may far outweigh the benefit of the lower transmit power.

In contrast, linear schemes make use of only linear matrix multiplications for the components of the signal processing. The advantage of linear processing or beamforming is that the complexity is much reduced, leading to a decrease in hardware demand. In addition to reducing the complexity of this multi-carrier system, linear processing tends to be more robust against channel uncertainty, than nonlinear processing like DPC. Furthermore, for a flat fading MIMO broadcast channel, ZF beamforming with time division multiple access (TDMA) has been shown to achieve a sum rate close to the optimal DPC scheme when the number of users is large [19].

For the SISO case, optimal orthogonal frequency division multiple access (OFDMA) downlink resource allocation has been developed in [16], which does not have the complexity of different encoding/decoding orders, since there can only be one user per subcarrier. [14] obtains subcarrier and bit allocations with a goal of minimizing the overall transmit power while maintaining a target BER for a multiuser MIMO-OFDM system. Similar to [16], in [14], there can only be one user per subcarrier. For each subcarrier, the user that achieves the maximum SNR is selected for this subcarrier. [14] and [16] are suitable for frequency-selective fading channels. Frequency-flat channels, if they occur, may result in an inability to guarantee user rates because the decision to select a particular user for one subcarrier would be repeated for all the subcarriers.

In [17], users are classified according to the spatial separability, which is calculated from the correlation between the users' spatial signatures. By grouping the users in this manner, subcarriers can be allocated to the users while ensuring that the highly correlated users would not use the same subcarriers. More specifically, the correlation between any 2 users in different groups is set less than a predefined threshold. Therefore, parallel interference cancellation at the BS during the uplink is assumed to remove all the interference between the users.

In this paper, an efficient method based on convex optimization theory is designed to minimize the total transmit power for MIMO-OFDM communications, subject to individual user rate constraints. This strategy requires only linear transmit and receive processing. Therefore it is applicable to both the downlink and the uplink. By considering the Lagrangian dual of the sum power objective function, the problem is broken

down into  $M$  individual subproblems, where  $M$  is the number of subcarriers. The complexity is thus reduced from one exponential in  $M$  to one linear in  $M$ . Given that  $M$  is typically large for multi-carrier systems, this represents a huge amount of savings. The supergradient of the dual function is then used to update the Lagrange multipliers in finite step sizes. The step sizes are adjusted based on the convergence behaviour in order to speed up the convergence of the algorithm. Furthermore, the algorithm is able to adapt to changing channel conditions. It has been found that methods based on dual decomposition could possibly suffer from a uniformity among the subcarriers, resulting in large oscillations within the algorithm. A solution based on a *dual proportional fairness* is proposed to tackle the event of frequency-flat fading. Simulation results show that with reasonable number of subcarriers, the duality gap is effectively zero, thereby substantiating the proposed solution.

Section II describes the channel model and the strategy of linear block diagonalization (LBD) [18] that separates the users spatially via linear beamforming. The optimal solution to resource allocation for power minimization is given in Section III. An efficient solution based on convex optimization is developed in Section IV. Adjustment of the step size for faster convergence and adaptation to changing channel conditions is discussed in Section V. To handle the event of flat fading channels, a modification based on a dual proportional fairness is introduced in Section VI. Simulation results are given in Section VII. Finally, conclusions are drawn in Section VIII.

*Notations:*

Vectors and matrices are denoted by boldface letters.  $(\cdot)^T$  and  $(\cdot)^H$  denote the transpose and conjugate transpose operations respectively.  $\mathbb{E}[\cdot]$  and  $\text{Tr}(\cdot)$  stand for the expectation and matrix trace operators respectively.  $\|\cdot\|_2$  denotes the vector Euclidean norm, while  $\mathbf{I}_N$  denotes the  $N \times N$  identity matrix.  $\mathbf{A} = \text{blkd}(\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_K)$  represents the block diagonal matrix of the form

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & 0 & \dots & 0 \\ 0 & \mathbf{A}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{A}_K \end{bmatrix}. \quad (1)$$

## II. CHANNEL MODEL AND TRANSMISSION STRATEGY

### A. Channel Model

In this section, a general description of the channel model is given. Consider a cellular-based MIMO-OFDM system with a BS communicating with  $K$  user terminals via  $M$  subcarriers. Suppose the BS is equipped with  $N_T$  antennas and the  $k$ -th user terminal has  $n_k$  antennas. Denote  $N_R = \sum_{k=1}^K n_k$  as the total number of receive antennas. Let  $\sigma_{k,m}$  indicate the presence of the  $k$ -th user on subcarrier  $m$ ;  $\sigma_{k,m} = 1$  if present and 0 if not. Therefore  $\{\sigma_{k,m}\}$  represents the user selection on each subcarrier. Let the rank of the channel matrix of user  $k$  on subcarrier  $m$  be denoted by  $\eta_{k,m}$ , where  $0 \leq \eta_{k,m} \leq \min(n_k, N_T), \forall m$ . The diagram of downlink transmission is shown in Fig. 1. The baseband input-output relationship is represented as

$$\mathbf{y}_d = \mathbf{H}_d \mathbf{x}_d + \mathbf{n}_d, \quad (2)$$

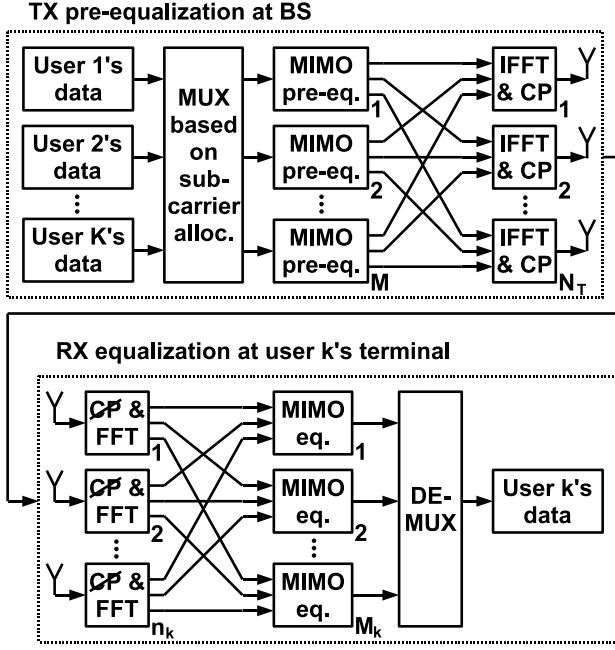


Fig. 1. Block diagram of MIMO-OFDM downlink.

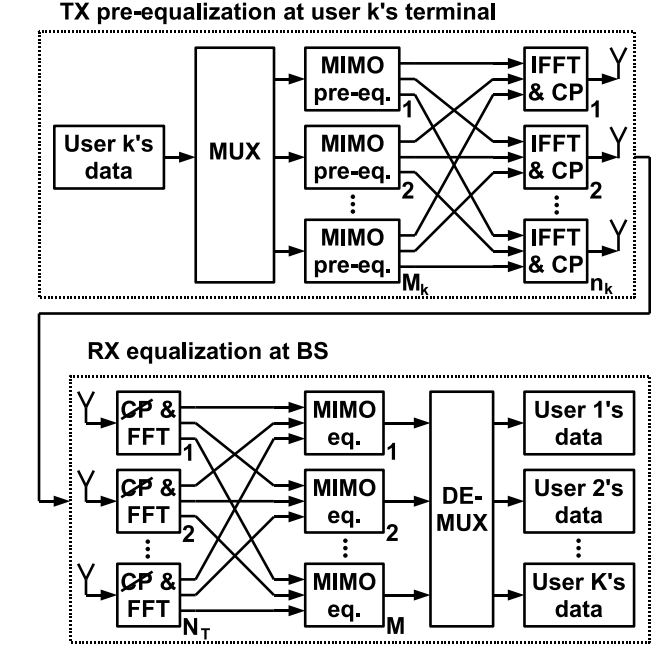


Fig. 2. Block diagram of MIMO-OFDM uplink.

where  $\mathbf{x}_d = [\mathbf{x}_{d,1}^T, \dots, \mathbf{x}_{d,M}^T]^T$  is the transmit signal vector,  $\mathbf{H}_d = \text{blkd}(\mathbf{H}_{d,1}, \dots, \mathbf{H}_{d,M})$  is the channel,  $\mathbf{y}_d = [\mathbf{y}_{d,1}^T, \dots, \mathbf{y}_{d,M}^T]^T$  is the receive signal vector, and  $\mathbf{n}_d$  is the  $MN_R \times 1$  noise vector. Assume that the noise is zero-mean, circularly symmetric complex Gaussian (CSCG) with  $\mathbb{E}[\mathbf{n}_d \mathbf{n}_d^H] = N_0 \mathbf{I}$ , and  $\mathbf{n}_d$  is independent of  $\mathbf{x}_d$ . For the  $m$ -th subcarrier, (2) can be interpreted as

$$\mathbf{y}_{d,m} = \mathbf{H}_{d,m} \mathbf{x}_{d,m} + \mathbf{n}_{d,m}, \quad (3)$$

where  $\mathbf{H}_{d,m} = [\mathbf{H}_{d,1,m}^T, \dots, \mathbf{H}_{d,K,m}^T]^T$  is the  $N_R \times N_T$  random MIMO channel and  $\mathbf{y}_{d,m} = [\mathbf{y}_{d,1,m}^T, \dots, \mathbf{y}_{d,K,m}^T]^T$  is the  $N_R \times 1$  receive signal vector on subcarrier  $m$ .

For the uplink, the block diagram is shown in Fig. 2, where the received signal is given by

$$\mathbf{y}_u = \mathbf{H}_u \mathbf{x}_u + \mathbf{n}_u, \quad (4)$$

with  $\mathbf{H}_u = \text{blkd}(\mathbf{H}_{u,1}, \dots, \mathbf{H}_{u,M})$  being the uplink channel matrix.  $\mathbf{x}_u = [\mathbf{x}_{u,1}^T, \dots, \mathbf{x}_{u,M}^T]^T$  is the transmit signal vector and  $\mathbf{n}_u$  is the  $MN_T \times 1$  noise vector. On the  $m$ -th subcarrier, we have

$$\mathbf{y}_{u,m} = \mathbf{H}_{u,m} \mathbf{x}_{u,m} + \mathbf{n}_{u,m}, \quad (5)$$

where  $\mathbf{x}_{u,m} = [\mathbf{x}_{u,1,m}^T, \dots, \mathbf{x}_{u,K,m}^T]^T$  is the  $N_T \times 1$  transmit signal vector,  $\mathbf{H}_{u,m}$  is the  $N_T \times N_R$  uplink MIMO channel, and  $\mathbf{y}_{u,m}$  is the  $N_T \times 1$  receive signal vector on subcarrier  $m$ . Similar to the downlink case, the noise vector  $\mathbf{n}_{u,m}$  is zero-mean CSCG with  $\mathbb{E}[\mathbf{n}_{u,m} \mathbf{n}_{u,m}^H] = N_0 \mathbf{I}_{N_T}$ .

### B. Equalization using Linear Block Diagonalization

This section describes the transmission scheme for the MIMO-OFDM channel, using linear transmit and receive equalization to block diagonalize the channel.

At each transmission slot, the BS decides on the subcarrier allocation and the transmit preprocessing for the downlink. When more than 1 user share a certain subcarrier, ZF linear block diagonalization (LBD) [18] can be used to separate the users spatially. This creates decoupled channels for all the users. The user terminals then perform channel estimation and receive processing, which can be based on ZF equalization. The user terminals are informed by the BS what transmit processing to employ for the uplink. The BS need only to perform ZF linear receive equalization, because all the user channels are completely decoupled.

Alternatively, if communication is by time division duplex (TDD), the reciprocity principle can be used at the BS to estimate the channel. The transmit matrix operations are identical to the receive matrix operations, greatly simplifying the communications procedure. Likewise, at the user terminals, each user can estimate its own channel because there is no interference between the users. The receive matrix operations are applied directly for the uplink transmission.

1) *Downlink Case:* Consider the transmission over one subcarrier during the downlink. For each user, singular value decomposition (SVD) is applied to the combined channel matrix of all the other users. The last few right singular vectors that correspond to zero singular values give the null space of this combined matrix. Next, each user's matrix is multiplied by the corresponding null space obtained earlier and SVD is performed on the resultant matrix. These two steps would give the transmit and receive equalization matrices. The different users' MIMO channels become completely decoupled, with no interference between the users.

When the number of users is large, all these mutual projections would make each user's subchannels very weak. The advantage of multicarrier MIMO communications is that all

the users do not need to share the same subcarrier. An easy way to exploit this is to constrain the system such that only 1 user occupies each subcarrier. Consequently, no projections are required because the other users are not expecting any data on this subcarrier and would ignore whatever signals they receive. Due to low complexity, this design would be suitable for low cost hardware implementation.

When the number of BS antennas is relatively large, compared to the number of users, the performance may be improved by allowing more than 1 user to share the same subcarrier. Next, the algorithm to calculate the transmit and receive equalization matrices is illustrated for this general case.

Suppose that there are  $K_m$  users in subcarrier  $m$ . Let the downlink channel on this subcarrier be denoted by  $\mathbf{H} = [\mathbf{H}_1^T, \dots, \mathbf{H}_{K_m}^T]^T$ . For user  $k$ , define  $\tilde{\mathbf{H}}_k = [\tilde{\mathbf{H}}_1^T, \dots, \tilde{\mathbf{H}}_{k-1}^T, \tilde{\mathbf{H}}_{k+1}^T, \dots, \tilde{\mathbf{H}}_{K_m}^T]^T$ . Perform SVD on each  $\tilde{\mathbf{H}}_k$ :

$$\tilde{\mathbf{H}}_k = \tilde{\mathbf{U}}_k \tilde{\mathbf{S}}_k \tilde{\mathbf{V}}_k^H = \tilde{\mathbf{U}}_k \tilde{\mathbf{S}}_k \left[ \tilde{\mathbf{V}}_k^{(1)} \tilde{\mathbf{V}}_k^{(0)} \right]^H, \quad (6)$$

where  $\tilde{\mathbf{U}}_k$  and  $\tilde{\mathbf{V}}_k$  are unitary matrices in which the columns are the left and right singular vectors of  $\tilde{\mathbf{H}}_k$  respectively.  $\tilde{\mathbf{S}}_k$  is a diagonal matrix, which may be rectangular, containing the singular values of  $\tilde{\mathbf{H}}_k$ . The columns of  $\tilde{\mathbf{V}}_k^{(1)}$  correspond to non-zero singular values whereas the columns of  $\tilde{\mathbf{V}}_k^{(0)}$  correspond to the zero singular values of  $\tilde{\mathbf{H}}_k$ . Therefore  $\tilde{\mathbf{V}}_k^{(0)}$  is the null space of  $\tilde{\mathbf{H}}_k$ . Each user's channel after mutual projections is

$$\check{\mathbf{H}}_k = \mathbf{H}_k \tilde{\mathbf{V}}_k^{(0)} = \check{\mathbf{U}}_k \check{\mathbf{S}}_k \check{\mathbf{V}}_k^H, \quad (7)$$

where the last equality represents the SVD of  $\check{\mathbf{H}}_k$ . As a result, on each subcarrier, for the group of users currently being served, their channels are completely decoupled and they observe no interference from one another. The transmit equalization matrix  $\mathbf{F}$  is defined as

$$\mathbf{F} = [\mathbf{F}_1, \dots, \mathbf{F}_{K_m}] \quad (8)$$

where  $\mathbf{F}_k^H \mathbf{F}_k = \mathbf{I}$ ,  $1 \leq k \leq K_m$ . For each user here,

$$\mathbf{F}_k = \tilde{\mathbf{V}}_k^{(0)} \check{\mathbf{V}}_k. \quad (9)$$

The receive equalization matrix for this user is

$$\mathbf{W}_k = \check{\mathbf{U}}_k. \quad (10)$$

In sections III and IV, the user subcarrier allocation and the subchannel power loading will be derived. Let  $\mathbf{P}$  be the diagonal power allocation matrix on a this subcarrier, where  $\mathbf{P} = \text{blkd}(\mathbf{P}_1, \dots, \mathbf{P}_k, \dots, \mathbf{P}_{K_m})$  and  $\mathbf{P}_k$  is the power allocation matrix for user  $k$ . Depending on where the diagonal elements of  $\mathbf{P}_k$  are zero, some spatial subchannels may not be used. The final input-output relationship for each user on this subcarrier may be expressed as

$$\begin{aligned} \mathbf{z}_k &= \mathbf{W}_k^H \mathbf{y}_k \\ &= \mathbf{W}_k^H (\mathbf{H}_k \mathbf{x}_k + \mathbf{n}_k) \\ &= \mathbf{W}_k^H (\mathbf{H}_k \mathbf{F}_k \mathbf{P}_k^{1/2} \mathbf{a}_k + \mathbf{n}_k) \\ &= \check{\mathbf{S}}_k \mathbf{P}_k^{1/2} \mathbf{a}_k + \mathbf{W}_k^H \mathbf{n}_k. \end{aligned} \quad (11)$$

Therefore, the data streams for each user are decoupled:

$$z_{k,l} = s_{k,l} \sqrt{\tilde{p}_{k,l}} a_{k,l} + \tilde{n}_{k,l}, \quad (12)$$

where  $a_{k,l}$  is the transmitted data symbol on subchannel  $l$ ,  $s_{k,l}$  and  $\tilde{p}_{k,l}$  are the  $l$ -th diagonal elements of  $\check{\mathbf{S}}_k$  and  $\mathbf{P}_k$  respectively,  $z_{k,l}$  is the received signal, and  $\tilde{n}_{k,l}$  is the zero-mean CSCG noise with variance  $N_0$ . Overall, for this subcarrier, we have

$$\begin{aligned} \mathbf{z} &= \mathbf{W}^H \mathbf{y} \\ &= \mathbf{W}^H (\mathbf{H} \mathbf{x} + \mathbf{n}) \\ &= \mathbf{W}^H (\mathbf{H} \mathbf{F} \mathbf{P}^{1/2} \mathbf{a} + \mathbf{n}) \\ &= \check{\mathbf{S}} \mathbf{P}^{1/2} \mathbf{a} + \mathbf{W}^H \mathbf{n}, \end{aligned} \quad (13)$$

where  $\mathbf{a} = [\mathbf{a}_1^T, \dots, \mathbf{a}_{K_m}^T]^T$  is the transmit data vector in which  $\mathbb{E}[\mathbf{a} \mathbf{a}^H] = \mathbf{I}$ ,  $\mathbf{W} = \text{blkd}(\mathbf{W}_1, \dots, \mathbf{W}_{K_m})$  is the receive equalization matrix,  $\check{\mathbf{S}} = \text{blkd}(\check{\mathbf{S}}_1, \dots, \check{\mathbf{S}}_{K_m})$  is the equivalent channel, and  $\mathbf{z} = [\mathbf{z}_1^T, \dots, \mathbf{z}_{K_m}^T]^T$  is the equalized signal vector at the receiver.

2) *Uplink Case*: The equalization scheme for the uplink can be derived by considering the *dual downlink*. For an uplink channel over one subcarrier,  $\mathbf{H}_u$ , define the dual downlink channel as

$$\mathbf{H}_d = \mathbf{H}_u^H. \quad (14)$$

The same steps as in the previous subsection can be used to derive  $\mathbf{F}$  and  $\mathbf{W}$ , by considering  $\mathbf{H}_d$  as  $\mathbf{H}$ . However, in this case,  $\mathbf{W}$  is the transmit equalization matrix while  $\mathbf{F}$  is the receive equalization matrix. The input-output relationship for this subcarrier is

$$\begin{aligned} \mathbf{z}_u &= \mathbf{F}^H \mathbf{y}_u \\ &= \mathbf{F}^H (\mathbf{H}_u \mathbf{x}_u + \mathbf{n}_u) \\ &= \mathbf{F}^H (\mathbf{H}_d^H \mathbf{W} \mathbf{P}^{1/2} \mathbf{a}_u + \mathbf{n}_u) \\ &= \check{\mathbf{S}} \mathbf{P}^{1/2} \mathbf{a}_u + \mathbf{F}^H \mathbf{n}_u, \end{aligned} \quad (15)$$

Again, the data streams of all the users are decoupled, just as in (12). The same power allocation can be used, giving the same data rates for all the users, be it the downlink or the uplink.

### III. OPTIMAL SOLUTION FOR POWER MINIMIZATION

In this section, the problem of power minimization given user rate requirements is formulated mathematically and the optimal solution is derived. While this is optimal, the complexity is huge because of an exhaustive search over a large set of possible subcarrier allocations.

The objective is to find the optimal subcarrier allocation  $\{\sigma_{k,m}\}$  and power allocation  $\{p_{k,m}\}$  that minimize the overall transmit power subject to satisfying each user's normalized data rate requirement  $\bar{R}_k$  bits per sec per Hz (bps/Hz).<sup>1</sup>

<sup>1</sup>For  $M$  subcarriers, each with bandwidth  $\omega$ , the overall rate for user  $k$  is  $M \bar{R}_k \omega$  bps.  $M \bar{R}_k$  bits are transmitted for user  $k$  in the duration of one OFDM symbol i.e. one channel use.

Mathematically, the optimization can be expressed as

$$\begin{aligned} & \underset{\{\sigma_{k,m}\}, \{p_{k,m}\}}{\text{minimize}} && \sum_{m=1}^M \sum_{k=1}^K p_{k,m} \\ & \text{subject to} && \sum_{m=1}^M r_{k,m} \geq M\bar{R}_k, \quad \forall k \\ & && p_{k,m} \geq 0, \quad \forall k, m \end{aligned} \quad (16)$$

where  $r_{k,m}$  is the rate of user  $k$  on subcarrier  $m$  and it can be written as

$$r_{k,m} = \sum_{l=1}^{\eta_{k,m}} \log_2 \left( 1 + \frac{\tilde{p}_{k,m,l} s_{k,m,l}^2}{\Gamma N_0} \right), \quad (17)$$

where  $s_{k,m,l}$  is the  $l$ -th diagonal element of user  $k$ 's equivalent channel  $\tilde{\mathbf{S}}_{k,m}$  on subcarrier  $m$  as in (12). Therefore  $\{s_{k,m,l}\}$  is dependent on the user selection  $\{\sigma_{k,m}\}$  on subcarrier  $m$ , where  $\sigma_{k,m}$  is as defined in Section II-A.  $\tilde{p}_{k,m,l}$  is the power loading on subchannel  $l$  for user  $k$  on the  $m$ -th subcarrier, and  $p_{k,m} = \sum_{l=1}^{\eta_{k,m}} \tilde{p}_{k,m,l}$ . If  $\sigma_{k,m} = 0$ , we set  $p_{k,m} = 0$ ,  $s_{k,m,l} = 0 \forall l$ , and  $r_{k,m} = 0$ . In (17),  $\Gamma$  is the SNR gap which can be represented as

$$\Gamma = -\frac{\ln(5 \text{ BER})}{1.5} \quad (18)$$

for an uncoded M-QAM modulation with a specified BER [13]. For practical systems that use error-correction coding, the SNR gap can be much smaller. If the subcarrier assignment  $\{\sigma_{k,m}\}$  is fixed, the power allocation can be found for each user separately. If user  $k$  is of interest, the problem becomes

$$\begin{aligned} & \underset{\{p_{k,m}\}}{\text{minimize}} && \sum_{m=1}^M p_{k,m} \\ & \text{subject to} && \sum_{m=1}^M r_{k,m} \geq M\bar{R}_k \\ & && p_{k,m} \geq 0, \quad \forall m \\ & && p_{k,m} = 0, \quad \text{if } \sigma_{k,m} = 0. \end{aligned} \quad (19)$$

Water-filling can be then carried out over user  $k$ 's eigenchannels across all the subcarriers to find the optimal power and rate allocation:

$$\tilde{p}_{k,m,l} = \max \left\{ \frac{\mu_k}{\ln 2} - \frac{\Gamma N_0}{s_{k,m,l}^2}, 0 \right\}, \quad (20)$$

$$\tilde{r}_{k,m,l} = \log_2 \left( \max \left\{ \frac{\mu_k s_{k,m,l}^2}{\ln 2 \Gamma N_0}, 1 \right\} \right), \quad (21)$$

where  $\frac{\mu_k}{\ln 2}$  is the water level such that

$$\sum_{m=1}^M \sum_{l=1}^{\eta_{k,m}} \tilde{r}_{k,m,l} = M\bar{R}_k. \quad (22)$$

To illustrate the water-filling,  $\frac{\mu_k}{\ln 2}$  can be interpreted as the common water level of the power or water poured over channels with river beds equal to  $\frac{\Gamma N_0}{s_{k,m,l}^2}$ . Starting with the maximum number of streams,  $\frac{\mu_k}{\ln 2}$  is evaluated for a decreasing number of streams until the point where the water level is above the highest river bed.

In order to obtain the globally optimal solution, an exhaustive search is needed over all the subcarrier assignments  $\{\sigma_{k,m}\}$  to find the minimum transmit sum power. Thus,  $K$  water-filling procedures over  $M\eta_k$  singular values have to be carried out for each of  $2^{KM}$  possibilities. Even if a constraint is imposed such that only 1 user occupies each subcarrier, there would be  $K^M$  possibilities to test.

#### IV. EFFICIENT SOLUTION FOR POWER MINIMIZATION

In this section, an efficient solution to the power minimization problem is derived based on a Lagrange dual decomposition. First, let us write the problem of (16) as the following optimization problem:

$$\begin{aligned} & \underset{\{r_{k,m}\}}{\text{minimize}} && f(\mathbf{r}) \\ & \text{subject to} && \sum_{m=1}^M \mathbf{r}_m \geq M\bar{\mathbf{R}}, \end{aligned} \quad (23)$$

where  $\mathbf{r} = [\mathbf{r}_1^T, \dots, \mathbf{r}_m^T, \dots, \mathbf{r}_M^T]^T$ , in which  $\mathbf{r}_m = [r_{1,m}, \dots, r_{K,m}]^T$ , is the rate allocation to be optimized.  $f(\cdot)$  is a  $\mathbb{R}^{MK} \rightarrow \mathbb{R}$  function that is not necessarily convex. The rate requirements are represented by the  $\bar{\mathbf{R}} = [\bar{R}_1, \dots, \bar{R}_K]^T$  and “ $\geq$ ” denotes a set of elementwise inequalities. Even though the objective function is not convex, it is still possible to transform this problem into a convex one, by forming the Lagrangian dual of the objective function. This is called the *dual method*. The original optimization is known as the *primal* problem, while the transformed problem is known as the *dual* problem. In the dual method, the Lagrangian of (23) is first evaluated:

$$\mathcal{L}(\mathbf{r}, \boldsymbol{\mu}) = f(\mathbf{r}) + \boldsymbol{\mu}^T \left( M\bar{\mathbf{R}} - \sum_{m=1}^M \mathbf{r}_m \right). \quad (24)$$

where  $\boldsymbol{\mu} = [\mu_1, \dots, \mu_K]^T$  is the vector of Lagrange multipliers. The dual function  $g(\boldsymbol{\mu})$  is defined as the unconstrained minimization of the Lagrangian.

$$g(\boldsymbol{\mu}) = \min_{\mathbf{r}} \mathcal{L}(\mathbf{r}, \boldsymbol{\mu}) = \mathcal{L}(\mathbf{r}^*, \boldsymbol{\mu}). \quad (25)$$

where  $\mathbf{r}^* = \arg \min_{\mathbf{r}} \mathcal{L}(\mathbf{r}, \boldsymbol{\mu})$ . The dual problem is therefore

$$\begin{aligned} & \underset{\boldsymbol{\mu}}{\text{maximize}} && g(\boldsymbol{\mu}) \\ & \text{subject to} && \boldsymbol{\mu} \geq \mathbf{0}. \end{aligned} \quad (26)$$

The dual function is always concave, independent of the convexity of  $f(\cdot)$ . Therefore efficient convex optimization techniques can be used to maximize  $g(\boldsymbol{\mu})$ . If the function  $f(\cdot)$  is convex, it turns out that solving the dual problem is equivalent to solving the primal problem, and both solutions are identical [4]. In our optimization of (16), the objective function is a pointwise minimum of several convex functions. This is clearly not convex. However, the solution to the dual problem is a lower bound for the optimal primal objective function value. The difference between the optimal primal and dual function values is termed the “duality gap.” It has been shown that for multicarrier systems with large  $M$ , the duality gap is negligible [20].

From the previous section, the Lagrangian of the optimization problem (16) is

$$\mathcal{L}_1 = \sum_{m=1}^M \sum_{k=1}^K p_{k,m} + \sum_{k=1}^K \mu_k \left( M\bar{R}_k - \sum_{m=1}^M r_{k,m} \right), \quad (27)$$

where  $\mu_k$  are the Lagrange multipliers as in (24) and  $r_{k,m}$  is given by (17). If the  $\mu_k$  are fixed, the user selection can be done on a per subcarrier basis as follows. Write (27) as

$$\mathcal{L}_1 = \sum_{m=1}^M \mathcal{L}_2(m) + \sum_{k=1}^K \mu_k M\bar{R}_k, \quad (28)$$

where

$$\mathcal{L}_2(m) = \sum_{k=1}^K (p_{k,m} - \mu_k r_{k,m}). \quad (29)$$

Consequently, the problem is decomposed into  $M$  independent subproblems. Assume that the user selection  $\{\sigma_{k,m}\}$  has been fixed. Considering one subcarrier,

$$\mathcal{L}_2(m) = \sum_{k=1}^K \sum_{l=1}^{\eta_k} \left( \tilde{p}_{k,m,l} - \mu_k \log_2 \left( 1 + \frac{\tilde{p}_{k,m,l} s_{k,m,l}^2}{\Gamma N_0} \right) \right). \quad (30)$$

$\mathcal{L}_2(m)$  can then be minimized for each user separately in order to calculate  $\tilde{p}_{k,m,l}$ . By applying the water-filling procedure, the power allocation and rate for the  $l$ -th subchannel of user  $k$  can be found:

$$\tilde{p}_{k,m,l} = \max \left\{ \frac{\mu_k}{\ln 2} - \frac{\Gamma N_0}{s_{k,m,l}^2}, 0 \right\}, \quad (31)$$

$$\tilde{r}_{k,m,l} = \log_2 \left( \max \left\{ \frac{\mu_k s_{k,m,l}^2}{\ln 2 \Gamma N_0}, 1 \right\} \right). \quad (32)$$

Consequently, a search over  $2^K$  possible user selections  $\{\sigma_{k,m}\}$  on subcarrier  $m$  can be carried out to find the best user selection that minimizes  $\mathcal{L}_2(m)$ .

A constraint of only one user per subcarrier would greatly simplify the search, since there would only be  $K$  possible selections to choose from. The user that minimizes  $\mathcal{L}_2(m)$  is selected. If  $\mathcal{L}_2(m) \geq 0$ , this user is dropped and eventually no users are allowed on this subcarrier. This is because a positive value of  $\mathcal{L}_2(m)$  does not serve to minimize  $\mathcal{L}_1$ . Overall, for  $M$  subcarriers, there would only be  $MK$  possibilities to test.

In a more general case, more than one user is allowed per subcarrier. On each subcarrier, once a certain user has been selected, the algorithm proceeds by finding the minimum  $\mathcal{L}_2(m)$  for  $\binom{K}{2}$  possible pairs of users. If this value of  $\mathcal{L}_2(m)$  is more than the value of  $\mathcal{L}_2(m)$  for a single user, the search stops here and only one user is selected for this subcarrier. However, if this value of  $\mathcal{L}_2(m)$  is lower than that of a single user, these two users are confirmed to be using the current subcarrier. The algorithm then proceeds to test all  $\binom{K}{3}$  possible triplets of users. The maximum number of user selections to examine would be  $2^K$ . Over all  $M$  subcarriers, there would be  $M \cdot 2^K$  possibilities to test.

When the number of subcarriers  $M$  is large, the duality gap is negligible [20]. For a certain channel realization, if the

duality gap happens to be zero, the efficient solution offered in this section coincides exactly with the optimal solution. The resource allocation would therefore be optimal, resulting in the least possible power. On the other hand, if the duality gap is not zero, this efficient solution is near-optimal in terms of sum power minimization for target rates.

On each subcarrier, a suboptimal search based on the greedy algorithm can be used to simplify the user selection process given above. As before,  $\mathcal{L}_2(m)$  is evaluated for each of the  $K$  users and the user that gives the minimum  $\mathcal{L}_2(m)$  is selected. Next,  $\mathcal{L}_2(m)$  is calculated for the case where one of the remaining  $K - 1$  users is added to the set. The user that gives the minimum value of  $\mathcal{L}_2(m)$  is selected. If this value of  $\mathcal{L}_2(m)$  is higher than the  $\mathcal{L}_2(m)$  found previously for a single user, this second user is dropped and eventually only one user would occupy this subcarrier.

However, if the current  $\mathcal{L}_2(m)$  value is lower than the previous  $\mathcal{L}_2(m)$  for a single user, these two users are confirmed to use the current subcarrier. The algorithm then proceeds to test if a third user is able to use this subcarrier and so on.

Finally, to complete this power minimization solution, the optimal Lagrange multipliers  $\boldsymbol{\mu}$  that maximize the dual function  $g(\boldsymbol{\mu})$  need to be found.  $g(\boldsymbol{\mu})$  can be maximized by updating  $\boldsymbol{\mu}$  along some search direction, all components at a time. The concavity of  $g(\boldsymbol{\mu})$  guarantees that the maximum can be found by a gradient-based search. Although  $g(\boldsymbol{\mu})$  is concave, it may not be differentiable at all points, so a gradient may not always exist. In spite of this, it is still possible to obtain a search direction by finding a supergradient [21], which is a generalization of a gradient. A supergradient at a point  $\hat{\boldsymbol{\mu}}$  is a vector  $\bar{\mathbf{d}}$  that satisfies

$$g(\tilde{\boldsymbol{\mu}}) \leq g(\hat{\boldsymbol{\mu}}) + \bar{\mathbf{d}}^T (\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}). \quad (33)$$

for every  $\tilde{\boldsymbol{\mu}} \neq \hat{\boldsymbol{\mu}}$ .

*Proposition 1:* For the optimization problem (16) with a dual function value  $g(\hat{\boldsymbol{\mu}}) = \mathcal{L}(\mathbf{r}^*, \hat{\boldsymbol{\mu}})$  at  $\hat{\boldsymbol{\mu}}$ , where  $\mathbf{r}^* = \arg \min_{\mathbf{r}} \mathcal{L}(\mathbf{r}, \hat{\boldsymbol{\mu}})$ , a valid supergradient at the point  $\hat{\boldsymbol{\mu}}$  is given by

$$\bar{\mathbf{d}} = M\bar{\mathbf{R}} - \sum_{m=1}^M \mathbf{r}_m^*. \quad (34)$$

*Proof:*

$$\begin{aligned}
g(\tilde{\boldsymbol{\mu}}) &= \min_{\mathbf{r}} \mathcal{L}(\mathbf{r}, \tilde{\boldsymbol{\mu}}) \\
&= \min_{\mathbf{r}} \sum_{m=1}^M f_m(\mathbf{r}_m) + \tilde{\boldsymbol{\mu}}^T \left( M\bar{\mathbf{R}} - \sum_{m=1}^M \mathbf{r}_m \right) \\
&\leq \sum_{m=1}^M f_m(\mathbf{r}_m^*) + \tilde{\boldsymbol{\mu}}^T \left( M\bar{\mathbf{R}} - \sum_{m=1}^M \mathbf{r}_m^* \right) \\
&= \sum_{m=1}^M f_m(\mathbf{r}_m^*) + \hat{\boldsymbol{\mu}}^T \left( M\bar{\mathbf{R}} - \sum_{m=1}^M \mathbf{r}_m^* \right) \\
&\quad + (\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}})^T \left( M\bar{\mathbf{R}} - \sum_{m=1}^M \mathbf{r}_m^* \right) \\
&= g(\hat{\boldsymbol{\mu}}) + \left( M\bar{\mathbf{R}} - \sum_{m=1}^M \mathbf{r}_m^* \right)^T (\tilde{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}), \quad (35)
\end{aligned}$$

thereby satisfying the supergradient definition (33). ■

A supergradient can be represented as a supporting hyperplane defined by the vector  $(-\mathbf{d}, 1)$  that touches the graph of  $g(\boldsymbol{\mu})$  at the point  $\hat{\boldsymbol{\mu}}$  such that the graph  $g(\boldsymbol{\mu})$  lies below this hyperplane for all  $\boldsymbol{\mu}$ .

In practice, a scaled version of the supergradient,  $\mathbf{d} = [d_1, \dots, d_K]^T = \frac{\bar{\mathbf{d}}}{M}$ , can be used, where

$$d_k = \bar{R}_k - \frac{1}{M} \sum_{m=1}^M r_{k,m}. \quad (36)$$

Therefore, starting from an initial value, the Lagrange multipliers are updated in the positive supergradient direction in order to maximize the dual function.

$$\mu_k(\tau + 1) = \max \{ \mu_k(\tau) + \delta d_k, 0 \}, \quad (37)$$

where  $\tau$  represents the iteration number and  $\delta$  is a small step size.  $\mu_k$  can be interpreted as the reward for user  $k$  to increase its rate. The direction of (37) suggests that if the rate of user  $k$  falls below its target rate, its rate reward  $\mu_k$  should be increased. On the other hand, if user  $k$  exceeds its rate requirement,  $\mu_k$  should be decreased. Furthermore, the rate reward should not fall below zero. Note that for minimization of a convex function, the corresponding generalization of the gradient is the subgradient, in which case, the update is in the negative subgradient direction.

During the optimization process, the dual rates for the users,

$$r_k = \sum_{m=1}^M r_{k,m}, \quad (38)$$

gradually approach the rate requirements  $M\bar{R}_k$ . However, at any point in time, the current subcarrier selections  $\{\sigma_{k,m}\}$  can be captured to solve for the optimal minimum power solution given target rates. As the optimization proceeds, this power value for guaranteed rates will tend to decrease and approach the dual function  $\mathcal{L}_1$ . Unlike algorithms such as steepest-descent, the dual function is not guaranteed to increase monotonically with each iteration. Therefore, the algorithm keeps track of the the subcarrier selection  $\{\sigma_{k,m}\}$  that provides the minimum sum power over all the previous iterations.

## V. ADAPTATION FOR EFFICIENT SOLUTION

The previous section has shown how efficient power minimization can be done using convex optimization techniques. For the Lagrange multiplier update, while any initial value of  $\boldsymbol{\mu}$  can be used, it would be better to start with an estimate of  $\boldsymbol{\mu}$  to shorten the convergence time. Furthermore, a good value of the step size  $\delta$  would also improve the convergence. Too small a step size would result in slow convergence while too large a step size results in low precision. In this section, algorithms are provided to estimate an initial value of  $\boldsymbol{\mu}$  and to update the step size adaptively for faster convergence.

An initial value of  $\boldsymbol{\mu}$  can be found if the subcarrier allocation is fixed cyclicly. Let user  $k$  take subcarriers  $qK + k$ ,  $q = 0, 1, 2, \dots$ . Then  $\mathcal{L}_1$  can be minimized by considering each user separately.

$$\mathcal{L}_1 = \sum_{k=1}^K \mathcal{L}_3(k), \quad \text{where} \quad (39)$$

$$\begin{aligned}
\mathcal{L}_3(k) &= \sum_{m=1}^M p_{k,m} + \mu_k \left( M\bar{R}_k - \sum_{m=1}^M r_{k,m} \right) \\
&= \sum_{m=1}^M \sum_{l=1}^{\eta_{k,m}} \tilde{p}_{k,m,l} + \mu_k \left( M\bar{R}_k - \sum_{m=1}^M \sum_{l=1}^{\eta_{k,m}} \tilde{r}_{k,m,l} \right). \quad (40)
\end{aligned}$$

Water-filling can be applied to calculate the power allocation:

$$\tilde{p}_{k,m,l} = \max \left\{ \frac{\mu_k}{\ln 2} - \frac{\Gamma N_0}{s_{k,m,l}^2}, 0 \right\}, \quad (41)$$

$$\tilde{r}_{k,m,l} = \log_2 \left( \max \left\{ \frac{\mu_k s_{k,m,l}^2}{\ln 2 \Gamma N_0}, 1 \right\} \right), \quad (42)$$

where  $\frac{\mu_k}{\ln 2}$  is the water level such that

$$\sum_{m=1}^M \sum_{l=1}^{\eta_{k,m}} \tilde{r}_{k,m,l} = M\bar{R}_k. \quad (43)$$

Let these values of  $\mu_k$  be the initial values  $\mu_k(1)$ . The initial step size can be chosen as

$$\delta(1) = \xi_1 \frac{\sum_{k=1}^K \mu_k(1)}{\sum_{k=1}^K \bar{R}_k}. \quad (44)$$

where  $\xi_1$  is a positive constant. The step size is adjusted adaptively as the algorithm proceeds, based on the performance of the convergence. Before going into the adaptation algorithm, thresholds are set for the maximum and minimum step size.

$$\delta_{\max} = \xi_{\max} \delta(1), \quad (45)$$

$$\delta_{\min} = \xi_{\min} \delta(1). \quad (46)$$

where the constants are such that  $\xi_{\max} > 1$  and  $0 < \xi_{\min} < 1$ . When the dual rates for all the users are observed to be moving in one direction, the step size  $\delta$  is increased:

$$\delta(\tau + 1) = \delta(\tau) \times \xi_2, \quad (47)$$

where the constant  $\xi_2 > 1$ , or else if a user's dual rate is oscillating, the step size  $\delta$  is decreased:

$$\delta(\tau + 1) = \delta(\tau) / \xi_3, \quad (48)$$

where the constant  $\xi_3 > 1$ . The conditions for these two actions can be defined mathematically. When

$$\begin{aligned} & [d_k(\tau - 1) > 0 \text{ and } d_k(\tau) > 0] \\ \text{or } & [d_k(\tau - 1) < 0 \text{ and } d_k(\tau) < 0] \end{aligned} \quad (49)$$

for all the users, the step size is increased. Else, when

$$\begin{aligned} & \left[ \begin{array}{l} r_k(\tau) - r_k(\tau - 1) < 0 \text{ and} \\ r_k(\tau - 1) - r_k(\tau - 2) > 0 \\ \text{for at least one user} \end{array} \right], \\ \text{or } & \left[ \begin{array}{l} r_k(\tau) - r_k(\tau - 1) > 0 \text{ and} \\ r_k(\tau - 1) - r_k(\tau - 2) < 0 \\ \text{for at least one user} \end{array} \right], \end{aligned} \quad (50)$$

the step size is decreased. If these two conditions are not satisfied, the step size remains as it is.

While any values of the parameters  $\xi_1, \xi_{\max}, \xi_{\min}, \xi_2$ , and  $\xi_3$  could work theoretically, specific values may be chosen to speed up the convergence. A suggested combination of the parameters is  $\xi_1 = 0.1, \xi_{\max} = 5, \xi_{\min} = 0.1, \xi_2 = 1.1$ , and  $\xi_3 = 2$ . The rationale for choosing these is as follows. A large initial value of  $\xi_1$  would result in large oscillations in the beginning, which would tend to stabilize as the step size is reduced. It is found that the given value of  $\xi_1$  would also result in a fast convergence except without large initial oscillations. In the initial stage of the algorithm, the dual rates are relatively far from the rate requirements and would approach the rate requirements without oscillations. This means that it makes sense to increase the step size to speed up the convergence. Once the dual rates are close to the rate requirements, they tend to oscillate around the rate requirements. Therefore the step size is reduced to increase the precision. However, oscillations generally do not eventually disappear in methods based on the supergradient, so a lower limit  $\xi_{\min}$  is set on the step size. In the trivial case of only one user, there are no oscillations during convergence. To prevent the step size from increasing without bound, an upper limit  $\xi_{\max}$  is set. As for the step size adaptation, a small value of  $\xi_2$  ensures that the algorithm would not suddenly go into large oscillations, and if oscillations do occur, a large value of  $\xi_3$  allows the oscillations to be brought down quickly. These benefits have to be traded off with the advantage of a large step size.

It is interesting to see how well this adaptive method based on the supergradient can perform. In the following, we will investigate how close the algorithm can get to the maximum of the dual function,  $g(\boldsymbol{\mu}^*)$ . When the Lagrange multipliers  $\boldsymbol{\mu}$  approach the optimal value  $\boldsymbol{\mu}^*$ , the dual rates  $\mathbf{r}$  tend to hover about the target rates  $M\bar{\mathbf{R}}$ , resulting in oscillations. It is therefore expected that the step size would be close to the minimum threshold  $\delta_{\min}$  due to the adaptation above. Furthermore, the Euclidean distance between  $\mathbf{r}/M$  and  $\bar{\mathbf{R}}$ , or equivalently the supergradient norm  $\|\mathbf{d}^{(\tau)}\|_2$ , would normally be small for a large iteration number  $\tau$ .

*Theorem 2:* Assume that  $\|\mathbf{d}^{(\tau)}\|_2 < d_1, \forall \tau > \tau_1$  and  $\delta^{(\tau)} < \delta_1, \forall \tau > \tau_2$  for some positive real numbers  $d_1$  and  $\delta_1$ , and some positive integers  $\tau_1$  and  $\tau_2$ . Also, assume  $\delta^{(\tau)} \geq \delta_{\min}, \forall \tau$ . Denote the maximum dual function value over all the previous iterations as  $g(\boldsymbol{\mu}_{\text{best}}^{(\tau)})$ . For any  $\epsilon > 0$ ,

it can be shown that  $\exists \tau_3$  such that

$$g(\boldsymbol{\mu}^*) - g(\boldsymbol{\mu}_{\text{best}}^{(\tau)}) < \frac{M\delta_1^2 d_1^2}{2\delta_{\min}} + \epsilon, \quad \forall \tau > \tau_3 \quad (51)$$

*Proof:*

$$\begin{aligned} & \|\boldsymbol{\mu}^{(\tau+1)} - \boldsymbol{\mu}^*\|_2^2 \\ = & \|\boldsymbol{\mu}^{(\tau)} - \boldsymbol{\mu}^*\|_2^2 \\ & + 2\delta^{(\tau)} \mathbf{d}^{(\tau)T} (\boldsymbol{\mu}^{(\tau)} - \boldsymbol{\mu}^*) + \delta^{(\tau)2} \|\mathbf{d}^{(\tau)}\|_2^2 \\ \leq & \|\boldsymbol{\mu}^{(\tau)} - \boldsymbol{\mu}^*\|_2^2 \\ & + \frac{2}{M} \delta^{(\tau)} (g(\boldsymbol{\mu}^{(\tau)}) - g(\boldsymbol{\mu}^*)) + \delta^{(\tau)2} \|\mathbf{d}^{(\tau)}\|_2^2, \end{aligned} \quad (52)$$

from the definition of the supergradient. Due to recursion, we have

$$\begin{aligned} & \|\boldsymbol{\mu}^{(\tau+1)} - \boldsymbol{\mu}^*\|_2^2 \\ \leq & \|\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^*\|_2^2 - \frac{2}{M} \sum_{t=1}^{\tau} \delta^{(t)} (g(\boldsymbol{\mu}^*) - g(\boldsymbol{\mu}^{(t)})) \\ & + \sum_{t=1}^{\tau} \delta^{(t)2} \|\mathbf{d}^{(t)}\|_2^2. \end{aligned} \quad (54)$$

Let  $\beta = \|\boldsymbol{\mu}^{(1)} - \boldsymbol{\mu}^*\|_2$ . Then

$$\begin{aligned} 0 \leq & \beta^2 - \frac{2}{M} \sum_{t=1}^{\tau} \delta^{(t)} (g(\boldsymbol{\mu}^*) - g(\boldsymbol{\mu}^{(t)})) \\ & + \sum_{t=1}^{\tau} \delta^{(t)2} \|\mathbf{d}^{(t)}\|_2^2. \end{aligned} \quad (55)$$

Since  $g(\boldsymbol{\mu}^*) - g(\boldsymbol{\mu}_{\text{best}}^{(t)}) \leq g(\boldsymbol{\mu}^*) - g(\boldsymbol{\mu}^{(t)})$ ,

$$\begin{aligned} \frac{2}{M} \sum_{t=1}^{\tau} \delta^{(t)} (g(\boldsymbol{\mu}^*) - g(\boldsymbol{\mu}_{\text{best}}^{(t)})) & \leq \beta^2 + \sum_{t=1}^{\tau} \delta^{(t)2} \|\mathbf{d}^{(t)}\|_2^2, \\ (g(\boldsymbol{\mu}^*) - g(\boldsymbol{\mu}_{\text{best}}^{(\tau)})) \frac{2\tau\delta_{\min}}{M} & \leq \beta^2 + \sum_{t=1}^{\tau} \delta^{(t)2} \|\mathbf{d}^{(t)}\|_2^2. \end{aligned} \quad (56)$$

Denote  $\tau_4 = \max\{\tau_1, \tau_2\}$  and define  $\tau_3$  as

$$\tau_3 = \left\lceil \max \left\{ \frac{M\beta^2}{\delta_{\min}\epsilon}, \frac{M \sum_{t=1}^{\tau_4} \delta^{(t)2} \|\mathbf{d}^{(t)}\|_2^2}{\delta_{\min}\epsilon} \right\} \right\rceil. \quad (58)$$

Then

$$\begin{aligned} & g(\boldsymbol{\mu}^*) - g(\boldsymbol{\mu}_{\text{best}}^{(\tau)}) \\ \leq & \frac{M\beta^2}{2\tau\delta_{\min}} + \frac{M \sum_{t=1}^{\tau_4} \delta^{(t)2} \|\mathbf{d}^{(t)}\|_2^2}{2\tau\delta_{\min}} \\ & + \frac{M \sum_{t=\tau_4+1}^{\tau} \delta^{(t)2} \|\mathbf{d}^{(t)}\|_2^2}{2\tau\delta_{\min}} \end{aligned} \quad (59)$$

$$\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} + \frac{M\tau\delta_1^2 d_1^2}{2\tau\delta_{\min}}, \quad \forall \tau > \tau_3 \quad (60)$$

$$= \frac{M\delta_1^2 d_1^2}{2\delta_{\min}} + \epsilon, \quad \forall \tau > \tau_3. \quad (61)$$

With the mentioned adaptations in place, the optimization algorithm in the previous section can be applied for time-varying



channels without a need to re-initialize  $\mu$ . This is because the relative channel strengths of different users would not tend to change drastically.  $\mu_k$ , which represents the rate reward for user  $k$ , would update to track the channel conditions. Similarly,  $\mu_k$  adapts to track user  $k$ 's rate requirements. When there is a change in the channel or the rate requirements, the thresholds  $\delta_{\max}$  and  $\delta_{\min}$  are recalculated and the last known best subcarrier allocation is reset.  $\mu$  and  $\delta$  are not re-initialized. It is suggested that the algorithm be run for a certain number of iterations before the actual usage of the subcarrier allocation, because it may take a few iterations for the sum power to fall, below that of a fixed subcarrier allocation for example.

## VI. DUAL PROPORTIONAL FAIRNESS

The optimization algorithm in Section IV is immediately applicable to harsh wireless channels. As the MIMO channel is frequency-selective in this case, the user selection on each subcarrier is optimized to provide the minimum overall transmit power. However, a problem arises for frequency-flat fading channels, if they ever occur. In a perfectly flat fading channel, user selection on one subcarrier is repeated for all the subcarriers. When this happens, only one or a few of the users are allocated subcarriers at any one time. This has serious consequences for the algorithm. The subcarrier allocation  $\{\sigma_{k,m}\}$  given by the optimization is unable to guarantee all the users' rate requirements.

In this section, a solution based on convex optimization theory is developed that can tackle the event of frequency-flat fading. This flat fading management is based on a concept that will be called *dual proportional fairness*. This is inspired by the principle of proportional fairness (PF) [22] in which there is a certain randomness to be exploited. While in PF, the nature of the fluctuating channel is used to design the time schedules, in dual PF, the nature of the fluctuating dual rates is utilized to design the subcarrier allocation.

### A. Principle of Dual Proportional Fairness

In the dual method of convex optimization, for example in power minimization, the Lagrange multipliers  $\mu$  represent a tangent plane in a graph of power versus user rates. In this graph, there are several power surfaces, each representing a different subcarrier allocation. The pointwise minimum of all these power surfaces represent the minimum sum power for any given tuple of user rate requirements. When the number of subcarriers is large, there are more power surfaces corresponding to various subcarrier allocations and the pointwise minimum of these power surfaces tend to assume a convex shape. During the optimization process, the tangent plane is in contact with this minimum surface. The coordinates at this contact point give the current dual rates for all the users. As the Lagrange multipliers get updated, the tangent plane adjusts and the point of contact shifts such that the dual rates approach the users' target rates. Convergence occurs when the dual rates hit the target rates and the minimum sum power is achieved.

Frequency-flat fading channels pose a problem because the points where the power surfaces can touch the tangent plane are collinear. For now, assume that only one user occupies each

subcarrier. Consider the case of 2 users. In simulations, it is impossible for the tangent plane to touch the centre power surface, corresponding to a subcarrier allocation of 50% to user 1 and 50% to user 2, without touching the other power surfaces. As a result, the algorithm oscillates between giving all the subcarriers to user 1 or all to user 2. Consequently, each user's rate swings between zero and a value larger than its rate requirement.

Based on this understanding, a flat fading management based on *dual proportional fairness* is proposed. In a flat fading scenario, the power allocations and rates for user  $k$  on all the subcarriers are identical:

$$p_{k,m} = \hat{p}_k, \quad \forall m \quad (62)$$

$$r_{k,m} = \hat{r}_k, \quad \forall m \quad (63)$$

$$p_k = M_k \hat{p}_k \quad (64)$$

$$r_k = M_k \hat{r}_k, \quad (65)$$

where  $M_k$  is the number of subcarriers allocated to user  $k$  and  $\sum_{k=1}^K M_k = M$ . Consider the case of two users. The possible coordinates given by the optimization algorithm are

$$(M\hat{r}_1, 0, M\hat{p}_1) \quad (66)$$

$$(0, M\hat{r}_2, M\hat{p}_2) . \quad (67)$$

Another coordinate, not given by the original optimization, is also possible:

$$(M_1\hat{r}_1, M_2\hat{r}_2, M_1\hat{p}_1 + M_2\hat{p}_2) . \quad (68)$$

It can be seen that these three coordinates are collinear. This concept can be extended to more than 2 users. The trick is now to find the right combination of  $\{M_k\}$  that minimizes the sum power. This can be found in the following three steps:

1. Identify the flat fading users.
2. Identify the flat fading groups.
3. Distribute the subcarriers proportionally for each fading group.

### B. Algorithm for Flat Fading Management

1) *Identify the flat fading users:* Flat fading users are identified as users with rates that oscillate largely or drop to zero:

$$\left\{ \begin{array}{l} r_k > 1.2 M\bar{R}_k \text{ at least once} \\ \text{and } r_k < 0.8 M\bar{R}_k \text{ at least once} \\ \text{or } r_k = 0 \text{ at least once} \end{array} \right\} \quad (69)$$

in the current and previous 9 iterations. Assume there are  $K_{ff}$  such users.

2) *Identify the flat fading groups:* For each flat fading user, look back to see when he had received a dual rate higher than his rate requirement. (If he had not, the flat fading management cannot be done right now.) Find out the minimum number of subcarriers user  $k$  needs to just fulfill his rate requirement. Let this be  $\bar{M}_k$ .

Next, consider all users pairwise. Take user 1 and user 2 for example. Find out where the subcarriers allocated to user 1,  $\Sigma_1$ , overlaps with the subcarriers of user 2,  $\Sigma_2$ . If they do

overlap, users 1 and 2 are in the same group  $G_v$ . The union of subcarriers is taken as the flat fading subcarriers of this group,  $\Sigma_{G_v}$ . Continue this process for all  $K_{ff}$  flat fading users. Users that are not interlinked in this manner are placed in separate flat fading groups. Assume there are  $K_v$  users in each fading group  $G_v$ .

3) *Distribute the subcarriers proportionally for each fading group:* Let there be  $\tilde{M}_v$  flat fading subcarriers in  $G_v$ . First assume the special case of flat fading over all the subcarriers. Users are allocated subcarriers cyclically until user  $k$  gets a maximum of

$$\text{round} \left[ \frac{\bar{M}_k}{\sum_{k \in G_v} \bar{M}_k} \tilde{M}_v \right] \quad (70)$$

subcarriers. To make sure all the subcarriers get allocated, the last user can get all the remaining subcarriers.

An additional modification to (70) allows the algorithm to handle the most general case of partially frequency-selective channels. Take for example the case of two users. In the graph of power versus user rates, only a subset of subcarrier allocations result in collinear points of contact with the tangent plane. This time, oscillations do occur but they are not between zero and very high rates. Instead, each user's dual rate oscillates above and below its rate requirement while its dual rate does not drop to zero. Practically, taking the current subcarrier allocation  $\{\sigma_{k,m}\}$  still allows the user rates to be guaranteed, but this is at an expense of higher transmit power that also oscillates largely. In the following, a modification to (70) is developed that allows smooth convergence for the general case of partially frequency-selective channels.

For each user, find the subcarriers that were allocated to this user for the current and previous 9 iterations. Let there be  $M_{k,\min}$  such subcarriers. Let  $\tilde{\Sigma}_{G_v}$  be the subcarriers of group  $G_v$  with the subcarriers corresponding to  $M_{k,\min}$  of all flat fading users removed. Let there be  $\tilde{M}_v$  flat fading subcarriers in  $\tilde{\Sigma}_{G_v}$ . These subcarriers are distributed in a similar manner as in the previous section. All flat fading users get allocated their respective  $M_{k,\min}$  subcarriers. The initial estimated number of subcarriers each flat fading user would get from  $\tilde{\Sigma}_{G_v}$  is

$$\bar{M}_k = \max \{ \bar{M}_k - M_{k,\min}, 0 \}. \quad (71)$$

Users are allocated subcarriers cyclically until user  $k$  gets a maximum of

$$\text{round} \left[ \frac{\bar{M}_k}{\sum_{k \in G_v} \bar{M}_k} \tilde{M}_v \right] \quad (72)$$

subcarriers. Again, to handle any rounding errors, the last user is allocated all the remaining subcarriers. Subcarriers that are not affected by the flat fading management are assigned the same subcarriers as given by the original solution without any flat fading management. For the purpose of adaptation, when the channel or rate requirements change, this algorithm is restarted. As in Section V,  $\mu$  and  $\delta$  are not re-initialized.

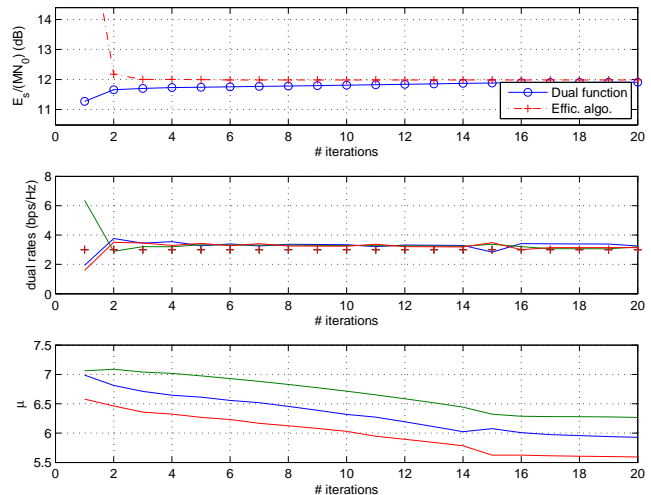


Fig. 3. Typical convergence behaviour of the efficient algorithm applied to a  $3 \times [3, 3, 3]$  MIMO system with  $M = 64$  subcarriers.

## VII. SIMULATION RESULTS

This section first shows the convergence behaviour of the proposed algorithm for certain typical scenarios. Following that, the performance of the efficient subcarrier allocation versus a fixed subcarrier allocation is examined. Other heuristic algorithms are also included for comparison.

Unless otherwise stated, the setup is a  $3 \times [3, 3, 3]$  MIMO system, where the base station has 3 antennas and there are 3 user terminals with 3 antennas each, and the rate requirement is  $\bar{R}_k = 3$  bps/Hz,  $\forall k$ , with an SNR gap of 3 dB. The number of subcarriers is  $M = 64$ . It is assumed that each subcarrier is occupied by at most one user only. The channel is frequency-selective with 17 taps and has a uniform power delay profile. The algorithm in Section VI-B is used in all the simulations. Step 1 involves an automatic identification of flat fading users. If there are flat fading users detected, steps 2 and 3 are then employed for flat fading management. For the graphs, the SNR is defined as  $\frac{\sum_{m=1}^M \sum_{k=1}^K p_{k,m}}{MN_0} = \frac{E_s}{MN_0}$ , where total transmitted signal energy is divided by total noise energy. Therefore the dual function is also scaled by  $\frac{1}{MN_0}$  for comparison.

Fig. 3 illustrates the typical convergence behaviour for these default settings. As can be seen, the sum power required for the efficient subcarrier allocation quickly drops to a near-optimal value, in just 2 iterations for this example. Note that this sum power is for *guaranteed* rates, as shown by the '+' symbols in the second subgraph. The dual function, on the other hand, corresponds to the dual rates denoted by the lines in the second subgraph. The power for the efficient allocation approaches the dual function value, showing that the duality gap is almost zero.

To see the concept of dual proportional fairness at work, consider a partially frequency-selective fading channel, with flat fading over 20 out of 64 subcarriers. As expected, the dual rates in Fig. 4 fluctuate over a wide range, suggesting that the sum power would be far from optimal. However, with the flat fading management, the algorithm easily obtains a near-optimal sum power in only 11 iterations. This is because the

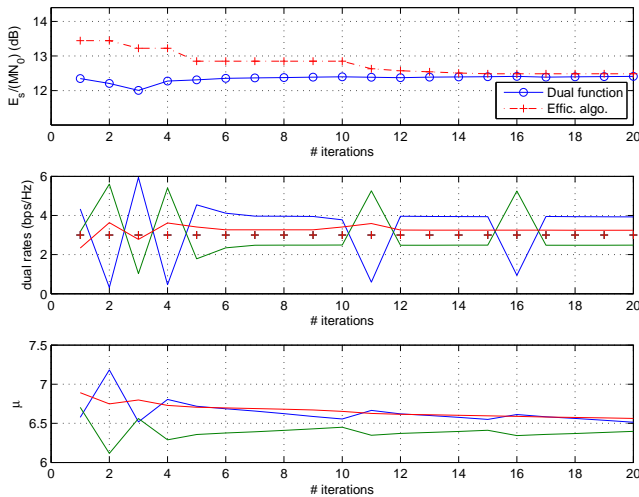


Fig. 4. Sample convergence for a partially frequency-selective channel, with flat fading over subcarriers 21 to 40.

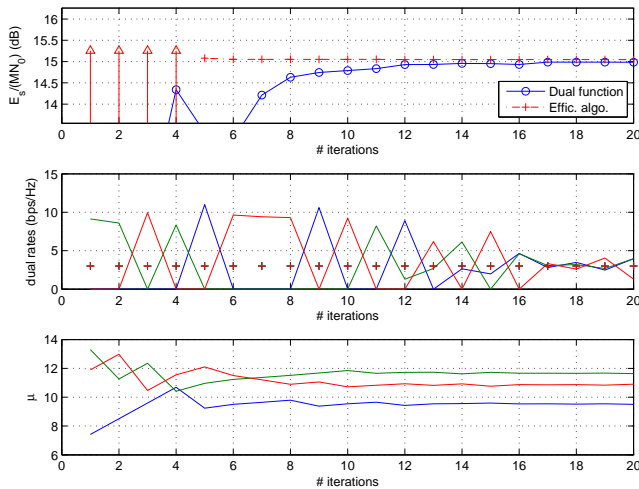


Fig. 5. Convergence behaviour for a weakly frequency-selective channel.

nature of the fluctuating dual rates are used to balance the subcarrier assignment between the users.

Fig. 5 shows the convergence behaviour of the algorithm applied to a channel with a power delay profile with only 2 taps: 0.999 and 0.001. This is an example of a channel with almost flat fading. Again, the dual rates fluctuate wildly, and the flat fading management is automatically started. Without flat fading management, it is often impossible to guarantee user rates because at least one user is not allocated any subcarriers, as can be seen by the zero dual rates. However, the proposed efficient allocation is able to attain a satisfactory sum power in just 4 iterations. The vertical lines in the first few iterations represent the instances where the proposed algorithm cannot give the solution as it is still evaluating the subcarrier allocation based on dual proportional fairness.

In the absence of prior channel knowledge, an equal number of subcarriers should be allocated to each user in a fixed scheme. To obtain some frequency diversity, a distributed cyclic subcarrier allocation is chosen due to its robustness

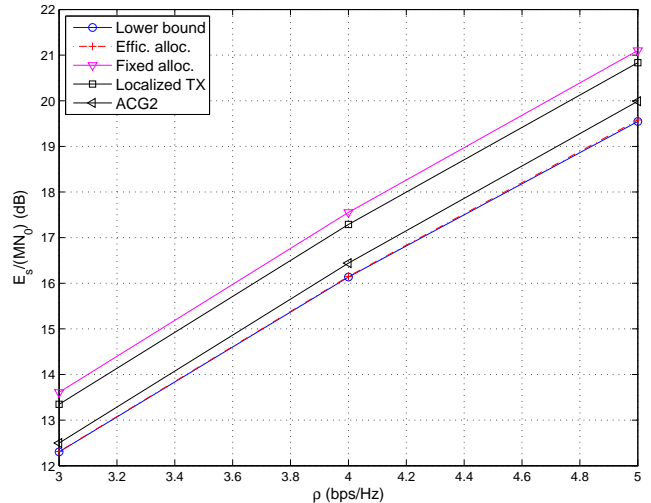


Fig. 6. Required total transmit power for various data rate requirements.

to frequency-selective fading. In this fixed allocation scheme, user  $k$  takes subcarriers  $qK + k$ ,  $q = 0, 1, 2, \dots$ . Another scheme, the “amplitude-craving greedy” (ACG) algorithm of [15] allocates subcarriers intelligently based on users’ rate requirements as well as channel strengths, for SISO-OFDM. In order to extend this heuristic algorithm to the MIMO-OFDM case, we modify the algorithm by substituting the SISO channel strength of [15] with the mean of the squared absolute values of the MIMO channel matrix elements i.e.  $\tilde{c}_{k,m} = \frac{\text{Tr}(\mathbf{H}_{k,m}^H \mathbf{H}_{k,m})}{N_T n_k}$ . This will be referred to as “ACG2” in the graphs.

A simple allocation scheme takes the form of localized transmission, where a block of consecutive subcarriers is allocated to each user. To achieve some multiuser diversity, the assignment is adapted based on the channel conditions of the different users. The first  $\hat{M} = \lfloor \frac{M}{K} \rfloor$  subcarriers are given to the user with the highest average channel strength  $\tilde{c}_{k,1} = \frac{\sum_{m=1}^{\hat{M}} \tilde{c}_{k,m}}{\hat{M}}$ . The next  $\hat{M}$  subcarriers are allocated to one of the remaining users with the highest channel strength, and so on. Finally, the last user gets all the remaining subcarriers. This is labelled “Localized TX” in the graphs.

The graphs include the “Lower bound” i.e.  $\frac{g(\mu)}{MN_0}$  for the solution obtained with optimal resource allocation, based on the fact that the value of the dual function  $g(\mu)$  from (25) is always a lower bound to the minimum transmit power  $\sum_{m=1}^M \sum_{k=1}^K p_{k,m}$ . Therefore, the solution  $\frac{E_s}{MN_0}$  achieved with optimal resource allocation is upper and lower bounded by the proposed “Effic. alloc.” and “Lower bound” respectively. If these two bounds coincide, the duality gap is zero and the proposed efficient allocation is also optimal.

In Fig. 6, the transmit power is plotted against the rate requirement  $\rho$ , where the rate requirement vector is  $\tilde{R}_k = \rho$  bps/Hz,  $\forall k$ . As expected, the sum power increases with the rate requirements while the efficient allocation performs uniformly better than the fixed allocation. At a common rate requirement of 4 bps/Hz for each user, the gain of the efficient subcarrier allocation over a fixed allocation is 1.4 dB.

Fig. 7 shows the graph of BER requirement versus the sum

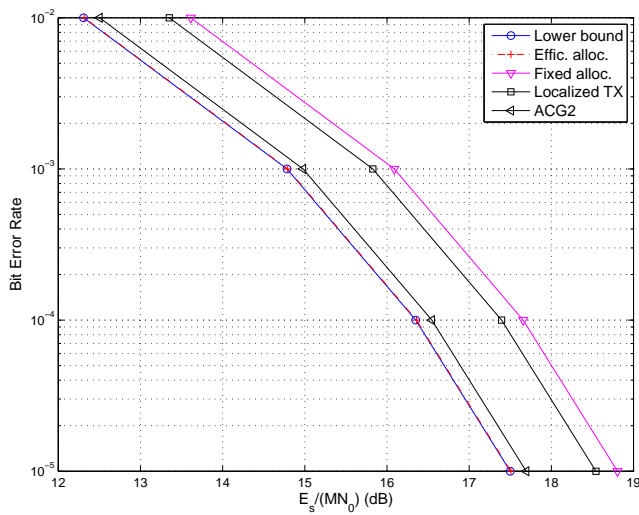


Fig. 7. BER versus sum power for the different subcarrier allocation schemes.

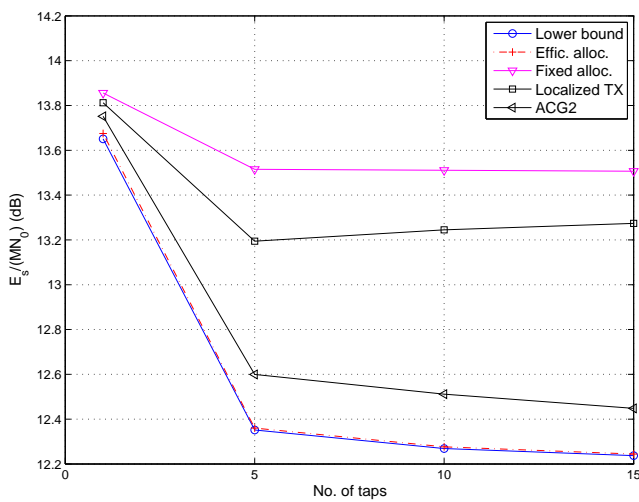


Fig. 8. Graph of sum power versus number of taps in power delay profile, showing the effect of channel frequency selectivity.

power. An uncoded M-QAM modulation is assumed in this case. At a BERs of  $10^{-3}$  to  $10^{-5}$ , the SNR gain appears relatively constant at 1.2 dB. This is due to the similar effect of the SNR gap on both the efficient and fixed allocation schemes.

The effect of channel frequency selectivity is tested in Fig. 8. The number of taps is varied from 1 to 15. The gain of the efficient subcarrier allocation grows as the channel becomes more frequency-selective. This is because a fixed subcarrier allocation would not be able to adapt to take advantage of the diverse channel conditions. With a flat fading channel, the gain is rather small, about 0.2 dB. This can be explained by the fact that with similar rate requirements and similar channel strengths among the users, a fixed allocation of subcarriers would serve just as well to distribute the subcarriers equally for all the users in a flat fading scenario.

Fig. 9 shows the effect of the number of antennas. The setup here is  $\bar{n} \times [\bar{n}, \bar{n}, \bar{n}]$ , where  $\bar{n}$  is varied from 1 to 4. As the number of antennas increase, the sum power required decreases, for the same target rates. This graph clearly shows

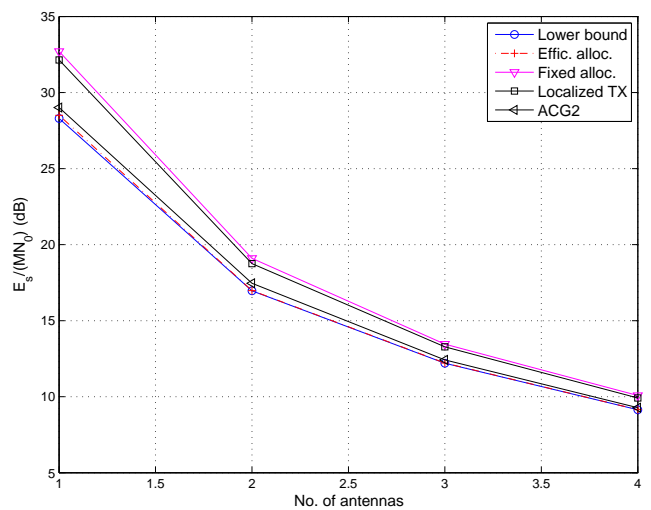


Fig. 9. Transmit power versus number of antennas  $\bar{n}$ , for a  $\bar{n} \times [\bar{n}, \bar{n}, \bar{n}]$  MIMO setup.

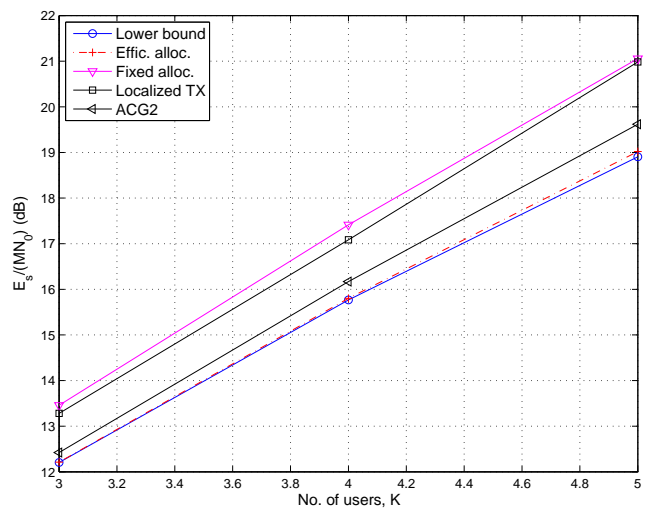


Fig. 10. Sum power for 3, 4, and 5 users in the system.

the advantage of MIMO communications over SISO communications. Even by just increasing the number of antennas  $\bar{n}$  from 1 to 2, the sum power can be decreased by over 10 dB.

Fig. 10 plots the performance with different number of users. Values of  $K$  range from 3 to 5. As the number of users increases, the sum power increases due to a higher sum rate requirement. It can be seen that the gain over a fixed subcarrier allocation also increases. This is because there is greater potential to exploit the multiuser diversity as there are more users introduced into the system. For example, with 5 users, the gain is 2 dB, compared to 1.2 dB with only 3 users.

For the performance comparisons so far, localized TX has a lower sum power than the fixed allocation. This is because by selecting the user with the highest channel strength for each block of subcarriers, some multiuser diversity is exploited. The ACG2 shows a further improvement from localized TX because both the number and positions of the subcarriers are adapted for each user.

In a general setting, users have differentiated rate require-

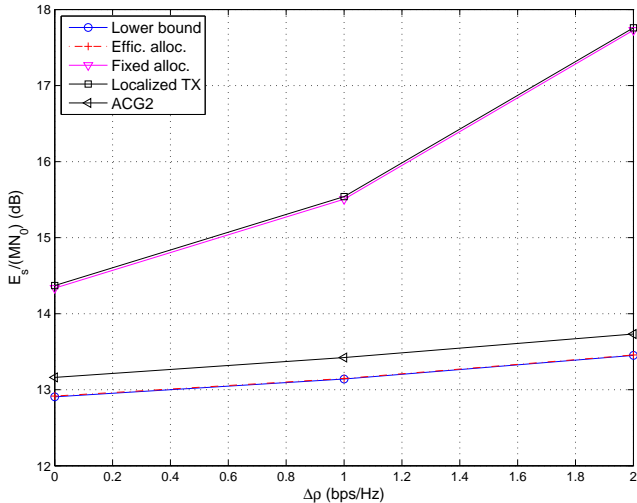


Fig. 11. Transmit power for differentiated rate requirements given by  $\bar{\mathbf{R}} = [3, 3 - \Delta\rho, 3 + \Delta\rho]^T$  bps/Hz, with channel strengths  $\mathbf{c} = [0.5, 1.5, 1]$ .

ments if they subscribe to services of different data rates. Additionally, for practical scenarios, user terminals may be placed at varying distances from the base station. This effect is represented by  $\mathbf{c} = [c_1, c_2, c_3]$ , where the variance of the channel matrix elements are scaled by  $c_1$ ,  $c_2$ , and  $c_3$  respectively for users 1, 2, and 3. Fig. 11 plots sum power versus  $\Delta\rho$ , where the target rate vector is  $\bar{\mathbf{R}} = [3, 3 - \Delta\rho, 3 + \Delta\rho]^T$  bps/Hz, and the channel strengths are  $\mathbf{c} = [0.5, 1.5, 1]$ . When  $\Delta\rho = 2$ , the gain of the proposed allocation is large, over 4 dB. This is because the fast adaptive subcarrier allocation is able to optimize the number and positions of subcarriers for each user. This time, the localized TX does not perform better than the fixed scheme because the large difference in channel strengths result in the users being selected in a fixed pattern. However, the ACG2 is still able to provide a low sum power because the number of subcarriers each user gets is decided by the users' target rates.

In Fig. 12, the channel strengths are given by  $\mathbf{c} = [1 - \Delta c, 1 + \Delta c, 1]$ , while the rate requirements are  $\bar{\mathbf{R}} = [3, 2, 4]^T$  bps/Hz. The transmit power is plotted against the variation in channel strength  $\Delta c$ . When  $\Delta c = 0.9$ , the gain over a fixed allocation is as large as 3 dB. Again, this is because under the optimal scheme, more subcarriers would be allocated to the user with the weaker channel in order to minimize the total transmit power, whereas the fixed allocation is not able to compensate for the different channel strengths. Fig. 13 examines the sum power as the number of subcarriers  $M$  increases. It can be seen that even with only 16 subcarriers, the duality gap is negligible. When  $M = 128$ , the duality gap becomes zero, and the proposed efficient algorithm is optimal.

In all these simulations, it can be seen that the efficient subcarrier allocation yields a large gain over a fixed subcarrier allocation. The gain tends to increase with a more frequency-selective channel or a greater number of users. The gains are largest for practical scenarios where there can be varied channel strengths or differentiated rate requirements. In general, the localized TX performs better than the fixed

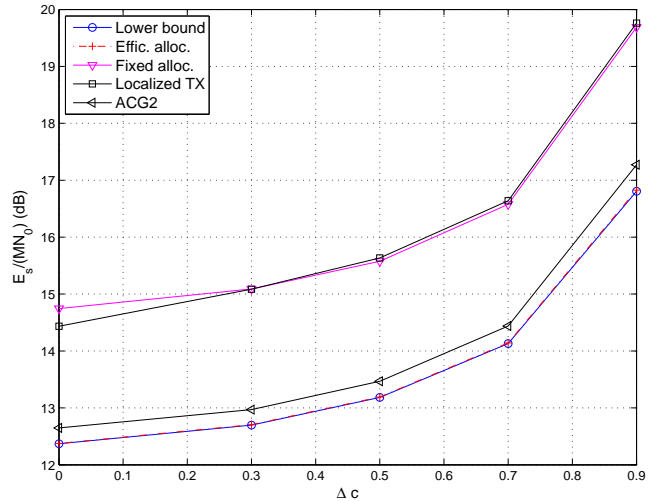


Fig. 12. Effect of different channel strengths among the users, where  $\mathbf{c} = [1 - \Delta c, 1 + \Delta c, 1]$ , with  $\bar{\mathbf{R}} = [3, 2, 4]^T$  bps/Hz.

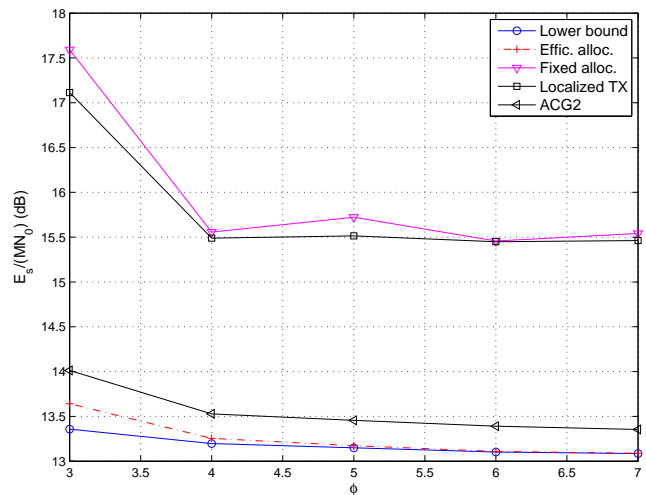


Fig. 13. Performance for different numbers of subcarriers  $M = 2^\phi$ , where number of taps  $= M/4 + 1$ ,  $\bar{\mathbf{R}} = [3, 2, 4]^T$  bps/Hz, and  $\mathbf{c} = [0.5, 1.5, 1]$ .

allocation, while the ACG2, in turn, performs better than the localized TX. Finally, the proposed efficient allocation consistently outperforms all the other schemes.

## VIII. CONCLUSION

High data rate communication is one of the key benefits of MIMO-OFDM. In order to utilize the system resources efficiently, fast and adaptive optimization algorithms are required. This paper has addressed the issue of optimal resource allocation to minimize the total transmit power while satisfying users' target rates. An efficient and adaptive algorithm, based on convex optimization theory, is proposed to obtain the subcarrier, power, and rate allocations that exploit the diversities of the system. To provide a low complexity implementation, only linear beamforming is carried out at the transmitter and the receiver. Therefore, this solution is immediately applicable to both the downlink and the uplink. Adaptation for this efficient resource allocation allows for fast

power convergence. When the duality gap for a particular channel realization is zero, this efficient solution coincides with the optimal minimum power solution, else this solution is near-optimal. To handle the event of a flat fading channel, a technique termed *dual proportional fairness* is employed to give good performance even in this scenario. Simulation results show a large performance improvement over a fixed subcarrier allocation.

#### ACKNOWLEDGMENT

For this work, the authors are grateful for the support provided by the Agency for Science, Technology and Research (A\*STAR) and the Institute for Infocomm Research (I<sup>2</sup>R), Singapore. The authors also appreciate the invaluable comments by the anonymous reviewers that helped to improve this manuscript. Moreover, the authors would like to thank Rui Zhang for rendering valuable advice.

#### REFERENCES

- [1] I. E. Telatar, "Capacity of Multi-antenna Gaussian Channels," *Bell Labs Technical Memorandum*, Jun. 1995.
- [2] G. J. Foschini and M. J. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," *Wireless Personal Commun.*, vol. 6, pp. 311-335, Mar. 1998.
- [3] K. B. Letaief and Y. J. Zhang, "Dynamic multiuser resource allocation and adaptation for wireless systems," *IEEE Wireless Commun.*, vol. 13, no. 4, pp. 38-47, Aug. 2006.
- [4] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [5] Z. Q. Luo and W. Yu, "An Introduction to Convex Optimization for Communications and Signal Processing," *IEEE J. Select. Areas Commun.*, vol. 24, no. 8, pp. 1426-1438, Aug. 2006.
- [6] P. Viswanath and D. Tse, "Sum Capacity of the Vector Gaussian Broadcast Channel and Uplink-Downlink Duality," *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1912-1921, Aug. 2003.
- [7] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658-2668, Oct. 2003.
- [8] W. Yu and J. Cioffi, "Sum Capacity of Gaussian Vector Broadcast Channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1875-1892, Sep. 2004.
- [9] G. Caire and S. Shamai (Shitz), "On the Achievable Throughput of a Multiantenna Gaussian Broadcast Channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691-1706, Jul. 2003.
- [10] C.-H. F. Fung, W. Yu, and T. J. Lim, "Multi-antenna downlink precoding with individual rate constraints: power minimization and user ordering," *Proc. Int. Conf. Commun. Systems*, pp. 45-49, Sep. 2004.
- [11] M. Mohseni, R. Zhang, and J. M. Cioffi, "Optimized Transmission for Fading Multiple-Access and Broadcast Channels with Multiple Antennas," *IEEE J. Select. Areas Commun.*, vol. 24, no. 8, pp. 1627-1639, Aug. 2006.
- [12] J. Lee and N. Jindal, "Symmetric Capacity of MIMO Downlink Channels," *IEEE Int. Symp. Inform. Theory*, pp. 1031-1035, Jul. 2006.
- [13] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218-1230, Oct. 1997.
- [14] Z. Hu, G. Zhu, Y. Xia, and G. Liu, "Multiuser subcarrier and bit allocation for MIMO-OFDM systems with perfect and partial channel information," *Proc. Wireless Commun. and Networking Conf.*, vol. 2, pp. 1188-1193, Mar. 2004.
- [15] D. Kivanc, G. Li, and H. Liu, "Computationally Efficient Bandwidth Allocation and Power Control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, pp. 1150-1158, Nov. 2003.
- [16] K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal Resource Allocation for OFDMA Downlink Systems," *Proc. Int. Symp. Inform. Theory*, pp. 1394-1398, Jul. 2006.
- [17] Y. J. Zhang and K. B. Letaief, "An Efficient Resource-Allocation Scheme for Spatial Multiuser Access in MIMO/OFDM Systems," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 107-116, Jan. 2005.
- [18] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-Forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 461-471, Feb. 2004.
- [19] T. Yoo and A. Goldsmith, "On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming," *IEEE J. Select. Areas Commun.*, vol. 24, no. 3, pp. 528-541, Mar. 2006.
- [20] W. Yu and R. Lui, "Dual Methods for Nonconvex Spectrum Optimization of Multicarrier Systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310-1322, Jul. 2006.
- [21] R. Freund, "15.084J / 6.252J Nonlinear Programming, Spring 2004," *MIT OpenCourseWare*.
- [22] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic Beamforming Using Dumb Antennas," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1277-1294, Jun. 2002.



**Winston W. L. Ho** (Student Member, IEEE, 2006) received the BEng (Hons) degree in electrical engineering from the National University of Singapore (NUS) in 2004. He is with the Institute for Infocomm Research (I<sup>2</sup>R), and is currently pursuing a PhD degree at the NUS, under a scholarship from the Agency for Science, Technology and Research (A\*STAR), Singapore. In 2003, during his half year industrial attachment, he worked at the Institute for Communications Research, now known as the I<sup>2</sup>R. His research interests include multiple antenna

systems, cooperative communications, multiuser systems, and communication theory.



**Ying-Chang Liang** (Senior Member, IEEE, 2000) received PhD degree in Electrical Engineering in 1993. He is now Senior Scientist in the Institute for Infocomm Research (I<sup>2</sup>R), Singapore, where he has been leading the research activities in the area of cognitive radio and cooperative communications and the standardization activities in IEEE 802.22 wireless regional networks (WRAN) for which his team has made fundamental contributions in physical layer, MAC layer and spectrum sensing solutions. He also holds adjunct associate professorship positions in Nanyang Technological University (NTU) and National University of Singapore (NUS), both in Singapore, and adjunct professorship position with University of Electronic Science & Technology of China (UESTC). He has been teaching graduate courses in NUS since 2004. From Dec 2002 to Dec 2003, Dr Liang was a visiting scholar with the Department of Electrical Engineering, Stanford University. His research interest includes cognitive radio, dynamic spectrum access, reconfigurable signal processing for broadband communications, space-time wireless communications, wireless networking, information theory and statistical signal processing.

Dr Liang is now an Associate Editor of IEEE Transactions on Vehicular Technology. He was an Associate Editor of IEEE Transactions on Wireless Communications from 2002 to 2005, Lead Guest-Editor of IEEE Journal on Selected Areas in Communications, Special Issue on Cognitive Radio: Theory and Applications, and Guest-Editor of COMPUTER NETWORKS Journal (Elsevier) Special Issue on Cognitive Wireless Networks. He received the Best Paper Awards from IEEE VTC-Fall'1999 and IEEE PIMRC'2005, and 2007 Institute of Engineers Singapore (IES) Prestigious Engineering Achievement Award. Dr Liang has served for various IEEE conferences as technical program committee (TPC) member. He was Publication Chair of 2001 IEEE Workshop on Statistical Signal Processing, TPC Co-Chair of 2006 IEEE International Conference on Communication Systems (ICCS'2006), Panel Co-Chair of 2008 IEEE Vehicular Technology Conference Spring (VTC'2008-Spring), TPC Co-Chair of 3rd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom'2008), Deputy Chair of 2008 IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN'2008), and Co-Chair, Thematic Program on Random matrix theory and its applications in statistics and wireless communications, Institute for Mathematical Sciences, National University of Singapore, 2006. Dr Liang is a Senior Member of IEEE. He holds six granted patents and more than 15 filed patents.