Project Number:	ftw. IST-026905
Project Title:	Multiple-Access Space-Time Coding Testbed
Project Coordinator:	C.F. Mecklenbräuker
Deliverable Number:	D1.1.1

Title of Deliverable:	Link layer adaptation and scheduling algorithms for MU-MIMO Systems
Workpackage:	WP-1
Nature:	R
Dissemination level:	PU
Editor:	M. Schubert
Authors:	A. Hottinen, M. Schubert, N. Vucic
Contractual Date of Deliverable:	Dec 31, 2006
Actual Date of Delivery:	Dec 23, 2006

Abstract:

We address the problem of multiuser MIMO scheduling over time and frequency domains. An additional optimization over the spatial domain is included by introducing a general model for multi-user interference coupling. We analyse the QoS feasible region and show under which conditions the boundary is achievable. It is shown for certain types of interference functions and coupling scenarios, that proportionally fair scheduling is a strictly convex optimization problem. This is exploited by a new algorithm, which combines utility optimization and adaptive transmitter optimization (e.g. beamforming). In addition, simple and computationally efficient integer programming algorithms are proposed for frequency domain scheduling in a MIMO channel.

Contents

Authors

Ari Hottinen
Nokia Research Center

tel.: +358 50 4836482
e-Mail: ari.hottinen@nokia.com

Martin Schubert
Fraunhofer Gesellschaft Heinrich Hertz Institut für Nachrichtentechnik

tel.: +49 30 31002 870
e-Mail: martin.schubert@hhi.fhg.de

Nikola Vucic
Fraunhofer Gesellschaft Heinrich Hertz Institut für Nachrichtentechnik

tel.: +49 30 31002 858
e-Mail: nikola.vucic@hhi.fhg.de

Chapter 1

Introduction

In a multiuser wireless system different users experience different and independent channel conditions. Depending on instantaneous (fast) fading realizations some users are able to access the channel more efficiently than others. When a new channel realization is observed, the relative channel quality among different users is likely to change. A channel-aware scheduler is able to utilize this phenomenon and realize a multiuser diversity gain which is in some cases analogous to a selection diversity gain in multi-antenna systems. Such scheduling is intimately related to the idea of multiuser diversity, as originally proposed in [35], where uplink temporal multiuser diversity was introduced from an information theory viewpoint.

Modern wireless systems, such as cdma2000 and WCDMA, already support channel-aware multiuser scheduling via channel quality feedback and related signalling capabilities. The current systems are designed essentially for single-input single-output (SISO) channels. However, in ongoing standardization in 3GPP Long-term evolution and in IEEE 802.16e (Mobile WiMAX) scheduling decisions in the time-frequency plane are required for users that potentially employ high-rate MIMO modulation.

Due to OFDM or multicarrier modulation, the upcoming wireless systems are likely to exploit frequency-domain scheduling, in addition to temporal scheduling. A frequency domain scheduler, as considered later in this document, needs to assign different users distinct subchannels or subcarriers. Moreover, due to multi-antenna assumption also spatial scheduling is relevant. This was considered e.g. in [24]. In general, it is sufficient that the channels that are assigned to different users are distinct in at least one of space,time, or frequency dimension. However, finding the appropriate assignment is a complex and computationally intensive optimization problem due to a large number of discrete parameters, and moreover, due to various user specific fairness and delay constraints.

1.1 Background

Consider a wireless communication system with $K > 1$ receivers (terminals, users) and a transmitter (e.g. base station), which is responsible for scheduling or assigning a distinct temporal transmission slot to the K users. In current wireless systems it is assumed, due to limited feedback capacity, that the transmitter does not necessarily know channel coefficients but only some derived information pertaining to the channel, such as channel quality or the value of some channel quality indicator. This is typical e.g. in systems using frequency-division-duplex (FDD), but in time-division-duplex systems also the transmitter has channel knowledge. Full channel state information (CSI) is assumed to be available at the receivers. In FDD downlink, the K terminals derive their respective CQIs based on downlink channel measurements, and signal them to the transmitter. The transmitter schedules transmissions to the users, targeting either to maximize throughput or minimize transmission power, with a quality of service constraint characterized e.g. by a maximum error rate.

If fairness is neglected, the scheduler should always pick the users that have the highest received signal power. This is exemplified in Fig. 1.1, which depicts the channel powers of four users during sixteen slots. It is observed that during the sixteen slots user 1 is scheduled three times, user 2 seven times, user 3 two times, and user 4 three times. Clearly, the solution is not fair, since channel access is dominated by user 2. On the other hand, user 1 is not allowed to access the channel between slots 3 and 15, thus imposing a thirteen slot delay to the service. This simple example demonstrates that an appropriate scheduling policy should consider a number of possibly conflicting quality-of-service criteria. Moreover, in practice the channel state information, used as an input to the scheduler, is quantized or imperfect, which deteriorates performance. As an example, in current 3G systems channel state information is quantized and signalled to the transmitter, so that minimal additional overhead is imposed to the system. A scheduling algorithm should be able to

- increase system throughput (sum capacity)
- support delay-throughput-differentiation (e.g. different delay and rate constraints)
- control fairness among users.

Fairness is considered as an important issue in scheduling. Fairness among users means that channel bandwidth is allocated in proportion of weights of

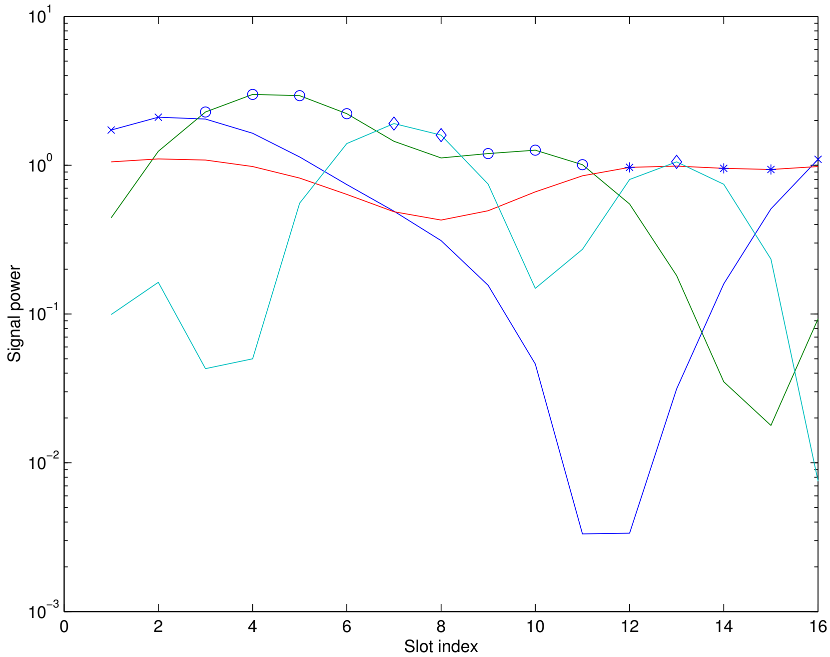

Figure 1.1: Example of multiuser diversity. The scheduling instants for users one to four are represented by symbols $\times$,$\circ$,$\diamond$, $\ast$, respectively.

the users. These weights can depend on the required data rates of different services such that high data rate users are given higher priority in bandwidth allocation. Moreover, user's weight may depend on the maximum delay of the service or the size of the data buffer.

In what follows, we briefly state some recently developed algorithms that attempt to solve the scheduling problem by taking into account one or several criteria given above.

1.1.1 Maximum Throughput (SNR) Scheduling

In Round robin (RR) scheduling a user is selected sequentially or randomly among K users and makes no use of channel state information. In contrast, maximum sum-rate or maximum throughput scheduler uses channel state information and transmits to the user with the best instantaneous supported data rate, using appropriate modulation and coding scheme. When the number of active users in the cell is large and the channels fluctuate rapidly, the maximum SNR scheduler can provide higher individual and total throughput than the round robin scheduler.

Formally, assume that the transmitter knows the maximum supported data rates $\{C_k\}$ of each user. In the maximum throughput scheduler the channel resources are given to user k^* with the rate

$$C_{k^*} = \max\{C_1, \cdots, C_K\}.$$

In a SISO channel this is analogous to selecting a user k^* with the best instantaneous received SNR. Due to the analogy, we refer to maximum throughput scheduler as maximum SNR scheduler to emphasize the connection to physical channel. In general, the SNR can also relate also post detection signal quality, as in related MIMO scheduling approaches in [27].

1.1.2 Delay-differentiated scheduling

Delay constraints make it necessary to introduce mechanisms for fairness. In this respect, the Max-SNR scheduler or a related proportionally fair scheduler can be generalized by using additional history-related parameters that account for temporal constraints. However, even if these parameters (e.g. shorter averaging window in proportionally fair schedulter) affects scheduling delays, such parameters are not directly related to any prescribed delay constraints. Delay-differentiated scheduling algorithms that explicitly account for delays are discussed in [25, 2]. The hybrid round-robin/proportional fairness solution in [2] is essentially supported by intuitive notions, where as [25] considers an optimization model that incorporates both throughput and delay constraints. Note, that the definition of proportional fairness used in [2] differs from the definition that will be used later in Chapter 2.

The multi-user delay differentiated scheduling problem can be formalized by considering a multi-access channel with T time slots and K users, where the scheduling decision variables are $x_k[t] \in \{0, 1\},\ k = 1, ..., K,\ t = 0, ..., T$ and the delay parameters are $\{b_k\}$. Using these parameters, a multiuser delay-penalized scheduling problem in a time window of size T can be stated as

$$\max \sum_{k=1}^{K} \sum_{t=1}^{T} E(b_k^t \phi_k[t]), \tag{1.1}$$

subject to

$$\sum_{k} x_k[t] \le 1\ , t = 1, ..., T \tag{1.2}$$

$$\sum_{t} x_k[t] \le R_k,\ k = 1, ..., K. \tag{1.3}$$

The above formulation allows for delay-differentiation with multiple service classes, each associated with its own delay-penalty parameter b_k. Note that in the special case where $T = K$, the data rates or utilities $\{\phi_k[t]\}$ are known for $T = 1, ..., K$, and where $R_k = 1,\ i = 1, .., K$, problem (1.1) reduces to the classical *assignment problem* in operations research [66]. We return to this integer program in connection with frequency scheduling later in this document.

1.2 MIMO Scheduling

Most research on scheduling was done for the conventional case of single antenna communication links. In this report we study the general MIMO case, where each user transmits over a multi-antenna MIMO link, as depicted in Fig. 1.2. As discussed before, the scheduling choice is made on the basis of channel qualities and additional fairness requirements, accounting for different priorities or queueing states.

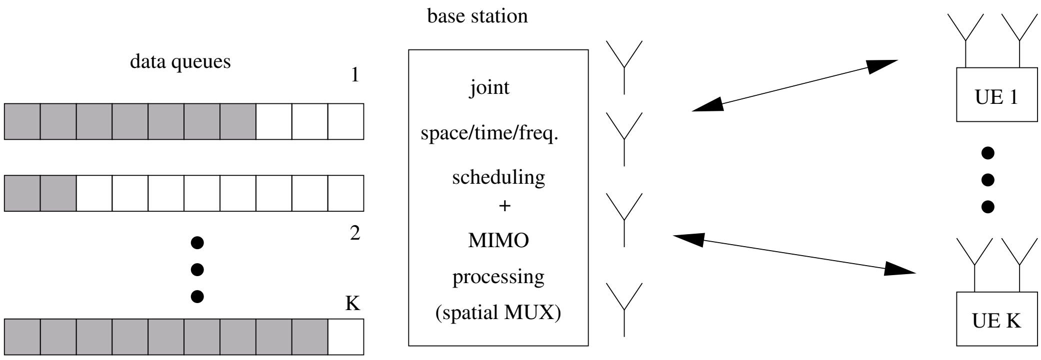

Figure 1.2: MIMO scheduling

However, preferred scheduling strategies can largely differ according to the chosen definition of optimality. With a single transmit antenna for each user and a single receive antenna at the base-station, Knopp and Humblet [35] showed that in order to maximise the sum ergodic capacity, at any given time instant, only the single user who has the best fading state should transmit. In this situation, a user needs to wait for his channel realization to become the best among all users before he could transmit.

For MIMO systems, the situation is more complicated. The presence of multiple antennas offers an additional degree of freedom. The optimum sum capacity is generally achieved by simultaneously transmitting multiple data streams at any given time instant. This is known from [57], and also follows from recent information-theoretical results on the sum-capacity of non-degraded broadcast and multiple access channels (e.g. [61, 69, 62, 64, 15, 70]).

The maximum sum-capacity provides an important benchmark for the overall performance of multiuser MIMO systems. However, by aiming at the sum capacity, it can happen that users with bad channel conditions remain inactive, since this strategy favours the strong links. Such a behaviour is not always desirable. In particular, possible latency constraints require a certain degree of *fairness*.

Fairness means that the scheduler can assign the available resource among the communication links in a flexible and efficient way. A good tradeoff should be found between the overall system efficiency and fairness. The actual strategy should depend on the current channel state, but also on other criteria, like the data queue length etc. The inclusion of all these aspects typically results in complex problem formulations with many degrees of freedom. This is one of the reasons why MIMO scheduling is only partly understood up to date.

Hence, the first year of the project was partly devoted to a fundamental analysis of the quality-of-service (QoS) region $\mathcal{Q}$ (the set of achievable QoS tuples). Here, QoS stand for a certain choice of performance measure, which will be specified later (examples are capacity, MMSE, delay, BER, etc.). The problem of scheduling can be regarded as the search for a "good" tradeoff point on the boundary of the QoS achievable region.

The focus in this part is on proportional fair scheduling schemes. But before we summarise our main results, we briefly review different notions of fairness used in the literature.

Chapter 2

General Framework for Proportional Fairness

The analysis of fairness issues in networks has its origin in the framework of wired networks (see [22, 34, 33, 14, 42, 53] for an overview on fairness related issues). Although we are free to define specialised notions of fairness for particular networks of interest, two fundamental fairness principles are established. These principles give rise to the majority of related fairness notions applicable to different network types (wired/ wireless), different network topologies (cellular/ ad-hoc networks) and different QoS parameters (e.g. the end-to-end delay in multi-hop ad-hoc networks or data-rate in cellular networks).

2.1 Fairness Principles

Two commonly used notions of fairness are discussed in this section. The first is the max-min principle. The second is utility optimization. The latter is a general approach, which includes other fairness definition, like proportional fairness as special cases.

2.1.1 Max-Min Fairness

One common fairness principle is max-min fairness. It consists of making the worst QoS parameter (of a route, link, etc.) as good as possible. In wired networks the max-min fair equilibrium of QoS values is the one, at which no QoS parameter q_i can be improved without degradation of any QoS parameter q_j, $j \neq i$, which is already inferior to q_i [34, 33, 14, 42, 38, 40, 53]. The same definition translates usually to the case of wireless multi-hop ad-hoc networks, when the QoS parameters are associated with routes (end-to-end QoS) [55, 45].

For the considered cellular wireless network model with minimum per-link service requirements $\boldsymbol{q} = [q_1, \ldots, q_K]^T$, we can formulate the min-max criterion in the form

$$\max_{(\mathrm{QoS}_1,\ldots,\mathrm{QoS}_K)\in\mathcal{Q}} \left(\min_{1\le k\le K} \frac{\mathrm{QoS}_k}{q_k} \right) \tag{2.1}$$

We refer to strategy (2.1) as weighted min-max fairness. This parallels the fairness definition in [45] with respect to end-to-end QoS. Pure min-max fairness neglects unequal per-link requirements and corresponds to the special case $\boldsymbol{q}^{\mathrm{req}} = c\mathbf{1}$, $\mathbf{1} := (1,\ldots,1)$, $c > 0$. In the behavioural and economic science such notion parallels ideal social fairness [59].

2.1.2 Weighted Utility Optimisation

Utility optimisation is another common approach for flexible bandwidth/rate sharing [34, 33, 14, 42, 40, 53]. It was also proposed in the context of wireless multi-hop and ad-hoc networks [45, 17, 37]. The single utilities are associated with links and correspond simply to link QoS parameters (see also [39, 18] and references therein).

In the discussed scheduling context, the scheduler chooses the links that maximise the weighted utility function

$$\min_{(\mathrm{QoS}_1,\ldots,\mathrm{QoS}_K)\in\mathcal{Q}} (\text{or max}) \left(\sum_k w_k \mathrm{QoS}_k \right) \tag{2.2}$$

for any given time slot. The degree of fairness is controlled by priority factors w_k being inversely proportional to the average QoS of preceeding time slots. We define $\boldsymbol{w} = [w_1, \ldots, w_K]^T$, with a normalisation $\sum_k w_k = \|\boldsymbol{w}\|_1 = 1$. Such a strategy can be considered as *proportional fairness*, although this notion is often reserved for special case of the generic problem (2.2), as will be discussed later.

Utility-based scheduling aims at the optimisation of some global performance measure (sometimes we also wish to minimize a cost function). In this respect it stands in opposition to the traditional power control approach, which aims at the most power-efficient achievement of minimum required QoS for each link. The latter approach is well-understood and extensively studied (see e.g. [23, 16, 31, 71, 68, 20, 56, 32, 26] and the references therein). The traffic type for which the utility-based scheduling is favourable is sometimes referred to as *elastic*, since no fixed per-link requirements have to be accounted for.

It is worth noting that there is a specific form of the utility optimisation problem, which is sometimes of special interest. This is the case when the weights in the utility are chosen as linear functions of buffer occupancies on the source nodes of the corresponding links (cellular case), or routes (multi-hop case), and the QoS parameters express the capacity of the corresponding links/ routes. It was shown originally in [54] (see also [43, 36, 7]), that the optimisation of such utility provides the largest stability region of the network. Hereby, the size of the stability region of the network can be seen, in broad terms, as a measure of robustness of the network with respect to arrival rates of bursty traffic on the physical layer [6].

For particular wired networks, min-max fairness and utility optimality of bandwidth sharing schemes were shown in [42, 40, 50] to be incompatible goals (see also the illustration in Fig. 2.1). However, such incompatibility

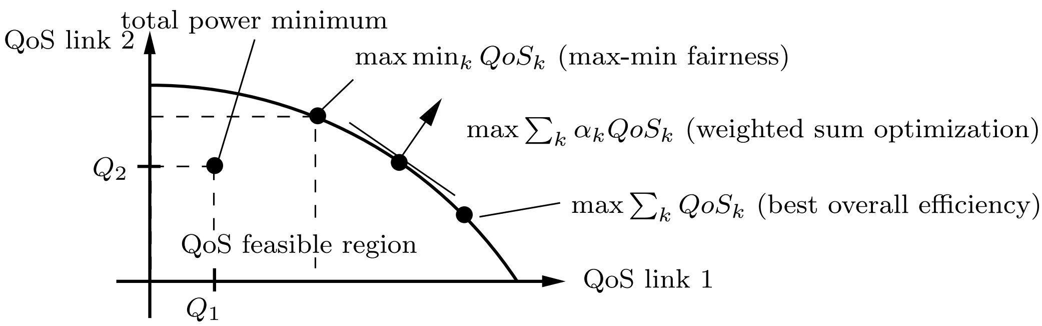

Figure 2.1: Trade-off between max-min fairness and weighted utility optimisation

is in general strongly topology-dependent. This follows from [53], where the corresponding conditions for compatibility/ incompatibility were stated and some examples of min-max fair and utility-optimal schemes were constructed. A kind of similar incompatibility was observed in [45] in the context of wireless multi-hop ad-hoc networks. To the best of our knowledge, the trade-off between min-max fairness and utility optimality has not been studied yet for cellular networks.

2.1.3 Proportional Fairness

Proportional fairness was introduced in [34]. This notion was established originally for wired networks (see also [41]), but is meanwhile well-understood also in the wireless context (see e.g. [63]). The proportional fair equilibrium $\boldsymbol{q}^{\text{pf}}$ of QoS parameters is the one, at which the difference to any other QoS vector $\boldsymbol{q}$ measured in the aggregated proportional change is nonnegative[1]. Precisely, with our model $\boldsymbol{q}^{\mathrm{pf}}$ is a proportional fair QoS vector if $\sum_{k=1}^{K} \frac{q_k - q_k^{\mathrm{pf}}}{q_k^{\mathrm{pf}}} \geq 0$, $\boldsymbol{q} \in \mathcal{Q}$. Interestingly, proportional fairness corresponds to the optimum of a specific utility function, with logarithmic QoS parameters [63, 4]. The motivation for the formulation of the proportional fairness principle was the observed significant utility-inefficiency (emphatic preferential treatment of small network flows [34, 33]) of a min-max fair allocation in wired networks.

2.2 Weighted Utility Optimisation

The objective (2.2) is generic and needs to be specified for practical purposes. Scheduling is typically based on channel information of all users in the system. An example is downlink transmit scheduling, with channel state information at the transmitter (CSIT). Examples are the HSDPA and 1xEV-DO standards. Thus, in the following we will assume a centralised scheduler with (partial) CSIT, so the impact of the transmission powers on the QoS is known and this knowledge can be exploited.

2.2.1 Interference Model

Exploiting channel knowledge, the region $\mathcal{Q}$ can be parametrised by the transmission powers, which are limited by a given sum-power constraint. The scheduling problem can be regarded as an allocation of the power resource $P_{\max}$ to the K communication links. The power allocation $\boldsymbol{p} = [p_1, \ldots, p_K]^T \in \mathbb{R}_+^K$, with $\|\boldsymbol{p}\|_1 \leq P_{\max}$, determines the individual QoS values. A power assignment $p_k = 0$ means that the kth link is inactive.

Using this parametrisation, the QoS of the kth link is

$$\mathrm{QoS}_k(\boldsymbol{p}) = \phi\Big(\frac{p_k}{\mathcal{I}_k(\boldsymbol{p})}\Big) \tag{2.3}$$

where $\mathcal{I}_k(\boldsymbol{p})$ is the interference of the kth link (possibly incorporating noise), and ϕ is a strictly monotonic bijective function, mapping the signal-to-interference-ratio (SIR) on the QoS. Some examples are BER: $\phi(x) = Q(\sqrt{x})$, MMSE: $\phi(x) = 1/(1+x)$, BER-slope for α-fold diversity: $\phi(x) = x^{-\alpha}$, or capacity: $\phi(x) = \log(1+x)$.

Our approach is based on a calculus for interference functions $\mathcal{I}(\boldsymbol{p})$, which will be specified later (see also [48]). But it is important to notice that $\mathcal{I}(\boldsymbol{p})$

[1]Clearly, in the case of QoS parameters increasing in service quality the non-negativity condition has to be replaced by the non-positivity condition.

does not only stand for the "interference power" as commonly assumed. This would not be an appropriate measure for MIMO system with multiple data streams per layer. However, the expression "interference" is understood in a more general sense here. For example, it is possible to define $\mathcal{I}$ in such a way that $p_k/\mathcal{I}_k(\boldsymbol{p})$ stands for the sum-MMSE [49, 3].

More detailed definitions of $\mathcal{I}$ will be discussed in the following. For the time being, it is enough to define $\mathcal{I}(\boldsymbol{p})$ by a few elementary common sense criteria. In particular, we call $\mathcal{I}$ an interference function if it fulfils the following axiomatic framework.

$$\begin{array}{ll} \text{A1} & \mathcal{I}(\boldsymbol{p}) \geq 0, \quad \boldsymbol{p} \in \mathbb{R}_+^K \\ \text{A2} & \mathcal{I}(\alpha \boldsymbol{p}) = \alpha \mathcal{I}(\boldsymbol{p}) \quad \alpha \in \mathbb{R}_+ \\ \text{A3} & \mathcal{I}(\boldsymbol{p}^{(1)}) \geq \mathcal{I}(\boldsymbol{p}^{(2)}) \text{ if } \boldsymbol{p}^{(1)} \geq \boldsymbol{p}^{(2)} \end{array}$$

These axioms form the basis of our theoretical framework. Note, that this approach generalises the "standard interference function" introduced by Yates [68], where the presence of a constant noise power was required. The model [68] is included in A1-A3 as a special case. In particular, one component of $\boldsymbol{p}$ can stand for noise, as discussed in [48].

With this model, our scheduling problem becomes

$$\min_{\boldsymbol{p} \in \mathbb{R}_+^K} (\text{or max}) \Big(\sum_k w_k \cdot \text{QoS}_k(\boldsymbol{p}) \Big) \quad \text{s.t.} \quad \|\boldsymbol{p}\|_1 \leq P_{\max} \tag{2.4}$$

2.2.2 Proportionally Fair Power Allocation

Proportional fairness corresponds to the optimum of a specific utility function, with logarithmic QoS parameters [34, 63, 4]. We focus on the special case where $\text{QoS}_k(\boldsymbol{p})$ is inversely proportional to the SIR, i.e., $\text{QoS}_k(\boldsymbol{p}) = \log(\mathcal{I}_k(\boldsymbol{p})/p_k)$. In this case, problem (2.4) can be rewritten as

$$F(\boldsymbol{w}) = \min_{\boldsymbol{s} \in \mathbb{R}^K} \sum_{k=1}^{K} w_k \cdot \log \frac{\mathcal{I}_k(\mathrm{e}^{\boldsymbol{s}})}{\mathrm{e}^{s_k}} \tag{2.5}$$

Following earlier work [52, 16, 31, 12, 51, 48] we use the substitution $\boldsymbol{p} = \mathrm{e}^{\boldsymbol{s}}$ for the power allocation (see e.g. [16]. Note, that the objective in (2.5) is inversely proportional to the SIR (like BER, ...), hence we call $F(\boldsymbol{w})$ a cost function.

Note, that the interference function $\mathcal{I}_k$ does not need to be linear. A non-linear model including beamforming was studied in [8]. It was shown (see e.g. [11, 5] and the references therein), that the problem (2.5) is convex if the interference functions are log-convex. Log-convex interference functions occur in various contexts. Examples are:

- The linear interference function

$$\mathcal{I}_k(\mathrm{e}^{\boldsymbol{s}}) = [\boldsymbol{V}\mathrm{e}^{\boldsymbol{s}}]_k, \quad k = 1, 2, \ldots, K \tag{2.6}$$

  is log-convex. Here, the interference coupling is characterised by an irreducible matrix $\boldsymbol{V} \geq 0$ [73].

- The worst-case interference function

$$\mathcal{I}_k(\mathrm{e}^{\boldsymbol{s}}) = \max_{z_k \in \mathcal{Z}_k} [\boldsymbol{V}(z)\mathrm{e}^{\boldsymbol{s}}]_k, \quad k = 1, 2, \ldots, K \tag{2.7}$$

  is log-convex. The function (2.7) can be used for robust power allocation. The parameter z_k stands for an uncertainty chosen from an uncertainty region $\mathcal{Z}_k$.

2.3 Basic Properties

We start by analysing properties of the underlying QoS region $\mathcal{Q}$. A thorough understanding of the quality-of-service (QoS) region is at the core of multiuser scheduling. It will turn out that under certain conditions the proportional fair scheduling problem is convex, and sometimes even strictly convex. Convexity is a desirable property, which allows for efficient algorithmic solutions.

2.3.1 The QoS Region

Let γ be the inverse function of ϕ, as introduced by (2.3). Then, $\gamma_k = \gamma(q_k)$ is the minimum SIR level needed to achieve a QoS target q_k. Thus, the QoS region $\mathcal{Q}$ is a one-to-one mapping of the SIR region

$$\mathcal{S} = \{\boldsymbol{\gamma} : C(\boldsymbol{\gamma}) \leq 1\} \tag{2.8}$$

The SIR region $\mathcal{S}$ is a sub-level set of the "indicator function"

$$C(\boldsymbol{\gamma}) = \inf_{\boldsymbol{p}>0} \left(\max_{1\leq k\leq K} \frac{\gamma_k(q_k) \cdot \mathcal{I}_k(\boldsymbol{p})}{p_k} \right) . \tag{2.9}$$

The min-max optimum $C(\boldsymbol{\gamma})$ provides a single measure for the quality of a multiuser system. It completely characterises the SIR region $\mathcal{S}$, and hence the QoS region $\mathcal{Q} = \phi(\mathcal{S})$. If $C(\boldsymbol{\gamma}) < 1$, then it can be observed from (2.9) that there exists a power allocation $\boldsymbol{p} > 0$ such that the SIR vector $\boldsymbol{\gamma} > 0$ can be supported.

The proportionally fair scheduler (2.4) finds an optimal trade-off point on the boundary of $\mathcal{Q}$. This is illustrated in Fig 2.2). For this example, we have assumed that QoS stands for a "cost" (e.g. BER), which we want to minimise. In this case the max in (2.4) is replaced by min, and the weighted utility maximisation problem becomes a weighted cost minimisation problem.

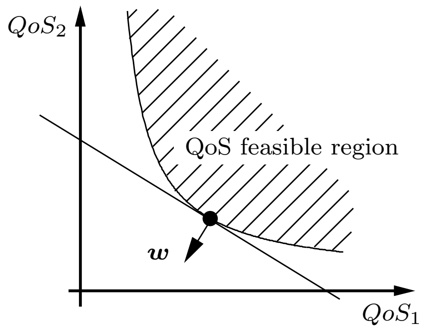

Figure 2.2: Weighted cost optimisation over the boundary of the QoS region

The structure of the region $\mathcal{S}$ (resp. $\mathcal{Q}$) determines the scheduling strategy. In order to explain this, consider the region depicted in Fig. 2.3. Different strategies result from different choices of the weights $\boldsymbol{w}$. We have three possible operating points A,B, and C. The points A and B correspond to policy which serves a single link at any given time. Whereas strategy C means that two users are scheduled simultaneously on the same resource. This corresponds to space division multiple access (SDMA).

This discussion shows that the geometry of the QoS region has an important impact on the chosen scheduling strategy. A better understanding of the QoS region is important for the development of scheduling algorithms.

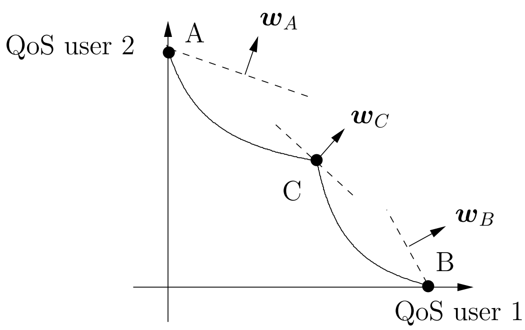

Figure 2.3: Example of a supportable QoS region, 2-user case.

2.3.2 Supportability of the QoS Region

Every QoS value q_k is associated with an SIR value $\gamma_k(q_k)$. Given the interference model A1-A4, the resulting QoS region is

$$\mathcal{Q} = \{\boldsymbol{q} : C\big(\boldsymbol{\gamma}(\boldsymbol{q})\big) \leq 1\} \tag{2.10}$$

For the linear interference model (2.6), the min-max optimum $C(\boldsymbol{\gamma})$ equals the Perron root of the weighted coupling matrix $(\text{diag}\{\boldsymbol{\gamma}\}\boldsymbol{V})$ (see e.g. [48]). Thus, the function $C(\boldsymbol{\gamma})$ generalises known concepts from power control theory [73]. It fully characterizes the SIR region, thus it plays a fundamental role for our analysis.

Our first result is to characterise the supportability of QoS $\boldsymbol{q} \in \mathcal{Q}$. We say that $\boldsymbol{q} > 0$ is supportable if there exists a power allocation $\boldsymbol{p} > 0$ such that $\text{QoS}_k(\boldsymbol{p}) = q_k$, $\forall k$. Or in other words: the infimum in (2.9) is achieved. This property is important, e.g., in order to ensure numerical stability of scheduling algorithms, which operate on the boundary of the region.

This effect was analysed in [5] for the linear model (2.6). It was shown that the boundary is supportable if the coupling matrix $\boldsymbol{V} \geq 0$ is irreducible [21]. However, such a characterisation is not possible if the interference is characterized by the axiomatic model A1-A4. In this case, a different way of characterizing interference coupling is required.

It was shown in [10] that the supportability of the boundary only depends on the combinatorial structure of a newly introduced "dependency matrix".

$$[\boldsymbol{D}_{\mathcal{I}}]_{kl} = \begin{cases} 1 & \text{if } \mathcal{I}_k \text{ depends on } p_l\text{, i.e., there} \\ & \text{exists a } \boldsymbol{p} > 0 \text{ such that} \\ & \mathcal{I}_k(\boldsymbol{p} + \delta \boldsymbol{e}_l) \text{ is not constant for} \\ & \text{some } \delta > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\boldsymbol{e}^{(l)}$ is the all-zero vector with the l-th component set to one. The matrix $\boldsymbol{D}_{\mathcal{I}}$ generalizes the coupling matrix $\boldsymbol{V}$, known from classical power control theory [73].

The following proposition shows that supportability depends on he structure of the dependency matrix $\boldsymbol{D}_{\mathcal{I}}$.

Proposition 1. *Let $\boldsymbol{D}_{\mathcal{I}}$ be irreducible, then there exists a $\boldsymbol{p} > 0$ such that*

$$C(\boldsymbol{\gamma})p_k = \gamma_k \mathcal{I}_k(\boldsymbol{p}) \quad \textit{for all } \gamma > 0. \tag{2.11}$$

This result will be used for the analysis of the proportional fair scheduler.

2.3.3 Analysis of Proportional Fairness

In order to understand the proportionally fair resource allocation problem (2.5), it is necessary to understand the structure of log-convex interference functions. Not only are our underlying interference functions $\mathcal{I}(\boldsymbol{p})$ log-convex, but also the function $C(\boldsymbol{\gamma})$ is log-convex. In addition, it fulfils all the properties A1-A3, thus $C(\boldsymbol{\gamma})$ itself is an "interference function". This observation is fundamental, since $C(\boldsymbol{\gamma})$ characterises the QoS region. By studying properties of log-convex interference functions, new insight on the QoS region is obtained.

We start by defining the set

$$\mathcal{L}(\mathcal{I}) = \left\{ \boldsymbol{w} \in \mathbb{R}^K_+ : f_\mathcal{I}(\boldsymbol{w}) > 0 \right\} . \tag{2.12}$$

where

$$f_\mathcal{I}(\boldsymbol{w}) = \inf_{\boldsymbol{p}>0} \frac{\mathcal{I}(\boldsymbol{p})}{\prod_{l=1}^K (p_l)^{w_l}} , \quad \boldsymbol{w} \in \mathbb{R}^K_+ . \tag{2.13}$$

The following result is fundamental for our analysis.

Proposition 2. *Every log-convex interference function $\mathcal{I}(\boldsymbol{p})$, characterised by A1-A4, with $\boldsymbol{p} > 0$, can be represented as*

$$\mathcal{I}(\boldsymbol{p}) = \max_{\boldsymbol{w}\in\mathcal{L}(\mathcal{I})} \left(f_\mathcal{I}(\boldsymbol{w}) \cdot \prod_{l=1}^K (p_l)^{w_l} \right) \tag{2.14}$$

This result shows that every log-convex interference function can be represented as a maximum over elementary functions, which itself are log-convex. These basic building blocks are used for further analysing proportional fairness. The existence of an optimal power allocation is characterized by the following result.

Proposition 3. *Let $\boldsymbol{D}_\mathcal{I}$ be irreducible, and $F(\boldsymbol{w}, \boldsymbol{\mathcal{I}}) = \inf_{\boldsymbol{q}\in\mathcal{Q}} \sum_k w_k q_k$. There exists a $\hat{\boldsymbol{q}} \in \mathcal{Q}$ such that*

$$F(\boldsymbol{w}, \boldsymbol{\mathcal{I}}) = \sum_k w_k \hat{q}_k$$

if and only if there exists a $\hat{\mathbf{p}}$ such that

$$F(\boldsymbol{w}, \boldsymbol{\mathcal{I}}) = \sum_k w_k \log \frac{\mathcal{I}_k(\hat{\mathbf{p}})}{\hat{p}_k} \tag{2.15}$$

2.3.4 Strict Convexity

In [5] the strict convexity of the function $f(\mathrm{e}^{\boldsymbol{s}}) = \sum_k w_k \log([\boldsymbol{V}\boldsymbol{p}]_k/p_k)$ was characterised. If $\boldsymbol{V}$ is irreducible, then $f(\boldsymbol{w})$ is strictly convex if and only if $\boldsymbol{V}\boldsymbol{V}^T$ is irreducible.

This result can be generalised. To this end, we need the following definition

Definition 1. Assume that for two arbitrary $\boldsymbol{p}^{(1)}$, $\boldsymbol{p}^{(2)}$, there exists at least one index l from the dependency set (indices for which $\boldsymbol{D}_{\mathcal{I}}$ is non-zero) such that $p_l^{(1)} \neq p_l^{(2)}$. An interference function $\mathcal{I}$ is called *separating* if

$$\mathcal{I}\big(\boldsymbol{p}(\lambda)\big) < \mathcal{I}\big(\boldsymbol{p}^{(1)}\big)^{1-\lambda} \cdot \mathcal{I}\big(\boldsymbol{p}^{(1)}\big)^{\lambda} \tag{2.16}$$

That is, $\mathcal{I}(\boldsymbol{p})$ is strictly convex on its domain.

Proposition 4. *Let $\mathcal{I}_1, \ldots, \mathcal{I}_K$ be separating interference functions. The matrix $\boldsymbol{D}_{\mathcal{I}}$ is assumed to be irreducible. Then,*

$$f(\boldsymbol{s}) = \sum_k w_k \log \frac{\mathcal{I}_k(\mathrm{e}^{\boldsymbol{s}})}{\mathrm{e}^{s_k}} \tag{2.17}$$

is strictly convex if and only if $\boldsymbol{D}_{\mathcal{I}}\boldsymbol{D}_{\mathcal{I}}^T$ is irreducible.

If the assumptions in Proposition 4 are fulfilled, then strict convexity of the proportional fair cost function (2.17) only depends on the combinatorial structure of $\boldsymbol{D}_{\mathcal{I}}$. Consider, e.g., the linear model with a coupling matrix $\boldsymbol{V}$, then only the positions of the non-zero entries matters. The actual value of these entries does not affect the strict convexity of functional. The proposition shows that this property holds for *arbitrary* log-convex interference functions.

Future work will show how strict convexity can be exploited for the development of fast iterative algorithms for resource allocation. The strictness shown in this paper implies the existence of iterations with super-linear convergence.

2.4 Algorithm for Proportional Fairness with Adaptive Transmission

The optimal scheduling policy depends on the quality of the effective channels, which are determined by the chosen MIMO transceivers. On the other hand, the transceiver optimisation depends on the chosen subset of communication links to be scheduled. Both parameters should be optimised jointly.

Assuming that $\mathcal{I}(\boldsymbol{p})$ is a convex or log-convex interference function, we can exploit the strict convexity of proportional fairness shown in the previous section (see also [10, 9]) in order to derive a new scheduling algorithm, which iteratively updates the transmit strategies and the power allocation. The design goal is proportional fairness (2.5), for given weights $\boldsymbol{w} = [w_1, \dots, w_K]^T$, which depend, e.g. on the current queue lengths and the average service rates.

A *transmit strategy* z is used at the base station. We have $z = \{z_1, \dots, z_K\}$, where z_k stands for the transmit strategy employed by the kth link. The strategy z_k is chosen from a compact set $\mathcal{Z}_k$, which contains all possible receive strategies of the kth user. It is assumed that the kth column of $\boldsymbol{\Psi}(z)$ only depends on z_k. This is typical, e.g. if z_k stands for a beamforming vector which is used for linear pre-equalization of the channel. Arbitrary given receivers are assumed. Then, the interference coupling can be characterized by a non-negative coefficient matrix $\boldsymbol{V}(z)$. The product $\boldsymbol{V}(z)\boldsymbol{p}$ is a vector which contains the interference values of all K links.

It is known from matrix theory, that $\rho\big(\text{diag}\{\boldsymbol{\gamma}\}\boldsymbol{V}(z)\big) = \rho\big(\text{diag}\{\boldsymbol{\gamma}\}\boldsymbol{V}^T(z)\big)$ for an arbitrary choice of z. As a consequence, the strategy z which provides the largest feasible region for the transmit-oriented scenario (independent columns), can equivalently be obtained by considering transpose coupling matrix $\boldsymbol{V}^T(z)$ instead. Both channels have the same SIR feasible region. This relationship was already observed in the context of power control [72], and was later used in the context of max-min-SIR downlink beamforming [47], where it was referred to as *duality*.

By exploiting this duality, we can find the optimum transmit strategy for any given power allocation. This is exploited by the following algorithm, which updates powers and transmit strategies in an alternating manner.

- for fixed z, allocate the powers $\boldsymbol{p}$ by optimising the proportional fairness criterion (2.5). This is a convex optimisation problem and can be solved by standard optimisation techniques.

- for fixed $\boldsymbol{p}$, optimise the transmit strategy $z \in \mathcal{Z}$ by individually minimising the interference terms resulting from a dual problem formulation.

The algorithm is guaranteed to converge to a (possibly local) optimum.

2.5 Conclusion

A framework for characterizing performance tradeoffs in a multiuser system was proposed in this chapter. This analysis provides a basis for a better understanding of struture of the QoS region and the closely related problem of proportional fairness. By proportinal fairness we mean the optimization of a weighted utility function, which serves as a criterion for scheduling. In the MIMO context, throughput-wise optimal transmission requires spatial multiplexing with residual interference [58]. This complicates the optimization the utility optimization problem. In this situation, it is useful to revert to a general and abstract model for characterizing the interference coupling.

We have characterized the QoS region, which is convex for certain types of log-convex interference functions. Convexity is a desirable property which allows for efficient algorithms. Since the scheduler operates on the boundary of the region, an important prerequisite is that there exists a power allocation such that boundary points can be actually achieved (and not just in an asymptotic sense). This has been fully characterized in this chapter. To this end, we have introduced the concept of a “dependency matrix”, which describes the interference coupling in the system. This concept can be seen as an extension of the coupling matrix used in power control theory.

Future work within MASCOT should focus on the application of the framework to MIMO related scheduling problems. One example was already given in Section 2.4. This should be extended, by considering additional system aspects and other transceiver designs developed in other workpackages.

Chapter 3

Frequency-Domain Scheduling in MIMO systems

In multiuser MIMO systems the previously discussed temporal scheduling solutions can still be used, possibly by replacing the channel SNR by an effective post-detection SNR matched to the given MIMO modulator [27]. However, MIMO modulation or spatial multiplexing adds another dimension to the scheduling problem. Indeed, users' signal streams may be transmitted using spatial multiplexing by assigning them to different transmit antennas [24]. While such a solution may be a natural for spatial multiplexing, more efficient high-rate matrix-modulation methods distribute the transmitted symbols inherently over all antennas. In this case, the spatial dimension is typically reserved for rate increase and users are again multiplexed using time-frequency slots.

In what follows, due to the recent adoption of OFDM/OFDMA in IEEE 802.16e and IEEE 802.11n [1], we focus on frequency-domain scheduling. For simplicity, we only consider frequency scheduling in systems where there is a single transmission mode and rate available. For OFDMA, scheduling is essentially analogous to subcarrier allocation. In the MIMO case, this is done by taking into account the structure of matrix modulation and channel awareness at transmitter. While the use of channel state information in assigning users to appropriate subcarriers or subchannels was considered e.g. in [19, 67, 46, 44, 13] in connection with SISO, the MIMO problem requires some additional work [29, 28].

In the following sections, we state the optimization model and sketch a concise signal model for a multi-user MIMO-OFDM system. The signal model is well-known and given below only to make the connection to resource allocation apparent. A connection between MIMO modulation methods and Channel Quality Indicators (CQI) is defined and they are used to make a joint subcarrier assignment problem for all users in a channel-aware multiuser OFDMA system.

3.1 Optimization Model

Assume we have K users that need to assigned to P orthogonal channel resources. For example, consider a typical multiuser downlink channel where the K users have distinct physical channels and the P channel resources correspond to P orthogonal subcarriers. Each user is assigned a different subcarrier. The orthogonality constraint is imposed in order to reduce the computational complexity, essentially to avoid the combinatorial explosion if an arbitrary subset of interfering users were to be selected per subcarrier. On the other hand, within a subcarrier there may be multiple interfering symbol streams, due to MIMO modulation.

One way to find an optimal pairing among the K users and P subcarriers is to pose the problem as the classical *assignment problem* [66]. We define a performance indicator (a.k.a Channel Quality Indicator (CQI)) $\phi_{k,p}$ for resource pair (k, p). For example, if the aggregate SNR over all users is to be maximized we let $\phi_{k,p} \doteq \mathrm{SNR[k, p]}$. As another example, if aggregate capacity maximization is of interest, we may set $\phi_{k,p} \doteq \log_2(1 + \mathrm{SNR}[k, p])$. The performance indicator or the SNR needs to be defined so that it matches with the selected MIMO modulation method. We return to this issue after posing the optimization model for generic performance indicators.

The performance indicators, regardless of how they are defined, are captured in matrix $\mathbf{\Phi} = [\phi_{k,p}]$. The assignment problem is posed as

$$\max \sum_k \sum_p \phi_{k,p} t_{k,p} \tag{3.1}$$

subject to

$$\sum_p t_{k,p} = 1, \forall k \tag{3.2}$$

$$\sum_k t_{k,p} = 1, \forall p, \tag{3.3}$$

$$t_{k,p} \geq 0, \forall k, p \tag{3.4}$$

Although decision variables in equation (3.4) are continuous, the optimal solution is known to be integral, where $t_{k,p} \in \{0, 1\}$. Variable $t_{k,p} = 1$ if pair (k, p) is assigned and $t_{k,p} = 0$ otherwise. The constraints (3.2)-(3.3) formalize the requirement that each input subchannel is assigned to exactly one output subchannel or user. Obviously, $\mathbf{T} = [t_{k,p}]$ is indeed a permutation matrix. The complexity of the classical primal-dual assignment algorithm for problem (3.1)-(3.4) is $O(n^4)$, where n is the dimensionality of the assignment problem [66].

Remark 1: Due to the constraints (3.2)-(3.3) all users are assigned an a priori defined number of subchannels. This constitutes constitutes means for controlling fairness among the K users. In practice, different users can be allocated a different number of subcarriers, in which case the model is generalized to a *transportation problem* [66], that holds the assignment problem as a special case. Fairness, defined in terms the number of assigned subcarriers, leads to a computationally efficient and well-known algorithmic solution. However, as a drawback, it is only indirectly related to performance, capacity or throughput (in terms of bits per channel use). Obviously, when viewed via a different criterion, the individual subchannels remain different.

Remark 2: Although the computational burden of the algorithms remains polynomial, it is still high if the number of subcarriers is high (e.g. P=2048 is some broadband systems). In such case, it may be possible to reduce the dimensionality of the problem, by local averaging. A viable lower-dimensional approximate model can be obtained by utilizing correlations between different elements of the indicator matrix $\mathbf{\Phi}$, in analogy with [28]. Using local averaging, we replace the original cost matrix with an approximate cost matrix. The approximate cost matrix may be computed as a (weighted) average the values of the utilities of c neighboring subcarriers. Algorithmically, this is implemented by defining a matrix

$$\mathbf{U} = \mathbf{I}_{P/c} \otimes \mathbf{1}_c \tag{3.5}$$

forming a reduced dimensional model

$$\bar{\mathbf{\Phi}} \leftarrow \mathbf{U}^T \mathbf{\Phi} \mathbf{U} \tag{3.6}$$

and using this as an input to the assignment problem. If the c neighboring the values of the cost matrix are similar the performance loss is marginal. In practice the number c has to be carefully selected and matched to reflect the frequency-selectivity of the channel.

3.2 Channel Assignment in MIMO-OFDM

In this section, we apply the optimization model in previous section to a multi-antenna OFDM system. Let $\mathbf{F}$ denote a $P \times P$ inverse DFT (IDFT) matrix, where $[\mathbf{F}]_{p,q} = 1/\sqrt{P}\exp(j2\pi(p-1)(q-1)/P)$. The DFT matrix, applied at the OFDM receiver is given by $\mathbf{F}^\dagger$, the transpose conjugate of $\mathbf{F}$. We assume that the signal is transmitted through a finite impulse response (FIR) channel of length L and that a cyclic prefix of length $L_c > L$ is used at the transmitter. Matrix $\mathbf{H}$ denotes a circulant convolution matrix with entries $[\mathbf{H}]_{p,q} = h((p-q) \mod P)$, where $h(l)$ designates the lth channel tap of a given transmit-receive antenna pair. Vector $\mathbf{x}$ represents the symbol vector and $\mathbf{n}$ complex iid gaussian noise.

In what follows, we define the performance indicator matrix for a particular idealized MIMO-OFDM channel using the notations above. In a multi-antenna context orthogonal modulation methods are available only for a limited set of antenna configurations and only when transmitting at most one symbol per channel use. As the symbol rate is increased by using $N_\mathrm{t} > 1$ transmit and $N_\mathrm{r} > 1$ receive antennas the symbols generally interfere with each other. For example, a conventional MIMO-OFDM system applies vector modulation by transmitting simultaneously N_t symbol vectors, with $\mathbf{F}\mathbf{x}_n$ transmitted from antenna n, where $\mathbf{x}_n$ is P dimensional symbol vector. With N_r receiver antennas the received baseband frequency-domain signal is of form

$$\mathbf{y} = \begin{bmatrix} \mathbf{D}_{1,1} & \cdots & \mathbf{D}_{1,N_\mathrm{t}} \\ \vdots & & \vdots \\ \mathbf{D}_{N_\mathrm{r},1} & \cdots & \mathbf{D}_{N_\mathrm{r},N_\mathrm{t}} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{N_\mathrm{t}} \end{bmatrix} + \mathbf{n} \tag{3.7}$$

where

$$\mathbf{D}_{m,n} = \mathrm{diag}(H_{m,n}(0), ..., H_{m,n}(P-1)),$$

with

$$H_{m,n}(p) = \sum_{l=0}^{L} h_{m,n}(l) \exp(-\mathrm{j}2\pi lp/P).$$

The model may be converted with appropriate permutations into a block diagonal form, where each block contains the symbols received by subcarrier p. Then, the received signal model for the pth $N_\mathrm{t} \times N_\mathrm{r}$ block is

$$\mathbf{y}[p] = \mathbf{E}[p]\mathbf{x}[p] + \mathbf{n}[p] \tag{3.8}$$

where $\mathbf{E}[p]$ is a (non-orthogonal) $N_\mathrm{t} \times N_\mathrm{r}$ MIMO channel matrix, as perceived at the output of pth frequency bin at receiver.

3.3 Channel Quality Indicator

The model above is covers conventional vector modulation (a.k.a BLAST) [65]. For many high-rate high-diversity MIMO-OFDM systems, the vector $\mathbf{x}$ is replaced by an $N_\mathrm{t} \times \mathrm{T}$ modulation matrix $\mathbf{X}$, where T is the block length. With

Q input symbols the symbol rate is Q/T. Thus, the signal model is

$$\mathbf{Y} = \mathbf{E}\mathbf{X} + \mathbf{N}, \tag{3.9}$$

where matrix $\mathbf{N}$ contains noise terms for T channel uses. For the purposes of decoding, or for defining channel quality indicators, it is convenient to vectorize the model, by explicitly taking into account the structure of the MIMO modulation matrix $\mathbf{X}$. In this subsection we omit the subcarrier index to simplify notations. Namely, for the frequency-domain MIMO channel we write $\mathbf{E} \doteq \mathbf{E}[p]$ and similarly for the other symbols.

The methodology for deriving the equivalent channel model is given in [30] using the basis matrices of matrix modulation. Rather than repeating the procedure here, we only an example and write explicitly the vectorized model for Double ABBA (DABBA) modulator [30], one of MIMO modulators proposed in UMTS Long-Term Evolution. It embeds $\mathbf{X}_A, \mathbf{X}_B, \mathbf{X}_C$ and $\mathbf{X}_D$ as four 2×2 STTD (Alamouti) blocks, encoding the symbol pairs (x_1, x_2), (x_3, x_4), (x_5, x_6) and (x_7, x_8), respectively. Using these matrices, the DABBA symbol rate two modulator $(Q = 8, T = 4)$ is written as

$$\mathbf{X}_{DABBA} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{X}_A + \mathbf{X}_C & \mathbf{X}_B + \mathbf{X}_D \\ \mathbf{X}_B - \mathbf{X}_D & \mathbf{X}_A - \mathbf{X}_C \end{bmatrix} \tag{3.10}$$

The equivalent model comprising vector symbol $\mathbf{x}$ [30] includes an equivalent channel correlation matrix

$$\mathbf{R}_{eq} = \mathbf{E}_{eq}^{\dagger} \mathbf{E}_{eq}$$

that arises when rewriting model (3.9) as

$$\mathbf{y} = \mathbf{E}_{eq}\mathbf{x} + \mathbf{n}. \tag{3.11}$$

In this particular example, the correlation matrix reads

$$\mathbf{R}_{eq} = \begin{bmatrix} P_1 + P_2 & S_1 & P_1 - P_2 & S_2^{\dagger} \\ S_1 & P_1 + P_2 & S_2 & P_2 - P_1 \\ P_1 - P_2 & S_2^{\dagger} & P_1 + P_2 & S_1 \\ S_2 & P_2 - P_1 & S_1 & P_1 + P_2 \end{bmatrix} \tag{3.12}$$

where

$$P_1 = (\sum_{j=1}^{N_\mathrm{r}} |e_{j,1}|^2 + |e_{j,2}|^2)\mathbf{I}_2, \tag{3.13}$$

$$P_2 = (\sum_{j=1}^{N_\mathrm{r}} |e_{j,3}|^2 + |e_{j,4}|^2)\mathbf{I}_2, \tag{3.14}$$

where $S_1 = S + S^\dagger$,$S_2 = S - S^\dagger$, with

$$S = \begin{bmatrix} \alpha & -\beta^* \\ \beta & \alpha^* \end{bmatrix}, \tag{3.15}$$

where $\alpha = \sum_{j=1}^{N_\mathrm{r}} e_{j,3}^* e_{j,1} + e_{j,2}^* e_{j,4}$ and $\beta = \sum_{j=1}^{N_\mathrm{r}} e_{j,4}^* e_{j,1} + e_{j,2}^* e_{j,3}$.

The equivalent channel can be used to define simple performance metrics that can be exploited by the scheduler. One possible way of determining the merit of assigning a subcarrier to a given user is to compute the effective signal-to-noise ratio at the output of a linear MIMO equalizer or filter.

As stated above, the model (for representative user and subcarrier) is

$$\mathbf{z} = \mathbf{R}_{eq}\mathbf{x} + \mathbf{n}. \tag{3.16}$$

Due to matched-filtering with $\mathbf{E}_{eq}^\dagger$ at the receiver, the noise term is correlated. After matched filtering, the symbol vector $\mathbf{x}$ is detected with a linear filter $\mathbf{L}$ operating on eq. (3.16). The post-detection SNR is easily derived via [60], essentially by invoking the Gaussian approximation using coefficients

$$\theta_{k',j} = (\mathbf{L}^\dagger \mathbf{R}_{eq})_{k',j},$$

$$\beta_{k'} = \frac{(\mathbf{L}^\dagger \mathbf{R}_{eq})_{k',k'}}{\sigma\sqrt{(\mathbf{L}^\dagger \mathbf{R}_{eq}\mathbf{L})_{k',k'}}},$$

and

$$\lambda_{k'}^2 = \frac{\beta_k'^2 \sum_{j\neq k'} \theta_{k',j}^2}{\theta_{k',k'}},$$

where k' is the symbols index. Using these notations, a computationally attractive and accurate approximation to the average error probability for a given subcarrier is

$$P_b = \frac{1}{Q}\sum_{k'=1}^{Q} Q\left(\frac{\beta_{k'}}{\sqrt{1+\lambda_{k'}^2}}\right). \tag{3.17}$$

The fraction $\theta_{k',j}/\theta_{k',k'}$ quantifies interference leakage between the k'th and jth symbol. Likewise, the total capacity in a given subcarrier can be approximated in a similar fashion or any other related CQI can be devised.

In general, the post-detection signal-to-noise-ratio (SNR) approximation is

$$\mathrm{SNR}_{k'} = \frac{\beta_{k'}^2}{1+\lambda_{k'}^2} \tag{3.18}$$

Regarding the approximation, clearly $\lambda_k^2 = 0, \forall k'$ for a decorrelating receiver. In special cases, e.g. for full diversity modulators where each symbol is treated equally (as in DABBA), all symbols attain the same SNR. Then we can characterize the performance of all symbols with only one number, e.g. aggregate SNR. With BLAST, each symbol generally has different SNR and it is more difficult to obtain a simple and relevant performance indicator (although one is proposed in [1]).

The description above relates to a representative subcarrier. Considering a K user system, with P subcarriers, we have a set of signal models

$$\mathbf{z}_k[p] = \mathbf{R}_{eq,k}[p]\mathbf{x}_k + \mathbf{n}_k[p] \quad k = 1, ..., K, \; p = 1, ..., P \tag{3.19}$$

and each is eventually associated with a suitable performance indicator, each individually as given above as the effective SNR used as an input in the optimization model in Section 3.1. It should be noted that the current formalism does not explicitly account for bit loading used in adaptive modulation and coding solutions. In further work, these can be included by using e.g. the approach in [19].

3.4 Performance

The use of the assignment algorithm results in fair channel-aware frequency domain scheduling. In this section we quantify the performance improvement for a MIMO modulator in both uplink and downlink using a CQI that models BER for QPSK modulation. The transmission rate is 4 bps/Hz, since symbol rate 2 modulator (DABBA) is used in all P subcarriers. Regarding definitions in subsequent Figures, the E_b refers to transmitted energy and N_o is noise power in each receive antenna. Thus, by doubling the number of receive antennas, a 3 dB SNR gain is realized.

We evaluate the performance with and without complexity (dimension) reduction, and compare the results with round-robin scheduling (TDMA), where each user is assigned all subcarriers when accessing the channel. From the results, it will be seen that the multiuser diversity gain is apparent in all these cases, despite the strict fairness constraints.

3.4.1 Downlink

The first example depicts the benefit of subcarrier assignment in conjunction with MIMO (DABBA) modulation in downlink. Here, we have 64 subcarriers, 4 users, 4 path iid Rayleigh channel with minimal delay spread, statistically identical for all users. The subcarriers are assigned to these 4 users adaptively, using the assignment algorithm. Figure 3.1 shows the results using a decorrelating detector (both the defining the elements of the assignment matrix and in detection) for a case with 4 tx in the transmitter and 2 rx antennas in each of the 4 terminals.

The use of approximate assignment using complexity reduction, via assigning simultaneously a set of neighboring subcarriers, is seen to deteriorate performance only sightly when compared to the case where the assigned subcarrier need not be next to each other in frequency domain.

3.4.2 Uplink

Figure 3.2 shows the results using a decorrelating detector for a case with 2 tx in each of the 4 transmitters, and 4 rx antennas in the 4 receiver, corresponding to the uplink case. The MIMO transmission matrix for two tx antennas is formed by puncturing even columns from DABBA transmission matrix ("punctured DABBA").

As in downlink, the use of approximate assignment using complexity reduction, via assigning simultaneously a set of neighboring subcarriers, is seen to deteriorate performance only sightly. The performance gain, when compared to downlink, is due to greater number of receive antennas.

3.5 Conclusion

The scheduling algorithms given in this Chapter were targeted for assigning subcarriers to different users in a multiuser MIMO-OFDMA system. In contrast to time-domain scheduling, the frequency domain counterpart benefits from the availability of accurate channel state information. Here, the frequency response was assumed to be known, which is obviously an idealistic assumption. In practice, the CQI is always estimated, quantized, subject to signalling errors, possibly outdated, etc. Taking these effect fully into account is a demanding task, and possibly leads to the development of stochastic integer programs, much is the same way as random channel in power control problem leads to stochastic linear programs [26]. The computational burden associated with the use of assignment methods given here is tolerable, especially when the assignment is done simultaneously for several neighboring subcarriers (via dimension reduction). In addition to frequency-domain scheduling, a practical system is likely to include temporal scheduling which selects a subset of users for frequency-domain scheduling. This motivates a study of delay-differentiated time-frequency scheduling as yet another future research topic. In short, scheduling problems in wireless systems numerous and many relevant problems are still to be addressed within MASCOT.

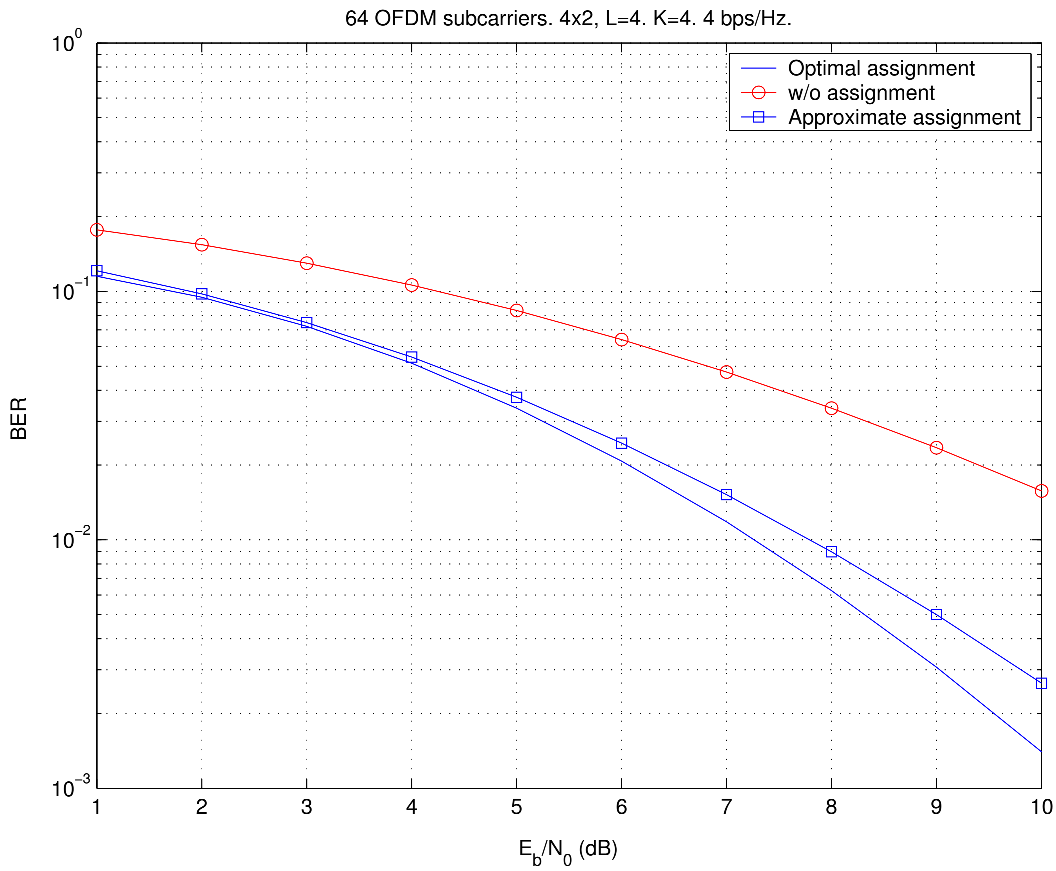

Figure 3.1: BER with subcarrier assignment in MIMO system using DABBA modulator (4 tx-2rx) in OFDMA with 4 users, 64 subcarriers in a four path channel. "Downlink"

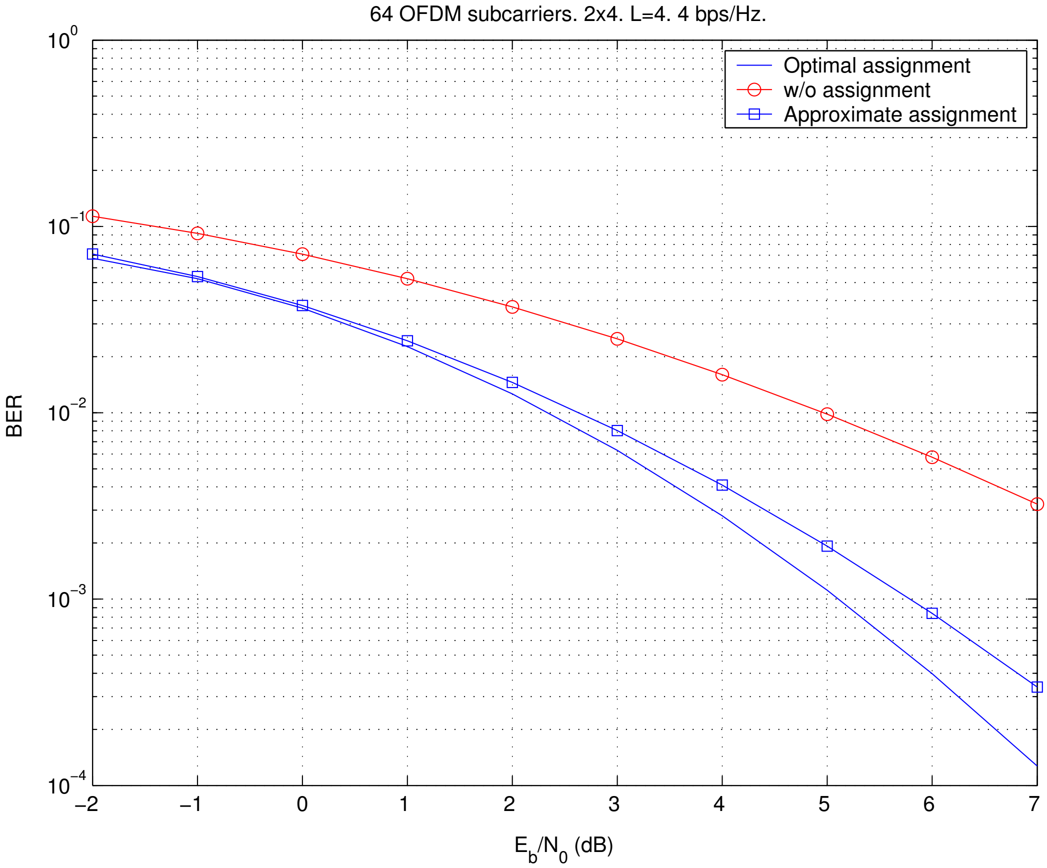

Figure 3.2: BER with subcarrier assignment in MIMO system using punctured DABBA modulator (2 tx-4rx) in OFDMA with 64 subcarriers in a four path channel. "Uplink"

Bibliography

[1] IEEE 802.16e standard D12. Air interface for fixed and mobile broadband wireless access system - amendment for physical and medium access control layers for combined fixed and mobile operation in licencesd bands. IEEE, October 2005. 22, 28

[2] G. Barriac and J. Holtzman. Introducing delay sensitivity into the proportional fair algorithm for cdma downlink scheduling. In *IEEE Seventh International Symposium on Spread Spectrum Techniques and Applications*, pages 652–656, Praque, Czech Republic, September 2002. 7

[3] H. Boche, E. A. Jorswieck, and A. Sezgin. Calculus for multiuser MIMO performance measures. In *Proc. Asilomar Conf. on Signals, Systems and Computers, Monterey*, October 2006. 14

[4] H. Boche and S. Stanczak. The Kullback-Leibler divergence and non-negative matrices. *IEEE Trans. Inform. Theory*, December 2006. 13, 14

[5] H. Boche and S. Stanczak. Strict convexity of the log-SIR region. In *IEEE Info. Theory Workshop (ITW), Chengdu, China*, 2006. 14, 17, 19

[6] H. Boche and M. Wiczanowski. The interplay of link layer and physical layer under mimo enhancement - benefits and challenges. *IEEE Communications Magazine*. to appear. 12

[7] H. Boche and M. Wiczanowski. Stability-optimal transmission policy for the multiple antenna multiple access channel in the geometric view. *EURASIP Signal Processing Journal, Special Issue on Advances in Signal Processing-assisted Cross-layer Designs*, (86):1815–1833, June 2006. 12

[8] Holger Boche and Martin Schubert. Resource allocation in multi-antenna systems – achieving max-min fairness by optimizing a sum of inverse SIR. *IEEE Trans. Signal Processing*, 54(6), June 2006. 14

[9] Holger Boche, Martin Schubert, Eduard Jorswieck, and Aydin Sezgin. A general framework for concave/convex interference coordination problems and network utility optimization. In *ITG/IEEE Workshop on Smart Antennas, Vienna, Austria*, 2007. 20

[10] Holger Boche, Martin Schubert, and Slawomir Stanczak. Proportional fair resource allocation for wireless networks. In *ITG/IEEE Workshop on Smart Antennas, Vienna, Austria*, 2007. 17, 20

[11] Holger Boche, Martin Schubert, Sławomir Stańczak, and Marcin Wiczanowski. An axiomatic approach to resource allocation and interference balancing. In *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP), Philadelphia, USA*, March 2005. 14

[12] Holger Boche and Sławomir Stańczak. Log-convexity of the minimum total power in CDMA systems with certain quality-of-service guaranteed. *IEEE Trans. Inform. Theory*, 51(1):374–381, January 2005. 14

[13] M. Bohge, J. Gross, and A. Wolisz. A new optimization model for dynamic power and sub-carrier allocations in packet-centric ofdma cells. In *Proc. of he 11th International OFDM-Workshop 2006 (InoWo '06)*, Hamburg, Germany, August 2006. 22

[14] J.-Y. Le Boudec. Rate adaptation, congestion control and fairness: A tutorial. Technical report, Tutorial, Ecole Polytechnique Federale de Lausanne (EPFL). 10, 11

[15] G. Caire and S. Shamai (Shitz). On the achievable throughput of a multi-antenna Gaussian broadcast channel. *IEEE Trans. Inform. Theory*, 49(7):1691–1706, July 2003. 8

[16] Daniel Catrein, Lorenz Imhof, and Rudolf Mathar. Power control, capacity, and duality of up- and downlink in cellular CDMA systems. *IEEE Trans. Commun.*, 52(10):1777–1785, 2004. 11, 14

[17] M. Chiang. To layer or not to layer: balancing transport and physical layers in wireless multihop networks. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, March 2004. 11

[18] M. Chiang and J. Bell. Balancing supply and demand of bandwidth in wireless cellular networks: Utility maximization over powers and rates. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 2, March 2004. 11

[19] M. Ergen, S. Coleri, and P. Varayia. QoS aware adaptive resource allocation techniques for fair scheduling in OFDMA based broadband wireless systems. *IEEE Tr. Broadcasting*, 9(4), December 2003. 22, 28

[20] G. J. Foschini and Z. Miljanic. A simple distributed autonomous power control algorithm and its convergence. *IEEE Trans. on Vehicular Technology*, 42(4):541–646, nov 1993. 11

[21] F. R. Gantmacher. *The Theory of Matrices, Vol. 2*. Chelsea Publishing Comp., New York, 1959. 17

[22] E. L. Hahne. Round-robin scheduling for max-min fairness in data networks. *IEEE J. Select. Areas Commun.*, 9(7):1024–1039, July 1991. 10

[23] S. Hanly and D. Tse. Power control and capacity of spread spectrum wireless networks. *Automatica*, 35(12):1987–2012, 1999. 11

[24] R.W. Heath, M. Airy, and A.J. Paulraj. Multiuser diversity for MIMO wireless systems with linear receivers. In *Proc. Asilomar Conf. on Signals, Systems and Computers*, volume 2, pages 1194–1199, 2001. 4, 22

[25] T. Heikkinen and A. Hottinen. Delay-differentiated scheduling in a fading channel. *(revision) Submitted to IEEE Tr. Wireless Comm.*, Dec 2006. 7

[26] T. Heikkinen and A. Prèkopa. Optimal power control in a wireless network using a model with stochastic link coefficients. *Naval Research Logistics*, 52(2):178–192, 2005. 11, 29

[27] A. Hottinen. Multiuser scheduling with matrix modulation. In *Proc. IEEE International Symposium on Signal Processing and Information Technology*, December 2003. 7, 22

[28] A. Hottinen and T. Heikkinen. Channel assignment in multi-antenna OFDM systems. In *European Wireless*, Athens, Greece, April 2-5 2006. 22, 24

[29] A. Hottinen and T. Heikkinen. Subcarrier allocation in a multiuser MIMO channel using linear programming. In *European Signal Processing Conference*, Florence, Italy, September 4-8 2006. 22

[30] A. Hottinen, O. Tirkkonen, and R. Wichman. *Multi-antenna Transceiver Techniques for 3G and Beyond*. Chichester: John Wiley and Sons, January 2003. 26

[31] Lorenz Imhof and Rudolf Mathar. Capacity regions and optimal power allocation for cdma cellular radio. *IEEE Trans. Inform. Theory*, 51(6):2011–2019, 2005. 11, 14

[32] S. Kandukuri and S. Boyd. Optimal power control in interference limited fading wireless channels with outage-probability specifications. *IEEE Trans. Wireless Commun.*, 1(1):46–55, January 2002. 11

[33] F. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: Shadow prices, proportional fairness and stability. *Journal of Operations Research Society*, 49(3):237–252, March 1998. 10, 11, 13

[34] F. P. Kelly. Charging and rate control for elastic traffic (corrected version). 8(1):33–37, January 1997. 10, 11, 12, 13, 14

[35] R. Knopp and P. A. Humblet. Information capacity and power control in single-cell multiuser communications. In *Proc. Int. Conf. on Communications, Seattle, USA*, June 1995. 4, 8

[36] M. Kobayashi and G. Caire. Joint beamforming and scheduling for a mimo downlink with random arrivals. In *Proc. IEEE Int. Symp. on Inf. Theory (ISIT), Seattle*, 2006. 12

[37] J.-W. Lee, M. Chiang, and A.R. Calderbank. Network utility maximization and price-based distributed algorithms for rate-reliability tradeoff. In *IEEE Infocom*, 2006. 11

[38] P. Marbach. Priority service and max-min fairness. *IEEE/ACM Trans. on Networking*, 11(5):733–746, June 2002. 10

[39] P. Marbach and R. Berry. Downlink resource allocation and pricing for wireless networks. In *IEEE Infocom 2002, (New York,NY)*, June 2002. 11

[40] L. Massoulie and J. Roberts. Bandwidth sharing: objectives and algorithms. *IEEE/ACM Trans. on Networking*, 3(10):320–328, June 2002. 10, 11, 12

[41] L. Massoulie and J. Roberts. Bandwidth sharing: objectives and algorithms. *IEEE/ACM Trans. on Networking*, 10(3):320 – 328, June 2002. 12

[42] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Trans. on Networking*, 5(8):556–567, October 2000. 10, 11, 12

[43] M. J. Neilly, E. Modiano, and C. E. Rohrs. Power allocation and routing in multibeam satellites with time-varying channels. *IEEE/ACM Trans. on Networking*, 11(1):138–152, February 2003. 12

[44] S. Pietrzyk and G.J.M. Janssen. Subcarrier and power allocation for QoS-aware OFDMA systems using embedded modulation. In *Proc. VTC 2000*, volume 6, pages 3202–3206, June 2004. 22

[45] B. Radunovic and J.-Y. Le Boudec. Why max-min fairness is not suitable for multi-hop wireless networks. In *Document, Ecole Polytechnique Federale de Lausanne (EPFL)*. 11, 12

[46] W. Rhee and J. H. Cioffi. Increase in capacity of multiuser OFDM systems using dynamic subchannel allocation. In *Proc. VTC 2000*, volume 2, pages 1085–1089, 2000. 22

[47] Martin Schubert and Holger Boche. A unifying theory for uplink and downlink multi-user beamforming. In *Proc. IEEE Intern. Zurich Seminar, Switzerland*, February 2002. 20

[48] Martin Schubert and Holger Boche. QoS-based resource allocation and transceiver optimization. *Foundation and Trends in Communications and Information Theory*, 2(6), 2006. 13, 14, 17

[49] Martin Schubert, Shuying Shi, and Holger Boche. Iterative transceiver optimization for linear multiuser MIMO channels with per-user MMSE requirements. In *Proc. European Signal Processing Conference (EUSIPCO)*, September 2006. (invited). 14

[50] R. Srinivasan and A. Somani. On achieving fairness and efficiency in high speed shared medium access. *IEEE/ACM Trans. on Networking*, 1(11):111–124, February 2002. 12

[51] S. Stanczak and H. Boche. Towards a better understanding of the quality-of-service tradeoff in multiuser multiple antenna systems. In *Smart Antennas – State-of-the-Art*. EURASIP, Hindawi Publishing Corp., 2005. to appear. 14

[52] C. W. Sung. Log-convexity property of the feasible SIR region in power-controlled cellular systems. *IEEE Communications Letters*, 6(6):248–249, June 2002. 14

[53] A. Tang, J. Wang, and S. H. Low. Is fair allocation always inefficient. In *Proc. INFOCOM*, 2004. 10, 11, 12

[54] L. Tassiulas and A. Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, 37:1936–1948, December 1992. 12

[55] L. Tassiulas and S. Sarkar. Maxmin fair scheduling in wireless networks. In *Proc. INFOCOM*, June 2002. 11

[56] S. Ulukus and R. Yates. Stochastic power control for cellular radio systems. *IEEE Trans. Commun.*, 46(6):784–798, June 1998. 11

[57] M. K. Varanasi and T. Guess. Achieving vertices of the capacity region of the Gaussian correlated-waveform multiple-access channel with decision feedback receivers. In *Proc. IEEE Int. Symp. on Inf. Theory (ISIT), Ulm, Germany*, page 270, June 1997. 8

[58] M. K. Varanasi and T. Guess. Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel. In *Proc. Asilomar Conf. on Signals, Systems and Computers, Monterey*, pages 1405–1409, November 1997. 21

[59] H. Varian. *The Economics of Information Technology, An Introduction*. Cambridge University Press, 2004. 11

[60] S. Verdú. *Multiuser Detection*. Cambridge University Press, 1998. 27

[61] S. Vishwanath, N. Jindal, and A. Goldsmith. On the capacity of multiple input multiple output broadcast channels. In *Proc. IEEE Int. Conf. on Comm. (ICC)*, 2001. 8

[62] P. Viswanath and D. Tse. Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality. *IEEE Trans. Inform. Theory*, 49(8):1912–1921, August 2003. 8

[63] P. Viswanath, D. N. C. Tse, and R. Laroia. Opportunistic beamforming using dumb antennas. *IEEE Trans. Inform. Theory*, 48:1277–1294, June 2002. 12, 13, 14

[64] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz). The capacity region of the Gaussian MIMO broadcast channel. In *Proc. 38th Annual Conference on Information Sciences and Systems (CISS)*, 2004. 8

[65] P.W. Wolniansky, G.J. Foschini, G.D. Golden, and R.A. Valenzuela. V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel. In *Proc. URSI Int. Symp. Sign., Syst. and Electr.*, pages 295–300, September 1998. 25

[66] L. Wolsey. *Integer programming.* Wiley, 1998. 8, 23, 24

[67] C. Y. Wong, C. Y. Tsui, R. C. Cheng, and B. Letaief. A real-time subcarrier allocation scheme for multiple access downlink OFDM transmission. In *Proc. VTC 1999*, volume 2, pages 1124–1128, 1999. 22

[68] Roy D. Yates. A framework for uplink power control in cellular radio systems. *IEEE J. Select. Areas Commun.*, 13(7):1341–1348, September 1995. 11, 14

[69] W. Yu and J. M. Cioffi. Sum capacity of Gaussian vector broadcast channels. *IEEE Trans. Inform. Theory*, 50(9):1875–1892, 2004. 8

[70] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi. Iterative water-filling for Gaussian vector multiple access channels. In *Proc. IEEE Int. Symp. on Inf. Theory (ISIT), Washington DC, USA*, June 2001. 8

[71] J. Zander. Distributed cochannel interference control in cellular radio systems. *IEEE Trans. on Vehicular Technology*, 41(3):305–311, August 1992. 11

[72] J. Zander and M. Frodigh. Comment on performance of optimum transmitter power control in cellular radio systems. *IEEE Trans. on Vehicular Technology*, 43(3):636, August 1994. 20

[73] J. Zander and S.-L. Kim. *Radio Resource Management for Wireless Networks*. Artech House, Boston, London, 2001. 15, 17