A FRAMEWORK FOR RESOURCE ALLOCATION IN OFDM BROADCAST SYSTEMS

Gerhard Wunder, Thomas Michel and Chan Zhou

Fraunhofer German-Sino Mobile Communications Lab, Heinrich-Hertz-Institut
Einstein-Ufer 37, D-10587 Berlin
{wunder,michel,zhou}@hhi.fhg.de

ABSTRACT

We consider an OFDM broadcast channel (BC) where resources have to be allocated among M users according to a scheduling policy. In this framework, we study four different schedulers, namely the maximum weight matching (MRMS), the minimum rates (MRS), the minimum sum power (MSPS) and the maximum throughput (MTS) scheduler. We present efficient algorithms and derive the corresponding stability regions. Moreover bounds on the average expected buffer length are presented.

1. INTRODUCTION

A key challenge of future wireless communication systems will be the smart and dynamic allocation of resources among multiple users depending on their individual requirements and the overall system state. This new *adaptivity paradigm* is about to win recognition widely and has already influenced state of the art systems such as UMTS HSDPA (high speed downlink packet access). Nevertheless, there are currently significant efforts to further enhance the downlink capacity of HSDPA within 3GPP long term evolution using OFDM as the downlink air interface. OFDM yields superior performance and higher implementation-efficiency compared to standard WCDMA technology. Furthermore, due to fine frequency resolution and a high number of degrees of freedom OFDM offers the possibility to apply such smart and flexible resource allocation schemes and to manage interference in a multi-cell environment [1]. On the other hand, the many degrees of freedom complicate the scheduling algorithms and a general framework providing optimal solutions in different settings is highly desirable.

As shown in [2] (even without modern duality tools) the OFDM BC channel can be viewed as a family of classical degraded BC channels. Thus, resource allocation generally reduces to finding the optimal power allocation over the parallel channels using a global Gaussian codebook. However, the capacity region of the OFDM BC as a *non-degraded* channel can be achieved with Costa Precoding over all subcarriers and a global en(de-)coding order which has to be specified. In recent work this general result has been applied to solve practical scheduling problems such as the sum power minimization problems or scheduling with rate requirements [3] using advanced Lagrangian and duality techniques.

In this paper we consider *optimal* resource allocation for OFDM broadcast channels (BC) in an ideal information-theoretic context. We introduce several scheduling policies in a common stability framework and analyze the performance in terms of throughput where we mean by throughput a combined measure of physical and medium access control (MAC) layer. We provide algorithmic solutions to each policy and indicate the supportable throughput region, i.e. the rates that can be (or not be) supported by this policy. These solutions serve as a general benchmark for more specific approaches (e.g. by using FDMA) and they also provide some intuition for good suboptimal solutions.

The remainder of this paper is arranged as follows. Section 2 presents the system model. The subsequent Section 3 contains a general analysis of stability and queue lengths. In Section 4, the schedulers are studied in detail while we conclude with some final comments in Section 5.

2. OFDM BROADCAST SYSTEM MODEL

2.1 Physical Layer

Supposing familiarity with OFDM we assume an OFDM BC with M users, K subcarriers, and a short term sum power constraint $\sum_{m,k=1}^{M,K}\mathbb{E}\{|x_{m,k}|^2\}=\sum_{m,k=1}^{M,K}p_{m,k}\leq\bar{P}$, where $x_{m,k}$ is the signal transmitted to user $m\in\mathscr{M}=\{1,...,M\}$ on subcarrier $k\in\mathscr{K}=\{1,...,K\}$ with power $p_{m,k}$ and $\mathbb{E}\{.\}$ stands for the expectation operator. Then, the system equation for each user on each subcarrier can be written as

$$y_{m,k}=h_{m,k}\sum_{j\in\mathscr{M}}x_{j,k}+n_{m,k},\quad m\in\mathscr{M},k\in\mathscr{K},\qquad(1)$$

where $y_{m,k}$ is the signal received by user m on subcarrier k, $n_{m,k}\sim\mathscr{CN}(0,\sigma^2)$ is circular symmetric additive white Gaussian noise with variance σ^2. Let $\mathbf{h}=[h_{1,1},\ldots,h_{1,K},h_{2,1},\ldots,...h_{M,K}]^T$ (we will do so as well for the rates, buffer states etc. by using common vector norms) denote the stacked vector of channel coefficients. We assume that these channel coefficients are related to a standard time-varying multipath model (describable by some power delay and Doppler profile) where the channel is approximately constant over the OFDM symbol. Furthermore, we assume that Costa Precoding is performed at the base station having full non-causal knowledge of all messages to be transmitted. Suppose that in the dual multiple access channel users can decode their messages using successive interference cancellation with the reverse coding order. Let $\pi\in\Pi$ be an arbitrary encoding order from the set of all $M!$ possible encoding orders, such that user $\pi(1)$ is encoded first, followed by user $\pi(2)$ and so on. Then the rate of user $\pi(m)$ can be expressed as

$$\tilde{r}_{\pi(m)}=\sum_{k=1}^{K}\log\left(1+\frac{g_{\pi(m),k}p_{\pi(m),k}}{1+g_{\pi(m),k}\sum_{n<m}p_{\pi(n),k}}\right).\qquad(2)$$

with $g_{m,k}=|h_{m,k}|^2/\sigma^2$ being the channel gain of user m on subcarrier k. The instantaneous capacity region of the OFDM

BC under a given sum power constraint $\bar{P}$ is given by

$$\mathscr{C}(\mathbf{h},\bar{P}) \equiv \bigcup_{\substack{\pi\in\Pi \\ \sum_{m,k=1}^{M,K} p_{m,k}\leq\bar{P}}} \left\{\mathbf{r} : r_{\pi(m)} \leq \tilde{r}_{\pi(m)}\ , m\in\mathscr{M}\right\} \quad (3)$$

where $\tilde{r}_{\pi(m)}$ is defined in equation (2) and $\mathbf{r}$ denotes the vector of rates. Now, the *ergodic capacity region* $\mathscr{C}_{erg}(\bar{P})$ is defined as the set of achievable rates averaged over the channel realizations subject to the short term sum power constraint $\bar{P}$. Furthermore, note that we know from recent results [2,4] that the capacity regions of OFDM multiple access channel and BC are identical (uplink-downlink duality) and thus solving a problem for the OFDM BC means to have the solution for the OFDM multiple access channel and vice versa. We will frequently use this duality in solving the various problems.

2.2 Medium access control layer

Assuming that transmission is time-slotted, data packets arrive randomly at the MAC and a buffer with finite length is reserved to store the incoming data for each user $m\in\mathscr{M}$. Simultaneously the data is read out from the buffers according to the system state, i.e. the random fading realization and the current queue lengths. Thus, the system can be modeled as a queuing system with random processes reflecting the arrival and departure of data packets.

Denoting the buffer state of the m-th buffer in time slot $n\in\mathbb{N}$ by $q_m(n)$ and arranging all buffer states in the vector $\mathbf{q}(n)\in\mathbb{R}_+^M$ the evolution of the queue system can be written as

$$\mathbf{q}(n+1) = [\mathbf{q}(n)+\mathbf{a}(n)-\mathbf{r}(n)]^+ . \quad (4)$$

The random vector $\mathbf{r}(n)\in\mathbb{R}_+^M$ describes the rates asserted to the individual users according to a specific scheduling policy $\mathscr{P}(\mathbf{q}(n),\mathbf{h}(n))$ and $\mathbf{a}(n)\in\mathbb{R}_+^M$ is a random vector denoting the data arrival process. Note that the process is reminiscent of random walk on the half line (with dependent increments) where we have rigorously used an uncountable state space formulation. Since the random variables $\mathbf{a}(n)$ are sampled at a given time interval T from M independent Poisson processes they are independent. Denoting the mean of the packet arrival rate of user m as λ_m and the constant packet size as s_m, the expected bit arrival rate for user m is given by $\rho_m = s_m\lambda_m$. On the other hand the random vector $\mathbf{r}(n)$ depends on the buffer and channel state.

3. STABILITY AND QUEUE LENGTHS

3.1 Throughput region

According to a MAC oriented terminology, we call the set of arrival rates ρ *stabilizable* by a specific scheduler the *throughput region* of the scheduling policy $\mathscr{P}(\mathbf{q}(n),\mathbf{h}(n))$. This convention brings up the question for a definition of stability, since in general there exist a variety of definitions. Loosely speaking, we define the stability region as the set of arrival vectors that can be supported avoiding buffer overflows. More precisely, we follow the notion of stability from [5, 6]. To this end define the overflow function $b_m(\zeta)$ for a user m

$$b_m(\zeta) = \limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n} 1_{[q_m(i)>\zeta]} \quad (5)$$

where $1_{[x>x_0]}$ is the indicator function. The overflow function reflects the fraction of time instances in which a certain queue level ζ is exceeded.

Definition 1 *A vector of arrival rates ρ is called* stabilizable by a specific scheduling policy $\mathscr{P}(\mathbf{q}(n),\mathbf{h}(n))$*, if the fraction of time instances exceeding a level ζ vanishes in the limit of ζ for all users:*

$$\lim_{\zeta\to\infty} b_m(\zeta) = 0 \quad \forall m\in\mathscr{M}. \quad (6)$$

Definition 2 *A vector of arrival rates ρ is called* stabilizable*, if there exists a scheduling policy $\mathscr{P}(\mathbf{q}(n),\mathbf{h}(n))$ such that*

$$\lim_{\zeta\to\infty} b_m(\zeta) = 0 \quad \forall m\in\mathscr{M}. \quad (7)$$

Further, we formalize the description from above:

Definition 3 *The set of all* stabilizable *arrival rate vectors of a scheduling policy $\mathscr{P}(\mathbf{q}(n),\mathbf{h}(n))$ is called its* throughput region *$\mathscr{S}_{\mathscr{P}}$.*

Now define the set of rates $\mathscr{C}_{erg}^{\mathscr{P}}$ according to

$$\mathscr{C}_{erg}^{\mathscr{P}} := \bigcap_{||\mathbf{q}||_1=1}\left\{r_1,...,r_M : \mathbf{q}^T\mathbf{r}\leq\mathbf{q}^T\mathbb{E}\left\{\mathbf{r}^{\mathscr{P}}(\mathbf{q},\mathbf{h})\right\}\right\} \quad (8)$$

where $\mathbf{r}^{\mathscr{P}}(\mathbf{q},\mathbf{h})$ is the rate allocated by policy $\mathscr{P}$ according to fading and buffer state $\mathbf{h}$ and $\mathbf{q}$, respectively. Subsequently we will use $\mathbf{r}_{erg}^{\mathscr{P}}(\mathbf{q}) = \mathbb{E}\{\mathbf{r}^{\mathscr{P}}(\mathbf{q},\mathbf{h})\}$. A quick inspection of (8) reveals that $\mathscr{C}_{erg}^{\mathscr{P}}$ is a convex set.

With these definitions at hand, we are interested in conditions - necessary and sufficient - , that characterize whether a system is stabilizable or not. This problem was studied in [7] for the MIMO case and for a finite number of fading states in [6]. With the subsequent theorem we present a stability proof with respect to $\mathscr{C}_{erg}^{\mathscr{P}}$ for continuous fading distributions.

Theorem 1 *Assume the fading process $\mathbf{h}(\cdot)$ to be ergodic and the scheduling policy regarding the channel to be sufficiently smooth. Suppose further that $\mathscr{C}_{erg}^{\mathscr{P}}$ contains all ergodic rate points assigned by policy $\mathscr{P}$ and that it is nonempty. A sufficient condition for a queuing system to be stabilizable under a scheduling policy $\mathscr{P}$, i.e. $\rho\in\mathscr{S}^{\mathscr{P}}$, is that the vector of arrival rates ρ lies strictly inside the set $\mathscr{C}_{erg}^{\mathscr{P}}$:*

$$\rho\in\operatorname{int}\left(\mathscr{C}_{erg}^{\mathscr{P}}\right). \quad (9)$$

Further, a necessary condition for a queuing system to be stabilizable is that the vector of arrival rates ρ lies not outside the set $\mathscr{C}_{erg}^{\mathscr{P}}$

$$\rho\notin\mathscr{C}_{erg}^{\mathscr{P}C} \quad (10)$$

where $\mathscr{A}^C$ is the complementary set of $\mathscr{A}$.

Proof 1 *First we prove the sufficiency part. By Theorem 9.4.1 in [8] it has to be shown (even without irreducibility) that the Lyapunov drift is nonpositive*

$$\Delta V(\mathbf{q}(n)) := \mathbb{E}\{V(\mathbf{q}(n+1)) - V(\mathbf{q}(n))|\mathbf{q}(n)\}\leq 0 \quad (11)$$

for some $\|\mathbf{q}(n)\| > B \in \mathbb{R}_+$ *where* V *with* $V(\mathbf{q}) \to \infty$ *as* $\|\mathbf{q}\| \to \infty$ *is any* Lyapunov *function. Omitting the time index* n *and choosing* $V(\mathbf{q}) := \|\mathbf{q}\|_2^2$ *we have by elementary calculus the sufficient condition*

$$\Delta V(\mathbf{q}) = \mathbb{E}\left(\|\mathbf{a}-\mathbf{r}\|_2^2 |\mathbf{q}\right) + 2\mathbb{E}\left(\mathbf{q}^T(\mathbf{a}-\mathbf{r})|\mathbf{q}\right) \leq 0.$$

Assuming the policy $\mathscr{P}$ *it is easy to see that*

$$\mathbb{E}\left(\mathbf{q}^T(\mathbf{a}-\mathbf{r})|\mathbf{q}\right) = \mathbf{q}^T\rho - \mathbf{q}^T\mathbf{r}_{erg}^{\mathscr{P}}$$

where $\mathbf{r}_{erg}^{\mathscr{P}}$ *is the boundary point of* $\mathscr{C}_{erg}^{\mathscr{P}}$ *that fulfills* $\mathbf{q}^T\mathbf{r}_{erg}^{\mathscr{P}} \geq \mathbf{q}^T\mathbf{r}'^{\mathscr{P}}_{erg}$ *for all other boundary points* $\mathbf{r}'^{\mathscr{P}}_{erg}$ *due to the convexity of* $\mathscr{C}_{erg}^{\mathscr{P}}$*. Taking some* ρ *strictly inside* $\mathscr{C}_{erg}^{\mathscr{P}}$ *with* $r_{erg_m}^{\mathscr{P}} - \rho_m \geq \beta, m \in \mathscr{M}$*, for some* $\beta > 0$ *we get*

$$\begin{aligned}\Delta V(\mathbf{q}) &= \mathbb{E}\left(\|\mathbf{a}-\mathbf{r}\|_2^2 |\mathbf{q}\right) + 2\mathbb{E}\left(\mathbf{q}^T(\mathbf{a}-\mathbf{r})|\mathbf{q}\right) \\ &= \xi - 2\beta||\mathbf{q}||_1\end{aligned} \tag{12}$$

with $\xi = \mathbb{E}(\|\mathbf{a}-\mathbf{r}\|_2^2|\mathbf{q})$*. Thus, the Lyapunov drift is nonpositive*

$$\Delta V(\mathbf{q}) \leq 0 \quad \forall ||\mathbf{q}||_1 > B \tag{13}$$

choosing $B = \xi/2\beta$ *concluding the first part.*

To prove the necessary condition is a bit more involving. All inequalities are component-wise in the following and without loss of generality assume finite channel coefficients. From [5] we have the necessary condition

$$\rho \leq \frac{1}{n}\sum_{i=1}^{n}\mathbf{r}(n) + \varepsilon, \quad n > n_0(\varepsilon) \tag{14}$$

for some $\varepsilon > 0$*. Further each admissible rate vector must lie inside the instantaneous rate region achievable with policy* $\mathscr{P}$*. Quantize the channel state space according to* $\mathbf{h}_j \in \mathscr{A} \subset \mathbb{R}_+^n$ *with* $j = 1, ..., |\mathscr{A}|$*. Define a quantization mapping* $f: \mathbb{R}_+^n \to \mathscr{A}$ *such that* $f(\mathbf{h}) \leq \mathbf{h}$ *component-wise with quantization levels* $\mathbf{h}_j \in \mathscr{A} \subset \mathbb{R}_+^n$*,* $j = 1, ..., |\mathscr{A}|$*. Then*

$$\begin{aligned}\rho &\leq \sum_{j\leq|\mathscr{A}|}\frac{1}{n}\sum_{i\in\chi_j^n}\mathbf{r}(i) + \varepsilon \\ &= \sum_{j\leq|\mathscr{A}|}\frac{1}{n}\sum_{i\in\chi_j^n}(\tilde{\mathbf{r}}(i) + \mathbf{r}_{\Delta\mathbf{h}}(i)) + \varepsilon\end{aligned} \tag{15}$$

where χ_j^n *is defined as the set of time slots where the channel is in state* j*,* $\tilde{\mathbf{r}}$ *is the quantized rate vector and* $\mathbf{r}_{\Delta\mathbf{h}}$ *is the quantization error. Now*

$$\begin{aligned}\rho &\leq \sum_{j\leq|\mathscr{A}|}\frac{|\chi_j^n|}{n}\frac{1}{|\chi_j^n|}\sum_{i\in\chi_j^n}\tilde{\mathbf{r}}(i) + \sum_{j\leq|\mathscr{A}|}\frac{|\chi_j^n|}{n}\max_{i\in\chi_j^n}\mathbf{r}_{\Delta\mathbf{h}}(i) + \varepsilon \\ &= \sum_{j\leq|\mathscr{A}|}\frac{|\chi_j^n|}{n}\mathbf{r}'_j + \sum_{j\leq|\mathscr{A}|}\frac{|\chi_j^n|}{n}\max_{i\in\chi_j^n}\mathbf{r}_{\Delta\mathbf{h}}(i) + \varepsilon\end{aligned} \tag{16}$$

where $\mathbf{r}'_j$ *lies in the convex hull of all admissible rate allocations in channel state* j*. Taking the limit the RHS of (16) converges*

$$\begin{aligned}&\limsup_{n\to\infty}\sum_{j\leq|\mathscr{A}|}\frac{|\chi_j^n|}{n}\mathbf{r}'_j + \sum_{j\leq|\mathscr{A}|}\frac{|\chi_j^n|}{n}\max_{i\in\chi_j^n}\mathbf{r}_{\Delta\mathbf{h}}(i) + \varepsilon \\ &\quad= \sum_{j\leq|\mathscr{A}|}\Pr(\mathbf{h}_j \leq \mathbf{h} < \mathbf{h}_j + \Delta\mathbf{h})\mathbf{r}'_j + \sup_n \mathbf{r}_{\Delta\mathbf{h}}(n) \\ &\quad= \sum_{j\leq|\mathscr{A}|}\int_{\mathbf{h}_j}^{\mathbf{h}_j+\Delta\mathbf{h}} dF_{\mathbf{h}}\mathbf{r}'_j + O(\Delta\mathbf{h}) \\ &\quad\leq \int_{\mathbf{h}}\mathbf{r}'(\mathbf{h})dF_{\mathbf{h}} + O(\Delta\mathbf{h})\end{aligned} \tag{17}$$

where $F_{\mathbf{h}}$ *is the fading distribution and* $\mathbf{r}'(\mathbf{h})$ *lies in the convex hull of all admissable rate allocations in fading state* $\mathbf{h}$ *according to the policy* $\mathscr{P}$ *(we have tacitly assumend that the regions increase with increasing* $\mathbf{h}$*). Since the second term can be made arbitrary small by choosing* $\Delta\mathbf{h}$ *small and, further, since* $\mathscr{C}_{erg}^{\mathscr{P}}$ *contains all admissable points the region coincides with the convex hull and the theorem follows.*

It is worth noting that stronger stability properties can be proved for the queue system. In fact, it follows from the Ljapunov technique that also the expected average queue size is finite [9] which is upper bounded next.

3.2 Bounds on expected queue lengths

To get any propositions on the delay - which is particularly interesting for QoS-constrained services - we study the expected average queue sizes. The following upper bound can be found:

Lemma 1 *The expected value of the sum of buffer lengths in the stationary regime is upper bounded by*

$$\begin{aligned}\limsup_{n\to\infty}\|\mathbf{q}(n)\|_1 &\leq \frac{\left(\|\rho\|_1 - \|\rho\|_2^2\right)}{2\max_{\mathbf{r}\in\mathscr{C}_{erg}^{\mathscr{P}}}\min_{m\in\mathscr{M}}(r_m - \rho_m)} \\ &+ \frac{\mathbb{E}\left(\max_{m\in\mathscr{M},\mathbf{r}\in\mathscr{C}(\mathbf{h},\bar{P})}\{r_m\}\cdot\max_{\mathbf{r}\in\mathscr{C}(\mathbf{h},\bar{P})}\{\|\mathbf{r}\|_1\}\right)}{2\max_{\mathbf{r}\in\mathscr{C}_{erg}^{\mathscr{P}}}\min_{m\in\mathscr{M}}(r_m - \rho_m)}\end{aligned}$$

Proof 2 *From a standard telescoping argument and the convergence of the first moment we know from Theorem 1 in [9]*

$$\limsup_{n\to\infty}\|\mathbf{q}(n)\|_1 \leq \frac{\xi}{2\beta}$$

In order to evaluate ξ *we have*

$$\begin{aligned}\xi =&\mathbb{E}\left(\|\mathbf{a}\|_2^2\right) - 2\sum_{m\in\mathscr{M}}\mathbb{E}(a_m)\mathbb{E}(r_m) + \mathbb{E}\left(\|\mathbf{r}\|_2^2\right) \\ \leq&\mathbb{E}\left(\|\mathbf{a}\|_2^2\right) - 2\sum_{m\in\mathscr{M}}(\mathbb{E}(a_m))^2 + \mathbb{E}\left(\|\mathbf{r}\|_2^2\right) \\ \leq&\left(\|\rho\|_1 - \|\rho\|_2^2\right) + \mathbb{E}\left(\max_{m\in\mathscr{M},\mathbf{r}\in\mathscr{C}(\mathbf{h},\bar{P})}\{r_m\}\cdot\max_{\mathbf{r}\in\mathscr{C}(\mathbf{h},\bar{P})}\{\|\mathbf{r}\|_1\}\right)\end{aligned}$$

where we used in the last step $\mathbb{E}(r_m) \geq \mathbb{E}(a_m)$*. The term* $\mathbb{E}(\|\mathbf{r}\|_2)$ *is upper bounded since the evaluation of* $\|\mathbf{r}\|_2^2$ *for* *some channel state is a non-convex problem. The denominator is clearly obtained from*

$$\beta = \max_{\mathbf{r}\in\mathscr{C}_{erg}^{\mathscr{P}}} \left\{ \min_{m\in\mathscr{M}} \{r_m - \rho_m\} \right\}. \tag{18}$$

In general, the average delay can be expressed asymptotically in the number of users and it is shown that the delay grows not faster than $O(\log\log M)^2$ as the number of users increases [9]. Lemma 1 can be carried over to the throughput regions of all presented schedulers.

4. SCHEDULING POLICIES

We present and analyze the four schedulers subsequently. Besides an efficient algorithmic solution we are interested in the achievable throughput region of each scheduler.

4.1 Maximum Rate Matching Scheduler (MRMS)

The MRM Scheduler is a stability optimal scheduler in the sense, that the stability region is maximized [6]. However, QoS and service specific criteria do not play any role and thus rates can not be guaranteed.

In each state n the MRMS calculates the rates according to

$$\mathbf{r}(n) = \underset{\mathbf{r}'(n)\in\mathscr{C}(\mathbf{h}(n),\bar{P})}{\arg\max} \mathbf{q}^T(n)\mathbf{r}'(n). \tag{19}$$

The solution to this problem has been studied by several authors and a very efficient algorithm exists [2] which is described in Algorithm 1. To this end we introduces the notion of *marginal utility functions* to characterize the revenue of each user m to the objective function

$$u_m^{(k)}(z) = \frac{q_m}{(1/g_{m,k}+z)} - \lambda \tag{20}$$

where λ is the Lagrangian parameter corresponding to the sum power constraint. The set of equations characterizing the solution is given by

$$r_m(n) = \int_0^\infty \sum_{k:u_m^{(k)}(z)=\left[\max_i u_i^{(k)}(z)\right]^+} \frac{1}{(1/g_{m,k}+z)}\,dz \tag{21}$$

$$P(\lambda) = \sum_{k=1}^{K}\left[\max_m\left(\frac{q_m}{\lambda} - \frac{1}{g_{m,k}}\right)\right]^+ \tag{22}$$

and $[a]^+ := \max(0,a)$. For a detailed study the reader is referred to [2, 10]. The solution to the system of equations (21)-(22) can be found by first solving equation (22) for the Lagrangian multiplier λ. This can be done by simple bisection, since the RHS of (22) is monotone in λ.

Algorithm 1 MRMS optimization

(1) solve (22) for Lagrangian factor λ
(2) determine the intersections of marginal utility functions (20) for all $k\in\mathscr{K}$
(3) calculate resulting rates $r_m(n)$ (21) for each user $m\in\mathscr{M}$

Concerning the achievable throughput region of the MRMS $\mathscr{S}_{MRMS}$ it follows immediately from Theorem 1 that the entire interior of the ergodic capacity region can be achieved $\mathscr{S}_{MRMS}\equiv \text{int}(\mathscr{C}_{erg}(P))$ since $\mathscr{C}_{erg}^{MRMS}\equiv\mathscr{C}_{erg}$. The stability of the boundary of $\mathscr{C}_{erg}(P)$ is an unsolved problem. The MRMS is thus throughput optimal. However, note that the MRMS is not the only scheduler achieving the maximum throughput region.

4.2 Minimum Rates Scheduler (MRS)

The MR scheduler was designed to overcome the shortcomings of the MRMS. This policy takes into account QoS constraints in form of guaranteeing a set of given minimum rates $\bar{\mathbf{r}} = [\bar{r}_1,...,\bar{r}_M]^T$ for some user subset $\mathscr{L}$ to assure e.g. real time services independent of the fading state h. Unfortunately the consideration of rate constraints is achieved at the expense of a reduced stability region.

The scheduler solves the following problem at each time instance t:

$$\mathbf{r}(n) = \underset{\mathbf{r}'(n)\in\mathscr{C}(\mathbf{h}(n),\bar{P}),r'_m(n)\geq\bar{r}_m,m\in\mathscr{L}}{\arg\max} \mathbf{q}^T(n)\mathbf{r}'(n). \tag{23}$$

An efficient algorithm solving the instantaneous problem in the dual MAC was presented in [3], which is summarized beneath. To this end define *effective noise* coefficients $n_{m,k}$ as

$$n_{m,k} := \log\left\{ e^{\sum_{n>m} r_{\pi_k(n),k}} \left[1/g_{\pi_k(m),k} + \sum_{j=1}^{m-1} 1/g_{\pi_k(j),k}\left(e^{r_{\pi_k(j),k}}-1\right) e^{\sum_{n=j+1}^{m-1} r_{\pi_k(n),k}}\right]\right\}^{-1}. \tag{24}$$

where π_k is a set of permutations such that the channel gains are ordered on each subcarrier decreasingly

$$g_{\pi_k(1),k} \geq g_{\pi_k(2),k} \geq \cdots \geq g_{\pi_k(M),k}. \tag{25}$$

Algorithm 2 MRS optimization

(1) check feasibility with Algorithm 3
(2) choose initial Lagrangian factors λ_+ and λ_-
while sum power constraint $\bar{P}$ is not met **do**
  while desired accuracy not reached **do**
    for $m = 1$ to M **do**
      (3) compute the coefficients $n_{m,k}$ (24) for user m
      (4) do water-filling according to his queue q_m (26)
      if $r_m(n) < \bar{r}_m$ **then**
        (5) choose μ_m in water-filling level such that $r_m(n) = \bar{r}_m$
      end if
    end for
  end while
  (6) increase (decrease) λ if $P > \bar{P}$ ($P < \bar{P}$) by bisection
end while

Further, the Karush-Kuhn-Tucker (KKT) optimality conditions can be written in a form allowing water-filling:

$$r_{\pi_k(m),k}(n) = \left[\log((q_m+\mu_m)/\lambda) + n_{m,k}\right]^+, \quad \forall m,k \tag{26}$$

The factor μ_m is the nonnegative Lagrangian multiplier of user m corresponding to his minimum rate requirement $\bar{r}_m$ and constitutes a revenue to the water-filling level. Note that the problem might be infeasible, if the feasible set is empty, i.e. the required minimum rates $\bar{\mathbf{r}}$ are not supportable with the given power budget $\bar{P}$. Thus, feasibility is checked by running Algorithm 3 first, which is presented in the next subsection.

In contrast to the MRMS, the MR Scheduler is not throuput optimal. The throughput region is generally limited by the minimum rate requirements $\bar{\mathbf{r}}$ if we consider only those fading states where the rate requirements can be fulfilled (see Fig. 1).

4.3 Minimum Sum Power Scheduler (MSPS)

In contrast to the MRMS and the MRS, the MSP scheduler is not concerned about stability issues; the primary objective is to achieve a set of user rates $\bar{\mathbf{r}} = [\bar{r}_1, \ldots, \bar{r}_M]^T$ with minimum sum power P. Obviously, this strategy has impact on the stability and the delay.

The policy of the MSPS is defined as follows:

$$\mathbf{p}(n) = \operatorname*{arg\,min}_{\bar{\mathbf{r}}(n) \in \mathscr{C}(\mathbf{h}(n), P)} P. \tag{27}$$

The optimization problem can be efficiently solved in the dual MAC similar to the MRS. In fact, the MSPS and MRS algorithms base on the same KKT conditions of an enhanced problem revealing the intimate connection of both [3]. Thus once again the water-filling structure can be exploited.

Algorithm 3 MSPS optimization (Iterative rate water-filling)

set $r_{m,k}(n) = 0 \quad \forall m \in \mathscr{M}, \forall k \in \mathscr{K}$
while desired accuracy is not reached **do**
  for $m = 1$ to M **do**
    (1) compute the coefficients $n_{m,k}$ (24) for user m
    (2) do water-filling with respect to the rates $r_{m,k}(n)$ for user m as in equation (26) setting $\lambda = 1$
  end for
end while

It can be shown that the throughput region is given by the interior of the hypercube defined by the set of affine inequalities

$$r_m < \bar{r}_m \quad \forall m \in \mathscr{M}. \tag{28}$$

4.4 Maximum Throughput Scheduler (MTS)

In contrast to the other schedulers, the MT Scheduler is purely physical layer oriented: The scheduler maximizes the overall throughput not taking into account other constraints according to

$$\mathbf{r}(n) = \operatorname*{arg\,max}_{\mathbf{r}' \in \mathscr{C}(\mathbf{h}(n), \bar{P})} \sum_{m=1}^{M} r'_m(n). \tag{29}$$

Despite its practical shortcomings in terms of fairness the MTS is nevertheless interesting for theoretical investigation. Achieving the maximum throughput with MTS comes at the cost of severe delays. This is illustrated in Fig. 1.

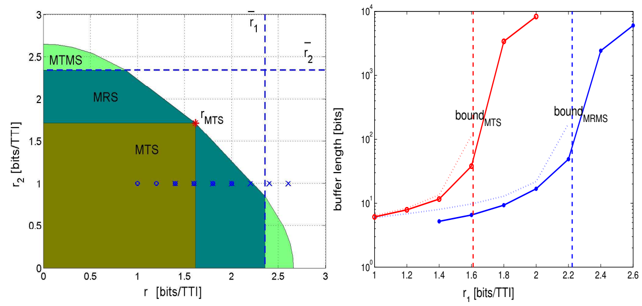

Figure 1: a) [left] Path in exemplary throughput regions b) [right] Simulated average buffer lengths (over 20000 Transmit Time Interval) and the bounds of e.g. MWMS and MTS

5. CONCLUSIONS

We presented a common stability framework and resource algorithms for OFDM broadcast channels. Clearly, this analysis has not come to its end and can be used for further steps towards investigation of more practical algorithms.

REFERENCES

[1] G. Wunder, C. Zhou, S. Kaminski, and H. E. Bakker, "Concept of an OFDM HSDPA Air Interface for UMTS Downlink," in *14th IST Mobile & Wireless Communication Summit*, Dresden, Jun 2005.

[2] D. Tse, "Optimal power allocation over parallel Gaussian broadcast channels," unpublished, available at http://www.eecs.berkeley.edu/˜dtse/broadcast2.pdf, 1998.

[3] T. Michel and G. Wunder, "Minimum rates scheduling for OFDM broadcast channels," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP)*, Toulouse, 2006.

[4] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 768–783, May 2004.

[5] M.J. Neely, E. Modiano, and C.E. Rohrs, "Power allocataion and routing in multibeam satellites wit time-varying channels," *IEEE/ACM Trans. on Networking*, vol. 11, pp. 138–152, Feb 2003.

[6] E. Yeh and A. Cohen, "Maximum throughput and minimum delay in fading multiaccess communications," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, July 2003.

[7] H. Boche and M. Wiczanowski, "Optimal scheduling for high speed uplink packet access," in *Proc. IEEE Vehicular Techn. Conf. (VTC)*, Milan, Italy, May 2004.

[8] S.P. Meyn and R.L. Tweedie, *Markov Chains and stochastic stability*, Springer Verlag, London, 1993.

[9] G. Wunder, C. Zhou, and T. Michel, "On Minimum Delay and Buffer Size in OFDM Broadcast Systems," 2006, Preprint.

[10] D.N.C. Tse and S.V. Hanly, "Multiaccess fading channels - part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2796–2815, Nov 1998.