Performance Analysis of Scheduling Policies for Delay-Tolerant Applications in Centralized Wireless Networks

Mohamed Shaqfeh and Norbert Goertz

Institute for Digital Communications
Joint Research Institute for Signal & Image Processing
School of Engineering and Electronics
The University of Edinburgh
Mayfield Rd., Edinburgh EH9 3JL, Scotland, UK
Email: {M.Shaqfeh, Norbert.Goertz}@ed.ac.uk

***Abstract*— Many scheduling schemes have been proposed in literature to control how different users access a wireless channel. Channel-aware schedulers exploit the measurements of instantaneous channel conditions of the different users to obtain throughput gains by proper selection of the users to transmit (or receive) in each channel state.**

In this paper, the performance of channel-aware scheduling policies which are applicable for delay-tolerant applications in centralized wireless networks are analyzed in a new mathematical framework. The framework is applied in numerical examples to compare the performance of different scheduling policies in terms of their efficiency in allocating the wireless resources. This efficiency is measured against the set of all possible operating points of the system, chosen, e.g., by the network operator.

***Index Terms* – centralized wireless networks, channel-aware scheduling, power allocation, delay-tolerant applications**

I. Introduction

An important feature of multiuser communication over fading channels is *multiuser diversity*: The independent fading characteristics of the users' channels increase the probability of having one or more users with very good channel condition at each time instance. The multiuser diversity gain can be obtained by exploiting the independent fading conditions, and scheduling only the users with good channels.

During the last decade, the design of multiuser scheduling schemes for wireless networks has been extensively studied. The algorithms which were state-of-the-art few years ago [1], are now outperformed by a lot of new schemes. Proposals in literature differ in their policies to choose the scheduled user based on channel conditions.

The objective of this work is to systematically compare, in a new universal framework, known channel-aware scheduling algorithms that are applicable for delay-tolerant applications in centralized wireless networks. Algorithms that consider strict delay-constraints are, however, not included in our analysis.

In centralized networks, the scheduling decisions for both the uplink and the downlink are taken at the wireless access point or the base station of the network. This unit is provided with a rich set of information such as the traffic load and the different QoS[1] requirements of the served traffic classes as well as the instantaneous channel conditions of the wireless links to active users. We assume in our analysis that the channel variations are not too fast so that the effect of the channel measurement delay [2] is negligible and the channel coefficients can be estimated at the receiver and be communicated to the transmitter with sufficient accuracy at low overhead. This also means that we can perform coherent detection at the receiver, i.e., we get no phase error and the "I" and the "Q" components are both scaled by the magnitude of the channel's fading coefficient which is the square root of the channel power gain.

For our analysis we assume general schedulers which may, e.g., be applied in OFDMA[2] systems with dynamic sub-carrier allocation (DSA) and adaptive modulation and coding (AMC). The OFDMA scheme enables the exploitation of multiuser diversity in the frequency domain (due to frequency-selective channels in the total transmission bandwidth) *and* in the time domain [3]. Therefore, OFDMA provides a rather general setting with other (e.g. single carrier) systems as special cases.

The scheduling decisions, i.e. the decisions which of the users are allowed to access the channel, are taken for each *channel block* over which the channel fading coefficients can be considered to be constant. Each channel block may consist of many *slots* (in time or frequency or both) over which the channel fading coefficients do not change as the slots belong to one channel block.

Some of the scheduling algorithms we include in a case study to demonstrate our new analysis were originally proposed and applied for single-carrier systems (e.g. the propor-

[1]Quality of Service
[2]Orthogonal Frequency Division Multiple Access

tional fair scheduler [4]). Since these scheduling algorithms are, on a sub-carrier level, also applicable to OFDMA systems, we include them in our analysis. For simplicity, we assume single-transmit single-receive antenna (SISO) systems. The scheduling schemes we discuss can, of course, also be used in a multiple-antenna (MIMO) framework but the analysis is more complicated, and this would distract from the main issues (scheduling and resource-allocation) we like to discuss.

In general, scheduling algorithms have two main objectives: efficient allocation of the scarce wireless resources and achieving suitable fairness criteria and QoS requirements of the different applications.

Since the network operator may have different objectives depending on the applications supported by the network and the way "fairness" is defined, one scheduling algorithm can be most suitable for a certain scenario or definition of fairness while it performs poorly in a different scenario. The key point of our work is to show that it is actually possible to compare scheduling algorithms in terms of their efficiency in allocating the wireless resources, i.e. power and bandwidth, across the *whole* range of possible operating points. Although there is a contradiction between the maximum total throughput and fairness and QoS constraints, there is *no* contradiction between the two objectives of (i) efficient resource allocation and (ii) achieving fairness with certain QoS requirements. It must, e.g., be accepted that there will be a loss in terms of total throughput when users with bad channels have to be served, but this is by no means a weakness of a particular scheduling scheme – it is rather a fundamental trade-off that is described by the theoretical limits of information theory, and the key question is how close to those limits a scheduling scheme performs. Efficient scheduling algorithms in our sense are those which operate on or close to the boundary of the capacity region, which is the set of long-term average achievable user rates for given average power constraints. Information-theory ([5], [6]) does not only provide such characterization of the capacity region, but it also provides guidelines for system designs leading to optimal performance, i.e., to achieving capacity limits. However, the optimal solutions are in most cases difficult if not impractical to implement. Thus, sub-optimal solutions which have close-to-optimal performance and at the same time lend themselves to an easy implementation are most favourable from a practical perspective. That is why there is a large number of scheduling algorithms suggested in the literature. In this paper, the channels' achievable rate limits when applying different power control policies are taken into consideration when comparing various scheduling policies.

In order to visualize the performance and efficiency of the scheduling algorithms, we analyze the two-user case, with the assumption of different long-term average channel qualities of the users. The achievable rates using the scheduling algorithms are plotted in order to compare the performance of the algorithms. Qualitatively, the results carry over to the M-user case.

The rest of the paper is organized as follows: first a summary of the scheduling algorithms which are included in the comparison is provided. Then, a description of the framework to evaluate the performance of the scheduling policies is given, followed by the numerical results of the performance analysis of the schedulers.

II. Scheduling Policies Proposed in Literature for use in a Performance Comparison Study

In literature, there exist scheduling schemes which use similar policies, but with different operating points[3] depending on the control parameters of the policy such as weighting factors and rate offsets, which are adjusted according to the fairness criteria and constraints for the given application. Examples of schemes to select an operating point include proportional user-rate ratios (throughput fairness) [7], proportional channel-access ratios (resource-sharing fairness) [8], [9], and utility function maximization [10].

In this section we provide a summary, along with a new unified algorithmic description, of scheduling policies proposed in the literature (all are applicable for both the uplink and the downlink) that lend themselves to a comparison in our framework. This also includes some novel extensions of the originally proposed schemes. Of course, our choice of schedulers is far from being exhaustive, but we believe our selection is sensible and provides a useful basis for comparison of other schemes as well. Moreover, our main contribution is a new framework (detailed in Section III) with which to analyze various types of schedulers; the specific scheduling algorithms we have picked are actually just examples used to demonstrate the new analysis.

Due to the objectives of this work, scheduling algorithms that explicitly take delay-constraints into account (such as the ones presented in [11], [12]) are not included in the mathematical analysis.

A. Scheduling Policies for Constant-Power Systems

All policies in this section will schedule exactly one user per channel block and, if scheduled, each user's power is the same. In what follows, and in particular in Figures 1 and 2 which present numerical results (details are discussed in Section IV), the different scheduling and power allocation policies are referred to by their *equation numbers* stated below.

1) User selection based on weighted feasible rates: Scheduling policies which schedule user m in channel block k according to a weighted value of the instantaneous feasible[4] rate $R_i[k]$ of the user are given by

$$m = \arg\max_i \mu_i R_i[k]\,. \quad (1)$$

[3]An operating point in a delay-tolerant system is defined by the vector of long-term average rates of the users.

[4]Below we will use the well-known capacity equation for an Additive White Gaussian Noise (AWGN) channel to estimate the rate $R_i[k]$ from a given power $P_i[k]$. The underlying assumptions are that we can approximate the fading multiuser channel by a block fading model and that the channel-coding blocksize can be made infinitely large over the "fading states" of the channel model and, hence, the Gaussian channel capacity can be achieved. Moreover, we assume (as discussed in Section I) that the channel power gain $h_i[k]$ is known at the transmitter, so that it can be exploited for scheduling decisions and/or transmit-power allocation.

 ISBN: 1-56555-320-9

The power allocated to the scheduled user is constant, i.e.

$$P_m[k] = \bar{P} \,. \tag{2}$$

This power allocation rule is used by all scheduling policies with a constant transmission-power constraint.

In the maximum sum-throughput scheduler[5], the weighting factors are all equal, i.e., $\mu_i = 1$. In the Proportional Fair (PF) Scheduler [14], μ_i is inversely proportional to the average throughput $T_i[k]$ of the user in a past window, i.e., $\mu_i = \frac{1}{T_i[k]}$. In [11] proportional fairness is suggested with payloads depending on the specific application. In [15] this policy is suggested in a generic form to maximize throughput relative to pre-specified target ratios.

The PF scheduler will not be included in the new analytical framework in Section III. The main reason is that we consider delay-tolerant applications: in this case any dynamic adaptation (e.g. as above by $T_i[k]$) of the scheduler parameters μ_i is actually counter-productive with respect to the achievable long-term average rates we are interested in. Further details are discussed in Section II-C.

2) User selection based on weighted channel quality: Those policies are given by

$$m = \arg\max_i \mu_i h_i[k] \,, \tag{3}$$

where $h_i[k]$ is the instantaneous power gain of user i's channel. In constant transmission power systems, this is equivalent to a policy which schedules the user with highest weighted received Signal-to-Noise ratio (SNR). In [16] this policy is suggested in two forms: maximum throughput ($\mu_i = 1$) and proportional fairness ($\mu_i = 1/\bar{h}_i$), where $\bar{h}_i$ is the long-term average channel power gain. In [9], the policy is suggested in a generic form to achieve pre-specified resource-sharing ratios.

3) User selection based on feasible rates with rate offset: In [8] it is suggested to maximize the average total system performance while satisfying pre-assigned "time-fraction" (channel-access rate) requirements of the users. The proposal is generic and applicable to any system performance measures. We include this scheduling concept in the comparison with the assumption that throughput is the system-performance measure to maximize. The resulting scheduling policy is given by

$$m = \arg\max_i \left(R_i[k] + v_i\right) \,, \tag{4}$$

where $R_i[k]$ is (as in Section II-A.1) the feasible rate for user i in channel block k and v_i is a rate offset which is adjusted such that pre-assigned resource-sharing constraints are achieved.

4) User selection based on the cumulative rate-density function: In [17] scheduling based on the cumulative density function (CDF) of user transmission rates $R_i(k)$ is suggested. The concept is to schedule the user whose rate is high enough, but least likely to take even larger values in other (e.g. future) blocks. This scheduling policy is given by

$$m = \arg\max_i \left(F_{R_i}\left(R_i[k]\right)\right)^{\frac{1}{w_i}} \,, \tag{5}$$

where $F_{R_i}(.)$ is the CDF of the user's feasible transmission rates. The weighting factors w_i are used to scan different operating points of the system based on the constraints of the applications.

[5]In [13] power control is used to achieve capacity. The same selection policy (1) maximizes the sum-capacity under a constant transmission-power constraint.

B. Scheduling Policies for Variable-Power Systems

1) User selection based on weighted feasible rates with waterfillig power allocation: In [18] "Proportional Fairness" with QoS provision in downlink OFDMA is suggested. The user selection in each sub-carrier is based on PF scheduling ([14], see also Section II-A) and the power $P_m[k]$ is allocated using the "water-filling approach" $P_m[k] = \max\left(\sigma^2\left[\frac{1}{\lambda_m} - \frac{1}{h_m[k]}\right], 0\right)$, with user-individual factors $\lambda_m = \frac{1}{\lambda T_m}$ that depend on the average rate T_m recently achieved for user m in a moving time-window of limited size. The factor λ is adjusted such that a specified average power constraint is met.

We use a generalized version of this concept: user selection is carried out by the general selection policy (1) based on weighted feasible rates, and the power allocation for user m (who is assumed to be scheduled in block k) is given by

$$P_m[k] = \sigma^2\left[\frac{\mu_m}{\lambda} - \frac{1}{h_m[k]}\right]^+ \tag{6}$$

where $[x]^+ \doteq \max(x, 0)$. The factor λ is again adjusted according to a long-term average power constraint, and μ_m are weighting factors used to pick a desired operating point. As in Section II-A.1, the same comments apply with repect to a dynamic adaptation of the weighting factors (this is further discussed in Section II-C).

It should be noted that for constant weights μ_i for all users this is the same as the suboptimal Time Division (TD) policy given in [5, Section III-C].

2) User selection based on weighted channel quality and simplified waterfilling power allocation: In [19] it is suggested to use a normalized-SNR-based user selection strategy with water-filling power control along the sub-carriers. However, the water-filling level λ is adjusted irrespectively of the user selection policy. For comparison, we consider a similar method in a generalized form: user selection is based on (3) and the power is controlled in the blocks according to

$$P_m[k] = \sigma^2\left[\frac{1}{\lambda} - \frac{1}{h_m[k]}\right]^+ \,. \tag{7}$$

Again λ has to be adjusted such that a long-term average power constraint is met. The difference to (6) is that the weighting factors μ_i used in the user selection are not used in (7).

C. Detrimental Effect of a Dynamic Adaptation of the Scheduler

In general all scheduling policies achieve their maximum performance when the control parameters (such as weighting factors or rate-offsets) are constant. But with constant control parameters it is impossible to control the delay and, hence, schedulers for delay-constrained applications often have to be dynamically adapted. A popular example is the Proportional Fair (PF) Scheduler [14], which uses (1) to take scheduling decisions but with the rate weighting factors adapted according to $\mu_i = \frac{1}{T_i[k]}$, where $T_i[k]$ is the "recently" achieved average rate in moving time-window. Such (or any other) dynamic adaptation will decrease the achievable long-term average rate: this immediately follows from the convexity of the achievable rate regions. We may think of two points (rate-tuples) on the boundary of the scheduler's rate region that are achieved for two different parameter settings. When the scheduler parameters dynamically change between those two parameter sets, we can get the achievable rates pro-rata by time-averaging the achieved rates in both cases. Hence, we obtain a point on the straight line connecting the two points on the boundary of the rate region. As the region is convex, any point on this line will lie inside the rate region but not on its boundary and, therefore, any dynamic adaptation of the scheduler is inherently sub-optimal. The detrimental effect on the achieved rate will be the larger the larger the rate-differences between the two points are. As we consider delay-tolerant applications, we will, therefore, not include any scheme in our theoretical analysis that uses dynamic adaptation of the scheduler parameters.

III. Mathematical Analysis of Achievable Rates of Different Scheduling Policies

When a scheduling policy allocates rate to a single user only in each channel block, the maximum possible achievable rate (bits/sec/Hz[6]) of user i who is scheduled in block k equals

$$R_i[k] = \log\left(1 + \frac{h_i[k]P_i[k]}{\sigma^2}\right) \tag{8}$$

for additive white Gaussian receiver noise with a variance of σ^2 – of course, (8) is the Shannon capacity for the AWGN channel. With AMC[7] a rate close to capacity can be achieved (see, e.g., [20]). In practice, wireless systems support a set of discrete rate values rather than a continuous range. However, our objective is not to evaluate the schedulers' achievable rates for some given set of practical modulation and coding schemes. Our goal is rather to evaluate the performances of scheduling schemes as such, without any system constraints that will change anyway from one application to another. Therefore, we use the idealisation (8) to relate "power" and "rate"; the relative performances[8] of various scheduling schemes will carry over into practice.

In constant power systems, $P_i[k]$ in (8) is constant in all blocks in which user i is scheduled, i.e., $P_i[k] = \bar{P}\ \forall\, k$. In systems applying power control, $P_i[k]$ will be a function of $h_i[k]$ (see, e.g., (6)).

Assuming that all the blocks have identical bandwidths and time durations, the average achievable rate (bits/sec/Hz) of user i, in the blocks in which user i is scheduled, equals

$$\tilde{\bar{R}}_i = \frac{1}{|\mathcal{S}_i|}\sum_{n\in\mathcal{S}_i}\log\left(1 + \frac{h_i[n]P_i[n]}{\sigma^2}\right) , \tag{9}$$

with $\mathcal{S}_i$ the set of indices of all channel blocks in which user i is scheduled.

Of course, user i does not transmit in all blocks but rather in a ratio ϱ_i of the total number of blocks. For a very large number of considered blocks this ratio converges against the probability that user i is scheduled and, hence, we set

$$\varrho_i = \Pr\{i \text{ is the scheduled user}\} . \tag{10}$$

Thus, the actually achievable long-term average rate (bits/sec/Hz) of user i is

$$\bar{R}_i = \varrho_i\,\tilde{\bar{R}}_i = \frac{\varrho_i}{|\mathcal{S}_i|}\sum_{n\in\mathcal{S}_i}\log\left(1 + \frac{h_i[n]P_i[n]}{\sigma^2}\right) . \tag{11}$$

The averaging in (11) over the realisations $h_i[n]$ from the set $\mathcal{S}_i$ of channel power gains can be replaced by an integration over a probability density function (PDF) of the random variable H_i by exploiting that the random process created by a time series of realisations of H_i from the set $\mathcal{S}_i$ is ergodic[9]:

$$\bar{R}_i = \varrho_i \int_0^\infty \tilde{f}_{H_i}(h_i)\log\left(1 + \frac{h_i\,P_i(h_i)}{\sigma^2}\right) dh_i \tag{12}$$

where

$$\tilde{f}_{H_i}(h_i) \doteq f_{H_i}(h_i | i \text{ is the scheduled user}) \tag{13}$$

is the conditional PDF of the channel power gain of user i, given that user i is scheduled (to transmit (uplink) or to receive (downlink)). The PDF (13) is different from that of the actual channel power gain since the user is transmitting with higher probability when the channel gain is larger. Furthermore, note that

$$\tilde{f}_{H_i}(h_i) = \frac{d\tilde{F}_{H_i}(h_i)}{dh_i} \tag{14}$$

with $\tilde{F}_{H_i}(h_i)$ the CDF[10] of the channel power gain, h_i, over the blocks in which user i is scheduled (to transmit or to

[6]This relates to each Hz of bandwidth on the radio-frequency bandpass channel. Bandwidth is defined as the width of a compact set of positive bandpass frequencies for which the signal spectrum is allowed to be non-zero. The occupied bandwidth in each real (I/Q) sub-channel of an equivalent complex baseband model is half the bandpass radio-frequency bandwidth.

[7]Adaptive Modulation and Coding

[8]The "absolute" performance of a combination of specific modulation and coding schemes can often be approximated by (8) as well. An "acceptable" residual bit or frame error rate will often be achieved by a practical scheme with some (fairly constant) power-offset against the theoretical "zero-error" curve given by (8).

[9]The scheduling decisions depend on the channel power gains of all users (and perhaps on a set of constant weighting factors), and all those channel power gains are assumed to form independent and ergodic random processes. Therefore, the new random process, which is created by considering the power gains only when the user is scheduled, will also be ergodic.

[10]cumulated density function

receive). We can write $\tilde{F}_{H_i}(h_i)$ equivalently as follows:

$$\begin{aligned}\tilde{F}_{H_i}(h_i) &= F_{H_i}(h_i|i \text{ is the scheduled user}) &(15)\\ &= \Pr\{H_i \le h_i|i \text{ is the scheduled user}\} &(16)\\ &= \frac{\Pr\{H_i \le h_i, i \text{ is the scheduled user}\}}{\Pr\{i \text{ is the scheduled user}\}} &(17)\\ &= \frac{\Pr\{H_i \le h_i, i \text{ is the scheduled user}\}}{\varrho_i} &(18)\end{aligned}$$

Thus, we obtain

$$\tilde{f}_{H_i}(h_i) = \frac{d\Pr\{H_i \le h_i, i \text{ is the scheduled user}\}}{\varrho_i\, dh_i}\,. \qquad (19)$$

Hence, we can rewrite (12) as

$$\bar{R}_i = \int_0^\infty v_{H_i}(h_i)\log\left(1 + \frac{h_i\, P_i(h_i)}{\sigma^2}\right) dh_i \qquad (20)$$

where

$$v_{H_i}(h_i) \doteq \frac{d}{dh_i}\Pr\{H_i \le h_i, i \text{ is the scheduled user}\}\,. \quad (21)$$

For adaptive power allocation systems, the power $P_i(h_i)$ in (20) contains λ which needs to be adjusted to maintain the average power constraint $\bar{P}$, i.e., λ is selected such that

$$\sum_i \bar{P}_i = \sum_i \int_0^\infty v_{H_i}(h_i)P_i(h_i)dh_i = \bar{P}\,. \qquad (22)$$

For systems with a constant-power constraint the situation is much simpler, as, if user i is scheduled in block k, $P_i(h_i[k]) = \bar{P}$ and, hence, $P_i[k] = \bar{P}$.

The next step in order to evaluate (20) and (22) is to derive the relation of $v_{H_i}(h_i)$ in terms of the known weighting factors μ_m and the known unconditional channel PDFs $f_{H_m}(h_m)$ of all users.

We consider policies that schedule no more than one user in each channel block (i.e., there is no time-sharing between any two users within a channel block with constant channel gain). In order to find the probability that a user is scheduled we define, as a novelty, continuous non-decreasing auxiliary functions $g_{ij}(h_i[k])$ which can be used as follows to take scheduling decisions:

$$\text{user } i \text{ is scheduled in block } k \quad \text{if and only if}$$
$$h_j[k] < g_{ij}(h_i[k]) \quad \forall j \neq i \qquad (23)$$

By (23), the functions $g_{ij}(h_i[k])$ are as yet only implicitly defined. We will describe below how to obtain them; of course, they will depend on the specific scheduling policy chosen. The functions $g_{ij}(h_i[k])$ describe the borders of the regions within the channel-gain vector-space $\boldsymbol{h} \doteq \{h_1, h_2, ...\}$ over which the different users are scheduled. In what follows we drop the block index k, as the functions $g_{ij}(.)$ actually describe how decisions are taken by some policy for a given set of channel coefficients[11] ("channel state").

[11] Of course the coefficients depend on the block index k and therefore the value of $g_{ij}()$ will also depend on k. The scheduling policy itself that is described by $g_{ij}()$ is, however, not dependent on the block index k.

TABLE I

$g_{ij}(h_i)$ FOR THE SCHEDULING POLICIES UNDER CONSIDERATION. EACH SCHEDULING POLICY IS CHARACTERIZED BY A SELECTION POLICY TO SELECT THE USER m IN BLOCK k, AND A POWER-ALLOCATION POLICY P_m FOR THE SCHEDULED USER. IN THE TABLE WE REFER TO THE EQUATION NUMBERS OF THESE POLICIES. THE BLOCK INDEX k IS OMITTED FOR BREVITY.

m	P_m	$g_{ij}(h_i)$
(1)	(2)	$\frac{\left(1+h_i\bar{P}\right)^{\frac{\mu_i}{\mu_j}}-1}{\bar{P}}$
(3)	(2)(7)	$\frac{\mu_i}{\mu_j}h_i$
(4)	(2)	$\left(\frac{\exp(v_i-v_j)\left[1+h_i\bar{P}\right]-1}{\bar{P}}\right)^+$
(5)	(2)	$F_{h_j}^{-1}\left(\left[F_{h_i}(h_i)\right]^{\frac{w_j}{w_i}}\right)$
(1)	(6)	$\frac{\lambda}{\mu_j}\left(\frac{\mu_i h_i}{\lambda}\right)^{\frac{\mu_i}{\mu_j}} : h_i > \frac{\lambda}{\mu_i}$ and $0 : h_i \le \frac{\lambda}{\mu_i}$

In general, the scheduling policies we consider have the format

$$m = \arg\max_i y_i(h_i) \qquad (24)$$

where y_i is an increasing function of h_i. Then the only possibility that user i is scheduled is iff[12] the channel power gains $h_j\ \forall j \neq i$ are below certain values which are specified by the $g_{ij}(h_i)$ functions.

For example, if the scheduler is applying policy (3), i.e. $y_i(h_i) = \mu_i h_i$, then user i is scheduled iff for every other user $j \neq i$

$$\mu_j h_j < \mu_i h_i \quad \Leftrightarrow \quad h_j < \frac{\mu_i}{\mu_j}h_i \doteq g_{ij}(h_i) \qquad (25)$$

This defines the function $g_{ij}(h_i)$ for this scheduling policy.

As another example, the scheduler may be applying policy (1), i.e. $y_i(h_i) = \mu_i R_i = \mu_i\log(1 + h_i\bar{P})$. Then user i is scheduled iff for every other user $j \neq i$

$$\begin{aligned}\mu_j\log\left(1+h_j\bar{P}\right) &< \mu_i\log\left(1+h_i\bar{P}\right) \Leftrightarrow &(26)\\ 1+h_j\bar{P} &< \exp\left[\frac{\mu_i}{\mu_j}\log\left(1+h_i\bar{P}\right)\right] \Leftrightarrow &(27)\\ h_j &< \frac{\left(1+h_i\bar{P}\right)^{\frac{\mu_i}{\mu_j}}-1}{\bar{P}} \doteq g_{ij}(h_i) &(28)\end{aligned}$$

Using the applied procedures in these examples, we can also obtain $g_{ij}(h_i)$ for all other scheduling policies under consideration. We summarize the results for $g_{ij}(h_i)$ in Table I.

Having defined the functions $g_{ij}(h_i)$, we can now go back

[12] if and only if

to calculate $v_{H_i}(h_i)$. We obtain from the definition (21) of $v_{H_i}(h_i)$ and from (23):

$$v_{H_i}(h_i) = \frac{d}{dh_i} \Pr\{H_i \leq h_i, i \text{ is scheduled}\} \quad (29)$$

$$= \frac{d}{dh_i} \Pr\{H_i \leq h_i,\ H_j < g_{ij}(H_i)\, \forall\, j\} \quad (30)$$

$$= \frac{d \int_0^{h_i} f_{H_i}(x) \prod_{j \neq i} \Pr\{H_j < g_{ij}(x)\} dx}{dh_i} \quad (31)$$

with (31) following from the independence of the channel power gains of the users. From the differentiation of the integral in (31) we obtain

$$v_{H_i}(h_i) = f_{H_i}(h_i) \prod_{j \neq i} \Pr\{H_j < g_{ij}(h_i)\} \quad (32)$$

$$= f_{H_i}(h_i) \prod_{j \neq i} F_{H_j}(g_{ij}(h_i)) \quad (33)$$

with f_{H_i} the (unconditional and stationary) PDF of user i's channel gains and $F_{H_j}(h_j) \doteq \int_0^{h_j} f_{H_j}(x)dx$ the unconditional CDF of the channel power gains for the users j.

Equations (20) and (22) can now be evaluated by using $v_{H_i}(h_i)$ according to (33). Note that (33) involves the known simple channel models (unconditional PDFs and CDFs) for the users' channel coefficients. The structural properties of the scheduling policy and the power allocation scheme are completely captured by the newly defined $g_{ij}(h_i)$-functions: by use in (20), these $g_{ij}(h_i)$-functions allow for a simple evaluation of the achievable rate of any scheduling policy with single-user selection in each block k (see Table I). By (if necessary numerical) evaluation of the integral (22), the $g_{ij}(h_i)$-functions also allow to adjust the control factor λ which is used, e.g., in power control policy (6). Note that the evaluation of (20) and (22) by means of the $g_{ij}(h_i)$-functions is a great simplification in comparison to a time-simulation of the scheduler with its associated power control strategy and time-averaging of the rates over "many" channel realisations to obtain a statistically significant values for the rate-averages. Moreover, the $g_{ij}(h_i)$-functions provide a useful tool for an analytical characterisation of scheduling policies; we provide results in Section IV that would, due to the extensive simulation time required, be very hard to obtain by simulation and time-averaging.

IV. Numerical Results for the Achievable Rates of Different Scheduling Policies

We now apply (33) to evaluate (20) and (22): we assume the magnitudes a_i of the users' channel coefficients to have Rayleigh or Rice distributions [21, pp. 45–48], [20, pp. 78–79]. As the channel power gain, h_i, is the square of the channel-coefficient's magnitude, we have to use the variable-substitution $h_i = a_i^2$ in the original Rayleigh/Rice PDFs. As $q(a_i) = a_i^2$ is monotonically increasing for $a_i > 0$, we use the standard rule $f_{H_i}(h_i) = \left.\frac{f_{A_i}(a_i)}{q'(a_i)}\right|_{a_i = q^{-1}(h_i)}$ with $q'(a_i) = 2a_i$ and $q^{-1}(h_i) = \sqrt{h_i}$ to obtain the PDF of the channel power gain h_i from the PDF of the channel coefficient's magnitude a_i. For use in (33) we obtain for the Rayleigh-case:

$$f_{H_i}(h_i) = \frac{1}{\bar{h}_i} \exp\left(-\frac{h_i}{\bar{h}_i}\right) \quad (34)$$

$$F_{H_i}(h_i) = 1 - \exp\left(-\frac{h_i}{\bar{h}_i}\right) \quad (35)$$

with $\bar{h}_i$ the long-term average channel power gain.

For the Rice-case we find

$$f_{H_i}(h_i) = \frac{\kappa+1}{\bar{h}_i} \exp\left(-\kappa - \frac{\kappa+1}{\bar{h}_i} h_i\right) I_0\left(2\sqrt{\frac{(\kappa+1)}{\bar{h}_i}\kappa h_i}\right) \quad (36)$$

$$F_{H_i}(h_i) = \int_0^{\frac{\kappa+1}{\bar{h}_i} h_i} e^{-(\kappa+x)}\, I_0\left(2\sqrt{\kappa\, x}\right) dx \quad (37)$$

with $I_0(.)$ the zero-th order modified Bessel function[13] of the first kind, and κ is the fading parameter that is defined as the ratio $\kappa \doteq h_{i,\mathrm{LOS}}/h_{i,\mathrm{NLOS}}$ of the average power gains h_{LOS} on the line-of-sight path and h_{NLOS} on the non-line-of-sight path: $\kappa = 0$ means we get Rayleigh fading, and $\kappa \to \infty$ means "no fading". The long-term average channel power gain is again denoted by $\bar{h}_i$.

We will consider the two-user case in two examples. This is for clarity only, as it is difficult to visualize and compare higher-dimensional rate regions. Of course, the mathematical concepts and the results from Section III can also be applied to the general M-user case. Moreover, the numerical results for the two-user case presented below provide interesting insights that will carry over to the M-user case.

Figures 1 and 2 show numerical results. Similar to Table I, we indicate by their *equation numbers* the user selection policy and power allocation policy of each scheduling scheme shown in the figures.

In Figure 1 the achievable rates (in bits/sec/Hz), over all possible operating points of the system, are depicted using the policies which schedule a single-user per channel block with constant transmit power. We consider a Rice fading channel for the first user ($\kappa = 10$), with a long-term average channel power gain that is 10dB higher than that of the second user, who has a Rayleigh fading channel. Figure 1 shows that the weighted feasible-rates policy (1) is the best. However, the weighted channel-gains policy (3) works almost as good for all operating points and, thus, is attractive for generic schedulers, as (3) does not need to assume a particular relation (such as the AWGN capacity equation) between "power" and "rate". Another advantage of this policy is that, unlike the weighted feasible-rate policy, it has a continuous probability distribution and thus, with probability of one, a single user is maximizing the scheduler metric.

[13] This function can be represented [21, p. 44] by the infinite series $I_0(x) = \sum_{k=0}^{\infty} \frac{1}{(2^k\, k!)^2} x^{2k}$, $x \geq 0$. As, for k sufficiently large, the denominator will dominate the result, a limited number of summands will suffice to get accurate results. This means that (36) can be evaluated without explicit use of any Bessel function and that (37) can be evaluated without any numerical integration.

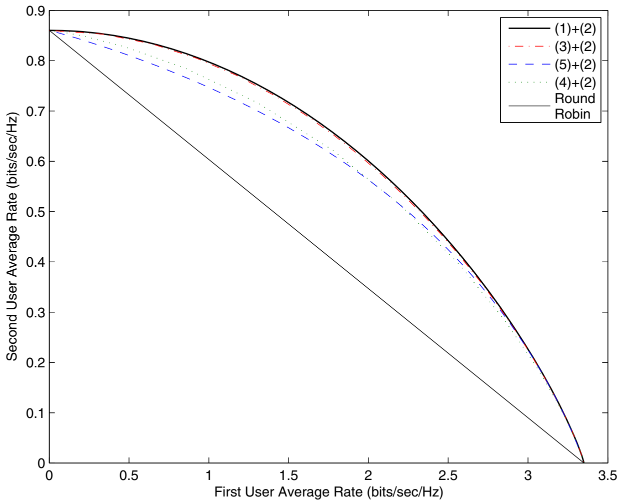

Fig. 1. Achievable long-term average rates of the constant-power-per-block scheduling policies for the two-user case. The scheduling policies are indicated by their equation numbers in the legends; (2) is the equation number of the constant-power allocation policy. The channel coefficient of the first user has a Rice-distribution ($\kappa = 10$), while Rayleigh fading applies to the second user. The average channel power gain of user 1 is 10 dB higher than that of user 2. The Round-Robin policy schedules user 1 in odd-numbered channel blocks and user 2 in even-numbered blocks, regardless of their channel coefficients; power allocation is also by (2).

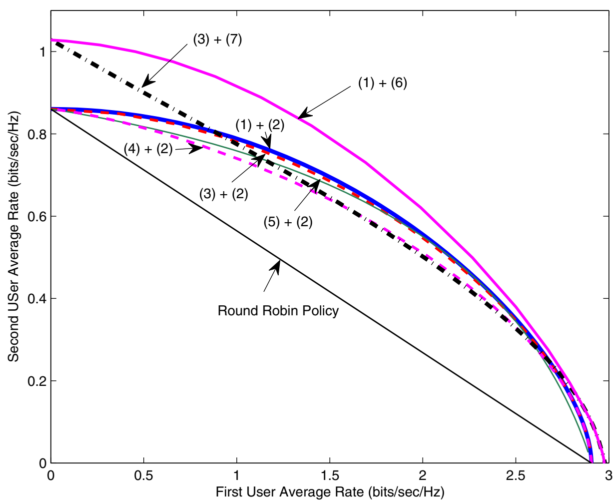

Fig. 2. The performance of various scheduling policies (section II) in a two-user case. The channels are assumed to be Rayleigh faded with 10 dB difference of the average channel power gains.

Policy (4) coincides with the limits achieved by policy (1) (which has the best possible performance for constant power) at the maximum sum-throughput point[14]. The latter is achieved by policy (1) when its weighting factors are all equal, and it is achieved by policy (4) when the rate offsets are all "zero". For all other points policy (4) has degraded performance in comparison to policy (1), and this degradation is larger when the system operates at low spectral efficiency (low sum-rate); the degradation at high spectral efficiency is however very small. Policy (5) is similar to policy (4) in that it has degraded performance compared to (1), but unlike (4) its best performance (again coinciding with policy (1)) is not at the maximum sum-throughput but rather on another operating point which depends on the fading channel models; that is why we observe different results in Figures 1, 2 for both policies.

In Figure 2 we investigate scheduling policies involving power control. We assume Rayleigh fading channels for both users, with 10dB higher average channel power gain for the first user. For comparison, we also include the constant-power policies investigated in Figure 1 already[15].

Figure 2 demonstrates that power control is helpful for the users with low-spectral efficiency (low rate), while using constant power is justified when operating at high spectral efficiency. As a compromise, we may use a limited number of different power levels (as long as this does not cause interference problems) to be better able to support users with bad channels. The power allocation using waterfilling in single-user systems means allocating more power when the channel is better. In multiuser communication, it also means allocating more power to the user with higher rate-reward μ. Thus, policy (7) is a bad choice. The selection policy (1), combined with power allocation policy (6), gives a better performance.

V. Conclusions

We have presented a new framework to evaluate the performance of scheduling policies which select at most one user per channel block. The mathematical framework was applied in numerical examples to compare the performances of scheduling policies known from literature. The performance is dependent on the user selection strategy as well as the power control policy. Although the exemplary analysis of the two-user case does not provide an exhaustive numerical evaluation of the scheduling policies considered, it highlights important results which, qualitatively, carry over to the general case. As a summary, we obtain the following results:

- There is no contradiction between efficient resource allocation and maximizing any network performance metric for delay-tolerant applications.
- A good scheduler uses a policy that closely achieves capacity for the operating point selected. Maximizing any utility function or maintaining fairness criteria should be done such that a suitable operating point is chosen on the capacity region's boundary.
- With known channel coefficients at both ends, the optimal scheduling policy does not depend on the statistics of the channel fading process.

[14]Policies (4) and (1) also coincide for the the trivial case that either user 1 or user 2 are scheduled all the time, but this is no longer a "multiuser case".

[15]We do not duplicate results, as the channels are different from those in Figure 1.

- For constant power and single user selection per block (1) is the best policy. If power control is applied, (6) should be used to adjust the power.
- Some scheduling policies are good for some operating points only, but they are not generic. There are policies which can be used in generic schedulers (such as (3)) because they have close-to-optimal performance for all set of operating points.
- Dynamic "on-line" variation of the weighting factors degrades the performance and will be avoided by a good scheduler if the applications are delay tolerant.

VI. Acknowledgements

The work reported in this paper has formed part of the Core 4 Research Program of the Virtual Centre of Excellence in Mobile and Personal Communications, Mobile VCE, www.mobilevce.com, whose funding support, including that of EPSRC, is gratefully acknowledged. Fully detailed technical reports on this research are available to Industrial Members of Mobile VCE. The authors would also like to thank for the support from the Scottish Funding Council for the Joint Research Institute with the Heriot-Watt University which is a part of the Edinburgh Research Partnership.

References

[1] Y. Cao and V. Li, "Scheduling algorithms in broad-band wireless networks," *Proceedings of the IEEE*, vol. 89, no. 1, pp. 76–87, Jan. 2001.

[2] D. Avidor, S. Mukherjee, J. Ling, and C. Papadias, "On asymptotically fair transmission scheduling over fading channels with measurement delay," *IEEE Transactions on Wireless Communications*, vol. 5, no. 7, pp. 1626–1633, July 2006.

[3] P. Svedman, S. Wilson, L. Cimini, and B. Ottersten, "Opportunistic beamforming and scheduling for OFDMA systems," *IEEE Transactions on Communications*, vol. 55, no. 5, pp. 941–952, May 2007.

[4] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users," *IEEE Communications Magazine*, vol. 38, no. 7, pp. 70–77, July 2000.

[5] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels – Part 1: Ergodic capacity," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 1083–1102, Mar. 2001.

[6] D. Tse and S. Hanly, "Multiaccess fading channels – Part 1: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2796–2815, Nov. 1998.

[7] Z. Shen, J. Andrews, and B. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2726–2737, Nov. 2005.

[8] X. Liu, E. Chong, and N. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 2053–2064, Oct. 2001.

[9] M. Shaqfeh and N. Goertz, "Channel-aware scheduling with resource-sharing constraints in wireless networks," *accepted for publication in Proceedings IEEE International Conference on Communications (ICC)*, 2008.

[10] G. Song and Y. Li, "Utility-based resource allocation and scheduling in OFDM-based wireless broadband networks," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 127–134, Dec. 2005.

[11] Z. Han, X. Liu, Z. Wang, and K. Liu, "Delay sensitive scheduling schemes for heterogeneous QoS over wireless networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 2, pp. 423–428, Feb. 2007.

[12] S. Ryu, B. Ryu, H. Seo, and M. Shin, "Urgency and efficiency based packet scheduling algorithm for OFDMA wireless system," in *Proceedings IEEE International Conference on Communications (ICC)*, May 2005, vol. 4, pp. 2779–2785.

[13] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proceedings IEEE International Conference on Communications (ICC)*, Seattle, WA, USA, June 1995, pp. 331–335.

[14] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[15] B. Borst and P. Whiting, "Dynamic rate control algorithms for HDR throughput optimization," in *INFOCOM*, Apr. 2001, vol. 2, pp. 976–985.

[16] L. Yang and M. Alouini, "Performance analysis of multiuser selection diversity," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 6, pp. 1848–1861, Nov. 2006.

[17] D. Park, H. Seo, H. Kwon, and B. Lee, "Wireless packet scheuling based on the cumulative distribution function of user transmission rates," *IEEE Transactions on Communications*, vol. 53, no. 11, pp. 1919–1929, Nov. 2005.

[18] T. Nguyen and Y. Han, "A proportional fairness algorithm with QoS provision in downlink OFDMA systems," *IEEE Communications Letters*, vol. 10, no. 11, pp. 760–762, Nov. 2006.

[19] Y. Ma, "Proportional fair scheduling for downlink OFDMA," in *Proceedings IEEE International Conference on Communications (ICC)*, June 2007, pp. 4843–4848.

[20] A. Goldsmith, *Wireless Communications*, Cambridge University Press, 2005.

[21] J. G. Proakis, *Digital Communications*, McGraw-Hill International Editions, third edition, 1995.