SCHEDULING AND RESOURCE ALLOCATION IN OFDM AND FBMC SYSTEMS: AN INTERACTIVE APPROACH AND PERFORMANCE COMPARISON

Qing Bai [#], Nikos Passas [*], and Josef A. Nossek [#]

[#] Institute for Circuit Theory and Signal Processing
Technische Universität München, Germany
{bai, josef.a.nossek}@tum.de

[*] Department of Informatics & Telecommunications
University of Athens, Greece
passas@di.uoa.gr

ABSTRACT

In modern wireless communication systems, scheduling and resource allocation are two closely related tasks of the medium access control (MAC) and physical (PHY) layers respectively. In this paper, we present a separate yet interactive design of the two functional modules, where the scheduler prioritizes data packets from different service flows according to their traffic characteristics and feedback from the resource allocator, while the resource allocator performs quality of service (QoS) constrained transmit power minimization given lists of prioritized packets from the scheduler. The simulation results of applying the model to both cyclic prefix based OFDM (CP-OFDM) and the filter bank based multicarrier (FBMC) systems demonstrate the superior performance of the latter.

1. INTRODUCTION

Due to their ability to overcome frequency selective fading and support high data rates, multicarrier systems have drawn much research and industrial attention. OFDM with cyclic prefix (CP) is by far the most popular special case of multicarrier systems and has been adopted in many current applications and standards. It has an efficient implementation by using the fast Fourier Transform (FFT) and requires very simple equalization as long as the CP exceeds the delay spread of the channel impulse response. However, the CP is purely redundant in terms of information and considerably reduces the bandwidth efficiency. On the other hand, FBMC systems provide a better spectral shaping of subcarriers than OFDM systems by careful design of the prototype filter, which not only simplifies equalization in the absence of CP, but also improves the robustness of the system against a potential carrier frequency offset (CFO). By employing offset quadrature amplitude modulation (OQAM), the full capacity of the transmission bandwidth can be achieved in FBMC systems.

Resource allocation in multicarrier systems refers mainly to the allocation of time slots and frequency bands per user, in order to transmit reliably an amount of data which will maintain the required QoS level per case. A resource allocation method in wireless multiuser systems consists of two main components, namely the *scheduler* and the *resource allocator* (RA). Generally the scheduler schedules the packet transmission in time, while the RA makes the real allocation of radio resources. The two components can be jointly optimized which could be optimal in terms of performance, but with a relatively high complexity. In this paper, we propose an interactive approach to the general resource allocation problem that aim to take advantage of the good frequency selectivity of FBMC to result in considerably improved performance compared to CP-OFDM. To achieve this goal, in the *cross-layer assisted* resource allocation procedure we employ adaptive subchannel size and take into account the effects of CFO in CP-OFDM and FBMC systems, which mainly include an attenuation of the desired signal and the introduction of *intercarrier interference* (ICI) and *intersymbol interference* (ISI).

The rest of the paper is organized as follows: in Sec. 2 the system structure, the functions of each component as well as the interactions between them are explained, and a summary of the proposed interactive approach is given. Details about the design of RA and the scheduler are described in Sec. 3 and Sec. 4, respectively, followed by simulation results shown in Sec. 5. Finally Sec. 6 concludes the paper.

2. SYSTEM STRUCTURE AND INTERACTIVE APPROACH

As described earlier, the addressed scheduling and resource allocation problem at the downlink of multiuser multicarrier systems is solved by two separate, yet interactive modules, *i.e.*, the RA working across the PHY and MAC layers, and the *scheduler* which works on top of RA. Consider the scenario of an isolated cell in which U users each with one service flow are to be served by the base station (BS). Due to the different characteristics of the applications that generate the traffic, the services flows have different QoS requirements in minimum reserved data rate, delay tolerance, *etc*. Thus at the data link layer of the BS arrive different amount of packets with various latency requirements from the U users. The problem we explore now is how to serve as many packets as possible by using the available radio resources.

We assume perfect channel state information at the transmitter (CSIT). The scheduling and resource allocation procedure is performed for each *Transmission Time Interval* (TTI) during which the wireless channel is assumed to stay constant. For every TTI, the scheduler receives a number of packets passed down from higher layers. Depending on the QoS requirements and previous statistics, the scheduler decides which packets are to be served and in which priority order, and provides the prioritized list of packets to the RA. The RA then looks for the specific subchannel assignment, power allocation and modulation and coding schemes (MCS) that could serve the list of packets best, under the current channel realization. To be more specific, for each TTI of length T_I, the scheduler provides the RA with a prioritized list of K_tot packets, where packet k is in the format of

User ID u_k	Latency requirement $\tau_k^{(\mathrm{rq})}$	Length b_k

The main task of the RA is then to find the resource alloca-

978-1-4244-6001-4/10/$26.00 ©2010 IEEE

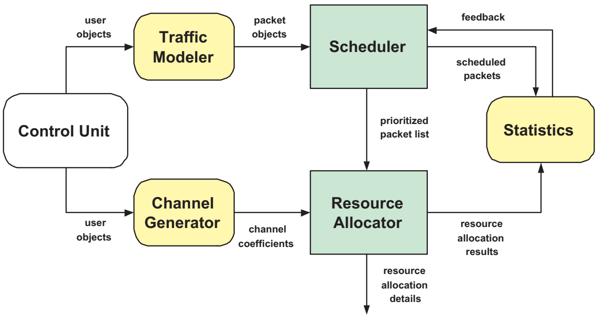


Figure 1: System Structure

tion that could serve the maximum number of packets from the list. We propose to solve this optimization by iteratively solving a series of transmit power minimization problems with given subsets of packets, and comparing the minimal resources required with the amount available. Since the packet list is prioritized by the scheduler, the generation of subsets of packets is straightforward and systematic. The basic RA algorithm is summarized in Algorithm 1, where P_{tot} denotes the total transmit power that is available in the system.

Algorithm 1 Resource Allocation Procedure by RA

Require: A prioritized packet list
Ensure: largest set of packets $\mathscr{P}_{\mathrm{S}}$ that could be served
  $\mathscr{P}_{\mathrm{S}} \leftarrow \{1,\ldots,K_{\mathrm{tot}}\}, K \leftarrow K_{\mathrm{tot}}$
  $P_{\min} \leftarrow$ minimum transmit power to serve $\mathscr{P}_{\mathrm{S}}$
  while $P_{\min} > P_{\mathrm{tot}}$ **do**
    $\mathscr{P}_{\mathrm{S}} \leftarrow \mathscr{P}_{\mathrm{S}} \setminus \{K\}, K \leftarrow K-1$
    $P_{\min} \leftarrow$ minimum transmit power to serve $\mathscr{P}_{\mathrm{S}}$
  end while
  if $K < K_{\mathrm{tot}} - 1$ **then**
    $K \leftarrow K+2$
    while $K \leq K_{\mathrm{tot}}$ **do**
      $P_{\min} \leftarrow$ minimum transmit power to serve $\mathscr{P}_{\mathrm{S}} \cup \{K\}$
      if $P_{\min} \leq P_{\mathrm{tot}}$ **then**
        $\mathscr{P}_{\mathrm{S}} \leftarrow \mathscr{P}_{\mathrm{S}} \cup \{K\}$
      end if
      $K \leftarrow K+1$
    end while
  end if

From the procedure description above, it can be seen that beside the two central components scheduler and RA, there are three auxiliary components needed in the system to make simulations and evaluations possible: a traffic modeler, a channel generator, and a statistics module. Moreover, a control unit is necessary for scenario setup and system initialization. The basic structure of the system and the interconnections between the components are shown in Fig. 1, and explained in the following.

- Control unit: creates the user objects and controls the simulation process. The properties of each user object describe the physical status (*e.g.*, distance from the BS) as well as the traffic characteristics (*e.g.*, service flow type, minimum reserved data rate, maximum latency) of the corresponding user. These user objects are assumed to be static over a large number of TTI's.
- Traffic modeler: simulates the data traffic of the users as various numbers of packets with different lengths and latency requirements, such that the generated traffic load is in accordance with each user's QoS requirement, and then passes these packet objects to the scheduler.
- Scheduler: schedules the input packets and put them into a prioritized list. The decision is informed to the statistics component.
- Channel generator: randomly generates realizations of Rayleigh fading channels of the users.
- Resource allocator: with a prioritized packet list as input for every TTI, outputs the resource allocation details such as a specific subchannel assignment and power allocation strategy, as well as the servabilities of the packets from the original list.
- Statistics: stores and processes the resource allocation results for consecutive TTI's, which helps the scheduler in providing the packet list and with the evaluation of the system performance.

3. RESOURCE ALLOCATOR DESIGN

Since the RA is responsible for iteratively solving the transmit power minimization problem associated with different packet lists, the focus of this section will be on the design of a resource allocation algorithm, which not only efficiently gives resource allocation strategies that minimize transmit power, but also takes into account the different features of specific multicarrier systems such that how the differences from a PHY aspect influence the QoS the systems are able to provide can be shown. Before we come to the algorithm part, the cross-layer system model is first introduced, which stems from [1] and lays the basis for cross-layer optimization.

3.1 System Model

The QoS parameter *latency* τ of a packet is defined as the delay it experiences until received correctly with an outage probability of no more than the predefined value $\pi^{(\mathrm{out})}$. Mathematically, let $f[m]$ be the probability that it takes exactly m TTI's to transmit a packet error-free, then $\tau = (M-1)(T_{\mathrm{R}} + T_{\mathrm{I}}) + T_{\mathrm{I}}$ where T_{R} represents *round trip delay*, and

$$M = \min_{M'} M' \quad \text{s.t.} \quad \sum_{m=1}^{M'} f[m] \geq 1 - \pi^{(\mathrm{out})}.$$

3.1.1 Channel Model

The downlink broadcast channel is modeled as frequency-selective fading over the total system bandwidth and frequency-flat fading over each subcarrier. A *subchannel* is defined as a chunk of N_{c} adjacent subcarriers, where N_{c} can be adaptive but the bandwidth of one subchannel must be smaller than the channel coherence bandwidth, so that the channel gains over one subchannel can be averaged. Assuming that one TTI contains N_{s} symbols for data transmission, we define the *minimum allocation unit* (MAU) as an allocation region of one subchannel in the frequency dimension by one TTI in the time dimension, which contains $N_{\mathrm{c}}N_{\mathrm{s}}$ symbols. This means that the assignment of one MAU is exclusive to one packet.

Let $H_{k,n}$ be the average channel coefficient of user u_k on the nth subchannel, and p_n be the power allocated on subchannel n which is equally distributed to each subcarrier in the subchannel. Also we let the white Gaussian noise power on one subcarrier at the receiver of user u_k be σ_k^2, which leads to a noise power of $N_c\sigma_k^2$ on one subchannel. When assigned to user u_k and neglecting ICI, the *signal-to-noise-ratio* (SNR) on subchannel n can be computed as

$$\gamma_{k,n} = \frac{|H_{k,n}|^2}{N_c\sigma_k^2} \cdot p_n. \tag{1}$$

In the next subsection we drop the subscripts k and n for simplicity.

3.1.2 FEC coding and modulation

We assume that modulation and coding is done on a per subchannel basis, and with reference to the WiMAX standard 8 modulation and coding schemes (MCS) are chosen as candidates, which are listed in Table 1.

Table 1: Modulation and Coding Schemes (MCS)

Index	Modulation Type	Alphabet Size A	Code Rate R	$R\log_2 A$
1	BPSK	2	1/2	0.5
2	QPSK	4	1/2	1
3	QPSK	4	3/4	1.5
4	16-QAM	16	1/2	2
5	16-QAM	16	3/4	3
6	64-QAM	64	2/3	4
7	64-QAM	64	3/4	4.5
8	64-QAM	64	5/6	5

We apply the *noisy channel coding theorem* [2] over each subchannel. Let the modulation alphabet and coding rate on the subchannel under consideration be $\mathscr{A} = \{a_1, \ldots, a_A\}$ and R respectively. The *cutoff rate* of the subchannel with SNR γ can be expressed as

$$R_0(\gamma, A) = \log_2 A - \log_2 \left[1 + \frac{2}{A}\sum_{m=1}^{A-1}\sum_{l=m+1}^{A} e^{-\frac{1}{4}|a_l - a_m|^2\gamma}\right].$$

The noisy channel coding theorem states that there always exists a block code with block length l and binary code rate $R\log_2 A \le R_0(\gamma, A)$ in bits per subchannel use, such that with maximum likelihood decoding the error probability $\tilde{\pi}$ of a code word satisfies

$$\tilde{\pi} \le 2^{-l(R_0(\gamma,A) - R\log_2 A)}.$$

In order to apply this upper bound to the extensively used turbo decoded convolutional code, quantitative investigations have been done in [1] and an expression for the *equivalent block length* is derived based on link level simulations as $n_{\text{eq}} = \beta \ln L$, where parameter β is used to adapt this model to the specifics of the employed turbo code, and L is the coded packet length. Consequently, the transmission of L bits is equivalent to the sequential transmission of L/n_{eq} blocks of length n_{eq} and has an error probability of

$$\pi = 1 - (1 - \tilde{\pi})^{\frac{L}{n_{\text{eq}}}} \le 1 - \left(1 - 2^{-n_{\text{eq}}(R_0(\gamma,A) - R\log_2 A)}\right)^{\frac{L}{n_{\text{eq}}}}.$$

3.1.3 Protocol

At the MAC layer an *automatic repeat request* (ARQ) protocol is employed. The data sequence transmitted in one MAU, which will be referred to as a *subpacket*, is used as the retransmission unit since it is independently decodable. We set a limit $\tilde{m}$ on the maximum number of transmissions and consider the case where the corrupted subpackets are simply abandoned at the receiver.

Denote the set of subchannels assigned to packet k as $\mathscr{S}_k$, and the number of information bits from packet k loaded on subchannel n as $B_{k,n}$. The completeness of the transmission of packet k requires

$$\sum_{n \in \mathscr{S}_k} B_{k,n} = b_k. \tag{2}$$

On the other hand, the latency time τ_k is determined by the largest subpacket error probability of packet k, denoted by $\pi_k = \max_{n\in\mathscr{S}_k} \pi_{k,n}$. Assuming that the subpacket error probability of a retransmitted subpacket is the same as that of its original transmission, then $f_k[m] = \pi_k^{m-1}(1-\pi_k)$ becomes a geometric series with ratio π_k. The latency time τ_k then follows from its definition.

The system parameters are summarized in Table 2, including some of their notations and the values used for simulations.

Table 2: System Parameters

Total bandwidth		10 MHz
Center frequency	f_c	2.5 GHz
FFT size	C	1024
Number of data subcarriers	N_d	
Number of subchannels	N	
Number of subcarriers per subchannel	N_c	
Transmission Time Interval (TTI)	T_I	2 ms
Number of data symbols per TTI	N_s	
Round Trip Delay (RTD)	T_R	10 ms
Maximum number of transmissions allowed	$\tilde{m}$	5
Turbo code dependent parameter	β	32
Outage probability	$\pi^{(\text{out})}$	0.01

3.2 Resource Allocation Algorithm

Due to the exclusive assignment of subchannels and the discrete MCS levels that are available, the transmit power minimization problem is combinatorial and computationally intractable when the number of subchannels and the number of packets are large. Therefore we proposed a suboptimal algorithm of low complexity based on the methods proposed in [4], which first looks for the optimal subchannel assignment with fixed MCS on each subchannel, and then chooses the MCS combination that leads to the minimum transmit power required with the obtained subchannel assignment. The algorithm allows for an adjustment phase in which the subchannel assignment could be amended.

3.2.1 A Three-step Approach

We formulate the transmit power minimization problem for a given packet list of length K as

$$\begin{aligned} &\min_{B\in\mathscr{B}} \quad \sum_{k=1}^{K}\sum_{n=1}^{N} \varphi_{k,n}(B_{k,n},\tau_k^{(\mathrm{rq})}) \\ &\text{s.t.} \quad \sum_{n=1}^{N} B_{k,n} = b_k, \quad k=1,\dots,K, \end{aligned} \tag{3}$$

where $B\in\mathbb{Z}_{+,0}^{K\times N}$ represents the bit-loading matrix which is an element of set $\mathscr{B}\subset\mathbb{Z}_{+,0}^{K\times N}$, which represents the set of matrices that have only one nonzero entry in each of their columns. The complete transmission of each packet is guaranteed by the equality constraint, and the latency requirements are contained in function $\varphi(B,\tau^{(\mathrm{rq})})$, which is defined as the minimum power needed for the successful transmission of B bits within latency time $\tau^{(\mathrm{rq})}$, *i.e.*,

$$\varphi(B,\tau^{(\mathrm{rq})}) \triangleq \min_{(A,R)\in\mathscr{M}} \left\lceil \frac{s}{N_s} \right\rceil \cdot \gamma(B,A,R)\cdot\frac{\sigma^2}{|H|^2}, \tag{4}$$

where $\gamma(B,A,R)$ is the SNR required to convey B bits with maximal $\left\lfloor \frac{\tau^{(\mathrm{rq})}-T_\mathrm{I}}{T_\mathrm{R}+T_\mathrm{I}} + 1 \right\rfloor$ transmissions when MCS (A,R) is employed, which can be obtained from a binary search on the cutoff rate curve, and $s=\left\lceil \frac{B}{R\log_2 A}\right\rceil$ is the number of symbols occupied in the MAU. Note that the minimization in (4) is independent of the channel realization, therefore the optimal MCS corresponding to each B and each maximal number of transmissions can be computed offline.

Subchannel Assignment (SA) In this step we assume the same MCS is used on every subchannel. A power matrix $P\in\mathbb{R}_+^{K\times N}$ can be computed, with its entry $p_{k,n}$ being the minimum power needed to achieve the required PEP of packet k on subchannel n. First it should be guaranteed that the minimum number of subchannels required, *i.e.*, $N_k^{(\mathrm{l})}=\left\lceil \frac{b_k}{5N_\mathrm{s}N_\mathrm{c}}\right\rceil$, is assigned to each packet. If $\sum_k N_k^{(\mathrm{l})}<N$, then the remaining subchannels can be assigned proportionally to the packets according to their lengths, as long as no packet takes more than $N_k^{(\mathrm{u})}=\left\lceil \frac{b_k}{0.5N_\mathrm{s}N_\mathrm{c}}\right\rceil$ subchannels, which is the maximum number packet k could possibly use. The SA problem is formulated as picking from each column of P one entry such that the kth row has the determined number of picked entries, and the sum of all picked entries is the minimum. This problem can be solved by relaxing the integer constraints and then employing linear programming.

Bit and Power Allocation (BPA) With the SA result $\{\mathscr{S}_k : k=1,\dots,K\}$ as input, bit and power allocation is no longer coupled among the users, *i.e.*, we have K decoupled minimization problems as

$$\min_{\{B_{k,n}:n\in\mathscr{S}_k\}} \sum_{n\in\mathscr{S}_k} \varphi_{k,n}(B_{k,n},\tau_k^{(\mathrm{rq})}) \quad \text{s.t.} \quad \sum_{n\in\mathscr{S}_k} B_{k,n}=b_k,$$

which can be solved by using Lagrangian dual methods. Since the objective function is nonconvex, the dual optimal solution is usually not the primal optimal, or even primal feasible. Let the dual optimal bit-loading be $\{B_{k,n}^* : n\in\mathscr{S}_k\}$. If $\sum_{n\in\mathscr{S}_k} B_{k,n}^*\neq b_k$, we can load or unload the extra bits one by one on the subchannel that leads to the minimum power increment or the maximum power decrement, until $\sum_{n\in\mathscr{S}_k} B_{k,n}^* = b_k$ is satisfied.

Adjustments The outcome of PA might indicate zero MCS on some subchannels, which means these subchannels are released from occupation and can be assigned to other packets. As higher MCS are much more power consuming than lower MCS, we find the subchannels using the relatively highest MCS as well as their possessors, and compare each alternative of assigning the empty subchannels to these packets.

3.2.2 Adaptive subchannel size

As one subchannel can be assigned to at most one packet, the number of subchannels is practically a hard limit on the number of packets that the system could serve. However, the MAU's may be under utilized if the loaded packets are small. In other words, to fix the number of subchannels, or equivalently, to fix the size of each subchannel, is in general inefficient. In order to better adapt to diverse traffic situations resulting in different number and different lengths of input packets, we make the number of subcarriers that make up one subchannel, *i.e.*, N_c, also an optimization variable.

Beside the possibility to serve more packets, the advantages of a small subchannel size also include having a finer frequency granularity and benefiting from multiuser diversity. On the other hand, a large subchannel size leads to more data symbols in one MAU which potentially provides a larger coding gain, as well as reduces the computations required to find a suboptimal resource allocation strategy. Based on these analysis, the optimal N_c should depend on the frequency selectivity of the channel and the input packet list. Note that in addition to the computational effort to find the optimal N_c, the performance gain of having N_c adaptive comes also at the cost of an extra overhead to inform the receiver the value of N_c.

Since the channel should be assumed constant over each MAU, the bandwidth of one subchannel can not exceed the coherence bandwidth of the multipath channel, *i.e.*, $\Delta f\cdot N_\mathrm{c}\leq B^{(\mathrm{coh})}$, where Δf and $B^{(\mathrm{coh})}$ denote the subcarrier spacing and the coherence bandwidth of the channel respectively. Hence $\lceil\frac{B^{(\mathrm{coh})}}{\Delta f}\rceil$ provides an upper bound for N_c. On the other hand, it is usually unnecessary and impractical to have a very high frequency resolution which requires more iterations to find the optimum N_c. As a result, an appropriate interval for two consecutive candidate N_c values should be set, which mainly depends on the ratio between pilot subcarriers and data subcarriers. Denote this interval as $N^{(\mathrm{I})}$, and determine the largest and smallest candidate values for N_c as $N_\mathrm{c}^{(\mathrm{u})}$ and $N_\mathrm{c}^{(\mathrm{l})}$. The search for optimum N_c for a given packet list is summarized in Algorithm 2.

Algorithm 2 Search for optimum N_c given a packet list

Initialize $N_\mathrm{c}\leftarrow N_\mathrm{c}^{(\mathrm{u})}$, $P_{\min}\leftarrow\infty$
$P\leftarrow$ minimum transmit power to serve the list given N_c
while $P<P_{\min}$ and $N_\mathrm{c}>N_\mathrm{c}^{(\mathrm{l})}$ **do**
  $P_{\min}\leftarrow P,\quad N_\mathrm{c}\leftarrow N_\mathrm{c}-N^{(\mathrm{I})}$
  $P\leftarrow$ minimum transmit power to serve the list given N_c
end while

3.2.3 Compensation for interference induced by CFO

Being sensitive to CFO is known to be one of the main drawbacks of multicarrier systems, which is mainly caused by Doppler shift due to mobility and the inherent difference between the oscillators at the transmitter and receiver. The effects of CFO on CP-OFDM systems have been extensively studied in the literature, *e.g.*, [3]. In our recent work [5], we have analyzed the effects of CFO on FBMC systems, where the *root-raised cosine* (RRC) filter with roll-off factor 1 is chosen to be the prototype filter. In this section we first summarize the results on CFO induced interference in both systems, and then discuss the method to compensate for this degradation in the resource allocation procedure.

Since we are investigating the interference situation at one receiver in the downlink, the packet index k is dropped in this section. Also, we use subscript c to index the subcarriers, which should be distinguished from subchannel index n. For both systems, we denote the normalized CFO as ε which is the CFO between the transmitter and the receiver with respect to the subcarrier spacing, *i.e.*, $\varepsilon \triangleq \frac{f_T - f_R}{\Delta f}$, where f_T and f_R are the carrier frequencies of the transmitter and receiver respectively, and Δf represents the subcarrier spacing. In addition, we assume that there is also a phase offset ϕ between the transmitter and the receiver, and the system is perfectly synchronized in the time domain. Furthermore, we restrict ε to be in the range $(-0.5, 0.5]$, as the integer part of the frequency offset does not affect the *signal-to-interference-plus-noise ratio* (SINR).

SINR expressions For coherent demodulation at the receiver, the phase rotation should be estimated and we assume that it is perfectly compensated. Moreover, we assume that the data symbols transmitted, whether on different subcarriers or at different times, are all statistically independent from each other. Let the power allocated on subcarrier c be p_c. The SINR on subcarrier c in CP-OFDM system can be expressed as

$$\mathrm{SINR}_c(\varepsilon) = \frac{|v_C(\varepsilon,0)|^2 |H_c|^2 p_c}{\sum\limits_{\substack{c'=0 \\ c' \neq c}}^{C-1} |v_C(\varepsilon, c'-c)|^2 |H_{c'}|^2 p_{c'} + \sigma^2}, \quad (5)$$

where $v_C(\varepsilon, c) = \frac{1}{C} \frac{\sin(\pi(\varepsilon - c))}{\sin \frac{\pi(\varepsilon - c)}{C}} e^{j\pi \frac{(\varepsilon - c)(C-1)}{C}}$.

On the other hand, due to the infinite impulse response of the prototype filter used in the FBMC system, the CFO causes not only ICI but also ISI to the desired signal. The SINR of the lth symbol on subcarrier c in FBMC system is expressed as

$$\mathrm{SINR}_{c,l}(\varepsilon) = \frac{\alpha_c^2(\varepsilon,0,0) p_c}{\sum\limits_{c'=0}^{C-1} \sum\limits_{l'=-\infty}^{+\infty} \alpha_{c'}^2(\varepsilon, \Delta c, \Delta l) p_{c'} - \alpha_c^2(\varepsilon,0,0) p_c + \frac{\sigma^2}{|H_c|^2}}, \quad (6)$$

where $\Delta c = c - c'$, $\Delta l = l - l'$, and

$$\alpha_c(\varepsilon, \Delta c, \Delta l) \triangleq \Re\left\{ e^{-j\frac{\pi}{2}(\Delta c + \Delta l)} e^{j\pi\Delta l(c-\varepsilon)} w(\varepsilon, \Delta c, \Delta l) \right\},$$

$$w(\varepsilon, \Delta c, \Delta l) \triangleq \int_{-\infty}^{+\infty} e^{j\pi\Delta l T f} H_{\mathrm{RRC}}\left(f - \frac{\varepsilon - \Delta c}{T}\right) H_{\mathrm{RRC}}(f)\,\mathrm{d}f,$$

where H_{RRC} denotes the frequency response of the prototype filter. Here it is assumed that the power allocation stays constant in time, which is reasonable because the weight factor $\alpha(\varepsilon, \Delta c, \Delta l)$ approaches 0 very fast with increasing Δl, which means the influence of symbols that are not close to the one of interest is negligible.

Compensation of residual CFO With the aid of pilot symbols, the receiver estimates and compensates for the CFO. However, this compensation could be imperfect even in the downlink, and the system should be able to live with the residual CFO. From the resource allocation point of view, what is obtained from subchannel assignment and bit allocation is now N required SINR values instead of N required SNR, which has a general form of

$$\gamma_c^{(\mathrm{rq})} = \frac{a_{c,c} p_c}{\sum_{c' \neq c} a_{c,c'} p_{c'} + \sigma^2},$$

where $a_{c,c'}$ are nonnegative scalers. Stacking all C equations we have a set of linear equations the solution of which gives the power allocation that is able to achieve all required SINR values on every subcarrier, *i.e.*,

$$p = \begin{bmatrix} a_{0,0} & -a_{0,1} & \cdots & -a_{0,C-1} \\ -a_{1,0} & a_{1,1} & \cdots & -a_{1,C-1} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{C-1,0} & -a_{C-1,1} & \cdots & a_{C-1,C-1} \end{bmatrix}^{-1} \cdot \sigma^2 \begin{bmatrix} \gamma_0^{(\mathrm{rq})} \\ \gamma_1^{(\mathrm{rq})} \\ \vdots \\ \gamma_{C-1}^{(\mathrm{rq})} \end{bmatrix}.$$

Note that at the transmitter, the worst case residual CFO is assumed for compensation, yet in the SA and BPA phase the impact of a potential residual CFO is neglected.

4. SCHEDULER DESIGN

The scheduler used in the present study assumes a WiMAX network, although it can be easily adapted to any similar cellular wireless network with differentiated services based on traffic classes. The system architecture of WiMAX consists of Base Stations (BSs), each one responsible for a specific cell area, and stationary Subscriber Stations (SSs). The communication path between SSs and the BS is divided into two directions: uplink (from SS to BS) and downlink (from BS to SS), multiplexed either with Time Division Duplex (TDD) or Frequency Division Duplex (FDD). Transmission parameters, including the modulation and coding schemes, may be adjusted individually for each SS on a frame-by-frame basis. A TDD frame has a fixed duration and is divided into a downlink subframe, and an uplink subframe. Each connection is associated with a single service flow and specifies a set of parameters that quantify its traffic behavior and QoS expectations. This set includes:

- minimum reserved traffic rate (in bits/sec),
- maximum sustained traffic rate (in bits/sec),
- maximum latency (in ms),
- tolerated jitter (maximum delay variation in ms),
- traffic priority (values 0-7, with 7 the highest), *etc.*

The respective IEEE 802.16 standard [6] defines four different services:

- *Unsolicited Grant Service* (UGS): This service supports real-time data streams consisting of fixed-size data packets transmitted at periodic intervals, such as Voice over IP without silence suppression. These applications require constant data rate allocation, so data rate requests are not required.

- *Real-time Polling Service* (rtPS): This service supports data streams consisting of variable-sized data packets that are transmitted at fixed intervals, such as MPEG video. These applications have specific data rate requirements, as well as a maximum acceptable latency. Late packets that miss the deadline are considered useless.
- *Non-real-time Polling Service* (nrtPS): This service is for non-real-time connections that require better than best effort service, *e.g.*, data rate intensive file transfer. These applications are time-insensitive but require a minimum data rate allocation.
- *Best Effort service (BE)*: This service is for best effort traffic with no QoS guarantee. The applications of this kind of service share the remaining resources after allocation to the rest of the services is completed. BE uses only contention mode.

In [7], a new service, referred to as *enhanced rtPS* (ertPS), is defined to better support real-time service flows that generate variable size data packets on a periodic basis, *e.g.*, VoIP with silence suppression.

The traffic scheduler located at the BS decides on the service order of packets from all active connections. Uplink scheduling is performed by the BS with the aim of providing each SS with enough data rate for uplink transmissions or opportunities for extra transmission requests. When additional data rate is needed, the SS utilizes its transmission opportunities during contention periods or when it is polled by the BS, depending on its agreed QoS characteristics, to pass its transmission requests. Downlink scheduling on the other hand, considers packets waiting for transmission at the BS as implicit requests for data rate allocation. Based on well-accepted studies, the scheduler has to combine the following properties:

- **Fast Data Scheduling**: The MAC scheduler must efficiently allocate available resources in response to bursty data traffic and time-varying channel conditions. The scheduler is located at each BS to enable rapid response to traffic requirements and channel conditions. The data packets are associated to traffic connections with well defined QoS parameters in the MAC layer so that the scheduler can correctly determine the packet transmission ordering over the air interface.
- **Scheduling for both downlink and uplink**: The scheduling service is provided for both downlink and uplink traffic. In order for the MAC scheduler to make an efficient resource allocation and provide the desired QoS in the uplink, the uplink must feedback accurate and timely information as to the traffic conditions and QoS requirements. Multiple uplink data rate request mechanisms, such as data rate request through ranging channel, piggyback request and polling are designed to support UL data rate requests. The uplink service flow defines the feedback mechanism for each uplink connection to ensure predictable uplink scheduler behavior. Furthermore, with orthogonal uplink sub-channels, there is no intra-cell interference. Uplink scheduling can allocate resource more efficiently and better enforce QoS.
- **Dynamic Resource Allocation**: The MAC supports frequency-time resource allocation in both downlink and uplink on a per-frame basis. The resource allocation is carried out at the beginning of each frame and therefore can be changed frame-by-frame in response to traffic and channel conditions. Additionally, the amount of resource in each allocation can range from one slot to the entire frame. The fast and fine granular resource allocation allows superior QoS for data traffic.
- **QoS Oriented**: The MAC scheduler handles data transport on a connection-by-connection basis. Each connection is associated with a single data service with a set of QoS parameters that quantify the aspects of its behavior. With the ability to dynamically allocate resources in both downlink and uplink, the scheduler can provide superior QoS for traffic in both directions. Particularly with uplink scheduling - the uplink resource is more efficiently allocated, performance is more predictable, and QoS is better enforced.

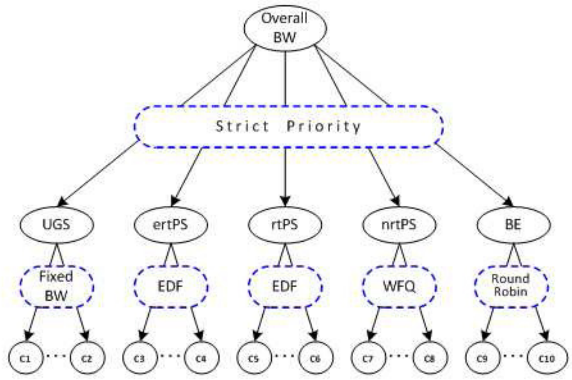


Figure 2: Services and priorities

In addition to whatever other factors the scheduler may consider, the following items can be taken into account for each active connection:

- the scheduling service specified for the connection (i.e., type of connection),
- the values assigned to the connection's traffic and QoS parameters,
- the availability of data for transmission (queue size).

A specific scheduling algorithm is not described in the IEEE 802.16 standard, because it is not included among the mandatory modules required for the standardized system's operation. On the other hand, the operation of the scheduler is important for the performance of the whole system, and this is why it attracts growing attention over the last couple of years. To efficiently support all types of connections (UGS, rtPS, ertPS, nrtPS and BE) as specified in the standard, the scheduler used in this study is based on ideas found in [8] and uses a combination of strict priority service discipline, earliest deadline first (EDF) and weight fair queue (WFQ) algorithms. The hierarchical structure of the bandwidth allocation is shown in Fig. 2.

The basic scheduling principles of the algorithms are as follows:

1. Overall data rate allocation: The data rate allocation per

traffic class follows strict priority, from highest to lowest: UGS, ertPS, rtPS, nrtPS and BE. One disadvantage of the strict priority service is that higher priority connections may lead lower priority connections to data rate starvation. To overcome this problem, a traffic policing module is included in each terminal, which forces the connections' data rate demands to stay within the traffic contract, as agreed during connection setup. This prevents the higher priority connections from using data rates more than their allocation, and allows for fair treatment of all traffic.

2. Data rate allocation for UGS connections: The scheduler allocates fixed data rates to UGS connections based on their fixed requirements. This policy is determined clearly by the IEEE 802.16 standard, without the need for real-time transmission requests.
3. Data rate allocation for ertPS and rtPS connections: The EDF service is adopted for these connections, to allow packets with the earliest deadline to be scheduled first. In case two packets belonging to two different service types (one of ertPS and one of rtPS) expire at exactly the same time, the scheduler will give priority to the ertPS packet, considering this packet of higher priority. Data rate needs are constantly updated through real-time transmission requests.
4. Data rate allocation for nrtPS connections: The weighted fair queue (WFQ) service is applied for this traffic class. For each nrtPS connection, the ratio of its average data to the total nrtPS average data rates is computed, and resources being left from the higher priority classes (UGS,ertPS and rtPS) are distributed according to the computed weights of the connections. No transmission requests are required on this case.
5. Data rate allocation for BE connections: The remaining resources are equally allocated among BE connections following the Round Robin model, without transmission requests.

The scheduler described above combines both simplicity and efficiency, since it can be easily implemented without the need for complex calculations, while it can provide service differentiation and QoS guarantees to all traffic classes. Simplicity was a critical requirement in our case, as the algorithm has to operate in real-time on a frame-by-frame basis. Nevertheless, it is expected that this will not sacrifice the algorithm's capability to operate under different traffic and channel conditions. Moreover, the scheduler can take advantage of the improved performance of FBMC compared to OFDM, by fairly supporting a larger number of connections.

5. SIMULATION RESULTS

Due to the less out-of-band emission it generates and the absence of CP, the FBMC system can employ more subcarriers (5% more than CP-OFDM is assumed) and more symbols per TTI (12.5% more than CP-OFDM is assumed which is a common fraction of CP to the number of data samples) for data transmission. The difference in the two systems are listed in Tab. 3. Moreover, we model the wireless channel as a frequency-selective fading channel consisting of 6 independent Rayleigh multipaths with an exponentially decaying power profile where the delay spread is 1 μs. The path loss in dB is computed as $PL(d) = 140.6 + 35.0 \log_{10} d$ following the COST-Hata model, where d is the distance between MS and BS in km, and all receiver noise levels are -174 dBm/Hz.

Table 3: System Parameters

Number of data subcarriers in CP-OFDM	N_d	720
Number of data subcarriers in FBMC	N_d	756
Number of data symbols per TTI in CP-OFDM	N_s	16
Number of data symbols per TTI in FBMC	N_s	18

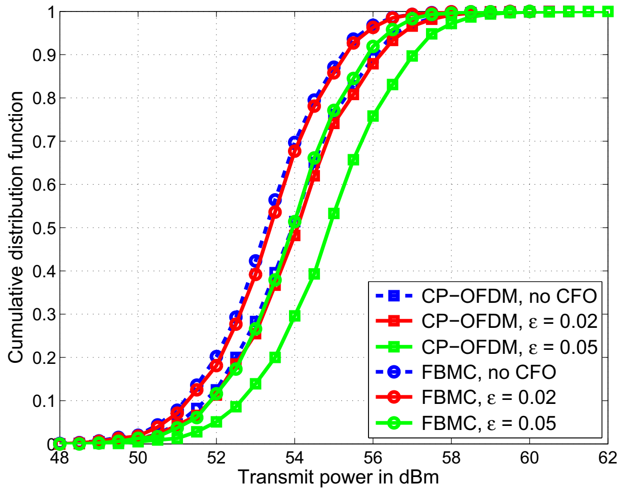
Firstly, to test the RA and compare the performances of CP-OFDM and FBMC systems, we design two lists of packets, one list of 40 small packets and the other of 20 larger packets with details given in Tab. 4, and use them as input to the RA. We assume that each packet comes from a different user, where the users are uniformly located in a cell of radius 2 km. The case that there are users with more than one packet to transmit can be easily accommodated.

Table 4: List of small and large packets

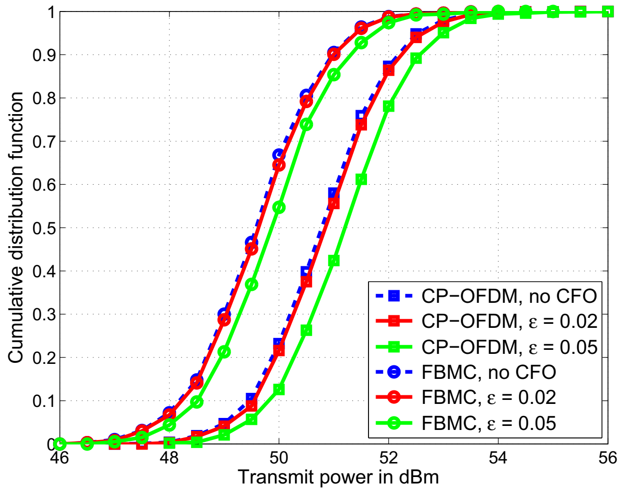
Packet k	b_k / bytes	$\tau_k^{(\mathrm{rq})}$ / ms	Packet k	b_k / bytes	$\tau_k^{(\mathrm{rq})}$ / ms
$1-20$	64	20	$1-10$	128	20
$21-40$	128	40	$11-20$	384	40

The two figures in Fig. 3 show the cumulative distributions of the minimum transmit power required in CP-OFDM and FBMC systems to serve the lists of large and small packets respectively, where two maximum residual CFO values $\varepsilon = 0.02$ and $\varepsilon = 0.05$ have been tested and the curves with no residual CFO are also drawn as a reference. On average, for the list of small packets, the FBMC system requires around 1.2 dBm less transmit power than the CP-OFDM system in the perfectly synchronized case, whereas for the list of large packets, the difference is 0.65 dBm. To compensate for the two residual CFO, the CP-OFDM system requires 0.0649 dBm and 0.4207 dBm more transmit power for the list of small packets, 0.1282 dBm and 0.9615 dBm more for the list of large packets, while the FBMC system requires 0.0427 dBm and 0.2863 dBm more transmit power for the list of small packets, 0.0903 dBm and 0.6576 dBm more for the list of large packets, respectively. This demonstrates from a QoS provisioning and resource allocation point of view, that the FBMC system is less sensitive to CFO and benefits more from multiuser diversity, and its advantage is even larger when the CFO or the number of packets increases.

To reveal the effectiveness of the proposed scheduling and resource allocation procedure in terms of differentiated QoS, we execute a scenario involving an increasing number of users, each one with one active connection per service type. Fig. 4 shows the input and output data rates in the system for the CP-OFDM case. For up to 6 users, the system manages to service all incoming traffic with no losses. After this point, BE, that is the type with the lower priority, starts facing denial of service, to allow transmission of higher priority traffic. As the load increases, nrtPS output data rate is reduced (after 8 users), leaving most of the capacity for the real-time service types. At the end of the simulation (18 users), the load is increased up to a point that even rtPS traffic starts facing losses. Fig. 5 shows the attained throughput per service type as a percentage of the offered traffic for



(a) With the list of large packets



(b) With the list of small packets

Figure 3: CDF of minimum transmit power in CP-OFDM and FBMC systems with different residual CFO

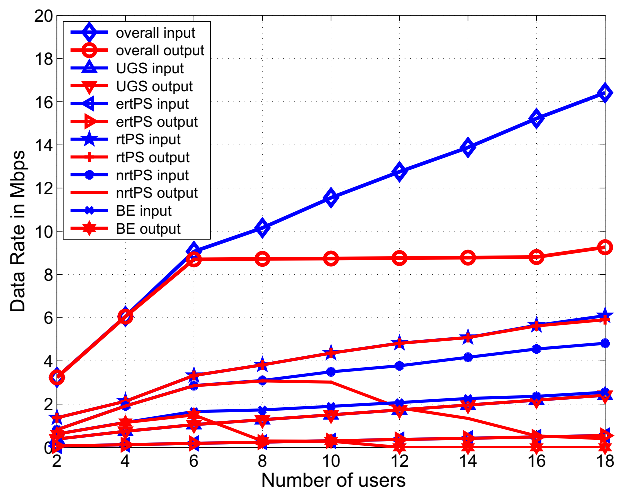


Figure 4: CP-OFDM input and output data rates

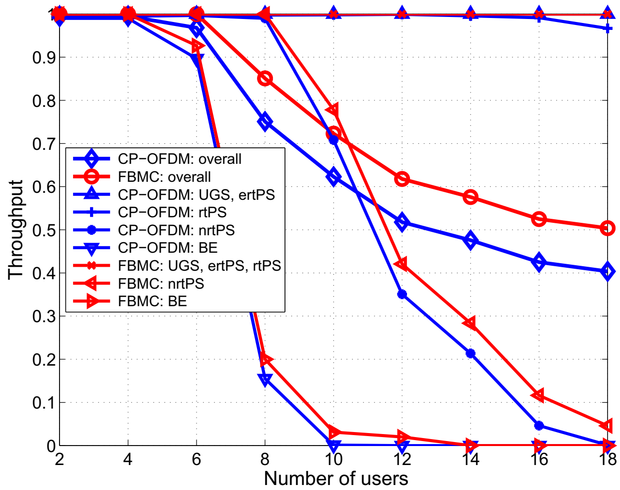


Figure 5: Throughput in CP-OFDM and FBMC systems

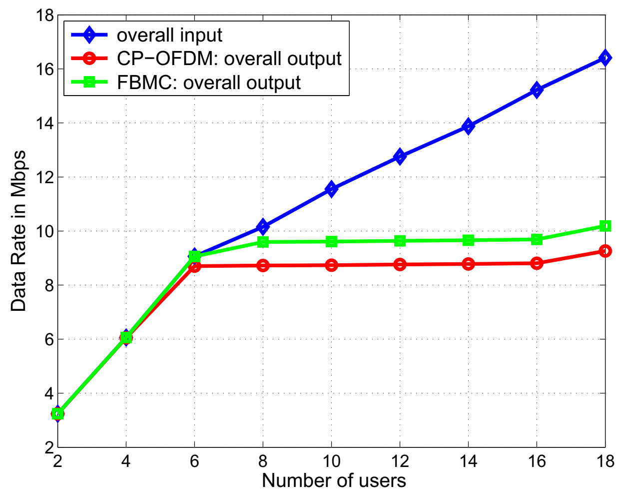


Figure 6: Input and output data rates in CP-OFDM and FBMC systems

both CP-OFDM and FBMC. Again, the differentiated treatment is clearly revealed forcing BE connections to reduce their throughput first, followed by nrtPS and rtPS. An overall throughput improvement of approximately up to 18% is attained with FBMC, compared to CP-OFDM, as a result of the improved operation of the physical layer. Finally, the comparative performance of FBMC and CP-OFDM in terms of overall data rates is shown in Fig. 6. The effectiveness of FBMC is indicated by an almost stable increase of the data rate for most of the cases.

6. CONCLUSIONS

In this paper the scheduling and resource allocation problem in the downlink of multicarrier systems is addressed. An interactive approach has been proposed where the scheduler prioritizes input packets from heterogeneous service flows, and the RA performs the allocation of subchannel, bit and power given the prioritized packet lists by cross-layer optimization. As FBMC systems have better frequency containment and could employ more subcarriers and symbols to transmit data compared to CP-OFDM systems, they are able to fulfill the QoS requirements of more users and exhibit a higher performance limit, which have been verified and evaluated with simulation results.

Acknowledgment

The authors wish to acknowledge the financial support from the 7th European Framework Programme under the PHYDYAS Project (contract no. 211887).

REFERENCES

[1] B. Zerlin, M. T. Ivrlač, W. Utschick, J. A. Nossek, I. Viering and A. Klein, *Joint optimization of radio parameters in HSDPA*, in IEEE 61st Vehicular Technology Conference VTC 2005-Spring, vol. 1, pp. 295-299.

[2] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley and Son, 1968.

[3] J. Lee, H. Lou, D. Toumpakaris and J. M. Cioffi, *Effect of carrier frequency offset on OFDM systems for multipath fading channels*, in IEEE Global Telecommunications Conference, Dec. 2004, vol. 6, pp. 3721-3725.

[4] Q. Bai and J. A. Nossek, *On Two Cross-layer Assisted Resource Allocation Schemes in Multiuser Multicarrier Systems*, Wireless Communication Systems, 6th International Symposium on, Sept. 2009, pp. 186-190.

[5] Q. Bai and J. A. Nossek, *On the Effects of Carrier Frequency Offset on Cyclic Prefix Based OFDM and Filter Bank Based Multicarrier Systems*, submitted to 2010 IEEE International Workshop on Signal Processing Advances for Wireless Communications.

[6] IEEE Std 802.16-2004, *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Access Systems*, October 2004.

[7] IEEE Std 802.16e, *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1*, Feb. 2006.

[8] K. Wongthavarawat and A. Ganz, *Packet Scheduling for QoS Support in IEEE 802.16 Broadband Wireless Access Systems*, International Journal of Communication Systems, vol. 16, pp. 81-96, 2003.