

# Downlink Resource Scheduling in an LTE System

Raymond Kwan<sup>†</sup>, Cyril Leung<sup>††</sup>, and Jie Zhang<sup>†</sup>

<sup>†</sup>*Centre for Wireless Network Designs (CWIND), University of Bedfordshire, Luton, UK*

<sup>††</sup>*Dept. of Electrical & Computer Engineering, The University of British Columbia,  
Vancouver, B.C., Canada*

*Emails: raymond.kwan@beds.ac.uk, cleung@ece.ubc.ca, Jie.zhang@beds.ac.uk*

## Abstract

The problem of allocating resources to multiple users on the downlink of a Long Term Evolution (LTE) cellular communication system is discussed. An optimal (maximum throughput) multiuser scheduler is proposed and its performance is evaluated. Numerical results show that the system performance improves with increasing correlation among OFDMA sub-carriers. It is found that a limited amount of feedback information can provide a relatively good performance. A sub-optimal scheduler with a lower computational complexity is also proposed, and shown to provide good performance. The sub-optimal scheme is especially attractive when the number of users is large, as the complexity of the optimal scheme may then be unacceptably high in many practical situations. The performance of a scheduler which addresses fairness among users is also presented.

## 1. Introduction

Orthogonal Frequency Division Multiplexing (OFDM) is an attractive modulation technique that is used in a variety of communication systems such as Digital Subscriber Lines (DSLs), Wireless Local Area Networks (WLANs), Worldwide Interoperability for Microwave Access (WiMAX) Andrews et al. (2007), and Long Term Evolution (LTE) cellular networks. In order to exploit multiuser diversity and to provide greater flexibility in resource allocation (scheduling), Orthogonal Frequency Division Multiple Access (OFDMA), which allows multiple users to simultaneously share the OFDM sub-carriers, can be employed. The problem of power and sub-carrier allocation in OFDMA systems has been extensively studied, e.g. Liu & Li (2005); Wunder et al. (2008), and the references therein.

What distinguishes packet scheduling in LTE from that in earlier radio access technologies, such as High Speed Downlink Packet Access (HSDPA), is that LTE schedules resources for users in both the time domain (TD) and the frequency domain (FD) whereas HSDPA only involves the time domain. This additional flexibility has been shown to provide throughput and coverage gains of around 40% Pokhariyal et al. (2006). Because packet scheduling for LTE involves scheduling users in both TD and FD, various TD and FD schemes have been proposed in Pokhariyal et al. (2006)-Monghal et al. (2008). Assume that we have packets for  $N_{users}$  users waiting in the queue and that resources can only be allocated at the beginning of a pre-defined time period known as the Transmission Time Interval (TTI) or scheduling period. In TD scheduling,  $U$  users from the total of  $N_{users}$  users are selected based on some priority metric. After the  $U$  users have been selected, appropriate subcarrier frequencies and

modulation and coding schemes (MCSs) are then assigned by the FD scheduler. Note that the metrics used for TD and FD scheduling can be different in order to provide a greater degree of design flexibility. Examples of TD/FD scheduling metrics have been proposed in Kela et al. (2008); Monghal et al. (2008).

In order to make good scheduling decisions, a scheduler should be aware of channel quality in the time domain as well as the frequency domain. Ideally, the scheduler should have knowledge of the channel quality for each sub-carrier and each user. In practice, due to limited signalling resources, sub-carriers in an OFDMA system are often allocated in groups. On the downlink in LTE systems, sub-carriers are grouped into resource blocks (RBs) of 12 adjacent sub-carriers with an inter sub-carrier spacing of 15 kHz Dahlman et al. (2008); *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8)* (2007). Each RB has a time slot duration of 0.5 ms, which corresponds to 6 or 7 OFDM symbols depending on whether an extended or normal cyclic prefix is used. The smallest resource unit that a scheduler can assign to a user is a Scheduling Block (SB), which consists of two consecutive RBs, spanning a sub-frame time duration of 1 ms Dahlman et al. (2008); *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8)* (2007) (see Fig. 1). From the perspective of downlink scheduling, the channel quality is reported by the user via a Channel Quality Indicator (CQI) over the uplink. If a single CQI value is used to convey the channel quality over a large number of SBs, the scheduler may not be able to distinguish the quality variations within the reported range of subcarriers. This is a severe problem for highly frequency-selective channels. On the other hand, if a CQI value is used to represent each SB, many CQI values may need to be reported back, resulting in a high signalling overhead. A number of CQI reporting schemes and associated trade-offs are discussed in Kolehmainen (2008).

Given a set of CQI values from different users, the multiuser scheduling problem in LTE involves the allocation of SBs to a subset of users in such a way as to maximize some objective function, e.g. overall system throughput or other fairness-sensitive metrics. The identities of the assigned SBs and the MCSs are then conveyed to the users via a downlink control channel. Studies on LTE-related scheduling have been reported in Kwan et al. (2008); Liu et al. (2007); Ning et al. (2006); Pedersen et al. (2007); Pokhariyal et al. (2007) and the references therein.

As pointed out in Jiang et al. (2007), the type of traffic plays an important role in how scheduling should be done. For example, Voice-over IP (VoIP) users are active only half of the time. Also, the size of VoIP packets is small, and the corresponding inter-arrival time is fairly constant. While dynamic scheduling based on frequent downlink transmit format signalling and uplink CQI feedback can exploit user channel diversity in both frequency and time domains, it requires a large signalling overhead. This overhead consumes time-frequency resources, thereby reducing the system capacity. In order to lower signalling overhead for VoIP-type traffic, *persistent scheduling* has been proposed *Discussion on Control Signalling for Persistent Scheduling of VoIP* (2006); *Persistent Scheduling in E-UTRA* (2007). The idea behind persistent scheduling is to pre-allocate a sequence of frequency-time resources with a fixed MCS to a VoIP user at the beginning of a specified period. This allocation remains valid until the user receives another allocation due to a change in channel quality or an expiration of a timer. The main disadvantage of such a scheme is the lack of flexibility in the time domain. This shortcoming has led to semi-persistent scheduling which represents a compromise between rigid persistent scheduling on the one hand, and fully flexible dynamic scheduling on the other. In semi-persistent scheduling, initial transmissions are persistently scheduled so as to

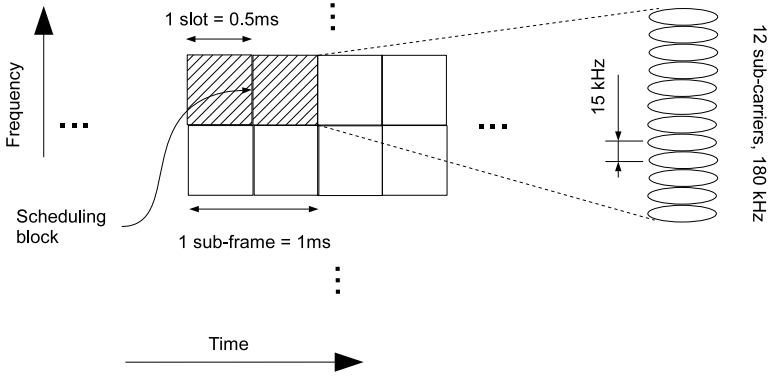


Fig. 1. LTE downlink time-frequency domain structure.

reduce signalling overhead and retransmissions are dynamically scheduled so as to provide adaptability. The benefits of semi-persistent scheduling are described in Jiang et al. (2007).

An important constraint in LTE downlink scheduling is that all SBs for a given user need to use the same MCS in any given TTI<sup>1</sup> Dahlman et al. (2008). In the rest of this chapter, we focus on the FD aspect of dynamic scheduling. Specifically, the challenging problem of multiuser FD scheduling is formulated as an optimization problem, taking into account this MCS restriction. Simpler sub-optimal solutions are also discussed.

## 2. System Model

The system model we will use to study the resource allocation problem is now described<sup>2</sup>. An SB consists of a number,  $N_{sb}$ , of consecutive OFDM symbols. Let  $L$  be the total number of sub-carriers and  $L_d(v) \leq L$  be the number of data-carrying sub-carriers for symbol  $v$ , where  $v = 1, 2, \dots, N_{sb}$ . Also, let  $R_j^{(c)}$  be the code rate associated with the MCS  $j \in \{1, 2, \dots, J\}$ ,  $M_j$  be the constellation size of the MCS  $j$  and  $T_s$  be the OFDM symbol duration. Then, the bit rate,  $r_j$ , that corresponds to a single SB is given by

$$r_j = \frac{R_j^{(c)} \log_2(M_j)}{T_s N_{sb}} \sum_{v=1}^{N_{sb}} L_d(v). \quad (1)$$

Let  $U$  be the number of simultaneous users, and  $N_{tot}$  be the total number of SBs that are available during each TTI. In addition, let  $\mathcal{N}_i$  be a subset of the  $N_{tot}$  SBs whose Channel Quality

<sup>1</sup> This applies in the non multiple-input-multiple-output (MIMO) configuration. For the MIMO configuration, a maximum of two different MCSs can be used for data belonging to two different transport blocks.

<sup>2</sup> The material in this section is based in part on Kwan et al. (2009b)

Indicator (CQI) values are to be reported by user  $i$ ; the size of  $\mathcal{N}_i$  is denoted by  $N_i$ . It is assumed that the  $N_i$  highest SB CQI values are fed back. Such a limited feedback scheme would necessitate a smaller bandwidth albeit at the cost of a degraded system performance. We also assume that the total available power is shared equally among the users. It is noted in Chung & Goldsmith (2001); Miki et al. (2007) that the throughput degradation resulting from such an assumption is small when adaptive modulation and coding (AMC) is used, as is the case in LTE.

Let  $\mathbf{x}_{i,n}, n = 1, 2, \dots, N_i$  be a real scalar or vector reported (via a feedback channel) by user  $i$  to indicate the collective channel qualities of all the sub-carriers within the  $n$ -th reported SB. The exact nature of  $\mathbf{x}_{i,n}$  depends on the feedback method adopted. Furthermore, let  $q_{i,\max}(\mathbf{x}_{i,n}) \in \{1, 2, \dots, J\}$  be the index of the highest-rate MCS that can be supported by user  $i$  for the  $n$ -th SB at CQI value  $\mathbf{x}_{i,n}$ , i.e.  $q_{i,\max}(\mathbf{x}_{i,n}) = \arg \max_j \left( R_j^{(c)} \log_2 \left( M_j \right) | \mathbf{x}_{i,n} \right)$ . Due to frequency selectivity, the qualities of the sub-carriers within a SB may differ; the indicator  $\mathbf{x}_{i,n}$  should provide a good collective representation of the qualities for all the sub-carriers within the  $n$ -th SB Blankenship et al. (2004); Ericsson (2003); Lampe & Rohling (2003). For convenience, we assume that the MCS rate  $R_j^{(c)} \log_2 \left( M_j \right)$  increases monotonically with  $j$  and that the rate of MCS 1 is zero. SBs whose CQI values are not reported back are assigned to MCS 1.

Let  $1 \leq q_{i,\max}(\mathbf{x}_{i,n}) \leq J$ , where  $q_{i,\max}(\mathbf{x}_{i,n}) = 1$  corresponds to the index of a zero bit rate, and  $J$  the index of the highest bit rate. The rate indices for the un-reported SBs are assigned zero bit rates, i.e.  $q_{\max,i}(\mathbf{x}_{i,n}) = 1, \forall n \notin \mathcal{N}_i$ . Fig. 2 shows the relationship between  $r_{i,j}$  and  $\mathbf{x}_{i,n}$  in the case when  $\mathbf{x}_{i,n}$  is assumed to represent an effective Signal-to-Interference and Noise Ratio (SINR) over a single SB. It might be noted that  $\mathbf{x}_{i,n}$  may not always take the form of a SINR Ericsson (2003)-Part 16: *Air Interface for Fixed and Mobile Broadband Wireless Access Systems* (2005). In Fig. 2,  $\mathbf{x}_{i,n}$  maps to  $j = q_{i,\max}(\mathbf{x}_{i,n}) = 3$ , which corresponds to  $r_{i,3}$ . It is important to note that  $q_{i,\max}(\mathbf{x}_{i,n})$  is assumed to be known at the scheduler once  $\mathbf{x}_{i,n}$  is known. The exact form of  $\mathbf{x}_{i,n}$  is not the focus of this paper.

As mentioned in Section 1, in a non-MIMO configuration, all SBs scheduled for a given user within the same TTI must use the same MCS. If MCS  $j$  is to be used for user  $i$ , then only certain SBs can be assigned to the user. For example, suppose  $N_i = 4$ , and

$$\begin{aligned} 1 \leq q_{i,\max}(\mathbf{x}_{i,2}) &< q_{i,\max}(\mathbf{x}_{i,1}) < q_{i,\max}(\mathbf{x}_{i,3}) \\ &< q_{i,\max}(\mathbf{x}_{i,5}) \leq J. \end{aligned} \quad (2)$$

Then, if MCS  $j = q_{i,\max}(\mathbf{x}_{i,3})$  is used, only SBs  $n = 3$  and 5 can be allocated to user  $i$  since only these SBs have good enough channel qualities to support an MCS index of  $q_{i,\max}(\mathbf{x}_{i,3})$  or higher. Selecting SBs  $n = 1$  or 2 with MCS  $j = q_{i,\max}(\mathbf{x}_{i,3})$  would result in unacceptably high error rates for these SBs. On the other hand, if  $j = q_{i,\max}(\mathbf{x}_{i,2})$ , all 4 SBs can be selected, at the expense of a lower bit rate for SBs 1, 3 and 5. This suggests that there is an optimal value of  $j$  which maximizes the total bit rate for user  $i$ .

### 3. Single User Optimization

In single user optimization, the aim is to determine the MCS (rate) index,  $j^*$  and the set of SBs to be allocated to user  $i$  so as to maximize the assigned bit rate,  $R_i$ , given the set of channel qualities  $\{q_{i,\max}(\mathbf{x}_{i,n}), n \in \mathcal{N}_i\}$ <sup>3</sup>.

<sup>3</sup> The material in this section is based in part on Kwan et al. (2009a)

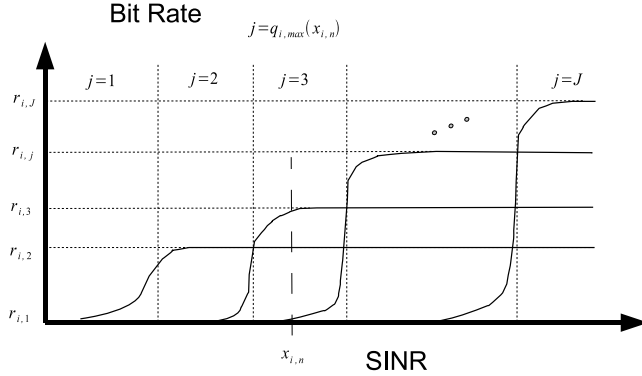


Fig. 2. Relationship between  $r_{i,j}$  and  $x_{i,n}$  in the case when  $x_{i,n}$  takes on the form of an effective SINR Kwan et al. (2009b)©[2009] IEEE.

Let  $Q_{max}(i) = \max_{n \in \mathcal{N}_i} \{q_{i,max}(\mathbf{x}_{i,n})\}$ , and let  $\mathbf{b}_i = [b_{i,1}, b_{i,2}, \dots, b_{i,Q_{max}(i)}]$  be the MCS vector for user  $i$ , where

$$b_{i,j} = \begin{cases} 0, & \text{if user } i \text{ is not assigned MCS } j \\ 1, & \text{if user } i \text{ is assigned MCS } j. \end{cases}$$

The optimal  $\mathbf{b}_i$ , which maximizes the total bit rate for user  $i$ , is obtained by solving Problem (P1).

$$(P0): \quad \max_{\mathbf{b}_i} \sum_{n \in \mathcal{N}_i} \sum_{j=1}^{q_{i,max}(\mathbf{x}_{i,n})} b_{i,j} r_j \quad (3)$$

subject to

$$\sum_{j=1}^{Q_{max}(i)} b_{i,j} = 1, \forall i, \quad (4)$$

$$b_{i,j} \in \{0,1\}, \forall i,j. \quad (5)$$

The formulation in (3) allows the selected bit rate for SB  $n$  to be less than what  $\mathbf{x}_{i,n}$  can potentially support, as may be the case if user  $i$  is assigned more than one SB during a TTI. Constraint (4) ensures that the MCS for user  $i$  can only take on a single value between 1 and  $Q_{max}(i)$ .

The above optimization problem can be easily solved as follows. Let  $\mathbf{R}^{(i)}$  be an  $N_i \times Q_{max}(i)$  matrix with  $(n,j)$ -th element  $\{r_{n,j}^{(i)} = r_j, j = 1, 2, \dots, q_{i,max}(\mathbf{x}_{i,n})\}$ . Denote the sum of the elements in the  $j$ -th column of  $\mathbf{R}^{(i)}$  by

$$c_j^{(i)} = \sum_{n=1}^{N_i} r_{n,j}^{(i)}. \quad (6)$$

Then the optimal MCS for user  $i$  is

$$j^* = \underset{1 \leq j \leq Q_{\max}(i)}{\operatorname{argmax}} c_j^{(i)} \quad (7)$$

and the corresponding maximum bit rate is  $c_{j^*}^{(i)}$ . The set,  $\mathcal{K}_i$ , of SBs allocated to user  $i$  is given by

$$\mathcal{K}_i = \{k \mid q_{i,\max}(\mathbf{x}_{i,k}) \geq j^*, k \in \mathcal{N}_i\}. \quad (8)$$

#### 4. Throughput Optimal Scheduler

In this section, the issue of multiuser scheduling is addressed, in which radio resources are jointly allocated in maximizing the total system throughput<sup>4</sup>.

##### 4.1 Multi-User Optimal Scheduling

If there are many users, the optimization problem becomes harder to solve. In addition, each SB can be allocated to at most one user Dahlman et al. (2008). Let

$$v_{i,n}(\mathbf{x}_{i,n}) = \sum_{j=1}^{q_{i,\max}(\mathbf{x}_{i,n})} b_{i,j} r_j \quad (9)$$

be the bit rate of SB  $n$  selected for user  $i$  given the channel quality  $\mathbf{x}_{i,n}$ , where  $b_{i,j} \in \{0,1\}$  is a binary decision variable. Let  $Q_{\max}(i) = \max_{n \in \mathcal{N}_i} \{q_{i,\max}(\mathbf{x}_{i,n})\}$ . The constraint

$$\sum_{j=1}^{Q_{\max}(i)} b_{i,j} = 1 \quad (10)$$

is introduced to ensure that the MCS for user  $i$  can only take on a single value between 1 and  $Q_{\max}(i)$ . The formulation in (9) allows the selected bit rate for SB  $n$  to be less than what  $\mathbf{x}_{i,n}$  can potentially support, as may be the case if user  $i$  is assigned more than one SB during a TTI. From (9) and (10), it can be seen that SB  $n$  might be selected for user  $i$  only if the MCS  $j^*$  chosen for user  $i$  satisfies  $j^* \leq q_{i,\max}(\mathbf{x}_{i,n})$ , as illustrated in Fig. 3.

The problem of jointly maximizing the sum of the bit rates for all users can be formulated as

$$(P1): \max_{\mathbf{A}, \mathbf{B}} \sum_{i=1}^U \sum_{n \in \mathcal{N}_i} a_{i,n} \sum_{j=1}^{q_{i,\max}(\mathbf{x}_{i,n})} b_{i,j} r_j \quad (11)$$

subject to (10) and

$$\sum_{i=1}^U a_{i,n} = 1, n \in \bigcup_{i=1}^U \mathcal{N}_i \quad (12)$$

$$a_{i,n}, b_{i,j} \in \{0,1\}, \forall i, j, n. \quad (13)$$

In problem (P1),  $\mathbf{A} = \{a_{i,n}, i = 1, \dots, U, n \in \mathcal{N}_i\}$ ,  $\mathbf{B} = \{b_{i,j}, i = 1, \dots, U, j = 1, \dots, Q_{\max}(i)\}$ , and  $a_{i,n}$  is a binary decision variable, with value 1 if SB  $n$  is assigned to user  $i$  and 0 otherwise. The objective in (11) is to select optimal values for  $\mathbf{A}$  and  $\mathbf{B}$  to maximize the aggregate bit rate  $\sum_{i=1}^U \sum_{n \in \mathcal{N}_i} a_{i,n} v_{i,n}(\mathbf{x}_{i,n})$ .

<sup>4</sup> The material in this section is based in part on Kwan et al. (2009b)

#### 4.2 Linearized Model

Problem (P1) is non-linear due to the product term  $a_{i,n}b_{i,j}$  in (11). Although solutions can be obtained using optimization techniques such as Branch-and-Bound Rardin (1998), a globally optimal solution cannot be guaranteed. To avoid this difficulty, the problem can be transformed into an *equivalent* linear problem by introducing an auxiliary variable  $t_{n,i,j} = a_{i,n}b_{i,j}$ . Then, Problem (P1) can be linearized as follows:

$$(P1') : \quad \max_{\mathbf{A}, \mathbf{B}, \mathbf{T}} \sum_{i=1}^U \sum_{n \in \mathcal{N}_i} \sum_{j=1}^{q_{i,max}(\mathbf{x}_{i,n})} t_{n,i,j} r_j \quad (14)$$

subject to (10), (12), (13) and

$$t_{n,i,j} \leq b_{i,j}, \quad (15)$$

$$t_{n,i,j} \leq a_{i,n}M, \quad (16)$$

$$t_{n,i,j} \geq b_{i,j} - (1 - a_{i,n})M, \quad (17)$$

where  $M$  is a large positive real value. Problem (P1') can then be solved using well-known integer linear programming techniques Rardin (1998).

#### 4.3 Multi-user Suboptimal Scheduling

In the optimal scheduler formulations in (P1) and (P1'), the MCSs, SBs, and users are jointly assigned. To reduce the computational complexity, the proposed sub-optimal scheduler performs the assignment in two stages. In the first stage, each SB is assigned to the user who can support the highest bit rate. In the second stage, the best MCS for each user is determined. The idea behind the sub-optimal scheduler is to assign a disjoint subset of SBs to each user, thereby reducing a joint multiuser optimization problem into  $U$  parallel single-user optimization problems. We will refer to this two-stage problem as Problem (P2).

Let  $\varphi_n$  be the index of the user which can support the highest-rate MCS for SB  $n$ , i.e.  $\varphi_n = \arg \max_{i \in \{1,2,\dots,U\}} q_{i,max}(\mathbf{x}_{i,n})$ . Furthermore, let  $\tilde{\mathcal{N}}_i$  be the (disjoint) set of SBs assigned to user  $i$ , i.e.  $\{n \text{ such that } \varphi_n = i\}$ . In the first stage, the sub-optimal scheduler determines  $\{\tilde{\mathcal{N}}_i\}_{i=1}^U$ . In other words, the best user is selected for each SB, and information regarding the set of SBs associated with each user is collected and represented by  $\tilde{\mathcal{N}}_i$ , as summarized in the following pseudo-code listing.

---

**Algorithm 1** Sub-optimal algorithm for solving Problem (P2).

---

```

1: for  $i = 1$  to  $U$  do
2:   for  $n \in \bigcup_{i=1}^U \tilde{\mathcal{N}}_i$  do
3:      $\varphi_n = \arg \max_i q_{i,max}(\mathbf{x}_{i,n})$ 
4:   end for
5:    $\tilde{\mathcal{N}}_i =$  the set of values of  $n$  such that  $\varphi_n = i$ 
6: end for
7: return  $\{\tilde{\mathcal{N}}_i\}_{i=1}^U$ .
```

---

Let  $Q'_{max}(i) = \max_{n \in \tilde{\mathcal{N}}_i} \{q_{i,max}(\mathbf{x}_{i,n})\}$  be the index of the highest MCS among all SBs associated with user  $i$ , and let the MCS vector,  $\mathbf{b}_i$ , for user  $i$  be

$$\mathbf{b}_i = [b_{i,1}, b_{i,2}, \dots, b_{i,Q'_{max}(i)}]. \quad (18)$$

Note that the length of  $\mathbf{b}_i$  may not necessarily be the same for all users. In the second stage, the sub-optimal scheduler determines  $\mathbf{b}_i$  which maximizes the total bit rate for user  $i$ . Similar to the approach in Section 4, the optimal  $\mathbf{b}_i$  can be obtained by solving the problem (P0) as described in section 3. Since the association between the SBs and the users are done in the first stage, the second stage only involves the selection of the best MCS given the selected SBs for each user. Compared to (P1) or (P1'), (P2) is a much simpler problem, due to the decoupled selection between SBs and MCSs.

#### 4.4 Numerical Results

For illustration purposes, we assume  $N_{tot} = 12$  SBs per TTI,  $L = 12$  sub-carriers per SB,  $N_1 = N_2 = \dots = N_U = N$  and that the normal cyclic prefix configuration is used Dahlman et al. (2008). The received amplitude for each subcarrier and user undergoes Nakagami-m fading Simon & Alouini (2005); unless otherwise indicated, a fading figure,  $m$ , value of 1 is assumed. The average signal-to-interference plus noise ratios (SINRs) for the  $U = 3$  users are 10 dB, 11 dB, and 12 dB respectively. It is assumed that the SINRs for all sub-carriers of each user are correlated, but identically distributed (c.i.d.), and that the resource blocks follow the localized configuration *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8)* (2007). The correlation coefficient between a pair of sub-carriers is given by  $\rho^{|i-j|}$ , where  $i$  and  $j$  are the sub-carrier indices. The SINR of each sub-carrier is assumed to be independent at the beginning of each scheduling period, and constant throughout the entire period. For simplicity, it is assumed that the set of MCSs consists of QPSK 1/2 and 3/4, 16-QAM 1/2 and 3/4, as well as 64-QAM 3/4 Andrews et al. (2007), and the L1/L2 control channels are mapped to the first OFDM symbol within each sub-frame. Furthermore, each sub-frame consists of 8 reference symbols Dahlman et al. (2008). The feedback method is based on the Exponential Effective SINR Mapping (EESM) Ericsson (2003), with parameter values obtained from Westman (2006). Let  $R_{tot}^*$  be the total bit rate defined in (11) or (14), and  $E[R_{tot}^*]$  be the value of  $R_{tot}^*$  averaged over 2500 channel realizations.

Fig. 4 shows the average total bit rate,  $E[R_{tot}^*]$ , as a function of  $\rho$  for  $N = 5$  and 12. It can be observed that the performance improves with the level of correlation among sub-carriers. Recall that the idea behind EESM is to map a set of sub-carrier SINRs,  $\{\Gamma_i\}_{i=1}^L$ , to a single effective SINR,  $\Gamma^*$ , in such a way that the block error probability (BLEP) due to  $\{\Gamma_i\}_{i=1}^L$  can be well approximated by that at  $\Gamma^*$  in additive white Gaussian noise (AWGN) Andrews et al. (2007); Ericsson (2003). The value of  $\Gamma^*$  tends to be skewed towards the weaker sub-carriers in order to maintain an acceptable BLEP. For small values of  $\rho$ , sub-carriers with large SINRs are not effectively utilized, leading to a relatively poor performance. For  $N = 5$ , the difference in  $E[R_{tot}^*]$  values achieved by the optimal scheduler at  $\rho = 0.1$  and  $\rho = 0.9$  is about 25%; the corresponding difference is similar for the sub-optimal scheduler.

Fig. 5 shows the average total bit rate,  $E[R_{tot}^*]$ , as a function of  $N$  for  $\rho = 0.5$  and 0.9. It can be seen that the performance improves with  $N$  as to be expected, but the rate of improvement decreases. There is little performance improvement as  $N$  increases beyond 8. For  $\rho = 0.5$ , the difference in  $E[R_{tot}^*]$  values achieved by the optimal scheduler at  $N = 6$  and  $N = 12$  is about 15%; the corresponding difference is about 10% for the sub-optimal scheduler.

Fig. 6 shows  $E[R_{tot}^*]$  as a function of the fading figure,  $m$ , with  $U = 3$ ,  $\rho = 0.9$  and  $N = 12$ . It can be seen that the throughput performance improves with  $m$ . As  $m$  increases, there is less variation in channel quality which allows a more efficient resource allocation among users. The difference in  $E[R_{tot}^*]$  values achieved by the optimal scheduler for  $m = 1$  and  $m = 10$  is about 15%; the corresponding difference is similar for the sub-optimal scheduler.



Fig. 7 shows  $E[R_{tot}^*]$  as a function of the number,  $U$ , of users for  $\rho = 0.9$  and  $N = 12$ . The average SINRs for all users are set to 10 dB. As  $U$  increases,  $E[R_{tot}^*]$  increases due to the more pronounced benefits from multiuser diversity. Fig. 8 shows the percentage gain in  $E[R_{tot}^*]$  for the optimal scheduler relative to the sub-optimal scheduler as a function of  $U$ . As  $U$  increases, it becomes increasingly likely that a given user will be assigned at most one SB in the first stage operation of the sub-optimal scheduler. In this event, the sub-optimal scheduler is actually optimal. It is therefore expected that the difference in performance between the optimal and sub-optimal schedulers will be small when  $U$  is large, as illustrated in Fig. 8. The result indicates that the sub-optimal scheduler is especially attractive for large values of  $U$  since it provides a significant reduction in complexity and its performance approaches that of the optimal scheduler.

## 5. Optimal Scheduler with Fairness

In Section 5<sup>5</sup>, the objective is to maximize system throughput. In this Section, we consider a proportional-rate scheduler which is intended to improve fairness among users.

### 5.1 Multi-User Optimal Scheduling

The joint optimization problem can be formulated as

$$(P3): \quad \max_{\mathbf{A}, \mathbf{B}} \sum_{i=1}^U \sum_{n \in \mathcal{N}_i} a_{i,n} \sum_{j=1}^{q_{i,max}(\mathbf{x}_{i,n}(t))} b_{i,j} \left( \frac{r_j}{\psi_i(t)} \right) \quad (19)$$

subject to (4), (12), (13). The term  $\psi_i(t)$  is given by

$$\psi_i(t) = \begin{cases} \bar{R}_i(t) & , \text{ PF Scheduling} \\ 1 & , \text{ Max-Rate Scheduling.} \end{cases} \quad (20)$$

In (20),  $\bar{R}_i(t) = (1 - \alpha)\bar{R}_i(t-1) + \alpha R_i(t-1)$  is the average bit rate up to time  $t-1$ ,  $\alpha \in [0, 1]$ , and  $R_i(t)$  is the bit rate assigned to user  $i$  at time  $t$ . It is known that the PF scheduler is asymptotically optimal Kelly (1997); Stolyar (2005). Other multi-user schedulers have been proposed in Bennett & Zhang (1996); Shakkottai & Stolyar (2002).

In problem (P3),  $\mathbf{A} = \{a_{i,n}, i = 1, \dots, U, n \in \bigcup_{i=1}^U \mathcal{N}_i\}$ ,  $\mathbf{B} = \{b_{i,j}, i = 1, \dots, U, j = 1, \dots, Q_{max}(i)\}$ , and  $a_{i,n}$  is a binary decision variable, with value 1 if SB  $n$  is assigned to user  $i$  and 0 otherwise. Problem (P3) is non-linear due to the product  $a_{i,n}b_{i,j}$  in (19). Although solutions can be obtained using optimization techniques such as Branch-and-Bound Rardin (1998), global optimality cannot be guaranteed. To overcome this difficulty, Problem (P2) can be transformed into an *equivalent* linear problem (P3') by introducing an auxiliary variable  $t_{n,i,j} = a_{i,n}b_{i,j}$ , i.e.

$$(P3'): \quad \max_{\mathbf{A}, \mathbf{B}, \mathbf{T}} \sum_{i=1}^U \sum_{n \in \mathcal{N}_i} \sum_{j=1}^{q_{i,max}(\mathbf{x}_{i,n}(t))} t_{n,i,j} \left( \frac{r_j}{\psi_i(t)} \right) \quad (21)$$

subject to (4), (12), (13), together with (15)-(17). Problem (P3') can then be solved using standard integer linear programming techniques Rardin (1998).

<sup>5</sup> The material in this section is based in part on Kwan et al. (2009a)

### 5.2 Multi-user Suboptimal Scheduling

The proposed sub-optimal multiuser scheduler consists of two stages. In the first stage, the scheduler determines the set,

$$\mathcal{Q}_{i,max}^{(t)}(\mathcal{N}_i) = \{q_{i,max}(\mathbf{x}_{i,n}(t)) | n \in \mathcal{N}_i\}, \quad (22)$$

of maximum rate indices, one for each SB for user  $i$  at time  $t$ . The users are then ranked according to their priority index values,  $\{\varphi_i, i = 1, 2, \dots, U\}$ ,

$$\varphi_i = \begin{cases} g(\mathcal{Q}_{i,max}^{(t)}(\mathcal{N}_i)) / \bar{R}_i(t) & , \text{ PF Scheduling} \\ g(\mathcal{Q}_{i,max}^{(t)}(\mathcal{N}_i)) & , \text{ Max-Rate Scheduling} \end{cases} \quad (23)$$

The first line in the RHS of (23) corresponds to proportional fair (PR) scheduling (Dahlman et al., 2008, Page 113), Kelly (1997); Wengert et al. (2005), whereas the second line corresponds to maximum rate scheduling (Dahlman et al., 2008, Page 111). The term  $g(\cdot)$  is a function which returns the highest bit rate that user  $i$  can support based on  $\mathcal{Q}_{i,max}^{(t)}(\mathcal{N}_i)$ , as discussed in Section 3, i.e.  $c_j^{(i)}(t) = g(\mathcal{Q}_{i,max}^{(t)}(\mathcal{N}_i))$ .

For notational convenience, let  $\varphi_{(1)} \geq \varphi_{(2)} \geq \dots \geq \varphi_{(U)}$  be the ranked version of  $\{\varphi_1, \varphi_2, \dots, \varphi_U\}$ , and  $\theta(j) \mapsto i$  be a function which maps the ordered user index  $j$  back to the original user index  $i$ . In the second stage, the allocation of resources is done in a sequential fashion, one user at a time, according to the following user order:  $\theta(1), \theta(2), \dots$ . Thus, starting with user  $\theta(1)$ , and the initial set of SBs,  $\mathcal{N}_{\theta(1)} = \mathcal{N}$ , where  $\mathcal{N}$  corresponds to the complete set of available SBs, the MCS index and the set of SBs,  $\mathcal{K}_{\theta(1)}$ , are determined as described in Section 3, and assigned to user  $\theta(1)$ . The remaining SBs,  $\mathcal{N}_{\theta(2)} = \mathcal{N}_{\theta(1)} - \mathcal{K}_{\theta(1)}$ , are then made available to user  $\theta(2)$ . The resource allocation process continues until all SBs have been assigned.

### 5.3 Numerical Results

For illustration purposes, we assume  $N_{tot} = 12$  SBs per TTI,  $L = 12$  sub-carriers per SB,  $N_1 = N_2 = \dots = N_U = N_{tot}$  and that the normal cyclic prefix configuration is used Dahlman et al. (2008). The fading amplitude for each subcarrier of any user follows the Nakagami-m model Simon & Alouini (2005), with a fading figure  $m$  equal to 1. It is assumed that the signal-to-interference plus noise ratios (SINRs) for all sub-carriers of any user are correlated, but identically distributed (c.i.d.), and that the resource blocks follow the localized configuration Dahlman et al. (2008). The correlation coefficient between a pair of sub-carriers is given by  $\rho^{|i-j|}$ , where  $i$  and  $j$  are the sub-carrier indices.

For simplicity, the SINR of a given sub-carrier is assumed to be independent at every scheduling period, and constant within a scheduling period. This independent assumption is reasonable for the purpose of comparing the long-term fairness for the Max-Rate and PF schedulers. The set of MCSs consists of QPSK 1/2 and 3/4, 16-QAM 1/2 and 3/4, as well as 64-QAM 3/4 Andrews et al. (2007), and the L1/L2 control channels are mapped to the first OFDM symbol within each sub-frame. Furthermore, each sub-frame consists of 8 reference symbols Dahlman et al. (2008). The feedback method is based on the Exponential Effective SINR Mapping (EESM) Ericsson (2003), with parameter values obtained from Westman (2006).

Let  $R_{tot}(t) = \sum_{i=1}^U R_i(t)$  be the total bit rate at time  $t$ , and  $\bar{R}_{tot} = \sum_{t=1}^T R_{tot}(t) / T$  be the corresponding value averaged over  $T = 2500$  channel realizations. Similarly, let  $\bar{R}_i = \sum_{t=1}^T R_i(t) / T$

be the average bit rate for user  $i$ , and  $\eta = \frac{(\sum_{i=1}^U \bar{R}_i)^2}{U \sum_{i=1}^U \bar{R}_i^2}$  be the Jain's fairness index Chiu & Jain (1989) for the average user bit rates. The value of  $\eta$  lies in the range  $[0, 1]$ ; an  $\eta$  value of 1 corresponds to all users having the same average (over  $T$  scheduling periods) bit rates.

Figs. 9 and 10 show the average total bit rate,  $\bar{R}_{tot}$ , and fairness index,  $\eta$ , as a function of  $\rho$  for three users, with average user SINRs of 14 dB, 15 dB, and 16 dB. It can be seen from Fig. 9 that the bit rates for all schedulers increase with  $\rho$ . This can be explained as follows. The motivation behind EESM is to map a set of sub-carrier SINRs,  $\{\Gamma_i\}_{i=1}^L$ , to a single effective SINR,  $\Gamma^*$ , in such a way that the block error probability (BLEP) due to  $\{\Gamma_i\}_{i=1}^L$  can be well approximated by that at  $\Gamma^*$  in additive white Gaussian noise (AWGN) Andrews et al. (2007); Ericsson (2003). The value of  $\Gamma^*$  tends to be skewed towards the weaker sub-carriers in order to maintain an acceptable BLEP. At a low value of  $\rho$ , sub-carriers with large SINRs are not effectively utilized, leading to a relatively poor performance. It can also be seen that the bit rate for the jointly optimal PF scheduler is almost as good as that for the jointly optimal Max-Rate scheduler. In comparison, the bit rates for the sequential Max-Rate and PF schedulers are about 5% and 10% lower. Fig. 10 shows that the fairness index,  $\eta$ , is significantly higher for the two PF schedulers than for their Max-Rate counterparts, indicating that the PF schedulers are quite effective in promoting fairness among users.

Similar plots are shown in Figs. 11 and 12, for user average SINRs of 10 dB, 15 dB, and 20 dB respectively. Here, the variation among user average SINRs is larger than in Figs. 9 and 10. Fig. 11 shows that there is now a larger gap between the bit rates for the jointly optimal PF and Max-Rate schedulers. This is due to the increased effort needed to maintain fairness. It can be seen from Fig. 12 that the two PF schedulers provide significantly better user fairness than the Max-Rate schedulers.

The average total bit rate and fairness index are plotted as a function of the number of users in Figs. 13 and 14 respectively, with an average SINR of 7 dB for all users. In this case, the results show that the jointly optimized Max-rate and PF schedulers provide similar performances. The sequential PF scheduler has a slightly lower throughput than the sequential Max-rate scheduler but a higher fairness index.

## 6. Conclusion

The problem of allocating resources to multiple users on the downlink in an LTE cellular communication system in order to maximize system throughput was studied. Numerical results show that both the correlation among sub-carriers and the amount of information fed back play important roles in determining the system throughput. It was found that a limited amount of feedback may be sufficient to provide a good performance. A reduced complexity sub-optimal scheduler was proposed and found to perform quite well relative to the optimal scheduler. The sub-optimal scheduler becomes especially attractive as the number of users increases. A scheduler which takes fairness among users into account was also discussed.

*Acknowledgment:* This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Grant OGP0001731, by the UBC PMC-Sierra Professorship in Networking and Communications, and a EU FP7 Marie Curie International Incoming Fellowship.

## 7. References

- Andrews, J. G., Ghosh, A. & Muhamed, R. (2007). *Fundamentals of WiMAX - Understanding Broadband Wireless Networking*, Prentice Hall.
- Bennett, J. C. R. & Zhang, H. (1996). WF2Q: Worst-case Fair Weighted Fair Queuing, *INFO-COM*, pp. 120–128.
- Blankenship, Y., Sartori, P. J., Classon, B. K., Desai, V. & Baum, K. L. (2004). Link Error Prediction Methods for Multicarrier Systems, *Proc. of IEEE Vehicular Technology Conference, Fall*, Vol. 6, pp. 4175 – 4179.
- Chiu, D. & Jain, R. (1989). Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks, *Computer Networks and ISDN Systems*.
- Chung, S. T. & Goldsmith, A. J. (2001). Degrees of Freedom in Adaptive Modulation: A Unified View, *IEEE Transactions on Communications* **49**(9): 1561 – 1571.
- Dahlman, E., Parkvall, S., Sköld, J. & Beming, P. (2008). *3G HSPA and LTE for Mobile Broadband*, Academic Press.
- Discussion on Control Signalling for Persistent Scheduling of VoIP* (2006). Technical Document R1-063275, 3rd Generation Partnership Project, Riga, Latvia.
- Ericsson (2003). System-Level Evaluation of OFDM - Further Considerations, TR R1-031303, TSG-RAN WG1 #35, Lisbon, Portugal.
- Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8)* (2007). Technical Report 3G TS36.211, 3rd Generation Partnership Project.
- Jiang, D., Wang, H., Malkamaki, E. & Tuomaala, E. (2007). Principle and Performance of Semi-persistent Scheduling for VoIP in LTE System, *Proc. of International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 2861 – 2864.
- Kela, P., Puttonen, J., Kolehmainen, N., Ristaniemi, T., Henttonen, T. & Moisio, M. (2008). Dynamic Packet Scheduling Performance in UTRA Long Term Evolution Downlink, *Proc. of Wireless Pervasive Computing*, pp. 308 – 313.
- Kelly, F. (1997). Charging and rate control for elastic traffic, *European Transactions on Telecommunications* **8**: 33 – 37.
- Kolehmainen, N. (2008). Channel Quality Indication Reporting Schemes for UTRAN Long Term Evolution Downlink, *Proc. of IEEE Vehicular Technology Conference, Spring*, pp. 2522 – 2526.
- Kwan, R., Leung, C. & Zhang, J. (2008). Multiuser Scheduling on the Downlink of an LTE Cellular System, *Research Letters in Communications*.
- Kwan, R., Leung, C. & Zhang, J. (2009a). Proportional Fair Multiuser Scheduling in LTE, *IEEE Signal Processing Letters* **16**(6): 461 – 464.
- Kwan, R., Leung, C. & Zhang, J. (2009b). Resource Allocation in an LTE Cellular Communication System, *Proc. of IEEE International Conference on Communications, Dresden, Germany*.
- Lampe, M. & Rohling, H. (2003). PER-Prediction for PHY Mode Selection in OFDM Communication Systems, *Proc. of IEEE Globecom*, Vol. 1, pp. 25 – 29.
- Liu, G., Zhang, J., Zhou, B. & Wang, W. (2007). Scheduling Performance of Real Time Service in Multiuser OFDM System, *Proc. of International Conference on Wireless Communications, Networking and Mobile Computing (WiCom)*, Shanghai, China, pp. 504 – 507.
- Liu, H. & Li, G. (2005). *OFDM-Based Broadband Wireless Networks, Design and Optimization*, John Wiley & Sons.
- Miki, N., Kishiyama, Y., Higuchi, K. & Sawahashi, M. (2007). Optimum Adaptive Modulation and Channel Coding Scheme for Frequency Domain Channel-Dependent Scheduling

- in OFDMA Based Evolved UTRA Downlink, *Proc. of IEEE Wireless Communications and Networking Conference WCNC*, pp. 1785 – 1790.
- Monghal, G., Pedersen, K. I., Kovacs, I. Z. & Mogensen, P. E. (2008). QoS Oriented Time and Frequency Domain Packet Schedulers for the UTRAN Long Term Evolution, *Proc. of IEEE Vehicular Technology Conference, Spring*, pp. 2532 – 2536.
- Ning, X., Ting, Z., Ying, W. & Ping, Z. (2006). A MC-GMR Scheduler for Shared Data Channel in 3GPP LTE System, *IEEE Proc. of Vehicular Technology Conference (VTC), Fall*, pp. 1 – 5.
- Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems (2005). TS IEEE Std 802.16e-2005, TSG-RAN WG1 #35.
- Pedersen, K. I., Monghal, G., Kovacs, I. Z., Troels, E., Pokhariyal, A., Frederiksen, F. & Mogensen, P. (2007). Frequency Domain Scheduling for OFDMA with Limited and Noisy Channel Feedback, *Proc. of IEEE Vehicular Technology Conference (VTC), Fall*, Baltimore, USA, pp. 1792 – 1796.
- Persistent Scheduling in E-UTRA (2007). Technical Document R1-070098, 3rd Generation Partnership Project, Sorrento, Italy.
- Pokhariyal, A., Kolding, T. E. & Mogensen, P. E. (2006). Performance of Downlink Frequency Domain Packet Scheduling For the UTRAN Long Term Evolution, *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*.
- Pokhariyal, A., Pedersen, K. I., Monghal, G., Kovacs, I. Z., Rosa, C., Kolding, T. E. & Mogensen, P. E. (2007). HARQ Aware Frequency Domain Packet Scheduler with Different Degrees of Fairness for the UTRAN Long Term Evolution, *Proc. of IEEE Vehicular Technology Conference (VTC), Spring*, Dublin, Ireland, pp. 2761 – 2765.
- Rardin, R. (1998). *Optimization in Operations Research*, Prentice Hall, Upper Saddle River, NJ.
- Shakkottai, S. & Stolyar, A. L. (2002). Scheduling for Multiple Flows Sharing a Time-Varying Channel: The Exponential Rule, *American Mathematical Society Translations* **207**: 185–202.
- Simon, M. K. & Alouini, M.-S. (2005). *Digital Communication over Fading Channels*, 2 edn, John Wiley & Sons.
- Stolyar, A. L. (2005). On the Asymptotic Optimality of the Gradient Scheduling Algorithm for Multiuser Throughput Allocation, *Operations Research* **53**(1): 12 – 25.
- Wengerter, C., Ohlhorst, J. & A.G.E Von Elbwert (2005). Fairness and Throughput Analysis for Generalized Proportional Fair Frequency Scheduling in OFDMA, *Proc. of IEEE Vehicular Technology Conference*.
- Westman, E. (2006). *Calibration and Evaluation of the Exponential Effective SINR Mapping (EESM) in 802.16*, Master's thesis, The Royal Institute of Technology (KTH).
- Wunder, G., Zhou, C., Bakker, H.-E. & Kaminsk, S. (2008). Throughput Maximization under Rate Requirements for the OFDMA Downlink Channel with Limited Feedback, *Article ID 437921, EURASIP Journal on Wireless Communications and Networking*.

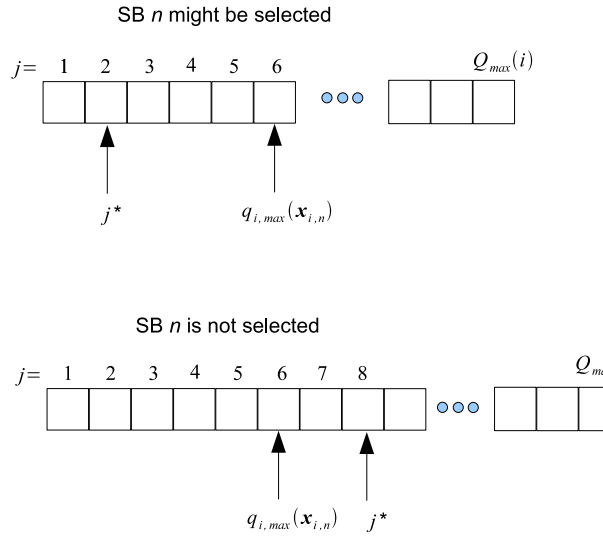


Fig. 3. MCS Selection. In the upper illustration, SB  $n$  could be selected, since  $j^* \leq q_{i,max}(x_{i,n})$ . On the other hand, when  $j^* > q_{i,max}(x_{i,n})$ , SB  $n$  cannot be selected, as shown in the bottom illustration Kwan et al. (2009b)©[2009] IEEE.

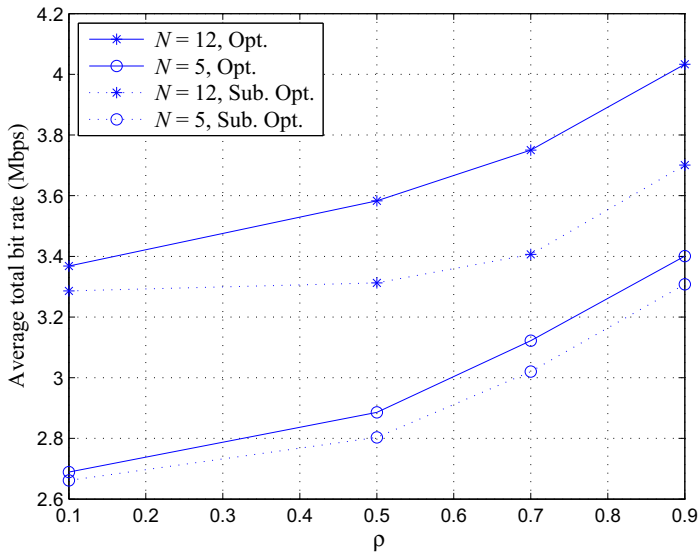


Fig. 4. Average total bit rate as a function of  $\rho$  with  $N = 5$  and  $12$  and  $U = 3$  Kwan et al. (2009b)©[2009] IEEE.

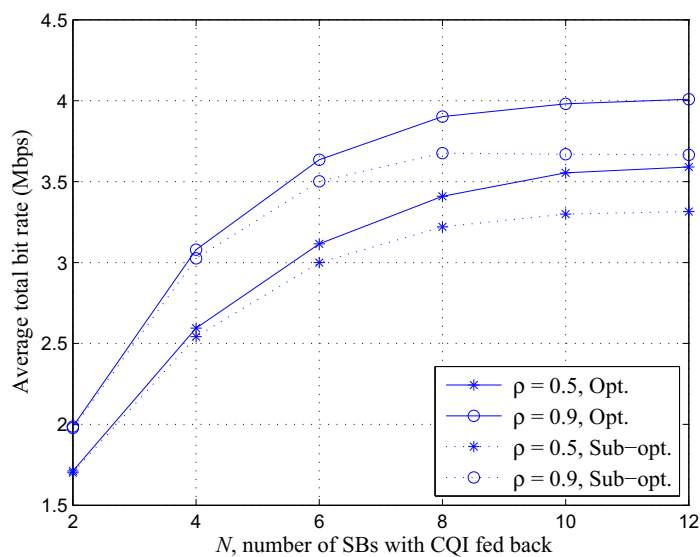


Fig. 5. Average total bit rate as a function of  $N$  for  $\rho = 0.5$  and  $0.9$  with  $U = 3$  Kwan et al. (2009b)©[2009] IEEE.

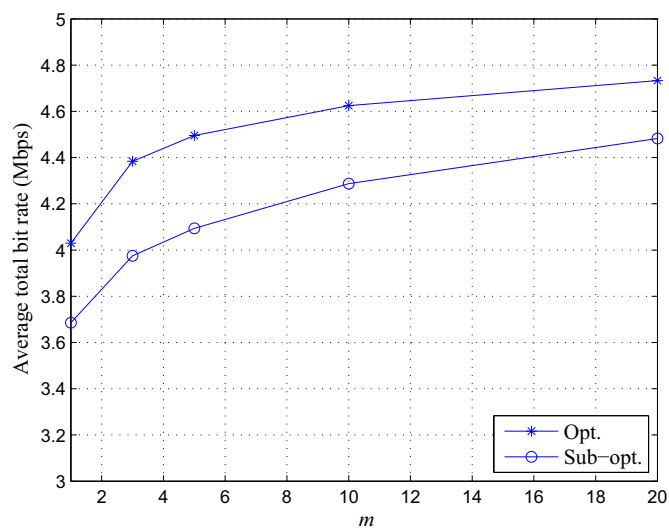


Fig. 6. Average total bit rate as a function of the fading figure,  $m$ , with  $U = 3$ ,  $\rho = 0.9$  and  $N = 12$  Kwan et al. (2009b)©[2009] IEEE.

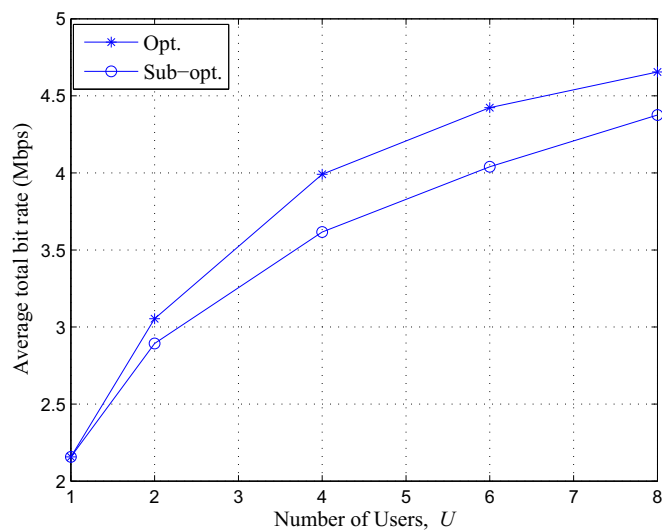


Fig. 7. Average total bit rate as a function of the number of users,  $U$ , with  $\rho = 0.9$  and  $N = 12$  Kwan et al. (2009b)©[2009] IEEE.

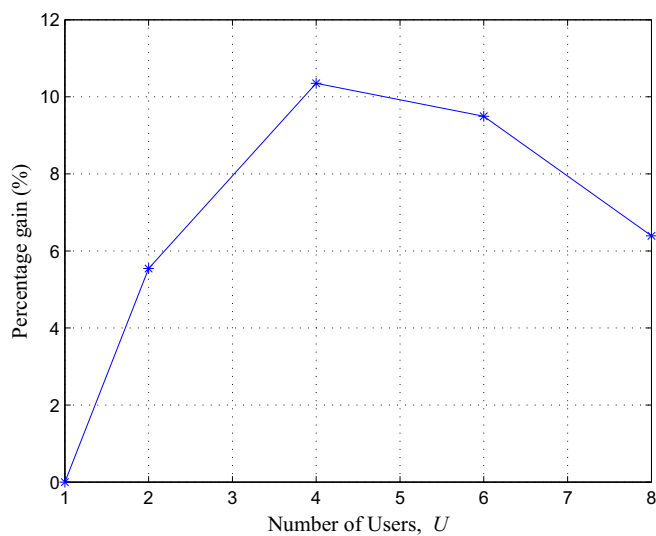


Fig. 8. Percentage gain of the optimal scheduler over the suboptimal scheduler as a function of the number,  $U$ , of users with  $\rho = 0.9$  and  $N = 12$  Kwan et al. (2009b)©[2009] IEEE.



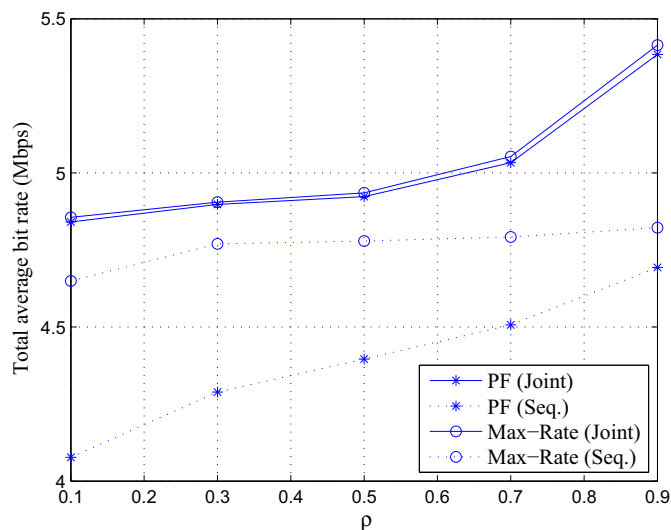


Fig. 9. Average total bit rate as a function of  $\rho$  for three users with average SINRs of 14 dB, 15 dB and 16 dB Kwan et al. (2009a)©[2009] IEEE.

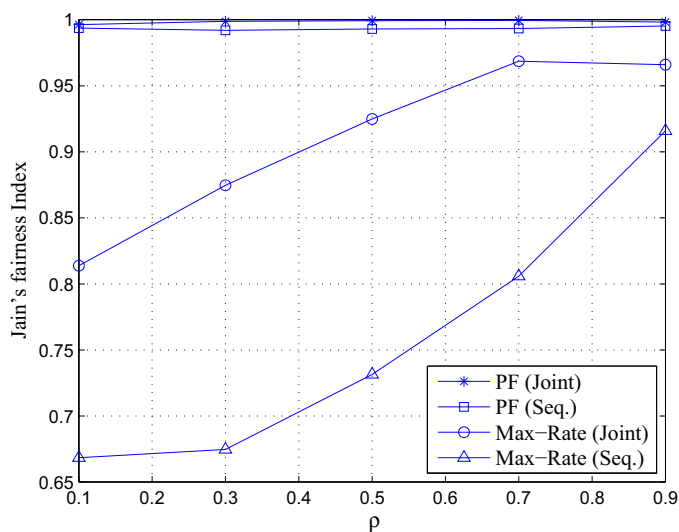


Fig. 10. Fairness index as a function of  $\rho$  for three users with average SINRs of 14 dB, 15 dB and 16 dB Kwan et al. (2009a)©[2009] IEEE.

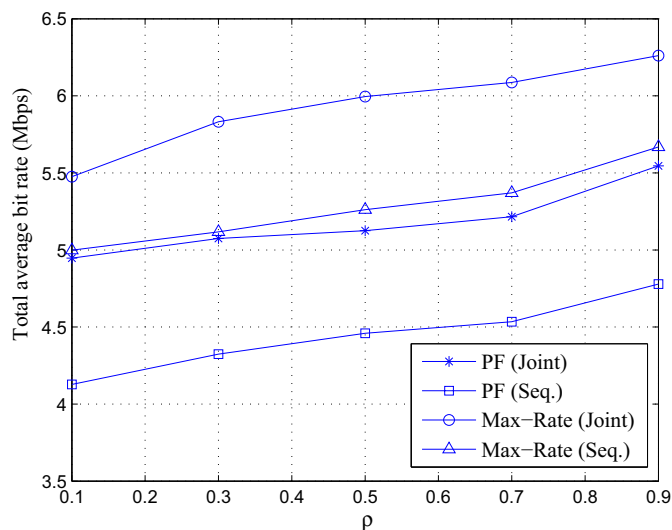


Fig. 11. Average total bit rate as a function of  $\rho$  for three users with average SINRs of 10 dB, 15 dB and 20 dB Kwan et al. (2009a)©[2009] IEEE.

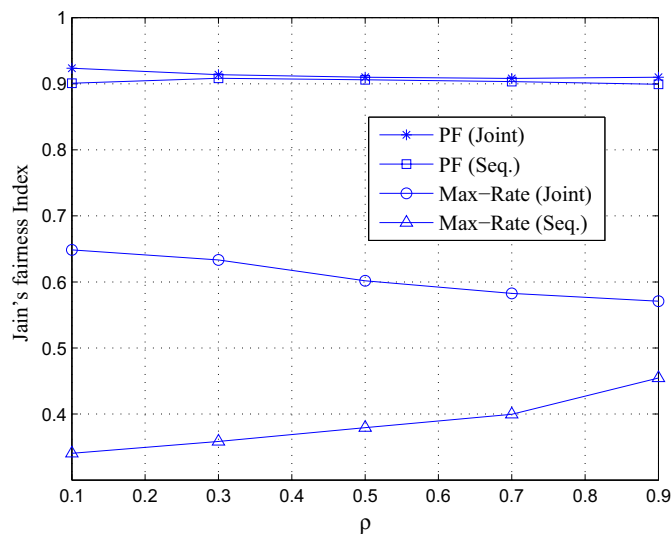


Fig. 12. Fairness index as a function of  $\rho$  for three users with average SINRs of 10 dB, 15 dB and 20 dB Kwan et al. (2009a)©[2009] IEEE.

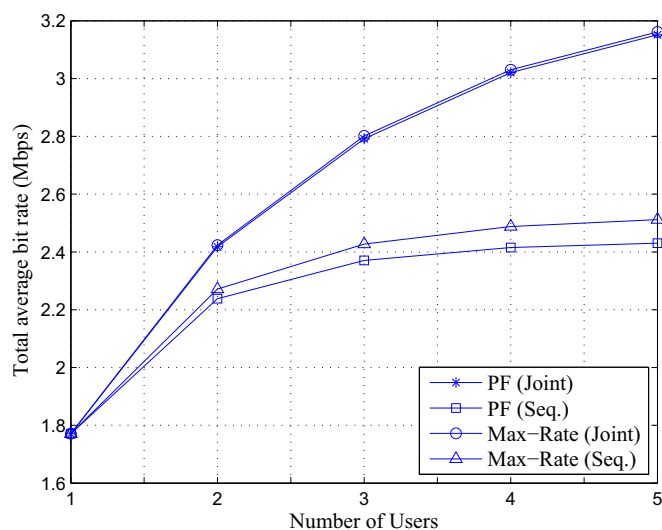


Fig. 13. Average total bit rate as a function of the number of users with  $\rho = 0.9$  and average SINRs of 7 dB for all users Kwan et al. (2009a)©[2009] IEEE.

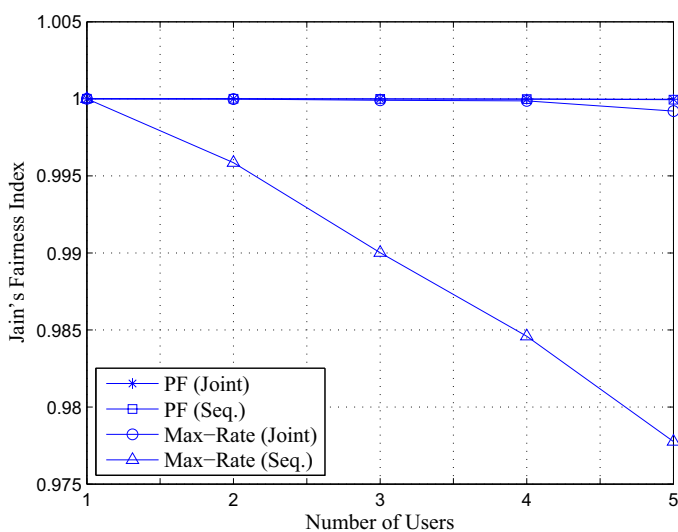


Fig. 14. Fairness index as a function of the number of users with  $\rho = 0.9$  and average SINRs of 7 dB for all users Kwan et al. (2009a)©[2009] IEEE.

