Time-Frequency Coupled Proportional Fair Scheduler with Multicarrier Awareness for LTE Downlink

D. Martin-Sacristan, J. F. Monserrat, D. Calabuig and N. Cardona
Universitat Politècnica de València - iTEAM Research Institute
Valencia, Spain
Email: {damargan,jomondel,dacaso,ncardona}@iteam.upv.es

***Abstract*—Proportional fair schedulers have been thoroughly used in Long Term Evolution (LTE) due to their ability to provide a good trade-off between cell spectral efficiency and user fairness. Current algorithms provide suboptimum solutions at a low computational cost, but present several drawbacks. This paper proposes a Coupled and Multicarrier Aware PFS (CMA-PFS) for LTE downlink that increases efficiency as compared with current algorithms with independent time and frequency domain scheduling, referred to as Decoupled PFS (D-PFS). The proposed algorithm includes new features such as tight coupling between time and frequency domain scheduling and multicarrier transmission awareness. Simulations have been conducted using an International Mobile Telecommunications Advanced (IMT-Advanced) compliant semi-static simulator. Results show that the CMA-PFS improves proportional fairness as compared with D-PFS that is reflected both in an increase of cell spectral efficiency (around $+2\%$) and a higher cell-edge user spectral efficiency (around $+10\%$) in a Single User MIMO (SU-MIMO) system.**

I. INTRODUCTION

Packet scheduling is the Long Term Evolution (LTE) resource management function whose main aim is to allocate radio resources to users [1]. Many scheduling algorithms have been proposed to optimize system performance. Among them, Proportional Fair Scheduling (PFS) provides a good trade-off between cell spectral efficiency and user fairness maximization due to a distribution of resources among users that takes into account both the channel quality and the previously experienced throughput [2], [3], [4]. Nowadays, most of the real implementations of schedulers are based on this PFS philosophy.

One peculiarity of LTE is that, in this technology, the scheduler decides not only on the resource allocation but also on the transmission characteristics. Specifically, in downlink, the scheduler chooses the modulation, coding, number of streams, spatial processing, priority between transmissions and/or retransmission and even the distribution of flows into Hybrid ARQ (HARQ) processes. LTE presents two particular characteristics related with the selection of the transmission format [5]. First, the standard indicates that all Resource Blocks (RBs) allocated to a user in a subframe must use the same Modulation and Coding Scheme (MCS). This MCS-constraint affects multicarrier transmission performance. Second, it is a common assumption in LTE-related research that the same number of RBs and even the same MCS used in the first transmission are used in case of retransmission since it reduces the complexity of the receiver. This will be referred to as the HARQ-constraint.

The high complexity of optimal PFS has encouraged the development of suboptimum scheduling algorithms. In LTE, most algorithms are based on decoupling the scheduling problem into two phases as proposed in [6]. In the first phase, the Time Domain Scheduling (TDS), a subset of users is selected following a given criterion. In the second phase, the Frequency Domain Scheduling (FDS), resources are allocated to the previously selected users. One positive feature of the algorithm described in [6] is that it is HARQ-aware, since it takes into account the HARQ-constraint and the resulting decoding gain due to retransmissions. One disadvantage of this algorithm is that, although TDS reduces FDS complexity, it also decreases multiuser diversity order in a non-optimized way since TDS and FDS are almost independent processes. Besides, FDS is unaware of multicarrier transmission implications.

This paper proposes a novel scheduling algorithm, suited for the downlink of LTE, referred to as Coupled and Multicarrier Aware PFS (CMA-PFS). This algorithm extends that proposed by Pokhariyal *et al.* [6], hereinafter referred to as Decoupled PFS (D-PFS). The CMA-PFS includes a higher coupling between TDS and FDS processes than D-PFS. Additionally, CMA-PFS is multicarrier-aware in the sense that MCS-constraint is considered to properly estimate the performance of multicarrier tranmsmissions while D-PFS assumes that this performance is just the addition of the performance of several single-carrier transmissions.

II. PROPORTIONAL FAIR SCHEDULING

A. Mathematical Model of Scheduling

We consider a system with a set of K users and a set, Z, of Q RBs. Let us define Ω_k as the set of RBs allocated to user k in a scheduling interval, fulfilling $\Omega_k \subseteq Z$. Consider P as the collection of all subsets of Z, that is, $A \in P \leftrightarrow A \subseteq Z$. Then $\Omega_k \in P$.

We assume that at the time of decision making, t, the scheduler knows the average data rate (hereinafter named throughput) experienced by user k until this moment, $T_k(t)$. We also assume that a Link Adaptation (LA) module provides $R_k(A,t)$, defined as the estimation of the data rate that user k can achieve in a transmission over the set of resources A during the scheduling interval that starts at time t.

Once the scheduling decision is made, a specific data rate is assigned to each user. Let us define $\hat{R}_k(t)$ as the assigned data rate to user k in the scheduling interval that starts at time t. Note that $\hat{R}_k(t) = 0$ if user k is not selected for transmission.

978-1-4799-8088-8/15/$31.00

Hereinafter, variable t is omitted if an undefined scheduling time is considered. Also, the set of resources A is omitted if a system consists of only 1 resource.

B. Proportional fairness

The concept of proportional fairness was originally proposed by Kelly [7]. Adapting the original concept, we could say that a vector of experienced throughputs $T^{PF}(t) = \{T_1^{PF}(t), ..., T_K^{PF}(t)\}$ is proportionally fair if it maximizes the following utility function:

$$U_{PF}(T(t)) = \sum_{k=1}^{K} \log(T_k(t)), \tag{1}$$

for all feasible $T(t)$ vectors, where it is worth noting that only bounded $T_k(t)$ values are feasible. Proportional fairness utility function provides a trade-off between maximization of the aggregate of user throughputs and maximization of the fairness among users.

A PFS was implemented for Qualcomm's High Data Rate (HDR) system in [2]. The PFS implementation assigns in each time t the channel to the k user that maximizes the next utility function:

$$U_{HDR}(k,t) = \frac{R_k(t)}{T_k(t)}. \tag{2}$$

In practice, an exponential moving average is used to obtain $T_k(t)$, instead of using an exact averaging of all the past throughput values [8]. At time $t+1$, the throughput is updated according to the following equation [2]:

$$T_k(t+1) = \frac{1}{W} \cdot T_k(t) + \left(1 - \frac{1}{W}\right) \cdot \hat{R}(k,t), \tag{3}$$

where W is a parameter equivalent to the length of a sliding average window.

It is shown in [4] that with sufficiently large value of t and W, $T_k(t)$ weakly converges to a constant value for a certain user k. Also, it is demonstrated that the vector of average throughput values converges to a proportional fair solution that solves the problem presented in Equation 1. In order to drawn such conclusion, fairly accurate rate predictions are assumed in [4]. Concerning the value of W, it has been shown that, in practice, it should be chosen to offer a good estimation of the average throughput, with the ability to track changes in the channel characteristics [4] (e.g. 100 slots in [3], 1000 slots in [2]).

In [4] an extension of the HDR proportional fair scheduling to multicarrier systems is discussed. A simple strategy is proposed in which the HDR PFS is applied independently to each RB. Therefore, each resource r is allocated to the user k maximizing the next utility function:

$$U_{MC_a}(k,r,t) = \frac{R_k(r,t)}{T_k(t)}. \tag{4}$$

Additionally, in [4] a more complete strategy is also presented in which it is considered that the data rate of a joint transmission over multiple resources can be different from the aggregate of the data rates of multiple transmissions performed over independent resources. The strategy can be translated to an algorithm that assigns at each time t the channel according to an allocation Ω_k to maximize the next utility function:

$$U_{MC_b}(t) = \sum_{k=1}^{K} \frac{R_k(\Omega_k,t)}{T_k(t)}. \tag{5}$$

Note that if the number of channels and/or the number of schedulable users is 1, both strategies are equivalent to that represented by Equation 2. It is worth noting that, according to [4], the latter strategy presents the same convergence properties as the original HDR PFS. However, the first strategy is clearly the worst one since it does not consider properly multicarrier transmission performance.

III. D-PFS ALGORITHM

This paper presents a new algorithm based on a LTE scheduler implementation that has become a *de facto* standard, the D-PFS algorithm. This algorithm is based on the work of Pokhariyal *et al.* [6]. Other studies have followed a similar approach (see [1], e.g.). The main characteristic of D-PFS operation is that it is divided into two phases: the TDS and the FDS.

First, the TDS selects up to K_S users from the K users requiring resources. This is achieved after a prioritization of users according to a proportional fair utility. The aim of this TDS selection is twofold: decrease FDS complexity and signaling overhead required by user multiplexing. Additionally TDS allows a high Quality of Service (QoS) control. A peculiarity of the D-PFS is that it is assumed that TDS does not know how FDS allocates resources (they are independent processes). Therefore, user utilities are calculated assuming full-bandwidth transmissions. Besides, in TDS, prioritization of users does not care about pending retransmissions.

After TDS, the FDS is executed. In this phase, resources are allocated to the K_S selected users. Users with retransmissions are prioritized (arguing the minimization of transmission delay) and their required resources are guaranteed. Specifically, assuming the presence of K_{S_R} users with pending retransmissions that jointly require Q_R of a total of Q RBs, $Q_N = Q - Q_R$ RBs are firstly allocated to the $K_{S_N} = K_S - K_{S_R}$ users with new data, guaranteeing the availability of Q_R blocks for retransmissions. Resources for new transmissions are allocated according to proportional fair utilities calculated per user and RB as in Equation 4. On the other hand, resources for retransmissions are distributed according to only the channel quality of each RB. Although this mechanism (that allocates resources for new transmission first) disfavors retransmissions, it is stated that this fact is counteracted by the decoding gain due to soft combining of HARQ retransmissions.

One drawback of D-PFS is related to the utility function used in the FDS phase. Note that the used utility function, Equation 4, is not the most complete one, Equation 5. They would be equivalent if and only if the achievable data rate in a multicarrier transmission was just the addition of the achievable data rates in multiple transmissions over single resources. Nevertheless, this is not true in LTE, among other reasons, because the same MCS must be used in all RBs allocated to a user. In practice, it means that, when the data rate estimate $R_k(A,t)$ is calculated over a set of resources A, the estimated data rate per resource $R_k(A,t)/|A|$ (being $|A|$ the number of resources in A) may be lower than the data rate estimated for some resources (underutilized) and higher for others (limiting resources). Two ideas arise from this analysis. First, if the value of $R_k(A,t)$ is highly conditioned by the resource with the worst quality it would be useful for the user to get rid of this RB. Moreover, another user could get the unallocated RB to improve its own utility value. The second idea is that, even if a user cannot improve its utility by getting rid of a RB, it is yet possible to achieve a better scheduling decision through the transfer of a RB from one user to another. This can be achieved via the transfer of a bad RB that produces an underutilization of capacity in other RBs, or through the transfer of a good RB with underutilized capacity that could be better used by another user.

The second drawback of the D-PFS is that TDS operation produces an obvious reduction in the diversity order of the final scheduling decision. The effect of this drawback would be minimized if TDS prioritized users in such a way that the final decision was similar to that of the strategy presented in Equation 5. In order to get the most of the opportunistic scheduling performed by FDS it could be positive that TDS prioritized users with very high utility peaks, as FDS tends to do. Nevertheless, the TDS prioritizes users according to full bandwidth metrics. In conclusion, it is envisaged that D-PFS could be improved by increasing the coupling between TDS and FDS, making TDS prioritize users in a way similar to how FDS allocates resources.

IV. NEW ALGORITHMS

Based on the analysis of the D-PFS, two algorithms incrementally including new features are presented. This incremental process is ended with the CMA-PFS algorithm.

A. Decoupled and Multicarrier Aware PFS

The first assessed algorithm is the Decoupled and Multicarrier Aware PFS (DMA-PFS). This algorithm is based on the D-PFS but includes the MCS-constraint awareness, and hence, multicarrier awareness. This algorithm is applied to the users with new transmissions. In summary, after the normal operation of the D-PFS FDS an iterative process is conducted to exchange resources among users, and even to leave RBs unallocated, with the aim of maximizing the scheduling utility.

The exact procedure is described in Algorithm 1. For each candidate user, the RB r_w with highest negative impact on its utility, in the sense that if this RB was liberated the utility would present the highest increase (even if this increase is negative), is found. This increase is referred to as 'donor user utility increase', ΔU_d. Next, it is found the user k_r, that being the recipient of r_w, would have the highest utility increase. This increase is referred to as 'recipient user utility increase', ΔU_r. Then, if total utility can be increased, the RB is allocated to the best recipient user or remains unallocated. Finally, for each unallocated RB it is found the user that getting this resource maximizes the user utility increase. If the maximum user utility increase is positive the resource is allocated. The process is repeated for all the users if, in a previous execution, any allocation change is done.

Algorithm 1 MCS constraint Aware Enhancement

```
Initialize $continue = true$
Let $K_{S_N}$ be the set of candidate users
Let $\Theta$ be the set of unallocated resources
while $continue = true$ do
    $continue = false$
    for each $k_d$ in $K_{S_N}$ do
        $r_w = \Omega_{k_d}(\text{argmax}_j\ U(k_d, \Omega_{k_d} - \Omega_{k_d}(j)))$
        $\Delta U_d = U(k_d, \Omega_{k_d} - r_w) - U(k_d, \Omega_{k_d})$
        $k_r =_{\text{argmax}_{k_t}} \ (U(k_t, \Omega_{k_t} + r_w) - U(k_t, \Omega_{k_t}))$
        $\Delta U_r = U(k_r, \Omega_{k_r} + r_w) - U(k_r, \Omega_{k_r})$
        if $\Delta U_d > 0$ or $\Delta U_d + \Delta U_r > 0$ then
            unallocate $r_w$: $\Omega_{k_d} \leftarrow \Omega_{k_d} - r_w$ and $\Theta \leftarrow \Theta + r_w$
            if $\Delta U_r > 0$ then
                allocate $r_w$: $\Omega_{k_r} \leftarrow \Omega_{k_r} + r_w$ and $\Theta \leftarrow \Theta - r_w$
            end if
            if $|\Theta| > 0$ then
                for each $r$ in $|\Theta|$ do
                    $k_r =_{\text{argmax}_{k_t}} \ (U(k_t, \Omega_{k_t} + r) - U(k_t, \Omega_{k_t}))$
                    $\Delta U_r = U(k_r, \Omega_{k_r} + r) - U(k_r, \Omega_{k_r})$
                    if $\Delta U_r > 0$ then
                        allocate $r$: $\Omega_{k_r} \leftarrow \Omega_{k_r} + r$ and $\Theta \leftarrow \Theta - r$
                    end if
                end for
            end if
            $continue = true$
        end if
    end for
end while
```

B. Coupled and Multicarrier Aware PFS

The second algorithm is the CMA-PFS. This algorithm is based on the DMA-PFS but includes a tighter coupling between TDS and FDS. It is said that TDS and FDS are coupled because some knowledge is incorporated to TDS about FDS allocation in order to calculate TDS priorities.

The exact procedure is described in Algorithm 2. First, utilities are calculated for each RB r and for each user k, assuming a new transmission will be performed independently over each RB. Then, each resource is tentatively allocated to the user with highest utility for this resource, identified by k_{best}. Users with tentatively allocated resources are grouped in a high priority group and their priorities, P_k^H, are calculated as the aggregate of the utilities of the resources tentatively allocated. Users without tentatively allocated resources are included in a low priority group and their priorities, P_k^L, are calculated as the maximum resource utility of each user. Finally, the K_S users with the highest priorities are selected from the high priority group and if there is still room for more users those come from the low priority group.

Note that HARQ-awareness is not considered in CMA-PFS.

Algorithm 2 TDS-FDS Coupling Enhancement

```
Calculate $U(k,r) = R_k(r)/T_k(r)$
for each $r$ do
    $k_{best} =_{\text{argmax}_k} \; U(k,r)$
    $\Omega_{k_{best}} \leftarrow \Omega_{k_{best}} + r$
end for
for each user $k$ do
    if $|\Omega_k| > 1$ then
        $P_k^H = \sum_{i=1}^{|\Omega_k|} U(k, \Omega_k(i))$
    else
        $P_k^L = \max(U(k,\Omega_k(1)), \ldots, U(k,\Omega_k(|\Omega_k|)))$
    end if
end for
select $K_S$ users
```

V. ASSESSMENT METHODOLOGY

Assessment methodology of this work was based on system level simulations made following the guidelines specified by the International Telecommunication Union (ITU) Radiocommunication Sector (ITU-R) for the performance analysis of International Mobile Telecommunications Advanced (IMT-Advanced) technologies [9]. The system simulation platform used in this assessment was calibrated in the framework of the Wireless World Initiative New Radio + (WINNER+) project [10] to ensure validity of results.

The proposed algorithms were tested in the Indoor hotspot (InH) environment extracted from [9]. It considers an indoor scenario with a long hall with adjacent offices. Users are pedestrians and two base stations operating at 3.4 GHz with an omnidirectional antenna setup are mounted on the ceiling of the corridor. The IMT-Advanced stochastic and geometric channel model was used in the assessment.

Concerning LTE configuration, Channel Quality Indicator (CQI), Precoding Matrix Indicator (PMI) and Rank Indicator (RI) were reported with 5 ms period and with a frequency granularity of 5 RBs. Using this feedback, a LA algorithm was conducted that included an Outer-Loop Link Adaptation (OLLA) to control CQI estimation errors maintaining the Block Error Rate (BLER) of new transmissions at a target value of 20% [11].

Base stations had 4 transmitting antennas, while users had 2 receiving antennas. Antennas were considered to be vertically polarized and uniformly spaced 0.5 wavelengths. Two Multiple-Input Multiple-Output (MIMO) schemes were considered: Single Input Multiple Output (SIMO) and Single User MIMO (SU-MIMO) using codebook-based closed-loop spatial multiplexing with dynamic rank adaptation based on user's feedback.

Additional assumptions and models of the simulation methodology are indicated in Table I.

Several performance indicators are used in this assessment. The Cell Scheduling Utility (CSU) is used to measure how proportionally fair is an algorithm. It is calculated as:

$$CSU = \frac{\sum_{c=1}^{C} \sum_{k=1}^{K_c} \log(T_{c,k})}{C}, \tag{6}$$

TABLE I: Simulation assumptions and models.

Simulation length / runs	2 s / 100 runs
Cell selection	1 dB handover margin
Traffic model	full buffer
Interference model	explicit model
CSI feedback	realistic
SINR estimation error	1 dB lognormal error per RB
Control channel overhead	3 OFDM symbols per subframe
PFS W	1000
PFS K_S	5

where C is the total number of cells simulated (in all simulation runs), K_c is the number of users served by cell c and $T_{c,k}$ is the average throughput experienced by user k of cell c during a simulation run. Cell Spectral Efficiency (CSE) measures the spectral efficiency of the whole system. This performance indicator is defined as:

$$CSE = \frac{\sum_{c=1}^{C} \sum_{k=1}^{K} T_{c,k}}{C \cdot \omega}, \tag{7}$$

where ω is the channel bandwidth used per cell. Cell-Edge User Spectral Efficiency (CEUSE) is defined as the 5% point of the Cumulative Density Function (CDF) of the User Spectral Efficiency (USE), being the USE defined as the throughput of a user divided by the channel bandwidth.

VI. RESULTS

The gains in terms of CSU achieved by DMA-PFS and CMA-PFS over the CSU of the D-PFS algorithm are represented in Figure 1, both for SIMO and SU-MIMO. Results show an improvement in CSU with both new algorithms. This means that the throughput distribution produced by both algorithms is more proportionally fair than that of the D-PFS. It also means that both multicarrier-awareness and coupling are positive features. CMA-PFS algorithm provides the best performance since it includes both features.

Concerning the rest of performance indicators, it would be desirable an increase of them. In fact, results shown in Figures 2 and 3 show that CEUSE and CSE are improved with the new algorithms. Indeed CEUSE presents very high gains, from 9% to 20% for the CMA-PFS. At the same time, CSE does not decrease but presents gains between 1.4% and 2%. This is a very remarkable fact since it is easy to improve the CEUSE of a system with a scheduling that gives higher priority to users with low throughput values. But it usually comes with the cost of a reduced CSE. Nevertheless, the proposed algorithms provide CEUSE and CSE gain at the same time.

Two additional facts are remarkable. First, results have been obtained in a realistic simulation setup. For example, rate estimation errors have been taken into account. Therefore, it is not strictly necessary to have perfect rate estimates to obtain performance improvements from the presented algorithm, what makes them really appealing. Second, both in SIMO and SU-MIMO performance gains have been obtained, what makes the algorithms widely useful in LTE deployments.

VII. CONCLUSIONS

A novel scheduling algorithm, named CMA-PFS, has been proposed for the downlink of LTE, based on a well-known

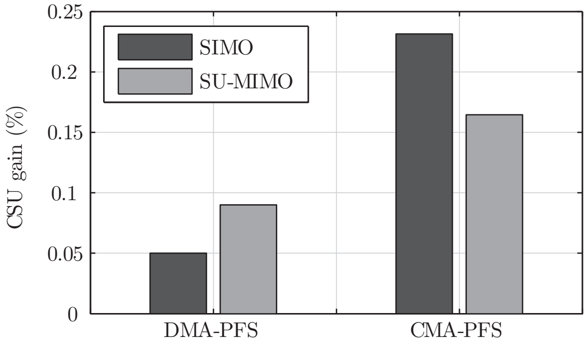

Fig. 1: CSU gain obtained with new algorithms

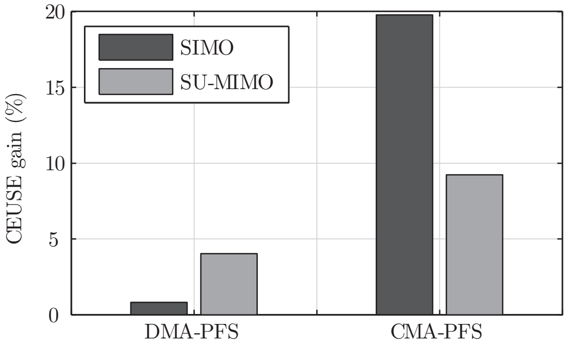

Fig. 2: CEUSE gain obtained with new algorithms

algorithm, the D-PFS. First, the advantages and disadvantages of D-PFS have been analyzed. Then, as a first step towards CMA-PFS, multicarrier awareness has been incorporated to D-PFS obtaining the DMA-PFS algorithm. Next, coupling between D-PFS has been considered to obtain the CMA-PFS. Simulations conducted in an indoor scenario have shown that the proposed algorithms provide an increase of proportional fairness to the system. Furthermore, this increase comes with a higher cell-edge user spectral efficiency at the same time that cell spectral efficiency is improved. This behavior has been shown for both SIMO and SU-MIMO schemes.

ACKNOWLEDGMENT

This work was supported by the Spanish Ministry of Science under the project TEC2011-27723-C02-02.

REFERENCES

[1] D. Martín-Sacristán, J. F. Monserrat, J. Cabrejas-Peñuelas, D. Calabuig, S. Garrigas, and N. Cardona, “On the way towards fourth-generation mobile: 3GPP LTE and LTE-advanced,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2009, pp. 1–10, Mar. 2009.

[2] A. Jalali, R. Padovani, and R. Pankaj, “Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system,” in *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, vol. 3, 2000, pp. 1854–1858.

[3] C. Wengerter, J. Ohlhorst, and A. von Elbwart, “Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA,” in *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, vol. 3, May/Jun. 2005, pp. 1903–1907.

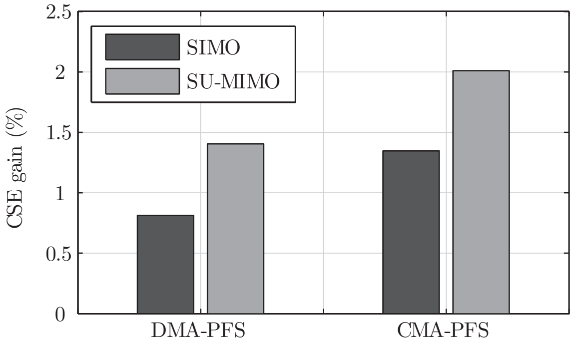

Fig. 3: CSE gain obtained with new algorithms

[4] H. Kushner and P. Whiting, “Convergence of proportional-fair sharing algorithms under general conditions,” *Wireless Communications, IEEE Transactions on*, vol. 3, no. 4, pp. 1250–1259, Jul. 2004.

[5] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures,” 3GPP, TS 36.213 v8.0.0, 2007.

[6] A. Pokhariyal, K. I. Pedersen, G. Monghal, I. Z. Kovacs, C. Rosa, T. E. Kolding, and P. E. Mogensen, “HARQ Aware Frequency Domain Packet Scheduler with Different Degrees of Fairness for the UTRAN Long Term Evolution,” in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, 2007, pp. 2761–2765.

[7] F. Kelly, “Charging and rate control for elastic traffic,” *European Transactions on Telecommunications*, vol. 8, pp. 33–37, 1997. [Online]. Available: http://www.statslab.cam.ac.uk/ frank/elastic.pdf

[8] J. Holtzman, “Asymptotic analysis of proportional fair algorithm,” in *Personal, Indoor and Mobile Radio Communications, 2001 12th IEEE International Symposium on*, vol. 2, Sep./Oct. 2001, pp. 33–37.

[9] ITU-R, “Guidelines for evaluation of radio interface technologies for IMT-Advanced,” International Telecommunications Union Radio (ITU-R), Report ITU-R M.2135-1, Dec. 2009, available online at http://www.itu.int/publ/R-REP/en.

[10] K. Safjan, V. D’Amico, D. Bultmann, D. Martin-Sacristan, A. Saadani, and H. Schoneich, “Assessing 3GPP LTE-advanced as IMT-advanced technology: the WINNER+ evaluation group approach,” *IEEE Communications Magazine*, vol. 49, no. 2, pp. 92–100, Feb. 2011.

[11] I. Z. Kovacs, K. I. Pedersen, T. E. Kolding, A. Pokhariyal, and M. Kuusela, “Effects of Non-Ideal Channel Feedback on Dual-Stream MIMO-OFDMA System Performance,” in *Vehicular Technology Conference, 2007. VTC-2007 Fall. 2007 IEEE 66th*, 2007, pp. 1852–1856.