

Evaluation of Utility-Based Adaptive Resource and Power Allocation for Real Time Services in OFDMA Systems

Yuri V. L. Melo, Emanuel B. Rodrigues, Fco. Rafael M. Lima, Tarcisio F. Maciel and Fco. Rodrigo P. Cavalcanti

Wireless Telecommunications Research Group (GTEL)
Federal University of Ceará (UFC), Fortaleza, Brazil
Email: {yurivictor, emanuel, rafaelm, maciel, rodrigo}@gtel.ufc.br

Abstract—This paper analyzes the performance of Adaptive Power Allocation (APA) based on (multi-level) water filling as a novel extension of a utility-based Radio Resource Allocation (RRA) framework designed for Real Time (RT) service provision in Orthogonal Frequency Division Multiple Access (OFDMA)-based systems. We evaluate capacity, satisfaction and fairness improvements that originate from exploiting the high degree of flexibility of RRA in the context of OFDMA systems and particularly, concentrate on optimal power allocation combined with utility-based scheduling. Results show that Delay-based Satisfaction Maximization (DSM) presents the best performance to maximize user satisfaction in comparison with classical algorithms and conclude also that benefit of APA is marginal compared to Equal Power Allocation (EPA).

I. INTRODUCTION

Efficient and innovative RRA techniques are essential to provide considerable efficiency gains in coverage, capacity and Quality of Service (QoS) for OFDMA-based broadband wireless networks. Currently, the area of power allocation in wireless systems has received interest from both academic and industrial researchers [1], because the demand for services with different QoS requirements, such as real time and non-real time services, is increasing. Due to the frequency-selective attenuation of the wireless channel, the transmit power per sub-carrier can be adapted in order to increase the spectral efficiency. The capacity can be maximized if more transmit power is applied to frequency areas with lower attenuation relative to the other frequencies. As different sub-carriers experience different fades and transmit different number of bits, the transmit power levels must be changed accordingly.

In [2], the authors use four types of utility functions that can represent most of the services in wireless systems. The proposed algorithm can deal with concave, convex, sigmoid and inverse sigmoid utility functions in a unified way and it can be applied to uplink and downlink power allocation. Water filling is an efficient method of power allocation applied in many engineering problems [3], however, this is an unfair solution, because it tends to allocate more power to users with good channel conditions that, in general, are located near the Base Station (BS).

This work was supported by the Innovation Center, Ericsson Telecomunicações S.A., Brazil, under EDB/UFC.33 Technical Cooperation Contract. Yuri Victor Lima de Melo is sponsored by CNPq - Brazil. Emanuel B. Rodrigues would like to thank CNPq and FUNCAP (Brazil) for a Postdoctoral Scholarship support.

In this paper, we are going to evaluate the impact of APA in the performance of utility-based RRA techniques, namely DSM, which are suitable for RT services and it was also compared with other algorithms as Satisfaction-Oriented Resource Allocation for Real Time Services (SORA-RT) [4], Rate Maximization (RM) [5], Modified Largest Weighted Delay First (MLWDF) [6] and Urgency and Efficiency-based Packet Scheduling (UEPS) [7].

The sections are organized as follows: Section II, the system model is addressed; Section III describes the general utility-based optimization problem considered in this work; Section IV is discussed about utility-based multi-level water filling solution and EPA; Section V presents the simulation results; and finally, in Section VI some conclusions are drawn.

II. SYSTEM MODELING

The considered scenario is based on an OFDMA-based Long Term Evolution (LTE) system and it is formed by a single hexagonal cell with a BS at its corner. In order to get statistically reliable results, each simulation follows a Monte Carlo approach and is composed by a sufficiently large number of drops in which User Equipment (UE) are positioned randomly. In each drop, UEs are static, i.e., there is no UE mobility. Downlink transmissions are based on Orthogonal Frequency Division Multiplexing (OFDM) using a normal cyclic prefix length, 14 OFDMA symbols per Transmission Time Interval (TTI), and 12 sub-carriers with 15 kHz of bandwidth each. Herein, this frequency-time block is designed as Resource Block (RB) and is the minimum allocable unit taking thus a TTI of 1 ms (two slots of 0.5 ms) and occupying 180 kHz. In our model, we assume that the channel coherence bandwidth is larger than the bandwidth of an RB, so that each RB experiences flat fading. We also assume low mobility for the fast fading, so that the channel remains constant during one TTI, but varies from one TTI to another. We also assume that the BS has perfect Channel State Information (CSI) for all users and all RBs. For our RT service, traffic consists of regular packets of 256 bits generated every 2 ms. The delay of each packet is accounted and it must respect the delay budget of the radio access network. If the packet arrives at the receiver later than this delay budget, it is discarded. The main parameters considered in the simulations are summarized in Table I.

Table I
SIMULATION PARAMETERS

Parameter	Value	Ref.
Number of cells	1	
Max. BS transmit power	1 W	
BS antenna pattern	Three-sectored	
Cell radius	500 m	
Carrier frequency	2 GHz	
System bandwidth	5 MHz	
Number of RBs	25	
Path loss	$128.1 + 37.6 \log_{10} d$	[8]
Large-scale fading	Zero-mean lognormal	[9]
Shadowing stand. deviation	8 dB	[9]
Small-scale fading	3GPP Typical Urban (TU)	[10], [11]
AWGN power	-123.24 dBm	
Link adaptation	LTE 15 MCS	[12]

d is the distance between communicating devices in meters

III. UTILITY-BASED OPTIMIZATION FORMULATION

Utility theory can be used in communication networks to quantify the benefit of the usage of certain resource as bandwidth and power; or to evaluate the degree to which a network satisfies delay requirements of UEs' applications. The general utility-based optimization problem considered in this work is formulated as

$$\max_{\rho_{j,k}, p_k} \sum_{j \in \mathcal{J}} U(x_j) \quad (1a)$$

$$\text{subject to } \rho_{j,k} \in \{0, 1\}, \forall j \in \mathcal{J} \text{ and } \forall k \in \mathcal{K}, \quad (1b)$$

$$\sum_{j \in \mathcal{J}} \rho_{j,k} = 1, \forall k \in \mathcal{K}, \quad (1c)$$

$$\sum_{k \in \mathcal{K}} p_k = P_t, \quad (1d)$$

$$p_k \geq 0, \forall k \in \mathcal{K}, \quad (1e)$$

where J is the total number of UEs in the set \mathcal{J} of UEs of the cell, K is the total number of RBs in the set \mathcal{K} of system resources to be assigned to UEs, $\rho_{j,k}$ is an assignment variable that assumes the value 1 if the RB k is assigned to the UE j and 0 otherwise, p_k is the transmit power allocated to RB k , P_t is total BS transmit power, $U(x_j)$ is a user utility function based on a generic variable x_j that can represent a resource usage or QoS metric associated to user j . Constraints (1b) and (1c) state that RBs are discrete and that the same RB cannot be shared by two or more users in the same TTI. Constraints (1d) and (1e) require that the total sum of the powers over all RBs must not surpass the total transmit power of the BS, and that these powers must be non-negative.

In order to achieve user satisfaction shaping, we propose to use a sigmoidal user utility function based on a generic QoS metric x_j of user j , as indicated below:

$$U(x_j[n]) = \frac{1}{1 + e^{\mu \cdot \sigma (x_j[n] - x_j^{\text{req}})}}, \quad (2)$$

where μ is a constant (-1 or 1) that determines if the sigmoid is an increasing or decreasing function, respectively, σ is a non-negative parameter that determines the shape of the

sigmoidal function, and $x_j[n]$ and x_j^{req} are the current QoS metric and QoS requirement of UE j , respectively.

The user marginal utility given by the utility-based weight plays an important role in the Dynamic Resource Assignment (DRA) algorithm, because the higher the weight, the higher the priority of the user to get a resource. The user-dependent utility-based weight based on a generic QoS metric x_j of user j is given by

$$w_j = \frac{\partial U(x_j[n])}{\partial x_j[n]} = \frac{\sigma \cdot e^{\mu \cdot \sigma (x_j[n] - x_j^{\text{req}})}}{\left(1 + e^{\mu \cdot \sigma (x_j[n] - x_j^{\text{req}})}\right)^2}, \quad (3)$$

In order to have a desired step-shaped sigmoid no matter the value of the QoS requirement, the fixed σ parameter must be a function of the QoS requirement x_j^{req} . The higher the value of σ , the closer to a step-shaped function the utility function will be. Otherwise, considering lower values of σ , the utility function becomes more linear. The sigmoid should be equal to a given value δ when the QoS metric $x_j[n]$ achieves a given proportion ρ of the QoS requirement x_j^{req} . Therefore, we have that

$$\sigma = \frac{\log \frac{1-\delta}{\delta}}{\rho \cdot x_j^{\text{req}}}, \quad (4)$$

It is possible to perform user satisfaction shaping for RT services through the sigmoidal user utility function $U(\cdot)$ in the optimization problem formulated in (1). In this paper a novel RRA technique based in sigmoidal user utility function is used, which is termed DSM. The considered optimization problem for RT services is the maximization of the total utility with respect to the users' Head Of Line (HOL) packet delays. The objective function becomes

$$\max_{\rho_{j,k}, p_k} \sum_{j \in \mathcal{J}} U_{\text{rt}}(d_j^{\text{hol}}[n]), \quad (5)$$

The HOL delay is the time that the oldest packet in the users' buffer has to wait before gaining access to the wireless channel. The DSM policy considers the UE' HOL delay as the QoS metric, which is calculated using

$$d_j^{\text{hol}}[n+1] = d_j^{\text{hol}}[n] + t_{\text{tti}} - \frac{1}{L} \cdot \left(\frac{R_j[n] \cdot t_{\text{tti}}}{S_p} \right), \quad (6)$$

where t_{tti} is the duration of the TTI in seconds, L is the packet arrival rate, S_p is the packet size, and $R_j[n]$ is the instantaneous achievable transmission rate on TTI n . In this queue model, we assume that the packet size S_p is sufficiently small, so that the queue can be represented ideally by a sequence of time slices with duration $1/L$ seconds each.

Looking at (6), firstly it can be seen that the HOL delay is always incremented by at least the duration of one TTI, no matter how many bits were transmitted in the current transmission interval.

This represents the passing of time in the system, which means that all packets in the queue will be one TTI older. Secondly, the decrement of the HOL delay depends on the number of time slices (duration of $1/L$ seconds each) that is

decremented due to the transmission in the current TTI. If the j^{th} UE has not been served by any resource in the n^{th} TTI, $R_j[n]$ is equal to zero and no time slices are decremented.

If the instantaneous transmission rate is such that the HOL packet is totally transmitted in the current TTI, it means that one time slice with duration of $1/L$ seconds should be decremented in the HOL delay. If the instantaneous transmission rate is sufficiently high so that many packets in the queue can be transmitted, the corresponding number of time slices should be decremented in the HOL delay.

IV. POWER ALLOCATION

After the resources have been dynamically allocated among UE using DSM, SORA-RT, RM, MLWDF and UEPS, the next step is realized APA or EPA. In this second stage of the problem-splitting technique, the APA sub-problem finds the optimum power solution based in [13], [14] and EPA.

A. Adaptive Power Allocation

The optimal power allocation has the solution in the form of a utility-based multi-level water filling, as indicated in (7):

$$p_k = \left(w_j \cdot \tilde{\mu} - \frac{1}{\xi_{j,k}} \right)^+, \quad (7)$$

where $(x)^+ \triangleq \max(0, x)$; p_k is the current optimal power allocated to RB k belonging to the set \mathcal{K}_j of RBs assigned to user j ; w_j is the user weight given by (3); $1/\xi_{j,k}$ is the inverse of the Channel-to-Noise Ratio (CNR) (channel quality) of RB k assigned to user j ; and $\tilde{\mu}$ is a non-negative variable that represents the water-level of the water filling problem.

In [3], the authors provide a family of exact algorithms to efficiently obtain water filling solutions. Before presenting an algorithm to obtain the water filling solution we redefine some variables in the following.

Notice that we have K RBs, which yields a total of K sub-channels with allocated power p_k given by the water filling solution (7). Let us assume more general water filling expressions such as $p_i = [a_i \cdot \tilde{\mu} - b_i]^+$, where a_i 's and b_i 's are arbitrary positive numbers, and $i \in \{1, \dots, K\}$ is the sub-channel index. We define $a_i = w_j$ and $b_i = (1/\xi_{j,k})$ for all sub-channels associated to each user j .

Therefore, the solution in equation (7) is re-written as

$$p_i = (a_i \cdot \tilde{\mu} - b_i)^+, \quad \forall i \in \{1, \dots, K\}, \quad (8)$$

Note that substituting (8) in (1d) and solving for $\tilde{\mu}$ we have that

$$\tilde{\mu} = \frac{P_t + \sum_{i=1}^K b_i}{\sum_{i=1}^K a_i}. \quad (9)$$

Assume also that

$$g(\tilde{\mu}) = \left(\sum_{i=1}^K a_i \right) \tilde{\mu} - \left(P_t + \sum_{i=1}^K b_i \right). \quad (10)$$

Finally, the algorithm to obtain the water filling solution is presented in Algorithm 1

- 1: Set $T = K$ and sort the set of pairs $\{(a_i, b_i)\}$ such that $\frac{a_i}{b_i}$ is in descending order. Furthermore, define $\frac{a_{T+1}}{b_{T+1}} = 0$.
- 2: If $\frac{b_T}{a_T} < \frac{b_{T+1}}{a_{T+1}}$ and $g\left(\frac{b_T}{a_T}\right) < 0$ then go to step (2) of this algorithm. Otherwise, do $T = T - 1$ and go to step (1) of this algorithm.
- 3: Solve $g(\tilde{\mu}) = 0$ for $\tilde{\mu}$, i.e., apply equation (9), and find the power allocation solution according to equation (8). Undo the sorting of step (1) of this algorithm.

Algorithm 1: Water filling algorithm for SISO.

B. Equal Power Allocation

EPA is characterized by the equal distribution of the total BS power among the resources. Notice that EPA can also be adopted in the power allocation step instead of APA. In some cases, adaptive power allocation in OFDMA-based systems provides limited gains in comparison with equal power allocation with much more complexity [15].

V. RESULTS

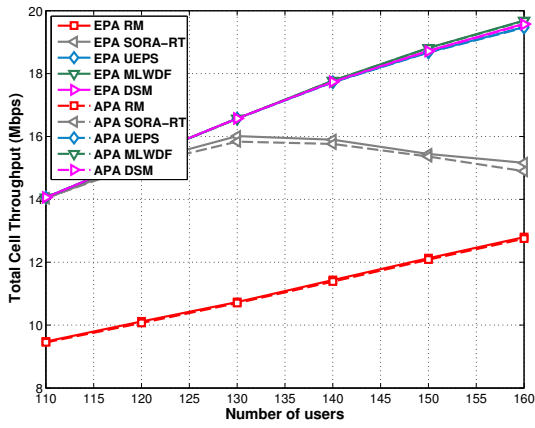
This section provides the performance of DSM, MLWDF, UEPS, SORA-RT and RM and also evaluate them with or without APA. The simulations took into account the main characteristics of an OFDMA system. The parameters considered are summarized in table II.

Table II
SIMULATION PARAMETERS IN THE EVALUATION OF ADAPTIVE POWER ALLOCATION FOR RT SERVICES

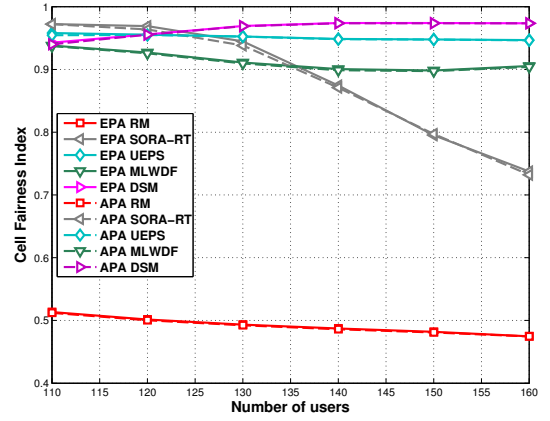
Service	Parameter	Value
RT	Packet size (S_p)	256 bits
	Packet interarrival time ($1/\lambda$)	2 ms
	FER threshold	2%
	HOL delay requirement (d_j^{req})	100 ms
	Delay budget	100 ms
	Parameter μ	1
	Parameter σ	138.135
	Parameter δ	0.01
	Parameter ρ	0.5
	Antenna configuration	SISO
	Simulation time span	30 s
	Number of independent simulation runs	10

Fig. 1(a) depicts the total cell throughput (system capacity) as a function of the number of RT users. The delay-aware policies have better performance of capacity because they are more successful at avoiding packet losses due to delays.

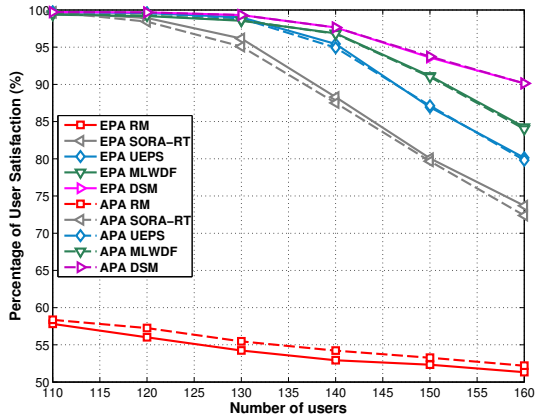
If more packets are successfully transmitted, the system capacity is higher. At first sight, one could expect that the pure opportunistic policy RM would present the highest system capacity. But in the scenario we are evaluating, this is not the case, as it can be seen in Fig. 1(a). The reason for that behaviour is because RM chooses few users with best channel quality to transmit, but the buffers of these users do not have so much data to transmit because of the nature of the RT traffic model considered in this work. Therefore, the RBs, which have



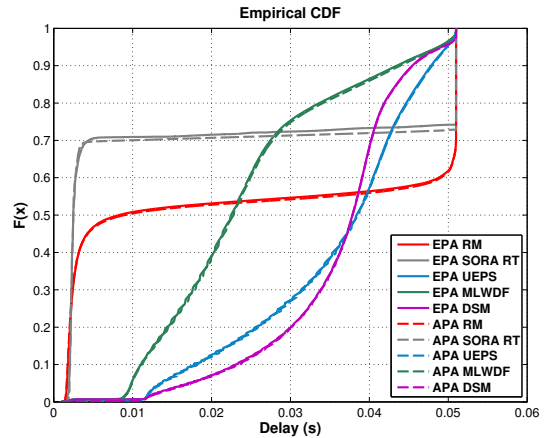
(a) Total cell throughput as a function of the number of users.



(b) Cell fairness index as a function of the number of users.



(c) User satisfaction as a function of the number of users.



(d) CDF of mean packet delays for 160 users.

Figure 1. Comparison of the main results between APA and EPA to RT service

a huge transmission capability, will not be efficiently used due to lack of data. That is why the system capacity provided by RM in this scenario is poor.

However, if we combine the opportunistic characteristic of RM with a proper delay-based component, just like the DSM policy does, we have a remarkable improvement in system capacity. The DSM technique, together with MLWDF and UEPS, shows the best results. It is interesting to notice that the SORA-RT algorithm initially shows a good performance, but suddenly starts to lose capacity when the offered load achieves 120 RT users. This is an indication that the SORA-RT algorithm fails at avoiding system congestion when the offered load increases. Also notice that the APA versions of the RRA techniques have approximately the same performance of their EPA counterparts.

The mean cell fairness index based on HOL delay is shown in Fig. 1(b). The RM provides the lowest fairness because it leaves many users unattended due to bad channel quality. On the other extreme, we have the proposed DSM algorithm, which is able to provide both the highest system capacity and fairness in a remarkable way. Other delay-aware algorithms, such as UEPS and MLWDF have also good performance in

Finally, the SORA-RT algorithm does not show good fairness results for high offered loads. When the system load increases, the algorithm heuristics is not good enough to avoid users from becoming unsatisfied, and so many users are neglected and fairness decreases. The fairness decrease in this case is associated with an increasing number of users that do not have chance to transmit. One more time, we cannot see any noticeable difference between the fairness levels achieved by EPA and APA.

Fig. 1(c) depicts the percentage of satisfied RT users. The algorithms that take into account the HOL delay in their formulations are those ones that provide the highest user satisfaction. The resource allocation criteria of these algorithms are based on the combination of two indicators: a QoS indicator that is a function of the HOL delay, and an efficient indicator that can be the achievable transmission rate (DSM) or the ratio between the transmission rate and the user throughput (MLWDF and UEPS). Comparing DSM and UEPS, which have the same QoS indicator (bell-shaped marginal utility function), it can be concluded that the achievable transmission rate is a better efficiency indicator for the maximization of user satisfaction, since DSM outperforms UEPS.

SORA-RT, which was especially designed to provide high user satisfaction levels, provides reasonable percentage of user satisfaction, but not the sufficient to surpass DSM, MLWDF or UEPS. Special attention must be given to the proposed DSM technique, which achieved its objective of maximizing user satisfaction. The combination of the bell-shaped delay-based indicator and the achievable transmission rate indicator proved to be the best option.

Regarding power allocation, one can observe that APA does not make any difference in terms of user satisfaction for DSM, MLWDF and UEPS. On the other hand, the water filling power allocation has a slight impact on the performance of the RM and SORA-RT techniques. While the former took advantage of the adaptive power allocation among RBs to improve user satisfaction, the latter was not able to translate this dynamic allocation into satisfaction gains.

The Cumulative Distribution Function (CDF) of mean packet delays for 160 RT users is shown in Fig. 1(d). We define the mean packet delay of an RT user as the average of the delays of all packets in the buffer of this user. We considered in the simulations an RT delay budget of 100 ms, i.e., packets are discarded in the transmitter if they are waiting for transmission for more than 100 ms. Let us suppose that a given user is in outage and his/her packets are arriving in a given time interval. After some time his/her buffer will have a constant length because the HOL packet will be discarded and a new packet will arrive. In this case, the mean packet delay will be approximately 50 ms, which is the average among all packet delays in his/her buffer. This is the reason why the CDF range is approximately from 1 ms (TTI duration) to 50 ms. After this initial explanation, let us analyze Fig. 1(d) from the capacity point of view. A saturation of the CDF in the maximum allowed mean packet delay indicates that the system is congested and many packets are being discarded at the transmitter. This is the case of RM and SORA-RT. These packet discards cause a capacity loss in the system, as illustrated in Fig. 1(a). According to Fig. 1(d) and Fig. 1(a), the DSM, MLWDF and UEPS present a few packet discards, and so system capacity grows linearly with the offered traffic.

Finally, we can also conclude that APA does not have an impact on the behaviour of the CDFs of mean packet delays, except for the SORA-RT technique. In this case, the water filling was even slightly harmful for some users, probably because it removed some power from the users with bad channel conditions and caused a little bit more packet discards.

VI. CONCLUSIONS

In this case-study, an adaptive power allocation technique, which is a constitutive block of the utility-based RRA framework proposed is evaluated by means of system-level simulations of an OFDMA network. DSM uses a decreasing sigmoidal function based on HOL delay with inflection point in the users' HOL delay requirement, which is usually equal to the RT delay budget of the system. Its main objective is to improve satisfaction among RT users. Regarding the DSM technique, when it is compared with classical algorithms, such

as MLWDF, UEPS, RM and SORA-RT, it presents simultaneously the highest satisfaction, fairness and system capacity. Although RM presents the lowest computation complexity among all policies, the complexity of the DSM technique is also low and approximately the same of MLWDF and UEPS, and lower than SORA-RT.

Finally, we conclude that adaptive power allocation presents approximately the same performance of equal power allocation with higher complexity. It seems that most of the influence on the investigated QoS metrics is exerted by the DRA algorithm, which is the first step in our optimization solution. Therefore, for the scenario investigated in this case-study, it is more advantageous to use equal power allocation instead of adaptive power allocation. A perspective for future work is to evaluate the utility-based APA technique in Single-User (SU)-Multiple Input Multiple Output (MIMO) scenarios and investigate if APA is able to present better results.

REFERENCES

- [1] E. B. Rodrigues and F. Casadevall, "Control of the trade-off between resource efficiency and user fairness in wireless networks using utility-based adaptive resource allocation," *IEEE Communications Magazine*, vol. 49, no. 9, pp. 90–98, 2011.
- [2] J.-W. Lee and J.-A. Kwon., "Utility-based power allocation for multi-class wireless systems," *IEEE Transactions On Vehicular Technology*, vol. 58, pp. 3813–3819, 2009.
- [3] D. P. Palomar and J. R. Fonollosa, "Practical Algorithms for a Family of Waterfilling Solutions," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 686–695, 2005.
- [4] F. R. M. Lima, S. Wänstedt, F. R. P. Cavalcanti, and W. C. Freitas, "Scheduling for Improving System Capacity in Multiservice 3GPP LTE," *Journal of Electrical and Computer Engineering*, no. 819729, 2010.
- [5] J. Jang and K. B. Lee, "Transmit Power Adaptation for Multiuser OFDM Systems," vol. 21, no. 2, pp. 171–178, Jan. 2003.
- [6] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 32, no. 2, pp. 150–154, Jun. 2001.
- [7] S. Ryu, B. Ryu, H. Seo, and M. Shin, "Urgency and efficiency based wireless downlink packet scheduling algorithm in OFDMA system," in *Proc. IEEE 61st Vehicular Technology Conference - VTC Spring*, vol. 3, May 2005, pp. 1456–1462.
- [8] UMTS, "Selection procedures for the choice of radio transmission technologies of the UMTS," Universal Mobile Telecommunications System (UMTS), Sophia Antipolis, France, Tech. Rep. TR 101 112 V3.2.0 - UMTS 30.03, Apr. 1998.
- [9] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Prentice Hall, 2002.
- [10] W. C. Jakes, *Microwave Mobile Communications*. John Wiley & Sons / The Institute of Electrical and Electronics Engineers (IEEE), 1994.
- [11] 3GPP, "Deployment aspects," Third Generation Partnership Project, Tech. Rep. TR 25.943 V9.0.0, Dec. 2009.
- [12] C. Mehlführer, M. Wrulich, J. C. Ikuno, D. Bosanska, and M. Rupp, "Simulating the Long Term Evolution Physical Layer," in *Proceedings of the European Signal Processing Conference*, Glasgow, Scotland, Aug. 2009. [Online]. Available: http://publik.tuwien.ac.at/files/PubDat_175708.pdf
- [13] L. M. C. Hoo, B. Halder, J. Tellado, and J. M. Cioffi, "Multiuser transmit optimisation for multicarrier broadcast channels: asymptotic FDMA capacity region and algorithms," *IEEE Transactions on Communications*, vol. 52, no. 6, pp. 922–930, Jun. 2004.
- [14] G. Song and Y. G. Li, "Cross-layer optimization for OFDM wireless networks - part I: Theoretical framework," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 614–624, Mar. 2005.
- [15] J. Gross and M. Bohge, "Dynamic Mechanisms in OFDM Wireless Systems: A Survey on Mathematical and System Engineering Contributions," Telecommunication Networks Group (TKN), Technical University Berlin, Tech. Rep. TKN-06-001, May 2006.