Radio Resource Allocation Algorithms for the Downlink of Multiuser OFDM Communication Systems

Sanam Sadr, Alagan Anpalagan and Kaamran Raahemifar

***Abstract*—This article surveys different resource allocation algorithms developed for the downlink of multiuser OFDM wireless communication systems. Dynamic resource allocation algorithms are categorized into two major classes: margin adaptive (MA) and rate adaptive (RA). The objective of the first class is to minimize the total transmit power with the constraint on users' data rates whereas in the second class, the objective is to maximize the total throughput with the constraints on the total transmit power as well as users' data rates. The overall performance of the algorithms are evaluated in terms of spectral efficiency and fairness. Considering the trade-off between these two features of the system, some algorithms attempt to reach the highest possible spectral efficiency while maintaining acceptable fairness in the system. Furthermore, a large number of RA algorithms considers rate proportionality among the users and hence, are categorized as RA with constrained-fairness. Following the problem formulation in each category, the discussed algorithms are described along with their simplifying assumptions that attempt to keep the performance close to optimum but significantly reduce the complexity of the problem. It is noted that no matter which optimization method is used, in both classes, the overall performance is improved with the increase in the number of users, due to multiuser diversity. Some on-going research areas are briefly discussed throughout the article.**

***Index Terms*—OFDM, radio resource management, adaptive subcarrier and power allocation, fairness, rate and margin adaptive algorithms.**

I. INTRODUCTION

THE FUTURE wireless communication systems should support a large number of users with flexibility in their quality of service (QoS) requirements. The challenges to ensure the fulfillment of these requirements arise from the limited availability of frequency spectrum, the total transmit power and the nature of the wireless channel. In broadband applications, the wireless channel encounters frequency selective-multipath fading which means that the transmitted signal is scattered, diffracted and reflected, and reaches the antenna as an incoherent superposition of many signals each as a poorly synchronized echo component of the desired signal. This phenomenon leads to severe intersymbol interference (ISI) both in time and frequency impacting the service quality and data rates. To solve this issue, intelligent radio resource management algorithms interacting in both the physical and the media access control (MAC) layers are essential with the ability to combat ISI.

Orthogonal frequency division multiplexing (OFDM) is one of the promising solutions to provide a high performance physical layer and has been widely adopted in standards by wireless industry. Examples include IEEE 802.11a and IEEE 802.11g Wireless Local Area Networks (WLANs). It is also the physical layer specification for IEEE 802.16 [1] fixed wireless metropolitan area networks (WMANs) which is later extended in IEEE 802.16e(WiMAX) [2] to accommodate high speed mobility and to support both fixed and mobile subscriber stations.

OFDM is based on the concept of multicarrier transmission. The idea is to divide the broadband channel into N narrowband subchannels each with a bandwidth much smaller than the coherence bandwidth of the channel. The high rate data stream is then split into N substreams of lower rate data which are modulated into N OFDM symbols and transmitted simultaneously on N orthogonal subcarriers [3]. The low bandwidth of the subchannels along with the frequency spacing between them are necessary to have flat fading orthogonal subcarriers with approximately constant channel gain during each transmission block.

In a single user system, the user can use the total power to transmit on all N subcarriers; the system is then optimized by exploiting the frequency selectivity of the channel and dynamically adapting the modulation type and transmit power on each subcarrier. These dynamic power allocation schemes [4], [5] have shown significant performance gain in terms of throughput compared to static schemes.

In a wireless network, on the other hand, new challenges arise as the number of the users in the system increases. These challenges include dynamic subcarrier allocation, adaptive power allocation, admission control and capacity planning [6]. The first two are also referred to as radio resource allocation and is the subject of this article.

In a multiuser OFDM system, there is a need for a multiple access scheme to allocate the subcarriers and the power to the users. In static subcarrier allocation schemes, each user is assigned predetermined time slots or frequency bands respectively regardless of the channel status. In other words, in non-adaptive fixed subcarrier allocation schemes, an independent dimension is allocated to each user without considering the channel status. In such systems, the optimization problem of

Manuscript received 11 January 2008; revised 8 July 2008. This work was supported in part by a grant from National Science and Engineering Research Council of Canada.

The authors are with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, Canada (e-mail: alagan@ee.ryerson.ca).

Digital Object Identifier 10.1109/SURV.2009.090307.

maximizing the total throughput of the system reduces to only power allocation or bit loading on the subcarriers. On the other hand, since the fading parameters for different users are mutually independent, the probability that a subcarrier is in deep fade for all users is very low. In other words, each subcarrier is likely to be in good condition for some users in the system. This is the principle of Orthogonal Frequency Division Multiple Access (OFDMA) scheme with adaptive power allocation in which subcarrier allocation itself plays a very significant role in maximizing the total throughput by using multiuser diversity. A survey of adaptive OFDMA system design problems including an overview of physical layer, MAC and radio resource management design issues are provided in [7].

In this article, we provide an overview of different adaptive resource (bandwidth and power) allocation algorithms and compare them in terms of their objectives, performance and complexity. First, we start by formulating the general problem of resource allocation in multiuser OFDM systems in a frequency selective fading channel and then discuss the idea behind major classes of algorithms. We then proceed to discuss the different techniques within each category.

II. System Model

In this section, the parameters and the basic assumptions regarding the system and the channel under consideration are introduced. The problem of resource allocation in a multiuser OFDM system is formulated next.

A. *FFT-based Transceiver*

In the downlink of a multiuser OFDM system, the basestation should communicate with multiple users with limited resources i.e., bandwidth and power. Fig. 1 shows the main building blocks in a multiuser OFDM transmitter with adaptive resource allocation.

Using the channel information, the transmitter applies the combined subcarrier, bit and power allocation algorithm to assign subcarriers to different users and the number of bits/OFDM symbol to be transmitted on each subcarrier. The power allocated to each subcarrier is then determined by the number of bits assigned as well as the corresponding modulation scheme. The complex symbols at the output of the modulators are transformed into an OFDM symbol in time domain by the inverse fast Fourier transform (IFFT) in the transmitter. Before transmission, a cyclic prefix which is the copy of the last portion of the data symbol is added to the time domain samples. The cyclic extension is sized appropriately to serve as a guard interval to ensure the orthogonality between the subcarriers. This type of multicarrier modulation is also referred to as *cyclic prefix OFDM* (CP-OFDM). The ISI could be eliminated provided that the amount of time dispersion from the channel is smaller than the duration of the guard interval. Other techniques to maintain the orthogonality of the subcarriers could be found in [7, Chapter II]. CP-OFDM is of particular interest as it gives the best performance-complexity trade-off [7].

Along with each OFDM symbol, the subcarrier and bit allocation information are sent to the receiver via a separate control channel; therefore, the users need only to decode the bits on their respective assigned subcarriers. In the receiver, the guard interval is removed and the time samples are transformed into modulated symbols by means of fast Fourier transform (FFT). Then, the bit allocation is used to configure the demodulators whereas the subcarrier information is used to extract the demodulated bits from the subcarriers assigned to each user.

The problem of resource allocation in an OFDMA system with N subcarriers and K users is to determine the elements of matrix $\mathbf{C}=[c_{k,n}]_{K\times N}$ specifying which subcarrier should be assigned to which user and vector $\mathbf{p}=[p_n]_{N\times 1}$ specifying how much power should be allocated to each subcarrier. An overview of the problem is shown in Fig. 2.

To determine the elements of matrix $\mathbf{C}$, it is assumed that subcarriers are not shared by different users based on the theorem proved in [8] showing that the data rate of a multiuser OFDM system is maximized when each subcarrier is assigned to $only$ one user that has the best channel gain for that subcarrier and the power is distributed among the subcarriers using water-filling [9]. The elements of vector $\mathbf{p}$ are determined based on the power constraints of the system. While only one total power constraint exists in the downlink of a multiuser system, multiple power constraints exist for the uplink depending on the number of users. Similar to downlink, the objective of resource allocation in uplink is to maximize the total data rate except for the additional concern regarding power control to minimize inter-cell interference for the users on the cell-edge. [10].

Since the requirements for the future services continue to evolve, it is of great importance that the developed techniques are flexible enough to handle as high bit rates as possible with various quality requirements [11]. To do so, many dynamic resource allocation algorithms and optimization techniques have been proposed for the downlink of an OFDMA system which differ in their objectives and constraints.

B. *Channel Characteristics*

In wireless signal transmission where the signal is transmitted over a radio channel, multipath fading is a common phenomenon especially in urban and suburban areas where the communication environment changes quickly. In those areas where there is no direct line-of-sight between the transmitter and the receiver, multipath reflections occur from different objects which result in the electromagnetic wave travelling along different paths of varying length. The interaction between those waves causes multipath fading with frequency selectivity where the fading parameter changes with frequency. As a result, the wireless channel is assumed to be wideband time-varying frequency-selective multipath fading.

An extensive overview of statistical analysis and information-theoretic and communications features of fading channels has been presented in [12]. One of the parameters to characterize and hence simplify the response complexity of multipath fading channels in frequency domain is the coherence bandwidth of the channel defined as a range of frequencies over which the channel can be considered flat [13]. By choosing the bandwidth of the subchannels much

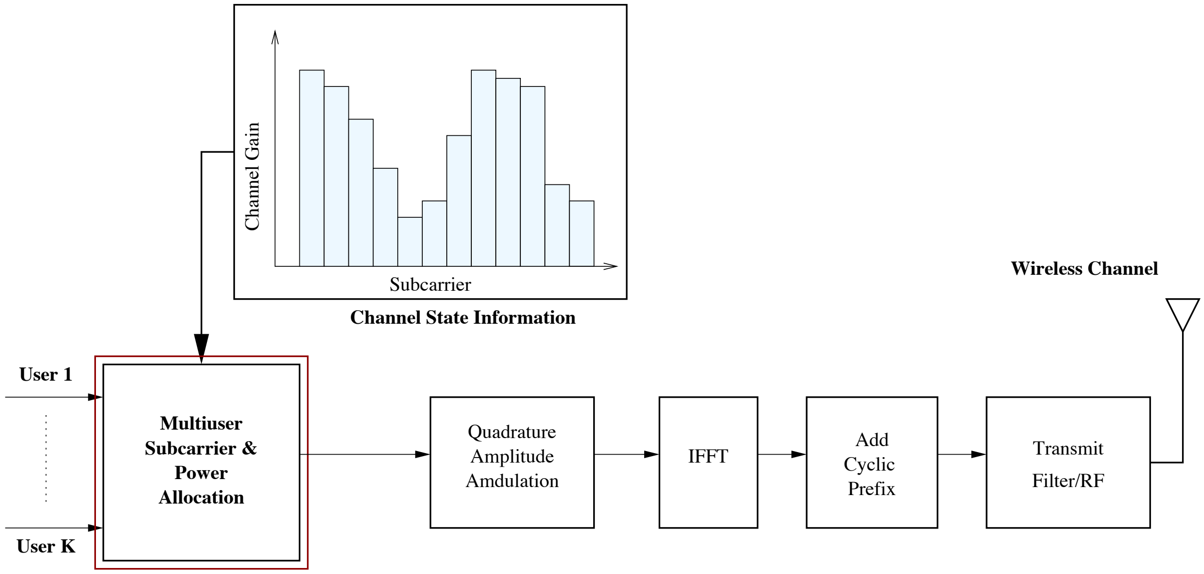


Fig. 1. Block diagram of a multiuser OFDM transmitter with adaptive subcarrier and power allocation.

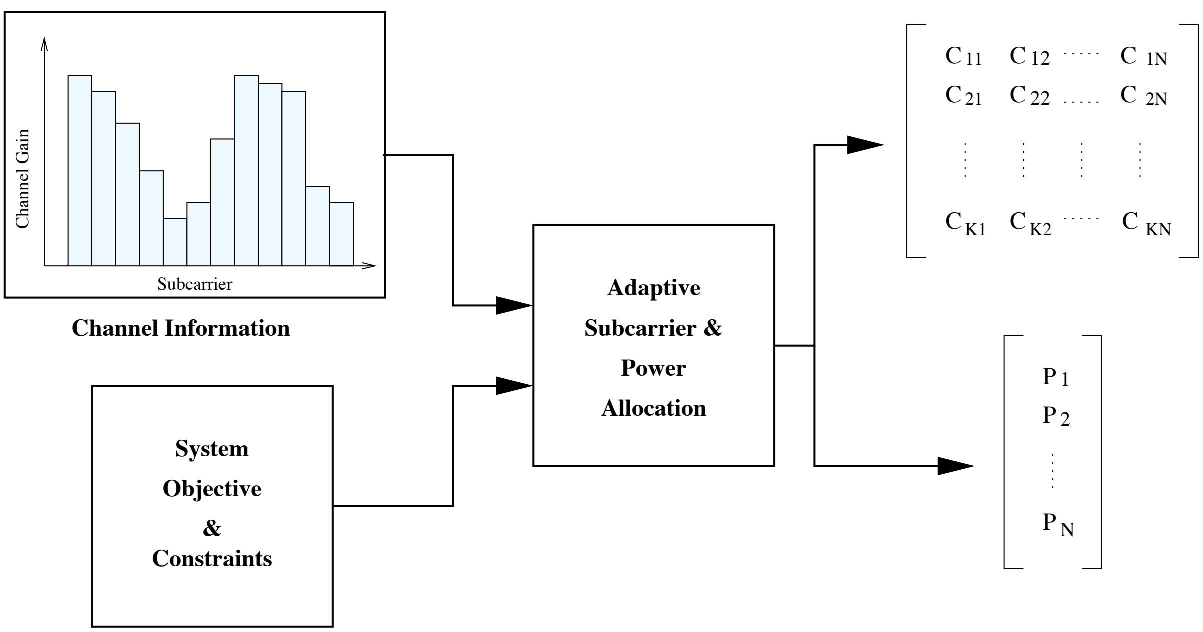


Fig. 2. Overview of the problem of resource allocation in OFDMA systems.

smaller than the coherence bandwidth of the channel, each subchannel can be assumed to undergo flat fading. One of the widely used models to explain the statistical nature of flat fading channels is Clark's model based on scattering [13]. According to this model, the fading parameter of the channel is considered to be a random variable with Rayleigh distribution. In modelling the channel, it is also assumed that additive white Gaussian noise (AWGN) is present for all subcarriers of all users.

The advantages of adaptive resource allocation in multiuser OFDM systems are partially due to multiuser diversity which is based on assigning each subchannel to the user with good channel gain on it. To do so, it is assumed that users perfectly estimate and feedback their channel information to the basestation and the channel condition is always available to the basestation in the beginning of each transmission block. Also, it is assumed that the fading rate of the channel is slow enough such that the time-varying channel can be considered quasi-static where the channel condition does not change within each OFDM transmission block. Otherwise, the resources would be assigned based on channel condition which has already changed hence making the expected performance outdated.

The algorithms discussed in this article are based on the assumption of perfect channel information at both the transmitter and the receiver. While this is a reasonable assumption in wireline systems where the channel remains invariant, in

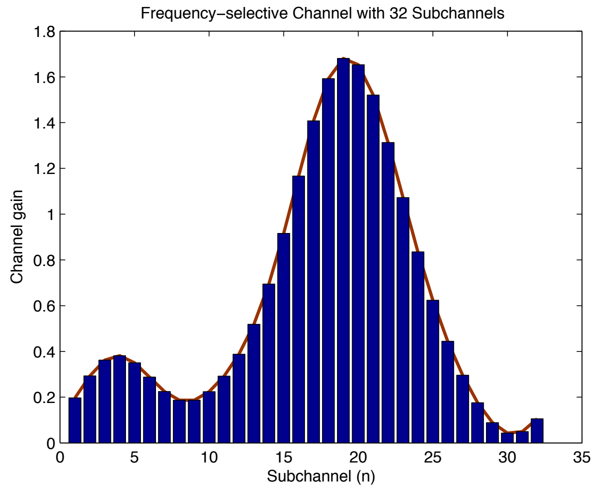


Fig. 3. Snapshot of a frequency selective channel with thirty two subcarriers, ($N = 32$).

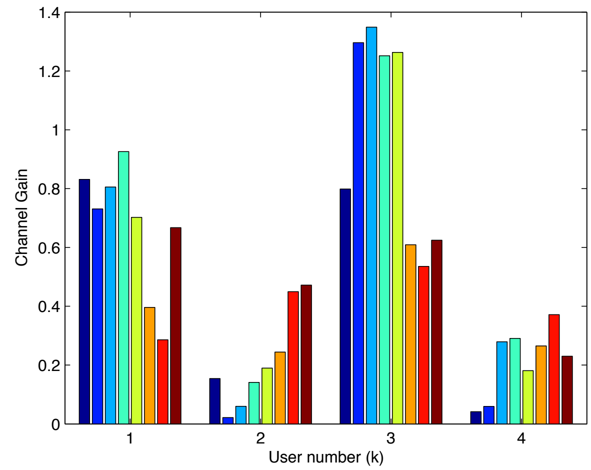


Fig. 4. Snapshot of a wireless channel with eight subcarriers and four users, ($N = 8, K = 4$).

wireless transmission, it is rarely possible for the transmitter to acquire perfect channel state information (CSI). This inaccuracy is due to channel estimation errors and channel feedback delay also referred to as channel mismatch errors. The latter is due to the variations of the wireless channel once it has been estimated. A typical channel estimation is accurate enough to justify the use of adaptive modulation and resource allocation; it is the delay in channel feedback that results in outdated channel information and invalidates the expected performance.

Channel prediction algorithms for flat fading channels have been investigated extensively in the past several years e.g., [14] whereas more recent studies have been focused on channel prediction methods in the context of OFDM [15–20]. While multiple estimates in time or frequency have been shown to improve the average spectral efficiency and robustness of the system to larger estimation error and longer delay [21], the feedback channel information would increase the system overhead; therefore, some authors [22], [23] have investigated channel prediction methods that reduce the amount of feedback overhead.

Along with the recent studies on channel prediction methods and algorithms, many researchers have investigated the impact of non-ideal CSI knowledge on design and performance of radio resource allocation strategies. Leke et al. [24] studied the impact of channel mismatch errors and derived analytical expressions for the probability of error of a multicarrier system in a Rayleigh fading channel in the presence of these errors. The effect of imperfect CSI on rate maximization of a single-user OFDM wireless system has been well studied in [25], [26]. It was shown in [26] that adaptive OFDM is more sensitive to channel estimation error than OFDM with uniform modulation. Also for relatively slow varying channels, the advantages of adaptive over non-adaptive OFDM is still significant even in the presence of channel estimation error.

The problem of adaptive resource allocation in a multiuser system with imperfect or partial CSI was addressed in [27], [28] where the ergodic sum utility is considered under QoS requirements and the total power constraint. Brah et al. [27] characterized the channel estimation error as additive Gaussian noise, independent of the channel itself. It was shown in [27] that even with imperfect CSI, adaptive resource allocation improves the performance in OFDMA systems.

In an OFDMA system, the subcarriers are assigned to the users based on instantaneous channel information and the constraints of the system. Fig. 3 shows a snapshot of a frequency selective channel with thirty two subcarriers. A snapshot of a wireless channel with eight subcarriers and four users is shown in Fig. 4.

These two snapshots indicate two important properties of subchannel gains in a multiuser frequency selective fading channel. Firstly, different subcarriers of each user suffer from different fading levels due to frequency selectivity of the channel a.k.a. frequency diversity (per user). Secondly, the subchannels of different users vary independently due to different locations of the users a.k.a. multiuser diversity. Using the channel information, the transmitter performs the subcarrier and power allocation to achieve the best performance in the system.

C. *Adaptive vs Fixed Modulation*

Each subcarrier in a multiuser OFDM system can potentially have a different modulation scheme and each modulation scheme provides a trade-off between spectral efficiency and BER. In those OFDM systems where a fixed modulation scheme is used over all subcarriers, the carrier modulation is designed such that it maintains acceptable performance when the channel quality is poor. Thus, these systems are effectively designed for the worst channel conditions. This results in such systems using BPSK or QPSK with poor spectral efficiency of 1 or 2 bits/s/Hz respectively. However, as mentioned earlier, each subcarrier in a multipath frequency selective channel experiences different fade which might result in received power variation of as much as 30dB [29]. Consequently, for a subcarrier with good channel gain, the modulation can be increased to 16-64 QAM significantly increasing the spectral efficiency of the overall system to 4-6 bits/s/Hz. In other words, using adaptive modulation, the subcarrier modulation is matched to its channel signal to noise ratio (SNR) maximizing the overall spectral efficiency.

There are several limitations with adaptive modulation. Adaptive modulation requires accurate channel estimates at the receiver and a reliable feedback path between the receiver and transmitter [30]. If the channel is changing faster than it can be estimated and fed back to the transmitter, adaptive techniques will perform poorly. Also, overhead information needs to be exchanged and updated regularly, as both the transmitter and receiver must know what modulation is being used which further increases the overhead with the mobility of the receiver.

D. *Efficiency and Fairness*

Efficiency and fairness are two crucial issues in resource allocation for wireless communication systems. Spectral efficiency is defined as the data rate per unit bandwidth and is calculated by dividing the total throughput of a system by its total bandwidth. Therefore, it takes into account the total data rate rather than each user's achieved data rate. A system might achieve the highest throughput, hence the highest spectral efficiency while being unfair to those users far away from the basestation or with bad channel conditions. Fairness, on the other hand, indicates how equally the resources are distributed among the users. There is always a trade-off between efficiency and fairness in wireless resource allocation.

Fairness could be defined in terms of different parameters of the system. It could be defined in terms of bandwidth where each user is assigned an equal number of subcarriers [31], or it could be in terms of power where each user is allocated equal portion of the power from the budget. It could also be in terms of data rate where the objective is to allocate the resources to the users such that all the users achieve the same data rate [32]. When the objective is to ensure rate proportionality among the users, it is called optimization with constrained-fairness [33].

Shen et al. [33] defined the fairness index in terms of rate proportional constraints with the maximum value of 1 to be the fairest case in which all users would achieve the same data rate. Based on this definition, a new parameter was defined in [34] to examine the performance of the system in maintaining proportional fairness. The new parameter is in terms of proportional rate constraint as well as the achieved data rate for each user and is a real number in the interval (0,1]. Again, the maximum value represents the case in which the achieved rate proportions among the users are the same as the predetermined rate proportional constraints.

In formulating the optimization problem of resource allocation, it is assumed that there are K active users in the system all the time, each with a different quality of service requirement in terms of data rate and BER. Also, when scheduled for transmission, they always have some data to transmit.

III. General Problem of Resource Allocation in Multiuser OFDM Systems

Consider a multiuser OFDM system with K users and N subcarriers. K = $\{1, 2, ..., K\}$ and N=$\{1, 2, ..., N\}$ are the set of users and subcarriers respectively. The data rate of the kth user R_k in bits/s is given by:

$$R_k = \frac{B}{N}\sum_{n=1}^{N} c_{k,n}\log_2(1+\gamma_{k,n}), \tag{1}$$

where B is the total bandwidth of the system and $c_{k,n}$ is the subcarrier assignment index indicating whether the kth user occupies the nth subcarrier. $c_{k,n}$=1 only if subcarrier n is allocated to user k; otherwise it is zero. The bandwidth of each subchannel is $\frac{B}{N} = \frac{1}{T}$ where T is the OFDM symbol duration. Note that as the symbol duration is increased, the relative amount of dispersion in time caused by multipath delay spread decreases [8]. $\gamma_{k,n}$ is the SNR of the nth subcarrier for the kth user and is given by:

$$\gamma_{k,n} = p_{k,n}H_{k,n} = \frac{p_{k,n}h_{k,n}^2}{N_0\frac{B}{N}}, \tag{2}$$

where $p_{k,n}$ is the power allocated for user k in subchannel n and $h_{k,n}$ and $H_{k,n}$ denote the channel gain and channel-to-noise ratio for user k in subchannel n respectively. $N_0\frac{B}{N}$ is the noise power on each subcarrier with N_0 as the power spectral density of AWGN.

Eq. (1) is the data rate achieved by the kth user in a zero margin system. In practical modulation schemes however, the effective SNR has to be adjusted according to the modulation scheme for a desired BER. The power loss which is the difference between the SNR needed to achieve a certain data transmission rate for a practical system and the theoretical limit is called the SNR gap. The BER for an AWGN channel with MQAM modulation and ideal coherence phase detection is bounded by [35]:

$$\text{BER} \le 2e^{-1.5\gamma/(M-1)}, \tag{3}$$

where $M = 2^r$ and r denotes the number of bits. γ is the SNR defined as in (2). If $r \ge 2$ and $0 \le \gamma \le 30$ dB, BER could be better approximated within 1 dB by [30]:

$$\text{BER} \le 0.2^{-1.5\gamma/(M-1)}. \tag{4}$$

Using (4), the number of bits r is given by:

$$r = \log_2\left(1+\frac{\gamma}{\Gamma}\right), \tag{5}$$

where Γ is the SNR gap and a function of BER:

$$\Gamma = \frac{-\ln(5\text{BER})}{1.5}. \tag{6}$$

From (1), the total data rate R_T of a zero margin system is given by:

$$R_T = \frac{B}{N}\sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n}\log_2\left(1+\gamma_{k,n}\right). \tag{7}$$

Knowing the modulation scheme, the effective SNR $\gamma_{k,n}$ is adjusted accordingly to meet the BER requirements.

The general form of the subcarrier and power allocation problem is shown below:

$$
\begin{aligned}
&\text{Objective :}\\
&\max_{c_{k,n},p_{k,n}} \quad R_T = \frac{B}{N}\sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n}\log_2\left(1+\frac{p_{k,n}h_{k,n}^2}{N_0\frac{B}{N}}\right),\\
&\qquad\qquad \text{or}\\
&\min_{c_{k,n},p_{k,n}} \quad P_T = \sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n}p_{k,n}.\\
&\text{subject to :}\\
&\qquad C1: c_{k,n}\in\{0,1\},\ \ \forall k,n\\
&\qquad C2: \sum_{k=1}^{K} c_{k,n}=1,\ \ \forall n\\
&\qquad C3: p_{k,n}\geq 0,\ \ \forall k,n\\
&\qquad C4: \sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n}p_{k,n}\leq P_{total},\\
&\qquad C5: \text{User Rate Requirements.}
\end{aligned}
\tag{8}
$$

The problem could be formulated with two possible objectives followed by various constraints (C1-C5). The first two constraints are on subcarrier allocation to ensure that each subchannel is assigned to only one user. C4 is only effective in problems where there is a power constraint P_{total} on the total transmit power of the system P_T (e.g., rate adaptive algorithms). C5 determines fixed or variable rate requirements of the users.

Since the requirements of the future services are still unknown, each class of algorithms has considered different objectives and constraints to further investigate the performance and flexibility of dynamic resource allocation in multiuser OFDM systems. In each class, the problem is formulated accordingly and the optimal solution is derived using different optimization techniques. Due to high computational complexity of the optimal solutions, they may not be practical in real time applications. As a result, suboptimal algorithms have been developed which differ mostly in:

- the approach they choose to split the procedure into several (preferably independent) steps to make the problem tractable and,
- their simplifying assumptions to reduce the complexity of the allocation process.

The performance of each algorithm highly depends on the formulation of the problem and the validity of these simplifying assumptions. Fig. 5 gives a summary of different classes of resource allocation algorithms developed in multiuser OFDM systems.

IV. Classes of Dynamic Resource Allocation in Multiuser OFDM Systems

Two major classes of dynamic resource allocation schemes have been reported in literature: 1) margin adaptive (MA) [36–38], 2) rate adaptive (RA) [8], [32], [33], [39–45]. The optimization problem in margin adaptive allocation schemes is formulated with the objective of minimizing the total transmit power while providing each user with its required quality of service in terms of data rate and BER. The rate adaptive objective is to maximize the total data rate of the system with the constraint on the total transmit power.

While the sum capacity of a system provides a good measurement of the spectral efficiency, it is not a valid indication of each user's satisfaction in a multipath fading channel. It was proved in [8] and [40] that the total throughput of a multiuser system is maximized if each subchannel is assigned to the user with the best channel gain on it and the power is distributed using water-filling policy. However, when the path loss differences among users are large, the users with higher channel gains will be allocated most of the resources while leaving less for the users with low channel gains. Therefore, rate adaptive algorithms are divided into two major groups based on the user rate constraints. In the first group, there is a fixed rate requirement for each user. The algorithms in this group (e.g., [39]) attempt to maximize the total throughput of the system while supporting each user with its fixed rate requirement. A large number of RA algorithms [8], [32], [33], [40–42], [44], [45] fall into the second group where they consider the concept of fairness or constrained-fairness among the users. In this category, while the objective is to maximize the total throughput within the power budget, the goal is to maintain the rate proportionality among the users according to proportional constraints rather than reaching a specific requested data rate.

In the following sections, major categories of resource allocation are discussed starting with the formulation of the problem in each category.

A. *Rate Adaptive Algorithms*

Song et al. [40] used the concept of utility functions $U(.)$ to formulate the problem of resource allocation in multiuser OFDM systems. Utility function maps the network resources a user utilizes into a real number and is a function of the user's data rate. The utility-based dynamic resource allocation problem is formulated as:

$$
\begin{aligned}
&\max_{c_{k,n},p_{k,n}} \quad \sum_{k=1}^{K} U_k(R_k)\\
&\text{subject to:}\\
&\qquad C1: S_i\cap S_j=\emptyset\quad \forall i,j\in\kappa,\ \ i\neq j\\
&\qquad C2: \bigcup_k S_k\subseteq\{1,2,...,N\}\\
&\qquad C3: p_{k,n}\geq 0\quad \forall k,n\\
&\qquad C4: \sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n}p_{k,n}\leq P_{total}
\end{aligned}
\tag{9}
$$

where $U_k(R_k)$ is the utility function for the kth user. R_k is defined as in (1), S_k is the set of subcarriers assigned to user k for which $c_{k,n}$ =1 and $\bigcup_k S_k$ is the union of all subcarrier sets. As in (8), the first two constraints ensure no sharing in subcarriers. The objective is to maximize the sum utility within the power budget P_{total}.

An extreme case of the problem formulated in (9) is obtained when there are infinite number of orthogonal subcarriers each with an infinitesimal bandwidth within the total available bandwidth B. In this extreme case, the whole bandwidth [0,B] is divided into several non-overlapping frequency sets that are

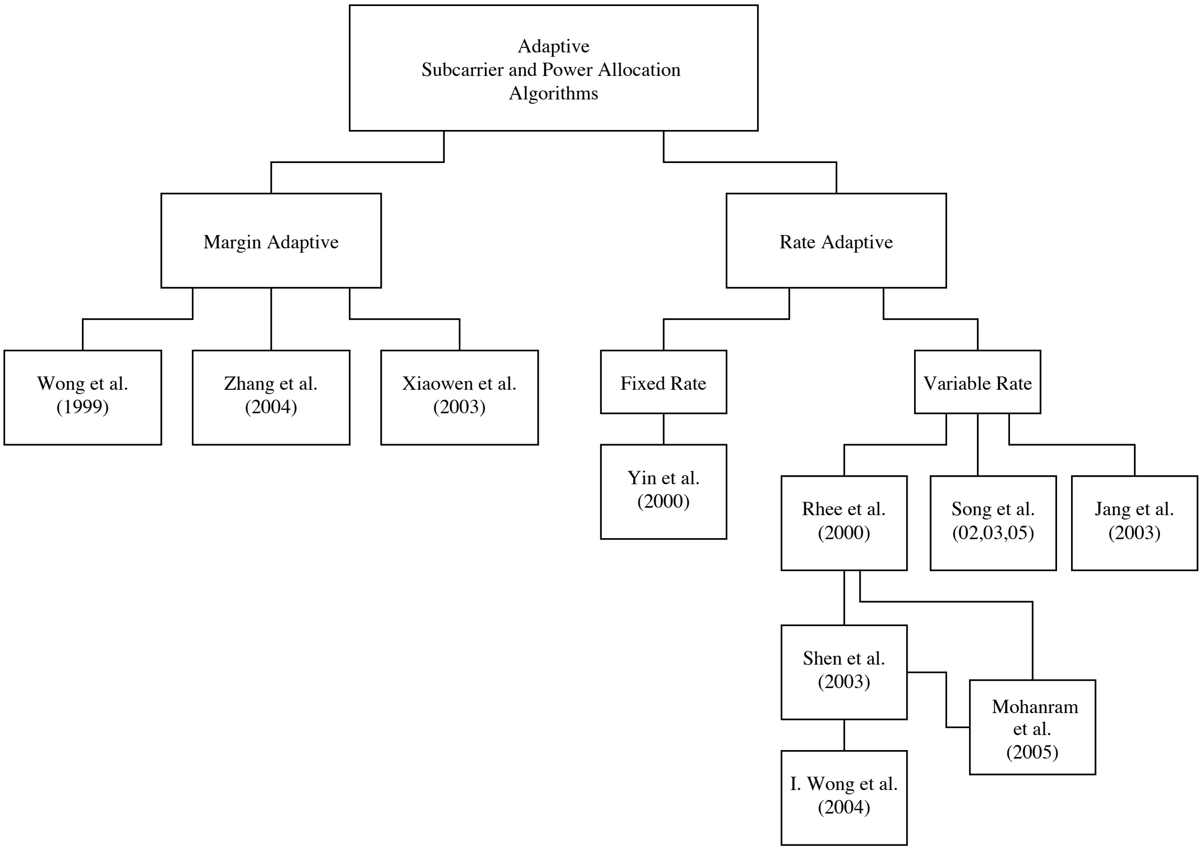


Fig. 5. Different classes of algorithms developed for resource allocation in multiuser OFDM systems.

assigned to the users. D_k is defined as the set or the union of frequency sets occupied by user k. The continuous data rate for the kth user in bits/s is given by:

$$R_k^c = \int_{D_k} C_k(f)df, \tag{10}$$

where $C_k = \log_2(1+\gamma_k(f))$ is the achievable rate for user k in bits/s/Hz and $\gamma_k(f)$ is the signal to noise ratio of the kth user as a continuous function of frequency f.

Song et al. investigated the extreme case in two theorems [41, theorem I and II]. (Also see [42] for the proof of Theorem I). Theorem I gives the optimal subcarrier allocation assuming a fixed power allocation on all the subcarriers. Theorem II gives the optimal power allocation given a fixed subcarrier allocation. Combining the results of the two theorems, the optimal frequency set and the power allocation for the extreme case are given by:

$$D_k^* = \bigcap_{i\in\kappa\ i\neq k} \{f \in [0,B] : U_i^{'}(R_i^*)C_i(f) \leq U_k^{'}(R_k^*)C_k(f)\} \tag{11}$$

with the power allocation satisfying:

$$\begin{cases} p^*(f) = \left[\frac{U_k^{'}(R_k^*)}{\lambda} - \frac{1}{\gamma_k(f)}\right]^+, \quad f \in D_k^* \\ \int_0^B p^*(f) = P_{total} \\ R_k^* = \int_{D_k} \log_2\left(1+p^*(f)\gamma_k(f)\right)df, \end{cases} \tag{12}$$

where $[x]^+ = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$.

In the above equations, D_k^* and R_k^* are the optimal frequency set and data rate for the kth user respectively and $p^*(f)$ denotes the optimal power allocation within the bandwidth. Interestingly, the power allocation is similar to the classical single user water-filling but with the level of water proportional to the marginal utility value at the optimal rate $U_k^{'}(R_k^*)$. Therefore, the power allocation is dependent on the choice of utility function. Furthermore, the power constraint is the total power of the system rather than a single user's total power.

When deriving an optimal solution for any problem, there is always the question of uniqueness of the optimal solution that should be taken into consideration. The same issue should be discussed for the utility-based optimization problem described in (9). It is shown in [41] that for concave utility functions, the necessary conditions summarized in (11) and (12) are also sufficient and the global optimality of the solution is guaranteed. Furthermore, if $U_k(R_k)$'s are all strictly concave, then there exists a unique global optimal solution satisfying the above conditions. However, the uniqueness of the optimal solution does not guarantee the uniqueness of the subcarrier or power allocation schemes leading to that solution. In other words, for strictly concave utility functions, there is only one optimal data rate set of $\{R_k\}_{k=1}^K$ denoting the achieved rate for different users, but there might be several frequency and power allocation schemes leading to the same data rate set.

If the utility function is chosen to be the data rate, i.e., $U_k(R_k) = R_k$, then the problem is transformed into:

$$\max_{c_{k,n},p_{k,n}} \quad R_T = \sum_{k=1}^{K} R_k. \tag{13}$$

With $U_k^{'} = 1$, the optimal solution is then given by:

$$\begin{cases} D_k^* = \{f \in [0, B] : \gamma_k(f) = \max_{i\in\kappa} \gamma_i(f)\} \\ p^*(f) = \left[\frac{1}{\lambda} - \frac{1}{\gamma_k(f)}\right]^+, \quad f \in D_k^* \\ \int_0^B p^*(f) = P_{total} \end{cases} \tag{14}$$

which is the continuous counterpart to the results in [8]. In other words, if the utility function is just data rate, the optimal power allocation is *independent* from the optimal subcarrier allocation. Each subcarrier is assigned to the user with the largest channel gain and the water-filling policy is applied. Furthermore, assignment of a subcarrier has no effect on the assignments of other subcarriers.

The case of resource allocation in practical OFDM wireless networks with finite number of subcarriers using utility functions is investigated in [43]. The optimization problem then turns from continuous to discrete with the data transmission rate of the kth user given by:

$$R_k = \frac{B}{N} \sum_{k\in D_k} C_k[n]. \tag{15}$$

where $C_k[n]$ is the achievable data rate for the kth user on the nth subcarrier in bits/s/Hz.

It is shown in [43] that the optimality conditions for the continuous case summarized in (11) and (12) also hold for the discrete frequency case. The only difference is that these conditions are sufficient for optimality but are not necessary anymore.

In [43], Song et al. developed several algorithms for dynamic subcarrier allocation with fixed power allocation (DSA), adaptive power allocation with fixed subcarrier allocation (APA) and joint DSA and APA with continuous rate adaptation which does not assume any specific data modulation. A greedy power allocation algorithm based on maximizing total utility for discrete rate adaptation has also been proposed in [43] where there are discrete modulation levels.

The most important decision to make in a utility-based optimization problem is to choose the utility function properly according to the *objective* of the system. Since in almost all wireless applications the most important factor to determine a user's satisfaction is its reliable data transmission rate, the utility function is chosen to be a non-decreasing function of the rate. If the utility function is chosen to be the data rate, for instance, the total capacity and hence the spectral efficiency is maximized using (14) (or its discrete counterpart given in [8] and [43]). However, with no constraint on users' minimum data rates, the users with poor channel conditions are penalized. To maintain fairness, the utility function should be chosen to prioritize the users with low data rate.

1) Rate Adaptive Algorithms with Fairness: One way to accomplish both efficiency and fairness is to use utility functions that are both increasing and marginally decreasing. As a result, the slope of the utility curve decreases with an increase in the data rate. Choosing a marginally decreasing utility function also guarantees its strictly concavity which ensures the global optimality as well as uniqueness of the optimal solution. A logarithmic utility function $U(R) = \ln(R)$ is both increasing and marginally decreasing. Therefore, a resource allocation policy using a logarithmic utility function is said to be proportionally fair [46]. Different types of utility functions have been proposed in [42], [43], [46], [47] depending on the type of application. Choosing the proper utility function which ensures both efficiency and fairness is better obtained through subjective survey rather than pure theoretical derivation.

The problem of maximizing the total throughput with fairness was formulated differently in [32] and [33]. Rhee and Cioffi [32] studied the max-min problem, whereby maximizing the worst user's capacity, it is assured that all users achieve the same data rate. Shen et al. [33] considered this problem by introducing proportional constraints among the users' data rates. Their proposed algorithms were further modified in [45] and [48] to reduce the complexity.

Considering the general resource allocation problem formulated in (8), the optimization problem with variable rate constraints is given by:

$$\begin{aligned} &\text{Objective :} \\ &\max_{c_{k,n},p_{k,n}} \quad \frac{B}{N}\sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n}\log_2\left(1+\frac{p_{k,n}h_{k,n}^2}{N_0\frac{B}{N}}\right), \\ &\text{subject to :} \\ &\qquad \text{C1} : c_{k,n} \in \{0,1\}, \ \ \forall k,n \\ &\qquad \text{C2} : \sum_{k=1}^{K} c_{k,n} = 1, \ \ \forall n \\ &\qquad \text{C3} : p_{k,n} \geq 0, \quad \forall k,n \\ &\qquad \text{C4} : \sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n}p_{k,n} \leq P_{total}, \\ &\qquad \text{C5} : R_1 : R_2 : ... : R_K = \alpha_1 : \alpha_2 : ... : \alpha_K, \end{aligned} \tag{16}$$

where the objective is to maximize the total rate within the total power constraint of the system while maintaining rate proportionality among the users indicated in C5. Here, $\{\alpha_1, \alpha_2, ...\alpha_K\}$ is the set of predetermined proportional constraints where α_k is a positive real number with $\alpha_{min} = 1$ for the user with the least required proportional rate. When all α_k terms are equal, the objective function in (16) is similar to the objective function of max-min problem introduced in [32].

In a system with K users and N subcarriers, each of N subcarriers is to be allocated to one of K users. In addition, the power allocated to each of K users should be optimized. Therefore, there are $K + N$ parameters to optimize to achieve the optimal solution and K^N possible subcarrier allocations assuming no subcarrier can be used by more than one user. Ideally, the subcarrier and power allocation should be carried out jointly which leads to high computational complexity necessitating suboptimal algorithms.

To solve this problem, a very simple but highly efficient subcarrier allocation algorithm was proposed in [32]. This

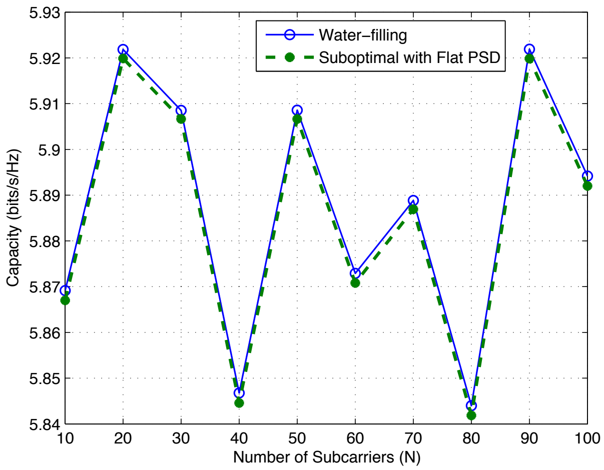


Fig. 6. Achieved data rate versus the total number of subcarriers, N.

algorithm is based on flat transmit power. It was suggested in [49] that in a single user water-filling solution, the total data rate of a zero margin system is close to capacity even with flat transmit power spectral density (PSD) as long as the energy is poured only into subchannels with good channel gains. This is a very important result since it completely eliminates the major step of power allocation concentrating mainly on subcarrier allocation. In addition to reducing the complexity, a flat transmit power would also be necessary in case there is a power mask constraint on each subcarrier as well as on the total transmit power [32].

Several authors have investigated the effect of such simplification on the performance of the system. One way to gain insight into the validity of the simplifying assumption of flat transmit power is to examine the performance of optimal (water-filling) and suboptimal power allocation (with flat transmit power) in a single user OFDM system. In [49], a very simple algorithm is proposed to find the best number and set of subcarriers with equal power for the objective of maximizing the overall data rate of a single user system under the total power constraint. The comparison of the achieved data rates of two power allocation methods is shown in Fig. 6. The first algorithm uses water-filling to distribute power among the subcarriers. In the second algorithm [49], the best set of subcarriers is selected and the total power is equally distributed among those subcarriers involved in data transmission. Both algorithms use the same channel information as the subchannel gains.

Based on this result, one may conclude that in a system with N subchannels, a flat transmit power over all the subcarriers would always give close to optimum performance. However, this is not always the case. The rate difference between the optimal and suboptimal power allocation is negligible only when in the suboptimal algorithm, the optimal number and set of subcarriers are chosen to transmit the data and the total power is equally distributed among these subcarriers while the rest of subchannels are allocated no power. In the case of multiuser OFDM systems, there is no doubt that the dynamic power allocation would increase the total throughput. The question however, is whether the achieved performance gain is high enough to justify the significant additional computational load.

Bohge et al. [50] discussed the performance of four different cases in a multiuser OFDM system with 48 subcarriers and 16 users uniformly distributed in a single cell. Combining flat and dynamic power allocation with fixed and adaptive subcarrier allocation, they considered four cases of flat transmit power with fixed subcarrier assignment, flat transmit power with adaptive subcarrier assignment, dynamic power allocation with fixed subcarrier assignment and dynamic power allocation with adaptive subcarrier assignment. In all the cases, adaptive modulation with a fixed target BER were considered.

Interesting results were reported in [50, Fig. 2]. In general, as the transmit power increases so does the total throughput of the system. However, if the total transmit power is low or the average attenuation among different users in the cell is high (i.e., a large cell is considered), the gain obtained from the dynamic power allocation has been shown to be quite significant. Also, dynamic power allocation has a larger performance improvement if it is applied with adaptive subcarrier allocation rather than fixed subcarrier assignment. Therefore, in large cells with high channel gain differences among the users, a fully dynamic approach (adaptive power and subcarrier allocation) has been recommended due to its high additional performance.

The reason that in multiuser OFDM systems, a flat transmit power might also perform well is that it is assumed that due to multiuser diversity, only subchannels with good channel gains are assigned to each user. Therefore, almost all the subchannels involved in data transmission are in good channel condition.

Based on this assumption, a flat transmit PSD was used in [32] indicating that the power allocated to each subcarrier is constant and equal to $\frac{P_{total}}{N}$. The resource allocation then reduces to only subcarrier allocation with N optimization parameters. In the process of subcarrier allocation, two goals take place alternatively: 1) maintaining fairness among the users by giving priority to the user with the least achieved rate to choose the next subcarrier; 2) increasing the total data rate by allocating the best available subcarrier to that user. The suboptimal algorithm proposed in [32] showed $50 \sim 130\%$ of capacity gain over a non-adaptive TDMA resource allocation scheme and less than 4% spectral efficiency loss compared to optimal solution with adaptive power allocation. This algorithm achieves acceptable fairness as long as the number of subcarriers is much larger than the number of users i.e., $N \gg K$.

Though acceptable fairness amongst users is achieved in [32], the frequency selective nature of a user's channel is not fully exploited by allocating power uniformly across all subcarriers. To improve its performance, Shen et al. [33] added a second step of adaptive power allocation to further enforce the rate proportionality among the users.

The two-step approach adopted in [33] is as follows: in the first step, the modified version of the algorithm outlined in [32] is employed for subcarrier allocation to achieve coarse proportional fairness. Hence, instead of giving priority to the user with the least achieved data rate R_k, priority is given

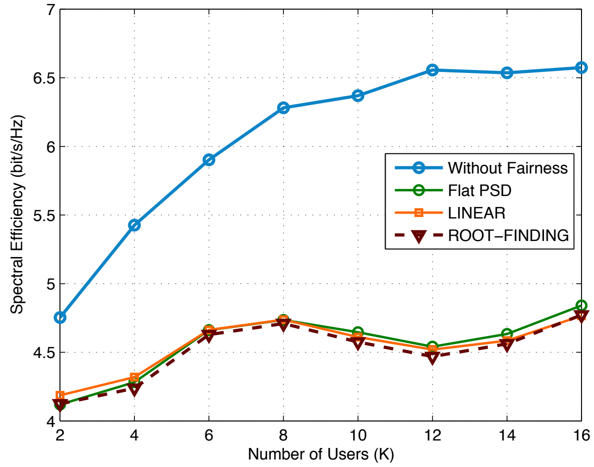


Fig. 7. Spectral efficiency versus the total number of users K, in an OFDMA system with 128 subcarriers and K = 2 $\sim$ 16.

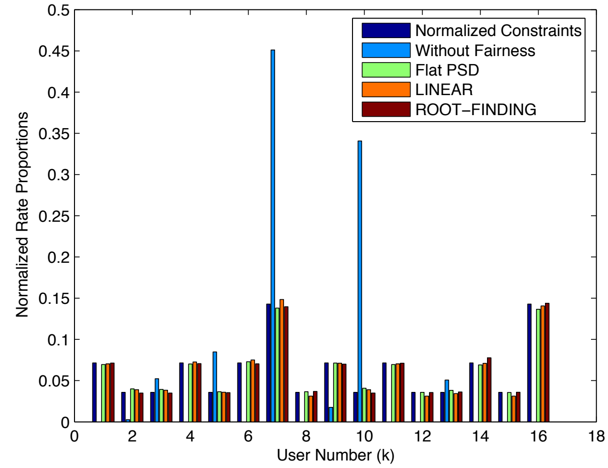


Fig. 8. Normalized capacity ratios per user for N = 128 and K = 16.

to the user with the least achieved proportional data rate i.e., $\frac{R_k}{\alpha_k}$. In this step, the achieved rate is calculated considering equal power on all the subcarriers. After subcarrier allocation is carried out, the problem is simplified into a maximization over continuous variables of power. In the second step, the power is reallocated between the users and then among the subcarriers through the use of water-filling to enforce the rate proportionality among the users. To find the kth user's power p_k, Lagrange multiplier techniques [51] are used to formulate and then solve the optimization problem resulting in K nonlinear equations with K unknowns. These equations can not be solved directly and numerical methods such as Newton-Raphson [52] and its variants are used. Two special cases were analyzed in [33] to reduce the complexity which are described below. In each case, the computational complexity of the algorithm was calculated to be $O(K)$.

1) High channel-to-noise ratio case: based on the fact that adaptive subcarrier allocation was used in the first step, it could be assumed that the best subchannels were chosen for each user and that they have relatively small channel gain differences among them. Furthermore, assuming that the basestation can provide a large amount of power and the channel-to-noise ratio is high, the SNR is much larger than one. With these two approximations, the system of K equations is transformed into a single nonlinear equation in one variable which could be solved using Newton's root finding method. This case was used in deriving the simulation results in Fig. 7 and Fig. 8 and is referred to as "Root-Finding".

2) Linear case: in this approximation, it is assumed that the proportion of subcarriers assigned to each user is approximately the same as the rate constraints (also assumed in [39]). In other words:

$$N_1 : N_2 : \ldots : N_K = \alpha_1 : \alpha_2 : \ldots : \alpha_K. \tag{17}$$

The linear case was further investigated in [45]. In the proposed algorithm, the subcarrier proportionality in (17) is not just an assumption; on the contrary, this proportionality is enforced through the subcarrier allocation in the first step. In this algorithm, although the user with the least proportional capacity is still getting priority to choose its best available subchannel, the number of subchannels to be assigned to each user N_k is determined by its rate constraints given by $N_k = \left\lfloor \frac{\alpha_k N}{\sum_{k=1}^{K} \alpha_k} \right\rfloor$. Once the kth user gets the allotment of N_k subcarriers, it will be assigned no more subchannels until all the users are assigned their pre-determined proportion of subcarriers. With this approximation, the system of K nonlinear equations (in the second step) turns into K *linear* simultaneous equations which could be written in matrix form. The total power p_k for each user k is obtained solving these K linear equations and the water-filling is applied to allocate the power to the assigned subcarriers of each user. In the simulation results, this method is referred to as "Linear".

Fig. 7 shows the comparison of four algorithms discussed above in terms of the total achieved data rate versus different numbers of users. In the simulation results, they are referred to as "Without Fairness" [8], "Flat PSD" [32], "Root-Finding" and "Linear" (proposed in [33] and [45] respectively). The comparison of these algorithms in terms of rate proportionality is shown in Fig. 8. The leftmost bars are the preset normalized constraints $\{\phi\}_{k=1}^{K}$, where $\phi_k = \frac{\alpha_k}{\sum_{k=1}^{K} \alpha_k}$. They indicate the desired rate proportionality among the users and are chosen randomly. The same normalization is used for achieved data rates.

A total of 1000 different channel realizations were used and the results were averaged. Also, an average SNR of 30dB (defined as $\frac{P_{tot}}{N_0 B}$) was chosen in the simulations.

The graph with the highest spectral efficiency in Fig. 7 belongs to [8]. In this algorithm, only the total data rate of the system is considered. While the highest spectral efficiency is achieved, some of the users are not allocated any of the subcarriers and hence achieve no data rate (e.g., users 1 and 4 in Fig. 8). The trade-off between the spectral efficiency and fairness is obvious.

Another trade-off is between achieving the exact rate proportionality and computational complexity. Out of the three algorithms that consider fairness, "Root-Finding" has achieved the exact rate proportionality determined by the leftmost bars. However, the complexity of this method is much higher compared to the other two algorithms with rough proportionality.

As the last observation, note that the total data rate of the system, without fairness, increases with the increase in the number of users as a result of multiuser diversity.

The suboptimal algorithms described above, either use fixed power allocation and perform only subcarrier allocation [32], or handle subcarrier and power allocation separately as in [33], [45] to reduce the complexity of the algorithm. However, subcarrier and power allocation have to be carried out jointly to achieve the optimal solution. A very subtle but effective change in Rhee's algorithm [32] was made by Mohanram et al. [48] to perform joint subcarrier and power allocation, hence avoiding the second step of power allocation outlined in [45] or [33]. In the algorithm proposed in [48], optimization of $N+K$ parameters is carried out by alternating between subcarrier and power allocation. The allocation procedure is the same as [32] but differs in updating user's achieved data rate to find the user with the minimum achieved rate. When a subcarrier is allocated to a user, the power allocated to that user is incremented by $\frac{P_{total}}{N}$, i.e., the power allocated to each user is proportional to the number of subcarriers currently allocated to that user. The total power allocated to the user is then distributed among the assigned subcarriers with water-filling policy resulting in higher user rate. This updated rate information is then used to find the user with the minimum achieved rate to choose the next available subcarrier. Since the power redistribution is needed when there are more than one subcarrier assigned to a user, the water-filling should be performed $N-K$ times.

The reported results are as follows: 1) the achieved total throughput [48, Fig. 1] is slightly higher compared to Shen's algorithm [33] (which has the optimal power allocation) and has up to 25% gain (for 12 users) compared to Rhee's algorithm [32] (with flat transmit power). It even achieves up to 28% gain (for 12 users) when combined with the power allocation algorithm proposed in [33]. 2) Combining the additional step of optimal power allocation proposed in [33] with the joint resource allocation in [48] does not improve the data rate of the worst user (the user with the minimum achieved data rate) [48, Fig. 2]. 3) Finally, the algorithm shows higher achieved gain in total throughput compared to [32] when the PSD of AWGN is higher. This could be explained by the fact that applying water-filling versus fixed power allocation yields larger gains at low SNRs [53].

2) Rate Adaptive Algorithms with Fixed Rate Requirements: When there is a fixed target data rate for each user, the optimization problem can be rephrased to find the best assignment matrix as well as the best power allocation vector such that:

$$\begin{aligned}
&\max_{c_{k,n},p_{k,n}} \quad R_T = \frac{B}{N}\sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n}\log_2\left(1+\frac{p_{k,n}h_{k,n}^2}{N_0\frac{B}{N}}\right),\\
&\text{subject to :}\\
&\qquad \text{C1} : c_{k,n}\in\{0,1\},\ \ \forall k,n\\
&\qquad \text{C2} : \sum_{k=1}^{K} c_{k,n}=1,\ \ \forall n\\
&\qquad \text{C3} : p_{k,n}\geq 0,\quad \forall k,n\\
&\qquad \text{C4} : \sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n}p_{k,n}\leq P_{total},\\
&\qquad \text{C5} : R_k\geq R_{k,min},\ \ k=1,2,...,K
\end{aligned} \tag{18}$$

where R_k and $R_{k,min}$ are the achieved and the minimum required data rate for the kth user respectively. The maximum rate could be still achieved by assigning each subcarrier to the user with the largest channel gain on it. However, although the total rate is maximized, the rate constraint of each user would not be satisfied. This problem was addressed by Yin et al. [39]. In [39], the problem has been partitioned into three steps: 1) the first step determines how many subcarriers N_k and how much power p_k are needed for each user; 2) the second step is the subcarrier allocation which determines the particular set of subcarriers for each user; 3) the third step is the bit loading which determines the number of bits for each subcarrier or in other words, is to determine the power allocation on each user's assigned subcarriers.

The complexity of this problem arises from the fact that the first two variables N_k and p_k are not independent, therefore certain simplifying assumption have been considered in each step. In the first step, it is assumed that the number of the subcarriers and the power allocated to a particular user depend on its rate requirement and its channel condition. Also, the amount of power assigned to each user should be proportional to the number of subcarriers allocated. In the first step, there is no power allocation to the subcarriers and only the total power of each user is determined. It is interesting to note that these two assumptions constitute the very basic assumptions of the algorithm in [45] where a flat transmit power were used on all the subcarriers. To calculate N_k and p_k in [39], each user's channel condition is assumed to be flat on all subcarriers represented only by its $average$ channel-to-noise ratio, hence neglecting the frequency diversity in the first step. Once N_k and p_k are determined, the exact subcarrier assignment is solved by applying Hungarian algorithm [54]. In the final step, the bits are loaded to the subcarriers knowing the total power of each user which is done by the known single user bit loading algorithm. It is in the second and the final step that each user's channel condition on each subcarrier is considered while being neglected in the first step.

If there is a mixture of users with fixed required data rates and users with variable rates, the objective remains the same and only C5 in (18) changes to:

$$\text{C5} : \begin{array}{l} R_1 : R_2 : ... : R_L = \alpha_1 : \alpha_2 : ... : \alpha_L, \\ R_k\geq R_{k,min}\quad k=L+1,L+2,...,K \end{array} \tag{19}$$

where it is assumed that among the first L users, rate proportionality should be maintained while the rest of the users require fixed minimum data rates. The problem of adaptive resource allocation with a mixture of variable and fixed data rate constraints was introduced in [32]. Suh et al. [55] considered this problem in the case of K users where only one of them requires fixed data rate and called it priority user.

B. Margin Adaptive Algorithms

In deriving the algorithms of this group, a given set of user data rates is assumed with a fixed QoS requirement. The optimization problem can then be formulated as:

$$\begin{aligned} \min_{c_{k,n},p_{k,n}} \quad & P_T = \sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n}p_{k,n}. \\ \text{subject to :} \quad & \\ & \mathrm{C1} : c_{k,n} \in \{0,1\}, \ \ \forall k,n \\ & \mathrm{C2} : \sum_{k=1}^{K} c_{k,n} = 1, \ \ \forall n \\ & \mathrm{C3} : p_{k,n} \geq 0, \quad \forall k,n \\ & \mathrm{C4} : \sum_{k=1}^{K}\sum_{n=1}^{N} c_{k,n}p_{k,n} \leq P_{total}, \\ & \mathrm{C5} : R_k \geq R_{k,min}, k = 1,2,...,K. \end{aligned} \tag{20}$$

with the rate requirements indicated in C5.

This problem was first addressed in [56] where the focus was only on subcarrier allocation and further in [36] where adaptive power allocation was also considered. To make the problem tractable, authors in [36] introduced a new parameter to the cost function taking values within the interval [0,1] which can be interpreted as the sharing factor for each subcarrier. With the help of the new parameter, it was shown that the optimization problem can be reformulated as a convex minimization problem over a convex set. Using the standard optimization techniques, the Lagrangian of the new problem is obtained along with the necessary conditions under which not only the minimum total transmit power occurs but also the data rate constraint of each user is satisfied. In order to find the set of Lagrange multipliers such that the individual data rate constraints are satisfied, an iterative search algorithm is used. The obtained set determines the optimal sharing factor of all the subcarriers for all users; however, each subcarrier is assigned to only one user that has the largest sharing factor on that subcarrier. This modification of the final solution leads to a slight deviation from the minimum transmit power.

The drawback of this approach is that the efficiency and the convergence of the algorithm depend critically on the step size and the initial point of the searching. Since this algorithm can be viewed as a counterpart of multiuser "water-filling" solution, the iterative computation is so complex. Also, it is prohibitively expensive and not suitable for real time applications due to its high complexity. One solution to simplify the algorithm is to assume that the channel is flat for certain number of subcarriers. More specifically, Yin et al. [39] assumed that each user's channel across all subcarriers is flat whereas Xiaowen et al. [38] proposed a blockwise subcarrier allocation algorithm where the subcarriers are divided into blocks with the channel being almost flat for each block.

The principle of blockwise subcarrier allocation algorithms is that subcarriers must be assigned to users in blocks. The number of blocks assigned to each user highly depends on the average channel gain for the user as well as its data rate requirement. This approach reduces the number of the optimization parameters based on the number of subcarriers grouped in each block. However, how to group the subcarriers without neglecting the frequency diversity is still an open question.

V. SCHEDULING AND RESOURCE ALLOCATION IN OFDMA SYSTEMS

Scheduling is another technique to deal with the scarcity of radio resources in order to support multiusers in a high speed wireless communication system. In Time Division Multiplexed (TDM) systems, time is divided into several time-slots and the scheduling decision is simply which user to schedule in a time-slot. This decision is made based on the users' current channel conditions and their quality of service requirements. Scheduling could be combined with other multiplexing techniques. When combined with OFDM, for instance, in addition to choosing the active users in a time-slot, the subcarriers and the power should also be allocated to the scheduled users.

In (8), the problem of resource allocation was formulated for each transmission block. However, when scheduling is considered, the time is first divided into TDM time-slots that contain an integer number of OFDM symbols [57]; then, the problem is formulated for each time-slot. In formulating the problem, the target data rate for each user is selected from a feasible rate region based on the channel condition in that time-slot; the resources are then allocated to the users to reach the target data rates. If a user is not scheduled in that time-slot, the corresponding data rate would be zero, hence no power is allocated to the unscheduled user. Therefore, the problem of scheduling and resource allocation is also referred to as $power$ $scheduling$ [58].

Many scheduling algorithms have been proposed in literature (e.g., [59] and the references therein). In these algorithms, a utility function is defined for each user to quantify fairness and other quality of service considerations in each time-slot. The utility is an increasing concave function of each user's average throughput up to that time-slot. A target rate vector (elements of which are the target data rates for the users) is then chosen such that its projection onto the gradient of the system's total utility is maximized. These algorithms are referred to as "gradient-based".

Recently, various opportunistic scheduling schemes have been developed for wireless networks. These scheduling schemes are classified as either "single-server" or "multi-server" based on the underlying multiple access scheme used [58]. An example of emerging wireless systems which uses a combination of TDM and OFDM is IEEE 802.16e (WiMAX) where scheduling and resource allocation is still an open research issue.

VI. CURRENT RESEARCH AREAS

Adaptive resource allocation methods have shown to offer higher user data rates due to the additional degree of freedom provided by multichannel systems. One way to create multiple channels is in frequency domain using multiple carrier frequencies with the methods and algorithms discussed in this article. The other way is in the spatial domain with multiple transmit and receive antennas. The latter is also referred to as multiple-input multiple-output (MIMO).

A. Multiple-Input-Multiple-Output (MIMO) OFDM

In the downlink MIMO-OFDM system, a basestation is communicating simultaneously to multiple users while both the basestation and the users are equipped with multiple antennas. The problem of resource allocation in a multiuser MIMO-OFDM system is similarly formulated but more challenging due to multiple antennas. The performance of MIMO-OFDM systems has been and still is an interesting research topic.

Adaptive subcarrier and power allocation algorithms (e.g., [60]) have been shown to significantly improve the spectral and power efficiency in a MIMO-OFDM system compared to conventional fixed subcarrier and power allocation. Simpler algorithms to reach the sum capacity of a Gaussian MIMO channel have been proposed in [61], [62]. A comparison of the potential maximum sum capacity of downlink MIMO-OFDMA and MIMO multicarrier CDMA (MIMO-MC-CDMA) in a single cell multiuser environment has been presented in [63]. It has been shown that for large number of antennas in some cases, the benefits of OFDMA over MC-CDMA are significantly reduced.

As in single-input single-output systems, several papers (e.g., [64], [65]) have discussed the impact of imperfect channel information on the performance of MIMO-OFDM systems and proposed new algorithms based on partial CSI. Further research should be carried out to develop low complexity resource allocation algorithms considering practical issues such as imperfect CSI as well as multi-cell environment.

B. Resource Allocation in Multi-cell Systems

In a single cell, when the average SNR of some users are very low, their rate requirements can not be met. If the user's average SNR received from its neighboring cell is relatively higher due to positive shadowing, such user can communicate with the neighboring basestation. This additional degree of freedom would decrease the probability of outage. Zhang et al. [66], for instance, discussed adaptive cell selection to reduce the probability of outage for those users with low average SNR that happen to be located near the boundary of the cell. As mentioned earlier, other issues such as power control in uplink become crucial when considering multi-cell environments.

VII. Conclusions

In this article, we presented an overview of algorithms in the literature to adaptively allocate the available resources in a multiuser OFDM communication system. Different classes of algorithms considered different objectives and attempted to obtain a solution that is close to optimum but at the same time is simple enough to be implemented. General observations have been made from the reported simulation results:

- **Increase in total throughput with higher number of users:** no matter which optimization method is used, in each class, the performance of the system is improved with the increase in the number of users. This is a result of multiuser diversity. With more users in the system, the chances that a subcarrier is in good condition for at least one user are getting higher resulting in that subcarrier carrying more bits. This will result in lower transmit power and higher spectral efficiency.
- **Trade-off between performance and complexity:** adaptive resource allocation would with no doubt result in better performance. However, it increases the number of optimization parameters of the system and results in higher computational complexity. Valid simplifying assumptions are derived by investigating the cases where the achieved performance gain is not significant compared to the computation burden. Such simplifying assumptions and methods include flat transmit power, blockwise subcarrier allocation or splitting the allocation procedure into separate steps.

References

[1] "IEEE standard for local and metropolitan area networks part 16: Air interface for fixed and mobile broadband wireless access systems," *IEEE, Tech. rep. 802.16*, October 2004.

[2] "IEEE standard for local and metropolitan area networks part 16 and amendment 2," *IEEE, Tech. rep. 802.16e*, February 2006.

[3] H. Schulze and C. Lueders, *Theory and Applications of OFDM and CDMA Wideband Wireless Communications*. John Wiley, 2005.

[4] P. Chow, J. Cioffi, and J. Bingham, "A practical discrete multitone transceiver loading algorithms for data transmission over spectrally shaped channels," *IEEE Trans. Commun.*, vol. 43, pp. 773–775, February 1995.

[5] R. Fischer and J. Huber, "A new loading algorithm for discrete multitone transmission," *in Proc. IEEE Globecom*, vol. 1, pp. 724–728, November 1996.

[6] S. H. Ali, Ki-Dong Lee, and V. C. Leung, "Dynamic resource allocation in OFDMA wireless metropolitan area networks [radio resource management and protocol engineering for IEEE 802.16]," *IEEE Wireless Commun. Mag.*, vol. 14, pp. 6–13, February 2007.

[7] M. Sternad, T. Svensson, and T. Ottosson et al., "Towards systems beyond 3G on adaptive OFDMA transmission," *Proc. IEEE*, vol. 95, pp. 2432–2455, December 2007.

[8] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 171–178, February 2003.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[10] L. K. (Anna) Tee, C. van Rensburg, and Jiann-An Tsai, "Uplink power control for an OFDMA mobile cellular system," *in Proc. 66th IEEE VTC*, pp. 357–361, Sept-Oct. 2007.

[11] M. Wahlqvist et al., "A conceptual study of OFDM-based multiple access schemes, part 1: Air interface requirements," *Technical Report Tdoc 117/96, ETSI STC SMG2 meeting No. 18, Helsinki, Finland*, pp. 1–6, May 1996.

[12] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, October 1998.

[13] T. S. Rappaport, *Wireless Communications*. Prentice Hall PTR, 2002.

[14] A. Duel-Hallen, S. Hu, and H. Hallen, "Long-range prediction of fading signals," *IEEE Signal Processing Mag.*, vol. 17, pp. 62–75, May 2000.

[15] A. Forenza and R. W. Heath, "Link adaptation and channel prediction in wireless OFDM systems," *in Proc. IEEE 45th MWSCAS*, vol. 3, pp. 211–214, August 2002.

[16] M. Sternad and D. Aronsson, "Channel estimation and prediction for adaptive OFDM downlinks [vehicular applications]," *in Proc. IEEE VTC*, vol. 2, pp. 1283–1287, October 2003.

[17] I. C. Wong, A. Forenza, R. W. Heath, and B. L. Evans, "Long range channel prediction for adaptive OFDM systems," *in Proc. IEEE ACSSC*, vol. 1, pp. 732–736, November 2004.

[18] I. C. Wong and B. L. Evans, "Joint channel estimation and prediction for OFDM systems," *in Proc. IEEE Globecom*, vol. 4, 28 Nov-2 Dec 2005.

[19] D. Schafhuber and G. Matz, "MMSE and adaptive prediction of time-varying channels for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 593–602, March 2005.

[20] I. C. Wong and B. L. Evans, "Low-complexity adaptive high-resolution channel prediction for OFDM systems," *in Proc. IEEE Globecom*, pp. 1–5, November 2006.

[21] S. Ye, R. S. Blum, and L. J. Cimini Jr., "Adaptive OFDM systems with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 5, pp. 3255–3265, November 2006.

[22] Z. Shen, J. G. Andrews, and B. L. Evans, "Short range wireless channel prediction using local information," *in Proc. IEEE ACSSC*, vol. 1, pp. 1147–1151, November 2003.

[23] I. C. Wong, A. Forenza, R. W. Heath, and B. L. Evans, "Long range channel prediction for adaptive OFDM systems," *in Proc. IEEE ACSSC*, vol. 1, pp. 732–736, November 2004.

[24] A. Leke and J. M. Cioffi, "Multicarrier systems with imperfect channel knowledge," *in Proc. IEEE PIMRC*, vol. 2, pp. 549–553, September 1998.

[25] Y. Yao and G. B. Giannakis, "Rate-maximization power allocation in OFDM based on partial channel knowledge," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1073–1083, May 2005.

[26] M. R. Souryal and R. L. Pickholtz, "Adaptive modulation with imperfect channel information in OFDM," *in Proc. IEEE ICC*, vol. 6, pp. 1861–1865, June 2001.

[27] F. Brah, L. Vanderdorpe, and J. Louveaux, "OFDMA constrained resource allocation with imperfect channel knowledge," *in Proc. IEEE Symposium on Communications and Vehicular Technology*, pp. 1–5, November 2007.

[28] I. C. Wong and B. Evans, "Optimal OFDMA subcarrier, rate, and power allocation for ergodic rates maximization with imperfect channel knowledge," *in Proc. IEEE ICASSP*, vol. 3, pp. 89–92, April 2007.

[29] E. Lawrey, "Multiuser OFDM," *in Proc. 5th ISSPA*, pp. 761–764, August 1999.

[30] A. J. Goldsmith and Soon-Ghee Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218–1230, October 1997.

[31] Y. Otani, S. Ohno, K. ann Donny Teo, and T. Hinamoto, "Subcarrier allocation for multi-user OFDM system," *in Proc. Asia-Pasific Conf. on Commun.*, pp. 1073–1077, 2005.

[32] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," *in Proc. IEEE VTC*, vol. 2, pp. 1085–1089, May 2000.

[33] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2726–2737, November 2005.

[34] S. Sadr, A. Anpalagan, and K. Raahemifar, "A novel subcarrier allocation algorithm for multiuser OFDM system with fairness: User's perspective," *in Proc. 66th VTC*, pp. 1772–1776, Sept-Oct 2007.

[35] G. J. Foschini and J. Salz, "Digital communications over fading radio channels," *Bell Syst. Tech. J.*, pp. 429–456, February 1983.

[36] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit and power allocation," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1747–1758, October 1999.

[37] G. Zhang, "Subcarrier and bit allocation for real-time services in multiuser OFDM systems," *in Proc. IEEE ICC*, vol. 5, pp. 2985–2989, June 2004.

[38] L. Xiaowen and Z. Jinkang, "An adaptive subcarrier allocation algorithm for multiuser OFDM system," *in Proc. IEEE VTC*, vol. 3, pp. 1502–1506, October 2003.

[39] H. Yin and H. Liu, "An efficinet multiuser loading algorithm for OFDM-based broadband wireless systems," *in Proc. IEEE Globecom*, vol. 1, pp. 103–107, November 2000.

[40] G. Song and Y. G. Li, "Utility-based joint physical-MAC layer optimization in OFDM," *in Proc. IEEE Globecom*, vol. 1, pp. 671–675, November 2002.

[41] G. Song and Y. G. Li, "Adaptive subcarrier and power allocation in OFDM based on maximizing utility," *in Proc. IEEE VTC*, vol. 2, pp. 905–909, April 2003.

[42] G. Song and Y. G. Li, "Cross-layer optimization for OFDM wireless networks-Part I: theoretical framework," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 614–624, March 2005.

[43] G. Song and Y. G. Li, "Cross-layer optimization for OFDM wireless networks-Part II: Algorithm development," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 625–634, March 2005.

[44] Z. Shen, J. G. Andrews, and B. L. Evans, "Optimal power allocation in multiuser OFDM systems," *in Proc. IEEE Globecom*, vol. 1, pp. 337–341, December 2003.

[45] I. C. Wong, Z. Shen, B. L. Evans, and J. G. Andrews, "A low complexity algorithm for proportional resource allocation in OFDMA systems," *in Proc. IEEE Workshop on Signal Processing Systems*, pp. 1–6, October 2004.

[46] F. P. Kelly, "Charging and rate control for elastic traffic," *European Trans. on Telecommun.*, vol. 8, pp. 33–37, 1997.

[47] Z. Jiang, Y. Ge, and Y. G. Li, "Max-utility wireless resource management for best effort traffic," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 100–111, January 2005.

[48] C. Mohanram and S. Bhashyam, "A sub-optimal joint subcarrier and power allocation algorithm for multiuser OFDM," *IEEE Commun. Lett.*, vol. 9, pp. 685–687, August 2005.

[49] P. S. Chow and J. M. Cioffi, "Bandwidth optimization for high speed data transmission over channels with severe intersymbol interference," *in Proc. IEEE Globecom*, vol. 1, pp. 59–63, December 1992.

[50] M. Bohge, J. Gross, and A. Wolisz, "The potential of dynamic power and sub-carrier assignments in multi-user OFDM-FDMA cells," *in Proc. IEEE Globecom*, vol. 5, pp. 2932–2936, December 2005.

[51] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 2nd ed., 1999.

[52] R. Baldick, *Applied Optimization: Formulation and Algorithms for Engineering Systems*. Cambridge University Press, 2006.

[53] T. L. Tung and K. Yao, "Channel estimaltion and optimal power allocation for a multiple-antenna OFDM system," *EURASIP J. Applied Signal Processing*, pp. 330–339, March 2002.

[54] E. D. Nering and A. W. Tucker, *Linear Programs and Related Problems*. Academic Press Inc., London, 1993.

[55] C. Suh, Y. Cho, and S. Yoon, "Dynamic subchannel and bit allocation in multiuser OFDM with a priority user," *in Proc. IEEE Eighth International Symposium on Spread Spectrum Techniques and Applications*, pp. 919–923, September 2004.

[56] C. Y. Wong, C. Y. Tsui, R. S. Cheng, and K. B. Letaief, "A real-time subcarrier allocation scheme fore multiple access downlink OFDM transmission," *in Proc. IEEE VTC*, vol. 2, pp. 1124–1128, September 1999.

[57] R. Agrawal, R. Berry, J. Huang, and V. Subramanian, "Optimal scheduling for OFDMA systems," *in Proc. IEEE ACSSC*, pp. 1347–1351, October-November 2006.

[58] J.-W. Lee, R. R. Mazmudar, and N. B. Shroff, "Opportunistic power scheduling for multiuser wireless systems with minimum performance constraints," *in Proc. INFOCOM*, vol. 2, pp. 1067–1077, March 2004.

[59] J. Huang, V. Subramanian, R. Arawal, and R. Berry, "Downlink scheduling and resource allocation for OFDM systems," *in Proc. IEEE ISS*, pp. 1272–1279, March 2006.

[60] C. Pan, Y. Cai, and Y. Xu, "Adaptive subcarrier and power allocation for multiuser MIMO-OFDM systems," *in Proc. IEEE ICC*, vol. 4, pp. 2631–2635, May 2005.

[61] Z. Shen, R. Chen, J. Andrews, R. Heath, and B. Evans, "Sum capacity of multiuser broadcast channels with block diagonalization," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 2040–2045, June 2007.

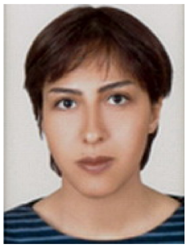
[62] F. S. Chu and K. C. Chen, "Radio resource allocation for mobile MIMO-OFDMA," *in Proc. IEEE VTC*, pp. 1876–1880, May 2008.

[63] E. Lo, P. Chan, V. Lau, R. Cheng, K. Letaief, R. Murch, and W. Mow, "Adaptive resource allocation and capacity comparison of downlink multiuser MIMO-MC-CDMA and MIMO-OFDMA," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1083–1093, March 2007.

[64] Z. Hu, G. Zhu, Y. Xia, and G. Liu, "Multiuser subcarrier and bit allocation for MIMO-OFDM systems with perfect and partial channel information," *in Proc. IEEE WCNC*, vol. 2, pp. 1188–1193, March 2004.

[65] P. Xia, S. Zhou, and G. Giannakis, "Adaptive MIMO-OFMA based on partial channel state information," *IEEE Trans. Signal Processing*, vol. 52, pp. 202–213, January 2004.

[66] Y. J. Zhang and K. B. Letaief, "Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM transmission," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 1566–1575, September 2004.



Sanam Sadr (S07) received the B.S. degree in electrical engineering from Tehran University, Tehran, Iran, in 2003 and the M.S. degree in electrical engineering from Ryerson University, Toronto, Canada, in 2007. She was awarded Ryerson Graduate Award for the 2005-2007 academic year and Ryerson Graduate Scholarship for the 2006-2007 academic year by the Department of Electrical and Computer Engineering and the School of Graduate Studies at Ryerson University respectively. Her research interests include multicarrier communication systems and resource allocation in multiuser environment.

Alagan S. Anpalagan received the B.A.Sc., M.A.Sc., and Ph.D. degrees in Electrical Engineering from the University of Toronto, Canada in 1995, 1997 and 2001 respectively. Since August 2001, he has been with the Ryerson University, Toronto, Canada, where he co-founded WINCORE laboratory in 2002 and leads the WAN (Wireless Access and Networking) R&D group. Currently, he is an Associate Professor and Program Director for Graduate Studies in the Department of Electrical and Computer Engineering at Ryerson University.

His research interests are in general, wireless communication, mobile networks and system performance analysis; and in particular, QoS-aware radio resource management, joint study of wireless physical/link layer characteristics, cross-layer resource optimization, and wireless sensor networking. He has published more than 80 papers and articles in international conferences and journals in his research area.

Dr. Anpalagans editorial duties include Guest Editor, Special Issues on Radio Resource Management in 3G+ Wireless Systems (2005-06), Fairness in Radio Resource Management for Wireless Networks; and Associate Editor, EURASIP Journal of Wireless Communications and Networking. He previously served as IEEE Toronto Section Chair (2006-07) and Communications Chapter Chair (2004-05) and Technical Program Co-Chair, IEEE Canadian Conference on Electrical and Computer Engineering (2008, 2004). He is an IEEE Senior Member and a Registered Professional Engineer in the province of Ontario, Canada.

Kaamran Raahemifar (S'91, M'99, SM'02) is currently an associate professor at Electrical & Computer Engineering Dept., Ryerson University, Toronto, Ontario, Canada. He received his B.Sc. (Eng.) degree from Sharif University of Technology, Tehran, Iran, in 1989. He received his MAsc degree from University of Waterloo, and his PhD degree from University of Windsor, Ontario, Canada in 1994 and 1998 respectively. He joined Electronic Workbench, Toronto, Ontario, as Chief Scientist in 1998. He joined Ryerson University in 1999. He is currently a member of Professional Engineers of Ontario, and a senior member of IEEE. He has published many conference and journal papers in the area of design and testing VLSI circuits for computational and biomedical applications. His current research interests are in the area of Hardware Implementation of Biomedical Signals, Application of Optimization in Engineering, and Modeling and Simulation of Epilepsy signals.