Carrier assignment algorithms for OFDM-based multi-carrier wireless networks with channel adaptation

Iordanis Koutsopoulos, *Member IEEE* and Leandros Tassiulas, *Senior Member IEEE*

***Abstract*—We study carrier assignment in a single-cell multi-user OFDM multi-carrier system so as to satisfy user rate requirements with minimal resources. Different users experience different quality in different carriers due to frequency selectivity of users' propagation channels and due to non-co-located user receivers that perceive different interference from neighboring cells across carriers. We study a static instance of the problem, specified by user carrier qualities and rate requirements. Adaptive modulation at the transmitter differentiates carriers for each user. In good quality carriers, the user satisfies per-frame rate requirements with few slots (or equivalently it satisfies its per-slot rate requirements with small occupied time slot portion). We study integral and fractional assignment, where a user is assigned to only one or several carriers. Fractional assignment is formulated as a linear programming problem. For integral assignment, we introduce two classes of iterative heuristics that use carrier reassignment to users and user substitution in carriers respectively and may be viewed as resulting from corresponding optimal fractional assignment algorithms. We use Lagrangian relaxation to obtain performance bounds and show that the two classes of heuristics arise from two relaxations. Our approach identifies efficient feasible solutions and is amenable to distributed implementation.**

***Index Terms*—Wireless OFDM systems, multi-carrier systems, carrier allocation, optimization.**

I. Introduction

Wireless network users and their rate demands steadily increase while sizes of allocated bandwidth chunks to various communication systems remain limited. Orthogonal frequency division multiplexing (OFDM) emerged as a multi-carrier access and signaling technique for providing high data rates per unit of available bandwidth. The spectrum is divided into orthogonal narrow-band subcarriers as in frequency division multiplexing. Appropriate subcarrier spacing preserves channel orthogonality despite subcarrier overlap in frequency and leads to high spectral efficiency. The user bit stream is split into bit subsets, the subsymbols. Each subsymbol modulates a subcarrier and several subsymbols of a user are transmitted in parallel over subcarriers. This transmission mode reduces effective symbol transmission rate and provides immunity to inter-symbol interference. OFDM is included in IEEE 802.11a/g standards for WLANs in distributed coordination function (DCF) with users connected in multiple hops, or in point coordination function (PCF) with single-hop connection to an access point (AP). OFDM is also considered for IEEE 802.11n WLAN standard, the current and evolving IEEE 802.16x WiMAX broadband wireless access standards, and in wireless personal area networks. In OFDM-based 802.11 and 802.16 system, a carrier can be thought of as a band of subcarriers and is referred to as a channel. For instance, IEEE 802.11a/g has 11 channels. Resource allocation at a channel level and in the presence of transceiver limitations is performed so that a user is allocated to a single carrier. Furthermore, in other currently employed multi-carrier systems such as 2G cellular ones and their successors and in wireless cable systems, frequency bands corresponding to different carriers do not overlap. The difference from OFDM is that a user is allocated to a single carrier frequency out of available ones.

The fundamental goal in a communications system, Quality of service (QoS), is perceived as an acceptable signal-to-interference-and-noise ratio (SINR) or bit error rate (BER) at the receiver at the physical layer and as certain rate or delay guarantees at higher layers. In order to ensure best user QoS, techniques that span several layers are employed. At the access and higher layers these are channel allocation, buffer management, routing and flow control. At physical layer, adaptation of modulation level [2] and channel coding rate controls the amount of sustainable interference for a maximum acceptable BER, while transmit power control changes SINRs. In [3], adaptive modulation is considered together with time slot allocation for one carrier. The base station (BS) receives user SINR measurements and searches for available slots with a given modulation level to support user rate requirements. If an adequate number of such slots cannot be found, a query for more slots and lower modulation level is made.

Most of the studies on multi-user OFDM systems focus on power control for maximizing instantaneous achievable rate [4]-[6]. For given subcarrier allocation to users and a total power constraint for each user, the optimal rate is achieved by power water-filling across subcarriers of each user, followed by user bit allocation. However, finding the subcarrier allocation and bit and power allocation that maximize total rate is a hard problem due to the combinatorial nature of subcarrier allocation and the different subcarrier qualities for users. In [4], this discrete allocation problem is studied for uplink and a power constraint for each user. The problem is relaxed to a continuous convex optimization one that is solved numerically. The corresponding problem for downlink with a total power constraint over all users is solved by assigning each subcarrier

Part of the material in this paper was presented in [1] at ICC 2001, Helsinki, Finland.

The authors are with the Department of Computer and Communications Engineering, University of Thessaly, Volos, GR 38221, Greece (email: jordan@uth.gr ; leandros@uth.gr).

to the user with the largest gain in it and then water-filling [5]. In [7] the authors present a utility maximization framework for OFDM resource allocation. In [8], the dual problem of [4] is studied, namely subcarrier, bit and power allocation for minimizing total power while satisfying rate constraints of users. The continuous relaxation of the integer programming problem leads to an iterative algorithm and a suboptimal solution. In these problems, continuous variables are interpreted as time portions where a subcarrier is occupied by a user. Another line of works [9], [10] give subcarrier allocation heuristics that find an initial assignment and then perform greedy iterative improvements on the objective value (which is usually power minimization) by subcarrier reassignments.

In multi-user OFDM systems with underlying TDMA frame, each user is served by the BS through allocated subcarriers and time slots [11]. In each subcarrier of fixed symbol rate, user rate depends on the modulation level in bits per symbol. In OFDM, a user experiences different frequency response in different subcarriers due to frequency selectivity of its propagation channel, induced by multi-path fading. Furthermore, since user receivers in the downlink are not co-located, users in different locations have different multi-path characteristics and perceive different interference from neighboring cells across carriers. The cumulative effect is that a user experiences different quality in different carriers. Given the situation above, the resource allocation controller should allocate each user to the subcarrier of best quality where it can receive data with the highest modulation level. A high modulation level means a large number of conveyed bits per symbol and therefore a small number of needed symbols to satisfy rate requirements. This is turn implies that the user needs to occupy a small portion of time slot to satisfy its rate requirements per slot, or, by proportion, it needs fewer slots per frame to satisfy per-frame rate requirements. The objective is to minimize the total amount of required resources that fulfil user requirements, whether resource refers to total required time slot portion across subcarriers or to total number of time slots per frame. By utilizing small amount of resources to accommodate all users, the allocation agent can respond better to a sudden channel quality deterioration or user load increase. Ideally, each user should be allocated exclusively to the subcarrier where it can receive data with highest modulation level so that it satisfies rate requirements with minimum amount of resources. However, it may happen that some subcarriers of good quality are very popular for assignment to several users and therefore not all users can be accommodated in these subcarriers. In that case, subcarriers of lower quality should be used for some users with the expense that users have to occupy more time slot portion and time slots per frame in these subcarriers to achieve their rates. Besides OFDM subcarrier allocation, carrier assignment arises in other multi-carrier systems where each user is allocated only to one carrier as discussed earlier.

We study carrier assignment with the objective to satisfy certain user rate requirements with minimal amount of resources. We consider a static problem instance, specified by user rate requirements and perceived carrier qualities. Under the assumption of time-invariant channel in a frame, rate requirements for each user are mapped to a number of required bits per time slot. Depending on carrier quality and modulation level for the user, these bits correspond to different occupied time slot portions in different carriers. The carrier assignment problem amounts to finding time slot portions for users across different carriers that satisfy user requirements with minimum amount of resources. If the assumption above is relaxed, a problem with N carriers and C slots per frame can be handled as NC carriers. With this work, we contribute to the current literature as follows: (i) we formulate the carrier allocation problem where adaptive modulation differentiates resource utilization of carriers by users, (ii) we study fractional and integral carrier assignment which capture the situation in OFDM or other multi-carrier systems and extend our approach to allocation of groups of carriers (iii) we rely on the optimal solution method for fractional user assignment to devise meaningful iterative heuristics for integral user assignment, and (iv) we use Lagrangian relaxation to obtain performance bounds and define a framework for designing efficient assignment algorithms. The structure of relaxation makes the algorithms amenable to distributed implementation as well.

The term "carrier" refers either to an OFDM subcarrier or to a carrier frequency in some generic multi-carrier system. The rest of the paper is organized as follows. In section II we provide the model, assumptions and problem statement. In section III we present algorithms for integral and fractional user assignment, in section IV we relate them to Lagrangian relaxation and in section V we present numerical results. Section VI concludes our study.

II. System Model and Problem Statement

A. System Model

We consider downlink transmission from a BS to K users with a multi-carrier system of N orthogonal carriers. A fixed transmission power is used, so that power allocation across carriers and users is not an issue. A TDMA frame of duration T_f sec is assumed, consisting of a part for control data such as allocation information and measurement exchange and C time slots of duration T_s for data transmission. Data arrive from higher layers in bits and need to be transmitted to users. A user i has rate requirements r_i in bits/sec at the access layer, which corresponds to r_iT_f bits per frame or r_iT_s bits per slot.

Carrier quality for a user depends on propagation factors, such as path loss, shadowing and multi-path. Path loss depends on carrier frequency. Multi-path channel characteristics of a user, such as path gains and delays give rise to frequency selectivity which is reflected to different values of the channel frequency response function across carriers of the user. Furthermore, the different locations of users and different amounts of carrier reuse in neighboring locations create user-dependent perceived amount of interference. We assume no mobility for user receivers and the surrounding environment, so that carrier coherence time is sufficiently large. For Doppler frequencies of less than 2km/h, carrier coherence time is of the order of several msec. We also assume that the amount of interference at a user receiver is similar across all slots in a carrier frame, which implies that neighboring locations

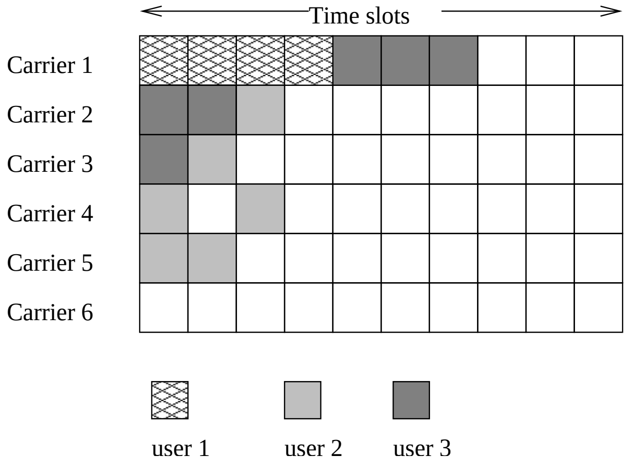

Fig. 1. Examples of carrier and time slot assignment to users.

use all slots of an employed carrier. Under these assumptions, carrier gains and interference factors for each carrier and user are time-invariant for a frame duration. The collective effect of propagation factors and interference is captured by carrier gains $\sqrt{g_{n,k}}$ and interference factors $\sqrt{I_{n,k}}$ respectively for each user $k = 1, \ldots, K$ and carrier $n = 1, \ldots, N$. Each user experiences the same quality in all slots of a given carrier, yet it sees different carrier qualities across carriers.

To satisfy user rate requirements, the allocation controller at the BS assigns a number of carriers and slots to users. For user i that is assigned to carrier j, a modulation level with b_{ij} bits per symbol is selected from a L_0-element set $\mathcal{M} = \{b_1, \ldots, b_{L_0}\}$ of available QAM or QPSK constellations with different number of bits per subsymbol, $b_1 < \ldots < b_{L_0}$. We do not consider receiver noise. The BER at the output of the detector of a user receiver in a carrier should satisfy BER $\leq \epsilon$, where ϵ is a prespecified value. For M-QAM modulation level with $M = 2^b$, $b \in \mathcal{M}$ and signal-to-interference ratio (SIR) γ, the maximum sustainable modulation level so that BER $\leq \epsilon$ is $\log_2(1+K\gamma)$ with $K = -1.5/\ln(5\epsilon)$ [2]. Each user measures received signal strength and interference at each carrier with pilot symbols transmitted by the BS in control slots at the beginning of the frame and provides that information to the BS controller in the uplink. For time-invariant carrier quality in a frame, the controller estimates attainable SIR per carrier and user as $g_{n,k}/I_{n,k}$ and computes maximum sustainable modulation level for each user and carrier.

In OFDM, the bit stream of each user i is divided into bit groups of variable size, each of which constitutes an OFDM symbol for the user. A given number, $S = T_s/T$, of OFDM symbols is transmitted in a slot, where T is the symbol period. The bits of an OFDM symbol of user i are divided into $N_i \leq N$ subgroups, where N_i is the number of carriers used for transmission to user i. The bits of each subgroup constitute a user subsymbol, and S such subsymbols modulate a carrier within a slot. If user i is allocated A_{ij} time slots in carrier j with modulation level b_{ij}, it gets a total of $n_i = S\sum_{j=1}^{N_i} b_{ij}A_{ij}$ bits per frame and rate n_i/T_f. In multi-carrier systems other than OFDM, a user i is assigned only to one carrier, it transmits a fixed number of S symbols in a slot with modulation b_i and is given A_i slots in that carrier, achieving rate $\frac{S}{T_f}b_iA_i$. Two types of user assignment can be seen: (i) *fractional* user assignment, where a user is allocated slots in several carriers to satisfy rate requirements, (ii) *integral* user assignment, where a user is allocated slots of only one carrier. An example with 3 users, 6 carriers and $C = 10$ time slots per frame is depicted in figure 1. User 1 is assigned only to carrier 1, user 2 uses slots in carriers $2, 3, 4, 5$ and user 3 occupies slots in carriers $1, 2, 3$. We have integral assignment for user 1 and fractional assignment for users 2 and 3.

Carrier gains $\{g_{n,k}\}$ for $n = 1, \ldots, N$ and $k = 1, \ldots, K$ and user rate requirements $\{r_k\}$, for $k = 1, \ldots, K$ may change between frames. In this work, we consider the static resource allocation problem on a frame basis and we are not concerned with dynamics across frames. We focus on assignment in one cell and do not address coordination among multiple BSs.

B. Problem Statement

User rate requirements can be expressed in terms of bits per frame or bits per slot. Each user perceives different quality in different carriers and thus needs different amounts of time resource to fulfil rate requirements in each carrier. On a frame basis, this amount pertains to the portion of a frame occupied by the user so that per-frame requirements are met. This portion is approximately given by the number of required time slots in the carrier frame by that user (assuming that in each time slot, all S symbols belong to that user), divided by C. On a time slot basis, the time resource is the portion of a carrier slot devoted to symbols of that user so that per-slot requirements are met. This portion is given by the number of symbols of a user in a slot, divided by S.

The amount of required time resources in a carrier by a user depends on its rate requirements and on the maximum sustainable modulation level of the user in the carrier which reflects carrier quality for the user. High modulation level means a large number of conveyed bits per symbol and therefore a small number of needed symbols to satisfy rate requirements. This implies that the user needs a small portion of time slot to satisfy its per-slot rate requirements. By proportion, it needs few slots per frame to satisfy per-frame rate requirements. Hence, more users can be accommodated in the system, and user capacity is increased. On the other hand, low modulation levels lead to low bit rates and much time resource is needed by users to satisfy rate requirements. However, high modulation levels are susceptible to interference, they require high SIR to maintain a given BER and can be used only in good quality carriers, as opposed to low modulation levels which can be used in lower quality carriers.

Allocating users to carriers so as to fulfil user requirements should be accomplished with minimal amount of resources so that the system is better prepared to cope with a potential traffic load increase or carrier quality deterioration. Each user should be assigned carriers of best quality, namely those where it can sustain high modulation levels. However, it may happen that some good quality carriers are preferable by several users and all users cannot achieve their (per-frame or per-slot) rate requirements. In that case lower quality carriers should be used for some users, with the expense that additional time resource will be consumed to satisfy rate requirements. The arising problem is stated as follows:

Problem (P): *Given K users with certain rate requirements and given N carriers, allocate users to carriers so that user rate requirements are satisfied, and the minimum total amount of resources is used, subject to a maximum allowable BER per user and carrier.*

We assume that user rate requirements can be satisfied at least with the lowest modulation level b_{L_0}, otherwise the problem is not meaningful. In the sequel, we present the problem formulation on a frame basis. Let α_{ij} denote the number of required slots so that user i satisfies its rate requirements r_i when assigned *only* to carrier j, for $i = 1, \ldots, K$ and $j = 1, \ldots, N$. This is computed as

$$\alpha_{ij} = \left\lceil \frac{r_i T_f}{S b_{ij}} \right\rceil , \tag{1}$$

where $\lceil x \rceil$ denotes the smallest integer that exceeds x. Define the long $(NK \times 1)$ vector $\boldsymbol{\alpha} = (\alpha_{ij} : i = 1, \ldots, K$ and $j = 1, \ldots, N)$. Vector $\boldsymbol{\alpha}$ completely specifies an instance of the problem. Let x_{ij} be the portion (percentage) of rate requirements of user i that are satisfied by assignment to carrier j and let $\mathbf{x} = (x_{ij} : i = 1, \ldots, K$ and $j = 1, \ldots, N)$ be the corresponding $(NK \times 1)$ vector. Problem (P) is stated formally as follows:

$$\min_{\mathbf{x}} Z(\mathbf{x}) = \sum_{i=1}^{K} \sum_{j=1}^{N} \alpha_{ij} x_{ij} \tag{2}$$

subject to the constraints:

$$\sum_{j=1}^{N} x_{ij} = 1, \quad \text{for } i = 1, \ldots, K, \tag{3}$$

$$\sum_{i=1}^{K} \alpha_{ij} x_{ij} \leq C, \quad \text{for } j = 1, \ldots, N, \tag{4}$$

$$0 \leq x_{ij} \leq 1 \quad \text{or} \quad x_{ij} \in \{0, 1\} \, \forall i, j. \tag{5}$$

Constraints (3) are assignment constraints and imply fulfilment of user rate requirements. Constraints (4) are capacity constraints. Finally, constraints (5) specify the range of values of variables x_{ij} for fractional or integral assignment. An assignment $\mathbf{x}$ is *feasible* if it satisfies (3)-(5). A feasible assignment $\mathbf{x}^*$ is *optimal* if $Z(\mathbf{x}^*) \leq Z(\mathbf{x})$ for all feasible assignments $\mathbf{x}$. With appropriate scaling of parameters α_{ij} and constraints, the problem above with C time slots becomes equivalent to the per-slot problem of finding the slot fractions assigned to different users in different carriers so as to satisfy per-slot user requirements with minimum amount of time resources under the premise that a user occupies different time slot fractions in different carriers. In what follows we present our approach and algorithms on a frame basis.

III. Integral and Fractional Carrier Assignment

A. Integral Assignment

In integral user assignment, which arises in multi-carrier systems other than OFDM, each user is assigned to only one carrier. A first question concerns identifying a feasible integral assignment of K users to N carriers for a problem instance specified by $\boldsymbol{\alpha}$. Consider a simple instance $\mathcal{I}$, where a user needs the same number of slots in all carriers, namely for every user i, it is $\alpha_{ij} = \alpha_i$ for $j = 1, \ldots, N$. Call each user i an item of size α_i and let each carrier of capacity C be a bin of size C. Then, the feasibility question is equivalent to the decision version of bin packing problem: "Given K items of sizes $\alpha_1, \ldots, \alpha_K$ and an integer N, is it possible to pack all items in N bins of size C?". This is known to be NP-Complete. Since instance $\mathcal{I}$ of our problem is equivalent to the decision version of bin packing, instance $\mathcal{I}$ is also NP-Complete.

Next, we show NP-Completeness for a general instance of the feasibility problem, where for each i it is $\alpha_{ij} \neq \alpha_{ik}$ for $j \neq k$. We use reduction to transform instance $\mathcal{I}$ to an instance $\mathcal{I}'$ of the general problem. We do that for $N = 2$ carriers and K items with sizes $\alpha_1, \alpha_2, \ldots, \alpha_K$. Assume without loss of generality that K is even, i.e, $K = 2\kappa$ for some integer κ. Given instance $\mathcal{I}$, we construct an instance $\mathcal{I}'$ of the general problem as follows. We have κ users with $\alpha_{i1} = \alpha_i$, and $\alpha_{i2} = \alpha_{\kappa+i}$, for $i = 1, \ldots, \kappa$. The carrier capacities in $\mathcal{I}'$ are C and $C + \sum_{i=1}^{\kappa} (\alpha_{i2} - \alpha_{i1})$ respectively. Then, instance $\mathcal{I}'$ is equivalent to $\mathcal{I}$, in the sense that a feasible allocation for $\mathcal{I}'$ exists if and only if a feasible allocation for $\mathcal{I}$ exists. Thus, the feasibility problem is NP-Complete. The optimization version of the problem is given by objective (2) subject to constraints (3), (4) and the integral constraint (5). This is the Generalized Assignment Problem (GAP) which is known to be NP-Hard [12].

B. Algorithms for Integral Assignment

It is desirable to design heuristic algorithms of reasonable complexity that provide feasible assignments with performance close to that of the optimal solution. In the sequel, we present two classes of algorithms, each of which follows a different rationale in allocation. Both consist of two phases: (i) initial and possibly infeasible assignment, and (ii) iterative improvements to reach an efficient feasible solution. In the next section, we show that these heuristics can be viewed as emerging from two Lagrangian relaxations. We assume that, at the beginning of the algorithm, the allocation controller obtains measurements and computes vector $\boldsymbol{\alpha} = (\alpha_{ij} : i = 1, \ldots, K$ and $j = 1, \ldots, N)$.

1) Class $\mathcal{A}_1$ of heuristics: In the first class of heuristics, the controller sorts parameters $\alpha_{i1}, \ldots, \alpha_{iN}$ for each user i in increasing order and constructs a preference list of carriers for each user. The algorithm starts by assigning each user to its best carrier in terms of required number of slots. A carrier is *overloaded* if its capacity constraint is violated, otherwise it is *underloaded*. If after initial assignment no carrier is overloaded, this is the optimal assignment. If all carriers are overloaded, no feasible assignment exists. The interesting and often arising case is that, after initial assignment, a set of carriers $\mathcal{S}_1$ may be overloaded and a set of carriers $\mathcal{S}_2$ are underloaded. In this case, reassignment is needed.

Fix attention to carriers j and k, where j is overloaded and k is underloaded after initial assignment. Users should be transferred from overloaded (and most preferable) carriers to underloaded (and less preferable) ones, if there exists sufficient residual capacity in the latter. Users must be transferred so that they induce minimal additional increase in slot occupancy. For each user i in overloaded carrier j, we define the *User-Carrier Transfer Factor (UCTF)* for the tentative transfer of i from carrier j to k as $\Lambda_i\left(j \to k\right) = \alpha_{ik}/\alpha_{ij}$, with $\Lambda_i\left(j \to k\right) \geq 1$. This factor captures transfer "efficiency". Among candidate users, we transfer the one that causes the minimal additional increase in used slots. User transfers with small UCTF values should occur first. Ties are broken by assigning an index to each user and selecting the smallest index. A feasible solution is an assignment of each user to a carrier so that all user requirements are satisfied and no carrier is overloaded. For $N = 2$ carriers, say j and k, let j be overloaded and k be underloaded after initial assignment. Let $\mathcal{U}_j$ be the set of users assigned to carrier j. The idea is to select user i_0 from carrier j, such that $i_0 = \arg\min_{i\in\mathcal{U}_j} \Lambda_i\left(j \to k\right)$ and transfer it to carrier k. User transfers are performed until either both carriers are underloaded or both are overloaded. In the former case we have a feasible solution, and in the latter case no feasible solution exists.

Consider now $N > 2$ carriers and let $N_1 = |\mathcal{S}_1|$ and $N_2 = |\mathcal{S}_2|$ be the number of overloaded and underloaded carriers. If $N_1 = 1$ and $N_2 > 1$, we start moving users from the overloaded carrier (say k) to underloaded ones. In that case, we must select a user i_0 in carrier j and transfer it to an appropriate underloaded carrier k_0, such that

$$(i_0, k_0) = \arg \min_{i\in\mathcal{U}_j, k\in\mathcal{S}_2} \Lambda_i\left(j \to k\right). \tag{6}$$

In the most general case with $N_1 > 1$ and $N_2 > 1$, we need to select a user i_0 in an overloaded carrier j_0 and move it to an underloaded carrier k_0 such that

$$(i_0, j_0, k_0) = \arg \min_{\substack{i\in\mathcal{U}_j \\ j\in\mathcal{S}_1, k\in\mathcal{S}_2}} \Lambda_i\left(j \to k\right). \tag{7}$$

User reassignments from overloaded to underloaded carriers terminate when all carriers become underloaded, or all become overloaded, or when no further reassignments from an overloaded to an underloaded carrier are possible because they lead to at least one overloaded carrier. There exist rare cases when reassignment of a user i from overloaded carrier j to k leads to infeasible assignment, while reassignment of user i' with $\Lambda_{i'}(j \to k) > \Lambda_i(j \to k)$ leads to a feasible assignment. The algorithm may consider this case by testing potential reassignment of all users in the overloaded carrier j in increasing order of UCTFs in instances when the algorithm seems to terminate with an infeasible solution. For $N = 2$ carriers, the algorithm above finds a feasible assignment whenever one exists. The complexity of the algorithm is computed as follows. Sorting N carriers for each user takes $O(N \log N)$ time and $O(KN \log N)$ time for K users. Picking the smallest element for each user takes $O(\log N)$ time. The computation of UCTFs per iteration is of complexity $O(N^2K)$. Usually, a bounded number of reassignments is needed before the algorithm terminates. Hence, the complexity of algorithms of class $\mathcal{A}_1$ is $O(N^2K + KN \log N)$.

2) Class $\mathcal{A}_2$ of heuristics: The algorithms of class $\mathcal{A}_2$ follow a different rationale. For each carrier j separately, the controller sorts parameters $a_{1j}, \ldots, a_{Kj}$ in increasing order and allocates users in carrier j in that order until no more users can be assigned to carrier j because of the capacity constraint. The procedure is repeated for all carriers. After initial assignment, there exist three types of users: users that are assigned to only one carrier, users that are assigned to more than one carrier and users that are not assigned to any carrier. Let $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3$ denote these sets of users and call them single-carrier, multiple-carrier and no-carrier users. Let $\mathcal{C}_i$ be the set of carriers to which a user $i \in \mathcal{U}_2$ is assigned. If $\mathcal{U}_3 = \emptyset$, the assignment is optimal. In this case, if $\mathcal{U}_2 \neq \emptyset$, for each user $i \in \mathcal{U}_2$ we just remove user i from all carriers except from the carrier where it uses the smallest number of slots. If $|\mathcal{U}_3| > \sum_{i\in\mathcal{U}_2} |\mathcal{C}_i|$, a feasible solution does not exist.

Consider the case where rearrangements are needed to find a feasible solution. Assignments for users in $\mathcal{U}_1$ should not change. The idea is to remove all slots of a user $i \in \mathcal{U}_2$ from an appropriate carrier $j \in \mathcal{C}_i$ in order to make room for a user $k \in \mathcal{U}_3$ to be assigned to carrier j. Preference should be given to users in $\mathcal{U}_2$ that use several slots in a carrier j (so that enough residual capacity is created by removing those slots) and to users in $\mathcal{U}_3$ that use few slots in j (so that the number of additionally used slots is as small as possible). We quantify these rules by User Substitution Factors (USFs) $\Phi_j(i \leftrightarrow k) = a_{kj}/a_{ij}$ for each no-carrier user k, each multiple-carrier user i and each carrier $j \in \mathcal{C}_i$. Clearly $\Phi_j(i \leftrightarrow k) \geq 1$. At each step we identify a user $k_0 \in \mathcal{U}_3$, a user $i_0 \in \mathcal{U}_2$ and a carrier $j_0 \in \mathcal{C}_{i_0}$ such that

$$(i_0, j_0, k_0) = \arg \min_{\substack{i\in\mathcal{U}_2, k\in\mathcal{U}_3 \\ j\in\mathcal{C}_i}} \Phi_j(i \leftrightarrow k), \tag{8}$$

and perform user substitution while satisfying the capacity constraint of carrier j after the substitution. The sequence of user substitutions ends with a feasible assignment when $\mathcal{U}_3| = \emptyset$. If at the end of the algorithm $\mathcal{U}_2$ is non-empty, then for each user $i \in \mathcal{U}_2$ we eliminate slots from all carriers i where is assigned , except from the one where it uses the smallest number of slots. The complexity is computed as follows. Sorting K users for each carrier takes $O(K \log K)$ time and $O(NK \log K)$ time for N carriers. Picking the best users for each carrier takes $O(\log K)$ time. The computation of USFs per iteration is of complexity $O(K^2N)$. Usually, there exists a bounded number of reassignments before the algorithm terminates. Hence, the complexity of algorithms of class $\mathcal{A}_2$ is $O(K^2N + NK \log K)$.

C. Fractional Assignment

Fractional user assignment arises in OFDM, where a user may be partially assigned to and use slots of more than one carrier. Vector $\mathbf{x}$ is real-valued and its elements satisfy $0 \leq x_{ij} \leq 1$ and are treated as a divisible fluid. Minimizing $Z(\mathbf{x})$ in (2) subject to constraints (3), (4) and the real-valued (5) is a Linear Programming (LP) problem with NK variables and $NK + N + K$ constraints that can be solved optimally with the Simplex method [13]. However, due to the large number of variables and constraints, this may involve large computational load. Our goal here is to use the rationale of Simplex to obtain useful insights and design efficient implementable algorithms.

Consider $N = 2$ carriers, each with C slots, and K users. Let α_i and β_i be the number of required slots for user i if allocated only to carrier 1 or 2 respectively. If x_i is the fraction of requirements of i that is satisfied by carrier 1, the corresponding fraction for carrier 2 is $1 - x_i$. The problem is to find the real vector $\mathbf{x} = (x_i : i = 1, \ldots, K)$ that minimizes total number of occupied slots. This is formulated as

$$\min_{\mathbf{x}} Z(\mathbf{x}) = \sum_{i=1}^{K} \left[\alpha_i x_i + \beta_i (1 - x_i)\right] \quad (9)$$

subject to:

$$\sum_{i=1}^{K} \alpha_i x_i \le C, \text{ and } \sum_{i=1}^{K} \beta_i (1 - x_i) \le C \quad (10)$$

$$0 \le x_i \le 1, \; i = 1, \ldots, K. \quad (11)$$

Objective (9) can be written as $Z(\mathbf{x}) = \sum_{i=1}^{K} (\alpha_i - \beta_i) x_i + \sum_{i=1}^{K} \beta_i$. Since we want to minimize it, we have that if $\alpha_i > \beta_i$, then x_i should be small or ideally zero so as to induce the smallest increase. Thus, user i should use carrier 2 as much as possible since this carrier has better quality. If $\alpha_i < \beta_i$, carrier 1 should be preferred, namely x_i should be large and ideally equal to 1 to reduce the objective function value.

An optimal algorithm for $N = 2$ carriers: The following algorithm finds the optimal solution for $N = 2$ carriers. Each user is initially assigned to the best carrier. If both capacity constraints are satisfied, then this is the optimal assignment. If no constraints are satisfied, no feasible solution exists. If one of the two constraints is satisfied, users are transferred from the overloaded carrier to the underloaded one such that they induce minimal additional slot increase. For each user i, this is captured by ratios α_i/β_i or β_i/α_i, depending on which carrier is overloaded. If carrier 1 is overloaded, users are transferred from carrier 1 to 2 in increasing order of ratios β_i/α_i until both constraints are satisfied. The last user in carrier 1 whose reassignment makes carrier 1 underloaded is assigned fractionally to both carriers. The fraction of this user's requirements in carrier 1 is such that the capacity constraint in carrier 1 is tightly satisfied. The remaining portion of its requirements is assigned to carrier 2.

This algorithm has some interesting properties. In the optimal solution, at most one user is fractionally assigned, and this is the last reassigned one from carrier 1 to 2. Other users use only one of the two carriers. Thus, at most one of the K variables x_i is fractional, while the rest are 0 or 1. In addition, if $\alpha_i > \beta_i$ for user i, we have non-zero x_i at the optimal solution only if carrier 2 is filled to its capacity. That is, a user is assigned to a lower quality carrier only if the more preferable carrier is filled to its capacity. The final value of the objective function is equal to the one obtained by the Simplex method. Furthermore, the formulation and algorithm above are related to the fractional Knapsack problem which go as follows. Given K items, each of weight w_i and value v_i, find portions x_i of each item i to maximize total value $\sum_i v_i x_i$ subject to a total weight constraint $\sum_i w_i x_i \le C$. There exists a greedy algorithm that solves optimally this problem. Items are selected in decreasing order of ratios v_i/w_i. If the weight constraint is violated during a reassignment, that item is selected fractionally so that weight constraint is tightly satisfied. The Knapsack in our case is the initially underloaded carrier, say carrier 2, and ratios α_i/β_i correspond to ratios v_i/w_i in the Knapsack.

The greedy reassignment ratios above for integral and fractional assignment is similar in flavor to the criterion for moving among basic feasible solutions in the Simplex method. Note however that in our case we do not move from one feasible solution to another. The extension of our rationale to $N > 2$ carriers is not straightforward. Users can be reassigned from appropriate overloaded carriers to underloaded ones in the spirit of the algorithm above. The user whose reassignment makes an overloaded carrier j underloaded is fractionally assigned to carriers j and k. However, a new preference factor needs to be defined after the reassignment, since fractionally assigned users may need to further split their requirements among carrier k and other carriers ℓ if k becomes overloaded. The algorithm either ends with a feasible assignment or comes up with an infeasible instance.

Remark: If the problem instance is infeasible, a meaningful objective would be to maximize total supportable rate when all carriers are fully utilized. This is expressed by the LP problem of maximizing $\sum_{i=1}^{K} \sum_{j=1}^{N} r_i x_{ij}$ subject to $\sum_{j=1}^{N} x_{ij} = 1$ for $i = 1, \ldots, K$, $\sum_{i=1}^{K} \alpha_{ij} x_{ij} = C$ for $j = 1, \ldots, N$ and $0 \le x_{ij} \le 1$, $\forall i, j$.

D. *Allocation of groups of carriers*

The algorithms above assign single carriers and apply to systems with small to medium number of carriers, such as IEEE 802.11a/g with 48 data carriers. In systems with a large number of carriers, such as WiMAX based with 1024 carriers, these methods are costly in terms of computational load and required bandwidth for carrier quality measurements. In addition, if SIR measurements can be obtained for only few carriers in a group, the group needs to be handled as one entity of quality that is determined by measurements of representative carriers. Lastly, user rate requirements may be high enough, such that the allocation of a group of carriers makes sense. It is thus meaningful to extend the algorithms to groups of carriers rather than individual carriers.

Let the N carriers be divided into G given disjoint groups $\mathcal{G}_j$, $j = 1, \ldots, G$. A group may have contiguous carriers in frequency, and the group size may be decided a priori, e.g. based on average coherence bandwidth of users, so that frequency responses of carriers in the same group are reasonably close to each other. Another possibility is that carriers in the same group are evenly spaced in frequency for diversity. For each user i, a group of carriers $\mathcal{G}_j$ is characterized collectively by a maximum sustainable modulation level b_{ij} based on SIR measurements in one or more carriers of the group. For example, b_{ij} may be determined by the minimum SIR in the group. All carriers in group $\mathcal{G}_j$ have the same sustainable modulation level, otherwise we resort to individual carrier allocation. For user i and carrier group $\mathcal{G}_j$, the assignment preference factor a_{ij} denotes the number of slots that are needed for user i to satisfy rate requirements if i is given only time slots from carriers in $\mathcal{G}_j$. Time slots from carriers in group $\mathcal{G}_j$ are allocated to a user by treating a carrier group as a single virtual carrier of $C|\mathcal{G}_j|$ slots, where $|\mathcal{X}|$ is the cardinality of set $\mathcal{X}$. The algorithms for integral and fractional carrier assignment can be easily extended to integral and fractional carrier group assignment, where a user can be allocated to one or more carrier groups.

IV. Performance bounds

Performance bounds are benchmarks for evaluating performance of different heuristics. Furthermore, the procedure of deriving a bound can draw guidelines for finding good feasible solutions and for designing low-complexity efficient heuristics. Lagrangian relaxation [14] can provide good enough (namely, high) lower bounds for minimization problems. In Lagrangian relaxation, a subset of the constraints of the original problem are relaxed. Each relaxed constraint is multiplied by a Lagrange multiplier and is added to the objective function. The problem with relaxed constraints is usually easier to solve than the original one. Given a set of Lagrange multipliers, the solution to the relaxed problem provides a lower bound on the value of the objective function of the original problem. The corresponding Lagrangian dual problem is to find the values of multipliers that maximize this lower bound. We now elaborate on integral assignment, for which two Lagrangian relaxations, LR1 and LR2 arise. Lagrangian relaxation can be applied to fractional assignment as well.

A. Lagrangian Relaxation LR1 and relation to our algorithm

In relaxation LR1 we relax capacity constraints. For a given Lagrange multiplier vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N)$, we define the Lagrangian

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}) &= \sum_{i=1}^{K}\sum_{j=1}^{N} \alpha_{ij}x_{ij} + \sum_{i=1}^{K}\sum_{j=1}^{N} \lambda_j(\alpha_{ij}x_{ij} - C) \\ &= -KC\sum_{j=1}^{N}\lambda_j + \sum_{i=1}^{K}\sum_{j=1}^{N}\alpha_{ij}(1+\lambda_j)x_{ij}. \end{aligned} \quad (12)$$

The relaxed original problem is

$$Z_I(\mathbf{x}) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) \quad (13)$$

$$\text{subject to:} \quad \sum_{j=1}^{N} x_{ij} = 1,\ i = 1, \ldots, K, \ \ \mathbf{x} \in \{0,1\}^{KN}$$

and the Lagrangian dual problem is

$$Z_{LD}^1 = \max_{\boldsymbol{\lambda} \geq \mathbf{0}} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) \quad (14)$$

$$\text{subject to:} \quad \sum_{j=1}^{N} x_{ij} = 1,\ i = 1, \ldots, K, \ \ \mathbf{x} \in \{0,1\}^{KN}.$$

For given $\boldsymbol{\lambda}$, problem (13) is solved by the assignment $\mathbf{x}(\boldsymbol{\lambda})$, such that $x_{ij^*}(\boldsymbol{\lambda}) = 1$, if $j^* = \arg\min_j[\alpha_{ij}(1+\lambda_j)]$ and $x_{ij}(\boldsymbol{\lambda}) = 0$, for $j \neq j^*$. The supergradient method [15, p.173-174] can be applied to solve (14). Each iteration of the supergradient method involves two kinds of updates: variable update and Lagrange multiplier adjustment. Variable update at step n involves applying the solution $\mathbf{x}^*(\boldsymbol{\lambda}(n))$ above for given multiplier vector $\boldsymbol{\lambda}(n)$. The multiplier adjustment at step $n+1$ is made according to $\lambda_j(n+1) = \left[\lambda_j(n) + \mu_n(\sum_{i=1}^{K}\alpha_{ij}x_{ij}(n) - C)\right]^+$ for $j = 1, \ldots, N$, where $x^+ = x$ if $x > 0$ and 0 otherwise, and μ_n is the step size. It turns out that the lower bound provided by LR1 after convergence of the supergradient method equals that of LP, namely $Z_{LD}^1 = Z_{LP}$.

LR1 can provide good feasible solutions and defines the class $\mathcal{A}_1$ of heuristics described above. Initially each user is assigned to the carrier where it uses minimum number of slots under no capacity constraints. Note that for $\boldsymbol{\lambda} = \mathbf{0}$, $\mathbf{x}(\mathbf{0})$ corresponds to the initial assignment with each user assigned to its best carrier. User reassignments from overloaded to underloaded carriers are reminiscent of variable updates and Lagrange multiplier adjustments of supergradient method.

B. Lagrangian Relaxation LR2 and associated algorithms

In relaxation LR2, we relax assignment constraints. Each carrier has a capacity constraint, but a user can be assigned to more than one carriers. For given $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)$, we have

$$\begin{aligned} \hat{L}(\mathbf{x}, \boldsymbol{\lambda}) &= \sum_{i=1}^{K}\sum_{j=1}^{N}\alpha_{ij}x_{ij} + \sum_{i=1}^{K}\lambda_i\Big(\sum_{j=1}^{N}x_{ij} - 1\Big) \\ &= \sum_{i=1}^{K}\lambda_i + \sum_{i=1}^{K}\sum_{j=1}^{N}(\alpha_{ij}+\lambda_i)x_{ij}. \end{aligned} \quad (15)$$

The relaxed original problem is

$$Z_I(\mathbf{x}) = \min_{\mathbf{x}} \hat{L}(\mathbf{x}, \boldsymbol{\lambda}) \quad (16)$$

$$\text{subject to:} \quad \sum_{i=1}^{K}\alpha_{ij}x_{ij} \leq C,\ j = 1, \ldots, N, \ \ \mathbf{x} \in \{0,1\}^{KN}$$

and the Lagrangian dual problem is

$$Z_{LD}^2 = \max_{\boldsymbol{\lambda}} \min_{\mathbf{x}} \hat{L}(\mathbf{x}, \boldsymbol{\lambda}) \quad (17)$$

$$\text{subject to:} \quad \sum_{i=1}^{K}\alpha_{ij}x_{ij} \leq C,\ j = 1, \ldots, N, \ \ \mathbf{x} \in \{0,1\}^{KN}.$$

For given $\boldsymbol{\lambda}$, the relaxed problem (16) becomes a set of N independent Knapsack problems, one for each carrier, that are solved at each iteration n for multiplier vector $\boldsymbol{\lambda}(n)$. The supergradient method here involves multiplier updates $\lambda_i(n+1) = \left[\lambda_i(n) + \mu_n(\sum_{j=1}^{N}x_{ij}(n) - 1)\right]^+$, $i = 1, \ldots, K$. The obtained lower bound from LR2 is higher that that of LP, namely $Z_{LD}^2 \geq Z_{LP}$ [15]. In analogy to LR1 and algorithm class $\mathcal{A}_1$, relaxation LR2 gives rise to class $\mathcal{A}_2$ of algorithms. Each carrier is treated separately and users are initially assigned to each carrier as in Knapsack to minimize incurred cost subject to a carrier capacity constraint. Substitution of multiple-carrier users with no-carrier users in appropriate carriers parallel the iterative variable updates and multiplier adjustments of the supergradient method.

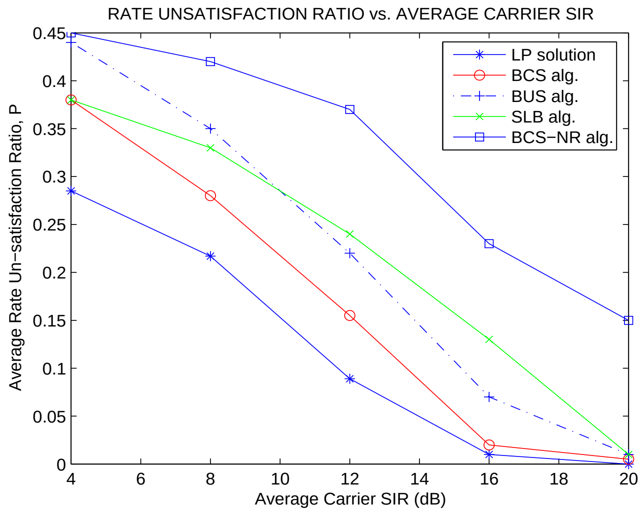

Fig. 2. Unsatisfied user rate requirements vs. average carrier SIR.

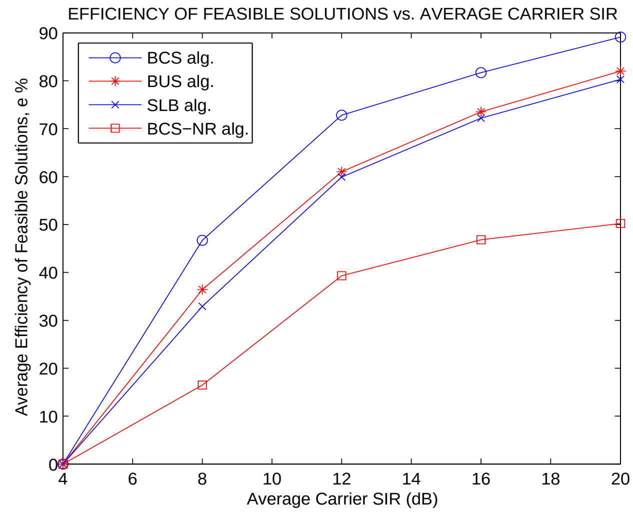

Fig. 3. Efficiency of feasible solutions vs. average carrier SIR.

V. Numerical results

We now evaluate the performance of our algorithms and compare it to derived performance bounds and other heuristics. A problem instance is specified by vector $\boldsymbol{\alpha}$. Different instances may lead to a feasible solution or may be infeasible. A first performance metric is the ability of an algorithm to identify feasible solutions. This ability is quantified by the *Rate Unsatisfaction Ratio*,

$$P = \frac{\text{Unsatisfied user requirements (bps)}}{\text{Total requirements (bps)}} = \frac{N_u}{\sum_{i=1}^{K} r_i}, \quad (18)$$

where N_u is the total user requirements that are not satisfied t the solution. This metric captures percentage of residual user requirements that remain unsatisfied. The number of users whose rate requirements are fully satisfied is also considered as a performance measure of similar nature. Another performance metric concerns proximity of the resulting value of the objective function to that of the optimal (LP) solution. This is captured by the *Efficiency* of a feasible solution,

$$e = \frac{\text{Number of used slots in LP solution}}{\text{Number of used slots in the feasible solution}}, \quad (19)$$

with $0 \leq e \leq 1$. A value of e close to 1 implies that an algorithm can find good feasible solutions which are close to the optimal one. We consider the following algorithms:

- Best Carrier Selection (BCS). This is the algorithm of class $\mathcal{A}_1$ based on user reassignments from subsection III-B1. The algorithm performs integral assignment.
- BCS algorithm with no reassignments of users (BCS-NR). Once users are assigned to their best carriers, no reassignments occur.
- Best User Selection (BUS). This is the algorithm of class $\mathcal{A}_2$ for integral assignment with user substitutions from subsection III-B2.
- Linear Programming (LP) solution. This performs fractional assignment and is found with MATLAB Simplex.
- Subcarrier Load Balancing (SLB) algorithm. The algorithm first assigns each user to its best carrier. User reassignments are performed such that carrier loads are as balanced as possible. Namely, a reassignment is performed if it minimizes the maximum load difference among carriers.

We consider a system with $N = 10$ carriers with $C = 10$ slots each and $K = 35$ users with same rate requirements, $r = 150$ kbps. The symbol, slot and frame durations are $T_{\text{sym}} = 10^{-5}$ sec, $T_s = 10^{-4}$ sec and $T_f = 10^{-3}$ sec. There exist $L_0 = 6$ modulation levels and modulation level i is associated with threshold γ_i dB that is computed for BER $= 10^{-3}$ per slot. Carrier quality for a user depends on interference and multipath. The effect of interference is captured by Gaussian random variables I, i.i.d for each user and carrier, with mean μ dB and variance σ_I^2 dB. The dependence of carrier n gain G_n on the multipath channel with L paths and channel response $h(t) = \sum_{\ell=0}^{L-1} A_\ell \delta(t-\tau_\ell)$ (where A_ℓ, τ_ℓ are the gain and delay of the ℓ-th path) is given by $G_n = |X_n|^2 = |\sum_{\ell=0}^{L-1} A_\ell e^{-j2\pi\tau_\ell(f_c+n/T)}|^2$, where f_c is the reference carrier frequency. Assuming large L, each of the two quadrature components of X_n is Gaussian with zero mean and variance $\frac{1}{2}(a^2+\sigma_A^2)$, where a, σ_A^2 are the mean and variance of A_ℓ irrespective of distribution. Furthermore, G_n follows exponential distribution with mean $a^2+\sigma_A^2$. For the experimental setup, the G_n's are generated for $L = 3$ by Gaussian A_ℓ's and τ_ℓ's uniformly distributed in $[0, T]$.

First, we study performance in terms of *capability to identify feasible solutions*. Figure 2 depicts ratio P for BCS, BCS-NR, BUS, CLB and LP methods as a function of different average SIR ratios which denote different carrier quality. For each average SIR value, we consider 200 scenarios, each of which is defined by different variables $\{G_n\}$ and I for each user. Results are averaged over these scenarios. The ability of all techniques to identify feasible solutions increases with average SIR as expected, since the percentage of infeasible instances decreases as users require fewer slots to fulfil their rate requirements and can be accommodated in carriers. The LP solution provides a lower bound on ratio P. Since the LP solution corresponds to fractional assignment, this assignment results in the largest number of feasible solutions. Fractional user assignment is very efficient for average SIR ≥ 14dB.

TABLE I
AVERAGE NUMBER OF SUPPORTABLE USERS FOR DIFFERENT ALLOCATION APPROACHES

Avg. SIR	LP	BCS	TDMA-based	FDMA-based
4dB	12.1	8.6	8.7	8.6
8dB	20.4	12.1	9.8	9.8
12dB	26.2	18.7	10	10
16dB	32.3	28.4	10	10
20dB	35	32.2	10	10

The BCS algorithm generates fewer feasible solutions than LP, since fewer users are accommodated and higher percentage of requirements is unsatisfied. For small SIR values (e.g. less than 8dB) the performance of BCS is inferior to that of LP by $33-35\%$, while for larger average SIRs, it approaches that of LP. For SIR > 15dB, integral and fractional assignment have very similar performance since the number of required slots decreases. The SLB algorithm belongs to the class of heuristics that do not use carrier quality as assignment criterion, and thus it is effective only in high SIRs. However, for low SIRs, the differences in carrier quality incur assignments to inappropriate carriers and lead to unnecessary consumption of slots. The BUS algorithm has performance similar to SLB for medium SIRs, while it is inferior to BCS. This implies that the class $\mathcal{A}_1$ of heuristics gives more feasible solutions than class $\mathcal{A}_2$. Finally, we draw the curve for the greedy one-shot algorithm BCS-NR that assigns once each user to its best carrier to show the significance of reassignments, especially in low and medium SIRs.

Figure 3 illustrates the *efficiency* e of feasible solutions for BCS, BUS, SLB and BCS-NR algorithms, captured by proximity to the LP solution. Feasible solutions for the heuristics are not generated for average SIR of 4dB. The quality of feasible solutions for BCS improves for larger SIRs. For moderate SIR values, the BCS solution is within $30-40\%$ from the optimal one, while for larger SIRs, it is within $10-20\%$. We also observe that the quality of solutions of the SLB algorithm is close to that of BCS. Indeed, since we focus on feasible solutions and carriers are filled almost up to their capacities, a feasible solution for both the BCS and SLB algorithms involves almost balanced assignments in different carriers. Furthermore, the performance of BUS is very similar to that of SLB and in several cases about $10-15\%$ lower than that of BCS. Finally, feasible solutions for algorithm BCS-NR result in at least twice the number of used slots compared to that of the optimal solution.

The proposed algorithms of classes $\mathcal{A}_1$ or $\mathcal{A}_2$ for integral assignment start with an initial assignment and iteratively improve it toward a good feasible one. Another type of algorithms is presented in [9], [10] for power minimization. These focus on carrier allocation in an OFDM symbol interval, within which at most one user can load bits to one carrier. They find an initial assignment and iteratively improve it by user reassignments. The difference is that resource allocation is performed step by step for each user in a round-robin fashion until either user requirements are satisfied or there is no unallocated carrier. We compare this approach to BCS and BUS algorithms. To make a fair comparison, we adjust the approach of [9],[10] to our problem as follows. We define a quantity of bits, $\kappa b_i x_i$, where b_i is the modulation level of user i in a carrier, x_i is the allocated portion of carrier capacity to that user and κ is constant. For each user i we create an ordered preference list of carriers. We go through users sequentially and for each user we consider the first carrier in its preference list. If the carrier has residual capacity and the user requirement is not satisfied, we assign a portion of capacity of that carrier to the user according to allowable modulation level, else we proceed to the next user. In the next round we consider the second carrier in each user's preference list and repeat the process. The procedure continues until either users satisfy requirements, or all user preference lists are exhausted. User reassignments are then performed in the following order. For each user pair (i, j) and carrier n originally used by user i, we compute the resource reduction factor $dZ_{i,j}(n)$ if the portion allocated to user i in carrier n is removed and another portion is allocated to user j in carrier n (subject to capacity constraints). We find the maximum of these factors over carriers n, $\Delta Z_{i,j}$. Similarly, we find $\Delta Z_{j,i}$. For each pair of users (i, j), we find $Z(i, j) = \Delta Z_{i,j} + \Delta Z_{j,i}$ and order them in descending order. We consider reassignments in that order, provided they cause reduction in used carrier capacity, until no reassignments can further improve performance. For SIR values in $[4, 20]$dB with step 4dB, the average P is measured as 0.38, 0.35, 0.21, 0.05 and 0.0 respectively, showing that this approach is inferior to BCS but better than BUS.

Finally, in order to show the benefits of performing the allocation in both time and frequency domains, we consider two fixed resource allocation schemes [8], namely

- TDMA-based, where each user is allocated a slot and uses exclusively all carriers within that slot; no other user can use carriers within that slot. Selected users i for allocation in time slots are those with best overall carrier quality, captured by term $\sum_{j=1}^{N} \alpha_{ij}$ for $i = 1, \ldots, K$.
- FDMA-based, where each user is allocated a set of carriers and uses slots in those carriers; no other user can use slots in these carriers. The user i allocated to each carrier j is the one with smallest α_{ij}.

In table I, we present comparative results for average number of supportable (i.e, fully satisfied) users for LP, BCS, TDMA-based and FDMA-based carrier assignment schemes and different average carrier SIRs. The fixed schemes cannot accommodate more than C and N users respectively, regardless of channel quality. For small SIRs, a slot for the TDMA-based scheme or a carrier for the FDMA scheme may be inadequate for accommodating even one user. For high SIRs, BCS achieves performance within 10% of the optimal LP one, while for small SIRs, it has performance similar to that of TDMA and FDMA based allocation. The performance benefits of LP and BCS that stem from performing the allocation in both the time and frequency domains become more pronounced for moderate and high SIRs.

VI. Conclusion

We considered the static carrier assignment problem that arises in OFDM or other multi-carrier systems and studied fractional and integral assignment of users to carriers. The objective was to satisfy user requirements (per-frame or per-slot) with minimal resources. We characterized the complexity of integral assignment and presented two greedy iterative heuristic algorithms. We used Lagrangian relaxation to provide performance bounds and to define a framework for efficient heuristics that identify good feasible solutions.

There exist several directions for future study. Per-carrier power control adds a new dimension to the problem in OFDM. Selective user assignment to carriers and power control could be used jointly to ensure an acceptable peak-to-average power ratio (PAPR) at each transmitter. Another issue is distributed implementation of algorithms in the uplink, so that no central coordination is needed. For example, in heuristic $\mathcal{A}_1$, the BS computes the Lagrange multiplier vector with the supergradient method and sends it to users. Each user then solves an independent optimization problem and selects a carrier. Users communicate their preferences to the base station, which then computes the new multiplier vector and so on. It would be interesting to investigate such practical distributed approaches and compare their performance to the optimal solution. Finally, another extension would be carrier allocation in the context of multicasting to some user multicast groups. On the one hand, efficient resource utilization implies that data be transmitted to users of a multicast group with the same carriers and slots. On the other hand, users of the same multicast group have different carrier quality. Thus, carrier quality across users is not consistent with multicast group membership. Resource allocation across multicast groups and users can be considered for future investigation.

VII. Acknowledgments

The authors acknowledge support from Greek Secretariat for Research and Technology through a PENED grant and support of European Commission NoE NEWCOM++ (FP7-ICT-216715).

References

[1] I. Koutsopoulos and L. Tassiulas, "Carrier assignment algorithms in wireless broadband networks with channel adaptation", *Proc. IEEE Intern. Conf. Commun. (ICC),* vol.5, pp. 1401-1405, 2001.

[2] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems", *IEEE Trans. Commun.,* vol.47, no.6, pp.884-894, June 1999.

[3] T. Ikeda, S. Sampei and N. Morinaga, "TDMA-based adaptive modulation with dynamic channel assignment for high-capacity communications systems", *IEEE Trans. Veh. Tech.,* vol.49, no.2, pp.404-412, March 2000.

[4] W. Yu and J.M. Cioffi, "FDMA capacity of Gaussian multi-access channels with ISI", *IEEE Trans. Commun.,* vol.50, no.1, pp.102-111, Jan. 2002.

[5] L.M.C. Hoo, B. Halder, J. Tellado and J.M. Cioffi, "Multiuser transmit optimization for multicarrier broadcast channels: asymptotic FDMA capacity region and algorithms", *IEEE Trans. Commun.,* vol.52, no.4, pp. 922-930, June 2004.

[6] W. Rhee and J.M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation", *Proc. Veh. Tech. Conf.,* pp.1085 -1089, 2000.

[7] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks-Part I: Theoretical framework", *IEEE Trans. Wireless Commun.,* vol.4, no.2, pp.614-624, March 2005.

[8] C.Y. Wong, R.S. Cheng, K. Ben Letaief and R.D. Murch, "Multiuser OFDM with adaptive subcarrier, bit and power allocation", *IEEE J. Select. Areas Commun.,* vol.17, no.10, pp.1747-1758, Oct. 1999.

[9] C.Y. Wong, C.Y. Tsui, R.S. Cheng and K.B. Letaief, "A Real-time subcarrier allocation scheme for multiple access downlink OFDM transmission", *Proc. IEEE Veh. Tech. Conference,* pp.1124-1128, 1999.

[10] Y.-F. Chen, J.-W. Chen and C.-P. Li, "A fast suboptimal subcarrier, bit, and power allocation algorithm for multiuser OFDM-based systems," *Proc. IEEE Intern. Conf. Commun. (ICC),* pp. 3212-3216, 2004.

[11] J. Chuang and N. Sollenberger, "Beyond 3G: Wideband wireless data access based on OFDM and dynamic packet assignment", *IEEE Commun. Mag.,*, vol.38, no.7, pp.78-87, July 2000.

[12] M.L. Fisher, R. Jaikumar and L.N. Van Wassenhove, "A multiplier adjustment method for the generalized assignment problem", *Management Sci.,* vol.32, no.9, pp.1095-1103, Sept. 1986.

[13] D.G. Luenberger, *Introduction to linear and non-linear programming,* Addison-Wesley, 1984.

[14] M.L. Fisher, "The Lagrangian relaxation method for solving integer programming problems", *Management Sci.,* vol.27, no.1, pp.1-18, Jan. 1981.

[15] L.A. Wolsey, *Integer Programming,* Wiley 1998.

PLACE PHOTO HERE

I ordanis Koutsopoulos (S '99, M '03) is Lecturer at the Department of Computer and Telecommunications Engineering, University of Thessaly, Greece. He was born in 1974 in Athens, Greece. He obtained the Diploma in Electrical and Computer Engineering from the National Technical University of Athens, Athens, Greece in 1997 and the M.S and Ph.D degrees in Electrical and Computer Engineering from the University of Maryland, College Park (UMCP) in 1999 and 2002 respectively. From 1997 to 2002 he was a Fulbright Fellow and a Research Assistant with the Institute for Systems Research (ISR) of UMCP. He has held internship positions with Hughes Network Systems, Germantown, MD, Hughes Research Laboratories LLC, Malibu, CA, and Aperto Networks Inc., Milpitas, CA, in 1998, 1999 and 2000 respectively. For the summer period of 2005 he was a visiting scholar with University of Washington, Seattle, WA. For the period 2005-2007 he was awarded a Marie Curie International Reintegration Grant (IRG). His research interests are in the field of networking with emphasis on wireless networks, cross-layer design, wireless network security and peer-to-peer networks.

PLACE PHOTO HERE

L eandros Tassiulas (S'89, M'91, SM/06, F07) obtained the Diploma in Electrical Engineering from the Aristotelian University of Thessaloniki, Thessaloniki, Greece in 1987, and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Maryland, College Park in 1989 and 1991 respectively.

He is Professor in the Dept. of Computer and Telecommunications Engineering, University of Thessaly, since 2002. He has held positions as Assistant Professor at Polytechnic University New York (1991-95), Assistant and Associate Professor University of Maryland College Park (1995-2001) and Professor University of Ioannina Greece (1999-2001). His research interests are in the field of computer and communication networks with emphasis on fundamental mathematical models, architectures and protocols of wireless systems, sensor networks, high-speed internet and satellite communications. Dr. Tassiulas is a Fellow of IEEE. He received a National Science Foundation (NSF) Research Initiation Award in 1992, an NSF CAREER Award in 1995, an Office of Naval Research, Young Investigator Award in 1997 and a Bodosaki Foundation award in 1999. He also received the INFOCOM 1994 best paper award and the INFOCOM 2007 achievement award.