Optimal and Near-Optimal Resource Allocation Algorithms for OFDMA Networks

Yuan-Bin Lin, Tai-Hsiang Chiu, and Yu T. Su, *Senior Member, IEEE*

Abstract—Given the availability of multiple orthogonal channels and multimedia transmission rate requirements from multiple wireless OFDMA users, we are interested in a joint channel, power and rate assignment scheme that satisfies the given requirements with the minimum total transmit power.

Algorithms for finding suboptimal and optimal solutions to sum power minimization resource allocation problems in OFDMA-based networks have been proposed. But the complexity of finding the optimal solution is prohibitively high. We present two efficient algorithms with which each channel (sub-carrier) is assigned to at most one user. The first approach, which gives near-optimal solutions, employs a dynamic programming (DP) based tree search and adopts a fair initial condition that offers every user all available channels and removes a channel from all but one user at each stage. Each removal is based on the criterion of the least total power increase. Using the DP-based solution as the initial upper bound and the partial path cost used in the DP approach as the lower bound for each visited node, we develop an efficient branch-and-bound based algorithm that guaranteed to lead to the optimal solution.

The average complexities of both algorithms are evaluated and effective schemes to further reduce the required complexity are proposed. We also provide performance and complexity comparisons with other suboptimal algorithms that are modified from the existing ones.

Index Terms—Radio resource allocation, OFDMA, broadband communication, dynamic programming, branch-and-bound.

I. INTRODUCTION

AS the demand for high data rate multi-media wireless communications increases, it also becomes more and more important that one takes into account the energy/spectral efficiency factor in designing an anti-fading transmission scheme for mobile terminals. A fast and proper adaptive algorithm in allocating both the physical and MAC layer resources is essential to provide high quality high rate multiuser transmissions. Because of its robustness against frequency-selective fading and its flexibility in appropriating the transmission resources, the OFDM-based Frequency Division Multiple Access (OFDMA) scheme in which each user is allocated a collection of time slots and sub-carriers for transmission, has been adopted in several industrial wireless communication

standards. If the allocation is predetermined and static, there may be unused sub-carriers and time slots if the designated users do not need so many signal dimensions.

When there are limited power and multiple orthogonal channels available for transmitting multiuser/multimedia signal, a proper channel and power allocation scheme is needed to minimize the average power consumption, co-channel interferences while meeting various users and media's rate requirement and maintaining the link quality. For an OFDMA system, this problem is complicated by the fact that a subcarrier (channel)¹ is bad, in deep fade and with low channel signal-to-noise ratio (SNR) for one user may be good (with high channel SNR) for another user. In [1], the authors proposed a suboptimal multiuser subcarrier/bit allocation scheme which minimizes the total transmit power with rate constraints. They relaxed the discrete-(integer-)rate constraint by allowing time-sharing use of a subcarrier by multiple users—an idealized assumption that was subsequently used by many investigators. [4] considered a continuous-rate version of the same problem but forbade the multiple-user-per-subcarrier scenario and suggested a method for computing the optimal solution. Obtaining the exact optimal solutions to either problem requires high computing complexity and become impractical for large channel and/or user constraints. There are many works that studied variations and extensions of [1] or [4]; each deals with different objective function (maximizing weighted sum rate [9], utility [12]), constraint (fairness [7], proportional rates [8]), or scenario (e.g., multi-cell [10], relay-aided [7]). A survey on various dynamic resource allocation (RA) solutions was recently reported in [6]. Further discussion on some suboptimal solutions will be given in Section V.A.2.

In this paper we present two efficient algorithms for solving the problems of [1] and [4], i.e., we are concerned with efficient subcarrier, power and rate assignments that satisfy multi-user multi-media requirements with the minimum total transmitted power. The first algorithm uses a dynamic programming (DP) approach; it is simple and offers near-optimal performance. The second algorithm invokes the branch-and-bound (B&B) principle, uses a good initial bound and tight lower bounds along with some complexity-reduction techniques. It gives the optimal solution with a moderate increase of complexity. Our discourse concentrates on the continuous-rate case but both algorithms can be used for the discrete-rate case with a minor modification (see Section IV.D). It is not difficult to see that, through suitable modifications, our algorithms can also be applied to solve a similar RA problem

Manuscript received February 16, 2008; revised August 3, 2008, April 17, 2009, and April 20, 2009; accepted May 8, 2009. The associate editor coordinating the review of this paper and approving it for publication was R. M. Buehrer.

Y.-B. Lin and Y. T. Su are with the Department of Communications Engineering, National Chiao Tung University, Hsinchu, 30056, Taiwan (e-mail: fred.yplin@gmail.com; ytsu@mail.nctu.edu.tw).

T.-H. Chiu is with MediaTek Inc., Hsinchu, 30078, Taiwan (e-mail: noah.chiu@mediatek.com).

This work was supported by the National Science Council of Taiwan under Contract 95-2221-E-009-160.

Digital Object Identifier 10.1109/TWC.2009.080221

¹We shall use the terms *subcarrier* and *channel* interchangeably throughout this paper.

of maximizing the aggregated throughput or weighted sum rate with individual power constraints. The rest of this paper is organized as follows. The ensuing section describes the operation scenarios of concern and gives an optimization problem formulation. Section III presents the proposed DP-based resource allocation algorithm and the B&B-based approach is given in Section IV. We also derive some useful properties and suggest design guidelines there. Numerical performance of the proposed algorithm and some existing suboptimal algorithms is presented in Section V. Finally, we give concluding remarks in Section VI and derive an optimal mono-rate (single user) power allocation (OMPA) algorithm in Appendix A.

II. SYSTEM ASSUMPTIONS AND PROBLEM FORMULATION

A. Basic assumptions

We assume that there are N orthogonal subcarriers. $\mathbf{C} = \{1, 2, \dots, N\}$ and d user data streams with the rate requirements $\mathbf{R} = \{R_j, j = 1, 2, \dots, d\}$ to be transmitted over an OFDMA downlink, where the required transmission rate of user j is denoted by R_j . We further assume that the base station and d user terminals are each equipped with single antenna. The base station assigns a set of subcarriers to each user and determines the power and number of bits per OFDM symbol to be transmitted on each subcarrier. The cyclic prefix (guard interval) is long enough to remove all intersymbol interference caused by multipath propagation. In addition, sharing the same subcarrier by different users is not allowed. The base station's resource allocation decision is sent to all users through a separate control channel. At each terminal, user can demodulate the signals over those subcarriers assigned to it. Denoting by c_{ij} the bit rate of the i th subcarrier which serves the j th user, we can express the maximum achievable rate (capacity) c_{ij} using transmitted power p_{ij} as

$$c_{ij} = W_i \log_2 \left(1 + \frac{p_{ij} |h_{ij}|^2}{\sigma_{ij}^2} \right), \quad 1 \leq i \leq N, \quad 1 \leq j \leq d \quad (1)$$

where W_i is the bandwidth for channel i , $|h_{ij}|^2$ and σ_{ij}^2 denote the channel gain and noise power of the i th channel which serves the j th user. The normalized capacity (rate) r_{ij} of the i th channel when used for serving the j th user is given by

$$r_{ij} = \frac{c_{ij}}{W_i} = \log_2 \left(1 + \frac{|h_{ij}|^2 p_{ij}}{\sigma_{ij}^2} \right) = \log_2 (1 + a_{ij} p_{ij}), \quad (2)$$

where $a_{ij} = |h_{ij}|^2 / \sigma_{ij}^2$ is the corresponding channel gain-to-noise ratio (GNR).

B. Problem formulation

Given the multi-user transmission requirements and channel state information (i.e., a_{ij} 's), one would like to find the subcarrier assignment and power allocation that minimize the total transmitted power. We define the $N \times d$ subcarrier assignment matrix $\mathbf{A} = [A_{ij}]$ by $A_{ij} = 1$ if the i th subcarrier is used to transmit the j th user; otherwise, $A_{ij} = 0$. As a subcarrier can only serve one user at a given time interval, A_{ij}

is either 1 or 0 and a legitimate channel assignment matrix $\mathbf{A}$ must satisfy

$$\sum_{j=1}^d A_{ij} \leq 1, \quad \sum_{i=1}^N A_{ij} \geq 1, \quad 1 \leq i \leq N, \quad 1 \leq j \leq d \quad (3)$$

For the downlink case, all signals are transmitted from the same base station, hence only the total transmitter power will be considered. Let $\mathbf{P}$ be the power allocation matrix with (i, j) th entry, p_{ij} , then the problem of concern becomes

$$\min_{\mathbf{P}, \mathbf{A}} \sum_{i=1}^N \sum_{j=1}^d A_{ij} p_{ij} \quad \text{s.t.} \quad \sum_{i \in C(j)} r_{ij} \geq R_j, \quad \sum_{j=1}^d A_{ij} \leq 1$$

where $C(j) = \{i | A_{ij} = 1, 1 \leq i \leq N\}$ (4)

Although in reality there is a total power constraint $\sum_{i=1}^N \sum_{j=1}^d p_{ij} \leq P_c$, we shall not consider this constraint to begin with. Solving the problem with the total power constraint follows a two-step procedure. In the first step we solve the unconstrained problem to obtain the required optimal total power and then check if the solution meets the total power constraint. The problem is solved if the constraint is satisfied; otherwise the problem does not have an admissible solution and one is forced to go to the second step. In the second step, one can prioritize users' transmission requests, modify (decrease) some rate requirements according to the corresponding latency requirements, or settle with a suboptimal channel/power allocation to accommodate the total power constraint. Which of these options is chosen depends on other system design considerations and the final solution is likely to be obtained by an *outer* iterative process. As far as this paper is concerned, however, the total transmit power constraint will not be discussed henceforth.

In the next section, we adopt a DP approach to derive a simple and practical solution which requires much lower complexity than that of [4] and, more importantly, offers near-optimal performance.

III. DYNAMIC PROGRAMMING BASED NEAR-OPTIMAL RESOURCE ALLOCATION

When $d = 1$ the optimal solution to (4) can be obtained by a water-filling process (for parallel Gaussian channels). The water-filling level, however, is difficult to determine. We present a very efficient algorithm called OMPA in Appendix A. Hence if the channel assignment is known, one can determine each user's optimal power allocation by using the proposed OMPA algorithm.

For the general case ($d \neq 1$), an obvious optimal solution to (4) is the exhaustive search over all possible channel assignments with the associated power allocation matrices computed by the OMPA algorithm (or water-filling method) to satisfy all users' rate requirement. Although this algorithm is guaranteed to yield the optimal solution, the searching process is prohibitively complicated, especially if the numbers of users and/or subcarriers are large. An improvement is suggested in [4] which first determines the "water-filling" levels and the channels for each user. Overbooking of channels is inevitable as every one wants the best channels. A complicated process is thus needed to resolve such conflicts and recompute the

“water-filling” levels iteratively. Although the optimal solution can be found, the complexity is still very high and is practical for small N and d only (e.g., the case $N = 8, d = 2$ was given in [4]). [1] considered a discrete-rate scenario but relaxed the discrete constraint to find a lower-bound solution of (4) iteratively. The quantized version of this solution gives a suboptimal subcarrier allocation, $\{C(j) : 1 \leq j \leq d\}$, where $C(j)$ is the j th user’s serving-channel set (SCS) that consists of the indices of the assigned channels. A single-user rate (bit) allocation algorithm is then applied to each $C(j)$. Numerical behavior of this approach was shown but no comparison with the optimal performance was given.

Other earlier suboptimal proposals [5], [6], [8] for solving (4) start with some initial subcarrier (channel) allocation and assign remaining available subcarriers sequentially according to some ad hoc criterion. Since a given channel has different GNRs when serving different users, [8] gives a channel to the user with strongest gain, i.e., the i th subcarrier is assigned to the k th user if $k = \arg \max_{1 \leq j \leq d} a_{ij}$. However, the ordering of the subcarriers or the user is arbitrary and it is highly likely that the best channels for two users are the same, say channel k , but the second best channel for the first user is much better than that for the second user. When the first user obtains channel k the second user can only use its second best channel which is much worse than channel k . If instead, the first user is given its second best channel which is not much worse than channel k while the second user is assigned channel k then the overall performance (required total power) will be much improved. On the other hand, [5] makes an initial SCS size $|C(j)|$ decision based on users’ average channel GNRs and rate requirements R_j ’s. The average GNR ignores frequency selectivity and the resulting algorithm is unlikely to find the optimal solution.

In contrast, our approach begins with the fair initial condition that all users are given the opportunity to take every subcarrier. The proposed channel allocation process consists of a series (N -level) of deletion decisions. At each level, a subcarrier is given to an user and is simultaneously removed from the SCSs of all other users, where the SCS for the j th user at the t th level, $C_t^s(j)$, is the set of all subcarriers allocated to serve user j then. Obviously, our fair initial condition implies that $C_0^s(j) = \{1, 2, \dots, N\}$, $\forall j$. We initially eliminate the constraint $C_t^s(i) \cap C_t^s(j) = \emptyset, \forall i \neq j, t = 0, 1, \dots, N$ and, at stage t , impose the constraint that $t \in C_t^s(j)$ for only one j (i.e., the t th channel can only be in one of SCS’s) so that the original single-user-per-subcarrier (SUPS) constraint is eventually re-installed and satisfied. Hence, in a sense what we adopt is a constraint relaxation approach.

In such a sequential assignment process the order of subcarriers may be important as once a subcarrier is assigned, no re-assignment is possible. A reasonable ordering is to sort (re-arrange) the N subcarriers in descending order of their maximum GNR, $a_i^* = \max_{1 \leq j \leq d} a_{ij}$ such that with the new channel order, channel 1 has the best GNR, followed by channel 2, channel 3, $\dots$, etc. Formally, this channel sorting is the permutation μ on the ordered integer set $\{1, 2, \dots, N\}$ which satisfies the inequality $a_{\mu^{-1}(1)}^* > a_{\mu^{-1}(2)}^* > \dots > a_{\mu^{-1}(N)}^*$, where μ^{-1} is the inverse mapping of μ .

Our DP-based algorithm can be described by a d -ary tree in

which there are d outgoing branches at the root (initial level) to represent possible assignment of the channel 1. Similarly, every node at any given level (height), say the t th level, has d outgoing branches (to d child nodes), each represents a possible channel-assignment (removal) decision and a tentative channel allocation. The channel allocation is tentative because only t channels are assigned and the remaining $N - t$ channels still belong to all SCSs and unassigned. If we associated each level’s decision with a cost, then at the k th level, we shall assign channel k to user i and remove it from the SCSs of all other users (branches) if the associated cost is minimized. Such a decision is equivalent to selecting the i th branch emitted from the surviving node at the $(k - 1)$ th level as the survival branch while all other $d - 1$ branches are terminated.

Given the initial fair channel allocation and the ultimate object of minimizing the required power, the cost for a decision at any level should be the minimum required power for the corresponding tentative channel allocation. Hence if we define the SCS collection at the t th level as $\mathbf{C}_t^s = (C_t^s(1), \dots, C_t^s(d))$, then the corresponding cost function J_t is

$$J_t(\mathbf{C}_t^s) = \sum_j g(R_j; C_t^s(j)) \quad (5)$$

in which each $g(R_j; C_t^s(j))$ is determined by applying the OMPA algorithm to solve the problem

$$\begin{aligned} \text{Given } C_t^s(j), \text{ find } g(R_j; C_t^s(j)) = \min \sum_{i \in C_t^s(j)} p_{ij} \\ \text{s.t. } \sum_{i \in C_t^s(j)} r_{ij} \geq R_j. \end{aligned} \quad (6)$$

$C_t^s(j)$ for each j is modified at each level so that the subcarrier and power assignment process is guaranteed to end at the N th level. As the minimum required power $g(R_j; C_t^s(j))$ for each j is a decreasing function of the cardinality $|C_t^s(j)|$ of its SCS, the cost J_t is an increasing function of t . At each level, however, we find the removal of the subcarrier from all but one $C_t^s(j)$ that results in minimum cost (total power) increase. As the collection $\{C_t^s(j)\} = \mathbf{C}_t^s$ allows multiple channel assignments, i.e., $C_t^s(j) \cap C_t^s(k) \neq \emptyset$, if $j \neq k$ and $t < N$, it does not satisfy the constraints (3) of a legitimate channel assignment matrix. But as the subcarriers are assigned to users one by one, at the end of the N th level, $\{C_N^s(j)\} = \mathbf{C}_N^s$ will correspond to a legitimate one. Therefore, the metric defined by (5)-(6) is simply the minimum total transmit power for a given rate-subcarrier assignment with various degrees of relaxation on the SUPS constraint.

Since a path in the tree that visits the j th child node at the k th level implies a channel assignment that gives the k th subcarrier to the j th user, an N -level path would represent a complete channel allocation. But not all paths are legitimate for a path may assign no serving-channel to an user. In particular, if at the end of the t th level there are still more than $N - t$ users without any serving-channel, i.e., whose SCS cardinality is equal to $N - t$, then there will be at least one user with an empty SCS at the end of the N th level. To avoid such a possibility and rule out all illegitimate channel assignments, we modify the cost function as

$$J_t(\mathbf{C}_t^s) = \min_{1 \leq k \leq d} \left\{ \sum_{j=1}^d g(R_j, C_t^s(j; k)) + \omega_t \left[\sum_{j=1}^d \delta(N - t - |C_t^s(j; k)|) \right] \right\} \stackrel{\text{def}}{=} \min_{1 \leq k \leq d} J_t^k(\mathbf{C}_t^s) \quad (7)$$

where

$$C_t^s(j; k) = \begin{cases} C_{t-1}^s(j) & , \quad j = k \\ C_{t-1}^s(j) \setminus \{t\} & , \quad j \neq k \end{cases} \quad (8)$$

$$\delta(x) = \begin{cases} 1 & , \quad x = 0 \\ 0 & , \quad \text{otherwise} \end{cases}$$

and

$$\omega_t(x) = \begin{cases} 0 & , \quad x \leq N - t \\ \infty & , \quad x > N - t \end{cases} \quad (9)$$

By adding the weight function $w_t(\cdot)$ in the cost function, we avoid continuously assigning channels to some users while other users might not be able to obtain any channel, although the probability of such an event is almost zero so long as $N > d$ and the GNR distributions $\{a_{ij}, i = 1, 2, \dots, N\}$ for each user are independent.

The resulting DP-based resource allocation (DPRA) algorithm, unlike other approaches [1][8][6], accomplishes channel and power (rate) allocations simultaneously and is listed in Table I. Early terminations and computational complexity reduction are possible if certain conditions are satisfied; see *Guidelines 4, 5* in the next section.

Table I: A dynamic programming based resource allocation (DPRA) algorithm

Step 1:	(Channel-sorting) Given N, d, a_{ij} and $\mathbf{R}$, find $a_i^* = \max_{1 \leq j \leq d} a_{ij}$ and re-arrange the channel indexes by decreasing magnitude of the maximum GNR such that $a_1^* > a_2^* > \dots > a_N^*$ with the new channel indexes.
Step 2:	(Initial channel allocation) Set $C_0^s(j) = \{i \mid 1 \leq i \leq N\}$, for $1 \leq j \leq d$.
Step 3:	(Sequential channel-power-rate assignment) for $t = 1 : N$ $k^* = \arg \min_{1 \leq k \leq d} J_t^k(\mathbf{C}_t^s)$ $J_t(\mathbf{C}_t^s) = J_t^{k^*}(\mathbf{C}_t^s)$ for $j = 1 : d$ if $j = k^*$ then $C_t^s(j) = C_{t-1}^s(j)$ else $C_t^s(j) = C_{t-1}^s(j) \setminus \{t\}$ end end
Step 4:	(Output) The final channel allocation is the N th level SCS collection $\mathbf{C}_N^s$. The power-rate allocation is obtained while computing $J_N(\mathbf{C}_N^s)$ through (5)-(7).

IV. AN OPTIMAL RESOURCE ALLOCATION ALGORITHM

The N -level tree shown in Fig. 1 is a graphic representation of the solution space of (4). The tree contains all possible—legitimate or illegitimate—channel assignments. At each level we allocate a channel so that each “complete” path L from the root node to a leaf node represents a candidate assignment and can be denoted by $L = (b_1, b_2, \dots, b_N)$, where b_i is the i th node visited by the path and, for brevity, the initial (root) node is not included in the notation. A partial path $l_n = (b_1, b_2, \dots, b_n), n < N$ is thus defined as the part of a complete path that starts at the root node and ends at some internal node.

Searching over the complete tree can certainly lead to the optimal solution but the complexity is of exponential order. The DPRA algorithm calls for the elimination of $d - 1$ child nodes at each level and promises to finish the tree-search process in N stages. As will be shown in Section V, this approach is very efficient in that it yields near-optimal solution with low complexity. However, there is no guarantee that the optimal solution will be obtained as it is possible that the optimal channel assignment path is discarded somewhere along the way, especially if SNR is low. Many other techniques can be used to reduce the prohibitive high complexity of searching the total solution space. We employ a simple linear programming technique called branch-and-bound (B&B) which has the potential of significant complexity reduction if the bounds are properly chosen. Besides presenting novel tight bounds, we also suggest a subcarrier-sorting procedure, which is crucial in reducing the search complexity, use a good initial upper bound and derive some useful properties and guidelines for further complexity reductions.

A. A branch-and-bound approach

In the search tree shown in Fig. 1, each parent node has d child nodes to enclose all possible solutions. Similar to our description of the DPRA algorithm, a path in Fig. 1 that passes through the j th child node at the k th level of the tree (i.e., $b_k = j$) represents a channel assignment that gives the k th subcarrier to the j th user and a (complete) path is legitimate only if it visits every candidate child node at least once. The B&B paradigm needs an upper bound B_u on any $v(l_N)$ (i.e., legitimate channel assignment or complete path) and a lower bound $B_l(l_t)$ associated with each partial path l_t of length t . The use of the upper bound for the cost (minimum required total power) combined with the lower bound which represents the current best solution value (associated with a partial path) enables the algorithm to prune the non-promising subtrees rooted at certain nodes and search parts of the complete tree only. These bounds should be updated as soon as possible to accelerate the searching process but the initial upper bound often plays an important role in the reducing the search complexity. A weak bound will not be capable of eliminating many visits to nodes that lie outside of the correct (optimal) path. To find a tight lower bound, we need the following fundamental definition.

Definition 1: The node value (cost) $v(l_t)$ of an internal node of the search tree is defined by (7) with each $C_t^s(j)$ obtained by removing from $C_0^s(j)$ the channels that have been

assigned to other users along the partial path l_t that ends at the current node.

Obviously, the node value so defined is a function of the node and the associated partial path. We thus denote the node value by $v(l_t)$ to emphasize such a dependence. To see that the node value is indeed a lower bound, we first notice that, like the cost function of the DPRA algorithm, it is a function of a channel allocation that is illegitimate and optimistic. The channel allocation is illegitimate because a subcarrier may be assigned to more than one user and it is optimistic since an user tends to own more than its share of subcarriers, resulting in reduced required power. In the search tree shown in Fig. 1, each parent node has d child nodes to enclose all possible solutions. When we search along a path to visit an internal node of the tree, we compute the associated "node value" by (7) with each $C_t^s(j)$ obtained by removing from $C_0^s(j)$ the channels that have been assigned to other users along the partial path l_t from the root node to the current node. Obviously, the node value so defined is a function of the node and the associated partial path. To emphasize such a dependence, we denote the node value by $v(l_t)$. As an user's SCS is a decreasing function of the partial path length in the sense that a child node's SCS is a subset of their parent node's, the node value of a child node must be equal to or greater than that of its parent node. In other words, the fact

$$C_0^s(j) = C(j) \supset C_1^s(j) \supset C_2^s(j) \supset \cdots \supset C_N^s(j), \forall j \quad (10)$$

implies

Property 1: Both $g(R_j, C_t^s(j))$ and the cost function $J_t(C_t^s)$ defined by (7) are increasing functions of t .

As every complete path is associated with a sequence of shrinking SCSs $\{C_0^s(j), C_1^s(j), \dots, C_N^s(j)\}$ and $C_N^s(j)$ is the cost of this path, we have

Property 2: The node value $v(l_t)$ defined by (7) is a lower bound for the cost of any complete path that coincides with the t -level partial path l_t .

Thus, if J_t at a parent node is not smaller than the upper bound, we are sure that there is no optimal solution in its child nodes and one should check other nodes of the same level. On the other hand, the order of visiting the d child nodes of a parent node should be based on their node values as the node value represent our current best estimate of all subsequent assignments. For the convenience of subsequent reference, we summarize these two observations, which often brings about significant search complexity reduction (see Table III) of a B&B-based resource allocation (BBRA) algorithm, as

Guideline 1: The order of visiting d child nodes of a given parent node should be the same as the ascending order of the magnitudes of the corresponding node values. In other words, one should visit the node with the least node value, followed by the one with second smallest node value, and so on.

Guideline 2: When visiting a node (say at the t th level) of a partial path l_t , we compute the node value $v(l_t)$ and compare it with the current upper bound B_u . If $v(l_t) < B_u$ then visit its first child node in the next level. Otherwise, searching on the subtree rooted at this node is terminated and the search should continue on the next unvisited child node of the same level or backtrack to the next unexplored nodes in the previous

level, where the order of d siblings descending from the same parent node is determined by *Guideline 1*.

Because only a complete path corresponds to a candidate solution, the depth-first-search (DFS) strategy is suitable for our B&B approach. The initial upper bound B_u^0 can be obtained by the DPRA algorithm. The ensuing DFS searching procedure tries to continuously separate the parent space into the subproblem (child) space. Therefore, we have

Guideline 3: Upon arriving at the final level, we check the resulting cost (node value) to see if B_u has to be updated. We then backtrack to the nearest parent node determined by *Guideline 1* and resume the searching process.

Note that the above three *Guidelines* are valid for general B&B approaches and are listed for the convenience of subsequent discussions.

Definition 1 and the above guidelines all assume that we compute the node values when transversing along a path based on the same principle used by the DPRA algorithm. In other words, every user is given all channels initially and, at each level along a path, a channel is assigned to the user associated with the selected child node and removed from all other users' SCSs. Such a procedure will not exclude any legitimate solution from the tree search. With this assumption, we note that the node values along a path may reach a steady state before the leaf node is visited. A necessary condition is

Property 3: Further traversing on a path will not change the node value if the set of remaining unassigned channels $\bigcap_j C_t^s(j) = C_U$ satisfy either (i) $\forall i \in C_U \Rightarrow r_{ij} = 0, \forall j$, or (ii) $\forall i \in C_U, i \in C_t^s(j)$ for only one j .

This property can be used to accelerate our search without missing the optimal solution.

Guideline 4: Besides those terminations specified by *Guideline 2*, early termination (of a path) is possible if one of the conditions in *Property 3* is satisfied.

Since computing the node value via (7) requires repeated calls to the OMPA subroutine, the search complexity can be reduced if we can minimize the numbers of calls. A careful examination of (7) and the search procedure reveals

Guideline 5: In computing the node value for the k th child node of a $(t-1)$ th level parent node, the fact $C_t^s(k) = C_{t-1}^s(k)$ implies that $g(R_k, C_t^s(k; k)) = g(R_k, C_{t-1}^s(k))$. Furthermore, if in computing the parent node's value we have $r_{tj} = 0$ for some j , then $g(R_j, C_t^s(j)) = g(R_j, C_{t-1}^s(j))$. For both cases there is no need to call the OMPA subroutine to compute the minimum required power. Finally, although for a fixed k , d OMPA calls are needed in computing each cost $g(R_j, C_t^s(j; k))$, $d-1$ of them can be reused for other k 's.

The last two guidelines can be used to reduce the computing complexity of the DPRA algorithm as well. In particular, *Guideline 5* implies that only d OMPA calls are needed to compute d child node values of a given parent node.

B. Sorting the serving channels

We have suggested a channel ordering for the DPRA algorithm according to the maximum GNR's. This channel indexing is simple but, according to our simulation, does not yield fast convergence. Like the DPRA algorithm, the order of the channels is very important. If our ordering (indexing)

of the channels is such that the i th ($i < N$) channel is so "bad" that it is not used in the final optimal solution (no user really wants it) then we have to check all its d child nodes in the next level. Simulations indicate that the channel ordering affects the search speed significantly. In view of *Guideline 1*, *Property 2* and given we have decided the first k channels, the $(k+1)$ th channel should be the most demanded one such that its assignment to a user (thus is removed from the SCSs of all other users) increases the costs (node values) of all other users most significantly. The channel-sorting algorithm based on this idea, is presented in Table IV.

We have several remarks on the above channel-sorting process.

- R1.** The sole purpose of this algorithm is channel-sorting and the corresponding channel assignments are auxiliary operations, not to be realized.
- R2.** Step 2 in Table IV defines the most demanding channel as the one that offers the highest sum rate and is requested by two or more users given the current SCS collection. When a channel offers the highest rate but serves only one user, it must render relatively low GNR for all other users, hence the decision of its order in the tree should be postponed.
- R3.** Step 4 deals with the ordering of those channels which, after several rounds of filtering the most demanded channels, are still requested by one user only.

With this channel-sorting procedure and in view of the properties and guidelines mentioned before along with our definition of the node value, we propose the BBRA algorithm of Table II.

Table II: A branch and bound based resource allocation (BBRA) algorithm

Step 1:	(Initialization) Use the DPRA algorithm to obtain the initial upper bound B_u^0 and the channel-sorting process in Table IV, to rearrange the channel order. Set the initial level at $\aleph = 1$
Step 2:	Visit the child nodes of the $\aleph$ th level according to <i>Guideline 1</i> and invoke <i>Guideline 2</i> . Set $\aleph \leftarrow \aleph + 1$ if no backtracking is needed; otherwise set $\aleph \leftarrow \aleph - 1$. <i>Property 3</i> should be used at every node visited to check the possibility of early termination of a candidate path.
Step 3:	Go to Step 2 if $\aleph < N$. If $\aleph = N$ then terminate the searching process if all nodes have been visited or been excluded from further consideration; otherwise invoke <i>Guideline 3</i> . Set $\aleph \leftarrow \aleph - 1$ and go to Step 2.

C. Complexity reduction techniques

To explore the effectiveness of various techniques implied by properties and guidelines on reducing the computing complexity, we have performed 10^6 simulated runs of the BBRA algorithm that incorporates (1) the channel-sorting process in Table IV and the combinations of (2) *Guideline 1*, (3) *Property 3* and (4) the fifth Guideline. The numbers of users

and channels are 5 and 128, respectively, and the normalized rate for each user is uniformly distributed in $[0, 3]$. The results are summarized in Table III with the complexity measured in terms of the number of calls to the OMPA algorithm.

Channel-sorting is most critical for with other conventional channel-sorting methods (e.g., that used by DPRA), the searching complexity often becomes greater than 10^6 . Hence it is always assumed as part of the initialization step in the BBRA algorithm. The reuse of existing OMPA results (i.e., *Guideline 5*) also brings about significant reduction as it is applicable in every node visit. Proper branching and early terminations help accelerating the search process a lot as well.

D. Application to integer constellation systems

With minor modifications, our algorithms remain valid and are applicable for solving a similar RA problem with integer constellation (discrete-rate) constraints. All we to have to do is inserting an SNR gap, which depends on the constellation size and the BER requirement, in the rate-power equation (2) and replacing the OMPA (water-filling procedure) algorithm by a known bit-loading algorithm, e.g., Campello's optimal algorithm whose complexity is upper-bounded by $O(N)$ [3].

A B&B approach was also suggested in [13] to solve a similar problem for integer constellation systems. Besides not having the attributes mentioned in the second paragraph of this section, their method differs from ours in at least two major aspects. First, their approach implies a tree structure that grows a $(dM+1)$ -ary sub-tree out of each node where M is the number of discrete rates allowed while we need only a d -ary sub-tree. In other words, [13] converts both user and rate selections into node selections, each node represents a fixed user/rate assignment for a given subcarrier but our tree search has to do with user selection only. Second, each node value (lower bound) of [13] is obtained by solving a linear programming problem after relaxing three major constraints, namely, (i) the SUPS, (ii) the single-rate-per-subcarrier, and (iii) the discrete-rate constraints. The first two relaxations are directly related to their tree structure and the last relaxation is needed to convert the integer linear programming problem into a (real) linear one which is much easier to solve. As a result, the lower bound so obtained is not very tight. In contrast, we use either the OMPA algorithm or Campello's algorithm [3] to perform the corresponding (provisional) optimal rate/power allocation once an user is selected (for using a subcarrier). The corresponding bounds do not have to remove constraints (ii) and (iii) mentioned above whence are much tighter and result in far less search complexity.

V. NUMERICAL RESULTS AND ALGORITHMIC COMPLEXITY

We report some simulated performance and complexities of the proposed algorithms and two suboptimal algorithms modified from existing ones in this section. As the performance of two proposed algorithms is almost identical, that of the BBRA algorithm is not shown and is used as the reference for comparison only. 10^5 runs, each with a different channel realization, are performed to obtain the numerical results presented here.

Table III: The effects of (1) channel-sorting in Table IV, (2) Guideline 1, (3) Property 3/Guideline 4, and (4) Guideline 5 on the computing complexity reduction of the BBRA algorithm; 10^6 runs are performed to obtain the statistics. The complexity is measured in terms of numbers of calls n_{op} to the OMPA algorithm. The complexity of the DPRA algorithm is also included for comparison purpose.

$d = 5, N = 128$	DPRA	(1)	(1)+(2)	(1)+(3)	(1)+(4)	(1)+(2)+(3)+(4)
$E[n_{op}]$	44.61	2587.2	773.98	578.09	116.19	88.32
$E[n_{op} n_{op} < 2 \times 10^5]$	44.61	1717.8	773.98	549.91	93.78	88.32
$\text{Prob}[n_{op} > 2 \times 10^5]$	0	0.0012	0	0.00005	0.00004	0
$\max\{n_{op}\}$	81	21622894	180132	1800602	981053	587

A. Relative efficiency performance

1) *Performance of BBRA and DPRA algorithms:* As only the GNRs a_i affect the performance we assume, without loss of generality, that $\sigma_{ij} = \sigma, \forall i, j$. We normalize the bandwidth of each sub-carrier (channel) such that $W = 1$ and set the normalized noise power lever σ^2 to be 1. We also normalize the Rayleigh-distributed channel gains $|h_{ij}|^2$ such that $E[|h_{ij}|^2] = 1$ and assume that channels are independently faded. These two normalization assumptions effectively imply $E[a_{ij}] = 0$ dB. Since the channel capacity or the normalized rate is a function of the product $p_{ij}a_{ij}$, the simulation results shown in Figs. 3–6 are scalable in the sense that a higher (lower) $E[a_{ij}]$ needs a proportionally lower (higher) minimum required power. The normalization of the channel bandwidth W has a similar purpose in interpreting the normalized data rates R_i which now have the unit of bits/sec/Hz. We also have the normalized sum rate as $\sum_{j=1}^d R_j$. Various normalized rate distributions with the same sum rate are examined.

Let J_{DP} and J_{BB} be the total required transmit power determined by the DPRA and BBRA algorithms, respectively, and define the relative efficiency (RE) of the former algorithm by

$$\eta = 1 - \frac{E[J_{DP}] - E[J_{BB}]}{E[J_{BB}]} \quad (11)$$

Since the power-rate allocation is solely determined by OMPA once the subcarrier assignment is fixed, we say two algorithms give the same solution if both suggest the same subcarrier allocation. Fig. 2 plots the probability that the DPRA algorithm converges to the optimal solution for several cases ($N = 64, 128$ $d = 5, 10, 15$). Since the BBRA algorithm is guaranteed to give the optimal solution, this probability is equal to $\text{Pr}[J_{DP} = J_{BB}]$. It is found that when $d \ll N$ (say $d/N < 0.1$) the probability that the DPRA algorithm yields the optimal solution is greater than 0.9 if the sum rate is less than 8 bits/sec/Hz. Although for other cases under investigation, this probability is smaller than 0.9, Fig. 3, which plots the RE of the DPRA algorithm, indicates that the corresponding solutions still lie very close to the optimal one. It is clear that the DPRA algorithm is capable of offering a near-optimal solution that even in the worst case (64 channels, 15 users and a normalized required sum rate of 20) it achieves a RE as high as 99.82%.

In general, the larger the number of the users d or the sum rate is, the less efficient the DPRA algorithm becomes. Such a behavior is consistent with the fact that, when d increases but N is fixed, a correct channel selection at each level become less likely so is the probability of obtaining the optimal channel allocation. On the other hand, the assumption of independent fading of channels implies that the probability of having “good” channels increases as N increases, and the

probability of correct or good decision at each level increases as well. Hence, for a fixed d , the probability of obtaining the optimal or near optimal solution is an increasing function of N and so is the RE (η).

Table IV: The channel-sorting algorithm

Step 1:	(Initialization) Set $t = 1$ and let the SCS and the assigned channel set (ACS) for user j be $C_t^s(j) = \{1, 2, \dots, N\}$ and $C_t^a(j) = \emptyset$, respectively.
Step 2:	(Find the most demanded channel) Compute the sum rate $R_i^s \stackrel{\text{def}}{=} \sum_{j=1}^d r_{ij}$ for each channel, where r_{ij} is obtained by applying the OMPA algorithm for each SCS $C_t^s(j)$, compute $\Psi(i) = \{j \mid 1 \leq j \leq d, r_{ij} \neq 0\}$, $C_A = \bigcup_{j=1}^d C_t^a(j)$, $C_h = \{i \mid 1 \leq i \leq N, \Psi(i) \geq 2\}$. If $ C_h \neq \emptyset$ and $\ell = \max_{i \in C_h} R_i^s$, then the ℓ th channel is re-indexed as channel t (i.e., $\mu(\ell) = t$) and assign this channel to user k if $k = \max_j r_{\ell j}$. Go to Step 4 if $ C_h = \emptyset$.
Step 3:	(Updating) The ACS for user k and the SCSs are updated by $C_t^a(k) \leftarrow C_t^a(k) \cup \{\ell\}$, $C_t^s(j) \leftarrow C_t^s(j) \setminus \{\ell\}$, $\forall j \neq k$, respectively. The tree level index is updated by $t \leftarrow t + 1$. If $t < N$, go to Step 2; otherwise, the sorting process is completed.
Step 4:	(Sorting the less demanded channels) If $C_t^s(j) \cap C_t^a(j) = \emptyset, \forall j$, go to Step 5; otherwise, for all j , $C_t^s(j) \cap C_t^a(j) \neq \emptyset$, modify the corresponding SCS by $C_t^s(j) \leftarrow C_t^s(j) \setminus \{j_m\}$, where $j_m = \arg \max_{i \in C_t^s(j) \cap C_t^a(j)} r_{ij}$, and go to Step 2.
Step 5:	(Sorting the remaining channels) The order (numbering) of the channels in the set $\{i \mid 1 \leq i \leq N, i \notin C_A\}$ is determined by the maximum GNR criterion used in the DPRA algorithm.

2) Performance of representative sub-optimal algorithms:

Although many RA schemes have been proposed, they assume different scenarios and costs. Those dealing with RA problems similar to (4) often follow a three-step procedure [5],[6]. (S1) *Resource allocation*—determine the resource (number of channels) to be given to each user based on some criterion. (S2) *Subcarrier assignment*—decide which subcarrier should serve which user. (S3) *Local optimization*—each user computes the optimal power allocation according to its channel set and rate requirement. The average GNR criterion, i.e., the BABS (bandwidth assignment based on GNR) algorithm [5], is perhaps the simplest and most popular choice for use in (S1). Such an approach treats the channel of concern as a

Fig. 1. A complete search tree for multiuser channel allocation. For the DPRA algorithm, only one child node survives at each level.

flat-faded wideband channel when determining the number of subcarriers an user is entitled to possess. It simplifies the resource allocation procedure by ignoring the selectivity of a wideband channel but is very likely to exclude the optimal solution from further consideration. Many methods were proposed to obtain the water-filling solution for (S3) with various degrees of precision. The main difference lies in (S2).

The first approach called the amplitude craving greedy (ACG) algorithm [5] sequentially assigns the subcarriers to the user with the largest GNR unless the channel number quota determined in (S1) has been exceeded. An alternate method called rate craving greedy (RCG) algorithm [5] finds the water-filling rate level for all users, assuming they have been given all subcarriers. The subcarriers are then assigned to the one with the highest achievable rate unless its channel number quota is exceeded. The original ACG and RCG algorithms can not be used to solve (4) as they are designed for discrete-rate constraints. Moreover, they use an approximate instead of exact water-filling solution. For the purpose of fair comparison, we modify both algorithms by using the rate-power equation (2) and the OMPA solution. The resulting algorithms are henceforth referred to as the modified ACG (MACG) and RCG (MRCG) algorithms, respectively.

We assume $\text{GNR} = 20$ dB for all subcarriers and $R_j = 5$ bits/sec/Hz for all j . Besides the independent-fading channel

Fig. 2. The DPRA algorithm's probability of correct convergence (i.e., the probability of obtaining the optimum subcarrier/power/rate allocation) in an OFDMA downlink.

Fig. 3. Average relative efficiency (η) performance of the DPRA algorithm.

model, for $N = 128$ we also consider the ITU Vehicular A model [11] which has been adopted by UMTS and WiMax forum as one of the reference channel models. The RE performance shown in Fig. 4 indicates that our DPRA algorithm does outperform both MACG and MRCG algorithms. It yields a near-optimal solution that even in the worst case ($N = 64, d = 14$), achieves a 99.92% RE while the the two modified algorithms give 91.74% and 92.92% efficiencies. Due to the fixed rate requirement, a larger N results in better performance for all suboptimal approaches. The increase of d , on the other hand, leads to reduced efficiency but DPRA is much more robust in the sense of maintaining almost constance RE for different d , N and channel conditions. The optimal allocation probabilities for these two suboptimal algorithms are not presented for they are simply too small.

Fig. 4. Average relative efficiency of the DPRA, MRCG, and MACG algorithms; $R_j = 5$ bps/Hz for all j , IF = independent fading, CF = correlated fading.

B. Complexity evaluation

The computing complexity of various RA methods is dominated the number of calls to the single-user (mono-rate) water-filling algorithm whether it is the OMPA or any other algorithm that computes the power or rate associated with each channel assigned to a user. An exhaustive search requires $O(Nd^N)$ single-tone power- or rate-level computing operations [4] or $O(d \cdot d^N)$ calls of OMPA.

1) *Average complexities of the proposed algorithms:* For the DPRA algorithm, at most $d^2 \times N$ calls of OMPA is needed. But the complexity of BBRA method is difficult to analyze directly for it depends on the channel order and the initial upper bound value. By using computer simulation, we estimate the average complexities, measured in terms of number of calls of the OMPA algorithm, of the DPRA and BBRA algorithms and present the results in Figs. 5-6. We assume that the normalized GNR is 0 dB and examine the required complexity for different numbers of users with the same normalized sum rate. A few observations on the last two figures can be made. First, because of *Guidelines 4* and *5*, the complexity of DPRA algorithm is limited to at most $dN + 2d$ OMPA calls. Second, although the complexities of both DPRA and BBRA algorithms increase with the number of users d , the latter is much more sensitive to this parameter. Finally, the average complexity of the BBRA algorithm is higher when there are 64 channels than if there are 128 channels. The reason for this interesting fact is that there are more good channels when $N = 128$ and, for a fixed sum rate requirement, as good channels tend to support higher data rates, fewer channels are needed and early terminations due to *Guideline 2* and *Guideline 4* occur more often.

2) *Complexities of other sub-optimal algorithms:* Since the complexity of water-filling is a function of the channel number involved, a more precise and fair comparison is counting the number of rate(power)-evaluation iterations, i.e. the [MR2]-[MR4] loop of the OMPA algorithm; see Appendix A. For BBRA or DPRA algorithms, the iteration number in

Fig. 5. Average complexities (numbers of calls to the OMPA algorithm) for the BBRA and DPRA algorithms in a 64-subcarrier OFDMA system.

Fig. 6. Average complexities (numbers of calls to the OMPA algorithm) for the BBRA and DPRA algorithms in a 128-subcarrier OFDMA system.

every call is upper-bounded by $\log_2 N$ as bisection search is in place. The complexity of the DPRA algorithms can be reduced by using *Guidelines 4*, *5* and the iteration number is thus upper-bounded by $(dN + 2d) \log_2 N$.

Fig. 7 shows the average complexity performance of our algorithms and the MACG, MRCG algorithms for $N = 64$ and 128. The system and channel parameter values used here are the same as those used in Fig. 4. As expected, the performance in correlated fading is worse than that in independent fading and the computation complexity of all algorithms are far less than the DRRA upper-bound, $(dN + 2d) \log_2 N$. Moreover, BBRA requires the highest average complexity, followed by DPRA, MRCG and the MACG algorithms. The complexity of the DPRA algorithm is about twice that of the RCG based algorithm but is far less than that of the BBRA algorithm. The DPRA algorithm, as mentioned before, yields near-optimal performance and is robust against the variations of the numbers of users and subcarriers.

Fig. 7. Average number of power (or rate) level evaluation iterations for various dynamic RA algorithms; $R_j = 5$ bps/Hz for all j , IF = independent fading, CF = correlated fading.

VI. CONCLUSION

OFDMA is an effective multiple access scheme in a wideband wireless mobile network. Besides its anti-fading capability, an OFDMA system can achieve high spectral efficiency in a multiuser environment by adaptively allocating subcarriers and time slots to the most suitable users with the minimum required transmit power. An efficient dynamic RA algorithm to solve the corresponding constrained optimization problem in real time is thus crucial for realizing this potential advantage.

Based on the principles of dynamic programming and branch-and-bound, we propose two algorithms—the DPRA and BBRA algorithms—which give either near-optimal or optimal solution. In contrast to the existing algorithms, which suffer from the shortcomings of requiring high complexity and/or unsatisfactory performance, the DPRA algorithm renders near-optimal performance with relative low complexity. Since the existing efficient algorithms are designed with a discrete-rate constraint and use some suboptimal water-filling solution, we make some modifications for fair comparisons. As expected, the resulting ACG and RCG based DPRA algorithms are shown to provide less satisfactory performance with reduced complexities. With proper reuse of the water-filling solution obtained in earlier stages, the average DPRA complexity can be further reduced and is insensitive to d , N and the required sum rate. The average complexity of the BBRA algorithm, on the other hand, is at least an order higher than that of the DPRA algorithm when the number of users is greater than 10 but is still much less than the known algorithms for obtaining the optimal solution.

Our numerical experiment in both independent and correlated fading environments have demonstrated that the near-optimal DPRA algorithm is suitable for real-time resource allocation application and the optimal BBRA algorithm is practical only if $d \leq 5$. Nevertheless, the latter algorithm offers the optimal solution and performance for large N and d with reasonable complexity, which has never been achieved before

and is needed for benchmarking and comparison purposes. Finally, we would like to mention that, although we restrict our discourse to the capacity (rate) constraint, the proposed algorithms are applicable to other constraints such as fixed BER or weighted capacity constraint by modifying the corresponding rate-power function.

APPENDIX A

AN OPTIMAL MONO-RATE POWER ALLOCATION ALGORITHM

Let us redefine the normalized channel capacity r_i by

$$r_i = \log_2 \left(1 + \frac{|h_i|^2 p_i}{\sigma_i^2} \right) \equiv \log_2 (1 + p_i a_i) \quad (\text{A.1})$$

where the subscript i denotes the i th channel and $|h_i|^2$, p_i , σ_i^2 are the corresponding channel gain, transmitted power, and noise power, respectively. In addition, the N orthogonal channels are sorted according to descending channel gain-to-noise ratio, e.g., $a_1 > a_2 > \dots, a_N$, $a_i \equiv |h_i|^2 / \sigma_i^2$. Note that because of (A.1), power and rate allocations are equivalent provided that a_i is known.

For the mono-rate case, (4) becomes

$$\min_{\mathbf{p}} \sum_{i=1}^N p_i, \quad \text{s.t.} \quad \sum_{i=1}^N r_i \geq R, \quad p_i \geq 0, \quad (\text{A.2})$$

The water-filling solution implies that only the strong channels (those whose reciprocal channel gains are below the water-filling level) will be used. Hence we assume that only the strongest x channels are used so that the power and rate for the weakest $N - x$ channels are identically zero, i.e., $p_i(x) = r_i(x) = 0$, $x < i \leq N$, where $p_i(x)$, $r_i(x)$ denote the power and rate of the i th channel when only the first x channels are activated. The optimization problem (A.2) then become that determining the optimal x . Define the Lagrange dual function as

$$f(\{r_i(x)\}, \{p_i(x)\}, \lambda) = \left[\sum_{i=1}^N p_i(x) \right] - \lambda \left[\sum_{i=1}^N r_i(x) - R \right] \quad (\text{A.3})$$

and omit the constraints $0 \leq p_i$ for the moment. Taking derivative with respect to r_i for $i = 1, 2, \dots, x$ we obtain $\lambda = \frac{e^{R/x} \ln 2}{\hat{a}(x)}$, where $\hat{a}(x) = \left(\prod_{j=1}^x a_j \right)^{1/x}$ and

$$r_i(x) = \frac{R}{x} + \log_2 \left(\frac{a_i}{\hat{a}(x)} \right), \quad i = 1, 2, \dots, x \quad (\text{A.4})$$

Obviously, it is possible $r_i(x) < 0$ as the constraint $p_i \geq 0$ has been removed.

Note that

$$r_x(x) = \frac{x-1}{x} \left[r_{x-1}(x-1) + \log_2 \left(\frac{a_x}{a_{x-1}} \right) \right] \quad (\text{A.5})$$

Using the fact that, $a_1 \geq a_2 \geq \dots \geq a_N$, we conclude that

Lemma 1: The sequence $\{r_x(x), x = 1, 2, \dots, x\}$ is monotonically decreasing.

To find the constrained solution we need the following definition.

Definition 2: An unconstrained solution $\mathbf{r}(x, N) = (r_1(x), r_2(x), \dots, r_x(x), 0_{1 \times (N-x)})$ is said to be admissible if the least rate $r_x(x) > 0$. The admissible active channel

number sets for the problem defined by (A.2) is defined by $\mathbf{F} = \{x | r_x(x) > 0, 1 \leq x \leq N\}$, where $r_x(x)$ is given by (A.4).

Lemma 2: The total transmitted power associated with the admissible unconstrained optimal rate assignment (A.4) is a decreasing function of the number of channels used. In other words, $N_1 < N_2 \implies \sum_{i=1}^{N_1} p_i(N_1) > \sum_{i=1}^{N_2} p_i(N_2)$, for $N_1, N_2 \in \mathbf{F}$.

Proof: To begin with, let us assume that $N_1 = m$ and $N_2 = N_1 + 1 = m + 1$, i.e., $N_2 - N_1 = 1$. If the Lemma is valid in this case, it will also be valid when $N_2 - N_1 > 1$.

$$p_i(x) = \frac{e^{r_i(x)} - 1}{a_i}, \quad 1 \leq i \leq x, \quad (\text{A.6})$$

The minimum required power for the case $x = m$ is given by

$$\begin{aligned} \tilde{P}_m &= \sum_{i=1}^m \frac{e^{r_i(m)} - 1}{a_i} = \sum_{i=1}^m \left[\frac{e^{R/m}}{\hat{a}(m)} - \frac{1}{a_i} \right] \\ &= m \cdot \frac{e^{R/m}}{\hat{a}(m)} - \sum_{i=1}^m \frac{1}{a_i} \end{aligned} \quad (\text{A.7})$$

where $\hat{a}(m) = [\prod_{i=1}^m a_i]^{\frac{1}{m}}$. The minimum required power for the case $x = m + 1$ can be expressed as a function of r_{m+1} .

$$\begin{aligned} \tilde{P}_{m+1} &= \tilde{P}_m + p_{m+1} = \sum_{i=1}^m \frac{e^{r'_i(m)} - 1}{a_i} + p_{m+1} \\ &= \sum_{i=1}^m \left[\frac{e^{(R-r_{m+1}(m+1))/m}}{\hat{a}(m)} - \frac{1}{a_i} \right] + p_{m+1} \\ &= m \cdot \frac{e^{(R-r_{m+1}(m+1))/m}}{\hat{a}(m)} \\ &\quad - \sum_{i=1}^m \frac{1}{a_i} + \frac{e^{r_{m+1}(m+1)} - 1}{a_{m+1}} \end{aligned} \quad (\text{A.8})$$

Expressing the difference between $\tilde{P}_m$ and $\tilde{P}_{m+1}$ as a function of $r_{m+1}(m+1)$, we obtain

$$\begin{aligned} g(r_{m+1}(m+1)) &= \tilde{P}_m - \tilde{P}_{m+1} \\ &= \frac{m}{\hat{a}(m)} \left(e^{R/m} - e^{(R-r_{m+1}(m+1))/m} \right) \\ &\quad - \frac{e^{r_{m+1}(m+1)} - 1}{a_{m+1}} \end{aligned} \quad (\text{A.9})$$

$$\begin{aligned} g'(r_{m+1}(m+1)) &= \frac{\partial g(r_{m+1}(m+1))}{\partial r_{m+1}(m+1)} \\ &= \frac{e^{(R-r_{m+1}(m+1))/m}}{\hat{a}(m)} \\ &\quad - \frac{e^{r_{m+1}(m+1)}}{a_{m+1}} \end{aligned} \quad (\text{A.10})$$

The solution of $g'(r_{m+1}(m+1)) = 0$, $r_{m+1}^*(m+1)$, is given by

$$\begin{aligned} r_{m+1}^*(m+1) &= \frac{R}{m+1} + \frac{m}{m+1} \cdot \ln \left[\frac{a_{m+1}}{\hat{a}(m)} \right] \\ &= \frac{R}{m+1} + \ln \left[\frac{a_{m+1}}{\hat{a}(m+1)} \right] \end{aligned} \quad (\text{A.11})$$

For $0 \leq r_{m+1}(m+1) < R$, the second derivative of $g(r_{m+1}(m+1))$

$$\begin{aligned} g^{(2)}(r_{m+1}(m+1)) &= \frac{-1}{\hat{a}(m)a_{m+1}} \\ &\quad \left[\frac{a_{m+1}}{m} e^{(R-r_{m+1}(m+1))/m} + \hat{a}(m)e^{r_{m+1}(m+1)} \right] \end{aligned} \quad (\text{A.12})$$

is always negative. Since $g'(r_{m+1}^*(m+1)) = 0$, $g'(r_{m+1}(m+1)) > 0$, for $0 \leq r_{m+1}(m+1) < r_{m+1}^*(m+1)$, the fact that $g(0) = 0$ then lead to the desired conclusion that $g(r_{m+1}^*(m+1)) > 0$. In other words, the minimum power for the case $x = m$ is larger than that for the case $x = m + 1$ which can be achieved with $r_{m+1}(m+1) = r_{m+1}^*(m+1)$. ■

The above two Lemmas suggest that the solution to the constrained optimization problem (A.2) can be found by repeatedly calculating the unconstrained solution (A.5) for $x = N, N-1, N-2, \dots$ until the constraints $p_i \geq 0$, $\forall 1 \leq i \leq x$ are satisfied. A similar but less efficient solution was proposed by Fischer and Huber [2] who iteratively recompute (A.5) by excluding all negative-rate channels and setting $x \leftarrow x - l$, where l is the number of negative-rate channels. Such an approach does not necessarily give the optimal solution and the issue of optimality was not addressed in [2]. Instead of sequentially decreasing x with a decrement of 1, we accelerate the process of locating the optimal x through a bisection search so that the optimal power allocation can be found in Table V.

The resulting algorithm will be referred to as the optimal mono-rate power allocation (OMPA) algorithm henceforth. Note the OMPA algorithm can easily be modified to solve the maximum sum-rate problem

$$\max \sum_i r_i, \quad \text{s.t.} \quad \sum_i p_i \leq P, \quad p_i \geq 0, \quad (\text{A.13})$$

Table V: An Optimal Mono-rate Power Allocation Algorithm

Step 1:	(Initialization) Given $a_i, 1 \leq i \leq N$, and R , set $upbound = N$, $lowbound = 1$, and $x^* = [(upbound + lowbound)/2]$.
Step 2:	(Update the lowest rate) $r_{x^*}(x^*) = \frac{R}{x^*} + \log_2 \left[\frac{a_{x^*}}{\hat{a}(x^*)} \right]$, where $\hat{a}(x^*) = \left(\prod_{j=1}^{x^*} a_j \right)^{1/x^*}$, the number of iterations.
Step 3:	If $r_{x^*}(x^*) \geq 0$, $lowbound \leftarrow x^*$, else $upbound \leftarrow x^*$
Step 4:	If $lowbound < upbound - 1$, $x^* \leftarrow [(upbound + lowbound)/2]$, go to Step 2; else $x^* \leftarrow lowbound$, $r_i(x^*) \leftarrow 0$, for $i > x^*$ and compute $r_i(x^*)$, for $1 \leq i < x^*$.

ACKNOWLEDGEMENT

The authors would like to thank the reviews for their detailed comments and helpful suggestions, and for bringing reference [13] to their attention.

REFERENCES

- [1] 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, Oct. 1999.
- [2] R. Fischer and J. B. Huber, "A new loading algorithm for discrete multitone transmission," in *Proc. IEEE Globecom*, vol. 1, pp. 724-728, Nov. 1996.
- [3] J. Campello, "Practical bit loading for DMT," in *Proc. ICC 1999*, Vancouver, Canada, June 1999.
- [4] K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *Proc. ISIT 2006*, Seattle, USA, July 2006.
- [5] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Wireless Commun.*, vol. 2, pp. 1150-1158, Nov. 2003.
- [6] M. Bohge, J. Gross, A. Wolisz, and M. Meyer, "Dynamic resource allocation in OFDM systems: an overview of cross-layer optimization principles and techniques," *IEEE Network*, vol. 21, pp. 53-59, Jan. 2007.
- [7] G. Li and H. Liu, "Resource allocation for OFDMA relay networks with fairness constraints," *IEEE J. Select. Areas Commun.*, vol. 24, no. 11, pp. 2061-2069, Nov. 2006.
- [8] 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, Nov. 2005.
- [9] M. Ergen, S. Coleri, and P. Varaiya, "QoS aware adaptive resource allocation techniques for fair scheduling in OFDMA based broadband wireless access systems," *IEEE Trans. Broadcasting*, vol. 49, pp. 362-370, Dec. 2003.
- [10] G. Li and H. Liu, "Downlink dynamic resource allocation for multi-cell OFDMA system," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3451-3459, Dec. 2006.
- [11] Wimax Forum, "WiMAX Forum mobile release 1.0 channel model," 2008. [Online]. Available: <http://www.wimaxforum.org/technology/documents>
- [12] C. Y. Ng and C. W. Sung, "Low complexity subcarrier and power allocation for utility maximization in uplink OFDMA systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1667-1675, May 2008.
- [13] Z. Mao and X. Wang, "Efficient optimal and suboptimal radio resource allocation in OFDMA system," *IEEE Trans. Wireless Commun.*, vol. 7, pp. 440-445, Feb. 2008.

Yuan-Bin Lin received the B.S. and M.S. degrees in communications engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1998 and 2000, respectively. From 2003 to 2005, he was a lecturer in the Ta Hwa Institute of Technology, Hsinchu, Taiwan. His current research interests include radio resource allocation, convex optimization and communication theory.

Tai-Hsiang Chiu received the B.S. and M.S. degrees in communications engineering from the National Chiao Tung University, Hsinchu, Taiwan, in 2005 and 2007, respectively. After fulfilling his mandatory military service he joined MediaTek Incorporation (Hsinchu, Taiwan) as an R&D engineer in the System Development Division of the Wireless Communications BU.

Yu T. Su received the B.S. and Ph.D. degrees in electrical engineering from Tatung Institute of Technology, Taipei, Taiwan and the University of Southern California, Los Angeles, USA, in 1974 and 1983, respectively. From 1983 to 1989, he was with LinCom Corporation, Los Angeles, USA, where he was a Corporate Scientist involved in the design of various measurement and digital satellite communication systems. Since September 1989, he has been with the National Chiao Tung University, Hsinchu, Taiwan, where he is an Professor at the Department of Communications Engineering. He was an Associative Dean of the College of Electrical and Computer Engineering from 2004 to 2007, and was the Head of the Communications Engineering Department from 2001 to 2003. He is also affiliated with the Microelectronic and Information Systems Research Center of the same university and served as a Deputy Director from 1997 to 2000. In 2005, he was appointed as the Area Coordinator of National Science Council's Telecommunications Programme. His main research interests include communication theory and statistical signal processing.