QoS Aware Adaptive Resource Allocation Techniques for Fair Scheduling in OFDMA Based Broadband Wireless Access Systems

Mustafa Ergen, Sinem Coleri, and Pravin Varaiya

***Abstract*—A system based on orthogonal frequency division multiple access (OFDMA) has been developed to deliver mobile broadband data service at data rates comparable to those of wired services, such as DSL and cable modems. We consider the resource allocation problem of assigning a set of subcarriers and determining the number of bits to be transmitted for each subcarrier in OFDMA systems. We compare simplicity, fairness and efficiency of our algorithm with the optimal and proposed suboptimal algorithms for varying values of delay spread, number of users and total power constraint. The results show that performance of our approach is appealing and can be close to optimal.**

We also consider another resource allocation scheme in which there is no fixed QoS requirements per symbol but capacity is maximized.

***Index Terms*—Adaptive modulation, broadband wireless networks, fair scheduling, IEEE 802.16, multiple access, multiuser diversity, multiuser OFDM, OFDMA, power control, resource allocation.**

I. Introduction

BROADBAND WIRELESS ACCESS (BWA) is an appealing system for providing flexible and easy deployment solution to high-speed communications. It is an alternative to wireline broadband access techniques such as copper line, coaxial cable, xDSL and cable modem [1], [2]. Visionaries predict a big market because of its distributed installation and semi ad hoc routing protocol that reduces the need for an infrastructure [3], [4].

Vector Orthogonal Frequency Division Multiplexing (VOFDM) is considered as a base setting for BWA systems by the Broadband Wireless Internet Forum (BWIF), one of the programs of the IEEE Industry Standards and Technology Organization (IEEE-ISTO) [1]. Some vendors offer BWA system with existing wireless LAN technologies such as IEEE802.11(a, b) and IEEE 802.16 group aims to unify the BWA solutions [5] 802.16 group issued standards in the 10–66 GHz bands and IEEE802.16a group was formed to develop standards to operate in the 2–11 GHz bands in which channel impairments, multipath fading and path loss become more significant with the increase in the number of subscribers. Improved and flexible multiple access methods are needed to cope with these impairments. Orthogonal Frequency Division Multiple Access (OFDMA) is a promising multiple access scheme that has attracted interest. OFDMA is based on OFDM and inherits its immunity to inter-symbol interference and frequency selective fading [6], [7].

Achieving high transmission rates depends on the ability of BWA system to provide efficient and flexible resource allocation. Recent studies [2], [8]–[13] on resource allocation demonstrate that significant performance gains can be obtained if frequency hopping and adaptive modulation are used in subcarrier allocation, assuming knowledge of the channel gain in the transmitter. The frequency hopping strategy resembles the interference cancellation of CDMA [14].

In multiuser environment, a good resource allocation scheme leverages multiuser diversity and channel fading [15]. It was shown in [16] that the optimal solution is to schedule the user with the best channel at each time. Although in this case, the entire bandwidth is used by the scheduled user, this idea can also be applied to OFDMA system, where the channel is shared by the users, each owing a mutually disjoint set of subcarriers, by scheduling the subcarrier to a user with the best channel among others. Of course, the procedure is not simple since the best subcarrier of the user may also be the best subcarrier of another user who may not have any other good subcarriers. The overall strategy is to use the peaks of the channel resulting from channel fading. Unlike in the traditional view where the channel fading is considered to be an impairment, here it acts as a channel randomizer and increases multiuser diversity [15].

The resource allocation problem has been recently considered in many studies. Almost all of them define the problem as a real time resource allocation problem in which Quality of Service (QoS) requirements are fixed by the application. QoS requirement is defined as achieving a specified data transmission rate and bit error rate (BER) of each user in each transmission. In this regard, the problem differs from the water-pouring schemes wherein the aim is to achieve Shannon capacity under the power constraint [8].

We introduce an iterative multiuser bit and power allocation scheme so that the QoS requirements of users are fulfilled. Our objective is to minimize the total transmit power by allocating subcarriers to the users and then to determine the number of bits transmitted on each subcarrier. Variable transmittable bits (i.e adaptive modulation) was considered in [8], [9]. Our scheme is simple and sufficiently fair to meet real time applications criteria in which a quick scheme is needed to allocate subcarriers before

Manuscript received February 24, 2003; revised August 6, 2003. Research supported by a grant from National Semiconductor.

The authors are with the Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkley, CA USA (e-mail: ergen@eecs.berkeley.edu; csinem@eecs.berkeley.edu; varaiya@eecs.berkeley.edu).

Digital Object Identifier 10.1109/TBC.2003.819051

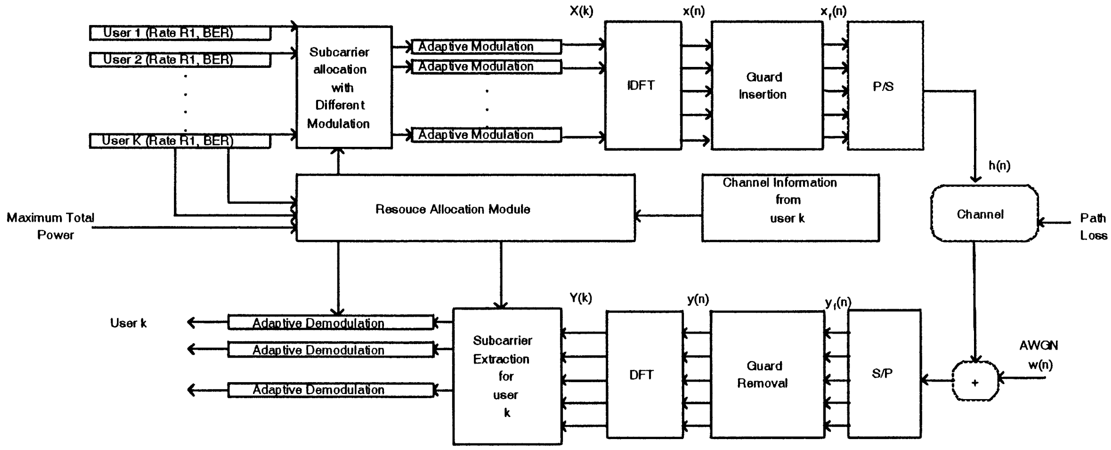


Fig. 1. Orthogonal frequency division multiple access system.

the channel changes and a fair scheme is needed to treat each user.

We also consider a continuous allocation scheme where the allocator uses the previous channel information per user for the current allocation. We try to extend the point-to-point version of proportional fair scheduling (as in [15]) to a point-to-multipoint version. In this scheme there is no fixed requirements per symbol, the aim is to maximize capacity.

In Section II we give the main features of an OFDMA system and state the resource allocation problem. In Section III we discuss about optimal solution. In Section IV we review the proposed suboptimal solutions. In Section V we introduce our resource allocation scheme. In Section VI we briefly introduce another resource allocation solution which aims to maximize the capacity. In Section VII we present the performance analysis of our system compared to optimal and proposed suboptimal schemes. In Section VIII we conclude the paper.

II. Orthogonal Frequency Division Multiple Access (OFDMA)

This section outlines the OFDMA system and states the resource allocation problem. Unlike in an OFDM system [6], K users are involved in the OFDMA system to share N subcarriers. The difference arises in the forming and an deforming of FFT block. The rest is the same as an OFDM system as seen in Fig. 1. Each user allocates nonoverlapping set of subcarrriers S_k where the number of subcarriers per user is $J(k)$ following the notation in [17]. We denote by $X_k(l)$ the l^{th} subcarrier of the FFT block belonging to k^{th} user. $X_k(l)$ is obtained by coding the assigned bits c with the corresponding modulation scheme. In the downlink the $X_k(l)$ are multiplexed to form the OFDM symbol[1] of length $(N+L)$ with the appended guard prefix L in order to eliminate ISI. At the uplink, the OFDM symbol is formed in the base station with a synchronization error.

$$x(l) = \sum_{k=0}^{K-1} \sum_{n=0}^{J(k)-1} X_k(n) e^{j\frac{2\pi}{N}(I_k(n))l}, \qquad (1)$$

[1]A OFDMA symbol is defined as one OFDM FFT block.

with $n = -L, \ldots, N-1$, where $I_k(n)$ denotes the subcarrier assigned to the k^{th} user. Resource allocation problem comes into the picture when associating the set of subcarriers to the users with different bits loaded into them. The received signal from the j^{th} user is

$$y_j(l) = x(l) \bigotimes h_j(l) + w(l), \qquad (2)$$

where $h_j(t)$ is the baseband impulse response of the channel between base station (BS) and j^{th} user. Equation (2) is the received signal $y(t)$ sampled at rate $1/T$. The first L samples are discarded and the N-point FFT is computed. The data of the j^{th} user is

$$Y_j(n) = \begin{cases} X_j(n)H_j\left(i_j(n)\right) + W(n), & \text{if } i_j(n) \in S_j \\ 0, & \text{otherwise,} \end{cases} \qquad (3)$$

where $H_j(n) := \sum_i h_j(i) exp(j2(\pi/N)ni)$ is the frequency response of the channel of k^{th} user.

In a perfectly synchronized system, the allocation module of the transmitter assigns subcarriers to each user according to some QoS criteria. QoS metrics in the system are rate and bit error rate (BER). Each user's bit stream is transmitted using the assigned subcarriers and adaptively modulated for the number of bits assigned to the subcarrier. The power level of the modulation is adjusted to overcome the fading of the channel. The transmission power for AWGN channel can be predicted. In addition the channel gain of subcarrier n to the corresponding user k should be known. The channel gain of the subcarrier is defined as

$$\alpha_{k,n} = H_k(n) * PL_k, \qquad (4)$$

where PL is the path loss, defined by

$$PL_k = PL(d_o) + 10\alpha \log_{10}\left(\frac{\mathrm{d_k}}{\mathrm{d_o}}\right) + \mathrm{X}_\sigma,$$

where d_o is the reference distance, d_k is the distance between transmitter and receiver, α is the path loss component and X_σ is a Gaussian random variable for shadowing with a standard deviation σ [10]. An example of channel gain can be seen in Fig. 2.

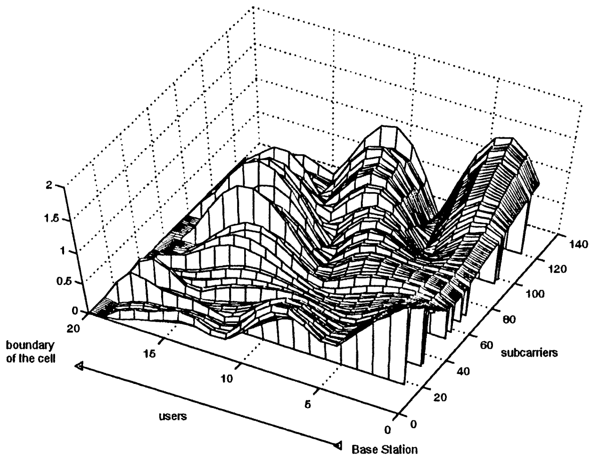


Fig. 2. An example of channel gain.

The problem above is called resource allocation in the OFDMA literature. The channel information is assumed to be known at transmitter and receiver [2], [8]–[13]. The channel is assumed to be reciprocal; BS is able to estimate the channel of all BS-to-mobile links based on the received uplink transmission as long as the channel variation is slow [11]. As a result, the resource allocation should be done within the coherence time [8].

With the channel information, the objective of resource allocation problem can be defined as maximizing the throughput subject to a given total power constraint regarding the user's QoS requirements. As we clarify further, BER_k of the transmission should not be higher than the required BER_k and data rate of every user should be equal to the requirement R_k.

Let's define $\gamma_{k,n}$ as the indicator of allocating the n^{th} subcarrier to the k^{th} user. The transmission power allocated to the n^{th} subcarrier of k^{th} user is expressed as

$$P_{k,n} = \frac{f_k(c_{k,n}, BER_k)}{\alpha_{k,n}^2}, \tag{5}$$

where $f_k(c_k, n)$ is the required received power with unity channel gain for reliable reception of $c_{k,n}$ bits per symbol [9]. We can formulate the resource allocation problem with an imposed power constraint as

$$\begin{aligned} \max_{c_{k,n},\gamma_{k,n}} R_k &= \sum_{n=1}^{N} c_{k,n}\gamma_{k,n} \text{ for all } k \\ \text{subject to } P_T &= \sum_{k=1}^{K}\sum_{n=1}^{N} \\ &\times \frac{f_k(c_{k,n}, BER_k)}{\alpha_{k,n}^2}\gamma_{k,n} \le P_{Max} \end{aligned} \tag{6}$$

where the limit on the total transmission power is expressed as P_{Max} for all $n \in \{1,\ldots,N\}$, $k \in \{1,\ldots,K\}$ and $c_{k,n} \in \{1,\ldots,M\}$.

If there is no power constraint, (6) is changed in order to minimize P_T subject to allocating R_k bits for all k (i.e problem is to find the values of the $\gamma_{k,n}$ and the corresponding $c_{k,n}$ while minimizing P_T) [8], [9], [12]. As it can be seen the cost function in our system is the power consumption matrix in (5). Rather than using $\alpha_{k,n}^2$ as in [8], [9], [12], we adopted using $P_{k,n}$ from [10] since in this case modulation type and BER get involved in the decision process.

III. Optimal Solution

In a multiuser environment with multiple modulation techniques, the solution to the problem is complicated since the optimal solution needs to pick the subcarriers in balance. We can classify the problem according to each set of bits assigned to a subcarrier. For a user k, $f_k(c_{k,n}) \in \{f_k(1, BER_k),\ldots,f_k(M, BER_k)\}$. We can construct M times $[K*N]$ power matrices $\{P^c\}$ for each c. For a constant c, $\{f(c)\}$ can be computed and the transmission power requirement can be found with (5). The dimension of the indicator function is incremented and represented by $\gamma_{k,n,c}$ and defined as follows [9]:

$$\gamma_{k,n,c} = \begin{cases} 1, & if\ c_{k,n} = c \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

The above problem can be solved with Integer Programming (IP). We refer to the IP approach as the optimal solution to the resource allocation problem. As stated in [9], the non linear approximation in [8], [13] requires more computation than the IP.

There are $K*N*M$ indicator variables and M power matrices where the entries of each matrix for a given c can be found from

$$P_{k,n}^c = \frac{f_k(c, BER_k)}{\alpha_{k,n}^2}. \tag{8}$$

Using (8) as an input, the cost function now can be written as

$$P_T = \sum_{k=1}^{K}\sum_{n=1}^{N}\sum_{c=1}^{M} P_{k,n}^c \gamma_{k,n,c} \tag{9}$$

and the description of the IP problem is

$$\min_{\gamma_{k,n,c}} P_T, \quad \text{for } \gamma_{k,n,c} \in \{0,1\} \tag{10}$$

subject to

$$R_k = \sum_{n=1}^{N}\sum_{c=1}^{M} c.\gamma_{k,n,c}, \quad \text{for all } k,$$

and

$$0 \le \sum_{k=1}^{K}\sum_{c=1}^{M} \gamma_{k,n,c} \le 1, \text{ for all } n.$$

Although the optimal solution gives the exact results, from an implementation point of view, it is not preferred since in a time varying channel, in order to allocate the subcarriers within the coherence time, the allocation algorithm should be fast and the IP complexity increases exponentially with the number of constraints. This real time requirement leads to searching suboptimal solutions that are fast and close to the optimal solution. Several suboptimal allocation schemes are proposed for different settings in the literature [8]–[12], [18]. Up to now suboptimal solutions differ in the modulation type. There are a few suboptimal schemes that use adaptive modulation, the rest assume fixed modulation, i.e. same number of bits are assigned to each subcarrier. We will describe current solutions and compare with our iterative solution.

IV. Suboptimal Solutions

In most attempts to simplify the resource allocation problem, the problem is decomposed into two procedures: A subcarrier allocation with fixed modulation, and bit loading. Subcarrier allocation with fixed modulation deals with one P^c matrix with fixed c and then by using bit loading scheme, the number of bits is incremented.

A. Subcarrier Allocation

We know that $f_k(x,y)$ is a convex function [8], [9]. We can start with $P^1_{k,n}$ and we can define new $\bar{R}_k$ with $\sum_{k=1}^{K}\bar{R}_k \leq N$ which can be obtained by decrementing R_k properly. Then the solution to this problem can be solved with Linear Programming or Hungarian problem. Although the Hungarian algorithm is proposed as an optimal solution for resource allocation with a fixed modulation in [10], [12], we consider it as a suboptimal solution for adaptive modulation.

1) Linear Programming: Linear programming is investigated in [9]. For comparison purposes, we briefly restate the problem description,

$$P_T = \min \sum_{k=1}^{K}\sum_{n=1}^{N} P^1_{k,n}\rho_{k,n} \quad \rho_{k,n} \in [0,1], \tag{11}$$

subject to

$$\sum_{n=1}^{N} \rho_{k,n} = \bar{R}_k \quad \forall\, k \in \{1,\ldots,K\}$$
$$\sum_{k=1}^{K} \rho_{k,n} = 1 \quad \forall\, n \in \{1,\ldots,N\}.$$

After linear programming, the $[K*N]$ allocation matrix has entries ranging between 0 and 1. The entries are converted to integers by selecting the highest $\bar{R}_k$ nonzero values from N columns for each k and assigning them to the k^{th} user.

2) Hungarian Algorithm: The problem described above can also be solved by an assignment method such as Hungarian algorithm [19]. The Hungarian algorithm works with square matrices. Entries of the square matrix can be formed by adding $\bar{R}_k$ times the row of each k. The problem formulation is as

$$P_T = \min \sum_{k=1}^{N}\sum_{n=1}^{N} P^1_{k,n}\rho_{k,n} \quad \rho_{k,n} \in \{0,1\} \tag{12}$$

and the constraints become

$$\sum_{n=1}^{N} \rho_{k,n} = 1 \quad \forall\, n\ \in \{1,\ldots,N\}$$
$$\sum_{k=1}^{N} \rho_{k,n} = 1 \quad \forall\, k\ \in \{1,\ldots,N\}$$

as stated in [12]. Although the Hungarian method has computation complexity $O(n^4)$ in the allocation problem with fixed modulation, it may serve as a base for adaptive modulation.

B. Bit Loading Algorithm

The bit loading algorithm (BLA) appears after the subcarriers are assigned to users that have at least $\bar{R}_k$ bits assigned. Bit loading procedure is as simple as incrementing bits of the assigned subcarriers of the users until $P_T \leq P_{Max}$. Following the notation of [8], [9], define $\Delta P_{k,n}(c)$ as the additional power needed to increment one bit of the n^{th} subcarrier of k^{th} user as represented in (13),

$$\Delta P_{k,n}(c_{k,n}) = \frac{[f(c_{k,n}+1) - f(c_{k,n})]}{\alpha^2_{k,n}} \tag{13}$$

The bit loading algorithm assigns one bit at a time with a greedy approach to the subcarrier $\{\arg\min_{k,n} \Delta P_{k,n}(c_{k,n})\}$.

BL Algorithm

Step 1) For all n, Set $c_{k,n} = 0$, $\Delta P_{k,n}(c_{k,n})$, and $P_T = 0$;
Step 2) Select $\bar{n} = \arg\min_n \Delta P_{k,n}(0)$;
Step 3) Set $c_{k,\bar{n}} = c_{k,\bar{n}} + 1$ and $P_T = P_T + \Delta P_{k,n}(c_{k,n})$;
Step 4) Set $\Delta P_{k,n}(c_{k,\bar{n}})$;
Step 5) Check $P_T \leq P_{Max}$ and R_k for $\forall k$, if not satisfied GOTO STEP 2.
Step 6) Finish.

It is a simple algorithm. Bits on the subcarriers are incremented one by one. If there is no power constraint, procedure runs for $\sum_{k=1}^{K} R_k$ times. This algorithm enable us to convert the fixed modulation schemes into adaptive modulation ones.

The Hungarian approach and LP approach with bit loading appear as two different suboptimal solutions to the resource allocation with adaptive modulation. We use these schemes as a reference in our simulations and call them GreedyHungarian and GreedyLP respectively in our simulations.

V. Iterative Solution

The GreedyLP and GreedyHungarian methods both first determine the subcarriers and then increment the number of bits on them according to the rate requirements of users. This may not be a good schedule in some certain cases: For instance, consider a user with only one good subcarrier and low rate requirement. The best solution for that user is allocating its good carrier with high number of bits. But if GreedyLP or GreedyHungarian is used, user may have allocated more than one subcarrier with lower number of bits and in some cases, its good subcarrier is never selected. Consider another scenario where a user does not have any good subcarrier (i.e. it may have a bad channel or be at the edge of the cell). In this case, rather than pushing more bits and allocating less subcarriers as in GreedyLP and GreedyHungarian, the opposite strategy is preferred since fewer bits in higher number of subcarriers give better result. Another difficulty arises in providing fairness. Since GreedyLP and GreedyHungarian are based on greedy approach, the user in the worst condition usually suffers. In any event, these are complex schemes and simpler schemes are needed to finish the allocation within the coherence time. To cope with these challenges, we introduce a simple, efficient and fair subcarrier allocation scheme with iterative improvement.

Our scheme is composed of two modules named scheduling and improvement modules. In the scheduling section, bits and subcarriers are distributed to the users and passed to the improvement module where the allocation is improved iteratively by bit swapping and subcarrier swapping algorithms.

A. Fair Scheduling Algorithm

We introduce a simple and mixed allocation scheme that considers fair allocation among users with adaptive modulation. The allocation procedure starts with the highest level of modulation scheme. In this way, it tries to find the best subcarrier of a user to allocate the highest number of bits. We can describe the strategy by an analogy: "The best strategy to fill a case with stone, pebble and sand is as follows. First filling the case with the stones and then filling the gap left from the stones with pebbles and in the same way, filling the gap left from pebbles with sand. Since filling in opposite direction may leave the stones or pebbles outside". With this strategy more bits can be allocated and the scheme becomes immune to uneven QoS requirements. The fair scheduling algorithm (FSA) runs greedy release algorithm (GRA) if there are nonallocated subcarriers after the lowest modulation turn and the rate requirement is not satisfied. GRA decrements one bit of a subcarrier to gain power reduction, which is used to assign higher number of bits to the users on the whole. FSA is described as follows;

FS Algorithm

Step 1) Set $c = M$, Select a k, and $P_T = 0$;
Step 2) Find $\bar{n} = arg\min_n P^c_{k,n}$;
Step 3) Set $R_k = R_k - c$ and $\rho_{k,\bar{n}} = 1$, Update P_T, Shift to the next k;
Step 4) If $P_T > P_{Max}$, Step Out and Set $c = c - 1$, GOTO STEP 2.
Step 5) If $\forall k$, $R_k < c$, Set $c = c - 1$, GOTO STEP 2.
Step 6) If $\{c == 1\}$, $\sum_{k=1}^{K}\sum_{n=1}^{N}\rho_{k,n} < N$, $P_T > P_{Max}$, Run "Greedy Release" and GOTO STEP 2.
Step 7) Finish.

B. Greedy Releasing Algorithm

The GRA tends to fill the un-allocated subcarriers. It releases one of the bits of the most expensive subcarrier to gain power reduction in order to drive the process. GRA works in the opposite direction of BLA. GRA is described as follows;

GR Algorithm

Step 1) Find $\{\bar{k}, \bar{n}, \bar{c}_{\bar{k},\bar{n}}\} = \arg\max_{k,n,c} P^c_{k,n}\rho_{k,n}\forall c$;
Step 2) Set $\bar{c}_{k,n} = \bar{c}_{k,n} - 1$, $P_T = P_T - \Delta P_{\bar{k},\bar{n}}(c_{\bar{k},\bar{n}})$;
Step 3) Set $c = c_{\bar{k},\bar{n}} - 1$;
Step 4) Finish.

C. Horizontal Swapping Algorithm

The horizontal swapping algorithm (HSA) aims to smooth the bit distribution of a user. When the subcarriers are distributed, the bit weight per subcarrier can be adjusted to reduce power. One bit of a subcarrier may be shifted to the other subcarrier of the same user if there is a power reduction gain. Therefore, variation of the power allocation per subcarrier is reduced and a smoother transmission is performed. HSA is described as follows;

HS Algorithm

Step 1) Set $P_C = \infty$
STEP1a: Find $\{\bar{k}, \bar{n}, \bar{c}_{\bar{k},\bar{n}}\} = \arg\max_{k,n,c}(P^c_{k,n}\rho_{k,n}) < P_C \forall c$;
Step 2: Define $n \in S_k$, where $\{\rho_{k,n} == 1\}$ for $\forall n$;
Step 3: Set $\Delta_{\dot{n}} = \max_n \Delta P_{\bar{k},\bar{n}}(c_{\bar{k},\bar{n}} - 1) - \Delta P_{\bar{k},\dot{n}}(c_{\bar{k},\dot{n}})$, $\dot{n} \in S_k$;
Step 4: Set $P_C = P^{\bar{c}}_{\bar{k},\bar{n}}$;
STEP4a: if $\Delta_{\dot{n}} > 0$, Set $P_T = P_T - \Delta_{\dot{n}}$
STEP4b: Set $c_{\bar{k},\bar{n}} = c_{\bar{k},\bar{n}} - 1$, $c_{\bar{k},\dot{n}} = c_{\bar{k},\dot{n}} + 1$ GOTO Step 1a;
Step 5: if $\{P_C == \min_{k,n,c}(P^c_{k,n}\rho_{k,n})\}$, Finish.

D. Vertical Swapping Algorithm

Vertical swapping is done for every pair of users. In each iteration, users try to swap their subcarriers such that the power allocation is reduced. There are different types of vertical swapping. For instance, in triple swapping, user i gives its subcarrier to user j and in the same way user j to user k and user k to user i. Pairwise swapping for fixed modulation is described in [10], [12] with a slight difference: the former uses power and the latter uses channel gain as a decision metric. We modified pairwise swapping to cope with adaptive modulation case. In this case, there is more than one class where each class is defined with its modulation (i.e number of bits loaded to a subcarrier) and swapping is only within the class. Each pair of user swap their subcarriers that belong to the same class if there is a power reduction. In this way, adjustment of subcarrier is done across users, to try to approximate the optimal solution. VSA is described as follows;

VS Algorithm

Step 1) $\forall$ pair of user $\{i, j\}$;
STEP1a: Find $\partial P_{i,j}(n) = P^{\dot{c}}_{i,n} - P^{\dot{c}}_{j,n}$ and $\Delta^{\hat{n}} P_{i,j} = \max \partial P_{i,j}(n)$, $\forall n \in S_i$;
STEP1b: Find $\partial P_{j,i}(n) = P^{\dot{c}}_{j,n} - P^{\dot{c}}_{i,n} \Delta^{\check{n}} P_{j,i} = \max \partial P_{j,i}(n)$, $\forall n \in S_j$;
STEP1c: Set $\Omega^{\hat{n},\check{n}} P_{i,j} = \Delta^{\hat{n}} P_{i,j} + \Delta^{\check{n}} P_{j,i}$;
STEP1d: Add $\Omega^{\hat{n},\check{n}} P_{i,j}$ to the $\{\Lambda\}$ list;
Step 2: Select $\Omega = \max_{(i,\hat{n}),(j,\check{n})} \Lambda$;
Step 3: if $\Omega > 0$, Switch subcarriers and $P_T = P_T - \Omega$ GOTO STEP 1a;
Step 4: if $\Omega \leq 0$, Finish.

VI. Resource Allocation Regarding Capacity

In the previous sections, we considered the problem where the QoS requirements per symbol is fixed. Another way to approach resource allocation is in terms of capacity [8]. Suppose there is no fixed requirements per symbol and the aim is to maximize capacity.

It has been shown in [15] that for point-to-point links, a fair allocation strategy maximizes total capacity and the throughput of each user in the long run, when the user's channel statistics are the same. This idea underlying the proposed fair scheduling algorithm is exploiting the multiuser diversity gain.

With a slight modification, we can extend the fair scheduling algorithm for point-to-point communication to an algorithm for point-to-multipoint communication. Suppose the user time varying data rate requirement $R_k(t)$ is sent by the user to the base station as feedback of the channel condition. We treat symbol time as the time slot, so t is discrete, representing the number of symbols. We keep track of average throughput $t_{k,n}$ of each user for a subcarrier in a past window of length t_c. The scheduling algorithm will schedule a subcarrier $\bar{n}$ to a user $\bar{k}$ according to the following criterion

$$\{\bar{k}, \bar{n}\} = \arg\max_{k,n} \frac{r_{k,n}}{t_{k,n}} \tag{14}$$

where $t_{k,n}$ can be updated using an exponentially weighted low-pass filter described in [15]. Here, we are confronted with determining the $r_{k,n}$ values. We can set $r_{k,n}$ to R_k/N, where N is the number of carriers. With this setting, the peaks of the channel for a given subcarrier can be tracked. The algorithm schedules a user to a subcarrier when the channel quality in that subcarrier is high relative to its average condition in that subcarrier over the time scale t_c. When we consider all subcarriers the fairness criterion match with the point-to-point case as follows

$$\bar{k} = \max_{k} \frac{R_k}{T_k}, \tag{15}$$

where $T_k = \sum_{n=1}^{N} t_{k,n}$. The theoretical analysis of fairness property of (15) for point-to-point communication is derived in [15]. We can apply those derivations for point-to-multipoint communication.

VII. Performance Analysis

We compare the performance of our iterative algorithm with the proposed suboptimal GreedyHungarian and GreedyLP schemes and optimal IP scheme. We adopt the M-ary quadrature amplitude modulation of 4-QAM, 16-QAM, and 64-QAM which are used to carry two, four, or six bits/subcarrier [8]. Required transmission power for c bits/subcarrier at a given BER with unity channel gain is

$$f(c, BER) = \frac{N_0}{3}\left[Q^{-1}\left(\frac{BER}{4}\right)\right]^2 (2^c - 1) \tag{16}$$

where $Q^{-1}(x)$ is the inverse function of

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{t^2}{2}} dt.$$

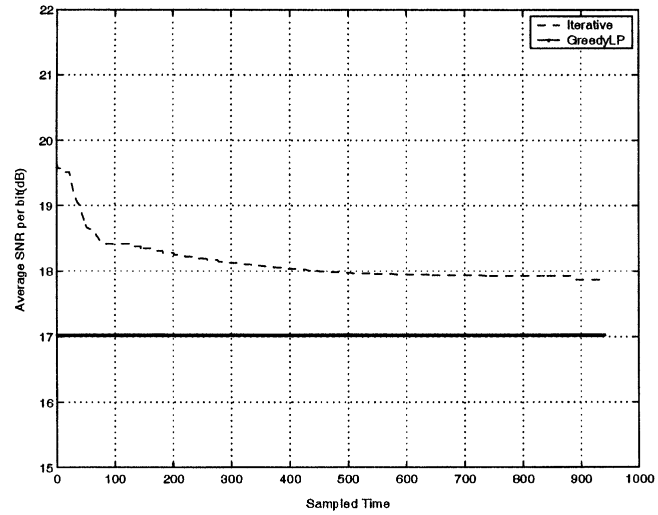


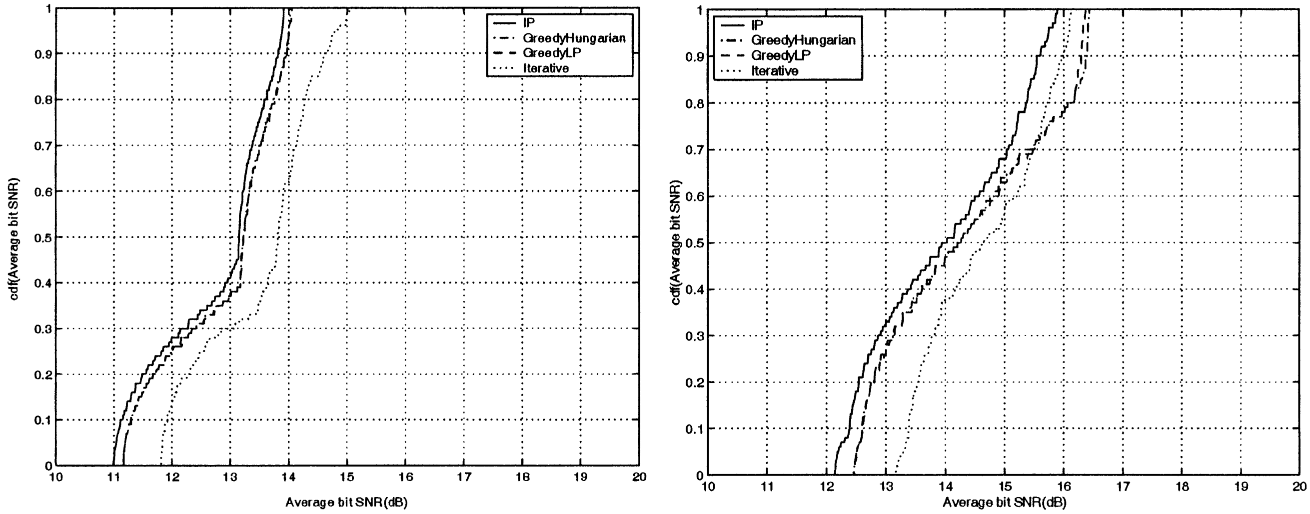
Fig. 3. Comparison of convergence of the iterative approach to the GreedyLP one.

We evaluate our scheme in Rayleigh fading channels video, [20], [21]. The power spectral density level N_o is equal to unity, and gain of Rayleigh channel $E|H_k(n)|^2$ is also equal to unity. We use 128 subcarriers with a total transmission rate between 480 bits/symbol and 768 bits/symbol. BER requirement of users is selected from the list $\wp = \{1e-2, 1e-4\}$. In the simulations, depending on the constraint, either the rate requirements are fulfilled when the transmit power is minimized or the power constraint is fulfilled when rates are maximized. BER requirement, on the other hand, is satisfied in both situations. We distinguish the settings by naming them with or without power constraint.

Fig. 3 shows the convergence of our scheme with iterative betterment. In each iterative step, the power is reduced keeping the total number of bits constant. The steepest decrease is observed in the HSA step since the power reduction in bit swapping is higher than the one in subcarrier swapping because of the exponential growth of the $f(x, y)$ function. It can be seen from the figure that iterative solution approximates the GreedyLP with time.

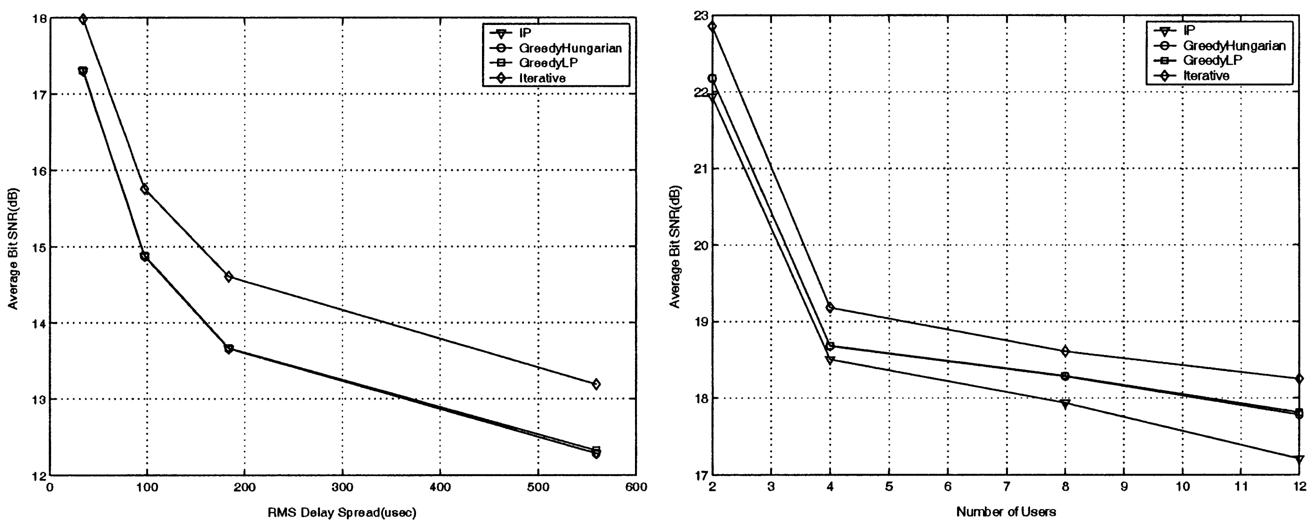
Fig. 4 presents the cumulative distribution functions of the average bit SNR for the cases with and without power constraint. There are four users in two sets of BER requirement and each user has rate requirement of 120 bits/symbol. It can be seen from the Fig. 4(a) that iterative approach approximates the optimal solution up to 0.9 dB when there is no power constraint. When there is a power constraint, as seen in Fig. 4(b), the iterative approach outperforms the GreedyHungarian and GreedyLP approach and is close to the IP solution within 0.3 dB. The reason why iterative solution gives better performance than the suboptimal solution is its tight power control scheme, which allows to transmit more number of bits. GRA is one of the most important module that decreases the variance of avg bit SNR and make the iterative approach perform better at the end by exchanging one high cost bit with more than one low cost bits i.e. lower level modulation.

Fig. 5 shows the performance of the schemes in various channel fading and multiuser diversity situations [15]. Fig. 5(a)



(a) (b)

Fig. 4. Comparison of the cumulative distribution function of the average bit SNR; (a) without power constraint; (b) with power constraint.



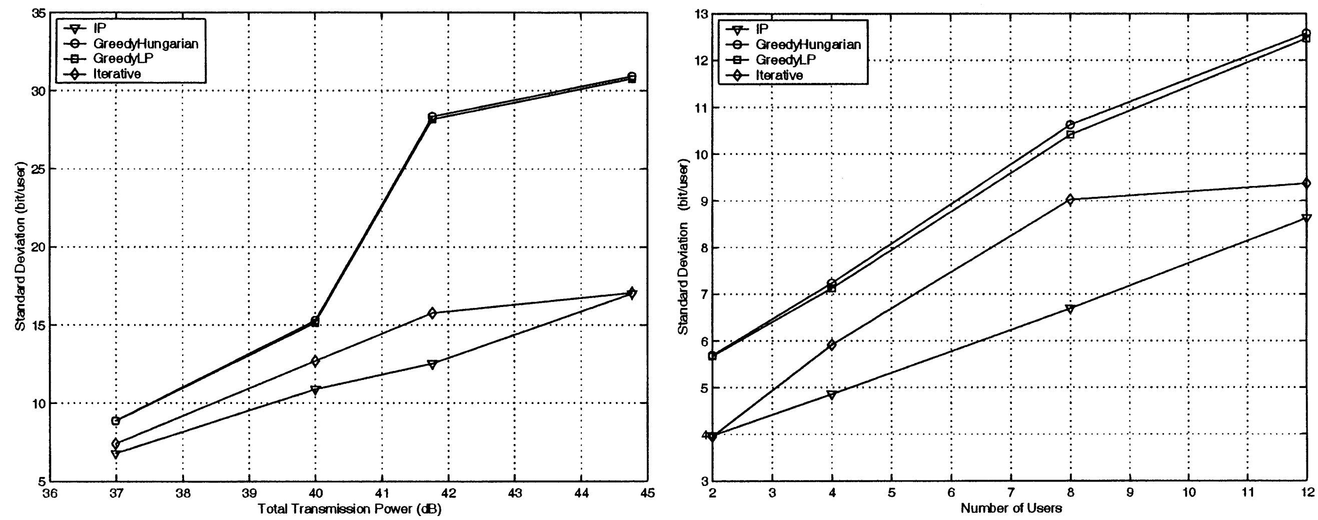
(a) (b)

Fig. 5. Average bit SNR versus channel fading and multiuser diversity; (a) average versus delay spread; (b) average bit SNR versus number of users.

presents the average bit SNR as a function of root mean square (RMS) delay spread for different resource allocation schemes. As RMS delay spread increases, the fading variation increases, so higher gains are obtained by adaptive allocation. We find that iterative approach is never more than 0.9 dB above the IP approach. Fig. 5(b) shows the average bit SNR versus the number of users where each has the same BER requirement of $1e-4$ and RMS value of 30 usec. As the number of users increases the probability of obtaining a good channel in the subcarriers increases. The iterative approach follows the lower bound within 0.9 dB and follows the GreedyHungarian and GreedyLP schemes within 0.5 dB.

Fig. 6 shows the standard deviation of the bit allocation of the users for different power constraint or different number of users. Each user has a BER requirement of $1e-4$ and total transmit rate is 480 bits/symbol which is equally distributed to each user. Each user has a 180 usec RMS delay spread. Fig. 6(a) presents the standard deviation of bit distribution among users under the total power constraint. It can be observed from the graph that iterative approach outperforms the GreedyHungarian and GreedyLP and is close to the IP. The FSA distributes the bits fairly compared to the greedy approach. The fairness property is an important metric for real time data if there is tight power control. The iterative solution maintains fairness. As the total transmit power increases, the significance of a power control scheme decreases as can be inferred from the graph. In Fig. 6(b) fairness is tested under varying number of users, the iterative approach again outperforms the GreedyHungarian and GreedyLP and closely follows the IP.

Fig. 7 shows the average data rates per subcarrier versus total power constraint when there are four users. Each user has a rate requirement of 192 bits/symbol (maximum rate) and BER requirement of $1e-4$. The performance of the iterative approach is close to that of the optimal and difference between suboptimal and iterative approaches decreases as the total transmit power increases.



(a) (b)

Fig. 6. Fair distribution of bits among users under the varying total power or number of users; (a) spectral efficiency versus total power; (b) standard deviation of bits/user versus number user.

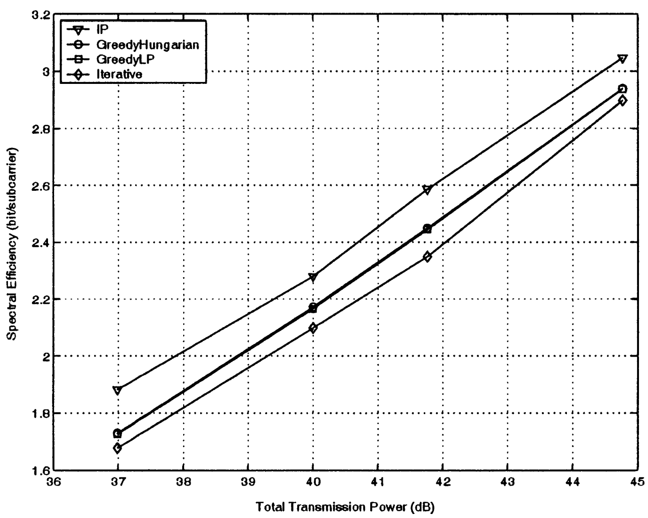


Fig. 7. Spectral efficiency versus total transmission power.

VIII. Conclusion

We considered the problem of resource allocation for adaptive modulation in OFDMA systems. The problem is considered in two different approaches: one maximizes the capacity and the other one satisfies fixed QoS criteria (i.e the rate and bit error rate requirements) in each symbol. Recent work has focused on developing algorithms to meet the QoS criteria [2], [8]–[13].

In an OFDMA system, subcarriers are distributed among users and number of bits transmitted in each subcarrier is adjusted according to the rate requirements of users to minimize total transmit power. It has been shown that resource allocation can be optimized by Integer Programming [9]. However, the optimal solution can not be implemented in real time. We proposed a simple suboptimal solution that fairly allocates and efficiently converges close to optimal meeting the QoS criteria per symbol.

The algorithm showed good performance in terms of tight power control, iterative betterment and fair scheduling among users when compared with the optimal solution and previously proposed suboptimal schemes. The proposed solution can also be applied to the uplink when there is perfect synchronization.

We also considered a possible resource allocation scheme when the objective is to maximize capacity, based on proportional fair scheduling algorithm for point-to-point communication introduced in [15].

References

[1] I. Koffman and V. Roman, "Broadband wireless access solutions based on OFDM access in IEEE 802.16," *IEEE Commun. Mag.*, Apr. 2002.

[2] I. Koutsopoulos and L. Tassiulas, "Channel state-adaptive techniques for throughput enhancement in wireless broadband networks," in *INFOCOM 2001*, vol. 2, 2001, pp. 757–766.

[3] M. Ergen and A. Puri, "MEWLANA-mobile IP enriched wireless local area network architecture," in *Proc. VTC 2002*, vol. 4, 2002, pp. 2449–2453.

[4] M. S. Corson, R. Laroia, A. O'Neill, V. Park, and G. Tsirtsis, "A new paradigm for IP-based cellular networks, IT professional," in *INFOCOM 2001*, vol. 3.

[5] *IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface*, IEEE Std 802.16-2001.

[6] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 223–229, Sept. 2002.

[7] R. Nogueroles, M. Bossert, A. Donder, and V. Zyablov, "Performance of a random OFDMA system for mobile communications," in *Proc. 1998 Int. Zurich Seminar on Broadband Communications*, 1998, pp. 37–43.

[8] 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, no. 10, pp. 1747–1758, Oct. 1999.

[9] I. Kim, H. L. Lee, B. Kim, and Y. H. Lee, "On the use of linear programming for dynamic subchannel and bit allocation in multinser OFDM," in *IEEE GLOBECOM '01*, vol. 6, 2001, pp. 3648–3652.

[10] S. Pietrzyk and G. J. M. Janssen, "Multiuser subcarrier allocation for QoS provision in the OFDMA systems," in *Proc. VTC 2002*, vol. 2, 2002, pp. 1077–1081.

[11] Y. Zhang and K. B. Letaief, "Multiuser subcarrier and bit allocation along with adaptive cell selection for OFDM transmission," in *ICC 2002*, vol. 2, 2002, pp. 861–865.

[12] 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," in *Proc. VTC 1999*, vol. 2, 1999, pp. 1124–1128.

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

[14] C. Ibars and Y. Bar-Ness, "Comparing the performance of coded multiuser OFDM and coded MC-CDMA over fading channels," in *IEEE GLOBECOM '01*, vol. 2, 2001, pp. 881–885.

[15] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Information Theory,* , vol. 48, no. 6, pp. 1277–1294, June 2002.

[16] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," in *ICC '95*, vol. 1, 1995, pp. 331–335.

[17] S. Barbarossa, M. Pompili, and G. B. Giannakis, "Time and frequency synchronization of orthogonal frequency division multiple access systems," in *ICC 2001*, vol. 6, 2001, pp. 1674–1678.

[18] D. Kivanc and H. Liu, "Subcarrier allocation and power control for OFDMA," in *Conf. Rec. 34th Asilomar Conf. Signals, Systems and Computers, 2000*, vol. 1, 2000, pp. 147–151.

[19] H. W. Khun, "The Hungarian method for the assignment problem," *Nav. Res. Logist. Quart.*, vol. 2, pp. 83–97, 1955.

[20] *Digital Video Broadcasting (DVB): Framing, Channel Coding and Modulation for Digital Terrestrial Television*, Draft ETSI EN300 744 V1.2.2 (2000-08).

[21] A. Bohdanowicz, G. J. M. Janssen, and S. Pietrzyk, "Wideband indoor and outdoor multipath channel measurements at 17 GHz," in *VTC 1999*, vol. 4, 1999, pp. 1998–2003.

Mustafa Ergen is a Ph.D. student in the Department of Electrical Engineering and Computer Science at University of California, Berkeley. He received his M.S. degree from University of California Berkeley in May 2002 and his B.S. degree from Middle East Technical University as a METU Valedictorian in June 2000. His research interests are in wireless networks and communication theory.

Sinem Coleri is a Ph.D. student in the Department of Electrical Engineering and Computer Science at University of California, Berkeley. She received her M.S. degree from the University of California Berkeley in December 2002 and her B.S. degree from Bilkent University in June 2000. Her research interests include communication theory, sensor networks, and *adhoc* networks.

Pravin Varaiya is Nortel Networks Distinguished Professor in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. From 1975 to 1992, he was also Professor of Economics at Berkeley. His research is concerned with communication networks, transportation, and hybrid systems.

He has taught at MIT and the Federal University of Rio de Janeiro, Varaiya has held a Guggenheim Fellowship and a Miller Research Professorship. He received an Honorary Doctorate from L'Institut National Polytechnique de Toulouse, and the Field Medal of the IEEE Control Systems Society. He is a Fellow of IEEE and a member of the National Academy of Engineering. He is on the editorial board of several journals, including "Discrete Event Dynamical Systems" and "Transportation Research—C". He has co-authored three books and more than 250 technical papers. The second edition of "High-Performance Communication Networks" (with Jean Walrand) was published by Morgan-Kaufmann in 2000. "Structure and interpretation of signals and systems" (with Edward Lee) was published in 2002 by Addison-Wesley.