Downlink Resource Management for Packet Transmission in OFDM Wireless Communication Systems

Jun Cai, *Student Member, IEEE*, Xuemin Shen, *Senior Member, IEEE*, and Jon W. Mark, *Life Fellow, IEEE*

***Abstract*—In this paper, an optimal downlink resource management scheme for heterogeneous packet transmission in orthogonal frequency-division multiplexing (OFDM) wireless communication systems is proposed. By making use of the channel impulse response and the properties of the OFDM physical layer, a resource management scheme is developed by integrating power distribution, subcarrier allocation, and the generalized processor sharing (GPS) scheduling. The scheme can: 1) maximize system throughput; 2) guarantee the required signal-to-noise ratio for heterogeneous traffic; 3) provide fairness to all the traffic admitted in the system; and 4) satisfy the total transmission power constraint. For practical implementation, a simplified power and subcarrier allocation algorithm, a robust H_∞ channel estimation algorithm, and a truncated GPS (TGPS) scheduling scheme are introduced. Simulation results show that the proposed resource management scheme exhibits good throughput performance.**

***Index Terms*—Generalized processor sharing, multiuser diversity, orthogonal frequency-division multiplexing, power distribution, subcarrier allocation, wireless communications.**

I. Introduction

The increasing demand for large volumes of multimedia services in wireless communication systems requires high transmission rates. However, high transmission rates may result in severer frequency-selective fading and intersymbol interference (ISI) if the bandwidth of the transmitted signal is large compared to the coherence bandwidth of the channel. Orthogonal frequency-division multiplexing (OFDM) has been proposed to combat these types of channel disturbance [1]–[4]. In an OFDM, the signal is transformed into a number of components, each with a bandwidth narrower than the coherence bandwidth of the propagation channel. Each of the OFDM signal components is modulated onto a distinct subcarrier. With OFDM, the signal fading across the bandwidth of each subcarrier is uniformly distributed, and OFDM is said to have transformed frequency-selective fading to flat fading. This feature makes OFDM an attractive multiple-access scheme for future multimedia wireless communication systems.

Manuscript received May 16, 2003; revised January 29, 2004; accepted June 1, 2004. The editor coordinating the review of this paper and approving it for publication is J. C. Hou. This work was supported by a grant from the Ministry of Energy, Science, and Technology (MEST) of the Province of Ontario, Canada, under the Ontario–Singapore Collaboration Research Programme.

The authors are with the Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Digital Object Identifier 10.1109/TWC.2005.850272

As the wireless access to the Internet becomes increasingly popular, the downlink (from the base station to the mobile users) may have to transport more traffic. However, supporting multimedia traffic, such as voice, video, and data, and making efficient use of the radio resource are very challenging tasks for the downlink OFDM wireless communication systems. This is due to the following: 1) the scarce radio resource and the limited base station transmission power; 2) the time-variant channel conditions resulting from the fading and user mobility; and 3) the diverse quality of service (QoS) requirements of multimedia users, in terms of delay, delay variance, throughput, and bit error rate. One of the promising approaches to efficiently support multimedia traffic in downlink OFDM systems is to employ a resource management scheme at the link layer which can dynamically allocate bandwidth to mobile users in accordance with the variation of traffic load and channel conditions. The resource management scheme should be efficient in utilizing the radio resources and be fair in scheduling services. By efficiency, it is meant that a user can get as much service as needed whenever there are available resources in the system. By fairness, it is meant that every user is guaranteed the agreed-upon service rate with QoS satisfaction, even though other users may be greedy in demanding bandwidth.

An ideal fair scheduling discipline is the well-known generalized processor sharing (GPS) [5]. The basic principle of GPS is to assign each user a fixed weight, instead of a fixed bandwidth, and to dynamically allocate bandwidth (or service rate) to all the users according to their weights and traffic load. With GPS, each user is guaranteed a minimum bandwidth proportional to its weight; in addition, if a user does not fully use its guaranteed bandwidth, the excess bandwidth can be distributed to other users in proportion to their weights. This results in perfect isolation of heterogeneous traffic flows and guaranteed bandwidth provision. The main drawback of GPS is that it is defined on virtual time and is not practically implementable. Several modified GPS fair scheduling schemes, such as packet GPS (PGPS), have been proposed for wireline packet networks [5], [6] and extended to wireless networks [7]–[10]. These modified GPS scheduling schemes are based on a time-scheduling approach, which entails extensive computation for the virtual time of each packet [6]. The time-scheduling approach is suitable for time division multiple access-based wireless networks [7], [8] and has been extended to code division multiple access (CDMA)-based wireless networks

[9], [10]. However, to the best of our knowledge, little research has been done on fair scheduling for packet-switched OFDM wireless systems. On the other hand, resource allocation schemes, which appropriately allocate power and transmission rates for each subcarrier, have been proposed to support multimedia traffic in single-user and multiuser OFDM systems [11]–[14]. In [15], a low-complexity power and subcarrier allocation algorithm is proposed. However, all these schemes assume persistent transmission and do not take fairness and burstiness of the traffic into account. Therefore, it is very important and challenging to develop an effective and efficient resource allocation scheme for multiuser packet-switched OFDM systems.

The effectiveness of the mechanisms used in the physical layer has a significant impact on the design and operation of upper layer protocols. For the OFDM system, its physical layer has the following properties.

1) Total bandwidth of the OFDM system is divided into many narrow bands such that information from different users can be transmitted in parallel. This feature enables the use of parallel-transmission-based scheduling schemes, for example, the GPS scheduling.
2) Different subcarriers of the same user experience different channel fading due to frequency selectivity, and channel fading experienced by different users are independent. These differences introduce the so-called multiuser diversity, the incorporation of which has been proven to offer significant capacity improvement [16].
3) Each subcarrier transmits at a fixed symbol rate. If the modulation scheme is fixed, all subcarriers have a fixed transmission bit rate.

These physical layer properties should be considered in developing resource management schemes at the link layer to determine the subcarrier and power allocation to a mobile user.

In this paper, an optimal downlink resource management scheme is proposed for heterogeneous packet transmission in OFDM wireless communication systems. By making use of the physical layer properties of the OFDM system and the channel impulse response, the resource management scheme is developed by integrating power distribution, subcarrier allocation, and GPS scheduling. The scheme can:

1) achieve maximum system throughput;
2) guarantee the required signal-to-noise ratio (SNR) for heterogeneous traffic;
3) provide fairness to all the traffic admitted in the system;
4) satisfy the total transmission power constraints.

To reduce the implementation complexity of the optimal resource management scheme, the overall optimization problem is decomposed into a hierarchy of two subproblems (scheduling and a combination of power and subcarrier allocation). Then, a distributed resource management scheme based on a simplified power and subcarrier allocation algorithm, a robust H_∞-based channel estimation algorithm, and a truncated GPS (TGPS) is introduced. Simulation results show that the proposed resource management scheme can achieve a much better performance in terms of system throughput and transmission delay than the conventional resource management scheme based on PGPS scheduling and can achieve similar fairness as GPS scheduling.

The remainder of this paper is organized as follows. In Section II, the OFDM system model is described. An optimal resource management problem by integrating power distribution, subcarrier allocation, and ideal GPS scheduling is formulated in Section III. A TGPS and a simplified power and subcarrier allocation algorithm are presented in Section IV. The H_∞ channel estimator for channel impulse response estimation is adopted in Section V. Simulation results are given in Section VI to demonstrate the performance of the proposed resource management scheme in terms of the system throughput and transmission delay under the homogeneous and heterogeneous traffic, respectively. Conclusions are given in Section VII.

II. System Model

Fig. 1 shows the structure of a downlink OFDM to support N users. At the base station transmitter, the serial data sequence from the scheduler, which is a sequence of samples occurring at interval T_s, is first serial-to-parallel (S/P) converted into M low-rate parallel streams to increase the symbol duration to $T = MT_s$. The low-rate streams can be represented by the symbols $b_m[k]$, where $m = 0, 1, \ldots, M-1$, $k = 1, 2, \ldots$. Through scheduling, $b_m[k]$ for different m may come from the same user or from different users. $b_m[k]$ for some m may be equal to zero, which means no transmission on these subcarriers at the kth epoch. In order to eliminate interference between parallel data streams, each of the low-rate data streams is modulated onto a distinct subcarrier belonging to an orthogonal set with subcarrier spacing $1/T$. The parallel streams are then multiplexed and a cyclic prefix is added to eliminate the effect of ISI. Thus, the signal transmitted during the kth epoch $y(t)$ can be written as

$$y(t) = \sum_{m=0}^{M-1} \sqrt{P_m[k]} b_m[k] e^{\frac{j2\pi mt}{T}}, \qquad -\rho + kT \le t \le (k+1)T \quad (1)$$

where $P_m[k]$ is the transmission power of the kth data symbol in the mth stream, M is the total number of subcarriers, and ρ is the length of the guard interval.

The transmitted signal $y(t)$ passes through the wireless channel that introduces signal distortion and additive noise. For each user, the wireless channel can be modeled as a multipath frequency-selective fading channel using a tapped delay line with time-varying coefficients and fixed tap spacing [17]. For user i, the channel impulse response can be represented as

$$h^i(t,\tau) = \sum_{l=0}^{\chi^i} h_l^i(t)\delta\left(\tau - \tau_l^i\right), \qquad i = 0, 1, \ldots, N-1 \quad (2)$$

where $h_l^i(t)$ and τ_l^i are the complex amplitude and the delay of the lth path, respectively. $\chi^i + 1$ is the total number of taps and $\tau_{\chi^i}^i$ represents the maximum multipath delay spread. For OFDM to be effective, the length of the cyclic prefix ρ should be

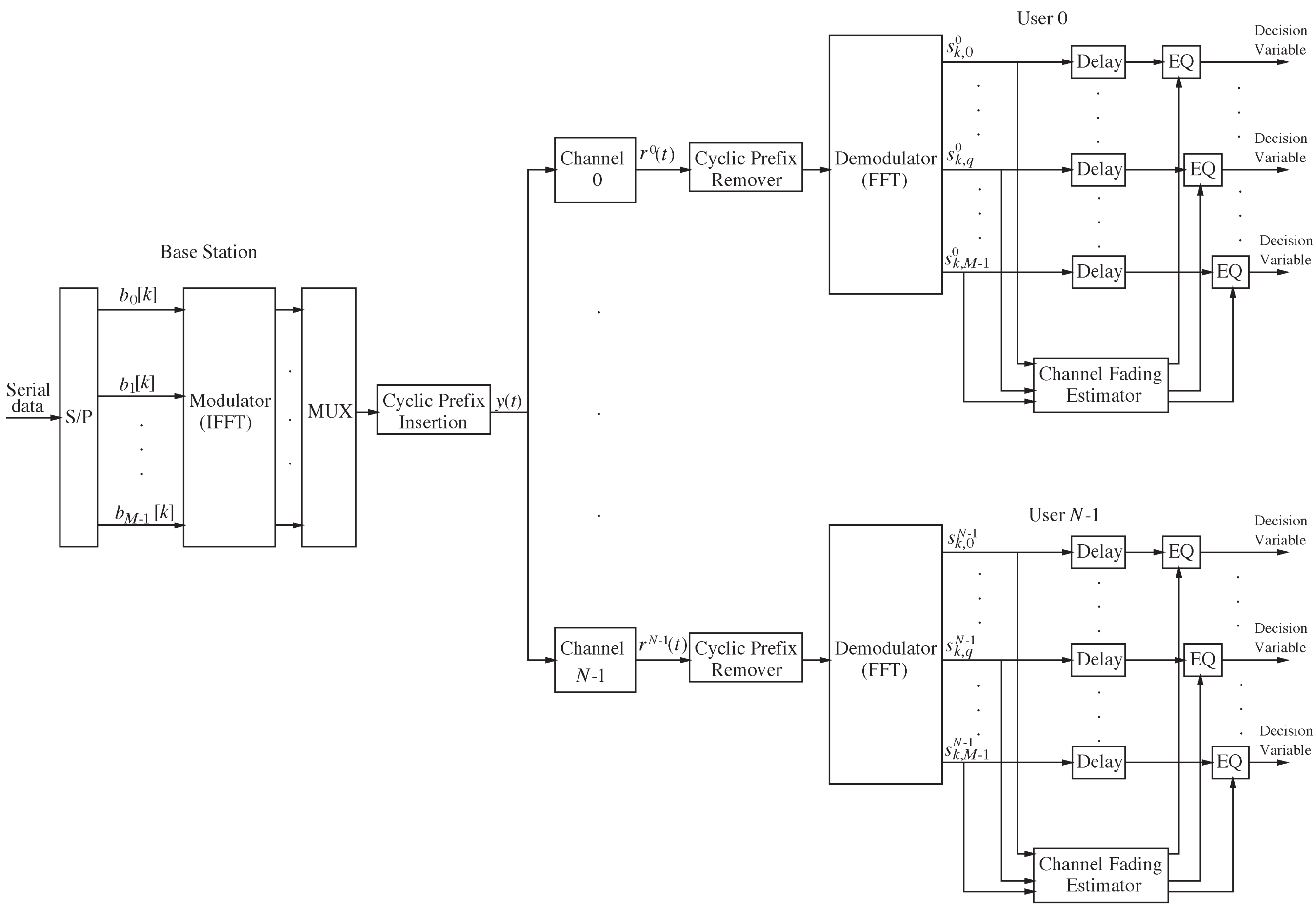

Fig. 1. Transceiver structure of the OFDM system.

larger than the maximum multipath delay spread of the channel. $h^i_l(t)$ can be modeled as a wide-sense stationary uncorrelated scattering process, with correlation function [18]

$$\begin{aligned}\phi_{h^i_l}(\Delta t) &\triangleq E\left\{h^i_l(t)\left[h^i_l(t-\Delta t)\right]^*\right\}\\ &= \left(\sigma^i_l\right)^2 \phi_{\mathrm{t}}(\Delta t) \qquad (3)\end{aligned}$$

where $*$ denotes complex conjugation, $(\sigma^i_l)^2$ is the variance of the channel fading at path l of the ith user, which is determined by the power delay profile of the channel and satisfies the normalized condition $\sum_{l=0}^{\chi^i}(\sigma^i_l)^2 = 1$, and $\phi_{\mathrm{t}}(\Delta t)$ is the normalized correlation function.

The received signal of the ith user $r^i(t)$ can be expressed as

$$\begin{aligned}r^i(t) &= \int h^i(t,\tau)y(t-\tau)\mathrm{d}\tau + n^i(t)\\ &= \sum_{l=0}^{\chi^i} h^i_l(t)y\left(t-\tau^i_l\right) + n^i(t) \qquad (4)\end{aligned}$$

where $n^i(t)$ is an independent identically distributed (i.i.d.) complex Gaussian background noise for any i.

At the ith user's receiver, the received signal is first demodulated after cyclic prefix removal to obtain the output of each subcarrier. For practical implementation, modulation and demodulation can be achieved by inverse fast Fourier transform (IFFT) and fast Fourier transform (FFT), respectively. In Fig. 1, the second index q at the output of the demodulator refers to the qth subcarrier, and $s^i_{k,q}$ is the output for the qth subcarrier in the kth epoch. Assume that the channel impulse response is quasi-static during one symbol interval, i.e., during the kth epoch, $h^i(t) \approx h^i(kT)$, for $kT \le t < (k+1)T$. Under this condition, the intercarrier interference (ICI) is negligibly small compared to the background noise. Thus, the qth subcarrier output $s^i_{k,q}$, $q \in \{0, 1, \ldots, M-1\}$, from the demodulator can be expressed as

$$\begin{aligned}s^i_{k,q} &= \frac{1}{T}\int_{kT}^{(k+1)T}\left[\sum_{l=0}^{\chi^i} h^i_l(kT)\sum_{m=0}^{M-1}\sqrt{P_m[k]}b_m[k] \times \mathrm{e}^{\frac{j2\pi m\left(t-\tau^i_l\right)}{T}} + n^i(t)\right]\mathrm{e}^{-\frac{j2\pi qt}{T}}\mathrm{d}t\\ &= \frac{1}{T}\sum_{l=0}^{\chi^i} h^i_l(kT)\sum_{m=0}^{M-1}\sqrt{P_m[k]}b_m[k]\mathrm{e}^{-\frac{j2\pi m\tau^i_l}{T}}\\ &\quad\times\int_{kT}^{(k+1)T}\mathrm{e}^{\frac{j2\pi(m-q)t}{T}}\mathrm{d}t + \frac{1}{T}\int_{kT}^{(k+1)T} n^i(t)\mathrm{e}^{-\frac{j2\pi qt}{T}}\mathrm{d}t\\ &= \sqrt{P_q[k]}b_q[k]H^i_{k,q} + v^i_{k,q} \qquad (5)\end{aligned}$$

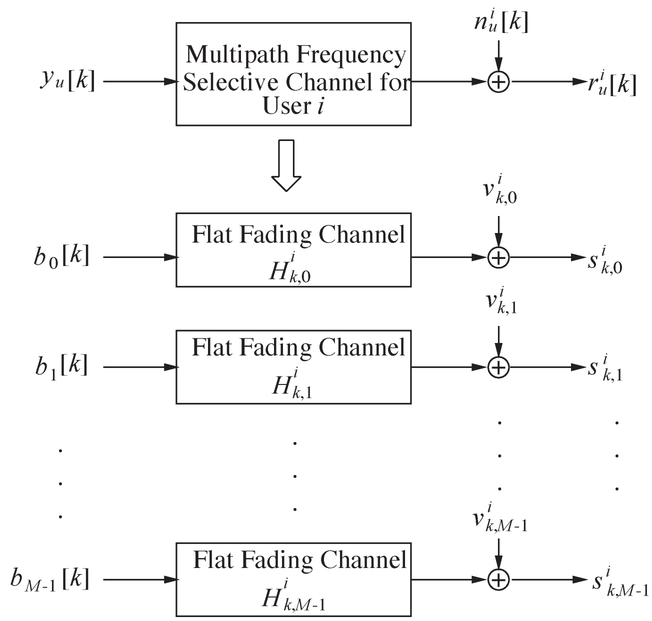

Fig. 2. Equivalent channel model for user i.

where

$$H_{k,q}^i = \sum_{l=0}^{\chi^i} h_l^i(kT)\mathrm{e}^{-\frac{j2\pi q\tau_l^i}{T}}$$

$$v_{k,q}^i = \frac{1}{T}\int_{kT}^{(k+1)T} n^i(t)\mathrm{e}^{-\frac{j2\pi qt}{T}}\mathrm{d}t. \qquad (6)$$

$v_{k,q}^i$ is an i.i.d. complex Gaussian random variable with zero mean and variance σ_v^2 for any i, q, and k. After demodulation, each user selects the appropriate subcarriers that contain the desired signals for detection.

From (5) and (6), the following can be observed. 1) For each OFDM downlink, a wideband multipath frequency-selective fading channel can be modeled as M separate narrowband flat fading channels (Fig. 2), each of which has a time-varying multiplicative complex channel fading gain $H_{k,q}^i$. This indicates that the OFDM system possesses the capability to support parallel transmission so that some parallel-transmission-based scheduling schemes, such as GPS scheduling, can be applied in OFDM downlinks. Moreover, parallel transmission in the frequency domain provides an opportunity to apply a more flexible resource allocation for the OFDM system. Since each subcarrier experiences flat fading, the channel gain can be defined as

$$\alpha_{k,q}^i = \left|H_{k,q}^i\right|^2, \qquad q = 0, 1, \ldots, M-1. \qquad (7)$$

2) The frequency selectivity of the channel fading results in nonuniformly distributed channel gains among all subcarriers, i.e., for a given user i, $\alpha_{k,q}^i$ may be different for a different subcarrier index q. When the distance between two subcarriers is larger than the coherence bandwidth of the channel, the channel gains of these two subcarriers may be considered independent. Furthermore, since each user experiences i.i.d. channel fading, $\alpha_{k,q}^i$ is independent for different i, $i = 0, 1, \ldots, N-1$. The difference in channel gains has the following impacts on the resource allocation. a) Multiuser diversity exists, where the diversity gain can be achieved by appropriate resource allocation at the link layer. Therefore, multiuser diversity should be taken into account. b) Since the channel gains are different for different subcarriers and different users, in addition to determining the number of subcarriers allocated to each user, resource allocation should also determine which set of subcarriers should be allocated to the given user in order to achieve maximum system throughput. 3) With the transmission symbol rate fixed for each subcarrier, transmission bit rate can only be changed by changing the modulation level or the number of subcarriers allocated. This is completely different from the CDMA system where the symbol duration, or the symbol rate, can be set dynamically. In addition, in an OFDM system, a symbol is the minimum unit for transmission. All these properties at the physical layer of the OFDM system introduce new requirements and features for designing resource management schemes at the link layer. In other words, efficient and effective resource management schemes for the OFDM wireless communication systems can be developed by making full use of these properties.

III. Optimal Resource Management

In this section, an optimal resource management scheme with ideal GPS scheduling to maximize OFDM system throughput while satisfying the total transmission power constraint and heterogeneous SNR requirements is developed. Since the discussion focuses on each OFDM symbol interval, in what follows, the symbol index k will also be used as the time index to denote the epoch corresponding to the symbol interval.

A. GPS Scheduling

GPS has been shown to offer fair scheduling [5]. GPS is ideal in the sense that it is designed in virtual time and hence not implementable. Consider N sessions[1] sharing a network link with a total link transmission rate Ψ. Each session i is associated with a positive weight η_i. Let $W_i(\tau,t)$ be the amount of session i traffic served during the interval $(\tau,t]$. Then, a GPS server is defined as a work-conserving service discipline for which

$$\frac{W_i(\tau,t)}{W_j(\tau,t)} \geq \frac{\eta_i}{\eta_j}, \qquad j = 0, 1, \ldots, N-1 \qquad (8)$$

for any session i that is continuously backlogged in the interval $(\tau,t]$ [5]. If both session i and session j are continuously backlogged in the interval $(\tau,t]$, then (8) holds with equality. It can be shown that in a GPS server, any backlogged session i is guaranteed a service rate R_i given by

$$R_i = \frac{\eta_i}{\sum_{j=0}^{N-1}\eta_j}\Psi. \qquad (9)$$

In fact, R_i is the service rate of session i when all sessions are busy. Whenever fewer than N sessions are active, the system

[1]Here, the terms session and user are used interchangeably, as these terms are being used quite freely in the literature.

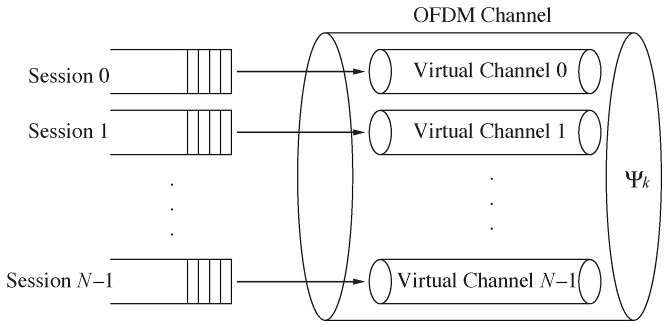

Fig. 3. A queuing model for the OFDM system.

resources (Ψ) will be shared among those busy sessions, and hence each busy session will be served at a rate greater than its guaranteed service rate.

There are several interesting features associated with GPS that make it an attractive scheduling scheme for integrated services networks. A significant feature is the flexibility to treat different classes of service differently according to their QoS requirements. Such flexibility is achieved by assigning different η_is to various classes of service. This in turn results in dissimilar service rates, and hence different performance specifications in terms of throughput and/or delay [9].

B. Optimal Resource Management

Consider an OFDM system with an infinitely large number of subcarriers such that the subcarrier (or bandwidth) allocation can be carried out at any small frequency band. Assume a fixed modulation level for all subcarriers. Then, the number of subcarriers or the system bandwidth completely defines the system transmission rate. Therefore, the considered OFDM system can be modeled as a queuing system, as depicted in Fig. 3, where each virtual channel is a combination of the subcarriers allocated to the given session and Ψ_k denotes the effective system bandwidth, defined as the bandwidth that the system can support at the kth epoch. Fig. 3 obviously indicates that the considered OFDM system can support both parallel transmission and fine granularity in resource allocation such that ideal GPS scheduling can be applied to achieve optimal fairness performance. Although in a wireless system Ψ_k is time variant, which is a consequence of the channel condition and the resource allocation algorithm applied, (8) and (9) still hold for time-variant Ψ_k since all the sessions share the bandwidth in parallel. The subcarrier allocation functions, each of which represents the subcarrier distribution of the given user over the whole bandwidth, can be defined as a continuous function in the frequency domain since the number of subcarriers M is infinite. Let $\{A_k^i(f), i = 0, 1, \ldots, N-1\}$, $f \in [0, W)$, denote the subcarrier allocation functions, where W is the total bandwidth of the system. $A_k^i(f)$ is an indicator function, i.e.,

$$A_k^i(f) = \begin{cases} 1, & \text{frequency } f \text{ allocated to session } i \\ 0, & \text{otherwise.} \end{cases}$$

The objective of resource management is to maximize the normalized system bandwidth utilization or the normalized throughput, defined as Ψ_k/W, subject to the constraints on total transmission power, users' SNR requirements, and satisfaction of GPS fair scheduling. Let the set of backlogged users be $\mathbf{\Lambda}$. Optimal resource management in the OFDM system can be described as follows:

$$\begin{cases} \max_{\{A_k^i(f)\}} \frac{\Psi_k}{W} = \max_{\{A_k^i(f)\}} \sum_{i\in\mathbf{\Lambda}} \int_0^W \frac{A_k^i(f)}{W} \mathrm{d}f \\ \text{subject to} \\ \sum_{i\in\mathbf{\Lambda}} \int_0^W \frac{\mathrm{SNR}_i}{\alpha_{k,f}^i} A_k^i(f)\mathrm{d}f \le \frac{P}{\sigma_v^2} \qquad \text{(Condition I)} \\ A_k^i(f)A_k^j(f) = 0 \\ \qquad \text{for any } i,j \in \mathbf{\Lambda} \text{ and } i \ne j \quad \text{(Condition II)} \\ \frac{\int_0^W A_k^i(f)\mathrm{d}f}{\int_0^W A_k^j(f)\mathrm{d}f} = \frac{\eta_i}{\eta_j}, \qquad i,j\in\mathbf{\Lambda} \qquad \text{(Condition III)} \end{cases} \tag{10}$$

where P is the base station target transmission power and $\{\mathrm{SNR}_i, i = 0, 1, \ldots, N-1\}$ is the SNR requirement of user i. In (10), Condition I means that the total transmission power should be less than the target transmission power while satisfying all users' SNR requirements; Condition II means that no more than one user transmit in the same subcarrier; and Condition III describes the requirement of the GPS scheduling. Since the noise power σ_v^2 is a constant, it has no effect on the solution of the optimization problem in (10). Without loss of generality, in what follows, σ_v^2 is normalized to 1 for simplicity.

If the optimal solutions of (10) could be obtained, they would act as the performance benchmark. Unfortunately, solving (10) is too time consuming for a practical system. In addition, both the ideal GPS scheduling, which is based on the fluid-flow model (i.e., the flow generated from each session is assumed to be a continuous flow), and the OFDM system with an infinite number of subcarriers are difficult to implement in practice. As a way to render the resource management problem implementable, a truncation of the ideal GPS is proposed in the next section.

IV. Practically Implementable Scheduling and Resource Management Schemes

In this section, a TGPS to facilitate implementation of the resource management scheme is proposed.

A. TGPS Scheduling

Application of GPS scheduling to the OFDM channel requires M to be infinitely large, or the width of the subbands to be infinitely small. For practical implementation, the number of subcarriers has to be finite. It is assumed that, on the basis that the width of the subbands is sufficiently small, a finite number of subcarriers can yield a relative good system performance.

Although GPS scheduling provides optimal fairness performance, it is not practically implementable for the practical OFDM systems with M subcarriers due to the following two aspects. 1) For practical OFDM systems, the total bandwidth is quantized to M subbands. This results in a limitation on the maximum number of sessions which can be transmitted simultaneously, i.e., the maximum number of parallel transmissions must be no greater than M. When the target transmission power is taken into account, the number of effective subcarriers, defined as the number of subcarriers that can be actually supported by the system, is even less than M due to severe channel fading. Thus, at certain instants when the total number of backlogged sessions is larger than the number of effective subcarriers, some backlogged sessions cannot obtain the bandwidth which should be guaranteed with ideal GPS scheduling. 2) A side effect of the quantization is that in the OFDM system, the minimal transmission unit is a symbol. Even when the total number of backlogged sessions is less than the number of effective subcarriers, the scheduling scheme may still not be able to guarantee the capacity allocation according to the predefined weights $\{\eta_i, i \in \mathbf{\Lambda}\}$, because the required capacity may not be an integer number of subcarriers. Under the GPS approach, the effective number of subcarriers for session i at time epoch k is a real number. To facilitate implementation of GPS in a practical OFDM system, we convert each user's effective number of subcarriers to an integer value by truncation and refer to the modified scheme as TGPS.

Consider a busy period beginning at time epoch $k = 0$. Let the set of all backlogged sessions be $\mathbf{\Lambda}$. Given the number of effective subcarriers $M^k_{\text{effective}}$ at any epoch k in a busy period, the effective number of subcarriers allocated to user i is $(\eta_i / \sum_{j\in\mathbf{\Lambda}} \eta_j) M^k_{\text{effective}}$, which is converted to an integer value by truncation

$$M_i^k = \left\lfloor \frac{\eta_i}{\sum_{j\in\mathbf{\Lambda}} \eta_j} M^k_{\text{effective}} \right\rfloor \qquad \forall i \in \mathbf{\Lambda} \tag{11}$$

where M_i^k denotes the number of subcarriers allocated to the ith session and $\lfloor x \rfloor$ denotes the floor function that takes the maximum integer less than or equal to x. Define the capacity loss of each backlogged session $\{\varepsilon_i^k, i \in \mathbf{\Lambda}\}$ as the capacity difference between the ideal GPS scheduling and M_i^k

$$\varepsilon_i^k = \frac{\eta_i}{\sum_{j\in\mathbf{\Lambda}} \eta_j} M^k_{\text{effective}} - M_i^k, \qquad i \in \mathbf{\Lambda}. \tag{12}$$

Since ε_i^k results from the quantization of the OFDM system in the frequency domain, using the same terminology as in amplitude quantization, ε_i^k is called the quantization error. Define the cumulative quantization error $\{\Delta M_i^k, i \in \mathbf{\Lambda}\}$ as the quantization error accumulated from epoch 0 to epoch k, where $\Delta M_i^0 = 0$. From (12), the cumulative quantization error can be obtained as

$$\Delta M_i^k = \Delta M_i^{k-1} + \varepsilon_i^k, \qquad i \in \mathbf{\Lambda}. \tag{13}$$

After the allocation from (11), there may be some subcarriers that are not allocated. Define the number of unallocated subcarriers as

$$M^k_{\text{left}} = M^k_{\text{effective}} - \sum_{i\in\mathbf{\Lambda}} M_i^k. \tag{14}$$

According to the work-conserving principle, the unallocated subcarriers need to be allocated to some sessions. Since the cumulative quantization errors reflect the total difference between the proposed TGPS scheduling and the ideal GPS scheduling, in order to minimize the difference, a compensation algorithm is introduced to allocate the unallocated subcarriers to the sessions with maximum cumulative quantization errors. The compensation algorithm in the kth epoch can be described as follows.

1) Select the session i that has maximum cumulative quantization error

$$\Delta M_i^k = \max_j \left\{\Delta M_j^k, j \in \mathbf{\Lambda}\right\}. \tag{15}$$

2) Allocate one unallocated subcarrier to session i and carry out the following upgrades

$$\begin{aligned} \Delta M_i^k &= \Delta M_i^k - 1 \\ M^k_{\text{left}} &= M^k_{\text{left}} - 1. \end{aligned} \tag{16}$$

3) Repeat steps 1 and 2 until $M^k_{\text{left}} = 0$.

During a busy period, in any epoch k, the proposed TGPS scheduling allocates the effective subcarriers to each backlogged session, calculates the cumulative quantization errors, and distributes the unallocated subcarriers using (11), (13), and (16) until one session completes its packet transmission and departs from the set $\mathbf{\Lambda}$. In that epoch, the session's cumulative quantization error is set to zero. The reason is that when the session completes the transmission of all its packets, the cumulative quantization error has been translated into the transmission delay of the packets. The error should not be reconsidered for future transmissions.

From the description of the proposed TGPS scheduling (11)–(16), we make the following observations.

1) Given the effective system capacity at the kth epoch $M^k_{\text{effective}}$, each backlogged session can be guaranteed the capacity allocation $\lfloor (\eta_i / \sum_{j\in\mathbf{\Lambda}} \eta_j) M^k_{\text{effective}} \rfloor$, which is a function of η_i, $i \in \mathbf{\Lambda}$. Compared with the ideal GPS scheduling, the difference of the guaranteed capacity allocation equals the cumulative quantization error $\Delta M_i^k \; \forall i \in \mathbf{\Lambda}$. The guarantee on capacity allocation indicates that the proposed TGPS can provide fairness among sessions, which is determined by the predefined weights $\{\eta_i, i \in \mathbf{\Lambda}\}$.
2) If $(\eta_i / \sum_{j\in\mathbf{\Lambda}} \eta_j) M^k_{\text{effective}}$ were an integer for all $i \in \mathbf{\Lambda}$, then TGPS would achieve the same fairness as the ideal GPS scheduling.
3) Given the total system bandwidth W, the quantization errors are determined by the number of the quantization levels, i.e., the total number of subcarriers M in the frequency domain. When $M \to \infty$, the TGPS scheduling approaches the ideal GPS.
4) The compensation algorithm (15) and (16) can compensate the quantization errors with the capacity allocation. The minimum compensation unit is one subcarrier.

B. *Power and Subcarrier Allocation Algorithm*

The information $M_i^k, i \in \mathbf{\Lambda}$, generated by the TGPS scheduling, is actually a function of the parameter $M^k_{\text{effective}}$. It can be determined by the power and subcarrier allocation algorithm, which is designed to determine which set of subcarriers should be allocated to the given user according to the information M_i^k so that the total transmission power is minimized.

Given the number of subcarriers required by each user to transmit at the current symbol duration M_i^k, $i \in \mathbf{\Lambda}$, the optimal power and subcarrier allocation can be derived from (10) as

$$\begin{cases} \min_{A_k^i(q)} \sum_{i\in\mathbf{\Lambda}} \sum_{q=0}^{M-1} \frac{\mathrm{SNR}_i}{\alpha_{k,q}^i} A_k^i(q) \\ \text{subject to} \\ \sum_{q=0}^{M-1} A_k^i(q) = M_i^k, \qquad i \in \mathbf{\Lambda} \\ \sum_{i\in\mathbf{\Lambda}} A_k^i(q) \le 1, \qquad q = 0, 1, \ldots, M-1. \end{cases} \tag{17}$$

Different from (10), the subcarrier allocation function $A_k^i(q)$, $i \in \mathbf{\Lambda}$, is a discrete two-valued function defined on $q \in \{0, 1, \ldots, M-1\}$. Although the optimization problem defined in (17) can be solved using linear programming, since a fast allocation algorithm is required for a practical OFDM system, a simplified power and subcarrier allocation algorithm for solving (17) is proposed based on the fact that the users with higher SNR requirement will dominate the power consumption. This can be explained as follows.

Consider an OFDM system with only two users, user 1 and user 2, having SNR requirements of SNR_1 and SNR_2, respectively. Assume $\mathrm{SNR}_2 > \mathrm{SNR}_1$. If subcarrier q is allocated to user 2, from the minimum power consumption criterion, it requires that

$$\frac{\mathrm{SNR}_1}{\alpha_q^1} > \frac{\mathrm{SNR}_2}{\alpha_q^2} \Longrightarrow \frac{\mathrm{SNR}_2}{\mathrm{SNR}_1} < \frac{\alpha_q^2}{\alpha_q^1}. \tag{18}$$

Equation (18) indicates that only when the channel gain of the user with high SNR requirement is more than $\mathrm{SNR}_2/\mathrm{SNR}_1$ times larger than that of the user with lower SNR requirement, allocating the subcarrier to the user with higher SNR requirement can achieve the minimum transmission power. In addition, the power consumption at a given subcarrier is determined by the function $g(\alpha) = \mathrm{SNR}/\alpha$. Its derivative with respect to α is $-\mathrm{SNR}/\alpha^2$. Therefore, increasing the channel gain will decrease more transmission power for the user with high SNR requirement than the user with a lower one since the function SNR/α has a larger absolute value of derivative at any given α for the user with higher SNR. Therefore, the user with high SNR requirement has a greater effect on the power consumption and should get a higher priority to occupy the subcarriers with the largest channel gains.

The simplified power and subcarrier allocation algorithm is presented as follows.

1) During any epoch k, the power and subcarrier allocation algorithm accepts the information $\{M_i^k, i \in \mathbf{\Lambda}\}$, the required number of subcarriers allocated to user i, from the TGPS scheduling.
2) The power and subcarrier allocation algorithm sorts the users in a nonincreasing order according to their SNR requirements. Then, the algorithm allocates the subcarriers to the users one by one.
3) At the jth step of subcarrier allocation, let Ω^j be the set of subcarriers that have not been allocated till the jth step and user l be the user with the maximum SNR requirement among the unallocated users. The power and subcarrier allocation algorithm selects M_l^k subcarriers with maximum channel gains from the set Ω^j and allocates them to user l.
4) After the subcarrier allocation, the algorithm calculates the power needed for each subcarrier.

C. Distributed Resource Management Scheme

1) The base station establishes one individual queue for each user. The arriving packet from each user is stored in its own individual queue in a first-in first-out order. At the beginning of each OFDM symbol interval k, set $M_{\text{effective}}^k = M$.
2) The base station operates TGPS scheduling to determine the number of subcarriers allocated to each backlogged user M_i^k, $i \in \mathbf{\Lambda}$, based on the given $M_{\text{effective}}^k$. The information M_i^k is transferred to the power and subcarrier allocation algorithm.
3) The base station operates the power and subcarrier allocation algorithm based on the information $\{M_i^k, i \in \mathbf{\Lambda}\}$ to determine the subcarrier distribution of each backlogged user and calculates the transmission power for each subcarrier.
4) The base station calculates the total transmission power required and compares it with the base station target transmission power. If the total power consumption is less than the target transmission power, the resource management procedure is complete.
5) Otherwise, the base station decreases $M_{\text{effective}}^k$ by 1 and repeats steps 2–4.
6) The base station selects M_i^k head symbols from the ith individual queue, $i \in \mathbf{\Lambda}$, and transmits them on the allocated subcarriers with the allocated transmission power.

V. Channel Estimation

The resource management schemes developed in the previous sections require the channel impulse response of the OFDM system. In this section, we present a robust estimation of the channel impulse response using an H_∞ approach.

In order to estimate the channel impulse response, the base station inserts some pilots to certain subcarriers at given symbol intervals. At the receiver end, each user estimates the individual channel impulse response based on these pilots using the pilot-assisted estimation algorithms and feeds back the estimates to the base station through some uplink (from the mobile users to the base station) signaling channels. The base station then carries out the proposed resource management schemes for the next symbol interval using the feedback channel impulse response. Here, no feedback errors are considered. In order to reduce the number of information fed back from the mobile users, which consists of the overhead of the system, only the

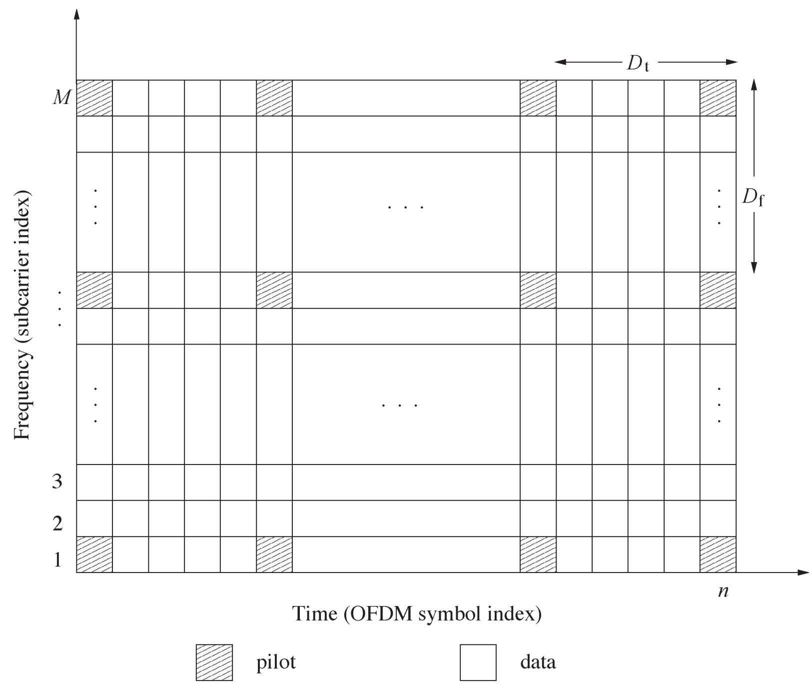

Fig. 4. Configuration of pilot arrangement.

channel estimates at the pilot positions need to be fed back to the base station, in which interpolation is carried out to obtain the channel estimates at the nonpilot subcarriers. Since the system is working under a slow fading channel, i.e., the channel fading is time invariant over a few OFDM symbols, the small feedback delay inherent in the OFDM system [21] should have little effect on the resource management performance.

Consider the pilot pattern shown in Fig. 4, where the known pilots are inserted in every $D_{\rm t}$ OFDM symbols and $D_{\rm f}$ subcarriers at the base station. In general, the values of $D_{\rm t}$ and $D_{\rm f}$ may significantly affect the estimation performance and they should be selected properly [22]–[24]. Since all users implement the same channel estimation algorithm, in the following, the user index i is omitted for simplicity. Without loss of generality, let

$$\sqrt{P_q[k]}b_q[k] = \beta, \qquad k \in \theta_{\rm t},\ q \in \theta_{\rm f}$$

where β is a constant; $\theta_{\rm t}$ and $\theta_{\rm f}$ are the sets of pilots in the time and frequency domains, respectively. Then, (5) becomes

$$\tilde{s}_{k,q} = H_{k,q} + \tilde{v}_{k,q}, \qquad k \in \theta_{\rm t},\ q \in \theta_{\rm f} \tag{19}$$

where $\tilde{s}_{k,q} = s_{k,q}/\beta$ and $\tilde{v}_{k,q} = v_{k,q}/\beta$. The channel estimation at each mobile user is to estimate $H_{k,q}$ based on the observations $\tilde{s}_{k,q}$.

From (3) and (6), the correlation function of the channel fading $\{H_{k,q},\ k \in \theta_{\rm t}$ and $q \in \theta_{\rm f}\}$ for different instants and subcarriers can be written as [19]

$$\begin{aligned}\phi_H[m,n] &\triangleq E\left[H_{k,q}H^*_{k-mD_{\rm t},q-nD_{\rm f}}\right]\\ &= \phi_{\rm t}[m]\phi_{\rm f}[n]\end{aligned} \tag{20}$$

where

$$\phi_{\rm t}[m] \triangleq \phi_{\rm t}(mD_{\rm t}T)$$
$$\phi_{\rm f}[n] \triangleq \sum_{l=0}^{\chi} \sigma_l^2 {\rm e}^{-\frac{j2\pi n D_{\rm f}\tau_l}{T}}.$$

Let

$$\boldsymbol{\phi}_{\rm f}[q] = \begin{bmatrix}\phi_{\rm f}[q]\\ \vdots\\ \phi_{\rm f}[0]\\ \vdots\\ \phi_{\rm f}[-K+1+q]\end{bmatrix}$$
$$\boldsymbol{\Phi}_{\rm f} = [\,\boldsymbol{\phi}_{\rm f}[0] \quad \boldsymbol{\phi}_{\rm f}[1] \quad \ldots \quad \boldsymbol{\phi}_{\rm f}[K-1]\,]$$

where $K = M/D_{\rm f}$ is the number of pilots in the frequency domain for any given $k \in \theta_{\rm t}$. Assuming that the correlation matrix $\boldsymbol{\Phi}_{\rm f}$ is diagonalizable, the eigendecomposition of $\boldsymbol{\Phi}_{\rm f}$ is

$$\boldsymbol{\Phi}_{\rm f} = \mathbf{U}^{\rm H}\mathbf{D}\mathbf{U} \tag{21}$$

where the superscript H denotes Hermitian transposition, $\mathbf{U}$ is a unitary matrix consisting of the eigenvectors of $\boldsymbol{\Phi}_{\rm f}$, and $\mathbf{D}$ is a $K \times K$ diagonal matrix with the diagonals consisting of $K_0(K_0 \le K)$ nonzero eigenvalues d_l, $l = 0, 1, \ldots, K_0 - 1$, and $K - K_0$ zeros. Let

$$\bar{\mathbf{s}}_k = [\,\tilde{s}_{k,0} \quad \tilde{s}_{k,D_{\rm f}} \quad \ldots \quad \tilde{s}_{k,(K-1)D_{\rm f}}\,]\,\mathbf{U}^{\rm H}$$
$$\mathbf{g}_k = [\,H_{k,0} \quad H_{k,D_f} \quad \ldots \quad H_{k,(K-1)D_{\rm f}}\,]\,\mathbf{U}^{\rm H}$$
$$\bar{\mathbf{v}}_k = [\,\tilde{v}_{k,0} \quad \tilde{v}_{k,D_{\rm f}} \quad \ldots \quad \tilde{v}_{k,(K-1)D_{\rm f}}\,]\,\mathbf{U}^{\rm H}.$$

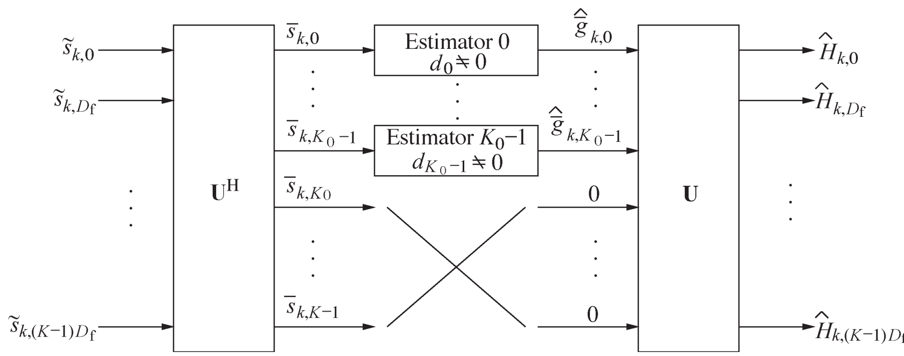

Fig. 5. Joint time–frequency channel estimator for OFDM system.

From (19) and (20), we have

$$\begin{cases} \bar{s}_{k,l} = g_{k,l} + \bar{v}_{k,l}, \qquad l = 0, 1, \ldots, K_0 - 1 \\ \phi_{g,l}[m] \triangleq E\left[g_{k,l} g^*_{k-mD_\mathrm{t},l}\right] = \phi_\mathrm{t}[m]\mathbf{U}_l \mathbf{\Phi}_\mathrm{f} \mathbf{U}_l^\mathrm{H} = d_l \phi_\mathrm{t}[m] \\ \sigma^2_{\bar{v}} \triangleq E\left[|\bar{v}_{k,l}|^2\right] = E\left[|\tilde{v}_{k,l}|^2\right] = \sigma^2_v \end{cases} \tag{22}$$

where $\bar{s}_{k,l}$, $g_{k,l}$, and $\bar{v}_{k,l}$ are the lth elements of $\bar{\mathbf{s}}_k$, $\mathbf{g}_k$, and $\bar{\mathbf{v}}_k$, respectively; $\mathbf{U}_l$ is the lth column of $\mathbf{U}$. Since the columns of $\mathbf{U}$ form a unitary system, i.e., $E[g_{k,l} g^*_{k,l-i}] = \mathbf{U}_l \mathbf{\Phi}_\mathrm{f} \mathbf{U}^\mathrm{H}_{l-i} = d_l \mathbf{U}_l \mathbf{U}^\mathrm{H}_{l-i} = 0$, for $i \neq 0$, from (19)–(22), we can obtain a channel estimator structure shown in Fig. 5, where the joint time–frequency estimation problem for $H_{k,q}$ is transformed to K_0 one-dimensional time-domain estimation problems for $g_{k,l}$.

For the slow-fading channel $g_{k,l}$, it can be approximated by an autoregressive process [25], [26]

$$g_{k,l} = \sum_{i=1}^{n} a_{i,l} g_{k-iD_\mathrm{t},l} + w_{k,l} \tag{23}$$

where n, $a_{i,l}$, and $w_{k,l}$ denote the order, the coefficient (tap-gain parameter), and the model noise, respectively. Without loss of generality, let the zeroth time-domain estimator be the reference and drop off the second index l. Combining (22) and (23), the one-dimensional time-domain channel estimation problem can be formulated by the following state-space model:

$$\mathbf{X}_k = \mathbf{A}\mathbf{X}_{k-1} + \mathbf{B}w_k \quad \text{(state equation)} \tag{24}$$

$$\bar{s}_k = \mathbf{C}\mathbf{X}_k + \bar{v}_k \quad \text{(measurement equation)} \tag{25}$$

where

$$\mathbf{X}_k = \left[g_{k-(n-1)D_\mathrm{t}}, g_{k-(n-2)D_\mathrm{t}}, \ldots, g_k\right]'$$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ a_n & a_{n-1} & a_{n-2} & \ldots & a_2 & a_1 \end{bmatrix}$$

$$\mathbf{C} = [0, 0, \ldots, 0, 1]$$

$$\mathbf{B} = [0, 0, \ldots, 0, 1]'$$

the superscript $'$ denotes matrix transposition. Since the knowledge of the channel fading statistics and the variance of the background noise are ordinarily unknown *a priori* in practical systems, in the following, an H_∞-based channel estimation algorithm [27], which does not require channel statistics, is adopted as the channel estimator. A brief description of the H_∞ estimator is given below.

Let $z_k = \boldsymbol{\xi}\mathbf{X}_k$, where $\boldsymbol{\xi}$ is a $1 \times n$ linear transformation operator and $\hat{z}_k$ be the estimate of z_k. Define the estimation error as

$$e_k \triangleq z_k - \hat{z}_k. \tag{26}$$

The design criterion of the H_∞ estimator is to provide a uniformly small estimation error for any w_k, $\bar{v}_k$, and initial condition $\mathbf{X}_0$. The measure of performance is defined as the transfer operator that transforms the w_k, $\bar{v}_k$ and the uncertainty of the initial condition $\mathbf{X}_0$ to the estimation error e_k. The objective function is

$$J \triangleq \frac{\sum_{i=0}^{\infty} |e_i|^2_Q}{|\mathbf{X}_0 - \hat{\mathbf{X}}_0|^2_{\mathbf{p}_0^{-1}} + \sum_{i=0}^{\infty}\left\{|\bar{v}_i|^2_{V_H^{-1}} + |w_i|^2_{W_H^{-1}}\right\}} \tag{27}$$

where $\hat{\mathbf{X}}_0$ is an *a priori* estimate of $\mathbf{X}_0$, $(\mathbf{X}_0 - \hat{\mathbf{X}}_0)$ represents the unknown initial condition error, and $Q \geq 0$, $\mathbf{p}_0 > 0$, $W_H > 0$ and $V_H > 0$ are weighting parameters. $\mathbf{p}_0$ denotes a positive definite matrix that reflects *a priori* knowledge on how close the initial guess $\hat{\mathbf{X}}_0$ is to $\mathbf{X}_0$. W_H and V_H are weighting variables. In practical systems, the values of W_H and V_H can be chosen as the estimates of the covariances of w_i and $\bar{v}_i$. The optimal estimate of z_k, among all possible $\hat{z}_k$'s (i.e., the worst case performance measure), should satisfy

$$\|J\|_\infty \triangleq \sup_{\bar{v}_k, w_k, \mathbf{X}_0} J \leq \gamma^{-1} \tag{28}$$

where "sup" stands for supremum and $\gamma(> 0)$ is a prescribed level of noise attenuation.

For the state-space model (24) and (25), with the performance criterion (28), there exists an H_∞ estimator for z_k if and only if there exists a stabilizing symmetric positive definite solution $\mathbf{\Gamma}_k$ to the following discrete-time Riccati-type equation

$$\mathbf{\Gamma}_{k+1} = \mathbf{A}\mathbf{\Gamma}_k\left(\mathbf{I} - \gamma\bar{\mathbf{Q}}\mathbf{\Gamma}_k + \mathbf{C}' V_H^{-1}\mathbf{C}\mathbf{\Gamma}_k\right)^{-1}\mathbf{A}' + \mathbf{B}W_H\mathbf{B}'$$

$$\mathbf{\Gamma}_0 = \mathbf{p}_0 \tag{29}$$

where $\mathbf{p}_0$ is the initial condition. If a solution $\mathbf{\Gamma}_k$ exists, then the H_∞ estimator is given by

$$\hat{z}_k = \boldsymbol{\xi}\hat{\mathbf{X}}_k, \qquad k = 1, 2, 3, \ldots \tag{30}$$

where

$$\hat{\mathbf{X}}_k = \mathbf{A}\hat{\mathbf{X}}_{k-1} + \mathbf{G}_k(\bar{s}_k - \mathbf{C}\mathbf{A}\hat{\mathbf{X}}_{k-1}), \qquad \hat{\mathbf{X}}_0 = [0]_{n\times 1} \tag{31}$$

and $\mathbf{G}_k$ is the gain of the H_∞ estimator given by

$$\mathbf{G}_k = \mathbf{\Gamma}_k \left(\mathbf{I} - \gamma\bar{\mathbf{Q}}\mathbf{\Gamma}_k + \mathbf{C}'V_H^{-1}\mathbf{C}\mathbf{\Gamma}_k\right)^{-1} \mathbf{C}'V_H^{-1}. \tag{32}$$

From (29)–(32), the following observations can be obtained to reveal a glimpse of the implementation complexity of the H_∞ algorithm compared to the Kalman and MMSE algorithms. 1) From a similar observer structure between the proposed H_∞ and the Kalman estimation algorithms, the H_∞ estimator has a similar hardware structure and computation complexity as the Kalman estimator. 2) For the H_∞ estimation algorithm, different estimation results can be obtained with different vector $\boldsymbol{\xi}$. For example, if we choose $\boldsymbol{\xi} = [1, 0, \ldots, 0]_{1\times n}$, the H_∞ estimation algorithm is designed to obtain the optimal estimation of $g_{k-(n-1)D_\mathrm{t}}$, which should give a better estimation of channel fading $g_{k-(n-1)D_\mathrm{t}}$ in the kth epoch since the estimation is based on the $\{\bar{s}_i\}$, $1 \leq i \leq k$. This estimation is equivalent to the fixed-lag smoothing problem. The only difference from the traditional fixed-lag smoothing problem is that no additional computation is required in this case. 3) Although the MMSE estimation algorithm proposed in [19] can endure some mismatch on the correlation function of the channel fading, information on coherence bandwidth of the fading channel and the variance of the background noise is still required. Obtaining this information accurately may greatly increase the complexity of the receiver design. For the proposed H_∞ algorithm, the inherent robustness reduces the dependence of the estimation performance on the accuracy of the parameter estimation, which significantly reduces the complexity of the receiver design.

The parameter γ in the algorithm is determined by guaranteeing $\mathbf{\Gamma}_{k+1}$ in (29) to be positive definite, i.e.,

$$\begin{aligned} &\mathbf{\Gamma}_k(\mathbf{I} - \gamma\bar{\mathbf{Q}}\mathbf{\Gamma}_k + \mathbf{C}'V_H^{-1}\mathbf{C}\mathbf{\Gamma}_k)^{-1} > 0 \\ \Rightarrow &\gamma^{-1} > \max\left\{\mathrm{eig}\left[\bar{\mathbf{Q}}\left(\mathbf{\Gamma}_k^{-1} + \mathbf{C}'V_H^{-1}\mathbf{C}\right)^{-1}\right]\right\} \end{aligned} \tag{33}$$

where $\max\{\mathrm{eig}(\mathbf{X})\}$ denotes the maximum eigenvalue of the matrix $\mathbf{X}$. It has been shown in [27] that very accurate channel estimates can be achieved by the proposed H_∞ estimation algorithm.

VI. Simulation Results and Discussions

In this section, simulation results are presented to demonstrate the performance of the proposed downlink resource management scheme for OFDM system.

A. Simulation Parameters

Consider an OFDM system using quaternary phase-shift keying (QPSK) [19]. The entire system bandwidth is 800 kHz, which is divided into 128 subbands. The four subcarriers on each end are used as guard tones and the rest (120 tones) are used to transmit data. To make the tones orthogonal to each other, the symbol duration is 160 μs. An additional 40 μs guard interval is used to provide protection from ISI due to channel multipath delay spread. This results in a total block length of $T = 200\ \mu\mathrm{s}$ and a subcarrier symbol rate $r_s = 5\ \mathrm{kBd/s}$, or equivalently, a bit rate $r_\mathrm{b} = 10\ \mathrm{kb/s}$ for QPSK. In general, the channels corresponding to different users have the same but independent statistics. The channel in the simulation is a two-path Rayleigh fading channel with delay 0 and 20 μs, respectively. The normalized Doppler fading rate is $f_\mathrm{d}T = 0.01$ and the delay power spectral density satisfies uniform distribution. Channel fading is generated using Jakes' model [28]. At the receiver end of each mobile user, H_∞-based channel estimation algorithm is adopted to obtain the channel fading information, where the SNR at the pilots is set to 40 dB. The parameters for channel estimation are the same as those in [27]. For performance comparison, without loss of generality, no feedback delay is considered.

Two classes of packetized traffic are considered: voice and video. Voice traffic is generated using an on–off model. During the on state, the voice packets arrive at a constant rate $R_\mathrm{vo} = 16\ \mathrm{kb/s}$. The voice activity factor is assumed to be 0.4. The SNR requirement for the voice traffic at the receiver end is set to 10 dB. Video traffic is modeled by an eight-state Markov-modulated Poisson process [20]. The average duration in each state is chosen to be 40 ms, which is equivalent to the length of one frame of the video sequence with a frame rate of 25 frames/s. The SNR requirement of the video traffic at the receiver end is set to be 14 dB. Both classes of traffic have the same packet length of 1000 bits/packet or 500 symbols/packet. The buffer for the individual queue is assumed infinite, i.e., no packet dropping is taken into account.

B. Simulation Results

We first simulate four homogeneous video flows with the same weights, $\eta_1 = \eta_2 = \eta_3 = \eta_4$. For comparison, the conventional resource management scheme based on PGPS scheduling [6] is also simulated. Since the target transmission power determines the number of effective subcarriers, in the simulation, it is regulated to reflect different system loads. Let the packet transmission delay be the time difference between the time when the packet arrives at the base station and the time when the packet transmission is finished. The performance of the various resource management schemes is compared using the following performance measures:

1) the packet maximum transmission delay;
2) the packet average transmission delay;
3) the system throughput (the total number of packets transmitted over the simulation duration).

Figs. 6–8 show the performance comparison in terms of maximum transmission delay, average transmission delay, and

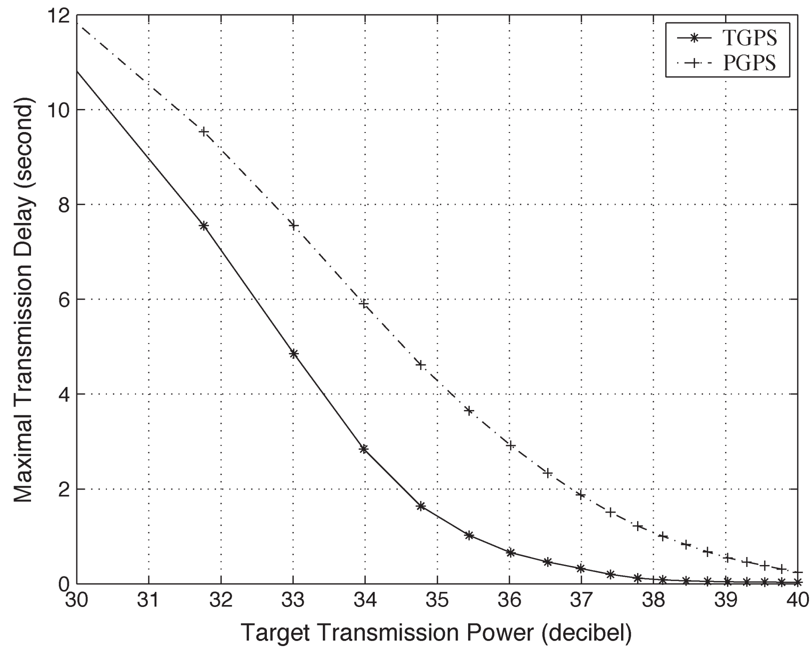

Fig. 6. Maximum transmission delay for $\eta_1 = \eta_2 = \eta_3 = \eta_4$.

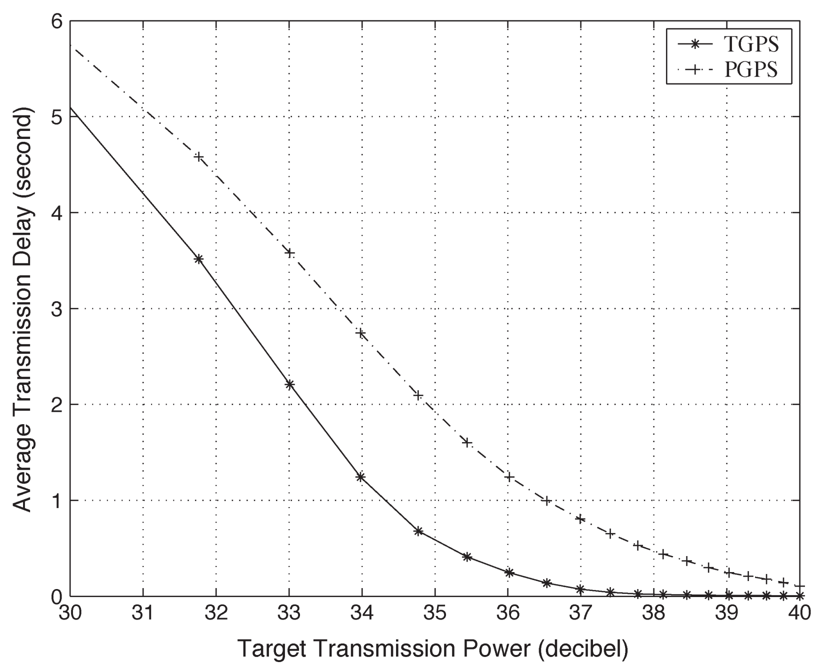

Fig. 7. Average transmission delay for $\eta_1 = \eta_2 = \eta_3 = \eta_4$.

system throughput, respectively. The main observations from the simulation results are as follows.

1) The proposed resource management scheme outperforms the conventional one based on PGPS scheduling in terms of maximum transmission delay, average transmission delay, and system throughput. The performance improvement is due to the parallel transmission properties of the TGPS scheduling, which allows the power and subcarrier allocation algorithm to make use of the multiuser diversity gain. The performance gain also indicates that multiuser diversity must be considered in a multiuser system. Lower transmission delay indicates that the proposed scheme is much more suitable to support the delay-sensitive traffic such as voice.
2) When the target transmission power becomes large enough, each user will experience the same system capacity, i.e., each user can use all the subcarriers, in any time epoch. No effects of channel fading difference exist between the users, and the two resource management schemes achieve the same performance for large target transmission power.
3) Figs. 6 and 7 also demonstrate the effects of the target transmission power on the transmission delay. At low

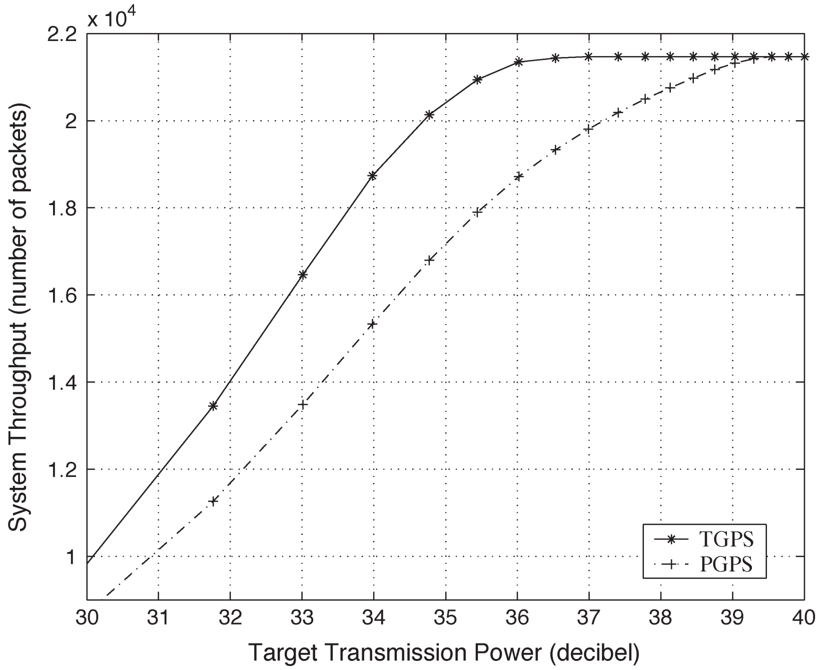

Fig. 8. System throughput for $\eta_1 = \eta_2 = \eta_3 = \eta_4$.

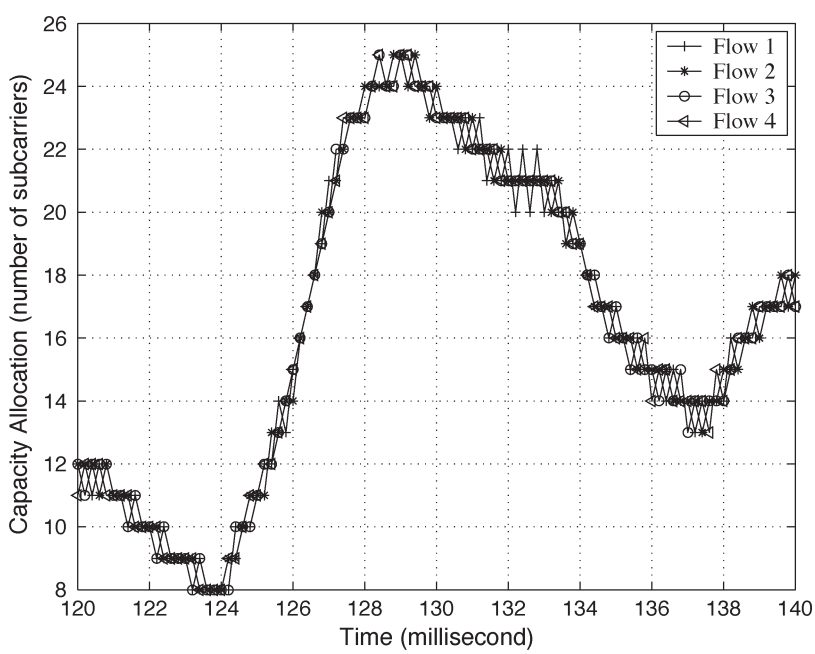

Fig. 9. Capacity allocation for $P = 30$ dB, $\eta_1 = \eta_2 = \eta_3 = \eta_4$.

target transmission power, only a small number of subcarriers can transmit simultaneously, i.e., the system works under a heavy load condition and the traffic experiences large maximum transmission delay and average transmission delay. With an increase in the target transmission power, the transmission delays decrease significantly until they reach a saturated value when the system capacity is independent of the target transmission power.

4) In Fig. 8, when the target transmission power is small, system throughput is mainly determined by the target transmission power, and increases with the increase of it. However, after the target transmission power is above a threshold, for example, 36 dB in our simulation, the system throughput of the proposed scheme is saturated; it is mainly determined by the traffic load and the total number of subcarriers in the frequency domain. The lower saturation threshold also indicates the multiuser diversity gain of our proposed scheme.

In order to evaluate the fairness of the proposed TGPS scheduling, the subcarrier allocation for each video flow over a time duration is shown in Fig. 9 where the target transmission

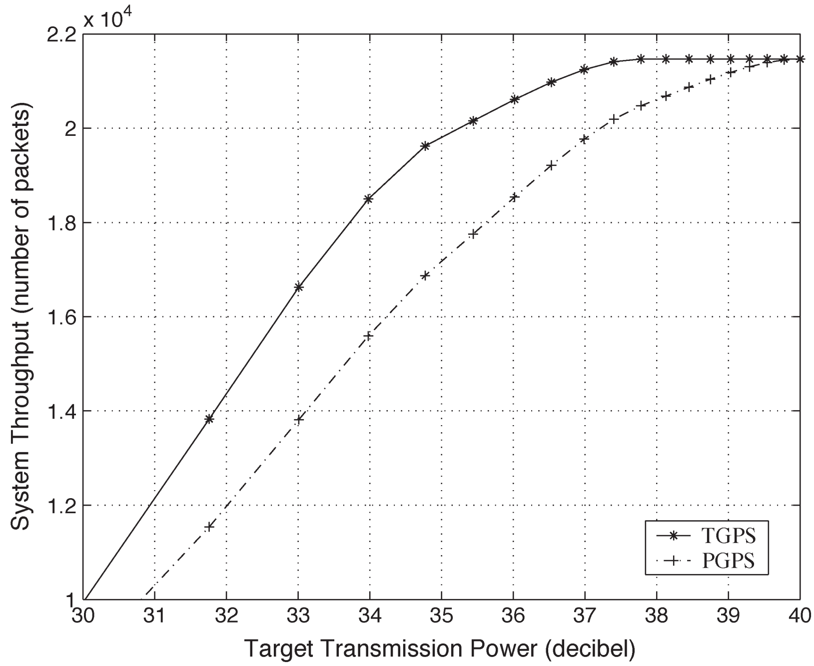

Fig. 10. System throughput for $\eta_1 = 1$, $\eta_2 = 2$, $\eta_3 = 3$, and $\eta_4 = 4$.

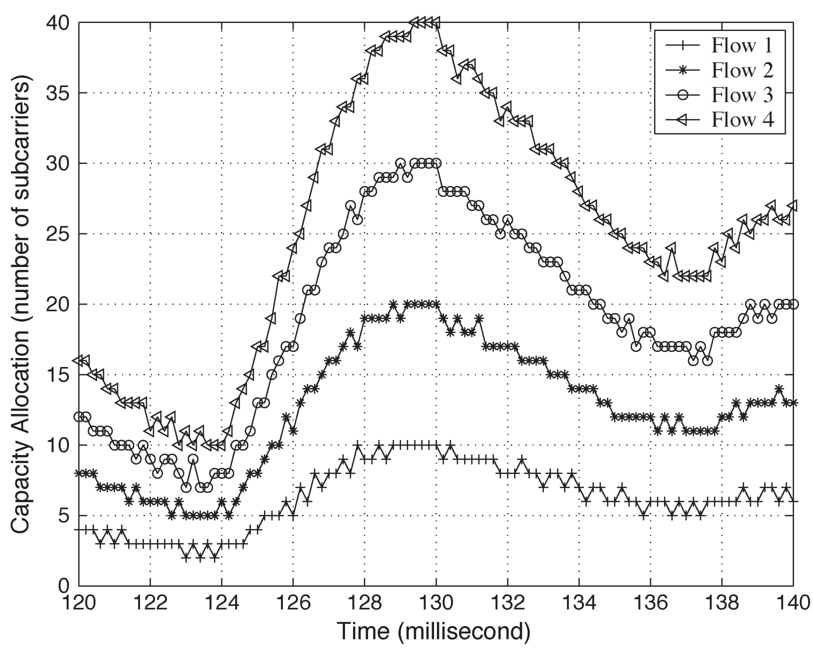

Fig. 11. Capacity allocation for $P = 30$ dB, $\eta_1 = 1$, $\eta_2 = 2$, $\eta_3 = 3$, and $\eta_4 = 4$.

power is set to be 30 dB. It can be observed that the subcarrier allocation of each flow is proportional to its weight as long as it is continually backlogged, except for a very small fluctuation due to the quantization error. Therefore, the proposed TGPS scheduling can achieve nearly the same performance as the ideal GPS in terms of fairness.

The same simulation is carried out for flows with different weights $\eta_1 = 1$, $\eta_2 = 2$, $\eta_3 = 3$, and $\eta_4 = 4$ and flows with significant different weights $\eta_1 = 100$, $\eta_2 = 10$, $\eta_3 = 9$, and $\eta_4 = 1$. The results are presented in Figs. 10–13. From the figures, the same performance gain can be observed as that under the same weight case and the fairness of the TGPS scheduling is further verified. One additional observation from Fig. 13 is that the capacity of flow 1 equals 0 during, for example, 130–131 ms. The reason is that during this period, at any OFDM symbol interval, the total number of symbols of flow 1 waiting for transmission is 0. According to the work-conserving principle, the unused subcarriers by flow 1 should be allocated to other backlogged users according to their weights. This explains why at some instants, the capacity allocated to flow 1 is less than that allocated to other flows even though the weight of flow 1 is the largest.

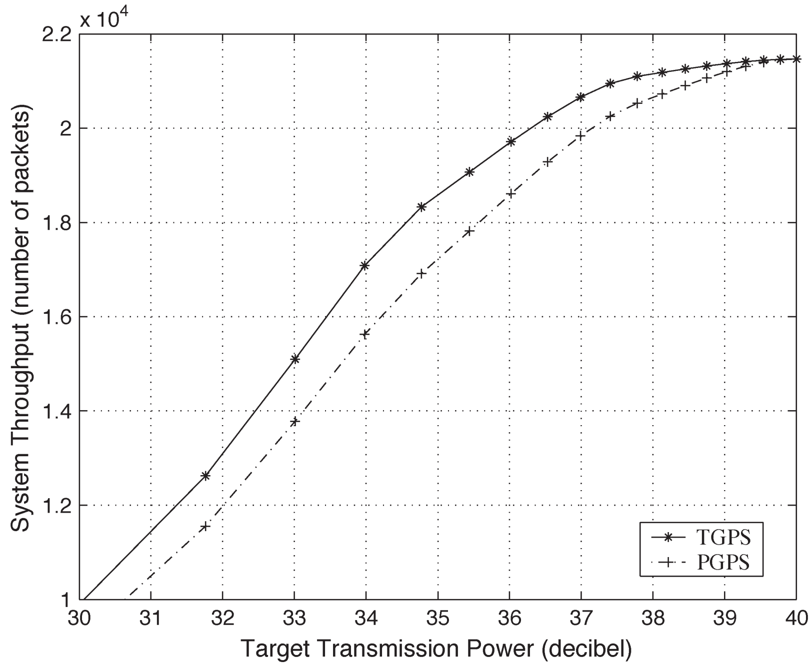

Fig. 12. System throughput for $\eta_1 = 100$, $\eta_2 = 10$, $\eta_3 = 9$, and $\eta_4 = 1$.

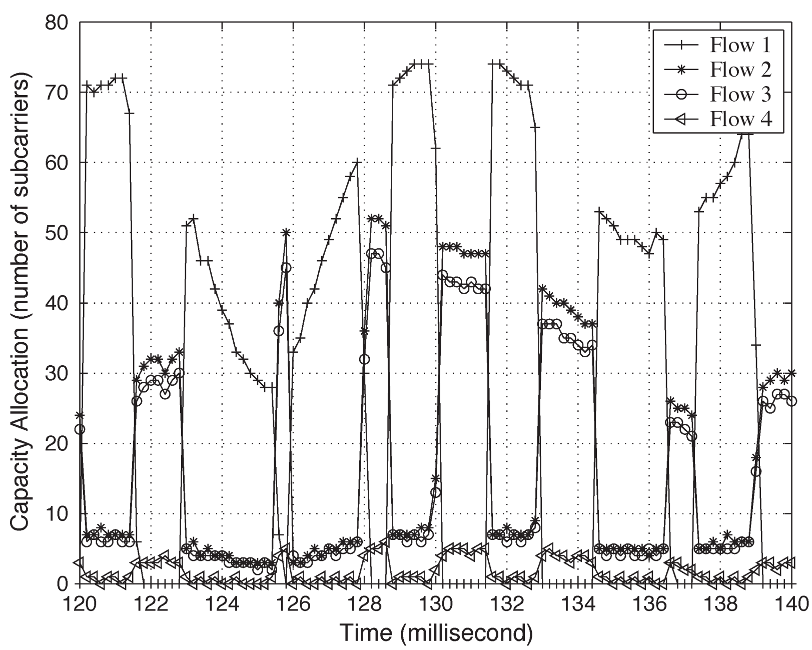

Fig. 13. Capacity allocation for $P = 30$ dB, $\eta_1 = 100$, $\eta_2 = 10$, $\eta_3 = 9$, $\eta_4 = 1$.

To demonstrate the performance of the proposed resource management scheme in a heterogeneous traffic environment, ten voice and four video flows are simulated. For heterogeneous traffic, the weights should be selected carefully based on each traffic's delay and bandwidth requirements [29], [30]. For simulation purpose, the weights of the voice and the video flows are set to be 2 and 1, respectively, in order to provide higher priority to the voice flows by assuming that the voice flows are more delay sensitive than the video flows. The higher priority to the voice flows can be explained as follows. The video flows usually have, e.g., K times larger bandwidth requirement than the voice flows. Normalizing the bandwidth requirement of both classes of traffic by the bandwidth requirement of the voice flows results in one unit normalized bandwidth requirement for the voice flows and K units normalized bandwidth requirement for the video flows. From the bandwidth requirement point of view, the weights ($\eta_{\rm vo} = 1$, $\eta_{\rm vi} = K$) give the same priority to both flows, where $\eta_{\rm vo}$ and $\eta_{\rm vi}$ are the weights of the voice and the video flows, respectively. Thus, other weight allocations ($\eta_{\rm vo} > 1$, $\eta_{\rm vi} < K$) will give higher priority to the voice flows. Without loss of generality, the target transmission power is set to be 33 dB. Simulation results are summarized in Table I. It

TABLE I
PERFORMANCE FOR HETEROGENEOUS TRAFFIC: $P = 33$ dB, $\eta_1 = \cdots = \eta_{10} = 2, \eta_{11} = \cdots = \eta_{13} = 1$

	Traffic Type	Maximum Delay (millisecond)	Average Delay (millisecond)	Throughput (packets)
Flow 1		5.7	2.9	292
Flow 2		3.2	2.4	313
Flow 3		5.7	2.6	295
Flow 4		4.2	2.7	315
Flow 5	Voice	5.7	3.1	297
Flow 6		5.7	3.4	275
Flow 7		3.7	2.5	289
Flow 8		4.9	2.7	269
Flow 9		5.7	3.2	283
Flow 10		3.8	2.9	294
Flow 11		2229.3	885.7	5682
Flow 12	Video	485.5	119	4697
Flow 13		1145.4	577.3	5223
Flow 14		504.8	212.7	5171

can be seen that a lower maximum transmission delay and a lower average transmission delay are achieved by the voice traffic flows, since they have larger weights than the video traffic flows.

VII. CONCLUSION

A downlink resource management scheme for packet transmission in OFDM wireless communication systems has been proposed. By integrating power distribution, subcarrier allocation, and GPS scheduling, the proposed resource management scheme maximizes the system throughput while satisfying the constraints on the target transmission power and fairness requirements. The proposed resource management scheme is formulated based on a knowledge of the channel impulse response. In the absence of this knowledge, an H_∞ channel impulse estimator is used to provide a reasonably accurate estimate of the channel impulse response. The ideal GPS algorithm is made practically implementable by truncation. The resource management scheme with TGPS exhibits better throughput performance than the PGPS. It is conjectured that the proposed resource management scheme is useful for future development of efficient OFDM systems supporting multimedia services.

REFERENCES

[1] J. A. C. Bingham, "Multicarrier modulation for data transmission: An idea whose time has come," *IEEE Commun. Mag.*, vol. 28, no. 5, pp. 5–14, May 1990.

[2] S. B. Weinstein and P. M. Ebert, "Data transmission by frequency-division multiplexing using the discrete Fourier transform," *IEEE Trans. Commun.*, vol. COM-19, no. 5, pp. 628–634, Oct. 1971.

[3] L. J. Cimini, Jr., "Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing," *IEEE Trans. Commun.*, vol. COM-33, no. 7, pp. 665–675, Jul. 1985.

[4] V. Mignone and A. Morello, "CD3-OFDM: A novel demodulation scheme for fixed and mobile receivers," *IEEE Trans. Commun.*, vol. 44, no. 9, pp. 1144–1151, Sep. 1996.

[5] A. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: The single-node case," *IEEE/ACM Trans. Netw.*, vol. 1, no. 3, pp. 334–357, Jun. 1993.

[6] D. Stiliadis and A. Varma, "Efficient fair queuing algorithms for packet-switched networks," *IEEE/ACM Trans. Netw.*, vol. 6, no. 2, pp. 175–185, Apr. 1998.

[7] Y. Cao and V. O. K. Li, "Scheduling algorithms in broadband wireless networks," *Proc. IEEE*, vol. 89, no. 1, pp. 76–87, Jan. 2001.

[8] S. Lu, V. Bharghavan, and R. Srikant, "Fair scheduling in wireless packet networks," *IEEE/ACM Trans. Netw.*, vol. 7, no. 4, pp. 473–489, Aug. 1999.

[9] M. A. Arad and A. L. Garcia, "A generalized processor sharing approach to time scheduling in hybrid CDMA/TDMA," in *Proc. IEEE Information Communications (INFOCOM)'98*, San Francisco, CA, pp. 1164–1171.

[10] L. Xu, X. Shen, and J. W. Mark, "Dynamic fair scheduling with QoS constraints in multimedia wideband CDMA cellular networks," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 60–73, Jan. 2004.

[11] B. S. Krongold, K. Ramchandran, and D. L. Jones, "Computationally efficient optimal power allocation algorithms for multicarrier communication systems," *IEEE Trans. Commun.*, vol. 48, no. 1, pp. 23–27, Jan. 2000.

[12] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 773–775, Feb./Mar./Apr. 1995.

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

[14] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 171–178, Feb. 2003.

[15] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, pp. 1150–1158, Nov. 2003.

[16] R. Knopp and P. Humblet, "Information capacity and power control in single cell multiuser communications," in *Proc. Int. Conf. Communications*, Seattle, WA, Jun. 1995, pp. 331–335.

[17] R. Steele, *Mobile Radio Communications*. New York: IEEE Press, 1992.

[18] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 1998.

[19] Y. Li, L. J. Cimini, and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 902–915, Jul. 1998.

[20] M. Schwart, *Broadband Integrated Networks*. Englewood Cliffs, NJ: Prentice Hall, 1996.

[21] M. S. Corson, R. Laroia, A. O'Neill, V. Park, and G. Tsirtsis, "A new paradigm for IP-based cellular networks," *IT Prof.*, vol. 3, no. 6, pp. 20–29, Nov./Dec. 2001.

[22] R. V. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. Norwood, MA: Artech House, 1999.

[23] Y. Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 4, pp. 1207–1215, Jul. 2000.

[24] P. Hoeher, S. Kaiser, and P. Robertson, "Two-dimension pilot-symbol-aided channel estimation by Wiener filtering," in *Proc. Int. Conf. Acoust., Speech and Signal Process. (ICASSP) 1997*, Munich, Germany, pp. 1845–1848.

[25] M. K. Tsatsanis, G. B. Giannakis, and G. Zhou, "Estimation and equalization of fading channels with random coefficients," *Signal Process.*, vol. 53, no. 2–3, pp. 211–229, Sep. 1996.

[26] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[27] J. Cai, X. Shen, and J. W. Mark, "Robust channel estimation for OFDM wireless communication systems—An H_∞ approach," in *Proc. IEEE Vehicular Technology Conf. '02 Fall*, Vancouver, BC, Canada, pp. 2106–2110.

[28] W. C. Jakes, *Microwave Mobile Communications*. New York: IEEE Press, 1993.

[29] A. Elwalid and D. Mitra, "Design of generalized processor sharing schedulers which statistically multiplex heterogeneous QoS classes," in *Proc. IEEE Information Communications (INFOCOM)'99*, New York, pp. 1120–1230.

[30] K. Kumaran, G. E. Margrave, D. Mitra, and K. R. Stanley, "Novel techniques for the design and control of generalized processor sharing schedulers for multiple QoS classes," in *Proc. IEEE Information Communications (INFOCOM) 2000*, Tel Aviv, Israel, pp. 932–941.

Jun Cai received the B.Eng. degree in radio techniques and the M.Eng. degree in communication and information systems from Xi'an Jiaotong University, Xi'an, China, in 1996 and 1999, respectively, and the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, ON, Canada, in 2004.

He is currently doing research as Postdoctoral Fellow in electrical and computer engineering at the University of Waterloo. His research interests include channel estimation, interference cancellation, and resource management in wireless communication systems.

Xuemin (Sherman) Shen received the B.Sc. degree from Dalian Maritime University, Dalian, China, in 1982 and the M.Sc. and Ph.D. degrees from Rutgers University, New Brunswick, NJ, in 1987 and 1990, respectively, all in electrical engineering.

From September 1990 to September 1993, he was first with Howard University, Washington D.C., and then the University of Alberta, Edmonton, AB, Canada. Since October 1993, he has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, where he is a Professor. His research focuses on mobility and resource management in interconnected wireless/wireline networks, UWB wireless communications systems, wireless security, and ad hoc and sensor networks. He is a coauthor of two books and has published more than 150 papers in wireless communications and networks, control, and filtering.

Dr. Shen was the Technical Cochair for IEEE Globecom '03 Symposium on Next Generation Networks and Internet, and ISPAN '04. He serves as the Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS; IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY; Dynamics of Continuous, Discrete and Impulsive—Series B: Applications and Algorithms; Wireless Communications and Mobile Computing (Wiley); and International Journal of Computer and Applications. He also serves as Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS, and IEEE Communications Magazine. He received the Premier's Research Excellence Award (PREA) from the Province of Ontario, Canada, for demonstrated excellence of scientific and academic contributions in 2003, and the Distinguished Performance Award from the Faculty of Engineering, University of Waterloo, for his outstanding contribution in teaching, scholarship, and service in 2002. He is a registered Professional Engineer of Ontario, Canada.

Jon W. Mark (M'62–SM'80–F'88–LF'03) received the B.A.Sc. degree from the University of Toronto, Toronto, ON, Canada, in 1962, and the M.Eng. and Ph.D. degrees from McMaster University, Hamilton, ON, Canada, in 1968 and 1970, respectively, both in electrical engineering.

From 1962 to 1970, he was an Engineer and then Senior Engineer with Canadian Westinghouse Company Ltd., Hamilton, Ontario, Canada. During the period between October 1968 and August 1970, he was on leave of absence from Canadian Westinghouse to pursue his Ph.D. studies at McMaster University under the auspices of an NRC PIER Fellowship. In September 1970, he joined the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, where he is currently a Distinguished Professor Emeritus. He served as Department Chairman during the period between July 1984 and June 1990. In 1996, he established the Centre for Wireless Communications (CWC) at the University of Waterloo and is currently serving as its founding Director. He was on sabbatical leave at the IBM Thomas Watson Research Center, Yorktown Heights, NY, as a Visiting Research Scientist (1976–1977); at AT&T Bell Laboratories, Murray Hill, NJ, as a Resident Consultant (1982–1983); at the Laboratoire MASI, Universite Pierre et Marie Curie, Paris, France, as an Invited Professor (1990–1991); and at the Department of Electrical Engineering, National University of Singapore, Singapore, as a Visiting Professor (1994–1995). He has worked in the areas of adaptive equalization, spread spectrum communications, antijamming secure communication over satellites, and ATM networks. His current research interests are in broadband and wireless communications and networks, including power control, resource allocation, mobility management, and end-to-end QoS provisioning in hybrid wireless/wireline networks.

Dr. Mark was an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS from 1983 to 1989. He served as the Technical Program Chairman of INFOCOM '89 and was a member of the Inter-Society Steering Committee of the IEEE/ACM TRANSACTIONS ON NETWORKING (1992–2003), an Editor of the ACM/Baltzer Wireless Networks journal (1997–2004), and an Associate Editor of Telecommunication Systems (1996–2004).