

Research Article

Joint Throughput Maximization and Fair Uplink Transmission Scheduling in CDMA Systems

Symeon Papavassiliou^{1,2} and Chengzhou Li³

¹Network Management and Optimal Design Laboratory (NETMODE), Institute of Communications and Computer Systems (ICCS), 9 Iroon Polytechniou Street, Zografou 157 73, Athens, Greece

²School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), 9 Iroon Polytechniou Street, Zografou 157 73, Athens, Greece

³LSI Corporation, 1110 American Parkway NE, Allentown, PA 18109, USA

Correspondence should be addressed to Symeon Papavassiliou, papavass@mail.ntua.gr

Received 9 July 2008; Revised 10 December 2008; Accepted 20 February 2009

Recommended by Alagan Anpalagan

We study the fundamental problem of optimal transmission scheduling in a code-division multiple-access wireless system in order to maximize the uplink system throughput, while satisfying the users quality-of-service (QoS) requirements and maintaining fairness among them. The corresponding problem is expressed as a weighted throughput maximization problem, under certain power and QoS constraints, where the weights are the control parameters reflecting the fairness constraints. With the introduction of the power index capacity, it is shown that this optimization problem can be converted into a binary knapsack problem, where all the corresponding constraints are replaced by the power index capacities at some certain system power index. A two-step approach is followed to obtain the optimal solution. First, a simple method is proposed to find the optimal set of users to receive service for a given fixed target system load, and then the optimal solution is obtained as a global search within a certain range. Furthermore, a stochastic approximation method is presented to effectively identify the required control parameters. The performance evaluation reveals the advantages of our proposed policy over other existing ones and confirms that it achieves very high throughput while maintains fairness among the users, under different channel conditions and requirements.

Copyright © 2009 S. Papavassiliou and C. Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The continuous growth in traffic volume and the emergence of new services have begun to change the structure and requirements of wireless networks. Future mobile communication systems will be characterized by high throughput, integration of services, and flexibility [1–5]. With the demand for high data rate and support of multiple quality of service (QoS), the transmission scheduling plays a key role in the efficient resource allocation process in wireless systems. The transmission scheduling determines the time instances that a mobile user may receive service, as well as the resources that should be allocated to support the requested service, in order to make the resource distribution fair and efficient.

The fundamental problem of scheduling the users transmission and allocating the available resources in a realistic uplink code-division multiple-access (CDMA) wireless

system that supports multirate multimedia services, with efficiency and fairness, is investigated and analyzed in this paper. A transmission scheduling method which achieves the maximum system throughput under the constraints of satisfying certain users QoS requirements and maintaining throughput fairness among them is provided and evaluated.

1.1. Related Work and Motivation. A class of scheduling schemes, namely, the opportunistic scheduling schemes, has been proven to be an effective approach to improve the system throughput by utilizing the multiuser diversity effect [6, 7] in wireless communications. Specifically, for a system with many users that have independent varying channels, with high probability there is a user with channel much stronger than its average SNR requirement. Therefore, the system throughput may be maximized by choosing

the user with “relatively best” channel for transmission at a given slot. However, some fairness constraints must be imposed on the scheduling policies to ensure the fair resource allocation.

It has been shown in [8] that scheduling users one-by-one can result in higher system throughput for high data rate traffic in the CDMA downlink. However, this work does not exploit the time-varying channel conditions. In [7, 9], a high-speed data rate scheme (HDR) is introduced, where the base station schedules the downlink transmission of a single user at a given time slot with the data rates and slot lengths varying according to the specific channel condition. In [10–12], a transmission scheduling scheme for multiple users, which considers both the channel condition and queueing delay/length, is proposed and shown to be throughput optimal if it is feasible. However, the fairness issue is not explicitly addressed in that work. In [13–15], a framework for opportunistic scheduling that maximizes the system performance by exploiting the time-varying channel conditions of wireless networks is presented. Three categories of scheduling problems—the temporal fairness, utilitarian fairness, and minimum-performance guarantee scheduling—are studied and optimal solutions are given.

Although the downlink transmission assignment is important for several applications, the efficient uplink transmission scheduling plays an important role as well, especially with the prevailing of multimedia communications and applications. It has been argued that the downlink scheduling method is not suitable to be applied to the uplink transmission scheduling, where simultaneous transmissions may result in higher throughput [16, 17]. The uplink transmission scheduling problem is more complicated and requires further consideration of additional elements to make the corresponding scheduling policies feasible [18]. The achievable throughput in such a case depends not only on the service access time, but also on the transmission powers and the corresponding users interference. In addition, multiple users can be scheduled simultaneously to transmit in the same time slot, which is a major difference from the wireline and TDMA-like scheduling schemes, making the respective scheduling processes either inapplicable or inefficient in the CDMA environment. The simple temporal fairness scheduling, where the main resource to be shared is the time, fails to provide rational fairness in this case. As a result, the throughput optimal and fair uplink transmission scheduling problem needs to jointly consider multiple factors such as access time, transmission power, channel conditions, and number of users to be scheduled at the same time. Heuristic approaches to address the problem of short-term fairness and demonstrate the tradeoff between fairness and throughput under some special cases have been introduced in [19–21].

Furthermore, how to maximize the throughput of uplink CDMA system was first analyzed in [16]. The sole purpose of uplink throughput maximization can be achieved by choosing the “best” K users in terms of their received power, when they transmit at their maximum power. However, such throughput maximization does not consider fairness, that is, the equal opportunity for all users to receiving service despite

their channel conditions. Therefore, among the objectives of our approach in this paper is to identify the actual “best” users that should transmit in order to maximize the throughput, when the fairness constraints are introduced and respected.

In [22], several scenarios of scheduling uplink CDMA transmission with voice and data services are analyzed. With the number of voice users and their corresponding transmission rates fixed, that work attempted to maximize the throughput of data service. It was shown that when the synchronization overhead is reasonable, a smaller number of simultaneous transmission users achieve higher system throughput and at the same time lower the average transmission power. However, in this case the achievable throughput is affected by the “weakest link.” Therefore, this approach can be regarded only as a static analysis that considers the relationship between the performance and the number of users chosen for transmission. The problem of identifying the actual set of users to transmit based on their channel conditions, which may reduce the impact of the “weakest link”, has not yet been investigated and is one of the main objectives of our paper.

In addition, the problem of uplink CDMA scheduling is further complicated by the fact that the conventional concept of capacity used in the wireline networks, for example, total bandwidth of the physical media, is not directly applicable in the CDMA systems. In this case, the actual system capacity is not fixed and known in advance, since it is a function of several parameters such as the number of users, the channel conditions, and the transmission powers.

Therefore, in summary the main contributions of this paper are as follows. (1) Jointly consider the factors of channel capacity, number of users and their interference, transmit power, and fairness requirements. (2) Formulate an optimization problem that stresses the fairness requirement under time-varying wireless environment and proves the existence of an optimal solution based on all constraints. (3) Exploit the power index concept and power index capacity, as a novel and effective way, to treat the fairness issue in the transmission scheduling policy under the considered uncertain and dynamic environment. (4) Devise a scheduling policy that achieves throughput fairness among the users and optimal system throughput under certain constraints.

1.2. Paper Outline. The rest of the paper is organized as follows. In Section 2, the system model that is used throughout our analysis is described, and the problem of the uplink scheduling in CDMA systems is rigorously formulated as a multiconstraint optimization problem. It is demonstrated that this problem can be expressed as a weighted throughput maximization problem, under certain power and QoS constraints, where the weights are the control parameters that reflect fairness constraints. Based on the concept of power index capacity, this optimization problem is converted into a simpler linear knapsack problem in Section 3.1, where all the corresponding constraints are replaced by the users power index capacities at some certain system power index. The optimal solution of the latter problem is identified in Sections 3.2 and 3.3, while

in Section 3.4, a stochastic approximation method is presented in order to effectively identify the required control parameters. Section 4 contains the performance evaluation of the proposed method, along with some numerical results and discussion, and finally Section 5 concludes the paper.

2. System Model and Problem Formulation

In this paper, we consider a single cell DS-CDMA system with $B(k)$ backlogged users at time slot k . The users channel conditions are assumed to change according to some stationary stochastic process, while the uplink transmission rate is assumed to be adjustable with the variable spreading gain technique [23]. Each user i is associated with some preassigned weight ϕ_i according to its QoS requirement. In the following for simplicity in the presentation, we omit the notation of the specific slot k from the notations and definitions we introduce. Let us denote by r_i the transmission rate of user i in the slot under consideration. We assume that the chip rate W for all mobiles is fixed, and hence the spreading gain G_i of user i is defined as $G_i = W/r_i$. Let us also denote by γ_i the required signal-to-interference and noise ratio (SINR) level of user i , by h_i the corresponding channel gain, and by p_i the user i transmission power at a given slot, which, however, is limited by the maximum power value p_i^{\max} . Therefore, the received SINR γ'_i for a user i is given by

$$\frac{h_i p_i G_i}{\alpha \sum_{j=1, j \neq i}^{B(k)} h_j p_j + W \eta_0} = \gamma'_i, \quad i = 1, 2, \dots, B(k), \quad (1)$$

where η_0 is the one-sided power spectral density of additive white Gaussian noise (AWGN), and α determines the proportion of the interference from other users received power. Without loss of generality in the following, we assume $\alpha = 1$. Obviously, to meet the SINR requirement, the received SINR γ'_i has to be larger than the corresponding threshold γ_i , that is, $\gamma'_i \geq \gamma_i$. In the following, we assume perfect power control in the system under consideration, while users are scheduled to transmit at the beginning of every fixed-length slot. The objective of the optimal scheduling policy Q^* is to find the optimal number of allowable users and their transmission rates, which achieves the maximum system throughput while maintaining the fairness property.

2.1. Problem Formulation. Let $R(k) = \sum_{i=1}^{B(k)} r_i(k)$ denote the total throughput in slot k . Our objective function is to maximize the expectation of $R(k)$ by selecting the optimal transmit power vector $(p_1, p_2, \dots, p_{B(k)})$ and transmit rate vector $(r_1, r_2, \dots, r_{B(k)})$, that is,

$$\max E \left(\sum_{i=1}^{B(k)} r_i \right) \quad (2)$$

subject to specific SINR, maximum transmit power, and fairness constraints as follows:

$$\begin{aligned} \frac{h_i p_i G_i}{\sum_{j=1, j \neq i}^{B(k)} h_j p_j + W \eta_0} &\geq \gamma_i, \quad \text{for } i = 1, 2, \dots, B(k), \\ p_i &\leq p_i^{\max}, \quad \text{for } i = 1, 2, \dots, B(k), \\ \frac{\bar{r}_i}{\phi_i} &= \frac{\bar{r}_j}{\phi_j} \quad \text{for } 1 \leq i, j \leq B(k), \end{aligned} \quad (3)$$

where $\bar{r}_i = E(r_i)$ denotes the mean throughput of user i in the corresponding backlogged period. It has been shown in [15, 24] that the above-constrained optimization problem can be considered as equivalent to the following problem (4), where Z is the minimal value among all \bar{r}_i/ϕ_i , that is, $Z = \min_i \{\bar{r}_i/\phi_i\}$. In (4), we transform the objective function (2) into finding the optimal transmit powers and rates that maximize the minimal normalized average rate Z . Therefore,

$$\begin{aligned} &\max Z, \\ \text{s.t. } &Z \leq \frac{\bar{r}_i}{\phi_i}, \quad 1 \leq i \leq B(k), \\ &\frac{h_i p_i W/r_i}{\sum_{j=1, j \neq i}^{B(k)} h_j p_j + W \eta_0} \geq \gamma_i \quad i = 1, 2, \dots, B(k), \\ &p_i \leq p_i^{\max}, \quad 1 \leq i \leq B(k). \end{aligned} \quad (4)$$

Apparently, the solution of the above problem will finally make $Z = \bar{r}_i/\phi_i$ for $1 \leq i \leq B(k)$ since one can always reduce its throughput for the benefit of other users in order to maximize Z . With the constraint $Z = \bar{r}_i/\phi_i$, the objective function then is generalized to

$$\max \sum_{i=1}^{B(k)} w_i \bar{r}_i, \quad (5)$$

where w_i is an arbitrary positive number. Here the crucial observation [24] is that the optimal scheduling policy will be the one that maximizes the sum of weighted throughputs and equalizes the normalized throughput. The maximization of mean-weighted rate in (5) is obtained by the maximization of the weighted rate in every slot, that is, $\max \sum_{i=1}^{B(k)} w_i r_i$ for every slot k . In conclusion, to obtain the optimal uplink throughput while keeping fairness, we must solve the following problem:

$$\max \sum_{i=1}^{B(k)} w_i r_i, \quad (6)$$

$$\begin{aligned} \text{s.t. } &\frac{h_i p_i W/r_i}{\sum_{j=1, j \neq i}^{B(k)} h_j p_j + W \eta_0} \geq \gamma_i, \quad i = 1, 2, \dots, B(k), \\ &p_i \leq p_i^{\max}, \quad 1 \leq i \leq B(k). \end{aligned} \quad (7)$$

The fairness constraint, that is, $\bar{r}_i/\phi_i = \bar{r}_j/\phi_j$, is represented by the choice of w_i . By adjusting the value of w_i , the user will get more or less opportunities to transmit data, and hence the corresponding normalized throughput is balanced. As we discuss later in this paper, the value of w_i can

be approximated by a stochastic approximation algorithm, which has already found its application in [14, 15] under similar situations. Note that since we assume perfect power control in the CDMA system under consideration, only the equality case of (7) is considered here.

The following Proposition 1 states that the optimal solution is achieved when a user either transmits at full power or does not transmit at all.

Proposition 1. *The optimal solution that maximizes the weighted throughput of problem (6) is such that*

$$p_i(k) \in \{0, p_i^{\max}\}, \quad \text{for } i = 1, 2, \dots, B(k). \quad (9)$$

Proof. In order to minimize the multiple access interference, users transmit with the minimum required power to meet the required threshold γ_i . Therefore, we consider the equality case of constraint (7). To maintain exactly the threshold γ_i for user i , the achievable transmit rate is represented as

$$r_i(k) = \frac{h_i p_i W}{\gamma_i \left(\sum_{j=1, j \neq i}^{B(k)} h_j p_j + W \eta_0 \right)}. \quad (10)$$

The objective function then becomes

$$Z = \sum_{i=1}^{B(k)} w_i r_i = \sum_{i=1}^{B(k)} \frac{w_i h_i W}{\gamma_i} \frac{p_i}{\sum_{j=1, j \neq i}^{B(k)} h_j p_j + W \eta_0}. \quad (11)$$

Differentiating twice with respect to the transmit power of a user m , we obtain

$$\frac{\partial^2 Z}{\partial p_m^2} = 2 \sum_{i=1, i \neq m}^{B(k)} \frac{w_i h_i W}{\gamma_i} \frac{p_i h_m^2}{\left(\sum_{j=1, j \neq i}^{B(k)} h_j p_j + W \eta_0 \right)^3}. \quad (12)$$

Since w_i is positive number, obviously (12) is nonnegative, while the objective function is a convex function of p_m . Hence, the optimal solution of this problem is that the transmit power obtains the value of its boundary, that is, either 0 or p_i^{\max} . \square

In Section 3, the corresponding optimization problem is transformed to an equivalent problem of a simpler form, which facilitates the identification of the optimal solution. However, in the following we first introduce the concept of power index capacity which is used to represent the corresponding constraints, under the problem transformation.

2.2. Power Index Capacity. It has been shown in [25] that by solving the constraints (7) and (8), the following inequality must be satisfied if there exists a feasible power assignment $\mathbf{p} = [p_1, p_2, \dots, p_{B(k)}]$ that meets the QoS requirements:

$$\begin{aligned} \sum_{i=1}^{B(k)} g_i &\leq 1 - \frac{\eta_0 W}{\min_{1 \leq i \leq B(k)} \{p_i^{\max} h_i (G_i / \gamma_i + 1)\}} \\ &= 1 - \frac{\eta_0 W}{\min_{1 \leq i \leq B(k)} \{p_i^{\max} h_i / g_i\}}, \end{aligned} \quad (13)$$

where

$$g_i = \frac{\gamma_i}{\gamma_i + G_i} \quad (14)$$

is defined as the power index of user i [26]. Relation (13) is the necessary and sufficient condition such that a power and rate solution is feasible under constraints (7) and (8) [25].

Let us regard $\sum_i g_i$ as the actual system load, which is the sum of power indices assigned to all backlogged users, while we assume that there is a target system load ψ . It should be noted that ψ here is not fixed but has value $0 \leq \psi < 1$. The meaning and motivation for the definition of the target system load ψ are that the system will attempt to provide the appropriate scheduling in order to make the actual system load $\sum g_i$ reach the target load (however, it serves as an upper bound and cannot be exceeded). For an arbitrarily selected ψ in the range of $0 < \psi < 1$, there exist two possible cases concerning the relationship between the actual system load $\sum g_i$ and the target system load ψ , the system can easily make the actual system load $\sum g_i$ reach the target load under consideration, that is, $\sum g_i = \psi$. On the other hand, when ψ is large, especially when it approaches to 1, it may be impossible for the actual achievable system load $\sum g_i$ to reach ψ due to the limitation imposed by (13). Let us assume that in time slot k the maximum system load this system can achieve based on all users channel states and all possible user schedulings is $\psi^* = \max(\sum g_i)$. We will now consider the two cases mentioned above, that is, $0 < \psi \leq \psi^*$ and $\psi > \psi^*$.

2.2.1. Target Load Is Less than or Equal to Maximum System Load. If we assume $0 < \psi \leq \psi^*$, then the system load can achieve the target load, $\sum_i g_i = \psi$. Therefore, (13) can be rewritten as follows:

$$\min_{1 \leq i \leq B(k)} \left\{ \frac{p_i^{\max} h_i}{g_i} \right\} \geq \frac{\eta_0 W}{1 - \psi}, \quad g_i \leq \psi, \quad (15)$$

$$\text{therefore } \frac{p_i^{\max} h_i}{g_i} \geq \frac{\eta_0 W}{1 - \psi} \quad \forall i, 1 \leq i \leq B(k).$$

For each individual user, there is a limitation on the maximum power index that it can reach, given by (15)

$$g_i \leq (1 - \psi) \frac{p_i^{\max} h_i}{\eta_0 W}, \quad g_i \leq \psi. \quad (16)$$

2.2.2. Target Load Is Larger than Maximum System Load. If the target load is larger than the maximum system load, that is, $\psi > \psi^*$, it means there will be no feasible transmission power solution in (7) and (8) to achieve this target load and therefore the relationship in (15) does not hold any more. In this case, we simply apply the power index restriction of (16) to each user. The consequence is that the final achieved system load becomes $\sum_i g_i < \psi^* < \psi$ since $g_i \leq (1 - \psi) p_i^{\max} h_i / \eta_0 W < (1 - \psi^*) p_i^{\max} h_i / \eta_0 W$.

In fact, unless all possible transmission user sets are searched, it is unknown in advance whether or not the actual system load $\sum g_i$ can reach the chosen ψ . Therefore, applying (16) to the case $\psi > \psi^*$ unifies the definition of the power index range, within which a user can be assigned a feasible power index without knowing the value of ψ^* . One key principle and rule regarding the algorithm proposed in this paper is to assign to an individual user a power index that is less than or equal to its power index capacity. In the power index assignment algorithm described in Section 3.2, the situation where $\sum g_i < \psi$ may occur. However, it should be noted here that as proven by Theorem 1 later in the paper, the global optimal solution must be the one satisfying $\sum g_i = \psi$. The target load range where $\psi > \psi^*$ is then not possible to be the optimal solution. The intentionally introduced restriction of (16) in the case of $\psi > \psi^*$ allows the algorithm to rule out such values of ψ due to the fact that $\sum g_i < \psi$ in this case.

2.2.3. Definition of Power Index Capacity. Hence, given the system load ψ the maximum possible power index g_i a user can accept in (15) is determined by the maximum transmit power p_i^{\max} and the channel gain h_i .

Definition 1. In a CDMA system with $B(k)$ backlogged users at time slot k , given the target system power index ψ , the maximum power index that does not violate (13) for a single user whose channel gain is h_i is defined as the power index capacity (PIC) $\pi_i(h_i, \psi)$ of this user.

From (15), it can be easily found that the PIC of user i is

$$\pi_i(h_i, \psi) = \min \left\{ (1 - \psi) \frac{p_i^{\max} h_i}{\eta_0 W}, \psi \right\}. \quad (17)$$

Note that in (17) the power index capacity is limited by the target system power index. This is reasonable since a power index capacity that is greater than ψ will have no practical meaning and application. Furthermore, since our focus in this paper is to find an optimal scheduling policy as well as the optimal system load ψ , the value of ψ in (17) is not determined in advance.

Intuitively, the power index represents the relationship between the transmission power and the corresponding interference that is caused to other users. If we considered that the total system power index is fixed to ψ , larger power index g_i for user i indicates that it has relatively higher signal-to-interference ratio compared to the other users with smaller power index, while at the same time it causes more interference to them. Accordingly, users with high-power indices may lower their transmission power to reduce the interference they may cause, which in turn means that they will have smaller power index to limit the intracell interference of the system, and therefore satisfy (13) that guarantees the existence of a feasible transmission power solution.

3. Problem Transformation and Optimal Solution

3.1. Problem Transformation. The corresponding constraints in terms of the power index can be represented as follows:

$$\max Z = \sum_{i=1}^{B(k)} w_i f_r(g_i, \gamma_i), \quad (18)$$

$$\sum_{i=1}^{B(k)} g_i \leq \psi, \quad (19)$$

$$g_i \leq \pi_i(h_i, \psi), \quad 1 \leq i \leq B(k), \quad (20)$$

$$0 \leq \psi < 1. \quad (21)$$

Note that in the objective function we represent the rate $r_i = f_r(g_i, \gamma_i)$ as a function of power index g_i , where

$$f_r(g_i, \gamma_i) = \frac{g_i}{1 - g_i} \frac{W}{\gamma_i}, \quad (22)$$

which converts the power index into transmission rate and can be easily derived from (14) by replacing G_i with W/r_i .

In the following, let $\mathbf{V} = \{v_1, v_2, \dots, v_i, \dots\}$ denote the set that contains all the power and rate vectors that satisfy constraints (7) and (8) and $v_i = \{p_{1,i}, p_{2,i}, \dots, p_{B(k),i}, r_{1,i}, r_{2,i}, \dots, r_{B(k),i}\}$. The elements $p_{j,i}$ and $r_{j,i}$ represent the transmit power and rate of the j th user in the i th vector. Similarly, we define another set \mathbf{V}' containing the power and rate vectors v'_i that satisfy constraints (19), (20), and (21). By definition, it is obvious that any power and rate vector $v_i \in \mathbf{V}$ is feasible. However, since in constraint (21), ψ may take a value, that is, close to 1, the required transmit power could also accordingly become larger than maximum allowable transmit power p_i^{\max} if we simply look at the result from (15). The following proposition states that if perfect power control is assumed, for any rate (or power index) vector that satisfies constraints (19), (20), and (21), there always exists a feasible transmit power vector.

Proposition 2. *If the power index assignment for all $B(k)$ backlogged users satisfies constraints (19), (20), and (21), there always exists a feasible transmit power assignment, that is, $p_i < p_i^{\max}$ for $1 \leq i \leq B(k)$.*

Proof. Let vector $\mathbf{g} = \{g_1, g_2, \dots, g_{B(k)}\}$ be the power index vector that satisfies constraints (19), (20), and (21). Denote $\psi = \sum_{i=1}^{B(k)} g_i$ the sum of all power indices in vector \mathbf{g} . From the definition of power index capacity, the power index capacity of each user is $\pi_i(h_i, \psi)$ and $g_i \leq \pi_i(h_i, \psi)$. Based on Definition 1 and (17), we have the following relation:

$$\psi \leq 1 - \frac{\eta_0 W \cdot \pi_i(h_i, \psi)}{p_i^{\max} h_i} \leq 1 - \frac{\eta_0 W \cdot g_i}{p_i^{\max} h_i}. \quad (23)$$

Hence, for any user i , the transmit rate may be chosen within range

$$p_i^{\max} \frac{g_i}{\pi_i(h_i, \psi)} \leq p_i \leq p_i^{\max}, \quad (24)$$

which still satisfies the above inequality and proves this proposition. The power control of the CDMA system will choose the minimal transmit power, that meets the required SINR. \square

The following proposition proves that the two sets \mathbf{V} and \mathbf{V}' contain the same elements, which means that (19), (20), (21) and (7), (8) impose the same constraints over our problem.

Proposition 3. *Any vector $v_i \in \mathbf{V}$ is also included in set \mathbf{V}' , while any vector $v'_i \in \mathbf{V}'$ is also included in set \mathbf{V} .*

Proof. Suppose that $v_i \in \mathbf{V}$, and therefore it satisfies constraints (7), (8). It is apparent that $p_{j,i} \leq p_j^{\max}$. Since, as shown earlier, constraints (7), and (8) can also be represented by (13) [25], v_i also satisfies (13). Using function (22), we can convert the rate vector $\{r_{1,i}, r_{2,i}, \dots, r_{B(k),i}\}$ into the corresponding power index vector $\{g_{1,i}, g_{2,i}, \dots, g_{B(k),i}\}$. Let $\psi = \sum_{j=1}^{B(k)} g_{j,i}$. For a feasible power and rate vector, with known ψ ($0 \leq \psi < 1$ [25]), we can find each user power index capacity $\pi_j(h_j, \psi)$. Since v_i satisfies (13), based on Proposition 2 and the definition of power index capacity, we conclude that $g_{j,i} \leq \pi_j(h_j, \psi)$. That means that the assigned powers and rates in v_i also satisfy the constraints (19), (20), and (21). Therefore, $v_i \in \mathbf{V}'$.

Let us consider vector $v'_i = \{p'_{1,i}, p'_{2,i}, \dots, p'_{B(k),i}, r'_{1,i}, r'_{2,i}, \dots, r'_{B(k),i}\} \in \mathbf{V}'$. As before, the rate vector part can be converted to corresponding power index vector $\{g'_{1,i}, g'_{2,i}, \dots, g'_{B(k),i}\}$. Let $\psi' = \sum_{j=1}^{B(k)} g'_{j,i}$ and hence $g'_{j,i} \leq \pi'_j(h_j, \psi')$ due to constraints (19), (20), and (21). Note that for the case where $\psi' > \sum_{j=1}^{B(k)} g'_{j,i}$, $\pi'_j(h_j, \psi) \geq \pi'_j(h_j, \psi')$. Based on the previous discussion, we can easily conclude that the power vector is feasible. Therefore,

$$\psi \leq 1 - \frac{\eta_0 W \cdot g'_{j,i}}{p'_{j,i} h_j}, \quad (25)$$

which satisfies (13), for user j , $1 \leq j \leq B(k)$. Therefore, $v'_i \in \mathbf{V}$. \square

The above proposition shows that the optimal solution can also be obtained with the new constraints since they define the same solution set. Please note that, as mentioned before, the fairness constraints in the original problem are replaced by parameters w'_i s. The choice of the proper values of w'_i s that maintain fairness is discussed in detail later in this paper.

Among the new constraints, the right-hand sides of inequalities (19) and (20) are not fixed values, but are functions of the selected target system load ψ . Hence, whether or not the final solution is feasible also depends on the choice of ψ . For any value of $\psi \in [0, 1)$, there could be many feasible solutions among which one will be the optimal. Moreover, there must exist an optimal system load ψ^* that can achieve the overall best solution. It is natural to regard the objective Z as the function of system load ψ , $Z = F(\psi)$, and thus Z is the local optimal result at some specific ψ . The maximum Z is achieved when $\psi = \psi^*$. The

ultimate objective of the proposed method is to find this optimal ψ^* and the optimal power index assignment vector under it.

In Sections 3.2 and 3.3, we propose a two-step approach to solve the optimization problem (17)–(20). More specifically, in the first step (Section 3.2), we assume a fixed ψ and then given that fixed parameter ψ we propose a simple method (greedy algorithm) trying to find the optimal set of users to receive service. However, this optimality is not a global optimality. In general, as mentioned before, ψ could get any value within the interval $[0, 1)$. The global optimal solution can be obtained when parameter ψ is chosen to be the optimal one ψ^* . The actual objective of the second step of our approach (Section 3.3) is to find this optimal ψ^* , by which the global optimal set of users that will be scheduled to receive service can be identified.

3.2. Greedy Algorithm for a Given System Load. Before obtaining the best system load, we first discuss how to find the local best solution. Assuming that the value of $\psi \in [0, 1)$ is known, the right-hand sides of (19) and (20) can be determined. Combining the two constraints together, we can express the optimization problem (18) by replacing g_i with $\pi_i(h_i, \psi)x_i$, $0 \leq x_i \leq 1$ as follows:

$$\begin{aligned} \max Z &= \sum_{i=1}^{B(k)} w_i f_r(\pi_i(h_i, \psi)x_i, \gamma_i), \\ \text{s.t.} \quad &\sum_{i=1}^{B(k)} \pi_i(h_i, \psi)x_i \leq \psi, \quad 0 \leq x_i \leq 1. \end{aligned} \quad (26)$$

Note that (26) is a nonlinear continuous knapsack problem with the x_i taking continuous values between 0 and 1. In general, solving this type of problem is proven to be difficult or even impossible in some cases [27]. However, Proposition 1 limits the transmit power of a user i , to either p_i^{\max} or 0 for the optimal solution. This condition provides a possible method to solve the above nonlinear knapsack problem. Without loss of generality, we suppose that the optimal solution is when the first K users transmit at their maximum power, $p_i = p_i^{\max}$, $1 \leq i \leq K$. The optimal system load is $\psi^* = \sum_{i=1}^K g_i$. The following theorem states that the power index of an individual user is equal to its power index capacity under ψ^* , that is, $g_i = \pi(h_i, \psi^*)$.

Theorem 1. *Let the optimal solution allow K users to transmit at their maximum power and the system achieves the system load ψ^* . The power index that an individual user received in this case is equal to its power index capacity, that is, $g_i = \pi(h_i, \psi^*)$.*

Proof. For those users whose transmit powers are zero, the corresponding power index capacities are also zero. Therefore, their power indices are zero as well. Without loss of generality, we assume that the K users under consideration are identified as follows: $1 \leq i \leq K$. Based on Proposition 1,

we have

$$\frac{h_i p_i^{\max} G_i}{\sum_{j=1, j \neq i}^{B(k)} h_j p_j + W \eta_0} = \gamma_i, \quad \text{for } 1 \leq i \leq K. \quad (27)$$

Performing some manipulations in these K equations, we have

$$\frac{h_i p_i^{\max}}{g_i} \left(1 - \sum_{i=1}^K g_i \right) = W \eta_0, \quad \text{for } 1 \leq i \leq K. \quad (28)$$

Letting $\psi^* = \sum_{i=1}^K g_i$, we obtain g_i as

$$g_i = \frac{h_i p_i^{\max} (1 - \psi^*)}{W \eta_0}. \quad (29)$$

From the definition of power index capacity, we find that $g_i = \pi(h_i, \psi^*)$. \square

With reference to the optimal solution of problem (26), we can prove the following theorem.

Theorem 2. *The optimal solution of the constrained optimization problem (26) can be obtained by solving the following linear 0-1 knapsack problem:*

$$\begin{aligned} \max Z &= \sum_{i=1}^{B(k)} w_i \frac{W}{\gamma_i} \frac{\pi_i(h_i, \psi)}{1 - \pi_i(h_i, \psi)} x_i, \\ \text{s.t. } \sum_{i=1}^{B(k)} \pi_i(h_i, \psi) x_i &\leq \psi, \quad x_i = \{0, 1\}. \end{aligned} \quad (30)$$

Proof. Since $f_r(x, \gamma_i) = (W/\gamma_i)(x/(1-x))$ for user i , we present the objective function of (26) as follows:

$$\max Z' = \sum_{i=1}^{B(k)} w_i \frac{W}{\gamma_i} \frac{\pi_i(h_i, \psi)}{1 - \pi_i(h_i, \psi)} x_i. \quad (31)$$

Based on Proposition 1, we know that the optimal solution is achieved when the transmit power of a user i is either p_i^{\max} or 0. According to Theorem 1, in terms of power index that means that users are assigned either their power index capacity or 0 for the chosen system load ψ . In the above relation (31), the solution for x_i is either 1 or 0. Therefore, we can modify (31) as follows without changing the final optimal solution:

$$\max Z = \sum_{i=1}^{B(k)} w_i \frac{W}{\gamma_i} \frac{\pi_i(h_i, \psi)}{1 - \pi_i(h_i, \psi)} x_i, \quad (32)$$

where $x_i = \{0, 1\}$. \square

Instead of solving for the optimal solution of the above integer knapsack problem (30), which is in principle NP-hard, we utilize a greedy algorithm (GA) in order to obtain an approximate solution. Let Z_a denote the result achieved by the approximate solution, while Z and Z_c denote the corresponding results of the optimal solutions for the integer

and continuous knapsack problems, respectively. It has been proven that $Z_a \leq Z \leq Z_c$ [28]. Furthermore, let

$$\alpha_i \triangleq \frac{W}{\gamma_i (1 - \pi_i(h_i, \psi))}, \quad (33)$$

which is a constant value for an individual user. Let us further suppose that all backlogged users are sorted in descending order according to $w_i(k)\alpha_i$, that is, $w_i(k)\alpha_i \geq w_j(k)\alpha_j$, for $i < j$. If it is not the case, these values can be sorted in $O(n \log n)$ time through an efficient procedure. Thus, the optimal continuous solution of problem (30) is given by

$$\begin{aligned} x_i &= 1, \quad \text{for } i < s, \\ x_j &= 0, \quad \text{for } j > s, \\ x_s &= \frac{\psi - \sum_{i < s} \pi_i(h_i, \psi)}{\pi_s(h_s, \psi)}. \end{aligned} \quad (34)$$

An algorithm that finds the critical point s within $O(n)$ time in a system with n users is provided in [28]. Based on solution (34), the greedy algorithm (GA) obtains the approximate solution U as follows:

$$U = \max\{U_1, U_2\}, \quad (35)$$

where

$$\begin{aligned} U_1 &= \begin{cases} x_i = 1, & \text{for } i < s, \\ x_j = 0, & \text{for } j \geq s, \end{cases} \\ U_2 &= \begin{cases} x_i = 1, & \text{for } i = s, \\ x_j = 0, & \text{for } i \neq s. \end{cases} \end{aligned} \quad (36)$$

It has been shown in [28] that in worst case $Z_a/Z = 1/2$. Let Z represent the result that corresponds to the integer solution of (32) when ψ is assigned a value from $[0, 1)$, and Z^* be the result when $\psi = \psi^*$. From the definition of ψ^* , we know that Z^* is the maximum value among all Z , that is, $Z^* = \max_{\psi} \{Z\}$. Based on Proposition 1 and the analysis in the previous subsection, it is easy to find that $\psi^* = \sum_i \pi_i(h_i, \psi^*) x_i$, $x_i = \{0, 1\}$. Therefore, when the optimal system power index ψ^* is chosen, $Z_a = Z = Z_c = Z^*$. Since $Z_a \leq Z \leq Z^*$ and the equality $Z_a = Z^*$ holds only when $\psi = \psi^*$, and the optimal solution can be obtained.

3.3. Optimal System Load. As we discussed in the last subsection the optimal solution of problem (26) depends on the selected system load ψ . Relation (17) shows that the power index capacity increases as ψ decreases. At the first point when $\pi_i = \psi$, the power index capacity reaches its largest value and then it decreases linearly following the value of ψ . Although a smaller value of ψ may increase the single user power index capacity at some range, the finally achieved objective function could be low due to the small system load ψ . On the other hand, setting large ψ reduces the individual user power index capacity as (17) indicates. The consequence of smaller power index capacity is that more users are required to share ψ , and probably a small objective function

should be used due to the concavity of function $f_r(x, \gamma_i)$ that converts the power index to throughput. Therefore, whether or not the objective function reaches its maximum value depends not only on the value of the system load ψ , but also on how it is shared among the candidate users. There must exist an optimal value of system load ψ^* that can achieve the maximum weighted rate.

Let the power index vector \mathbf{g} denote the optimal solution, which can be found through the method described in the previous section for a given specific value of ψ . Apparently, \mathbf{g} is a function of ψ . The objective function (18) is the sum of individual weighted rates that are obtained from \mathbf{g} using function $f_r(x, \gamma_i)$. Therefore, Z can also be regarded as a function of ψ . Let $FZ(\psi)$ be the function that gives the maximum value of the sum of weighted rates at ψ . Then the original optimization problem can be rewritten as follows:

$$\begin{aligned} \max Z &= FZ(\psi), \\ \text{s.t. } &0 \leq \psi < 1. \end{aligned} \quad (37)$$

The optimal solution ψ^* of the above problem and its corresponding power index assignment by (34) with $\psi = \psi^*$ provides the final optimal solution of (18).

Problem (37) is a simple unconstrained maximization problem that searches for the maximum Z within the interval $[0, 1)$. The disadvantage of (37) is that it does not have an explicit expression. Hence, algorithms that rely on the first- or second-order derivatives will not be applicable in this case. Therefore, the searching process depends on the result of (34). Note that every time when a new value of ψ is chosen, the order of $w_i(k)\alpha_i$ may be different from that of previous ψ .

The time of calculating the best result for a newly chosen ψ , including the time of reordering the users (if needed), is easily obtained as $O(n \log n) + O(n) = O(n \log n)$ if n is assumed to be large enough. Moreover, there are many possible local maximum points within the range $0 \leq \psi < 1$. The final optimal ψ must be a global best value. Although in [29] many searching algorithms on how to locate the minimum/maximum solution within a range are described, to make these algorithms effective there must be only one extreme point in the specified range. However, in general it is not possible to know the range which contains only the global optimal value. Thus, an exhaustive search within $[0, 1)$ would be needed. However, the following proposition provides a lower bound ψ^0 with respect to the searching range instead of 0 in order to restrict the corresponding feasible searching range.

Proposition 4. *The lower bound of the feasible searching range is given by*

$$\psi^0 = \min_{1 \leq i \leq B(k)} \left(\frac{\zeta_i}{1 + \zeta_i} \right), \quad \text{where } \zeta_i \triangleq p_i^{\max} \frac{h_i}{\eta_0} W. \quad (38)$$

Proof. With the decrease of the target system load ψ , the individual power index, provided by (14), will keep increasing till ψ reaches the point ψ_i for user i , that is $(1 - \psi_i)\zeta_i = \psi_i$. With respect to user i , if $\psi \leq \psi_i$ its power index

$\pi_i(h_i, \psi) = \psi$. ψ_i is given by $\psi_i = \zeta_i / (1 + \zeta_i)$, which varies with different users since their ζ_i are not likely the same. Let ψ^0 be the minimum among all ψ_i 's. Once $\psi < \psi^0$ all backlogged users will have the same power index capacities $\pi_i(h_i, \psi) = \psi$. Define a small increment $\Delta\psi$ and let $\psi' = \psi + \Delta\psi < \psi^0$. Apparently, for all users their power indices will all have small increment $\Delta\psi$ such that $\pi_i(h_i, \psi') = \psi + \Delta\psi$. Maintaining the previous power index assignment and giving $\Delta\psi$ to any backlogged user will help increase the objective function (18). We hence can keep adding $\Delta\psi$ to ψ till it reaches $\psi^0 = \Delta\psi + \psi$, which proves this proposition. \square

Since the optimal ψ can reside between ψ^0 and 1, we need to calculate a series of sample values after every interval $\Delta\psi$. Apparently, the smaller the $\Delta\psi$, the more samples we get and thereby the more accurate is the obtained result. On the other hand, it also increases the required computational time and power.

Therefore, in practice we only use reasonably small $\Delta\psi$ in order to reduce the corresponding computational power and complexity, while still obtain a good approximation of the optimal solution. It should be noted though that in theory when $\Delta\psi$ becomes infinitely small the above methodology can be used to find the optimal solution. Specifically, there exists an algorithm with complexity of $O(n^4 \log n)$ that guarantees the finding of the optimal solution, however its high complexity limits its applicability for real-time computations and can be used only for benchmarking purposes. Let us assume that the order in (34) is known and fixed. Under this condition, there are only $B(k)$ possible results satisfying the optimal condition in Proposition 1, that is, try the maximum transmission power in the fixed order with number of users from 1 to $B(k)$. The maximum result is the optimal one. For any two users in the possible system load range from $(0, 1)$, their order of $w_i(k)\alpha_i$ will change at most three times. Therefore, there are totally $1.5B(k)(B(k) - 1)$ order changes for $B(k)$ users. Every order change requires first the sorting operation and then the comparison operation that have complexity of $O(n \log n)$ and $O(n)$, respectively, which makes the overall complexity of this method $O(n^4 \log n)$.

The optimal algorithm is described as follows.

(1) Find the m points of target system load, $x_1 < x_2 < \dots < x_m$, between $[0, 1)$, where the users change their orders in $w_i(k)\alpha_i$. Such points represent actually any point that for any two users i and j , $w_i(k)\alpha_i = w_j(k)\alpha_j$, which is,

$$w_j(k)(1 - \pi_i(h_i, \psi)) = w_i(k)(1 - \pi_j(h_j, \psi)). \quad (39)$$

Based on the definition of power index capacity in (17), the above equation will have at most three solutions.

(2) Once the order is fixed, sort all $B(k)$ users by $w_i(k)\alpha_i$ in descending order. The value α_i can be calculated using any number between $[x_i, x_{i+1})$ since the order will be the same within this range.

(3) Perform $B(k)$ rounds of calculation of objective function (6). In round i , let the largest i users transmit with their largest transmit powers.

(4) Compare the result of round $(i + 1)$ to that of round i . If the result in round $(i + 1)$ is less than round i , then stop

the calculation. In that case, the result of round i is the best result in this order between x_i and x_{i+1} .

(5) The largest result obtained in step (4) is the global optimal solution.

Once the order is fixed in the range $[x_i, x_{i+1}]$ at step (2), the method provided in Section 3.2 that finds the best local solution can be applied here, which will provide the largest n , $1 \leq n \leq B(k)$, users with this fixed order. The only difference is that the target system load is not provided directly by a specific known value ψ , but lies within a specific range. Based on Proposition 1, according to which the users allowed to transmit will use their maximum transmission power, we perform $B(k)$ rounds of calculation in step (3) and compare the results to find the optimal n users.

3.4. Fairness Conditions. As mentioned before, fairness is controlled by the vector $\mathbf{w} = \{w_1, w_2, \dots, w_{B(k)}\}$. When changing the values of w_i , we are actually pursuing a set of optimal fixed values $\mathbf{w}^* = \{w_1^*, w_2^*, \dots, w_{B(k)}^*\}$ that balance the rate of users with varying channel conditions and hence keep fairness. Since we do not know in advance the exact distribution of the channel conditions, and the number of users may also change, it is difficult to obtain vector \mathbf{w}^* in advance. Therefore, a real-time algorithm is required that is capable of converging w_i toward w_i^* , while maintaining the asymptotic fairness. Stochastic approximation algorithm has been proven to be effective in estimating such parameters. Note that this algorithm has been implemented in [14, 15] in order to solve similar problems. Generally, the stochastic approximation algorithm is a recursive procedure for finding the root of a real-value function $f(x)$. In many practical cases, the form of function $f(x)$ is unknown. Therefore, the result with the input variable x cannot be obtained directly. Instead, the observations of the results, sometimes with noise, will be taken. It has been proven that the root of $f(x)$ can be estimated with the observation $Y_n = f(x_n)$ by the following procedure:

$$x_{n+1} = x_n - \varepsilon_n Y_n, \quad (40)$$

where $\varepsilon_n > 0$, $\varepsilon_n \rightarrow 0$. We can simply let $\varepsilon_n = 1/n$. In most situations, the value of $f(x_n)$ may not be directly available, but instead the $f(x_n) + e_n$, where e_n is the observation noise. In this case, the above approximation approach still applies, with the observed value replaced by $Y_n = f(x_n) + e_n$. The convergence of x_n to the root requires $E(e_n) = 0$.

Here, we define our function $f(\mathbf{w}) = \{f(w_1), f(w_2), \dots, f(w_{B(k)})\}$ as follows:

$$f(w_i) = \frac{E[r_i(n)]}{E[\sum_j r_j(n)]} - \frac{\phi_i}{\sum_j \phi_j}, \quad (41)$$

whose root w_i^* will make $f(w_i) = 0$ which satisfies the fairness condition (3). The noise observation Y_n in our case is:

$$Y_n = \frac{r_i(n)}{E[\sum_j r_j(n)]} - \frac{\phi_i}{\sum_j \phi_j}. \quad (42)$$

It is easy to prove that the mean of noise $E[e_n] = E[f(w_i) - Y_n] = 0$. Therefore, the value of w_i^* is then recursively obtained by

$$w_i(n+1) = w_i(n) - \frac{Y_n}{n}. \quad (43)$$

However, Y_n need to know the mean of total system throughput $E[\sum_j r_j(n)]$. We use a smoothed value $\bar{R}(n)$ to approximate $E[\sum_j r_j(n)]$ and update $\bar{R}(n)$ as follows:

$$\bar{R}(n) = \bar{R}(n-1)\beta + (1-\beta)\sum_j r_j(n-1), \quad (44)$$

where β is the smooth factor which determines how the estimated $\bar{R}(n)$ follows the change of actual achieved system throughput. In the remaining of the paper, throughout the performance evaluation of our approach, the value $\beta = 0.999$ is chosen. The numerical results presented in Sections 4.2.2 and 4.2.3, with respect to the convergence of w_i 's and the achievable fairness, demonstrate that such a method is very effective in approximating the optimal values of w_i^* and therefore controlling and maintaining fairness.

4. Performance Evaluation

In this section, we evaluate the performance of the proposed method in terms of the achievable fairness and throughput, via modeling and simulation. Furthermore, to better understand the performance of the proposed scheduling algorithm—in the following we refer to as throughput maximization and fair scheduling (MAX-FAIR)—we compare it with the maximum throughput (MAX) scheme [16], which achieves the maximum total uplink throughput by allowing only the best k users in terms of their received power to transmit, and with the HDR algorithm [7, 9], which is a single user scheduling algorithm. The principles and operation of HDR basically refer to a proportional fair scheduling scheme, which can be used in the uplink scheduling to demonstrate the one-at-a-time proportional fair scheduling. Following the HDR principles the transmission of a single user at a given time slot is scheduled, with the data rates and slot lengths varying according to the specific channel condition. In the MAX scheme parameter, k is determined by iteratively comparing the throughput of best i users, $1 \leq i \leq N$, where N is the total number of users. The throughput achieved by MAX scheme is regarded as the upper bound throughput in the uplink CDMA scheduling. On the other hand, since HDR achieves temporal fairness, we consider it here to mainly observe the difference between temporal fairness and throughput fairness and their corresponding advantages in specific cases.

4.1. Model and Assumptions. Throughout our numerical study, we consider a single cell DS-CDMA multirate system with multiple active users. All active users are continuously backlogged during the simulation and generate packets with average size of 320 bytes. The maximum transmission power is assumed the same for all users, that is, $p_i^{\max} = 2$ Watts,

while the system chip rate is $W = 1.2288 \times 10^6$ chip/s as defined in IS-95 and the required SINR is $\gamma_i = 8$ dB for data service, the same for all users. The transmission time is divided into 1 millisecond equal length slots, which is the algorithm scheduling interval, while the simulation lasts for 1.7×10^5 slots.

To study the impact of the channel condition variations on the system throughput and fairness performance, we model the channels through an 8-state Markov-Rayleigh fading channel model [30]. According to this model, the channel has equal steady-state probabilities of being in any of the eight states. We also assume that the coherent time is much larger than the length of a time-slot, hence the channel state is assumed to be constant within a time slot. At the beginning of each time slot, the channel model decides to transit to a new state, which can only be itself or one of its neighbor states, that is, from state s to $s+1$, or $s-1$. Table 1 summarizes the state transition probabilities for all the eight states.

Furthermore, four different cases with respect to the ranges of the average SNRs that are assigned to the various users are considered. Specifically, Table 2 presents the corresponding ranges and lists the assignment of the average SNRs for each user for a seven-user scenario, under all these cases. The four different cases represent four different scenarios with respect to the SNR as follows (from top to bottom): large SNR range with low SNR users, low SNR, middle SNR, and high SNR. In the next subsection, we evaluate the performance of MAX-FAIR, MAX, and HDR methods under all four cases and compare their corresponding achieved throughput and fairness.

In most of the numerical results presented in the next subsection, unless otherwise is explicitly indicated, all users are assumed to have the same weight. Such a scenario allows us to better understand and compare the achievable performances of the various scheduling schemes, when users have different channel conditions. However, the operation and effectiveness of the proposed MAX-FAIR policy is also demonstrated in an environment, where users present different weights.

4.2. Numerical Results and Discussion. The numerical results presented in Sections 4.2.1 and 4.2.2 refer mainly to the impact of some of the parameters associated with the proposed MAX-FAIR algorithm on its operation and achievable performance and allow us to obtain a better understanding of its operational characteristics and properties. Then in Sections 4.2.3 and 4.2.4, comparative results about the achievable throughput and fairness of the MAX-FAIR, MAX and HDR algorithms are presented.

4.2.1. Finite System Power Index Samples. Figure 1 shows the sensitivity of the weighted throughput achieved by the MAX-FAIR algorithm as a function of the number of samples used to obtain these values. The last point in the horizontal axis corresponds to the optimal value. It should be noted that in the vertical axis, the depicted weighted throughputs are normalized over the optimal value. Moreover, the different

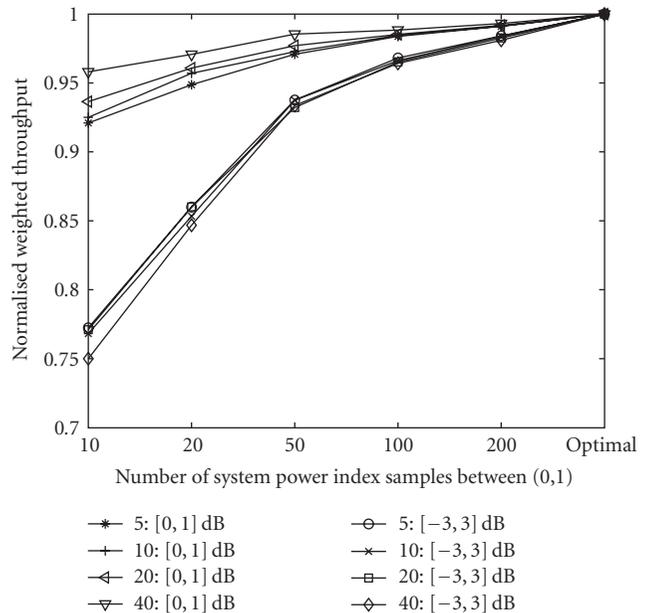


FIGURE 1: The impact of number of samples on the weighted throughput (MAX-FAIR).

curves provided in this figure correspond to different combinations of the SNR ranges and the number of active users. As can be seen, the more samples we choose, the closer is the obtained maximum value to the optimal value, which clearly presents the tradeoff between the accuracy and the required computational power, as discussed before in Section 3.3. For instance, we observe that in the cases with small SNR range (e.g., [0,1] dB), even 20 samples are sufficient to get satisfactory results, while for the cases with larger SNR range (e.g., [-3,3] dB), more samples may be required.

Furthermore, as it can be observed from this figure, for the case of [0,1] dB, the larger the number of active users in the system, the less sensitive is the achievable maximum result to the number of samples (i.e., the slope of the corresponding curve becomes smoother as the number of active users increases). On the other hand, when there are users with high SNR values (e.g., [-3,3] dB), the increasing number of active users makes the achieved throughput drop slightly for small number of samples. This difference in the system behavior is closely related to a different number of simultaneously served users, under different SNR ranges and channel conditions, as depicted by the different observed service patterns in Figure 2.

Specifically, in Figure 2, we present the probabilities of the number of simultaneously served users in each scheduling cycle. For this experiment, we consider 40 backlogged users in the system and perform 200 trials. In each trial, users are randomly assigned the SNRs in the designated SNR range, following the 8-state model [30] described in Section 4.1. We observe that when there are users having high SNR values, for example, in the cases of [-3,3] dB and [2,4] dB, only a small number of users (at most 2 in this experiment), are served concurrently. However, in the case

TABLE 1: Channel state transition probability.

	$s = 1$	$s = 2$	$s = 3$	$s = 4$	$s = 5$	$s = 6$	$s = 7$	$s = 8$
$p_{s,s}$	0.9304	0.8419	0.8170	0.8216	0.8349	0.8590	0.8945	0.9616
$p_{s,s-1}$	0	0.069	0.0879	0.0894	0.0876	0.0777	0.0637	0.0384
$p_{s,s+1}$	0.0696	0.0891	0.0951	0.089	0.0775	0.0633	0.0418	0

TABLE 2: Simulation cases with different SNR(dB) distribution.

	1	2	3	4	5	6	7
Case: [-3, 3]	-3	-3	-3	0	0	0	3
Case: [-4, -2]	-4	-4	-4	-3	-3	-3	-2
Case: [0, 1]	0	0	0	1	1	1	1
Case: [2, 4]	2	2	2	3	3	3	4

that all users have small SNR values, for example, in the case of [-4, -2] dB, the number of simultaneously served users increases significantly (it is distributed between 4 and 17 in our case as can be seen by Figure 2). Such user distribution indicates that in the case that a single user cannot consume all the system resources (e.g., the case where users have low SNR values), more users will be scheduled simultaneously in order to achieve a more efficient resource utilization and as a result increase the total system throughput. This also demonstrates the advantage of our proposed scheduling algorithm over the one-by-one scheduling algorithms that have been proposed in literature. As a result, with respect to Figure 1, for the case of [0,1] dB, multiple users are scheduled to reach the maximal throughput. Increasing the number of active users enables the system to schedule more available candidates to achieve higher throughput, and therefore the achievable result is less sensitive to the number of samples. However, for the case [-3,3] dB at most only 1 or 2 users are scheduled for simultaneous transmission. In the following experiments and numerical results, we adopt the accuracy of 100 samples, which is sufficient to reach 95% of the optimal-weighted throughput.

4.2.2. Parameter Convergence by Stochastic Approximation. As described in Sections 2.1 and 3.4, parameters w_i 's are used to represent the fairness constraints in our optimization problem formulation. Figure 3 shows the dynamic change of parameters w_i 's as the system and time evolve, for two different cases that correspond to two different SNR ranges. A seven-user scenario is considered, while for demonstration purposes for each case the corresponding values of only two representative users are presented—one user with strong channel and one user with weak channel. As mentioned before, all the users are assigned the same weight in order to more clearly demonstrate the influence of the channel conditions on w_i 's. It can be seen by this figure that the converged values of w_i 's have the effect of compensating users with the weak channels and reducing the priority of users with strong channels in the scheduling policy. In fact, the converged values of w_i 's will make both users (weak and strong) to gain proper system resources and therefore achieve fair throughput. Please note that it is the relative values of w_i 's

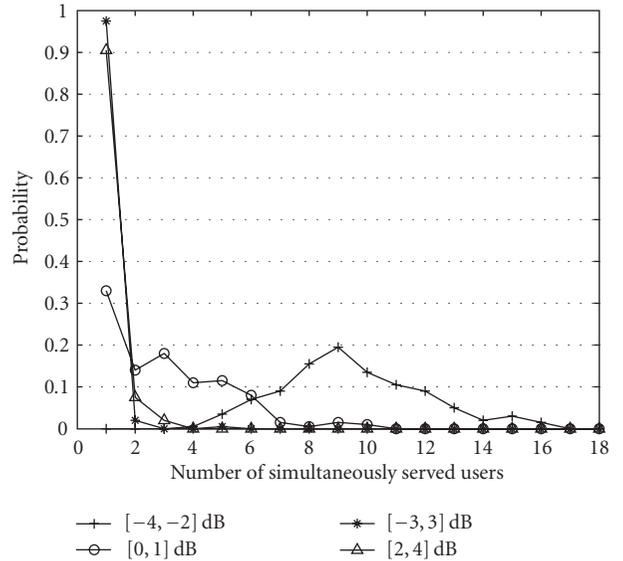


FIGURE 2: The service pattern under different channel conditions (i.e., SNRs) (MAX-FAIR).

that control the priority of accessing the system resources, and not their absolute values. Furthermore, it should be noted that the lower the average SNR of a weak user, the larger the gap between the weak user and a strong user, which has negative impact on the achievable system throughput, as we will see in the following subsection.

4.2.3. Throughput and Fairness Performance. Figure 4 shows the average throughputs of all the users under the MAX-FAIR, MAX, and HDR methods, for a seven-user scenario where the average SNR range is [-3,3] dB and the corresponding average SNR assignments to the seven users are as shown in Table 2. In order to better demonstrate the tradeoff between the computational complexity and the achievable throughput of MAX-FAIR approach, we obtained the corresponding results under two different cases with respect to the number of power index samples (i.e., 20 and 100 samples). As observed in this figure the MAX-FAIR with 100 power index samples achieves slightly higher throughput, however it requires five times the computational power of the MAX-FAIR with 20 power index samples.

When compared to other two scheduling schemes, MAX-FAIR presents the best throughput-fairness performance (balances the achievable throughput of all users) despite the variable channel conditions of the different users, which indicates that the fairness is well maintained under the proposed scheduling algorithm. As mentioned before in the paper, the main objective of HDR is to achieve

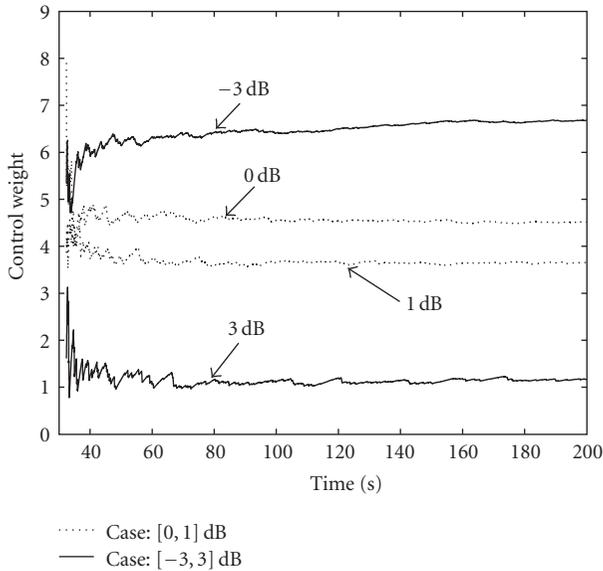


FIGURE 3: The convergence of w_i 's for different users and different SNR ranges (MAX-FAIR).

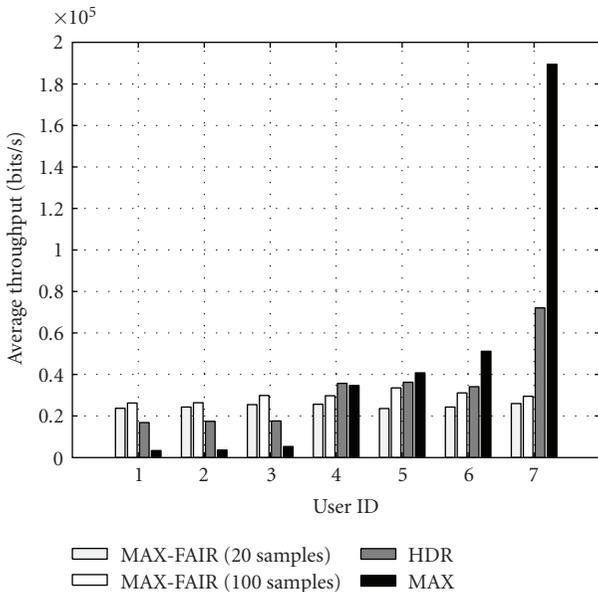


FIGURE 4: Average throughput for the $[-3, 3]$ dB case.

temporal fairness. Therefore, under HDR scheduling each user throughput is closely related to its channel conditions. That is why in Figure 4 we observe that users 1, 2, and 3 have smaller throughput than users 4, 5, and 6, while user 7 has the largest throughput under the HDR scheme. Under the MAX algorithm, user 7 consumes most of the system resources and achieves much higher throughput than the rest of the users due to the fact that the objective of MAX algorithm is to achieve the highest possible total system throughput, without however considering the fairness issue. In Figure 5, we further measure and evaluate the fairness performance by the standard deviation of the average throughput under all

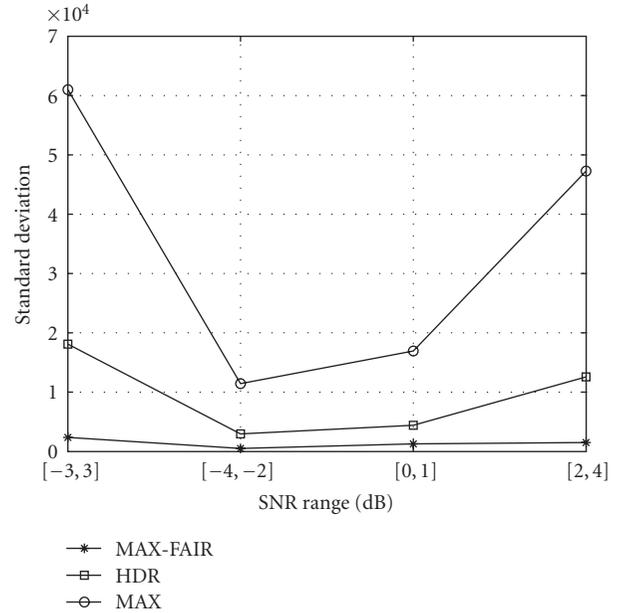


FIGURE 5: Standard deviation of achievable average throughputs.

the four different SNR cases. Among the three algorithms, MAX-FAIR algorithm has the smallest deviation for all the different cases under consideration, while the corresponding values change only slightly from case to case. We also find that in general the standard deviation increases as the SNRs become higher. This happens because small fluctuation of w_i results in larger throughput change, if all the users have higher SNR levels.

Figure 6 compares the corresponding average system throughputs of the three algorithms under evaluation, for the different SNR ranges (cases). As we expected, MAX-FAIR outperforms HDR in most cases due to the simultaneous scheduling of multiple users, as has been demonstrated in Figure 2, and consequently results in higher resource utilization. However, in the case of SNR range of $[-3, 3]$ dB, MAX-FAIR achieves slightly lower throughput than the HDR. The reason of that resides in the different fairness criteria considered and satisfied in these two algorithms, namely, the throughput fairness and temporal fairness. If we examine again Figure 3, we notice that users that have low average SNR (-3 dB) (e.g., users 1, 2, and 3) finally converge to a high w_i , which enables them to have equal opportunity to transmit under the MAX-FAIR scheduling policy. Due to their weak channel conditions, their average throughputs will be low and hence the total system throughput will become lower because of the satisfaction of the throughput fairness constraint. However, as explained before since access time is not the only resource to be shared among the users in these systems, considering throughput fairness instead of temporal fairness is more meaningful in these systems and environments, despite the slightly lower total throughput that can be achieved in some cases under this consideration. One possible alternative solution is to relax the fairness constraint if the QoS permits it. Our experiments have demonstrated

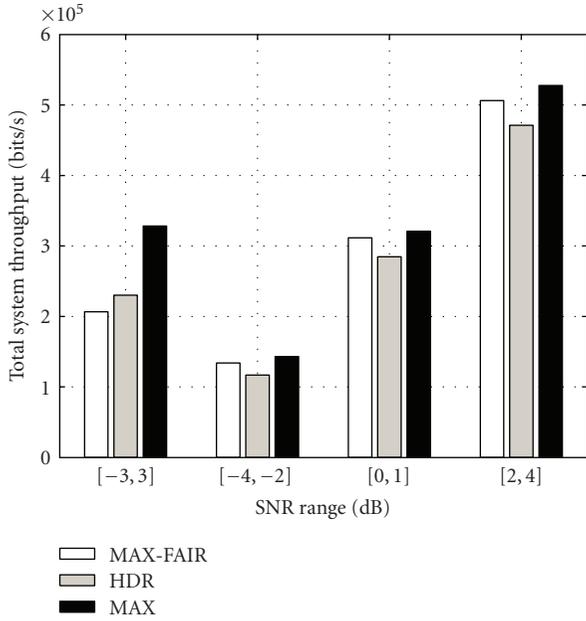


FIGURE 6: Achieved system throughput under different SNR ranges.

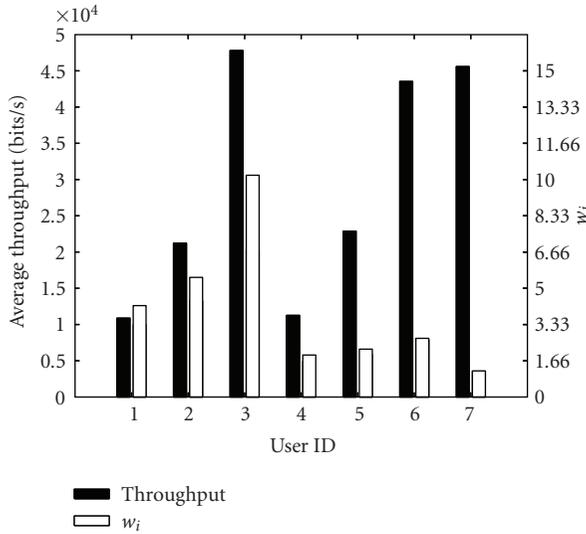


FIGURE 7: Average throughput under different QoS requirements (weights) by MAX-FAIR.

that after relaxing fairness to 85% of its original requirement, the MAX-FAIR catches up and outperforms the HDR.

In order to obtain a more in-depth understanding of the MAX-FAIR fairness operation, in Figure 7, we present the achieved average throughputs for all the seven users under MAX-FAIR scheme, for a scenario where the SNR range is assumed to be [-3,3]dB, and the users are assigned different weights. The different weights can be considered as the mapping of different QoS requirements. In this scenario, users 1 and 4 have weight 1, users 2 and 5 have weight 2, while users 3, 6, and 7 have weight 4. Figure 7 demonstrates that the MAX-FAIR successfully

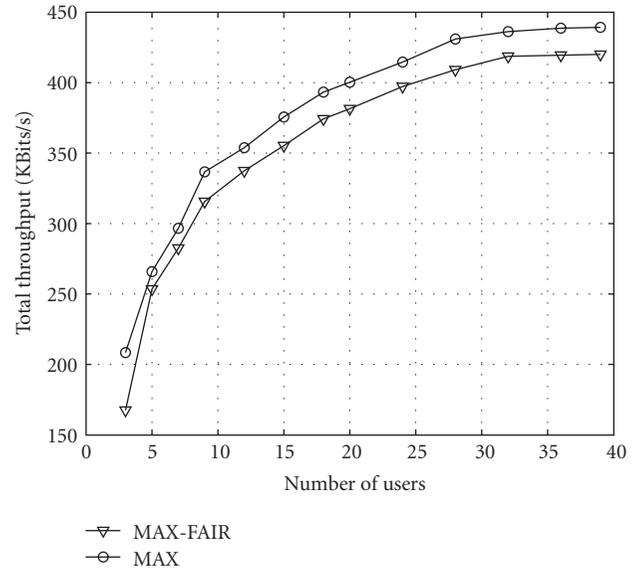


FIGURE 8: System throughput as a function of the number of backlogged users.

schedules the transmissions and distributes the resources so that the various users achieve throughput according to their corresponding assigned weights. Specifically users with weights 2 and 4 obtain, respectively, two times and four times the throughput achieved by users with weight 1. In this figure, we also present (on the right-hand side vertical axis) the converged values of parameters w_i 's. Here, the different values of w_i 's reflect both the channel condition variations and the weight differences. Please note that the relationship between w_i and weight is not linear due to the nonlinearity between the allocated resources and throughput.

4.2.4. Number of Users. Figure 8 shows the achieved total system throughput under MAX and MAX-FAIR algorithms as a function of the number of backlogged users, for the case where the users SNRs are located within [0,1]dB range. Please note that as mentioned before MAX algorithm provides the maximum uplink transmission throughput without considering the fairness property, and therefore is assumed to provide the upper bound throughput in uplink scheduling. From this figure, we can clearly observe the great advantage of the proposed MAX-FAIR approach and its ability to achieve very high throughput, while still maintaining fairness. When the number of backlogged users reaches a certain level, for example, 35 in this experiment, the throughput becomes flat for both MAX-FAIR and MAX, which means that the chances of improving the throughput by opportunistic scheduling with multiple users have been fully utilized.

5. Conclusions

In this paper, the CDMA uplink throughput maximization problem, while maintaining throughput fairness among the

various users, was considered. It was shown that such a problem can be expressed as a weighted throughput maximization problem, under certain power and QoS requirements, where the weights are the control parameters that reflect the fairness constraints. A stochastic approximation method was presented in order to effectively identify the required control parameters. The numerical results presented in the paper, with respect to the convergence of the control parameters and the achievable fairness, demonstrated that this method is very effective in approximating the optimal values and therefore controlling and maintaining fairness. Furthermore, the concept of power index capacity was used to represent all the corresponding constraints by the users power index capacities at some certain system power index. Based on this, the optimization problem under consideration was converted into a binary knapsack problem, where the optimal solution can be obtained through a global search within a specific range.

The performance of the proposed policy in terms of the achievable fairness and throughput was obtained via modeling and simulation and was compared with the performances of other scheduling algorithms. The corresponding results revealed the advantages of the proposed policy over other existing scheduling schemes and demonstrated that it achieves very high throughput, while satisfies the QoS requirements and maintains fairness among the users, under different channel conditions and requirements.

Acknowledgment

This work has been partially supported by EC EFIPSANS Project (INFISO-ICT-215549).

References

- [1] F. Adachi, M. Sawahashi, and H. Suda, "Wideband DS-CDMA for next-generation mobile communications systems," *IEEE Communications Magazine*, vol. 36, no. 9, pp. 56–69, 1998.
- [2] K. D. Wong and V. K. Varma, "Supporting real-time IP multimedia services in UMTS," *IEEE Communications Magazine*, vol. 41, no. 11, pp. 148–155, 2003.
- [3] R. Berezdivin, R. Breinig, and R. Topp, "Next generation wireless communications concepts and technologies," *IEEE Communications Magazine*, vol. 40, no. 3, pp. 108–116, 2002.
- [4] N. R. Sollenberger, N. Seshadri, and R. Cox, "The evolution of IS-136 TDMA for third-generation wireless services," *IEEE Personal Communications*, vol. 6, no. 3, pp. 8–18, 1999.
- [5] J. Ramis, L. Carrasco, G. Femenias, and F. Riera-Palou, "Scheduling algorithms for 4G wireless networks," *IFIP International Federation for Information Processing*, vol. 245, pp. 264–276, 2007.
- [6] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1277–1294, 2002.
- [7] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi, "CDMA/HDR: a bandwidth-efficient high-speed wireless data service for nomadic users," *IEEE Communications Magazine*, vol. 38, no. 7, pp. 70–77, 2000.
- [8] F. Berggren, S.-L. Kim, R. Jäntti, and J. Zander, "Joint power control and intracell scheduling of DS-CDMA nonreal time data," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 1860–1870, 2001.
- [9] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," in *Proceedings of the 51st IEEE Vehicular Technology Conference (VTC '00)*, vol. 3, pp. 1854–1858, Tokyo, Japan, May 2000.
- [10] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "CDMA data QoS scheduling on the forward link with variable channel conditions," Bell Labs Technical Memorandum 10009626-000404-05TM, Bell Labs, Paris, France, April 2000.
- [11] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–154, 2001.
- [12] S. Shakkottai and A. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: the exponential rule," Tech. Rep., Bell Labs, Paris, France, 2000.
- [13] X. Liu, E. K. P. Chong, and N. B. Shroff, "Transmission scheduling for efficient wireless utilization," in *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '01)*, vol. 2, pp. 776–785, Anchorage, Alaska, USA, April 2001.
- [14] X. Liu, E. K. P. Chong, and N. B. Shroff, "A framework for opportunistic scheduling in wireless networks," *Computer Networks*, vol. 41, no. 4, pp. 451–474, 2003.
- [15] Y. Liu and E. Knightly, "Opportunistic fair scheduling over multiple wireless channels," in *Proceedings of the 22nd Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '03)*, vol. 2, pp. 1106–1115, San Francisco, Calif, USA, March 2003.
- [16] S. A. Jafar and A. Goldsmith, "Adaptive multirate CDMA for uplink throughput maximization," *IEEE Transactions on Wireless Communications*, vol. 2, no. 2, pp. 218–228, 2003.
- [17] L. Xu, X. Shen, and J. W. Mark, "Dynamic bandwidth allocation with fair scheduling for WCDMA systems," *IEEE Wireless Communications*, vol. 9, no. 2, pp. 26–32, 2002.
- [18] C. Li and S. Papavassiliou, "Fair channel-adaptive rate scheduling in wireless networks with multirate multimedia services," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 10, pp. 1604–1614, 2003.
- [19] J. Cho and D. Hong, "Tradeoff analysis of throughput and fairness on CDMA packet downlinks with location-dependent QoS," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 1, pp. 259–271, 2005.
- [20] S. S. Kulkarni and C. Rosenberg, "Opportunistic scheduling policies for wireless systems with short term fairness constraints," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '03)*, vol. 1, pp. 533–537, San Francisco, Calif, USA, December 2003.
- [21] C. Li and S. Papavassiliou, "Opportunistic scheduling with short term fairness in wireless communication systems," in *Proceedings of the Conference on Information Sciences and Systems (CISS '04)*, pp. 167–172, Princeton, NJ, USA, March 2004.
- [22] S. Ramakrishna and J. M. Holtzman, "A scheme for throughput maximization in a dual-class CDMA system," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 6, pp. 830–844, 1998.
- [23] T. Ottosson and A. Svensson, "On schemes for multirate support in DS-CDMA systems," *Wireless Personal Communications*, vol. 6, no. 3, pp. 265–287, 1998.

- [24] S. Borst and P. Whiting, "Dynamic rate control algorithms for HDR throughput optimization," in *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '01)*, vol. 2, pp. 976–985, Anchorage, Alaska, USA, April 2001.
- [25] A. Sampath, P. S. Kumar, and J. M. Holtzman, "Power control and resource management for a multimedia CDMA wireless system," in *Proceedings of the 6th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '95)*, vol. 1, pp. 21–25, Toronto, Canada, September 1995.
- [26] C. Li and S. Papavassiliou, "Joint throughput maximization and fair scheduling in uplink DS-CDMA systems," in *Proceedings of IEEE/Sarnoff Symposium on Advances in Wired and Wireless Communication*, pp. 193–196, Princeton, NJ, USA, April 2004.
- [27] D. S. Hochbaum, "A nonlinear Knapsack problem," *Operations Research Letters*, vol. 17, no. 3, pp. 103–110, 1995.
- [28] S. Martello and P. Toth, *Knapsack Problems: Algorithms and Computer Implementations*, John Wiley & Sons, New York, NY, USA, 1990.
- [29] R. Miller, *Optimization: Foundations and Applications*, John Wiley & Sons, New York, NY, USA, 2000.
- [30] H. Wang and N. Moayeri, "Finite-state Markov channel—a useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163–171, 1995.