DIMACS Technical Report 2003-09
April 2003

Uplink Scheduling in CDMA Packet-Data Systems

by

Krishnan Kumaran, Lijun Qian[1]

[1]The authors are from the Math Research at Bell Labs, Lucent Tech., Murray Hill, NJ. E-mail: {kumaran,ljqian}@research.bell-labs.com.

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs-Research, Bell Labs, NEC Research Institute and Telcordia Technologies (formerly Bellcore).

DIMACS was founded as an NSF Science and Technology Center, and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

Uplink scheduling in wireless systems is gaining importance due to arising uplink intensive data services (ftp, image uploads etc.), which could be hampered by the currently in-built asymmetry in favor of the downlink. In this work, we propose and study algorithms for efficient uplink packet-data scheduling in a CDMA cell. The algorithms attempt to maximize system throughput under transmit power limitations on the mobiles assuming instantaneous knowledge of user queues and channels. However no channel statistics or traffic characterization is necessary. Apart from increasing throughput, the algorithms also improve fairness of service among users, hence reducing chances of buffer overflows for poorly located users.

The major observation arising from our analysis is that it is advantageous on the uplink to schedule "strong" users one-at-a-time, and "weak" users in larger groups. This contrasts with the downlink where one-at-a-time transmission for all users has shown to be the preferred mode in much previous work. Based on the optimal schedules, we propose less complex and more practical approximate methods, both of which offer significant performance improvement compared to one-at-a-time transmission, and the widely acclaimed *Proportional Fair* (PF) algorithm, in simulations. When queue content cannot be fed back, we propose a simple modification of PF, *Uplink PF* (UPF), that offers similar improvement.

1 Introduction

Data scheduling in wireless networks is a widely studied topic due to the impending explosion of high speed wireless data services in third generation (3G) systems. For natural reasons associated with the expected traffic characteristics, most of the previous research has focused on the forward-link/downlink, i.e. base to mobile communication (see [1, 2, 3] and references therein for examples, or proceedings of the last several IEEE INFOCOMs!). The traffic is expected to be dominated by web browsing and file downloads. As a result, current wireless data systems employ highly asymmetric link designs (e.g. HDR) with skinny uplinks and fat downlink pipes [4]. However, it has also often been pointed out that there could be a proportional increase in reverse-link/uplink traffic in the form of acknowledgments, feedback etc. along with the growth of other services like ftp, image/data uploads etc. which require high data rates on the uplink. These considerations have resulted in some research on the subject of uplink scheduling [5, 6, 7, 8, 9, 10], although small compared to the literature on downlink scheduling. There is also some amount of related work from information theoretic point of view, see [11, 12] and references therein, on the subject of wireless multiaccess channels. We discuss the relationship of our work to these approaches in greater detail below.

In this work, we consider the problem of optimal scheduling of uplink user transmissions in a single CDMA cell. We assume that the system operates in a TD/CDMA manner, with time-slotted scheduling of transmissions, assisted by periodic feedback of channel and/or congestion information through control channels. The base station then controls user transmission rates via downlink signaling. While our general goals are similar to those of the prior work on uplink scheduling, we formulate the problem in a different manner that enables efficient computation of throughput optimal schedules. Moreover, we observe that the optimal schedules are reasonably well approximated by some very simple scheduling rules that may be implemented with modest effort in 3G systems. The optimal schedules, as well as our proposed approximations, seem to possess an intuitively appealing property that is consistent with traditional observations about wireless transmission: *it is advantageous for users with weak channels to transmit simultaneously, and for users with strong channels to transmit one-at-a-time*[1]. The intuition behind this is that the added interference at the base from simultaneous transmissions of weakly recieved users to each other is small compared to the extraneous interference, and thereby does not affect the user SIRs and data rates significantly. On the contrary, for users recieved strongly at the base, the penalty in terms of SIR and data rate with simultaneous transmission can be quite significant.

As mentioned earlier, there is some previous work on the subject of this paper, scheduling for CDMA uplink [7, 10, 9]. However, the previous work addresses somewhat different problems, and does not offer the solutions that we do below to maximize uplink scheduling gains while meeting rate/QoS requirements among competing users. Perhaps the closest work that arrives at roughly similar qualitative conclusion of "One-strong or many-weak" is [13], which

[1]Hereafter, we refer to users with low recieved power at the base even when transmitting at peak transmit power as "weak" users, and the strongly recieved users at the base as "strong" users.

solves a dynamic programming formulation of a static uplink scheduling problem. Some useful observations regarding the optimality of "bang-bang" control (each user is either silent or transmiting at full power in each slot) were made in [14] and subsequent work, but the authors do not solve the single-cell optimal scheduling problem conclusively. The approach of maximizing rate sum or SIR sum, taken in [10, 9] is often a practically poor option, since it mostly benefits users in favorable locations while ignoring the performance seen by the disadvantaged users. As for the information theoretic approaches of [11, 12], these works deal extensively with various issues such as optimal power control, ergodic capacity, delay limited capacity etc., but in all cases with constraints on *long term average transmit power* and assuming continuously backlogged users with known channel statistics. Our work focuses on air interface scheduling with *instantaneous constraints on transmit power* and without these assumptions, but using feedback-based knowledge of instantaneous channel conditions and user queue. While the need to eliminate channel statistics assumptions is obvious, the introduction of instantaneous power limits reflects the limitations of most current mobile devices more accurately than long-term averages. Also, the assumption of continuous backlog for all users is unrealistic, as data traffic is typically bursty in nature.

Our main focus is on optimizing user QoS measures (queue lengths/delays) using knowledge about users' instantaneous queue length (or delay) status and instantaneous channel condition. As a by-product, our methods also provide greater degree of fairness, in addition to improved throughput, compared to widely used scheduling algorithms (e.g. *Proportional Fair* (PF), Max C/I, etc.). Furthermore, we also consider the effect of limited orthogonality among the user codes through the *orthogonality factor*, which we show tends to encourage simultaneous user transmission due to reduced interference. Based on observations from the optimal method, we also propose a simple modification of the widely used PF algorithm, termed *Uplink PF* (UPF), that demonstrates substantial performance improvement in simulations. UPF uses the same information as PF, but is designed to encourage simultaneous transmission by "weak" users, which greatly improves performance.

The rest of the presentation begins with the derivation of the feasible SNR region on the uplink in Section 2, implicitly with $f = 1$. This is followed by a formulation of the optimal scheduling problem in terms of the user rates and the feasible SNR region in Section 3. We then show how the optimal solution can be obtained with low computational effort later in Section 3. We then discuss the influence of choosing other values for $f \in [0, 1]$ in Section 4, and give a procedure to solve for the optimal schedule when $0 < f < 0.5$. This is followed by a discussion of approximate scheduling algorithms derived from the optimal solution in Section 5, combined with a description of related benchmark algorithms. Finally, in Section 6, simulation results are presented to compare all the algorithms discussed before conclusions in Section 8.

2 Feasible Rate Region for Single Cell Uplink

Consider the uplink of a single CDMA cell serving N users. Let P_i be the instantaneous received power and γ_i the SIR of the ith user. For simplicity, we express P_i in units of the total interference $I+\sigma$, where σ is thermal noise and I is the total instantaneous interference from other sources, such as other cells in the network. In order to meet the SIR requirement of all users, we must then have for each i

$$\frac{P_i}{\sum_{j\in\{1,\ldots,N\},j\neq i} P_j + 1} \geq \gamma_i \tag{1}$$

The feasible SIR vectors γ specified by (1) above has been derived in many previous papers [5, 8, 6], and we recall it below to point out the specific aspects utilized later in our scheduling algorithm. Given the peak received power of the ith user $\bar{P}_i$, computed using the path gain G_i and peak transmit power $\bar{P}_i^t$ as $\bar{P}_i = \bar{P}_i^t G_i$, we may change variables to $\theta_i = \frac{P_i}{\bar{P}_i}$ to rewrite (1) as

$$\frac{\theta_i \bar{P}_i}{\sum_{j\in\{1,\ldots,N\},j\neq i} \theta_j \bar{P}_j + 1} \geq \gamma_i \tag{2}$$

A given SIR vector γ_i is feasible if (2) can be satisfied with equality with $0 \leq \theta_i \leq 1$ for all i. We hence examine the solution to the set of linear equations $\frac{\theta_i \bar{P}_i}{\gamma_i} = \sum_{j\in\{1,\ldots,N\},j\neq i} \theta_j \bar{P}_j + 1$ which can be further rewritten as

$$\theta_i \bar{P}_i (1 + \frac{1}{\gamma_i}) = \sum_{j\in\{1,\ldots,N\}} \theta_j \bar{P}_j + 1 \tag{3}$$

It can be seen by inspection that the solution is of the form $\theta_i \bar{P}_i (1 + \frac{1}{\gamma_i}) = C$ where C is a global parameter. The value of C can be obtained by substituting the postulated solution in (3) to obtain $C = C \sum_j \frac{\gamma_j}{\gamma_j + 1} + 1$ which gives the final solution

$$\theta_i = \frac{\gamma_i}{\bar{P}_i(1+\gamma_i)} \frac{1}{\left[1 - \sum_{j\in\{1,\ldots,N\}} \frac{\gamma_j}{1+\gamma_j}\right]} \tag{4}$$

Defining $\alpha_i = \frac{\gamma_i}{1+\gamma_i}$ we see that

$$\theta_i = \frac{\alpha_i/\bar{P}_i}{1 - \sum_j \alpha_j} \tag{5}$$

Clearly, $0 \leq \alpha_i < 1$. Since we require $0 \leq \theta_i \leq 1$, equations (5) result in the following feasibility conditions to meet the required SIRs.

$$\sum_j \alpha_j + \frac{\alpha_i}{\bar{P}_i} \leq 1 \qquad \forall\ i \tag{6}$$

Note the simple linear form of the feasible SIRs (6) in terms of the α_i, about which we make the following observations:

- When there is no power limitation, i.e. $\bar{P}_i$ are arbitrarily large for each i, equations (6) collapse into the single condition $\sum_j \alpha_j \leq 1$, which is the simple, single-cell version of the well-known stability condition for uplink power control. In this case, as we later show, the feasible SIR region has a fully concave boundary which is dominated by its convex hull composed by time-sharing single user transmissions.

- When the power limitations are severe, i.e. $\bar{P}_i$ are small for each i, the constraints (6) approach independent box constraints $0 \leq \alpha_i \leq \bar{P}_i$ for each i, and simultaneous transmission is favored.

- For intermediate cases, where some $\bar{P}_i$ are large and others small, the optimal scheduling strategy involves time-sharing over different subsets of simultaneously transmitting users. An interesting observation regarding the optimal strategy, that we prove below, is that all transmitting users would always transmit at full power, i.e. at $\bar{P}_i$.

In the subsequent section, we analyze the last item above, which is the most likely practical scenario, to determine the optimal transmitting sets.

3 Uplink Scheduling

Our overall scheduling approach mirrors the downlink scheduling algorithms proposed in [15, 16, 1, 3], where a weighted sum of user rates is maximized for each scheduling interval. This choice has provable stability properties shown in much previous work in various contexts involving data scheduling and resource allocation. The weights may be chosen to optimize one of many possible performance measures, including average queue length, delay, or corresponding percentiles, and other similar criteria. A version of the algorithm that guarantees *queue stability*, i.e. boundedness of queue lengths when feasible, is specified as the rate choice that satisfies

$$\mathbf{R}^* = \arg\max_{\mathbf{R}\in\mathcal{R}} \mathbf{Q}\cdot\mathbf{R}$$

where $\mathbf{R}$, $\mathbf{Q}$ are rate and queue vectors of the user set respectively, and $\mathcal{R}$ is the *rate region*, or the set of feasible rate vectors. Minimum/maximum instantaneous rate guarantees may be satisfied by restricting the rate region $\mathcal{R}$ appropriately. Thus, the general optimal scheduling problem can be solved if one has a technique to solve for $\mathbf{R}^*$ in

$$\mathbf{R}^* = \arg\max_{\mathbf{R}\in\mathcal{R}} \mathbf{w}\cdot\mathbf{R}. \tag{7}$$

for arbitrary given weights $\mathbf{w}$.

To formulate (7) for uplink CDMA scheduling, we require a relationship between rate and SIR as $R \triangleq f(\gamma)$ for each user. We assume this relationship to be *concave* in the argument γ, as is the case for the Shannon formula for the AWGN Gaussian channel where $R = \beta\log(1+\gamma)$. Since (6) are linear in $\alpha \triangleq \gamma/(1+\gamma)$, it is more convenient to consider the R, α relationship $R = g(\alpha)$ which is now *convex* for the Shannon formula as $g(\alpha) = \beta\log[1/(1-\alpha)]$. (7) then becomes the following optimization problem:

$$\max \sum_{i=1}^{N} w_i g_i(\alpha_i) \tag{8}$$

subject to

$$\sum_{j=1}^{N} \alpha_j + \frac{\alpha_i}{\bar{P}_i} \leq 1 \quad , \quad \alpha_i \geq 0 \qquad \forall i. \tag{9}$$

Typically $g_i(\cdot) = g(\cdot) \;\; \forall \; i$ are identical functions, as is the case for the Shannon formula, but our results remain unaffected even if they were all different, as long as they stay *convex*. Before discussing methods to solve (8), (9), we observe some useful properties of the optimal solution.

Theorem 1 *The optimal schedule has the property that each transmitting user transmits at full power, i.e. $P_i = 0$ for some subset $\mathcal{S}$ of the users and $P_i = \bar{P}_i$ for the complementary set $\bar{\mathcal{S}}$.*

Equations (9) specify $2N$ constraints on the feasible α_i. From standard theorems on convex maximization with linear constraints, it is easy to see that the optimum occurs at *corner point* of (9) due to the joint-convexity of (8) in the α_i. Corner points of (9) have exactly N of the $2N$ constraints binding, i.e., some subset of the α_i are null, while the complementary set saturate their respective constraints in the first equation of (9). Combining this observation with (5) results in $\theta_i = 1$ for the complementary set, thus proving the theorem.

Theorem 2 *Without power constraints, i.e. $\bar{P}_i = \infty \quad \forall \; i$, the optimal schedule picks a single user at each scheduling interval.*

In this case, the first equation of (9) reduces to the single constraint $\sum_j \alpha_j \leq 1$, reducing the number of constraints to $N+1$ including the non-negativity constraints. From an argument similar to the proof of Theorem 1, we see that exactly $N-1$ of the α_i are 0 and only one is set at unity on any corner point. One of these is the optimal solution, namely $i^* \triangleq \arg\max_i w_i g_i(\alpha_i)$.

We now present the solution to (8), (9), which determines the subset $\mathcal{S}, \bar{\mathcal{S}}$ for each scheduling interval. Defining $\Lambda \triangleq \sum_{j=1}^{N} \alpha_j$, we rewrite (8), (9) as

$$\max \sum_{i=1}^{N} w_i g_i(\alpha_i) \tag{10}$$

subject to

$$\begin{aligned} \sum_{j=1}^{N} \alpha_j &= \Lambda \\ 0 \leq \quad \alpha_i &\leq \bar{P}_i(1-\Lambda) \qquad \forall i. \end{aligned} \tag{11}$$

For any fixed value of Λ, (10) and (11) have a simple greedy solution for $g_i(\cdot)$ convex. The idea behind the procedure is illustrated in Figure 1, where the convex rate curve is replaced by a straight line joining the endpoints. This is valid on account of Theorem 1, which implies that the α_i always take one of the bounding values in the constraints of eqs. (11) if Λ is appropriately chosen. The algorithm to construct the solution is outlined below:

Algorithm **OPT**:

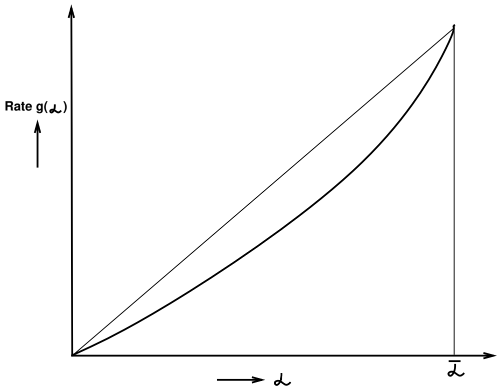


Figure 1: Rate vs. α for $f \in [0.5, 1]$ for fixed Λ.

1. Let $\bar{\alpha}_i = min\{\Lambda, \bar{P}_i(1-\Lambda)\}$. Order the users according to decreasing value of the quantity $v_i = \frac{w_i}{\bar{\alpha}_i} g_i(\bar{\alpha}_i)$.
2. Assign $\alpha_i = \bar{\alpha}_i$ for the user with the highest value of v_i.
3. Update as $\Lambda \to \Lambda - \bar{\alpha}_i$ and repeat with remaining users. The user ordering will not change if $\Lambda \geq \bar{P}_i(1-\Lambda)$ for all remaining users. Else the users must be reordered by recomputing the v_i with the new Λ.
4. Stop if $\Lambda = 0$ and set $\alpha_i = 0$ for all remaining users.

The optimal solution to (8), (9) is then obtained by searching over all $\Lambda \in [0, 1]$ with sufficiently fine granularity.

Further, it is also clear from the original problem that $\Lambda^* = \sum_{i \in \mathcal{S}} \bar{P}_i(1-\Lambda^*)$ for the optimal value Λ^*, where $\mathcal{S}$ is the optimal transmitting set. Thus, the role of Λ is mainly in ordering the users in the best manner out of the $N!$ possibilities, and once an ordering is chosen, it is simple to check the N different values $\Lambda_k = \frac{\sum_{i<=k} \bar{P}_i}{1+\sum_{i<=k} \bar{P}_i} \quad i \in \{1, ..., N\}$ for optimality. Further, an ordering change only takes place when $\Lambda = \Lambda_{ij}$ specified by the solution to $v_i = v_j$ for some $i \neq j$. This condition hence specifies atmost $N(N-1)/2$ ordering changes out of the $N!$ possibilities, some of which may not lie in $[0, 1]$ and can hence be discarded. From these considerations, the following computationally simpler algorithm can be devised.

1. Compute, and sort in increasing order, the Λ_{ij} satisfying $v_i = v_j$ for some $i, j \in \{1, ..., N\}$, $\Lambda_{ij} \in (0, 1)$. Denote this list as $\{\Lambda_k : \; m \in \{1, .., M\}\} \triangleq \{0, \Lambda_1, ..., 1\}$.
2. For each interval $[\Lambda_m, \Lambda_{m+1}]$, determine the user ordering in some interior point, say the midpoint, according to decreasing values of v_i.
3. Evaluate the objective (8) for the current ordering by successively including users from the top of the order.

4. Pick the best objective over all the intervals and user sets examined.

The complexity of the above algorithm for N users is $\mathcal{O}(N^3 \log N)$, and is guaranteed to give the optimal solution.

Rate Limits: All of the above can be repeated when there are specified upper and lower bounds on individual user rates, i.e. $R_i^{min} \leq R_i \leq R_i^{max}$. This condition can be transformed to $\alpha_i^{min} \leq \alpha_i \leq \alpha_i^{max}$ using the functions $g_i(\cdot)$. The lower rate limits may arise due to real-time services, and the upper limits may arise from transmitter capabilities or current queue content. These limits further favor simultaneous transmission, and must be included in the formulation for QoS, and power efficiency reasons. Our approach can accommodate rate limits in the optimization by modifying (11) as $\alpha_i^{min} \leq \alpha_i \leq \min\{\bar{P}_i(1 - \Lambda)\ ,\ \alpha_i^{max}\}$, but we don't describe the somewhat more elaborate solution procedure here. Simple inspection reveals the changes necessary in the previous algorithms, which entail only a small computational overhead.

4 Role of Partial Orthogonality in Uplink Scheduling

We now consider the effect of partial orthogonality in the Walsh codes assigned to users. This effect is traditionally modeled by means of an orthogonality factor f, which specifies the fraction of power transmitted to other users that appears as direct interference to a particular user. The SIR equation (1) is then modified as

$$\frac{P_i}{1 + f \sum_{j \in \{1,\ldots,N\}, j \neq i} P_j} \geq \gamma_i. \tag{12}$$

It is easy to see that most of the previous analysis goes through unchanged if one sets $\gamma_i \to f\gamma_i$ and $(P_i, \bar{P}_i) \to (fP_i, f\bar{P}_i)$. Thus, the feasible SIR vectors are specified by the following modified version of equation (6)

$$\sum_j \alpha_j + \frac{\alpha_i}{f\bar{P}_i} \leq 1 \tag{13}$$

with the α_i now defined as

$$\alpha_i = \frac{f\gamma_i}{1 + f\gamma_i}.$$

Further, the optimization problem (8), (9) changes as

$$\max \sum_{i=1}^{N} w_i g_i(\alpha_i) \tag{14}$$

subject to

$$\sum_{j=1}^{N} \alpha_j + \frac{\alpha_i}{f\bar{P}_i} \leq 1 \quad , \quad \alpha_i \geq 0 \qquad \forall i. \tag{15}$$

where $g_i(\cdot)$ now also depend on f. It is instructive to consider the Shannon formula, for which

$$g(\alpha) = \beta \log \left[1 + \frac{\alpha}{f(1-\alpha)}\right] \tag{16}$$

The feature that impacts scheduling in this case is that $g(\alpha)$ is convex for $f \in [0.5, 1]$, but not so for $f \in [0, 0.5)$! In the latter case, $g(\alpha)$ begins as a concave increasing curve and has an inflexion point given by $\alpha^* = (1-2f)/[2(1-f)]$ where convexity sets in, see Figure 2. The effect of this on scheduling is to favor simultaneous transmission of users, and at power levels smaller than the peak. This is due to the fact that concave maximization problems typically have their optima in the interior of the feasible region. It seems reasonable that other choices of the function $g_i(\cdot)$ will reveal similar traits, since reducing f to very small values eliminates interference, and eventually favors simultaneous transmission by all users in an interference-free manner.

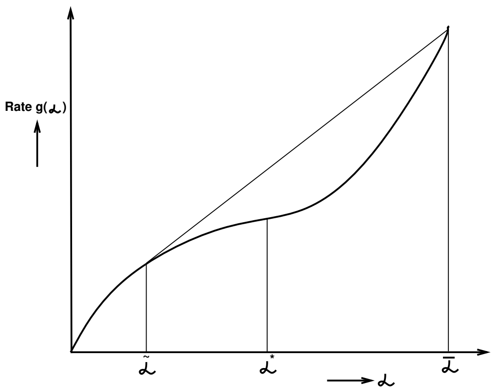


Figure 2: Rate vs. α for fixed Λ.

The optimal scheduling methods discussed earlier continue to apply for $f \in [0.5, 1]$ since $g(\cdot)$ is still convex in this range for the Shannon rate formula. In the following, we address the problem of optimal scheduling, as stated in (14), (15), for $f \in [0, 0.5)$. Figure 2 shows the following critical values of α which will be used in the solution procedure discussed later:

- The upper limit for each user $\bar{\alpha}_i \triangleq \min(\Lambda, f\bar{P}_i(1-\Lambda), \alpha_i^{max})$.
- The inflection point $\alpha^* \triangleq (1-2f)/[2(1-f)]$.
- The critical tangent originating at $\tilde{\alpha}$ obtained by solving

$$\frac{g(\bar{\alpha}) - g(\tilde{\alpha})}{\bar{\alpha} - \tilde{\alpha}} = \left.\frac{dg}{d\alpha}\right|_{\tilde{\alpha}}.$$

The algorithm to obtain the optimum for given Λ then proceeds as follows:

1. Compute α^*, $\bar{\alpha}_i$, $\tilde{\alpha}_i$ for current value of Λ.

2. Divide users into 2 classes:
   - Class I : $\{i \ : \ \bar{\alpha}_i \leq \alpha^*\}$. These users have $\tilde{\alpha}_i = \bar{\alpha}_i$.
   - Class I : $\{i \ : \ \bar{\alpha}_i > \alpha^*\}$. These users have $\tilde{\alpha}_i < \alpha^* < \bar{\alpha}_i$.
3. If Class II is empty, all users operate in the concave portion of their curves, which leads to a standard concave maximization problem with a Lagrange multiplier solution specified as follows:
   - Set $\alpha_i(\lambda) = \min(\bar{\alpha}_i, \alpha_i^c)$ where α_i^c is the solution to $\frac{dg_i}{d\alpha_i}\Big|_{\alpha_i^c} = \lambda$.
   - λ is then determined by solving $\sum_i \alpha_i(\lambda) = \Lambda$.
4. If Class II is non-empty, pick the user i^* with the largest value of the measure $w_i \triangleq Q_i \frac{dg}{d\alpha}\Big|_{\tilde{\alpha}}$ and set $\alpha_{i^*} = \bar{\alpha}_{i^*}$.
5. Set $\Lambda \leftarrow \Lambda - \alpha_{i^*}$ and go back to step 1 above.

The above algorithm then needs to be repeated over a set of values of Λ to obtain the best final objective. Clearly, this procedure is more computationally intensive compared to the $f \in [0.5, 1]$ case, and we include it here for completeness. In practice, the coding employed results in $f \approx 0.6$, and the simpler solution applies. Even for this easier case, however, the optimal solution is too burdensome to use in practice. Hence we devote the following sections to develop simpler approximations motivated by the structure of the optimal solution which demonstrate comparable performance.

5 Low-Complexity Approximation of Optimal Uplink Scheduling

In this part, we restrict to the $f \in [0.5, 1]$ case, and attempt to provide a greedy, low-complexity, approximate solution to the convex maximization discussed before.

5.1 QRP Algorithm

Recall that the sorting measure used in the optimal solution for fixed Λ was of the form $v_i = Q_i g_i(\alpha_i)/\bar{P}_i$ which suggests a greedy algorithm that ranks users by the same measure without Λ. We hence propose the following simple scheduling scheme that may be more suitable for practical implementation:

QRP algorithm:

1. Sort users in decreasing order of the measure $v_i = \frac{Q_i R_i^0}{\bar{P}_i}$ assuming no interference from other users while computing R_i^0.

2. Add user i, in order starting from the top of the list, while maintaining and updating the value of $\mathcal{O} \triangleq \sum_{j<i} Q_i R_i$, where R_i now takes into account interference from all added users.

3. Stop if adding the next user reduces $\mathcal{O}$, and allow transmission of all added users at their peak powers and rates as computed.

As we will see in our simulations section, this simple algorithm captures most of the benefits of optimal uplink scheduling, and has the properties alluded to in the introduction. In other words, the chosen user sets from the above algorithm tend to be one of the following types:

- A single "strong" user with high $\bar{P}_i$.
- A group of "weak" users with low $\bar{P}_i$, and often high Q_i.

This observation is consistent with the common intuition relating to the nature of interference in CDMA systems.

5.2 The Uplink Proportional Fair (UPF) Algorithm

Proportional Fair (PF) scheduling algorithm was proposed and implemented by QualComm for 3G1X EVDO (HDR) downlink. PF algorithm provides fairness among users such that in the long run each user receive the same number of time slots of services. At the same time, PF also takes advantage of channel variations (user diversity).

However, since PF schedule user one-at-a-time, it need to be modified for uplink. The motivation is the same as QRP algorithm, i.e., allow single "strong" user or group of "weak" users to transmit.

UPF algorithm:

1. Sort users in decreasing order of the measure $v_i = \frac{R_i^0}{\bar{R}_i}$ assuming no interference from other users while computing R_i^0. $\bar{R}_i$ is the average rate of user i, updated through a low pass filter in each scheduling interval [17].

2. Add user i, in order starting from the top of the list, while maintaining and updating the value of $\mathcal{O} \triangleq \sum_{j<i} R_i$, where R_i now takes into account interference from all added users.

3. Stop if adding the next user reduces $\mathcal{O}$, and allow transmission of all added users at their peak powers and rates as computed.

5.3 Other Sub-optimal Algorithms and Benchmarks

One benchmark algorithm is the optimal algorithm given in previous sections. It gives the best possible performance. Another benchmark algorithm is the MaxQR algorithm. It serves the user (one-at-a-time) with the maximum queue length and data rate product, $argmax_i Q_i R_i$. This algorithm serves as a lower bound. We will compare with this algorithm to evaluate the gains of different sub-optimal algorithms. Round Robin and fully simultaneous transmission are considered too far from optimal and perform very poorly in most of the cases, and are thus ignored here.

We will provide some other sub-optimal algorithms, which perform less well than the above proposed QRP algorithm. However, they offer simplicity in implementation by using less processing and signaling power. One example of such algorithm is the *Average-SIR algorithm*.

Calculate average SIR, γ_{avg}, among all the users (with non-empty queue). There are two ways to compute users SIR at this stage.

A1. Assume users (with non-empty queue) transmit one-at-a-time.

A2. Assume users (with non-empty queue) transmit simultaneously.

We choose the 2nd approach since it gives clearer difference between "strong" users and "weak" users. Another reason is that it takes co-existence (namely, the orthogonalization factor) into consideration. Then apply MaxQR algorithm to the group of users below average SIR and each user above average SIR. The detailed steps are listed below:

Average-SIR algorithm:

Step 1. Denote the group of users below average SIR as a set B. Re-compute the rates of users belonging to B (only users in B will transmit simultaneously). Denote the group of users above average SIR as a set C. Re-compute the rates of users belonging to C (users in C will transmit one-at-a-time).

Step 2. For users in B, let $q_r[0] = \sum_{i \in B} Q_i R_i$; For the ith user (i start at 1) in C, let $q_r[i] = Q_i R_i$.

Step 3. Choose user/users to serve by $argmax_i q_r[i]$, $i = 0, 1, 2, ...$ If $i = 0$, serve all users in B. Otherwise, serve one user in C with maximum queue length and data rate product.

Note that we could also replace the SIR in the above algorithm by other quantities, for example, data rate or received power. The performance will be similar.

6 Simulation Results

In order to quantify the performance gain by applying optimal/sub-optimal scheduling algorithms, a discrete-event simulator has been used to evaluate them in a single cell CDMA system. All users have the same load and traffic pattern on the uplink. The uplink is implemented as a slot based (Time Division) data transmission mechanism, for example, in 3G1xEV-DO (HDR). For simulation purposes, we assume omni-transmission in the CDMA cell, however all our results apply to a sector in a sectorized cell.

Since every mobile user experiences the same uplink load, we will use the time-averaged queue length as the criterion to compare different uplink scheduling algorithms. Individual as well as total average queue lengths are considered for comparison.

In the simulation we further make the following assumptions:

1. The scheduling decision is made by the base station for every time slot We use 1.6667 msec time slot as defined in 3G1xEV-DO (HDR).

2. The location of the mobiles are assumed to be uniformly distributed in the cell area.

3. It is assumed that the link gains have the following form

$$G_i(k) = d_i^{-4}(k)A_i(k)B_i(k) \tag{17}$$

   where $d_i(k)$ is the distance from the ith mobile to the base station at time instant k, A_i is a log-normal distributed stochastic process (shadowing). B_i is a fast fading factor (Rayleigh distributed).

4. It is assumed that the cell diameter is 2 km. $d_i(k)$ is a 2-D uniformly distributed random variable.

5. It is assumed that the standard deviation of A_i is 8 dB, [18].

6. It is assumed that the Doppler frequency is 8 Hz, corresponding to pedestrian mobile users, [18].

7. It is assumed that all users share 1.25 MHz bandwidth.

8. It is assumed that the uplink traffic of each mobile user is Poisson with the same inter-arrival time 0.05 sec.

9. It is assumed that packet length is exponentially distributed with mean 1024 bits.

10. Simulation time = 10 minutes.

11. $f = 1.0$ for all our experiments.

12. Discrete rate sets, as in 3G1xEV-DV: $9.6Kbps$, $19.2Kbps$, ...$2.4Mbps$ roughly in powers of 2.

Scheduling Alg.	Total Avg. Queue Length
PF	66845
Max QR	70688
UPF	34317
Average-SIR	42895
QRP	34219
OPT	27663

Table 1: Comparison of uplink scheduling algorithms in terms of the total averaged queue length.

6.1 Example Run for Fixed Load, 40 Users

To obtain an idea of the performance gains seen in a typical simulation run, we consider a cell with 40 users offering Poisson traffic at mean rate 20Kbps/user. The results are summarized in the following table. From Table 6.1, we observe that using QRP scheduling increases efficiency by more than 50% comparing with using Max QR scheduling, and is reasonably close to OPT (about 15numbers UPF also performs well, but as we will see later in the varying load case, it breaks down at heavy loads. However, UPF performs much better than PF, which is the state-of-the-art algorithm for downlink scheduling. Another observation is that the Average-SIR scheduling algorithm also performs well (more than 40% efficient comparing with using Max QR scheduling). A feature of the Average-SIR scheduling algorithm is that it requires even less processing than the QRP scheduling algorithm, which is very attractive for implementation. We will discuss implementation details in the next section.

Figure 3(a) compares two queue sensitive, low implementation complexity algorithms, namely MaxQR and QRP. The total average queue length of all 40 users is shown in Figure 3(a). Note that QRP scheduling has more than 50% gain over Max QR algorithm. Furthermore, the queue length of a typical "weak" mobile user (see Figure 3(c)), and the queue length of a typical "strong" mobile user (see Figure 3(b)), clearly indicate that QRP algorithm improves fairness among mobile users as well as increases throughput. It shows that QRP algorithm, which allowing either multiple "weak" mobile users transmitting simultaneously or a single "strong" mobile user transmitting, achieve most performance gain by reducing the queue lengths of the "weak" mobile users.

An interesting plot (Figure 4(a)) shows how the performance gain of QRP over MaxQR changes with the percentage of "weak" mobile users. When most users are at favorable location (i.e. "strong"), the gain from simultaneous transmission is low, and QRP does not provide much gains over MaxQR. However, as the cell is loaded with mostly "weak" users, the gain is substantial (as high as 55%). The latter scenario, of course, is where the gain is most needed.

Figure 4(b) depicts the distribution of number of simultaneously transmitting users for a

(a)
(b)
(c)

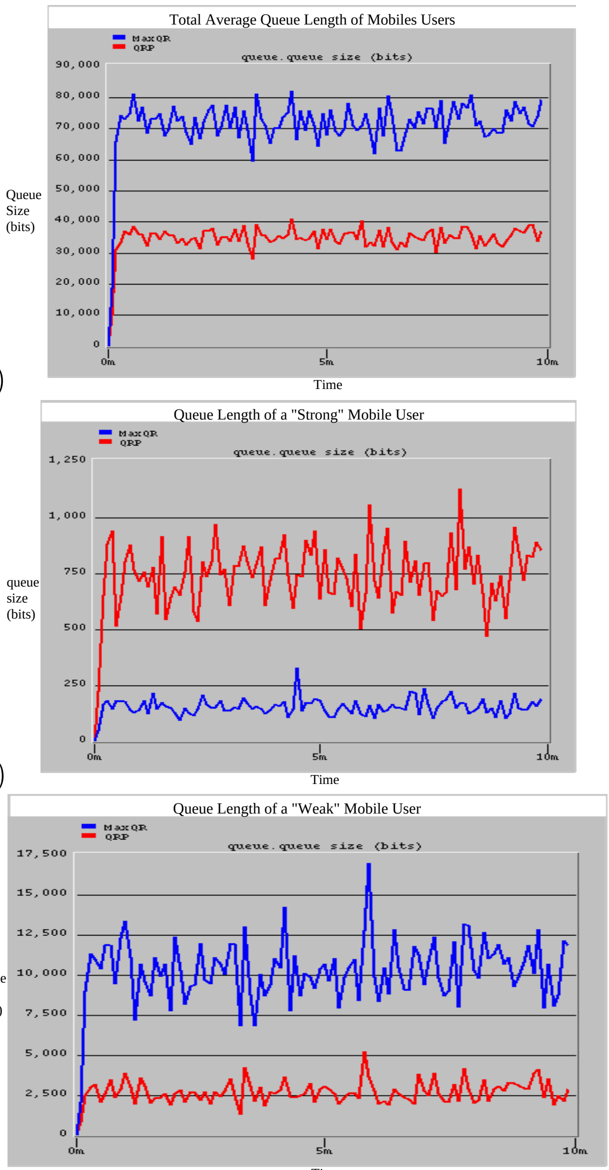


Figure 3: Queue length of (a) all users (b) a typical “strong” mobile user and (c) a typical “weak” mobile user: Max QR vs. QRP scheduling

single run with 20 users employing the QRP algorithm. It shows that there are a significant number of slots with single user transmission, coresponding to the selection of a "strong" user by the algorithm. However, the majority of the slots have multiple (probably "weak") users transmitting, which is the main source of throughput gain from the algorithm over MaxQR. The plot also implicitly reflects the fact that distributing users uniformly in the cell area results in a large fraction of distant, i.e. "weak", users. Rarely do more than 8 of the 20 users transmit simultaneously in this scenario, even though about twice that number classified as "weak" users. Figure 4(c) plots a histogram of the average service rate for each slot (= Total Transmission rate during slot/Number of transmitting users). Again notice that the majority of the slots have low per-user rates (left peak), some fraction have moderately high rates (middle peak), and a small fraction have very high rates(rightmost peak). This supports our earlier observations about the nature of the scheduled user sets. The jagged structure is on account of the discreteness of the rate set.

6.2 Performance Under Varying Load

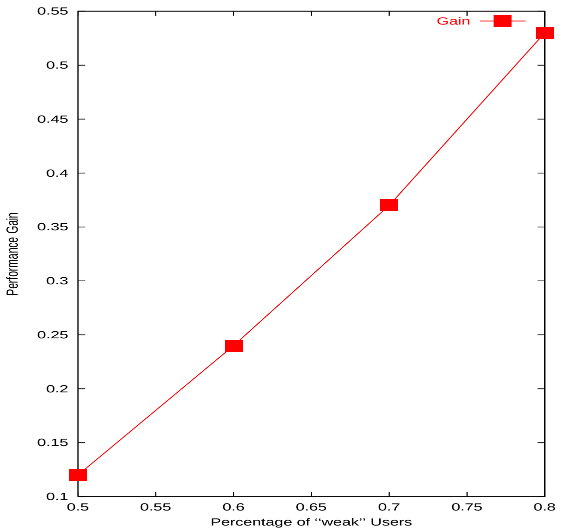
In order to evaluate the performance of different scheduling algorithms, we ran simulations for 20 mobile users in a single cell, under different offered load from 20 kbps to 50 kbps uniformly for each mobile user, for each of the 5 different scheduling algorithms, namely, MaxQR, Proportional Fair (PF), Uplink PF(UPF), QRP and the optimal algorithm (OPT). The results are shown in Figure 5 in terms of (a) total average queue length, (b) queue length of a typical "strong" user and (c) queue length of a typical "weak" user.

Note that, while OPT is uniformly better by almost all metrics, QRP is quite competitive, and even offers better performance than OPT for the "weak" users under heavy load at the cost of some increase in "strong" user and average queue lengths. PF is good for average queue length at lighter loads, but offers poor service to the "weak" users compared to OPT and QRP. However, UPF performs close to optimal under light loads, and degrades more gracefully than PF under heavy loads. Thus, UPF offers uniform improvement over PF by taking advantage of simultaneous transmissions. Also, as expected, both PF and UPF perform worse than the other algorithms at heavy loads, since they do not take queue length into account and can therefore become unstable sooner.

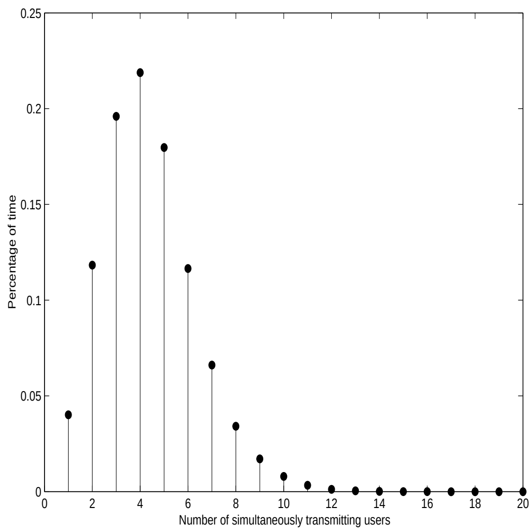
7 Further Discussions

In this section, we will discuss the implementation details of the proposed algorithms. For example, In the 3G1xEV-DV systems, in order to implement the sub-optimal algorithms, including the QRP algorithm, UPF algorithm, MaxQR algorithm and the Average-SIR algorithm, the following enhancements to the system are needed:

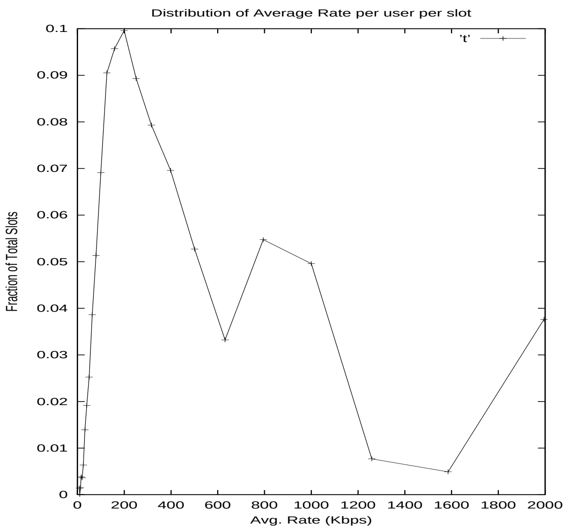
(1). Each active mobile station reports its queue length at every time slot through uplink signaling. This information is not needed for PF, UPF and average SIR, which makes these algorithms attractive for implementation. The other algorithms can also be modified to cope



(a)

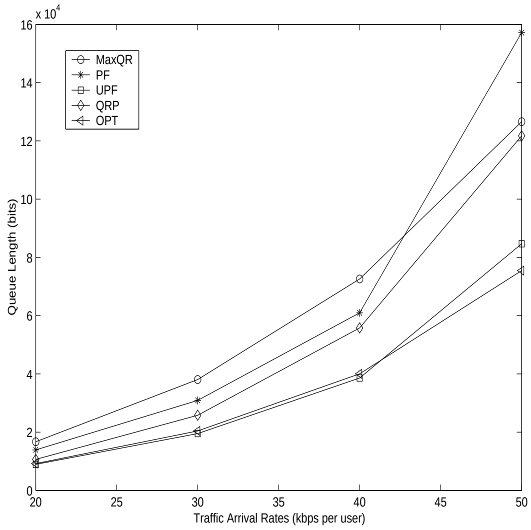


(b)

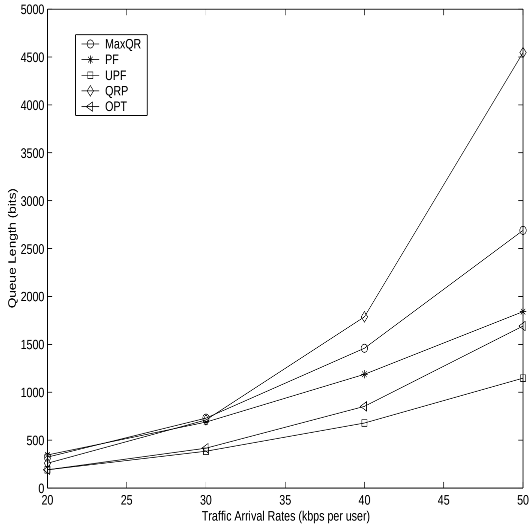


(c)

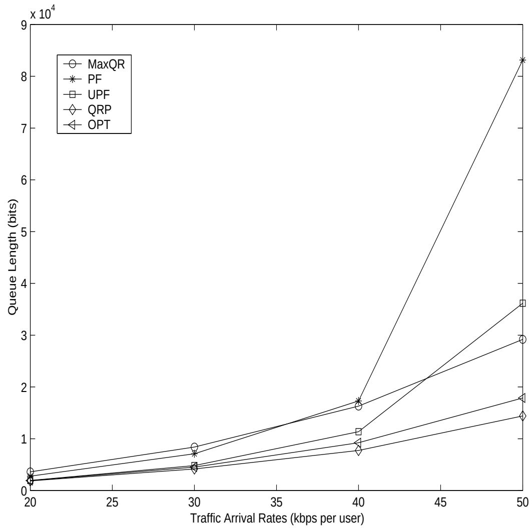
Figure 4: (a) Performance gain of QRP over MaxQR vs. percentage of "weak" mobile users (b) Distribution of simultaneously transmitting users. (c) Distribution of avg. rate per user over slots.



(a)



(b)



(c)

Figure 5: (a) Total average queue length (b) Queue length of a typical "strong" user (c) Queue length of a typical "weak" user, applying MaxQR, PF, UPF, QRP and OPT vs. offered traffic.

with the lack of queue length information, but the performance studies of these modified algorithms will be discussed in a subsequent paper.

(2). Base station decides which mobile station(s) will transmit in the next time slot based on the reported queue length and measured SIR, using either QRP, UPF, Average-SIR or Max QR algorithms.

(3). Base station notifies the chosen mobile station(s) through downlink signaling. Since every mobile station transmits with maximum power, we may replace the power control bit(s) by scheduling bit(s), to indicate whether the mobile station should transmit or not. This requires only a minor change within the standard. Another concern is that the measured (received) SIR at the base station receiver represents the uplink quality in the past time slot(s). In the scheduling algorithm, a predicted uplink quality is needed. There are two ways to accommodate this: If the mobile user is experiencing mainly slow (shadow) fading, it is reasonable to assume that uplink quality will not change abruptly most of the time, because the time-scale of slow fading is much larger than the time-scale of scheduling (time slot). Then we could use a filter with forgetting factor to update SIR measurements. If the fast fading effect is significant, then a channel predictor may be used to predict uplink quality. One choice of such predictors could be discrete-time Kalman filter or H_∞ filter [8]. The first case would be typical in reality since usually data service users have low mobility.

8 Conclusion

We have proposed optimal uplink scheduling algorithms for a single CDMA cell, and related efficient approximate algorithms for practical implementation. Simulations demonstrate substantial performance improvement with these algorithms, and we also discuss the implementation requirements they entail. Further research is required to address multi-cell systems, and also to incorporate the effects of soft-handoff, which may have a significant influence in uplink scheduling.

Acknowledgement

The second author would like to thank DIMACS' hospitality, especially the help from Professor Fred Roberts, and also to acknowledge support through National Science Foundation grant EIA0205116 to Rutgers University.

References

[1] N. Kahale and P. E. Wright, "Dynamic Global Packet Routing in Wireless Networks," in *Proc. of INFOCOM'97*, 1997.

[2] P. Bender et al, "CDMA/HDR A Bandwidth Efficient High Speed Wireless Data Service for Nomadic Users," *IEEE Communications Magazine*, 2000.

[3] D. M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, “CDMA Data QoS Scheduling on the Forward Link with Variable Channel Conditions,” *Bell Labs Technical Memo No.10009626-000404-05TM*, 2000.

[4] E. Esteves, “On the Reverse Link Capacity of CDMA2000 High Rate Packet Data Systems,” in *Proc. of ICC’02*, may 2002, vol. 3.

[5] S.V. Hanly, *Information capacity of Radio Networks*, Ph. D. Thesis, Cambridge University, England, 1993.

[6] A. Sampath and J. Holtzman, “Access Control of Data in Integrated Voice/Data CDMA Systems: Benefits and Tradeoffs,” *IEEE Journal on Selected Areas in Comm.*, vol. 15, no. 8, oct. 1997.

[7] S. Ramakrishna, *Optimal Scheduling of CDMA Systems*, Ph. D. Thesis, WINLAB, Rutgers University, Piscataway, NJ, 1996.

[8] L. Qian, *Optimal Power Control in Cellular Wireless Systems*, Ph. D. Thesis, Rutgers University, Piscataway, NJ, 2001.

[9] C. W. Sung and W. S. Wong, “Power Control and Rate Management for Wireless Multimedia CDMA Systems,” *IEEE Trans. on Comm.*, vol. 49, no. 7, july 2001.

[10] S. Ulukus and L. J. Greenstein, “Throughput Maximization in CDMA Uplinks Using Adaptive Spreading and Power Control,” in *Proc. of IEEE 6th Intl. Symposium on Spread-Spectrum Tech. and Appl.*, 2000.

[11] S. Hanly and D. Tse, “Multiaccess Fading Channels - Part II: Delay Limited Capacities,” *IEEE Trans. on Info. Theory*, vol. 44, no. 7, nov. 1998.

[12] J. Luo, L. Lin, R. Yates, and P. Spasojevic, “Service Outage Based Power and Rate Allocation,” in *Proc. of CISS’01*, mar. 2001.

[13] R. Jäntti and Seong-Lyun Kim, “Transmission Rate Scheduling for the Non-Real-Time Data in a Cellular CDMA System,” *IEEE Communications Letters*, vol. 5, no. 5, May 2001.

[14] R. Agrawal and R. Leelahakriengkrai, “Scheduling in Multimedia DS-CDMA Wireless Networks,” in *Proc. of INFORMS National Meeting*, Oct. 1998.

[15] L. Tassiulas and A. Ephremides, “Stability properties of constrained queueing systems and scheduling for maximum throughput in multihop radio networks,” *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1949, Dec. 1992.

[16] L. Tassiulas, “Scheduling and performance limits of networks with constantly changing topology,” *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 1067–1073, May 1997.

[17] A. Jalali, R. Padovani, and R. Pankaj, "Data Throughput of CDMA-HDR a High Efficiency-High Data Rate Personal Communication Wireless System," in *Proc. of IEEE VTC'00*, may 2000, vol. 3.

[18] G. Stuber, *Principles of Mobile Communication*, Kluwer Academic Publishers, Second Edition 2001.