Self-organizing Dynamic Fractional Frequency Reuse in OFDMA Systems

Alexander L. Stolyar
Bell Labs, Alcatel-Lucent
Murray Hill, NJ 07974
stolyar@research.bell-labs.com

Harish Viswanathan
Bell Labs, Alcatel-Lucent
Murray Hill, NJ 07974
harishv@research.bell-labs.com

June 18, 2007

Abstract

We describe an algorithm for sub-carrier and power allocation that achieves out-of-cell interference avoidance through dynamic fractional frequency reuse (FFR) in downlink of cellular systems based on orthogonal frequency division multiple access (OFDMA). The focus in on the constant-bit-rate (CBR) traffic type flows (e.g., VoIP). Our approach is based on the continuous "selfish" optimization of resource allocation by each sector. No a priori frequency planning and/or inter-cell coordination is required. We show, both analytically (on a simple illustrative example) and by simulations (of a more realistic system), that the algorithm leads the system to "self-organize" into efficient frequency reuse patterns.

1 Introduction

Several techniques with different degrees of complexity can be considered for out-of-cell interference mitigation in orthogonal frequency division multiple access (OFDMA) systems. Most of these schemes involve transmitting in any given cell over a portion of the spectrum that is smaller than the entire available bandwidth while neighboring cells employ a different portion of the spectrum. The main goal of these techniques is to enhance the overall system performance, for example in terms of the total number of VoIP users that can be supported, or in terms of the throughput of "best effort"-type traffic. (In some cases, the goal may be to improve throughput of cell-edge users, even at the expense of average sector throughput.) The interference mitigation schemes differ primarily in how the sub-carrier assignment and power transmitted on the different sub-carriers in a given sector are determined and whether it is adapted as a function of the users being served, and whether it is coordinated across different cells. At the high end of complexity one can envision coordinated scheduling across multiple cells where the bandwidth allocations (and perhaps even the actual transmitted signal) are jointly determined across the cells on a per scheduling interval basis. Such schemes typically rely on very low latency communication between base stations and involve significant signaling between base stations, making them difficult to implement. Schemes based on frequency partitioning and/or distributed coordination that *do not involve signaling between base*

stations are more practical. In this paper we describe an approach for a *dynamic* frequency allocation/partitioning, which does not require any signaling or other explicit coordination between base stations. Thus, the proposed approach lies on the "extreme end" of all distributed coordination techniques. In this paper we focus on the case of supporting constant-bit-rate (CBR, e.g. VoIP) user traffic. Extension of the techniques to other traffic types and a combination of multiple types is a subject of on-going research.

In an OFDMA system the transmission band is divided into a number of sub-carriers and information is transmitted by modulating each of the sub-carriers. Further, time is divided into slots consisting of a number of OFDM symbols and transmissions are scheduled to users by assigning a set of sub-carriers on specific slots. Scheduling can be either persistent or non-persistent. For a CBR user (traffic flow), a natural choice is persistent scheduling, when a user is assigned specific frequency resources at specific times periodically so that multiple encoder packets can be transmitted until resources are released. The frequency resources scheduled are usually logical sub-carriers. The logical sub-carriers are mapped to physical sub-carriers for transmission. The mapping can change from time to time and is referred to as frequency hopping. Frequency hopping is employed to achieve interference averaging.

OFDMA systems supporting *fractional frequency reuse* (FFR) for interference mitigation divide frequency and time resources into several *resource sets*. (Later in the paper they are referred to as *subbands*.) Frequency hopping of sub-carriers is restricted to be within the sub-carriers of a resource set so that users scheduled on a certain resource set experience interference only from users scheduled in in neighboring sectors in the same resource set. Typically, each resource set is reserved for a certain reuse factor and is associated with a particular transmission power profile. For example, suppose we have three sectors covering a certain area, and there are four resource sets. Then, resource set 4 can be reserved for "good geometry" users (those close to their base station, with less interference from other sectors) in all sectors, and resource sets 1,2,3, for "bad geometry" users (farer from their base station, more interference from other sectors) in sectors 1, 2, 3, respectively. As a result, we have 1/3 reuse for bad geometry users and 1/1 - universal - reuse for good geometry users. This is an example of a fractional frequency reuse. Note that FFR can also be "soft" reuse in the sense that although all resource sets are utilized in all cells, a reuse patten is created through non-uniform transmission of power across the different resource sets - most of the power is transmitted on a subset of the resource sets while a small portion of the power is transmitted on the remaining resource sets.

Fixed FFR does not adapt to traffic dynamics in the sense that the frequency reuse achieved is not adjusted based on interference conditions experienced by the users. Instead of a fixed partition of the available bandwidth leading to fixed frequency reuse it is possible to achieve dynamic frequency reuse through prioritized use of sub-carriers in the adjacent sectors. Interference avoidance is achieved by assigning different priority orders in the neighboring sectors so that when transmitting to cell edge users using sub-channelization, the neighboring interfering sectors transmit on different sub-carriers. Such a scheme was described in [1]. For such an approach it is necessary to do some a priori frequency planning to determine the priorities and also to dynamically adjust the thresholds under which users are assigned to different resource sets.

In this paper we describe an alternative approach that systematically and dynamically achieves a frequency reuse efficient for a given user spatial distribution. (In fact, in our simulation experiments, the "automatically" achieved FFR patterns appear to outperform even those "hand-crafted" a priori.) The proposed algorithm does not require any a priori frequency planning. The key idea of our approach, described in more detail in Section 2, is that each sector constantly performs a "selfish"

optimization of the assignment of its users to resource sets, with the objective of optimizing its own performance. In our case, the objective will be to minimize power usage. This optimization is done based on the interference levels reported by users for different resource sets, and is performed "continuously" via a computationally efficient *shadow scheduling algorithm* (described in Section 3).

To study the basic dynamics of sectors' power and user allocations under our algorithm, we consider a *fluid model* of the system, where individual users are replaced by "drops of fluid." We define this model, and establish its basic properties, such as existence of the Nash equilibrium points, in Section 4.1. We also consider a very simple special case of the fluid system in Section 4.2, for which we prove that the dynamical system in fact does reach the desired efficient Nash equilibrium points. This phenomenon of the system "automatically" reaching an efficient equilibrium point is then observed in our extensive simulations of a realistic system in various scenarios, under the actual proposed algorithm.

As mentioned earlier, we believe that our basic approach is applicable not only for CBR traffic but also for other traffic types. For different traffic types, the "selfish" optimization objective of each sector may be different. This is a subject of current and future research.

The paper is organized as follows. In Section 1.1 we briefly discuss some related work. In Section 2 we describe the system model under consideration and provide the overview of the algorithm. In Section 3 we describe the algorithm in detail. In Sections 4.1 and 4.2 we present analytical results on the fluid model for the general case and additional specific results for the one-dimensional case, respectively. In Section 5 numerous simulation results are presented to demonstrate how the algorithm works in a realistic setting. We conclude with a summary and discussion of future work in Section 6.

1.1 Related Work

Fractional frequency reuse in the context of OFDMA systems has mainly been discussed in cellular network standardization fora such Third Generation Partnership Project (3GPP) and Third Generation Partnership Project 2 (3GPP2) [2], [3]. Numerous papers have been published on scheduling in OFDMA systems. However, most of these papers do not consider the effect of out-of-cell interference. Dynamic distributed resource allocation in the context of Gaussian interference channels has been considered in [4] and [5] where a game-theoretic view has been presented. Paper [6] investigates the effect of selfish behavior in spectrum sharing systems motivated by unlicensed band systems. None of these papers consider the model of this paper with multiple interfering base stations each serving several, differently located users. (As a result, even within same cell, different users experience different interference levels in different resource sets.) Centralized, coordinated resource allocation in the context of cellular systems has been considered in [7], [8], [9]. The focus of this paper, on the other hand, is on distributed allocation algorithms.

2 System model and algorithm overview

We consider the downlink of an OFDMA cellular system in which users are assigned a set of sub-carriers at specific time slots for transmission of packets. The OFDMA system further supports FFR by division of sub-carriers into sub-bands. Before describing the model formally, we need to mention

two important aspects of a real system, that motivate and justify our model assumptions.

First, for delay sensitive traffic such as VoIP traffic considered in this paper, channel-aware scheduling in mobile systems does not provide significant benefits and hence frequency hopping is used for channel and interference averaging. Frequency hopping is employed so that the signals transmitted in the different sub-carriers assigned to a user experience different interference levels. This is because each interfering base station transmits at different power levels on the different sub-carriers. Frequency hopping enables code symbols of a code block taken together to experience average interference levels. In OFDMA systems that support fractional frequency reuse it is necessary to restrict the interference averaging to be within each of the resource sets (or, sub-bands) since the goal is to employ interference avoidance to enhance performance. Interference avoidance is achieved by having neighboring sectors transmitting at different power levels in the different sub-bands. Thus signal transmitted to a user scheduled in a given sub-band should only experience interference from the same sub-band. However, it is still necessary to achieve interference averaging within each sub-band since the transmitted power levels could be different in the different sub-carriers within a sub-band. This is achieved by restricting frequency hopping for a given transmission to be across sub-carriers within the sub-band this transmission is assigned to. The hopping pattern is chosen at random so that they are different for the different sectors.

Another important aspect of the system, which is assumed in the model described below, is the so called channel quality indicator feedback that is sent by the mobiles back to the base station for the purpose of resource allocation. The feedback is used to determine the transmission modulation format, channel code rate, number of sub-carriers, and transmit power level required to meet the target data rate. As mentioned earlier, for real time CBR traffic with strict delay constraints such as VoIP, frequency selective scheduling is not employed, and the supportable transmission data rate depends only on the number of sub-carriers, transmit and interference power levels. Average channel quality, in the form of instantaneous rates computed for a nominal transmit power averaged over several 10s of slots (each slot is of the order or 1 msec) is assumed to be fed back periodically. This feedback is essential - it determines each user's current power and sub-carrier requirements in different sub-bands. (These are the p_{ij} and m_{ij} defined later.)

Now we describe our model of an OFDMA system. We have K cells (sectors) $k \in \mathcal{K} = \{1, \ldots, K\}$, and J sub-bands $j \in \mathcal{J} = \{1, \ldots, J\}$ in the system. We assume that each sub-band consists of a fixed number c of sub-carriers, referred to as sub-band capacity. The time is slotted, so that transmissions within each cell are synchronized, and do not interfere with each other. To simplify the exposition we assume that the resources sets discussed earlier correspond directly to sub-bands in the frequency domain and span the entire time period. Extension to more general resource sets is straightforward. A transmission in a cell, assigned to a sub-band, causes interference to only those users in other cells, that are assigned to the same sub-band; there is no inter-sub-band interference.

Consider one of the cells, which needs to support N CBR-type flows, say VoIP. Then, for each user $i \in \mathcal{I} = \{1, \ldots, N\}$, the cell's base station (BS) can choose which sub-band j to assign it to. Given the other-cell interference levels, *currently observed* by user i, the BS "knows" (i.e., can estimate from user feedback) that it would need to allocate m_{ij} sub-carriers and average power p_{ij}, if this user is to be assigned to sub-band j. Since other-cell interference is not constant in time (it depends on the user-to-sub-band assignments in other cells, and the actual powers those users require), the values of m_{ij} and p_{ij} change with time. However, these "parameters" depend on time-average interference levels (over the intervals of the order of 1 second), and therefore they do not

change "fast" (i.e., from slot to slot). Any user-to-sub-band assignment the cell employs at a given time should be such that sub-band capacities are not exceeded:

$$\sum_{i \in A(j)} m_{ij} \leq c, \quad \forall j, \tag{1}$$

where $A(j)$ is the set of users assigned to sub-band j, and the total power used in all sub-bands is below the maximum available level p^*:

$$\sum_{j} \sum_{i \in A(j)} p_{ij} \leq p^*. \tag{2}$$

A good user-to-sub-band assignment strategy, from the overall system performance point of view, would be one "producing" user-to-sub-band assignments in the cells, allowing the system to support as many users as possible with the constraints (1)-(2) being satisfied in each cell.

Key idea behind our proposed approach. Assume the typical situation when it is the power constraints (2) that limit system capacity. Then, a natural "selfish" strategy for each cell is to try to minimize its own total power usage, given the current values of m_{ij} and p_{ij} for its current users:

$$\min \sum_{j} \sum_{i \in A(j)} p_{ij}, \tag{3}$$

subject to (1)-(2). (The minimization in (3) is over all possible assignments $\{A(j),\ j \in \mathcal{J}\}$.)

Key intuition why the proposed decentralized approach should produce a good system-wide performance. Suppose that each cell does try to minimize its own total power usage, as described by (3). Then, we expect the following to happen. "Edge users" in a cell (those further away from their own BS), will have generally larger requirements m_{ij} and p_{ij}, and - more importantly - m_{ij} and p_{ij} will be relatively smaller in those "good" sub-bands j where neighboring cells happen to allocate less power. "Inner users" of the cell (those close to their own BS) generally have smaller requirements m_{ij} and p_{ij}; in addition, they are less affected by the interference from neighboring cells and, consequently, the inner users' values of m_{ij} and p_{ij} are much less dependent on the sub-band j. As a result, the cell (trying to solve (3)) will have a tendency to put its edge users into its good sub-bands j. Therefore, the cell will allocate larger powers to its good sub-bands, thus making those sub-bands "bad" for the neighboring cells. Neighboring cells then (while trying to minimize their own total powers) will "avoid" putting their edge user into those sub-bands, making them even "better" for the cell under consideration, and so on. It is intuitive that the system "settles" into a user-to-sub-band allocation pattern, generally requiring less power in all cells, because neighboring cells will automatically "separate" their edge users into different sub-bands. Section 4.2 illustrates this general phenomenon in an one-dimensional setting that confirms this intuitive explanation.

3 Assignment algorithm run by each cell

In this section we describe our algorithm which essentially reduces to *how* each BS is going to solve problem (3), (1), (2). This question, however, is of crucial importance, because, to make our approach practical, the following requirements need to be observed:

- the algorithm has to be computationally very efficient;
- it should *not* result in a large number of users being re-assigned from sub-band to sub-band in each time slot;
- we want an algorithm which adapts automatically to "evolving" set of users $\mathcal{I}$ and their "evolving" parameters m_{ij} and p_{ij}.

The algorithm we propose is run separately in each cell (that is, at its corresponding BS).

First of all, it is convenient to slightly generalize objective in (3), as follows:

$$\min \sum_j \sum_{i \in A(j)} a_j p_{ij}, \tag{4}$$

where parameters $a_j > 0$ give some a priori "unit costs of power" in different sub-bands j. (These may be different in different cells.) Using unequal a_j may be beneficial if we want to a priori "bias" the system towards certain configurations. The default values can be all $a_j = 1$, which is in fact what is done in all simulations presented later. Thus, the problem the cell needs to solve is (4), subject to (1)-(2).

Remark. The optimization objective (4) is reasonable, but certainly not the only reasonable, "individual" sector objective to induce an efficient system-wide pattern. For example, one may try to explicitly include sub-carrier usage into the objective. We plan to explore alternative objectives in future work.

If we relax integrality constraints in (4)-(1)-(2), we obtain the following linear program:

$$\min_{\{z_{ij}\}} \sum_i \sum_j a_j p_{ij} z_{ij}, \tag{5}$$

$$\sum_i m_{ij} z_{ij} \leq c, \quad \forall j, \tag{6}$$

$$\sum_i \sum_j p_{ij} z_{ij} \leq p^*, \tag{7}$$

$$z_{ij} \geq 0, \quad \forall i, j, \tag{8}$$

$$\sum_j z_{ij} = 1, \quad \forall i. \tag{9}$$

The meaning of z_{ij} is the "fraction" of user i that is placed in sub-band j.

Approximating problem (4)-(1)-(2) with linear program (5)-(9) is reasonable if m_{ij}'s are typically much smaller than c, which is in fact the case e.g. for VoIP users. In this case, typically, a solution to (5)-(9) will assign the "entire" user i to one of the sub-bands, i.e. $z_{ij} = 1$. A small number of users i, however, will be "split", meaning $0 < z_{ij} < 1$ for several j; we deal with this problem as described below.

To solve the linear program (5)-(9), we apply a simple "shadow scheduling" algorithm, defined next, which is a special case of the *Greedy Primal-Dual* (GPD) algorithm in [11].

SHADOW ALGORITHM:

For the cell under consideration, BS maintains virtual ("fictitious") queue Q_j for each sub-band - these are to "keep track" of the constraints (6). It also maintains virtual queue Z, which keeps track of the total power constraint (7). $\beta > 0$ is a small parameter, which controls the tradeoff between responsiveness of the algorithm and its accuracy.

In each time slot (or less frequently, depending on the accuracy we want):
1. For each user i, we identify queue

$$j \in B(i) \doteq \arg\min a_j p_{ij} + \beta Q_j m_{ij} + \beta Z p_{ij},$$

and for this j perform the following updates:

$$Q_j = Q_j + m_{ij} \quad \text{and} \quad Z = Z + p_{ij}.$$

This has the interpretation of "routing" one unit of flow i traffic to queue j (and consuming amount p_{ij} of power).
2. For each j, we update $Q_j = \max\{Q_j - c, 0\}$. Interpretation: "server" j serves c units of work from its queue.
3. Update $Z = \max\{Z - p^*, 0\}$.

The initial state is, for example, $\beta Q_j = 1$ for all j and $Z = 0$. (As we discuss below, the shadow algorithm runs "continuously", even as its "parameters" p_{ij} and m_{ij} gradually change with time. Therefore, the choice of initial state - at the system start-up or reset - is not crucial.)

END ALGORITHM

Shadow algorithm solves (5)-(9) in the following sense. Let ζ_{ij} be the average fraction of time slots, in which user i is "routed" to sub-band j by the shadow algorithm. Then, the set of ζ_{ij} is an approximate solution to (5)-(9); the smaller the β the more accurate the approximation (which becomes exact [11] as $\beta \to 0$).

In reality, it is typically impractical (or impossible) to "split" a user between several sub-bands, so the main information we obtain from the current "routing decision" for user i is this: user i "should be" assigned to one of the sub-bands from the set $B(i)$. Since β is not infinitely small, there is always some "jitter" in the queue values, and so the set $B(i)$ of good sub-bands for flow i is defined more robustly, as

$$B(i) = \{j \mid a_j p_{ij} + \beta Q_j m_{ij} + \beta Z p_{ij} < (1+\delta)[\min_k a_k p_{ik} + \beta Q_k m_{ik} + \beta Z p_{ik}]\},$$

where $\delta > 0$ is some parameter.

Shadow algorithm solves problem (5)-(9) asymptotically assuming the set of users and all m_{ij} and p_{ij} are constant. But, in practice these change - relatively slowly - with time. The algorithm therefore will "track" (if we choose reasonable parameters) the optimal solution of (5)-(9). (This fact is one of the advantages of using an iterative shadow algorithm - we do not need to keep solving the linear problem "from scratch," but rather "track" the solution by doing simple updates of variables.)

Now, as mentioned earlier, we do not want frequent reassignments. To avoid this, we introduce another parameter $\Delta > 0$, and we reassign flow i from its current sub-band $j \notin B(i)$ to another sub-band $j' \in B(i)$ only if the gain is "significant":

$$a_{j'} p_{ij'} + \beta Q_k m_{ij'} + \beta Z p_{ij'} < (1-\Delta)[a_j p_{ij} + \beta Q_j m_{ij} + \beta Z p_{ij}].$$

4 Analysis of the algorithm

4.1 A "fluid" model of user and power allocation

When each sub-band can typically support a large number of users, and power allocations to individual users are small, the following model, where user population is viewed as a continuous "fluid" with a certain spatial distribution, is relevant. The implicit additional assumption leading to this model is that the subcarrier requirements for any user in any sub-band is the same, $m_{ij} = \epsilon$, and the power requirement for a user is such that its received Signal-to-Interference-and-noise-ratio (SINR) meets a certain target level, so that the power requirement is also of the order ϵ. It is also assumed that when a cell cannot meet the SINR targets for all its users, due to the total cell power limitation p^*, then all actual transmitted powers are scaled down by the same factor so that the total power is p^*.

Consequently, we adopt the following conventions in the fluid model. Given the SINR target $\nu > 0$, the transmit power assigned to a small amount δ of fluid, located at a given point, is proportional to $\nu\delta$. More specifically, the transmit power is given by $\nu\delta(N_0 + I)/g$, where N_0 is the noise power and I is the interference power at that location and g is the channel gain from the transmitter to the receiver at that location. We define *nominal* power to be the power that *would need to be* allocated to all users in a cell (in all sub-bands) to achieve the target SINR ν of all users. If the nominal power is such that it exceeds the total available transmit power, p^*, in the cell, we assume that the transmitted powers are scaled down so that the total actual power is exactly p^*. We refer to the scaled nominal powers as *actual* powers.

More formally, the fluid model is as follows. We have K cells $k \in \mathcal{K} = \{1, \ldots, K\}$, and J sub-bands $j \in \mathcal{J} = \{1, \ldots, J\}$ in the system. Denote by $S \subseteq R^2$ the region where users can be located. (Region S does not have to be 2-dimensional.) The distribution of users, assigned to cell k, is given by a measure $\rho^{(k)}$ on S. (S is assumed Lebesgue measurable, as a subset of R^2, with the induced Lebesgue σ-algebra on it.) Let measure $\rho_j^{(k)}$ denote the distribution of cell-k users assigned to sub-band j, so that

$$\sum_j \rho_j^{(k)} = \rho^{(k)}, \quad \text{for any } k \in \mathcal{K}.$$

We require that the mass ("number") of users that can be assigned to any sub-band in a cell is upper bounded by a fixed number, which we set to be 1 without loss of generality. (We can always scale the user distributions up or down, while scaling ν in opposite direction, to obtain an equivalent model.) In other words,

$$\rho_j^{(k)}(S) \leq 1, \quad \forall j, k, \tag{10}$$

Consequently, we always assume that $\rho^{(k)}(S) \leq J$ for all cells k, so that (10) is feasible.

We denote by $g^{(k)}(s) \leq 1$, $s \in S$, $k \in \mathcal{K}$, the propagation gain from base station (BS) of cell k to point s, and assume that $g^{(k)}(s)$ is continuous in s for each k. In addition, assume that all propagation gains are uniformly bounded away from 0, i.e. $g^{(k)}(s) \geq \epsilon$ for some $\epsilon > 0$.

Now, suppose the *user allocation* $\rho = \{\rho_j^{(k)}, \ j \in \mathcal{J}, \ k \in \mathcal{K}\}$, (satisfying (10)) is fixed. Denote by $f_j^{(k)}$ and $h_j^{(k)}$, the nominal and actual power, respectively, allocated by BS k in sub-band j. Then, we

say that the *power allocation* (f,h), where $f=\{f_j^{(k)},\ \ j\in\mathcal{J},\ k\in\mathcal{K}\}$ and $h=\{h_j^{(k)},\ \ j\in\mathcal{J},\ k\in\mathcal{K}\}$, corresponds to the user allocation ρ if it satisfies the following conditions:

$$f_j^{(k)}=\int_S \nu[N_0+I_j^{(k)}(s)]\frac{1}{g^{(k)}(s)}\rho_j^{(k)}(ds),\quad \forall k,\ \forall j, \tag{11}$$

where

$$I_j^{(k)}(s)\equiv\sum_{\ell\neq k} h_j^{(\ell)}g^{(\ell)}(s), \tag{12}$$

and

$$h_j^{(k)}=\begin{cases} f_j^{(k)}, & \text{if } f_\Sigma^{(k)}\le p^*,\\ f_j^{(k)}\frac{p^*}{f_\Sigma^{(k)}}, & \text{if } f_\Sigma^{(k)}>p^*,\end{cases} \tag{13}$$

where

$$f_\Sigma^{(k)}=\sum_j f_j^{(k)}. \tag{14}$$

We say that cell k *achieves target SNR* if $f_\Sigma^{(k)}\le p^*$.

Lemma 4.1 *For any user allocation ρ there exists a corresponding power allocation (f,h).*

Proof. An actual power allocation h, satisfying (11) -(14), can be viewed as a fixed point of the operator (naturally defined by (11)-(14)) mapping the set of all possible h into itself. Such a fixed point exists by Brower's fixed point theorem (cf. [10]). All conditions of Brower's theorem are easily verified; in particular, the continuity of propagation gains $g^{(k)}(s)$ is used to verify the continuity of the operator. ∎

Remark. Lemma 4.1 *cannot* be obtained from standard results regarding solutions to a *power control problem.* (See [12] and references therein.) This is due to the "power capping" condition (13). (In the terminology of [12], this makes "interference function" non-monotone, and therefore "non- standard".

Definition 4.1 *Suppose a set of positive parameters, $a_j^{(k)},\ j\in\mathcal{J},\ k\in\mathcal{K}$, is fixed. (These are the "unit costs of power" for each sub-band in each cell.) We say that a pair of user allocation (assignment) ρ and a corresponding power allocation (f,h) is a* Nash equilibrium *of the system, if the following holds for each cell k: Given the actual power allocation $\{h_j^{(k')},\ \ k'\neq k,\ j\in\mathcal{J}\}$ in all other cells is fixed, the user allocation $\{\rho_j^{(k)},\ j\in\mathcal{J}\}$ in cell k minimizes the total weighted nominal power $\sum_j a_j^{(k)}f_j^{(k)}$ (with $f_j^{(k)}$ as given by (11) and (12)) in this cell.*

Remark. The Nash equilibrium as defined above is in fact a Nash equilibrium in the game theoretic sense, for the game described roughly as follows. "Player" (cell) k's strategy is a pair of

its user allocation $\{\rho_j^{(k)}, \quad j \in \mathcal{J}\}$ and nominal power allocation $\{f_j^{(k)}, \quad j \in \mathcal{J}\}$. After all players pick their strategies, the player k "payoff" is $[-\sum_j a_j^{(k)} f_j^{(k)}]$ if its nominal power allocation in each sub-band is sufficient for all its "users" to achieve target SINR, and is $-\infty$ otherwise. Each player's goal is to maximize its payoff.

Lemma 4.2 *For any user distributions $\rho^{(k)}$, $k \in \mathcal{K}$, (before user assignment to sub-bands), there exists a Nash equilibrium (ρ, f, h) of the system.*

Proof is essentially similar to that of Lemma 4.1, except now we use Kakutani's fixed point theorem (generalizing Brower's theorem, cf. [10]). In more detail, we define a multi-valued operator, mapping an actual power allocation h to a *subset* of such allocations, as follows. Given actual power allocation by all other cells, for a given cell k we find the set of its user distributions such that they, first, minimize the cell's total weighted nominal power allocation $\sum_j a_j^{(k)} f_j^{(k)}$ and, second (among those already chosen), minimize the total nominal power allocation $f_\Sigma^{(k)}$. This set of user distributions then determines the corresponding *set* of actual power allocations by the cell. (Using, the continuity of propagation gains, it is easy to see that this latter set is convex and compact.) Such set is determined for each cell, and their direct product is the image of the original allocation h. The defined mapping has a fixed point by Kakutani theorem. This fixed point determines a Nash equilibrium (ρ, f, h). ∎

Consider now the "uniform power cost" case when all $a_{ij} = 1$. *In this case*, we can always construct a trivial Nash equilibrium, based on *load balancing* (LB), as follows. Suppose, each cell splits its users equally among all sub-bands in all locations, i.e., $\rho_j^{(k)} = (1/J)\rho^{(k)}$. Now, we determine the power allocation separately for each band, by solving (for a fixed sub-band, say $j = 1$) equations (11), (12), and $h_j^{(k)} = \min\{f_j^{(k)}, p^*/J\}$. (We limit the power in a sub-band to $1/J$ of the cell power limit.) This solution is in fact unique, which in this case does follow from the standard power control theory [12]. The resulting user and power allocation satisfies all conditions (11)-(14), and is trivially a Nash equilibrium; indeed, each point in a cell experiences equal interference levels in all sub-bands - therefore, the total nominal power of a cell will not change no matter how this cell would change its user allocation to sub-bands.

Of course, even in the uniform power cost case, the load balancing Nash equilibrium, constructed just above, is not necessarily unique. We demonstrate this on a simple example in the next section. Moreover, our hope is that, typically, if BS in each cell tries to "selfishly" minimize its own total power usage, the resulting system dynamics is such that the system "automatically" finds "better" - requiring less power - Nash equilibria. We show that this is indeed the case for the simple system in the next section. We believe, however, that such behavior of the simple system demonstrates a general phenomenon: selfish power minimization by the cells leads to better system operating points. This intuition will be confirmed by the simulation results, shown later in Section 5.

4.2 Symmetric 1-dimensional system

Consider a simple symmetric 1-dimensional system, shown in Figure 1. This system is a (very) special case of the general system of Section 4.1; and we additionally assume the uniform power cost case: all $a_{ij} = 1$. There are two "cells", 1 and 2. The total "number" (mass) of users in each cell is

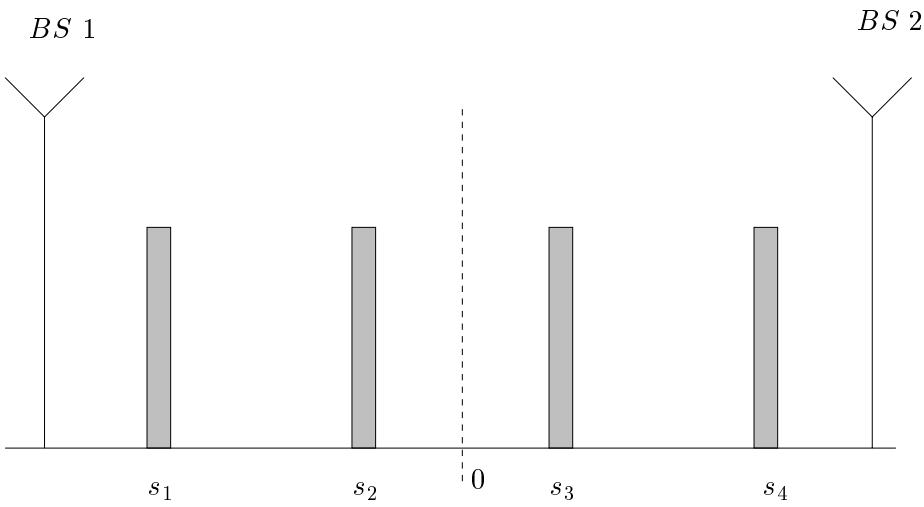

Figure 1: Symmetric 1-dimensional system.

2. In cell 1, the user distribution consists of two atoms, each of mass 1, one at point s_1 (closer to the BS 1) and another at point s_2 (closer to the boundary between cells, which is at 0); the location of BS 2 and the distribution of users in cell 2 are symmetric to those in cell 1 w.r.t. the boundary between the cells (at point 0). The users located at s_1 and s_4 we will call "inner" users, and those at s_2 and s_3 - "edge" users. We assume that propagation gain is strictly monotone decreasing in distance, and so $1 \geq G_1 > G_2 > G_3 > G_4 > 0$, where gains G_1, G_2, G_3, G_4, are are from BS 1 to points s_1, s_2, s_3, s_4, respectively. (By symmetry, those are also the gains from BS 2 to s_4, s_3, s_2, s_1, respectively.)

There are 2 sub-bands, A and B. Recall that in any sub-band of any cell the mass of users is limited to 1 at any time. Therefore, in each cell, each of the two sub-bands contains *exactly* mass 1 of users. Consequently, the entire allocation of users to sub-bands in our system is given by pair $(x, y) \in [0, 1]^2$, where $x = \rho_A^{(1)}(\{s_2\})$ is the number of edge users (located at s_2) assigned to sub-band A in cell 1, and y is the number of edge users (located at s_3) assigned to sub-band B in cell 2.

Recall that $f_j^{(k)}$ and $h_j^{(k)}$ denote the nominal and actual power, respectively, allocated by BS k in sub-band j. (Here $k = 1, 2$, and $j = A, B$.) A power allocation (f, h) corresponding to user allocation (x, y) satisfies conditions (11)-(14). (We will not write the corresponding specialized versions.)

4.2.1 "Load balancing" Vs. "Interference avoidance"

Consider user allocation $(1/2, 1/2)$. This is the "load balancing" (LB) allocation, defined earlier - each cell equally splits its users between the two sub-bands in each location. By symmetry, for this user allocation, the solution to (11)-(14) is unique and is such that $f_A^{(1)} = f_B^{(2)} = f_B^{(1)} = f_A^{(2)} = c > 0$, and $h_A^{(1)} = h_B^{(2)} = h_B^{(1)} = h_A^{(2)} \leq c$, with the last inequality being strict iff $2c > p^*$, i.e., when the total nominal power in each cell exceeds the limit p^*.

The user allocation $(1, 1)$ [and symmetric one $(0, 0)$] can be called "interference avoidance" (IA) allocation, because here, in each sub-band, the edge user of one cell are "matched" with inner users of another. It is intuitive that, typically (assuming propagation gains "correlate enough" with the

distance), the IA allocation should require less total power from each cell. This in fact can be formalized as follows.

Theorem 4.1 *Consider IA user allocation* $(1,1)$*, and assume that a solution to (11)-(14) exists such that* $h = f$*. Such a solution is then unique, and we must have* $f_A^{(1)} = f_B^{(2)} = c_{edge} > c_{in} = f_B^{(1)} = f_A^{(2)}$*. (Consequently, IA user allocation* $(1,1)$ *along with its unique power allocation is a Nash equilibrium.) Then,*

$$c_{edge} + c_{in} < 2c.$$

In other words, the LB user allocation is worse than IA in that it either requires greater total power, or cannot achieve the target SINR, or both.

Proof. Denote $\bar{c} = (c_{edge} + c_{in})/2$. It will suffice to show the following. Suppose cell 2 allocates actual power $\bar{c}$ in each sub-band, and the user allocation is $(1/2, 1/2)$. Then, the nominal power allocation in cell 1 in each sub-band (say, sub-band A), as defined by (11)-(12), is strictly greater than $\bar{c}$. (This implies that if we solve the entire system (11)-(14) for the allocation $(1/2, 1/2)$, the nominal power c for each (cell,sub-band) pair would have to be greater than $\bar{c}$.)

Let us denote by c_1 and c_2 the powers that would have to be allocated (by cell 1 in sub-band A) to the (mass 1/2) users at point s_1, and to the (mass 1/2) users at point s_2, respectively. If we specialize equations (11)-(14) to our case, we can write

$$r_{in} \doteq \frac{c_1}{c_{in}} = \frac{1}{2}\frac{N + \bar{c}G_4}{N + c_{edge}G_4} < r_{edge} \doteq \frac{c_2}{c_{edge}} = \frac{1}{2}\frac{N + \bar{c}G_3}{N + c_{in}G_3}.$$

The inequality above is because $c_{in} < \bar{c} < c_{edge}$, $G_4 < G_3$, and r_2 is strictly increasing in G_3. We need to show that

$$c_1 + c_2 = r_{in}c_{in} + r_{edge}c_{edge} > \bar{c} = (1/2)c_{in} + (1/2)c_{edge}.$$

Now, since $c_{in} < c_{edge}$ and $r_{in} < r_{edge}$, it will suffice to show that

$$r_{in} + r_{edge} \geq 1.$$

This is in fact true, because

$$r_{in} + r_{edge} \geq \frac{1}{2}\frac{N + \bar{c}G_4}{N + c_{edge}G_4} + \frac{1}{2}\frac{N + \bar{c}G_4}{N + c_{in}G_4} \geq 1,$$

where the right relation is by Jensen inequality. ∎

Remark. Theorem 4.1 for the 1-dimensional system can be easily generalized as follows. Suppose the distributions $\rho^{(1)}$ and $\rho^{(2)}$ in the cells 1 an 2 are symmetric with respect to the cell boundary (see Figure 2), and are such that the total mass of users in each cell does not exceed 2. The IA allocation in this case is such that cell 1 places as many of its edge users as possible (up to the total mass 1) into one of the sub-bands, say A. On Figure 2 these are the users located to the right of point s_* (and up to the cell boundary). The remaining users (those to the left of s_* on Figure 2), are placed in the other sub-band B. The allocation in cell 2 is symmetric, with as many edge users as possible placed in sub-band B. Suppose, the total power required for this IA allocation does not exceed p^*, the maximum. Then, the total nominal power required for the LB allocation, $\rho_A^{(k)} = \rho_B^{(k)} = (1/2)\rho^{(k)}$, $k = 1, 2$, is greater than that for the IA allocation.

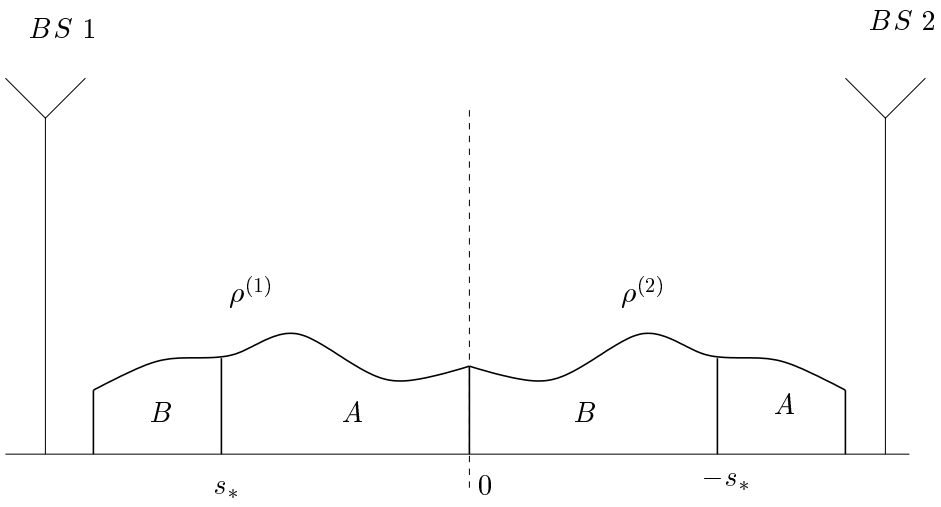

Figure 2: Generalization of Theorem 4.1.

4.2.2 Dynamics

In this section, let us further assume that for any (x, y) there exist a solution to (11)-(14) such that the nominal and actual power allocations are equal, $h = f$. Such a solution is then unique, by standard power control theory [12], and moreover the simple properties summarized in the following lemma hold.

Lemma 4.3 *(i) For any (x, y) relations (11)-(14) uniquely determine power allocation $h = f$. Moreover, the dependence of h on (x, y) is continuous.*
(ii) As functions of (x, y), $h_A^{(1)}$ and $h_A^{(2)}$ are strictly increasing in x and strictly decreasing in y; symmetrically, $h_B^{(1)}$ and $h_B^{(2)}$ are strictly increasing in y and strictly decreasing in x.
(iii) If $x = y$, $h_A^{(1)} = h_B^{(2)}$ and $h_B^{(1)} = h_A^{(2)}$. If $x = y = 1/2$, then $h_A^{(1)} = h_B^{(2)} = h_B^{(1)} = h_A^{(2)}$. If $x = y > 1/2$, then $h_A^{(1)} = h_B^{(2)} > h_B^{(1)} = h_A^{(2)}$.
(iv) If $x > y \geq 1/2$, then $h_A^{(1)} > h_B^{(1)}$. Symmetrically, if $y > x \geq 1/2$, then $h_B^{(2)} > h_A^{(2)}$.

Now, let us consider a dynamic system, where user allocation $(x(t), y(t))$ is a function of continuous time $t \geq 0$. The dynamics is such that, in a small time interval $[t, t + dt]$, each cell can reassign up to dt mass of its users from sub-band B to A in one location, and the matching mass from A to B in another location. The "direction" of reassignment depends on the current interference levels from the neighbor cell in each sub-band, and is such that the cell's total power would decrease (assuming the other cell keeps it power allocation unchanged).

Remark. Such a dynamical system models the dynamics of the actual shadow scheduling based algorithm described earlier if we additionally assume that there is a fixed limit on the number of users that can be reassigned within one time slot. Also, in our one-dimensional system, all "users" in one location are completely indistinguishable. As a result, as long as power transmitted by a neighbor cell is less in one of the sub-bands, say A, the shadow algorithm will try to reallocate its edge users to sub-band A "as fast as it can," while reallocating the matching number of inner users to sub-band B.

More formally, let $h_j^{(k)}(t)$ denote the actual power, allocated at time t by BS k in sub-band j. These powers are determined uniquely by the user allocation $(x(t), y(t))$. We define the dynamic system to be such that both $x(t)$ and $y(t)$ are Lipschitz continuous with (Lipschitz) constant 1. Function $x(t)$ satisfies the condition that $x'(t) = 1$ if $h_A^{(2)}(t) < h_B^{(2)}(t)$, and $x'(t) = -1$ if $h_A^{(2)}(t) > h_B^{(2)}(t)$; and $y(t)$ satisfies the symmetric conditions.

Lemma 4.4 *(i) If $x(t) > y(t) \geq 1/2$, then $y'(t) = 1$. (ii) If $1 > x(t) = y(t) > 1/2$, then $x'(t) = y'(t) = 1$. (iii) In either case (i) or (ii) above, $[x(t) \wedge y(t)]' = 1$, and consequently $[x(\tau) \wedge y(\tau)]' = 1$ for all $\tau \geq t$ as long as $x(\tau) \wedge y(\tau) < 1$.*

Proof follows directly from the definition of the dynamic system and the properties of the dependence of h on (x, y) given in Lemma 4.3. ∎

The LB assignment $(1/2, 1/2)$ is obviously an equilibrium point of the dynamic system, that is $(x(t), y(t)) \equiv (1/2, 1/2)$ is a valid trajectory. Note, however, that the trajectory starting at $(x(0), y(0)) = (1/2, 1/2)$ and such that $(x(t), y(t)) = (1/2+t, 1/2+t)$, $0 \leq t \leq 1/2$, and $(x(t), y(t)) = (1, 1)$, $t > 1/2$ is also valid. (By symmetry, so is the trajectory with $(x(t), y(t))$ decreasing at rate $(-1, -1)$ until it hits $(0, 0)$ and stays there.) Therefore, LB assignment $(1/2, 1/2)$ is not a *stable* equilibrium. Note also that if the initial state is such that $x(0) + y(0) = 1$, then the trajectory is unique in the interval $[0, |x(0) - 1/2|]$, in which both $x(t)$ and $y(t)$ change at rate 1 (or -1) until $(x(t), y(t))$ reaches $(1/2, 1/2)$; and then trajectory is non-unique.

On the other hand, both IA assignments $(1, 1)$ and $(0, 0)$ are stable equilibrium states. The following result shows that, in essence (i.e. except for "contrived" initial states), those two points form the global attraction set.

Theorem 4.2 *Consider the dynamic system $(x(t), y(t))$, defined in this section.*
(i) Both $(1, 1)$ and $(0, 0)$ are Lyapunov stable equilibrium points.
(ii) If $x(0) + y(0) \neq 1$, then the trajectory reaches either state $(1, 1)$ or $(0, 0)$ within time at most 1 and stays in this state thereafter.

Proof. Statement (i) for point $(1, 1)$ follows from Lemma 4.4(iii). (If $(x(t), y(t))$ is in a neighborhood of $(0, 0)$, we can use system symmetry, i.e. work with $u = 1 - x$ and $w = 1 - y$ instead of x and y.)

To prove (ii), we consider two cases. Case (a): initial state $(x(0), y(0))$ is such that $h_A^{(1)}(0) \geq h_B^{(1)}(0)$ and $h_A^{(2)}(0) \leq h_B^{(2)}(0)$. [The case when the inequalities are reversed is treated analogously, using system symmetry.] In this case, it is easy to see that we must have $x(0) \geq 1/2$ and $y(0) \geq 1/2$. Moreover, the case $x(0) = y(0) = 1/2$ is impossible, because $x(0) + y(0) \neq 1$. Therefore, either $x(0) > 1/2$ or $y(0) > 1/2$. Assuming for concreteness that $x(0) \geq y(0)$, we see that, at $t = 0$ we are in the conditions of either (i) or (ii) of Lemma 4.4. By Lemma 4.4(iii), the trajectory converges to $(1, 1)$ within time at most $1/2$ and stays there thereafter.

Case (b): the initial state $(x(0), y(0))$ is such that $h_A^{(1)}(0) \geq h_B^{(1)}(0)$ and $h_A^{(2)}(0) \geq h_B^{(2)}(0)$. [The case when the inequalities are reversed is treated analogously, using system symmetry.] In this case, starting time 0, $x'(t) = -1$ and $y'(t) = 1$ until the earliest time t_0, when either $h_A^{(1)}(t_0) = h_B^{(1)}(t_0)$

Parameter	Assumption
Cell Layout	Three Sectors
Inter-site distance	2.5 Km
Path Loss Model	$L = 133.6 + 35 \log_{10}(d)$
Shadowing	Log Normal with 8.9 dB Std. Dev.
Penetration Loss	10 dB or 20 dB
Noise Bandwidth	1.25 Mhz
BS Power	40 dBm
BS Antenna Gain	15 dB
Rx Antenna Gain	0 dB
Rx Noise Figure	7 dB
Channel Model	No fast fading

Table 1: Propagation parameter values used in the simulation results

or $h_A^{(2)}(t_0) = h_B^{(2)}(t_0)$ - we assume the latter for concreteness. We have $x(t_0) + y(t_0) \neq 1$ because $x(0) + y(0) \neq 1$. Thus, the state $(x(t_0), y(t_0))$ is of the same type as the initial state in the case (a) considered earlier. In remains to show that $t_0 \leq 1/2$. This is true, because otherwise $x(t_0) < 1/2$ and $y(t_0) > 1/2$, which would imply $h_A^{(2)}(t_0) < h_B^{(2)}(t_0)$. ∎

5 Simulation results

In this section we describe results from a simulation study of the algorithm. We first describe the system model employed for the simulation study and then discuss the simulation results for various scenarios.

5.1 System model for simulations

A three sector network with base stations laid out as shown in Figure 3 is used to study the behavior of the algorithm under a variety of traffic distribution scenarios. Users are distributed in the triangular region covered by the three sectors. Standard propagation parameters as listed in Table 5.1 are used to determine the received signal power level for a given transmit power level. The penetration loss is set to either 10 dB or 20 dB which results in a cell edge SNR (signal to thermal noise ratio, when there is no interference from surrounding cells, assuming total available power is distributed over the entire bandwidth) of 20 dB or 10 dB, respectively. Sensitivity of the performance gain to the cell edge SNR is captured in this fashion.

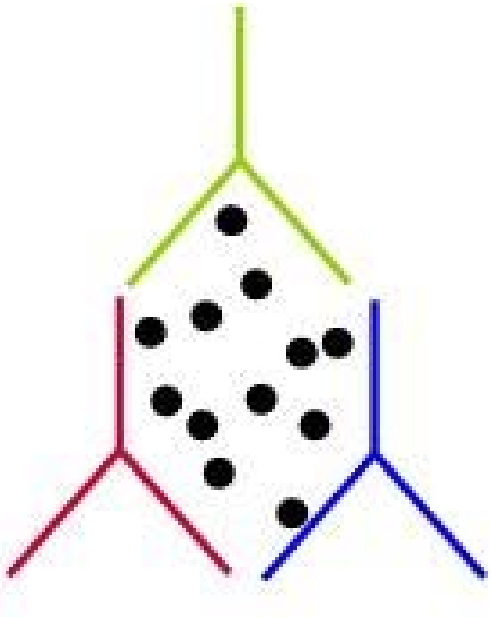

Figure 3: Lay out of sectors and sample user locations.

5.1.1 Traffic model

A constant bit rate traffic model in which, for all users in the active state, fixed length packets of size 128 bits arrive once every 20 slots is adopted. Users transition from active to inactive state according to exponentially distributed waiting times in each state. No packets arrive when a user is in the inactive state. The mean time in each state is 200 slots.

5.1.2 Frequency hopping

In our simulation, we consider an OFDMA system with 48 sub-carriers divided into a number of sub-bands with the same number of sub-carriers in each sub-band. Typically, we consider three sub-bands with 16 sub-carriers in each sub-band. Random frequency hopping is implemented from slot to slot by permuting the sub-carrier indices independently across the different sub-bands and sectors.

5.1.3 Transmission rate computation and persistent scheduling

Each user that is scheduled for transmission in a given slot is assigned a set of sub-carriers, say $\mathcal{C}_i$, and transmit power P_i divided equally across the $|\mathcal{C}_i|$ sub-carriers. A user is assigned resources in only a single sub-band. Given such an allocation, the number of bits transmitted in a slot is

computed using the Shannon formula for channel capacity as

$$R_i = \sum_{j \in \mathcal{C}_i^{\pi_n}} \log_2 \left(1 + G_{ni} \frac{P_i / |\mathcal{C}_i|}{N_0 + \sum_{b \neq n} G_{bi} P^{j,b}} \right) \tag{15}$$

where n is the serving sector of user i, $\mathcal{C}_i^{\pi_n}$ is the set of physical sub-carrier locations that are obtained as a result of the permutation π_n of the logical sub-carriers $\mathcal{C}_i$ for frequency hopping, $P^{j,b}$ is the power transmitted by sector b in the physical sub-carrier j, and G_{ni}, G_{bi} are the channel gains to the user from the serving and interfering base stations, respectively.

Since we are focused on CBR traffic in this paper, we adopt a persistent scheduling approach in which users in the active state that periodically receive packets are assigned resources in a sub-band in periodic fashion, i.e., in every k^{th} slot for some k. In our simulations $k = 5$. The actual number of sub-carriers assigned and power transmitted are adjusted according to the power control algorithm described below. Furthermore, users, although persistently scheduled, may be re-allocated to other sub-bands if dictated by the shadow algorithm.

5.1.4 Channel quality indicator (CQI) feedback

A key requirement for the algorithm is the feedback of channel quality in the form of data rate that can be supported within each of the sub-bands. In our simulation study we do not employ frequency selective sub-band scheduling. The feedback is used to determine the transmission rate, number of sub-carriers, and transmit power level. Since frequency selective scheduling is not employed and no fast fading is simulated, the supportable transmission rate depends only on the transmit and interference power levels. Average channel quality, in the form of instantaneous rates computed for a nominal transmit power averaged over the last 50 slots (implemented through a moving average filter with a parameter of 1/50), is assumed to be fed back every slot in the simulations. In practice, for diversity scheduling, CQI averaged across sub-bands can be sent frequently for rate and power adjustments while per sub-band CQI averaged over many time slots that provides information about interference power levels in the different sub-bands could be sent relatively infrequently.

5.1.5 Incremental redundancy and power control

The simulation also captures incremental redundancy and power control. A single hybrid automatic repeat request (HARQ) interlace period consists of 20 slots, which are divided into five interlaces with up to 4 transmissions per interlace. Bandwidth resources are allocated for each active user in a given interlace in a persistent fashion. If the transmission of the packet is successful prior to 4 transmissions, a new packet transmission is initiated provided there are packets in the buffer to be sent. If there are no packets to be sent then the resources assigned to the user are unused until a new packet arrives for transmission. Power control is employed to guarantee that packets are transmitted within the targeted number of transmissions. Each time a packet is successfully transmitted prior to the target number of transmissions, the power spectral density, i.e., the transmit power for each sub-carrier, is reduced by 0.05 dB while if the packet is not transmitted even after the target number of retransmissions the power spectral density (PSD) is stepped up by 0.5 dB. The PSD has a maximum and minimum limit set at 5 and 0.5, times the nominal PSD, respectively, where nominal PSD is the PSD when total available power is uniformly distributed across all the sub-carriers. When the

maximum PSD is reached the number of sub-carriers assigned is incremented by 1 and the PSD value corresponding to the new number of sub-carriers is calculated. Similarly, when the power control drives the PSD to the minimum value a sub-carrier is freed and the new PSD calculated.

5.1.6 Computation of m_{ij} and p_{ij} for the Shadow algorithm

When the activity state for user i changes from inactive to active and the first packet arrives at the sector, p_{ij}, m_{ij} are calculated for each sub-band j as follows. The PSD is set to the maximum possible PSD and the minimum number of sub-carriers required to transmit a packet over the target number of transmissions is calculated for each sub-band j. For this computation the channel quality indicator fed back by each user on a per sub-band level is used. Once this is determined then the PSD is adjusted for the chosen number of sub-carriers. m_{ij} is the scaled version of the number of sub-carriers, with the scale factor being one over the number of interlaces, 5 in the simulation, in an interlace period. Similarly, p_{ij} is the PSD computed times the computed number of sub-carriers normalized by the number of interlaces. Subsequently in each iteration as the power control updates the transmit PSD, the p_{ij} computed is scaled by a common scale factor across all the sub-bands so that the p_{ij} of the assigned sub-band corresponds to the current transmit PSD.

5.2 Results

5.2.1 Performance for uniform user distribution

We compare the performance of the shadow algorithm for three sub-bands to that of universal reuse with a single sub-band and no interference coordination. Users are distributed uniformly in all three sectors in all of the results in this section. Comparison is performed on the basis of the maximum number of users that can be supported in each sector. This maximum number is obtained from the cumulative distribution functions (CDFs) of the total sector transmit power and the mean queue size.

Figure 4 shows the normalized average transmission power in each sub-band and the average number of sub-carriers assigned in each of the three sub-bands for the case of 10 dB cell edge SNR. Recall that the total number of sub-carriers in each sub-band is 16. From the figure it is clear that the shadow algorithm achieves a desirable frequency plan. Note that the frequency plan achieved is *not* the strict 1/3 reuse, which is inferior to universal reuse.

Figure 5 shows the complementary cumulative distribution functions of the total sector power normalized by the maximum available sector power and the mean queue size over the duration of the simulation normalized by the packet size for the different users. The criteria for determining if a given number of users can be supported by the system are (a) the one-percentile point in normalized total sector power plot should not exceed 1 for valid system operation and (b) the one-percentile point in the normalized mean queue size plot should not exceed about 2. The latter is because of the fact that for real-time traffic such as VoIP traffic, over the air one way delay of up to 80 ms may be tolerated. This translates to a maximum delay of about 4 packets. Since we are using mean delay, 2 packets delay has been used as the criterion. Based on the above criterion it is clear from the figure that the shadow algorithm can support about 145 users while the universal frequency reuse with a single sub-band approach can only support 120 users. Thus the gain is about 20%.

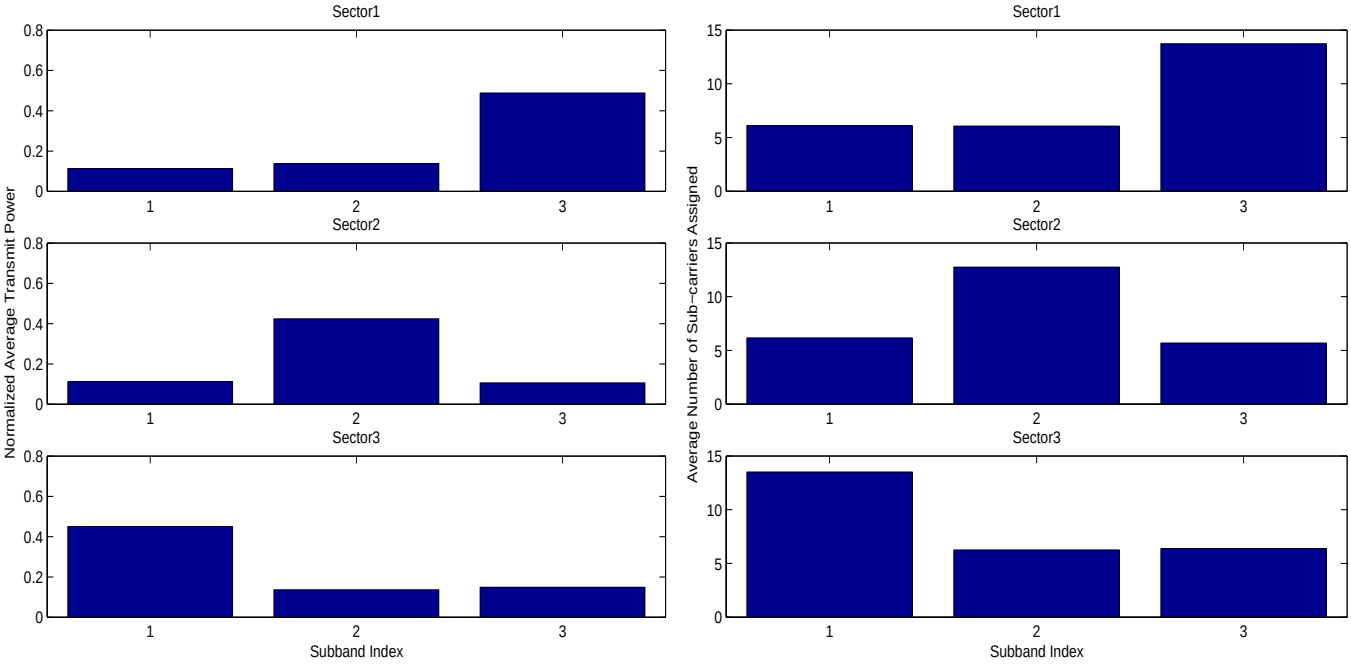

Figure 4: (a) Normalized average transmission power in the different sub-bands and (b) Average number of sub-carriers assigned in he different sub-bands

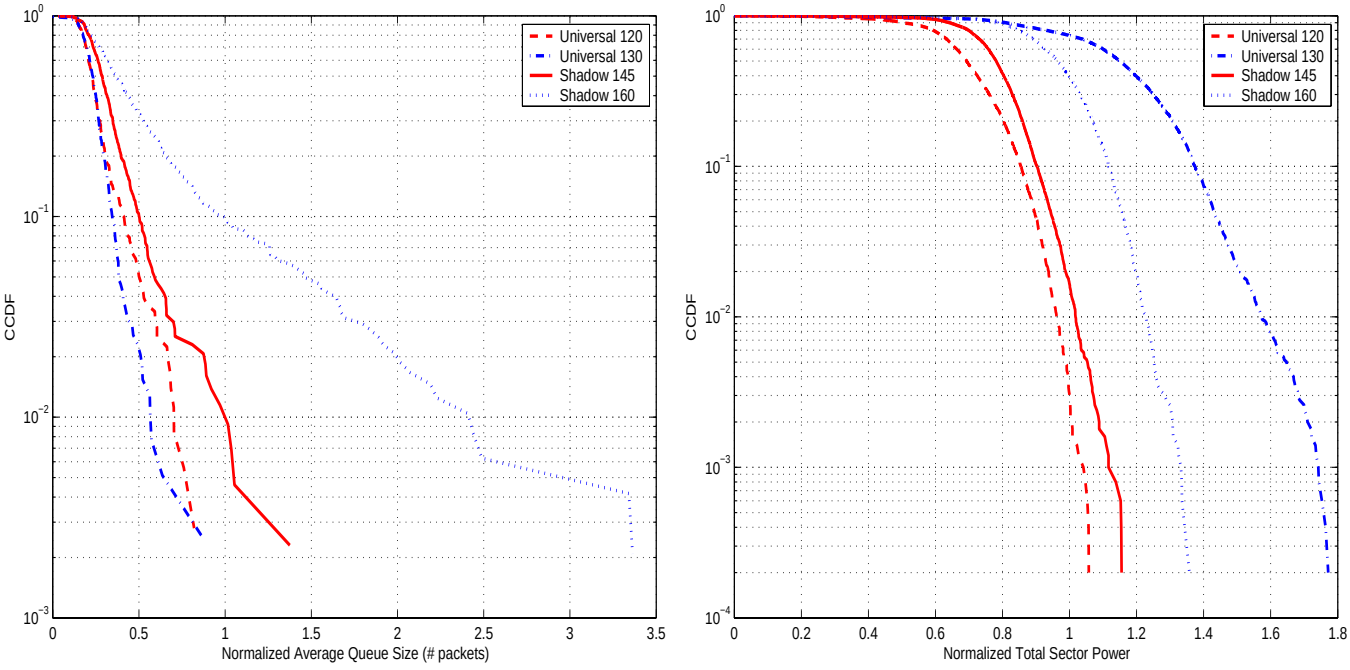

Figure 5: Complementary cumulative distribution functions of the total sector power and mean queue size for 10 dB cell edge SNR

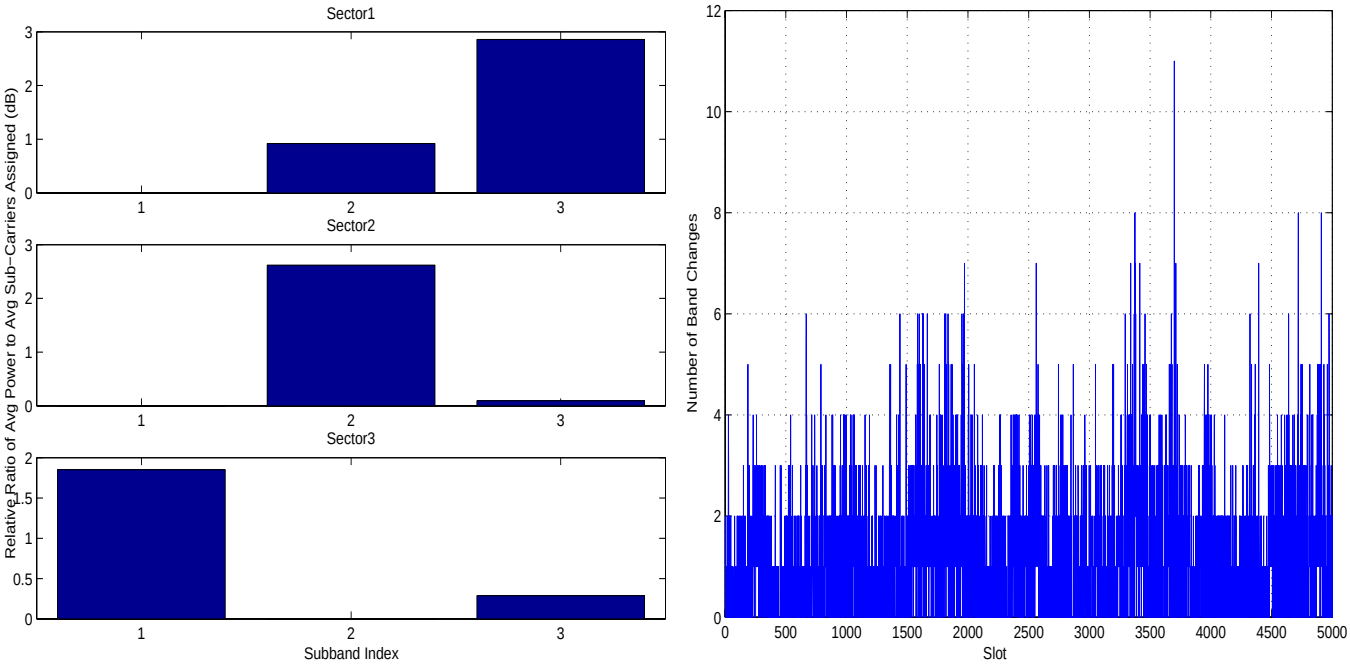

Figure 6: (a) Relative transmit power ratios (in dB) in the different bands in the three sectors (b) Number of Band Changes

In Figure 6 on the left we show the relative power spectral density across the three sub-bands. This is calculated as follows. First the average transmit power level in each sub-band is normalized by the average number of sub-carriers assigned in that sub-band. The resulting power spectral densities are then normalized by the smallest of the three power spectral density values to obtain the relative values which are then plotted in dB. This shows that there is a significant difference in the transmit power spectral densities across the three sub-bands.

On the right in Figure 6 we show the total number of users across the three sectors that require re-assignment to a different band to make the allocation more efficient as a function of the index slot. This number is important because reassigning a user to a different band incurs additional signaling overhead. Thus, the smaller the number of reassignments the better. From the figure it is clear that the number of users reassigned is a small fraction, less than 3%, of the total number of users.

Figure 7 shows the complementary cumulative distribution functions for the 20 dB cell edge SNR case. From these plots we conclude that the gain for the shadow algorithm over universal reuse in this case is about 30%, larger than that in the 10 dB case as expected because of higher out-of-cell interference levels.

5.2.2 Performance for different number of sub-bands

In this section we show the behavior of the shadow algorithm for the same simulation set up as in the previous section but with four and six sub-bands instead of three sub-bands. We observe in Figures 8 and 9, showing the normalized average sector power transmitted in each sub-band and the average number of sub-carriers assigned in each sub-band, that the shadow algorithm produces

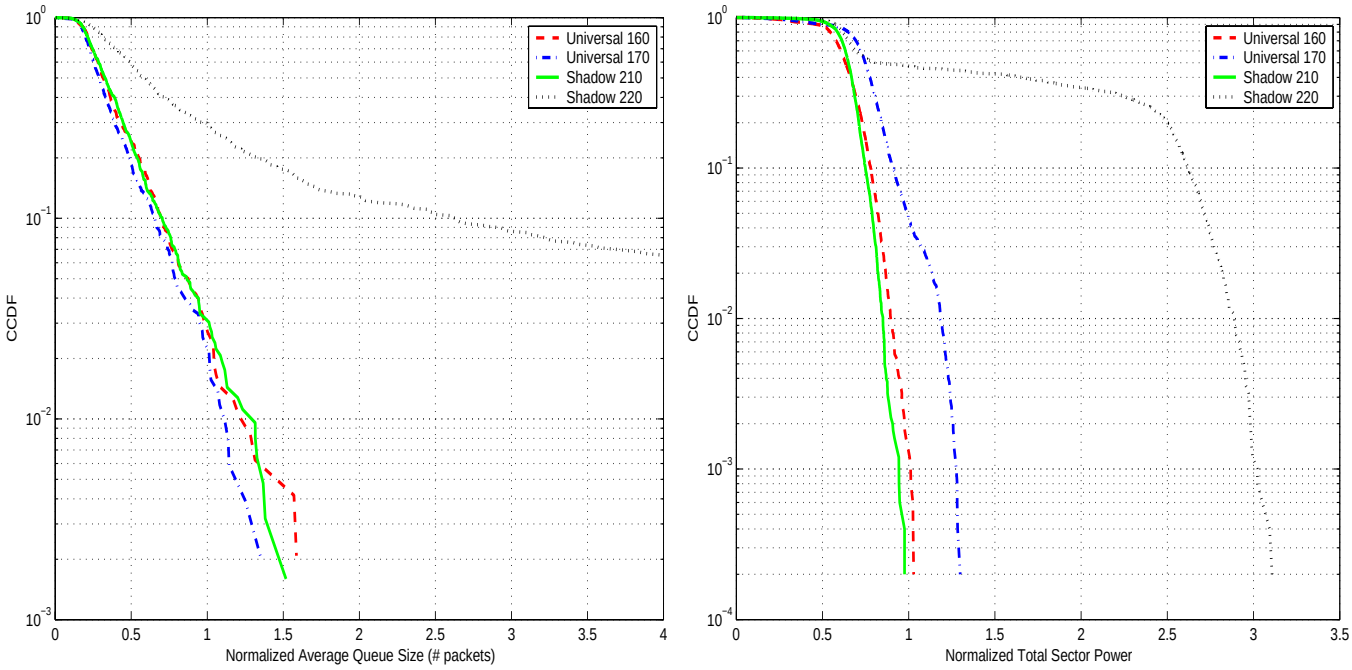

Figure 7: Complementary cumulative distribution functions of the total sector power and mean queue size for 20 dB cell edge SNR

efficient frequency plans independent of the number of sub-bands. Furthermore, based on the total sector power and average delay CDFs not shown here, we determined that the maximum number of users that can be supported with these larger number of sub-bands is nearly the same as that with three sub-bands. It is slightly smaller because of the reduced statistical multiplexing from dividing the total number of sub-carriers into more sub-sets each with fewer sub-carriers. Thus it is best to set the number of sub-bands such that there is sufficient flexibility to achieve the desired reuse but no larger than that.

5.2.3 Performance for non-uniform user distributions

In this section we show how the shadow algorithm automatically selects the appropriate frequency plans when distributions of users within the sectors are not uniform. In Figure 10 the normalized average transmit power and average number of sub-carriers used in the three sub-bands are shown for the case in which users are distributed close to the base station in Sector 1 while users are distributed near the edge of the sector in Sectors 2 and 3. The results correspond to the penetration loss of 10 dB or cell edge SNR of 20 dB. Center and edge distribution of users is achieved by selecting in Sector 1 users that have average full power SINR of 10 dB or higher and by selecting in Sectors 2 and 3 users that have full power SINR that is 2 dB or lower. As can be seen in the figure Sector 1 utilizes all the three bands nearly uniformly while Sectors 2 and 3 transmit a significant fraction of their power in one of the sub-bands most heavily, thus avoiding each other. This is expected since Sector 1 users do not experience significant interference from Sectors 2 and 3 because they are located close to the center of their own sector. Conversely, for the same number of users, significant interference is not caused by Sector 1 transmission to users in Sectors 2 and 3. Sectors 2 and 3 transmissions, on the other hand, interfere with each other. Thus running the shadow algorithm results in a reasonable

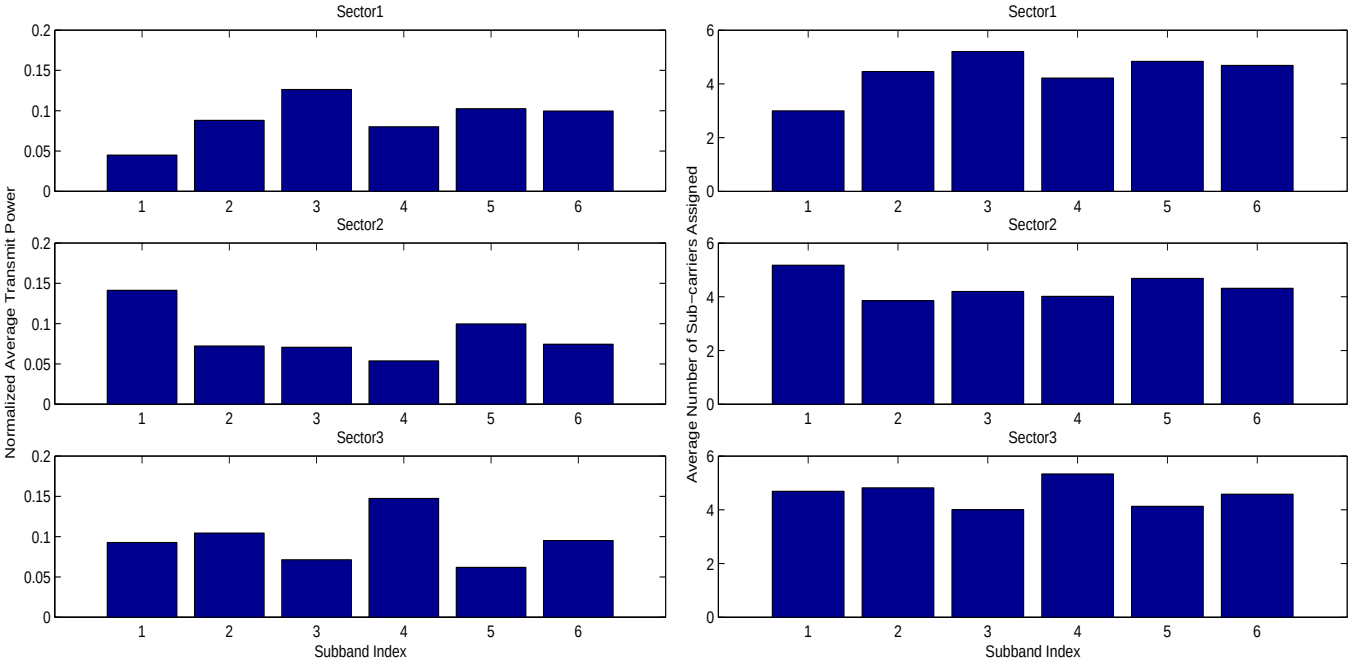

Figure 8: (a) Normalized average transmission power in the different sub-bands and (b) Average number of sub-carriers assigned in he different sub-bands for 6 sub-bands case

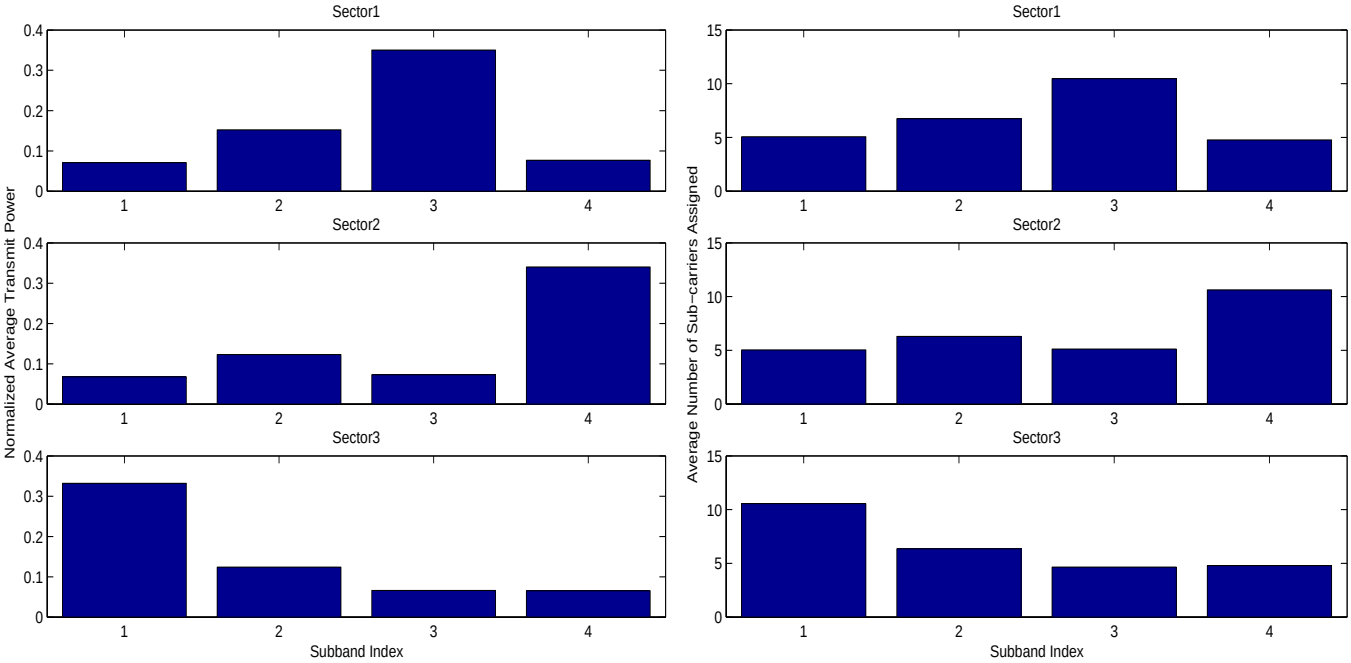

Figure 9: (a) Normalized average transmission power in the different sub-bands and (b) Average number of sub-carriers assigned in he different sub-bands for 4 sub-bands case

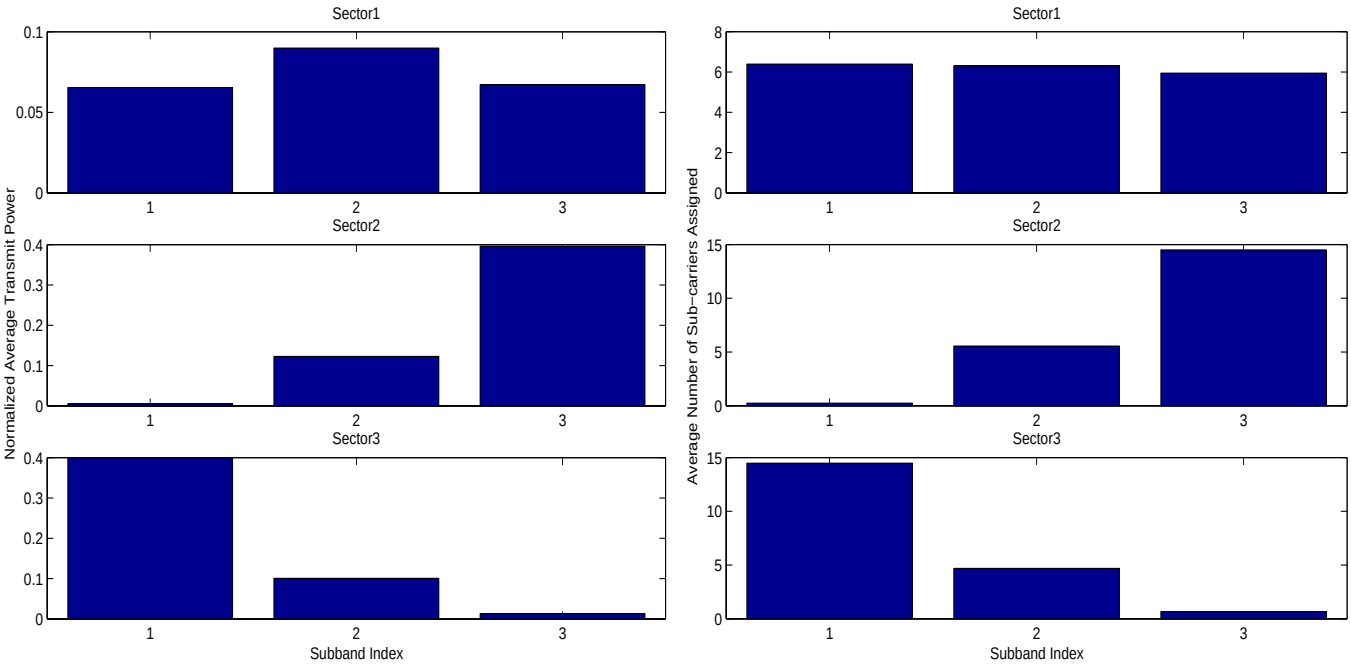

Figure 10: (a) Normalized average transmission power in the different sub-bands and (b) Average number of sub-carriers assigned in the different sub-bands for users distributed only in the center of Sector 1 and only in the edge in Sector 2 and Sector 3

reuse pattern. The gain in terms of the number of users that can be supported over a single band universal reuse approach is about 45%.

In Figure 11 the normalized average transmit power and average number of sub-carriers used in the three sub-bands are shown for the case in which users are distributed near the edge of the sector in all three sectors. As before edge users are characterized by full power SINR that is 2 dB or lower. In this case, again as expected, the shadow algorithm results in a frequency reuse plan that is closer to 1/3 reuse, i.e., using separate frequency bands in the three sectors. This is expected because all three sectors' transmissions interfere with that of each other because users are at the edge. The gain in terms of the number of users that can be supported over a single band universal reuse approach is about 87% for the case of 20 dB cell edge SNR.

In Figure 12 results are shown for the case in which users are distributed near the base station in the sector in all three sectors. As before sector center users are characterized by full power SINR that is 10 dB or higher. In this case, there is relatively small amount interference between sectors and use of all three bands in all sectors is efficient. The shadow algorithm once again results in the expected reuse pattern, which is close to the universal frequency reuse plan, as seen from the figure.

6 Summary and future work

Interference avoidance in OFDMA systems through intelligent resource allocation is a technique to improve overall system capacity. The general approach to the interference avoidance problem, taken in this paper, is to make each sector "selfishly" pursue its own performance objective, which naturally

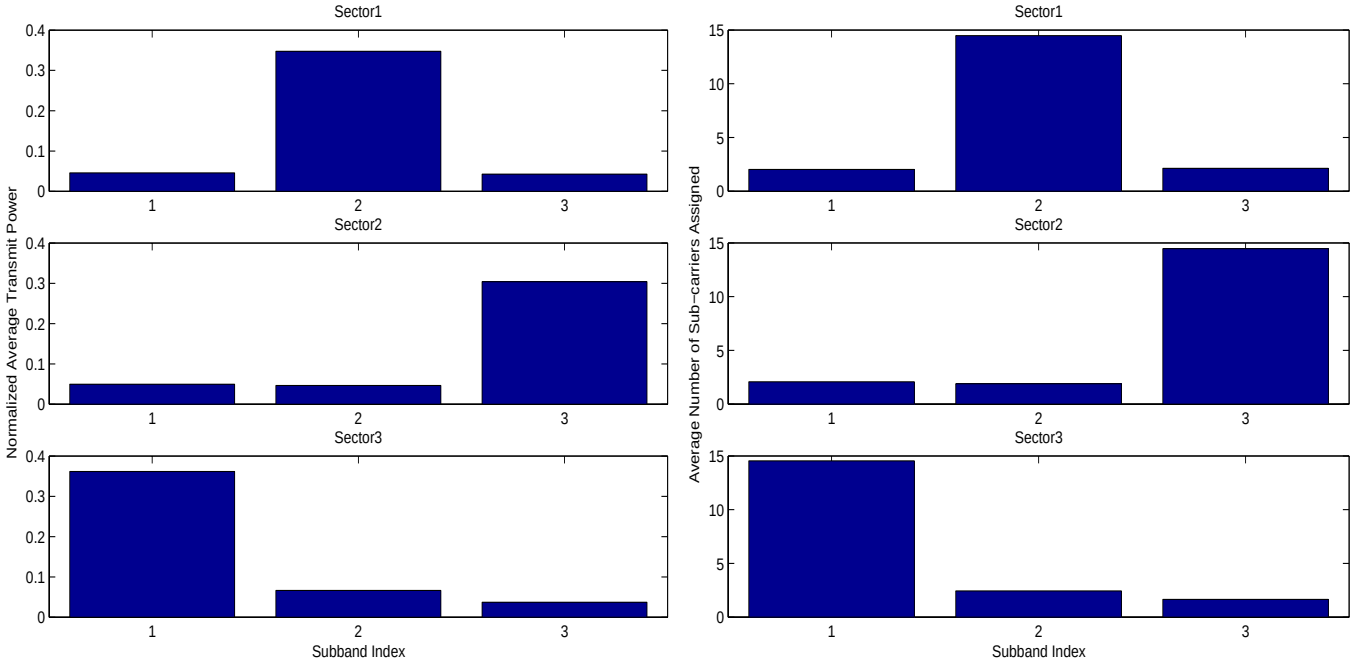

Figure 11: (a) Normalized average transmission power in the different sub-bands and (b) Average number of sub-carriers assigned in the different sub-bands for users distributed only in the edges of the three sectors

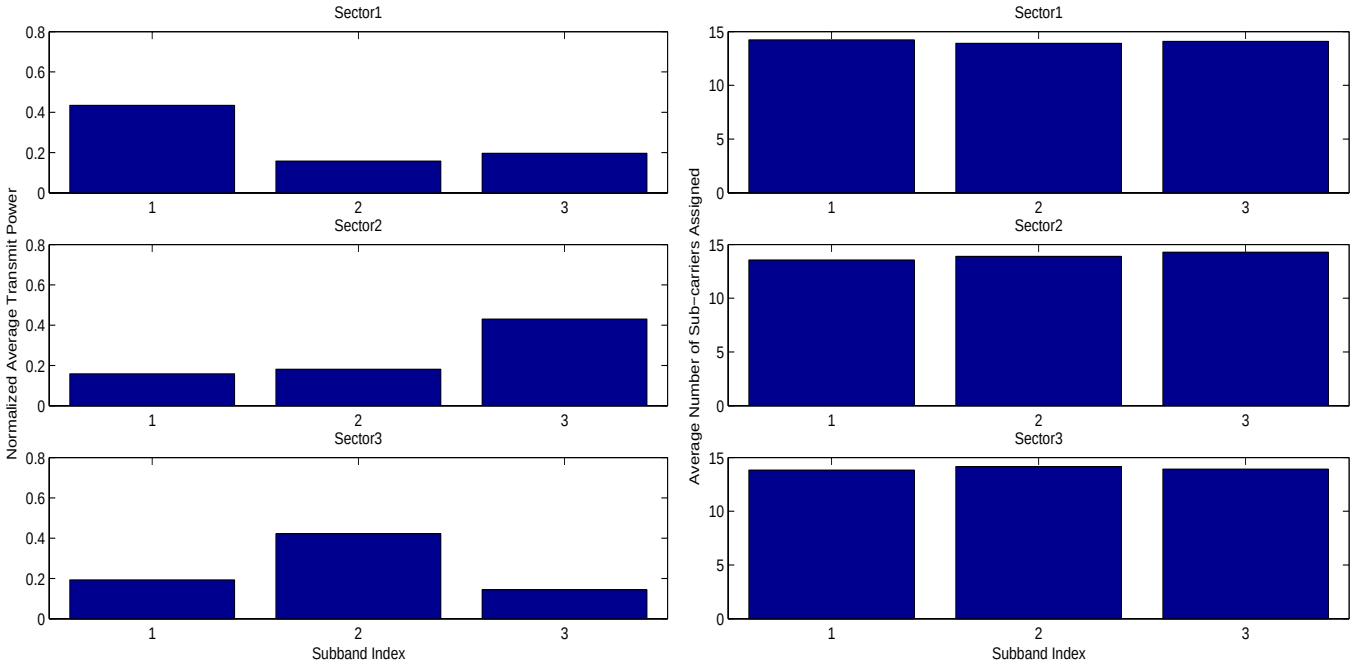

Figure 12: (a) Normalized average transmission power in the different sub-bands and (b) Average number of sub-carriers assigned in the different sub-bands for users distributed only in the centers of the three sectors

leads to completely distributed algorithms. We proposed a specific distributed algorithm that only uses the information fed back from the users, and demonstrated that the algorithm indeed achieves efficient interference avoidance, without any prior planning. Extensive simulations showed efficient FFR patterns matched to the user distributions are achieved automatically. For a fluid model of the system behavior under our algorithm, we proved the existence of Nash equilibria in a general setting; convergence results were also proved for a symmetric one-dimensional case.

While the algorithm was described for CBR type traffic, a similar approach can also be adopted for delay tolerant best effort traffic. The specific algorithm for best effort traffic, and algorithms for a combination of best effort and CBR traffic types existing simultaneously are topics for further research.

References

[1] S. Das and H. Viswanathan, "Interference mitigation through intelligent scheduling," in *Proceedings of the Asilomar Conference on Signals and Systems*, Asilomar, CA, November 2006.

[2] Third Generation Partnership Project 2, "Ultra Mobile Broadband Technical Specifications,", http://www.3gpp2.org, March 2007

[3] Third Generation Partnership Project, Radio Access Network Work Group 1 Contributions, http://www.3gpp.org, September 2005

[4] S. T. Chung, S. J. Kim, J. Lee, and J.M. Cioffi, "A game theoretic approach to power allocation in frequency-selective Gaussian interference channels," *Proceedings of the IEEE International Symposium on Information Theory*, pp 316-316, July 2003

[5] E. Altman, K. Avrachenkov, and A. Garnaev, "Closed form solutions for water-filling problem in optimization and game frameworks," 2007, submitted.

[6] R. Etkin, A. Parekh, and D. Tse, "Spectrum sharing for unlicensed bands," *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2005

[7] A. Gjendemsjo, D. Gesbert, G. E. Oien, and S. G. Kiani, "Optimal power allocation and scheduling for two-cell capacity maximization," *Proceedings of the IEEE RAWNET (WiOpt)*, April 2006

[8] S. Das, H. Viswanathan, and G. Rittenhouse, "Dynamic load balancing through coordinated scheduling in packet data systems," *Proceedings of INFOCOM*, 2003

[9] T. Bonald, S. C. Borst, and A. Proutiere, "Inter-cell scheduling in wireless data networks," *Proceedings of European Wireless Conference*, 2005

[10] L. V. Kantorovich and G. P. Akilov. *Functional Analysis*, 2nd ed. Pergamon Press, New York, 1982.

[11] A. L. Stolyar. Maximizing Queueing Network Utility subject to Stability: Greedy Primal-Dual Algorithm. *Queueing Systems*, 2005, Vol.50, No.4, pp.401-457.

[12] R. D. Yates. A Framework for Uplink Power Control in Cellular Radio Systems. *IEEE Journal on Selected Areas in Communications*, 1995, Vol.13, No.7, pp.1341-1347.