Dynamic association for load balancing and interference avoidance in multi-cell networks — Source link ⧉

Kyuho Son, Song Chong, Gustavo de Veciana

Institutions: KAIST, University of Texas at Austin

Related papers:

- User Association for Load Balancing in Heterogeneous Cellular Networks

- Generalized Proportional Fair Scheduling in Third Generation Wireless Data Networks

- Distributed α -Optimal User Association and Cell Load Balancing in Wireless Networks

- A survey on 3GPP heterogeneous networks

- Cell Association and Interference Coordination in Heterogeneous LTE-A Cellular Networks

Dynamic Association for Load Balancing and Interference Avoidance in Multi-cell Networks

Kyuho Son, *Student Member, IEEE*, Song Chong, *Member, IEEE*
and Gustavo de Veciana, *Senior Member, IEEE*

Abstract

Next-generation cellular networks will provide higher cell capacity by adopting advanced physical layer techniques and broader bandwidth. Even in such a next-generation system, boundary users would suffer from low throughput due to severe inter-cell interference (ICI) and unbalanced user distributions among cells, unless additional schemes to mitigate this problem are employed. In this paper, we tackle this problem by jointly optimizing partial frequency reuse (PFR) and load-balancing schemes in a multi-cell network. We formulate this problem as a network-wide utility maximization problem and propose optimal offline and efficient online algorithms to solve this. Our online algorithm turns out to be a simple mixture of inter- and intra-cell handover mechanisms for existing users and admission control and cell-site selection mechanisms for newly arriving users. A remarkable feature of the proposed algorithm is that it uses a notion of *expected throughput* as the decision making metric, as opposed to signal strength in conventional systems. Extensive simulations demonstrate that our online algorithm can not only closely approximate network-wide proportional fairness (PF) but also provide two types of gain, *interference avoidance (IA)* gain and *load balancing (LB)* gain, which yield $20 \sim 100\%$ throughput improvement of boundary users (depending on traffic load distribution), while not penalizing total system throughput. We also demonstrate that this improvement cannot be achieved by conventional systems using universal frequency reuse and signal strength as the decision making metric.

Index Terms

Inter-cell interference (ICI), interference avoidance, load balancing, handover, admission control, multi-cell network, network utility maximization.

I. INTRODUCTION

To support higher data rates, several next-generation wireless broadband systems based on OFDMA (Orthogonal Frequency Division Multiple Access) are currently being standardized including: IEEE 802.16/ WiMAX (Wireless Interoperability for Microwave Access) [1] and 3GPP LTE (Long Term Evolution) [2]. In these promising systems, downlink signals originating from the same base station (BS) do not interfere with each other because subbands are allocated orthogonally across users. By contrast, signals

from different BSs may interfere and as a consequence, inter-cell interference (ICI) is a major source of performance degradation. In particular, users at the cell edge (or simply, boundary users) may have low signal to interference plus noise ratio (SINR) because such locations suffer severely from ICI. In addition, in real-world systems, users are not evenly distributed across cells, yielding load imbalance between cells, which is the second major source of system-wide performance degradation. Especially, the performance of boundary users in a hot spot cell is mostly affected by this load imbalance so that they might be unable to get services. To guarantee a reasonable system-wide quality of service (QoS) irrespective of users' geographical locations and enhance cell coverage, effective ICI mitigation and load balancing schemes are required.

There has been several research on multi-cell networks, which can be classified into two types. The first is a traditional load balancing problem [3]–[6], and the second is ICI mitigation problem [7]–[11] attracting much attention recently. However, little work has been done to jointly consider both load balancing and ICI mitigation so far.

For load balancing, a cell breathing technique was investigated in [3] and [4]. It contracts (or expands) the coverage of congested (or under-loaded) cells by reducing (or raising) the power level, and therefore the load becomes more balanced. Sang *et al.* [5] proposed an integrated framework consisting of a MAC-layer cell breathing technique and load-aware handover/cell-site selection to deal with load balancing. Bu *et al.* [6] were the first to rigorously consider a mathematical formulation of network-wide proportional fairness (PF) [12] where dynamic associations between users and BSs are decision variables. They showed that the general problem is NP-hard and proposed efficient optimal offline and online heuristic algorithms to approximately solve the problem. However, none of these works had considered ICI mitigation schemes in conjunction with their proposed load balancing schemes, which have an extra potential to further increase the system-wide performance.

For mitigating ICI, a brute-force approach is the use of traditional frequency reuse schemes [7] with a reuse factor greater than one. The more ICI is mitigated by using the higher reuse factor, the less resource is available at each cell. Frequency reuse will be effective in improving the throughput of low SINR users at the cell edge but less effective to high SINR users in the inner region of the cell so that it can waste frequency resources unless selectively applied.

More elaborate work on mitigating ICI has been done by [8]–[11]. Li *et al.* [8] formulated an optimization problem to maximize the system throughput in a multi-cell OFDMA system. In their solution, a RNC (Radio Network Controller) coordinates the interference among multiple cells so that each cell utilizes not all but around 80% of its subbands to avoid the dominant ICI. Bonald *et al.* [9] examined the capacity gains achievable by inter-cell time resource sharing in CDMA/HDR systems. They formulated an optimization problem which coordinates the activity phases of BSs so as to provide higher data rate for boundary users by mitigating ICI. Even though cell selection for load balancing is also considered and studied in [11] in a three-cell system with an arbitrary traffic distribution, numerical results are only available for limited cases, that is to say, they do not give a clear answer to general multi-cell networks with heterogeneous traffic distribution. In both the [8], [9], it is noteworthy that using only partial resources (frequency and time, respectively) is essential to obtain potential performance gains associated with mitigating ICI.

In our work, partial frequency reuse (PFR), a practical ICI mitigation scheme is considered in conjunction with load balancing. Unlike traditional frequency reuse schemes, where all users share the same reuse factor, such as 1 (universal), 3, 4 or 7, PFR allows users in different channel conditions to enjoy different reuse factors. In this scheme, the entire system bandwidth is divided into two groups of subbands: *inner band* (with universal reuse factor) and *outer band* (with reuse factor greater than one). Each cell uses all the subbands in the inner band and a portion of the subbands in the outer band. For example, in type 1 cells in Fig. 1, users in the inner region of the cell are allowed to use the entire inner band and users at the cell edge are allowed to use a portion of the outer band, i.e., band O1. According to the current state of 3GPP LTE [13], ICI mitigation approaches are classified into three types: 1) inter-cell interference randomization, 2) inter-cell interference cancellation and 3) inter-cell interference coordination/avoidance. The PFR scheme explained above belongs to the last category, i.e., aims at avoiding ICI by selectively restricting downlink frequency resources in a coordinated way among multiple cells.

In this paper, we extend the Bu's work [6] to multi-cell networks using PFR and jointly optimize load balancing and ICI mitigation to achieve network-wide proportional fairness. We assume that each BS has limited frequency resources based on an ICI pre-coordination of PFR and independently runs a proportionally fair scheduler. In this setting, we concentrate on the following association (long-term binding relationship) problems:

- *Inter-cell association*: To which BS should each user be associated?

- *Intra-cell association*: Should a user be allocated to the inner or outer band?

The remainder of this paper is organized as follows. In Section II, we present our system model. In Section III, we present a formulation of network-wide proportional fairness in a multi-cell network using PFR. In Section IV and V, we present offline optimal and heuristic online algorithms that use the *expected throughput* as a key metric in making association decisions. In Section VI, we discuss our simulations for a two-tier cellular system, which demonstrate that our online algorithm can closely achieve network-wide PF. We also show that they perform better than conventional systems with universal frequency reuse where each user connects to the BS offering the best signal strength. In Section VII, we present further discussion and conclusions.

II. SYSTEM MODEL

We consider a downlink network with PFR consisting of a set of BSs and users. We shall make the following assumptions for the remainder of this paper:

1) Each user generates persistent data traffic and has an infinitely backlogged queue.

2) Every time slot, each user can be associated with only one BS using either inner or outer band.

3) Each user knows instantaneous achievable rates of both inner and outer bands from all BSs to itself.

4) Each BS knows instantaneous achievable rates of both inner and outer bands from itself to all users.

5) Each time slot, each BS schedules two users, one for inner band and the other for outer band.

6) Each BS allocates power equally to all the subbands being used.

Remark: Assumptions 3) and 4) require network-wide channel estimations and feedbacks. As you will see in Section V, however, our proposed online algorithm substantially reduces this overhead by considering neighboring BSs only. Assumption 5) implies that we schedule users on a per-band basis rather than a per-subband basis. We make this assumption to further reduce the channel estimation and feedback overhead since per-band scheduling does not require per-subband channel information. As the channel becomes more frequency-selective, per-subband scheduling will outperform per-band scheduling. Note, however, that the intra-cell scheduling is not the focus of this paper and our result later will not prohibit us from assuming per-subband scheduling. The equal power assumption in 6) has been frequently used for implementation simplicity as well as analytical tractability in downlink resource allocation problems

[8], [14]. Moreover, equal power allocation is near optimal in many cases especially in high SINR regime [15]–[17].

A. Resource partitioning

Fig. 1 depicts an example of frequency partitioning in a regular multi-cell network using PFR where the entire bandwidth is divided into two subband groups: inner and outer bands. For ease of presentation, all boundaries of cells, inner and outer regions are depicted by straight lines. In reality, they would be irregular due to shadowing in the environment as well as the distribution of users. We let α denote the portion of the entire bandwidth that is reserved for the inner band. Each BS can use, the whole inner band for users in the inner region of the cell, while using only a portion of the outer band for boundary users for ICI mitigation. The inner and outer regions are logical concepts, which are dynamically changed depending on channel conditions and the distribution of users.

We let RF denote reuse factor of outer band, that is, the fraction of the outer band which can be used in each cell. Note that this is equivalent to the inverse of the traditional definition of a reuse factor. As shown in Fig. 1 by using a reuse factor $RF = 1/3$ in the outer band, we can remove the first-tier of ICI. We can mitigate the dominant ICI by choosing $RF = 6/7$ (Fig. 6 in [9]). Many other patterns can be designed. Note that the smaller RF we use, the less ICI users are likely to experience. However, the total resource available in each cell, $\alpha + RF(1 - \alpha)$, is also reduced. Because of this tradeoff between the degree of ICI mitigation and total resource available, choosing an optimal resource partitioning (α, RF) is critical. In this paper, we assume that a proper resource partitioning is given and fixed, and instead we focus on the association problem, i.e., decide which BS and band each user should be associated with.

B. Link model

The sets of BSs and users in the network are denoted by N and K, respectively. The set of bands, consisting of inner and outer bands, is denoted by $B = \{in, out\}$. The received SINR at time slot t for user $k \in K$ from BS $i \in N$ on band $b \in B$ can be written as:

$$\gamma_{i,k}^b(t) = \frac{g_{i,k}^b(t)p_i^b}{\phi \, g_{i,k}^b(t)p_i^b + \sum_{j \in L_i^b, j \neq i} g_{j,k}^b(t)p_j^b + N_0^b} \, , \tag{1}$$

where

- $g_{i,k}^b(t)p_i^b$ is the signal strength received by the user k from BS i at time slot t with p_i^b and $g_{i,k}^b$ representing the transmit power of BS i on band b and the channel gain between BS i and user k on band b, respectively. The channel gain takes into account the path loss, log-normal shadowing and fast fading.

- L_i^b is the set of BSs allowed to use the same inner or outer band b as BS i.

- N_0^b is the additive white Gaussian noise (AWGN) on band b. Without loss of generality, we assume the noise level is the same for all users.

- ϕ is the orthogonality (or self-interference) factor that models transmitter and receiver non-linearities and limits the maximum SINR. For our simulations, we set ϕ to be 0.01 which correspondingly bounds the maximum SINR by 20dB.

Given p_i^b, the instantaneous achievable rate at time slot t for user k from BS i on band b as given by $R_{i,k}^b(t) = BW^b \log_2\left(1 + \gamma_{i,k}^b(t)\right)$ bps where BW^b is the bandwidth of band b. Note that once a resource partitioning (α, RF) is fixed, p_i^b is fixed due to Assumption 6).

III. PROBLEM FORMULATION

A. General utility maximization problem - P1

We shall start by discussing a general utility maximization problem in the multi-cell network setting. Given p_i^b, the set of instantaneous achievable rates of user k from BS i on band b, denoted by $\{R_{i,k}^b(t),\ t \in \mathbb{Z}\}$ where $\mathbb{Z}$ is the set of nonnegative integers, is assumed to be a stationary ergodic process. These processes are independent, but not necessarily identically distributed across users, bands and BSs. Each user can be served from any BS on any band, and these may vary at every time slot. Define an assignment indicator vector at time slot t by $\mathbf{I}(t) = \left(I_{i,k}^b(t) \mid i \in N, k \in K, b \in B\right)$ where $I_{i,k}^b(t)$ is equal to 1 if BS i assigns band b to user k at time slot t, and 0 otherwise. There are two constraints on assignments:

$$\sum_{k \in K} I_{i,k}^b(t) = 1, \quad \forall i \in N \text{ and } \forall b \in B, \tag{2}$$

$$\sum_{i \in N} \sum_{b \in B} I_{i,k}^b(t) \leq 1, \quad \forall k \in K. \tag{3}$$

The first constraint in Eq. (2) ensures that every time slot each BS schedule two users, one for inner band and the other for outer band. The second constraint in Eq. (3) ensures that each user receives

data from at most one BS on either inner or outer band. There is also a power constraint given by $\sum_{b \in B} p_i^b = p^{max}, \forall i \in N$ where p^{max} is the transmit power budget of a BS.

Let $S_k(t) = \sum_{i \in N} \sum_{b \in B} \left[R_{i,k}^b(t) I_{i,k}^b(t) \right]$ be the data rate assigned to user k at time slot t. We define the estimated long-term throughput for user k up to time slot t by the sample average $R_k(t) = \frac{1}{t} \sum_{\tau=1}^{t} S_k(\tau)$. Letting $\epsilon_t = 1/(t+1)$, this can be written in a recursive form: $R_k(t+1) = R_k(t) + \epsilon_t \left[S_k(t+1) - R_k(t) \right]$. We let $\mathbf{R}(t) = (R_k(t) \mid k \in K)$ denote the vector of the estimated long-term throughputs.

Next we define achievable steady-state region $\mathcal{R}$ by the set of all possible long-term throughput vectors that are achievable under all possible assignment policies. Each long-term throughput vector $\mathbf{r} = (r_k \mid k \in K) \in \mathcal{R}$ corresponds to an assignment policy. Then, we can easily show that the region $\mathcal{R}$ is a closed bounded convex set (We omit the proof here, see Section III in [18]). Our objective is to maximize the aggregate utility over the convex feasible set $\mathcal{R}$:

$$\textbf{[P1]} \quad \max \sum_{k \in K} U_k(r_k) \tag{4}$$

$$\text{s.t.} \quad \mathbf{r} \in \mathcal{R}, \tag{5}$$

where $U_k(\cdot)$ is an increasing, strictly concave and continuously differentiable utility function for user k.

Broadly speaking, we are interested in finding optimal assignment $\mathbf{I} = (\mathbf{I}(t), t \in \mathbb{Z})$ which maximizes the aggregate utility $U(\mathbf{R}(t)) = \sum_{k \in K} U_k(R_k(t))$ as $t \to \infty$. Consider the aggregate utility drift as a myopic view of optimization:

$$
\begin{aligned}
U(\mathbf{R}(t+1)) - U(\mathbf{R}(t)) &= \sum_{k \in K} \left[U_k\left(R_k(t) + \epsilon_t \left[S_k(t+1) - R_k(t) \right]\right) - U_k(R_k(t)) \right] \\
&= \sum_{k \in K} U_k'(R_k(t)) S_k(t+1) \epsilon_t - \sum_{k \in K} U_k'(R_k(t)) R_k(t) \epsilon_t + O(\epsilon_t^2),
\end{aligned}
$$

where the second equation follows from the first order Taylor expansion in the neighborhood of $\epsilon_t \approx 0$. Since only the first term in the above depends on future decisions, selecting $\mathbf{I}(t+1)$ which maximizes the first term will maximize $U(\mathbf{R}(t+1)) - U(\mathbf{R}(t))$. Therefore, we obtain the following so-called *gradient-based scheduler*:

$$\mathbf{I}(t+1) = \arg \max_{\mathbf{I}(t+1)} \sum_{k \in K} U_k'(R_k(t)) \left[\sum_{i \in N} \sum_{b \in B} R_{i,k}^b(t+1) I_{i,k}^b(t+1) \right]. \tag{6}$$

Recently, Stolyar [18] showed that the sample average vector $R_k(t)$ asymptotically converge to the unique optimal solution to **[P1]** when the gradient-based scheduling algorithm is used.

Finding optimal assignment $\mathbf{I}(t+1)$ in Eq. (6) is equivalent to solving the maximum-weight bipartite matching problem (MWMP) with nonnegative weights $U'_k(R_k(t))R^b_{i,k}(t+1)$. This can be solved in polynomial-time by the well-known Hungarian method whose complexity is $O(|K^3|)$ while the complexity of exhaustive search is $O(|K|^{2|N|})$ [19]. Although it takes polynomial-time, its realization is still too complex due to the large number of users, $|K|$, in the system. Moreover, gathering global knowledge (achievable data rates of inner and outer bands for every user-BS pair) at a central node *for each time slot* makes its realization even more difficult in practice.

B. *Utility maximization problem with network-wide proportional fairness objective - P2*

Finding optimal assignment $\mathbf{I}(t+1)$ at every time slot in **[P1]** is a network-wide packet scheduling problem which requires centralized computation at every time slot with global information gathering. In contrast to this, packet scheduling in practice is typically undertaken by individual BSs independently provided that associations between users and BSs are given. In this subsection, we reformulate the network-wide utility (NUM) maximization problem in order to take into account this practical scheduling concern. We assume that each BS independently runs a packet scheduler at every time slot for a given set of users who are currently associated with it, as typical in practice. Instead we will change the association between BSs and users occasionally in a longer time scale (longer than given time slot) in order to maximize the network-wide aggregate utility. Thus, our focus from here is to find optimal association between BSs and users for a given intra-cell scheduling.

Consider a network with a fixed number of users, i.e., no user arrivals or departures.[1] Let $\mathbf{x} = (x^b_{i,k} \mid i \in N, k \in K, b \in B)$ denote the association vector, where

$$x^b_{i,k} = \begin{cases} 1, & \text{if user } k \text{ is associated with BS } i \text{ on band b} \\ 0, & \text{otherwise .} \end{cases} \tag{7}$$

Since a user can be associated with only one BS on either inner or outer band, we have a unique association constraint: $\sum_{i \in N} \sum_{b \in B} x^b_{i,k} = 1, \ \forall k \in K$. For a given $\mathbf{x}$, each BS i runs the following intra-cell scheduler to select the optimal user k^{b*}_i in each band $b \in B$ at every time slot:

$$k^{b*}_i(t+1) = \arg\max_{k \in K^b_i} U'_k(R_k(t))R^b_{i,k}(t+1), \tag{8}$$

---

[1]Our online algorithm in the forthcoming section still works even if the system is dynamic where users are mobile, arrive and depart.

where $K_i^b = \{k \mid x_{i,k}^b = 1, k \in K\}$ is the set of all users who are associated with BS i on band b.

Following procedure analogous to that used in [20], the average throughput $\mathbf{R}(t)$ corresponding to the scheduler in Eq. (8) converges weakly to the unique equilibrium point $\mathbf{r}^*$ of the following ODE (ordinary differential equation) as $\tau \to \infty$:

$$\frac{dr_k(\tau)}{d\tau} = \sum_{i \in N} \sum_{b \in B} x_{i,k}^b \left[\int R_{i,k}^b I_{i,k}^b \, f(R_{i,k}^b) \, dR_{i,k}^b \right] - r_k(\tau), \ \ \forall k \in K, \tag{9}$$

where $I_{i,k}^b$ are equal to 1 if $R_{i,k}^b U_k'(r_k) \geq R_{i,j}^b U_j'(r_j)$ for all $j \in \left(l \mid x_{i,l}^b = 1 \right)$, and 0 otherwise; $f(\cdot)$ is the probability density function of achievable rate $R_{i,k}^b$ given p_i^b. The symbol $R_{i,k}^b$ is used as the canonical value of the components $R_{i,k}^b(t)$, i.e., $R_{i,k}^b$ is the random variable with the same distribution as the stationary stochastic process $R_{i,k}^b(t)$. To explicitly determine the stationary expectation of long-term throughput, i.e., find a solution to the above fixed point equation is generally quite hard. However, in the case where all users have the *log utility function*, have Rayleigh fading channels, and the feasible rate is linear in the SINR, it can be shown (See Section III in [20]) that

$$r_k^* = \sum_{i \in N} \sum_{b \in B} x_{i,k}^b \left[\frac{G(\sum_{k \in K} x_{i,k}^b) \, E[R_{i,k}^b]}{\sum_{k \in K} x_{i,k}^b} \right], \tag{10}$$

where $E[R_{i,k}^b]$ is the statistical average of achievable data rate $R_{i,k}^b$. And $G(y) = \sum_{k=1}^y \frac{1}{k}$ represent a multi-user diversity gain (scheduling gain) depending only on the number of users competing with the same resource. Note that the same result in Eq. (10) can be derived by another method [21] using the fact that PF scheduler assigns an equal fraction of slots to users, i.e., temporal fairness. Therefore, the problem of associating users with a BS and band can be formulated as the following utility maximization problem with the network-wide PF as an objective in the multi-cell network: [2]

$$\textbf{[P2]} \quad \max_{\mathbf{x}} \ \sum_{k \in K} \log\left(r_k\right) \tag{11}$$

$$\text{s.t.} \ \sum_{i \in N} \sum_{b \in B} x_{i,k}^b \ = \ 1, \qquad\qquad\qquad k \in K,$$

$$x_{i,k}^b \ \in \ \{0, 1\}, \qquad\qquad\qquad i \in N, k \in K, b \in B,$$

$$y_i^b \ = \ \sum_{k \in K} x_{i,k}^b, \qquad\qquad\qquad i \in N,$$

$$r_k \ = \ \sum_{i \in N} \sum_{b \in B} x_{i,k}^b \left[\frac{G(y_i^b) E[R_{i,k}^b]}{y_i^b} \right], \ \ k \in K.$$

[2]To keep the notation simple, we suppress the superscript from r_k^* to r_k.

This problem should be solved whenever the arrival and departure of users occur or channel conditions are severely changed by users' mobility. Once associations are determined by solving **[P2]**, each BS can execute the PF scheduler independently among its associated users. The form of PF scheduler can be easily obtained by putting $U'_k(R_k(t)) = 1/R_k(t)$ into Eq. (8).

IV. OFFLINE ALGORITHM

Next we present an offline algorithm for **[P2]** that finds the optimal associations. The following proposition states the interesting property of **[P2]**.

Proposition 1. *If the number of inner/outer users in each BS is fixed, then* **[P2]** *can be reduced to a maximum weighted matching problem (MWMP) which can be solved in polynomial time.*

Proof. If $\mathbf{y} = \left(y_i^b \mid i \in N, b \in B\right)$ is fixed, then $w_{i,k}^b = \frac{G(y_i^b)E[R_{i,k}^b]}{y_i^b}$ is fixed. The problem is equivalent to finding a maximum weight bipartite matching between virtual $2|N|$ BSs (each BS is split into two BSs, inner BS and outer BS) and $|K|$ users each with a nonnegative weight $w_{i,k}^b$. This can be solved in a polynomial time by the well-known Hungarian method $O(|K|^3)$. ∎

If the number of BSs is a constant $|N|$, then the number of all possible configurations for $\mathbf{y}$ is $O(|K|^{2|N|})$. For each $\mathbf{y}$ configuration, we solve the above MWMP. Because both the number of all configurations and the computational complexity of MWMP are polynomial, the total running time of the proposed offline algorithm is also polynomial, $O(|K|^{2|N|+3})$. However, it is computationally too complex, e.g., $O(|K|^{2|N|+3}) = O(190^{41}) \approx 2.7 \times 10^{93}$ when the number of BSs is 19 (two-tier system) and each BS has only 10 users. In addition, the feedback overhead associated with collecting all users' average achievable data rates to a central node is excessive. In order to overcome these computational and feedback overheads, we consider the design of a heuristic online algorithm.

V. ONLINE ALGORITHM

The objective in solving **[P2]** is to determine the association of each user, which is naturally related to handover and the cell site selection. Conventional algorithms for handover and cell-site selection are based on signal strength. Each user selects the best BS with the strongest mean channel quality, and it binds to the inner (or outer) band if the mean channel quality is higher (or lower) than a certain threshold. However, such a decision does not maximize the total utility since the satisfaction of a user depends on its

actual throughput r_k rather than the signal strength. Moreover r_k depends not only on the signal strength but on the population of users served by the BS. Even if the signal strength is high, the actual throughput may be low if many users are competing for the same resource. This observation motivates us to develop a new algorithm to bind users to BSs and bands. The following properties are essential in design of our online algorithm.

Proposition 2 (Intra-cell handover condition). *Assume a user k is binding to BS i through band b and the numbers of inner/outer band users in BS i are large enough. Then changing the band currently being used to band $\bar{b}$ will improve the value of the network-wide objective function in Eq. (11) if*

$$R_{i,k}^b = \frac{G(y_i^b)E[R_{i,k}^b]}{y_i^b} < \frac{G(y_i^{\bar{b}} + 1)E[R_{i,k}^{\bar{b}}]}{y_i^{\bar{b}} + 1} = R_{i,k}^{\bar{b}}, \tag{12}$$

where the overline represents changing the current associated band. Thus, $\bar{b}$ is equal to out-band if a user is currently associated with in-band, and in-band, otherwise; $R_{i,k}^b$ is the *expected* long-term throughput according to Eq. (10) when user k is binding to BS i through band b.

Proposition 3 (Inter-cell handover condition). *Assume a user k is binding to BS i and numbers of users in BS i and BS j are large enough. Then moving the user to another BS j will improve the value of the network-wide objective function in Eq. (11) if*

$$R_{i,k}^b = \frac{G(y_i^b)E[R_{i,k}^b]}{y_i^b} < \frac{G(y_j^b + 1)E[R_{j,k}^b]}{y_j^b + 1} = R_{j,k}^b . \ ^3 \tag{13}$$

Note that the left and right hand sides of Eqs. (12), (13) are the *expected average throughput* of user k before and after the intra/inter-cell handover, respectively.

There is a reason why such simple conditions are obtained: when the number of users in each cell is large enough, the increment and decrement in the total utility (except user k) in cell i and j are almost the same and counterbalance each other. This makes the net increment of utility only depends on the handover of user k. Please refer to Appendix for detailed proofs. In a similar way, we can obtain a condition for admission control.

Proposition 4 (Admission condition). *Assume a new user k arrives to the system. Then admitting the user k to BS i on band b will improve the value of the network-wide objective function if*

$$R_{i,k}^b = \frac{G(y_i^b + 1)E[R_{i,k}^b]}{y_i^b + 1} > e, \tag{14}$$

---

[3] The majority of inter-cell handovers will be between outer bands due to geographical adjacency.

where the constant e is base of the natural logarithm.

These propositions suggest the importance of the *expected throughput (load-aware metric)*. Based on this observation, we suggest a heuristic online algorithm for **[P2]**. It is a simple mixture of intra/inter-cell handovers and cell-site selection with an admission control that use the *expected throughput* as a key metric in making association decisions instead of the signal strength. In our offline algorithm, whenever the arrival and departure of users occur or average channel gains are severely changed by users' mobility, we have to solve **[P2]** again. Our online algorithm, however, keeps track of these dynamics (arrival and departure, mobility) and gradually changes users' associations following the steepest utility-increasing direction. The following are the details of our proposed procedure.

A. Intra-cell handover

Step 1. (Measurement & Report) Every user k binding to BS i periodically measures and reports the average achievable data rate in the band which is not in use. It is not necessary to report that of the band in use since the instantaneous achievable rate is sent to the BS each time slot to enable scheduling.

Step 2. (Decision) If many users change their serving bands at the same time, this may result in oscillations, thus BS i chooses only the user k_i^* that achieves the largest benefit by changing its band, [4],

$$k_i^* = \arg\max_k \phi_{i,k}, \tag{15}$$

where $\phi_{i,k} = R_{i,k}^{\bar{b}}/R_{i,k}^b$, $k = 1, \ldots, (y_i^{in} + y_i^{out})$ for all users in BS i and $\phi_{i,0} = \phi^{intra}$. We introduce a hysteresis $\phi^{intra} \geq 1$ to reduce possible ping-pong effects [22].

Step 3. (Notification) If $k_i^* = 0$, then either changing the band for any user cannot increase the value of the network-wide objective function or hysteresis precludes such a change. Thus, nothing occurs. Otherwise, the BS i notifies user k^* to change its intra-cell association.

The intra-cell handover is just the procedure to change the band currently being used, rather than a real handover, as such, it brings minimal system overhead. To accommodate channel variation due to mobility, the periodicity of intra-cell handovers should be performed on a short time scale (< 1 sec), and hysteresis, if used at all, should be small ($\phi^{intra} \approx 1$).

---

[4]If more than one user achieve the same largest benefit, then a suitable random tie-breaking rule is used. The same is true for inter-cell handover in the next subsection.

B. Inter-cell handover

Step 1. (Measurement & Report) The central node periodically receives the following information from all BSs, and broadcasts this information so that every BS has the knowledge of its neighboring cells:

- y_i^b: the number of users associated with BS i through band b;

- $G(y_i^b)$: the multi-user diversity gain associated with BS i and band b.

Every BS i announces to all its associated users the above information for current and neighboring cells. Every user calculates expected throughputs from neighboring cells in addition to the current cell. Then, only users k, expecting higher throughput by changing BS i into j, report the highest ratio $\phi_k = R_{j,k}/R_{i,k}$ and the index of target cell j to the central node through the BS i.

Step 2. (Decision) To avoid the oscillation problems, the central node chooses the user k^* that achieves the largest benefit by changing its serving BS. We introduce hysteresis with $\phi_0 = \phi^{inter} \geq 1$ to reduce possible ping-pong effects.

$$k^* = \arg\max_k \phi_k \ . \tag{16}$$

Step 3. (Notification) If $k^* = 0$, then either moving any user to another BS cannot increase the value of the network-wide objective function or hysteresis precludes such a switch. Otherwise, the central node notifies the user k^*, its original BS i and target BS j to handle the inter-cell handover.

In contrast to the intra-cell handover, the inter-cell handover is true a handover and brings additional system overheads. Thus the periodicity of inter-cell handovers should be large, i.e. time scales > 1 sec and also hysteresis should be implemented enough.

C. Cell-site selection with an admission control

Step 1. (Measurement & Report) A newly arriving user k measures average achievable data rates from several BSs. It reports this information to the central node through BSs.

Step 2. (Decision) The central node chooses the best BS i^* and band b^* that gives the highest expected throughput to the user k,

$$(i^*, b^*) = \arg\max_{(i,b)} \ \phi_i^b, \tag{17}$$

where $\phi_i^b = R_{i,k}^b$, $i = 1, \ldots, N$ and $\phi_0^b = e$ according to admission condition in Eq. (14).

Step 3. (Notification) If $i^* = 0$, the user k is rejected since admitting the user k will deteriorate the value

of network-wide objective function. Otherwise, the central node notifies the user i to associate with the optimal BS i^* and band b^*.

D. Multi-user diversity gain estimation

Under the assumptions made in Section III, the multi-user diversity gain can be written as $G(y) = \sum_{k=1}^{y} \frac{1}{k}$; it only depends on the number of users sharing the same resource. We make use of this property for mathematical tractability. However, in our simulations, we use the estimation of multi-user diversity gain to calculate the expected throughput more accurately. Below we describe a detailed procedure for estimating the multi-user diversity gain $G_i^b(t)$ at time t associated with BS i and band b.

1) A scheduler module, associated with BS i and band b, receives the instantaneous achievable rate $\{R_{i,k}^b(t)\}$ from all its associated users $k \in K_i^b$ at every time slot.

2) It takes an average of data rate for each user over fixed-length time window W:

$$\overline{R_{i,k}^b}(t) = \frac{1}{W} \sum_{\tau=t-W+1}^{t} R_{i,k}^b(\tau), \; \forall k \in K_i^b. \tag{18}$$

3) It also takes an average of the data rate of the selected cases for each user over the same time window:

$$\overline{R_{i,k}^b}^*(t) = \frac{\sum_{\tau=t-W+1}^{t} R_{i,k}^b(\tau) I_{i,k}^b(\tau)}{\sum_{\tau=t-W+1}^{t} I_{i,k}^b(\tau)}, \; \forall k \in K_i^b. \tag{19}$$

4) We obtain the multi-user diversity gain by taking an average of the ratio of Eq. (19) to Eq. (18) for all its associated users:

$$G_i^b(t) = \frac{1}{y_i^b} \sum_{k \in K_i^b} \frac{\overline{R_{i,k}^b}^*(t)}{\overline{R_{i,k}^b}(t)}. \tag{20}$$

VI. PERFORMANCE EVALUATION

A. Simulation setup

The two-tier multi-cell network composed of 19(=N) hexagonal cells shown in Fig. 1 was considered, where the distance between BSs is 2km. We let $\rho_t = \frac{\lambda_t}{\mu_t}$ denote the traffic load of tier t. In any cell at tier t, users arrive according to a Poisson process with rate λ_t and depart the system after a holding time that is exponentially distributed with mean of $1/\mu_t$=60sec. We assume that users have infinitely backlogged queues during their lifetime. Initially, users are uniformly distributed in each cell and they move based on the random waypoint model in which we fix the speed at 3km/h. All BSs have the same maximum

transmission power p^{max}=20dBm which is divided into p^{in} and p^{out} proportional to the size of inner and outer bands: $p^{in} = \frac{\alpha}{\alpha + (1 - \alpha)RF} p^{max}$ and $p^{out} = p^{max} - p^{in}$. In modeling the propagation environment, a path loss $\Gamma(d_{km}) = -130 - 35 \log_{10}(d_{km})$, log-normal shadowing with a standard deviation σ_s=8dB, and Jakes' Rayleigh fading [23] models were adopted. There is a shadowing correlation of 0.5 among paths from several BSs. We have assumed that each user sees interference from other cells up to two-tiers using a wrap-around technique. The system bandwidth is 10MHz and the time slot is 5ms, this conforms with the IEEE 802.16e standard. The periods for intra- and inter-cell handovers are t^{intra}=0.1sec and t^{inter}=1sec, and ϕ^{intra}=1.01 and ϕ^{inter}=1.10 are used for hysteresis, respectively. For each given parameter set, we ran simulations over 720000 time slots (3600 sec).

B. Comparison of online algorithm with optimal

We randomly picked 100 different static (no user arrivals/departures or mobility) scenarios varying the resource partitioning and the number of users in each cell. For each scenario the optimal association was obtained by the offline algorithm, and we evaluated the heuristic-based online algorithm. Fig. 2 exhibits the CDF for the performance ratios, which are defined as the ratio between performance values obtained from the online algorithm and that from the optimal offline algorithm. As can be seen, the performance ratios of the total utility exceed 98% for all scenarios. Similarly, our online algorithm achieves a total throughput which is identical to that of optimal algorithm.[5] Thus we can conclude that *our online algorithm, which is efficient and easy to implement, is a good approximation of the offline algorithm.*

C. Interference avoidance gain

Next we move to dynamic scenarios, where users arrive and departure at/from uniformly distributed locations and also have mobility during their lifetime. We assume that every cell in the network has the same traffic load $\rho = 40$. We fix $RF = 1/3$ and evaluate the system performance by varying the portion α of the spectrum devoted to the inner band. If $\alpha = 1$, then the system is operating under universal frequency reuse since each cell uses all resources. Also if $\alpha = 0$, then the system, operates under a reuse factor of 3 since each cell uses only $1/3$ of the total spectrum. So these two points can be obtained by

---

[5]Note occasionally the total throughput achieved by an online algorithm exceeds that of the offline algorithm, however, the total utility of the throughput is always lower because this is not an optimal point.

traditional frequency reuse schemes. The middle points $(0< \alpha <1)$ are newly achieved by employing PFR and our intra-cell handover to determine the optimal user binding to inner and outer bands.

Fig. 3(a) plots the CDF of the SINRs seen by users. The SINR curve moves consistently to the right as the inner band portion α decreases. Note that only the SINR of boundary users using the ICI mitigated outer band is improved about 4∼8dB using PFR. The lower α, the more ICI avoidance gain can be obtained. However, the amount of resource available in each cell is also reduced. To observe the real gains of PFR, we examine the throughput seen by each user in Fig. 3(b). As α decreases, there is a decreasing and increasing trend for the throughput of inner region users (high throughput users) and for the throughput of boundary users (low throughput users), respectively. It is worthy of note that the increment of throughput at the cell boundary becomes smaller as α decreases. In the extreme case (α=0), the throughput of edge users is even lower than α=6/8 case due to the lack of total resource available. Meanwhile, the increment of throughput in the inner region of cell becomes larger as α decreases. Therefore, it is very critical to choose a proper α, such that the performance of boundary users will be improved as much as possible while that of users associated with the inner region of cell is not excessively degraded.

Fig. 3(c) shows the total and 5th percentile average throughput together. The 5th percentile average throughput is equal to the average of the lowest 5% throughput of users. This can be regarded as a representative performance metric of boundary users.[6] As explained above, the decrease of α from α=1 increases the 5th percentile throughput until α=3/8, but decreases in the region $\alpha <3/8$. At the same time, the total throughput remains almost the same until α=6/8, but decreases consistently after this point. From several results in Fig. 3, we can conclude that α=6/8 is a good choice.

D. Load balancing gain

Now let us consider a network with a heterogeneous user distribution. Cells in Tier 1 and 2 have $\rho = 40$, while the cell in Tier 0 has ρ_0 in [20,100]. The following four schemes are evaluated to determine the performance gains: *interference avoidance (IA)* and *load balancing (LB)*.

---

[6]Actually, *the cell boundary* is not clearly defined because a user, located closer to a BS i than another user associated with BS i, may be associated with another BS j due to shadowing. Moreover, if we adopt load balancing, then the boundary may be load dependent. Nevertheless, low throughput users in each cell are likely to locate in near its boundary. This is the reason why we regard the 5th percentile throughput as a representative for the performance of boundary users.

1) N/A: Universal frequency reuse (α=1); The inter-cell association is determined based on the best signal strength BS. This is a conventional scheme.

2) LB: Universal frequency reuse (α=1); Our proposed cell-site selection and inter-cell handover are used for balancing load. Note that this LB case, considering only load balancing scheme without ICI mitigation scheme, can be regarded as Bu's work [6].

3) IA: PFR with (RF, α)=(1/3,6/8) for ICI mitigation; The inter-cell association is determined based on the best signal strength BS. Our proposed intra-cell handover is used for optimal band selection.

4) IA + LB: PFR with (RF, α)=(1/3,6/8) for ICI mitigation; Our proposed intra- and inter-cell handovers as well as cell-site selection are used.

Fig. 4(a) shows the average number of users in each tier. When Tier 0 is under-loaded ($\rho_0 < 40$), the LB moves edge users from Tier 1 to Tier 0. On the other hand, when Tier 0 is over-loaded ($\rho_0 > 40$), LB moves edge users from Tier 0 to Tier 1. For example, when $\rho_0 = 100$, it shifts about 20 users to the Tier 1. Thus, the average numbers of users are 80(=100-20) and 43($\simeq$ 40+20/6) for Tier 0 and Tier 1, respectively. The same is true for Fig. 4(b).

To observe the benefits of IA and LB in terms of user performance at cell edge, the 5th percentile throughput is investigated in Figs. 4(c)(d). Let us examine the LB gain first. The LB shifts edge users in the hot-spot cell to the neighbor cells. This gives two advantages to the hot-spot cell: 1) the number of users competing for the same resource is reduced; and 2) boundary users become closer to the BS than before. For these reasons, the 5th percentile throughput for the hot-spot cell increases. On the other hand, 5th percentile throughput decreases for the under-loaded cell. Thus, as illustrated in Figs. 4(c)(d), the LB reduces the gap between 5th percentile throughputs and thus improves boundary (the worst) user performance in the system by a factor of $10{\sim}40\%$. The gain realized by LB increases with the heterogeneity of user distribution.

Now let us consider IA scenario: By comparing Fig. 4(c) with (d), one can easily notice the difference of relative level. The 5th percentile throughput in the latter case, adopting load balancing, is 20% better than that in the former case. This gain depends on the value of α as we explained in the previous subsection.

Figs. 4(e)(f) show the cell throughput for each tier and total throughput which is defined as the sum of throughputs of all cells in Tier 0 and 1. For the cases of N/A and IA where load balancing is not used,

the cell throughput of Tier 0 slowly increases with the number of users in Tier 0 due to the multi-user diversity gain. When the LB is used, the cell throughputs of Tier 0 and 1 are improved and degraded for the over-loaded case. This is because LB shifts edge users from Tier 0 to 1, which upgrades and degrades the average channel quality in Tier 0 and 1, respectively. This relation is reversed for the under-loaded case. Meanwhile, the total throughput deteriorates a little as the heterogeneity increases. On the whole, however, total throughput remains almost same for all cases. In brief, *IA and LB gains improve the performance of cell boundary users while keeping the total throughput almost unchanged.*

To strengthen our claim that IA and LB help users at the cell edge, let us compute the QoS violation probability with a minimum throughput requirement, $r_k \geq m$. In Fig. 5, we plot the percentage of users whose average throughput is lower than a threshold for m=100kbps. As expected, the violation percentage decreases when IA or LB schemes are used. When both schemes are used, it decreases significantly. This means that the system can accommodate more satisfied users which boost the revenue of service provider.

VII. CONCLUSION

Next-generation broadband systems can provide the higher capacity, but users at cell edge still suffer from low throughput due to severe ICI and load imbalance. Therefore, to guarantee a QoS for boundary users and more balanced data rate among all users, PFR and load-balancing are considered in this paper. The contributions of the paper can be summarized as follows:

– We have formulated this problem as a NUM problem in multi-cell networks. Though we found the offline optimal polynomial time algorithm, it is still too complicated to be applied in practice. To overcome such overheads, we developed a practical online algorithm.

– A remarkable feature of the proposed online algorithm is that it uses a notion of *expected throughput*. Using this metric in making user binding decisions, users can take into consideration channel conditions as well as traffic loads among cells.

– Our online algorithm can not only closely achieve network-wide proportional fairness but also bring two types of performance gain: IA and LB gains. In brief, *they improve the performance of users at the cell edge while keeping the total throughput almost unchanged.*

To achieve our objective, additional inter-cell handover events may occur. If system designers are concerned about this overhead, then they may use our *load-aware metric* only for new arrivals or during

handovers due to the mobility, rather than periodically. This requires more time for the convergence, but the number of inter-cell handover events remains almost the same as before. Throughout the paper we have determined a good choice for α by simulations. As future work, we intend to develop a (possibly adaptive) approach which can be used in a practical system.

APPENDIX

A. Proof of Proposition 2 and 3: Inter/Intra-handover conditions

Let us consider the net increment of utility for the inter-cell handover when a user k moves form BS i to BS j.

$$
\triangle U = \left[\log \frac{G(y_j^b+1)E[R_{j,k}^b]}{y_j^b+1} - \log \frac{G(y_i^b)E[R_{i,k}^b]}{y_i^b} \right] \\
+ \sum_{l \in K_i - \{k\}} \left[\log \frac{G(y_i^b-1)E[R_{i,l}^b]}{y_i^b-1} - \log \frac{G(y_i^b)E[R_{i,l}^b]}{y_i^b} \right] + \sum_{l \in K_j} \left[\log \frac{G(y_j^b+1)E[R_{j,l}^b]}{y_j^b+1} - \log \frac{G(y_j^b)E[R_{j,l}^b]}{y_j^b} \right],
\tag{21}
$$

where $K_i = \{k \mid x_{i,k} = 1, k \in K\}$ and $K_j = \{k \mid x_{j,k} = 1, k \in K\}$. The first term of Eq. (21) means the utility increment for the user k by changing the serving BS. And the second and the last term mean the aggregate utility increment of BS i by losing the user k and the decrement of BS j by adding user the k, respectively. Using the $\lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x = e$ and the Euler's approximation to harmonic series $G(M) = \sum_{m=1}^{M} \frac{1}{m} \simeq \gamma + \log(M)$ where $\gamma = 0.5772 \cdots$ is the Euler-Mascheroni constant, we can obtain the following equations for $y_i^b, y_j^b \gg 1$:

$$
\left(\frac{y_i^b}{y_i^b-1} \right)^{y_i^b-1} \simeq e, \quad \left(\frac{y_j^b}{y_j^b+1} \right)^{y_j^b} \simeq \frac{1}{e}, \quad \left(\frac{G(y_i^b)}{G(y_i^b-1)} \right)^{y_i^b-1} \simeq 1, \quad \left(\frac{G(y_j^b)}{G(y_j^b+1)} \right)^{y_j^b} \simeq 1 .
\tag{22}
$$

By putting Eq. (22) into Eq. (21), the aggregate utility increment of BS i decrement of BS j become 1 and -1 regardless of the numbers of users as long as they are large, and they cancel each other. Thus, we have the inter-cell handover condition Eq. (13) (i.e., $\triangle U > 0$). Following the same procedure, we can derive the intra-cell handover condition Eq. (12) as well. ∎

B. Proof of Proposition 4: Admission Condition

Let us consider the net increment of utility when a newly user k arrives to BS i.

$$
\triangle U = \log \frac{G(y_i^b+1)E[R_{i,k}^b]}{y_i^b+1} + \sum_{k \in K_i} \left[\log \frac{G(y_i^b+1)E[R_{i,k}^b]}{y_i^b+1} - \log \frac{G(y_i^b)E[R_{i,k}^b]}{y_i^b} \right] .
\tag{23}
$$

By putting Eq. (22) into Eq. (23), we have the admission condition Eq. (14) (i.e., $\triangle U > 0$). ∎

REFERENCES

[1] Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, IEEE Std. 802.16e, 2005.

[2] 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Feasibility study for evolved UTRA and UTRAN (Release 7), 3GPP Std. Tech. Rep. 25.912, Jun. 2006.

[3] S. V. Hanly, "An algorithm for combined cell-site selectino and power control to maximize cellular spread spectrum capacity," *IIEEE Journal on Selected Areas in Communications*, vol. 13, no. 7, pp. 1332-1340, Sep. 1995.

[4] S. Das, H. Viswanathan and G. Rittenhouse, "Dynamic load balancing through coordinated scheduling in packet data systems," *in Proc. IEEE Infocom 2003*, Mar. 2003.

[5] A. Sang, X. Wang, M. Madihian and R. Gitlin, "Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems," *in Proc. ACM Mobicom 2004*, Sep. 2004.

[6] T. Bu, L. Li and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," *in Proc. IEEE Infocom 2006*, Apr. 2006.

[7] V. H. MacDonald, "The cellular concept," *Bell System Technical Journal*, vol. 58, no. 1, pp. 15-41, Jan. 1979.

[8] G. Li and H. Liu, "Downlink radio resource allocation for multi-cell OFDMA system," *IEEE Trans. Wireless Communications*, vol. 5, no. 12, pp. 3451-3459, Dec. 2006.

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

[10] J. Cho, J. Mo and S. Chong, "Joint network-wide opportunistic scheduling and power control in multi-cell networks," *in Proc. IEEE WoWMoM 2007*, Jun. 2007.

[11] S. Liu and J. Virtamo, "Inter-Cell Coordination with Inhomogeneous Traffic Distribution," *in Proc. NGI 2006*, pp. 64-71, Apr. 2006.

[12] F. Kelly, A. Maullo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237-252, Jul. 1998.

[13] 3GPP TR 25.814 v.1.2.2, "Physical layer aspects for evolved UTRA," Mar. 2006.

[14] Z. Zhang, Y. He, and E. K. P. Chong, "Opportunistic downlink scheduling for multiuser ofdm systems," *in Proc. IEEE Wireless Communications and Networking Conference*, Mar. 2005.

[15] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser ofdm system," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 171-178, Feb. 2003.

[16] P. Viswanath, D. N. C. Tse, and V. Anantharam, "Asymptotically optimal water-filling in vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 47, no. 1, pp. 241.267, Jan. 2001.

[17] E. Biglieri, J. Proakis and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. on Information Theory*, vol. 44, pp. 2619-2692, Oct. 1998.

[18] A. L. Stolyar, "On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation," *Operations Research*, vol. 53, no. 1, pp. 12.25, Jan. 2005.

[19] C. Papadimitriou and K. Steiglitz, "Combinatorial optimization: algorithms and complexity," *Prentice-Hall*, Englewood Cliffs, NJ, 1982.

[20] H. J. Kushner and P. A. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Trans. Wireless Communications*, vol. 3, no. 4, pp. 1250-1259, Jul. 2004.

[21] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," *in Proc. Infocom 2003*, Mar. 2003.

[22] G. P. Pollini, "Trends in handover design," *IEEE Commun. Mag.*, vol. 34, pp. 82-90, Mar. 1996.

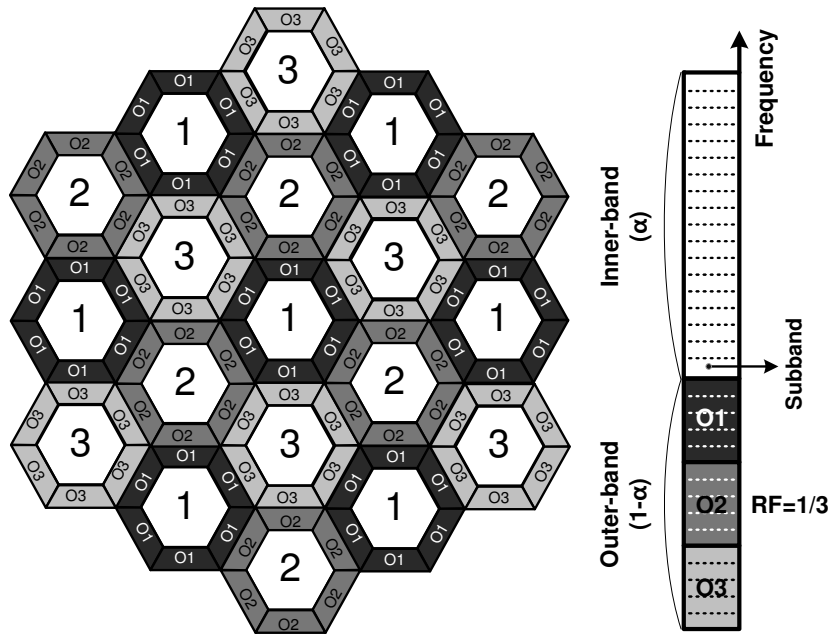[23] W. C. Jakes, *Microwave mobile communication*, Wiley, 1974.

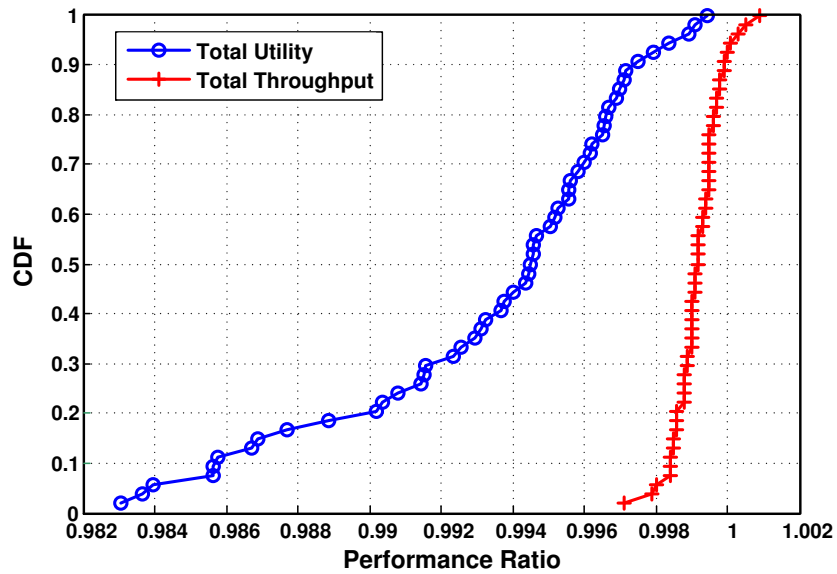Fig. 1. Examples of frequency partitioning: $RF = 1/3$, the first-tier ICI mitigation.

Fig. 2. Performance ratios between the optimal and the online algorithm: total utility and total throughput.
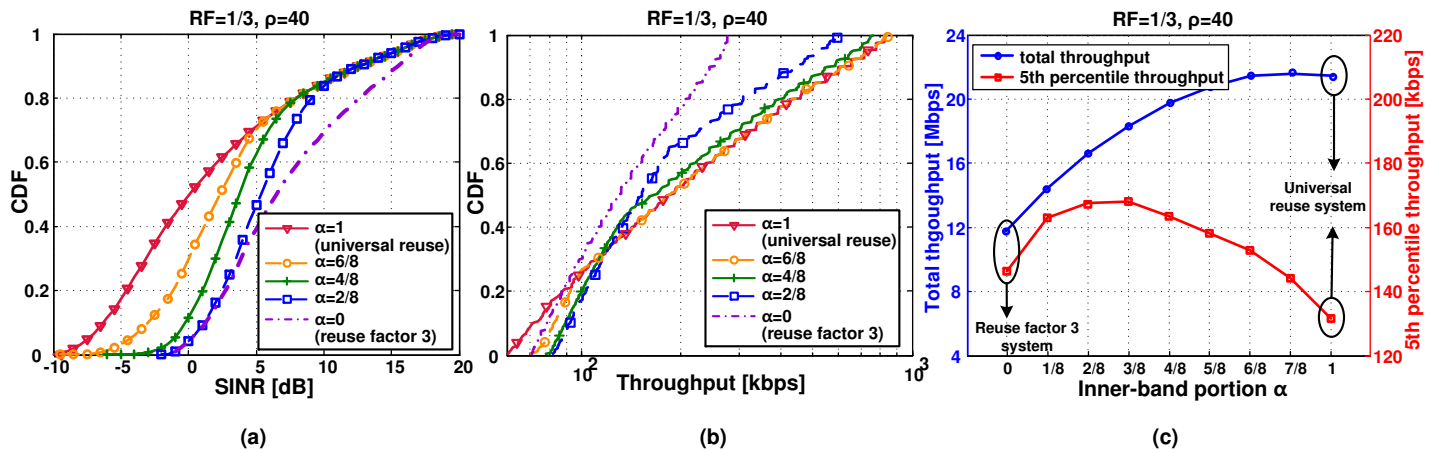
Fig. 3.   Impact of the inner band portion α on system performances: (a) CDF of SINR, (b) CDF of throughput (c) 5th percentile and total throughput.
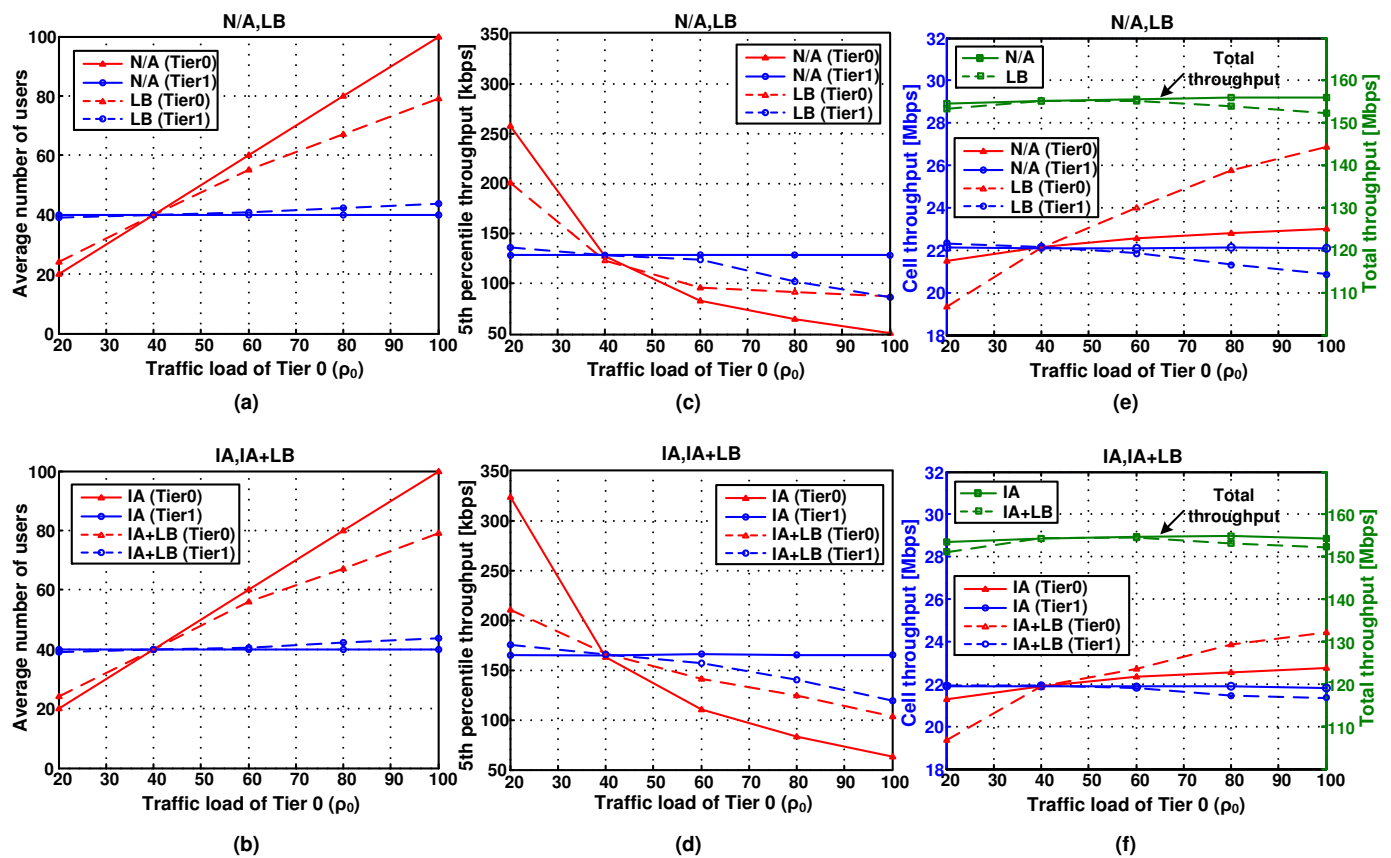

Fig. 4.   System performances with different four schemes (N/A, LB, IA, IA+LB): (a),(b) number of users, (c),(d) 5th percentile throughput and (e),(f) cell and total throughput.
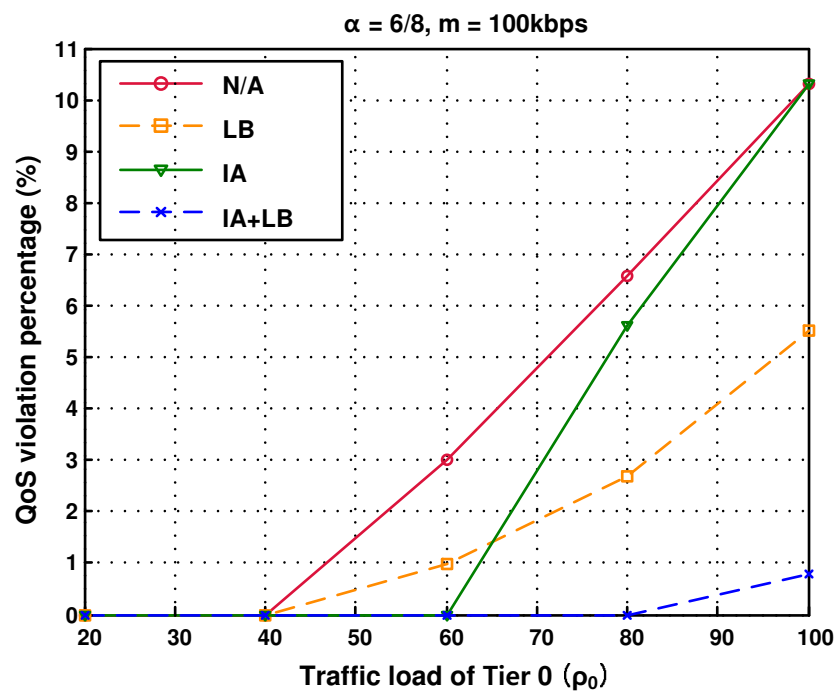
Fig. 5.   QoS violation percentage with different four schemes (N/A, LB, IA, IA+LB).