Resource Allocation in Autonomous Cellular Networks

Bijan Golkar, *Student Member, IEEE* and Elvino S. Sousa, *Fellow, IEEE*

***Abstract*—In a traditional cellular network the base stations (BSs) are deployed regularly according to a pre-determined pattern modeled by hexagons. The deployment of a larger number of BSs improves the network performance. However, a dense regular deployment of BSs is prohibitively expensive and in most cases not feasible. Alternatively the network can grow in an organic fashion by the deployment of BSs according to the traffic demand. A high traffic demand in a given locality is reflected in a higher number of BSs deployed in the area. In this paper, we propose a practical framework for the resource allocation of cellular networks with an irregular BS deployment pattern. To this end, a network clustering technique is proposed which forms clusters of coordinating BSs. The coordinated resource allocation among the BSs within each cluster is devised to achieve proportional fairness. The performance of the proposed framework is evaluated with regular as well as irregular deployment of BSs. The results are compared against standard resource allocation techniques and show promising results.**

***Index Terms*—Autonomous infrastructure wireless network, coordinated resource allocation, self-adaptation, base station clustering.**

I. Introduction

A CELLULAR network is formed by the deployment of a set of base stations (BSs) across the network coverage area. In a typical cellular network, BSs are (more or less) uniformly deployed based on a hexagonal pattern [1]. In a typical urban cellular network, each BS covers an area with a radius in the range of $100 - 1000$ meters. In order to satisfy the ever-growing demand for higher transmission rates, new network infrastructures should be considered.

A network with the same regular deployment pattern, however with a much smaller cell radius, can potentially satisfy the future requirements. Scaling down the size of each cell does not substantially affect the signal-to-interference-plus-noise ratio (SINR) of each link. However, with the reduction of the cell coverage area, the number of served terminals per BS drops significantly. Each BS allocates more system resources to each terminal, which in turn results in a larger aggregate throughput for the end-user. However, a significant reduction in the cell size translates into a highly dense regular deployment of BSs. The deployment of such an infrastructure will be prohibitively expensive. More importantly, such a dense deployment will not be feasible due to restrictions on the locations of the base stations.

Manuscript received October 30, 2012; revised June 20, 2013; accepted August 13, 2013. The associate editor coordinating the review of this paper and approving it for publication was A. Nosratinia.

The authors are with the Deparment of Electrical and Computer Engineering, University of Toronto 10 King's College Road, Toronto, Ontario, Canada, M5S 3G4 (e-mail: {bijan, sousa}@comm.utoronto.ca).

Digital Object Identifier 10.1109/TWC.2013.121686

Alternatively, the network can be constructed by a BS deployment based on local traffic demand. The network will no longer have a regular deployment pattern. In an autonomous cellular network, the role of adaptation in efficient resource allocation becomes increasingly important. Not only does it improve the performance but more importantly allows the network to scale in an organic fashion. In this paper, we propose a practical adaptive resource allocation framework for these networks. Cellular networks with regular and irregular BS deployments are referred to as *traditional* and *autonomous cellular networks* respectively.

A. Related work

Bonald *et al.* introduce coordination among cells as a new notion in scheduling of cellular networks [2]. For the first time, resource allocation is performed in two phases, i.e. two time scales. The first phase is in charge of coordinating the activity of interfering BSs in order to maximize transmission data rates. The second phase is responsible for load balancing in order to divert traffic from heavily-loaded cells to lightly-loaded cells. Li and Liu extend the two-level resource allocation framework to a multi-cell OFDMA system [3]. A radio network controller assigns the available spectrum to BSs in a centralized fashion during the first phase. In the second phase each BS independently assigns the available frequency channels to the terminals based on the actual traffic conditions of the active terminals (e.g. terminals buffer sizes). This paper offers the first — although computationally intensive — two-level resource allocation framework for OFDMA cellular networks, based on which several other solutions have been proposed (e.g. [4]–[9]).

Most of the techniques proposed over the past decade assume full knowledge of the channel state information (CSI). Although this simplifies the problem tremendously, it jeopardizes the practicality of the proposed technique. The authors of [6] emphasize on this aspect. They propose a two-level framework based on graph theory. In the first level, inter-cell interference is reduced with no precise knowledge of out-of-cell interference. The second level allocates the system resources according to instantaneous channel gains.

Another important aspect in practicality is the degree of centrality in the proposed technique. Centralized techniques assume full coordination among all BSs and hence achieve better performance. This, however, is at the expense of unrealistic level of inter-BS communication. In [10], a distributed resource allocation method is proposed. An interference metric is defined based on the interference among BSs. Accordingly for each BS i, two sets of in-neighbor and out-neighbor BSs

are introduced. An in-neighbor BS is defined as a BS which creates interference on the terminals served by BS i and an out-neighbor BS is one whose terminals are affected by simultaneous transmission from BS i. The proposed technique is essentially a coordinated power control algorithm, runs once every TTI and has multiple iterations. At each iteration, inter-BS communication is among a limited number of BSs. This reduces over-head signaling significantly. In particular, at each BS individual CSI from only the in-neighbor set is required. Nevertheless, it is assumed that each BS knows fast fading aggregate interference levels due to simultaneous transmission of all BSs at each of its serving terminals. Furthermore, although the proposed technique is distributed, the decision on terminating the iterative algorithm should be taken at a central entity (alternatively the algorithm can be performed a pre-determined number of times). Moreover, the results provided in the paper are based on re-forming the neighbor sets every TTI. This in turn potentially increases the over-head signaling and complexity by a significant amount.

B. Contributions

A scheduling cell is defined as a set of BSs whose resource allocation is coordinated. The cells adapt to actual traffic distribution across the network. In regions where high interference among BSs call for more coordination (e.g. hot-spots), a larger cluster of BSs form a scheduling cell. Likewise regions which serve sparse population of terminals and/or the terrain obstructs interference among BSs, smaller scheduling cells are created. A coordinated resource allocation method tailored to the irregular structure of the adaptive scheduling cells is proposed. The corresponding resource allocation optimization problem was formulated in [11] and a solution was proposed. This solution is modified according to our assumptions, requirements and objectives. It is important to note that the resource allocation across the scheduling cells is distributed. In other words, the resource allocation in each scheduling cell is performed with no knowledge from the rest of the network.

The paper is organized as follows. Section II introduces the system model. In Section III the network clustering problem is defined and the network clustering algorithm is developed. Section IV discusses the resource allocation framework, presents the coordinated resource allocation method and elaborates on the different time scales of adaptation. In Section V the performance evaluation results are discussed. Section VI concludes the paper and provides possible future research directions.

Notation: Bold upper case letters ($\mathbf{X}$) denote matrices, lower case letters ($\mathbf{x}$) are adopted for column vectors and the Latex Mathcal font ($\mathcal{X}$) is used for sets. $\mathbf{x}^T$, $|\mathcal{X}|$, $\bar{\mathcal{X}}$ denote the transpose of $\mathbf{x}$, the cardinality of $\mathcal{X}$ and the complement of $\mathcal{X}$ respectively. The set of real numbers is denoted with $\mathbb{R}$, the empty set with $\emptyset$ and an $M \times N$ matrix of zeros with $\mathbf{0}_{M\times N}$.

II. System model

In order to provide service to K wireless terminals, A BSs are deployed. Each BS i has a backbone connection to the wired network and a transmission power budget P_i[1]. Each terminal l is assigned to and communicates with BS $a(l)$ with the strongest large-scale fading channel gain. Thus BS i serves the set of terminals $\mathcal{K}_i = \{l|a(l) = i\}$ such that $\sum_{i=1}^{A} |\mathcal{K}_i| = K$. Wireless communication between BSs and terminals is established on the available frequency spectrum, a set of N (frequency) resources.

Unless otherwise stated, it is assumed that the BS power budget P_i is uniformly distributed across the available sub-carriers. Assuming each frequency resource has M sub-carriers, the power budget for transmission on a sub-carrier at BS i is

$$p_i = \frac{P_i}{MN}. \tag{1}$$

Let us consider a network with one BS and K terminals and assume $\mathbf{I}(t) \in \{0,1\}^{K\times N}$ to be the frequency channel allocation matrix for time slot t. $I_{ln}(t)$ is equal to one if frequency channel n is assigned to terminal l in time slot t and $I_{ln}(t) = 0$ otherwise. The data rate of terminal l in time slot t is

$$r_l(t) = \sum_{n=1}^{N} I_{ln}(t) r_{ln}(t), \tag{2}$$

where $r_{ln}(t)$ is the supported data rate on frequency channel n by terminal l in time slot t.

Let $U_l(t)$ be the utility function of terminal l in time slot t. The resource allocation problem at the beginning of time slot t can be formulated as follows [12]:

$$\max_{\mathbf{I}(t)} \sum_{l=1}^{K} U_l\left(r_l(t)\right) \tag{3}$$
$$\text{subject to} \begin{cases} \sum_{l=1}^{K} I_{ln}(t) = 1 & \forall n \\ I_{ln}(t) \in \{0,1\} & \forall n, l \end{cases},$$

where n and l are the indices for frequency channels and terminals respectively. The first constraint in (3) guarantees that each resource is allocated to exactly one transmission per time slot. The second constraint dismisses the option of time-sharing during a time slot. In a cellular network with A BSs, the scheduling problem can be generalized as follows:

$$\max_{\{\mathbf{I}_1(t),\cdots,\mathbf{I}_A(t)\}} \sum_{l=1}^{K} U_l\left(r_l(t)\right) \tag{4}$$
$$\text{subject to} \begin{cases} \sum_{l\in\mathcal{K}_i} I_{ln}(t) \le 1 & \forall n, i \\ I_{ln}(t) \in \{0,1\} & \forall n, l \end{cases},$$

where $\mathbf{I}_i(t)$ is the frequency channel allocation matrix of BS i for time slot t.

III. Network clustering

The optimal solution to (4) requires full coordination among all BSs. This degree of coordination in a cellular network, however, is not practical. Alternatively, the resource allocation

[1]In this paper, it is assumed that all BSs have the same power budget. The results, however, are applicable to the more general case where the BSs have distinct power budgets.

problem can be broken down into smaller sub-problems and solved locally. Instead of coordinating the resource allocation among all BSs, coordination can be among a subset (cluster) of BSs. At each time slot, the coordinating BSs adjust their transmission power levels to achieve the best performance. Inevitably BSs outside each cluster induce interference on all links in the cluster. Hence larger clusters of BSs would result in a better performance due to a lower number of potentially interfering BSs. The improvement in performance comes at the price of higher complexity for coordination among a larger number of BSs. This Section develops a framework to form clusters of coordinating BSs based on the BS locations and the traffic distribution.

Let us define the network topology by the interference matrix $\mathbf{H} = [h_{li}] \in \mathbb{R}^{K \times A}$, where h_{li} denotes the channel gain between terminal l and BS i. High channel gains between terminals and BSs result in dense topologies and call for coordination in resource allocation. On the other hand, a sparse topology is formed as a result of low channel gains between terminals and interfering BSs. Naturally, the size of a scheduling cell[2] should depend on the network topology. Although a large scheduling cell implies high coordination and hence a better system performance, it in turn increases the implementation cost and complexity. On the other hand, too small a scheduling cell results in excessive interference levels among scheduling cells.

DEFINITION 4.1 (Network clustering problem) Let $\mathcal{A}$ be the set of BSs in the network. A partition of $\mathcal{A}$ is defined as the set of subsets $\mathcal{A}(1), \mathcal{A}(2), \cdots, \mathcal{A}(C)$ such that

$$\begin{cases} \bigcup_{c=1}^{C} \mathcal{A}(c) = \mathcal{A} \\ \mathcal{A}(i) \cap \mathcal{A}(j) = \emptyset, \forall i \neq j \end{cases} . \tag{5}$$

Clustering is the process of finding such a partition subject to some criteria. The clustering problem imposes boundaries among scheduling cells. Each boundary delineates the lack of coordination among certain BSs. This in turn means, unlike traditional frequency allocation schemes, transmission to terminals at the boundary between two adjacent scheduling cells will no longer be perfectly orthogonal. The level of interaction among scheduling cells heavily depends on the network topology. An intelligent choice of boundaries among scheduling cells reduces the inter-cell interference to acceptable levels. More importantly in a clustered network, practical coordinated resource allocation methods in autonomous cellular will be possible.

Let us model the cellular network as a fully connected weighted graph $(\mathcal{V}, \mathcal{E})$. The set of vertices $\mathcal{V}$ represents the set of BSs $\mathcal{A}$. Each edge connects two BSs in the network. The weight on the edge is defined based on the notion of interference among BSs. This notion is introduced in Section III-A which gives rise to the definition of the *similarity index*. In Section III-B, a clustering algorithm is proposed which partitions the BSs into scheduling cells.

[2]The size of a scheduling cell is defined as the number of BSs in the scheduling cell.

A. Similarity index

Resource allocation in a cellular network is performed in a time-slotted fashion. At each time slot, each BS serves a number of terminals on the available frequency channels. The SINR of a terminal on a given frequency channel can be written as

$$\text{SINR} = \frac{P_D}{P_I + \eta_0 W}, \tag{6}$$

where P_D is the received signal from the desired BS at the desired terminal. P_I denotes the interference at the desired terminal and $\eta_0 W$ is the background noise power. The interference is caused by simultaneous transmission (i.e. at the same time slot and on the same frequency channel) from all or a subset of BSs in the network.

In general, the quality of a link improves with higher SINR. Let us define the *performance level of a terminal* by the corresponding SINR level. Each BS serves a set of terminals with a unique set of performance levels. Simultaneous transmission from BS j has an impact on the performance levels of the terminals assigned to BS i. If the impact is significant, coordination in resource allocation between BS i and BS j is required.

The *performance level of a BS* is defined by an increasing function of the performance levels of the terminals assigned to it. The performance level of BS i can be defined as the arithmetic average of the performance levels of its terminals. This definition has one major problem. The SINR levels of the terminals assigned to a BS can take a wide range of values. In some scenarios, due to exceptionally good channel gains (or unrealistic simulation models or flaws in channel estimations), a few terminals experience unexpectedly high SINR levels. The resource allocation decisions, however, should not be influenced by or be based on the channel gains of these terminals. Albeit, the arithmetic average SINR level is skewed toward these high values. The median SINR level, on the other hand, is defined as the midpoint where half of the SINR levels are above and half of the SINR levels are below this value. Essentially, by defining the performance level of a BS as the median value, the effect of unrealistically high SINR levels is mitigated.

Based on the interference coming from BS j, the SINR at terminal l assigned to BS i can be written as

$$\text{SINR}_{lij} = \frac{h_{li} p_i}{h_{lj} p_j + \eta_0 W}, \tag{7}$$

where h_{li} denotes the channel gain between terminal l and BS i, p_i is the transmit power of BS i. With no interference from BS j, the signal-to-noise ratio (SNR) for terminal l is

$$\text{SNR}_{li} = \frac{h_{li} p_i}{\eta_0 W}. \tag{8}$$

Let us form the SINR vector $\mathbf{x}_{ij}$ by the SINR levels of all terminals assigned to BS i with interference from BS j as calculated in (7). Similarly, let $\mathbf{y}_i$ be the vector of SNR_{li} levels of the terminals assigned to BS i as calculated in (8). The loss in performance due to the effect of interference from BS j on BS i is quantified by the following index:

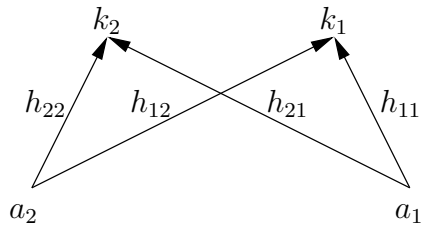


Fig. 1. A cellular network with two BSs and two terminals

$$c_{ij} = \frac{f(\mathbf{y}_i)}{f(\mathbf{x}_{ij})}, \tag{9}$$

where $f(\mathbf{y}_i)$ is the performance level of BS i without interference from BS j. Similarly, $f(\mathbf{x}_{ij})$ is the performance level of BS i with interference from BS j. Based on the discussions earlier in this Section, the performance level of each BS is defined as the median of the performance levels of all terminals assigned to the BS.

By exchanging i and j in (7), (8) and (9) c_{ji} is defined. A larger c_{ij} and/or c_{ji} implies higher levels of interference between BSs i and j, and in turn translates into a higher urgency for coordination between the two BSs. The similarity index between BS i and BS j is defined as

$$s_{ij} = \frac{1}{2}\left(c_{ij} + c_{ji}\right). \tag{10}$$

As an example let us consider the cellular network in Figure 1. BS 1 is serving terminal 1 and BS 2 is serving terminal 2. Based on (9) we have

$$\begin{cases} c_{12} = \frac{\text{SNR}_{11}}{\text{SINR}_{112}} = \frac{h_{12}p_2}{\eta_0 W} + 1 \\ c_{21} = \frac{\text{SNR}_{22}}{\text{SINR}_{221}} = \frac{h_{21}p_1}{\eta_0 W} + 1 \end{cases}. \tag{11}$$

Depending on the values of the channel gains h_{12} and h_{21}, the two indices can be very different. Hence the similarity index between BS 1 and BS 2 is defined as

$$\begin{aligned} s_{12} &= s_{21} \\ &= \frac{1}{2}\left(c_{12} + c_{21}\right) = \frac{h_{12}p_2 + h_{21}p_1}{2\eta_0 W} + 1. \end{aligned} \tag{12}$$

It should also be noted that the similarity index is defined as a function of large-scale fading channel gains (i.e. path-loss and shadow fading). As a result it can be easily applied to multiple antenna systems, since the MIMO channel structure usually affects only the small-scale fading gains.

The network similarity matrix $\mathbf{S} = [s_{ij}] \in \mathbb{R}^{A\times A}$ is formed by the similarity indices of all pairs of BSs in the network. This matrix quantifies the level of interference between any pair of BSs in the network. In order to form this matrix, the channel gains between all terminals and BSs are required. This information is readily available from the BS assignment stage, where the channel gains of each terminal to the BSs in its vicinity are compared and the terminal is assigned to the BS with the strongest channel gain. The estimated channel gains are reported from each terminal to the corresponding BSs. The reported channel gains are then communicated from the BSs to a central control unit where the similarity matrix is formed. In Section III-B, this matrix is employed to develop the network clustering algorithm.

B. Clustering

Parallel computing efficiently distributes computation loads across multiple processors. Firstly, the computations should be assigned to the processors in such a way that there exists little or no inter-processor communication. Secondly, the computation load should be balanced across the processors. The load distribution can be stated as the problem of finding the computation load assignment which minimizes the inter-processor communication, while keeping the computation loads assigned to the processors as equal as possible.

The cellular network clustering problem is analogous to the load distribution problem in parallel computing. Scheduling cells take the place of processors and inter-cell interference takes the place of inter-processor communication. Instead of minimizing the inter-processor communication, the interference levels among the scheduling cells are to be minimized. A balanced distribution of the computation load is replaced by a balanced distribution of the resource allocation load across the scheduling cells, which in other words, equalizes the resource allocation complexity across the scheduling cells.

The load distribution problem in parallel computing has been extensively studied in the context of graph partitioning. There are several ways to pose partitioning a graph as a mathematical problem. The min-cut technique is a straightforward method which attempts to minimize the inter-dependencies among clusters. Let $(\mathcal{V}, \mathcal{E})$ be a fully connected weighted graph with N nodes $\{v_i\}_{i=1}^{N}$ and weight matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$. The connectivity between two sets $\mathcal{E}, \mathcal{F} \subset \mathcal{V}$ is defined as

$$c(\mathcal{E}, \mathcal{F}) = \sum_{v_e\in\mathcal{E}, v_f\in\mathcal{F}} w_{ef}, \tag{13}$$

where w_{ef} is the element on row e and column f of the weight matrix $\mathbf{W}$. The min-cut technique generates a Q-partition of the graph $\{\mathcal{V}_1, \cdots, \mathcal{V}_Q\}$ by minimizing the connectivity of each cluster $\mathcal{V}_q$ from the rest of the network $\bar{\mathcal{V}}_q$. Mathematically this problem is expressed as

$$\min_{\{\mathcal{V}_1,\cdots,\mathcal{V}_Q\}} \sum_{q=1}^{Q} c(\mathcal{V}_q, \bar{\mathcal{V}}_q). \tag{14}$$

The min-cut solution often tends to separate individual vertices from the graph which disturbs the desirable balance in the size of the clusters. In order to avoid the generation of too small clusters, the problem is modified as

$$\min_{\{\mathcal{V}_1,\cdots,\mathcal{V}_Q\}} \sum_{q=1}^{Q} \frac{c(\mathcal{V}_q, \bar{\mathcal{V}}_q)}{|\mathcal{V}_q|}. \tag{15}$$

By normalizing the connectivity $c(\mathcal{V}_q, \bar{\mathcal{V}}_q)$ by the cardinality of the corresponding set $|\mathcal{V}_q|$, we are essentially enforcing the formation of reasonably large clusters.

The solution of problem (15) has been shown to be NP hard. In what follows, we elaborate on how the problem can be relaxed into an easy-to-solve standard linear algebra problem.

This technique is known as the spectral clustering technique [13].

1) Spectral clustering: Let $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_Q] \in \mathbb{R}^{N \times Q}$ be the indicator matrix such that

$$u_{ij} = \begin{cases} \frac{1}{\sqrt{|\mathcal{V}_j|}} & , \quad v_i \in \mathcal{V}_j \\ 0 & , \quad v_i \notin \mathcal{V}_j \end{cases} \tag{16}$$

and $\mathbf{L}$ be the Laplacian of the weight matrix $\mathbf{W}$ as follows:

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \tag{17}$$

where $\mathbf{D} = [d_{ij}] \in \mathbb{R}^{N \times N}$ is a diagonal matrix with $d_{ii} = \sum_{j=1}^{N} w_{ij}$. It is easy to show that

$$\mathbf{u}_q^T \mathbf{L} \mathbf{u}_q = \frac{c(\mathcal{V}_q, \bar{\mathcal{V}}_q)}{|\mathcal{V}_q|}. \tag{18}$$

Hence,

$$\sum_{q=1}^{Q} \frac{c(\mathcal{V}_q, \bar{\mathcal{V}}_q)}{|\mathcal{V}_q|} = \sum_{q=1}^{Q} \mathbf{u}_q^T \mathbf{L} \mathbf{u}_q = \text{trace}(\mathbf{U}^T \mathbf{L} \mathbf{U}) \tag{19}$$

and the partitioning problem can be formally written as

$$\min_{\{\mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_Q\}} \text{trace}(\mathbf{U}^T \mathbf{L} \mathbf{U}), \tag{20}$$

where $\mathbf{U}$ is defined in (16).

By relaxing the constraint on $\mathbf{U}$, the optimization problem can be written as

$$\begin{aligned} &\min_{\{\mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_Q\}} \text{trace}(\mathbf{U}^T \mathbf{L} \mathbf{U}) \\ &\text{subject to} \quad \mathbf{U} \in \mathbb{R}^{N \times Q}. \end{aligned} \tag{21}$$

From the Rayleigh-Ritz theorem, it is well-known that the optimum $\mathbf{U}$ is formed by the first Q eigen-vectors[3] of $\mathbf{L}$ as its columns. Hence, $\mathbf{U}$ can be written as $[\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_Q]$ where $\mathbf{u}_1$ is the eigen-vector corresponding to the smallest eigen-value. For any eigen-vector $\mathbf{u}_q = [u_{1q}, u_{2q}, \cdots, u_{Nq}]^T$, we have

$$\mathbf{u}_q^T \mathbf{L} \mathbf{u}_q = \sum_{n=1}^{N} \sum_{m=1}^{N} (u_{nq} - u_{mq})^2 w_{nm}. \tag{22}$$

Without loss of generality the eigen-vectors are assumed to be normalized, i.e. $\mathbf{u}_q^T \mathbf{u}_q = 1$. Hence,

$$\lambda_q = \sum_{n=1}^{N} \sum_{m=1}^{N} (u_{nq} - u_{mq})^2 w_{nm}. \tag{23}$$

Let us consider two vertices n and m with a high w_{nm}. For a small eigen-value λ_q, the term $(u_{nq} - u_{mq})^2$ should be small. Since $\mathbf{U}$ is constructed from the eigen-vectors with the smallest Q eigen-values of $\mathbf{L}$, the term $(u_{nq} - u_{mq})^2$ should be small for all $q = 1, \cdots, Q$. This in turn translates into two geometrically close points in a Q-dimensional vector space. Following this argument, each vertex of the graph is represented by a point in the Q-dimensional vector space with coordinates $\mathbf{g}_n \in \mathbb{R}^Q$. In order to form a Q-partition $\{\mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_Q\}$, the points $\{\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_N\}$ in the Q-dimensional vector space are partitioned based on the K-clustering technique. The pseudo-code for the spectral clustering is provided in Algorithm 1. The choice of K-clustering for the last stage of the spectral clustering method is purely due to the simplicity of this technique as will be discussed in the following.

[3]Corresponding to the Q smallest eigen-values.

Algorithm 1 Spectral clustering

Require: Parameter Q.

1: Compute the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ is a diagonal degree matrix with $d_{ii} = \sum_{j=1}^{N} w_{ij}$.
2: Compute the first Q eigen-vectors $\mathbf{u}_1, \cdots, \mathbf{u}_Q$ of $\mathbf{L}$.
3: Let $\mathbf{U} \in \mathbb{R}^{N \times Q}$ be the matrix containing the vectors $\mathbf{u}_1, \cdots, \mathbf{u}_Q$ as columns.
4: Let $\mathbf{g}_n \in \mathbb{R}^Q$ be the vector corresponding to the row n of matrix $\mathbf{U}$.
5: Cluster the points $\mathbf{g}_1, \cdots, \mathbf{g}_N$ into $\mathcal{V}_1, \cdots, \mathcal{V}_Q$ clusters with the K-clustering algorithm.
6: Form clusters: $\mathcal{V}_q = \{n | g_n \in \mathcal{V}_q\}$.

Let us consider a graph of N vertices with the coordinate set $\{\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_N\}$. Given a Q-partition, each cluster $\mathcal{V}_q$ can be represented by a cluster-head $\mathbf{m}_q$ defined as

$$\mathbf{m}_q = \frac{1}{|\mathcal{V}_q|} \sum_{v \in \mathcal{V}_q} \mathbf{g}_v. \tag{24}$$

It should be noted that the cluster-head can be any point in the vector space and does not necessarily have to be chosen from the vertices.

Similarly given a set of cluster-heads $\{\mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_Q\}$, the corresponding Q-partition can be formed by assigning each vertex of the graph to the cluster-head with the smallest Euclidean distance.

The K-clustering problem is defined as finding the set of Q cluster-heads which minimizes the squared distance from any vertex to its cluster-head. The K-clustering technique is an iterative algorithm to solve this problem: In the initialization step, the vertices are randomly clustered into Q clusters $\left\{\mathcal{V}_1^{(0)}, \mathcal{V}_2^{(0)}, \cdots, \mathcal{V}_Q^{(0)}\right\}$. The corresponding cluster-head set $\left\{\mathbf{m}_1^{(0)}, \mathbf{m}_2^{(0)}, \cdots, \mathbf{m}_Q^{(0)}\right\}$ is formed according to (24). Given the set of cluster-heads, vertex v with coordinates $\mathbf{g}_v$ is reassigned to cluster q^* such that

$$q^* = \arg \min_{1 \leq q \leq Q} ||\mathbf{g}_v - \mathbf{m}_q||^2. \tag{25}$$

Hence, a new set of Q clusters $\left\{\mathcal{V}_1^{(1)}, \mathcal{V}_2^{(1)}, \cdots, \mathcal{V}_Q^{(1)}\right\}$ is formed. In the subsequent iterations the set of cluster-heads is updated and the vertices are correspondingly reassigned. The procedure is repeated for W iterations. The pseudo-code of this technique is provided in Algorithm 2. It should be noted that

- the solution of the K-clustering algorithm is a partition of the graph with not exactly but at most Q clusters. In some instances, one or more cluster-heads may not have any

vertex associated to them. In this case, a smaller number of clusters are formed.

- the K-clustering technique is very sensitive to the initialization step. In other words, with a different initial random partition, the technique may result in a very different solution.

Algorithm 2 K-clustering

Require: Parameter Q, W.

1: Randomly cluster $\{\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_N\}$ into Q clusters $\left\{\mathcal{V}_1^{(0)}, \mathcal{V}_2^{(0)}, \cdots, \mathcal{V}_Q^{(0)}\right\}$.
2: **for** $w = 1$ to W **do**
3: For each cluster $\mathcal{V}_q^{(w-1)}$ find the coordinates of the cluster-head $\mathbf{m}_q^{(w)}$ based on (24).
4: Reassign node $\mathbf{g}_v$ to cluster q^* based on (25), and form the Q-partition $\left\{\mathcal{V}_1^{(w)}, \mathcal{V}_2^{(w)}, \cdots, \mathcal{V}_Q^{(w)}\right\}$.
5: **end for**

2) Network clustering algorithm: In traditional cellular networks, pre-determined static frequency re-use clusters are employed to decrease the level of interference among BSs. In the proposed network clustering approach, the key idea is essentially the same. Instead of the static frequency re-use clusters, it is the role of the network clustering algorithm to adaptively form scheduling cells such that the interference among clusters are reduced to acceptable levels. By developing the network clustering algorithm, a new level of adaptation is introduced into the resource allocation framework of cellular networks.

In the network graph, each vertex represents a BS. The vertices are all connected and weighted according to the similarity matrix introduced in Section III-A. Thus a fully connected and weighted graph is constructed. The network clustering algorithm forms a partition of the graph which translates into the formation of scheduling cells. It should be noted that the scheduling cells are not completely isolated from each other. In other words, interference among the scheduling cells still exits. Nevertheless, as will be seen from the performance evaluation results, an appropriate level of coordination among the BSs, reduces the level of interference among the scheduling cells to acceptable levels such that independent coordination in the scheduling cells becomes possible.

The proposed network clustering algorithm is based on the spectral clustering technique. By applying the spectral clustering method on the network graph, a partition $\mathcal{P} = \{\mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_R\}$ is formed with $R \leq Q$. Due to the intrinsic sensitivity of the K-clustering technique to initialization, the spectral clustering will be performed for L randomly chosen independent initial partitions. Thus L independent partitions of the network $\left\{\mathcal{P}^{(1)}, \mathcal{P}^{(2)}, \cdots, \mathcal{P}^{(L)}\right\}$ are formed. An association matrix $\mathbf{D} = [d_{ij}] \in [0,1]^{A \times A}$ is defined based on the L partitions as follows:

$$d_{ij} = \frac{\text{\# of times BS } i \text{ and } j \text{ are in the same cluster}}{L}. \quad (26)$$

BSs i and j with an association factor d_{ij} larger than threshold T join the same scheduling cell. The network clustering strategy is provided in Algorithm 3.

Algorithm 3 Forming the scheduling cells

Require: Parameters Q and T.

1: Initialize the association matrix: $\mathbf{D} \leftarrow \mathbf{0}_{A \times A}$.
2: **for** $l = 1$ to L **do**
3: Cluster the BSs based on Algorithm 1 with $\mathbf{W} = \mathbf{S}$.
4: **if** i and j are members of the same cluster **then**
5: $d_{ij} \leftarrow d_{ij} + \frac{1}{L}$.
6: **end if**
7: **end for**
8: **if** $d_{ij} > T$ **then**
9: **if** BS i or j is already part of a SC **then**
10: Join the other node to that SC.
11: **else**
12: Form a new SC with BSs i and j as members.
13: **end if**
14: **end if**
15: BSs which are not members of any SC form stand-alone SCs.

The proposed clustering algorithm has two parameters Q and T. Parameter Q defines the maximum number of clusters in the K-clustering problem and hence limits the maximum number of clusters in each iteration l. A higher Q results in a lower number of BSs per scheduling cell which translates into less coordination in resource allocation. Parameter T adjusts the tendency of BSs to join one cluster. A larger T requires two BSs to have a larger association factor to join the same scheduling cell and as a result it also directly impacts the size of the scheduling cells. With a T of unity, no two BSs will join the same scheduling cell as $d_{ij} < 1$ for any $i \neq j$. On the other hand, a T of zero translates into full coordination among all BSs. This is due to the fact that $d_{ij} > 0$ for any $i \neq j$. Figure 2 shows the average scheduling cell size as a function of Q and T. Regardless of Q, a smaller T results in a higher degree of coordination as expected. In a sense, the level of coordination that is affordable – based on the available network infrastructure (e.g. back-haul capacity) – is reflected in an appropriate choice of T. Parameter Q can then fine tune the level of BS coordination based on the desired trade-off between complexity and performance.

IV. Resource allocation framework

The proposed resource allocation framework is comprised of two levels of abstraction. In the first level (level A), the network is clustered into scheduling cells. In the second level (level B), resource allocation is performed. Resource allocation is coordinated among the BSs forming each scheduling cell and is performed independently across scheduling cells. It is based on a coordinated proportional fair strategy developed in Section IV-A. Section IV-B discusses the time scales of adaptation in this framework.

A. Coordinated resource allocation

A scheduling cell c with the BS set $\mathcal{A}(c)$ and the terminal set $\mathcal{K}(c)$ is studied. For ease of notation the index c is dropped

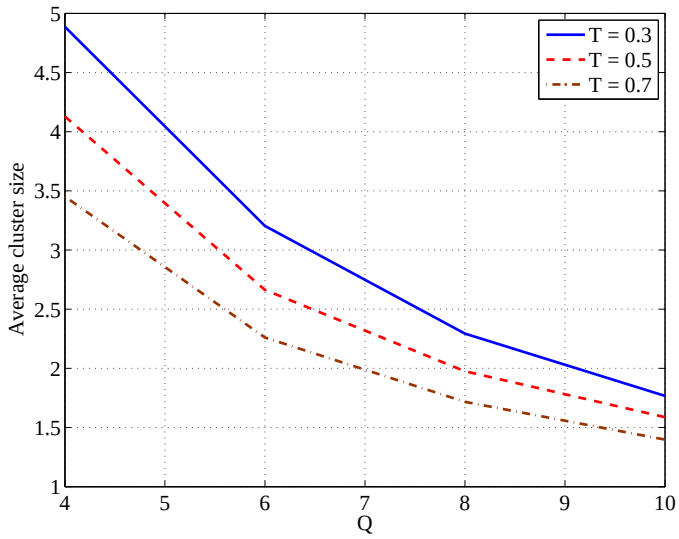


Fig. 2. The average scheduling cell size as a function of clustering parameters Q and T: With a fixed Q, a larger T results in smaller scheduling cells. In a network where higher complexity can be afforded, the system designer can increase T to achieve a higher coordination among BSs.

throughout the Section. In time slot t the data rate of terminal l is given by (2). After time slot t, the average rate of terminal l is updated according to

$$\begin{aligned}\bar{r}_l(t) &= \frac{(t-1)\bar{r}_l(t-1) + r_l(t)}{t} \\ &= \left(1 - \frac{1}{t}\right)\bar{r}_l(t-1) + \frac{1}{t} r_l(t),\end{aligned} \tag{27}$$

where the average rate at the beginning of the first time slot is set to zero for all terminals. The resource allocation problem at the beginning of time slot t is formulated as

$$\max \sum_{k_l \in \mathcal{K}} \log \bar{r}_l(t), \tag{28}$$

where $\mathcal{K}$ is the set of terminals in the scheduling cell. It can be shown that assuming $\frac{1}{t-1} \cdot \frac{r_l(t)}{\bar{r}_l(t-1)} \ll 1$ the following approximation holds:

$$\begin{aligned}\log \bar{r}_l(t) \approx \log &\left[\left(1 - \frac{1}{t}\right)\bar{r}_l(t-1)\right] \\ &+ \frac{1}{t-1} \cdot \frac{r_l(t)}{\bar{r}_l(t-1)}.\end{aligned} \tag{29}$$

At the beginning of time slot t the first term is known (constant). Hence, the optimization problem (28) can be approximated by

$$\max \sum_{k_l \in \mathcal{K}} \frac{r_l(t)}{\bar{r}_l(t-1)}. \tag{30}$$

In a multi-carrier system this problem can be generalized as follows:

$$\max \sum_{k_l \in \mathcal{K}} \sum_{n=1}^{N} \frac{I_{ln}(t) r_{ln}(t)}{\bar{r}_l(t-1)}. \tag{31}$$

The power budget of the BSs is assumed to be uniformly distributed across the available frequency channels. In a given time slot and frequency channel, a BS either transmits with a fixed transmit power or not transmit at all. With the assumption of equal power distribution across the available frequency channels, the optimization problem in (31) is reduced to a per-frequency channel problem as follows:

$$\max \sum_{k_l \in \mathcal{K}} \frac{r_{ln}(t)}{\bar{r}_l(t-1)}. \tag{32}$$

In the remainder of this Section, a system with one frequency channel is assumed. Let $\mathbf{p}(t) = [p_1(t), \cdots, p_{|A|}(t)]^T$ be the transmission power vector of the BSs in time slot t. Each terminal receives a desired signal from the serving BS and interfering signals from all or a subset of BSs in the network. With a fixed power vector $\mathbf{p}(t)$, each BS knows the received signal power levels from all BSs in the scheduling cell at each of its assigned terminals. Hence, BS i can estimate[4] the supported data rate of the terminals it is serving. Accordingly, each BS allocates the frequency channel to terminal $k_a^*(t)$ based on the following proportional fairness criterion:

$$k_a^*(t) = \arg \max_{k_l \in \mathcal{K}_a} \frac{r_l(t)}{\bar{r}_l(t-1)} \quad , \forall a \in \mathcal{A} \tag{33}$$

For a given power vector $\mathbf{p}(t)$, the set of the chosen terminals $\{k_a^*(t)\}_{a \in \mathcal{A}}$ forms the co-channel user set $\mathcal{K}(\mathbf{p}, t)$ in time slot t. In a scheduling cell with A BSs, there are 2^A possible transmission power vectors. At time slot t, for each power vector $\mathbf{p}(t) \in \{0, P_{\max}\}^{2^A \times 1}$ the corresponding co-channel user set $\mathcal{K}(\mathbf{p}, t) = \{k_a^*(t)\}_{a \in \mathcal{A}}$is formed according to (33). The optimal power vector at time slot t is then the one which satisfies

$$\mathbf{p}^*(t) = \arg \max_{\mathbf{p}} \sum_{k_l \in \mathcal{K}(\mathbf{p}, t)} \frac{r_l(t)}{\bar{r}_l(t-1)} \tag{34}$$

The complexity of this approach increases exponentially with the number of BSs in a scheduling cell. However, this complexity is tunable by adjusting the average size of the scheduling cells as will be discussed in Section V.

B. Time scales of adaptation

The coordinated resource allocation at level B is performed once every transmission time interval, also known as a sub-frame (SF) in the LTE terminology. The network clustering at level A re-forms the scheduling cells at a larger time scale, i.e. once every U sub-frames. The system parameter U primarily depends on the rate of variations in the wireless communication environment. For a fast time-varying network, a small U translates into fast re-formation of the scheduling cells. Parameter U is essentially a design parameter which should be adjusted according to the environment. As an example, let us assume that the shadow fading changes with a displacement in the order of tens of wavelengths. For a system with a

[4]It should be noted that due to the lack of knowledge of out-of-cell interference, the actual supported data rate of each terminal is not available. However, an estimate can be obtained based on the in-cell interference which is known due to the coordination within each scheduling cell.

center frequency of 2GHz and a terminal velocity of 30Kmph, a terminal travels ten wavelengths in 0.18 seconds which is equivalent to 18 frames in the LTE system. Consequently, a U of a few tens of sub-frames will be a reasonable choice. The pseudo-code of the framework is presented in Algorithm 4.

Algorithm 4 Proposed framework pseudo-code

Require: Parameters Q, U.

loop
  $\{$LEVEL A$\}$
  Update the set of active BSs $\mathcal{A}$ and active terminals $\mathcal{K}$.
  Update the wide-band interference matrix $\mathbf{H} = [h_{li}]^{K\times A}$.
  Assign terminal l to BS $a(l)$ such that $a(l) = \arg\max_{a_i\in\mathcal{A}} h_{li}$.
  Form the similarity matrix $\mathbf{S} = [s_{ij}]^{A\times A}$, with s_{ij} in (10).
  Apply the adaptive clustering Algorithm 3 on similarity matrix $\mathbf{S}$ and form the scheduling cells $(\mathcal{A}(c), \mathcal{K}(c))$.
  $\{$LEVEL B$\}$
  for SUBFRAME$= 1$ to U **do**
    Perform coordinated resource allocation in each scheduling cell as discussed in Section IV-A.
  end for
end loop

V. Numerical results

In Section V-A, we study the performance of a traditional cellular network. Section V-B discusses the performance of the proposed framework in an autonomous cellular network.

A. Traditional cellular network

The BSs are deployed regularly according to the hexagonal pattern with an inter-site-distance of 500 meters. Channel gains are modeled based on the ITU recommendations for urban-macro environments [14]. The simulation parameters are provided in Table I. For simplicity, a system with one frequency channel is considered. In order to comply with the LTE air interface technology, the system bandwidth has been chosen to be equal to that of a sub-carrier in the standard. In addition, each time slot has been chosen to be equal to a sub-frame of 1ms. In each time slot 14 transmissions (in time) occur for the scheduled terminal corresponding to the 14 OFDM symbols in a resource block.

The effect of the clustering parameter T is discussed in section III-B2. In the numerical results a fixed T of 0.7 has been considered. The choice of this value is simply because it results in a shorter simulation time due to a lower degree of coordination among BSs. With a smaller T, better performance is expected at the expense of higher complexity.

The performance of the proposed framework is compared with the following classical techniques:

- *Universal frequency reuse (UFR)*: The entire system spectrum is available to all BSs. At each time slot, each BS schedules one terminal on the available frequency channel based on the proportional fairness criterion.

TABLE I
System parameters

System parameter	Value
Carrier frequency	2GHz
Bandwidth (W)	15KHz
Sub-frame size (time slot duration)	1ms
Number of simulated sub-frames (t_s)	40
BS power budget ($P_{\max}$)	18.22dBm
Noise figure at terminal	7dB
Background noise PSD	-174dBm/Hz
Terminal velocity	30Kmph
Number of BSs (M)	21
Total number of terminals	$4, 6, 8, 10$ per BS
Traffic model	Full buffer
Clustering algorithm threshold (T)	0.7

- *Static cluster size of* 3: The cellular network with 21 BSs is clustered into 7 scheduling cells of 3 BSs. Scheduling is based on the coordinated approach discussed in Section IV-A.

In Figure 3 the average size of clusters is plotted as a function of the clustering parameter Q. A higher Q results in a higher number of clusters. Essentially Q is a system parameter, equivalent to the frequency reuse factor of the cellular concept, which adjusts the level of BS coordination and hence the complexity in resource allocation.

The system performance is evaluated with Monte-Carlo simulations. In the ITU report 2135, a *drop* is defined as an independent deployment of terminals and BSs [14]. During the simulation of a drop, the shadow fading and slow fading channel gains are assumed to be constant. The mobility of terminals, however, change the channel gains from one sub-frame to the next according to the Doppler effect. In this paper, 1000 drops have been considered and the performance of each drop is evaluated over 40 sub-frames.

Let $\bar{\eta}_l$ be the average spectral efficiency of terminal l defined as

$$\bar{\eta}_l = \frac{\bar{r}_l(t_s)}{W}. \tag{35}$$

The cell spectral efficiency is defined as the sum spectral efficiency of all terminals in the network normalized by the number of BSs M.

$$\bar{\eta} = \frac{\sum_{l=1}^{K} \bar{\eta}_l}{M} \tag{36}$$

Let $\mathbf{s} = [\bar{\eta}_1, \cdots, \bar{\eta}_K]$ denote the vector of average spectral efficiencies of all terminals in the network. The cell edge spectral efficiency is defined as the 5% point of the cumulative distribution function of $\mathbf{s}$.

The performance of the proposed technique is evaluated for two values of 4 and 8 for Q, corresponding to average cluster sizes of 3.3 and 1.7 respectively. The cell spectral efficiency performance results are provided in Figure 4. With a Q of 4, the cell spectral efficiency is on average 10.7% higher than that of UFR. Compared to a static cluster size of 3, the spectral efficiency has improved by 2.4%. With a Q of 8, the spectral efficiency is 5.5% greater than that of UFR. It achieves 97.7% of the cell spectral efficiency with a static cluster size of 3. In other words, by reducing the coordination from 3 to 1.7 BSs (43.4%), the loss in performance is only 2.3%.

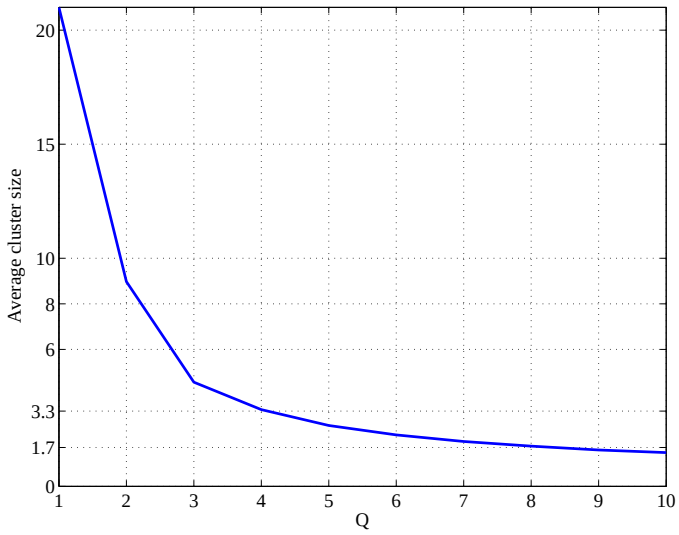


Fig. 3. Average cluster size as a function of the clustering parameter Q: Increasing Q decreases the average cluster size. This in turn results in a lower level of coordination among BSs.

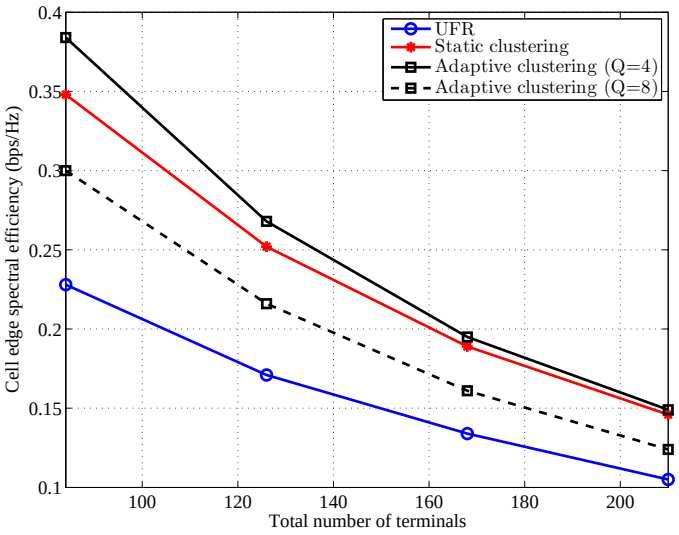


Fig. 5. Cell edge spectral efficiency as a function of the total number of terminals in the network: Despite the lack of coordination among clusters, these results show that the cell edge performance is not compromised.

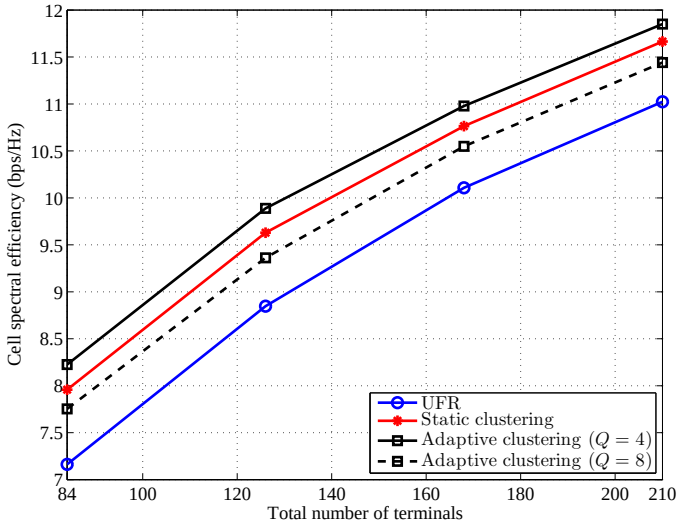


Fig. 4. Cell spectral efficiency as a function of the total number of terminals in the network: UFR achieves the worst performance due to excessive level of interference among BSs. For a Q of 8 the proposed technique is better than UFR. With a Q of 4 the coordination is increased to an average of 3.3 which results in a better performance. A proper choice for Q achieves a desirable balance between complexity and performance.

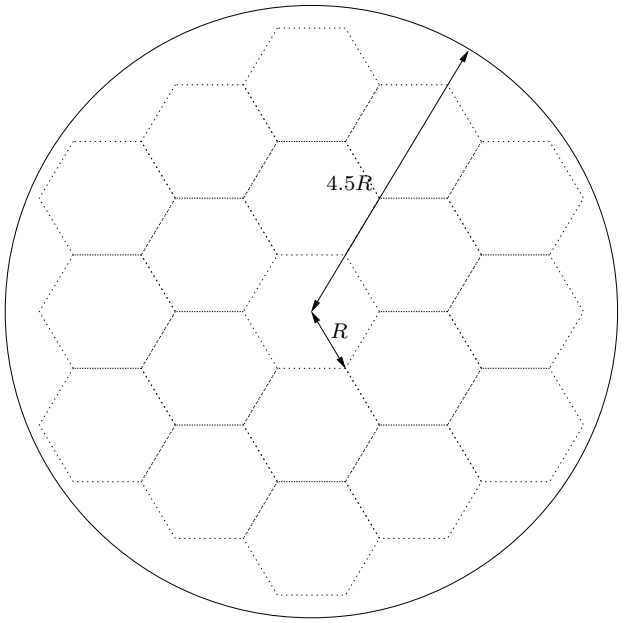


Fig. 6. Network coverage area for a cellular network with 19 BSs: In the traditional network the BSs are deployed at the center of the illustrated hexagons. In the autonomous cellular network the same number of BSs are randomly deployed in the network coverage area, modeled with a circle of radius $4.5R$.

The cell edge spectral efficiency performance is illustrated in Figure 5. With a Q of 4 the cell edge spectral efficiency is on average 53.1% higher than that of UFR. Compared to a static cluster size of 3, the cell edge spectral efficiency has improved by 5.5%. With a Q of 8, the cell edge spectral efficiency is 24% great than that of UFR. It achieves 85.5% of the cell spectral efficiency of static clustering.

B. Autonomous cellular network

Due to the irregularity in the infrastructure of an autonomous cellular network, most of the classical resource allocation techniques are not applicable. In fact, UFR is the only classical method which can be considered. Figure 6 illustrates the coverage area of a traditional cellular network with 19 BSs. In the autonomous cellular network, the same number of BSs is deployed in the same coverage area, modeled with a circle of radius $4.5R$. A random two-dimensional uniform distribution is considered for the deployment of BSs and terminals. The same system parameters of Section V-A are adopted with the exception that in this Section the cellular network is formed by 19 BSs (instead of 21).

In Figure 7, the average size of a scheduling cell is plotted as a function of the clustering parameter Q. A Q of 8 results in an average scheduling cell size of 1.7 BSs. With a Q of

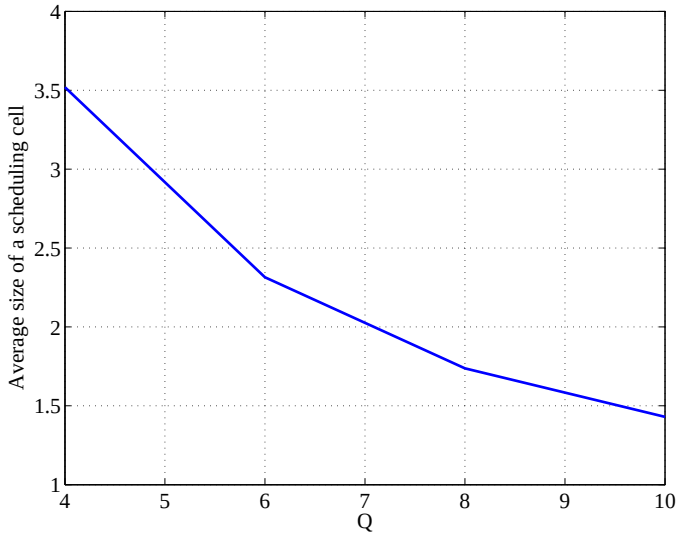


Fig. 7. Average size of the scheduling cells as a function of the clustering parameter Q: The proposed technique clusters the network regardless of the BS deployment pattern. Similar to the traditional network, increasing Q results in a lower level of coordination among BSs.

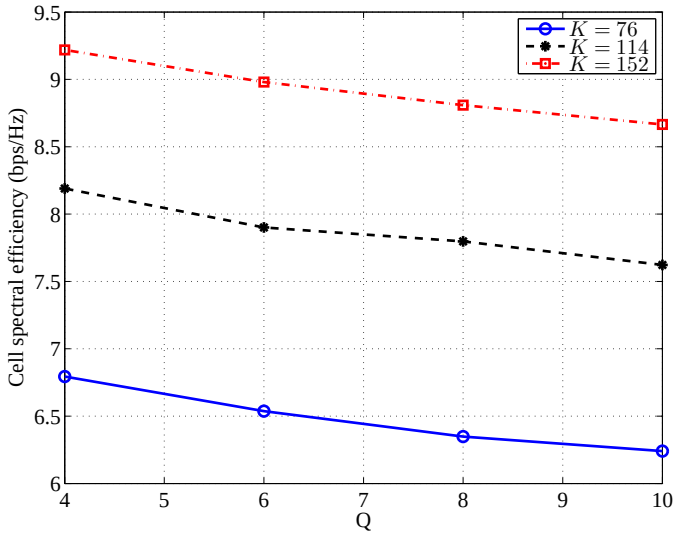


Fig. 8. Cell spectral efficiency as a function of the clustering parameter Q: As the number of terminals in the network increases, a higher spectral efficiency is observed. This can be explained by the larger multi-user diversity gain. With a fixed number of terminals, a smaller Q improves the performance due to a higher level of coordination among BSs.

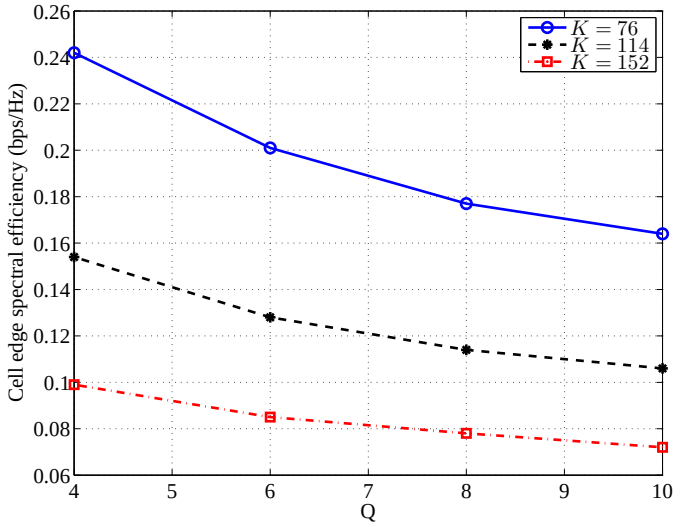


Fig. 9. Cell edge spectral efficiency as a function of the clustering parameter Q

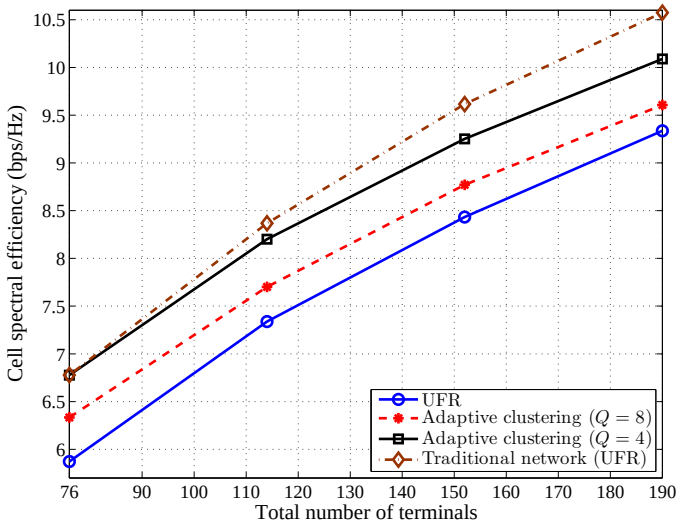


Fig. 10. Cell spectral efficiency as a function of the total number of terminals in the network: The performance of two network with regular (traditional) and irregular (autonomous) BS deployment pattern are compared. Based on the system designer requirements, the technique allows a tunable trade-off between complexity and performance. Decreasing Q increases coordination among BSs and consequently improves the performance.

4 the level of coordination increases to among an average of 3.5 BSs.

The cell spectral efficiency is plotted as a function of Q in Figure 8. The performance is evaluated for an average of 4, 5 and 6 terminals per BS. Based on Figure 7, decreasing Q from 10 to 4 is equivalent to increasing the coordination level by a factor of 2.5. For a network with 152 active terminals, this results in a 6.4% improvement in the cell spectral efficiency. The cell edge spectral efficiency improves by 35.5% as illustrated in Figure 9.

In the rest of this section, the system performance is evaluated for Q equal to 4 and 8 corresponding to an average cluster size of 3.5 and 1.7 BSs respectively. The cell spectral efficiency is plotted as a function of the total number of terminals in Figure 10. With a Q of 8, the cell spectral efficiency is on average 4% higher than that of UFR. Increasing the coordination to among an average of 3.5 BSs (with Q of 4) the cell spectral efficiency improvement increases to 11%. The performance results of a traditional cellular network are also provided as a point of reference. In the traditional network, UFR is considered for resource allocation and is denoted by 'Traditional network (UFR)'. The results show that the autonomous network with no coordination, i.e. UFR, incurs a penalty of 14% for deployment irregularity in the cell spectral efficiency. With a Q of 8 the penalty is reduced to 8% which is further reduced to 3% with a Q of 4.

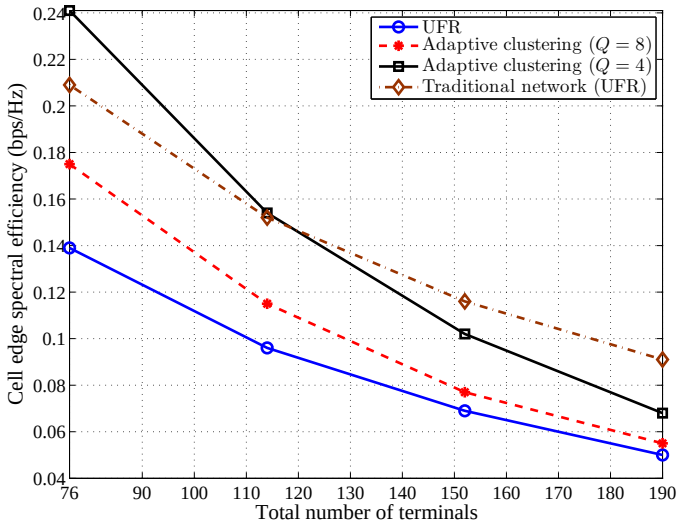


Fig. 11. Cell edge spectral efficiency as a function of the total number of terminals in the network

The cell edge spectral efficiency performance results are provided in Figure 11. With a Q of 8, the cell edge spectral efficiency is improved by 17% when compared to UFR. With a Q of 4 an improvement of 54% is observed in the cell edge spectral efficiency. When compared to a traditional network, a Q of 8 achieves 72% of the cell edge spectral efficiency and Q of 4 achieves 94% of the cell edge spectral efficiency. This is while the autonomous network with no coordination, i.e. UFR, achieves 61% of the cell edge spectral efficiency of the traditional network. Table II reports the performance evaluation results as a percentage of the performance in the traditional network.

VI. Summary and Future Work

In this paper we have developed an adaptive framework for resource allocation in cellular networks with irregularly deployed BSs. A scheduling cell is defined by a set of coordinating BSs. A network clustering method is proposed. Accordingly, the scheduling cells are formed based on the network infrastructure (i.e. distribution of BSs and terminals). Scheduling cells adapt to the time variations in the network. The size of each scheduling cell is a function of the density of BS and terminal deployment in each locality. The average size of scheduling cells is tunable in order to achieve a desirable balance between performance and complexity. A coordinated resource allocation scheme is developed which achieves proportional fairness among the terminals in each scheduling cell.

In section III-A, we have introduced the notion of similarity between two BSs. The similarity index quantifies the level of interference among BSs and plays an important role in the formation of the scheduling cells. It should be noted that there are many ways to define the similarity index. The defined index in this paper is dynamically updated based on the channel gains of the active terminals. One might consider static indices solely based on the location of the BSs. Furthermore the proposed clustering technique is based on the spectral clustering approach. Similar methods, such as the page rank algorithm [15], should also be studied.

TABLE II
Resource allocation performance comparison

Resource allocation method	Cell spectral efficiency(%)	Cell edge spectral efficiency (%)
UFR	86	61
Proposed ($Q = 8$)	92	72
Proposed ($Q = 4$)	97	94

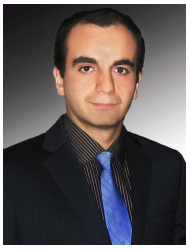
In an autonomous cellular network, the deployment of extra BSs would naturally improve the system performance. If we assume there are a limited number of extra BSs to be deployed, where is the best location to deploy them? The strategic positioning of additional BSs given the available network infrastructure is an interesting problem to pursue. In addition, we have assumed all terminals have the same quality of service (QoS) requirements. The volume of traffic is directly translated into the number of active terminals. The framework can be extended to cases where multiple classes of terminals with different QoS requirements are considered.

In this paper, each terminal communicates with only one BS. Alternatively, each terminal can communicate with more than one BS in the scheduling cell. Coordinated transmission from the BSs in a cell to each terminal can potentially improve the system performance. However, whether the additional complexity would result in considerable improvements or not should be carefully examined. Furthermore, we have assumed that the power budget of each BS is equally distributed across the available frequency channels. The frequency channels are then assigned based on a terminal scheduling policy. Although this is a common practice in OFDMA-based systems, power control is still an area which should be thoroughly studied in autonomous cellular networks.

References

[1] V. H. Macdonald, "The cellular concept," *Bell System Technical Journal*, vol. 58, no. 1, pp. 15–41, 1979.
[2] T. Bonald, S. Borst, and A. Proutière, "Inter-cell scheduling in wireless data networks," *European wireless proceedings*, 2005.
[3] G. Li and H. Liu, "Downlink radio resource allocation for multi-cell OFDMA system," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3451–3459, Dec. 2006.
[4] K. Son, Y. Yi, and S. Chong, "Utility-optimal multi-pattern reuse in multi-cell networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 142–153, Jan. 2011.
[5] M. Rahman and H. Yanikomeroglu, "Enhancing cell-edge performance: a downlink dynamic interference avoidance scheme with inter-cell coordination," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1414–1425, Apr. 2010.
[6] R. Y. Chang, Z. Tao, J. Zhang, and C.-C. J. Kuo, "Multicell OFDMA downlink resource allocation using a graphic framework," *IEEE Trans. Veh. Technol.*, vol. 58, no. 7, pp. 3494–3507, Sept. 2009.
[7] S. H. Ali and V. C. M. Leung, "Dynamic frequency allocation in fractional frequency reused OFDMA networks," *IEEE Trans. Commun.*, vol. 8, no. 8, pp. 4286–4295, Aug. 2009.
[8] G. Fodor, C. Koutsimanis, A. Rácz, N. Reider, A. Simonsson, and W. Müller, "Intercell interference coordination in OFDMA networks and in the 3GPP long term evolution system," *Journal of Communications*, vol. 4, no. 7, pp. 445–453, Aug. 2009.
[9] A. L. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse for best effort traffic through distributed inter-cell coordination," in *Proc. 2009 IEEE INFOCOM Conference*, pp. 1287–1295.

[10] H. Zhang, L. Venturino, N. Prasad, P. Li, S. Rangarajan, and X. Wang, "Weighted sum-rate maximization in multi-cell networks via coordinated scheduling and discrete power control," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1214–1224, June 2011.
[11] L. Venturino, N. Prasad, and X. Wang, "Coordinated scheduling and power allocation in downlink multicell OFDMA networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 2835–2848, July 2009.
[12] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks-part II: algorithm development," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 625–634, Mar. 2005.
[13] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, Aug. 2007.
[14] *ITU-R M.2135, Requirements related to technical performance for IMT-Advanced radio interface(s)*, Std., 2008. [Online]. Available: http://www.itu.int/publ/R-REP-M.2135-2008/en
[15] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: bringing order to the web (1998)," Stanford Digital Library Technologies Project, Tech. Rep., 1998.



Bijan Golkar received the B.Sc. degree in electrical engineering from KNToosi university, Tehran, Iran in 2005. He received the M.A.Sc degree in 2007 from Carleton University, Ottawa, Canada and the Ph.D. degree from University of Toronto in 2013. During his PhD studies he worked on self-configurable and adaptive techniques for resource allocation in autonomous cellular networks. He has also developed a practical model for the generation of shadow fading gains for realistic performance evaluation of these networks. His main research interests include wireless communications and broadband wireless networks.

Elvino S. Sousa received his B.A.Sc. in engineering science, and the M.A.Sc. in Electrical Engineering from the University of Toronto in 1980 and 1982 respectively, and his Ph.D. in electrical engineering from the University of Southern California in 1985. Since 1986 he has been with the department of Electrical and Computer Engineering at the University of Toronto where he is now a Professor and the Jeff Skoll Professor in Computer Network Architecture. He has performed research in CDMA and wireless systems since 1983. His current interests are in the areas of broadband wireless systems, smart antenna systems, autonomous infrastructure wireless networks, cognitive radio, self configurable wireless networks, user deployed networks, and cognitive networks. He was the founder of wireless communications at the University of Toronto and is the director of the wireless lab, which has undertaken research in wireless systems for the past 25 years. He has been invited to give lectures and short courses on spread spectrum, CDMA, and wireless systems in many countries. He was the technical program chair for PIMRC 95, vice-technical program chair for Globecom '99, Technical program co-chair for WPMC 2010 (Recife, Brazil) and Technical program co-chair for PIMRC 2011 (Toronto, Canada), and has been involved in the technical program committee of numerous international conferences. He is a past chair of the IEEE Technical committee on Personal Communications and was awarded the inaugural TCPC award for his contributions in the field. He has spent sabbatical leaves at Qualcomm and Sony CSL/ATL, where he was the holder of the Sony sabbatical chair. He has been awarded the Queen Elizabeth II Golden Jubilee Medal.