Three-Stage Resource Allocation Algorithm for Energy-Efficient Heterogeneous Networks

Cemil Can Coskun, *Student Member, IEEE*, Kemal Davaslioglu, *Member, IEEE*, and Ender Ayanoglu, *Fellow, IEEE*

***Abstract*—In this paper, we investigate the energy efficiency of downlink transmissions in heterogeneous networks. Our objective is to satisfy the rate requirement of users while maximizing the energy efficiency of the network. We employ the fractional frequency reuse (FFR) scheme to increase the energy efficiency of downlink transmissions and to eliminate outages for the cell-edge users. We formulate the joint cell-center boundary selection for FFR, scheduling, and power allocation problems. This formulation gives us a mixed discrete (selection of the cell-center boundary selection for FFR and scheduling) and continuous (power allocation) optimization problem which is hard to solve jointly. In order to solve this problem, we propose a three-stage resource allocation algorithm. In the first stage, we propose a dynamic method to determine the cell-center region boundaries. In the second stage, we employ the Lagrangian-directed scheduler algorithm to incorporate the rate requirements of users. The third stage solves the power allocation subproblem using the Levenberg–Marquardt method combined with dual decomposition. In order to make the base stations further reduce intercell interference, interference pricing mechanism is applied. This scheme penalizes the utility of a base station with the interference it creates. The performance of the proposed algorithm is simulated in a long-term evolution (LTE) network simulation tool. Our numerical results reveal that significant gains in terms of energy efficiency can be achieved with the proposed algorithm. The outage probability is also significantly reduced.**

***Index Terms*—Energy efficiency, heterogeneous cellular networks, Levenberg–Marquardt algorithm, long term evolution (LTE), power control, resource allocation, scheduling.**

I. Introduction

With the proliferation of the mobile devices, such as smart phones, tablets, and laptops, ubiquitous, fast, and reliable wireless connections are needed. A recent study in [1] estimates that the number of mobile devices will reach 11.6 billion by 2020. According to the same source, this will be accompanied by an increase in mobile data traffic volume reaching a compounded annual growth rate of 53% between 2015 and 2020. Motivated by the financial and ecological concerns, network operators are pursuing energy-efficient solutions to keep their energy consumption at reasonable levels while satisfying this demand. These solutions have been studied in the literature under the general description "Green Communications," see, e.g., [2] and the references therein.

Resource allocation problem is an important research area in the field of game theory with wide applications in wireless communication networks, see, e.g., [3]–[8]. This problem arises when multiple agents or players with conflicting interests compete for the same resources. For example, resource allocation in heterogeneous networks (HetNets) reflects a case such that while each base station tries to maximize its utility by increasing its transmit power, this creates an excessive interference in the system for the neighboring cells. In this paper, we employ interference pricing, which is introduced in [4]–[6]. Each base station is penalized according to the interference it creates. The penalty factor increases if the interference leads to outage of users in other sectors. The pricing algorithm defines the utility as the energy efficiency of a sector minus the created interference to other sectors. This makes macrocell base stations (MBSs) decrease their transmission power, and the energy efficiency of the network is further improved. In addition, due to the drop of the interference level, the outage probability of users decreases.

Energy-efficient wireless networks have been widely investigated in the literature, see, e.g., [9]–[18] and the references therein. In [9], Xiong *et al.* formulate energy efficiency function with predetermined weights to user rates and propose three resource allocation algorithms: optimal, near optimal, and suboptimal. A suboptimal resource allocation algorithm is proposed in [10] in which each user is first assigned with one subcarrier and then the rest of the resources are assigned to minimize the power consumption. They show that energy efficiency function is quasi-concave for given subcarrier assignment. Then, the multilevel water-filling algorithm is used for power allocation. The scheduling problem in a single-cell network is transformed into a fractional program in [11]. The water-filling algorithm is used for power allocation. In [12], an energy-efficient water-filling algorithm is proposed. Energy-per-goodbit metric is defined for the optimization problem. However, all of these works investigate the problem in single-cell single-tier networks. The proposed algorithm in this paper investigates resource allocation problem in multicell multitier networks. The following works investigate the energy-efficient wireless networks in multicell networks. An iterative energy-efficient link adaptation

Manuscript received April 18, 2016; revised September 12, 2016 and November 28, 2016; accepted January 22, 2017. Date of publication January 31, 2017; date of current version August 11, 2017. This work was supported in part by the National Science Foundation under Grant 1307551. The review of this paper was coordinated by R. Jäntti.

C. C. Coskun and E. Ayanoglu are with the Center for Pervasive Communications and Computing, Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697-2625 USA (e-mail: ccoskun@uci.edu; ayanoglu@uci.edu).

K. Davaslioglu was with the Center for Pervasive Communications and Computing, Department of Electrical Engineering and Computer Science, University of California, Irvine, CA 92697-2625 USA. He is now with Intelligent Automation Inc., Rockville, MD 20855 USA (e-mail: kemald@usf.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2017.2661698

framework is presented in [13]. This paper considers the circuit power and transmit power jointly to determine the optimal solution that maximizes the energy efficiency in a noncooperative manner. A similar link adaptation algorithm that considers interference to determine the transmit power levels is proposed in [14]. In [15], the energy-efficient resource allocation in an OFDMA network has been investigates. It is shown that the maximum energy efficiency and spectral efficiency are achieved at the same point for the low-power regime. In [16], Venturino *et al.* show that energy efficiency maximization problem is equivalent to spectral efficiency maximization problem for small values of transmission power. However, rate requirements of users are not considered in [16]. In [13]–[16], universal frequency reuse (FR) has been employed and HetNet deployments have not been considered. Contrary to these works, in what follows, we present an algorithm that determines the MBS and picocell transmit power levels together in networks with fractional FR (FFR).

Several interference cancellation and mitigation techniques have been investigated in the literature, such as FFR, opportunistic scheduling, and almost blank subframes [19], [20]. In next-generation networks, such as long-term evolution (LTE) and 5G, FFR has been identified as an efficient, and at the same time, a low-complexity method to mitigate intercell interference [21]. Although the FFR scheme has been mostly studied in the literature for single-layer networks, see, e.g., [22] and [23], its performance has not been investigated to its full potential in multitier networks. Recently, energy efficiency of heterogeneous cloud radio access networks is investigated in [24], and the soft-FFR method is employed to mitigate the interference between the high and low power nodes. It has been shown that heterogeneous cloud radio networks provide significant performance gain over both HetNets and cloud networks. In [25], Huq *et al.* propose an energy-efficient resource scheduling algorithm for heterogeneous coordinated multi-point transmissions. The proposed algorithm in [25] converges quickly which is crucial in practical wireless systems. A novel multitier FFR scheme has been proposed in [26] and [27], which have investigated its throughput and outage probability in [26] and its spectral efficiency in [27] for constant cell-center boundaries. In this paper, we will also employ the same FFR scheme, but with dynamic cell-center boundaries to maximize the energy efficiency.

The problems of energy efficiency maximization and satisfying rate constraints do not always coincide. On the contrary, these problems usually contradict with each other. In order to satisfy rate requirements of users, base stations may need to increase their transmission power to energy-inefficient levels. In the literature, these two problems are studied in [13], [18], and [12]. However, none of these works considers intercell interference conditions that increase the complexity of the problem. In this paper, we address this problem in a multicell environment in which the intercell interference is very critical.

The contributions of this paper are as follows. We study energy efficiency maximization of HetNets with rate constraints using the FFR scheme and propose a dynamic algorithm to select cell-center radius, assign frequency resources, and solve the power allocation problem. First, a dynamic method is proposed to determine cell-center radius. Second, frequency resources are assigned to users considering the interference conditions and user rate requirements. As the majority of the network traffic is generated indoors [28], the first stage is typically solved at a large time scale. However, in this paper, we combine the first and second stage and solve cell-center radius selection problem in small scale. Last, a Levenberg–Marquardt method-based approach is implemented to solve the power allocation problem. Dual decomposition techniques are used for minimum rate constraints of users. This reactive approach helps us further improve the energy efficiency and satisfy the rate requirements of users. Note that the third stage is updated more frequently at a smaller time scale. We have previously proposed a gradient ascent based power allocation method to solve the energy efficiency maximization problem without considering rate constraints of users, constant cell-center boundaries, and constant frequency assignments in [29]. In this paper, we propose a Levenberg–Marquardt method-based approach to extend the framework to consider Quality-of-Service constraints and to increase the speed of convergence.

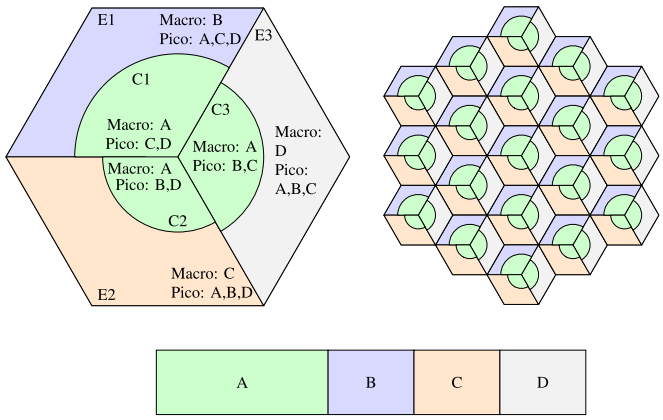


Fig. 1. Dynamic cell-center region boundaries and spectrum assignments in a multitier FFR scheme. The network layout assumes a uniform 19 cell hexagonal grid in which the MBSs have three sector antennas and pico-BSs employ omnidirectional antennas.

The remainder of this paper is organized as follows. Section II introduces the system model and base station power consumption models. Section III formulates the energy-efficient resource allocation problem in HetNets. The proposed solution methods are presented in Section IV. Simulation results are presented to evaluate the performance gains in Section V and concluding remarks are presented in Section VI.

II. System Model

In this section, we first present the system model and the multitier FFR scheme. We then study a linearized base station power consumption model that will be used later to formulate the energy-efficient resource allocation problem in Section III.

Consider a cellular layout of 19 hexagonal cells, as depicted in Fig. 1. As is commonly done in the literature, we will use this layout together with the wrap-around technique, e.g., [29] and [30], to model a cellular network of infinite dimensions. Assume that macrocells employ three-sector antennas and picocell base stations (pico-BSs) have omnidirectional antennas. Spectrum allocation in the macrocell and picocell tiers is fundamental to determine the interference conditions and user rates at each tier.

In densely deployed networks, intercell interference becomes a significant problem, limiting the system performance. The FFR scheme provides a solution by assigning the frequency resources in a coordinated manner such that high interference conditions are avoided. In this paper, we employ the FFR scheme depicted in Fig. 1. We refer to the macrocell-associated users as MUEs and picocell-associated users as PUEs. The MUEs can be categorized into cell-center and cell-edge users depending on factors, such as their received reference power, path loss, or traffic load within the sector [29]. In this process, the variable $r_{th,s}$ determines the cell-center region boundary of sector s. In Section IV-A, we elaborate the selection process of $r_{th,s}$ in detail. In terms of spectrum allocation, the MBSs can allocate subcarriers in subband A to their cell-center users, whereas the cell-edge MUEs are assigned to either one of the remaining three subbands. For example, in Sector 1, subband B is allocated to the cell-edge users. Pico-BSs in the cell-center region are assigned to orthogonal channels with respect to the subbands that MBS operates at. This is to reduce the cross-tier interference as the cell-center pico-BSs are close to the MBS. For example, cell-center pico-BSs operate at subbands C and D. Note that, in a typical LTE deployment, the MBSs and pico-BSs have around 16 dB transmit power difference [31]. This would have detrimental effects for the cell-center PUEs in the downlink if they were not assigned to orthogonal channels with the MBS. For the cell-edge pico-BSs, subband A can also be reused in order to increase the throughput. For example, cell-edge pico-BSs operate at subbands A, C, and D in Sector 1. Note that the multitier FFR scheme depicted in Fig. 1 favors the cell-center PUEs and cell-edge MUEs that would have been exposed to severe interference if, for example, universal FR had been employed. This also enables operation at a lower interference region such that the energy efficiency of the downlink transmissions can be significantly increased. This fact is demonstrated later in Section V through numerical simulations. We test our findings in an LTE scenario. In LTE, the smallest scheduling granularity is per resource block (RB) in which an RB consists of 12 subcarriers [31].

In this paper, we employ constant power allocation across subbands. Let N_A, N_B, N_C, and N_D denote the total number of subcarriers in subbands A, B, C, and D, respectively. The total number of subcarriers is denoted by N, i.e., $N = N_A + N_B + N_C + N_D$. We introduce two variables, ε and β, to determine transmission power levels of base stations. The parameter ε denotes the ratio of the downlink transmissions of cell-edge MUEs to cell-center MUEs. This parameter is introduced to favor either cell-center or cell-edge users that do not satisfy their minimum rate requirements. This parameter is only defined for the MBSs. The variable β scales the transmission power of the base stations. This parameter introduces energy savings into the system. The corresponding spectrum and power utilization scheme is illustrated in Fig. 2. The downlink transmission per subcarrier of an MBS M for cell-center MUEs in Sector 1, P_M, is given by

$$P_M = \frac{\beta_s P_{\max,M}}{N_A + \varepsilon N_B} \tag{1}$$

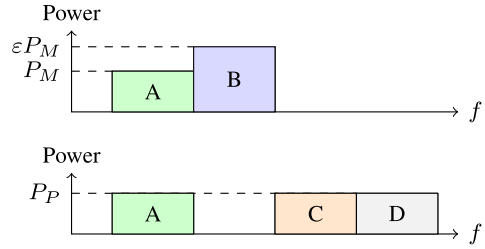


Fig. 2. Illustration of the proposed spectrum and power utilization of an MBS and a cell-edge pico-BSs in Sector 1 with the multitier FFR scheme.

where $P_{\max,M}$ is the maximum transmit power of an MBS. Similarly, for the cell-edge MUEs, the MBS transmit power per subcarrier is εP_M. It is straightforward to obtain the expressions for Sectors 2 and 3 by replacing N_B with N_C and N_D, respectively. In the picocell tier, the downlink transmission across subcarriers is also considered to be constant. For example, the transmit power per subcarrier of a pico-BS P, P_P, in Sector 1 can be expressed as

$$P_P^C = \frac{\beta_P P_{\max,P}}{N_C + N_D} \text{ and } P_P^E = \frac{\beta_P P_{\max,P}}{N_A + N_C + N_D} \tag{2}$$

where P_P^C and P_P^E denote the transmit power of a cell-center and a cell-edge pico-BS per subcarrier, respectively. The maximum transmit power of a pico-BS is represented by $P_{\max,P}$. For pico-BSs in Sectors 2 and 3, a similar expression can be obtained by simply replacing the values of N_C and N_D with the respective subband values.

Let sector s consists of K_s active users, and $\mathcal{K}_{M,s}$ and $\mathcal{K}_{P,s}$ denote the sets of MUEs and sets of PUEs, respectively. Let k be the index of user and n be the index of the subcarrier. The vector $\boldsymbol{C}_{M,s}$ consists of binary variables denoting whether or not MUE is in the cell-center region, i.e., $\boldsymbol{C}_{M,s}^k = 1$ if MUE k is located in the cell center region and $\boldsymbol{C}_{M,s}^k = 0$ otherwise, where $\boldsymbol{C}_{M,s}^k$ is the kth entry of the vector $\boldsymbol{C}_{M,s}$. Likewise, the vector $\boldsymbol{C}_{P,s}$ consists of binary variables $\boldsymbol{C}_{P,s}^k \in \{0,1\}$ denoting whether or not PUE is in the cell-center region. The matrix $\boldsymbol{F}_{M,s}$ is $|\mathcal{K}_{M,s}| \times N$ and its (n,k)th element denotes whether or not the subcarrier is assigned to MUE by a value of 1 or 0, respectively. The matrix $\boldsymbol{F}_{P,s}$ is $|\mathcal{K}_{P,s}| \times N$ and its (n,k)th element denotes whether or not the subcarrier is assigned to PUE by a value of 1 or 0, respectively. The matrices $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are $N \times |\mathcal{K}_{M,s}|$, and the matrices $\boldsymbol{R}_3$ and $\boldsymbol{R}_4$ are $N \times |\mathcal{K}_{P,s}|$. The (n,k)th element of $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ is the throughput of MUE k on subcarrier n when the user k is located in the cell-center and cell-edge regions, respectively. Likewise, the (n,k)th element of $\boldsymbol{R}_3$ and $\boldsymbol{R}_4$ denotes the throughput of PUE k on subcarrier n when the user k is located in the cell-center and cell-edge regions, respectively. The aggregate throughput of a sector s can be expressed as

$$\begin{aligned} R_s = &\sum_{k \in \mathcal{K}_{M,s}} (\boldsymbol{C}_{M,s}^k \boldsymbol{F}_{M,s}^{(k,:)} \boldsymbol{R}_1^{(:,k)} + (1 - \boldsymbol{C}_{M,s}^k) \boldsymbol{F}_{M,s}^{(k,:)} \boldsymbol{R}_2^{(:,k)}) \\ &+ \sum_{k \in \mathcal{K}_{P,s}} (\boldsymbol{C}_{P,s}^k \boldsymbol{F}_{P,s}^{(k,:)} \boldsymbol{R}_3^{(:,k)} + (1 - \boldsymbol{C}_{P,s}^k) \boldsymbol{F}_{P,s}^{(k,:)} \boldsymbol{R}_4^{(:,k)}). \end{aligned} \tag{3}$$

Note that $\boldsymbol{X}^{(n,:)}$ corresponds to the nth row vector of the matrix $\boldsymbol{X}$. Likewise, $\boldsymbol{X}^{(:,k)}$ is the kth column vector of the matrix $\boldsymbol{X}$. Depending on the associated tier, region, and subband, the throughput terms in (3) can be expanded using the definitions in (1) and (2) as

$$\begin{aligned}
\boldsymbol{R}_1^{(n,k)} &= \Delta_n \log_2\left(1+\frac{P_M g_{k,M}^{(n)}}{I_1^{(k,n)}+N_0\Delta_n}\right),\\
\boldsymbol{R}_2^{(n,k)} &= \Delta_n \log_2\left(1+\frac{\varepsilon_s P_M g_{k,M}^{(n)}}{I_2^{(k,n)}+N_0\Delta_n}\right),\\
\boldsymbol{R}_3^{(n,k)} &= \Delta_n \log_2\left(1+\frac{P_P^C g_{k,P}^{(n)}}{I_3^{(k,n)}+N_0\Delta_n}\right),\\
\boldsymbol{R}_4^{(n,k)} &= \Delta_n \log_2\left(1+\frac{P_P^E g_{k,P}^{(n)}}{I_4^{(k,n)}+N_0\Delta_n}\right) \qquad (4)
\end{aligned}$$

where $\boldsymbol{X}^{(n,k)}$ is the entry on the nth row and kth column of the matrix $\boldsymbol{X}$. The channel gain between user k and MBS M and between user k and pico-BS P on subcarrier n are denoted by $g_{k,M}^{(n)}$ and $g_{k,P}^{(n)}$, respectively. The interference incurred by user k on subcarrier n is denoted by $I_j^{(n,k)}$ in (4), where $j=1,2,3,4$. The value of $I_j^{(n,k)}$ can be calculated as follows:

$$I_j^{(n,k)} = \begin{cases} \sum\limits_{\substack{M'\neq M\\ M'\in\mathcal{B}_M^{(n)}}} P_{M'}^{(n)} g_{k,M'}^{(n)} + \sum_{P\in\mathcal{B}_P^{(n)}} P_P^{(n)} g_{k,P}^{(n)} & \text{for } j=1,2\\ \sum_{M'\in\mathcal{B}_M^{(n)}} p_{M'}^{(n)} g_{k,M'}^{(n)} + \sum\limits_{\substack{P'\neq P\\ P'\in\mathcal{B}_P^{(n)}}} P_{P'}^{(n)} g_{k,P'}^{(n)} & \text{for } j=3,4 \end{cases} \qquad (5)$$

where $\mathcal{B}_M^{(n)}$ and $\mathcal{B}_P^{(n)}$ denote the sets of MBSs and pico-BSs that transmit on subcarrier n, respectively. The sets of interfering MBSs and pico-BSs differ based on the sector, associated tier, and whether the UE is in the cell-center or cell-edge region. The transmit power of an MBS M and a pico-BS P are represented by $p_M^{(n)}$ and $p_P^{(n)}$, respectively. Likewise, these power levels depend on the sector and transmission band. The thermal noise power per Hz and bandwidth of a subcarrier are represented by N_0 and Δ_n, respectively, $\Delta_n = 15$ kHz for LTE systems [31]. Notice that for the cell-edge PUEs that are assigned to Band A, the interference term in $\boldsymbol{R}_4^{(n,k)}$ includes the intrasector interference contributions from the MBS M to a cell-edge PUE k in the same sector, that is $P_M g_{k,m}^{(n)}$. Including these terms enables us to balance the cross-tier interference while maximizing the sector energy efficiency. In [32] an extensive survey on resource allocation methods with cooperating BSs has been presented. Our paper, in particular, can be categorized as a case where base stations share their channel state information (CSI), which are obtained through feedback channels, to coordinate their power and scheduling assignments among cooperating cells. In this paper, we assume that there exists backhaul capacity to exchange CSI information among base stations.

Several base station power consumption models are proposed in the literature, see, e.g., [33]–[37]. In [33], Holtkamp *et al.* provide a parameterized power model that especially considers the effects of transmission bandwidth and number of transmission chains. In [34] and [35], the circuit power consumption is defined as a function of sum rate. In [36], the power consumption of base station is defined as a summation of the transmit power dependent power consumption and static power consumption. In [37], Isheden and Fettweis propose a power model for various types of base stations and the individual contributions of the different equipments, such as the baseband unit, radio frequency transceiver, power amplifier, power supply unit, and cooling devices, are considered for total power consumption. This model also captures the power consumption during the sleep mode, which is very crucial for next-generation networks [2]. In this paper, base stations go into sleep mode when they are not serving any user. Therefore, we employ the model in [37] in this paper. The power consumption at MBS M is given by

$$P_{\text{MBS},M} = \begin{cases} N_M^{TRX}\left(P_{0,M}+\Delta_M\beta_s P_{\max,M}\right), & \text{if } 0<\beta\leq 1,\\ N_M^{TRX} P_{\text{sleep},M}, & \text{if } \beta = 0 \end{cases} \qquad (6)$$

where $P_{\text{MBS},M}$ and P_M^{TX} are the total power consumption at an MBS and RF transmit power, respectively. N_M^{TRX} is the number of transceiver chains and $P_{0,M}$ is the power consumption at the minimum nonzero output power at an MBS. The slope of the load-dependent power consumption at an MBS is denoted by Δ_M. When an MBS does not transmit, it is considered to be in the sleep mode and its power consumption is captured in $P_{\text{sleep},M}$. In this model, the total power consumption depends on the transmit power or the load. Therefore, it is referred to as the load-dependent power consumption model [37]. Similarly, the power consumption of pico-BS P is given by

$$P_{\text{PBS},P} = \begin{cases} N_P^{TRX}\left(P_{0,P}+\Delta_P\beta_P P_P^{TX}\right), & \text{if } 0<P_P^{TX}\leq P_P^{\max},\\ N_P^{TRX} P_{\text{sleep},P}, & \text{if } P_P^{TX}=0 \end{cases} \qquad (7)$$

where $P_{\text{PBS},P}$, $P_{0,P}$, $P_{\text{sleep},P}$, and P_P^{TX} denote the total power consumption, power consumption at the minimum nonzero output power, power consumption in sleep mode, and RF transmit power at pico-BS P, respectively. The number of transceiver chains at the pico-BSs is denoted by $N_{TRX,P}$. The slope of the load-dependent power consumption at pico-BS is denoted by Δ_P. In Table I, we present the values of the linearized power consumption model parameters for MBSs and pico-BSs. Using (6) and (7), the power consumed in sector s can be expressed as

$$\psi_s(\beta_s,\boldsymbol{\beta}_{P,s}) = P_{\text{MBS}} + \sum_{P\in\mathcal{B}_{P,s}} P_{\text{PBS,P}} \qquad (8)$$

where $\boldsymbol{\beta}_{P,s}$ is the vector of all β parameters of all pico-BSs in sector s. The set of pico-BSs in sector s is denoted by $\mathcal{B}_{P,s}$.

Using the aggregate throughput and power consumption expressions in (3) and (8), the energy efficiency of sector s, in

TABLE I
SIMULATION PARAMETERS

Parameter	Setting
Channel bandwidth	10 MHz
Total number of RBs	50 RBs
Freq. selective channel model (CM)	Extended Typical Urban CM
UE to MBS PL model	$128.1 + 37.6\log_{10}(d)$
UE to pico-BS PL model	$140.7 + 36.7\log_{10}(d)$
Effective thermal noise power, N_0	-174 dBm/Hz
UE noise figures	9 dB
MBS and pico-BS antenna gain	14 dBi and 5 dBi
UE antenna gain	0 dBi
Antenna horizontal pattern, $A(\theta)$	$-\min(12(\theta/\theta_{3\,\mathrm{dB}})^2, A_m)$
A_m and $\theta_{3\,\mathrm{dB}}$	20 dB and 70°
Penetration loss	20 dB
Macrocell and picocell shadowing	8 dB and 10 dB
Inter-site distance	500 m
Minimum MBS to user distance	50 m
Minimum pico-BS to user distance	10 m
Minimum pico-BS to MBS distance	75 m
Minimum pico-BS to pico-BS distance	40 m
Traffic model	Full buffer
Power consumption parameters	MBS: (130 W, 75 W, 46 dBm, 4.7)
$(P_0, P_{\mathrm{sleep}}, P_{\mathrm{max}}, \Delta)$	Pico-BS: (56 W, 39 W, 30 dBm, 2.6)

bits/Joule, is defined as

$$\eta_s(\varepsilon_s, \beta_s, \boldsymbol{\beta}_{P,s}) = \frac{R_s}{\psi_s(\beta_s, \boldsymbol{\beta}_{P,s})}. \tag{9}$$

Note that the MBS transmissions on Band A determine the cross-tier interference for the cell-edge PUEs in the same sector as well as the intercell interference. However, as the downlink transmissions of the MBS and cell-center pico-BSs in the same sector are orthogonal, the MBS transmissions do not affect the throughput of cell-center PUEs in the same sector. In the following section, we formulate the energy efficiency maximization problem.

III. JOINT ENERGY-EFFICIENT RESOURCE ALLOCATION PROBLEM

In this section, we develop the framework for a utility-based resource allocation algorithm in which the objective is to maximize the energy efficiency while satisfying the rate requirement of users. The energy efficiency maximization problem can be formulated as

$$\max_{\mathbf{x},\boldsymbol{\beta}_P,\mathbf{C},\mathbf{F}} \sum_{s\in\mathcal{S}} \eta_s(\mathbf{x}_s, \boldsymbol{\beta}_{P,s})$$

$$\text{s.t. } \boldsymbol{C}^k_{M,s}\boldsymbol{F}^{(k,:)}_{M,s}\boldsymbol{R}^{(:,k)}_1 + (1-\boldsymbol{C}^k_{M,s})\boldsymbol{F}^{(k,:)}_{M,s}\boldsymbol{R}^{(:,k)}_2 \geq R_{\min,k}, \tag{10a}$$

$$\text{for all } k \in \mathcal{K}_{M,s}, s \in \mathcal{S}$$

$$\boldsymbol{C}^k_{P,s}\boldsymbol{F}^{(k,:)}_{P,s}\boldsymbol{R}^{(:,k)}_3 + (1-\boldsymbol{C}^k_{P,s})\boldsymbol{F}^{(k,:)}_{P,s}\boldsymbol{R}^{(:,k)}_4 \geq R_{\min,k},$$

$$\text{for all } k \in \mathcal{K}_{P,s}, s \in \mathcal{S} \tag{10b}$$

$$\sum_{k\in\mathcal{K}_{M,s}\boldsymbol{C}^k_{M,s}=1} \boldsymbol{F}^{(k,n)}_{M,s} = 1 \text{ and } \sum_{k\in\mathcal{K}_{M,s}\boldsymbol{C}^k_{M,s}=0} \boldsymbol{F}^{(k,n)}_{M,s} = 0 \tag{10c}$$

$$\text{for all } n \in \mathcal{N}^C_{M,s}, s \in \mathcal{S}$$

$$\sum_{k\in\mathcal{K}_{M,s}\boldsymbol{C}^k_{M,s}=0} \boldsymbol{F}^{(k,n)}_{M,s} = 1 \text{ and } \sum_{k\in\mathcal{K}_{M,s}\boldsymbol{C}^k_{M,s}=1} \boldsymbol{F}^{(k,n)}_{M,s} = 0 \tag{10d}$$

$$\text{for all } n \in \mathcal{N}^E_{M,s}, s \in \mathcal{S}$$

$$\sum_{k\in\mathcal{K}_{M,s}} \boldsymbol{F}^{(k,n)}_{M,s} = 0 \quad \text{for all } n \notin \mathcal{N}^C_{M,s} \cup \mathcal{N}^E_{M,s}, s \in \mathcal{S} \tag{10e}$$

$$\sum_{k\in\mathcal{K}^p_{P,s}\boldsymbol{C}^k_{P,s}=1} \boldsymbol{F}^{(k,n)}_{P,s} = 1 \text{ and } \sum_{k\in\mathcal{K}^p_{P,s}\boldsymbol{C}^k_{P,s}=0} \boldsymbol{F}^{(k,n)}_{P,s} = 0 \tag{10f}$$

$$\text{for all } n \in \mathcal{N}^C_{P,s}, s \in \mathcal{S}$$

$$\sum_{k\in\mathcal{K}^p_{P,s}\boldsymbol{C}^k_{P,s}=0} \boldsymbol{F}^{(k,n)}_{P,s} = 1 \sum_{k\in\mathcal{K}^p_{P,s}\boldsymbol{C}^k_{P,s}=1} \boldsymbol{F}^{(k,n)}_{P,s} = 0 \tag{10g}$$

$$\text{for all } n \in \mathcal{N}^E_{P,s}, s \in \mathcal{S}$$

$$\sum_{k\in\mathcal{K}^p_{P,s}} \boldsymbol{F}^{(k,n)}_{P,s} = 0 \quad \text{for all } n \notin \mathcal{N}^C_{P,s} \cup \mathcal{N}^E_{P,s}, s \in \mathcal{S} \tag{10h}$$

$$\boldsymbol{\varepsilon} \succeq \mathbf{0} \text{ and } \mathbf{0} \preceq \boldsymbol{\beta} \preceq \mathbf{1} \tag{10i}$$

$$0 \preceq \boldsymbol{\beta}_P \preceq \mathbf{1} \tag{10j}$$

where $R_{\min,k}$ is the minimum rate constraint of user k. The sets of subcarriers assigned to cell-center and cell-edge regions of MBS M in sector s are denoted by $\mathcal{N}^C_{M,s}$ and $\mathcal{N}^E_{M,s}$, respectively. Likewise, the sets of subcarriers assigned to cell-center and cell-edge regions of pico-BSs in sector s are denoted by $\mathcal{N}^C_{P,s}$ and $\mathcal{N}^E_{P,s}$, respectively. The set of sectors in the simulations area is $\mathcal{S}$. The set of PUEs that are associated with pico-BS P in sector s is denoted by $\mathcal{K}^p_{P,s}$. The vectors $\boldsymbol{\varepsilon}$ and $\boldsymbol{\beta}$ consist of ε_s and β_s for all sectors in the network. The vector $\boldsymbol{\beta}_P$ consists of $\boldsymbol{\beta}_{P,s}$ that denotes the power control variable for each pico-BSs in the sector s. The notation $\boldsymbol{\varepsilon} \succeq \mathbf{0}$ means that each element of $\boldsymbol{\varepsilon}$ is greater than or equal to 0. Throughout the rest of the paper, the vector $\mathbf{x}_s$ will be used for the (ε_s, β_s) couple and the vector $\mathbf{x}$ will be used for the vector couple $(\boldsymbol{\varepsilon}, \boldsymbol{\beta})$. Constraints (10a) and (10b) ensure that minimum rate constraints of users are satisfied. Constraints (10c) and (10d) guarantee that cell-center MBS resources assigned to cell-center MUEs and cell-edge MBS resources are assigned to cell-edge MUEs, respectively. Constraint (10e) ensures that MBSs do not assign any resources that are not available to themselves. Constraints (10f) and (10g) guarantee that cell-center pico-BS resources assigned to cell-center PUEs and cell-edge pico-BS resources assigned to cell-edge PUEs, respectively. Constraint (10h) ensures that pico-BSs do not assign any resources that are not available to themselves. Constraints (10i) and (10j) state that parameters $\boldsymbol{\varepsilon}$, $\boldsymbol{\beta}$, and $\boldsymbol{\beta}_{P,s}$ are within the given limits.

This resource allocation problem needs to be jointly solved over the cell-center radius, frequency, and power domains to obtain the optimum solution. The problem is combinatorial over the first two domains and is nonconvex over the power allocation domain [38], [39]. Therefore, obtaining the global solution to this problem requires an exhaustive search that is fairly impractical. For the power allocation problem, we have showed in [29] that for the same cell-center radius and frequency domain allocation, the energy efficiency function of a sector is quasi-concave over the power levels when the interference conditions are constant. To benefit from this property in the following section, we propose a three-stage algorithm in which the joint problem is divided into separate domains.

IV. Proposed Solution

Our formulation in (10) enables us to develop an energy-efficient resource allocation algorithm. Similar resource allocation algorithms have been studied in the literature, see, e.g., [13], [14], and of [3, Ch. 3]. In our proposed algorithm, we decouple the main problem into three subproblems. First, we start by selecting three candidate cell-center boundaries such that the cell-center and cell-edge MUEs are selected and the corresponding information is sent to the pico-BSs, identifying their regions and subbands. Second, we solve the frequency allocation problem for all candidate cell-center boundaries. Note that, the power control parameters that are obtained during the last time instant is used to calculate the Lagrangian function in this process. Then, among these candidate cell-center boundaries, the one that maximizes the Lagrangian function is selected. Last, we calculate the power control parameters for MBS and pico-BSs, consecutively. After each sector solves its power allocation subproblem, the interference prices and power levels are distributed among the network. By using this information, each sector first resets the cell-center radius, resolves the frequency allocation, and finally recalculates the power levels. The proposed algorithm is presented under the heading Algorithm 1. In the sequel, we discuss each stage of the proposed algorithm in detail.

Algorithm 1: Proposed Energy-Efficient Resource Allocation Algorithm.

1: **Initialize:** $r_{th,s}^{(0,c)} = r_{r,s}/2 \quad (\varepsilon_s^{(0)}, \beta_s^0, \boldsymbol{\beta}_{P,s}^{(0)}) = [1, 1, \mathbf{1}]$
2: $r_{th,s}^{(t,c)} = r_{th,s}^{(t-1,c)}$
3: **Stage 1:** Each sector determines $r_{th,s}^{(t,c-1)}$ and $r_{th,s}^{(t,c+1)}$ by using the cell-center radius selection algorithm in Section IV-A.
4: **for** $n := -1$ to 1 **do**
5: **Stage 2:** For cell-center radius, $r_{th,s}^{(t,c+n)}$, the frequency assignment algorithm is solved by using the algorithm in Section IV-B.
6: $\mathcal{L}_s^{(n)}$ which is described in Section IV-C is calculated for the given frequency assignment.
7: Note that dual variables and power control parameters in previous time instant are used during the Lagrangian function calculation.
8: **end for**
9: The cell-center radius and the frequency assignments that maximize the Lagrangian function are selected.
10: **Stage 3:** For this cell-center radius and frequency assignment, the power control parameters are determined by the proposed power control algorithm in Section IV-C.
11: Go to Step 2 and repeat.

A. Cell-Center Region Boundaries

The first stage of the proposed algorithm determines the cell-center region boundary of each sector. It is described in [27] that over 10% throughput improvement can be achieved by proper selection of the constant cell-center region boundaries. Further improvements can be obtained by selecting the boundaries dynamically in each sector. For single-layer networks, the cell-center boundary categorizes its users into the cell-center and cell-edge MUEs. MUEs can typically be distinguished into cell-center and cell-edge users based on the reference signal received power (RSRP) or reference signal received quality (RSRQ) measurements fed back from UEs to MBSs. The RSRP signal indicates the path loss between the serving base station and users, whereas the RSRQ measurement gives the ratio of the reference signal to the interference. Since it is described in [40] that the performance of both schemes are similar, in this paper, we also distinguish users into cell-center and cell-edge users based on the RSRP measurements. If the RSRP of a user is higher than a threshold, it is considered to be in the cell-center region, and vice versa.

For two-tier networks, the cell-center boundary also determines the available RBs for each pico-BS. For example, consider the multitier FFR scheme depicted in Fig. 1. If a pico-BS is located in the cell-center region, the number of available RBs for this pico-BS reduces. This frequency allocation alleviates the cross-tier interference between the cell-center MUEs and cell-edge PUEs. On the other hand, if a pico-BS is located in the cell-edge region, all subbands, except the subband that is used by cell-edge MUEs, are available to this pico-BS. Therefore, when a pico-BS is located in the cell-center region, the available RBs and correspondingly the total throughput of the pico-BS significantly decreases.

In two-tier networks, due to the high power difference of the reference signal between the MBSs and pico-BSs, PUEs are typically located close to the pico-BS [41]. For this reason, PUEs typically have low path loss values, or equivalently, high channel gains to their serving pico-BSs. When the number of subbands available to the pico-BS is decreased, it can be expected that the total throughput of PUEs will be reduced. However, due to the reduction in the cross-tier interference, the signal-to-interference-plus-noise ratio of the cell-center MUEs increases subsequently. In [29], we proposed two different cell-center selection algorithms. The first algorithm (CSSA2) is proposed to maximize the throughput and the energy efficiency of the network at the cost of fairness. On the other hand, the second one (CSSA3) maximizes fairness among users at the cost of throughput. In CSSA2, the MUE that is closest to the MBS is selected to be in the cell-center region, while the rest of the MUEs are in the cell-edge region. This algorithm achieves higher energy

efficiency and throughput while it leads to the starvation of some users, due to the lack of resources for cell-edge users. There is even the possibility of some users not getting any resources when the network is crowded. On the other hand, CSSA3 achieves significantly higher fairness by sacrificing the energy efficiency of the network. We will evaluate the performance of CSSA2 and CSSA3 in Section V. To benefit the advantage of both algorithms, we propose a new cell-center radius selection method (CSSA1) that dynamically updates the radius. In heterogeneous networks, shifting the cell-center radius is only effective if the new radius changes the region of an MUE or a pico-BS. Otherwise, the change will be useless. Therefore, CSSA1 compares the Lagrangian function of the previous cell-center radius with two cell-center radii: one more MUE or pico-BS is included in the cell-center region and one MUE or pico-BS is excluded from the cell-center region. The MUE or pico-BS that is going to be included in the cell-center region is the closest MUE or pico-BS to the cell-center radius and located in the cell-edge region for the previous cell-center radius. Likewise, the MUE or pico-BS that is going to be excluded from the cell-center region is the closest MUE or pico-BS to cell-center radius and located in the cell-center region. Note that RSRP measurements are used to determine the users or pico-BSs that are going to change their region. In addition, only the region of one MUE or pico-BS changes at each iteration. After that, CSSA1 selects the cell-center radius that maximizes the Lagrangian function among these three radii. In order to calculate the Lagrangian function, the frequency assignment problem needs to be solved. Therefore, the frequency assignment problem that is going to be explained in Section IV-B has to be solved for all these three radii. When the rate requirements of the users are small, CSSA1 shrinks the cell-center radius and improves the energy efficiency of the network. On the other hand, when the rate requirements of the users increase, a more fair cell-center radius selection helps satisfy the rate requirements of users.

B. Frequency Assignment Problem

In the second stage of the problem, we determine the frequency assignments. There are many scheduling methods discussed in the literature, see, e.g., [31], [42]. Each scheduler has an efficiency and fairness tradeoff. In general, both of these utilities cannot be increased at the same time. In the LTE radio protocol stack, scheduling is handled by the medium access control layer [31]. The scheduler in each base station is responsible from the distribution of frequency resources and this is left to the network operator for implementation. Green scheduling schemes have been surveyed in [42] and references therein. These schemes can significantly improve energy efficiency of the system and decrease the transmission powers. In this paper, we propose a novel scheduling algorithm to satisfy the rate requirements of users and to maximize the energy efficiency of the sector. For fairness purposes, we first assign one RB to each user. Then, the rest of the resources are assigned to users that provide the highest improvement to the Lagrangian function which is described in Section IV-C. When the rate requirements of the users are not satisfied, the dual prices significantly decrease the value of the Lagrangian function. Therefore, the proposed algorithm favors those users that could not satisfy the rate requirement with the current assignment. On the other hand, when the rate requirements of all users are satisfied, the proposed algorithm assigns RBs to the user that provides the highest increase to the Lagrangian function. This user is usually the one who has the best average channel gain among all users. We describe the scheduler algorithm next.

1) Lagrangian-Directed Scheduler (LDS): In the LDS, RBs are distributed among users to maximize the Lagrangian function. Note that, we run this algorithm twice in the MBSs, one for the cell-center MUEs and one for the cell-edge users, due to the fact that these users do not compete for the same RB sets. In LTE standards, the smallest granularity is RB. Therefore, we assign all subcarriers in one RB to one user during the scheduling process. In the proposed scheduling algorithm, the scheduler first assigns one RB to each user. The RB that has the best average channel gain among the available RBs is assigned to the user. After that, if there are still unassigned RBs, the proposed algorithm calculates the increase in the Lagrangian function for an RB and this RB is assigned to the user that provides the highest improvement to the Lagrangian function. This process continues until all RBs are assigned to the users.

In Section V, we will compare the proposed LDS with the equal bandwidth (EBW) scheduler and max–min fair (MMF) scheduler whose descriptions are given below.

2) EBW Scheduler: The EBW scheduler distributes RBs equally among users in the following manner. Assume that there are K users and N_{RB} RBs. These RBs are assigned to users such that $\lfloor N_{RB}/K \rfloor + 1$ RBs to $K_h = \mod(N_{RB}, K)$ users and $\lfloor N_{RB}/K \rfloor$ RBs to $K_i = K - K_h$ users.

3) MMF Scheduler: In the MMF scheduler, RBs are distributed among users to maximize the minimum throughput. We adopt the scheduler that has been described in [43]. The primary difference between these two algorithms is the scheduler granularity. The one described in [43] allocates subcarriers to users, however, the smallest granularity in the LTE standard is RB. Therefore, we need to adopt the algorithm given in [43] to work at RB level. In the MMF scheduler, each user is assigned the RB with the best average channel gain. We then remove this user from the user set and the corresponding RB from the RB set. This process continues until one of the sets is empty. If the user set is empty first, then the users are sorted according to their actual rates with current assignment. Then, the RB that has the best channel is assigned to the first user and users are sorted again according to updated rates. This process continues until all RBs are assigned.

C. Power Control Problem

The third stage of the proposed solution determines the power assignments such that the optimal power levels are assigned to each subband in order to maximize the energy efficiency and satisfy the rate requirements in the sector. Given the frequency assignments from the previous stage, it remains to solve this power control problem. As we discussed earlier, this problem is nonconvex over the power allocation subdomain and the solution requires exhaustive search over all power control parameters. However, we showed that energy efficiency function

$\eta_s(\mathbf{x}_s, \boldsymbol{\beta}_{P,s})$ is quasi-concave over the power control parameters ε_s and β_s in [29]. Therefore, when we divide the power allocation problem into $|S|$ subproblems such that each sector maximizes its own energy efficiency while satisfying the rate requirements of its users, where the number $|S|$ corresponds to the number of sectors in the network, each problem is quasi-concave and has a unique maximum over the the power control parameters ε_s and β_s. In the maximization problem, the received interference is assumed to be constant. In addition, the region of the users in the sector and frequency assignment of the users have to be determined before calculating the power levels. After these steps, by using convex optimization techniques, the optimal β_s, ε_s, and $\boldsymbol{\beta}_{P,s}$ that maximize the energy efficiency of the sector s and satisfy the rate constraints of users can be obtained.

The pico-BSs are expected to become significantly dense to meet the increasing rate demands [44]. Therefore, updating the power control parameters of MBSs and all pico-BSs in the sector concurrently requires significant data exchange between base stations. This may create congestion in the backhaul network. In addition, this process requires significant computation time and may not be obtained in real time especially in denser networks. Therefore, we split the power control problem into two subproblems. In the first part of the algorithm, MBS determines optimum β_s and ε_s. During this process, the power control parameters of pico-BSs, $\boldsymbol{\beta}_{P,s}$, are assumed to be constant. In the second part of the process, pico-BSs concurrently calculate their power control parameters for the determined β_s and ε_s in the first part. Note that each pico-BS assumes that the transmission power level of the other pico-BSs in the sector is not changing during this process.

1) MBS Power Control Problem: In the first part, MBSs determine the optimum power control parameters in each sector. Due to the fact that each MBS tries to satisfy the rate requirement of its users, the transmission power of the MBSs may increase imprudently. This not only reduces the energy efficiency of the network, but also augments the intercell interference and elevates the outage probabilities of the users in the other sectors. Therefore, we study a pricing mechanism to account for the interference caused to the users associated with other base stations. This pricing mechanism reduces the intercell interference by giving incentives to MBSs for decreasing their transmit powers.

The pricing function penalizes the utility of an MBS based on the interference it creates. If the interference leads to outage of the users in other sector, the penalty factor increases. Let $\theta_s(\mathbf{x}_s)$ denote the pricing function of sector s. The energy efficiency maximization problem with pricing function per sector can be formulated as

$$\begin{aligned}
\max_{\mathbf{x}_s}\ & \eta_s(\mathbf{x}_s, \boldsymbol{\beta}_{\boldsymbol{P,s}}) - \theta_s(\mathbf{x}_s) \\
\text{s.t. }\ & \boldsymbol{C}^k_{M,s}\boldsymbol{F}^{(k,:)}_{M,s}\boldsymbol{R}^{(:,k)}_1 + (1-\boldsymbol{C}^k_{M,s})\boldsymbol{F}^{(k,:)}_{M,s}\boldsymbol{R}^{(:,k)}_2 \geq R_{\min,k}, \\
& \quad \text{for all } k \in \mathcal{K}_{M,s} \\
& \boldsymbol{C}^k_{P,s}\boldsymbol{F}^{(k,:)}_{P,s}\boldsymbol{R}^{(:,k)}_3 + (1-\boldsymbol{C}^k_{P,s})\boldsymbol{F}^{(k,:)}_{P,s}\boldsymbol{R}^{(:,k)}_4 \geq R_{\min,k}, \\
& \quad \text{for all } k \in \mathcal{K}_{P,s} \\
& \varepsilon_s \geq 0 \text{ and } 0 \leq \beta_s \leq 1.
\end{aligned} \tag{11}$$

Several pricing functions have been proposed in the literature. For example, in [7] and [8], Saraydar *et al.* propose to use $\theta_s(\mathbf{x}_s) = c_s\beta_s P_M^{\max}$, where c_s is a constant. The cost function proposed in [7] and [8] penalizes the utility of an MBS with the total amount of power it transmits. In this paper, we pursue an alternative approach and penalize the interference that an MBS creates. This type of pricing function was first proposed in [4]–[6]. In this approach, the pricing function is defined as

$$\theta_s(\mathbf{x}) = \sum_{n\in\mathcal{N}_M} p_M^{(n)} \sum_{l\neq k, l\in\mathcal{K}^{(n)}} \frac{\partial \eta_{s'}}{\partial p_M^{(n)}} \tag{12}$$

where the transmit power of MBS M on subcarrier n is denoted by $p_M^{(n)}$. The set of subcarriers that MBS M allocates for the cell-center and cell-edge regions is denoted by $\mathcal{N}_M$. Let $\mathcal{K}^{(n)}$ denote the set of users that are assigned to subcarrier n. Then, the set of users that MBS m interferes on subcarrier n are represented by $l \in \mathcal{K}^{(n)}$. Let a user l be in sector s', then the energy efficiency of the sector s' is denoted by $\eta_{s'}$. The term $\partial\eta_{s'}/\partial p_M^{(n)}$ denotes the derivative of energy efficiency of sector s' with respect to the transmit power of MBS M of sector s. Thus, the penalty function (12) characterizes the marginal change in the utility of a neighboring sector s' per unit power change in MBS M of sector s. In addition, this approach prevents the base stations to increase their transmission power to very high levels when the minimum rate requirement of one or more of their users cannot be satisfied without causing outage of the other users. In LTE standards, the X2 interface provides a fast and reliable backhaul link between base stations [31]. In this paper, we use this interface for three reasons: First, pico-BSs send the CSI for their users to MBS where these are processed. Second, the CSI of users is distributed among MBSs. Third, for the pricing method, the same interface distributes the interference prices between MBSs. The Lagrangian of the problem in (11) can be written as

$$\begin{aligned}
\mathcal{L}(\mathbf{x}_s, \boldsymbol{\beta}_{P,s}, \boldsymbol{\lambda}, \nu_s, \tau_s, \rho_s) &= \eta_s(\mathbf{x}_s, \boldsymbol{\beta}_{P,s}) - \theta_s(\mathbf{x}_s) \\
&- \sum_{k\in\mathcal{K}_{M,s}} \lambda_{k,s}\left(R_{\min,k} - \left(\boldsymbol{C}^k_{M,s}\boldsymbol{F}^{(k,:)}_{M,s}\boldsymbol{R}^{(:,k)}_1 + (1-\boldsymbol{C}^k_{M,s})\right.\right. \\
&\left.\left.\boldsymbol{F}^{(k,:)}_{M,s}\boldsymbol{R}^{(:,k)}_2\right)\right) - \sum_{k\in\mathcal{K}_{P,s}} \lambda_{k,s}\left(R_{\min,k} - \left(\boldsymbol{C}^k_{P,s}\boldsymbol{F}^{(k,:)}_{P,s}\boldsymbol{R}^{(:,k)}_3\right.\right. \\
&\left.\left. + (1-\boldsymbol{C}^k_{P,s})\boldsymbol{F}^{(k,:)}_{P,s}\boldsymbol{R}^{(:,k)}_4\right)\right) + \nu_s\beta_s + \tau_s(1-\beta_s) + \rho_s\varepsilon_s.
\end{aligned} \tag{13}$$

For simplicity, we will use $\mathcal{L}_s$ for $\mathcal{L}(\mathbf{x}_s, \boldsymbol{\beta}_{P,s}, \boldsymbol{\lambda}, \nu_s, \tau_s, \rho_s)$ throughout the rest of the paper.

Since the downlink transmissions of an MBS M in sector s, $p_M^{(n)}$, are characterized by ε_s and β_s, we need to adopt (12). In addition, we need to include the effect of interference to the rate constraints of users in other sectors. Therefore, the pricing

function can be written as

$$\theta_s(\mathbf{x}_s) = \mathbf{x}_s^T \sum_{\substack{s' \in \mathcal{S} \\ s' \neq s}} \Bigg(\nabla_{\mathbf{x}_s} \eta_{s'}(\mathbf{x}_{s'}, \boldsymbol{\beta}_{P,s'}) + \sum_{k \in \mathcal{K}_{M,s'}} \lambda_{k,s'} \Big(\boldsymbol{C}_{M,s'}^k \boldsymbol{F}_{M,s'}^{(k,:)} \boldsymbol{R}_1^{(:,k)} + (1 - \boldsymbol{C}_{M,s'}^k) \boldsymbol{F}_{M,s'}^{(k,:)} \boldsymbol{R}_2^{(:,k)} \Big) + \sum_{k \in \mathcal{K}_{P,s'}} \lambda_{k,s'} \Big(\boldsymbol{C}_{P,s'}^k \boldsymbol{F}_{P,s'}^{(k,:)} \boldsymbol{R}_3^{(:,k)} + (1 - \boldsymbol{C}_{P,s'}^k) \boldsymbol{F}_{P,s'}^{(k,:)} \boldsymbol{R}_4^{(:,k)} \Big) \Bigg). \tag{14}$$

Hence, the pricing function in (14) reflects the marginal costs of the variables ε and β.

For the solution, we will employ the Levenberg–Marquardt method. The Levenberg–Marquardt method is a modification to the Newton method. The Newton method premultiplies the gradient ascent direction by the inverse of the Hessian matrix. The motivation of the Newton method is to find a suitable direction based on the quadratic approximation of a function, whereas the gradient ascent method seeks to find a linear approximation of a function. Consider the Lagrangian function in (13). Its quadratic approximation evaluated at $\mathbf{y}_s^{(l)} = (\varepsilon_s^{(l)} \beta_s^{(l)})^T$ can be expressed as

$$g(\mathbf{y}) = \mathcal{L}_s + \nabla \mathcal{L}_s^T (\mathbf{y} - \mathbf{y}_s^{(l)}) + \frac{1}{2} (\mathbf{y} - \mathbf{y}_s^{(l)})^T \nabla^2 \mathcal{L}_s (\mathbf{y} - \mathbf{y}_s^{(l)}), \tag{15}$$

where $\nabla^2 \mathcal{L}_s$ is the Hessian matrix of $\mathcal{L}$ evaluated at $\mathbf{y}_s^{(l)}$. Note that we are going to use $\mathbf{y}_s^{(l)}$ for $(\varepsilon_s^{(l)} \beta_s^{(l)})$ pair for the Newton iteration l and the parameter $\mathbf{x}_s^{(t)}$ will be used for the same pair at time instant t. The parameter updates that maximize $g(\mathbf{y})$ are given by

$$\mathbf{y}_s^{(l+1)} = \mathbf{y}_s^{(l)} - \mu_l (\nabla^2 \mathcal{L}_s^{(l)})^{-1} \nabla \mathcal{L}_s^{(l)} \tag{16}$$

where the Newton search direction is $\mathbf{d}_l^N = -(\nabla^2 \mathcal{L}_s^{(l)})^{-1} \nabla \mathcal{L}_s^{(l)}$. In general, convergence of the Newton method is not guaranteed [45]. This is due to the fact that the Hessian can be singular or the search direction, $\mathbf{d}_l^N$, may not even give an ascent direction. Therefore, even when the inverse of the Hessian matrix exists, it does not necessarily imply that $\mathcal{L}_s^{(l+1)}$ is greater than $\mathcal{L}_s^{(l)}$. However, when the starting point $\mathbf{y}_s^{(0)}$ is close enough to the optimal $\mathbf{y}^*$ such that $\nabla \mathcal{L}_s^{(l)*} = 0$ and $\nabla^2 \mathcal{L}_s^{(l)*}$ is full rank, then the Newton method converges to the optimal $\mathbf{y}^*$ [45], [46]. In order to address the convergence problem of the Newton method, several methods have been proposed in the literature, see [45, Ch. 8] and [46, Ch. 5.2.4]. In this paper, we employ the Levenberg–Marquardt method due to its guarantee of convergence. With the Levenberg–Marquardt method, the parameter updates are given by

$$\mathbf{y}_s^{(l+1)} = \mathbf{y}_s^{(l)} - \mu_l (\nabla^2 \mathcal{L}_s^{(l)} - \xi \mathbf{I})^{-1} \nabla \mathcal{L}_s^{(l)} \tag{17}$$

where $\mathbf{d}_l^{LM} = -(\nabla^2 \mathcal{L}_s^{(l)} - \xi \mathbf{I})^{-1} \nabla \mathcal{L}_s^{(l)}$ is the search direction evaluated at $\mathbf{y}_s^{(l)}$ and $\mathbf{I}$ is the identity matrix. The constant ξ ensures all the eigenvalues of $\mathbf{D} = (\nabla^2 \mathcal{L}_s^{(l)} - \xi \mathbf{I})$ are negative such that $\mathbf{D}$ is negative definite. It is called as the damping or the Levenberg–Marquardt parameter [45]. If the largest eigenvalue of $\nabla^2 \mathcal{L}_s^{(l)}$ is negative, then ξ will be equal to zero and the Levenberg–Marquardt method reduces to the Newton method, that is $\mathbf{d}_{(l)}^{LM} = \mathbf{d}_{(l)}^N$. Under this condition, quadratic convergence is achieved. If the largest eigenvalue of the Hessian is non-negative, then we take $\xi = \omega_{\max} + \sigma$, where $\omega_{\max}$ is the largest eigenvalue of $\nabla^2 \mathcal{L}_s^{(l)}$ and $\sigma > 0$ is a sufficiently large number. This operation forces the Hessian to be negative definite. In our simulations, this offset is taken as $\sigma = 1$. The proposed approach using the interference pricing method is depicted under the heading Algorithm 2, where $l_{\max}$ is the maximum number of iterations and ϵ is a sufficiently small positive number to determine when to exit the algorithm. The parameter $\alpha_{k,s}$ in Step 10 is the positive scalar step size. In addition, the controlled increase mechanism in Step 15 is used to update the power levels. When the power level of the MBSs changes largely between two consecutive time instants, the interference pricing mechanism does not accurately model the effect of the interference over the utilities of the other sectors [5]. The controlled increase mechanism in Step 15 prevents large changes of the power levels. In this step, the selection of small ζ slows down the convergence of the algorithm. On the other hand, large ζ may cause large changes in the power levels. Therefore, an adaptive algorithm is used to select ζ in this paper. The parame-

Algorithm 2: Proposed Power Control Algorithm with Pricing.

1: % Each sector solves (11) by using the Levenberg-Marquardt Method
2: **for** $l := 1$ to $l_{\max}$ **do**
3: **if** $\omega_{\max} = \max(\text{eig}(\nabla_{\mathbf{y}}^2 \mathcal{L}_s^{(l)})) < 0$ **then**
4: $\xi = 0$.
5: **else**
6: $\xi = \omega_{\max} + \sigma$.
7: **end if**
8: $\mathbf{d}_l^{LM} = -(\nabla^2 \mathcal{L}_s^{(l)} - \xi \mathbf{I})^{-1} \nabla \mathcal{L}_s^{(l)}$.
9: Update the power control parameters, $\mathbf{y}_s^{(l+1)}$, using
$$\mathbf{y}_s^{(l+1)} = \mathbf{y}_s^{(l)} + \mu_l \mathbf{d}_l^{LM},$$
10: Update the Lagrange multiplier, $\lambda_{k,s}^{(l+1)}$ for all $k \in \mathcal{K}_{M,s}^C$ and $\mathcal{K}_{M,s}^E$, using
$$\lambda_{k,s}^{(l+1)} = \Big[\lambda_{k,s}^{(l)} + \alpha_{k,s} \Big(R_{\min,k} - (\boldsymbol{C}_{M,s}^k \boldsymbol{F}_{M,s}^{(k,:)} \boldsymbol{R}_1^{(:,k)} + (1 - \boldsymbol{C}_{M,s}^k) \boldsymbol{F}_{M,s}^{(k,:)} \boldsymbol{R}_2^{(:,k)}) \Big) \Big]^+.$$
11: **if** $\left| \nabla \mathcal{L}_s^T \mathbf{d}_l^{LM} \right| \leq \epsilon$ **then**
12: **Break**
13: **end if**
14: **end for**
15: $\mathbf{x}_s^{(t+1)} = (1 - \zeta) \mathbf{x}_s^{(t)} + \zeta \mathbf{y}_s^{(l)}$
16: *Price Update:* Each user calculates interference prices and feeds these values back to its base station.
17: Interference prices are distributed among base stations.

ter ζ is equal to $t/(2t+1)$, where t is the time instant [3]. When t goes to infinity ζ will converge to $1/2$.

Note that the Levenberg–Marquardt method guarantees convergence regardless of the starting point [45, p. 312]. Using the Levenberg–Marquardt method, the parameters are updated in Step 9 of Algorithm 2. The expression $\sqrt{(\nabla\mathcal{L}_s^{(l)})^T\mathbf{d}_l^{LM}}$ is called as the Newton increment [47]. It is used as a stopping criterion in iterative line search algorithms [47]. Note that this stopping criterion is also the condition to check whether a search direction $\mathbf{d}_l^{LM}$ is an ascent direction or not, that is, to check if $\nabla\mathcal{L}_s^{(l)T}\mathbf{d}_l^{LM} > 0$ is true. The loop terminates when the convergence condition is satisfied or the maximum number of iterations is reached.

In general, the main computational effort in Algorithm 2 is at Step 3 and Step 8 where the eigenvalues of matrix $\nabla^2\mathcal{L}_s^{(l)}$ and the inverse of matrix $\nabla^2\mathcal{L}_s^{(l)} - \xi\mathbf{I}$ are calculated [45]. Using classical approaches such as the Gauss–Jordan elimination method, the computational complexity of calculating eigenvalues and taking the inverse of an $n \times n$ matrix is $\mathcal{O}(n^3)$. For large-scale problems, this operation becomes prohibitively complex. Consider the centralized resource allocation approach where we solve for 57 (ε, β) pairs, one pair for each sector. Taking the inverse of this big matrix would require $\mathcal{O}(114^3)$ floating point operations (flops). Fortunately, with the proposed distributed algorithm, we only need to calculate the eigenvalues of $\nabla^2\mathcal{L}_s^{(l)}$ and take the inverse of $\nabla^2\mathcal{L}_s^{(l)} - \xi\mathbf{I}$, which are both 2×2 matrices, and these calculations are straightforward. The computational complexity of finding the maximum eigenvalue and updating power control parameters is $\mathcal{O}(n)$. Those steps require significantly less amount of time. The computational complexity of the updating the Lagrange multipliers is independent from the number of power control parameters and it depends on the number of users in the sector. Under these conditions, the computational complexity of both the gradient and Levenberg–Marquardt methods are on the same order per iteration step. In addition, due to the distributed nature of the algorithm, increasing the number of sectors in the simulation area does not change the required time due to the parallel processing. However, the total required time increases linearly due to the fact that more sectors run the proposed algorithm. Along with similar computational complexity, the Levenberg–Marquardt based-method has significantly faster convergence rate compared to the gradient-based method. The convergence properties of the algorithm are inherited from the detailed analysis in [29]. Also, note that the expressions of the gradient and Hessian of the energy efficiency function η_s are presented in detail in the [29, Appendix].

2) Picocell Power Control Problem: In the second part of the power problem, each pico-BS determines its own power control parameter separately. When the power control parameters for MBS are selected, the problem will be similar to the MBS power control problem. However, instead of two power control parameters, we only need to determine one power control parameter for each pico-BS. During this process, we assume that all power control parameters of other pico-BSs in sector s are constant. Then, the energy efficiency maximization problem with pricing function for pico-BS P can be formulated as

$$\begin{aligned}
\max_{\beta_{P,s}} \quad & \eta_s(\mathbf{x}_s, \boldsymbol{\beta}_{P,s}) - \theta_s(\beta_{P,s}) \\
\text{s.t.} \quad & \boldsymbol{C}_{M,s}^k \boldsymbol{F}_{M,s}^{(k,:)} \boldsymbol{R}_1^{(:,k)} + (1 - \boldsymbol{C}_{M,s}^k)\boldsymbol{F}_{M,s}^{(k,:)} \boldsymbol{R}_2^{(:,k)} \geq R_{\min,k}, \\
& \quad \text{for all } k \in \mathcal{K}_{M,s} \\
& \boldsymbol{C}_{P,s}^k \boldsymbol{F}_{P,s}^{(k,:)} \boldsymbol{R}_3^{(:,k)} + (1 - \boldsymbol{C}_{P,s}^k)\boldsymbol{F}_{P,s}^{(k,:)} \boldsymbol{R}_4^{(:,k)} \geq R_{\min,k}, \\
& \quad \text{for all } k \in \mathcal{K}_{P,s} \\
& 0 \leq \beta_{P,s} \leq 1.
\end{aligned} \tag{18}$$

The pricing function $\theta_s(\beta_{P,s})$ will be similar to the one in (14) except the derivatives are calculated with respect to $\beta_{P,s}$. Therefore, we do not repeat it here for simplicity.

The Lagrangian of the problem in (18) can be written as

$$\begin{aligned}
\mathcal{L}_s = \eta_s(\mathbf{x}_s, \boldsymbol{\beta}_{P,s}) - \theta_s(\beta_{P,s}) - \sum_{k \in \mathcal{K}_{M,s}} \lambda_{k,s} \Big(& R_{\min,k} \\
& - \Big(\boldsymbol{C}_{M,s}^k \boldsymbol{F}_{M,s}^{(k,:)} \boldsymbol{R}_1^{(:,k)} + (1 - \boldsymbol{C}_{M,s}^k)\boldsymbol{F}_{M,s}^{(k,:)} \boldsymbol{R}_2^{(:,k)} \Big) \Big) \\
- \sum_{k \in \mathcal{K}_{P,s}} \lambda_{k,s} \Big(& R_{\min,k} - \Big(\boldsymbol{C}_{P,s}^k \boldsymbol{F}_{P,s}^{(k,:)} \boldsymbol{R}_3^{(:,k)} \\
& + (1 - \boldsymbol{C}_{P,s}^k)\boldsymbol{F}_{P,s}^{(k,:)} \boldsymbol{R}_4^{(:,k)} \Big) \Big) + \nu_{P,s}\beta_{P,s} + \tau_s(1 - \beta_{P,s}).
\end{aligned} \tag{19}$$

The optimum $\beta_{P,s}$ can be calculated by using the Levenberg–Marquardt method based algorithm similar to the algorithm that is described in Section IV-C1. In Section IV-C1, the algorithm determines two power control parameters concurrently. However, we only determine one power control parameter in each pico-BS. Therefore, the complexity of this algorithm is lower than the one in Section IV-C1. In addition, this process is done concurrently at all pico-BSs in sector s and increasing the number of pico-BSs in sector does not affect the required time for this calculation.

In this paper, we are maximizing the energy efficiency of the network while satisfying the rate constraints of users. This particular approach is beneficial for applications such as Voice-over-IP, video call, streaming, and real-time gaming applications that require minimum rate to perform properly. As we will show next in Section V, when we enforce higher rates, the energy efficiency of the network may decrease.

V. Simulation Results

In this section, first, we show the convergence behavior of the Levenberg–Marquardt method at the first time instant. Second, we evaluate the performance of the proposed algorithm in terms of energy efficiency, outage probability, and power consumption for different rate constraints. Third, we compare the performance of the dynamic cell-center radius with constant cell-center region boundary selection algorithms. Fourth, we assess the effect of the scheduler on the energy efficiency and the outage probabilities. Fifth, we study the efficiency of the

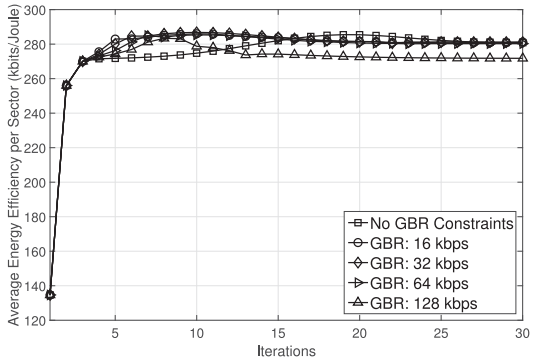


Fig. 3. Average energy efficiency of the network at the first time instant for lower GBR requirements of users.

proposed algorithm by comparing it with an exhaustive search algorithm over all possible cell-center radii and power levels.

For the FFR method, the spectrum allocation scheme is such that 14 RBs are assigned to subband A. The remaining 36 RBs are divided into 3 equal segments and assigned to subbands B, C, and D. The simulation layout is illustrated in Fig. 1. The simulation area consists of 19 hexagonal cells with wrap-around edges. Single antenna transmission is considered, i.e., $N_{TRX,M} = 1$ and $N_{TRX,P} = 1$ for all sectors. Two pico-BSs are randomly generated in each sector. Although we present the results for two pico-BS case, the proposed algorithm can easily be implemented for a network that has more pico-BSs per sector. A total of 20 users are generated in each sector. First, two users are generated within the radius of 40 m for each pico-BS. Then, the rest of the users are uniformly distributed in the sector area. The highest RSRP method is used for the cell association [41]. The simulation parameters, distance constraints in generating new nodes, and the base station power consumption values are given in Table I [48]. Furthermore, the initial values are chosen as $(\boldsymbol{\varepsilon}^{(0)}, \boldsymbol{\beta}^{(0)}, \boldsymbol{\beta}_P^{(0)}) = (\mathbf{1}, \mathbf{1}, \mathbf{1})$.

In the first part of the simulations, we assume that all MUEs in the network have the same guaranteed-bit-rate (GBR) requirements, i.e., $R_{\min,k} = R_{\min}$, for all $k \in \mathcal{K}_{M,s} \cup \mathcal{K}_{P,s}$ and $s \in \mathcal{S}$. Six different rate constraints are considered, $R_{\min}$ is equal to 16, 32, 64, 128, 256, and 512 kb/s (kbps). In the second part of the simulations, we compare the performance of the proposed dynamic cell-center radius selection algorithm with the ones that are described in [29]. In the third part of the simulations, we study the importance of the scheduler. The LDS, the EBW scheduler, and the MMF scheduler described in Section IV-B are considered for no rate constraint and various rate constraint cases.

Fig. 3 depicts the average energy efficiency function of the network for each iteration at first time instant when $R_{\min}$ is equal to 0, 16, 32, 64, and 128 kbps. The average energy efficiency of the sector increases significantly at the first iteration, then the Levenberg–Marquardt method based updates the power control parameters every iteration until the convergence. Convergence behavior is independent from the rate requirements of users and it took around 20–25 iterations.

Fig. 4(a) shows the energy efficiency of the network for the proposed algorithm when $R_{\min}$ is equal to 0, 16, 32, 64, and

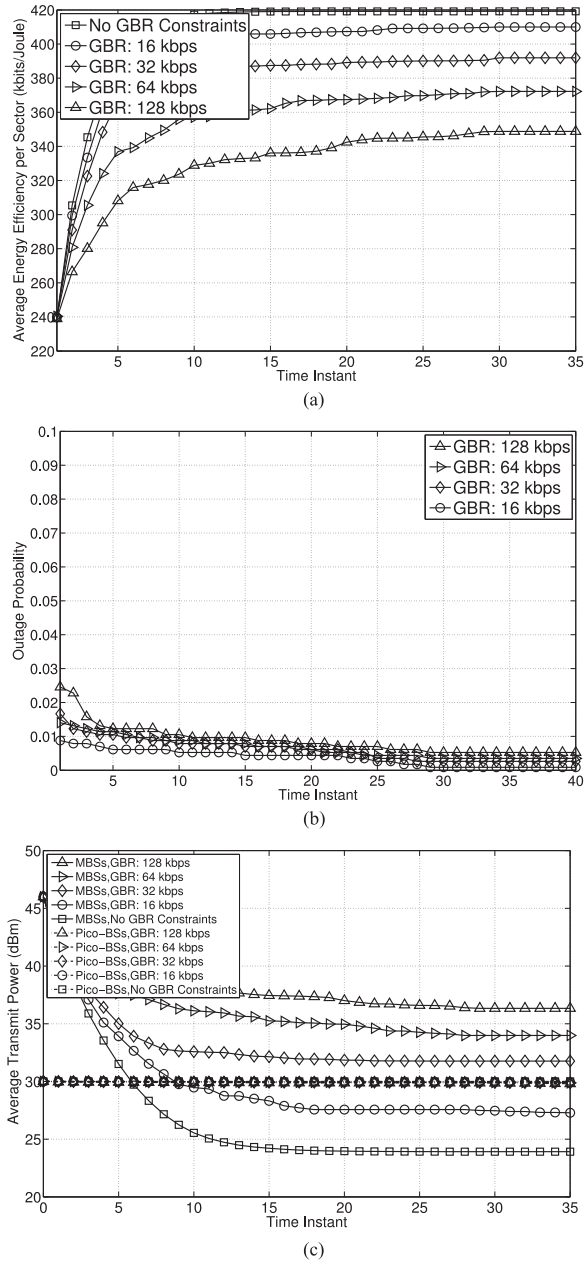


Fig. 4. (a) Average energy efficiency per sector, (b) the outage probabilities, and (c) average transmission power for lower GBR requirements of users.

128 kbps. Note that in this range of $R_{\min}$ values, the algorithm uniformly improves the energy efficiency of the system. Not surprisingly, the improvement decreases as $R_{\min}$ increases. For the case without rate constraints, the energy efficiency of the network is 2.2%, 6.9%, 12.6%, and 20.2% higher than the cases with $R_{\min}$ is equal to 16, 32, 64, and 128 kbps, respectively. For the lower rate constraints, the reduction in energy efficiency is marginal due to the fact that the rate constraints of most of the users are already satisfied at the energy-efficient optimum power level. Therefore, only small changes are required to satisfy the rate constraints of the rest of the users. In Fig. 4(b), the outage probabilities of the network using the proposed algorithm are shown for the same rate constraints. A user is assumed to be

in outage if the actual rate of the user is less than its rate requirement. The outage probabilities are very low for these rate constraints, even for the early time instances that the transmission power levels of the base stations have not converged. The highest outage probability is 0.5% when the rate requirement is 128 kbps. The efficiency of the proposed algorithm in terms of outage will become clear in the sequel when higher rate constraints are forced. Fig. 4(c) illustrates the power savings of the proposed algorithm. The average transmission power of the MBSs decreases every time instant until the convergence. Due to the fact that the rate constraints of the users are low, the proposed algorithm provides significant power savings. The average transmission power decreases more than 20 dB when users do not have any rate requirements. When the rate constraints are enforced, it leads to higher average transmission power and that is the main reason of the smaller gain in energy efficiency by using the algorithm. On the other hand, pico-BSs always transmit at the full power for the given rate constraints. Due to the fact that the interference from the pico-BSs to MBS does not cause outages of the users and pico-BSs reach their most energy-efficient case when they are transmitting at the full power, the pico-BSs transmit at the full power.

In Fig. 5(a), the energy efficiency of the network is illustrated for higher rate constraints, 256 and 512 kbps. Different from the previous cases, the average energy efficiency of the network now decreases in each time instant. The energy efficiency of the network drops 4.9% and 27.3% from the first time instant to the last time instant when $R_{\min}$ is equal to 256 and 512 kbps, respectively. The reason behind that is threefold. First, MBSs increase their transmission power to energy-inefficient levels in order to satisfy the higher minimum rate requirements of users. Second, the increased transmission power elevates the interference to the users that are in other sectors and using the same subchannels. Due to the elevated interference, base stations increase power levels to be able to support the minimum rate requirements, which gives rise to more interference and a chain reaction starts. Third, pico-BSs decrease their transmission power in order to decrease the interference to MBSs. The result is the reduction in energy efficiency manifested in Fig. 5(a). In Fig. 5(b), the outage probabilities of MUEs are shown for the higher rate constraints. As expected, the outage probability of users increases for the higher rate requirements as compared to lower rates shown in Fig. 4(b). At the first iteration, the outage probabilities are 7.9% and 27.0% for GBR requirements 256 and 512 kbps, respectively. After 40 time instants, the outage probabilities become 1.8% and 12.4% for the GBR requirements in the same order as before. When the rate requirement of users increases, energy-efficient power levels cause significant outages and that needs to be adjusted to decrease outage probabilities to tolerable levels. If we continue to increase the rate requirements of users, that requires further sacrifice from the energy efficiency of the network. This points out to the tradeoff between the network energy efficiency and outage probability, which becomes more severe at higher GBR requirements. Nevertheless, it is worthwhile to state that while an outage probability of 27.0% is unacceptable, 12.4% can be acceptable, pointing out to the fact that although the algorithm does not improve energy efficiency for these high

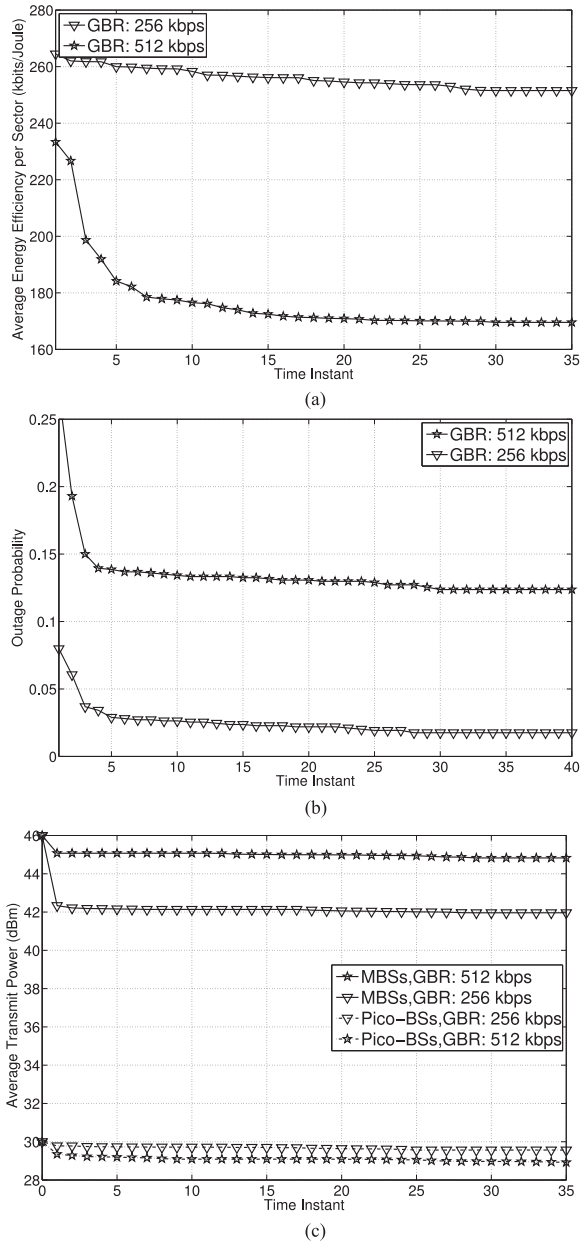


Fig. 5. (a) Average energy efficiency per sector (b) the outage probabilities, and (c) average transmission power for higher GBR requirements of users.

rates, it is useful for a substantial reduction in outage probability. In Fig. 5(c), the average transmission power of base stations is shown. When higher rate constraints are enforced, the average transmission power of MBSs reaches the lowest level at the first time instant, and then it increases each time instant until the convergence. In order to satisfy the rate requirements of the users, MBSs need to transmit at higher levels compared to lower rate cases. Therefore, compared to Fig. 4(c), the power savings of the network is significantly lower. For example, when the GBR requirement is equal to 256 kbps, the power saving of the MBSs is less than 5 dB. The power savings are even lower when the rate constraints are 512 kbps. The decline in the power savings is the main cause of the decrease in energy efficiency. Yet, it is worthwhile to note that there is a gain in transmitted

TABLE II
OUTAGE PROFILE OF THE DIFFERENT USER TYPES

	Nr. of users (Outage probability)		
Minimum GBR (kbps)	Cell-center MUE	Cell-edge MUE	PUE
-	348 (0%)	616 (0%)	176 (0%)
16	358 (0%)	606 (0.2%)	176 (0%)
32	366 (0%)	598 (0.5%)	176 (0%)
64	396 (0.2%)	568 (0.5%)	176 (0%)
128	447 (0.4%)	517 (0.8%)	176 (0%)
256	482 (1.7%)	482 (2.5%)	176 (0.5%)
512	517 (10%)	447 (15.2%)	176 (12%)

power, although it does not translate into improved energy efficiency. The transmission power of the pico-BSs decreases when the higher rate requirement constraints are enforced. When the rate requirements of the users increase, some of the MUEs will be in outage. In order to satisfy the rate requirement of these MUEs, pico-BSs decrease their transmission powers and consequently the interference they create to these users. For example, when the GBR requirement is equal to 512 kbps, the pico-BSs decrease their transmission power more than 1 dB.

In Table II, we further investigate the relationship between the outage probability and cell-center region boundaries. The users in the network are divided into three categories, cell-center MUEs, cell-edge MUEs, and PUEs. Based on the minimum rate requirements, the cell-center regions are dynamically adjusted. As we increase the GBR requirements, we observe that the number of cell-center MUEs increases. As discussed in Section IV-A, the proposed cell-center radius selection algorithm behaves similar to CSSA3 to share resources more fairly among users. Although the FFR scheme decreases the interference that cell-edge MUEs suffer from, due to the fact that they are located away from the MBS, they encounter higher outage probabilities. When the minimum rate requirement of users is less than 256 kbps, all PUEs in the network satisfy their rate requirement. However, when the rate requirement is 512 kbps, 12% of PUEs cannot satisfy their rate requirement. As shown in Fig. 5(c), the transmission power of the pico-BSs decreases when the rate requirement is 512 kbps in order to decrease the interference on cell-center MUEs. Therefore, 12% of the PUEs is in outage when GBR is 512 kbps.

In Fig. 6(a) and (b), we investigate the effect of the cell-center radius selection algorithms in terms of energy efficiency and outage probabilities. The proposed cell-center radius selection algorithm (CCSA1) is compared with the ones that are proposed in [29]. Our simulation results show that CSSA1 performs significantly better than CSSA3 in terms of energy efficiency when the rate constraints are below 128 kbps. For example, when the GBR requirement is equal to 16 kbps, CSSA1 performs 16% better than CSSA3 in terms of energy efficiency. CSSA1 converges to CSSA3 to meet the rate constraints of more users when the rate constraints become aggressive. For example, when the rate constraints are 512 kbps, the selected cell-center radii are the same for these two algorithms for most of the sectors in the network. On the other hand, CSSA1 performs worse than

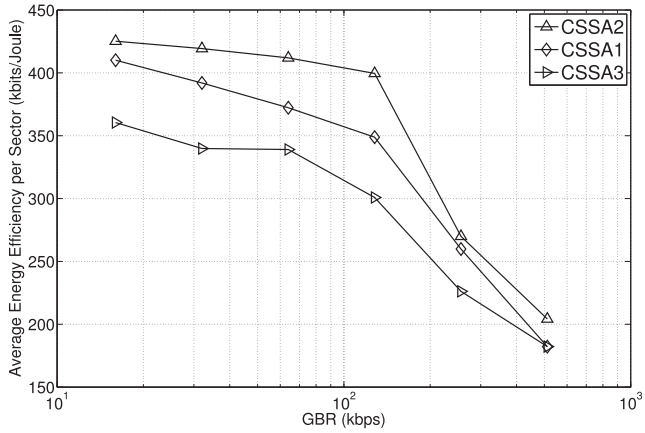
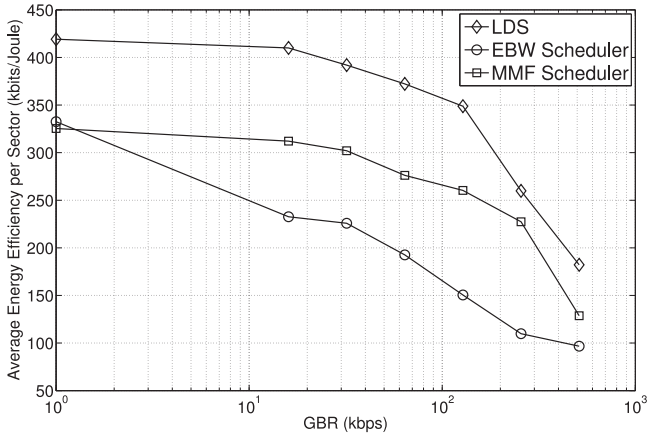


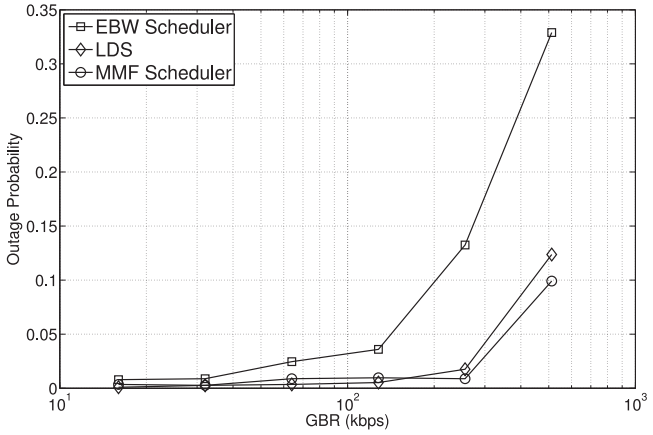
Fig. 6. (a) Average energy efficiency per sector and (b) the outage probabilities for different cell-center radius selection algorithms.

CSSA2 in terms of energy efficiency when users have nonzero rate requirements. CSSA1 selects a cell-center radius such that each user is assigned at least one RB when users have nonzero rate requirements. However, CSSA2 does not consider the rate requirements of the users, for example, 19.8% of the users are not assigned any resources for the given network. While this approach increases the energy efficiency of the network, it causes unacceptable outage probabilities. For example, the energy efficiency of CSSA1 is 3.7% less than CSSA2 when the GBR requirement is 256 kbps. However, the outage probability of CSSA1 is 1.8%, while CSSA2 is 21%. Thus, the outage probability of CSSA2 is not acceptable. The proposed cell-center region selection algorithm benefits from the advantages of both algorithms. It selects more aggressive cell-center radii when the rate constraints are small. On the other hand, it selects more fair cell-center radii when the rate constraints are higher.

Fig. 7(a) shows the average energy efficiency per sector for the LDS, the EBW, and the MMF schedulers. The LDS performs significantly better than the other two schedulers for all GBR requirements. For example, when no GBR requirement is enforced, LDS performs 26% and 28.8% better than the EBW and MMF scheduler, respectively. When no GBR requirement is enforced, the dual prices will always be zero. Therefore, after



(a)



(b)

Fig. 7. (a) Average energy efficiency per sector and (b) outage probability of the network for the LDS, EBW, and MMF schedulers.

assigning one RB to each user, the LDS will assign the rest of the RBs to the user that has the best channel. This approach significantly increases the energy efficiency of the network. On the other hand, the EBW scheduler performs better than the MMF scheduler when no GBR requirement is enforced. The EBW scheduler disregards the rate requirement of users and the channel quality between the user and the base station during the resource allocation process. Each user obtains equal amount of RBs. Therefore, users with better channel qualities get significantly better rates than the other users. On the other hand, the MMF scheduler assigns most of the resources to the users with worse channel conditions to maximize the minimum throughput. Therefore, when there are no GBR requirements, the throughput and corresponding energy efficiency of the EBW scheduler are better than the ones for the MMF scheduler. When we enforce rate requirements, LDS still performs significantly better than the EBW and MMF scheduler. For example, when the GBR requirement is 128 kbps, LDS performs 34% and 132% better than the MMF and the EBW schedulers, respectively. When the minimum rate requirements of users increase, LDS starts to assign resources to users with worse channels in MBSs due to the dual prices. On the other hand, the minimum rate requirement

TABLE III
SIMULATION RESULTS

Methods	Average Sector Energy Efficiency (kb/J) (Outage Probabilities) 0 kbps	128 kbps	512 kbps	Required Time (min)
The Proposed Algorithm	130.81(0%)	128.48(0%)	121.16(14%)	1.45
The Exhaustive Search Algorithm	131.11(0%)	128.71(0%)	121.19(14%)	9855

of the PUEs can be easily satisfied with fewer RBs. Therefore, the dual prices will still be zero and most of the RBs are going to be assigned to the user with the best channel. This approach increases the overall energy efficiency of the network. In addition, when users have minimum rate requirements, higher power transmission levels are necessary for the EBW scheduler to satisfy the rate requirements of users that are in outage. Therefore, the energy efficiency of the sector decreases. In addition, the intercell interference becomes the more significant problem because of these high transmission levels. Due to the fact that the MMF scheduler assigns more resources to the user with worse channel conditions, less power might suffice to satisfy the rate requirement of these users. Therefore, when users have nonzero minimum rate requirements, the MMF scheduler performs better than the EBW scheduler in terms of energy efficiency. In Fig. 7(b), we compare the outage probabilities of the network. Due to the fact that MBSs need to transmit in higher transmission power levels to satisfy the rate requirements, the intercell interference becomes a significant problem for the EBW scheduler. Due to the higher intercell interference and fewer RB allocation to users with worse channel conditions, the outage probability of the EBW scheduler becomes significantly higher than the one for the LDS and MMF scheduler. For example, the outage probability of the EBW scheduler is 13.2%, whereas the one for the LDS is 1.7% and the MMF scheduler is 0.9% when the GBR requirement is 256 kbps. When the rate requirement of the user increases the outages of the LDS becomes higher than the MMF scheduler. The LDS assigns the RB to the user that provides highest increment to the Lagrangian function. Therefore, more users will be in outage than the MMF scheduler to have higher energy efficiency. These results indicate the significance of the scheduler selection.

Table III shows the average energy efficiency per sector for the proposed algorithm and the exhaustive search algorithm. For the purposes of this table, the proposed algorithm and the exhaustive search algorithm are implemented over three sectors that are using the same frequency bands (i.e., shown with the same color in Fig. 1) while the rest of the sectors transmit in full power. The exhaustive search algorithm searches over all possible β and ε pairs and cell-center radii over these three sectors. Due to the fact that searching over the frequency domain increases exponentially with the number of RBs, we use

the EBW scheduler for both the proposed algorithm and the exhaustive search algorithm and use the same frequency allocation over all searches. In Table III, the performance of the proposed algorithm and the exhaustive search algorithm are very similar to each other. In addition, when we increase the minimum rate requirements, the difference between the proposed algorithm and the exhaustive search algorithm decreases. The difference between the proposed algorithm and the exhaustive search algorithm is approximately 0.23% in terms of energy efficiency when there are no minimum rate constraints. The difference becomes 0.18% and 0.02% when the minimum rate requirements are 128 kbps and 512 kbps, respectively. In addition, the outage probabilities of the network are the same for the proposed algorithm and the exhaustive search algorithm and they are 0% and 14% when the rate requirements are 128 kbps and 512 kbps, respectively. In our simulation tool, while the proposed algorithm obtains the results in 1.45 min, the exhaustive search algorithm requires 9855 min to obtain the results. As we discussed in Section IV, when we increase the number of sectors that implement the proposed algorithm, the required time increases linearly. On the other hand, the required time increases exponentially for the exhaustive search algorithm and it requires decades to obtain results for the 19 cell hexagonal grid layout depicted in Fig. 1.

VI. Conclusion

In this paper, we studied the energy efficiency of HetNets. We have proposed an energy-efficient resource allocation algorithm in which the cell-center radius selection, scheduling, and power allocation problems are decoupled. The proposed algorithm maximizes the sector energy efficiency while satisfying the rate requirement of users. The interference pricing mechanism is introduced to prevent selfish behavior of base stations. The proposed algorithms employ the Levenberg–Marquardt method to solve the power allocation problem. Furthermore, the effect of cell-center radius selection is investigated. Based on our simulation results, we demonstrate that significant energy savings can be achieved, while the outage probability is also reduced.

There is a range of GBR requirements our approach improves energy efficiency, outage probabilities, and reduces average transmit power, substantially in some cases. For the tested traffic mixes of the same GBR requirements, this range is $0, 16, 32, 64$, and 128 kbps within the transmission parameters studied. Within the experimental parameters, for GBR requirements of 256 and 512 kbps, the energy efficiency becomes worse but the outage probability improves. The improvement is such that the outage probability can move from an unacceptable level to acceptable one. The transmit power improves but not uniformly, unlike the parameters in the previous case. We note that in these simulation setups, the GBR requirements for all users are considered the same. In a real-life situation, there will be a mix of GBR requirements, and the effects of these higher GBR requirements will be less. The newly introduced cell-center selection algorithm (CSSA1) performs better than the two we proposed earlier, namely CSSA2 and CSSA3. The LDS performs better in terms of both energy efficiency and outage probability. Interference pricing mechanism reduces outages significantly. The proposed approach achieves the same energy efficiency as the exhaustive search, whereas the exhaustive search takes an unacceptable amount of time.

References

[1] Cisco Systems, Inc., "Cisco Visual Networking Index: Global mobile data traffic forecast update, 2015–2020," White Paper, Cisco Systems, Inc., San Jose, CA, USA, Feb. 2016.

[2] K. Davaslioglu and E. Ayanoglu, "Quantifying potential energy efficiency gain in green cellular wireless networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2065–2091, Fourth Quarter 2014.

[3] Z. Han, D. Niyato, W. Saad, T. Başar, and A. Hjørungnes, *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*. Cambridge, U.K.: Cambridge Univ., 2012.

[4] J. Huang, R. Berry, and M. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1074–1084, May 2006.

[5] C. Shi, R. Berry, and M. Honig, "Distributed interference pricing for OFDM wireless networks with non-separable utilities," in *Proc. Annu. Conf. Inf. Sci. Syst.*, Mar. 2008, pp. 755–760.

[6] C. Shi, R. Berry, and M. Honig, "Monotonic convergence of distributed interference pricing in wireless networks," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2009, pp. 1619–1623.

[7] C. Saraydar, N. B. Mandayam, and D. Goodman, "Pricing and power control in a multicell wireless data network," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 10, pp. 1883–1892, Oct. 2001.

[8] C. Saraydar, N. B. Mandayam, and D. Goodman, "Efficient power control via pricing in wireless data networks," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 291–303, Feb. 2002.

[9] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-efficient resource allocation in OFDMA networks," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3767–3778, Dec. 2012.

[10] Z. Zheng, L. Dan, S. Gong, and S. Li, "Energy-efficient resource allocation for downlink OFDMA systems," in *Proc. IEEE Int. Conf. Commun. Workshops*, Jun. 2013, pp. 391–395.

[11] X. Xiao, X. Tao, and J. Lu, "QoS-aware energy-efficient radio resource scheduling in multi-user OFDMA systems," *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 75–78, Jan. 2013.

[12] R. S. Prabhu and B. Daneshrad, "An energy-efficient water-filling algorithm for OFDM systems," in *Proc. IEEE Int. Conf. Commun.*, May 2010, pp. 1–5.

[13] G. Miao, N. Himayat, and G. Li, "Energy-efficient link adaptation in frequency-selective channels," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 545–554, Feb. 2010.

[14] G. Miao, N. Himayat, G. Li, and S. Talwar, "Distributed interference-aware energy-efficient power optimization," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1323–1333, Apr. 2011.

[15] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3618–3631, Oct. 2012.

[16] L. Venturino, A. Zappone, C. Risi, and S. Buzzi, "Energy-efficient scheduling and power allocation in downlink OFDMA networks with base station coordination," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 1–14, 2015.

[17] E. Hossain, V. K. Bhargava, and G. P. Fettweis, *Green Radio Communication Networks*. Cambridge, U.K.: Cambridge Univ., 2012.

[18] C. Isheden, Z. Chong, E. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2946–2957, Aug. 2012.

[19] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, "Interference coordination and cancellation for 4G networks," *IEEE Commun. Mag.*, vol. 47, no. 4, pp. 74–81, Apr. 2009.

[20] H. Holma and A. Toskala, *LTE-Advanced: 3GPP Solution for IMT-Advanced*. West Sussex, U.K.: Wiley, Ltd, 2012.

[21] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, "Interference coordination and cancellation for 4G networks," *IEEE Commun. Mag.*, vol. 47, no. 4, pp. 74–81, Apr. 2009.

[22] M. Assaad, "Optimal fractional frequency reuse (FFR) in multicellular OFDMA system," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2008, pp. 1–5.

[23] Z. Xu, G. Li, C. Yang, and X. Zhu, "Throughput and optimal threshold for FFR schemes in OFDMA cellular networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2776–2785, Aug. 2012.

[24] M. Peng, K. Zhang, J. Jiang, J. Wang, and W. Wang, "Energy-efficient resource assignment and power allocation in heterogeneous cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5275–5287, Nov. 2015.

[25] K. M. S. Huq, S. Mumtaz, and J. Rodriguez, "QoS to QoS-efficient resource scheduling for HetNet comp," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2015, pp. 5954–5960.

[26] P. Lee, T. Lee, J. Jeong, and J. Shin, "Interference management in LTE femtocell systems using fractional frequency reuse," in *Proc. Int. Conf. Adv. Comm. Techno.*, Feb. 2010, vol. 2, pp. 1047–1051.

[27] N. Saquib, E. Hossain, and D. I. Kim, "Fractional frequency reuse for interference management in LTE-Advanced HetNets," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 113–122, Apr. 2013.

[28] M. Heath and A. Brydon, "Wireless network traffic 2008–2015: Forecasts and analysis," White Paper, Analysys Mason, London, U.K., 2008.

[29] K. Davaslioglu, C. Coskun, and E. Ayanoglu, "Energy-efficient resource allocation for fractional frequency reuse in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5484–5497, Oct. 2015.

[30] C. Raman, G. J. Foschini, R. A. Valenzuela, R. D. Yates, and N. B. Mandayam, "Power savings from half-duplex relaying in downlink cellular systems," in *Proc. Allerton Conf. Commun. Control Comput.*, Sep. 2009, pp. 748–753.

[31] S. Sesia, I. Toufik, and M. Baker, *LTE—The UMTS Long Term Evolution: From Theory to Practice*. West Sussex, U.K.: Wiley, 2009.

[32] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[33] H. Holtkamp, G. Auer, V. Giannini, and H. Haas, "A parameterized base station power model," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2033–2035, Nov. 2013.

[34] C. Isheden and G. P. Fettweis, "Energy-efficient multi-carrier link adaptation with sum rate-dependent circuit power," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2010, pp. 1–6.

[35] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation for secure OFDMA systems," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2572–2585, Jul. 2012.

[36] F. Richter, A. Fehske, and G. Fettweis, "Energy efficiency aspects of base station deployment strategies for cellular networks," in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2009, pp. 1–5.

[37] G. Auer *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[38] L. Hoo, B. Halder, J. Tellado, and J. Cioffi, "Multiuser transmit optimization for multicarrier broadcast channels: Asymptotic FDMA capacity region and algorithms," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 922–930, Jun. 2004.

[39] W. Yu, "Multiuser water-filling in the presence of crosstalk," in *Proc. Inf. Theory Appl. Workshop*, Jan. 2007, pp. 414–420.

[40] A. Najjar, N. Hamdi, and A. Bouallegue, "Efficient frequency reuse scheme for multi-cell OFDMA systems," in *Proc. IEEE Symp. Comput. Commun.*, Jul. 2009, pp. 261–265.

[41] K. Davaslioglu and E. Ayanoglu, "Interference-based cell selection in heterogeneous networks," in *Proc. Inf. Theory Appl. Workshop*, San Diego, CA, USA, Feb. 2013, pp. 1–6.

[42] T. Yang, F. Heliot, and C. H. Foh, "A survey of green scheduling schemes for homogeneous and heterogeneous cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 11, pp. 175–181, Nov. 2015.

[43] W. Rhee and J. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. IEEE Veh. Technol. Conf.*, May 2000, vol. 2, pp. 1085–1089.

[44] Qualcomm, "Rising to meet the 1000x mobile data challenge," White Paper, Qualcomm, San Diego, CA, USA, Jun. 2012.

[45] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*. New York, NY, USA: Wiley, 1993.

[46] P. M. Pardalos and M. G. C. Resende, Eds., *Handbook of Applied Optimization*. New York, NY, USA: Oxford Univ., 2002.

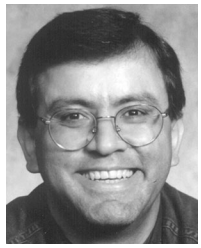
[47] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ., 2012.

[48] 3GPP, TR 36.814, "Further advancements for E-UTRA physical layer aspects (Release 9)," Mar. 2010.

Cemil Can Coskun (S'07) received the B.S. and M.S. degrees in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 2010 and 2012, respectively. He is currently working toward the Ph.D. degree at the Department of Electrical Engineering and Computer Science, University of California, Irvine, CA, USA.

His research interests include the area of resource allocation in wireless networks, small cell base station deployment in wireless networks, green communications, and energy efficiency and spectral efficiency tradeoff in wireless networks.

Kemal Davaslioglu (S'07–M'17) received the B.S. degree from Bilkent University, Ankara, Turkey, in 2008 and the M.S. degree from Bogazici University, Istanbul, Turkey, in 2010, both in electrical and electronics engineering. He received the Ph.D. degree in electrical and computer engineering from the University of California, Irvine, CA, USA, in 2015. He has held internships at Broadcom Corp., Irvine, CA, USA, working on 10-Gigabit Ethernet systems from 2011 to 2012, and on 60 GHz system design and channel characterization in 2015. In 2016, he was a postdoctoral scholar in the Electrical Engineering Department of the University of South Florida, Tampa, FL, USA. Since 2017, he works at Intelligent Automation Inc., Rockville, MD, USA. His research interests are on the resource allocation in wireless networks, green communications, signal processing for wireless communications, and ultra-wideband communications.



Ender Ayanoglu (S'82–M'85–SM'90–F'98) received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1986.

He was with the Communications Systems Research Laboratory, Holmdel, NJ, USA, part of AT&T Bell Laboratories, until 1996, and Bell Laboratories, Lucent Technologies, from 1996 to 1999. From 1999 to 2002, he was a Systems Architect with Cisco Systems, Inc., San Jose, CA, USA. Since 2002, he has been a Professor with the Department of Electrical Engineering and Computer Science, University of California, Irvine, CA, USA, where he served as the Director of the Center for Pervasive Communications and Computing and the Conexant-Broadcom Endowed Chair from 2002 to 2010. His past accomplishments include invention of 56K modems, characterization of wavelength conversion gain in wavelength division multiplexed systems, and diversity coding.

Dr. Ayanoglu served as the Founding Chair of the IEEE-ISTO Broadband Wireless Internet Forum, an industry standards organization, from 2000 to 2001. He served on the Executive Committee of the IEEE Communications Society Communication Theory Committee from 1990 to 2002 and the Chair from 1999 to 2002. From 1993 to 2014, he was an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS. He served as the Editor-in-Chief of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2004 to 2008 and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS—Series on Green Communications and Networking from 2014 to 2016. Since 2014, he has been a Senior Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS. Since 2016, he has been serving as the Founding Editor-in-Chief of the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING. He was the recipient of the IEEE Communications Society Stephen O. Rice Prize Paper Award in 1995, the IEEE Communications Society Best Tutorial Paper Award in 1997, and the IEEE Communications Society Communication Theory Technical Committee Outstanding Service Award in 2014.