Energy- and Spectrum-Efficient User Association in Millimeter-Wave Backhaul Small-Cell Networks

Agapi Mesodiakaki, Ferran Adelantado, Luis Alonso, Marco Di Renzo, Christos Verikoukis

Energy and Spectrum Efficient User Association in Millimeter Wave Backhaul Small Cell Networks

Agapi Mesodiakaki, *Member, IEEE,* Ferran Adelantado, *Member, IEEE,* Luis Alonso, *Senior Member, IEEE,* Marco Di Renzo *Senior Member, IEEE*, Christos Verikoukis, *Senior Member, IEEE*

***Abstract*—Macrocells are expected to be densely overlaid by small cells (SCs) to meet the increasing capacity demands. Due to their dense deployment, some SCs will not be connected directly to the core network and thus they may forward their traffic to the neighboring SCs until they reach it, thereby forming a multi-hop backhaul (BH) network. This is a promising solution, since the expected short length of BH links enables the use of millimeter wave (mmWave) frequencies to provide high capacity BH. In this context, user association becomes challenging due to the multi-hop BH architecture and therefore new optimal solutions should be developed. Thus, in this paper, we study the user association problem aiming at the joint maximization of network energy and spectrum efficiency, without compromising the user quality of service. The problem is formulated as an ε-constraint problem, which considers the transmit energy consumption both in the access network, i.e., the links between the users and their serving cells, and the BH links. The optimal Pareto front solutions of the problem are analytically derived for different BH technologies and insights are gained into the energy and spectrum efficiency trade-off. The proposed optimal solutions, despite their high complexity, can be used as a benchmark for the performance evaluation of user association algorithms. We also propose a heuristic algorithm, which is compared with reference solutions under different traffic distribution scenarios and BH technologies. Our results motivate the use of mmWave BH, while the proposed algorithm is shown to achieve near-optimal performance.**

***Index Terms*—Backhaul, cell selection, context-awareness, green communications, LTE-Advanced, millimeter wave.**

I. Introduction

THE mobile data traffic is expected to grow significantly during the next few years, which results in an urgent need for mobile operators to maintain capacity growth. Serving more traffic leads to increased energy consumption, and therefore, how to minimize the energy consumption becomes also important. In parallel, the spectrum scarcity problem stresses the need for spectral efficient s olutions. T he aforementioned goals can be summarized into the joint maximization of energy and spectrum efficiency, w hich c onstitutes a fundamental design objective for next generation cellular networks.

A. Mesodiakaki is with the Department of Computer Science, Karlstad University, Karlstad, Sweden. E-mail: agapi.mesodiakaki@kau.se

F. Adelantado is with Open University of Catalonia, Barcelona, Spain. E-mail: ferranadelantado@uoc.edu

L. Alonso is with the Signal Theory and Communications Department, Technical University of Catalonia, Spain. E-mail: luisg@tsc.upc.edu

M. Di Renzo is with the Laboratoire des Signaux et Systèmes, CNRS, CentraleSupélec, Univ Paris Sud, Université Paris-Saclay, Gif-sur-Yvette, France. E-mail: marco.direnzo@l2s.centralesupelec.fr

C. Verikoukis is with Telecommunications Technological Centre of Catalonia, Castelldefels, Spain. E-mail: cveri@cttc.es

To that end, the dense deployment of small cells (SCs), overlaying the existing macrocell networks, is a promising solution. The SC deployment reduces the distance between user equipments (UEs) and base stations (BSs)[1] and, consequently, i) the area spectral efficiency (bps/Hz/m^2) increases, and ii) the energy consumption in the access network (AN), i.e., the links between the UEs and their serving BSs, decreases. Hence, dense deployment of SCs is expected during the next years, with SC radius being eventually of the order of 50 meters [1].

However, the dense deployment of SCs also poses new challenges. Due to the high number of deployed SCs, the direct connection of all SCs to the core network becomes complicated. Fiber connections, which have been traditionally considered as the best backhaul (BH) solution, are prohibitive in this case due to their high deployment cost [2]. A promising solution lies in exploiting the existing connection between the macrocell site and the core network (most of the times it is a fiber connection), and to provide core network connectivity to SCs through the macrocell site [3]. Still, in order to connect the SCs to the macrocell site (thus providing them core network connectivity), new cheap wireless BH solutions are required.

In addition, this wireless BH is expected to provide high-capacity services from the SCs to the core network, in order to meet the expected traffic demands of the order of Gbps [1]. Therefore, a promising solution for high capacity wireless BH connections between the SCs and the core network lies in using millimeter wave (mmWave) frequencies, due to their high bandwidth availability [2]. It has been shown, however, that mmWave frequencies are capable of providing good coverage only if the transmission distance is shorter than 200 meters [1]–[3]. Otherwise, links may not be established. In parallel, small wavelengths enable highly directive antennas to compensate the high path loss with the use of *pencil* beams [2]. Since the macrocell radius is even in dense deployments of the order of 500 meters, this implies that a multi-hop architecture of point-to-point line-of-sight (LOS) links is needed, in order to allow each of the SCs to reach the macrocell site [3], [4].

In this context, user association becomes challenging due to the multi-hop BH architecture [5] and therefore new optimal solutions need to be developed aiming at the joint energy and spectrum efficiency maximization of the network.

A. State-of-the-art and Contribution

The user association problem has received a lot of research attention, since it impacts both the network and UE perfor-

[1]In this paper, we will use the term BS to refer to a macrocell BS and/or a SC BS (i.e., an eNodeB (eNB) and/or a Home eNB in LTE-A, respectively).

mance. In LTE-Advanced, the user association is based on the reference signal received power (RSRP), which measures the average received power over the resource elements that carry cell-specific reference signals within certain bandwidth [4]. Although RSRP maximizes the signal-to-interference-plus-noise ratio (SINR) of UEs, it was shown that it does not significantly increase the overall throughput, since few users get connected to SCs [6]. Thus, range expansion (RE) (also known as biasing) was introduced, whereby UEs are actively pushed onto SCs [6]. In this case, although a UE may be associated with a BS not providing the best SINR, better load balancing is achieved between SCs and macrocell.

In [7], the authors propose a low-complexity distributed algorithm that converges to a near-optimal solution and they show that a per-tier biasing loses little, if the bias values are chosen carefully. In [8], the joint user association and resource allocation problem is studied. The authors aim to find the optimal association so that the total resources required to satisfy the given UE traffic demands are minimized. Focusing also on the joint spectrum allocation and user association problem, in [9], a proportionally fair utility function based on the coverage rate is defined. The authors associate the UEs with BSs based on the biased downlink received power, while stochastic geometry is used to model the placement of BSs. In [10], the authors formulate two different user association problems. The first one is based on a sum utility of long-term rate maximization with rate quality of service (QoS) constraints, and the second on minimizing a global outage probability with outage QoS constraints.

Taking into account the BH, in [11], the authors model a BH-aware BS assignment problem as a multiple-choice multi-dimensional Knapsack problem. In the considered framework, they impose constraints on both AN and BH resources. The main idea behind their algorithm is to distribute traffic among BSs according to a load balancing strategy, considering both AN and BH load status. Yet, the proposed algorithm, reduces the BH congestion at the expense of lower spectral efficiency, since some UEs may be assigned to non-optimal BSs in terms of RSRP. In [12], a load-balancing based mobile association framework is proposed under both full and partial frequency reuse, and pseudo-optimal solutions are derived using gradient descent method. In [13], a new theoretical framework is introduced to model the downlink user association problem, while upper bounds are derived for the achievable sum rate and minimum rate using convex optimization. In [14], a joint user association and resource allocation optimization problem is proposed, which is shown to be NP-hard. Therefore, the authors develop techniques to obtain upper bounds on the system performance. In [15], the joint problem of downlink user association and wireless BH bandwidth allocation is studied in two-tier cellular heterogeneous networks (HetNets). According to the considered architecture, SCs are connected through wireless BH with the macrocell BS. The problem is formulated as a sum logarithmic user rate maximization problem, and wireless BH constraints are also considered.

However, the aforementioned approaches either consider only the AN [4], [6]–[10], thus totally overlooking the BH capacity constraints and energy impact, or do not take into account the energy consumption of the network and hence, their energy efficiency cannot be guaranteed [11]–[15].

To that end, in this paper, we study the user association problem aiming at the joint energy and spectrum efficiency maximization, while taking into account both the AN and BH and without compromising the UE throughput demands. Preliminary results of this research have been published in [16]. However, in this paper, we provide the following contributions:

- The aforementioned problem is formulated as an ε-constraint problem [17], where the total transmit power consumption of AN and BH is the objective to be minimized and the amount of spectrum resources needed is set as constraint, with its upper bound denoted by ε.
- We study the trade-off between energy and spectrum efficiency analytically for different BH technologies by solving the ε-constraint problem for all different ε. Thereby, we derive the Pareto front solutions of the problem, i.e., the set of optimal solutions for all ε values, which can be used as a benchmark for the performance evaluation of user association algorithms.
- Due to the high complexity of the derived optimal solutions, which increases for a higher number of UEs and BSs, we also propose a low-complexity user association algorithm, which aims at the maximization of the energy efficiency given a specific spectral efficiency target. The algorithm is able to select any point of the Pareto front, by accordingly tuning a single parameter, i.e., the spectral efficiency target. Moreover, for each UE, it considers the total transmit power consumption needed (both AN and BH) to serve its traffic. This association metric relaxes the assumption of [16] that all BH links are homogeneous, by considering the actual transmit power consumption of each BH link and not just the number of hops.
- Finally, we compare the energy and spectrum efficiency of the proposed algorithm with existing user association solutions as well as with the derived optimal solutions under different spectral efficiency targets, traffic distribution scenarios and BH technologies. Our results motivate the use of mmWave frequencies to provide high capacity BH, while the proposed algorithm is shown to achieve notable performance gains.

The rest of the paper is organized as follows: In Section II, the system model is presented. In Section III, the problem formulation and the solution methodology are provided. In Section IV and Section V, the proposed algorithm is described and compared, respectively, with existing user association algorithms as well as with the analytical solutions derived in Section III. Finally, Section VI concludes the paper.

II. System Model

Without loss of generality and in accordance with the scenarios proposed by 3GPP [18], we focus our analysis on a single eNB sector, overlaid with multiple SCs. In particular, we consider a set of BSs, denoted by $\mathcal{C}$, which includes one eNB (j=0) and $C-1$ SCs (j=1...C-1), with C representing the cardinality of the set $\mathcal{C}$. The SCs are divided in N_{cl} clusters (k=1...N_{cl}), as depicted in Fig. 1, with SC_k denoting the

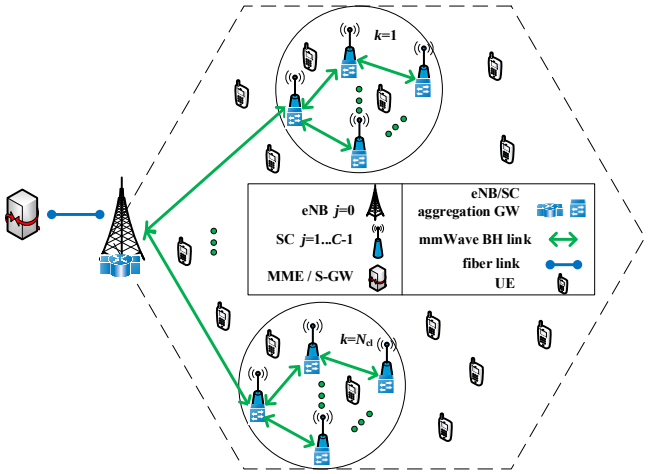


Fig. 1. System model.

number of SCs in cluster k [18]. We study the downlink and make the following assumptions:

- Each SC is connected to the core network through the eNB aggregation gateway either directly or through one or more SC aggregation gateways [3]–[5].
- There is a fiber connection between the core network and the eNB site, and a set of point-to-point LOS mmWave BH links between the eNB site and the SCs, denoted by $\mathcal{L}$=$\{\mathcal{L}_1,\mathcal{L}_2,...\mathcal{L}_l,...\mathcal{L}_{C-1}\}$. Each mmWave BH link l is represented by a set $\mathcal{L}_l$ that includes all cells j that backhaul their traffic through it (i.e., $\forall j \in \mathcal{L}_l$).
- Flat slow-fading channels are considered [14]. Therefore, we assume that the total transmission power of each BS is equally distributed among its subcarriers [4].
- We consider a set of $\mathcal{N}$ UEs (i=1,...,N) with strict guaranteed bit rate (GBR) QoS requirements, denoted as $r_{i,net}$, based on their service/application [19].
- Each UE can be associated only with one BS at a time.
- There is a maximum number of spectrum resource units available to each BS j, i.e., physical resource blocks (PRBs)[2], denoted by $c_{j_{max}}$.

In the following, the most important parameters involved in the total network energy efficiency calculation are derived. The SINR calculation is given in Section II-A, while both AN and BH power consumption models are provided in Section II-B.

A. SINR calculation

The signal-to-noise ratio (SNR) received by UE i from BS j is given by [20]

$$SNR_{ij\,(dB)} = P_{j_{PRB}\,(dBm)} + G_{T_{x_j}\,(dBi)} - L_{cb_j\,(dB)} - L_{p_{ij}\,(dB)} - L_{f_{ij}\,(dB)} - N_{th\,(dBm)} - NF_{(dB)} \quad (1)$$

with $P_{j_{PRB}}$=$10log_{10}(P_{j_{max}}/c_{j_{max}})$ being the power allocated by BS j to a PRB, where $P_{j_{max}}$ is its maximum transmission power (mW), and $c_{j_{max}}$ is the maximum number of PRBs allocated to it. The parameter $G_{T_{x_j}}$ is the antenna gain of BS j and L_{cb_j} is the cable loss between the radio RF connector and the antenna. The path loss between UE i and BS j is denoted by $L_{p_{ij}}$, while $L_{f_{ij}}$ represents the losses due to shadowing. Finally, N_{th} stands for the thermal noise and NF is the noise figure. The SINR of UE i from BS j is given by

$$SINR_{ij\,(dB)} = SNR_{ij\,(dB)} - 10log_{10}\left(\frac{I_{ij(mW)}}{N_{total\,(mW)}} + 1\right) \quad (2)$$

where I_{ij} is the total interference experienced by UE i, when associated with BS j, which depends on the applied frequency allocation scheme. Due to the constant power allocation, the $SINR_{ij}$ of UE i from BS j can be estimated *a priori*[3], and be given as an input to the problem. Hence, the proposed work can be applied regardless of the employed channel allocation scheme. Still, although it is out of the scope of this paper, the combination of our proposal with a sophisticated channel allocation could further improve the system performance. Finally, $N_{total} = 10^{(N_{th\,(dBm)}+NF_{(dB)})/10}$ denotes the total noise power (mW) experienced by UE i.

B. Power consumption models

The total network power consumption can be divided into the power consumed in the BSs (i.e, in the AN) and in the BH links. The first is given by [16], [21]

$$P_{AN_{(W)}} = \sum_{j\in\mathcal{C}} \left(P^{AN}_{j_{stat\,(W)}} + P^{AN}_{j_{var\,(W)}}\right) \quad (3)$$

where $P^{AN}_{j_{stat}}$ is the fixed power consumption of BS j attributed to e.g., power supply, cooling, and baseband unit operation [21] and $P^{AN}_{j_{var}}$ is the load-dependent power consumption of BS j. Without loss of generality, we assume ideal electronics in terms of power efficiency and therefore the load dependent power consumption part becomes equal to the radio frequency (RF) transmit power consumption part, which is given by [16]

$$P^{AN}_{j_{var\,(W)}} = P^{AN}_{j_{RF\,(W)}} = \sum_{i\in\mathcal{N}} (P_{j_{PRB\,(W)}})\lceil c_{ij}\rceil a_{ij} = \sum_{i\in\mathcal{N}} \left(\frac{P_{j_{max}}}{c_{j_{max}}}\right) \left\lceil \frac{r_{i,net}}{(1-BLER)(1-k_{ov})} \frac{1}{b\log_2(1+SINR_{ij})} \right\rceil a_{ij} \quad (4)$$

where c_{ij} represents the number of PRBs needed for the association of UE i with BS j and $r_{i,net}$ is the rate demand of UE i. The parameter $BLER$ stands for the block error rate (BLER), i.e, for the number of erroneous blocks divided by the total number of received blocks [22] and k_{ov} is the percentage of overhead bits (e.g., cyclic prefixes, reference signals) [23]. From now on, we will denote as $r_i = \frac{r_{i,net}}{(1-BLER)(1-k_{ov})}$, the total rate needed for the satisfaction of $r_{i,net}$. Parameter b is the bandwidth of a PRB and $\lceil\cdot\rceil$ is the ceiling function operator. The denominator of the third fraction is derived by Shannon's theorem and represents the maximum rate that can be achieved with effective $SINR_{ij}$ [23] and bandwidth equal to b. Finally, a_{ij} is the association vector (equal to 1 when the UE i is associated with BS j and 0 otherwise).

[2]Please note that 1 PRB is equal to 12 subcarriers in the frequency domain and 0.5 ms in the time domain [4].

[3]Please note that the SINR after the UE association may differ from the estimated one, as the interference experienced by the UE depends on resource allocation i.e., whether neighboring BSs allocate the same PRBs to other UEs, and consequently on user association. In this work, to overcome this problem, the worst-case scenario in terms of generated interference is considered.

Similar to the AN, the power consumption of the BH links consists of a fixed and a variable part [21], and thus equals to

$$P_{BH(W)} = \sum_{\mathcal{L}_l \in \mathcal{L}} P^{BH}_{\mathcal{L}_{l\,stat}(W)} + P^{BH}_{\mathcal{L}_{l\,var}(W)} \tag{5}$$

Under the assumption of ideal electronics, the load dependent power consumption of a BH link $\mathcal{L}_l$, i.e., $P^{BH}_{\mathcal{L}_{l\,var}}$, equals to the RF transmit power consumption, which is given by [20], [24]

$$P^{BH}_{\mathcal{L}_{l\,RF\,(dBm)}} = SINR^{trg}_{\mathcal{L}_{l\,(dB)}} + \overbrace{L_{p_{o\,(dB)}} + L_{p_{\mathcal{L}_l\,(dB)}} + IL_{(dB)}}^{\alpha_{\mathcal{L}_l}} \overbrace{+N_{th\,(dBm)} + NF_{(dB)} - G_{T_{x\,\mathcal{L}_l\,(dBi)}} - G_{R_{x\,\mathcal{L}_l\,(dBi)}}}^{} \tag{6}$$

where L_{p_o} is the path loss at 1 m distance and $L_{p_{\mathcal{L}_l}}$=$20log_{10}(4\pi\frac{d_{\mathcal{L}_l}}{\lambda})$ is the path loss at distance $d_{\mathcal{L}_l}$ equal to the length of the link. Moreover, λ is the signal wavelength (e.g., for 60 GHz, $\lambda = 0.005\ m$) and IL is the implementation loss that may account for e.g., distortion, intermodulation and/or phase noise. The over-braced equation, which is derived by subtracting from the total losses, the transmitter and receiver antenna gains of the BH link, will be denoted from now on by $\alpha_{\mathcal{L}_l}$. Finally, assuming that link adaptation is employed [20], $SINR^{trg}_{\mathcal{L}_l}$ corresponds to the (minimum) target SINR that is needed so that the aggregated BH link traffic is successfully transmitted and can be given by [20]

$$SINR^{trg}_{\mathcal{L}_{l\,(dB)}} = 10log_{10}\left(2^{\frac{\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{L}_l} r_i a_{ij}}{B_{\mathcal{L}_l}}} - 1\right) \tag{7}$$

where $B_{\mathcal{L}_l}$ is the bandwidth of the BH link $\mathcal{L}_l$ and $\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{L}_l} r_i a_{ij}$ is the aggregated traffic that passes through it. For mmWave, the generated interference is negligible due to high path loss, and thus $SINR^{trg}_{\mathcal{L}_l}$=$SNR^{trg}_{\mathcal{L}_l}$.

III. Problem formulation

The problem under study aims at the joint maximization of the network energy and spectrum efficiency, without compromising the UE QoS (i.e., the UE throughput demands). The energy efficiency (bits/Joule) is expressed as the total number of successfully transmitted useful bits divided by the total energy consumption or equivalently as the total *goodput* of the network divided by the total power consumption (i.e., the sum of the power consumed in the AN and in the BH links). Under the condition that the specific UE throughput demands are satisfied, the network energy efficiency maximization is equivalent to power consumption minimization, while the spectral efficiency maximization is equivalent to PRBs minimization.

The aforementioned problem is a non-convex multi-objective problem. Therefore, for its formulation, we employ the ε-constraint method, which is able to find any Pareto optimal solution even for non-convex problems [17]. According to it, one of the objectives is included in the utility function to be optimized (i.e., minimization of the total power consumption), while the others (i.e., minimization of the total number of required PRBs) are converted into constraints by setting an upper bound to them. Given that the fixed power consumption[4] is independent of the user association decision, the minimization of the total power consumption is equivalent to the minimization of the traffic-dependent part (i.e., the RF transmit power consumption in our case). Therefore, our study from now on focuses on this part, as depicted in (8).

Hence, the first term of the objective function in (8) represents the total RF transmit power consumption of the AN and the second of the BH links. We remind that a_{ij}[5] denotes the association vector that is equal to 1 when the UE i is associated with BS j and 0 otherwise (8a). Each UE can be associated only with one BS at a time (8b). The total number of PRBs used by BS j, denoted by c_{ij}, cannot exceed the maximum number that is allocated to it (8c). The RF transmit power consumption of the BH link $\mathcal{L}_l$ cannot exceed a maximum value, denoted by $P_{BH_{max}}$ (8d). The parameter $s_{\mathcal{L}_l j}$ is 1 if the traffic of the BS j passes through the BH link $\mathcal{L}_l$ and 0 otherwise (8e). Finally, constraint (8f) refers to the total number of PRBs and thus to the network spectrum efficiency.

$$\begin{aligned}
\underset{a_{ij}}{argmin} \quad & f_1(a_{ij}) = \sum_{j\in\mathcal{C}} \overbrace{\sum_{i\in\mathcal{N}} P_{j_{PRB}} c_{ij} a_{ij}}^{P^{AN}_{j_{RF}(W)}} + \\
& + \sum_{\mathcal{L}_l\in\mathcal{L}} \overbrace{\left(2^{\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} a_{ij} s_{\mathcal{L}_l j} \frac{r_i}{B_{\mathcal{L}_l}}} - 1\right) 10^{\frac{\alpha_{\mathcal{L}_l} - 30}{10}}}^{P^{BH}_{\mathcal{L}_{l\,RF}(W)}} \\
\text{s.t.} \quad & a)\ a_{ij} \in \{0,1\},\ \forall i \in \mathcal{N},\ \forall j \in \mathcal{C} \\
& b)\ \sum_{j\in\mathcal{C}} a_{ij} = 1, \forall i \in \mathcal{N} \\
& c)\ \sum_{i\in\mathcal{N}} a_{ij} c_{ij} \le c_{j_{max}},\ \forall j \in \mathcal{C} \\
& d) P^{BH}_{\mathcal{L}_{l\,RF}} \le P_{BH_{max}}\ \forall \mathcal{L}_l \in \mathcal{L} \\
& e)\ s_{\mathcal{L}_l j} \in \{0,1\},\ \forall \mathcal{L}_l \in \mathcal{L},\ \forall j \in \mathcal{C} \\
& f) f_2(a_{ij}) = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{C}} a_{ij} c_{ij} \le \varepsilon
\end{aligned} \tag{8}$$

Theorem 1. *The solution of the ε-constraint problem in* (8) *is weakly Pareto optimal.*

Proof. Let $a^\star_{ij}$ be a solution of the ε-constraint problem. Let us assume that $a^\star_{ij}$ is not weakly Pareto optimal. In this case there exists some other a_{ij} such that $f_k(a_{ij}) < f_k(a^\star_{ij})$ for k=1,2. This means that $f_2(a_{ij}) < f_2(a^\star_{ij}) \le \varepsilon$. Hence, a_{ij} is feasible with respect to the ε-constraint problem. While in addition $f_1(a_{ij}) < f_1(a^\star_{ij})$, we have a contradiction to the assumption that $a^\star_{ij}$ is a solution of the ε-constraint problem. Thus, $a^\star_{ij}$[6] has to be weakly Pareto optimal. □

Although, according to Theorem 1, every solution of the ε-constraint problem is weakly Pareto optimal, there is no Pareto optimal solution, since there is no solution that optimizes both objectives simultaneously. Therefore, it is reasonable to search for a *good* trade-off between the two objectives

[4]Still, the inclusion of the fixed power would impact all the algorithms by equally increasing their power consumption.

[5]Due to the binary association vector and the non-linear objective function and contraints, the problem is a 0-1 non-linear integer programming problem.

[6]Please note that, in the rest of the paper, a_{ij} is omitted.

instead. To that end, the increase of ε leads to a relaxation of the spectral efficiency constraint (i.e., f_2) and consequently to a more energy efficient solution. On the contrary, the decrease of ε improves the spectral efficiency of the solution by degrading its energy efficiency. The set of solutions for the subproblems resulting from the variation of ε define the Pareto front, hereafter denoted by $\mathcal{F}$. In practice, due to the high number of subproblems and the difficulty to establish an efficient variation scheme for the ε-vector, this approach has mostly been integrated within heuristic and interactive schemes. However, due to the nature of (8), it is possible to derive the exact Pareto front with the use of an iterative algorithm [25]. The idea is to construct a sequence of ε-constraint problems based on a progressive reduction of ε.

Let $\vec{\phi}^I = (\phi_1^I, \phi_2^I)$ be the ideal point, where $\phi_1^I = \min(f_1)$ and $\phi_2^I = \min(f_2)$ stand for the minimum value of f_1 and f_2, respectively. Equivalently, let $\vec{\phi}^N = (\phi_1^N, \phi_2^N)$ be the nadir point, with ϕ_1^N and ϕ_2^N being the minimum values of f_1 and f_2, when $f_2 = \phi_2^I$ and $f_1 = \phi_1^I$, respectively, i.e., $\phi_1^N = \min\{f_1 : f_2 = \phi_2^I\}$ and $\phi_2^N = \min\{f_2 : f_1 = \phi_1^I\}$. Thus, (ϕ_1^I, ϕ_2^N) is the solution of the Pareto front that minimizes the RF transmit power consumption (i.e., f_1) without spectral efficiency constraints, whereas (ϕ_1^N, ϕ_2^I) is the solution in $\mathcal{F}$ that minimizes the total number of PRBs used (i.e., f_2).

Lemma 1. *Both (ϕ_1^I, ϕ_2^N) and (ϕ_1^N, ϕ_2^I) belong to $\mathcal{F}$, i.e., $(\phi_1^I, \phi_2^N) \in \mathcal{F}$ and $(\phi_1^N, \phi_2^I) \in \mathcal{F}$.*

Proof. Let us assume that $(\phi_1^I, \phi_2^N) \notin \mathcal{F}$. Then, $\exists\ \vec{f} = (f_1, f_2) \in \Phi$: $(f_1, f_2) \succ (\phi_1^I, \phi_2^N)$, where Φ denotes the *objective space* and the expression $\vec{f} = (f_1, f_2) \succ (\phi_1^I, \phi_2^N)$ denotes that (f_1, f_2) dominates (ϕ_1^I, ϕ_2^N). In general, we say that $\vec{f} = (f_1, f_2)$ dominates $\vec{f'} = (f_1', f_2')$, with $\vec{f}, \vec{f'} \in \Phi$ if and only if (iff) $f_1 \leq f_1'$ and $f_2 \leq f_2'$, where at least one inequality is strict. Thus, $\vec{f} = (f_1, f_2) \succ (\phi_1^I, \phi_2^N)$ is true when a) $f_1 < \phi_1^I$ and $f_2 < \phi_2^N$ or b) $f_1 < \phi_1^I$ and $f_2 = \phi_2^N$ or c) $f_1 = \phi_1^I$ and $f_2 < \phi_2^N$. Since a) and b) contradict the definition of an ideal point and since c) contradicts the definition of a nadir point, then $(\phi_1^I, \phi_2^N) \in \mathcal{F}$. The proof of $(\phi_1^N, \phi_2^I) \in \mathcal{F}$ is analogous. □

Lemma 2. *For each $(f_1, f_2) \in \Phi$, if $(f_1, f_2) \in \mathcal{F}$, then $\phi_1^I \leq f_1 \leq \phi_1^N$ and $\phi_2^I \leq f_2 \leq \phi_2^N$.*

Proof. As proved in Lemma 1, $(\phi_1^I, \phi_2^N) \in \mathcal{F}$, and thus it is non-dominated. Since $\phi_1^I = \min(f_1)$, $f_1 \geq \phi_1^I, \forall\ (f_1, f_2) \in \mathcal{F}$. Moreover, if $f_2 > \phi_2^N$, $(\phi_1^I, \phi_2^N) \succ (f_1, f_2)$ and $(f_1, f_2) \notin \mathcal{F}$. Hence, $f_1 \geq \phi_1^I$ and $f_2 \leq \phi_2^N\ \forall\ (f_1, f_2) \in \mathcal{F}$. The proof for $\phi_2^I \leq f_2 \leq \phi_2^N$ is analogous. □

According to Lemma 1 and Lemma 2, Algorithm 1 generates the exact Pareto front of the problem described in (8).

Theorem 2. *Algorithm 1 generates one feasible solution for each point of the Pareto front.*

Proof. Let us denote the sequence of solutions of Algorithm 1 by $\{\vec{f}_1^*, \ldots, \vec{f}_m^*, \ldots, \vec{f}_M^*\}$, where, e.g., $\vec{f}_m^* = \big((f_m^*)_1, (f_m^*)_2\big)$, with 1, 2 denoting the first and the second objective, respectively. We have to prove that if $\vec{f} \in \Phi \setminus \{\vec{f}_1^*, \ldots, \vec{f}_m^*, \ldots, \vec{f}_M^*\}$, then $\vec{f} \notin \mathcal{F}$. Let us assume that there is a solution $\vec{f'} = (f_1', f_2') \in \Phi \setminus \{\vec{f}_1^*, \ldots, \vec{f}_m^*, \ldots, \vec{f}_M^*\}$ such that $\vec{f'} \in \mathcal{F}$. By Lemma 2, for the first objective we have $\phi_1^I \leq f_1' \leq \phi_1^N$. Thus, either a) $f_1' = (f_m^*)_1$ for a given $m = 1 \ldots M$ or b) $(f_{m-1}^*)_1 < f_1' < (f_m^*)_1$ and $(f_{m-1}^*)_2 < f_2' \leq (f_m^*)_2$ for a given $m = 1 \ldots M$. In the first case (i.e., case a), f_2' must be lower than $(f_m^*)_2$ for $\vec{f'}$ to be efficient. However, since $\Delta = 1$ and the second objective is integer by definition, $\exists\ \varepsilon^{m'}$ that will eventually reach a value for which the optimum of the corresponding ε-constraint problem is $\vec{f'}$ for $m+1 \leq m' \leq M$, that is $\vec{f'} \in \{\vec{f}_{m+1}^*, \ldots, \vec{f}_M^*\}$, which contradicts the hypothesis. Regarding the second case (i.e., case b), f_2' must be such that $(f_{m-1}^*)_2 < f_2' \leq (f_m^*)_2$, which is impossible since $\vec{f}_m^*$ is the optimal value of problem (8), with $\varepsilon^m = \varepsilon^{m-1} - \Delta$, $\Delta = 1$, and the second objective is integer. □

Algorithm 1 Exact Pareto front calculation of problem (8)

1: Calculate the ideal and nadir points, $\vec{\phi}^I$ and $\vec{\phi}^N$.
2: Add $\vec{f}_1^* = (\phi_1^I, \phi_2^N)$ to $\mathcal{F}$.
3: Set $m = 2$.
4: Set $\varepsilon^m = \phi_2^N - \Delta$, with $\Delta = 1$.
5: **while** $\varepsilon^m \geq \phi_2^I$ **do**
6: Solve problem (8) and add the optimal solution value $\vec{f}_m^* = \big((f_m^*)_1, (f_m^*)_2\big)$ to $\mathcal{F}$.
7: Set $\varepsilon^{m+1} = (f_m^*)_2 - \Delta$.
8: Set $m = m + 1$.
9: **end while**
10: Remove dominated points if required.

Some dominated solutions may be generated by the sequence of subproblems derived according to Theorem 2. However, since all dominated points can be identified, one can simply exclude the non-efficient solutions to obtain the exact Pareto front. Furthermore, although Algorithm 1 limits the number of subproblems, a subproblem may be very hard to solve. This stems from the fact that an exhaustive search would require the examination of C^N possible solutions, which results in prohibitive complexity ($\mathcal{O}(n^n)$), as the number of BSs, C, and the number of UEs, N, increase. Therefore, alternative algorithms, available in the literature, should be used, able to come up with very close to the optimal solutions with acceptable computational complexity [17]. In this work, we applied a meta-heuristic method [26], which has been shown to lead to high-quality solutions (the average gap is less than 1% with respect to best-known solutions) in almost real time. The applied method uses biased randomization together with an iterated local search meta-heuristic algorithm. Although the meta-heuristic algorithm involves lower complexity than $\mathcal{O}(n^n)$[7], it still requires a high number of iterations (50000 in our case). Therefore, there is need for low-complexity algorithms, able to achieve solutions close to the Pareto front.

[7]Meta-heuristics have no predefined end, and thus big O notation cannot be used to describe their complexity. Yet, they can be compared empirically (through number of objective function evaluations/iterations).

Algorithm 2 Proposed energy efficient user association algorithm

Input: $\mathcal{N}, \mathcal{C}, SINR_{ij}, r_{i,net}, b, c_{j_{max}}, \mathcal{L}_l, \mathcal{L}$

1: Calculate c_{ij} as in (4)
2: $Candidates_i \leftarrow j : c_{ij} \leq min(c_{ij_{min}} + c_{thres}, c_{j_{max}})$
3: Sort all UEs i by $Candidates_i$ size in ascending order
4: Calculate $P_{tot_{ij}} = P_{AN_{ij}} + P_{BH_{ij}}$ using (4), (6) $\forall j \in Candidates_i$
5: Sort the UEs with the same candidate number by the $P_{tot_{ij}}$ difference among their candidates in descending order
6: Sort $Candidates_i$ by $P_{tot_{ij}}$ in ascending order
7: Choose the candidate with the minimum $P_{tot_{ij}}$
8: **if** the chosen BS has sufficient spectrum resources **then**
9: Associate the UE to it
10: Update remaining spectrum resources
11: **else**
12: Move to the next candidate and repeat the process
13: **end if**

IV. Proposed heuristic Algorithm

In this section, we propose an algorithm that aims at a *good* trade-off between energy and spectrum efficiency, while inducing low complexity in the system. The proposed algorithm takes into account the available context-aware information, i.e., the UEs' measurements (SINR) and requirements ($r_{i,net}$), the HetNet architecture ($s_{\mathcal{L}_l j}$) and the available spectrum resources of each BS ($c_{j_{max}}$) to efficiently associate the UEs.

This context-aware information, which can be divided into information being reported by the network and information being reported by the UEs, can be easily available to all nodes in a LTE-A network (i.e., eNBs and/or SCs) [4], [27]. In particular, the information being reported by the network does not impose additional constraints, since the standard defines the X2 logical interface to allow the exchange of information among BSs (eNBs and/or SCs) [4]. Moreover, the information about the network architecture (s_{L_l}) requires very limited or nil update due to its static nature. Hence, the only additional information to be exchanged is the current traffic of each BH link. Regarding the information being reported by UEs, each UE is required to measure the SINR received from the neighboring BSs. For such a purpose, Release 8 has already defined the radio resource management (RRM) measurement set, i.e., the set of BSs from which a UE measures and reports parameters, such as RSRP or reference signal received quality (RSRQ). Later on, in order to support coordinated multi-point (CoMP), Release 10 defined a subset of the RRM measurement set, namely CoMP measurement set, to allow the UEs to measure and report short-term channel state information [27]. Thereby, the aforementioned mechanisms guarantee the availability of the required information.

The proposed algorithm, which is summarized in Algorithm 2, aims at the maximization of the energy efficiency given a specific spectral efficiency target. From this point on, we will refer to it as energy efficient (EE) user association algorithm.

As shown in Algorithm 2, EE considers as candidate cells for a UE i the set of cells, denoted by $Candidates_i$, that satisfy its rate requirements with fewer PRBs (c_{ij}) than a target c_{thres}=δ $c_{ij_{min}}$ (line 2). The spectral efficiency target is defined by the tuning parameter δ >0, which controls the deviation in the number of needed PRBs from the association that requires the fewest. For instance, selecting δ=0, and thus, c_{thres}=0, would result in the maximum spectral efficiency, while δ >0 would decrease the spectral efficiency accordingly in favor of higher energy efficiency. Note also that a BS j cannot be included in the candidates of a UE i, if $SINR_{ij}$ is too low and hence c_{ij} >$c_{j_{max}}$. To ensure that all the UEs will be associated, EE sorts the UEs by their number of candidates and starts with the UEs with the fewest candidates (line 3).

In order to maximize the network energy efficiency, EE calculates for each UE i and candidate cell j the total RF transmit power consumption needed for the traffic of the UE i to be served, denoted by $P_{tot_{ij}}$=$P_{AN_{ij}}$+$P_{BH_{ij}}$, (line 4). EE then sorts the UEs with the same $Candidates_i$ size by the difference in $P_{tot_{ij}}$ between the candidate cells in descending order, i.e., starting with the UE with the maximum difference between the first and the second candidate (line 5). Thereafter, EE sorts the candidate cells of each UE i by $P_{tot_{ij}}$ in ascending order (line 6) and associates the UE to the candidate cell, which involves the minimum power consumption, as long as it has sufficient spectrum resources to serve it (lines 8). Otherwise, it moves to the next candidate (line 12). Every time a UE is associated with a BS j, the algorithm updates the remaining spectrum resources of j. Contrary to the algorithm providing the exact Pareto front solutions, presented in Section III, the proposed heuristic algorithm is much less complex, i.e., $\mathcal{O}(n \log n)$ [28].

EE may be executed in each eNB sector at a specific time interval based on the dynamics of the UE traffic, so that the system performance is optimized. If a new UE becomes active in the meantime (i.e., after the last execution of the algorithm and before the next one), its association can be decided by EE given the associations of the rest of the UEs. In particular, Algorithm 2 is applied, excluding lines 3 and 5. Thereby, the proposed algorithm can provide high network scalability.

V. Simulation Results

A. Simulation Scenario

In the extensive simulations we executed in MATLAB®, we considered an eNB sector area, as depicted in Fig. 1, that overlaps with N_{cl}=2 clusters. Each cluster consists of 4 SCs (SC_1=SC_2=4) according to 3GPP [18]. Moreover, according to [18], the SC clusters are uniformly distributed within the eNB sector, and the SCs of each cluster are uniformly dropped within the cluster area. The minimum distance between two SCs is 20 m and between the eNB and a SC cluster center is 105 m. The minimum distance of a UE from the eNB is 35 m and from a SC is 5 m. In addition, in each cluster, one SC (the one being the closest to the eNB) is considered one hop away from the eNB site and thus plays the role of the aggregator of the cluster traffic, two SCs (the ones being the closest to the aggregator) are considered two hops away from the eNB site and the last SC is considered three hops away and connected to the closest two-hop-away SC of the cluster (see Fig. 1).

In order to gain further insights into the benefits of mmWave, we consider three different BH technologies: i) LOS

TABLE I
SIMULATION VALUES

Parameter	Value	Parameter	Value	Parameter	Value
f_{AN}	2.0 GHz	$P_{eNB_{max}}$	46 dBm	NF_{UE}	9 dB
B_{eNB}, B_{SC}	10 MHz	$P_{SC_{max}}$	30 dBm	NF_{BH}	5 dB
$c_{eNB_{max}}$, $c_{SC_{max}}$	50	$P_{BH_{max}}$	47 dBm	N_{th}	-174 dBm/Hz
$L_{p_{eNB}}$	69.55+26.16 logf_{AN}-13.82 logh_{eNB}-C_H+(44.9- 6.55 logh_{eNB}) log$d$, d in km	$G_{T_{xeNB}}$	17 dBi	h_{eNB}	25 m
$L_{p_{SC}}$	69.55+26.16 logf_{AN}-13.82 logh_{SC}+(44.9- 6.55 logh_{SC}) log$d$, d in km	$G_{T_{xSC}}$	5 dBi	h_{SC}	2.5 m
L_{p_o}	57.5 dB	$G_{T_{xUE}}$	0 dBi	h_m	1.5 m
C_H	0.8+ (1.1 logf_{AN}- 0.7) h_m -1.56 logf_{AN}	IL	2 dB		

mmWave links (f_{BH1}= 60 GHz) of B_{BH1}= 200 MHz channel bandwidth [24], ii) LOS microwave[8] links (f_{BH2}= 28 GHz) of B_{BH2}= 28 MHz [29], and iii) sub-6GHz (f_{BH3}= 3 GHz) of B_{BH3}= 10 MHz [30]. For a fair comparison, the path loss models of the provided references are used, while the antenna gains are selected equal to 37, 24 and 19 dBi, respectively. Due to the static BH nature, we assume that frequency planning among adjacent BH links is performed during the deployment phase, so that the generated interference is mitigated.

In each realization (1000 in total), we consider N UEs of different GBR requirements. Specifically, 60% of UEs demand 1.024 Mbps, 30% 2.560 Mbps and 10% 3.328 Mbps [19]. The following UE traffic distribution scenarios are considered:

- Uniform: the UEs are uniformly distributed in the sector area of radius R= 500 m.
- Hotspot: 2/3 of UEs are uniformly dropped within the clusters (in a radius r= 70 m from cluster center) and 1/3 of UEs are uniformly dropped in the eNB sector [18].

The proposed work, as explained in Section II-A, is independent of the employed channel allocation scheme. Therefore, for the sake of simplicity and without loss of generality, we assume that inter-sector interference is mitigated through some form of fractional frequency reuse scheme or sophisticated frequency allocation [31] and that the channels allocated to the eNB are orthogonal to the channels allocated to SCs. However, SCs belonging to different clusters reuse the same bands, thus interfering to each other.

The rest of the simulation parameters are summarized in Table I, where the subscript x={eNB, SC} refers to the eNB or to a SC, respectively. Then, f_{AN} denotes the frequency used in the AN, while B_x is the bandwidth allocated to x and h_x is the antenna height of x. The parameter h_m is the mobile antenna height, while C_H is the antenna height correction factor and d is the distance between the BS and the UE. According to LTE, $BLER$= 0.1 [22] and k_{ov}= 0.13 [23]. The slow fading is modeled by a log-normal random variable with zero mean and deviation 8 dB for the eNB and 10 dB for the SC signal.

B. Pareto front solutions

Following the general description of Section V-A, we consider two different simulation scenarios, as depicted in Fig. 2 a) and b). In the first scenario, the UEs are uniformly distributed, while in the second they form hotspots. To that end, in Fig. 3 a) and b) the exact Pareto front points of the problem in (8) are depicted for the considered BH technologies.

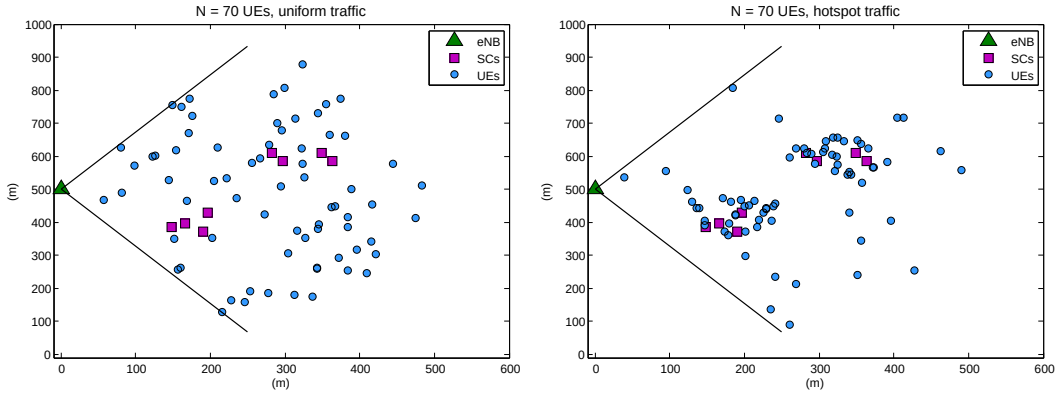
As already mentioned, the number of PRBs and the power consumption are two metrics that can not be minimized at the same time, and thus a *good* trade-off between them has to be found. Hence, each Pareto front point corresponds to a dominant solution of the ε-constraint problem for a different ε value, as described in Theorem 2. In general, in multi-objective optimization, none of the Pareto front solutions is better than the others. However, depending on the preference for each of the conflicting objectives, a Pareto front solution may be more preferable than another. For instance, in (8), the preference for one objective (f_1 or f_2) may vary based on the network state. In scenarios where spectral efficiency becomes important, e.g., in highly loaded scenarios, the operators may select a point near the right extreme Pareto front solution to maximize the spectral efficiency (f_2= ϕ_2^I). On the other hand, when the spectrum resources do not limit the system (except for (8c)), the operators could select a point near the left Pareto front solution (f_1=ϕ_1^I), thus minimizing the energy consumption.

In the considered example, for all BH technologies when f_1=ϕ_1^I (maximum energy efficiency), most UEs are associated to SCs to minimize the AN power consumption (the AN power consumption is much higher when a UE is associated to the eNB than to a SC). Moreover, the UE association with the SC that involves the minimum BH power consumption (e.g, the one with the fewest hops or shortest BH links) is favored. Therefore, when $f_1 = \phi_1^I$ (maximum energy efficiency), the number of required PRBs is higher in the uniform scenario (than in the hotspot), since the UEs are located further from the SC cluster centers. On the contrary, when $f_2 = \phi_2^I$ (maximum spectral efficiency), more UEs are associated to the eNB to reduce the required PRBs at the expense of higher AN energy consumption. Thus, the AN power consumption increase is higher in the uniform scenario for all BH technologies, as more UEs are associated to the eNB.

Regarding the rest of the Pareto front points, we notice that in the uniform scenario, for the same RF transmit power consumption as in the hotspot, more PRBs are required for all BH technologies. This stems from the fact that the UEs located in a hotspot mostly get associated with SCs both to use fewer PRBs, and to have much less AN power consumption. On the contrary, when the UEs are uniformly distributed, they are located further from the SC clusters and thus to decrease the RF transmit power consumption, a proportional PRB increase is needed. This results in a steeper Pareto front curve for the hotspot scenario, i.e., the hotspot Pareto front points provide better trade-offs between the two objectives than the uniform.

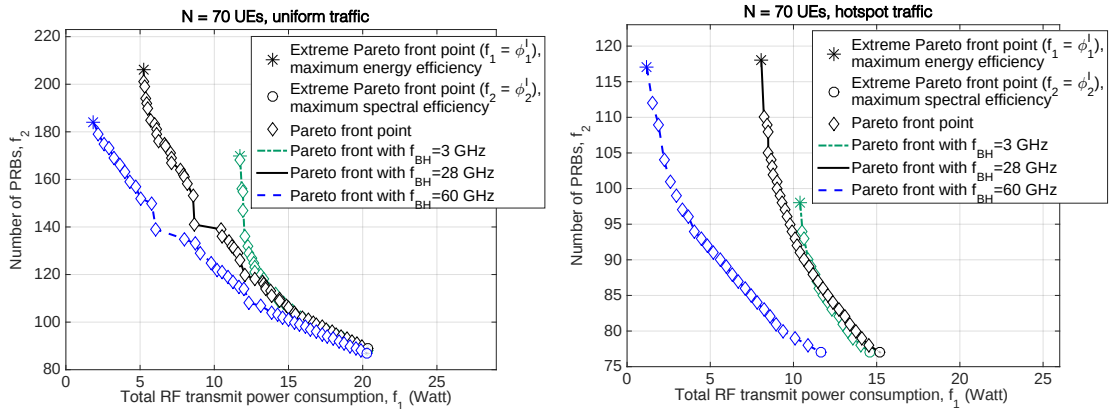
Among the different BH technologies, mmWave presents the best performance, since its Pareto front is shifted on the left. This implies that mmWave can provide better trade-offs than the rest of the BH technologies. Although mmWave experiences the highest path loss, it is able to send high amount

[8]In [29], 28 GHz is considered as mmWave. Still, in this paper, we adopt the SC Forum categorization (Fig. 5-2 in [1]).



(a) (b)

Fig. 2. Snapshots of (a) uniform and (b) hotspot traffic distribution scenarios with N=70 UEs.



(a) (b)

Fig. 3. Pareto front: number of PRBs vs. total RF transmit power consumption for N=70 UEs with (a) uniform and (b) hotspot traffic for different BH technologies with frequency equal to 3 GHz, 28 GHz and 60 GHz, respectively.

of data without increasing the transmitted power due to its high bandwidth availability. Therefore, mmWave achieves the highest performance gains for hotspot scenarios, where higher BH traffic is generated. On the contrary, the available bandwidth of sub-6GHz is very limited. Consequently, for higher BH traffic, a significant increase in the RF transmit power of the BH links is required, so that the SINR at the receiver increases. Thereby, higher order modulation and coding schemes can be used, which result in higher spectral efficiency. However, the much higher RF transmit power consumption results in lower energy efficiency. Still, 3 GHz outperforms 28 GHz for low BH traffic, e.g., when only the UEs very closely located to SCs are associated with them, as it presents lower path loss.

C. Performance Evaluation

In this section, we compare the performance of the proposed algorithm with both the state-of-the-art and the optimal (yet complex) solutions of Algorithm 1 for all BH technologies. The algorithms under study are summarized in the following.

- ε-constraint[9]: the two extreme Pareto front analytical solutions of the ε-constraint problem described in Section III. In particular, we refer with ε-constraint EE to the extreme Pareto front solution that maximizes the energy efficiency and with ε-constraint SE to the Pareto front solution that maximizes the spectral efficiency.
- EE: the proposed energy efficient algorithm, described in Section IV, with c_{thres}=0, 1, 2.
- BH-aware: the association algorithm proposed in [16].
- RSRP: a UE is associated with the BS from which it receives the strongest reference signal [4].
- Range expansion (RE): a $bias = 13$ dB is added to the RSRP if the signal comes from a SC [6], [31].
- Minimum path loss (MPL): a UE is associated with the BS from which it has the minimum path loss ($L_{ij} = L_{p_{ij}} + L_{f_{ij}}$) [14], independently of its received power.

In Fig. 4, the average network energy efficiency is depicted for all algorithms and BH technologies versus the number of UEs, N, under uniform traffic. In general, it can be noticed that mmWave achieves much higher energy efficiency than the rest of the BH technologies (i.e., 40% higher than 3 GHz and 2 times higher than 28 GHz) for all algorithms and N values. This is due to its high bandwidth availability which results in much lower BH power consumption (of the order of mW).

Regarding the user association algorithms, it is reminded that ε-constraint EE shows the maximum energy efficiency that can be achieved independently of the spectral efficiency,

[9]Under overloaded network conditions, when (8) had no feasible solution, we were relaxing constraint (8b) and solving the relaxed problem for ε-constraint EE and ε-constraint SE, while measuring the blocking probability in each case (i.e., percentage of UEs that were not served).

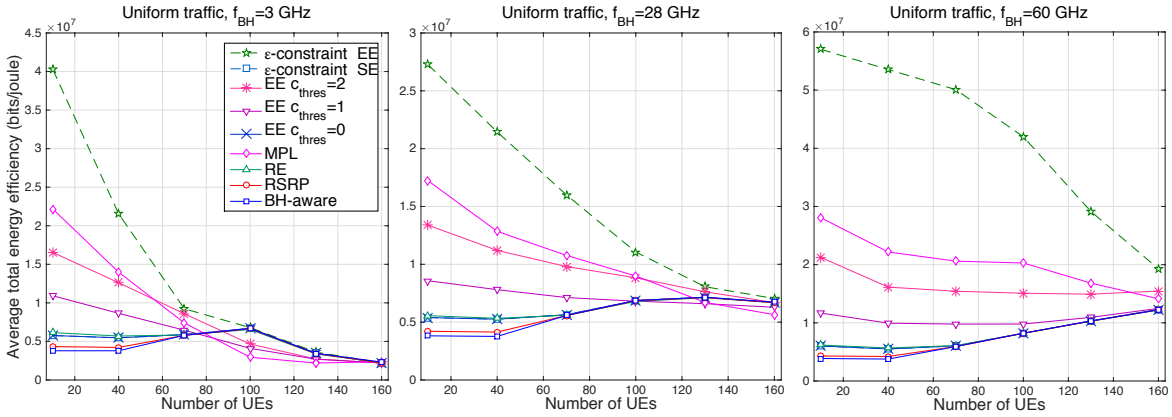


Fig. 4. Average total network energy efficiency for different N values and BH technologies (3, 28 and 60 GHz), when the UEs are uniformly distributed.

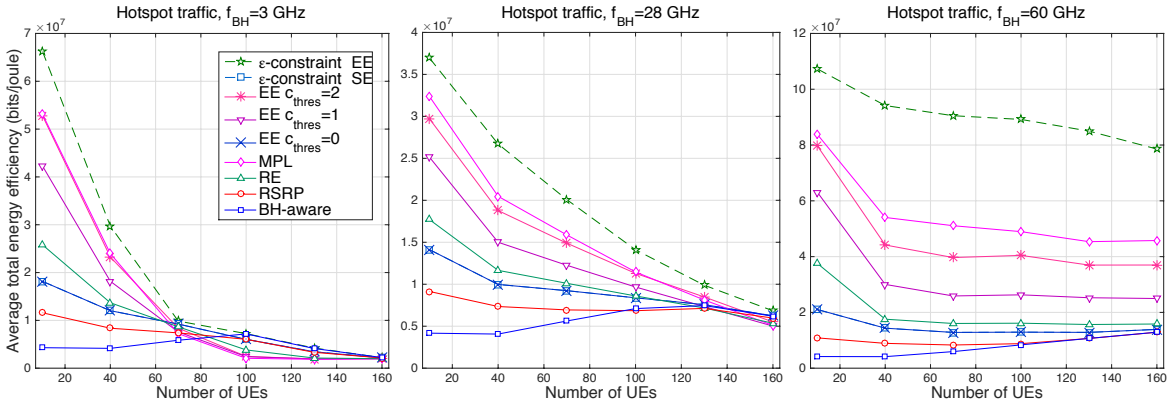


Fig. 5. Average total network energy efficiency for different N values and BH technologies (3, 28 and 60 GHz), when the UEs form hotspots.

while ε-constraint SE corresponds to the maximum energy efficiency given that the spectral efficiency is maximized. These solutions achieve better performance than the state-of-the-art (e.g., ε-constraint EE shows up to 10.5, 7, and 14.5 times higher energy efficiency for 3, 28 and 60 GHz, respectively). However, unlike the rest of the algorithms, they present very high complexity, which increases with an increasing number of UEs and BSs, as discussed in Section III.

As the network traffic increases, the gap between ε-constraint EE and ε-constraint SE decreases, until the network reaches saturation and thus the most energy-efficient solution is also the most spectrum-efficient one. As shown in Fig. 4, the network reaches saturation at an earlier point, i.e., for lower traffic, in 3 GHz compared to the other technologies. This stems from the fact that the maximum BH capacity (in terms of maximum RF transmit power) is reached earlier for 3 GHz, due to the much lower bandwidth availability at this frequency.

In Fig. 4, it can be also noticed that for all BH technologies the proposed low complexity ($\mathcal{O}(n \log n)$) EE algorithm outperforms the state-of-the-art (except for MPL for low traffic, which associates most UEs to SCs, leading to lower AN power consumption, at the expense, however, of much lower spectral efficiency, as it will be shown later on), while achieving similar performance to the ε-constraint solutions. Nevertheless, the selection of the parameter value c_{thres} is important. EE with $c_{thres} = 0$ achieves equal performance to the ε-constraint SE, while as c_{thres} increases and the system is not overloaded, the performance of the algorithm in terms of energy efficiency is improved at the expense of lower spectral efficiency. However, when the system becomes saturated a higher threshold would result in lower energy efficiency, since not all the traffic demands of UEs could be served (i.e., non-zero blocking probability). Thus, for maximum performance, the threshold should be adapted dynamically based on the network conditions, i.e., a high threshold value should be selected in low traffic scenarios, and a low value otherwise.

As for the rest of the algorithms, they achieve lower performance for all BH technologies. In particular, BH-aware gives priority to the candidate cell with the fewest hops to reach the core network and thus most of the UEs get connected to the eNB. Hence, similar to RSRP, for low values of N, although the BH energy consumption is zero, there is high AN energy consumption (we remind that the power per subcarrier is much higher for the eNB than for a SC). On the contrary, EE takes into account the possibility of having heterogeneous BH links and adapts the user association accordingly. Thus, it presents lower dependency on the employed scenario. Regarding RE, it achieves almost the same performance as EE with c_{thres}=0, as there are more UEs associated with SCs, resulting in lower AN energy consumption. However, this comes at the expense of much lower spectral efficiency, as it will be shown later on.

Accordingly, in Fig. 5, the average network energy efficiency of all algorithms is depicted in a hotspot scenario for all BH technologies. In this scenario, mmWave achieves even higher gains than in the uniform (i.e., 60% higher than 3 GHz and 3 times higher than 28 GHz) for all algorithms and N

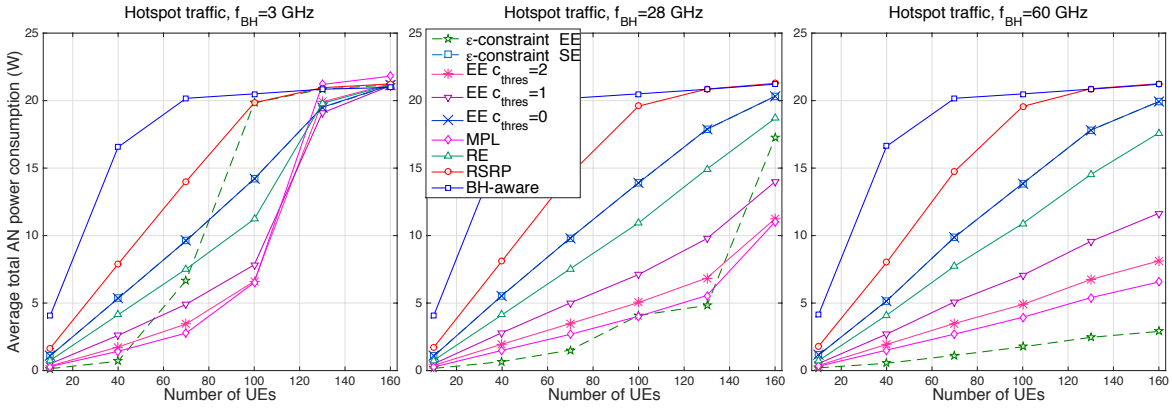


Fig. 6. Average total RF transmit power consumption in the access network for different N values and BH technologies with hotspot traffic.

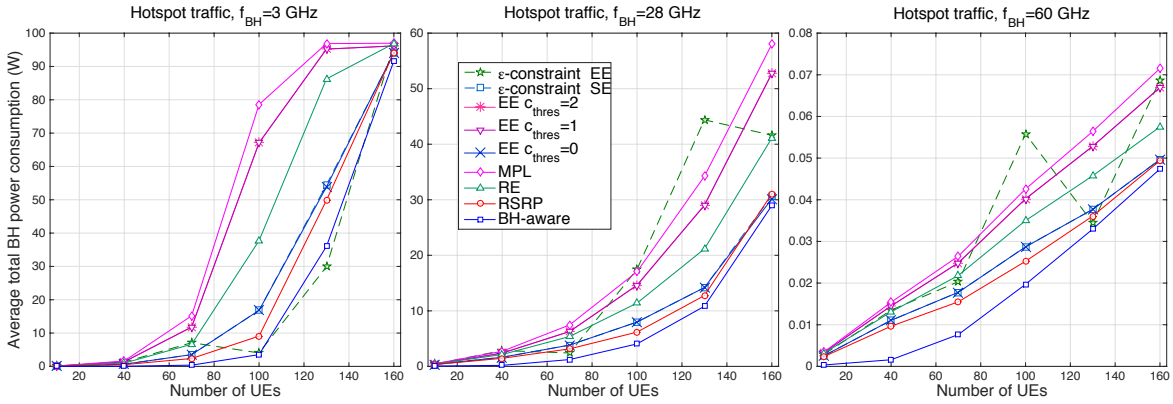


Fig. 7. Average total RF transmit power consumption in the backhaul network for different N values and BH technologies with hotspot traffic.

values. This stems from the fact that, in hotspot scenarios, the BH traffic increases and thus the available bandwidth of each technology becomes more important. Due to the same reason, in the hotspot scenario, the gap between ε-constraint EE and ε-constraint SE for 3 GHz decreases at a higher rate. However, it reaches saturation at a later point, i.e., for higher traffic, than in the uniform scenario, as the AN in hotspot scenarios gets saturated at a lower rate (most UEs get connected to SCs). This is the reason why, in 60 GHz, where BH capacity is not the network bottleneck, the gap decreases more smoothly than in the uniform. The proposed solutions also achieve higher gains (up to 15.5, 9, and 26 times higher energy efficiency for 3, 28 and 60 GHz, respectively) compared to the state-of-the-art than in the uniform scenario. To gain further insights into that, the AN and BH power consumption of all algorithms and BH technologies are depicted in Fig. 6 and 7, respectively.

As it can be observed in Fig. 6, the AN power consumption increases as N increases for all algorithms and BH technologies. For BH-aware, the AN power consumption increases initially at a high rate, as more UEs get connected to the eNB. Yet, for very high traffic the eNB becomes saturated, and thus more UEs get associated with SCs, which results in a smoother AN power consumption increase. As depicted in Fig. 7, the BH energy consumption also increases for all algorithms (except for ε-constraint EE), as N increases, since higher BH traffic is generated and thus higher energy consumption.

In general, ε-constraint EE favors the user association that minimizes the total RF transmit power consumption at a specific instant, and thus, it presents a different behavior than the rest of the algorithms. In particular, for low traffic, ε-constraint EE favors the association of most UEs with SCs and especially with the SC cluster located closer to the core network to minimize both the AN and BH power consumption as well as the number of PRBs required. For higher traffic, however, which differs for different technologies (N=100 UEs for 3 GHz and N=70, 160 UEs for 28 GHz and N=130 UEs for 60 GHz), the BH aggregated traffic increases a lot (we remind that the power consumption of a BH link increases in an exponential way with the traffic that passes through the link) and therefore the association of a portion of UEs with the eNB is preferable in order to avoid a significant increase in the BH power consumption. This is also due to the fact that the association with the eNB at this point gives the possibility of switching off one or even both SC clusters (in the case all UEs can be served by the eNB), thus resulting in higher energy efficiency gain. It is worth noting that the most energy-consuming links in the considered model are the links that are one hop away from the core network, which not only aggregate all the traffic of the cluster but also may be much longer than the rest of the BH links. Therefore, the complete switch off of a cluster corresponds to the highest energy efficiency gain.

Regarding the rest of the algorithms, it is shown that ε-constraint SE achieves a good balance between AN and BH power consumption and so does the proposed algorithm. MPL presents high energy efficiency for low traffic, as more UEs are associated to SCs than in RSRP and RE, resulting in lower AN power consumption and higher spectral efficiency than in the uniform scenario, as it will be shown later on. For high

TABLE II
AVERAGE NETWORK SPECTRAL EFFICIENCY (SE) AND BLOCKING PROBABILITY (BP)

	Uniform						Hotspot					
	3 GHz		28 GHz		60 GHz		3 GHz		28 GHz		60 GHz	
User association algorithm	**SE (bps/Hz)**	**BP (%)**	**SE (bps/Hz)**	**BP (%)**	**SE (bps/Hz)**	**BP (%)**	**SE (bps/Hz)**	**BP (%)**	**SE (bps/Hz)**	**BP (%)**	**SE (bps/Hz)**	**BP (%)**
ε-constraint SE	3.50	2.08	3.53	0	3.53	0	4.34	2.42	4.33	0.02	4.34	0
EE $c_{thres}=0$		2.29						2.59		0.07		
BH-aware		2.68		0.16		0.21	4.32	2.72		0.09		
RSRP	3.39	2.36	3.47	0.03	3.47	0.01	4.14	3.0	4.19	0.1	4.20	
RE	3.24	2.40	3.31		3.30		3.87	4.17	3.85		3.89	
MPL	1.95	3.75	1.99	1.21	1.98	1.21	2.96	4.62	3.03	0.05	3.08	
EE $c_{thres}=1$	3.01	2.27	3.10	0	3.12	0	3.83	3.96	3.79	0.03	3.81	
EE $c_{thres}=2$	2.61	2.40	2.64	0.03	2.67		3.53	4.24	3.49	0.04	3.53	
ε-constraint EE	2.21	2.08	1.70	0	1.58		3.74	2.42	2.78	0.02	2.63	

traffic in 3 GHz, however, the BH power consumption of MPL increases a lot (see Fig. 7) due to low bandwidth availability at this frequency band leading to very low energy efficiency.

In Table II, the average network spectral efficiency as well as the average blocking probability is presented for all algorithms and BH technologies. As it can be observed, the considered algorithms that aim at the maximization of the spectral efficiency (i.e., ε-constraint SE, EE with $c_{thres}=0$, BH-aware) achieve the highest spectral efficiency for all BH technologies, since the UEs are connected to the BSs that require the minimum spectrum resources for their QoS requirements to be fulfilled. On the contrary, RSRP and RE achieve slightly lower spectral efficiency, as the UEs, under high traffic load conditions, may be connected to BSs that require more spectrum resources. MPL, unlike the rest of the algorithms, presents much lower spectral efficiency, since it associates the UEs independently of their SINR. Hence, it is very likely that a UE is associated to a BS with low SINR, thus requiring more spectrum resources to achieve the same throughput. This holds also for EE $c_{thres}=1,2$ since energy efficiency is increased at the expense of lower spectral efficiency, which becomes even lower in ε-constraint EE, where energy efficiency is maximized. However, in overloaded networks (e.g., hotspot traffic in 3 GHz), ε-constraint EE achieves higher spectrum efficiency in order to ensure lower blocking probability.

In terms of blocking probability, the optimal solutions ε-constraint EE and ε-constraint SE present always the highest performance. The performance of the EE algorithm, however, depends on the selected threshold value and the employed scenario, as already explained. Specifically, under network overloading conditions, higher threshold values result in lower spectral efficiency as well as higher blocking probability.

Regarding the different BH technologies, notice that for 60 GHz all algorithms present the lowest blocking probability. This stems from the fact that, contrary to the other BH technologies, mmWave links do not become the network bottleneck due to their very high bandwidth availability. On the other hand, 3 GHz shows the worst performance especially for hotspot scenarios where the BH traffic is higher.

VI. CONCLUSION

We studied the user association problem in a HetNet, where several SCs forward their traffic through the BH to the neighboring SCs until it reaches the core network. We aimed at the joint maximization of network energy and spectrum efficiency, without compromising the UE QoS. The problem was formulated as an ε-constraint problem, which considers both the AN and BH energy consumption. The trade-off between energy and spectrum efficiency was analytically studied by deriving the exact Pareto front points of the problem for different BH technologies. The provided solutions can be used as a benchmark for the performance evaluation of user association algorithms. Moreover, a low-complexity adaptive algorithm was proposed, which was shown to be able to select any point of the Pareto front, by accordingly modifying the spectral efficiency target c_{thres}, and thus, to achieve a *good* trade-off between the aforementioned metrics. The proposed algorithm was also compared with existing user association solutions under different BH technologies. Our results indicated that i) the proposed algorithm achieves notable energy and spectrum efficiency gains and that ii) mmWave is a promising solution for high capacity and low energy consumption multi-hop BH.

ACKNOWLEDGMENT

This work was supported by the Research Projects 2014-SGR-1160, 2014-SGR-1551, CellFive (TEC2014-60130-P), DEFINE-5G (TEC2014-60258-C2-2-R), and CROSSFIRE (MITN-317126).

REFERENCES

[1] "Backhaul Technologies for Small Cells: Use Cases, Requirements and Solutions," Small Cell Forum, 049.01.01, Feb. 2013.

[2] L. Wei, R. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5G wireless systems," *IEEE Wireless Commun.,* vol. 21, no. 6, pp. 136-143, Jan. 2015.

[3] "Small Cell Backhaul Requirements," NGMN Alliance, v. 1.0, Jun. 2012.

[4] *E-UTRA and E-UTRAN; Overall Description; Stage 2,* 3GPP TS 36.300, v. 11.5.0, Rel. 11, Mar. 2013.

[5] A. Mesodiakaki, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy-efficient User Association in Cognitive Heterogeneous Networks," *IEEE Commun. Mag.,* vol. 52, no. 7, pp. 22-29, Jul. 2014.

[6] "DL pico/macro HetNet performance: cell selection," Alcatel-Lucent, R1-100945, Meeting 60, Feb. 2010.

[7] Q. Ye *et al.,* "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706-2716, Jun. 2013.

[8] C. Liu, P. Whiting, and S. V. Hanly, "Joint Resource Allocation and User Association in Downlink Three-tier Heterogeneous Networks," in *Proc. IEEE GLOBECOM,* Dec. 2014, pp. 4232-4238.

[9] Y. Lin, W. Bao, W. Yu, and B. Liang, "Optimizing User Association and Spectrum Allocation in HetNets: A Utility Perspective," *IEEE J. Sel. Areas Commun.,* vol. 33, no. 6, pp. 1025-1039, May 2015.

[10] H. Boostanimehr and V. K. Bhargava, "Unified and distributed QoS-driven cell association algorithms in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1650-1662, Mar. 2015.

[11] H. Galeana-Zapien and R. Ferrus, "Design and Evaluation of a Backhaul-aware Base Station Assignment Algorithm for OFDMA-based Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3226-3237, Oct. 2010.

[12] Q. Li, R. Q. Hu, G. Wu, and Y. Qian, "On the optimal mobile association in heterogeneous wireless relay networks," in *Proc. IEEE INFOCOM,* Mar. 2012, pp. 1359-1367.

[13] S. Corroy, L. Falconetti, and R. Mathar, "Cell Association in Small Heterogeneous Networks: Downlink Sum Rate and Min Rate Maximization," in *Proc. IEEE WCNC,* Apr. 2012, pp. 888-892.

[14] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous wireless cellular networks," *IEEE Trans. Wireless Commun.,* vol. 12, no. 1, pp. 248-257, Oct. 2012.

[15] N. Wang, E. Hossain, and V. K. Bhargava, "Joint Downlink Cell Association and Bandwidth Allocation for Wireless Backhauling in Two-Tier HetNets with Large-Scale Antenna Arrays," *IEEE Trans. Wireless Commun.*, Jan 2016.

[16] A. Mesodiakaki, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy-efficient Context-aware User Association for Outdoor Small Cell Heterogeneous Networks," in *Proc. IEEE ICC,* Jun. 2014, pp. 1614-1619.

[17] K.-M. Miettinen, *Nonlinear Multi-objective Optimization,* Kluwer Academic Publishers, Boston, 1999.

[18] *Small cell enhancements for E-UTRA & E-UTRAN-Physical layer aspects*, 3GPP TR 36.872, v. 1.0.0, Rel. 12, Aug. 2013.

[19] *Study on Small Cell enhancements for E-UTRA and E-UTRAN; Higher layer aspects*, 3GPP TR 36.842, v. 12.0.0, Rel. 12, Dec. 2013.

[20] H. Holma and A. Toskala, Eds., *LTE for UMTS: OFDMA and SC-FDMA based radio access,* Wiley, Mar. 2009.

[21] 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.

[22] *Terminal conformance specification; Radio transmission & reception (FDD)*, 3GPP TS 34.121, v. 6.1.0, Rel. 6, Jun. 2005.

[23] P. Mogensen *et al.,* "LTE Capacity compared to the Shannon Bound," in *Proc. IEEE VTC,* Apr. 2007, pp. 2306-2311.

[24] K.-C. Huang and Z. Wang, *Millimeter Wave Communication Systems,* Wiley, Mar. 2011.

[25] J.-F. Berube, M. Gendreau, and J.-Y. Potvin, "An exact ε-constraint method for bi-objective combinatorial optimization problems: application to the traveling salesman problem with profits," *European J. Operational Research, Elsevier,* vol. 194, no. 1, pp. 39-50, Apr. 2009.

[26] A.-A. Juan, I. Pascual, D. Guimarans, and B. Barrios, "Combining biased randomization with iterated local search for solving the multidepot vehicle routing problem," *Int. Trans. Operational Research,* pp. 1-21, May 2014.

[27] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects,* 3GPP TR 36.814, v. 9.0.0, Rel. 9, Mar. 2010.

[28] T.-H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to Algorithms,* The MIT Press, 3rd ed., 2009.

[29] G. R. MacCartney Jr., J. Zhang, S. Nie, and T. S. Rappaport, "Path Loss Models for 5G Millimeter Wave Propagation Channels in Urban Microcells," in *Proc. IEEE GLOBECOM,* Dec. 2013, pp. 3948-3953.

[30] V.S. Abhayawardhana *et al.,*"Comparison of Empirical Propagation Path Loss Models for Fixed Wireless Access Systems," in *Proc. IEEE VTC,* May 2005, pp. 73-77.

[31] D. Lopez-Perez, X. Chu, and I. Guvenc, "On the expanded region of picocells in heterogeneous networks," *IEEE J. Sel. Topics Signal Process.,* vol. 6, no. 3, pp. 281-294, Jun. 2012.