

# Downlink Cell Association and Load Balancing for Joint Millimeter Wave-Microwave Cellular Networks

Omid Semiari<sup>†</sup>, Walid Saad<sup>†</sup>, and Mehdi Bennis<sup>‡</sup>

<sup>†</sup>Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA,  
Emails: {osemiari,walids}@vt.edu

<sup>‡</sup> Centre for Wireless Communications, University of Oulu, Finland, Email: bennis@ee.oulu.fi

**Abstract**—The integration of millimeter-wave base stations (mmW-BSs) with conventional microwave base stations ( $\mu$ W-BSs) is a promising solution for enhancing the quality-of-service (QoS) of emerging 5G networks. However, the significant differences in the signal propagation characteristics over the mmW and  $\mu$ W frequency bands will require novel cell association schemes cognizant of both mmW and  $\mu$ W systems. In this paper, a novel cell association framework is proposed that considers both the blockage probability and the achievable rate to assign user equipments (UEs) to mmW-BSs or  $\mu$ W-BSs. The problem is formulated as a one-to-many matching problem with *minimum quota constraints* for the BSs that provides an efficient way to balance the load over the mmW and  $\mu$ W frequency bands. To solve the problem, a distributed algorithm is proposed that is guaranteed to yield a Pareto optimal and two-sided stable solution. Simulation results show that the proposed matching with minimum quota (MMQ) algorithm outperforms the conventional max-RSSI and max-SINR cell association schemes. In addition, it is shown that the proposed MMQ algorithm can effectively balance the number of UEs associated with the  $\mu$ W-BSs and mmW-BSs and achieve further gains, in terms of the average sum rate.

## I. INTRODUCTION

The integration of cellular networks with millimeter-wave (mmW) communication links is a promising solution to meet the high data traffic requirements of tomorrow's wireless services [1]–[6]. However, mmW communication is known to be inherently intermittent, due to the susceptibility of its links to signal blockage, due to shadowing by human, buildings, and other obstacles. To this end, mmW base stations (mmW-BSs) must coexist with the conventional microwave base stations ( $\mu$ W-BSs) to provide  $\mu$ W connectivity for users, when a reliable mmW communication is not feasible [3], [4].

Such integrated mmW- $\mu$ W networks introduce new challenges for cellular resource management. In particular, the association of user equipments (UEs) to the BSs must now account for the presence of two radio access technologies (RATs) with significantly different propagation environments. In fact, conventional approaches such as maximum signal-to-interference-plus-noise-ratio (max-SINR) and maximum signal strength indicator (max-RSSI) may result in significantly unbalanced load distributions and may not be directly applicable to the multi-RAT setting. That is due to three key reasons: a) mmW links are highly intermittent and have a higher path loss than  $\mu$ W, b) mmW communication is mostly limited by noise

rather than interference, and c) more bandwidth is available at mmW band compared to the  $\mu$ W frequency band.

The problem of cell association with load balancing has been extensively studied in heterogeneous cellular networks [3], [4], [7]–[10]. The work in [7] studies the performance of the max-SINR cell association for heterogeneous networks (HetNets) with load balancing via cell range expansion (CRE). The authors in [8] propose a cell association approach based on convex optimization to find a load-aware distributed cell association algorithm for HetNets. Moreover, in [9], a game-theoretic approach is adopted for network selection in HetNets, using an evolutionary game approach. For mmW networks, the work in [10] presents a distributed algorithm that yields a fair cell association. A stochastic geometry framework is used in [3] for the decoupled uplink-downlink cell association for traditional macrocells and mmW small cell networks. In addition, the authors in [4] study resource allocation for mmW- $\mu$ W networks where cell association is decoupled in the uplink for mmW users.

The existing works in [7]–[10] have focused on  $\mu$ W or mmW networks, separately and in isolation, and thus, they cannot be applied to integrated mmW- $\mu$ W cellular networks. In addition, the authors in [3] and [4] consider max-RSSI cell association. However, max-RSSI is not a proper association metric for integrated mmW- $\mu$ W networks, since it does not properly reflect the achievable rate of the users. Indeed, this rate depends on the allocated bandwidth and the interference, which are completely different between mmW and  $\mu$ W.

The main contribution of this paper is to introduce a novel cell association framework with load balancing for integrated mmW- $\mu$ W cellular networks. First, we show that conventional max-SINR and max-RSSI cell associations can result in significant unbalanced load in mmW and  $\mu$ W networks. Then, we formulate the proposed cell association problem as a *matching game with minimum quota constraints*, in which the BSs can adjust their minimum quota, in terms of the number of UEs they serve, to balance the network's load. For this game, we show that classical matching solutions such as in [11] and [12] cannot be applied. In contrast, to solve our problem, we propose a novel distributed algorithm that allows UEs to submit association requests to either the mmW-BS or  $\mu$ W-BS that maximizes its average achievable rate. To achieve a balanced load, BSs approve UEs' requests such that the quota constraints are met. We show that the proposed algorithm yields a Pareto optimal (PO) and stable solution for the UEs.

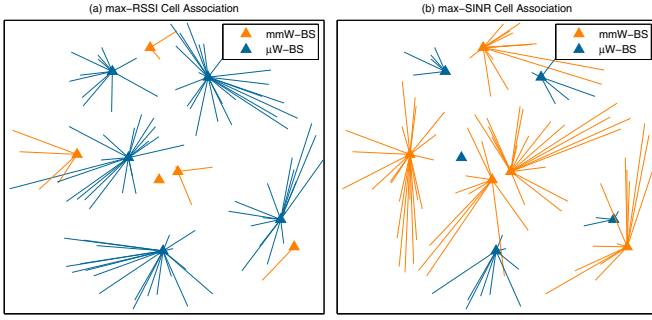


Fig. 1. Cell association for an integrated  $\mu$ W-mmW network using (a) max-RSSI, and (b) max-SINR approaches. The triangles show the BSs and the orange and blue colors represent, respectively, the mmW and  $\mu$ W links.

Simulation results show the effectiveness of our approach in integrated mmW- $\mu$ W networks.

The rest of this paper is organized as follows. Section II presents the problem formulation. Section III formulates the problem as a matching game. Section IV presents the proposed algorithm. Simulation results are analyzed in Section V. Section VI concludes the paper.

## II. SYSTEM MODEL

Consider the downlink of a cellular network, composed of a set  $\mathcal{N}_1$  of  $N_1$  mmW-BSs and a set  $\mathcal{N}_2$  of  $N_2$   $\mu$ W-BSs. In this network, a set  $\mathcal{M}$  of  $M$  UEs are deployed and must be assigned to one mmW-BS or  $\mu$ W-BS. UEs and BSs are distributed uniformly and randomly within a planar area with radius  $r$  centered at  $(0, 0) \in \mathbb{R}^2$ . UEs are equipped with both mmW and  $\mu$ W RF interfaces allowing them to manage their traffic at both frequency bands.

### A. Propagation Model at mmW and $\mu$ W Frequency Bands

Each mmW link between mmW-BS  $n \in \mathcal{N}_1$  and UE  $m \in \mathcal{M}$ , located at  $\mathbf{y}_n \in \mathbb{R}^2$  and  $\mathbf{y}_m \in \mathbb{R}^2$ , respectively, is characterized by the transmit power  $p_n$ , channel gain  $g(\mathbf{y}_m, \mathbf{y}_n)$  and the antenna gain  $\psi(\mathbf{y}_m, \mathbf{y}_n)$ . Assuming that the total power  $p_n$  is distributed uniformly over the mmW bandwidth, the achievable rate per unit of bandwidth for a UE  $m$  assigned to mmW-BS  $n$  is given by:

$$c_{m,n}^{\text{mmW}} = \log_2 \left( 1 + \frac{p_n \psi(\mathbf{y}_m, \mathbf{y}_n) g(\mathbf{y}_m, \mathbf{y}_n)}{w_1 N_0} \right), \quad (1)$$

where  $w_1$  is the mmW bandwidth,  $g(\mathbf{y}_m, \mathbf{y}_n)$  is the link channel gain, and  $N_0$  is the noise power spectral density. Hereinafter, we represent  $c_{m,n}^{\text{mmW}}$  by  $c_{m,n}^{\text{LoS}}$  and  $c_{m,n}^{\text{NLoS}}$ , respectively, if the link is line-of-sight (LoS) and non-line-of-sight (NLoS). Here,  $g(\mathbf{y}_m, \mathbf{y}_n) = L(\mathbf{y}_m, \mathbf{y}_n)^{-1}$ , where the path loss  $L(\mathbf{y}_m, \mathbf{y}_n)$  in dB follows the model of [5]:

$$L(\mathbf{y}_m, \mathbf{y}_n) = b_1 + a_1 10 \log_{10}(\|\mathbf{y}_m - \mathbf{y}_n\|) + \chi, \quad (2)$$

where  $a_1$  represents the slope of the best linear fit to the propagation measurement in mmW frequency band and  $b_1$  is the path loss (in dB) for 1 meter of distance. In addition,  $\chi$  models the deviation in fitting (in dB) which is a Gaussian random variable with zero mean and variance  $\xi_1^2$ .

For each UE-BS pair  $(m, n)$ , let  $\zeta_{m,n}$  be a Bernoulli random variable with success probability  $\rho_{m,n}$  that indicates whether the mmW link is LoS,  $\zeta_{m,n} = 1$ , or NLoS,  $\zeta_{m,n} = 0$ . Different path loss parameters in (2) are considered for the LoS and NLoS links, as listed in Table. I.

At  $\mu$ W band, the achievable rate per unit of bandwidth for a UE  $m \in \mathcal{M}$  associated with  $\mu$ W-BS  $n \in \mathcal{N}_2$  is given by:

$$c_{m,n}^{\mu\text{W}} = \log_2 \left( 1 + \frac{p_n g(\mathbf{y}_m, \mathbf{y}_n)}{\sum_{n' \neq n} p_{n'} g(\mathbf{y}_m, \mathbf{y}_{n'}) + w_2 N_0} \right), \quad (3)$$

where the total power  $p_n$  is distributed uniformly over the  $\mu$ W bandwidth,  $w_2$ , and the channel gain is characterized by parameters,  $a_2, b_2$  and  $\xi_2$ , similar to (2).

### B. Problem Formulation

The cell association problem can be defined as a decision policy  $\pi$  which, for any UE-BS pair  $(m, n)$ , it outputs a binary variable  $x_{m,n} \in \{0, 1\}$ , where  $x_{m,n} = 1$  indicates that UE  $m$  is assigned to BS  $n$ , otherwise,  $x_{m,n} = 0$ . Further, we define the BS  $n$ 's load,  $\kappa_n$  as

$$\kappa_n = \sum_{m=1}^M x_{m,n}. \quad (4)$$

Using (4), the *maximum load difference* can be defined as the difference of the load for the BSs with the maximum and minimum number of associated UEs, as follow:

$$\Delta_\kappa(\pi) = \max(\kappa_n) - \min(\kappa_n). \quad (5)$$

In (5), a smaller  $\Delta_\kappa(\pi)$  implies better load balancing. In general, it is desirable to achieve uniform loads for all BSs, i.e.,  $\Delta_\kappa(\pi) = 0$ . However, by using conventional cell association approaches, such as max-SINR and max-RSSI schemes [7], [8], as shown in Fig. 1, the network will exhibit a severely unbalanced load. In fact, the max-RSSI scheme assigns most of the UEs to  $\mu$ W-BS, due to the smaller path loss over the  $\mu$ W frequency band. On the other hand, the max-SINR scheme assigns most of the UEs to the mmW-BSs, due to the directional transmissions and less interference. In addition, in Figs. 4 and 5, we show by simulations that the CRE techniques used in small cell networks [7] may not effectively improve load balancing in mmW- $\mu$ W networks, due to the large gap in the RSSI and SINR values for mmW and  $\mu$ W links. Our joint mmW- $\mu$ W cell association problem is thus given by:

$$\underset{\mathbf{x}}{\text{maximize}} \quad \sum_{n=1}^N \sum_{m=1}^M x_{m,n} U_{m,n}(\mathbf{x}), \quad (6a)$$

$$\text{s.t.} \quad \sum_{n \in \mathcal{N}} x_{m,n} \leq 1, \quad \forall m \in \mathcal{M}, \quad (6b)$$

$$\kappa_n \leq q_n^{\text{max}}, \quad \forall n \in \mathcal{N}, \quad (6c)$$

$$\kappa_n \geq q_n^{\text{min}}, \quad \forall n \in \mathcal{N}, \quad (6d)$$

$$x_{m,n} \in \{0, 1\}, \quad (6e)$$

where  $\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2$  is the set of all  $N = N_1 + N_2$  BSs and  $U_{m,n}$  denotes the utility of the UE  $m$  associated to the BS  $n$ . Moreover,  $q_n^{\text{max}}$  and  $q_n^{\text{min}}$  denote, respectively, *the maximum and the minimum quotas* for BS  $n$  which represent the maximum and the minimum number of UEs that it can serve.

We let  $0 \leq q_n^{\min} \leq q_n^{\max}$  and  $\sum_{n \in \mathcal{N}} q_n^{\min} \leq M \leq \sum_{n \in \mathcal{N}} q_n^{\max}$  to ensure that a feasible solution exists. As we elaborate later in Section IV, constraints (6c)-(6d) are introduced to balance the network's load. Next, we make the following observation:

**Remark 1:** With  $q_n^{\min} = 0$  and  $q_n^{\max} = M$  for  $\forall n \in \mathcal{N}$ , the optimization problem (6a)-(6e) does not incorporate load balancing.

The cell association for an arbitrary UE depends on the associations of the other UEs, due to the quota constraints (6c)-(6d). In addition, the utility of a UE may depend on whether the associated BS is a mmW-BS or a  $\mu$ W-BS.

### III. CELL ASSOCIATION AS A MATCHING GAME WITH MINIMUM QUOTAS

The downlink association problem in (6a)-(6e) is a 0-1 integer programming problem for assigning UEs to BSs which does not admit a closed-form solution and has exponential complexity [13]. In fact, for such a cell association problem, an exhaustive search requires a comparison of  $O(N^M)$  assignments, which cannot adapt to the dynamics of dense cellular networks, particularly, when using the mmW frequency band.

In this regard, centralized cell association schemes require the BSs to send the network information to the radio network controller (RNC). Such implementations carried out by the RNC are updated at relatively long timescales. This can be detrimental for the mmW UEs that frequently experience NLoS transmissions. To this end, we propose a distributed solution for the mmW- $\mu$ W cell association problem.

#### A. Cell Association as a Matching Game: Preliminaries

To solve the problem in (6a)-(6e), we propose a novel solution based on *matching theory*, a mathematical framework that provides a decentralized solution with tractable complexity for combinatorial problems, such as the one in (6a)-(6e) [11], [12]. A *matching game* is essentially a two-sided assignment problem between two disjoint sets of players in which the players of each set are interested to be matched to the players of the other set, according to *preference relations*. In our model, over each cell association time frame, the set of BSs,  $\mathcal{N}$ , and the set of UEs,  $\mathcal{M}$ , are the two sets of players of the matching game. A preference relation  $\succ$  is defined as a complete, reflexive, and transitive binary relation between the elements of a given set. Here, we let  $\succ_m$  be the preference relation of UE  $m$  and denote  $n \succ_m n'$ , if UE  $m$  prefers BS  $n$  more than  $n'$ . Similarly, we use  $\succ_n$  to denote the preference relation of BS  $n \in \mathcal{N}$ .

To define the preference relations, we can introduce individual utility functions for each UE and BS, using which they can rank one another. In the proposed cell association problem, the preference relations of UEs will depend only on the local average achievable rate information, while the BSs will use network-wide information to distribute the loads and maximize the sum utility. In fact, matching-based cell association provides a suitable framework to balance the load by properly adjusting the maximum and minimum BS quotas.

#### B. Cell Association as a Matching Game

Each cell association policy  $\pi$  determines the allocation of a subset of UEs to each BS. Thus, the problem can be defined as a *one-to-many matching game*:

**Definition 1:** Given two disjoint finite sets of players  $\mathcal{M}$  and  $\mathcal{N}$ , the cell association policy,  $\pi$ , can be defined as a one-to-many *matching relation*,  $\pi: \mathcal{N} \rightarrow \mathcal{M}$  that satisfies 1)  $\forall n \in \mathcal{N}, \pi(n) \subseteq \mathcal{M}$ , 2)  $\forall m \in \mathcal{M}, \pi(m) \in \mathcal{N}$ , and 3)  $\pi(m) = n$ , if and only if  $m \in \pi(n)$ .

In fact,  $\pi(m) = n$  implies that  $x_{m,n} = 1$ , otherwise  $x_{m,n} = 0$ . One can easily see from Definition 1 that the proposed matching game inherently satisfies the constraints in (6b) and (6e). In addition,  $\pi$  is a *feasible matching*, if it satisfies the quota constraints, i.e.,  $|\pi(m)| \in \{0, 1\}$  and  $q_n^{\min} \leq \kappa_n = |\pi(n)| \leq q_n^{\max}$ , where  $|\cdot|$  denotes the set cardinality. Next, we define suitable utility functions.

#### C. Utility and Preference Relations of the UEs and BSs

For mmW links, a UE may experience multiple LoS/NLoS transmissions with different rates during the time that cell association is not updated. Thus, the utilities of UEs to BSs must be a function of the average rate. Here, we define the utility function of UE  $m$  for BS  $n$  as:

$$U_m(n) = \log [f(\mathbf{k}_{m,n})c_{m,n}^{\text{LoS}} + (1 - f(\mathbf{k}_{m,n}))c_{m,n}^{\text{NLoS}}] \quad n \in \mathcal{N}_1 \\ + \log [c_{m,n}^{\mu\text{W}}] \quad n \in \mathcal{N}_2, \quad (7)$$

where,

$$n \in \mathcal{N}_i = \begin{cases} 1 & \text{if } n \in \mathcal{N}_i, \\ 0, & \text{if } n \in \mathcal{N}_{j \neq i}, \end{cases} \quad (8)$$

and  $\mathbf{k}_{m,n}$  is a vector composed of elements,  $k_{m,n}(t')$  where  $t' = t - 1, t - 2, \dots, 0$ , is the number of successful LoS transmissions from mmW-BS  $n$  to UE  $m$  and  $f(k_{m,n}(t))$  is a metric that each UE uses to estimate the LoS probability  $\rho_{m,n}$  for cell association at time  $t$ . In practice, the UEs can update a moving average of the number of LoS transmissions from each mmW-BS by using:

$$f(\mathbf{k}_{m,n}(t)) = \lambda \frac{k_{m,n}(t)x_{m,n}}{k} + (1 - \lambda)f(\mathbf{k}_{m,n}(t - 1)), \quad (9)$$

where  $\lambda$  is a constant smoothing factor between 0 and 1 and  $k$  is the number of transmission slots within the time window in which the association policy  $\pi$  is not updated. Using the utilities in (7), the preference relations of UEs are:

$$n \succ_m n' \Leftrightarrow U_m(n) \geq U_m(n'), \quad (10)$$

for  $\forall m \in \mathcal{M}$ , and  $\forall n, n' \in \mathcal{N}$ .

We note that assigning UEs to their most preferred BS may not admit a feasible matching in general. In other words, in order to satisfy the minimum quotas of the BSs, some UEs may have to be assigned to a lower ranked BS. Therefore, a suitable mechanism is required at the level of the BSs to determine which UEs must be assigned to the BSs with unsatisfied minimum quotas. To this end, all BSs must use the same preference profile, known as *master list (ML)*,  $\succ_{\text{ML}}$  with  $\succ_n \equiv \succ_{\text{ML}}$ ,  $\forall n \in \mathcal{N}$ , as follow:

$$m \succ_{\text{ML}} m' \Leftrightarrow U_{\text{ML}}(m) \geq U_{\text{ML}}(m'), \quad (11)$$



---

**Algorithm 1** Proposed Cell Association Algorithm
 

---

**Inputs:**  $\succ_{\text{ML}}, \succ_m, \forall m \in \mathcal{M}, q_n^{\text{max}}, q_n^{\text{min}}, \forall n \in \mathcal{N}$ .

**Outputs:**  $\pi, \mathbf{x}$ .

- 1: Initialize:  $\pi(m) = \emptyset, \forall m \in \mathcal{M}, \mathcal{M}' = \mathcal{M}$ .
  - 2: Choose the UE  $m^* \in \mathcal{M}'$  that has the highest rank in ML profile, i.e.,  $m^* \succ_{\text{ML}} m, \forall m \in \mathcal{M}'$ .
  - 3: Let  $\pi(m^*) = n$ , where  $n$  is the most preferred BS based on  $\succ_{m^*}$  with  $\kappa_n < q_n^{\text{max}}$ . Moreover, add  $m^*$  to  $\pi(n)$  and remove it from  $\mathcal{M}'$ .
  - 4: **repeat** Steps 3 to 4
  - 5: **until**  $\sum_{n \in \mathcal{N}} [q_n^{\text{min}} - \kappa_n]^+ = |\mathcal{M}'|$ .
  - 6: **while**  $\mathcal{M}' \neq \emptyset$  **do**
  - 7:   Choose the UE  $m^* \in \mathcal{M}'$  that has the highest rank in ML profile, i.e.,  $m^* \succ_{\text{ML}} m, \forall m \in \mathcal{M}'$ .
  - 8:   Let  $\pi(m^*) = n$ , where  $n$  is the most preferred BS based on  $\succ_{m^*}$  with  $\kappa_n < q_n^{\text{min}}$ . Add  $m^*$  to  $\pi(n)$  and remove it from  $\mathcal{M}'$ .
  - 9: **end while**
- 

where,

$$U_{\text{ML}}(m) = \{U_m(n') | U_m(n') \geq U_m(n), \forall n \in \mathcal{N}\}. \quad (12)$$

In fact, (12) implies that BSs give higher priority to a UE that can achieve higher utility by being assigned to its preferred BS. This allows maximizing the sum utility in (6a). To form the ML in practice, BSs only require to exchange the ordering of their nearby UEs to neighboring BSs.

#### IV. PROPOSED CELL ASSOCIATION AND LOAD BALANCING ALGORITHM

To solve the proposed cell association matching problem, we consider two important concepts of *Pareto optimality* and *two-sided stability*. A PO matching is defined as follow [14]:

**Definition 2:** A cell association policy,  $\pi$ , is *Pareto optimal*, if there is no other feasible matching policy  $\pi'$  such that  $\pi'$  is preferred by all UEs over  $\pi$ ,  $\pi' \succeq_m \pi$ , for all  $m \in \mathcal{M}$ , and strictly preferred over  $\pi$ ,  $\pi' \succ_m \pi$ , for some UEs  $m \in \mathcal{M}$ .

In fact, PO is a widely adopted notion of efficiency for distributed mechanisms where each entity, here each UE, aims to maximize its own utility. Furthermore, the concept of two-sided *stable matching* between UEs and BSs is defined as follows [11]:

**Definition 3:** A UE-BS pair  $(m, n) \notin \pi$  is said to be a *blocking pair* of the matching  $\pi$ , if and only if  $m \succ_n m'$  for some  $m' \in \pi(n)$  and  $n \succ_m \pi(m)$ . Matching  $\pi$  is *stable*, if there is no blocking pair.

A stable cell association policy ensures fairness for the UEs. That is, if a UE  $m$  envies the assignment of another UE  $m'$ , then  $m'$  must be preferred by the BS  $\pi(m')$  to  $m$ , i.e., the envy of UE  $m$  is not justified. When  $\succ_n \equiv \succ_{\text{ML}}, \forall n \in \mathcal{N}$ , as in our problem, the two-sided stable  $\pi$  is also known as *ML-fair* matching.

For matching problems with no minimum quota, i.e.,  $q_n^{\text{min}} = 0$ , the well-known *deferred acceptance (DA)* algorithm is used to find a stable matching such as in [11], [12], and [15]. However, with minimum quotas, DA is no longer guaranteed to find a feasible solution.

**Proposition 1:** For cell association problems with minimum quota constraints, the standard DA algorithm may not admit a feasible solution.

*Proof:* We prove this using an example. Let  $\mathcal{M} = \{m_1, m_2, m_3\}$  and  $\mathcal{N} = \{n_1, n_2, n_3\}$ , with ML profile

$m_1 \succ_{\text{ML}} m_2 \succ_{\text{ML}} m_3$ . In addition, assume  $q_n^{\text{min}} = 1$ ,  $q_n^{\text{max}} = 2$  for all BSs, and  $n_1 \succ_{m_i} n_2 \succ_{m_i} n_3$ , for all  $m_i \in \mathcal{M}$ . The DA algorithm for the UE-proposed solution yields  $\pi(n_1) = \{m_1, m_2\}$ ,  $\pi(n_2) = \{m_3\}$ , and  $\pi(n_3) = \emptyset$ , which does not satisfy the minimum quota constraint for  $n_3$ . ■

Therefore, a new algorithm must be developed to solve the problem. To this end, we propose the matching with minimum quota (MMQ) algorithm shown in Algorithm 1, which is designed based on [14]. The proposed algorithm proceeds as follows. After initialization, in step 2, UE  $m^*$  with the highest rank in the ML profile requests a connection with its most preferred BS  $n$ . If  $\kappa_n$  is less than its maximum quota, UE  $m^*$  will be accepted by BS  $n$ . This procedure continues in Steps 3 and 4 for the remaining UEs until the number of UEs is equal to the required number of UEs for meeting the minimum quota constraints, i.e.,  $\sum_{n \in \mathcal{N}} [q_n^{\text{min}} - \kappa_n]^+ = |\mathcal{M}'|$ , where  $[x]^+ = \max(x, 0)$ . Next, in Step 7, the most preferred UE based on the ML profile must be assigned only to its most preferred BS from the subset of  $\mathcal{N}$  with  $\kappa_n < q_n^{\text{min}}$ . In fact, our algorithm allows each UE to be assigned to its most preferred BS, as long as the minimum and maximum quota constraints are not violated. The algorithm terminates once all the UEs are assigned to a BS. The proposed, distributed matching algorithm exhibits the following properties:

**Theorem 1:** Algorithm 1 is guaranteed to yield a feasible PO and stable matching between UEs and BSs.

*Proof:* If the cell association  $\pi$ , given by Algorithm 1 is not PO, a UE  $m$  must exist that can benefit by being assigned to another BS  $n$ , i.e.,  $n \succ_m \pi(m)$ . There are two possible cases to consider. First,  $n \succ_m \pi(m)$  and  $m \notin \pi(n)$  imply that UE  $m$  has applied to BS  $n$  prior to  $\pi(m)$  and is rejected, due to  $\kappa_n = q_n^{\text{max}}$  and  $m' \succ_{\text{ML}} m$ , for all  $m' \in \pi(n)$ . Therefore, adding  $m$  to  $\pi(n)$  does not yield a feasible solution. Second, UE  $m$  is assigned to  $\pi(m)$  to satisfy minimum quota constrain for  $\pi(m)$ . This means re-allocating  $m$  to BS  $n$  will violate the minimum quota criterion for  $\pi(m)$  and is not feasible. Therefore, the given solution is feasible Pareto optimal.

To prove the stability, we note that if UE  $m$  prefers to be assigned to BS  $\pi(m')$ , that implies  $m' \succ_{\text{ML}} m$ , otherwise,  $\pi(m) = \pi(m')$ . Hence, no blocking pair exists and the solution is stable. ■

We must note that Pareto optimality and stability cannot be inherently achieved if the BSs do not follow the ML preference profile. In fact, for  $\succ_n \neq \succ_{\text{ML}}$ , there is no algorithm in general that can guarantee a feasible PO and stable solution [14].

#### V. SIMULATION RESULTS

For simulations, we consider a network with  $N_1 = 10$  mmW-BSs,  $N_2 = 10$   $\mu$ W-BSs, and up to  $M = 100$  UEs located uniformly and randomly over an area with diameter  $r = 1$  km. The main parameters are summarized in Table I. The average probability of LoS for each mmW BS-UE pair is sampled from a uniform distribution,  $\rho_{m,n} \in [0, 1]$ . All statistical results are averaged over a large number of independent runs.

TABLE I  
SIMULATION PARAMETERS

Notation	Parameter	Value
$p_n$	Transmit power	30 dBm
$(\omega_1, \omega_2)$	Bandwidth	(1 GHz, 10 MHz)
$(\xi_{1,LoS}, \xi_{1,NLoS}, \xi_2)$	Standard deviation of path loss	(5.2, 7.6, 10) [3]
$(a_{1,LoS}, a_{1,NLoS}, a_2)$	Path loss exponent	(2,4,3) [3]
$(b_1, b_2)$	Path loss at 1 m	(70, 38) dB
$\psi$	Antenna gain	18 dBi [3]
$N_0$	Noise power spectral density	-174 dBm/Hz
$M$	Number of UEs	From 10 to 100
$q_n^{\max}$	Maximum quota	$M$

We compare the performance of the proposed MMQ algorithm with both conventional max-SINR and max-RSSI approaches. We also consider a CRE with bias factor  $\gamma_{\text{RSSI}}$  and  $\gamma_{\text{SINR}}$ , respectively, for the max-RSSI and max-SINR schemes for further comparisons. To calculate the rates, the total bandwidth at each BS is allocated equally to the associated UEs. That is,  $r_{m,n}^{\text{mmW}} = \frac{w_1}{\kappa_n} c_{m,n}^{\text{mmW}}$ , where  $r_{m,n}^{\text{mmW}}$  denotes the achievable rate for UE  $m$  associated with mmW-BS  $n$ . Moreover,  $r_{m,n}^{\mu\text{W}} = \frac{w_2}{\kappa_n} c_{m,n}^{\mu\text{W}}$ , where  $r_{m,n}^{\mu\text{W}}$  denotes the achievable rate for UE  $m$  assigned to  $\mu\text{W}$ -BS  $n$ . In [8], it is shown that for logarithmic utilities, as in (7), uniform resource allocation maximizes the sum utility.

Fig. 2 shows the average sum-rate for the proposed MMQ approach, compared to max-RSSI and max-SINR approaches versus the number of UEs. The bias factors are chosen such that near uniform loads are achieved for all the BSs. The minimum quotas for  $\mu\text{W}$ -BSs are chosen randomly from 0 to  $\lfloor M/N_2 \rfloor$ , with  $\lfloor \cdot \rfloor$  denoting the floor operand. The results show that the proposed approach achieves up to 14% and 18% improvements compared to, respectively, the max-SINR and the max-RSSI schemes, for  $M = 50$ . This is due to the fact that the achievable rate is a nonlinear function of the SINR or RSSI metrics. Hence, average SINR or RSSI, with respect to  $\zeta_{m,n}$ , cannot be used to find the average achievable rate. However, the proposed approach directly relies on the average achievable rate, as shown in (7).

Fig. 3 shows the optimal minimum quota for  $\mu\text{W}$ -BSs that yields the maximum average sum-rate, as the number of UEs varies, for different values of  $N_1 = N_2$ . The minimum quota for mmW-BSs is zero, since the load of the mmW-BSs are higher than  $\mu\text{W}$ -BSs. The results show that the optimal minimum quota increases, as  $M$  increases, since more UEs must be associated with the  $\mu\text{W}$ -BSs. Moreover, the optimal  $q_n^{\min}$  decreases as  $N_1$  and  $N_2$  increase, since more BSs are available. For  $M = 100$  UEs, we observe that the optimal minimum quotas are  $q_n^{\min} = 8$ , for all  $n \in N_2$ , which implies that 80% of the UEs must be assigned to the  $\mu\text{W}$ -BSs. Hence, if sum rate is considered as the optimality criterion, the result does not yield a balanced network.

In Fig. 4, the maximum load difference  $\Delta_{\kappa}$ , is evaluated for the proposed algorithm compared to max-RSSI approach with CRE under biasing values ranging from 0 to 60 dB. The results show that, as biasing increases, the load balancing decreases and then increases. For all biasing values,  $\Delta_{\kappa}$  for the max-RSSI approach is significantly larger than the proposed MMQ algorithm. In fact, we observe that the proposed MMQ

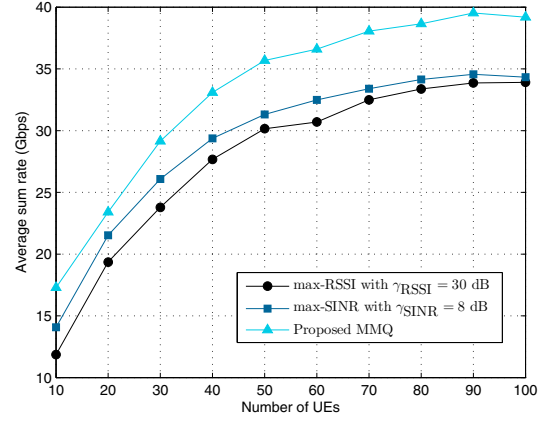


Fig. 2. The average sum-rate (Gbps) versus the number of UEs  $M$ .

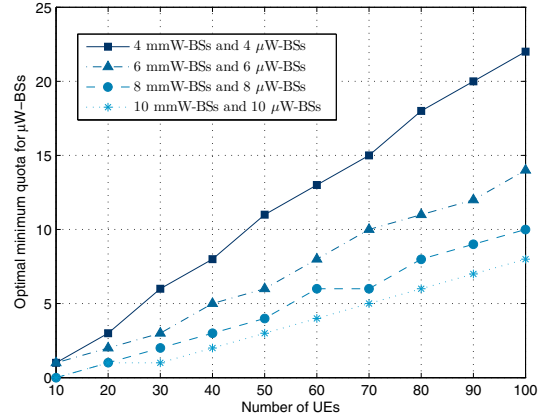


Fig. 3. The optimal quota values for  $\mu\text{W}$ -BSs versus the number of UEs.

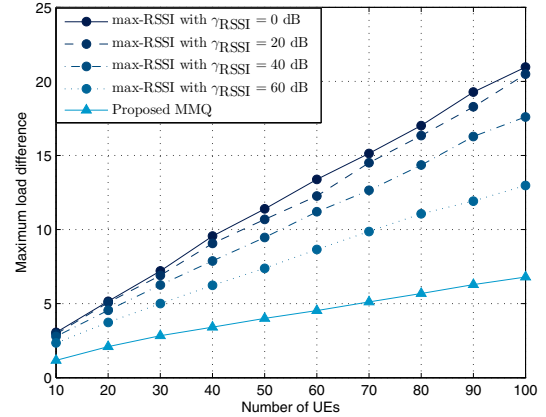


Fig. 4. The maximum load difference,  $\Delta_{\kappa}$ , for the proposed MMQ approach, compared to the max-RSSI with CRE.

algorithm substantially improves the load balancing, reaching up to 48% compared to the max-RSSI with  $\gamma_{\text{RSSI}} = 40$  dB for  $M = 70$ . This improvement is due to the fact that the CRE with biasing cannot precisely control the number of UEs re-associated from  $\mu\text{W}$ -BSs to the mmW-BSs. However, in the proposed approach, the BSs can directly control the number of associated UEs by adjusting their minimum quotas.

Fig. 5 compares the maximum load difference resulting from the proposed MMQ algorithm, compared to the max-

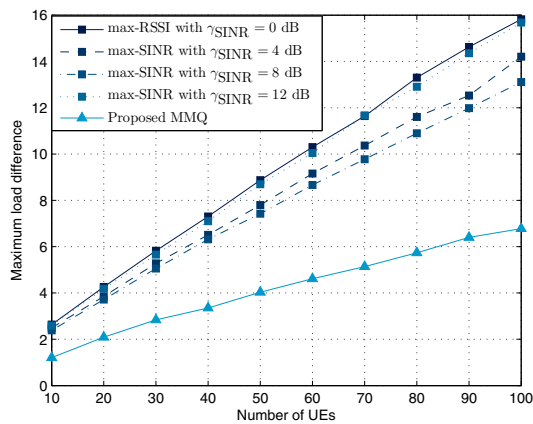


Fig. 5. The maximum load difference,  $\Delta\kappa$ , for the proposed MMQ approach, compared to the max-SINR with CRE.

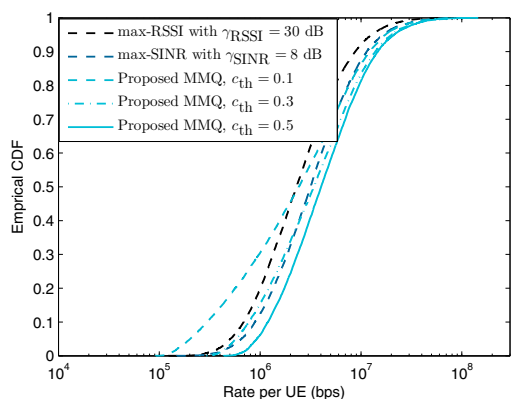


Fig. 6. The empirical CDF of the rate per UE over  $\mu$ W frequency band for  $M = 100$  UEs.

SINR approach with CRE. We observe that, as  $\gamma_{\text{SINR}}$  is increased from 0 to 8 dB, the maximum load difference decreases. However, for  $\gamma_{\text{SINR}} > 8$  dB, the load difference increases, since a larger number of UEs is being assigned to the  $\mu$ W-BSs. Moreover, Fig. 5 shows that for all network sizes, the proposed approach substantially outperforms the max-SINR approach with CRE. In fact, the proposed approach decreases  $\Delta\kappa$  by 47%, compared to max-SINR with  $\gamma_{\text{SINR}} = 8$  dB for  $M = 70$ . Here, we can once again see that the minimum quota constraints allow BSs to control the load more precisely, compared to max-SINR with CRE.

In Fig. 6, the statistics of the average rate per UE are shown over the  $\mu$ W band, compared to max-RSSI and max-SINR. Here, we focus on the average rate for  $\mu$ W links, since the mmW links achieve higher rates, due to the available bandwidth. The results show that, an inherent byproduct of any load balancing technique is the fact that some of the UEs will eventually be associated with an unpreferred  $\mu$ W-BS to satisfy the minimum quota constraints. Such UEs will then trade off rate for load balancing. To this end, parameter  $c_{\text{th}}$  is defined as a utility threshold for UEs. That is, the UE  $m$  is assigned to an unpreferred  $\mu$ W-BS  $n$ , if  $U_m(n) \geq c_{\text{th}}$ .  $c_{\text{th}}$  allows controlling the tradeoff between a highly balanced load and a low average rate for the cell edge UEs. Fig. 6 shows

that for  $c_{\text{th}} = 0.5$ , the proposed MMQ algorithm outperforms the max-RSSI and the max-SINR approaches with CRE.

## VI. CONCLUSIONS

In this paper, we have proposed a novel cell association and load balancing framework for small base stations operating at mmW and  $\mu$ W frequency bands. We have formulated the problem as a one-to-many matching game with minimum quotas. To solve this game, we have proposed a distributed algorithm that considers the average LoS probability in addition to the achievable rate, while assigning UEs to the BSs. We have shown that the proposed algorithm yields a Pareto optimal and stable association policy. Simulation results have shown that the proposed MMQ algorithm outperforms the conventional max-RSSI and max-SINR schemes in terms of both performance and load balancing.

## REFERENCES

- [1] F. Boccardi, R.W. Heath, A. Lozano, T.L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, February 2014.
- [2] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, March 2014.
- [3] H. Elshaer, M. N. Kulkarni, F. Boccardi, J. G. Andrews, and M. Dohler, "Downlink and uplink cell association with traditional macrocells and millimeter wave small cells," *arXiv:1601.05281*, Jan 2016.
- [4] J. Park, S. L. Kim, and J. Zander, "Tractable resource management with uplink decoupled millimeter-wave overlay in ultra-dense cellular networks," *arXiv:1507.08979*, March 2016.
- [5] A. Ghosh, R. Ratasuk, P. Moorut, T. S. Rappaport, and S. Sun, "Millimeter-Wave enhanced local area systems: A high-data-rate approach for future wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1152–1163, June 2014.
- [6] O. Semiari, W. Saad, and M. Bennis, "Context-aware scheduling of joint millimeter wave and microwave resources for dual-mode base stations," in *Proc. of IEEE International Conference on Communications, Mobile and Wireless Networks Symposium*, Kuala Lumpur, Malaysia, May 2016.
- [7] I. Guvenc, "Capacity and fairness analysis of heterogeneous networks with range expansion and interference coordination," *IEEE Communications Letters*, vol. 15, no. 10, pp. 1084–1087, October 2011.
- [8] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, June 2013.
- [9] D. Niyato and E. Hossain, "Dynamics of network selection in heterogeneous wireless networks: An evolutionary game approach," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 2008–2017, May 2009.
- [10] G. Athanasiou, P. C. Weeraddana, C. Fischione, and L. Tassiulas, "Optimizing client association for load balancing and fairness in millimeter-wave wireless networks," *IEEE/ACM Transactions on Networking*, vol. 23, no. 3, pp. 836–850, June 2015.
- [11] A. E. Roth and M. A. O. Sotomayor, *Two-sided matching: A study in game-theoretic modeling and analysis*, Cambridge University Press, 1992.
- [12] E. Jorswieck, "Stable matchings for resource allocation in wireless networks," in *Proc. of 17th International Conference on Digital Signal Processing (DSP)*, Corfu, Greece, July 2011.
- [13] K. Seong, M. Mohseni, and J. M. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *IEEE International Symposium on Information Theory*, Seattle, Washington, July 2006.
- [14] D. E. Fragiadakis, A. Iwasaki, P. Troyan, S. Ueda, and M. Yokoo, "Strategyproof matching with minimum quotas," in *Proc. of 11th International Conference on Autonomous Agents and Multiagent Systems*, Valencia, Spain, June 2012.
- [15] O. Semiari, W. Saad, S. Valentin, M. Bennis, and B. Maham, "Matching theory for priority-based cell association in the downlink of wireless small cell networks," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, May 2014.