Joint Millimeter Wave and Microwave Resources Allocation in Cellular Networks with Dual-Mode Base Stations

Omid Semiari¹, Walid Saad¹, and Mehdi Bennis²

¹Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA, Email: osemiari@vt.edu, walids@vt.edu
²Center for Wireless Communications, University of Oulu, Finland, Email: bennis@ee.oulu.fi

Abstract

The use of dual-mode base stations that can jointly exploit millimeter wave (mmW) and microwave (µW) resources is a promising solution for overcoming the uncertainty of the mmW environment. In this paper, a novel dual-mode scheduling framework is proposed that jointly performs user applications (UA) selection and scheduling over µW and mmW bands. The proposed scheduling framework allows multiple UAs to run simultaneously on each user equipment (UE) and utilizes a set of context information, including the channel state information (CSI) per UE, the delay tolerance and required load per UA, and the uncertainty of mmW channels, to maximize the quality-of-service (QoS) per UA. The dual-mode scheduling problem is then formulated as an optimization with minimum unsatisfied relations (min-UR) problem which is shown to be challenging to solve. Consequently, a long-term scheduling framework, consisting of two stages, is proposed. Within this framework, first, the joint UA selection and scheduling over µW band is formulated as a one-to-many matching game between the µW resources and UAs. To solve this problem, a novel scheduling algorithm is proposed and shown to yield a two-sided stable resource allocation. Second, over the mmW band, the joint context-aware UA selection and scheduling problem is formulated as a 0-1 Knapsack problem and a novel algorithm that builds on the Q-learning algorithm is proposed to find a suitable mmW scheduling policy while adaptively learning the UEs’ line-of-sight probabilities. Furthermore, it is shown that the proposed scheduling framework can find an effective scheduling solution, over both µW and mmW, in polynomial time. Simulation results show that, compared with conventional scheduling schemes, the proposed approach significantly increases the number of satisfied UAs while improving the statistics of QoS violations and enhancing the overall users’ quality-of-experience.
I. INTRODUCTION

Communication at high frequency, millimeter wave (mmW) bands is seen as promising approach to overcome the scarcity of the radio spectrum while providing significant capacity gains for tomorrow’s wireless cellular networks [1]–[3]. However, field measurements [2] have shown that the availability of mmW links can be highly intermittent, due to various factors such as blockage by different obstacles. Therefore, meeting the stringent quality-of-service (QoS) constraints of delay-sensitive applications, such as HDTV and video conferencing, becomes more challenging at mmW frequencies compared to conventional sub-6 GHz frequencies [2], [4]–[9].

Such strict requirements can be achieved by deploying dual-mode small base stations (SBSs) that can support high data rates and QoS by leveraging the available bandwidth at both mmW and microwave (µW) frequency bands [7]. Indeed, in order to provide robust and reliable communications, mmW networks must coexist with small cell LTE networks that operate at the conventional µW band [6]–[10]. However, differences in signal propagation characteristics and in the available bandwidth lead to a significant difference in the achievable rate and the QoS over mmW and µW frequency bands, thus, yielding new challenges for joint mmW-µW user scheduling [10], [11]. In addition, QoS provisioning in dual-mode mmW-µW networks requires overcoming two key challenges: 1) a joint scheduling over both frequency bands is required, since resource allocation over one band will affect the allocation of the resources over the other frequency band and 2) the QoS constraints per user application (UA) will naturally dictate whether the traffic should be served via mmW resources, µW resources, or both. Therefore, robust and efficient scheduling algorithms for dual-mode SBSs are required that exploit context information per UA, including the channel state information (CSI), maximum tolerable delay, and the required load to maximize users’ quality-of-experience (QoE).

To address these scheduling challenges, a number of recent works have emerged in [5]–[10] and [12]–[15]. The work in [7] provides an overview on possible mmW-µW dual-mode architectures that can be used to transmit control and data signals, respectively, at µW and mmW frequency bands. To cope with the intermittent mmW link quality, the authors in [12] formulate the handover decision problem as a Markov decision process (MDP) in mmW networks. In addition, the work in [13] studies the problem of radio access technology (RAT) selection and
traffic aggregation where each user can simultaneously be connected to multiple BSs. In [14], the authors develop an RAT selection scheme for mmW-µW networks via a multi-armed bandit problem that aims to minimize the cost of handoffs for the UEs.

Although interesting, the body of work in [7] and [12]–[14] does not address the scheduling problem in mmW-µW networks. In fact, [12]–[14] focus only on the cell association problem without taking into account, explicitly, the joint allocation of mmW and µW resources. Moreover, existing works such as in [12] are solely focused on the mmW network, while completely neglecting the impact of the communications over the µW frequencies.

In [5], the authors propose an energy-efficient resource allocation scheme for cellular networks, leveraging both µW and unlicensed 60 GHz mmW bands. In [10], the resource allocation problem for mmW-µW networks is studied under a model in which the cell association is decoupled in the uplink for mmW users. However, this work does not consider any QoS constraint in mmW-µW networks. The problem of QoS provisioning for mmW networks is studied in [6], [8], [9], and [15]. In [6], the authors propose a scheduling scheme that integrates device-to-device mmW links with 4G system to bypass the blocked mmW links. The work in [8] presents a mmW system at 60 GHz for supporting uncompressed high-definition (HD) videos for WLANs. In [9], the authors defined and evaluated important metrics to characterize multimedia QoS, and designed a QoS-aware multimedia scheduling scheme to achieve the trade-off between performance and complexity.

Nonetheless, [8] and [9] do not consider multi-user scheduling and multiple access in dual-mode networks. In addition, conventional scheduling mechanisms, such as in [5], [6], [8], and [9], identify each UE by a single traffic stream with a certain QoS requirement. In practice, however, recent trends have shown that users run multiple applications simultaneously, each with a different QoS requirement. Even though the applications at a single device experience the same wireless channel, they may tolerate different delays and QoS, thus, resulting in different user’s quality-of-experience (QoE). Accounting for precise, application-specific QoS metrics is particularly important for scheduling mmW resources whose channel is highly variable, due to large Doppler spreads and short channel coherence time. In fact, conventional scheduling approaches fail to guarantee the QoS for multiple applications at a single UE.

In [15], we studied the problem of the resource management with QoS constraints for dual-
mode mmW-$\mu$W SBSs. However, our previous work assumes that mmW resources are allocated to the users opportunistically and the SBS does not dynamically determine the probability of being at a line-of-sight (LoS) link with the UEs. Such a resource allocation approach may not be efficient, since it may schedule users with low probability of LoS over the mmW band which effectively reduces the available resources for the users with high LoS probability. In addition, we will extend the results provided in [15] by proposing a Q-learning model to dynamically learn the average LoS probability per user and exploit this information to enhance the scheduling over mmW band.

The main contribution of this paper is to propose a novel, context-aware scheduling framework for enabling a dual-band base station to jointly and efficiently allocate both mmW and $\mu$W resources to user applications. This proposed context-aware scheduler allows each user to seamlessly run multiple applications simultaneously, each having its own distinct QoS constraint. To this end, the proposed scheduling problem is formulated as an optimization problem with minimum unsatisfied relations (min-UR) with the goal is to maximize the number of satisfied user applications (UAs). To solve this NP-hard problem, a novel scheduling framework is proposed that considers a set of context information composed of the UAs’ tolerable delay, required load, and the average line-of-sight (LoS) probability, to jointly select and schedule UAs over the $\mu$W and the mmW frequency bands. This proposed approach is composed of two stages. First, over $\mu$W band, the scheduler assigns priority to UAs based on the context information of each UA in order to maximize the number of satisfied UAs. The resource allocation problem at $\mu$W band is modeled as a matching game that aims to assign resource blocks (RBs) to the candidate UAs. To solve this game, a novel algorithm is proposed that iteratively solves the UA selection and the resource allocation problems. We show that the proposed algorithm is guaranteed to yield a two-sided stable matching between UAs and the $\mu$W RBs. In the second stage, over the mmW band, the scheduler assigns priority, based on the context information, to the remaining UAs that were not scheduled over the $\mu$W band. In particular, we propose a Q-learning model that assigns positive and negative rewards, respectively, to LoS and NLoS transmissions to users over past time slots. This learning model allows determining which UAs are more preferred to be scheduled over the mmW band in a way to maximize the average reward of each UA. This information will be used jointly with other context information for UA selection. Consequently,
over the mmW band, we show that the scheduling problem can be cast as a 0-1 Knapsack problem. To solve this problem, we then propose a novel algorithm that allocates the mmW resources to the selected UAs. Consequently, we show that the proposed, two-stage scheduling framework can solve the context-aware dual-band scheduling problem in polynomial time with respect to the number of UAs. Simulation results show that the proposed approach significantly improves the QoS per application, compared to the proportional fair and round robin schedulers.

The rest of this paper is organized as follows. Section II presents the problem formulation. Section III presents the proposed context-aware scheduling solution over the \( \mu \)W band. The proposed context-aware scheduling solution over the mmW band is proposed in Section IV. Simulation results are analyzed in Section V. Section VI concludes the paper.

II. System Model

Consider the downlink of a dual-mode SBS that operates over both \( \mu \)W and mmW frequency bands. The coverage area of the SBS is a planar area with radius \( d \) centered at \((0,0) \in \mathbb{R}^2\). Moreover, a set \( \mathcal{M} \) of \( M \) UEs is deployed randomly and uniformly within the SBS coverage. UEs are equipped with both mmW and \( \mu \)W RF interfaces which allow them to manage their traffic at both frequency bands. The antenna arrays of mmW transceivers can achieve an overall beamforming gain of \( \psi(y_1,y_2) \) for a LoS UE located at \((y_1,y_2) \in \mathbb{R}^2\). Meanwhile, the transceivers at \( \mu \)W frequency have conventional single element, omni-directional antennas. In our model, each UE \( m \in \mathcal{M} \) runs \( \kappa_m \) UAs. We let \( \mathcal{A} \) be the set of all UAs with \( A = \sum_{m \in \mathcal{M}} \kappa_m \) as the total number of UAs across all UEs.

A. Channel Model and Multiple Access at mmW and \( \mu \)W Frequency Bands

The downlink transmission time is divided into time slots of duration \( \tau \). For the \( \mu \)W band, we consider an orthogonal frequency division multiple access (OFDMA) scheme in which multiple UAs can be scheduled over \( K_1 \) resource blocks (RBs) in the set \( \mathcal{K}_1 \) at each time slot with duration \( \tau \). Therefore, the achievable \( \mu \)W rate for an arbitrary UA \( a \) at RB \( k \) and time slot \( t \) is:

\[
R_a(k,t) = w_1 \log_2 \left( 1 + \frac{p_{k,1}|g_{kt}|^2 10^{-\frac{L_1(y_1,y_2)}{10 \log_{10} w_1 N_0}}}{w_1 N_0} \right). \tag{1}
\]
Here, \( w_1 \) is the bandwidth of each RB at \( \mu \)W band, and \( g_{kt} \) is the gain of the Rayleigh fading channel over RB \( k \) at time slot \( t \). The total transmit power at \( \mu \)W band, \( P_1 \), is assumed to be distributed uniformly among all RBs such that \( p_{k,1} = P_1/K_1 \). The path loss \( L_1(y_1, y_2) \) follows the log-distance model with parameters \( \alpha_1, \beta_1, \) and \( \xi_1 \) that represent, respectively, the path loss exponent, the path loss at 1 meter distance, and the variance of the shadowing for the \( \mu \)W band.

Over the mmW band, the SBS uses a time division multiple access (TDMA) scheme to schedule UAs \([6]\). We let \( G_{t,2} \) be the set of UAs that must be scheduled over the mmW band at slot \( t \). During each time slot, for UAs that are assigned to the mmW band, the SBS transmits to UA \( a \in G_{t,2} \) an OFDM symbol of duration \( \tau_{a,t} \) composed of \( K_2 \) resource blocks (RBs). In practice, the mmW transceivers must align their beams during a beam training phase, in order to achieve the maximum beamforming gain \([16]\). This training phase will introduce a non-negligible overhead on the TDMA system, which can become particularly significant as the number of mmW users increases. Hence, a beam training overhead time \( \tau' < \tau \) is considered per transmission to a UA over the mmW band. In practice, duration of \( \tau' \) can reach up to 1.54 milliseconds, depending on the beam resolution \([16]\). Therefore, the effective time for data transmission to UAs in \( G_{t,2} \) will be \( \sum_{a \in G_{t,2}} \tau_{a,t} = \tau - |G_{t,2}| \tau' \), where \( |G_{t,2}| \) denotes the cardinality of the set \( G_{t,2} \).

The large-scale channel effects over the mmW links follow the popular model of \([3]\):

\[
L_2(y_1, y_2) = \beta_2 + \alpha_2 \log_{10}(\sqrt{y_1^2 + y_2^2}) + \chi, \tag{2}
\]

where \( L_2(y_1, y_2) \) is the path loss at mmW frequencies for all UAs associated with a UE located at \( (y_1, y_2) \in \mathbb{R}^2 \). In fact, \((2)\) is known to be the best linear fit to the propagation measurement in mmW frequency band \([3]\), where \( \alpha_2 \) is the slope of the fit and \( \beta_2 \), the intercept parameter, is the pathloss (dB) for 1 meter of distance. In addition, \( \chi \) models the deviation in fitting (dB) which is a Gaussian random variable with zero mean and variance \( \xi_2^2 \). Overall, the total achievable \( \text{mmW rate} \) for UA \( a \) at time slot \( t \) is given by

\[
R_a(t) = \begin{cases} 
\sum_{k=1}^{K_2} w_2 \log_2 \left( 1 + \frac{p_{k,2} \psi(y_1, y_2) |h_{kt}|^2 10^{-\frac{L_2(y_1, y_2)}{10}}}{w_2 N_0} \right), & \zeta_{at} = 1, \\
0, & \zeta_{at} = 0,
\end{cases} \tag{3}
\]

where \( \zeta_{at} = 1 \) indicates that a LoS link is feasible for UA \( a \), otherwise, \( \zeta_{at} = 0 \) and the link is blocked by an obstacle. In fact, \( \zeta_{at} \) is a Bernoulli random variable with probability of success \( \rho_a \),
| Notation | Description | Notation | Description |
|----------|-------------|----------|-------------|
| $M$      | Number of UEs | $\mathcal{M}$ | Set of UEs |
| $\kappa_m$ | Number of UAs per UE $m$ | $A$ | Total number of UAs |
| $\tau$  | Time slot duration | $\tau'$ | Beam training overhead |
| $L$     | Path loss | $\pi_t$ | Scheduling decision at time slot $t$ |
| $K_1$   | Number of $\mu$W RBs | $K_1$ | Set of $\mu$W RBs |
| $K_2$   | Number of mmW RBs | $K_2$ | Set of mmW RBs |
| $G_{t,1}$ | Set of UAs to be scheduled over $\mu$W band | $G_{t,2}$ | Set of UAs to be scheduled over mmW band |
| $w_1$   | Bandwidth of $\mu$W RBs | $w_2$ | Bandwidth of mmW RBs |
| $g_{kt}$ | $\mu$W channel over RB $k$ at time slot $t$ | $\zeta \in \{0, 1\}$ | $\zeta = 1$ if link is LoS, otherwise, $\zeta = 0$. |
| $h_{kt}$ | mmW channel over RB $k$ at time slot $t$ | $p_{k,1}$ | Transmit power over $\mu$W RB $k$ |
| $\rho_a$ | Average LoS probability of the link for UA $a$ | $p_{k,2}$ | Transmit power over mmW RB $k$ |
| $P_1$   | Total transmit power over $\mu$W band | $P_2$ | Total transmit power over mmW band |
| $r$     | Q-learning reward vector | $J$ | Number of QoS classes |
| $\mathcal{A}$ | Set of all UAs across all UEs | $\mathcal{A}_j$ | Set of UAs with $j$-th QoS class |
| $b_a$   | Total required bits for UA $a$ | $b_{a}^{\text{rec}}(t)$ | Received bits by UA $a$ during time slot $t$ |
| $\lambda_{t,1}$ | Number of satisfied UAs over $\mu$W band | $\lambda_{t,2}$ | Number of satisfied UAs over mmW band |

and is identical for all UAs that are run by the same UE. Moreover, $w_2$ is the bandwidth of each RB, $h_{kt}$ is the Rayleigh fading channel coefficient at RB $k$ of slot $t$, and $N_0$ is the noise power spectral density. Furthermore, $p_{k,2}$ denotes the SBS transmit power at RB $k$ of mmW frequency band. The total transmit power at mmW band, $P_2$, is assumed to be distributed uniformly among all RBs, such that $p_{k,2} = P_2/K_2$.

Let $G_{t,1}$ be the set of UAs that must be scheduled over the $\mu$W band at time slot $t$. During each time slot, a UA can be scheduled only at one frequency band, i.e., $G_{t,1} \cap G_{t,2} = \emptyset$. The proposed dual-band multiple access scheme is shown in Fig. [I] where each color identifies a single, distinct UA.

B. Traffic Model with QoS Constraints

We assume a non-full buffer traffic model in which an arbitrary UA $a$ has a total of $b_a$ bits of data to receive. In addition, each UA has an application-specific tolerable delay which specifies its QoS class, as formally defined next.
Definition 1. The QoS class, $A_j$, is defined as the set of all UAs stemming from all UEs that can tolerate a maximum packet transmission delay of $j$ time slots.

Each UA in our system can belong to one out of a total of $J$ QoS classes, $A_j$, $j = 1, \cdots, J$ with $\bigcup_{j=1}^{J} A_j = A$, and $A_j \cap A_{j'} = \emptyset$, $j \neq j'$. Due to system resource constraints, not all UAs can be served instantaneously and, thus, a transmission delay will be experienced by the UAs. In essence, to transmit a data stream of size $b_a$ bits to UA $a \in A_j$, an average data rate of $b_a/j\tau_2$ during $j$ consecutive time slots is needed, otherwise, the UA experiences an outage due to the excessive delay.

The scheduling decision $\pi_t$ at a given slot $t$ is a function that outputs two vectors $x_t$ and $\tau_t$ that determine, respectively, the resource allocation over $\mu$W and mmW bands. In fact, $x_t$ includes the variables $x_{akt} \in \{0, 1\}$ with $a \in A$, $k \in K_1$ where $x_{akt} = 1$ indicates that $\mu$W RB $k$ is allocated to UA $a$ at slot $t$, otherwise, $x_{akt} = 0$. In addition, each element $\tau_{at} \in [0, \tau], a \in A$, of $\tau_t$ determines the allocated time to UA $a$ over mmW band. The required bits for UA $a$ at slot $t$, $b_a^{\text{req}}(t)$, depend on the number of bits received during previous slots, $\sum_{t'=0}^{t-1} b_a^{\text{rec}}(t')$. In other words, $b_a^{\text{req}}(t) = b_a - \sum_{t'=0}^{t-1} b_a^{\text{rec}}(t')$, with $b_a^{\text{req}}(1) = b_a$ and $b_a^{\text{rec}}(0) = 0$. For a given policy $\pi_t$, the required load at time slot $t + 1$, $b_a^{\text{req}}(t + 1)$, can be written recursively as

$$b_a^{\text{req}}(t + 1) = b_a^{\text{req}}(t) - b_a^{\text{rec}}(t) = b_a^{\text{req}}(t) - \left[ \tau \sum_{k=1}^{K_1} R_a(k, t)x_{akt} + R_a(t)\tau_{at}\zeta_{at} \right].$$

(4)
From (4), we observe that policy $\pi_t$ depends on the scheduling decisions during previous time slots $\{\pi_1, \pi_2, \ldots, \pi_{t-1}\}$. Thus, we define $\pi = \{\pi_1, \pi_2, \ldots, \pi_t, \ldots, \pi_J\} \in \Pi$ as a long-term scheduling policy, where $\Pi$ is the set of all possible scheduling policies.

Next, we use (4) to formally define the QoS criterion for any UA $a \in A_j$ as

$$\mathbb{1}(a \in A_j; \pi) = \begin{cases} 1 & \text{if } \sum_{t=1}^{j} b_{a}^{\text{req}}(t') \geq b_a, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbb{1}(a \in A_j; \pi) = 1$ indicates that under policy $\pi$, enough resources are allocated to UA $a \in A_j$ to receive $b_a$ bits within $j$ slots, while $\mathbb{1}(a \in A_j; \pi) = 0$ indicates that UA $a$ is going to experience an outage. We define the outage set $O^\pi = \{a|\mathbb{1}(a \in A_j; \pi) = 0, j = 1, \ldots, J\}$ as the set of UAs in outage.

Prior to formulating the problem, we must note the following inherent characteristics of dual-mode scheduling: 1) If mmW link with a high LoS probability is not feasible for a UE, scheduling over the mmW band can cause outage to the associated UAs, specifically for delay-intolerant UAs, 2) larger range of supported rates is available for UAs compared to the conventional single-band systems. Hence, for some UAs, the required rate exceeds the achievable rate at $\mu$W band. Therefore, effective dual-mode scheduling should not only rely solely on CSI, but it must also leverage UA-specific metrics, herein referred to as context information as formally defined next.

**Definition 2.** At any slot $t$, the tuple $C = (A_{j\geq t}, b^{\text{req}}(t), \rho)$ defined as context information, is composed of the delay constraints of UAs, $A_{j\geq t} = \bigcup_{j=t}^{J} A_j$, the required load per UA, $b^{\text{req}}(t) = \{b_{a}^{\text{req}}(t)|a \in A_{j\geq t}\}$, and the LoS probability of each UA, $\rho = \{\rho_a|a \in A_{j\geq t}\}$.

Note that exploiting the context information at any time slot $t$ properly links the scheduling policy $\pi_t$ to the history, since from (5), $C$ at slot $t$ depends on $\pi_{t'}, t' = 1, \ldots, t - 1$.

**C. Problem Formulation**

Our goal is to find a scheduling policy $\pi^* \in \Pi$ that satisfies (5) for as many UAs as possible over $J$ time slots. The general long-term scheduling problem for slots $t = 1, \ldots, J$ can be solved separately at each slot $t$ to find $\pi^*_t(C, \text{CSI}) = (x^*_t, \tau^*_t)$, while the time-dependency of scheduling decisions is captured by exploiting the context information. Therefore, the scheduling problem
at an arbitrary slot $t$ can be formulated as follows:

$$\argmax_{x_t, \tau_t} \lambda_{t,1} + \mathbb{E}[\lambda_{t,2}],$$  \hspace{1cm} (6a)

subject to

$$\tau \sum_{k=1}^{K_1} R_a(k, t) x_{akt} \geq b_{a}^{\text{req}}(t), \quad \forall a \in \mathcal{A}_t \cap \mathcal{G}_{t,1},$$  \hspace{1cm} (6b)

$$R_a(t) \tau_{at} \zeta_{at} \geq b_{a}^{\text{req}}(t-1), \quad \forall a \in \mathcal{A}_t \cap \mathcal{G}_{t,2},$$  \hspace{1cm} (6c)

$$x_t \in \mathcal{X} = \left\{ x_{akt} \in \{0,1\} \mid \sum_{a \in \mathcal{A}_t} x_{akt} \leq 1, \sum_{k=1}^{K_1} x_{akt} \leq K_1, \forall a \in \mathcal{A}_{j \geq t} \right\},$$  \hspace{1cm} (6d)

$$\tau_t \in \mathcal{Y} = \left\{ \tau_{at} \in [0, \tau] \mid \sum_{a \in \mathcal{G}_{t,2}} \tau_{at} + |\mathcal{G}_{t,2}| \tau' \leq \tau, \forall a \in \mathcal{G}_{t,2} \right\},$$  \hspace{1cm} (6e)

$$x_t, \tau_t \in \mathcal{Z} = \left\{ x_t \in \mathcal{X}, \tau_t \in \mathcal{Y} \mid \sum_{k=1}^{K_1} x_{akt} \tau_{at} = 0 \right\},$$  \hspace{1cm} (6f)

where $\lambda_{t,1}$ and $\lambda_{t,2}$ denote, respectively, the number of satisfied UAs scheduled at $\mu$W and mmW bands at slot $t$. Given a decision policy $\pi_t$, $\lambda_{t,2}$ is a random variable that depends on $\zeta_{at}$ at the mmW band. In fact, the expectation in (6a) is taken over $\zeta_{at}$, for all $a \in \mathcal{G}_{t,2}$. However, $\lambda_{t,1}$ is deterministic, if the slot duration $\tau$ is smaller than the $\mu$W channel coherence time.

The problem (6a)-(6f) falls into a class of optimization problems, referred to as minimum unsatisfied relations (Min-UR), which are known to be NP-hard [17]. Although linear systems with equality or nonequality constraints can be solved in polynomial time, using an adequate linear programming method, least mean squared methods are not appropriate for infeasible
systems when the objective is to maximize satisfied relations [17].

With this in mind, we propose a two-stage solution that solves (6a)-(6e) in polynomial time, as illustrated in Fig. 2. The scheduling at μW band is considered first in order to reliably schedule as many UAs as possible with small required loads over the μW band. To this end, a hierarchy scheme is proposed based on the UAs’ QoS class, CSI, and the required loads. Moreover, the UAs selection and scheduling are jointly done at SBS using an iterative algorithm.

Then, for the remaining UAs that were not scheduled at the μW band, we propose a joint UA selection criterion and scheduling algorithm that introduces a hierarchy to the UAs, based on the context information, and maximizes the number of satisfied UAs.

III. CONTEXT-AWARE UA SELECTION AND RESOURCE ALLOCATION AT μW BAND

Before scheduling at mmW band, the goal of the scheduler is to first find an allocation $x^*_t$ at each slot $t$ over μW band that satisfies

$$\arg\max_{x^*_t} \lambda_{t,1}(x^*_t),$$

s.t. (6b), (6d).

The downlink scheduling problem in (7a)-(7b) is an inconsistent combinatorial problem of matching users to resources which does not admit a closed-form solution and has an exponential complexity [18]. Hence, the solution of (7a)-(7b) depends on which UAs are chosen to be scheduled at μW band, i.e., the set $G_{t,1}$. To this end, we introduce a hierarchy for UA selection by grouping the different UAs into the following sets:

$$G_{t,1}^{(1)} = \{ a \in A_j | j = t, b_{ra, eq}(t) > 0 \},$$

$$G_{t,1}^{(2)} = \{ a \in A_j | j > t, b_{ra, eq}(t) > 0 \}.$$

In fact, the UAs in $G_{t,1}^{(1)}$ have higher priority than $G_{t,1}^{(2)}$, since they must be served during the current time slot, otherwise, they will experience an outage. In addition, for UAs of the same set, the UA that satisfies the following has the highest priority:

$$a^* = \arg\min_a \frac{b_{ra, eq}(t)}{\sum_{k \in K_1} R_{a}(k, t)},$$

where (10) selects UA $a^*$ that minimizes the ratio of the required load to the achievable rate. To ensure that the constraints set for the selected UAs $a \in G_{t,1}$ is feasible, i.e. $\lambda_{t,1}(x) = |G_{t,1}|$, the
UA selection has to be done jointly while solving (7a)-(7b). Following, we propose a framework that solves (7a)-(7b) for a given $G_{t,1}$.

A. Scheduling as a Matching Game: Preliminaries

For a selected set of UAs at $\mu$W band, $G_{t,1}$, we propose a novel resource allocation scheme at $\mu$W band based on matching theory, a mathematical framework that provides polynomial time solutions for complex combinatorial problems such as the one in (7a)-(7b) [19], [20]. A matching game is defined as 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. At each time slot $t$ of our scheduling problem, $K_1$ and $G_{t,1}$ are the two sets of players. A preference relation $\succ$ is defined as a complete, reflexive, and transitive binary relation between the elements of a given set. Here, we let $\succ_a$ be the preference relation of UA $a$ and denote $k \succ_a k'$, if player $a$ prefers RB $k$ over RB $k'$. Similarly, we use $\succ_k$ to denote the preference relation of RB $k \in K_1$.

In the proposed scheduling problem, the preference relations of UAs depend on both the rate and the QoS constraint which will be captured via well-designed, individual utility functions for UAs and SBS resources, as defined later in this section.

B. Scheduling at $\mu$W band as a Matching Game

Each scheduling decision $\pi_{t,1}$ determines the allocation of RBs to UAs during time slot $t$ over the $\mu$W band. Thus, the scheduling problem at $\mu$W frequency band can be defined as a one-to-many matching game:

**Definition 3.** Given two disjoint finite sets of players $G_{t,1}$ and $K_1$, the scheduling decision at time slot $t$, $\pi_{t,1}$, can be defined as a matching relation, $\pi_{t,1} : G_{t,1} \rightarrow K_1$ that satisfies 1) $\forall a \in G_{t,1}, \pi_{t,1}(a) \subseteq K_1$, 2) $\forall k \in K_1, \pi_{t,1}(k) \in G_{t,1}$, and 3) $\pi_{t,1}(k) = a$, if and only if $k \in \pi_{t,1}(a)$.

In fact, $\pi_{t,1}(k) = a$ implies that $x_{akt} = 1$, otherwise $x_{akt} = 0$. Therefore, $\pi_{t,1}$ is indeed the scheduling decision that determines the allocation at $\mu$W band. One can easily see from the above definition that the proposed matching game inherently satisfies the constraint (6d). Next, we need to define suitable utility functions to determine the preference profiles of UAs and RBs. Given matching $\pi_{t,1}$, we define the utility of UA $a$ for $k \in K_1$ at time slot $t$ as:
\[ \Psi_a(k,t;\pi_{t,1}) = \begin{cases} 0 & \text{if } \sum_{k' \in \pi_{t,1}(a)} R_a(k',t) \geq b^\text{req}_a(t), \\ R_a(k,t) & \text{otherwise.} \end{cases} \tag{11} \]

The utility of \( \mu W \) RBs \( k \in K_1 \) for UA \( a \in G_{t,1} \) is simply the rate
\[ \Phi_k(a,t) = R_a(k,t). \tag{12} \]

Using these utilities, the preference relations of UAs and RBs at a given time slot \( t \) will be
\[ k \succ_a k' \iff \Psi_a(k,t;\pi_{t,1}) \geq \Psi_a(k',t;\pi_{t,1}) \tag{13} \]
\[ a \succ_k a' \iff \Phi_k(a,t) \geq \Phi_k(a',t), \tag{14} \]
for \( \forall a,a' \in G_{t,1} \), and \( \forall k,k' \in K_1 \). Given this framework, we propose a joint UA selection and matching-based scheduling algorithm that maximizes \( \lambda_{t,1} \).

C. Proposed Context-aware Scheduling Algorithm at \( \mu W \) Band

To solve the proposed game and find a suitable outcome, we use the concept of two-sided stable matching between UAs and RBs, defined as follows [19]:

**Definition 4.** A pair \( (a,k) \notin \pi_{t,1} \) is said to be a blocking pair of the matching \( \pi_{t,1} \), if and only if \( a \succ_k \pi_{t,1}(k) \) and \( k \succ_a \pi_{t,1}(a) \). Matching \( \pi_{t,1} \) is stable, if there is no blocking pair.

A stable scheduling decision, \( \pi_{t,1} \), ensures fairness for the UAs. That is, if a UA \( a \) envies the allocation of another UA \( a' \), then \( a' \) must be preferred by the RB \( \pi_{t,1}(a') \) to \( a \), i.e., the envy of UA \( a \) is not justified. For conventional matching problems, the popular deferred acceptance (DA) algorithm is normally used to find a stable matching [19]–[21]. However, DA cannot be applied directly to our problem because it assumes that the quota for each UA is fixed. The quota is defined as the maximum number of RBs that a UA can be matched to. In our problem, however, quotas cannot be predetermined, since the number of RBs needed to satisfy the QoS constraint of a UA in (7b) depends on the channel quality at each RB, as well as the context information. In fact, the adopted utility functions in (11) depend on the current state of the matching. Due to the dependency of UAs’ preferences to the state of the matching, i.e. \( x_{akt} \) variables, the proposed game can be classified as a matching game with externalities [21]. For matching games with externalities, DA may not converge to a two-sided stable matching. Therefore, a new matching algorithm must be found to solve the problem.
Algorithm 1 Context-Aware UA Selection and Resource Allocation Algorithm at $\mu$W Band

Inputs: $G^{(1)}_{t,1}$, $G^{(2)}_{t,1}$, $b^{eq}(t)$, $R_a(k,t)$.

Outputs: $a$; $G_{t,1}$.

Initialize: $G_{t,1} = \emptyset$,
1: $G'_{t,1} = G^{(1)}_{t,1}$, $K_a = K_1$, $\forall a \in G'_{t,1}$.
2: Add UA $a^* \in G'_{t,1}$ with smallest $b^{eq}(t)/\sum_{k \in K_1} R_a(k,t)$ to $G_{t,1}$ and remove it from $G'_{t,1}$.
3: Update the preference ordering of UAs $a \in G_{t,1}$ and RBs $k \in K_1$, using (11) and (12).
4: Using $\succ_a$, a UA $a \in G_{t,1}$ is tentatively assigned to its most preferred RB in $K_a$.
5: From the tentative list of UA applicants plus $\pi_{t,1}(k)$ for RB $k$, only the most preferred UA, based on $\succ_k$, is assigned to $k$. Next, $k$ is removed from the applicants’ $K_a$ sets.
6: Each UA $a$ updates $b^{eq}(t)$ and $\succ_a$ based on (13).
7: repeat Steps 3 to 6
8: until $\mathbf{1}_a(x) = 1$, or $K_a = \emptyset$, $\forall a \in G_{t,1}$.
9: if $\exists a \in G_{t,1}$, $\mathbf{1}_a(x) \neq 1$ then
10: Remove $a^*$ from $G_{t,1}$ and go to Step 3.
11: end if
12: if $\exists k$, $\sum_{a \in G^{(1)}_{t,1} \cup G^{(2)}_{t,1}} x_{akt} = 0$ then let $G'_{t,1} = G^{(2)}_{t,1}$ and go to Step 2.
13: end if

To this end, we propose a novel context-aware scheduling algorithm shown in Algorithm 1. Algorithm 1 first allocates the RBs to the UAs in $G^{(1)}_{t,1}$. At every iteration, each UA $a^*$ given by (10) is added to the set $G_{t,1}$ of the matching game. In Steps 4 to 10, the algorithm assigns RBs $k \in K_1$ to UAs $a \in G_{t,1}$ as follows. Each UA $a \in G_{t,1}$ is tentatively assigned to its most preferred RB $k \in K_a$. Next, from the tentative list of candidate UAs as well as current assignment $\pi_{t,1}(k)$, the scheduler allocates RB $k$ only to the most preferred UA, based on $\succ_k$. The RB $k$ is removed from the set $K_a$ corresponding to each candidate UA $a \in G_{t,1}$. Based on the allocated RBs, the UAs update $b^{eq}(t)$ and $\succ_a$. This process is repeated until the rate constraints for UAs are satisfied, $\mathbf{1}_a(x) = 1$, or $K_a = \emptyset$ for UAs $a \in G_{t,1}$. Then, if some of the RBs are left unallocated, the algorithm follows Steps 2 to 14 to add UAs from $G^{(2)}_{t,1}$ to $G_{t,1}$.

**Theorem 1.** The proposed Algorithm 1 is guaranteed to yield a two-sided stable matching between UAs and $\mu$W RBs.

**Proof:** The convergence of the Algorithm 1 at each slot is guaranteed, since a UA never applies for a certain RB twice. Hence, at the worst case scenario, all UAs will apply for all
RBs once, which yields $K_a = \emptyset, \forall a \in A$. Next, we show that, once the algorithm converges, the resulting matching between UAs and RBs is two-sided stable. Assume that there exists a pair $(a,k) \notin \pi_{t,1}$ that blocks $\pi_{t,1}$. Since the algorithm has converged, we can conclude that at least one of the following cases is true about $a$: $1_a(x) = 1$, or $K_a = \emptyset$.

The first case, $1_a(x) = 1$ implies that $a$ does not need to add more RBs to $\pi_{t,1}(a)$. In addition, $a$ would not replace any of $k' \in \pi_{t}(a)$ with $k$, since $k' \succ_a k$. Otherwise, $a$ would apply earlier for $k$. If $a$ has applied for $k$ and got rejected, this means $\pi_{t,1}(k) \succ_a k$, which contradicts $(a,k)$ to be a blocking pair. Analogous to the first case, $K_a = \emptyset$ implies that $a$ has got rejected by $k$, which means $\pi_{t,1}(k) \succ_a k$ and $(a,k)$ cannot be a blocking pair. This proves the theorem. □

Given $G_{t,1}$ by Algorithm 1 at $\mu$W band, the scheduling problem at slot $t$ is now reduced to choosing a subset of unscheduled UAs and allocate mmW resources to them such that the number of satisfied UAs is maximized.

IV. CONTEXT-AWARE UA SELECTION AND RESOURCE ALLOCATION AT MMW BAND

We let $G_{t,2}' = \{a \in A_j \mid a \notin G_{t,1}, b_a^{\text{req}}(t) > 0\}$ be the set of UAs that have not been scheduled over the $\mu$W band. Here, the scheduling problem over the mmW band at slot $t$ can be formulated as a stochastic min-UR problem as follows:

$$\arg\max_{\tau_t} \mathbb{E} [\lambda_{t,2}(\tau_t)],$$

s.t. (6c), (6e), (6f). (15a)\hspace{1cm} (15b)

Here, we note that $\zeta_{at}$ in (6c) is a Bernoulli random variable with success probability $\rho_a$. Hence, for any allocation $\tau_t$, the number of satisfied constraints $\lambda_{t,2}$ is a random variable. Although the exact distribution of $\lambda_{t,2}$ may not be found for a general infeasible problem as (15a)-(15b), we can approximate the distribution of outage ratio at slot $t$, $P_{out,t} = 1 - [(\lambda_{t,1} + \lambda_{t,2})/A_{j=t}]$, as follows:

**Proposition 1.** Let $\tau_t$ be a feasible solution for the subset of constraints in (15b) associated with UAs $a \in G_{t,2}' \subseteq G_{t,2}'$. Given $\lambda_{t,1}$ and $0 \leq P_{th} < 1 - \lambda_{t,1}/A_{j=t}$, where $P_{th}$ is an outage threshold, the CDF of the outage ratio at slot $t$, $F_{P_{out,t}}(P_{th})$ can be approximated by,

$$F_{P_{out,t}}(P_{th}) \approx 1 - \frac{\Gamma([(1 - P_{th})A_t - \lambda_{t,1} + 1], \lambda_{ave})}{[(1 - P_{th})A_t - \lambda_{t,1}]!},$$

(16)
where \([\cdot]\) is the floor function, \(\Gamma(\cdot, \cdot)\) is the incomplete gamma function, and
\[
\lambda_{\text{ave}} = \mathbb{E}[\lambda_{t,2}(\tau_t)] = \sum_{a \in \mathcal{G}_{t,2}} \rho_a. \tag{17}
\]

**Proof:** We can write \(\lambda_{t,2}\) as the sum of Bernoulli random variables \(\zeta_{at}\), i.e.,
\[
\lambda_{t,2}(\tau_t) = \sum_{a \in \mathcal{G}_{t,2}} \zeta_{at}. \tag{17}
\]
Hence, using Le Cam’s theorem, the distribution of \(\lambda_{t,2}\) follows Poisson distribution, i.e.,
\[
\mathbb{P}[\lambda_{t,2}(\tau_t) = k] \approx \frac{\lambda_{\text{ave}}^k e^{-\lambda_{\text{ave}}}}{k!}, \tag{18}
\]
where \(\mathbb{E}[\lambda_{t,2}(\tau_t)]\) is the sum of expected values of \(\zeta_{at}\) for selected UAs in \(\mathcal{G}_{t,2}\) as given by (17). Moreover, the approximation error is bounded by
\[
\sum_{k=0}^{\infty} \left| \mathbb{P}[\lambda_t = k] - \frac{\lambda_{\text{ave}}^k e^{-\lambda_{\text{ave}}}}{k!} \right| < 2 \sum_{a \in \mathcal{G}_{t,2}} \rho_a^2, \tag{19}
\]
where \(\lambda_t = \lambda_{t,1} + \lambda_{t,2}\). Next,
\[
F_{P_{\text{out},t}}(P_{th}) = \mathbb{P}(P_{\text{out},t} \leq P_{th}) = \mathbb{P}\left(1 - \frac{\lambda_{t,1} + \lambda_{t,2}}{A_t} \leq P_{th}\right) \tag{20}
\]
\[= 1 - \mathbb{P}\left(\lambda_{t,2} \leq \lfloor (1 - P_{th})A_t - \lambda_{t,1} \rfloor\right) \tag{21}
\]
\[\approx 1 - \frac{\Gamma\left(\lfloor (1 - P_{th})A_t - \lambda_{t,1} \rfloor, \lambda_{\text{ave}}\right)}{\lfloor (1 - P_{th})A_t - \lambda_{t,1} \rfloor!}, \tag{22}
\]
where (22) follows the CDF of the Poisson distribution.

From (17) and (19), we can see that the objective function increases as UAs with higher \(\rho_a\) are satisfied, however, the approximation of the distribution becomes less accurate.

We note that if LoS probabilities \(\rho_a\) are known by the SBS, the proposed scheduling problem over the mmW band becomes equivalent to a 0-1 stochastic Knapsack optimization problem [22]. However, in practice, the explicit values of \(\rho_a\) may not be available at the SBS. In Section IV-B, we will introduce a learning approach using which the SBS can determine if \(\rho_a \geq \rho_{th}\), where \(\rho_{th}\) is a constant value. By learning which UAs satisfy \(\rho_a \geq \rho_{th}\), the SBS assigns priority to the UAs that are more likely to be at a LoS link from the SBS. This information along with the QoS classes of UAs will allow the scheduler to group UAs into the following non-overlapping subsets:

\[
\mathcal{G}^{(1)}_{t,2} = \{a \in \mathcal{A}_{j=t} \cap \mathcal{G}'_{t,2} | \rho_a \geq \rho_{th}\}, \tag{23}
\]
\[
\mathcal{G}^{(2)}_{t,2} = \{a \in \mathcal{A}_{j=t} \cap \mathcal{G}'_{t,2} | \rho_a < \rho_{th}\}, \tag{24}
\]
Algorithm 2 Context-Aware UA Selection and Resource Allocation Algorithm at mmW Band

Inputs: $G^{(i)}_{t,2}, i = 1, \ldots, 4, b^{req}(t), R_a(t)$.
Output: $\tau; G_{t,2}$.

1: \textbf{Initialize:} $G_{t,2} = \emptyset$.
2: \textbf{for} $i = 1; i \leq 4; i++ \textbf{do}$
3: \hspace{1em} \textbf{for} $j = 1 : |G^{(i)}_{t,2}| \textbf{do}$
4: \hspace{2em} Find UA $a^* \in G^{(i)}_{t,2}$ from (27), set $\tau_{a^*,t} = b^{req}(t)/R_{a^*}(t)$ and add $a^*$ to $G_{t,2}$.
5: \hspace{1em} \textbf{if} (6e) is not satisfied \textbf{then}
6: \hspace{2em} Remove $a^*$ from $G_{t,2}$. Break.
7: \hspace{1em} \textbf{end if}
8: \hspace{1em} \textbf{end for}
9: \textbf{end for}

$G^{(3)}_{t,2} = \{a \in A_{j>t} \cap G'_{t,2} | \rho_a \geq \rho_{th}\}$, \hspace{1em} (25)

$G^{(4)}_{t,2} = \{a \in A_{j>t} \cap G'_{t,2} | \rho_a < \rho_{th}\}$, \hspace{1em} (26)

In fact, the SBS will adopt a greedy approach that assigns priority to sets $G^{(i)}_{t,2}$ with $i = 1$ as highest and $i = 4$ as lowest priority. That is due to the fact that UAs in $G^{(1)}_{t,2}$ cannot tolerate further delays. Moreover, they belong to UEs with high possibility of LoS access to SBS. In addition, UAs in $G^{(2)}_{t,2}$ are in second priority, since they cannot tolerate more delay, while having a low $\rho_a$. Moreover, UAs $G^{(3)}_{t,2}$ are assigned to a third priority, since they can tolerate more delays and have high probability to be at LoS mmW link with SBS. The least priority is assigned to UAs in $G^{(4)}_{t,2}$ as they can tolerate further delays, while having low $\rho_a$.

Furthermore, for the UAs of the same set, the highest priority is given to a UA $a^*$ that satisfies:

$$a^* = \arg\min_a \frac{b^{req}(t)}{R_a(t)}.$$ \hspace{1em} (27)

In other words, the SBS selects the UA that requires the least time resource to be satisfied. Similar to $\mu$W band scheduling, the SBS must ensure that the constraints set for selected UAs $a \in G_{t,2}$ is feasible. Therefore, the UA selection has to be done jointly while solving (15a)-(15b). Next, we propose a joint UA selection and scheduling algorithm at mmW band.
A. Proposed Context-aware Scheduling Algorithm over the mmW Band

Over the mmW band, the objective is to serve as many UAs as possible in order to offload more traffic from the µW band, subject to UAs delay constraints. With this in mind, we propose Algorithm 2 to solve (15a)-(15b). The algorithm follows the priority criterion introduced in (24)-(26). Starting with the set $G_t^{(1)}$, the scheduling process is a 0-1 Knapsack problem composed of $|G_t^{(1)}|$ items all with the same benefit $\rho_{th}$ and weights equal to the required time $\tau_{a,t} = \frac{b_{req}(t)}{R_a(t)}$. This problem can be simply solved by sorting the required time in increasing order and adding UAs one by one to the set $G_t^{(1)}$. The algorithm follows the process for the remaining sets and converges, once the entire mmW time slot duration is allocated and no additional time is available for more UAs.

The above solution requires the SBS to determine for which UAs the condition $\rho_a \geq \rho_{th}$ is satisfied. Next, we introduce a learning scheme that enables the UEs to obtain this information by monitoring successful LoS transmissions from the SBS over time and send it to the SBS. Clearly, $\rho_a$ is the same for the UAs that run at an arbitrary UE, since they experience the same wireless channel.

B. Q-learning Model to Evaluate the Average LoS Probability

In a real-world cellular network, the UEs will be surrounded by many objects and, thus, the SBS may never know in advance whether an LoS mmW link will be available or not. Therefore, scheduling UAs of UEs that are experiencing a high blockage probability not only wastes network resources, it may drastically degrade QoS for delay intolerant UAs.

In practice, $\rho_a$ depends on many parameters such as the distance between the UE and the SBS, or blockage by human or other surrounding objects. Although finding a closed-form relation of $\rho_a$ with these parameters may not be feasible in general, the UEs can learn whether they have a high LoS probability based on transmissions from the SBS over time. The UEs will then update and send this information to the SBS at each time slot. Clearly, a simple averaging over time would not work, since the environment is dynamic and $\rho_a$ may change over time. To this end, we propose a learning framework, based on Q-learning (QL) [23], in order to determine UAs with $\rho_a \geq \rho_{th}$ without knowing the actual $\rho_a$ values. QL is a reinforcement learning algorithm that determines optimal policy without detailed modeling of the system environment [23], [24].
The proposed QL model is formally defined by the following key elements:

- **Agents**: UEs $m \in M$.

- **States**: Depending on whether a UA of a given UE is being scheduled over $\mu W$ or mmW bands, there are three possible states for the UA: 1) UA is served by the SBS over a LoS mmW link ($S_1$), 2) UA is scheduled over mmW band, but no LoS link is possible ($S_2$), and 3) UA is scheduled over $\mu W$ band ($S_3$).

- **Action**: At any state, a UE can make a decision $d$ chosen from a set $D = \{d_1, d_2\}$ where $d_1$ and $d_2$, respectively, stand for whether to schedule this user’s UAs at the current frequency band or switch to the other frequency band.

- **State transition probability**: $T(S_i, d, S_j)$ denotes the probability of transition from state $S_i$ to $S_j$ if decision $d \in D$ is chosen by the UE. Hence, $T(S_i, d_2, S_3) = 1$ for $i = 1, 2$, and $T(S_3, d_2, S_1) = 1 - T(S_3, d_2, S_2) = \rho_a$. In addition, $T(S_3, d_1, S_3) = 1$ and $T(S_i, d_1, S_1) = 1 - T(S_i, d_1, S_2) = \rho_a$ for $i = 1, 2$.

- **Reward**: The UE receives rewards $r = [r_1, -r_2, r_3]$, respectively, for each of its UAs being at states $S_1$, $S_2$, and $S_3$, where $r_2 > r_1 > r_3 > 0$. The rewards are assumed the same for all UAs $a \in A$. The reward values affect both the convergence and the policy. For instance, for large negative rewards, i.e., $r_2 \gg r_3$, the optimal policy for the UA is to choose $\mu W$, even for large $\rho_a$ values. The long-term reward for choosing mmW band by UA $a \in A$ is $r_1\rho_a - (1 - \rho_a)r_2$. Therefore, we can set $r$ such that only for $\rho_a \geq \rho_{th}$, mmW band be preferred by UA $a$. That is, $r_1\rho_a - (1 - \rho_a)r_2 \geq r_3$ which implies

$$
\rho_{th} = \frac{r_3 + r_2}{r_1 + r_2}, \quad r_2 > r_1 > r_3 > 0.
$$

(28)

At any time slot, each UA that is selected for scheduling will explore one of the three states. Consequently, this UA’s corresponding UE will achieve a reward associated with the current residing state. We note that the UEs do not have any prior knowledge about the transition probabilities in advance. However, QL provides a model-free approach that instead of estimating $\rho_a$, it allows UE to find the best decision while residing at each state. This is done by the notion of Q-values $Q(S, d)$ which represents the value of decision $d$ while being at state $S$. Starting from an initial Q-values, UA can find true values via an iterative process as follows:

$$
Q(S, d) \leftarrow (1 - \alpha)Q(S, d) + \alpha \left[ r(S') + \gamma \max_{d'} Q(S', d') \right], \quad (29)
$$
where $\alpha$ and $\gamma$ are predetermined constants. It can be shown that updating the Q-table based on (29) maximizes the long-term expected reward: 
$$\bar{r} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} r(S(t))$$
Moreover, given the converged $Q$ values, the following sufficient condition can be used to find a subset of UAs with $\rho_a \geq \rho_{th}$:

$$Q(S_i, d_2) \leq Q(S_i, d_1), i = 1, 2 \quad \text{and} \quad Q(S_3, d_1) \leq Q(S_3, d_2) \Rightarrow \rho_a \geq \rho_{th}. \quad (30)$$

We note that if there is only one UE that is running only one UA, the criterion given by (29) leads to making optimal decisions in terms of maximizing the expected reward. However, the multi-user resource allocation cannot be done only based on $\rho_a \geq \rho_{th}$ criterion. On the one hand, assigning mmW resources only to UAs with high $\rho_a$ and small required load will result low spectral efficiency. Moreover, UAs with small $\rho_a$ and large $b_{req}(t)$ will not meet their delay requirement, if they are scheduled over the $\mu$W frequency band. However, even with small $\rho_a$, it is still probable for these UAs to be served over a LoS mmW link. Therefore, multi-user scheduling enforces SBS to exploit per UA context information, i.e., required load per UA, delay constraint, as well as UEs-SBS channel diversity.

C. Complexity Analysis of the Proposed Two-stage Solution

With this in mind, we can make the following observation with regard to the proposed two-stage solution in Algorithms 1 and 2 for the original problem in (6a)-(6f).
TABLE II
SIMULATION PARAMETERS

| Notation          | Parameter                           | Value  |
|-------------------|-------------------------------------|--------|
| $P_1, P_2$        | Transmit power                       | 30 dBm |
| $(\Omega_1, \Omega_2)$ | Available Bandwidth             | (10 MHz, 1 GHz) |
| $\omega$         | Bandwidth per RB                    | 180 KHz |
| $(\xi_1, \xi_2)$ | Standard deviation of mmW path loss | (10, 5.2) [3] |
| $(\alpha_1, \alpha_2)$ | Path loss exponent             | (3, 2) [3] |
| $(\beta_1, \beta_2)$ | Path loss at 1 m                  | (38, 70) dB |
| $\psi$           | Antenna gain                        | 18 dBi |
| $\tau$           | Time slot duration                  | 10 ms  |
| $\tau'$          | Beam-training overhead              | 0.1 ms |
| $N_0$             | Noise power spectral density        | $-174$ dBm/Hz |
| $J$               | Number of QoS classes               | 5 [25] |
| $\kappa$         | Number of UAs per UE               | 3      |
| $r$               | Reward vector                       | $[3, -16, 1]$ |

**Theorem 2.** The proposed long-term scheduling algorithm composed of Algorithm 1 and Algorithm 2 solves the problem (6a)-(6f) in polynomial time with respect to the number of UAs.

**Proof:** First, we analyze the complexity of Algorithm 1. For each slot $t$, let $A_{j \geq t} = |A_{j \geq t}|$ be the number of UAs that can be selected to be scheduled at $\mu$W band. At most, the algorithm must find the solution for $A_{j \geq t}$ number of matchings. In addition, each matching has the complexity of $O(K_1)$, since in the worst case, each UA must be re-allocated to $K_1$ RBs by SBS. Hence, the complexity of Algorithm 1 at each slot $t$ is $O(K_1 A_{j \geq t})$ and the total complexity from slot $t = 1$ to $t = J$ is $O(K_1 \sum_{j=1}^{J} j A_j)$.

Next, we analyze the complexity of Algorithm 2. At each slot $t$, there are at most $A_{j \geq t}$ UAs to be scheduled at mmW ($G_{t,1} = \emptyset$). Therefore, the Algorithm 2 must converge after $A_{j \geq t}$ resource allocations, where each allocation is a special case of the 0-1 Knapsack problem. Hence, the total complexity of the Algorithm 2 from slot $t = 1$ to slot $t = J$ is $O(\sum_{j=1}^{J} j A_j)$.

From the above results, the overall complexity of the proposed long-term scheduling is $O \left( (K_1 + 1) \sum_{j=1}^{J} j A_j \right)$.  


V. Simulation Results

For simulations, we consider an area with diameter $d = 200$ meters with the SBS located at the center. UEs are distributed uniformly within this area with a minimum distance of 5 meters from the SBS. Each UE has $\kappa$ UAs chosen randomly and uniformly from $J$ QoS classes. The main parameters are summarized in Table III. All statistical results are averaged over a large number of independent runs. We compare the performance of the proposed context-aware algorithm with two well-known resource allocation approaches:

- Proportional Fair Scheduler with minimum rate requirement (PF-MRR): The PF scheduling for multi-carrier systems with minimum rate requirement is different than the conventional approach. In [26], a simple approach is proposed to implement PF-MRR which we modify to apply to the dual-mode system. At $\mu$W band, RB $k$ is assigned to the UA $a \in A_t$ that satisfies

$$a = \arg\max_{a \in A_t} \frac{R_a(k, t)}{\bar{R}_a(t) + R_a^{req}(t)},$$  \hspace{1cm} (31)

where $\bar{R}_a(t)$ is the achieved average rate up to time slot $t$, and $R_a^{req}(t) = b_a^{req}(t)/\tau$ is the required average rate at slot $t$ to meet the QoS constraint of UA $a$. UAs $a \in A_{t \geq t}$ with unsatisfied rate requirement are scheduled at mmW band where $\tau_{a,t} = \frac{b_a^{req}(t)}{R_a(t)}$ is allocated to the UA $a = \arg\max_{a \in A_{t \geq t}} \frac{R_a(t)}{\bar{R}_a(t) + R_a^{req}(t)}$, while $\sum_a \tau_{a,t} = \tau - |G_{t,2}|\tau'$. 

- Round Robin Scheduler (RR): At $\mu$W band, the scheduler allocates equal number of RBs to each $a \in A_t$. Unsatisfied UAs $a \in A_{t \geq t}$ are scheduled at mmW band with $\tau_{a,t} = \frac{\tau - |G_{t,2}|\tau'}{|G_{t,2}|}$. 

A. Quality-of-Experience of the Users

Fig. 4 shows a snapshot of a given network realization in which specific UEs are represented by circles. Each UE is associated with $\kappa = 3$ UAs, each having a required load of $b_a = 1$ Mbits. The results from this figure show each user’s satisfaction by indicating how many UAs per UE are satisfied. In Fig. 4, the colors red, yellow, and green are used, respectively, to indicate one, two, and three satisfied UAs per UE. Moreover, circles with no color represent UEs with no serviced UA. Clearly, in Fig. 4 we can see that the proposed approach significantly improves the overall system performance by providing service to more UEs, compared to the RR scheme. In addition, we observe that, although the same number of UEs are served in both context-aware
and PF-MRR schemes, the proposed context-aware outperforms PF-MRR by satisfying the QoS needs of more applications, which naturally leads to a higher quality-of-experience per user.

B. Outage Probability vs Number of UEs

The overall outage probability, $P_{\text{out}}$, is defined as the ratio of the number of QoS violations over the total number of UAs which will be given by:

$$P_{\text{out}}(\pi) = 1 - \frac{1}{A} \left( \sum_{t=1}^{J} \lambda_{t,1}(\pi) - \sum_{t=1}^{J} \lambda_{t,2}(\pi) \right)$$

(32)

$$= 1 - \frac{1}{A} \sum_{t=1}^{J} \sum_{a \in A_t} 1(a; \pi) = 1 - \frac{1}{A} |O^\pi|,$$

Since $P_{\text{out}}$ is a random variable, we will study whether the proposed scheduling policy $\pi^* \in \mathcal{H}$ guarantees $\mathbb{P}(P_{\text{out}}(\pi^*) \geq P_{\text{th}}) \leq \epsilon$, where $P_{\text{th}}$ is the maximum tolerable outage probability and $\epsilon$ is a pre-defined threshold. This can be written as $F_{P_{\text{out}}}(P_{\text{th}}) \geq 1 - \epsilon$, where $F_{P_{\text{out}}}(\cdot)$ is the cumulative distribution function (CDF) of $P_{\text{out}}$.

Fig. 5 shows the outage probability as the number of UEs varies, for the three considered approaches. Fig. 5 show that the outage probability increases as the number of UAs increases. In fact, the results show the number of UAs that can be satisfied for a given outage threshold. Clearly, the proposed algorithm outperforms the PF-MRR and RR scheduling approaches. For example, for a 0.01 outage probability, the proposed context-aware approach satisfies up to 150 UAs, considering $\kappa = 3$ UAs per UE. However, the baseline approaches fail to achieve this performance. In fact, the outage probability is always greater than 0.02 for both the RR and the PF-MMR approaches over all network sizes. Finally, from Fig. 5 we can clearly see that the
proposed approach can always guarantee the QoS for 120 UAs, which is four times greater than the number of satisfied UAs resulting from the PF-MRR approach.

C. Impact of Q-Learning

Fig. 6 shows the gain of the proposed QL approach. The QL gain is defined as the respective number of satisfied UAs with and without QL. The results presented in Fig. 6 show that more
gain is achievable as the number of UAs increases. This stems from the fact that, as the number of UAs increases, it becomes more probable that more number of UEs be at a LoS connection with the BS. Fig. 6 shows that the QL-based information allows scheduling UAs with higher LoS probabilities. More interestingly, Fig. 6 shows that the gain increases as the required load per UA increases. This is due to the fact that with more strict QoS constraints, it is become more important to allocate mmW resources only to the UAs with higher probability of LoS.

Fig. 7a and Fig. 7b show both random state exploration and the resulting long-term rewards. The average LoS probabilities $\rho_a = 0.8$ and $\rho_a = 0.2$ are considered, respectively, in Fig. 7a and Fig. 7b. The results in Fig. 7a show that up to 110 iterations is needed for the QL algorithm to converge, for $\rho_a = 0.8$. However, Fig. 7b shows that the algorithm will converge within less than 190 iterations for $\rho_a = 0.2$. Moreover, the average reward is higher in Fig. 7a, since the UAs with $\rho_a = 0.8$ are often served over mmW LoS links, while in Fig. 7b, mmW links are frequently blocked.

In Fig. 8a and Fig. 8b, the average reward and state transitions are shown when the optimal QL policy is followed, respectively, for $\rho_a = 0.8$ and $\rho_a = 0.2$. Clearly, when LoS probability is high, the optimal policy is to schedule the UA over mmW band, as shown in Fig. 8a. In addition, compared to Fig. 7 we can see that the QL policy will substantially increase the average reward compared to the random frequency band selection. For example, for $\rho_a = 0.2$, the average reward
Fig. 8. State transition and average reward resulted by the optimal policy.

Fig. 9. Performance comparison between scheduling approaches versus the required bit per UA $b_a$. Parameters $M = 30$ and $\rho_a = 1$ for all $a \in A$ are used.

is increased from $-5$ in Fig. 7b to 1 in Fig. 8b.

D. Outage Probability vs the Required Load

In Fig. 9 we show the outage probability as the required load per UA varies, for the three scheduling approaches. In this figure, we can see that the outage probability decreases as the required load per UA decreases. In addition, from Fig. 9 we can see that the proposed context-aware approach yields significant gains, compared to the PF-MRR and RR schemes. In fact, the proposed approach guarantees the required loads up to 2 Mbits per UA, for 0.01 outage
probability. However, the baseline PF-MRR and RR approaches can guarantee less than 0.1 Mbits load per UA for the same outage probability.

E. Statistics of the Outage Probability

The empirical CDF of the outage probability is shown in Fig. 10 for $M = 30$ UAs with $b_a = 1$ Mbits required load. From Fig. 10 we can see that the proposed context-aware approach substantially improves the statistics of the outage, compared with PF-MRR and RR approaches. For example, the probability that $P_{\text{out}}$ be less than 0.2 is only 8% for PF-MRR approach, while this value is 60% for the proposed approach.

F. Effect of Beam Training Overhead

In Fig. 11 the effect of the beam training overhead on the outage probability is shown. Here, we observe that $\tau'$ will significantly affect the performance. From Fig. 11 we can clearly see that as $\tau'$ increases, the remaining time for data transmissions to UAs decreases which results in a higher outage probability. Fig. 11 shows that, in the absence of beam training overhead, the outage probability is always less than 0.3. However, for $\tau' = 0.6$ ms, the outage probability will always be less than 0.5.
Outage probability

Fig. 11. The CDF of the outage probability for $M = 30$ UAs and $b_a = 1$ Mbits. For the cell edge UEs, $\rho_a$ is sampled randomly from $[0, 1]$ and for others $\rho_a = 1$.

G. Number of Iterations

Fig. 12 shows number of iterations resulting from the proposed scheduling approach as the number of UEs varies for different number of UAs per UE. Clearly, the number of iterations increases almost linearly with the number of UEs. From this figure, we can see that even for large network size up to 30 UEs and 60 UAs, the proposed framework is relatively fast, as it
converges within 205 number of iterations.

VI. CONCLUSIONS

In this paper, we have proposed a novel context-aware scheduling framework for dual-mode small base stations operating at mmW and µW frequency bands. To this end, we have developed a two-stage UA selection and scheduling framework that takes into account various network and UA specific context information to make scheduling decisions. Over the µW band, we have formulated the context-aware scheduling problem as a one-to-many matching game. To solve this game, we have proposed a novel algorithm for joint UA selection and resource allocation and we have shown that it yields a two-sided stable matching between µW resources and UAs. Next, we have proposed a joint UA selection and scheduling to allocate mmW resources to the unscheduled UAs. The scheduling problem over mmW band is formulated as a 0-1 Knapsack problem and solved using a suitable algorithm. Moreover, we have proved that the proposed two-stage dual-mode scheduling framework can solve the problem in a polynomial time. Simulation results have shown the various merits and performance advantages of the proposed context-aware scheduling compared to the PF-MRR and RR approaches.

REFERENCES

[1] F. Boccardi, R. Heath, A. Lozano, T. 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] 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.

[4] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, “Millimeter wave cellular networks: A MAC layer perspective,” *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3437–3458, October 2015.

[5] D. W. K. Ng, M. Breiling, C. Rohde, F. Burkhardt, and R. Schober, “Energy-efficient 5G outdoor-to-indoor communication: SUDAS over licensed and unlicensed spectrum,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3170–3186, May 2016.

[6] J. Qiao, X. Shen, J. Mark, Q. Shen, Y. He, and L. Lei, “Enabling device-to-device communications in millimeter-wave 5G cellular networks,” *IEEE Communications Magazine*, vol. 53, no. 1, pp. 209–215, January 2015.

[7] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, June 2011.
[8] H. Singh, J. Oh, C. Kweon, X. Qin, H. Shao, and C. Ngo, “A 60 GHz wireless network for enabling uncompressed video communication,” *IEEE Communications Magazine*, vol. 46, no. 12, pp. 71–78, December 2008.

[9] D. Wu, J. Wang, Y. Cai, and M. Guizani, “Millimeter-wave multimedia communications: challenges, methodology, and applications,” *IEEE Communications Magazine*, vol. 53, no. 1, pp. 232–238, January 2015.

[10] 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*, 2016.

[11] J. Wells, *Multi-Gigabit Microwave and Millimeter-Wave Wireless Communications*. Artech House, 2010.

[12] M. Mezzavilla, A. Dhananjay, S. Panwar, S. Rangan, and M. Zorzi, “An MDP model for optimal handover decisions in mmwave cellular networks,” *arXiv:1507.00387*, Feb. 2016.

[13] S. Singh, S. Yeh, N. Himayat, and S. Talwar, “Optimal traffic aggregation in multi-RAT heterogeneous wireless networks,” *arXiv:1603.08062*, Mar. 2016.

[14] M. Wang, A. Dutta, S. Buccapatnam, and M. Chiang, “Smart exploration in hetnets: Minimizing total regret with mmwave,” in *Proc. IEEE International Conference on Sensing, Communication and Networking*, London, UK, June 2016.

[15] 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.

[16] T. Nitsche, A. B. Flores, E. W. Knightly, and J. Widmer, “Steering with eyes closed: Mm-wave beam steering without in-band measurement,” in *Proc. of 2015 IEEE Conference on Computer Communications*, Hong Kong, April 2015.

[17] E. Amaldi and V. Kann, “On the approximation of minimizing non zero variables or unsatisfied relations in linear systems,” *Theoretical Computer Science*, vol. 209, pp. 237–260, December 1998.

[18] K. Seong, M. Mohseni, and J. M. Cioffi, “Optimal resource allocation for OFDMA downlink systems,” in *Proc. of IEEE International Symposium on Information Theory*, Seattle, Washington, July 2006.

[19] A. E. Roth and M. A. O. Sotomayor, *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge University Press, 1992.

[20] 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.

[21] Y. Gu, W. Saad, M. Bennis, M. Debbah, and Z. Han, “Matching theory for future wireless networks: fundamentals and applications,” *IEEE Communications Magazine*, vol. 53, no. 5, pp. 52–59, May 2015.

[22] B. C. Dean, *Approximation algorithms for stochastic scheduling problems*. Boston: Ph.D. thesis, Massachusetts Institute of Technology, 2005.

[23] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. MA: MIT Press, Cambridge, 1998.

[24] M. Simsek, M. Bennis, and A. Czyliwik, “Dynamic inter-cell interference coordination in HetNets: A reinforcement learning approach,” in *Proc. of IEEE Global Communications Conference*, Anaheim, USA, Dec 2012.

[25] J. Tang and X. Zhang, “Cross-layer-model based adaptive resource allocation for statistical QoS guarantees in mobile wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2318–2328, June 2008.

[26] L. Qian, N. Song, D. R. Vaman, X. Li, and Z. Gajic, “Power control and proportional fair scheduling with minimum rate constraints in clustered multihop td/cdma wireless ad hoc networks,” in *Proc. of IEEE Wireless Communications and Networking Conference*, Las Vegas, USA, April 2006.