Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-022-03739-7

ORIGINAL RESEARCH

# WiGig access point selection using non-contextual and contextual multi-armed bandit in indoor environment

**Ehab Mahmoud Mohamed**[1,2]

Received: 13 May 2021 / Accepted: 19 January 2022


**Abstract**
Millimeter wave (mmWave) band, i.e., 30 ~ 300 GHz, supports multi-gigabit communication making it a main component of fifth generation (5G) and future six generation (6G) wireless communications. Wireless gigabit (WiGig) is the standardized 60 GHz mmWave band for WLAN applications. MmWave has intermittent short-range transmissions necessitating the installation of multiple WiGig access points (APs) using antenna beamforming training (BT) to fully cover a target indoor area. WiGig user equipment (UE) should select the best AP among the installed ones maximizing its achievable data rate. Conventionally, UE should exhaustively search the best AP having the highest received power using BT with all available APs, which reduces the attainable throughput in consequence. In this paper, the problem of WiGig AP selection is formulated as a multi-armed bandit (MAB) game, where, the UE is considered as the player aiming to maximize its long-term average throughput, i.e., the reward, through playing over the available APs, i.e., the arms of the bandit. Non-contextual MAB algorithms, namely upper confidence bound (UCB) and Thompson sampling (TS) are adopted to address the formulated problem. Moreover, as standardized WiGig devices are multi-band capable containing 2.4/5 GHz Wi-Fi and 60 GHz mmWave bands, Wi-Fi signal characteristics are used as contexts of the mmWave links to further enhance the WiGig AP selection policy. Thus, contextual MAB (CMAB) algorithms, namely linear UCB (LinUCB) and contextual TS (CTS) are also suggested. Simulation analyses confirm the superior performance of the CMAB algorithms over the non-contextual ones in addition to the conventional approaches accompanied with high convergence rates. For example, at no blockage and using too narrow beams of $\theta_{-3\text{dB}} = 10^{\circ}$, the proposed CTS, LinUCB, TS, and UCB schemes obtain 98.7%, 96.8%, 89%, 84% of the optimal performance, while the benchmark schemes obtain 49%, and 1.8%, respectively.

**Keywords** Wireless fidelity · Millimeter wave · Upper confidence bound · Thompson sampling · Linear upper confidence bound · Contextual Thompson sampling

## 1 Introduction

The utilization of the millimeter wave (mmWave), i.e., 30 ~ 300 GHz, band is considered as a promising approach for providing high data rates capacity for fifth generation (5G) and future six generation (6G) wireless communications (Sakaguchi et al. 2015; Mohamed et al. 2017). This is due to the large vacant frequencies available in this band enabling multi-gigabit transmissions (Mohamed et al. 2020a). For standardization activities, the communication in the 60 GHz band is ratified by IEEE 802.11ad Standard (2012) and IEEE 802.11ay (Ghasempour et al. 2017) standards, i.e., wireless gigabit (WiGig) standards, for wireless personal area networks (WPAN) and wireless local area networks (WLANs) applications. Standardized WiGig devices are multi-band capable containing 2.4/5 GHz Wi-Fi and 60 GHz mmWave bands. Due to its highly operating frequency, mmWave is characterized by high path loss and oxygen absorption (Rappaport et al. 2013a, b). Nevertheless, mmWave band is subjected to path blockage, even human shadowing can severely affect the mmWave path. To overcome these harsh channel impairments towards realizing mmWave communications, both academia and industry advocated the use of antenna beamforming. Thus,

✉ Ehab Mahmoud Mohamed
ehab_mahmoud@aswu.edu.eg

[1] Electrical Engineering Department, College of Engineering, Prince Sattam Bin Abdulaziz University, Wadi Ad Dawasir 11991, Saudi Arabia

[2] Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan 81542, Egypt

Published online: 19 February 2022

steerable multi-input multi-output (MIMO) antenna arrays using structured codebooks are typically used in mmWave transceiver (Abdelreheem et al. 2018; Alkhateeb et al. 2014). This results in a burden overhead due to the beamforming training (BT) process required for transmit (TX)/receive (RX) beam alignment (Mohamed et al. 2019). As a result, mmWave is typically characterized by intermittent short-range transmission limiting its coverage area to be few meters around a mmWave access point (AP) especially in indoor environment where human and obstacles blocking are severe (Sakaguchi et al. 2015; Mohamed et al. 2017). Thus, to fully cover a certain environment with mmWave transmissions, several WiGig APs should be installed in the target area, and the WiGig user equipment (UE) should select one of those APs to construct the WiGig link (Sakaguchi et al. 2015). Typically, AP selection is done based on maximum received signal strength (RSS) criterion, i.e., the AP having the maximum received power is selected (Sakaguchi et al. 2015). Although this technique works well in the sub-6 GHz band, it consumes a considerable overhead in the case of mmWave band. This is because the UE should perform BT with all available APs before selecting the one associated with the maximum RSS, which will highly decrease the achievable throughput in consequence. For example, about 1.8 ms is required for completing the BT process between a WiGig TX and RX using 32 different beam settings applying the ratified IEEE 802.11ad BT mechanism (Hosoya et al. 2015). If there are ten available APs, about 18 ms will be required to complete the BT process before selecting the one having the maximum RSS. This problem becomes more sever in the case of using too many beams having small beamwidths, e.g., $10^{\circ}$, where even low complexity BT techniques still need a considerable BT overhead for WiGig AP selection.

Multi-armed bandit (MAB) is a powerful online tool, where an agent tries to maximize her profit through playing over multiple bandit arms (Auer et al. 2002; Gutowski et al. 2018). By observing the achievable rewards, the player compromises between always playing with the arm giving the maximum average reward so for or investigating new ones, formally known as the *exploitation-exploration* dilemma (Audibert et al. 2009). Based on the rewards distribution, the MAB problem can be categorized as *stochastic* MAB, where the arms' rewards are drawn from independent and identical distributions (i.i.d), or *adversarial* MAB, where the arms' rewards are drawn from unknown distributions (Mohamed et al. 2020b; Hashima et al. 2021). Varity of MAB algorithms exist in literature to implement the MAB game in its both categories such as epsilon-greedy, upper confidence bound (UCB) (Francisco et al. 2019), Thompson sampling (TS) (Kaufmann et al. 2012), exponential weight algorithm for exploration and exploitation (EXP3) (Seldin et al. 2012), etc. Contextual MAB (CMAB) is a powerful type of MAB games, where the features of the arms, i.e., contexts, are utilized by the player while observing the achievable rewards as given in Gutowski et al. (2018). By relating the arms' contexts with their achievable rewards, the player can highly enhance its arm selection policy over non-contextual counterparts. Linear UCB (LinUCB) (Lihong et al. 2010) and contextual TS (CTS) (Agrawal and Goyal 2013) are two well-known algorithms used for realizing CMAB games by linearly relating the arms' context vectors and their achievable rewards.

In this paper, the WiGig AP selection problem is considered as a MAB game, where the WiGig UE only selects one AP at a time. The motivation behind this proposal comes from the potency of online learning, where UE reaches the optimal WiGig AP successively over time. This contributes to highly reducing the BT overhead, as only one AP is tested at a time, which maximizes the achievable spectral efficiency accordingly. Moreover, multi-connectivity and distributed antenna system (DAS) are not needed any more for WiGig UE association thanks to this online learning approach. Furthermore, motivated by standardized multi-band WiGig devices containing both 2.4/5 GHz Wi-Fi and 60 GHz WiGig interfaces, CMAB approach is proposed using Wi-Fi signal information as contexts of WiGig APs. Thus, the main contributions of this paper can be summarized as follows:

- The WiGig AP selection problem is formulated as a MAB game, where the UE is acting as the player aiming to maximize its achievable spectral efficiency, i.e., the reward of the MAB game, via playing over the available WiGig APs acting as the arms of the bandit. This will highly enhance the achievable spectral efficiency while keeping the BT overhead at the minimum level. UCB and TS algorithms are suggested to address the formulated problem to select the sub-optimal WiGig AP over the time horizon.
- Thanks to the standardized multi-band capable WiGig devices along with the direct relationship between Wi-Fi and mmWave link statistics as proved in Mohamed et al. (2020a) and in Mohamed et al. (2019), Wi-Fi signal statistics are used as contexts of the mmWave links. The proposed context vector consists of the instantaneous Wi-Fi received power, its average and variance up to time $t$. Thus, WiGig AP selection will be formulated as a CMAB game, and LinUCB, CTS algorithms are suggested to address it.
- Extensive numerical simulations are conducted to prove the effectiveness of the proposed MAB/CMAB based WiGig AP selection algorithms over the conventional maximum RSS and random selection approaches in terms of average throughput, average energy efficiency,

and average total delay. Moreover, the conducted simulations prove the high performance of the CMAB based schemes over the non-contextual MAB based ones in terms of both the overall performance and the convergence rates.

The reminder of this paper is organized as follows, Section II gives the related works, and Section III introduces the system model including the used Wi-Fi/ mmWave channel models and WiGig AP selection optimization problem. Section IV gives the proposed MAB and CMAB based WiGig AP selection including the suggested UCB, TS, LinUCB and CTS algorithms. Section V gives the conducted numerical simulations followed by the concluded remarks in Section VI. Table 1 summarizes the nomenclature used throughout this paper.

## 2 Literature review

Maximum RSS based AP selection faces numerous problems in the sub 6 GHz WLAN, such as load imbalance and intense channel contention. Consequently, numerous research works exist in literature to efficiently address the sub 6 GHz WLAN AP selection problem. For example, the authors in Chen et al. (2013) considered the problem of AP selection in variable channel-width WLANs under certain quality of service requirements. A game theoretic approach is utilized where the AP selection problem is considered as an evolutionary game. In Peng et al. (2019), a scheme based on the achievable normalized throughput is proposed for enabling the UE to select among the available Wi-Fi APs. The authors in Kim et al. (2017) proposed a Wi-Fi AP selection scheme based on the minimum interference level while considering the APs' traffic load. In Dwijaksara et al. (2019), a multi-objective optimization is formulated for optimal AP selection while considering various techniques, such as multi-cast transmission, multi-user MIMO, and AP sleeping. Readers interested in other techniques of sub 6 GHz WLAN AP selection are advised to check the related works stated in Peng et al. (2019) and Dwijaksara et al. (2019).

The AP selection problem in mmWave is unique from that in sub 6 GHz due to the following reasons (1) In the case of sub 6 GHz band, there is no intermittent transmissions due to path blocking, no need for BT, and the channel coefficients can be easily estimated which is not the case in mmWave. (2) The interference is not a dominant factor in the case of mmWave due to the use of beamforming, which reduces packet collisions. Due to its uniqueness, researchers investigated other approaches for mmWave small cell (AP) association (selection), which are mainly based on multi-connectivity (DAS) in case of outdoors (indoors), respectively (see Liu et al. 2016; Ožkoç et al. 2021). In both approaches, to overcome the intermittent transmissions of mmWave, the UE is permitted to connect with multiple small cells (APs) simultaneously, which confirms a continuous coverage for mmWave UE. Recent works in mmWave user association are given in Liu et al. (2021) and Liu et al. (2020), where multi-connectivity is utilized. In Liu et al. (2021), the user association was formulated to maximize the obtained throughput while assuring ultra-reliable low latency mmWave communication. In Liu et al. (2020), a machine learning approach is proposed for mmWave user association using multi-connectivity, where the problem is formulated as a multi-label classification. In Zhang et al. (2012), Yoo et al. (2019) and Zhang et al. (2021), DAS was proposed for providing AP diversity in the case of WiGig WLAN. In Zhang et al. (2012), the AP selection was formulated as a partially observed Markov decision process to obtain the optimal one within the DAS. In Yoo et al. (2019), three switched diversity combining schemes were introduced to combine the signals from the 60 GHz DAS. In Zhang et al. (2021), three different AP selection schemes are given while considering the cross-correlation coefficient among the RSS of the APs. Table 2

**Table 1** Nomenclature

| | | | |
|---|---|---|---|
| mmWave | Millimeter wave | UCB | Upper confidence bound |
| 5G | Fifth generation | TS | Thompson sampling |
| 6G | Six generation | CMAB | Contextual multi-armed bandit |
| WiGig | Wireless gigabit | LinUCB | Linear upper confidence bound |
| AP | Access point | CTS | Contextual thompson sampling |
| BT | Beamforming training | WPAN | Wireless personal area network |
| UE | User equipment | WLAN | Wireless local area network |
| MAB | Multi-armed bandit | MIMO | Multi-input multi-output |
| RSS | Received signal strength | LoS | Line-of sight |
| DAS | Distributed antenna system | SBIFS | Short beamforming interframe space |
| MBIFS | Medium beamforming interframe space | | |

**Table 2** Comparisons of AP selection methodologies

| References | Methodology | Pros | Cons |
|---|---|---|---|
| Chen et al. (2013) | Game theoretic approach | Considering variable channel-width WLANs under certain quality of service (QoS) requirements | Applicable for sub 6 GHz |
| Peng et al. (2019) | Analyzing the achievable normalized throughput | Maximizing the achievable normalized throughput | Applicable for sub 6 GHz |
| Kim et al. (2017) | Load aware and interference minimization | Load balancing among installed Aps | Applicable for sub 6 GHz |
| Dwijaksara et al. (2019) | Multi-objective optimization considering multiple parameters | Load balancing among installed APs and energy saving | Applicable for sub 6 GHz |
| Liu et al. (2021) | Multi-connectivity using queueing theory and matrix geometric method | Applicable for mmWave, and exactly matches the performance of the exhaustive search | High management and controlling complexities due to multi-connectivity and the need for CSI estimation |
| Liu et al. (2020) | Machine learning approach using multi-label classification | Applicable for mmWave, and no need for mmWave CSI estimation | High management and controlling complexities due to multi-connectivity |
| Zhang et al. (2012) | Observed Markov decision process | Applicable for mmWave by modelling mmWave blockage as a Markovian process | High complexity due to full beamforming training with the multiple APs and distributed antenna system |
| Yoo et al. (2019) | Diversity combining schemes | Applicable for mmWave by utilizing the composite fading and withstanding blockage | High complexity due to full beamforming training with the multiple APs and distributed antenna system |
| Zhang et al. (2021) | AP selection using Spearman's rank-order correlation | Applicable for mmWave by considering the cross-correlation coefficient among the RSS from the multiple APs | High complexity due to full beamforming training with the multiple APs and distributed antenna system |

summarizes some of the existing AP selection schemes in both sub 6 GHz and mmWave bands.

The MAB based WiGig AP selection scheme proposed in this paper is different from the multi-connectivity and DAS approaches from the following aspects (1) In the proposed scheme, the APs are operating autonomously, where in the other approaches, a central controller is needed which complicates the network architecture. (2) In the proposed scheme, the UE does not need to connect with multiple APs simultaneously as in the other approaches, which highly relaxes the management and controlling complexities. (3) In both approaches, the problem of mmWave BT overhead was highly relaxed by assuming that the mmWave channel can be accurately estimated and the beamforming can be accurately performed in accordance, which is not a practical assumption. Instead, in the proposed MAB based WiGig AP selection scheme, the UE plays over the distributed autonomously operating APs to select the one having the maximum spectral efficiency and experiencing no blockage in consequence. This is done in timely basis thanks to modeling the problem as an online MAB game.

## 3 System model

Figure 1 shows the system model of a WiGig WLAN installed in an indoor office environment, where a multi-band WiGig APs containing 2.4/5 GHz Wi-Fi and 60 GHz mmWave bands are deployed inside it. A multi-band WiGig UE is located inside this area, where some APs can cover it and the others are undergoing path blocking due to shelfs, desks, etc. The UE should select one of the covering APs to establish the WiGig communication link. In the followings, we will present the used Wi-Fi link model, mmWave link model including the mmWave blocking probability, and the WiGig AP selection optimization problem.

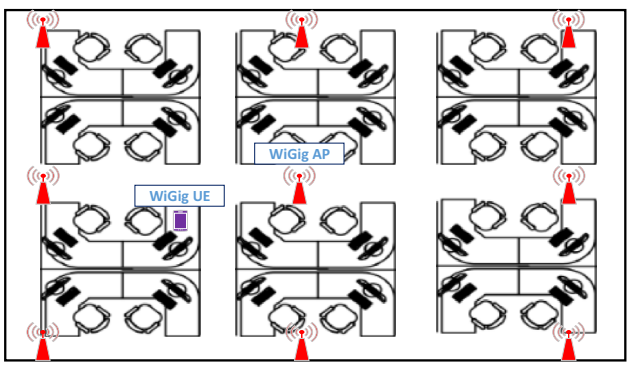



**Fig. 1** WiGig WLAN system model in an office indoor environment

### 3.1 Wi-Fi link model

For Wi-Fi, Wi-Fi link model given in Gao et al. (2007) operating at 5.25 GHz will be utilized, where the Wi-Fi received power, $\mathrm{P}_r^w$, at a separation distance $r$ from the Wi-Fi TX is given by:

$$\mathrm{P}_r^w[\mathrm{dBm}] = \mathrm{P}_t^w[\mathrm{dBm}] - \beta_w - 10n_w\log_{10}(r) - \delta_w, \tag{1}$$

where $\mathrm{P}_t^w[\mathrm{dBm}]$ indicates the Wi-Fi TX power in dBm, and the path loss at a reference distance $r_0$ is given by $\beta_w$. $n_w$ is the path loss exponent while $\delta_w \sim \mathcal{N}(0, \sigma_w)$ is the log-normal shadowing term with zero mean and standard deviation of $\sigma_w$.

### 3.2 MmWave link model and blockage probability

Following the mmWave link models given in Mohamed et al. (2020a) and Mohamed et al. (2019), the mmWave received power, $\mathrm{P}_r^m$, at a separation distance $r$ from a mmWave TX due to line-of sight (LoS) and non-LoS (NLoS) paths, including the effect of TX/RX beamforming gain, can be expressed as:

$$\mathrm{P}_r^m = \mathrm{P}_t^m A_{TX}(\theta,\theta_{TX})A_{RX}(\varphi,\varphi_{RX})\left(\underbrace{\varepsilon\left(\mathbb{P}_{\mathrm{LoS}}(r)\right)10^{-0.1\beta_m^{\mathrm{LoS}}}r^{-n_m^{\mathrm{LoS}}}\mathcal{Z}_{m,\mathcal{LN}}^{\mathrm{LoS}}}_{\Lambda_{\mathrm{LoS}}} + \underbrace{\upsilon\left(\mathbb{P}_{\mathrm{NLoS}}(r)\right)10^{-0.1\beta_m^{\mathrm{NLoS}}}r^{-n_m^{\mathrm{NLoS}}}\mathcal{Z}_{m,\mathcal{LN}}^{\mathrm{NLoS}}}_{\Lambda_{\mathrm{NLoS}}}\right) \tag{2}$$

where $\mathrm{P}_t^m$ is the mmWave TX power in watt. $A_{TX}(\theta,\theta_{TX})$ and $A_{RX}(\varphi,\varphi_{RX})$ are the TX and RX beamforming gain, where $\theta$ and $\varphi$ are the angle of departure (AoD) and angle of arrival (AoA), and $\theta_{TX}$ and $\varphi_{RX}$ are the boresight angles of the TX and RX beams. Utilizing the 2D steerable antenna model given in Mohamed et al. (2020a) and Mohamed et al. (2019) with the Gaussian main loop profile, $A_{TX}(\theta,\theta_{TX})$ can be written as:

$$A_{TX}(\theta,\theta_{TX}) = A_0\exp\left(-4\ln(2)\left(\frac{\theta-\theta_{TX}}{\theta_{-3\mathrm{dB}}}\right)^2\right), A_0 = \left(\frac{1.6162}{\sin\left(\frac{\theta_{-3\mathrm{dB}}}{2}\right)}\right)^2, \tag{3}$$

where $A_0$ is the maximum beamforming gain, and $\theta_{-3\mathrm{dB}}$ is the half power beamwidth. Same definition is applied for $A_{RX}(\varphi,\varphi_{RX})$, except that $\theta$, $\theta_{TX}$, and $\theta_{-3\mathrm{dB}}$ are replaced by $\varphi$, $\varphi_{RX}$ and $\varphi_{-3\mathrm{dB}}$, respectively.

In (2), $\Lambda_{\mathrm{LoS}} = \varepsilon\left(\mathbb{P}_{\mathrm{LoS}}(r)\right)10^{-0.1\beta_m^{\mathrm{LoS}}}r^{-n_m^{\mathrm{LoS}}}\mathcal{Z}_{m,\mathcal{LN}}^{\mathrm{LoS}}$ and $\Lambda_{\mathrm{NLoS}} = \upsilon\left(\mathbb{P}_{\mathrm{NLoS}}(r)\right)10^{-0.1\beta_m^{\mathrm{NLoS}}}r^{-n_m^{\mathrm{NLoS}}}\mathcal{Z}_{m,\mathcal{LN}}^{\mathrm{NLoS}}$ are the LoS and NLoS components of the received power. $\beta_m^{\mathrm{LoS}}$/ $\beta_m^{\mathrm{NLoS}}$ are LoS/ NLoS path losses at a reference distance $r_0$, and $n_m^{\mathrm{LoS}}/n_m^{\mathrm{NLoS}}$ are their path loss exponents. $\mathcal{Z}_{m,\mathcal{LN}}^{\mathrm{LoS}} \sim \mathcal{LN}\left(0, \left(\delta_{m,\mathcal{LN}}^{\mathrm{LoS}}\right)^2\right)$ and $\mathcal{Z}_{m,\mathcal{LN}}^{\mathrm{NLoS}} \sim \mathcal{LN}\left(0, \left(\delta_{m,\mathcal{LN}}^{\mathrm{NLoS}}\right)^2\right)$ are LoS and NLoS log-normal shadowing terms with zero mean and variances of $\left(\delta_{m,\mathcal{LN}}^{\mathrm{LoS}}\right)^2$ and $\left(\delta_{m,\mathcal{LN}}^{\mathrm{NLoS}}\right)^2$, respectively. From Singh et al. (2015), $\delta_{m,\mathcal{LN}}^{\mathrm{LoS}} = 0.1\sigma_m^{\mathrm{LoS}}\ln 10$ and $\delta_{m,\mathcal{LN}}^{\mathrm{NLoS}} = 0.1\sigma_m^{\mathrm{NLoS}}\ln 10$ giving that $\sigma_m^{\mathrm{LoS}}$ and $\sigma_m^{\mathrm{NLoS}}$ indicate the standard deviations of LoS and NLoS normal shadowing. $\varepsilon\left(\mathbb{P}_{\mathrm{LoS}}(r)\right)$ and $\upsilon\left(\mathbb{P}_{\mathrm{NLoS}}(r)\right)$ are two Bernoulli numbers indicating the availability of the LoS and NLoS paths. That is $\varepsilon\left(\mathbb{P}_{\mathrm{LOS}}(r)\right)$ and $\upsilon\left(\mathbb{P}_{\mathrm{NLoS}}(r)\right)$ are equal to 1 with probabilities $\mathbb{P}_{\mathrm{LoS}}(r)$ and $\mathbb{P}_{\mathrm{NLoS}}(r)$, which are the LoS and NLoS availabilities as functions of the separation distance $r$, and $\mathbb{P}_{\mathrm{NLoS}}(r) = 1 - \mathbb{P}_{\mathrm{LoS}}(r)$. In this paper, we follow the mmWave LoS blockage model given in Wu et al. (2018), which is applicable for both indoors and outdoors. In this model, blockers of cylindrical shapes are randomly distributed between mmWave TX and RX as a Poisson point process (PPP). Thus, the probability that there are no blockers intersecting the LoS path between TX and RX as a function of their separation distance $r$ can be expressed as:

$$\mathbb{P}_{\mathrm{LoS}}(d) = \Delta e^{-\omega r}, \tag{4}$$

$\Delta = e^{-\pi\lambda\xi\mathbb{E}(\Omega)^2}$ and $\omega = 2\lambda\xi\mathbb{E}(\Omega)$, where the density of the obstacles is represented by $\lambda$, and radius of the obstacles is represented by $\Omega$, which is a R.V. $\xi$ is the thinning factor, which is equal to 1 for indoors, and $\mathbb{E}(.)$ represents the average value or expectation operation. For detailed derivations of (4) and its associated parameters, readers are advised to refer to Wu et al. (2018). Typically, the LoS is the most dominant in mmWave communications as the LoS path is almost 20 dB higher than the NLoS counterpart (Mohamed et al. 2020a). Thus, for the sake of simplicity and without loss of generality, in this paper, only the LoS component of (2) will be considered.

### 3.3 WiGig AP selection optimization problem

Suppose that there is a WiGig UE located inside an indoor area of $N$ deployed WiGig APs that can cover it at its current location. The UE will conduct BT with $M$ candidate APs out of the total $N$ APs and chooses the best AP maximizes its achievable throughput, $\Psi$, in in bps, as follows:

$$i^* = \max_{\forall i\in\varnothing_M}\left(\Psi_i\right) = \max_{\forall i\in\varnothing_M}\frac{WT_D}{T_D + MT_{BT}}\left(\eta_i\widehat{\psi}_i\right), \tag{5}$$

s.t.
(1) $\eta_i \in (0,1)$
(2) $T_D, T_{BT} \in R^+$
(3) $\varnothing_M \subset \varnothing_N$

where W is the available bandwidth, and $\varnothing_M$ is the space of the $M$ candidate APs while $\varnothing_N$ is the space of all deployed APs. $T_D, T_{BT} \in R^+$ are the time duration of data transmissions and the time required for completing the BT process, respectively, where $R^+$ represents the set of positive real numbers. $\psi_i = \eta_i \widehat{\psi}_i$ is the available spectral efficiency in bps/HZ, where $\eta_i$ is the percentage of achievable spectral efficiency $\widehat{\psi}_i$ that can be achieved by the $UE - AP_i$ WiGig link. $\eta_i \in (0,1)$ depends on the available resources for the $UE - AP_i$ link such as the number of available time slots, etc. $\widehat{\psi}_i$ can be expressed as:

$$\widehat{\psi}_i = \log_2\left(1 + \frac{\mathrm{P}^m_{r,i}}{\sigma_0^2 + I_{j\neq i}}\right), \tag{6}$$

where $\mathrm{P}^m_{r,i}$ is the WiGig power received at UE from $AP_i$ assuming downlink transmissions, $\sigma_0^2$ is the noise power, and $I_{j\neq i}$ is the interference power at the UE coming from the other operating APs. Due to the high TX/RX directionality of the mmWave transmissions, $I_{j\neq i}$ can be neglected for the sake of simplicity.

Conventionally, UE exhaustively searches over all $N$ APs using BT to select the best one based on maximum RSS criterion, i.e., $M = N$, which highly increases the BT overhead and reduces the achievable throughput in consequence as given in (5). In the followings, solutions based on MAB formulation to the AP selection problem will be introduced based on both non-contextual and contextual MAB approaches. In both approaches, only one AP is selected at a time, i.e., $M = 1$, which is comparable to the optimal policy, where the best AP is selected at once using one BT operation. Thus, the aim of this MAB formulation is to maximize the available spectral efficiency $\psi_i$ while keeping the BT overhead at the lowest level.

## 4 Proposed MAB and CMAB approaches

In this section, the WiGig AP selection problem is formulated as a MAB game, where the WiGig UE acts as the player of the game aiming to maximize its available spectral efficiency, i.e., the payoff, through sequentially selecting among the deployed WiGig APs act as the arms of the bandit. This MAB formulation can be expressed as:

$$\max_{\mathbb{I}(1),\ldots,\mathbb{I}(T_H)} \frac{1}{T_H} \sum_t \sum\nolimits_i {}^{\mathbb{I}_{i_t}} \frac{WT_D}{T_D + T_{BT}} \left(\eta_{i_t} \widehat{\psi}_{i_t}\right), \tag{7}$$

s.t.
(1) $T_H \in (0, Z^+)$
(2) $\eta_{i_t} \in (0,1)$
(3) $T_D, T_{BT} \in R^+$
(4) $\sum_i \mathbb{I}_{i_t} = 1, 1 \le i \le N,$

where $T_H \in (0, Z^+)$ is the time horizon of the MAB game, and $Z^+$ represents the set of positive integer numbers. $\eta_{i_t} \widehat{\psi}_{i_t}$ is the available spectral efficiency results from selecting WiGig AP $i$ at time $t$, i.e., $AP_{i_t}$. $\mathbb{I}_{i_t}$ is the linkage indicator, which means that AP $i$ is selected at time $t$ by the UE. The fourth constraint $\sum_i \mathbb{I}_{i_t} = 1, 1 \le i \le N$ indicates that only one AP is allowable to be selected by the UE at time $t$.

**Algorithm 1:** UCB Policy for WiGig AP Selection

**Output:** $AP_{i_t^*}$
**Input:** $\varnothing_N$
**Initialization**: Selecting each AP for once then:
- Obtain its corresponding reward $\psi_{i_t}$.

**For** $t = N+1, \ldots, T_H$
1. Select a WiGig AP and obtain its corresponding reward
   - $i_t^* = \arg\max_{\forall i \in \varnothing_N} \left(\mathbb{E}\left(\psi_{i_{t-1}}\right) + \sqrt{\frac{2\,ln(t)}{x_{i_{t-1}}}}\right)$
   - Obtain $\psi_{i_t^*}$
2. Update $x_{i_t^*}$ and $\mathbb{E}\left(\psi_{i_t^*}\right)$
   - $x_{i_t^*} = x_{i_{t-1}^*} + 1$
   - $\mathbb{E}\left(\psi_{i_t^*}\right) = \frac{1}{x_{i_t^*}} \sum_{j=1}^{x_{i_t^*}} \psi_{i_j^*}$

**END For**

Two approaches are proposed to address this MAB formulation, one is context-free MAB, and the other is contextual based MAB, i.e., CMAB. The CMAB one takes advantage of the multi-band capability of the WiGig devices along with the direct relationship between Wi-Fi and mmWave link statistics as given in Mohamed et al. (2020a) and Mohamed et al. (2019).

## 4.1 Non-contextual MAB approach

Herein, we will introduce the proposed UCB and TS policies for addressing the formulated MAB based WiGig AP selection problem.

### 4.1.1 Proposed UCB policy

UCB is one of the famous context-free MAB algorithms that can efficiently address the *exploitation-exploration* dilemma. It frequently exploits the arm having the maximum average reward while exploring the less utilized ones. Algorithm 1 gives the proposed UCB algorithm for WiGig AP selection problem. At the beginning of the algorithm, each AP will be tested once and its corresponding reward $\psi_{i_t}$ is obtained. At each time $t$, the WiGig AP out of $\varnothing_N$ maximizing the following equality will be selected by the algorithm:

$$i_t^* = \arg\max_{\forall i \in \varnothing_N} \left( \mathbb{E}\left(\psi_{i_{t-1}}\right) + \sqrt{\frac{2ln(t)}{x_{i_{t-1}}}} \right), \tag{8}$$

where $x_{i_{t-1}}$ gives the number of times $AP_i$ is selected up to $t-1$ and $\mathbb{E}\left(\psi_{i_{t-1}}\right)$ is its corresponding average spectral efficiency. In (8), $\mathbb{E}\left(\psi_{i_{t-1}}\right)$ represents the exploitation term while $\sqrt{\frac{2ln(t)}{x_{i_{t-1}}}}$ represents the exploration term of the UCB algorithm. After selecting $i_t^*$ for constructing the WiGig link at time $t$, its corresponding spectral efficiency is obtained $\psi_{i_t^*}$. Then, its corresponding values of $x_{i_t^*}$ and $\mathbb{E}\left(\psi_{i_t^*}\right)$ are updated as given in Algorithm 1.

### 4.1.2 Proposed TS policy

TS constructs prior/posterior distributions for the arms' rewards as it is based on a pure Bayesian policy. Typically, TS outperforms UCB especially when the assumed probabilistic model of the arms' rewards matches their actual distributions. The parameters of the probabilistic model are initialized at the beginning of the TS algorithm for each arm. Samples are arbitrarily drawn from these distributions, and the arm with the highest random sample is played. After obtaining the reward corresponding to this played arm, its model parameters are updated for the next round of arms' selection while those related to the non-selected arms are not changed. Algorithm 2 summarizes the proposed TS based WiGig AP selection algorithm, where Gaussian distributions are assumed for the APs' rewards, i.e., spectral efficiencies $\psi_{i_t}$. Thus, the rewards are assumed to be drawn from $\mathcal{N}\left(\mathbb{E}\left(\psi_{i_t}\right), \sigma_{i_t}^2\right)$, where $\mathbb{E}\left(\psi_{i_t}\right)$ is the mean of the normal distribution and $\sigma_{i_t}^2$ is its variance. This assumption comes from the log-normal shadowing term in the mmWave received power given in (2). The APs' parameters, $\mathbb{E}\left(\psi_{i_t}\right)$, $\sigma_{i_t}^2$ and the number of arm selections $x_{i_t}$, are initialized at the beginning of the TS algorithm, and set to 1, 0, and 0, respectively. At each round of the TS algorithm, random samples, $\tau_{i_t}$, are taken from the prior spectral efficiency distributions $\mathcal{N}\left(\mathbb{E}\left(\psi_{i_{t-1}}\right), \sigma_{i_{t-1}}^2\right)$ of the APs. Then, the AP having the maximum sample is selected as follows:

$$i_t^* = \underset{\varnothing_N}{\text{argmax}}\left(\tau_{i_t}\right), \tag{9}$$

After playing the selected AP, $AP_{i_t^*}$, its reward $\psi_{i_t^*}$ is obtained. Then, its associated parameters $x_{i_t^*}$, $\mathbb{E}\left(\psi_{i_t^*}\right)$, and $\sigma_{i_t^*}^2$ are updated for posterior distributions calculations and next round of arms' selection as given in Algorithm 2.

**Algorithm 2:** TS Policy for WiGig AP Selection

**Output:** $AP_{i_t^*}$
**Inputs:** $\emptyset_N$
**Initialize:** $t = 0, \mathbb{E}(\psi_{i_t}) = 0, x_{i_t} = 0, \sigma_{i_t}^2 = 1 \; \forall i \in \emptyset_N$
**For** $t = 1, 2, 3, \ldots, T_H$

1. Sample $\tau_{i_t}, \forall i \in \emptyset_N$, from normal distributions $\mathcal{N}(\mathbb{E}(\psi_{i_{t-1}}), \sigma_{i_{t-1}}^2)$
2. Select a WiGig AP and obtain its corresponding reward
   - $i_t^* = \arg\max_{\emptyset_N}(\tau_{i_t})$
   - Obtain $\psi_{i_t^*}$
3. Update $x_{i_t^*}$, $\mathbb{E}(\psi_{i_t^*})$ and $\sigma_{i_t^*}^2$
   - $x_{i_t^*} = x_{i_{t-1}^*} + 1$
   - $\mathbb{E}(\psi_{i_t^*}) = \frac{1}{x_{i_t^*}} \sum_{j=1}^{x_{i_t^*}} \psi_{i_j^*}$
   - $\sigma_{i_t^*}^2 = \frac{1}{x_{i_t^*}+1}$

**END For**

### 4.2 CMAB approach

CMAB is an efficient variant of the MAB game, where the contexts of the arms along with its previous selections are considered when selecting the best arm for the next round as shown in Gutowski et al. (2018). Thus, the decision is taken not only based on the previous rewards' observations, as in the case of non-contextual MAB, but also based on the current/past context vectors of the arms. The job of the learner is to relate the context vectors and the rewards, so she can anticipate the best arm expected to give the best reward just by exploring the arms' context vectors. By efficiently relating the context vectors and the achievable rewards, CMAB has better performance guarantee than context-free counterparts. Varity of CMAB applications exist in literature, such as articles/movie recommendations, clinical trials, etc.

Utilizing the multi-band capability of WiGig devices and the direct relationship between Wi-Fi and mmWave link statistics as given in Mohamed et al. (2020a) and Mohamed et al. (2019), the characteristics of the Wi-Fi RSS are used as contexts of the mmWave spectral efficiency. In this regard, the constructed context vector of AP $i$ at time $t$, $\boldsymbol{b}_{i_t}$, includes the instantaneous Wi-Fi RSS at time $t$ in addition to its average and variance values till time $t$. This can be expressed as:

$$\boldsymbol{b}_{i_t} = \left[\mathrm{P}_{r,i_t}^{w}, \mathbb{E}\left(\mathrm{P}_{r,i_t}^{w}\right), var\left(\mathrm{P}_{r,i_t}^{w}\right)\right]^T, \tag{10}$$

where $(.)^T$ means transpose. As we previously explained, CMAB policy resides in expecting the value of $\psi_{i_t}$ giving the context vector $\boldsymbol{b}_{i_t}$ and the history of arms selections information $\mathcal{H}_{t-1}$, where linear relationship is assumed as follows:

$$\mathbb{E}\left[\psi_{i_t} \backslash \boldsymbol{b}_{i_t}, \mathcal{H}_{t-1}\right] = \boldsymbol{b}_{i_t}^T \mathbf{Q}_i^* \tag{11}$$

where $\mathbf{Q}_i^*$ is the optimal parameter of this relationship and $\mathcal{H}_{t-1} = \left\{ i_j, \psi_{i_j}, \boldsymbol{b}_{i_j}, i \in \varnothing_N, 1 \le j \le t-1 \right\}$. Thus, the task of the CMAB algorithms is to anticipate the value of $\mathbf{Q}_i^*$, i.e., $\widehat{\mathbf{Q}}_i$, through successive online learning. LinUCB given in Lihong et al. (2010) and CTS given in Agrawal et al. (2013) are known as efficient CMAB algorithms using two different procedures for approximating $\mathbf{Q}_i^*$, which will be utilized to address the problem of WiGig AP selection in this paper.

### 4.3 Proposed LinUCB policy

LinUCB was firstly introduced by the authors in Lihong et al. (2010) when working on personalized news articles recommendation while considering the context information of the users and their selected articles. In this work, two algorithms were introduced: namely the disjoint LinUCB and the hybrid LinUCB. In the disjoint LinUCB, the context vectors are not shared among the played arms while they are shared in the hybrid one. In this paper, we utilize the disjoint LinUCB as it is the most fitted to the mmWave AP selection problem as the APs are autonomously operating without central management. In LinUCB, $\mathbf{Q}_i^*$ is estimated using ridge regression as follows (Lihong et al. 2010):

$$\widehat{\boldsymbol{Q}}_i = \boldsymbol{G}_i^{-1} \boldsymbol{c}_i \tag{12}$$

$$\boldsymbol{G}_i = \boldsymbol{B}_i^T \boldsymbol{B}_i + \mathbf{I}_d \tag{13}$$

$$\boldsymbol{c}_i = \psi_i \boldsymbol{b}_i \tag{14}$$

where $\mathbf{B}_i$ is the matrix of context vectors of size $k \times l$ containing the past $k$ context vectors of length $l$, and $\mathbf{I}_d$ is the identity matrix of size $l \times l$, where $l = 3$ as given in (10). It is proved in Walsh et al. (2009) that the following inequality holds, which is utilized by the authors in Lihong et al. (2010) when developing the LinUCB algorithm:

$$\left| \boldsymbol{b}_{i_t}^T \widehat{\mathbf{Q}}_i - \mathbb{E}\left[\psi_{i_t} \backslash \boldsymbol{b}_{i_t}\right] \right| \leq \alpha \sqrt{\boldsymbol{b}_{i_t}^T \mathbf{G}_i^{-1} \boldsymbol{b}_{i_t}} \tag{15}$$

where $\alpha$ is a design parameter for tuning the accuracy of the approximation. Thus, the AP maximizing the following equation is selected for constructing the WiGig communication link at time $t$.

$$i_t^* = \underset{\varnothing_N}{\operatorname{argmax}} \left( \boldsymbol{b}_{i_t}^T \widehat{\mathbf{Q}}_i + \alpha \sqrt{\boldsymbol{b}_{i_t}^T \mathbf{G}_i^{-1} \boldsymbol{b}_{i_t}} \right) \tag{16}$$

where the term $\boldsymbol{b}_{i_t}^T \widehat{\mathbf{Q}}_i$ is the expected payoff demonstrating the exploitation term, while the term $\alpha \sqrt{\boldsymbol{b}_{i_t}^T \mathbf{G}_i^{-1} \boldsymbol{b}_{i_t}}$ is its standard deviation representing the exploration term. Algorithm 3 summarizes the proposed LinUCB policy for WiGig AP selection, where the inputs to this algorithm are the value of $\alpha$ and the space of all available relays $\varnothing_N$. Also, for initialization $\mathbf{G}_i \leftarrow \mathbf{I}_d$, $\boldsymbol{c}_i \leftarrow 0_{l\times 1}$. After selecting $AP_{i_t^*}$, its available spectral efficiency $\psi_{i_t^*}$ is obtained, and then its parameters are updated for the next round of the AP selection as given in Algorithm 3.

### 4.4 Proposed CTS policy

CTS is the contextual MAB variant of TS, where it f ollows the same pure Bayesian strategy. Thus, CTS constructs prior/posterior distributions for $\widehat{\mathbf{Q}}_i$ given $\boldsymbol{b}_{i_t}$ based on a predefined probabilistic model (Agrawal et al. 2013). As mmWave spectral efficiency comes from normal distributions due to the log-normal shadowing term in (2), multivariate gaussian distributions are considered for $\widehat{\mathbf{Q}}_i$, i.e., $\widehat{\mathbf{Q}}_i \sim \mathcal{N}\left(\mathbb{E}\left(\widehat{\mathbf{Q}}_i\right), h^2 \mathbf{B}_i^{-1}\right)$, where $\mathbb{E}\left(\widehat{\mathbf{Q}}_i\right)$ and $h^2 \mathbf{B}_i^{-1}$ are its means and variance. $\mathbf{B}_i$ is previously defined, $h$ is a tuning design parameter, and $\mathbb{E}\left(\widehat{\mathbf{Q}}_i\right)$ is given as (Agrawal et al. 2013):

$$\mathbb{E}\left(\widehat{\mathbf{Q}}_i\right) = \mathbf{B}_i^{-1} \boldsymbol{c}_i, \tag{17}$$

where $\boldsymbol{c}_i$ is given in (14). Algorithm 4 summarizes the proposed CTS policy for WiGig AP selection. The inputs to the algorithm are the design parameter $h$ and $\varnothing_N$. The AP's parameters are initialized at the beginning of the algorithm as follows $\mathbf{B}_i \leftarrow \mathbf{I}_d$, $\mathbb{E}\left(\widehat{\mathbf{Q}}_i\right) \leftarrow 0_{l\times 1}$, $\boldsymbol{c}_i \leftarrow 0_{l\times 1}$ for $\forall i \in \varnothing_N$. Then, at each time step, random $l$ dimensional samples, $\widetilde{\mathbf{Q}}_{i_t}$, are drawn from the prior multi-variate distributions of each candidate AP, i.e., $\widetilde{\mathbf{Q}}_{i_t} \sim \mathcal{N}\left(\mathbb{E}\left(\widehat{\mathbf{Q}}_{i_{t-1}}\right), h^2 \mathbf{B}_{i_{t-1}}^{-1}\right)$. Then, the AP having the maximum sample value will be selected by the UE for constructing the WiGig link as follows:

**Algorithm 3:** LinUCB Policy for WiGig AP Selection

**Output:** $AP_{i_t^*}$
**Input:** $\alpha \in \mathbb{R}^+$, $\varnothing_N$
**Initialization**: $\mathbf{G}_i \leftarrow \mathbf{I}_d$, $\boldsymbol{c}_i \leftarrow \mathbf{0}_{l\times 1}$ for $\forall i \in \varnothing_N$
**For** $t = 1, 2, 3, \ldots, T_H$

1. Observe $\boldsymbol{b}_{i_t}$ and Obtain $\widehat{\mathbf{Q}}_i \leftarrow \mathbf{G}_i^{-1} \boldsymbol{c}_i$ for $\forall i \in \varnothing_N$
2. Select a WiGig AP and obtain its corresponding reward
   - $i_t^* = \arg\max_{\varnothing_N} \left( \boldsymbol{b}_{i_t}^T \widehat{\mathbf{Q}}_i + \alpha \sqrt{\boldsymbol{b}_{i_t}^T \mathbf{G}_i^{-1} \boldsymbol{b}_{i_t}} \right)$
   - Obtain $\psi_{i_t^*}$
3. Update $\mathbf{G}_{i_t^*}$ and $\boldsymbol{c}_{i_t^*}$
   - $\mathbf{G}_{i_t^*} \leftarrow \mathbf{G}_{i_t^*} + \boldsymbol{b}_{i_t^*} \boldsymbol{b}_{i_t^*}^T$
   - $\boldsymbol{c}_{i_t^*} \leftarrow \boldsymbol{c}_{i_t^*} + \psi_{i_t^*} \boldsymbol{b}_{i_t^*}$

**End For**

$$i_t^* = \underset{\emptyset_N}{\text{argmax}}\left(\boldsymbol{b}_{i_t}^T \tilde{\mathbf{Q}}_{i_t}\right), \tag{18}$$

After selecting $AP_{i_t^*}$, its available spectral efficiency $\psi_{i_t^*}$ is obtained, and then its parameters are updated for posterior distributions calculations and next round of AP selection as given in Algorithm 4.

The proposed MAB algorithms are online learning approaches, which have a limitation of taking some time before converging to the optimal solution. In highly dynamic environment, where the channel is highly varying, the channel may change during the online learning process causing convergence limitations. However, in the considered WiGig WLAN shown in Fig. 1, the users are usually stationary or walking at very low speed. This relaxes the problem of highly varying channel affecting the convergence of the proposed MAB algorithms. The study of the proposed MAB algorithms under time varying channel conditions is out of scope of the current work and will be left for future investigations.

**Algorithm 4:** CTS Policy for WiGig AP Selection

**Output:** $AP_{i_t^*}$
**Input:** $h \in \mathbb{R}^+$, $\emptyset_N$
**Initialization**: $\mathbf{B}_i \leftarrow \mathbf{I}_d$, $\mathbb{E}(\widehat{\mathbf{Q}}_i) \leftarrow \mathbf{0}_{l\times 1}$ , $\boldsymbol{c}_i \leftarrow \mathbf{0}_{l\times 1}$ for $\forall i \in \emptyset_N$,
**For** $t = 1, 2, 3, \ldots, T_H$
1. Sample $\tilde{\mathbf{Q}}_{i_t}$ from distribution $\mathcal{N}\left(\mathbb{E}(\widehat{\mathbf{Q}}_{i_{t-1}}), h^2\mathbf{B}_{i_{t-1}}^{-1}\right)$ and observe $\boldsymbol{b}_{i_t}$.
2. Select a WiGig AP and obtain its corresponding reward
   - $i_t^* = \arg\max_{\emptyset_N}\left(\boldsymbol{b}_{i_t}^T \tilde{\mathbf{Q}}_{i_t}\right)$
   - Obtain $\psi_{i_t^*}$
3. Update $\mathbf{B}_{i_t^*}$, $\boldsymbol{c}_{i_t^*}$ and $\mathbb{E}(\widehat{\mathbf{Q}}_{i_t^*})$
   - $\boldsymbol{B}_{i_t^*} \leftarrow \boldsymbol{B}_{i_{t-1}^*} + \boldsymbol{b}_{i_t^*}\boldsymbol{b}_{i_t^*}^T$
   - $\boldsymbol{c}_{i_t^*} \leftarrow \boldsymbol{c}_{i_{t-1}^*} + \psi_{i_t^*}\boldsymbol{b}_{R_{i,t}^*}$
   - $\mathbb{E}(\widehat{\boldsymbol{Q}}_{i_t^*}) = \boldsymbol{B}_{i_t^*}^{-1}\boldsymbol{c}_{i_t^*}$

**End For**

## 5 Numerical analysis

In this section, numerical simulations are conducted to bound the performances of the proposed MAB/CMAB based WiGig AP selection algorithms. The simulated indoor environment, as shown in Fig. 1, consists of 16 multi-band WiGig APs equally separated in a large conference hall of size 40x40m$^2$. The multi-band WiGig UE is uniformly dropped within this area. For calculating $T_{BT}$, we used the BT fame structure standardized by IEEE 802.11ad, where $T_{BT}$ can be expressed as (IEEE 802.11ad Standard 2012):

$$T_{BT} = ((\text{BO}\times 8)/\mathcal{R}_{\text{MCS0}}^m)\times N_{beams} + \left(N_{beams} - 1\right)\times T_{\text{SBIFS}} + T_{\text{MBIFS}}, \tag{19}$$

where BO is number of octets used in the sector level sweep phase, $\mathcal{R}_{MCS0}^m$ indicates the MCS0 WiGig rate in bps, $T_{\text{SBIFS}}$ is the time duration of the short beamforming interframe space (SBIFS), $T_{\text{MBIFS}}$ is the time duration for medium beamforming interframe space (MBIFS), and $N_{beams}$ is the total number of trained beams. Other simulation parameters are given in Table 3.

As benchmark schemes, we compared the proposed MAB/CMAB schemes with optimal, conventional, and random AP selection approaches. In the optimal one, an oracle algorithm is assumed, which can select the WiGig AP having the maximum achievable spectral efficiency at once by just using one BT operation, i.e., $M = 1$. In the conventional one, all deployed APs are exhaustively searched using BT, i.e., $M = N$, and the best AP having the maximum RSS is selected. This conventional one mimics multi-connectivity and the DAS based approaches existing in literature. In the random selection policy, a WiGig AP is arbitrary selected based on uniform distribution, i.e., $M = 1$. For performance evaluations, the following performance metrices are used.

**Table 3** Simulation parameters

| Parameter | Value |
|---|---|
| $W$, $L_D$ | 2.16 GHz, 1 Gbit |
| $P_t^W$, $P_t^m$ | 20 dBm, 10 dBm |
| $\sigma_W$, $\beta_W$,$n_W$ | 6 dB, −47.4 dB, 6 |
| $\sigma_m^{\text{LoS}}$, $\beta_m^{\text{LoS}}$, $n_m^{\text{LoS}}$ | 10.3 dB, 54.9 dB, 2.22 |
| $\xi$, $\Omega$ | 1, uniform [0.3–0.6] m |
| $T_D$, $T_H$,$T_w$ | 1 ms, 1000, 3.6 usec |
| $\alpha$, $h$ | 0.2 and 0.1 |
| $\eta_{i_t}$ | Unfirmly distribution in the range (0, 1) |
| $\sigma_0^2$ | $-174+10\log_{10}(W)+7$ |
| BO, $\mathcal{R}_{\text{MCS0}}^m$ | 48, 27.5 Mbps |
| $T_{\text{SBIFS}}$, $T_{\text{MBIFS}}$ | 1 usec, 9 usec |
| AP and UE antenna height | 4 and 0.75 m |

- Average throughput, $\Psi$ in bps, which is defined:

$$\Psi = \frac{1}{T_H} \sum_{t} \frac{WT_D}{T_D + T_{BT}} \left( \psi_{i_t^*} \right), \tag{20}$$

- Average energy efficiency, EE in bps/J, which is defined as:

$$\mathrm{EE} = \frac{\Psi}{P_t^m \left( T_D + MT_{BT} \right)} \tag{21}$$

In the case of the proposed CMAB schemes, the term $P_t^w T_w$ is added to the dominator of (21) to indicate the power consumed by the Wi-Fi signaling, where $T_w$ is the duration of the Wi-Fi control frame used for Wi-Fi RSS measurement as given in Table 3.

- Average total delay, $D$ in sec, which is defined as:

$$D = \frac{L_D}{\Psi}, \tag{22}$$

where $L_D$ is the size of the TX data in bits as given in Table 3.

### 5.1 Adjusting LinUCB and CTS Parameters

In this part of simulations analysis, the values of the design parameters $\alpha$ and $h$ of the LinUCB and CTS algorithms will be adjusted. Figures 2 and 3 represent the average throughput performances in [Gbps] of the proposed LinUCB and CTS algorithms against the design parameters $\alpha$ and $h$ at different values of blockage density $\lambda$ using $\theta_{-3\mathrm{dB}}$ of $20^{\circ}$. $\lambda = 0$ means no LoS blockage at all, and $\lambda = 0.1$ means that

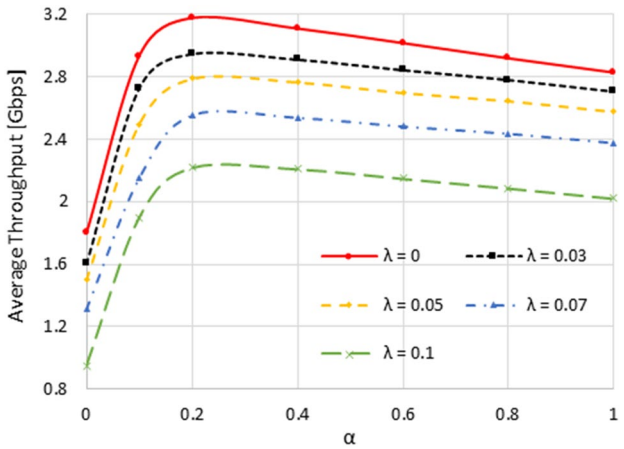



**Fig. 2** Adjusting $\alpha$ value of LinUCB at different values of $\lambda$

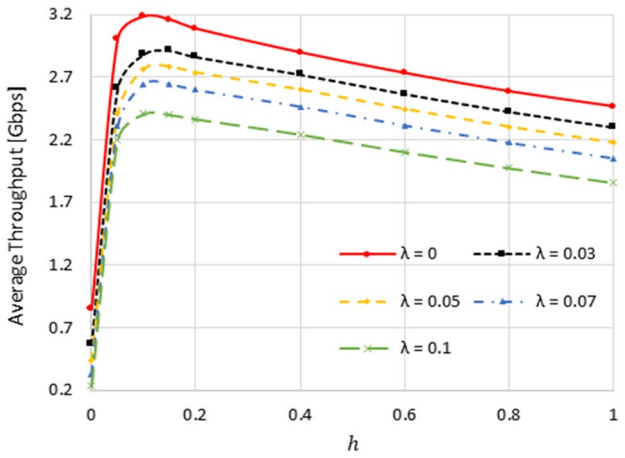



**Fig. 3** Adjusting $h$ value of CTS at different values of $\lambda$

the LoS blocking probability is equal to 70%. Generally, as the value of $\lambda$ is increased, the average throughput performances of both schemes are decreased. This comes from the low available spectral efficiency due to the low mmWave received power influenced by the harsh blockage environment. When $\alpha = 0$ and $h = 0$, only exploitation happens through terms $\boldsymbol{b}_{i_t}^T \widehat{\mathbf{Q}}_i$ in the case of LinUCB and $\mathbb{E}\left( \widehat{\mathbf{Q}}_{i_t} \right)$ in the case of CTS without any exploration, which causes a sharp decrease in the average throughput performances of both schemes as shown in Figs. 2 and 3. As the values of $\alpha$ and $h$ are increased, the average throughput of both schemes are increased due to the increase in the exploration terms $\alpha \sqrt{\boldsymbol{b}_{i_t}^T \mathbf{G}_i^{-1} \boldsymbol{b}_{i_t}}$ and $h^2 \mathbf{B}_{i_t}^{-1}$. However, as we further increase $\alpha$ and $h$, the exploration terms tend to be the most dominant, which results in decreasing the average throughput of both schemes again as shown in Figs. 2 and 3. From these figures, $\alpha = 0.2$ and $h = 0.1$ are selected as sufficient values for the proposed LinUCB and CTS algorithms as given in Table 3.

### 5.2 Performance comparisons

In this part of numerical simulations, the performances of the WiGig AP selection schemes involved in the comparisons are presented. The performances are measured against the used $-3$ dB beamwidth $\theta_{-3\mathrm{dB}}$ and against the blockage density $\lambda$.

#### 5.2.1 *Performance comparisons against the value of* $\theta_{-3dB}$

Figures 4, 5 and 6 represent the average throughput, average energy efficiency and average total delay performances

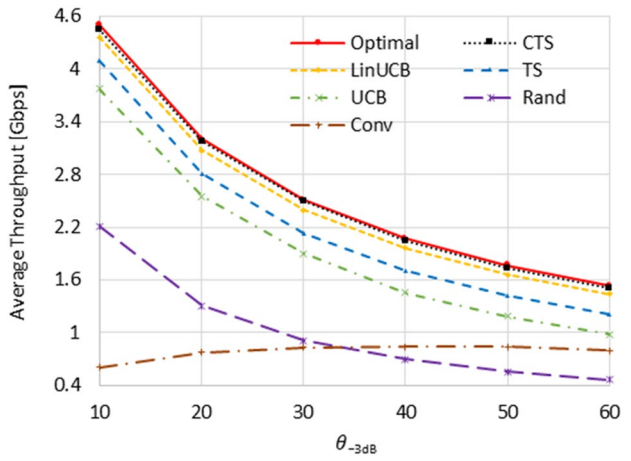



**Fig. 4** Average throughput comparisons against $\theta_{-3\,\mathrm{dB}}$ at $\lambda = 0$

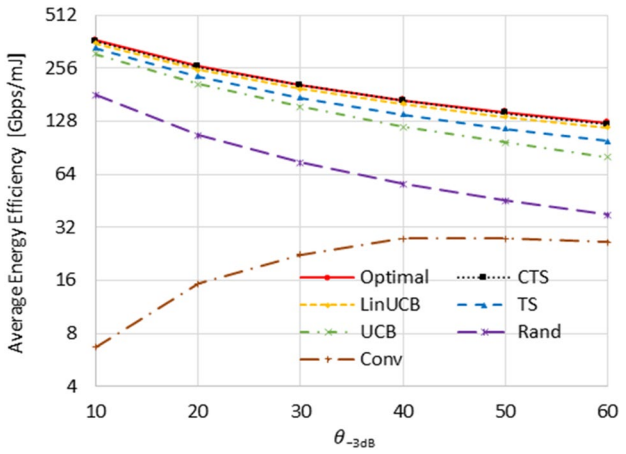



**Fig. 5** Average energy efficiency comparisons against $\theta_{-3\,\mathrm{dB}}$ at $\lambda = 0$

of the schemes involved in the comparisons against $\theta_{-3\mathrm{dB}}$ at $\lambda = 0$. From Fig. 4, as $\theta_{-3\mathrm{dB}}$ is increased, the average throughput performances of all compared schemes are decreased except the conventional one. Generally, using wide beams will decrease $\widehat{\psi}_i$ in (6) and $T_{BT}$ in (19) due to the decrease in $P^m_{r,i}$ and the use of low number of beams, respectively. Two opposite effects on average throughput come from decreasing $\widehat{\psi}_i$ and $T_{BT}$; one results in decreasing the average throughput and the other results in increasing it due to the decrease in the BT overhead $MT_{BT}$, respectively. As $M = 1$ for all compared schemes except the conventional one, the effect of decreasing $\widehat{\psi}_i$ becomes the most dominant and the average throughput of all schemes are decreased when increasing $\theta_{-3\mathrm{dB}}$. However, for the conventional scheme, as $M = N$, the effect of $MT_{BT}$ becomes the most dominant, and the average throughput is increased when increasing the value of $\theta_{-3\,\mathrm{dB}}$ as shown in Fig. 4. This is the reason why, random scheme has better performance than conventional at low values of $\theta_{-3\,\mathrm{dB}}$, and vice versa for high $\theta_{-3\,\mathrm{dB}}$ values. From this figure, CMAB schemes have the best performances especially CTS nearly matches the optimal one. Moreover, thanks to the context information provided by the Wi-Fi RSS statistics, CMAB schemes, i.e., LinUCB and CTS, outperform the context-free ones, i.e., UCB and TS. At $\theta_{-3\mathrm{dB}} = 10^{\circ}$, CTS, LinUCB, TS, UCB, Rand, and Conv schemes obtain about 99%, 97%, 89%, 84%, 49%, and 13.6% of the optimal performance, respectively. These values become 98%, 94%, 78.5%, 64%, 30% and 52% when $\theta_{-3\mathrm{dB}} = 60^{\circ}$.

Like the average throughput performances, the average energy efficiencies in [Gbps/mJ] of all compared schemes are decreasing when increasing $\theta_{-3\mathrm{dB}}$ except the conventional one as shown in Fig. 5. Due to the high BT overhead of the conventional scheme, it consumes a high power getting its energy efficiency even lower than the random scheme at all

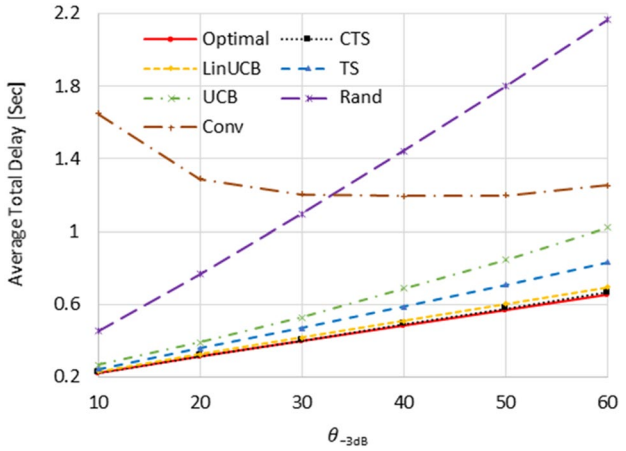



**Fig. 6** Average total delay comparisons against $\theta_{-3\,\mathrm{dB}}$ at $\lambda = 0$

tested $\theta_{-3\mathrm{dB}}$ values. Still, the CMAB has the best average energy efficiency performance, and it outperforms the non-contextual MAB. At $\theta_{-3\mathrm{dB}} = 10^{\circ}$, CTS, LinUCB, TS, UCB, Rand, and Conv schemes obtain about 98.7%, 96.8%, 89%, 84%, 49%, and 1.8% of the optimal performance, respectively. These values become 97.7%, 93.8%, 78.5%, 64%, 30% and 21% when $\theta_{-3\mathrm{dB}} = 60^{\circ}$. As shown by these results, the Wi-Fi power consumption has a negligible effect on the average energy efficiency of the CMAB schemes.

As given by Fig. 6, the average total delay performances of all compared schemes are increasing when increasing $\theta_{-3\mathrm{dB}}$ except the conventional one. Yet, at $\theta_{-3\mathrm{dB}} = 60^{\circ}$, the average total delay of the conventional scheme tends to slightly increase again influenced by the minor decrease in average throughput given by Fig. 4. This is due to the dominancy of the low spectral efficiency of this too wide beam. Again, CMAB schemes nearly match the optimal performance, and they are better than the non-contextual

MAB schemes. Also, at low $\theta_{-3dB}$ values, the average total delay performance of the random scheme is better than the conventional one, and vice versa at high values of $\theta_{-3dB}$. This comes from their average throughput performances as previously explained. At $\theta_{-3dB} = 10^{\circ}$, CTS, LinUCB, TS, UCB, Rand, and Conv have 1.01, 1.03, 1.1, 1.2, 2, and 7.4 times increase in average total delay over the optimal performance, respectively. These values become 1.02, 1.07, 1.3, 1.6, 3.3 and 1.92 when $\theta_{-3dB} = 60^{\circ}$.

#### 5.2.2 *Performance comparisons against the value of λ*

Figures 7, 8 and 9 give the average throughput, average energy efficiency and average total delay performances of the schemes involved in the comparisons against $\lambda$ at $\theta_{-3dB} = 20^{\circ}$.

From Fig. 7, as the value of $\lambda$ is increased, i.e., harsh blockage environments, the average throughput performances of all schemes involved in comparisons are decreased influenced by the low received mmWave power affected by the high probability of LoS blockage. It is interesting to notice that although the average throughput of the random scheme is better than that of the conventional scheme at low values of $\lambda$, the opposite happens at high values of $\lambda$. This is because, at low values of $\lambda$, i.e., low blockage probabilities, the random selection scheme will select a non-blocked AP with a high probability, while the opposite happens at high values of $\lambda$. Still, the proposed CMAB schemes have the best performances at all tested $\lambda$ values, and the CTS nearly matches the optimal performance. Again, the CMAB algorithms, i.e., CTS and LinUCB, show better performance than their context-free MAB counterparts, i.e., TS and UCB, thanks to the utilization of the Wi-Fi RSS context information. At $\lambda = 0$, CTS, LinUCB, TS, UCB, Rand, and Conv schemes obtain about 99%, 96%, 87%, 80%, 41%, and 24% of the optimal performance, respectively. These values become 98%, 95%, 93%, 84%, 17% and 24% when $\lambda = 0.1$.

Like the average throughput and due to the same reasoning, the average energy efficiency in Gbps/mJ of all compared schemes are decreasing with increasing the value of $\lambda$ as shown in Fig. 8. Also, the average energy efficiency of the conventional scheme is lower than that of the random scheme at all tested $\lambda$ values due to its high-power consumption resulting from its high BT overhead. Still the CMAB schemes have the superior performance over other schemes involved in comparisons. Also, the CMAB schemes show better average energy efficiency than the non-contextual MAB ones. From Fig. 8, at $\lambda = 0$, CTS, LinUCB, TS, UCB, Rand, and Conv schemes obtain about 99%, 96%, 87%, 80%, 41%, and 6% of the optimal performance, respectively. These values become 98%, 95%, 93%, 84%, 17% and 6% when $\lambda = 0.1$.

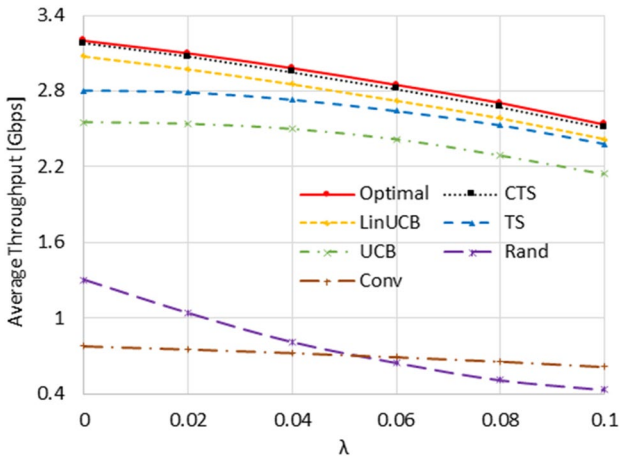



**Fig. 7** Average throughput comparisons against $\lambda$ at $\theta_{-3\ dB}$ of 20°

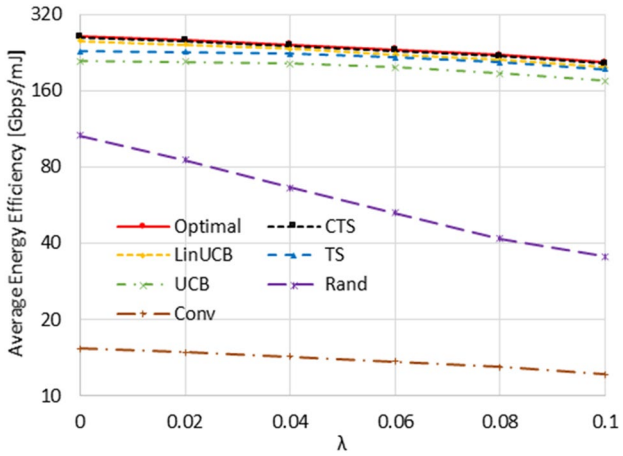



**Fig. 8** Energy efficiency comparisons against $\lambda$ at $\theta_{-3\ dB}$ of 20°

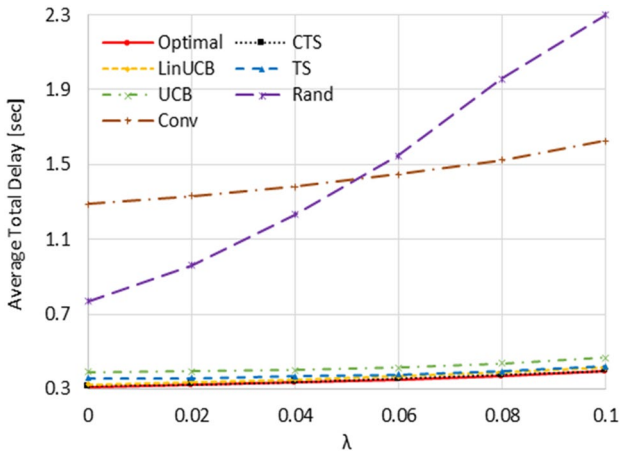



**Fig. 9** Average total delay comparisons against $\lambda$ at $\theta_{-3\ dB}$ of 20°

The average total delay of all compared schemes is increasing when increasing the value of $\lambda$ as shown in Fig. 9. Influenced by their average throughput performances, the average total delay of the random selection is better than that of the conventional scheme at low values of $\lambda$, and the opposite happens at high values of $\lambda$. Also, CMAB schemes show the best average total delay performance, which is better than the context-free MAB ones, where the CTS nearly matches the optimal performance. At $\lambda = 0$, CTS, LinUCB, Rand, and Conv have 1.007, 1.04, 1.14, 1.25, 2.46, and 4.12 times increase in average total delay over the optimal performance, respectively. These values become 1.01, 1.05, 1.07, 1.18, 5.38 and 4.12 when $\lambda = 0.1$.

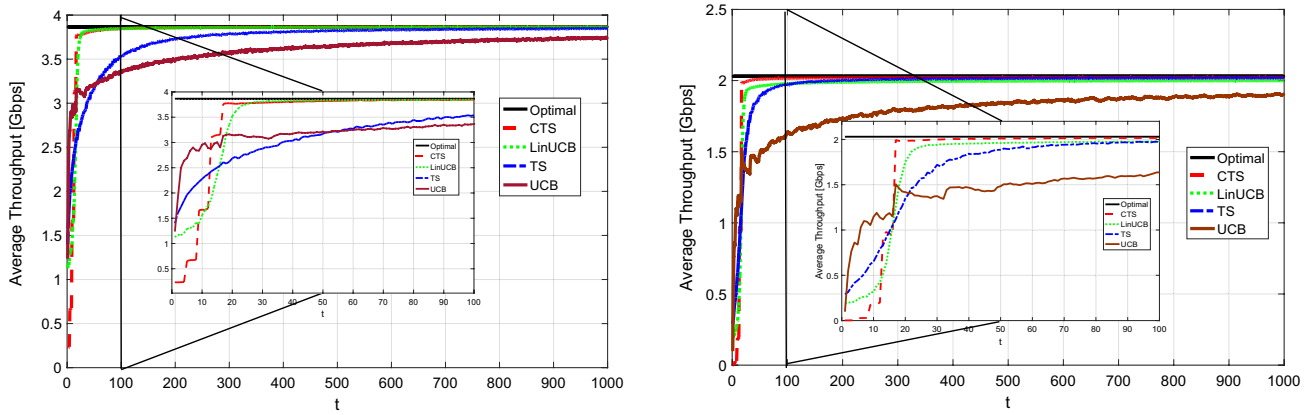
### 5.3 Convergence analysis

The convergence of the MAB algorithms is of the main concern as the best MAB scheme should reach the optimal value within few number of trails. Thus, in this part of numerical simulations, we will study the convergence of the proposed CMAB and MAB algorithms towards the optimal value against the time horizon. Figure 10 shows the average throughput convergence of the proposed CTS, LinUCB, TS and UCB algorithms against the number of trails $t$ under different simulation conditions. The main strength of the proposed MAB/CMAB schemes is that they reach the optimal performance at high value of $t$. However, at low value of $t$, the proposed MAB/CMAB algorithms suffer from bad performances, which is considered as weakness of the proposed algorithms. At low values of $t$, non-contextual MAB schemes show better convergence rates than their CMAB counterparts as shown in Fig. 10. This is because, at low values of $t$, CMAB schemes start to learn the relationship between Wi-Fi contexts of the deployed APs and their achievable rewards. After CMAB schemes learn this relationship effectively, they converge faster than the context-free ones. Also, faster convergence happens when $\lambda = 0.1$ and $\theta_{-3\text{ dB}} = 60^{\circ}$. This is because at these conditions, low power is received by the WiGig UE reducing the number of candidate APs, which results in fast reaching the optimal performance. At both simulation conditions, CTS has the fastest convergence while UCB has the slowest one. For example, at $\lambda = 0$ and $\theta_{-3\text{ dB}} = 60^{\circ}$, CTS, LinUCB and TS reach the optimal value after $t = 20$, 100 and 200, respectively. However, the UCB never reach the optimal value even after $t = 1000$.

Although this paper introduced the applications of online learning especially MAB and CMAB hypothesis in WiGig AP selection problem, other recent optimization techniques such as Chimp optimization given in Khishea et al. (2020) can be investigated, which is left as future research directions.

## 6 Conclusion

In this paper, the problem of WiGig AP selection was explored and formulated as a MAB game. In this formulation, the WiGig UE acts as the player, the candidate APs act as the arms of the bandit, and the available spectral efficiencies are the rewards. This MAB formulation enables the selection of the optimal WiGig AP over the time horizon while keeping the BT overhead at the minimum level. Based on this new formulation, non-contextual MAB schemes, i.e., TS and UCB, were adopted to address the WiGig AP selection problem. As WiGig devices are multi-band capable containing Wi-Fi and mmWave bands besides the direct relation between Wi-Fi and mmWave link




**(a)** At $\lambda = 0$ and $\theta_{-3\text{dB}} = 10^{\circ}$. **(b)** At $\lambda = 0.1$ and $\theta_{-3\text{dB}} = 60^{\circ}$.

**Fig. 10** Average throughput convergence comparisons at different values of $\lambda$ and $\theta_{-3\text{ dB}}$

statistics, Wi-Fi signal information was used as contexts of the WiGig APs. Thus, the problem was also considered as a CMAB game enabling the utilization of CTS and LinUCB CMAB algorithms to address it. Extensive numerical simulations proved the effectiveness of this MAB formulation through adopting CTS, LinUCB, TS and UCB algorithms over the conventional maximum RSS based and random AP selection schemes. Moreover, CMAB schemes show better performance and faster convergence rates than their context-free counterparts.

As extension to the current work, multi-user- multi-APs association will be considered. In this case multi-player MAB game will be proposed while considering load balance and interference management among the installed APs. Also, other online learning approaches such as Q-learning will be investigated and compared with the proposed MAB algorithms. The application of deep reinforcement learning in both single and multi-user association scenarios is another promising future research direction to the existing work. Finally, the adaptation of the proposed MAB algorithms for dynamic channel conditions needs more investigations.

## References


Abdelreheem A, Mohamed EM, Esmaiel H (2018) Location-based millimeter wave multi-level beamforming using compressive sensing. IEEE Commun Lett 22(1):185–188

Agrawal S, Goyal N (2013) Thompson sampling for contextual bandits with linear payoffs. In: the 30th International Conference on Machine Learning, ICML: 1–9

Alkhateeb A, El Ayach O, Leus G, Heath RW Jr (2014) Channel estimation and hybrid precoding for millimeter wave cellular systems. IEEE J Sel Top Signal Process 8(5):831–846

Audibert JY, Munos R, Szepesvari C (2009) Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. Theor Comput Sci 410(19):1876–1902

Auer P, Cesa-Bianchi N, Fischer P (2002) Finite-time analysis of the multiarmed bandit problem. Mach Learn 47(2):235–256

Chen X, Yuan W, Cheng W, Liu W, Leung H (2013) Access point selection under QoS requirements in variable channel-width WLANs. IEEE Wireless Commun Lett 2(1):114–117

Dwijaksara MH, Jeon WS, Jeong DG (2019) User association for load balancing and energy saving in enterprise WLANs. IEEE Syst J 13(3):2700–2711

Francisco VI, Marcial-Romero J, Valdovinos R (2019) A comparison between UCB and UCB-tuned as selection policies GGP. IFS 36(5):5073–5079

Gao X, Zhang J, Liu G et al (2007) Large-scale characteristics of 5.25 GHz based on wideband MIMO channel measurements. IEEE Antenn Wirel Propag Lett 6:263–266

Ghasempour Y, da Silva CR, Cordeiro C, Knightly EW (2017) IEEE 802.11 ay: next-generation 60 GHz communication for 100 Gb/s Wi-Fi. IEEE Commun Magaz 55(12):186–192

Gutowski N, Amghar T, Camp O, Chhel F (2018) Context enhancement for linear contextual multi-armed bandits. In: 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)

Hashima S, ElHalawany BM, Hatano K, Wu K, Mohamed EM (2021) Leveraging machine-learning for D2D communications in 5G/beyond 5G networks. Electronics 10(2):1–16

Hosoya K, Prasad N, Ramachandran K et al (2015) Multiple sector ID capture (MIDC): a novel beamforming technique for 60 GHz band multi-Gbps WLAN/PAN systems. IEEE Trans Antennas Propag 63(1):81–96

IEEE 802.11ad Standard (2012) Enhancements for very high throughput in the 60 GHz band

Kaufmann E, Korda N, Munos R (2012) Thompson sampling: An asymptotically optimal finite-time analysis. In: International Conference on Algorithmic Learning Theory: 199–213

Khishea M, Mosavib MR (2020) Chimp optimization algorithm. Expert Syst Appl 149(11338):1–26

Kim H, Lee W, Bae M, Kim H (2017) Wi-Fi seeker: a link and load aware AP selection algorithm. IEEE Trans Mob Comput 16(8):2366–2378

Lihong L, Wei C, Langford J, Schapire R (2010) A contextual-bandit approach to personalized news article recommendation. In: the19th international conference on World wide web: 661–670

Liu R, Yu G (2021) User association for millimeter-wave ultra-reliable low-latency communications. IEEE Wireless Commun Lett 10(2):315–319

Liu D, Wang L, Chen Y, Elkashlan M, Wong K, Schober R, Hanzo L (2016) User association in 5G networks: a survey and an outlook. IEEE Commun Surveys Tuts 18(2):1018–1044

Liu R, Lee M, Yu G, Li GY (2020) User association for millimeter-wave networks: a machine learning approach. IEEE Trans Commun 68(7):4162–4174

Mohamed EM, Sakaguchi K, Sampei S (2017) Wi-Fi coordinated WiGig concurrent transmissions in random access scenarios. IEEE Trans Veh Technol 66(11):10357–10371

Mohamed EM, Abdelghany MA, Zareei M (2019) An efficient paradigm for multiband WiGig D2D networks. IEEE Access 7:70032–70045

Mohamed EM, ElHalawany BM, Khallaf HS, Zareei M, Zeb A, Abdelghany MA (2020a) Relay probing for millimeter wave multi-hop D2D networks. IEEE ACCESS 8:30560–30574

Mohamed EM, Hashima S, Aldosary A, Hatano K, Abdelghany MA (2020b) Gateway selection in millimeter wave UAV wireless networks using multi-player multi-armed bandit. Sensors 20(14):1–22

Ožkoç MF, Koutsaftis A, Kumar R, Liu P, Panwar SS (2021) The impact of multi-connectivity and handover constraints on millimeter wave and terahertz cellular networks. IEEE J Select Areas Commun 39(6):1833–1853

Peng M, He G, Wang L, Kai C (2019) AP selection scheme based on achievable throughputs in SDN-enabled WLANs. IEEE Access 7:4763–4772

Rappaport TS, Sun S, Mayzus R, Zhao H et al (2013a) Millimeter wave mobile communications for 5G cellular: it will work! IEEE Access 1:335–349

Rappaport TS, Gutierrez F, Ben-Dor E et al (2013b) Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications. IEEE Trans Antenn Propag 61(4):850–1859

Sakaguchi K, Mohamed EM, Kusano H et al (2015) Millimeter-wave wireless LAN and its extension toward 5G heterogeneous networks. IEICE Trans Commun E98.B(10):1932–1948

Seldin Y, Szepesvari C, Auer P, Yadkori Y A (2012) Evaluation and analysis of the performance of the EXP3 algorithm in stochastic environments. In: the 10th European Workshop on Reinforcement Learning: 103–116

Singh S, Kulkarni MN, Ghosh A, Andrews JG (2015) Tractable model for rate in self-backhauled millimeter wave cellular networks. IEEE J Selec Areas Commun 33(10):2196–2211

Walsh TJ, Szita I, Diuk C, Littman ML (2009) Exploring compact reinforcement-learning representations with linear regression. In: the 25th Conference on Uncertainty in Artificial Intelligence: 1–11

Wu S, Atat R, Mastronarde N, Liu L (2018) Improving the coverage and spectral efficiency of millimeter-wave cellular networks using device-to-device relays. IEEE Transc Commun 66(5):2251–2265

Yoo S K, Zhang L, Cotton S L, Quoc Ngo H (2019) Distributed antenna systems used for indoor UE to access point communications at 60 GHz. In: the 13th European Conference on Antennas and Propagation (EuCAP): 1–5

Zhang X, Zhou S, Wang X et al (2012) Improving network throughput in 60GHz WLANs via multi-AP diversity. In: 2012 IEEE International Conference on Communications (ICC): 4803–4807

Zhang L, Cotton S L, Yoo S, Fernández M, Scanlon W G (2021) Access point selection strategies for indoor 5G millimeter-wave distributed antenna systems. In: the 5th European Conference on Antennas and Propagation (EuCAP): 1–5