Article

Distribution of Multi MmWave UAV Mounted RIS Using Budget Constraint Multi-Player MAB

Ehab Mahmoud Mohamed [1,2,*], Mohammad Alnakhli [1], Sherief Hashima [3,4] and Mohamed Abdel-Nasser [2]

1 Department of Electrical Engineering, College of Engineering in Wadi Addawasir, Prince Sattam Bin Abdulaziz University, Wadi Addawasir 11991, Saudi Arabia
2 Department of Electrical Engineering, Aswan University, Aswan 81542, Egypt
3 Computational Learning Theory Team, RIKEN-Advanced Intelligence Project (AIP), Fukuoka 819-0395, Japan
4 Engineering Department, Nuclear Research Center, Egyptian Atomic Energy Authority, Inshas, Cairo 13759, Egypt
* Correspondence: ehab_mahmoud@aswu.edu.eg

Abstract: Millimeter wave (mmWave), reconfigurable intelligent surface (RIS), and unmanned aerial vehicles (UAVs) are considered vital technologies of future six-generation (6G) communication networks. In this paper, various UAV mounted RIS are distributed to support mmWave coverage over several hotspots where numerous users exist in harsh blockage environment. UAVs should be spread among the hotspots to maximize their average achievable data rates while minimizing their hovering and flying energy consumptions. To efficiently address this non-polynomial time (NP) problem, it will be formulated as a centralized budget constraint multi-player multi-armed bandit (BCMP-MAB) game. In this formulation, UAVs will act as the players, the hotspots as the arms, and the achievable sum rates of the hotspots as the profit of the MAB game. This formulated MAB problem is different from the traditional one due to the added constraints of the limited budget of UAVs batteries as well as collision avoidance among UAVs, i.e., a hotspot should be covered by only one UAV at a time. Numerical analysis of different scenarios confirm the superior performance of the proposed BCMP-MAB algorithm over other benchmark schemes in terms of average sum rate and energy efficiency with comparable computational complexity and convergence rate.

Keywords: UAV mounted RIS; MP-MAB; Millimeter wave; hotspot

1. Introduction

Millimeter wave (mmWave), i.e., $30 \sim 300$ GHz band, constitutes the corner millstone of the current fifth generation (5G) and the upcoming six generation (6G) networks [1]. This comes from its sizeable available spectrum. However, its high operating frequency causes mmWave signal to be weak and subject to bad channel conditions [2]. This makes it prone to path blockage and human shadowing. Nevertheless, Oxygen absorption highly degrades the quality of the mmWave link [3]. Therefore, antenna beamforming is recommended as an effective solution for overwhelming mmWave channel impairments. This can be conducted using beamforming training (BT) by means of steering antenna elements utilizing structured codebooks [4].

In dense hotspot scenarios containing numerous numbers of mmWave users, mmWave coverage should be extended and strengthened to fully cover hotspot users and overcome their mutual path blockage. In this regard, reconfigurable intelligent surface (RIS) [5] can provide an effective solution. It is a talented 6G approach that can smartly reconfigure the wireless communication channel [5]. This means that an RIS board can effectively control the mmWave channel by reinforcing the received signal in some directions and weakening it in other directions [6]. Thus, it can provide additional non-line of sight (NLoS) paths to mmWave users inside their hotspots. This can be conducted by passively controlling the incident Electromagnetic wave (EM) on the RIS board by adjusting the phase shifts (PSs) of

its antenna array [6]. By this way, the complicated RF chains in the conventional relaying systems are highly relaxed [7]. Due to its cheapness and ease of installation, researchers investigated the application of RIS in numerous wireless communications systems [8–12]. These applications extend and strengthen the mmWave coverage, as presented in this paper [13,14]. However, in the case of numerous hotspots scenario, it will be challenging to install RIS boards nearby each hotspot, especially in the case of momentarily hotspots such as stadiums, theaters, markets, etc. In these scenarios, unmanned aerial vehicles (UAVs) will provide a practical and cost-effective solution, where the RIS boards will be attached to the UAVs. Then, these multi UAV mounted RIS will be distributed in the region of hotspots, and each UAV will serve a particular hotspot area. Recently, UAVs received significant attention in wireless communication due to their flying and maneuvering capabilities. For example, UAVs can be used as airborne base stations (BSs) to provide wireless connectivity in remote and post-disaster/stricken areas [15]. In addition, they can be used as relays to extend the coverage of mobile base stations (BSs) [16]. Moreover, data collection such as aerial photography, traffic, and environmental monitoring can be conducted quickly by UAVs [17,18].

Herein, the distribution of the UAVs mounted RIS among the hotspots becomes challenging due to their limited battery capacity. Therefore, each UAV should cover a hotspot, maximizing its achievable data rate while minimizing its flying and hovering energy consumptions. This problem is a non-polynomial (NP) time problem as its complexity increases in an NP behavior by increasing the number of UAVs and hotspots. In addition, the constraint of limited UAV battery capacity should be maintained while solving the problem. Furthermore, collision among UAVs should be avoided, i.e., no more than one UAV is permitted to cover a particular hotspot at a time. These constraints further complicate the optimization problem.

In this paper, online learning is used to efficiently address the problem of multi UAV mounted RIS distribution by considering it as a centralized budget constraint multiplayer multi-armed bandit (BCMP-MAB) game. As a robust online learning tool, MAB can efficiently handle the fundamental exploitation-exploration learning trade off. In this context, a MAB player challenges the maximization of his profit via consistently exploiting the highest reward arm or exploring the less selected ones [19,20]. This should be conducted while the player only observes the achievable rewards of the played arms [19]. Thus, the main contributions of this paper can be summarized as follows:

- UAVs mounted RIS are used to extend and strengthen the coverage of mmWave in highly dense hotspot areas containing considerable numbers of users. The distribution of the UAVs among the hotspots is formulated as an optimization problem to maximize the sum data rates of the hotspots while minimizing the flying and hovering energy consumptions of the UAVs.
- The aforementioned optimization problem is reformulated as a centralized budget constraint MP-MAB game, where the players are the UAVs, the arms of the bandit are the hotspots, and the rewards are the achievable hotspots' data rates. The proposed BCMP-MAB differs from the conventional MP-MAB game due to the added battery budget and UAVs collision-free constraints. The centralized nature of the proposed BCMP-MAB is used to avoid collisions among UAVs, and the budget constraint is used to take into account the limited battery capacity of UAVs when selecting the best hotspots at a time. To avoid such collisions, the UAV-hotspot selection process is made autonomously and sequentially by UAVs during the bandit game through centralized orchestration and information about the currently uncovered hotspots provided by the mmWave BS.
- The proposed BCMP-MAB algorithm shows greater performance than other benchmark schemes via extensive numerical simulations under different scenarios.

The remainder of this paper is organized as follows: the literature review is given in Section 2. The proposed system model is presented in Section 3. The proposed BCMP-

MAB algorithm is introduced in Section 4, and Section 5 gives the concluding remarks of this paper.

2. Related Works

Recently, few research works investigated the applications of RIS in mmWave communications. In [6,14], the coauthors of this paper proposed two-stage MAB schemes to find the optimal mmWave link from BS to RIS and from RIS to user equipment (UE), which maximizes the achievable data rate at the UE. Static and adaptively mixed relay RIS topologies were studied in [11], and they outperformed traditional benchmarking RIS architectures. Amplify and forward (AF) relay employing RIS-aided mmWave was investigated in [13], and a precise expression of signal-to-noise power ratio (SNR) was given. Moreover, the authors developed the AF relay's optimal power allocation approach to acquire the ideal PSs for optimizing the end-to-end SNR. In [21], a dual methodology is proposed for active precoders at the mmWave BS and passive precoders at the RIS to maximize the achievable spectrum efficiency of the RIS-aided mmWave system. In [22], a mathematical framework was proposed to analyze the coverage of RIS enabled mmWave system. Moreover, Federated learning (FL) was used in [23] to optimize the performance of RIS-assisted mmWave. In [24,25], channel estimation of RIS-enabled mmWave was considered, whereas in [24], the cascaded nature of the mmWave RIS channel was utilized, while in [25], atomic norm minimization was adopted. In [26], hybrid precoding approach was proposed to adjust the analog/digital precoders of the mmWave BS as well as the PSs of the RIS board. A deep learning-empowered compressive sensing approach was proposed in [27] to adjust the precoders and the PSs of both mmWave BS and RIS. The authors of [28] investigated machine learning (ML) based beam management of RIS aided mmWave communication. In [29], passive precoding, power allocation as well as user association of RIS-aided mmWave are jointly optimized using sequential fractional programming (SFP) and forward-reverse auction (FRA) techniques. In [30], the best proper beams and reflection coefficients of RIS-assisted mmWave were investigated. Despite the existing literature on RIS aided mmWave communications, few investigated the RIS-enabled UAV for mmWave communications. In [31], the coauthors of this work studied the trajectory planning of UAV-mounted RIS over multiple hotspot areas via single-player MABs to maximize its achievable data rate while minimizing its energy consumption over its path from one hotspot to another. However, this work considered only one UAV setting, plus the problem of optimal UAVs distribution was not inspected. In [32], the authors showed the superior advantage of UAV-mounted RIS over that based on fixed RIS in enhancing the coverage of mmWave users. In addition, they used deep reinforcement learning (DRL) to model the environment and optimize the performance of mmWave UAV-mounted RIS system. In [33], the authors extended the work to jointly optimize the precoding matrix at the BS, the PSs at the RIS, and the location of the UAV-mounted RIS to maximize the total sum rate. However, in these two papers, only one UAV scenario was studied, and the optimal distribution of UAVs was not deemed. In [34], fixed RIS attached to a building is used to enhance the secrecy rate of the mmWave UAV communication. However, in this paper, no UAV-mounted RIS was proposed, and only fixed RIS was used to assist the UAV flying BS. In [35], fixed multiple RIS boards were used to aid UAV-enabled mmWave cellular communications. In this regard, the RIS deployment, user scheduling, beamforming vectors, and RIS phases were jointly optimized to maximize the system's sum rate. In [36], fixed RIS board was used as an auxiliary to enhance the performance of UAV-enabled mmWave communications. In this regard, the power-delivering capability as well as the fading characteristics of RIS were studied. Again, no UAV-mounted RIS was implemented in [35,36].

Table 1 summarizes the research work conducted in RIS assisted mmWave UAV communications. Thus, to the best of our knowledge, no current research work considered the optimal distribution of multi UAV mounted RIS over hotspot areas such as the work presented in this paper.

Table 1. Literature review comparison in RIS assisted mmWave UAV communications.

Reference	Objective	Single/Multi-UAV	Fixed/Mounted
Mohamed, E. M. et al. 2022 [31]	Optimizing the trajectory of UAV mounted RIS	Single	Mounted
Zhang, Q. et al. 2019 [32]	Optimizing the performance of UAV mounted RIS	Single	Mounted
Zhang, Q. et al. 2019 [33]	Optimize the precoding matrix at the BS, the PSs at the RIS	Single	Mounted
Guo, X. et al. 2019 [34]	Enhance the secrecy rate of the mmWave UAV communication.	Single	Fixed
Jiang, L. et al. 2019 [35]	Multiple RIS boards were used to aid UAV-enabled mmWave cellular communications	Single	Fixed
Xiong, B. et al. 2019 [36]	An RIS board was used as an auxiliary to enhance the performance of UAV-enabled mmWave communications	Single	Fixed

3. System Model and Optimization Problem Formulation

This section will detail the proposed system model, the utilized channel models, and the optimization problem formulation of multi UAV distribution among hotspots.

3.1. Proposed System Model

Figure 1 shows the proposed system model of mmWave UAV-mounted RIS for hotspot area coverage. In this model, multiple RIS boards attached to UAVs are used to strengthen the coverage of mmWave BS at hotspots containing different numbers of UEs, such as stadiums, markets, etc. Every hotspot has a varied traffic demand based on the traffic needs of its associated users. The UAV-mounted RIS will provide an additional mmWave path from the mmWave BS to the mmWave users inside the hotspot, as shown by the dashed red lines in Figure 1, where the green lines indicates the direct path from the mmWave BS to the stadium hotspot area. Typically, the number of hotspots is higher than the number of UAVs. Thus, each hotspot should be served by only one UAV at a time. Based on the information of the uncovered hotspots provided by the mmWave BS, a free UAV should autonomously decide which hotspot from the uncovered ones it should fly towards and cover. Herein, we do not consider a fully centralized network, where the mmWave BS fully controls the UAV-hotspot selection, in order to prevent the high backhauling overhead, especially when using high number of UAVs. However, a fully centralized network will be the subject of our future investigations. The UAV-hotspot selection should maximize the achievable sum rate of the hotspots based on the specification of the attached RIS board while minimizing the flying and hovering energy consumptions of the UAV. After selecting a specific hotspot, the UAV informs the mmWave BS of its selection, and the mmWave BS controls the PSs of its attached RIS towards the chosen hotspot location, then considers this hotspot as covered. In this paper, we will focus on the optimal distribution of UAVs mounted RIS over the hotspots, while issues related to mmWave RIS channel estimation, joint BS active beamforming and RIS passive beamforming adjustment, and the effect of UAV turbulence on mmWave RIS channels are out of scope of this paper. These issues are already addressed in some of the research works as given in [24,26]. In the following two sub-sections, we will give the used mmWave channel models and the optimization problem formulation of the UAV-hotspot distributions, respectively.
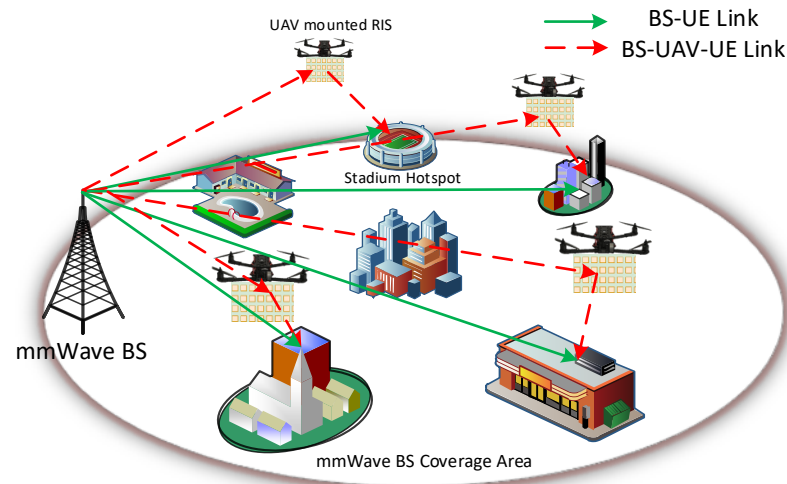
Figure 1. Proposed system model of multi mmWave UAV mounted RIS hotspot area coverage.

3.2. MmWave Channel Models

The received (RX) power P_{r,nk_m} at UE k in hotspot, m consists of two components. One component directly comes from the LoS path from the mmWave BS (B), and the other comes from the NLoS provided by UAV n. This can be represented mathematically through (1) as follows:

$$P_{r,nk_m} = P_{r,Bk_m} + P_{r,Bnk_m} \tag{1}$$

where P_{r,Bk_m} indicates the direct LoS power component received by UE k_m from BS, and P_{r,Bnk_m} indicates the NLoS one traced through UAV n. For P_{r,Bk_m}, the mmWave terrestrial link model given in [37] is utilized, where P_{r,Bk_m} can be expressed as follows:

$$P_{r,Bk_m} = P_t A_{t,Bk_m}(\theta_{t,Bk_m}, \theta_{-3dB}) A_{r,k_mB}(\phi_{r,k_mB}, \phi_{-3dB}) \left(\frac{\eta(\mathbb{P}_{LoS}(d_{Bk_m}))}{L_{LoS}(d_{Bk_m})} + \frac{\chi(\mathbb{P}_{NLoS}(d_{Bk_m}))}{L_{NLoS}(d_{Bk_m})} \right) \tag{2}$$

In (2), P_t is the transmit (TX) power of the mmWave BS. $\eta(\mathbb{P}_{LoS}(d_{Bk_m}))$ and $\chi(\mathbb{P}_{NLoS}(d_{Bk_m}))$ are two Bernoulli random variables with probabilities $\mathbb{P}_{LoS}(d_{Bk_m})$ and $\mathbb{P}_{NLoS}(d_{Bk_m}) = 1 - \mathbb{P}_{LoS}(d_{Bk_m})$ indicating the LoS and NLoS probabilities as functions of the separation distance d_{Bk_m} between mmWave BS and UE k_m as shown in Figure 2, respectively. $A_{t,Bk_m}(\theta_{t,Bk_m}, \theta_{-3dB})$ and $A_{r,k_mB}(\phi_{r,k_mB}, \phi_{-3dB})$ indicate the TX and RX beamforming gains of mmWave BS and UE k_m, respectively. Herein, θ_{t,Bk_m} and ϕ_{r,k_mB} indicate the boresight angles of the TX and RX beams, while θ_{-3dB} and ϕ_{-3dB} are their $-3dB$ beamwidths. By utilizing the 2D steerable antenna model with Gaussian's main loop profile given in [37], $A_{t,Bk_m}(\theta_{t,Bk_m}, \theta_{-3dB})$ can be expressed as follows:

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

where A_0 is the maximum antenna gain. For $A_{r,k_mB}(\phi_{r,k_mB}, \phi_{-3dB})$, the same equation given in (3) can be used except that θ_{t,Bk_m} and θ_{-3dB} are replaced by ϕ_{r,k_mB} and ϕ_{-3dB}, respectively.
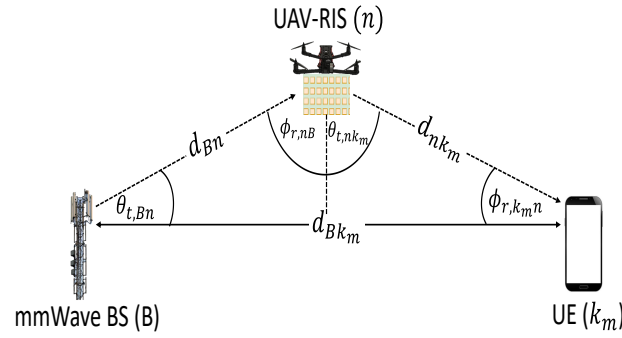
Figure 2. Schematic diagram of the mmWave BS, UAV mounted RIS, UE communication links.

In (2), $L_{LoS}(d_{Bk_m})$ and $L_{NLoS}(d_{Bk_m})$ are the path losses of the LoS and NLoS paths as functions of the separation distance d_{Bk_m}. They can be expressed as follows:

$$10\log_{10}\left(L_v(d_{Bk_m})\right) = \beta_v + 10\alpha_v\log_{10}\left(d_{Bk_m}\right) + \varepsilon_v, \tag{4}$$

where $v \in \{LOS, NLOS\}$, $\beta_v = 82.02 - 10\alpha_v\log_{10}(d_0)$ is the path loss at a reference distance d_0. α_v is path loss exponent, and $\varepsilon_v \frown \mathcal{N}(0, \delta_v)$ is the log-normal shadowing with zero mean and standard deviation of δ_v. Readers are advised to refer to [37] for the details behind these equations as well as their associated parameters.

The authors in [38] investigated the mmWave RX power received at UE from mmWave BS through far-field RIS board, like the case of UAV-mounted RIS deemed in this paper. They considered that all antenna elements of the RIS board will experience the same gain towards its center due to the far-field effect. Thus, P_{r,Bnk_m} can be expressed as:

$$P_{r,Bnk_m} = P_t\left(\left(\frac{\lambda}{4\pi}\right)^4 (Q_n\Gamma)^2 A_{t,Bn}(\theta_{t,Bn}, \theta_{-3dB}) G_{r,nB}(\phi_{r,nB}) G_{t,nk_m}(\theta_{t,nk_m}) A_{r,k_mn}(\phi_{r,k_mn}, \phi_{-3dB})(d_{Bn}d_{nk_m})^{-\alpha}\right) \tag{5}$$

where P_t and λ are the TX power and the wavelength of the mmWave BS signal. Q_n indicates the number of antenna elements of the RIS board attached to UAV n, and Γ is the amplitude reflection coefficient of the RIS elements. α is the path loss exponent, and d_{Bn} and d_{nk_m} are the separation distances between BS and UAV n, and between UAV n and UE k_m, as shown in Figure 2, where the schematic diagram of the UAV-mounted RIS communications links is presented. $A_{t,Bn}(\theta_{t,Bn}, \theta_{-3dB})$ and $A_{r,k_mn}(\phi_{r,k_mn}, \phi_{-3dB})$ are the TX and RX beamforming gains from mmWave BS to UAV n, and from UE k_m to UAV n, respectively. Whereas $\theta_{t,Bn}$ and ϕ_{r,k_mn} are the boresight angles of the beams, as shown in Figure 2, and θ_{-3dB} and ϕ_{-3dB} are their $-3dB$ beamwidths. The values of $A_{t,Bn}(\theta_{t,Bn}, \theta_{-3dB})$ and $A_{r,k_mn}(\phi_{r,k_mn}, \phi_{-3dB})$ can be calculated using (3) employing their parameters. In (5), $G_{t,nk_m}(\theta_{t,nk_m})$ and $G_{r,nB}(\phi_{r,nB})$ are TX and RX beamforming gains from UAV n to UE k_m and at UAV n from mmWave BS, respectively. Whereas θ_{t,nk_m} and $\phi_{r,nB}$ are the boresight angles of the beams as shown in Figure 2. $G_{t,nk_m}(\theta_{t,nk_m})$ can be expressed as [38]:

$$G_{t,nk_m}(\theta_{t,nk_m}) = 4\cos(\theta_{t,nk_m}), \tag{6}$$

The same equation can be applied to calculate $G_{r,nB}(\phi_{r,nB})$, but θ_{t,nk_m} should be replaced by $\phi_{r,nB}$. The equation given in (6) matches field measurements well, as stated in [38]. For the detailed derivation of (5), including its associated parameters, readers are advised to check [38] along with its cited references. Thus, the spectral efficiency of UE k_m served by UV n can be expressed as:

$$\psi_{nk_m} = \log_2(1 + P_{r,nk_m}/\sigma_0), \tag{7}$$

where σ_0 indicates the AWGN noise power.

3.3. Optimization Problem Formulation of UAV-Hotspot Distribution

Assume that there is a set of $\varnothing_M$ hotspots with a total number of M hotspots are distributed in mmWave BS area. In addition, there is a set of $\varnothing_N$ UAVs with a total number of N UAVs, where $N \leq M$, are flying to cover some of these hotspots by providing additional mmWave links to their associated UEs. These UAVs should be distributed among the hotspots for maximizing their achievable data rates while minimizing UAVs' flying and hovering energy consumptions. This should be conducted under the constraint that each uncovered hotspot should be covered by only one UAV at a time. Mathematically speaking, this optimization problem can be formulated as follows:

$$\mathbf{I}^*_{MN} = \underset{\forall \mathbf{I}_{MN} \in \mathbb{I}_{MN}}{\arg\max} \left(W \sum_{n=1}^{N} \sum_{m=1}^{M} I_{mn} \Psi_{mn} \right), \tag{8a}$$

$$s.t.$$

$$\mathbb{I}_{MN} \in \{0,1\}^{M \times N} \tag{8b}$$

$$\sum_{m=1}^{M} I_{mn} = 1, \forall n \in \varnothing_N \tag{8c}$$

$$\sum_{n=1}^{N} I_{mn} < 2, \forall m \in \varnothing_M \tag{8d}$$

$$E_{nm} \leq E_b \tag{8e}$$

where $\mathbf{I}^*_{MN}$ is the optimal UAV-hotspot assignment matrix, and $\mathbb{I}_{MN} \in \{0,1\}^{M \times N}$ is the space of all available assigned matrices. W is the available bandwidth, and Ψ_{mn} is the total spectral efficiency in bps/Hz of hotspot m when covered by UAV n, which can be expressed as:

$$\Psi_{mn} = \sum_{k_m=1}^{K_m} \psi_{nk_m}, \tag{9}$$

where K_m is the total number of users contained in hotspot m. The constraints (8c) and (8d) are used to guarantee that each hotspot is covered by only one UAV if it is accessible. The fourth constraint means that the available energy of UAV n, i.e., E_{nm}, needed to cover hotspot m is bounded by its battery capacity E_b, where E_{nm} equals:

$$E_{nm} = P_f T_{f_n} + P_h T_{h_n}, \; T_{f_n} = \frac{d_{nm}}{V_f}, \; T_{h_n} = \frac{R_m}{W \Psi_{mn}} \tag{10}$$

where P_f and P_h are the flying and hovering UAV powers while T_{f_n} and T_{h_n} are the flying and hovering periods. T_{f_n} is equal to the separation distance between the current position of UAV n and the location of its chosen hotspot m, i.e., d_{nm}, divided by its flying speed V_f. T_{h_n} is equal to the traffic needs of hotspot m, i.e., R_m in bits, divided by the available data rate in bps when covered by UAV n, i.e., $W \Psi_{mn}$. Herein, $R_m = \sum_{k_m=1}^{K_m} R_{k_m}$, where R_{k_m} is the traffic need of UE k_m in bits. The complexity of the optimization problem given in (8) is of order $\mathcal{O}\left(\frac{M!}{(M-N)!}\right)$, i.e., the number of permutations of N over M. Thus, this problem is an NP time problem, and the budget constraint in (8e) further complicates it.

4. Proposed BCMP-MAB Algorithm

In this section, to address the previous optimization problem, we will reformulate it as a time sequential optimization problem with the aim of maximizing the sum rates of the hotspots sequentially over time. Then, an online learning algorithm based on the MAB hypothesis will be envisioned to address the formulated time sequential problem efficiently.

4.1. MAB Concept

In the MAB game, a player plays over a bandit's arms to maximize his achievable reward through his observations of the played arms [19]. The arms' rewards may come from identical and independent distribution (i.i.d), and the MAB game will be classified as a stochastic MAB, or from random distributions and the MAB game will be classified as an adversarial MAB [19]. Exploitation and exploration are two main phases conducted by MAB players. In the exploitation phase, the best arm having the highest observed reward is selected, while in the exploration phase, the less selected arms are utilized [20]. In some cases, arms' selection comes with paying cost, defined as budget constraint MAB. In this MAB game, the player tries to maximize his achievable profit while minimizing the paying cost of his selected arm [39]. In addition, the MAB games can be classified as single-player MAB (SP-MAB) or MP-MAB based on the number of players involved in the game. In the case of MP-MAB, collisions between players may happen, i.e., two or more players select the same arm simultaneously. Based on the collision model, the arm's reward may be shared among the collided players or none of them gain a bonus. To prevent collisions among the players, some information should be shared among the players in a centralized manner. That is, if the current players' selections are known beforehand, the new player will try to avoid their selections and play the game with the free arms only.

4.2. UAV-Hotspot Distribution Optimization Problem Reformulation

Based on the previously explained MAB hypothesis, the optimization problem given in (8) can be reformulated as a time-sequential BCMP-MAB game as follows:

$$\mathbf{I}^*_{MN} = \underset{\forall \mathbf{I}_{MN,t} \in \mathbb{I}_{MN}}{\arg\max} \left(\frac{W}{T_H} \sum_{t=1}^{T_H} \sum_{n=1}^{N} \sum_{m=1}^{M_{n,t}} I_{mn,t} \Psi_{mn,t} \right),\tag{11a}$$

$s.t.$

$$T_H \in Z^+ \tag{11b}$$

$$\varnothing_{M_{n,t}} \subset \varnothing_M \tag{11c}$$

$$\mathbb{I}_{MN} \in \{0,1\}^{M \times N} \tag{11d}$$

$$\sum_{m=1}^{M} I_{mn,t} = 1, \forall n \in \varnothing_N \tag{11e}$$

$$\sum_{n=1}^{N} I_{mn,t} < 2, \forall m \in \varnothing_M \tag{11f}$$

$$E_{nm} \leq E_b \tag{11g}$$

where

$$\Psi_{mn,t} = \sum_{k_m=1}^{K_m} \psi_{nk_m,t}, \; \psi_{nk_m,t} = \log_2\left(1 + P_{r,nk_m,t}/\sigma_0\right), \tag{12}$$

Herein, $1 \leq t \leq T_H$ where $T_H \in Z^+$ indicates the total time horizon, and Z^+ is the set of positive integers. In (11), $\mathbf{I}_{MN,t}$ indicates the UAV-hotspot assignment matrix at time t, and $\Psi_{mn,t}$ is the total spectral efficiency in bps of hotspot m when covered by UAV n at time t. Constraint (11c), i.e., $\varnothing_{M_{n,t}} \subset \varnothing_M$ means that the set of uncovered hotspots available for UAV n at time t, with a total number of $M_{n,t}$, is a subset of $\varnothing_M$. This information is sent to UAVs from mmWave BS through the dedicated control link. The 4th and 5th constraints (11e) and (11f) mean that only one UAV should cover one hotspot at time t. Thus, the sequential optimization problem given in (11) suggests a time-by-time selection of $I_{mn,t}$. At every time t, UAVs select their corresponding hotspots sequentially based on the uncovered hotspot information sent by BS. In other words, if UAV n selects hotspot m at time t, then this hotspot will be removed from the set of uncovered hotspots available for UAV $n+1$ selection, i.e., $\varnothing_{M_{n+1,t}} = \varnothing_{M_{n,t}}/\{m\}$.

4.3. Proposed BCMP-MAB Algorithm

Algorithm 1 gives the proposed BCMP-MAB algorithm, which is inspired by the cost-subsidy MAB algorithm given in [39]. This algorithm is a budget constraint version of the famous upper confidence bound (UCB) MAB algorithm [40], where the cost of the selected arm is considered while choosing the best bandit's arm. In the original version of cost-subsidy MAB algorithm given in [39], after several rounds of pure exploration, the player calculates the UCB values, and the lower confidence bound (LCB) values of the candidate arms. Then, the arms with UCB values greater than or equal $(1 - \rho)$ multiplied by the maximum LCB value are enumerated. Herein, $\rho \in \{0, 1\}$ is a design parameter of the cost-subsidy algorithm. Among these enumerated candidate arms, the arm with the lowest cost is selected to played.

Algorithm 1: Proposed BCMP-MAB Algorithm

```
    Output:   I*_MN,t .
    Input:    ∅_M , ∅_N , ρ .
    Initialization:  At t = 0, set X_mn,t = 0, Ψ̄_mn,t = 0, I_mn,t = 0, ∀∅_M and ∀∅_N
 1  Pure Exploration Phase:
 2  Explore all available hotspots m ∈ ∅_M and obtain their achievable data
     rates.
 3  for t=1, 2,...,(M+N)τ do
 4  │   Temp=mod (t − 1, M + N)+1
 5  │   for n=1:Temp do
 6  │   │   1. m*_{n,t} = mod(Temp − n, M)+1
 7  │   │   2. UAV n flies towards hotspot m*_{n,t}.
 8  │   │   3. Obtain its achievable data rate Ψ_{m*_n n,t}  and traffic need R_{m*_n n,t}.
 9  │   │   4. X_{m*_n n,t} = X_{m*_n n,t−1}+1.
10  │   │   5. Ψ̄_{m*_n n,t} = (1/X_{m*_n n,t}) Σ_{r=1}^{X_{m*_n n,t}} Ψ_{m*_n n,r}.
11  │   end
12  end
13  Selection Phase:
14  for t=(M+N)τ +1, ...., T_H do
15  │   ∅_{M_1,t} = ∅_M,
16  │   for n=1, 2,...,N do
17  │   │   1. γ^{UCB}_{m_n n,t} = Ψ̄_{m_n n,t} + √(2 ln(t)/X_{m_n n,t})  ∀m_n ∈ ∅_{M_n,t}
18  │   │   2. γ^{LCB}_{m_n n,t} = Ψ̄_{m_n n,t} − √(2 ln(t)/X_{m_n n,t})  ∀m_n ∈ ∅_{M_n,t}
19  │   │   3. γ^{LCB}_{max} = max_{m_n} γ^{LCB}_{m_n n,t}
20  │   │   4. Fs_n(t) = {m_n : γ^{UCB}_{m_n n,t} ≥ (1 − ρ)γ^{LCB}_{max}, ρ ∈ {0, 1}
21  │   │   5. E_{m_n n,t} = P_h T_{h_{m_n n,t−1}} + P_f T_{f_{m_n n,t}}, T_{h_{m_n n,t−1}} = R_{m_n n,t−1}/(W Ψ_{m_n n,t−1}), ∀m_n ∈ ∅_{M_n,t}
22  │   │   6. m*_{n,t} = arg min_{m_n ∈ Fs_n(t)} (E_{m_n n,t})
23  │   │   7. UAV n flies towards hotspot m*_{n,t}
24  │   │   8. Set I_{m*_n n,t}=1
25  │   │   9. X_{m*_n n,t} = X_{m*_n n,t−1} + 1
26  │   │   10. Ψ̄_{m*_n n,t} = (1/X_{m*_n n,t}) Σ_{r=1}^{X_{m*_n n,t}} Ψ_{m*_n n,r}
27  │   │   11. ∅_{M_{n+1},t} = ∅_{M_n,t}/m*_{n,t}
28  │   end
29  end
```

The inputs to the proposed BCMP-MAP algorithm are $\varnothing_M, \varnothing_N$ and ρ, while the output is $\mathbf{I}^*_{MN,t}$, i.e., the UAV-hotspot selection matrix at time t. For initialization, at $t = 0$ and for $\forall \varnothing_M$ and $\forall \varnothing_N$, the number of times hotspot m is selected by UAV n, $X_{mn,t}$, is set to 0. The average spectral efficiency of UAV n when covering hotspot m, $\bar{\Psi}_{mn,t}$, is set to 0, and the element $I_{mn,t}$ is set to 0. The first phase of the BCMP-MAB algorithm is a pure exploration, where each UAV n should visit every hotspot m and obtains its achievable data rate and traffic need, which happens for τ rounds. That is for $1 \le t \le (M + N)\tau$,

a temporary number *Temp* is selected as given in Algorithm 1, where mod indicates the modulo operation, where $M + N$ is used to assure the circulation of the N UAVs over the M hotspots. Then, for $1 \leq n \leq Temp$, a hotspot $m_{n,t}^*$ is selected for UAV n based on the equation given in Algorithm 1. Afterwards, UAV n flies towards it and obtains its achievable data rate $\Psi_{m_n^* n,t}$ and traffic need $R_{m_n^* n,t}$. Then, the number of selections $X_{m_n^* n,t}$, and average spectral efficiency $\bar{\Psi}_{m_n^* n,t}$ are updated as given in Algorithm 1. UAVs visit the hotspots in a circular shift manner by the means of the mod operations, where only one UAV covers one hotspot at a time.

After the pure exploration phase, hotspots selection is accomplished in the second phase of Algorithm 1. In this phase, for $(M + N)\tau + 1 \leq t \leq T_H$, the set of uncovered hotspots available for the first UAV $\varnothing_{M_1,t}$ is set to equal $\varnothing_M$. This first UAV can be selected at random by the mmWave BS. As we previously explained, the BS controls the sequential UAV-hotspot selection to avoid collisions among UAVs and satisfy constraints 3 and 4 in (8). Then, for $1 \leq n \leq N$, UCB and LCB values for UAV n for $\forall m_n \in \varnothing_{M_n,t}$ are calculated as follows:

$$\gamma_{m_n n,t}^{UCB} = \bar{\Psi}_{m_n n,t} + \sqrt{2 \ln(t)/X_{m_n n,t}} \,\forall m_n \in \varnothing_{M_n,t}, \tag{13}$$

$$\gamma_{m_n n,t}^{LCB} = \bar{\Psi}_{m_n n,t} - \sqrt{2 \ln(t)/X_{m_n n,t}} \,\forall m_n \in \varnothing_{M_n,t}, \tag{14}$$

Then, the maximum of $\gamma_{m_n n,t}^{LCB}$ is determined as follows:

$$\gamma_{max}^{LCB} = \max_{m_n} \gamma_{m_n n,t}^{LCB} \tag{15}$$

A feasibility group of candidate hotspots for UAV n is constructed as follows:

$$Fs_n(t) = \{m_n : \gamma_{m_n n,t}^{UCB} \geq (1 - \rho)\gamma_{max}^{LCB}\}, \tag{16}$$

From $Fs_n(t)$, the hotspot $m_{n,t}^*$ characterized with the minimum flying and hovering energy consumptions is selected by UAV n at time t as given in Algorithm 1. This hotspot selection is conducted autonomously by UAV n based on its corresponding $\varnothing_{M_n,t}$ sent by BS. In calculating the expected hovering time of hotspot m_n, as the UAV has no prior knowledge about the spectral efficiency and traffic needs of hotspots, it uses its previous observations $\Psi_{m_n n,t-1}$ and $R_{m_n n,t-1}$ at time $t - 1$, as given in Algorithm 1. After selecting $m_{n,t}^*$, UAV n will fly towards it and cover it. Then, its corresponding element $I_{m_n^* n,t}$ in the UAV-hotspot selection matrix is set to 1, and its associated parameters $X_{m_n^* n,t}$ and $\bar{\Psi}_{m_n^* n,t}$ are updated as given in Algorithm 1. Moreover, it will be removed from the set of available hotspots $\varnothing_{M_{n+1},t}$ for the next UAV $n + 1$ as given in Algorithm 1, which can be selected at random by the BS. The set of $\varnothing_{M_{n+1},t}$ will be collected and sent to the UAV $n + 1$ by BS to schedule the UAV-hotspot selection process to be conducted one by one to prevent UAVs collision, as previously explained.

5. Numerical Analysis

In this section, Monto Carlo (MC) numerical simulations are conducted to prove the effectiveness of the proposed BCMP-MAB algorithm over other benchmarks in different scenarios. In the undertaken simulations, a simulation area of 25 km^2 is established, where 100 hotspots are uniformly distributed inside it. Different number of UAV mounted RIS are used to cover these hotspots based on the conducted simulation scenario. Each attached RIS board has a random number of antenna elements. The altitude of the UAVs is set to 6 m. In addition, each UAV has two statuses, the flying status when it flies towards a hotspot with a flying speed of V_f = 5 Km/h, and hovering status when it covers a hotspot. In addition, each hotspot contains random number of UEs with spontaneous traffic needs, as given in Table 2, which summarizes the simulation parameters used in the conducted numerical simulations unless otherwise stated. For comparisons, random (Rand) selection, where UAV n arbitrarily selects its associated hotspot is provided. In addition, the performance of the nearest hotspot selection is given, where UAVs always choose their nearest hotspots.

Moreover, the performance of naïve UCB is shown, where only $\gamma_{m_n n,t}^{UCB}$ is calculated as given in Algorithm 1, and the hotspot corresponding to the maximum $\gamma_{m_n n,t}^{UCB}$ value is selected. In addition, the maximum rate-based (max rate) selection is given, where UAV always selects the hotspot maximizing its achievable data rate in bps irrespective of its energy consumption. In all compared schemes, i.e., Rand, nearest, naïve UCB and max rate, no UAVs collision avoidance as well as UAV energy minimization are considered. This means that two or more UAVs can cover the same hotspot, where the achievable hotspot data rate is shared among them.

Table 2. Simulation Parameters.

Parameter	Value
P_t, P_f, P_h	1, 4, 2 Watts [31]
V_f	5 Km/h [31]
W	2.16 GHz [41]
λ	0.005 [41]
Γ	0.9 [38]
M	100
Q_n	Uniformly random in the range [32, 512]
d_0	5 m [41]
$\alpha_{LoS}, \alpha_{NLoS}, \alpha$	2.2, 3.88, 2 [41]
$\delta_{LoS}, \delta_{NLoS}$	10.3, 14.6 [41]
$\theta_{-3dB}, \phi_{-3dB}$	$30°$
ρ	0.6
T_H	1000
$\sigma_0(dBm)$	$-174 + 10log10(W) + 10$ [31]
R_{k_i}	Uniformly random in the range [10, 70] Gbit [31]
τ	$(T_H/M)^{2/3}$ [39]

5.1. Adjusting the Value of ρ

In this part of numerical analysis, we will adjust the value of ρ, where N is set to 20 UAVs. Figures 3 and 4 give the average sum rate in Gbps and average energy consumption in Joule of the UAVs using BCMP-MAB algorithm against the value of ρ. When ρ is equal to 0, only candidate hotspots with high UCB values, i.e., high average spectral efficiencies, are picked in $Fs_n(t)$ group of UAV n as given in Algorithm 1. This results in high average sum rates but at the expense of high UAV energy consumption. On the other hand, when ρ is equal to 1, all available hotspots are included in $Fs_n(t)$. This results in low UAV energy consumption as the lowest energy consumption hotspot will always be selected by UAV n, but the expense of low average sum rate. From both figures, $\rho = 0.6$ is chosen as a sufficient value of ρ as given in Table 1. This is because, at $\rho = 0.6$, the average sum rate is slightly reduced by 93.8% from its maximum value, while the average energy consumption is highly reduced to 68% from its maximum value.
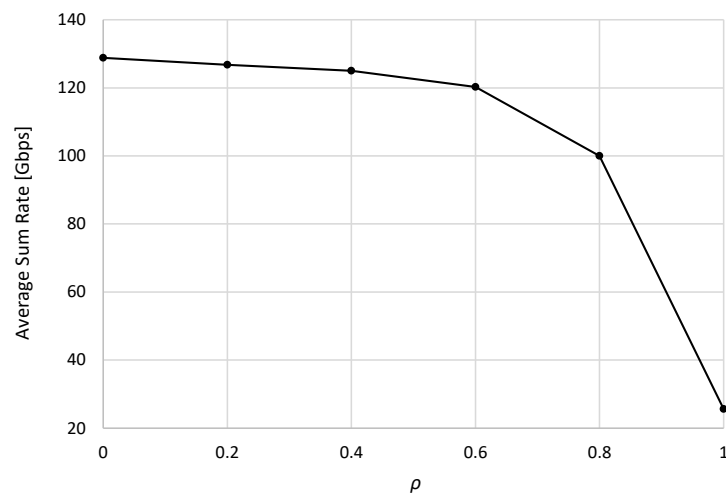
Figure 3. Average sum rate against the value of ρ.
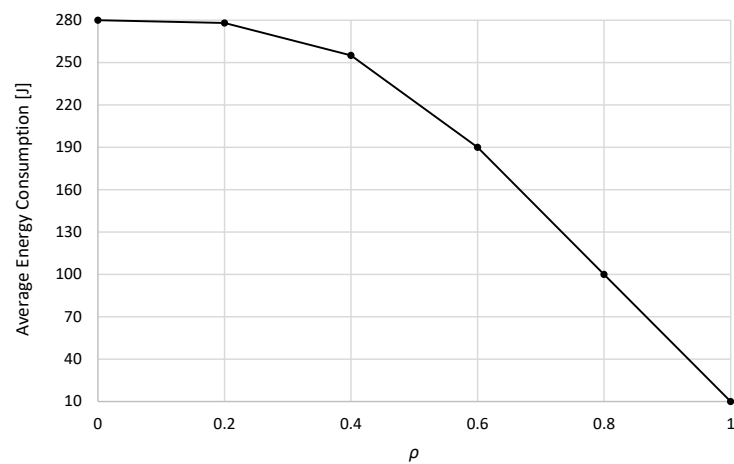
Figure 4. Average energy consumption against the value of ρ.

5.2. Performance Against Number of UAVs

Figure 5 shows the average sum rate against the number of used UAV-mounted RIS. It clearly appears that the proposed BCMP-MAB has the best performance among the schemes involved in the comparison. This comes from compromising between maximizing the achievable data rate while minimizing UAVs energy consumption. Nevertheless, the proposed UAV scheduling mechanism orchestrated by mmWave BS eliminates collisions, i.e., reward sharing among UAVs, hence maximizing the average sum rate compared to other benchmarks. It is interesting to notice that the average sum rate performance of naïve UCB matches that of max rate. This is because UAV always selects the hotspot with the highest average data rate for both schemes. Moreover, both schemes show lower average sum rate performances than the proposed BCMP-MAB because there is no collision avoidance mechanism in these schemes. Thus, multiple UAVs can cover the same hotspot and share its achievable data rate among them, while many other hotspots are left uncovered. Rand and nearest show the worst performance, and Rand is slightly better than the nearest due to randomness in selecting the associated hotspot. At $N = 10$, the proposed BCMP-MAB shows a higher average sum rate than naïve UCB/max rate, Rand and nearest by 1.36, 9.52 and 10.19 times, respectively. However, at $N = 100$, these values become 1.35, 2.12, and 2.54 times, respectively.
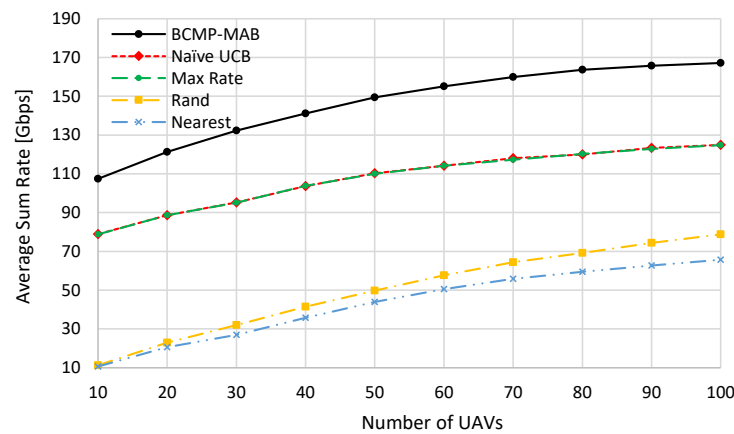
Figure 5. Average sum rate against number of UAVs.

Figure 6 shows the energy efficiency performances in Gbps/J of the schemes involved in the comparisons. As shown by this figure, the proposed BCMP-MAB has the best energy efficiency due to compromising between maximizing the achievable data and minimizing the energy efficiency, and maintaining collision free among UAVs. In addition, Nearest shows better energy efficiency than naïve UCB, max rate, and Rand. This is because it highly reduces the flying energy consumption of UAVs due to selecting the nearest hotspot to them. Again, naïve UCB and max rate show almost the same energy efficiency performance due to their identical objective. At $N = 10$, about 47 and 265, and 59.3 times higher energy efficiency than naïve UCB/max rate, Rand and nearest are obtained by the proposed BCMP-MAB algorithm, respectively. These values become 54.6, 71.5, and 18.76 times at $N = 100$, respectively.
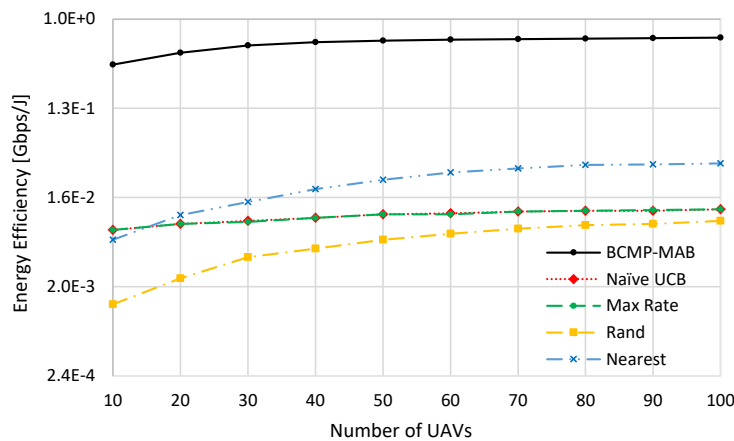

Figure 6. Average energy efficiency against number of UAVs.

5.3. Performance Against TX Power

In this part of numerical analysis, we will bound the performance of the schemes involved in the comparisons against the TX power P_t dBm. Figure 7 shows the average sum rate performances of the schemes involved in the comparisons against P_t dBm using $N = 20$. From this figure, the proposed BCMP-MAB has the best performance, especially for high P_t values due to its spectral efficiency maximization combined with UAVs collision avoidance. Still, naïve UCB and max rate schemes show almost the same performance, and rand-based selection outperforms the nearest based on their policies. At $P_t = 10$ dBm, the proposed BCMP-MAB algorithm obtains average sum rate higher than naïve UCB/max rate, Rand, and nearest by 1.44, 6.35, and 9.4 times, respectively. However, at $P_t = 60$ dBm, about 2.52, 3.13, and 3.14 times higher average sum rate than naïve UCB/max rate, Rand and nearest, are obtained, respectively.
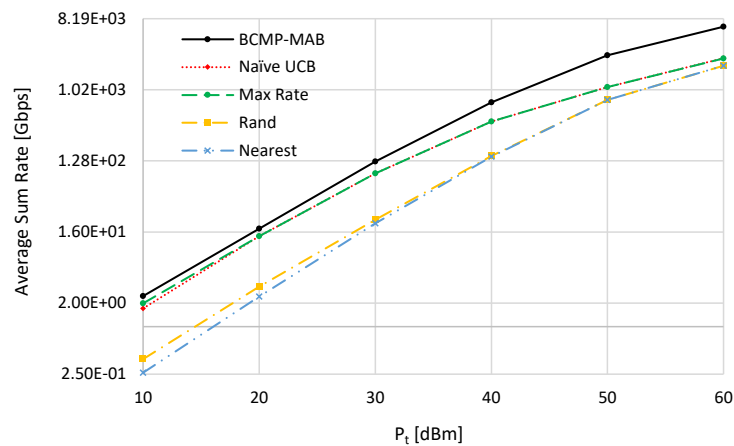
Figure 7. Average sum rate against P_t.

Figure 8 gives the average energy efficiency of the schemes involved in the comparison against P_t. As explicitly shown, the proposed BCMP-MAB has the best performance for all tested P_t values. In addition, naïve UCB and max rate almost have the same performance due to the aforementioned reasons. Rand has the worst energy efficiency performance especially for high values of P_t. It is interesting to notice that the nearest hotspot-based selection has lower energy efficiency performance than naïve UCB, max rate, and even Rand at low P_t values, while it shows better performance than those at high P_t values. This is because at very low P_t values, the hovering time will be considerable. This makes the hovering energy consumption larger than the flying energy consumption and has the most dominant effect. Thus, all compared schemes, without energy minimization features, will influence almost the same energy consumption values. Hence, the average sum rate of these schemes will have the dominant effect in differentiating among their energy efficiency performances. However, at high values of P_t the opposite happens, i.e., the flying energy consumption will be higher than the hovering energy consumption. Consequently, the nearest scheme will have lower energy consumption than naïve UCB, max rate, and Rand, which results in improving its energy efficiency over them, as shown in Figure 8. At $P_t = 10$ dBm, the energy efficiency of the proposed BCMP-MAB is 52.36, 221.76, and 342.72 times higher than naïve UCB/max rate, Rand, and nearest, respectively. At $P_t = 60$ dBm, these values become 6419.5, 4291.8, and 425.2 times, respectively.
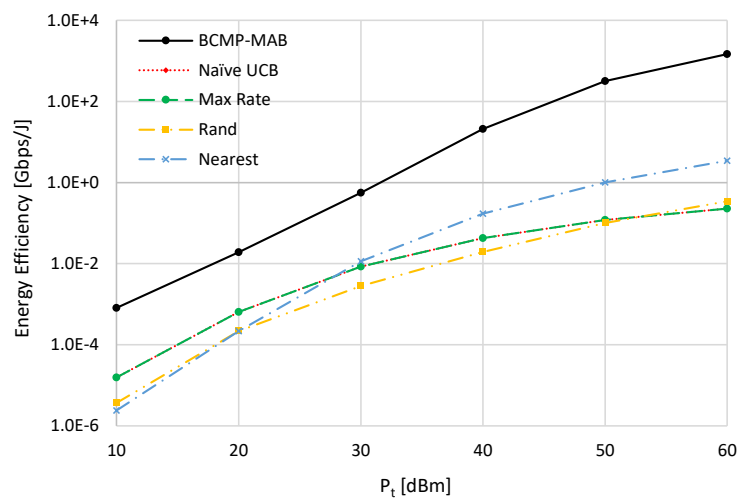

Figure 8. Average energy efficiency in bps/J against P_t.

5.4. Convergence Analysis

Figures 9 and 10 show the convergence performances of the schemes involved in the comparisons against the time horizon. In the conducted simulations, P_t is set to 1 Watt, $M = 100$, $N = 20$ in Figure 9 while $N = 100$ in Figure 10. From both figures, the proposed BCMP-MAB shows fast convergence comparable to the convergence rate of the naïve UCB. For both cases, at $t = 30$, the average sum rate of the proposed BCMP-MAB and naïve UCB reached about 96% of their maximum values at $t = 1000$. In addition, the average sum rate of the naïve UCB converges to that belongs to the max rate scheme.
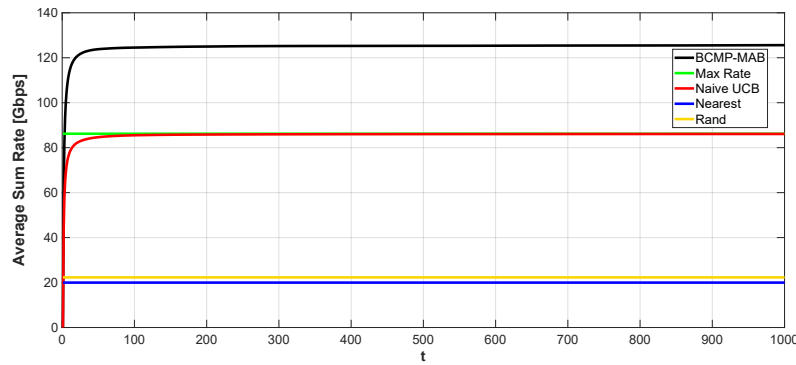
Figure 9. Average sum rate convergence against the time horizon using $M = 100$, and $N = 20$.
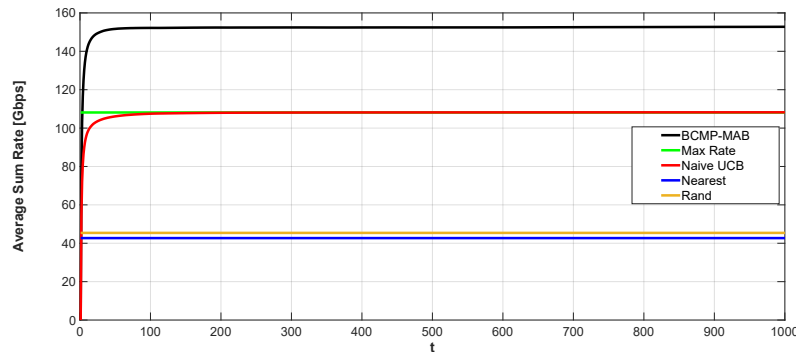
Figure 10. Average sum rate convergence against the time horizon using $M = 100$, and $N = 50$.

5.5. Computational Complexity Comparisons

In the naïve UCB algorithm, the computational complexity comes from calculating the UCB values of hotspots for each UAV and updating their corresponding parameters with computational complexity of $O(N(M+1))$ [31,40]. The computational complexity of the proposed BCMP-MAB consists of two parts. The first part comes from the pure exploration phase, where each UAV should visit all available hotspots several times and update their corresponding parameters with computational complexity of $O(N)$. The second part comes from the hotspot selection phase, which is like naïve UCB except that both UCB and LCB values are calculated. Then, the parameters of the selected hotspots are updated. For simplicity of computational complexity calculation, let us neglect the elimination of the previously selected hotspots. Thus, the upper bound of the computational complexity of the second BCMP-MAB phase can be approximated as $O(N(2M+1))$ [39]. Consequently, the upper bound of the total computational complexity of the proposed BCMP-MAB can be written as $O(N) + O(N(2M+1))$. For the nearest and the maximum rate hotspot-based selections, the distances between UAVs and hotspots and the expected rates between them are calculated, respectively. Then, the selection decision is taken individually for each UAV with total computational complexity of $O(NM)$ for both schemes. For random based selection, each UAV selects a random hotspot out of M total hotspots with total computational complexity of $O(N)$. Thus, the computational complexity of the proposed BCMP-MAB is approximately double the naïve UCB, max rate, and nearest,

while it is almost $2M$ times the random selection. Yet, the performance improvements in energy efficiency using the proposed BCMP-MAB are larger than the degradations in computational complexity, as given in the above simulation results.

6. Conclusions

In this paper, we proposed multi-UAV mounted RIS to cover dense hotspots using mmWave links. The problem of distributing UAVs among the hotspots was formulated as an optimization problem with the aim of maximizing the achievable hotspots sum rate while minimizing both UAVs' flying and hovering energy consumptions. To efficiently address this problem within its constraints, it is reformulated as a time sequential budget constraint MAB problem. Then, a BCMP-MAB algorithm was proposed to address it, where UAVs functioned as the players, hotspots as the bandit's arms and achievable rate as the rewards. Moreover, collision avoidance among UAVs and UAVs budget constraint were considered while maximizing the achievable sum rate. The proposed algorithm showed superior average sum rate and energy efficiency compared to naïve UCB, max rate, random, and nearest-based hotspot selection. For example, at $N = 10$ and $P_t = 10$ dbm, the proposed BCMP-MAB shows a higher energy efficiency than naïve UCB/max rate, Rand and nearest by 47 and 265, and 59.3 times, respectively. In addition, at $N = 20$ and $P_t = 60$ dbm, these values become 6419.5, 4291.8, and 425.2, respectively. These significant enhancements come with only double the computational complexity of the naïve UCB, max rate, and nearest, while it is almost $2M$ times the computational complexity of the random selection. Although we proposed the BCMP-MAB algorithm to address the multi-UAV mounted RIS distribution problem, other solutions such as Q-learning and DRL are applicable as well. However, more investigations are needed to study their realization as well as bounding their performances against the proposed BCMP-MAB scheme. Moreover, the turbulence effect of UAVs due to the rotation of the propellers in conjunction with RIS communication will be one of our future research directions.

Author Contributions: Conceptualization, E.M.M. and S.H.; methodology, E.M.M.; software, E.M.M.; validation, E.M.M., S.H. and M.A.-N.; formal analysis, E.M.M.; investigation, E.M.M.; resources, E.M.M. and M.A.; data curation, E.M.M.; writing—original draft preparation, E.M.M. and S.H.; writing—review and editing, M.A.-N.; visualization, M.A.-N.; supervision, E.M.M.; project administration, E.M.M. and M.A.; funding acquisition, E.M.M. and M.A. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sakaguchi, K.; Mohamed, E.M.; Kusano, H.; Mizukami, M.; Miyamoto, S.; Rezagah, R.E.; Takinami, K.; Takahashi, K.; Shirakata, N.; Peng, H.; et al. Millimeter-wave Wireless LAN and its Extension toward 5G Heterogeneous Networks. *IEICE Trans. Commun.* **2015**, *98-B*, 1932–1948. [CrossRef]
2. Mohamed, E.M.; Sakaguchi, K.; Sampei, S. Wi-Fi Coordinated WiGig Concurrent Transmissions in Random Access Scenarios. *IEEE Trans. Veh. Technol.* **2017**, *66*, 10357–10371. [CrossRef]
3. Rappaport, T.S.; Sun, S.; Mayzus, R.; Zhao, H.; Azar, Y.; Wang, K.; Wong, G.N.; Schulz, J.K.; Samimi, M.; Gutierrez, F. Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! *IEEE Access* **2013**, *1*, 335–349. 2260813. [CrossRef]
4. Abdelreheem, A.; Mohamed, E.M.; Esmaiel, H. Adaptive location-based millimetre wave beamforming using compressive sensing based channel estimation. *IET Commun.* **2019**, *13*, 1287–1296. [CrossRef]
5. ElMossallamy, M.A.; Zhang, H.; Song, L.; Seddik, K.G.; Han, Z.; Li, G.Y. Reconfigurable Intelligent Surfaces for Wireless Communications: Principles, Challenges, and Opportunities. *IEEE Trans. Cogn. Commun. Netw.* **2020**, *6*, 990–1002. [CrossRef]
6. Mohamed, E.M.; Hashima, S.; Hatano, K.; Aldossari, S.A. Two-Stage Multiarmed Bandit for Reconfigurable Intelligent Surface Aided Millimeter Wave Communications. *Sensors* **2022**, *22*, 2179. [CrossRef]

7. Björnson, E.; Özdogan, Ö.; Larsson, E.G. Intelligent Reflecting Surface Versus Decode-and-Forward: How Large Surfaces are Needed to Beat Relaying? *IEEE Wirel. Commun. Lett.* **2020**, *9*, 244–248. [CrossRef]

8. Cui, Z.; Guan, K.; Zhang, J.; Zhong, Z. SNR Coverage Probability Analysis of RIS-Aided Communication Systems. *IEEE Trans. Veh. Technol.* **2021**, *70*, 3914–3919. [CrossRef]

9. Pei, X.; Yin, H.; Tan, L.; Cao, L.; Li, Z.; Wang, K.; Zhang, K.; Björnson, E. RIS-Aided Wireless Communications: Prototyping, Adaptive Beamforming, and Indoor/Outdoor Field Trials. *IEEE Trans. Commun.* **2021**, *69*, 8627–8640. [CrossRef]

10. Tang, W.; Li, X.; Dai, J.Y.; Jin, S.; Zeng, Y.; Cheng, Q.; Cui, T.J. Wireless communications with programmable metasurface: Transceiver design and experimental results. *China Commun.* **2019**, *16*, 46–61. [CrossRef]

11. Nguyen, N.T.; Vu, Q.D.; Lee, K.; Juntti, M. Hybrid Relay-Reflecting Intelligent Surface-Assisted Wireless Communications. *IEEE Trans. Veh. Technol.* **2022**, *71*, 6228–6244. [CrossRef]

12. Zhao, D.; Lu, H.; Wang, Y.; Sun, H. Joint Passive Beamforming and User Association Optimization for IRS-assisted mmWave Systems. In Proceedings of the GLOBECOM 2020—2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; pp. 1–6. [CrossRef]

13. Du, H.; Zhang, J.; Cheng, J.; Ai, B. Millimeter Wave Communications With Reconfigurable Intelligent Surfaces: Performance Analysis and Optimization. *IEEE Trans. Commun.* **2021**, *69*, 2752–2768. [CrossRef]

14. Mohamed, E.M.; Hashima, S.; Anjum, N.; Hatano, K.; Shafai, W.E.; Elhlawany, B.M. Reconfigurable intelligent surface-aided millimetre wave communications utilizing two-phase minimax optimal stochastic strategy bandit. *IET Commun.* **2022**, *16*, 2200–2207. [CrossRef]

15. Mohamed, E.M.; Hashima, S.; Aldosary, A.; Hatano, K.; Abdelghany, M.A. Gateway Selection in Millimeter Wave UAV Wireless Networks Using Multi-Player Multi-Armed Bandit. *Sensors* **2020**, *20*, 3947. [CrossRef] [PubMed]

16. Zhan, P.; Yu, K.; Swindlehurst, A.L. Wireless Relay Communications with Unmanned Aerial Vehicles: Performance and Optimization. *IEEE Trans. Aerosp. Electron. Syst.* **2011**, *47*, 2068–2085. [CrossRef]

17. Mkiramweni, M.E.; Yang, C.; Li, J.; Zhang, W. A Survey of Game Theory in Unmanned Aerial Vehicles Communications. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 3386–3416. [CrossRef]

18. Mozaffari, M.; Saad, W.; Bennis, M.; Nam, Y.H.; Debbah, M. A Tutorial on UAVs for Wireless Networks: Applications, Challenges, and Open Problems. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 2334–2360. [CrossRef]

19. Auer, P.; Cesa-Bianchi, N.; Fischer, P. Finite-time Analysis of the Multiarmed Bandit Problem. *Mach. Learn.* **2004**, *47*, 235–256.:1013689704352. [CrossRef]

20. Audibert, J.Y.; Munos, R.; Szepesvari, C. Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. *Theor. Comput. Sci.* **2009**, *410*, 1876–1902. [CrossRef]

21. Yang, F.; Wang, J.B.; Zhang, H.; Lin, M.; Cheng, J. Intelligent Reflecting Surface Assisted mmWave Communication Using Mixed Timescale Channel State Information. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 5673–5687. [CrossRef]

22. Chen, Y.; Wang, Y.; Jiao, L. Robust Transmission for Reconfigurable Intelligent Surface Aided Millimeter Wave Vehicular Communications With Statistical CSI. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 928–944. [CrossRef]

23. Li, L.; Ma, D.; Ren, H.; Wang, D.; Tang, X.; Liang, W.; Bai, T. Enhanced reconfigurable intelligent surface assisted mmWave communication: A federated learning approach. *China Commun.* **2020**, *17*, 115–128. [CrossRef]

24. Liu, Y.; Zhang, S.; Gao, F.; Tang, J.; Dobre, O.A. Cascaded Channel Estimation for RIS Assisted mmWave MIMO Transmissions. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 2065–2069. [CrossRef]

25. He, J.; Wymeersch, H.; Juntti, M. Channel Estimation for RIS-Aided mmWave MIMO Systems via Atomic Norm Minimization. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 5786–5797. [CrossRef]

26. Pradhan, C.; Li, A.; Song, L.; Vucetic, B.; Li, Y. Hybrid Precoding Design for Reconfigurable Intelligent Surface Aided mmWave Communication Systems. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1041–1045. [CrossRef]

27. Taha, A.; Alrabeiah, M.; Alkhateeb, A. Enabling Large Intelligent Surfaces With Compressive Sensing and Deep Learning. *IEEE Access* **2021**, *9*, 44304–44321. [CrossRef]

28. Jia, C.; Gao, H.; Chen, N.; He, Y. Machine learning empowered beam management for intelligent reflecting surface assisted MmWave networks. *China Commun.* **2020**, *17*, 100–114. [CrossRef]

29. Zhao, D.; Lu, H.; Wang, Y.; Sun, H.; Gui, Y. Joint Power Allocation and User Association Optimization for IRS-Assisted mmWave Systems. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 577–590. [CrossRef]

30. Wang, W.; Zhang, W. Joint Beam Training and Positioning for Intelligent Reflecting Surfaces Assisted Millimeter Wave Communications. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 6282–6297. [CrossRef]

31. Mohamed, E.M.; Hashima, S.; Hatano, K. Energy Aware Multiarmed Bandit for Millimeter Wave-Based UAV Mounted RIS Networks. *IEEE Wirel. Commun. Lett.* **2022**, *11*, 1293–1297. [CrossRef]

32. Zhang, Q.; Saad, W.; Bennis, M. Reflections in the Sky: Millimeter Wave Communication with UAV-Carried Intelligent Reflectors. In Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 9–13 December 2019; pp. 1–6. [CrossRef]

33. Zhang, Q.; Saad, W.; Bennis, M. Distributional Reinforcement Learning for mmWave Communications with Intelligent Reflectors on a UAV. In Proceedings of the GLOBECOM 2020—2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020; pp. 1–6. [CrossRef]

34. Guo, X.; Chen, Y.; Wang, Y. Learning-Based Robust and Secure Transmission for Reconfigurable Intelligent Surface Aided Millimeter Wave UAV Communications. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1795–1799. [CrossRef]

35. Jiang, L.; Jafarkhani, H. Reconfigurable Intelligent Surface Assisted mmWave UAV Wireless Cellular Networks. In Proceedings of the ICC 2021—IEEE International Conference on Communications, Montreal, QC, Canada, 14–18 June 2021; pp. 1–6. [CrossRef]

36. Xiong, B.; Zhang, Z.; Jiang, H.; Zhang, J.; Wu, L.; Dang, J. A 3D Non-Stationary MIMO Channel Model for Reconfigurable Intelligent Surface Auxiliary UAV-to-Ground mmWave Communications. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 5658–5672. [CrossRef]

37. Mohamed, E.M.; Hashima, S.; Hatano, K.; Aldossari, S.A.; Zareei, M.; Rihan, M. Two-Hop Relay Probing in WiGig Device-to-Device Networks Using Sleeping Contextual Bandits. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1581–1585. [CrossRef]

38. Ntontin, K.; Boulogeorgos, A.A.A.; Selimis, D.G.; Lazarakis, F.I.; Alexiou, A.; Chatzinotas, S. Reconfigurable Intelligent Surface Optimal Placement in Millimeter-Wave Networks. *IEEE Open J. Commun. Soc.* **2021**, *2*, 704–718. [CrossRef]

39. Sinha, D.; Abinav Sankararaman, K.; Kazerouni, A.; Avadhanula, V. Multi-Armed Bandits with Cost Subsidy. In Proceedings of The 24th International Conference on Artificial Intelligence and Statistics, Virtual, 13–15 April 2021; Volume 130, pp. 3016–3024.

40. Francisco-Valencia, I.; Marcial-Romero, J.R.; Valdovinos, R.M. A comparison between UCB and UCB-Tuned as selection policies in GGP. *J. Intell. Fuzzy Syst.* **2019**, *36*, 5073–5079. [CrossRef]

41. Mohamed, E.M.; Hashima, S.; Hatano, K.; Fouda, M.M. Cost-Effective MAB Approaches for Reconfigurable Intelligent Surface Aided Millimeter Wave Relaying. *IEEE Access* **2022**, *10*, 81642–81653. [CrossRef]