

Article

Two-Stage Multiarmed Bandit for Reconfigurable Intelligent Surface Aided Millimeter Wave Communications

Ehab Mahmoud Mohamed^{1,2,*} , Sherief Hashima^{3,4} , Kohei Hatano^{3,5}  and Saud Alhajaj Aldossari¹ 

- ¹ Electrical Engineering Department, College of Engineering at Wadi Addwasir, Prince Sattam Bin Abdulaziz University, Wadi Addwasir 11991, Saudi Arabia; s.alhajaj@psau.edu.sa
- ² Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan 81542, Egypt
- ³ Computational Learning Theory Team, RIKEN-Advanced Intelligent Project, Fukuoka 819-0395, Japan; sherief.hashima@riken.jp (S.H.); hatano@inf.kyushu-u.ac.jp (K.H.)
- ⁴ Engineering and Scientific Equipment's Department, Egyptian Atomic Energy Authority, Cairo 13759, Egypt
- ⁵ Faculty of Arts and Science, Kyushu University, Fukuoka 819-0395, Japan
- * Correspondence: ehab_mahmoud@aswu.edu.eg

Abstract: A reconfigurable intelligent surface (RIS) is a promising technology that can extend short-range millimeter wave (mmWave) communications coverage. However, phase shifts (PSs) of both mmWave transmitter (TX) and RIS antenna elements need to be optimally adjusted to effectively cover a mmWave user. This paper proposes codebook-based phase shifters for mmWave TX and RIS to overcome the difficulty of estimating their mmWave channel state information (CSI). Moreover, to adjust the PSs of both, an online learning approach in the form of a multiarmed bandit (MAB) game is suggested, where a nested two-stage stochastic MAB strategy is proposed. In the proposed strategy, the PS vector of the mmWave TX is adjusted in the first MAB stage. Based on it, the PS vector of the RIS is calibrated in the second stage and vice versa over the time horizon. Hence, we leverage and implement two standard MAB algorithms, namely Thompson sampling (TS) and upper confidence bound (UCB). Simulation results confirm the superior performance of the proposed nested two-stage MAB strategy; in particular, the nested two-stage TS nearly matches the optimal performance.

Keywords: millimeter wave; reconfigurable intelligent surface; multiarmed bandit; Thompson sampling; upper confidence bound



Citation: 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. <https://doi.org/10.3390/s22062179>

Academic Editors: Wen-Cheng Lai, Han-Chieh Chao, Adam W. Skorek, Małgorzata Kujawska, Lidia Dobrescu, Sheng-Lyang Jang, Yi Wu, Vijayakumar Varadarajan, Hao Wang and Rashmi Bhardwaj

Received: 4 December 2021

Accepted: 4 March 2022

Published: 10 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A reconfigurable intelligent surface (RIS) is a promising technology to extend the coverage of the communication systems by means of passive antenna arrays [1]. This can be done by configuring the phase shifts (PSs) of the antenna elements to reflect the incoming electromagnetic wave towards an intended destination. Compared with the conventional amplify and forward (AF) and decode and forward (DF) relays, RIS has the advantages of low cost and ease of installation as no RF chains are needed [2]. Millimeter wave (mmWave) communication, i.e., 30~300 GHz band, is another promising technology for fifth-generation (5G) wireless communications and beyond due to its vacant frequencies enabling multi-Gbps transmissions [3–6]. However, due to its high operating frequencies, mmWave is characterized by a short-range transmission with increased susceptibility to path blockage [7]. This necessitates the use of directional antennas in the form of antenna beamforming training (BT) [8–11].

A symbiotic relationship exists between both technologies. On one side, RIS is considered an efficient solution for mmWave challenges, where RIS can extend the mmWave coverage and route around blockages. On the other side, mmWave can directly tune its beam direction towards the RIS location, and the RIS reflects this beam towards the intended mmWave receiver (RX) via adjusting its PSs. However, jointly optimizing the PSs of both mmWave transmitter (TX) and RIS antenna elements is challenging due to the

complex estimation of the massive mmWave channel state information (CSI) at both RIS and mmWave users. Moreover, as RIS is entirely passive, it does not support any channel estimation operations.

In this paper, for practical realization of the mmWave–RIS communication system and to avoid the estimation of the massive mmWave CSI, antenna codebooks are suggested for both mmWave TX and RIS. Without loss of generality, we follow the mmWave codebook design given in [12,13], where the codebooks are generated with 90-degree phase resolution without amplitude adjustment. We choose this codebook design due to its simplicity and standardization by mmWave standards, i.e., Wireless Gigabit (WiGig) standard [12,13]. Furthermore, to maximize the spectral efficiency at the intended mmWave user, PS vectors of both mmWave TX and RIS should be jointly optimized. However, this will consume a considerable beamforming training (BT) overhead, especially when using many PS vectors of sizeable antenna arrays. To efficiently address this problem, an online learning approach is proposed using a single-player multiarmed bandit (MAB) game.

MAB is an online learning strategy where an agent tries to maximize its profit via playing over the bandit's available arms. The agent attempts to compromise between consistently exploiting the arm, giving the maximum profit so far, or exploring new ones, known as *exploitation–exploration* trade-off. Based on the reward's distribution, MAB games can be categorized as *stochastic* or *adversarial* MAB, where in the former the rewards come from a pre-known distribution while in the latter the rewards come arbitrarily. A complete survey of the MAB approach, including its categories and algorithms, can be found in [14,15]. In the formulated MAB game, the mmWave TX will be the player, the available space of the candidate PS vectors will be the arms of the bandit game, and the achievable spectral efficiency at the mmWave user will be the reward. Thus, one set of joint PS vectors can be tested at a time, which highly relaxes the BT overhead. Furthermore, a nested two-stage MAB methodology is proposed to reduce the complexity and increase the convergence rate of the proposed MAB game. Thus, the main contributions of this paper can be summarized as follows:

- The RIS-assisted mmWave communication system is considered, where an optimization problem is formulated to jointly adjust the PS vectors of both mmWave TX and RIS.
- Discrete PSs in the form of codebook design are suggested to relax the complicated CSI estimation problem at both mmWave TX and RIS. In this design, the PSs are assumed to be generated with 90-degree phase resolution with constant amplitude like the codebook design used by WiGig standards [12,13].
- A stochastic single-player MAB game is constructed to jointly optimize the PS vectors of mmWave TX and RIS. This facilitates the adjustment of both PS vectors successively in a time-by-time fashion, which highly reduces the required BT overhead. Typically, the only available information for a MAB player is its reward observation, without any details about the environment. Thus, considering mmWave PSs optimization as a MAB game eliminates the need for CSI estimation as the observed achievable spectral efficiency, i.e., the reward of the game, is the only needed information. This information can be easily obtained via the feedback channel between the mmWave RX and TX. Moreover, the suggested codebook design facilitates the implementation of the MAB game, where the PS vectors are considered as its arms. To reduce the complexity of the arm optimization as we have two sets of arms, one belonging to the mmWave TX and the other to the RIS, a nested two-stage MAB game is proposed in this paper. In this approach, the PS vector of mmWave TX is adjusted in the first MAB stage, and based on it, the PS vector of the RIS is modified in the next stage and vice versa over the time horizon. Thus, at each trial, the player needs to only explore one set of the PS vectors, either that belonging to the mmWave TX or that belonging to the RIS, which reduces the computational complexity of the constructed MAB game. Two common MAB algorithms, namely Thompson sampling (TS) [16] and upper confidence bound (UCB) [17], are used to implement the proposed nested two-stage MAB and compare their performances under the mmWave–RIS environment.

- Numerical analyses are conducted to prove the effectiveness of the proposed mmWave–RIS communication system over benchmarks against the optimal performance under different simulation scenarios.

The rest of this paper is constructed as follows: Section 2 summarizes the related works, and Section 3 discusses the system model and the problem formulation. Section 4 proposes the antenna codebook design and the nested two-stage MAB approach. Section 5 delivers the numerical analysis, followed by the concluding remarks in Section 6.

2. Literature Review

One way to overcome the continuously increasing capacity in the wireless communication systems is to control the channel itself to develop an intelligent radio environment besides other existing solutions (diversity, high-frequency waves, etc.). RIS is a programmable arrangement that controls the propagation of electromagnetic waves by varying its surface's electric and magnetic characteristics. Furthermore, RIS can sense the radio environment by installing intelligent surfaces within the wireless system environment, which entirely or partially controls the features of the radio channels. Hence, RIS-assisted systems can improve the reliability and energy efficiency of wireless systems [1]. Lately, RIS has drawn much consideration as an up-and-coming technology that can suit future wireless systems demands [18], i.e., 6G and beyond. Hence, RIS has promoted wireless applications such as RIS-aided wireless power transfer [19], RIS-aided mobile edge computing [20], RIS-aided physical layer security [21,22], RIS-aided UAV communications [23,24], and mobility and handover management for RIS-aided wireless communications in high-speed trains [25].

There are limited related research works investigating the impact of RIS deployment in mmWave networks. A general tractable model for the coverage performance of the RIS-assisted mmWave networks focused on RIS and base station (BS) densities using stochastic geometry was proposed in [26]. A privacy-preserving design paradigm combining federated learning (FL) with RIS in the mmWave communication system was proposed in [27]. A deep learning algorithm was proposed in [28] to set up a relation between CSI and RIS configurations for better optimal communication rate performance. An efficient cascaded channel estimation model for an RIS-assisted mmWave MIMO system, with the wideband effect on the transmission model, was considered in [29]. A hybrid precoding (HP) design for the RIS-aided multiuser (MU) mmWave communication systems was investigated in [30]. Artificial intelligence (AI)-empowered mmWave communications, especially using RIS, were studied in [31]. To the best of our knowledge, all the current research works on RIS-assisted mmWave assume perfect CSI information. Based on it, the PS vectors of BS and RIS are adjusted to maximize the achievable spectral efficiency at the RX. Without this CSI information, these PS vectors cannot be optimized due to the assumption of continuous PS. However, perfect CSI is a strong assumption violating the RIS hypothesis of being utterly passive without any channel estimation functionality.

3. System Model

Figure 1 shows the system model of the RIS-assisted mmWave communication, where RIS is used to connect the mmWave BS with a single-antenna mmWave user equipment (UE) by routing around the blocker. RIS is equipped with a uniform planar array (UPA) of M antenna elements, while mmWave BS is equipped with a uniform linear array (ULA) of N antenna elements. An RIS controller is used to control the PSs of RIS antenna elements based on the selected PS vector. In addition, mmWave BS and RIS are connected through a dedicated communication link for controlling and information exchange. As a result, the received signal at the UE can be expressed as follows:

$$x = \mathbf{h}_{RU}^H \Phi_i \mathbf{H}_{BR} \mathbf{f}_j s + n, \quad (1)$$

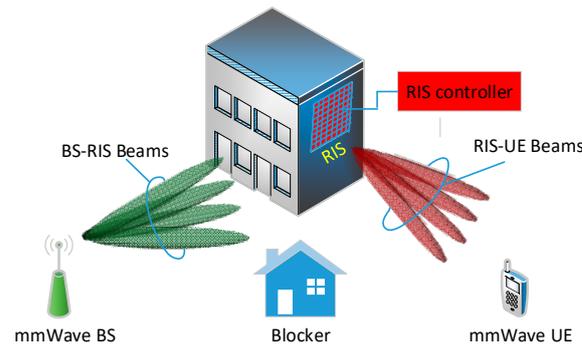


Figure 1. RIS-assisted mmWave communication system.

In (1), s is the transmitted symbol, and x is the received one where $\mathbb{E}[ss^H] = P$, and P is the TX power. $(\cdot)^H$ means Hermitian transpose and $n \sim \mathcal{CN}(0, \sigma_0^2)$ is the complex additive white Gaussian noise (AWGN) with zero mean and variance of σ_0^2 . $\mathbf{f}_j \in \mathbb{C}^{N \times 1}$ is the analog precoder vector of size $N \times 1$ applied at the mmWave BS, and $\Phi_i \in \mathbb{C}^{M \times M}$ is a diagonal matrix of size $M \times M$ containing the PSs of the RIS antenna elements in its diagonal. $\{i, j: 1 \leq i \leq |\mathcal{R}|, 1 \leq j \leq |\mathcal{F}|\}$ are the indices of the used Φ and \mathbf{f} , where \mathcal{R} and \mathcal{F} are their finite sets. $\mathbf{H}_{BR} \in \mathbb{C}^{M \times N}$ is the channel matrix of size $M \times N$ between BS and RIS, while $\mathbf{h}_{RU} \in \mathbb{C}^{M \times 1}$ is the channel vector of size $M \times 1$ between the RIS and UE. Following the geometric channel models with limited scatterers given in [30], \mathbf{H}_{BR} and \mathbf{h}_{RU} can be expressed as follows:

$$\mathbf{H}_{BR} = \sqrt{\frac{MN}{L_{BR}}} \sum_{l=1}^{L_{BR}} \zeta_l \mathbf{\Lambda}_R(\chi_l^{(AoA)}, \delta_l^{(AoA)}) \mathbf{\Lambda}_B(\chi_l^{(AoD)}), \quad (2)$$

$$\mathbf{h}_{RU} = \sqrt{\frac{M}{L_{RU}}} \sum_{l=1}^{L_{RU}} v_l \mathbf{\Lambda}_R(\theta_l^{(AoD)}, \phi_l^{(AoD)}), \quad (3)$$

where L_{BR} and L_{RU} are the number of channel paths between BS and RIS and between RIS and UE, respectively. $\zeta_l \sim \mathcal{CN}(0, \sigma_{\zeta_l}^2)$ and $v_l \sim \mathcal{CN}(0, \sigma_{v_l}^2)$ are the complex path gains of the l -th path in L_{BR} and L_{RU} , respectively. $\mathbf{\Lambda}_R(\chi_l^{(AoA)}, \delta_l^{(AoA)})$, and $\mathbf{\Lambda}_B(\chi_l^{(AoD)})$ are the response vectors of the l -th path array at the RIS and BS, where $\chi_l^{(AoA)}$ ($\delta_l^{(AoA)}$) and $\chi_l^{(AoD)}$ are the azimuth (elevation) angle of arrival (AoA) and angle of departure (AoD), respectively. In addition, $\mathbf{\Lambda}_R(\theta_l^{(AoD)}, \phi_l^{(AoD)})$ is the response vector of the l -th path at the RIS, where $\theta_l^{(AoD)}$ and $\phi_l^{(AoD)}$ are the corresponding azimuth and elevation AoD. Generally, for any θ and ϕ , $\mathbf{\Lambda}_R(\theta, \phi)$ can be expressed as follows [30]:

$$\mathbf{\Lambda}_R(\theta, \phi) = \frac{1}{\sqrt{M}} \left[1, \dots, e^{j \frac{2\pi}{\lambda} d (p \sin(\theta) + q \cos(\phi))}, \dots \right]^T, \quad (4)$$

where d is the antenna spacing and λ is the carrier wavelength and $0 \leq \{p, q\} \leq (\sqrt{M} - 1)$. By analogy, $\mathbf{\Lambda}_B(\chi_l^{(AoD)})$ is defined as follows [30]:

$$\mathbf{\Lambda}_B(\theta) = \frac{1}{\sqrt{N}} \left[1, \dots, e^{j \frac{2\pi}{\lambda} d n \sin(\chi_l^{(AoD)})}, \dots \right]^T, \quad (5)$$

where $0 \leq n \leq (N - 1)$.

The mmWave–RIS optimization problem aims to jointly optimize Φ and f for maximizing the achievable spectral efficiency ψ in bps/Hz at the UE. Mathematically speaking this can be expressed as follows:

$$\{i^*, j^*\} = \max_{i,j} (\psi_{\Phi_i f_j}), \quad (6)$$

where

$$\psi_{\Phi_i f_j} = \log_2 \left(1 + \frac{P \left(\mathbf{h}_{RU}^H \Phi_i \mathbf{H}_{BR} f_j \right) \left(\mathbf{h}_{RU}^H \Phi_i \mathbf{H}_{BR} f_j \right)^H}{\sigma_0^2} \right)$$

Herein, $\psi_{\Phi_i f_j}$ is the spectral efficiency at the UE resulting from using Φ_i and f_j , and the indices of the optimal values of Φ and f are represented by $\{i^*, j^*\}$. Most of the existing literature assumes perfect CSI information; i.e., \mathbf{H}_{BR} and \mathbf{h}_{RU} are well known at both BS and RIS. Based on that, both Φ and f can be jointly adjusted using different iterative techniques [30–32]. However, this is a strong assumption as it is too difficult to estimate \mathbf{H}_{BR} and \mathbf{h}_{RU} due to the use of massive antenna elements at both BS and RIS. Furthermore, RIS should be utterly passive without any channel estimation functionality.

4. Proposed Antenna Codebook and MAB Approach

In this section, antenna codebook and MAB approach are suggested for the mmWave–RIS system to overcome mmWave CSI estimation and jointly adjust the PS vectors of BS and RIS.

4.1. Antenna Codebook Design

To eliminate the need for CSI estimation, discrete PSs are considered for both mmWave BS and RIS, where they constitute the antenna codebook of both. In this context, we will utilize the antenna codebook of WiGig standards for PS design at both BS and RIS [12,13]. This codebook-based beam switching involves fixed beam patterns and can be realized using a predefined pool of antenna weight vectors maintained at TX and RX. Columns of a codebook matrix specify the beamforming weight vector that corresponds to a unique beam pattern. The TX–RX beam pattern pair that optimizes a certain cost function is searched during beamforming according to an agreed criterion. Codebooks support a variety of antenna array geometries and offer flexibility in terms of the number, size, and spacing between antenna elements. For phased array antennas, the columns of the codebook matrix specify the discrete PSs applied to individual antenna elements. The patterns may be generated without amplitude adjustment to obtain processing power savings. As a guiding principle, the columns of the codebook are made orthogonal to each other so that multiple beams can be generated simultaneously without significant interference. These beams can also be synthesized to create a wider beam. Thus, in this codebook design, the PS vectors for $K \leq A$, where A is the total number of antenna elements and K is the total number of PS vectors (i.e., beam directions), are given by column vectors of the following matrix:

$$\mathbf{V}(a, k) = j^{\lfloor \frac{a \times \text{mod}(k+K/2, K)}{K/4} \rfloor}, \quad (7)$$

$$a = 0, \dots, A - 1, k = 0, \dots, K - 1$$

In the case that $K = M/2$, the PS vector at $k = 0$ becomes

$$\mathbf{V}(a, 0) = (-j)^{\text{mod}(a, K)}, a = 0, \dots, A - 1 \quad (8)$$

Thus, the columns in \mathbf{V} are the available space for constructing f and the diagonal of Φ .

4.2. Proposed Nested Two-Stage MAB Approach

Jointly optimizing the values of Φ and f using the prementioned codebook design will consume a considerable BT overhead due to the search over $|\mathcal{R}||\mathcal{F}|$ different candidate beams. Instead, an online learning approach is proposed to successively obtain the optimal solution over the time horizon. This results in considerably reducing the BT overhead as only one pair Φ_i and f_j will be tested at a time. In this context, an online single-player MAB game is constructed to address this problem efficiently. In this formulation, the BS is considered as the player of the bandit game; the available joint values of Φ_i and f_j are the arms of the bandit; and the achievable spectral efficiency at the UE, i.e., $\psi_{\Phi_i f_j}$, is the reward. This MAB-based optimization problem can be mathematically formulated as follows:

$$\max_{\mathbb{I}(1), \dots, \mathbb{I}(T_H)} \frac{1}{T_H} \sum_t \sum_{i,j} \mathbb{I}_{i,j,t} \left(\psi_{\Phi_i f_j} \right), \quad (9)$$

s.t.

- (1) $T_H \in (0, Z^+)$,
- (2) $\sum_{i,j} \mathbb{I}_{i,j,t} = 1$,

where T_H indicates the time horizon and Z^+ is the set of all positive integers. $\psi_{\Phi_i f_j}$ is the spectral efficiency resulting from using Φ_i and f_j combination at time t , i.e., Φ_i and f_j . $\mathbb{I}_{i,j,t}$ is a selection indicator, which is equal to 1 if the combination Φ_i and f_j is selected at time t and 0 otherwise. The constraint $\sum_{i,j} \mathbb{I}_{i,j,t} = 1$ means that only one Φ and f combination is allowed to be selected at time t . To reduce the computational complexity of the constructed MAB game, a nested two-stage MAB strategy is proposed. In the proposed algorithm, the value of Φ is adjusted in the first MAB stage for a particular value of f . Then, based on the adjusted value of Φ , the value of f is adjusted in the second MAB stage, and so on over the time horizon. In this context, two common MAB algorithms are proposed to implement the suggested nested two-stage MAB approach, namely TS and UCB algorithms.

4.2.1. Proposed Nested Two-Stage TS Algorithm

TS is based on a pure Bayesian strategy [16], where prior/posterior distributions are considered for the arms' rewards. The parameters of the assumed probabilistic model are initialized for each arm at the beginning of the algorithm. Then, random samples are taken from the constructed distributions, and the arm related to the highest random sample is selected and played. After obtaining the rewards corresponding to the played arm, its parameters are updated for the posterior distribution of the next round of the MAB game. In the proposed TS algorithm, normal distributions are considered for the spectral efficiency corresponding to each value of Φ and f at time t , i.e., $\psi_{\Phi_i f_j}$, where $1 \leq i \leq |\mathcal{R}|$, $1 \leq j \leq |\mathcal{F}|$. This means that $\psi_{\Phi_i f_j} \sim \mathcal{N}\left(\bar{\psi}_{\Phi_i f_j}, \sigma_{\Phi_i f_j}^2\right)$, where $\bar{\psi}_{\Phi_i f_j}$ and $\sigma_{\Phi_i f_j}^2$ are the mean and variance of the assumed normal distribution, and $\bar{\psi}_{\Phi_i f_j}$ is the average value of $\psi_{\Phi_i f_j}$. This assumption comes from the AWGN term given in (1). Algorithm 1 gives the proposed nested two-stage TS algorithm, where the inputs to the algorithm are the spaces of codebooks \mathcal{R} and \mathcal{F} and the outputs are the adjusted values of Φ_i^* and f_j^* . At the beginning of the algorithm, i.e., $t = 1$ the average spectral efficiencies $\bar{\psi}_{\Phi_i f_j}$, the variances $\sigma_{\Phi_i f_j}^2$, and the number of selections $Z_{\Phi_i f_j}$ corresponding to all values of Φ_i and f_j are set to 0, 1, and 0, respectively. In addition, a PS vector f , i.e., f_j^* , is initialized by picking it uniformly from its corresponding PS codebook. For $2 \leq t \leq T_H$, where T_H is the total time horizon, nested two-stage TS algorithms are performed as follows:

Algorithm 1. Nested Two-Stage TS Algorithm**Output:** Φ_{i^*} and f_{j^*} **Input:** \mathcal{R}, \mathcal{F} **Initialization:** At $t = 1$,

1. $\bar{\psi}_{\Phi_{i^*} f_{j^*}} = 0, \sigma_{\Phi_{i^*} f_{j^*}}^2 = 1, Z_{\Phi_{i^*} f_{j^*}} = 0, 1 \leq i \leq |\mathcal{R}|, 1 \leq j \leq |\mathcal{F}|$
2. Select a value of f_{j^*} at random from its finite set \mathcal{F}

For $t = 2, \dots, T_H$ ❖ **First Stage TS**

1. Based on the value of $f_{j_{t-1}^*}$, select a value of $\Phi_{i_t^*}$ and obtain its corresponding reward

 $\psi_{\Phi_{i_t^*} f_{j_{t-1}^*}}$

- Sample $\tau_{\Phi_{i_t^*} f_{j_{t-1}^*}}, 1 \leq i \leq |\mathcal{R}|$ from normal distributions $\mathcal{N}(\bar{\psi}_{\Phi_{i_t^*} f_{j_{t-1}^*}}, \sigma_{\Phi_{i_t^*} f_{j_{t-1}^*}}^2)$
 - Evaluate i_t^* , and select $\Phi_{i_t^*}$
- $$i_t^* = \arg \max_i (\tau_{\Phi_{i_t^*} f_{j_{t-1}^*}})$$

• Obtain $\psi_{\Phi_{i_t^*} f_{j_{t-1}^*}}$

2. Update $Z_{\Phi_{i_t^*} f_{j_{t-1}^*}}, \bar{\psi}_{\Phi_{i_t^*} f_{j_{t-1}^*}}$, and $\sigma_{\Phi_{i_t^*} f_{j_{t-1}^*}}^2$

- $Z_{\Phi_{i_t^*} f_{j_{t-1}^*}} = Z_{\Phi_{i_{t-1}^*} f_{j_{t-1}^*}} + 1$

- $\bar{\psi}_{\Phi_{i_t^*} f_{j_{t-1}^*}} = \frac{1}{Z_{\Phi_{i_t^*} f_{j_{t-1}^*}}} \sum_{r=1}^{Z_{\Phi_{i_t^*} f_{j_{t-1}^*}}} \psi_{\Phi_{i_t^*} f_{j_{t-1}^*}}$

- $\sigma_{\Phi_{i_t^*} f_{j_{t-1}^*}}^2 = \frac{1}{Z_{\Phi_{i_t^*} f_{j_{t-1}^*}} + 1}$

❖ **Second Stage TS**

3. Based on the value of $\Phi_{i_t^*}$, select a value of $f_{j_t^*}$ and obtain its corresponding reward

 $\psi_{\Phi_{i_t^*} f_{j_t^*}}$

- Sample $\tau_{\Phi_{i_t^*} f_{j_t^*}}, 1 \leq j \leq |\mathcal{F}|$ from normal distributions $\mathcal{N}(\bar{\psi}_{\Phi_{i_t^*} f_{j_t^*}}, \sigma_{\Phi_{i_t^*} f_{j_t^*}}^2)$
- Evaluate j_t^* , and select $f_{j_t^*}$

$$j_t^* = \arg \max_j (\tau_{\Phi_{i_t^*} f_{j_t^*}})$$

• Obtain $\psi_{\Phi_{i_t^*} f_{j_t^*}}$

4. Update $Z_{\Phi_{i_t^*} f_{j_t^*}}, \bar{\psi}_{\Phi_{i_t^*} f_{j_t^*}}$, and $\sigma_{\Phi_{i_t^*} f_{j_t^*}}^2$

- $Z_{\Phi_{i_t^*} f_{j_t^*}} = Z_{\Phi_{i_t^*} f_{j_{t-1}^*}} + 1$

- $\bar{\psi}_{\Phi_{i_t^*} f_{j_t^*}} = \frac{1}{Z_{\Phi_{i_t^*} f_{j_t^*}}} \sum_{r=1}^{Z_{\Phi_{i_t^*} f_{j_t^*}}} \psi_{\Phi_{i_t^*} f_{j_t^*}}$

- $\sigma_{\Phi_{i_t^*} f_{j_t^*}}^2 = \frac{1}{Z_{\Phi_{i_t^*} f_{j_t^*}} + 1}$

END For

In the first stage and based on the value of $f_{j_{t-1}^*}$, a value of $\Phi_{i_t^*}$ is selected and its corresponding reward $\psi_{\Phi_{i_t^*} f_{j_{t-1}^*}}$ is obtained. This is done by sampling the prior distributions of $\psi_{\Phi_{i_t^*} f_{j_{t-1}^*}}$, i.e., $\tau_{\Phi_{i_t^*} f_{j_{t-1}^*}} \sim \mathcal{N}(\bar{\psi}_{\Phi_{i_t^*} f_{j_{t-1}^*}}, \sigma_{\Phi_{i_t^*} f_{j_{t-1}^*}}^2), 1 \leq i \leq |\mathcal{R}|$. Then, the index i_t^* corresponding to the maximum $\tau_{\Phi_{i_t^*} f_{j_{t-1}^*}}$ is selected as follows:

$$i_t^* = \arg \max_i (\tau_{\Phi_{i_t^*} f_{j_{t-1}^*}}), \quad (10)$$

Next, the value of Φ matrix corresponding to this index, i.e., $\Phi_{i_t^*}$, is obtained. Afterward, its corresponding reward $\psi_{\Phi_{i_t^*} f_{j_t^*}}$ is achieved, and its model parameters $Z_{\Phi_{i_t^*} f_{j_t^*}}$,

$\bar{\psi}_{\Phi_{i_t^*}^* f_{j_{t-1}^*}}$, and $\sigma_{\Phi_{i_t^*}^* f_{j_{t-1}^*}}^2$ are updated for its posterior distribution as given in Algorithm 1, where the methodology presented in [33] is used for updating $\bar{\psi}_{\Phi_{i_t^*}^* f_{j_{t-1}^*}}$ and $\sigma_{\Phi_{i_t^*}^* f_{j_{t-1}^*}}^2$.

In the second stage TS and based on $\Phi_{i_t^*}^*$ coming from the first stage, a value of $f_{j_t^*}$ is adjusted, and its corresponding reward $\psi_{\Phi_{i_t^*}^* f_{j_t^*}}$ is obtained. In this procedure, random samples $\tau_{\Phi_{i_t^*}^* f_{j_{t-1}^*}} \sim \mathcal{N}\left(\bar{\psi}_{\Phi_{i_t^*}^* f_{j_{t-1}^*}}, \sigma_{\Phi_{i_t^*}^* f_{j_{t-1}^*}}^2\right)$, $1 \leq j \leq |\mathcal{F}|$, are taken, and the index j_t^* corresponding to the maximum sample value is chosen as follows:

$$j_t^* = \arg \max_j \left(\tau_{\Phi_{i_t^*}^* f_{j_{t-1}^*}} \right), \tag{11}$$

Again, the value of f matrix corresponding to this index, i.e., $f_{j_t^*}$, is obtained. Then, its corresponding reward $\psi_{\Phi_{i_t^*}^* f_{j_t^*}}$ is achieved, and its model parameters are updated for the posterior distribution of the next round of the MAB game as given in Algorithm 1.

4.2.2. Proposed Nested Two-Stage UCB Algorithm

UCB is based on increasing the confidence of the chosen arm by decreasing its uncertainty. This is done by exploiting the arm with the maximum achievable average reward so far while exploring the less selected ones. Algorithm 2 summarizes the proposed nested two-stage UCB algorithm. Like the TS algorithm, the inputs to the algorithm are the spaces of codebooks \mathcal{R} and \mathcal{F} , and the outputs are the adjusted values of $\Phi_{i_t^*}$ and $f_{j_t^*}$. For initialization, at $t = 1$, the average spectral efficiencies $\bar{\psi}_{\Phi_{i_t} f_{j_t}}$ corresponding to all values of Φ_{i_t} and f_{j_t} are set to uniform random values in the range of $[0, 1]$, where $1 \leq i \leq |\mathcal{R}|$, $1 \leq j \leq |\mathcal{F}|$, and their corresponding numbers of selections, i.e., $Z_{\Phi_{i_t} f_{j_t}}$, are set to 1. Moreover, a PS vector f , i.e., $f_{j_t^*}$, is picked uniformly from its corresponding PS codebook, i.e., \mathcal{F} . For $2 \leq t \leq T_H$, nested two-stage UCB algorithms are conducted as follows:

In the first UCB stage, based on the value of $f_{j_{t-1}^*}$ and index of the Φ matrix, i_t^* is selected based on the UCB policy as follows [17]:

$$i_t^* = \arg \max_i \left(\bar{\psi}_{\Phi_{i_{t-1}^*} f_{j_{t-1}^*}} + \sqrt{\frac{2 \ln(t)}{Z_{\Phi_{i_{t-1}^*} f_{j_{t-1}^*}}}} \right) \tag{12}$$

where $\bar{\psi}_{\Phi_{i_{t-1}^*} f_{j_{t-1}^*}}$ represents the exploitation term, while the term $\sqrt{\frac{2 \ln(t)}{Z_{\Phi_{i_{t-1}^*} f_{j_{t-1}^*}}}}$ represents the exploration term of the UCB policy. After selecting $\Phi_{i_t^*}$, its corresponding reward $\psi_{\Phi_{i_t^*} f_{j_{t-1}^*}}$ is obtained and its related parameters $Z_{\Phi_{i_t^*} f_{j_{t-1}^*}}$ and $\bar{\psi}_{\Phi_{i_t^*} f_{j_{t-1}^*}}$ are updated as given in Algorithm 2. Based on the selected value of $\Phi_{i_t^*}$, the value of $f_{j_t^*}$ is adjusted in a nested manner via the second stage UCB as given in Algorithm 2. Then, its corresponding reward $\psi_{\Phi_{i_t^*} f_{j_t^*}}$ is obtained and its related parameters $Z_{\Phi_{i_t^*} f_{j_t^*}}$ and $\bar{\psi}_{\Phi_{i_t^*} f_{j_t^*}}$ are updated as given in Algorithm 2.

Algorithm 2. Nested Two-Stage UCB Algorithm

Output: Φ_{i^*} and f_{j^*}

Input: \mathcal{R}, \mathcal{F}

Initialization: At $t = 1$

1. Set $\bar{\psi}_{\Phi_i f_{j_i}}$ based on uniform random in the range $[0, 1]$, and $Z_{\Phi_i f_{j_i}} = 1, 1 \leq i \leq |\mathcal{R}|, 1 \leq j \leq |\mathcal{F}|$
2. Select a value of $f_{j_i^*}$ at random

For $t = 2, \dots, T_H$

❖ **First Stage UCB**

1. Based on the value of $f_{j_{t-1}^*}$, select a value of $\Phi_{i_t^*}$ and obtain its corresponding reward

$$\psi_{\Phi_{i_t^*} f_{j_{t-1}^*}}$$

- $i_t^* = \arg \max_i \left(\bar{\psi}_{\Phi_i f_{j_{t-1}^*}} + \sqrt{\frac{2 \ln(t)}{Z_{\Phi_i f_{j_{t-1}^*}}}} \right)$

- Obtain $\psi_{\Phi_{i_t^*} f_{j_{t-1}^*}}$

2. Update $Z_{\Phi_{i_t^*} f_{j_{t-1}^*}}$ and $\bar{\psi}_{\Phi_{i_t^*} f_{j_{t-1}^*}}$

- $Z_{\Phi_{i_t^*} f_{j_{t-1}^*}} = Z_{\Phi_{i_{t-1}^*} f_{j_{t-1}^*}} + 1$

- $\bar{\psi}_{\Phi_{i_t^*} f_{j_{t-1}^*}} = \frac{1}{Z_{\Phi_{i_t^*} f_{j_{t-1}^*}}} \sum_{r=1}^{Z_{\Phi_{i_t^*} f_{j_{t-1}^*}} \psi_{\Phi_{i_t^*} f_{j_{t-1}^*}}$

❖ **Second Stage UCB**

3. Based on the value of $\Phi_{i_t^*}$, select a value of $f_{j_t^*}$ and obtain its corresponding reward

$$\psi_{\Phi_{i_t^*} f_{j_t^*}}$$

- $j_t^* = \arg \max_j \left(\bar{\psi}_{\Phi_{i_t^*} f_{j_{t-1}^*}} + \sqrt{\frac{2 \ln(t)}{Z_{\Phi_{i_t^*} f_{j_{t-1}^*}}}} \right)$

- Obtain $\psi_{\Phi_{i_t^*} f_{j_t^*}}$

4. Update $Z_{\Phi_{i_t^*} f_{j_t^*}}$ and $\bar{\psi}_{\Phi_{i_t^*} f_{j_t^*}}$

- $Z_{\Phi_{i_t^*} f_{j_t^*}} = Z_{\Phi_{i_t^*} f_{j_{t-1}^*}} + 1$

- $\bar{\psi}_{\Phi_{i_t^*} f_{j_t^*}} = \frac{1}{Z_{\Phi_{i_t^*} f_{j_t^*}}} \sum_{r=1}^{Z_{\Phi_{i_t^*} f_{j_t^*}} \psi_{\Phi_{i_t^*} f_{j_t^*}}$

END For

5. Numerical Analysis

In this section, Monto Carlo (MC) numerical simulations are conducted to prove the effectiveness of the proposed nested two-stage MAB algorithms compared to the random PS selection, where values of Φ and f are picked uniformly, against the optimal performance. The optimal performance is obtained by testing all available candidate pairs of Φ and f and selecting the best one maximizing ψ . Table 1 summarizes the utilized simulation parameters unless otherwise stated. In addition, it is assumed that the line-of-sight (LoS) path is 10 dB higher than the other paths [34].

Table 1. Simulation parameters.

Parameter	Value
P	10 dBm [2]
BW	2.16 GHz [2]
L_{BR}	5
L_{RU}	5
T_H	1000
d	$\lambda/2$
σ_0^2 (dBm)	$-174 + 10\log_{10}(BW) + 10$ [4]

Figure 2 shows the spectral efficiency performances of the compared schemes i.e., nested two-stage TS, UCB, and random at no blockage against the used number of PS vectors (K), where $K = M = N$. Generally speaking, as the number of K increases, the spectral efficiencies of all schemes increase due to the increase in the received power affected by the increment in the beamforming gain. Although the proposed nested two-stage MAB algorithms do not need CSI estimation and only use the observed spectral efficiency, they have good performances against the optimal performance compared to random PS selection. Moreover, the proposed nested two-stage TS algorithm outperforms all other compared schemes due to its Bayesian policy, which constructs prior/posterior distributions to the achievable spectral efficiency. As the assumed normal distribution highly matches the actual distribution of the attainable spectral efficiency, the proposed nested two-stage TS outperforms UCB-based one. Random PS selection shows the worst performance because it selects the PS vectors arbitrarily without any optimization objective. At $K = 4$, about 98.5%, 97%, and 85.3% of the optimal performance are obtained by the proposed nested two-stage TS, UCB, and random selection, respectively. These values become 94.3%, 86%, and 71.7% when $K = 64$, where the number of alternative beam pairs is highly increased. As TS is a Bayesian-based approach, its performance is still near the optimal one, while the performance of the other two schemes is highly degraded compared to the case of $K = 4$.

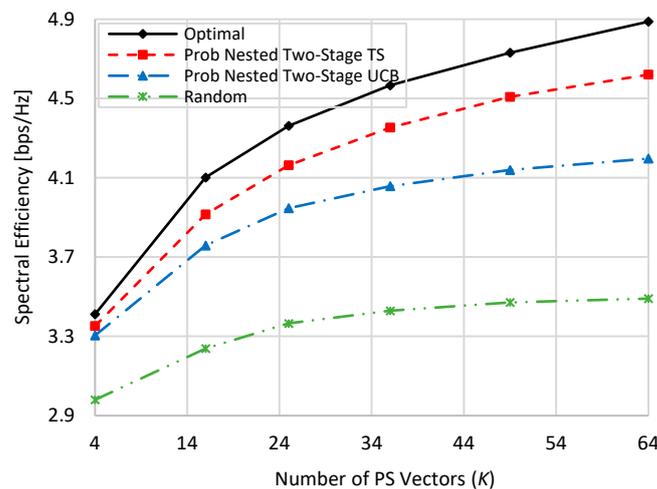


Figure 2. Spectral efficiency against the used number of beams when $M = N = K$ at zero blockage.

Figure 3 shows the spectral efficiency of the schemes involved in the comparison at 80% blockage, where it simulates a harsh blockage environment. In this context, four paths out of the five channel paths between BS and RIS and between RIS and UE undergo blockage, including the LoS path. Compared to Figure 2, more than 50% decrease in spectral efficiency occurs in this harsh blockage environment compared to the zero blockage case. This is due to the low power received from the only surviving path out of the five paths. Despite this harsh blockage environment, the proposed two-stage MAB algorithms still show good spectral efficiency performance against the optimal one. Yet, the proposed two-stage TS outperforms other schemes due to the aforementioned reason. At $K = 4$, about 98.6%, 95.7%, and 72% of the optimal performance are obtained by the proposed nested two-stage TS, UCB, and random selection, respectively. These values become 95.35%, 84.4%, and 58.3% at $K = 64$, respectively. By comparing these ratios with those given in the previous paragraph and shown in Figure 2, it is clearly shown that the performance of random PS selection is highly degraded compared to the optimal performance due to the blockage effect. However, the proposed two-stage MAB algorithms almost have the same ratios of the optimal performance even in this harsh blockage environment.

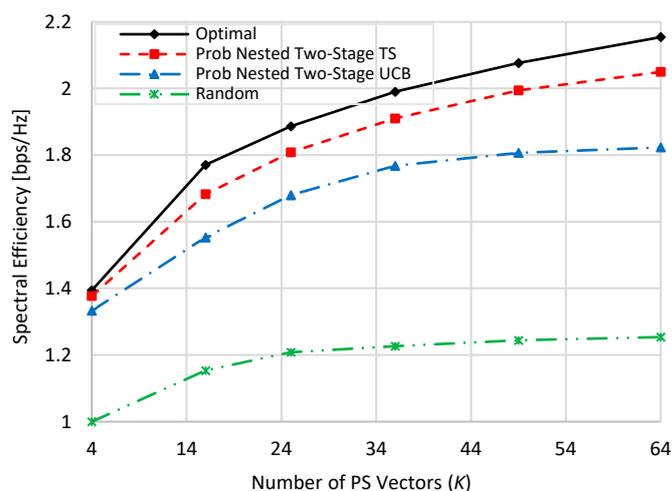


Figure 3. Spectral efficiency against the used number of PS vectors when $M = N = K$ at 80% blockage.

This means that the proposed algorithm can efficiently withstand the blockage effect due to its hypothesis of maximizing the achievable spectral efficiency irrespective of the environmental conditions.

Figures 4 and 5 show the spectral efficiency performance of the compared schemes against the number of PS vectors (K) while M and N are different and fixed, where $N = 36$ and $M = 64$, at zero and 80% blockage, respectively. Again, the spectral efficiencies of all compared schemes are decreased under harsh blockage effect. By comparing Figure 4 with Figures 2 and 5 with Figure 3, the spectral efficiencies of $K = 4, 16$, and 36 in Figures 4 and 5 are slightly higher than those represented in Figures 2 and 3, respectively. However, the spectral efficiencies of $K = 49$ and 64 in Figures 4 and 5 are less than those given in Figures 2 and 3 due to the use of a lower number of antenna elements at both BS and RIS. Typically, the half-power beamwidth is inversely proportional to the used number of antenna elements. Thus, increasing the number of antenna elements for generating the same codebook pattern, i.e., the same number of beams, has two opposite effects. On one hand, it generates narrower beams with larger antenna gains [12,13], and on the other hand, it is more vulnerable to phase shift errors. This increases the gain loss at the maximum gain [12] due to the increase in the interbeam null angles [13]. This is the reason why spectral efficiency is not highly increased when increasing the used number of antenna elements for generating the same number of PS vectors. Interested readers can check the detailed analysis given in [12,13] in this regard. Still, the proposed nested two-stage TS has the best performance that nearly matches the optimal performance due to the prementioned reasoning. At $K = 4$ and zero (80%) blockage, about 99% (97.4%), 97.5% (92%), and 83.5% (63.6%) of the optimal performance are obtained by the proposed nested two-stage TS, UCB, and random selection, respectively. These values become 97% (95%), 87% (87%), and 74.5% (61%) when $K = 64$, respectively. Again, the random selection is highly affected by the blockage effect more than the proposed nested two-stage MAB schemes.

Figures 6 and 7 show the spectral efficiency against the TX signal-to-noise ratio (SNR), i.e., $10 \log_{10} \left(\frac{P}{\sigma_0^2} \right)$, by changing the value of TX power P at zero blockage, when $K = M = N = 16$, and when $N = 36, M = 64$, and $K = 16$. Generally, as the value of P is increased, the spectral efficiency of all schemes is linearly increased. In addition, the spectral efficiencies in the case of $N = 36$ and $M = 64$ are slightly higher than those in the case when $N = 16$ and $M = 16$ due to the increase in beamforming gain resulting from increasing the number of antenna elements. From Figures 6 and 7 and at TX SNR = 10 dB, about 93.8% (94%), 84.9% (86.9), and 51% (49.7%) of the optimal performance are obtained by the proposed nested two-stage TS, UCB, and random selection, respectively. These values become 96.5% (96%), 94.2% (94.16%), and 81.4% (78.8%) at SNR = 100 dB, respectively.

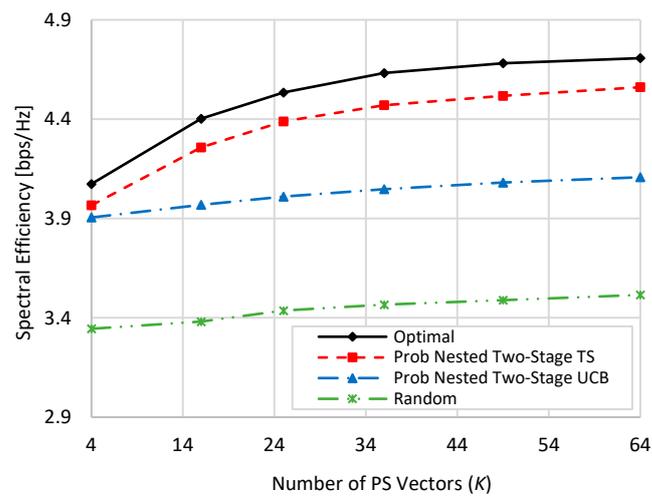


Figure 4. Spectral efficiency against the used number of PS vectors when $N = 36$ and $M = 64$ at zero blockage.

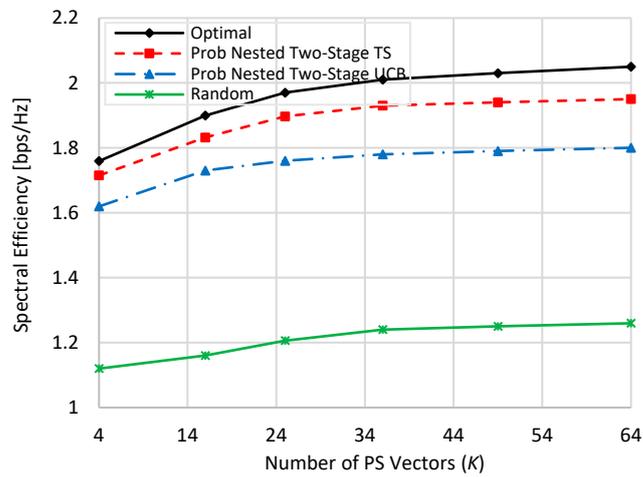


Figure 5. Spectral efficiency against the used number of PS vectors when $N = 36$ and $M = 64$ at 80% blockage.

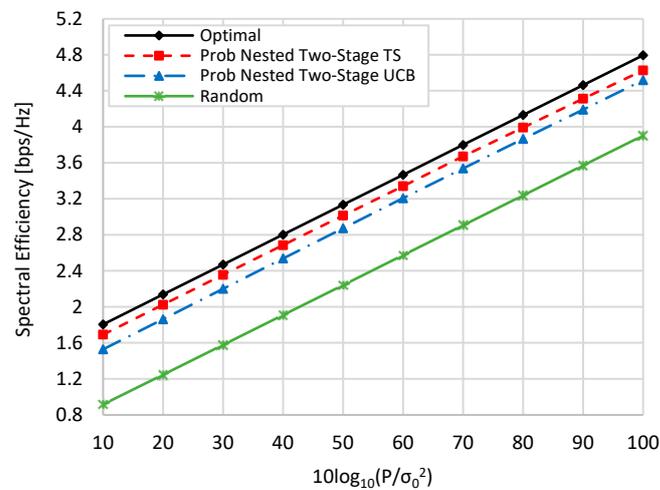


Figure 6. Spectral efficiency against TX SNR when $M = N = K = 16$ at zero blockage.

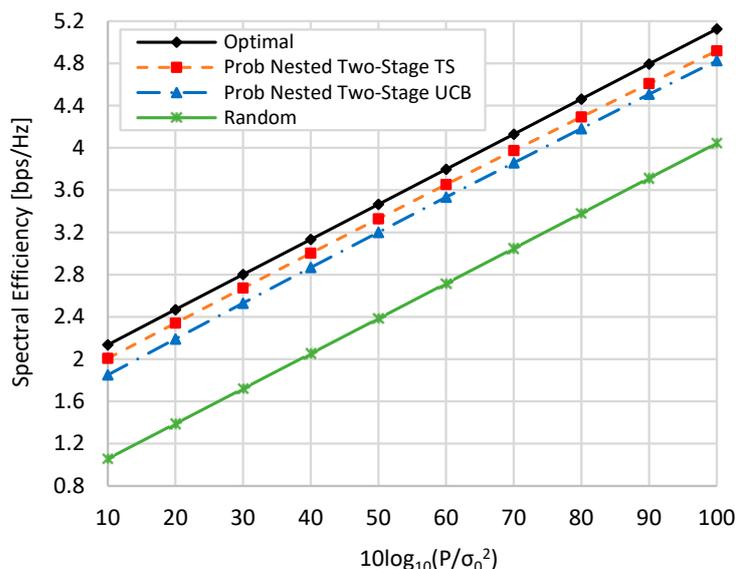


Figure 7. Spectral efficiency against TX SNR when $N = 36$, $M = 64$, and $K = 16$ at zero blockage.

Figures 8 and 9 show the spectral efficiency convergence rate of the proposed nested two-stage MAB schemes against the time horizon t using $K = N = M = 16$ at zero and moderate blockage effect with blockage probability of 50%, respectively. Due to the effect of blockage, the spectral efficiency performance given in Figure 9 is lower than that represented in Figure 8. From these figures, it can be seen that the proposed nested two-stage TS has a faster convergence rate than the UCB-based one due to its Bayesian strategy. Interestingly, the proposed nested two-stage UCB has better convergence than the TS-based one at low values of t , where the TS algorithm starts to build up the prior/posterior distributions of the achievable reward. As these distributions are constructed, the TS converges faster than UCB, as shown by these figures. At $t = 400$, about 99% (99%) and 95.12% (94.8%) of the optimal performance are obtained using the proposed nested two-stage TS and UCB-based one at zero (50%) blockage, respectively.

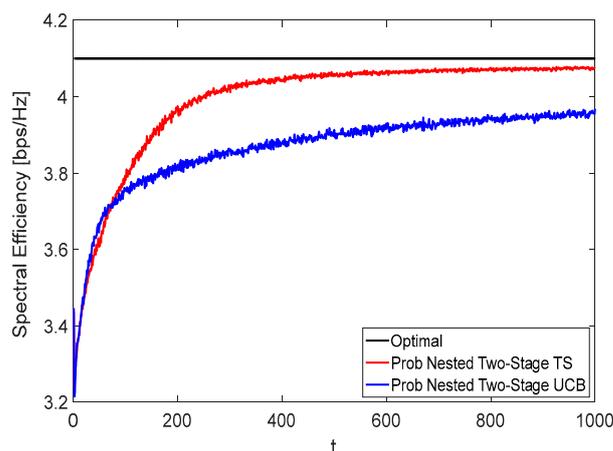


Figure 8. Spectral efficiency convergence at no blockage using $K = N = M = 16$.

Figures 10 and 11 show the spectral efficiency convergence rate using $N = 36$, $M = 64$, and $K = 16$ at zero and 50% blockage, respectively. By comparing Figure 8 with Figures 9 and 10 with Figure 11, it is interesting to observe that the spectral efficiency and the convergence rate performances represented by Figures 10 and 11 are better than those represented by Figures 8 and 9, respectively. This comes from the increased number of antenna elements. At $t = 400$, about 99.4% (99.1%) and 95.5% (95%) of the optimal performance

are obtained using the proposed nested two-stage TS and UCB-based one at zero (50%) blockage, respectively.

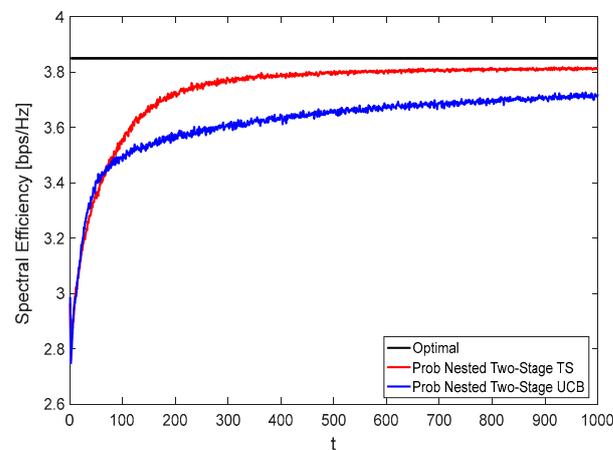


Figure 9. Spectral efficiency convergence at 50% blockage using $K = N = M = 16$.

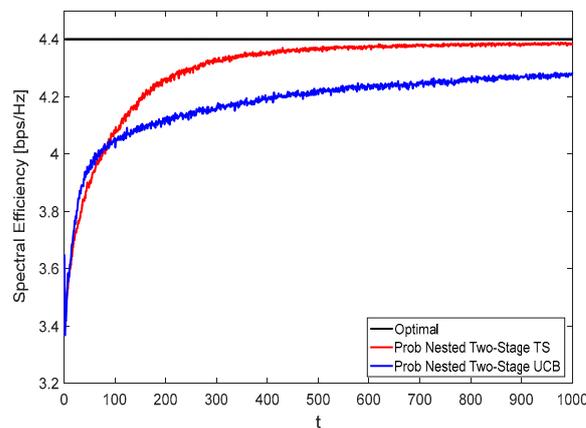


Figure 10. Spectral efficiency convergence at zero blockage using $N = 36$, $M = 64$, and $K = 16$.

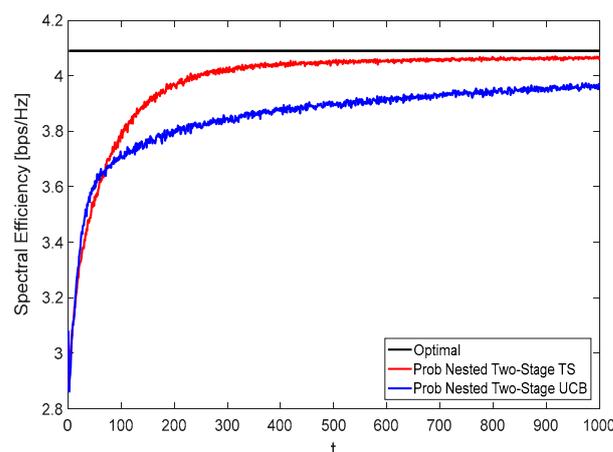


Figure 11. Spectral efficiency convergence at 50% blockage using $N = 36$, $M = 64$, and $K = 16$.

The suggested scheme of the perfect CSI-based approach presented in [30] reaches about 87~88% of the upper bound performance in the highest SNR scenario. This comes while assuming perfect mmWave CSI information, which is impractical in real scenarios. However, the proposed nested two-stage TS reaches about 94~99% of the optimal performance in the different simulation scenarios. Figure 12 shows the spectral efficiency ratio of

the proposed nested two-stage TS, nested two-stage UCB, and the scheme proposed in [30] compared to the random performance against TX SNR. For fair comparisons, we used the same simulation parameters given in [30], i.e., $N = 48$, $M = 100$, and $K = 6$, and the same TX SNR values. As shown by this figure, the spectral efficiency ratio of the proposed nested two-stage TS has the best performance. In addition, both MAB schemes outperform the scheme presented in [30]. At SNR = -25 dB, the spectral efficiency ratios of the proposed nested two-stage TS, nested two-stage UCB, and the scheme given in [30] become 5.5, 4.8, and 4, respectively. This means that about 37.5% and 20% improvements in spectral efficiency performance are obtained by the proposed MAB schemes over the scheme presented in [30]. This comes without any need for knowing the CSI of both mmWave BS and RIS.

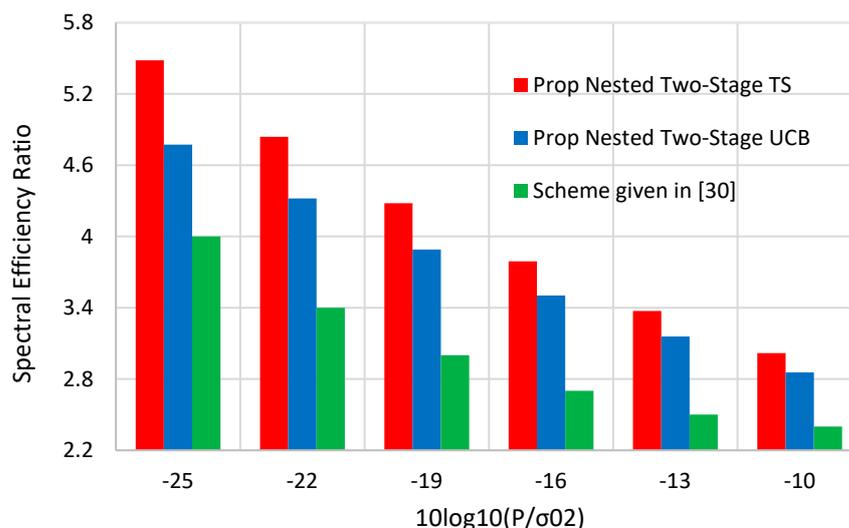


Figure 12. Spectral efficiency comparisons between the proposed nested two-stage MAB schemes and the scheme proposed in [30].

The complexity analysis clearly shows that the proposed nested two-stage MAB scheme has low BT and computational complexities compared to the optimal solution. This is because the optimal strategy explores all available $\{\mathcal{R}, \mathcal{F}\}$ pairs, which obtains its BT and computational complexities of order $\mathcal{O}(|\mathcal{R}||\mathcal{F}|)$. However, in the proposed MAB approach, the sets \mathcal{R} and \mathcal{F} are explored alternatively at every time t . Thus, the BT complexity of the proposed schemes is of order $\mathcal{O}(1)$. Regarding the computational complexities, for the proposed nested two-stage TS, the primary source of computational complexity comes from sampling a 1-dimensional Gaussian random variable and updating its related parameters with the complexity of $\mathcal{O}(|\mathcal{R}| + |\mathcal{F}| + 1)$. In addition, the computational complexity of the proposed nested two-stage UCB comes from selecting the optimal PS and updating its corresponding parameters with the same computational complexity order of $\mathcal{O}(|\mathcal{R}| + |\mathcal{F}| + 1)$. For example, when $|\mathcal{F}| = 36$ and $|\mathcal{R}| = 64$, the BT and computational complexities of the optimal solution are of order $\mathcal{O}(2304)$ while the BT and computational complexities of the proposed nested two-stage MAB approach will be $\mathcal{O}(1)$ and $\mathcal{O}(101)$, respectively. This means that about 99.96% and 96% reductions in BT and computational complexities are obtained, respectively. Consequently, the proposed nested two-stage MAB approach has a near-optimal performance with much lower complexity.

6. Conclusions

In this paper, we have explored RIS-assisted mmWave communications. To avoid estimating the massive mmWave CSI at both RIS and UE, we proposed using antenna codebooks. Moreover, the problem of jointly optimizing the PS vectors at both mmWave BS and RIS was formulated as a MAB game, which contributes to relaxing the required BT overhead. In this context, a nested two-stage MAB strategy was suggested, and nested

two-stage TS and UCB algorithms were proposed to implement the proposed strategy. Simulation analyses confirm the superior performance of the proposed two-stage TS compared to the UCB-based one. Moreover, the proposed nested two-stage MAB schemes outperform random selection and other benchmarks with a high convergence rate and low BT and computational complexities.

Author Contributions: Conceptualization, E.M.M.; Investigation, E.M.M. and K.H.; Methodology, E.M.M. and S.H.; Software, E.M.M. and S.H.; Supervision, S.A.A.; Validation, E.M.M.; Writing—original draft, E.M.M. and S.H.; Writing—review & editing, K.H. and S.A.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia, for funding this research work through the project number IF-PSAU-2021/01/18041.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. 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](#)]
2. 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](#)]
3. 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](#)]
4. Hashima, S.; Hatano, K.; Kasban, H.; Mahmoud Mohamed, E. Wi-Fi Assisted Contextual Multi-Armed Bandit for Neighbor Discovery and Selection in Millimeter Wave Device to Device Communications. *Sensors* **2021**, *21*, 2835. [[CrossRef](#)] [[PubMed](#)]
5. 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**, *10*, 1932–1948. [[CrossRef](#)]
6. Mohamed, E.M.; Abdelghany, M.A.; Zareei, M. An efficient paradigm for multiband WiGig D2D networks. *IEEE Access* **2019**, *7*, 70032–70045. [[CrossRef](#)]
7. Hemadeh, I.A.; Satyanarayana, K.; El-Hajjar, M.; Hanzo, L. Millimeter-Wave Communications: Physical Channel Models, Design Considerations, Antenna Constructions, and Link-Budget. *IEEE Commun. Surv. Tutor.* **2017**, *20*, 870–913. [[CrossRef](#)]
8. Abdelreheem, A.; Mohamed, E.M.; Esmail, H. Location-based millimeter wave multi-level beamforming using compressive sensing. *IEEE Commun. Lett.* **2018**, *22*, 185–188. [[CrossRef](#)]
9. Alkhateeb, A.; El Ayach, O.; Leus, G.; Heath, R.W., Jr. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE J. Sel. Top. Signal Process.* **2014**, *8*, 831–846. [[CrossRef](#)]
10. Abdelreheem, A.; Mohamed, E.M.; Esmail, H. Adaptive location-based millimetre wave beamforming using compressive sensing-based channel estimation. *IET Commun.* **2019**, *13*, 1287–1296. [[CrossRef](#)]
11. ElHalawany, B.M.; Hashima, S.; Hatano, K.; Wu, K.; Mohamed, E.M. Leveraging Machine Learning for Millimeter Wave Beamforming in Beyond 5G Networks. *IEEE Syst. J.* **2021**, 1–2. [[CrossRef](#)]
12. *Specifications for High Rate Wireless Personal Area Networks (WPANs) Amendment 2: Millimeter-Wave-Based Alternative Physical Layer Extension*; IEEE 802.15.3c Standard; IEEE: Piscataway, NJ, USA, 2009.
13. Wang, J.; Lan, Z.; Sum, C.S.; Pyo, C.W.; Gao, J.; Baykas, T.; Rahman, A.; Funada, R.; Kojima, F.; Lakkis, I.; et al. Beamforming Codebook Design and Performance Evaluation for 60GHz Wideband WPANs. In Proceedings of the IEEE Vehicular Technology Conference Fall, Anchorage, AK, USA, 20–23 September 2009; pp. 1–6.
14. Auer, P.; Cesa-Bianchi, N.; Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* **2002**, *47*, 235–256. [[CrossRef](#)]
15. Audibert, J.-Y.; Munos, R.; Szepesvári, C. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theor. Comp. Sci.* **2009**, *410*, 1876–1902. [[CrossRef](#)]
16. Agrawal, S.; Goyal, N. Further optimal regret bounds for thompson sampling. In Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics (AISTATS), Scottsdale, AZ, USA, 29 April–1 May 2013.
17. Francisco-Valencia, I.; Marcial-Romero, J.R.; Valdovinos-Rosas, R.M. A comparison between UCB and UCB-tuned as selection policies in GGP. *J. Intell. Fuzzy Syst.* **2019**, *36*, 5073–5079. [[CrossRef](#)]

18. Basar, E.; di Renzo, M.; de Rosny, J.; Debbah, M.; Alouini, M.-S.; Zhang, R. Wireless Communications Through Reconfigurable Intelligent Surfaces. *IEEE Access* **2019**, *7*, 116753–116773. [[CrossRef](#)]
19. Yang, H.; Yuan, X.; Fang, J.; Liang, Y.-C. Reconfigurable Intelligent Surface Aided Constant-Envelope Wireless Power Transfer. *IEEE Trans. Signal Process.* **2021**, *69*, 1347–1361. [[CrossRef](#)]
20. Li, Z.; Chen, M.; Yang, Z.; Zhao, J.; Wang, Y.; Shi, J.; Huang, C. Energy Efficient Reconfigurable Intelligent Surface Enabled Mobile Edge Computing Networks With NOMA. *IEEE Trans. Cogn. Commun. Netw.* **2021**, *7*, 427–440. [[CrossRef](#)]
21. Tang, Z.; Hou, T.; Liu, Y.; Zhang, J.; Zhong, C. A Novel Design of RIS for Enhancing the Physical Layer Security for RIS-aided NOMA Networks. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 2398–2401. [[CrossRef](#)]
22. Khoshafa, M.H.; Ngatched, T.M.N.; Ahmed, M.H. Reconfigurable Intelligent Surfaces-Aided Physical Layer Security Enhancement in D2D Underlay Communications. *IEEE Commun. Lett.* **2021**, *25*, 1443–1447. [[CrossRef](#)]
23. Wang, W.; Tian, H.; Ni, W. Secrecy Performance Analysis of IRS-Aided UAV Relay System. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 2693–2697. [[CrossRef](#)]
24. Mi, Y.; Song, Q. Energy Efficiency Maximization for IRS-Aided WPCNs. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 2304–2308. [[CrossRef](#)]
25. Liu, H.; Zhang, J.; Wu, Q.; Jin, Y.; Chen, Y.; Ai, B. RIS-Aided Next-Generation High-Speed Train Communications: Challenges, Solutions, and Future Directions. *arXiv* **2021**, arXiv:2103.09484.
26. Chen, Y.; Wang, Y.; Jiao, L. Robust Transmission for Reconfigurable Intelligent Surface Aided Millimeter Wave Vehicular Communications with Statistical CSI. *IEEE Trans. Wirel. Commun.* **2021**, *21*, 924–944. [[CrossRef](#)]
27. 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](#)]
28. Taha, A.; Alrabeiah, M.; Alkhateeb, A. Enabling Large Intelligent Surfaces with Compressive Sensing and Deep Learning. *IEEE Access* **2021**, *9*, 44304–44321. [[CrossRef](#)]
29. 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](#)]
30. 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.
31. 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](#)]
32. Jiang, Z.-M.; Rihan, M.; Zhang, P.; Huang, L.; Deng, Q.; Zhang, J.; Mohamed, E.M. Intelligent reflecting surface aided dual-function radar and communication system. *IEEE Syst. J.* **2021**, 1–12. [[CrossRef](#)]
33. Wilhelmi, F.; Cano, C.; Neu, G.; Bellalta, B.; Jonsson, A.; Barrachina-Muñoz, S. Collaborative spatial reuse in wireless networks via selfish multiarmed bandits. *Ad. Hoc. Netw.* **2019**, *88*, 129–141. [[CrossRef](#)]
34. Rappaport, T.S.; Gutierrez, F.; Ben-Dor, E.; Murdock, J.N.; Qiao, Y.; Tamir, J.I. Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications. *IEEE Trans. Antenn. Propag.* **2013**, *61*, 1850–1859. [[CrossRef](#)]