Received: 28 April 2022 | Revised: 7 July 2022 | Accepted: 26 July 2022

IET Communications

DOI: 10.1049/cmu2.12474

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Reconfigurable intelligent surface-aided millimetre wave communications utilizing two-phase minimax optimal stochastic strategy bandit

Ehab Mahmoud Mohamed[1,2] | **Sherief Hashima**[3,4] | **Nasreen Anjum**[5] | **Kohei Hatano**[3,6] | **Walid El Shafai**[7,8] | **Basem M. Elhlawany**[9,10]

[1]Electrical Engineering Department, College of Engineering, Prince Sattam Bin Abdulaziz University, Wadi Addwasir, Saudia Arabia

[2]Electrical Engineering Department, Aswan University, Aswan, Egypt

[3]Computational Learning Theory team, RIKEN-Advanced Intelligence project, Fukuoka, Japan

[4]Engineering and Scientific Equipment's Department, Nuclear Research Center, Egyptian Atomic Energy Authority, Cairo, Inshas, Egypt

[5]Cyber & Technical Computing School of Computing and Engineering, University of Gloucestershire, Cheltenham, Gloucestershire, UK

[6]Faculty of Arts and Science, Kyushu University, Fukuoka, Japan

[7]Security Engineering Lab, Computer Science Department, Prince Sultan University, Riyadh, Saudi Arabia

[8]Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufa University, Menouf, Egypt

[9]School of Computer Science, Shenzhen University, Shenzhen, China

[10]Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt

Correspondence
Basem M. ElHalawany, Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt.
Email: basem.mamdoh@feng.bu.edu.eg

Funding information
JSPS KAKENHI, Grant/Award Numbers: JP19H04174, JP21K14162

Abstract

Millimetre wave (mmWave) communications, that is, 30 to 300 GHz, have intermittent short-range transmissions, so the use of reconfigurable intelligent surface (RIS) seems to be a promising solution to extend its coverage. However, optimizing phase shifts (PSs) of both mmWave base station (BS) and RIS to maximize the received spectral efficiency at the intended receiver seems challenging due to massive antenna elements usage. In this paper, an online learning approach is proposed to address this problem, where it is considered a two-phase multi-armed bandit (MAB) game. In the first phase, the PS vector of the mmWave BS is adjusted, and based on it, the PS vector of the RIS is calibrated in the second phase and vice versa over the time horizon. The minimax optimal stochastic strategy (MOSS) MAB algorithm is utilized to implement the proposed two-phase MAB approach efficiently. Furthermore, to relax the problem of estimating the channel state information (CSI) of both mmWave BS and RIS, codebook-based PSs are considered. Finally, numerical analysis confirms the superior performance of the proposed scheme against the optimal performance under different scenarios.

1 | INTRODUCTION

Millimetre wave (mmWave) band, that is, 30 to 300 GHz, has a large swath of available spectrum suitable for fifth generation (5G) and beyond 5G (B5G) applications [1]. Nevertheless, it suffers from a fragile channel due to its high operating frequency, which results in intermittent short-range transmission. Thus, antenna beamforming in the form of beamforming

training (BT) is typically used in conjunction with mmWave transmissions [2]. Therefore, extending the mmWave coverage is one of its main challenges. Recently, reconfigurable intelligent surface (RIS) has become a promising technology used to extend the coverage of wireless communications by just reflecting the incident electromagnetic waves using passive antenna arrays [3]. This can be done by adjusting the phase shifts (PSs) of the RIS antenna elements. Therefore, a win–win relationship exists between mmWave and RIS. From one side, RIS can be used to extend the mmWave coverage and route around blockages. On the other side, the mmWave signal can be directed towards the RIS thanks to mmWave BT.

One of the main challenges of RIS-assisted mmWave communication is jointly optimizing the PS vectors of mmWave base station (BS) and RIS to maximize the spectral efficiency at the intended receiver. This challenge comes from the difficulty of estimating mmWave channel state information (CSI) of mmWave BS and RIS due to the use of massive antenna arrays as well as the passivity of RIS without any channel estimation functionality.

There are limited related research works investigating the impact of RIS deployment in mmWave networks. In [4], the coverage of mmWave-RIS system was studied by means of stochastic geometry. A federated learning (FL)-based mmWave-RIS system was proposed for the privacy-preserving design paradigm in [5]. For CSI estimation, the authors in [6] assumed semi-definite passive RIS, where the active antenna elements are used for CSI estimation through compressive sensing deep learning. In [7], the authors proposed cascaded channel estimation for mmWave-RIS to reduce the highly complicated joint channel estimation. In [8], the problem of hybrid precoding (HP) design of the multi-user mmWave-RIS as well as designing the PS matrix of the RIS was addressed by cascaded iterative algorithms. In [9], convolutional neural network (CNN) is utilized to estimate the cascaded mmWave-RIS channel using two stages. Artificial intelligence (AI)-enabled mmWave-RIS was explored in [10]. In [11], RIS is used to assist both dual function radar and communication systems. UAV-mounted RIS was investigated in [12] and [13]. In [12], the trajectory planning of UAV-mounted RIS was considered to maximize its achievable data rate constraint by its limited battery capacity. In [13], resource management, UAV trajectory planning, RIS PS, and mmWave BS beamforming were optimized to minimize the total transmit power. To the best of our knowledge, all existing research works considering the design of PS vectors of mmWave-RIS assumed perfect CSI information, which is a strong assumption violating the RIS hypothesis of being completely passive.

In this paper, an online learning approach is proposed to efficiently address the problem of jointly optimizing the PS vectors of mmWave BS and RIS. In this context, the problem is considered a single player multi-armed bandit (MAB) game. MAB is an efficient online learning policy, where the player attempts to maximize its profit via playing over the available arms of the bandit. The player tries to compromise between consistently exploiting one arm with the highest profit so-far or exploring new ones, called *exploitation-exploration* trade-off. In this mmWave-RIS scenario, the player will be the mmWave BS, the arms of the bandit will be the joint available PS vectors of mmWave BS and RIS, and the profit will be the spectral efficiency at the intended receiver. To mitigate the complexity of the constructed MAB game, a two-phase MAB strategy is proposed. In the first phase, the PS vector of the mmWave BS is adjusted, and based on it, the PS vector of the RIS is adjusted in the second phase, and vice versa over the time horizon. To implement the proposed MAB strategy, a minimax optimal stochastic strategy (MOSS) is leveraged, which is one of the most efficient bandit schemes [14]. The motivation behind selecting MOSS comes from its adaptability to both stochastic and adversarial environments, which goes in harmony with the mmWave-RIS setting. Moreover, antenna codebooks are considered for both mmWave BS and RIS, which facilitates the implementation of the MAB game without the need for CSI estimation. Numerical analysis confirms the superior performance of the proposed MAB-based mmWave-RIS system over benchmarks accompanied with a high convergence rate.

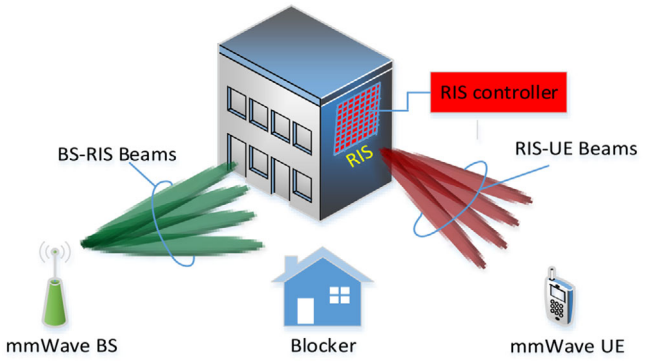

FIGURE 1 RIS-assisted mmWave communication system. mmWave, millimetre wave; RIS, reconfigurable intelligent surface.

2 | SYSTEM MODEL AND PROBLEM FORMULATION

In the mmWave-RIS system model shown in Figure 1, the mmWave BS is equipped with a uniform linear array of N antenna elements, and the RIS is equipped with uniform planner array (UPA) of M antenna elements. The mmWave BS tries to connect with a single antenna mmWave user equipment (UE) via the RIS by routing around the blocker. Through the RIS controller, the mmWave controls the PS vector of the RIS using a dedicated communication link. The received signal at the UE can be expressed as

$$x = \boldsymbol{h}_{RU}^{H} \boldsymbol{\Phi}_i \mathbf{H}_{BR} \boldsymbol{f}_j s + \epsilon, \tag{1}$$

$$1 \leq i \leq |\Omega| \text{ and } 1 \leq j \leq |\mathcal{F}|$$

where x and s indicate the received and transmitted symbols, respectively, where $\mathbb{E}[ss^H] = P$, where $(.)^H$ means Hermitian

transpose and P is the TX power. $\epsilon \sim \mathcal{CN}(0, \sigma^2)$ is the complex additive white Gaussian noise (AWGN), where σ^2 is the noise power. The analog precoder at the mmWave BS is represented by $\boldsymbol{f}_j \in \mathbb{C}^{N\times 1}$, while $\boldsymbol{\Phi}_i$ represents a diagonal matrix of size $M \times M$containing the RIS PS vector of size $M \times 1$ in its diagonal. i and j represent the indices of the used $\boldsymbol{\Phi}$ and $\boldsymbol{f}$, where Ω and $\mathcal{F}$ are the finite sets of available PS matrices and vectors of RIS and BS, respectively. $\boldsymbol{H}_{BR} \in \mathbb{C}^{M\times N}$ is the channel matrix of size $M \times N$ between BS and RIS, while $\boldsymbol{h}_{RU} \in \mathbb{C}^{M\times 1}$ is the channel vector of size $M \times 1$ between the RIS and UE. Following the mmWave channel models with a limited number of scatterers given in [8], $\boldsymbol{H}_{BR}$ and $\boldsymbol{h}_{RU}$ can be expressed as follows:

$$H_{BR} = \sqrt{\frac{MN}{L_{BR}}} \sum_{l=1}^{L_{BR}} \xi_l \boldsymbol{\Lambda}_R \left(\chi_l^{(AoA)}, \delta_l^{(AoA)}\right) \boldsymbol{\Lambda}_B \left(\chi_l^{(AoD)}\right), \tag{2}$$

$$\boldsymbol{h}_{RU} = \sqrt{\frac{M}{L_{RU}}} \sum_{l=1}^{L_{RU}} \nu_l \boldsymbol{\Lambda}_R \left(\theta_l^{(AoD)}, \phi_l^{(AoD)}\right), \tag{3}$$

The number of channel paths between BS and RIS and between RIS and UE are represented by L_{BR} and L_{RU}, with complex gains of $\xi_l \sim \mathcal{CN}(0, \sigma_{\xi_l}^2)$, and $\nu_l \sim \mathcal{CN}(0, \sigma_{\nu_l}^2)$, respectively. The response vectors of the lth path between BS and RIS are represented by $\boldsymbol{\Lambda}_R(\chi_l^{(AoA)}, \delta_l^{(AoA)})$ and $\boldsymbol{\Lambda}_B(\chi_l^{(AoD)})$, 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. Likewise, the response vector of the lth path between RIS and UE is represented by $\boldsymbol{\Lambda}_R(\theta_l^{(AoD)}, \phi_l^{(AoD)})$, where $\theta_l^{(AoD)}$ and $\phi_l^{(AoD)}$ are the azimuth and elevation AoD. Generally, $\boldsymbol{\Lambda}_R(\theta, \phi)$ is defined as

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

where d is the antenna spacing and λ is the carrier wavelength and $0 \le \{p, q\} \le (\sqrt{M} - 1)$. By analogy, $\boldsymbol{\Lambda}_B(\chi_l^{(AoD)})$ is expressed as

$$\boldsymbol{\Lambda}_B\left(\chi_l^{(AoD)}\right) = \frac{1}{\sqrt{N}} \left[1, \ldots, e^{j\frac{2\pi}{\lambda} dn\sin\left(\chi_l^{(AoD)}\right)}, \ldots\right]^T, \tag{5}$$

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

To maximize the spectral efficiency ψ in bps/Hz at the UE, the PS vector $\boldsymbol{f}_j \in \mathcal{F}$ and the PS diagonal matrix $\boldsymbol{\Phi}_i \in \Omega$ should be jointly optimized as follows:

$$\{i^*, j^*\} = \max_{i,j} \left(\psi_{\boldsymbol{\Phi}_i \boldsymbol{f}_j}\right), \tag{6}$$

where

$$\psi_{\boldsymbol{\Phi}_i \boldsymbol{f}_j} = \log_2\left(1 + \frac{P\left(\boldsymbol{h}_{RU}^H \boldsymbol{\Phi}_i H_{BR} \boldsymbol{f}_j\right)\left(\boldsymbol{h}_{RU}^H \boldsymbol{\Phi}_i H_{BR} \boldsymbol{f}_j\right)^H}{\sigma^2}\right)$$

$\psi_{\boldsymbol{\Phi}_i \boldsymbol{f}_j}$ is the spectral efficiency corresponding to $\boldsymbol{\Phi}_i$ and $\boldsymbol{f}_j$ combination, while $\{i^*, j^*\}$ is the indices of the selected optimal combination. The challenge of this optimization problem comes from the difficulty of estimating $\boldsymbol{H}_{BR}$ and $\boldsymbol{h}_{RU}$ to jointly adjust $\boldsymbol{\Phi}_i$ and $\boldsymbol{f}_j$ due to the massive antenna elements used by BS and RIS and the passivity of the RIS. Even if $\boldsymbol{H}_{BR}$ and $\boldsymbol{h}_{RU}$ can be perfectly estimated, the problem presented in (6) is still challenging to solve jointly as shown in [8]. To address this problem, the work presented in [8] assumes perfect CSI information, and the values of $\boldsymbol{\Phi}_i$ and $\boldsymbol{f}_j$ are jointly adjusted using an iterative heuristic method [8], which seems impractical in real scenarios due to the prementioned reasons.

3 | PROPOSED TWO-PHASE MAB APPROACH

In this section, a two-phase MAB approach is proposed for mmWave-RIS system to overcome mmWave CSI estimation and jointly adjust the PS vectors of BS and RIS with low complexity. Towards that antenna codebooks are assumed for antenna arrays of both.

3.1 | Antenna codebook design

In this letter, to eliminate the need to estimate CSI for adjusting the PS vectors of mmWave BS and RIS, the antenna codebook of WiGig standards is utilized [15]. In this codebook design, the PS vectors for $K \le A$ are given by the column vectors of the following matrix:

$$V(a, k) = j^{\text{floor}\left\{\frac{a\times \text{mod}(k+K/2,K)}{K/4}\right\}}, \tag{7}$$

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

where A is the total number of antenna elements, and K is the total number of PS vectors (i.e. beam directions). In the case that $K = M/2$, the PS vector at $k = 0$ becomes

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

Thus, the columns in $\boldsymbol{V}$ are the available space for constructing $\boldsymbol{f}$ and the diagonal of $\boldsymbol{\Phi}$, that is, $\mathcal{F}$ and Ω.

3.2 | MAB hypothesis

MAB is considered stateless reinforcement learning (RL) techniques that can efficiently handle self-decision-making problems better than the famous Q-leaning algorithm [16]. This is because it is lower complicated than Q-learning as it eliminates the need for the memory required to save the sequential states of the Q-learning algorithm. Generally, the MAB problem is a purely online learning, in which the player strives to gain the maximum reward from multiple arms of slot machines [17]. Precisely, the MAB problem aims to detect and select, through finite trials, the arm that maximizes the long-term reward. The player has no prior knowledge about the MAB game except its observable rewards from playing with the bandit arms. This unique feature of the MAB game motivates us to use it as an efficient solution for jointly adjusting the mmWave RIS PSs as it eliminates the need for mmWave CSI estimation. During the MAB game, the player compromises between always exploiting the arm giving the highest reward so far or exploring the less selected ones. MAB games can be categorized as stochastic MAB when the rewards come from independent and identical distributions (i.i.d) or adversarial MAB when the rewards come from unknown distribution. Several algorithms are used to implement the MAB game with different regret bound such as UCB1 [18], MOSS [14], Thompson sampling (TS) [19], and EXP3 [20]. Due to its efficiency, MAB approach was used to address many of wireless communication challenges as given in [21–23].

3.3 | Proposed two-phase MOSS algorithm

Based on the previous codebook design, the values of $\boldsymbol{\Phi}$ and $\boldsymbol{f}$ should be jointly optimized for maximizing ψ at the receiver. However, this will consume a considerable BT overhead due to the need to test $|\Omega||\mathcal{F}|$ different beam pairs, which reduces the achievable throughput. Instead, an online learning approach is proposed in this paper, where the problem is considered a time sequential optimization problem. In this context, the optimal values of $\boldsymbol{\Phi}$ and $\boldsymbol{f}$, that is, $\boldsymbol{\Phi}_{i^*}$ and $\boldsymbol{f}_{j^*}$ will be selected successively over time empowered by the potency of the online learning as follows:

$$\max_{\mathbb{I}(1),\ldots,\mathbb{I}(T_H)} \frac{1}{T_H} \sum_{t} \sum_{i,j} \mathbb{I}_{i_t j_t} \left(\psi_{\boldsymbol{\Phi}_{i_t} \boldsymbol{f}_{j_t}} \right), \tag{9}$$

s.t.

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

where $T_H \in (0,\ Z^+)$ represents the time horizon, and Z^+ is the set of positive integers. $\psi_{\boldsymbol{\Phi}_{i_t} \boldsymbol{f}_{j_t}}$ is the spectral efficiency results from selecting the combination $\boldsymbol{\Phi}_i$ and $\boldsymbol{f}_j$ at time t, that is, $\boldsymbol{\Phi}_{i_t}$ and $\boldsymbol{f}_{j_t}$. $\mathbb{I}_{i_t j_t}$ is a selection indicator, which is equal 1 when the combination $\boldsymbol{\Phi}_i$ and $\boldsymbol{f}_j$ is selected at time t, and zero otherwise. The second constraint in (9) indicates that only one $\boldsymbol{\Phi}$ and $\boldsymbol{f}$ combination is allowed to be selected at time t. To solve this problem, an MAB approach is adopted, where the mmWave BS will be the player of the MAB game, the spaces of $\boldsymbol{\Phi}_i$ and $\boldsymbol{f}_j$ are the arms of the bandit game, and the spectral efficiency is the profit. Additionally, to reduce the complexity and speed up the convergence, as the spaces of $\boldsymbol{\Phi}$ and $\boldsymbol{f}$ are too large, a two-phase MAB hypothesis is proposed. In the first phase, the value of $\boldsymbol{\Phi}$ is adjusted based on a particular value of $\boldsymbol{f}$. In the second phase, the value of $\boldsymbol{f}$ is re-adjusted based on the value of $\boldsymbol{\Phi}$ obtained in the first phase and vice versa over T_H.

To efficiently implement the proposed two-phase MAB strategy, the highly efficient MOSS MAB algorithm is adopted, where Algorithm 1 summarizes the proposed two-phase MOSS algorithm. MOSS [14] is a modified variant of the upper confidence bound (UCB) family, where it relies on the prior knowledge of the horizon. Herein, the confidence interval is identified according to the number of plays for each arm as well as number of actions/arms and the time horizon.

Algorithm 1 Two Phase MOSS Algorithm

Output: $\boldsymbol{\Phi}_{i^*}$, $\boldsymbol{f}_{j^*}$

Input: $\mathcal{F}$, Ω

Initialization: At $t = 1$

1. Set $\bar{\psi}_{\boldsymbol{\Phi}_{i_t} \boldsymbol{f}_{j_t}}$ based on uniform random in the range [0,1] and $Z_{\boldsymbol{\Phi}_{i_t} \boldsymbol{f}_{j_t}} = 1,\ 1 \le i \le |\Omega|,\ 1 \le j \le |\mathcal{F}|$
2. Select a value of $\boldsymbol{f}_{j_1^*}$ at random from its finite set $\mathcal{F}$

For $t = 2, \ldots, T_H$

❖ **First Phase MOSS**

1. Based on the value of $\boldsymbol{f}_{j_{t-1}^*}$, select a value of $\boldsymbol{\Phi}_{i_t^*}$ and obtain its corresponding reward $\psi_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}}$
   - $i_t^* = \arg\max_i \left(\bar{\psi}_{\boldsymbol{\Phi}_{i_{t-1}} \boldsymbol{f}_{j_{t-1}^*}} + \sqrt{\frac{\max\left(\log\left(\frac{t}{Z_{\boldsymbol{\Phi}_{i_{t-1}} \boldsymbol{f}_{j_{t-1}^*}}} \right), 0 \right)}{Z_{\boldsymbol{\Phi} \boldsymbol{f}_{j_{t-1}^*}}}} \right)$
   - Obtain $\psi_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}}$
2. Update $Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}}$ and $\bar{\psi}_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}}$
   - $Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}} = Z_{\boldsymbol{\Phi}_{i_{t-1}^*} \boldsymbol{f}_{j_{t-1}^*}} + 1$
   - $\bar{\psi}_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}} = \frac{1}{Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}}} \sum_{r=1}^{Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}}} \psi_{\boldsymbol{\Phi}_{i_r^*} \boldsymbol{f}_{j_{t-1}^*}}$

❖ **Second Phase MOSS**

3. Based on the value of $\boldsymbol{\Phi}_{i_t^*}$, select a value of $\boldsymbol{f}_{j_t^*}$ and obtain its corresponding reward $\psi_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_t^*}}$
   - $j_t^* = \arg\max_j \left(\bar{\psi}_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}}} + \sqrt{\frac{\max\left(\log\left(\frac{t}{Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}}}} \right), 0 \right)}{Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}}}}} \right)$
   - Obtain $\psi_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_t^*}}$
4. Update $Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_t^*}}$ and $\bar{\psi}_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_t^*}}$
   - $Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_t^*}} = Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}} + 1$
   - $\bar{\psi}_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_t^*}} = \frac{1}{Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_t^*}}} \sum_{r=1}^{Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_t^*}}} \psi_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_r^*}}$

END For

It achieves the order optimal cumulative regret on the finite instances. It divides the horizon by the number of arms to attain minimax optimality. MOSS is adaptable to both stochastic and adversarial setups with better performance than UCB1 [14, 24]. More precisely, the arm index that has been pulled more than horizon/number of arms times is the mean of the rewards collected from the arm. Regarding other arms, their indices are UCB on their mean rewards, which hold with high probability.

The inputs and outputs to the algorithm are the codebook spaces $\boldsymbol{\Omega}$ and $\boldsymbol{\mathcal{F}}$, and the values of $\boldsymbol{\Phi}_{i^*}$ and $\boldsymbol{f}_{j^*}$, respectively. For initialization, at $t = 1$, the expected spectral efficiencies $\bar{\psi}_{\boldsymbol{\Phi}_{i_t} \boldsymbol{f}_{j_t}}$ corresponding to all values of $\boldsymbol{\Phi}_i$ and $\boldsymbol{f}_j$ are set to uniform random values in the range [0,1], and their corresponding number of selections, that is, $Z_{\boldsymbol{\Phi}_{i_t} \boldsymbol{f}_{j_t}}$ are set to 1. Moreover, a PS vector $\boldsymbol{f}_{j_t^*}$ is picked randomly from its corresponding PS codebook, $\boldsymbol{\mathcal{F}}$. For $2 \le t \le T_H$, a two-phase MOSS algorithm is conducted. In the first MOSS phase, based on the value of $\boldsymbol{f}_{j_{t-1}^*}$, i_t^* is selected based on the MOSS policy as follows:

$$i_t^* = \arg\max_i \left(\bar{\psi}_{\boldsymbol{\Phi}_{i_{t-1}} \boldsymbol{f}_{j_{t-1}^*}} + \sqrt{\frac{\max\left(\log\left(\frac{t}{Z_{\boldsymbol{\Phi}_{i_{t-1}} \boldsymbol{f}_{j_{t-1}^*}}} \right), 0 \right)}{Z_{\boldsymbol{\Phi}_{i_{t-1}} \boldsymbol{f}_{j_{t-1}^*}}}} \right), \tag{10}$$

where $\bar{\psi}_{\boldsymbol{\Phi}_{i_{t-1}} \boldsymbol{f}_{j_{t-1}^*}}$ is the average spectral efficiency corresponding to the combination of $\boldsymbol{\Phi}_i$ and the selected $\boldsymbol{f}_{j^*}$ vector selected at time $t-1$. Also, $Z_{\boldsymbol{\Phi}_{i_{t-1}} \boldsymbol{f}_{j_{t-1}^*}}$ is its corresponding number of selections. After evaluating i_t^*, its corresponding $\boldsymbol{\Phi}_{i_t^*}$ and $\psi_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}}$ are obtained and its related parameters, $Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}}$ and $\bar{\psi}_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}}$, are updated as follows:

$$Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}} = Z_{\boldsymbol{\Phi}_{i_{t-1}^*} \boldsymbol{f}_{j_{t-1}^*}} + 1 \tag{11}$$

$$\bar{\psi}_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}} = \frac{1}{Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}}} \sum_{r=1}^{Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}}} \psi_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}^*}} \tag{12}$$

Based on the selected $\boldsymbol{\Phi}_{i_t^*}$, the value of $\boldsymbol{f}_{j_t^*}$ is adjusted in a nested manner via the second phase MOSS with the same way as given in Algorithm 1. Then, its corresponding reward, $\psi_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_t^*}}$, is obtained and its related parameters, $Z_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_t^*}}$ and $\bar{\psi}_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_t^*}}$, are updated as given in Algorithm1.

4 | NUMERICAL ANALYSIS

In this section, Monto-Carlo numerical simulations are conducted to prove the effectiveness of the proposed two-phase MOSS algorithm compared to random selection under different simulation scenarios. In random selection, the values of $\boldsymbol{\Phi}_i$ and $\boldsymbol{f}_j$ are randomly selected. Also, the optimal performance is given, where all $\boldsymbol{\Phi}_i$ and $\boldsymbol{f}_j$ combinations are tested, and the optimal one having the maximum spectral efficiency is chosen. As a benchmark scheme, two-phase UCB1 is employed in the comparisons. It is exactly like Algorithm 1, except that UCB1 equations [18] given in (10) and (11) are used instead of MOSS equations given in (9) and Algorithm 1 to evaluate i_t^* and j_t^*, as follows:

$$i_t^* = \arg\max_i \left(\bar{\psi}_{\boldsymbol{\Phi}_{i_{t-1}} \boldsymbol{f}_{j_{t-1}^*}} + \sqrt{\frac{2\log(t)}{Z_{\boldsymbol{\Phi}_{i_{t-1}} \boldsymbol{f}_{j_{t-1}^*}}}} \right), \tag{13}$$

$$j_t^* = \arg\max_j \left(\bar{\psi}_{\boldsymbol{\Phi}_{i_t^*} \boldsymbol{f}_{j_{t-1}}} + \sqrt{\frac{2\log(t)}{Z_{\boldsymbol{\Phi}_{i_{t-1}} \boldsymbol{f}_{j_{t-1}^*}}}} \right), \tag{14}$$

The used simulation parameters are summarized in Table 1

TABLE 1 Simulation parameters

Parameter	Value
P	10 dBm [2]
BW	2.16 GHz [2]
L_{BR}, L_{RU}	5,5
T_H	1000
$\mathcal{A}_0A, \mathcal{A}_0D$	Uniform random in the range $[0,2\pi]$
σ^2 (dBm)	$-174 + 10\log_{10}(BW) + 10$
d	$\lambda/2$
$\sigma_{\xi_l}^2, \sigma_{\nu_l}^2$	10 dB when $l = 1$, and 1 dB for $2 \le l \le 5$ [25]

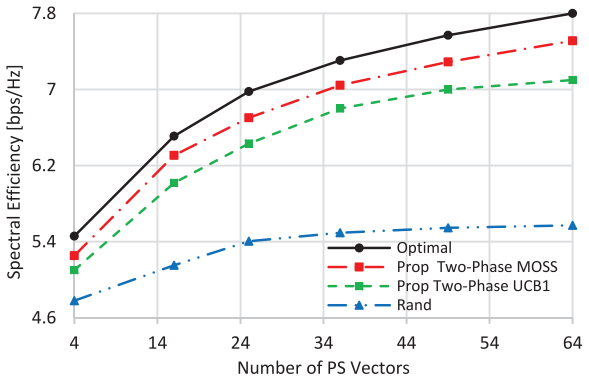

FIGURE 2 Spectral efficiency against the number of beams using $N = M = 16$

4.1 | Performance comparisons

Figure 2 shows the spectral efficiency of the schemes involved in the comparisons against the number of used PS vectors K, using $N = 16$ and $M = 16$. Generally, the spectral efficiency of all compared schemes is increased with increasing K due to the increase in the beamforming gain. Also, the proposed two-phase MOSS algorithm nearly matches the optimal performance which is better than two-phase UCB1. However, the

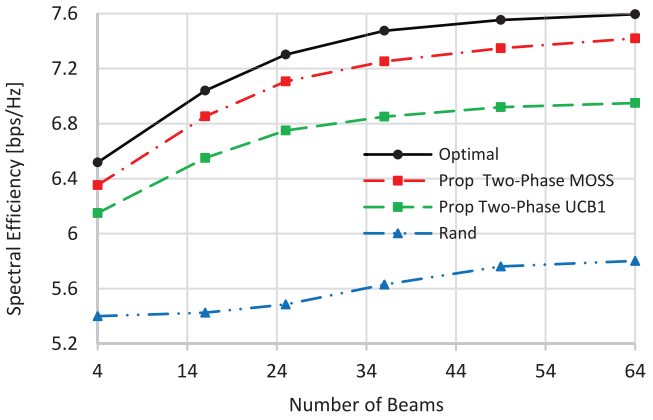

FIGURE 3 Spectral efficiency against the number of beams using $N = 36$ and $M = 64$

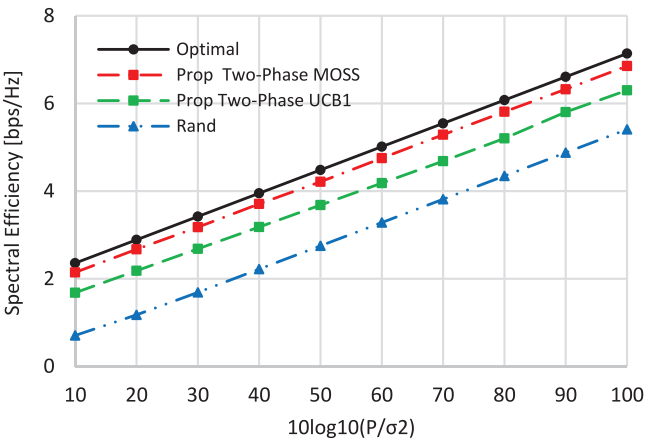

FIGURE 4 Spectral efficiency against 10log10(P/σ^2) using $N = 36$, $M = 64$, and $K = 16$

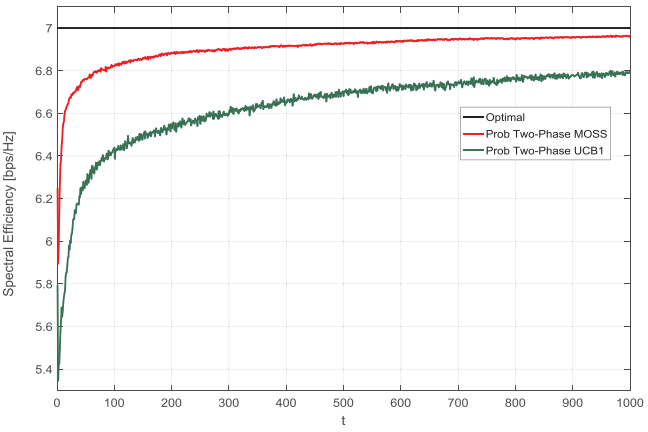

FIGURE 5 Spectral efficiency convergence using $N = 36$, $M = 64$, and $K = 16$

random selection scheme shows bad performance. This comes from the proposed online learning approach, which tries to reach the optimal performance successively over the time horizon along with the better performance of the MOSS strategy over UCB1. As the values of $\boldsymbol{\Phi}_i$ and $\boldsymbol{f}_j$ are selected randomly in the random selection scheme, its performance is too far from the optimal one. At $K = 4$ (64), the proposed two-phase MOSS, two-phase UCB1, and random selection achieve about 96.3% (96.15%), 93.6% (88.9%), and 87.5% (71.4%) of the optimal performance, respectively. It is noted that the random performance becomes far from the optimal one at higher values of K due to the increased number of PS vectors combinations. However, the proposed MOSS scheme is not affected by the number of used beams and always nearly matches the optimal performance.

Figure 3 shows the spectral efficiency of the schemes involved in the comparisons against K while $N = 36$ and $M = 64$. Again, the spectral efficiency of all schemes is increased when increasing K. By comparing Figures 3 and 2, for $K < 36$, the spectral efficiencies in Figure 3 are higher than those in Figure 2 due to the increased number of used antenna elements and vice versa for $K > 36$. Yet, the proposed two-phase MOSS scheme is better than UCB1 and nearly matches the optimal performance, while the random selection is far from it. At $K = 4$ (64), the proposed two-phase MOSS, two-phase UCB1, and random selection achieve about 97% (97.4%), 94.45% (91.6%), and 83% (76.3%) of the optimal performance, respectively.

Figure 4 shows the spectral efficiency of the schemes involved in the comparisons against increasing the TX power, that is, P. As P is increased, the spectral efficiency of all methods is increased due to the increase in the received power. Yet, the proposed two-phase MOSS scheme is better than UCB1, where it nearly matches the optimal performance at all tested P values, while the random selection is far from the optimal performance. For example, at TX SNR of 10 (100) dB, the proposed two-phase MOSS, two-phase UCB1, and random selection achieve about 91% (95.7%), 71.2% (88.73%), and 29.8% (76%) of the optimal performance, respectively. It is noted that deficient performance compared to the optimal one is obtained at low value of TX SNR for the random selection, while the proposed MAB schemes have good performance even in low TX SNR conditions.

Figure 5 gives the spectral efficiency convergence of the proposed two-phase MOSS and two-phase UCB1 algorithms against the time horizon using $N = 36$, $M = 64$, and $K = 16$. The proposed two-phase MOSS algorithm shows faster convergence than UCB1 towards the optimal performance. At $t = 400$, about 99% (95%) of the optimal performance is obtained by the proposed two-phase MOSS (UCB1) scheme, respectively.

Compared to the perfect CSI-based approach presented in [8], their suggested scheme reaches 87% to 88% of the upper bound performance in the highest SNR scenario. This comes while assuming perfect CSI information, which is impractical in real situations. However, the proposed two-phase MOSS reaches about 91% to 99% of the optimal performance in the different simulation scenarios. Moreover, using $N = 48$, $M = 64$, and SNR $= -12$ dB, the spectral efficiency of their proposed scheme reaches 2.25 of the random PS selections. However, by simulating the same parameters, about 3.4 improvement over the random PS selection is obtained by the proposed scheme. This comes without the need for knowing the CSI of both mmWave BS and RIS.

As the proposed scheme eliminates the need for mmWave CSI estimation, it considerably reduces the pilot overhead. In [26] and [27], it is stated that pilot overhead required for RIS-based CSI estimation should be $\geq MN$. As an illustrative example, suppose that M and N are equal to 64, then the pilot overhead should $\geq$ 3904 symbols, which is too large compared to the total frame length. This value will be incredibly increased for massive mmWave BS and RIS containing hundreds of antenna elements. Instead, in the proposed online learning approach, mmWave BS just needs to send the index of the used $\mathbf{\Phi}$ matrix to the RIS through the dedicated control channel between them at time t, which consumes negligible overhead compared to that required by perfect CSI based approach. For example, if the number of PS vectors is equal to 64, then only 6 bits need to be transmitted between BS and RIS. Moreover, the overhead of the proposed scheme is constant irrespective of the SNR conditions.

4.2 | Complexity analysis

The time complexity of the joint mmWave BS-RIS PSs estimation algorithms comes from two sources. The first source is due to the BT process among BS, RIS, and UE, and the second source is due to the computational complexity of the used algorithm. The first source is considered the major source of time complexity as one BT process between mmWave TX and RX may consume about 50 msec as given in [28]. However, the second source, i.e., computational complexity, is based on instructions execution time, which is negligible by considering the high-speed mmWave BS platforms.

For BT time complexity, the proposed two-phase MOSS algorithm and UCB1 scheme have much lower complexity than the optimal strategy. This is because the optimal strategy explores all available $\{\Omega, \mathcal{F}\}$ pairs, which gets its BT complexity of order $\mathcal{O}(|\Omega||\mathcal{F}|)$. However, in the proposed MAB approach, one $\mathbf{\Phi}_i, \boldsymbol{f}_j$ combination is tested using BT at every time t . Therefore, the BT complexity of the proposed scheme is of order $\mathcal{O}(1)$. Similarly, the BT complexity of the random selection is of order $\mathcal{O}(1)$ as only one random $\mathbf{\Phi}_i, \boldsymbol{f}_j$ combination is tested at time using BT.

For computational complexity, the primary source of computational complexity of the proposed two-phase MOSS/UCB1 schemes come from selecting the optimal PSs at every time t from the space of all available $\{\Omega, \mathcal{F}\}$ pairs and updating its corresponding parameters with complexity order of $\mathcal{O}(|\Omega| + |\mathcal{F}| + 1)$. For the optimal solution, its computational complexity is of order $\mathcal{O}(|\Omega||\mathcal{F}|)$ as it full searches all available $\{\Omega, \mathcal{F}\}$ pairs. The computational complexity of the random selection comes from generating a random number in the range $\{1, |\Omega||\mathcal{F}|\}$, and based on it, a combination of $\mathbf{\Phi}_i, \boldsymbol{f}_j$ is selected with computational complexity order of $\mathcal{O}(1)$. Table 2 summarizes the time complexity comparisons among the schemes involved in the assessments.

As a numerical example, let $|\mathcal{F}| = 36$ and $|\Omega| = 64$, then BT and computational complexities of the optimal solution are of order $\mathcal{O}(2304)$. However, BT and computational complexities of the proposed MAB schemes will be $\mathcal{O}(1)$ and $\mathcal{O}(101)$, that is, about 99.96% and 96% reductions in BT and computational complexities are obtained. Thus, the proposed two-phase MAB approach has a near-optimal performance with a much lower BT and computational complexities.

TABLE 2 Complexity analysis of RIS-user association algorithms

Algorithm	BT complexity	Computational complexity
Optimal	$\mathcal{O}(\|\Omega\|\|\mathcal{F}\|)$	$\mathcal{O}(\|\Omega\|\|\mathcal{F}\|)$
Two-phase MOSS	$\mathcal{O}(1)$	$\mathcal{O}(\|\Omega\|\|\mathcal{F}\| + 1)$
Two-phase UCB	$\mathcal{O}(1)$	$\mathcal{O}(\|\Omega\|\|\mathcal{F}\| + 1)$
Random	$\mathcal{O}(1)$	$\mathcal{O}(1)$

5 | CONCLUSION

In this paper, the problem of RIS-aided mmWave communication was explored. The main issue of the mmWave-RIS system is to jointly optimize the PS vectors of both mmWave and RIS for maximizing the spectral efficiency at the intended receiver. To efficiently address this problem while overcoming the problem of CSI estimation, a two-phase MAB approach was proposed. Besides, an antenna codebook was suggested for both mmWave BS and RIS. Numerical simulations prove that the proposed scheme achieves almost 91% to 99% of the optimal performance under different simulation scenarios with about 99.96% and 96% reductions in BT and computational complexities, and it outperforms other benchmarks.

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Numbers JP19H04174 and JP21K14162, respectively.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILIBILITY STATEMENT

Data available on request due to privacy/ethical restrictions.

ORCID

Ehab Mahmoud Mohamed https://orcid.org/0000-0001-5443-9711

Basem M. Elhlawany https://orcid.org/0000-0002-5900-6541

REFERENCES

1. Mohamed, E.M., Sakaguchi, K., Sampei, S.: Wi-Fi coordinated WiGig concurrent transmissions in random access scenarios. IEEE Trans. Veh. Techn. 66(11), 10357–10371 (2017)
2. Abdelreheem, A., Mohamed, E.M., Esmaiel, H.: Location-based millimeter wave multi-level beamforming using compressive sensing. IEEE Commun. Lett. 22, 185–188 (2018)
3. ElMossallamy, M.A., et al.: Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities. IEEE Trans. Cogn. Commun. Network. 6(3), 990–1002 (2020)
4. Chen, Y., Wang, Y., Jiao, L.: Robust transmission for reconfigurable intelligent surface aided millimeter wave vehicular communications with statistical CSI. IEEE Trans. Wirel. Commun. 21(2), 928–944 (2021)

5. Lixin, L., et al.: Enhanced reconfigurable intelligent surface assisted mmwave communication: A federated learning approach. China Commun. 17(10), 115–128 (2020)
6. Taha, A., Alrabeiah, M., Alkhateeb, A.: Enabling large intelligent surfaces with compressive sensing and deep learning. IEEE Access 9, 44304–44321 (2021)
7. Liu, Y., et al.: Cascaded channel estimation for RIS assisted mmWave MIMO transmissions. IEEE Wirel. Commun. Lett. 10(9), 2065–2069 (2021)
8. Pradhan, C., et al.: Hybrid precoding design for reconfigurable intelligent surface aided mmWave communication systems. IEEE Wirel. Commun. Lett. 9(7), 1041–1045 (2020)
9. Shtaiwi, E., et al.: RIS-assisted mmWave channel estimation using convolutional neural networks. In: 2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), 2021, pp. 1–6, https://doi.org/10.1109/WCNCW49093.2021.9419974.
10. Jia, C., et al.: Machine learning empowered beam management for intelligent reflecting surface assisted MmWave networks. China Commun. 17(10), 100–114 (2020)
11. Jiang, Z.M., et al.: Intelligent reflecting surface aided dual-function radar and communication system. IEEE Syst. J. 16(1), 475–486 (2021)
12. Mohamed, E.M., Hashima, S., Hatano, K.: Energy aware multi-armed bandit for millimeter wave-based UAV mounted RIS networks. IEEE Wirel. Commun. Lett. 11(6), 1293–1297 (2022). https://doi.org/10.1109/LWC.2022.3164939
13. Khalili, A., et al.: Resource management for transmit power minimization in UAV-Assisted RIS HetNets supported by dual connectivity. IEEE Trans. Wirel. Commun. 21(3), 1806–1822 (2022). https://doi.org/10.1109/TWC.2021.3107306.
14. Audibert, J.-Y., Bubeck, S'ebastien: Minimax Policies for Adversarial and Stochastic Bandits. *COLT 2009* (2009).
15. Specifications for High Rate Wireless Personal Area Networks (WPANs) Amendment 2: Millimeter-Wave-Based Alternative Physical Layer Extension. IEEE 802.15.3c Standard, (Oct. 2009)
16. Jang, B., et al.: Q-learning algorithms: A comprehensive classification and applications. IEEE Access 7, 133653–133667 (2019)
17. Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multiarmed bandit problem. Machine Learn. 47(2), 235–256 (2002)
18. Francisco Valencia, I., Marcial-Romero, J., Valdovinos, R.: A comparison between UCB and UCB-Tuned as selection policies GGP. J. Intell. Fuzzy Sys. 36, 5073–5079 (2019)
19. Kaufmann, E., Korda, N., Munos, R.: Thompson sampling: An asymptotically optimal finite-time analysis. In: Algorithmic Learning Theory (ALT), pp. 199–213. Springer, Berlin/Heidelberg (2012)
20. Seldin, Y., et al.: Evaluation and analysis of the performance of the EXP3 algorithm in stochastic environments. Eur. Workshop Reinf. Learn. 24, 103–116 (2012)
21. Mohamed, E.M., et al.: Gateway selection in millimeter wave UAV wireless networks using multi-player multi-armed bandit. Sensors 20, 3947 (2020)
22. Mohamed, E.M., et al.: Sleeping contextual/non-contextual Thompson sampling MAB for mmWave D2D two-hop relay probing. IEEE Trans. Veh. Technol. 70(11), 12101–12112 (2021)
23. Mohamed, E.M.: WiGig access point selection using non-contextual and contextual multi-armed bandit in indoor environment. J. Ambient Intell. Humanized Comput. (2022), https://doi.org/10.1007/s12652-022-03739-7
24. Bubeck, S., Cesa-Bianchi, N.: Regret analysis of stochastic and nonstochastic multi-armed bandit problems. Found. Trends Mach. Learn. 5, 1–122 (2012)
25. Rappaport, T.S., et al.: Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications. IEEE Trans. Antenn. Propag. 61, 1850–1859 (2013)
26. de Araújo, G.T., de Almeida, A.L.F., Boyer, R.: Channel estimation for intelligent reflecting surface assisted MIMO systems: A tensor modeling approach. IEEE J. Sel. Topics Signal Process. 15(3), 789–802 (2021)
27. Lin, T., et al.: Channel estimation for IRS-assisted millimeter-wave MIMO systems: Sparsity-inspired approaches. IEEE Trans. Commun. 20(6), 4078–4092 (2022)
28. Mohamed, E.M., et al.: Relay probing for millimeter wave multi-hop D2D networks. IEEE Access 8, 30560–30574 (2020)

How to cite this article: 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. 1–8 (2022). https://doi.org/10.1049/cmu2.12474