

Article

Spectrum Allocation and User Scheduling Based on Combinatorial Multi-Armed Bandit for 5G Massive MIMO

Jian Dou ^{1,*}, Xuan Liu ¹, Shuang Qie ¹, Jiayi Li ¹ and Chaoliang Wang ²

¹ China Electric Power Research Institute, Beijing 100192, China; liuxuan@epri.sgcc.com.cn (X.L.); qieshuang@epri.sgcc.com.cn (S.Q.); lijiaiyi@epri.sgcc.com.cn (J.L.)

² State Grid Zhejiang Electric Power Co., Ltd., Hangzhou 311121, China; wang_chaoliang@zj.sgcc.com.cn

* Correspondence: doujian@epri.sgcc.com.cn

Abstract: As a key 5G technology, massive multiple-input multiple-output (MIMO) can effectively improve system capacity and reduce latency. This paper proposes a user scheduling and spectrum allocation method based on combinatorial multi-armed bandit (CMAB) for a massive MIMO system. Compared with traditional methods, the proposed CMAB-based method can avoid channel estimation for all users, significantly reduce pilot overhead, and improve spectral efficiency. Specifically, the proposed method is a two-stage method; in the first stage, we transform the user scheduling problem into a CMAB problem, with each user being referred to as a base arm and the energy of the channel being considered a reward. A linear upper confidence bound (UCB) arm selection algorithm is proposed. It is proved that the proposed user scheduling algorithm experiences logarithmic regret over time. In the second stage, by grouping the statistical channel state information (CSI), such that the statistical CSI of the users in the angular domain in different groups is approximately orthogonal, we are able to select one user in each group and allocate a subcarrier to the selected users, so that the channels of users on each subcarrier are approximately orthogonal, which can reduce the inter-user interference and improve the spectral efficiency. The simulation results validate that the proposed method has a high spectral efficiency.

Keywords: massive MIMO; spectrum allocation; user scheduling; multi-armed bandit



Citation: Dou, J.; Liu, X.; Qie, S.; Li, J.; Wang, C. Spectrum Allocation and User Scheduling Based on Combinatorial Multi-Armed Bandit for 5G Massive MIMO. *Sensors* **2023**, *23*, 7512. <https://doi.org/10.3390/s23177512>

Academic Editors: Wei Wang, Qiang Ye, Yiliang Liu, Yanyu Cheng and Matthew Hamilton

Received: 26 July 2023

Revised: 23 August 2023

Accepted: 28 August 2023

Published: 29 August 2023



Copyright: © 2023 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

The development of new applications requires that the data rate be significantly increased. However, a significant increase in data traffic represents a huge challenge to the scarce spectrum resources [1,2]. Massive MIMO has become a key technology in 5G communication system, which greatly improves the system capacity and reduces latency by equipping base stations (BSs) and/or users with a large number of antennas [3,4]. Millimeter wave technology has received significant attention due to its large bandwidth. The combination of millimeter wave and massive MIMO can effectively improve the bandwidth and spectrum utilization of the system [5,6]. On the one hand, the millimeter wavelength is very short, which is conducive to the deployment of a large number of antennas in the system. On the other hand, antenna arrays in massive MIMO systems can bring greater freedom to the system, and using simple linear precoding at the BS can enable the system to achieve a high transmission performance [7,8]. Thus, massive MIMO can compensate for the propagation loss of millimeter wave channels. Therefore, millimeter wave technology is very suitable for massive MIMO systems. In order to fully utilize the gains brought by millimeter wave massive MIMO, it is usually necessary to obtain the channel state information (CSI).

If instantaneous CSI is not available, then non-coherent (NC) techniques are considered key solutions for massive MIMO systems [9–12]. The designs of NC techniques for massive MIMO systems are mainly focused on energy detection [9,10] or phase detection [11,12].

Manolakos et al. proposed a soft-output detector where the transmitter modulated information during symbol amplitude determination, and the receiver obtained the average received energy [9]. Ngo et al. calculated the marginal posterior of each transmitted signal using the channel distribution and proposed a novel soft-output multi-user detector [10]. Baeza et al. proposed a new constellation design for a multi-user noncoherent large-scale single-input multiple-output uplink system based on M -ary differential phase-shift keying [11]. Additionally, non-orthogonal constellation-based schemes to reduce the complexity and increase the spectral efficiency were proposed in [12]. If instantaneous CSI is available, then massive MIMO systems are said to be coherent and have high spectral efficiency due to the available instantaneous CSI [10]. In such scenarios, the symbols are usually from a scalar constellation, such as the quadrature amplitude modulation. In this paper, we consider coherent massive MIMO systems.

The massive MIMO system can serve multiple users on a single time and frequency resource. There have been many studies on the design of transmission schemes for single-carrier massive MIMO systems. Due to the large bandwidth provided by millimeter waves, multiple subcarriers can be utilized to transmit information. How to allocate subcarriers to maximize the spectral efficiency of the system is worth studying. In orthogonal frequency-division multiplexing (OFDM) systems, there are multiple subcarriers to which the spectrum should be allocated. A simple way to achieve this is to randomly allocate the subcarriers to the users. However, this random scheme achieves a relatively low spectral efficiency. Zhang et al. considered a subcarrier allocation method in which the user moves, and each user is allocated one subcarrier [13]. In order to further improve the system's spectral efficiency, Anand et al. proposed allocating multiple subcarriers to each user to ensure service quality [14]. However, the above studies did not consider the user scheduling problem.

The number of users that a BS can serve in one slot is limited, the system cannot provide services for all candidate users when the number of users is large. Therefore, the user scheduling scheme plays a crucial role in the overall performance of the system [15]. In massive MIMO systems, a simple scheduling scheme is a random scheduling scheme; it randomly selects multiple users for communication. This method does not consider channel information and, therefore, demonstrates poor performance. The authors in [16] proposed a sub-optimal user scheduling algorithm, which performs QR decomposition on the channel and schedules based on the results of QR decomposition, not only reducing the computational complexity but also improving the spectral efficiency. Shehata et al. proposed an iterative method to select terminals in turn by using the eigenvalue and condition number of the channel matrix between the BS and terminals [17]. Bu et al. considered the user scheduling and spectral allocation problem in massive MIMO systems, and proposed a reinforcement learning-based algorithm to improve the spectral efficiency [18]. However, all of these methods require the instantaneous CSI of the terminal. When there are many terminals, it is difficult to obtain the instantaneous CSI of all terminals simultaneously, meaning that these methods are no longer applicable. In addition, due to the slower variation of the statistical CSI compared to instantaneous CSI, user scheduling through statistical CSI can avoid obtaining instantaneous CSI, reduce pilot overhead, and improve spectral efficiency [19]. However, the estimation of statistical CSI also requires a large overhead. When the terminal moves fast, the changes in statistical CSI will also become fast, causing the system performance to deteriorate.

This paper proposes a user scheduling and spectrum allocation scheme based on combinatorial multi-armed bandit (CMAB) for a multi-user millimeter wave massive MIMO OFDM system. The proposed CMAB-based scheme avoids the need for channel estimation for all users, which can reduce pilot overhead and improve spectral efficiency. Our approach utilizes the multi-path channel model in millimeter waves [20]. The proposed scheme is a two-stage scheme. In the first stage, we transform the user scheduling problem into a CMAB problem, and the upper confidence bound (UCB) strategy is used to train and obtain the scheduled users to avoid channel estimation for each user and to improve

the spectral efficiency. Each user is regarded as a base arm in the CMAB problem, and multiple users are regarded as the super arm. To improve the system's spectral efficiency, we aim to maximize the receiving energy. A linear UCB algorithm with low computational complexity is proposed. By setting the UCB value of each base arm, the UCB value of the super arm consists of a linear combination of the UCB value of each base arm, and we select the super arm to maximize the UCB value and the received energy. In addition, it is theoretically proven that the regret of the proposed scheme grows logarithmically over time. In the second stage, the instantaneous CSI estimated in previous slots is used to calculate the statistical CSI, and subcarrier allocation is performed based on the statistical CSI. By grouping the statistical CSI, such that the statistical CSI between different groups is approximately orthogonal, we are able to select one user in each group and allocate one subcarrier to the selected users, resulting in the users in each subcarrier being approximately orthogonal and, thus, improving the spectral efficiency. The simulation results show that the proposed method can significantly improve the spectral efficiency compared with previous methods.

2. Preliminaries

2.1. System Model

Consider a millimeter wave uplink massive MIMO OFDM system in a single cell with multiple users, where the BS is equipped with a uniform linear array of antennas serving multiple single-antenna users. The communication bandwidth B is divided into F subcarriers with subcarrier spacing $\frac{B}{F}$. This paper adopts a digital combiner design scheme at the BS [21]. Let $\mathbf{h}_{k,f,t}$ be the channel vector between the BS and user k in the t -th slot at the f -th subcarrier and $s_{k,f,t}$ be the transmitted signal from user k . Then the received signal at the BS at the f -th subcarrier is

$$\mathbf{y}_{f,t} = \mathbf{W}_{f,t} \sum_{k=1}^K \mathbf{h}_{k,f,t}^H s_{k,f,t} + \mathbf{n}_{f,t}, \quad (1)$$

where $\mathbf{W}_{f,t} \in C^{M \times 1}$ is the digital combiner in the t -th slot at the f -th subcarrier, and $\mathbf{n}_{f,t}$ is the noise. Each element in $\mathbf{n}_{f,t}$ follows a Gaussian distribution whose mean is zero and variance is σ^2 .

In this paper, we consider a multi-path channel model in millimeter wave communications, and the channel is narrowband time-varying [13]. The channel vector between the BS and user k is given by [21]

$$\mathbf{h}_{k,f,t} = \sum_{l=1}^{L_{k,f,t}} \rho_t(\theta_{k,f,l,t}) \mathbf{a}(\theta_{k,f,l,t}), \quad (2)$$

where $L_{k,f,t}$ is the number of paths of user k at the f -th subcarrier, $\rho_t(\theta_{k,f,l,t}) \sim CN(0, \sigma_{k,f,t}^2 \mathbf{I})$ is the complex path gain, $\mathbf{a}(\theta_{k,f,l,t}) \in C^{M \times 1}$ is the steering vector, and the i -th element of $\mathbf{a}(\theta_{k,f,l,t})$ is $e^{-j2\pi\lambda \frac{i-1}{d} \cos \theta}$. λ is the wavelength, $d = \lambda/2$ is the antenna spacing, and $L_{k,f,t}$ is an integer uniformly distributed in $\{1, 2, \dots, L_{max}\}$. The power gain $\sigma_{k,f,t}^2$ is normalized in $[0, 1]$. Usually, the statistical CSI remains unchanged for a long time. Suppose that the statistical CSI remains unchanged in T slots; this means that the angles $\theta_{k,f,l,t}$ remain unchanged in T slots, and the covariance $\sigma_{k,f,t}^2$ also remains unchanged. Then, in T slots, the channel vector in (2) can be transformed into

$$\mathbf{h}_{k,f,t} = \sum_{l=1}^{L_{k,f}} \rho_t(\theta_{k,l}) \mathbf{a}(\theta_{k,l}), \quad (3)$$

where $\rho_t(\theta_{k,l}) \sim CN(0, \sigma_k^2 \mathbf{I})$, and σ_k^2 is normalized in $[0, 1]$.

In the user scheduling problem, the conventional schemes design the user scheduling based on the instantaneous CSI. Then, it is necessary for the BS to obtain the instantaneous CSI of all users in real time. Due to the large number of users, it takes a lot of time to estimate the channel, which will lead to a decrease in the system performance. This paper proposes a method based on multi-armed bandit for user scheduling and proposes a sub-carrier allocation method based on statistical CSI. In the process of spectrum allocation and user scheduling, this paper assigns each user a subcarrier and serves U users on each subcarrier. Key notations used in this work are listed in Table 1.

Table 1. Key notations.

Notations	Parameters
B	Bandwidth
F	The number of subcarriers
$\mathbf{h}_{k,f,t}$	The channel vector between the BS and user k in slot t at subcarrier f
$s_{k,f,t}$	The transmitted signal from user k in slot t at subcarrier f
$\mathbf{W}_{f,t}$	The digital combiner in slot t at subcarrier f
$\mathbf{n}_{f,t}$	Gaussian noise vector in slot t at subcarrier f
$L_{k,f,t}$	The number of paths of user k at the f -th subcarrier
$\theta_{k,f,l,t}$	The angle of the l -th path of user k in slot t at subcarrier f
ρ	The complex path gain
$\mathbf{a}(\cdot)$	The steering vector
λ	the wavelength
d	Antenna spacing
σ^2	The variance of the Gaussian noise
$\sigma_{k,f,t}^2$	The power gain of user k in slot t at subcarrier f
σ_k^2	The power gain of user k
$\theta_{k,l}$	The angle of the l -th path of user k in the t -th slot
A_i	The action
$Q_t(\cdot)$	The value of the action in slot t
R_i	The reward in slot i
$\mathcal{A}_{f,t}$	The set of selected users in slot t at subcarrier f
$\tilde{\mathbf{H}}_f(t)$	The channel matrix of the selected users in slot t at subcarrier f
\mathcal{A}_t	The set of users selected in slot t
\mathbf{X}_f	The pilot signal at subcarrier f
\mathbf{Y}_f	The received signal at subcarrier f
\mathbf{N}_f	The noise matrix in slot t at subcarrier f
$u_i(t)$	The set of users selected in slot t
$\bar{u}_{\mathcal{A}_t}$	The UCB value of the super arm \mathcal{A}_t
$\bar{R}_{i,t}$	The mean reward of action i slot t
\mathbf{D}_f	The discrete Fourier transform matrix at subcarrier f
$\hat{\mathbf{h}}_{k,f,t}^a$	The channel vector $\hat{\mathbf{h}}_{k,f,t}$ in the angular domain

Table 1. Cont.

Notations	Parameters
\mathbf{R}_k^a	The statistical CSI on the angular domain
$d_c(\mathbf{X}, \mathbf{Y})$	The chord distance between \mathbf{X} and \mathbf{Y}
$\mathbf{U}_X(\mathbf{U}_Y)$	The eigenvectors corresponding to the non-zero eigenvalues of $\mathbf{X}(\mathbf{Y})$

2.2. The Description of the CMAB

The MAB problem [22] revolves around selecting actions with the goal of maximizing rewards and expectations for the selected actions. After each action is executed, a reward is obtained, and the expectation of the reward is called the value of the action. The value of the action at the t -th slot is $Q_t(A_i)$, which can be calculated as

$$Q_t(A_i) = \frac{\sum_{i=1}^{t-1} R_i \cdot \mathbf{1}_{A_i=1}}{\sum_{i=1}^{t-1} \mathbf{1}_{A_i=1}}, \quad (4)$$

where R_i denotes the reward of the action A_i in the i -th slot. When the action A_i at the i -th slot is chosen, then $\mathbf{1}_{A_i=1}$ is 1, otherwise, $\mathbf{1}_{A_i=1}$ is 0. Then, it is possible to evaluate the quality $Q_t(A_i)$ of actions and develop strategies for selecting actions, including ϵ greedy strategies, UCB strategies, etc. [22]. Unlike the MAB problem, in the CMAB problem [23], multiple actions need to be selected each time, and the reward expectation for multiple actions needs to be maximized. Each action is called a base arm, while the multiple actions selected each time are called a super arm. A super arm consists of multiple base arms.

3. Two-Stage Spectrum Allocation and User Scheduling Scheme

3.1. The User Scheduling Problem Formulation

In each time slot, we select UF users among K users for communication. Each sub-carrier serves U users, and there are F subcarriers in total. The purpose is to enable the selected users to obtain a large spectral efficiency. The spectral efficiency is related to the signal-to-interference-plus-noise ratio (SINR). The received energy is composed of the energy of the received useful signal and the energy of the Gaussian noise. By maximizing the received energy, the energy of the received useful signal will be increased, resulting in improved SINR and spectral efficiency. It should be noted that the received energy is affected by the channel energy. The larger the channel energy, the greater the received energy. Therefore, in order to improve the spectral efficiency of the system, in this paper, we hope that the channel energy of the selected UF users is large. Assuming that the set of selected users in the t -th slot at the subcarrier f is $\mathcal{A}_{f,t}$, let $\tilde{\mathbf{H}}_f(t)$ be the channel matrix of the selected users in the t -th slot at the subcarrier f . We define $\|\tilde{\mathbf{H}}_f(t)\|_F^2$ as the equivalent channel energy of the selected users. The problem of maximizing the channel energy can be formulated as

$$\max \sum_{t=1}^T \sum_{f=1}^F \|\tilde{\mathbf{H}}_f(t)\|_F^2 \quad (5)$$

In the above problem, due to the large number of users, it is difficult to estimate the instantaneous CSI of all users simultaneously. Without the CSI of all terminals, traditional optimization-based scheduling schemes are not feasible. This paper adopts a method based on a CMAB to select users.

We first transform the user scheduling problem into a CMAB problem. We define K users as K base arms of the CMAB problem, denoted as $1, 2, \dots, K$. Each user (base arm) has a channel energy. As shown in Section 2, the CSI of user k corresponding to the subcarrier f at the t -th slot is $\mathbf{h}_{k,f,t}$, and let $\|\mathbf{h}_{k,f,t}\|_2^2$ be the reward of the base arm k . In each time slot, the selected users on all subcarriers are marked as a super arm, which is denoted as \mathcal{A}_t . At this point, the reward for the super arm \mathcal{A}_t is

$$r_{\mathcal{A}_t} = \sum_{f=1}^F \left\| \tilde{\mathbf{H}}_f(t) \right\|_F^2 = \sum_{k \in \mathcal{A}_t} \left\| \mathbf{h}_{k,f,t} \right\|_2^2. \quad (6)$$

We select the users in each time slot based on a strategy to maximize the total reward. Assuming that the set of users selected for the t -th slot is \mathcal{A}_t , and the set of the users selected on the subcarrier f is $\mathcal{A}_{f,t}$, we have $\mathcal{A}_t = \bigcup_f \mathcal{A}_{f,t}$.

When the users scheduled in each subcarrier are obtained, the coherent transmission scheme is shown as follows. Due to the lack of interference between signals on different subcarriers, when estimating the channel, the pilot signal sent by each user on the subcarrier f to the base station is \mathbf{X}_f . Therefore, the signal received at the base station on the subcarrier f in the t -th slot is

$$\mathbf{Y}_f(t) = \tilde{\mathbf{H}}_f(t)\mathbf{X}_f + \mathbf{N}_f(t), \quad (7)$$

where $\mathbf{N}_f(t)$ is the noise matrix in the t -th slot at subcarrier f , and each element of $\mathbf{N}_f(t)$ is a Gaussian noise with zero mean and the variance is σ^2 . Then we use the zero forcing criterion to estimate the instantaneous CSI. By multiplying (7) with \mathbf{X}_f^H , we can obtain

$$\mathbf{Y}_f(t)\mathbf{X}_f^H = \tilde{\mathbf{H}}_f(t) + \mathbf{N}_f(t)\mathbf{X}_f^H. \quad (8)$$

Then the estimation of $\tilde{\mathbf{H}}_f(t)$ is $\hat{\mathbf{H}}_f(t) = \mathbf{Y}_f(t)\mathbf{X}_f^H$. After the channel estimation is obtained, the MMSE detection operation is adopted for the channels on each subcarrier, so as to reduce the interference and improve the spectral efficiency of the system. When using minimum mean square error (MMSE) detection, the combining matrix $\mathbf{W}_f(t)$ is

$$\mathbf{W}_f(t) = \left(\hat{\mathbf{H}}_f^H(t)\hat{\mathbf{H}}_f(t) + \sigma^2\mathbf{I} \right)^\dagger \hat{\mathbf{H}}_f^H(t), \quad (9)$$

where $(\cdot)^\dagger$ is the pseudo-inverse operation. The spectral efficiency of the system is

$$R = \sum_{f=1}^F \sum_{u=1}^U \log_2 \left(1 + \frac{\left| \tilde{\mathbf{h}}_{f,u}^H(t)\mathbf{w}_{f,u}(t) \right|^2}{\sum_{l \neq u} \left| \tilde{\mathbf{h}}_{f,u}^H(t)\mathbf{w}_{f,l}(t) \right|^2 + \sigma^2} \right), \quad (10)$$

where $\mathbf{w}_{f,u}$ is the u -th column of $\mathbf{W}_f(t)$, $\tilde{\mathbf{h}}_{f,u}(t)$ is the u -th column of $\tilde{\mathbf{H}}_f(t)$, and σ^2 is the variance of the Gaussian noise.

3.2. Linear UCB User Scheduling Algorithm

From above, it can be seen that the channels of each user cannot be obtained simultaneously, so traditional optimization methods using the statistical CSI cannot be used. We propose the CMAB method for user scheduling. For the above model, it can be seen that in the multi-path channel model, the reward of user k (basic arm k) is

$$\left\| \mathbf{h}_{k,f,t} \right\|_2^2 = \sum_{l=1}^{L_{k,t}} \left| \rho_t(\theta_{k,l}) \right|^2, \quad (11)$$

where $\rho_t(\theta_{k,l}) \sim \text{CN}(0, \sigma_{k,l}^2)$ is a Chi-squared distributed random variable with the degree of freedom 1, and it also follows a sub-exponential distribution with parameter $(\sigma_{k,l}^2, 4\sigma_{k,l})$ [24]. According to the characteristics of sub-exponential distribution,

$\sum_{l=1}^{L_{k,t}} \left| \rho_t(\theta_{k,l}) \right|^2$ is also a sub-exponential distribution with parameter $\left(\sum_{l=1}^{L_{k,t}} \sigma_{k,l}^2, \max_l 4\sigma_{k,l} \right)$.

From Section 2, $\sigma_{k,l}^2$ is normalized in $[0, 1]$.

Due to the large number of user combinations, the number of super arms grows exponentially with the number of users. In order to reduce the complexity of the

algorithm—considering that the UCB strategy can provide a better performance—we propose a linear UCB strategy. This strategy has a complexity that is linear with the total number of users. Since the reward for the super arm is the linear sum of each combination arm, in the linear UCB strategy, the UCB value of the combination arm is set to the sum of the UCB values of the base arm. Considering that the reward of each arm obeys sub-exponential distribution, the UCB value of each arm is defined as

$$u_i(t) = \begin{cases} \bar{R}_{i,t-1} + \bar{c} \max\left\{\frac{16(U+1)\ln(t)}{m_{i,t-1}}, \sqrt{\frac{2L_k(U+1)\ln(t)}{m_{i,t-1}}}\right\} & m_{i,t-1} \neq 0, \\ \infty & m_{i,t-1} = 0, \end{cases} \quad (12)$$

where $\bar{c} = \max \sigma_{k,f,t}^2$, $m_{i,t}$ denotes the number of the action i that has been selected, and $\bar{R}_{i,t} = \frac{\bar{R}_{i,t-1}m_{i,t-1} + r_{i,t-1}}{m_{i,t-1} + 1}$ denotes the mean reward. From $\sigma_{l,k}^2 \in [0, 1]$, we have $\bar{c} = 1$. Then we define the UCB value of the super arm \mathcal{A}_t as

$$\bar{u}_{\mathcal{A}_t} = \sum_{i \in \mathcal{A}_t} u_i(t). \quad (13)$$

It can be seen that the UCB values of each super arm can be calculated based on the UCB values of the base arm, greatly reducing the computational complexity. In the t -th slot, we select \mathcal{A}_t to maximize the UCB value of the super arm, and the problem can be described as

$$\max_{\mathcal{A}_t} \bar{u}_{\mathcal{A}_t} = \max_{\mathcal{A}_t} \sum_{u \in \mathcal{A}_t} u_i(t). \quad (14)$$

The above problem is an unconstrained discrete optimization problem. In order to obtain the optimal solution, we calculate the UCB values for all actions, and then sort the UCB value $u_i(t)$. Finally, the users corresponding to the FU -largest UCB values are selected for communication. The user scheduling algorithm is shown in Algorithm 1.

Algorithm 1: The CMAB-based user scheduling algorithm

Input: $m_{i,t-1}, \bar{R}_{i,t-1}$;

- 1: Use (12) to calculate the UCB values of all the base arms;
- 2: Sort the UCB values of all the base arms;
- 3: Select the users corresponding to the FU -largest UCB values; the set of these users are \mathcal{A}_t ;
- 4: For each action $a \in \mathcal{A}_t$, set $m_{a,t} = m_{a,t-1} + 1$ and $\bar{R}_{a,t} = \frac{\bar{R}_{a,t-1}m_{a,t-1} + r_{a,t-1}}{m_{a,t-1} + 1}$;

Output: \mathcal{A}_t .

Theorem 1. *The regret of the proposed linear UCB scheduling algorithm is upper-bounded by $O(\ln t)$.*

Proof. Firstly, we define a variable $B_{i,t}$ to record the number of times the base arm i has been selected in the t -th slot. Considering the t -th slot, if the optimal combination arm \mathcal{A}^* is selected, the value $B_{i,t}$ will not change. If a non-optimal super arm \mathcal{A}_t is selected, then $B_{i,t}$ is added by 1 in the t -th slot, where $i = \min_{j \in \mathcal{A}_t} m_{j,t}$. It is easy to prove that in the t -th slot, the number of times a non-optimal combination arm is selected is $\sum_{i=1}^M B_{i,t}$. Then, we define a variable $V_{i,t}$. If the value $B_{i,t}$ at the t -th slot is increased by 1, $V_{i,t}$ is 1. If the value $B_{i,t}$ remains unchanged, then $V_{i,t}$ is set to 0. Let l be a positive integer, then we have

$$B_{i,t} = \sum_{p=1}^t \mathbf{1}\{V_{i,p} = 1\} \leq l + \sum_{p=1}^t \mathbf{1}\{V_{i,p} = 1, B_{i,p-1} \geq l\}, \quad (15)$$

where $\mathbf{1}\{a\}$ is a function. If event a is true, then $\mathbf{1}\{a\} = 1$. When event a is false, then $\mathbf{1}\{a\} = 0$. If $V_{i,p} = 1$, a non-optimal action is selected, and

$$B_{i,t} \leq l + \sum_{p=1}^t \mathbf{1} \left\{ \sum_{j \in \mathcal{A}^*} \bar{R}_{j,p-1} + v_{j,p-1} \leq \sum_{j \in \mathcal{A}_p} \bar{R}_{j,p-1} + v_{j,p-1}, B_{i,p-1} \geq l \right\} \leq l + \sum_{p=1}^t \mathbf{1} \left\{ \sum_{j \in \mathcal{A}^*} \bar{R}_{j,p} + v_{j,p} \leq \sum_{j \in \mathcal{A}_{p+1}} \bar{R}_{j,p} + v_{j,p}, B_{i,p} \geq l \right\}, \quad (16)$$

where

$$v_{j,p} = \bar{c} \max \left\{ \frac{16(M+1) \ln(p)}{m_{j,p}}, \sqrt{\frac{2K(M+1) \ln(p)}{m_{j,p}}} \right\}. \quad (17)$$

From $|\mathcal{A}^*| = U$, $|\mathcal{A}_{p+1}| = U$, and $l \leq B_{i,p} \leq m_{i,p}$, we have

$$\begin{aligned} & \sum_{p=1}^t \mathbf{1} \left\{ \min_{0 < m_{h_1,t_1}, m_{g_2,t_2}, \dots, m_{h_{Q^*},t_{Q^*}} \leq p} \sum_{j=1}^{Q^*} \bar{R}_{h_j,t_j} + v_{h_j,t_j} \right. \\ & \leq \max_{l \leq m_{g_1,t_1}, m_{g_2,t_2}, \dots, m_{g_{Q^p},t_{Q^p}} \leq p} \sum_{j=1}^{Q^p} \bar{R}_{g_j,t_j} + v_{g_j,t_j} \left. \right\} \\ & \leq l + \sum_{p=1}^{\infty} \sum_{m_{h_1,t_1}=1}^p \dots \sum_{m_{h_{Q^*},t_{Q^*}}=1}^p \sum_{m_{g_1,t_1}=l}^p \dots \sum_{m_{g_{Q^p},t_{Q^p}}=l}^p \\ & \mathbf{1} \left\{ \sum_{j=1}^{Q^*} \bar{R}_{h_j,t_j} + v_{h_j,t_j} \leq \sum_{j=1}^{Q^p} \bar{R}_{g_j,t_j} + v_{g_j,t_j} \right\}. \end{aligned} \quad (18)$$

If $\sum_{j=1}^{Q^*} \bar{R}_{h_j,t_j} + v_{h_j,t_j} \leq \sum_{j=1}^{Q^p} \bar{R}_{g_j,t_j} + v_{g_j,t_j}$, then one of the following three events must be satisfied.

$$\varepsilon_1 : \sum_{j=1}^{Q^*} \bar{R}_{h_j,t_j} \leq R_{\mathcal{A}^*} - \sum_{j=1}^{Q^*} v_{h_j,t_j}, \quad (19)$$

$$\varepsilon_2 : \sum_{j=1}^{Q^p} \bar{R}_{g_j,t_j} \geq R_{\mathcal{A}_{p+1}} + \sum_{j=1}^{Q^p} v_{g_j,t_j}, \quad (20)$$

$$\varepsilon_3 : R_{\mathcal{A}^*} < R_{\mathcal{A}_{p+1}} + \sum_{j=1}^{Q^p} v_{g_j,t_j} + \sum_{j=1}^{Q^*} v_{h_j,t_j}, \quad (21)$$

where $R_{\mathcal{A}^*} = \sum_{j=1}^{Q^*} R_{h_j}$, $R_{\mathcal{A}_{p+1}} = \sum_{j=1}^{Q^p} R_{g_j}$ and R_i is the mean reward of the action i . For the event ε_1 , we have

$$P\{\mathcal{E}_1\} \leq \sum_{j=1}^{Q^*} P\left\{ \bar{R}_{h_j,t_j} \leq R_{h_j} - v_{h_j,t_j} \right\}. \quad (22)$$

Since $\|\mathbf{h}_{k,f,t}\|_2^2$ follows a sub-exponential function with parameter $\left(\sum_{l=1}^{L_k} \sigma_{k,l}^2, \max_l 4\sigma_{k,l} \right)$.

Using the Hoeffding inequality, we have

$$P\left\{ \bar{R}_{h_j,t_j} \leq R_{h_j} - c_{h_j,t_j} \right\} \leq 2e^{-\min\left\{ \frac{m_{h_j,t_j}^2 c_{h_j,t_j}^2}{\tau^2}, \frac{m_{h_j,t_j} c_{h_j,t_j}}{8\tau} \right\}}. \quad (23)$$

From the above, we have $c_{h_j,t_j} \geq \bar{c} \sqrt{\frac{2K(M+1)\ln t}{m_{h_j,t_j}}}$, and then $e^{-\frac{m_{h_j,t_j} c_{h_j,t_j}^2}{\bar{c}^2}} \leq t^{-\frac{2(M+1)Kc^2}{\bar{c}^2}}$. Since $Kc^2 \geq \bar{c}^2$, we can obtain $e^{-\frac{m_{h_j,t_j} c_{h_j,t_j}^2}{\bar{c}^2}} \leq t^{-2(M+1)}$, $c_{h_j,t_j} \geq \bar{c} \frac{16(M+1)\ln(t-1)}{m_{i,t-1}}$ and $e^{-\frac{m_{h_j,t_j} c_{h_j,t_j}^2}{8\bar{c}^2}} \leq t^{-\frac{2(M+1)\bar{c}}{\bar{c}}}$. Thus we have $e^{-\frac{m_{h_j,t_j} c_{h_j,t_j}^2}{\bar{c}^2}} \leq t^{-2(M+1)}$, and

$$(P\{\mathcal{E}1\}) \leq Mt^{-2(M+1)}. \quad (24)$$

Similarly, we have $(P\{\mathcal{E}2\}) \leq Mt^{-2(M+1)}$. Let $\Delta_{\mathbf{a}(t+1)} = R_{\mathbf{a}^*} - R_{\mathbf{a}(t+1)}$, and

$$l \geq \max \left\{ \left\lceil \frac{32\bar{c}(M+1)\ln t}{\frac{\Delta_{\mathbf{a}(t+1)}}{M}} \right\rceil, \left\lceil \frac{8\bar{c}(M+1)\ln t}{\left(\frac{\Delta_{\mathbf{a}(t+1)}}{M}\right)^2} \right\rceil \right\}. \quad (25)$$

Then we have

$$\sum_{j=1}^{Q^p} \frac{16\bar{c}(M+1)\ln t}{m_{h_j,t_j}} \leq M \frac{16\bar{c}(M+1)\ln t}{l} \leq M \frac{\bar{c}(M+1)\ln t}{2\bar{c}(M+1)\ln t} \frac{\Delta_{\mathbf{a}(t+1)}}{M} = \frac{\Delta_{\mathbf{a}(t+1)}}{2}, \quad (26)$$

and

$$\sum_{j=1}^{Q^p} \sqrt{\frac{2\bar{c}(M+1)\ln t}{m_{i,t}}} \leq M \sqrt{\frac{2\bar{c}(M+1)\ln t}{l}} \leq M \sqrt{\frac{2\bar{c}(M+1)\Delta_{\mathbf{a}(t+1)}^2 \ln t}{8\bar{c}(M+1)M^2 \ln t}} = \frac{\Delta_{\mathbf{a}(t+1)}}{2}. \quad (27)$$

Then we can obtain $\sum_{j=1}^{Q^p} c_{g_j,t_j} \leq \frac{\Delta_{\mathbf{a}(t+1)}}{2}$. Meanwhile, we can obtain $\sum_{j=1}^{Q^*} c_{h_j,t_j} \leq \frac{\Delta_{\mathbf{a}(t+1)}}{2}$.

Thus,

$$R_{\mathbf{a}^*} - R_{\mathbf{a}(p+1)} - \sum_{j=1}^{Q^p} c_{h_j,t_j} - \sum_{j=1}^{Q^*} c_{h_j,t_j} \geq 0. \quad (28)$$

Then, we have $P\{\mathcal{E}3\} = 0$. Note that

$$\sum_{p=1}^{\infty} \sum_{m_{h_1,t_1}=1}^p \cdots \sum_{m_{h_{Q^*,t_{Q^*}}}=1}^p \sum_{m_{g_1,t_1}=l}^p \cdots \sum_{m_{g_{Q^p,t_{Q^p}}}=l}^p 2Mp^{-2(M+1)} \leq \sum_{p=1}^{\infty} 2Mp^{-2} \leq 1 + \frac{\pi^2}{3}M. \quad (29)$$

Thus, $(Z_{i,t}) \leq l + 1 + \frac{\pi^2}{3}M = O(\ln t)$, and then

$$\text{Regret} \leq \Delta_{\max} \sum_{i=1}^M (Z_{i,t}) = O(\ln t), \quad (30)$$

where $\Delta_{\max} = \max_{\mathbf{a}} \Delta_{\mathbf{a}}$. \square

3.3. Spectrum Allocation Algorithm Based on Statistical CSI Grouping

After obtaining the scheduled users, it is necessary to further allocate the subcarriers to the scheduled users. The goal of subcarrier allocation is to maximize the spectral efficiency. We hope that the user channels on the same subcarrier are approximately orthogonal to each other. If the channels between users are approximately orthogonal, then the inter-user interference will be low, and the spectral efficiency of the users will increase.

Since the instantaneous CSI cannot be obtained, we use the statistical CSI to allocate the subcarriers. Due to the estimation of each user's channel in the previous slots, the statistical CSI of the users can be calculated based on the instantaneous CSI in the previous slots. Using the statistical CSI to allocate the subcarriers greatly reduces the user's overhead

and improves the spectral efficiency. Suppose the estimation of user k 's channel vector in the t -th slot at subcarrier f is $\hat{\mathbf{h}}_{k,f,t}$. If this is the case, then its angular domain expression can be obtained as

$$\hat{\mathbf{h}}_{k,f,t}^a = \mathbf{D}_f \hat{\mathbf{h}}_{k,f,t} \quad (31)$$

where \mathbf{D}_f is the discrete Fourier transform matrix at the f -th subcarrier. This is due to the fact that one user's channel vector is different on the different subcarriers but the angular domain representation of one user's channel vector on the different subcarriers are the same. In order to better group the statistical CSI, the calculation method for the statistical CSI on the angular domain (CSIA) is defined as

$$\mathbf{R}_k^a = \frac{1}{t} \sum_{l=1}^t \hat{\mathbf{h}}_{k,f,t}^a \left(\hat{\mathbf{h}}_{k,f,t}^a \right)^H. \quad (32)$$

After obtaining the CSIA of each user, the next step is to design a subcarrier allocation method. When allocating subcarriers, it is desired that the user channels on each subcarrier are approximately orthogonal. As a result, the serving users are divided into groups, where the channel vectors in different groups are approximately orthogonal. We select one user in each group to obtain one user as the service user on a subcarrier; thus, the users on each subcarrier are more orthogonal, which can lead to higher system spectral efficiency. Considering that the K-means method has good performance in user grouping, this article uses the K-means method to group user statistics into CSI groups. Before using the K-means algorithm, we define a chord distance between \mathbf{X} and \mathbf{Y} as

$$d_c(\mathbf{X}, \mathbf{Y}) = \left\| \mathbf{U}_X \mathbf{U}_X^H - \mathbf{U}_Y \mathbf{U}_Y^H \right\|_F^2, \quad (33)$$

where \mathbf{U}_X and \mathbf{U}_Y denote the eigenvectors corresponding to the non-zero eigenvalues of \mathbf{X} and \mathbf{Y} , respectively. Then we use an improved K-means algorithm to group the statistical CSI. The difference between the improved K-means algorithm and the K-means algorithm is that the number of users in each group in our algorithm is set to U . The details of the spectrum allocation based on the improved K-means grouping algorithm are shown in Algorithm 2.

Algorithm 2: The spectrum allocation based on the improved K-means grouping method

Input: $\mathcal{A}_t, \hat{\mathbf{h}}_{k,f,t'}, t' < t$;

- 1: Use (31) and (32) to obtain the CSIA for each user;
 - 2: Randomly select U users (u_1, u_2, \dots, u_U) from the set \mathcal{A}_t ;
 - 3: **repeat**
 - 4: Compute the chordal distance $d_c(\mathbf{R}_i^a, \mathbf{R}_l^a)$ between all users with the selected U users ($i = 1, \dots, UF, l = u_1, \dots, u_U$);
 - 5: **for** $l = 1:U$ **do**
 - 6: Select $F - 1$ users from the remaining users that are closest to u_l and the index set of these users is \mathcal{S}_l ;
 - 7: Calculate the center point \mathbf{R}_l^a of the statistical CSIA of the users in \mathcal{S}_l ;
 - 8: **end for**
 - 9: **until** Convergence
 - 10: **for** $f = 1:F$ **do**
 - 11: Select one user from each group and allocate these users to the f -th subcarrier, the index sets of these users are $\mathcal{A}_{f,t}, f = 1, 2, \dots, F$;
 - 12: **end for**
- Output:** $\mathcal{A}_{f,t}$.
-

4. Simulation Results

We validate the performance of the proposed method through simulation results in this section. Considering the millimeter wave communication scenario, it is assumed that there are a total of K mobile users in the cell. The base station is equipped with M antennas, and the maximum number P_{max} of paths is set to 4. The number of subcarriers is $F = 128$. The path angles of each user are randomly distributed between $[0, 180^\circ]$. The subcarrier spacing is 30 kHz, the number of symbols in each slot is 14, and the slot time is 0.5 ms. Due to the mobile users, the statistical CSI coherence time length is set to 1000 slots (1000 time slots correspond to a range of 5 m at a speed of 10 m/s). After 1000 time slots, the statistical CSI of the users change; that is, the path angle and gain of each user change. We compare the proposed CMAB-based method with statistical CSI-based methods [9], random spectrum allocation, and user scheduling methods. In the method based on statistical CSI, it is assumed that the number of time slots used to estimate statistical CSI is 200, and user scheduling is carried out in the next 800 time slots.

Figures 1–3 show the spectral efficiency of the proposed method under different signal-to-noise ratios (SNRs). The number of antennas at BS in Figures 1 and 2 is 64, and the number of antennas at BS in Figure 3 is 128. The total number of users is $K = 2000$. The number of scheduled users in Figure 1 at each subcarrier is 4; that is, the total number of scheduled users in Figure 1 is 512. The number of scheduled users at each subcarrier in Figures 2 and 3 is 8. Figures 1–3 show that the CMAB-based method proposed in this paper has a higher spectral efficiency than the statistical CS-based method and the random method. For the scheduling method based on statistical CSI, estimating the channel covariance matrix (CCM) incurs a large pilot overhead, leading to poor average spectral efficiency. The method proposed in this paper does not estimate the statistical CSI and can improve the spectral efficiency. In addition, compared to the random scheduling scheme, the proposed CMAB-based scheme accounts for partial CSI, providing a better channel quality and, thus, improving the spectral efficiency of the system.

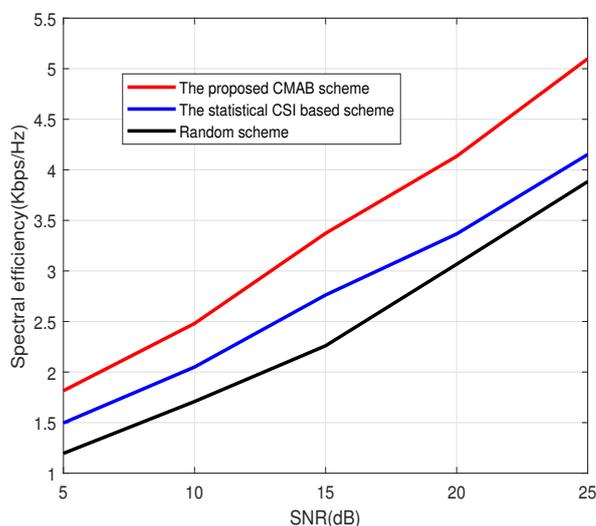


Figure 1. Spectral efficiency vs. SNR. $M = 64$, $U = 4$.

Figure 4 depicts the average spectral efficiency under different numbers of scheduled users per subcarrier. The number of BS antennas is set to 64, and the signal-to-noise ratio is set to 25 dB. The total number of users is 6000. From Figure 4, it can be observed that the method based on CMAB proposed in this paper performs better than the method based on statistical CSI and the random method under a different number of users. In addition, as the number of users increases, the spectral efficiency of all algorithms will increase first and then decrease. When the number of users is small, the inter-user interference is small, and the spectral efficiency will increase as the number of users increases due to user gain. When the number of users is large, the interference between users will increase. Because

the number of users is large, user interference will play a major role, and the spectral efficiency of the system will decrease due to the large inter-user interference. Therefore, when the number of users is large, the spectral efficiency of the system will decrease with the increasing number of users.

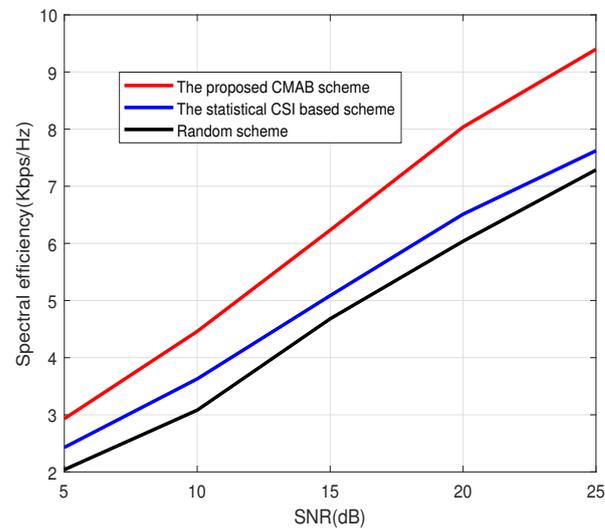


Figure 2. Spectral efficiency vs. SNR. $M = 64$, $U = 8$.

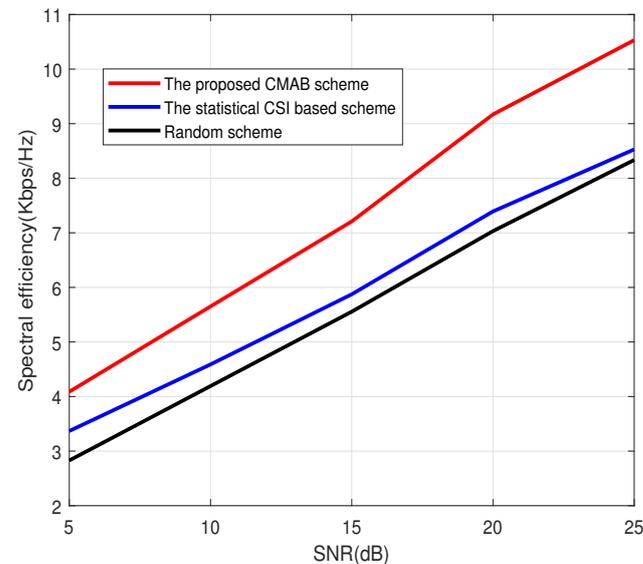


Figure 3. Spectral efficiency vs. SNR. $M = 128$, $U = 8$.

Figure 5 shows the spectral efficiency of the proposed CMAB-based user scheduling and subcarrier allocation scheme in each time slot. The number of base station antennas is 64, and the signal-to-noise ratio is set to 25 dB. The total number of terminals is 2000, and the user numbers scheduled per subcarrier are four and eight. The spectral efficiency of the proposed CMAB scheme increases first and then converges. The convergence speed when the number of users is four is faster than the convergence speed when the number of users is eight. This is because the smaller the number of users, the less time needed to explore. When the number of terminals is large, the spectral efficiency of the system is improved due to user gain, which corresponds to Figure 3; that is, eight users per subcarrier have higher spectral efficiency than four users per subcarrier.

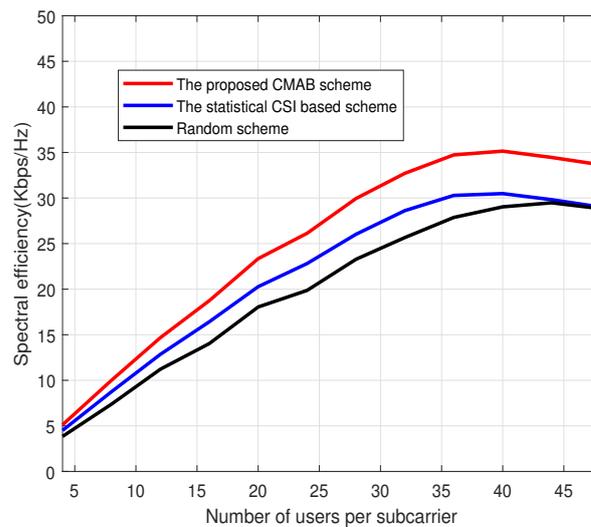


Figure 4. Spectral efficiency vs. users. $M = 64$.

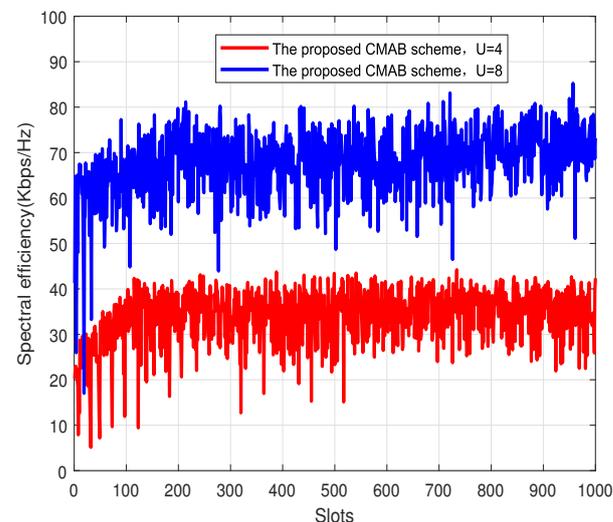


Figure 5. Spectral efficiency vs. slots. $M = 64$.

5. Conclusions

In this paper, we propose a spectrum allocation and user scheduling scheme based on CMAB for millimeter wave massive MIMO systems. This scheme uses CMAB to schedule the users and uses the statistical CSI to allocate the subcarriers, which can effectively improve the system's spectral efficiency. Simulation results show that the proposed CMAB-based scheme can effectively improve the spectral efficiency compared with the existing schemes. In future work, we will consider downlink massive MIMO systems. Considering that channel reciprocity does not hold in frequency-division duplex (FDD) massive MIMO systems, a downlink CSI cannot be obtained using the uplink pilot. Thus, designing a transmission scheme for downlink in FDD massive MIMO systems is more challenging. Moreover, it should be noted that a hybrid architecture (including the digital and analog parts) can reduce hardware complexity; due to this, we will design a user scheduling and spectrum allocation method in a downlink FDD massive MIMO system with hybrid architecture.

Author Contributions: Conceptualization, J.D. and X.L.; methodology, J.D. and X.L.; software, J.D.; validation, X.L., S.Q. and J.L.; formal analysis, S.Q. and J.L.; writing—original draft, J.D.; writing—review and editing, X.L., S.Q. and J.L.; visualization, J.L.; supervision, C.W.; project administration, C.W.; funding acquisition, C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the State Grid Corporation of China Technology project, grant number 5700-202255476A-2-0-KJ.

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.

Abbreviations

The following abbreviations are used in this manuscript:

MIMO	multiple-input multiple-output
MAB	multi-armed bandit
CMAB	combinatorial multi-armed bandit
UCB	upper confidence bound
CSI	channel state information
BS	base station
NC	non-coherent
TDD	time-division multiplex
OFDM	orthogonal frequency-division multiplexing
MMSE	minimum mean square error
CSIA	channel state information on the angular domain
SNR	signal-to-noise ratio
CCM	channel covariance matrix
FDD	frequency-division duplex

References

1. Wu, Q.; Wang, W.; Li, Z.; Zhou, B.; Huang, Y.; Wang, X. Spectrum chain: A disruptive dynamic spectrum-sharing framework for 6G. *Sci. China Inf. Sci.* **2023**, *66*, 130302. [[CrossRef](#)]
2. Wang, W.; Qi, N.; Jia, L.; Li, C.; Tsiftsis, T.A.; Wang, M. Energy-efficient UAV-relaying 5G/6G spectrum sharing networks: Interference coordination with power management and trajectory design. *IEEE Open J. Commun. Soc.* **2022**, *3*, 1672–1687. [[CrossRef](#)]
3. Yang, X.; Ma, S.; Jin, S. Antenna selection for asymmetrical uplink and downlink transceivers in massive MIMO systems. *IEEE Trans. Wirel. Commun.* **2023**, *in press*. [[CrossRef](#)]
4. Larsson, E.G.; Edfors, O.; Tufvesson, F.; Marzetta, T.L. Massive MIMO for next generation wireless systems. *IEEE Commun. Mag.* **2014**, *52*, 186–195. [[CrossRef](#)]
5. Liu, S.; Huang, X. Sparsity-aware channel estimation for mmWave massive MIMO: A deep CNN-based approach. *China Commun.* **2021**, *18*, 162–171. [[CrossRef](#)]
6. Yuan, M.; Wang, H.; Yin, H.; He, D. Alternating optimization based hybrid transceiver designs for wideband millimeter-wave massive multiuser MIMO-OFDM systems. *IEEE Trans. Wirel. Commun.* **2023**, *in press*. [[CrossRef](#)]
7. Agiwal, M.; Roy, A.; Saxena, N. Next generation 5G wireless networks: A comprehensive survey. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 1529–1543. [[CrossRef](#)]
8. Anarakifirooz, E.; Loyka, S. The robustness of favorable propagation in massive MIMO to location and phase errors. *IEEE Trans. Wirel. Commun.* **2023**, *in press*. [[CrossRef](#)]
9. Baeza, V.M.; Armada, A.G. Orthogonal versus Non-Orthogonal multiplexing in Non-Coherent Massive MIMO Systems based on DPSK. In Proceedings of the Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Porto, Portugal, 8–11 June 2021; pp. 101–105.
10. Manolakos, A.; Chowdhury, M.; Goldsmith, A. Energy-based modulation for noncoherent massive SIMO systems. *IEEE Trans. Wirel. Commun.* **2016**, *16*, 7831–7846. [[CrossRef](#)]
11. Baeza, V.M.; Armada, A.G.; Zhang, W.; El-Hajjar, M.; Hanzo, L. A noncoherent multiuser large-scale SIMO system relying on M-ary DPSK and BICM-ID. *IEEE Trans. Veh. Technol.* **2018**, *67*, 1809–1814. [[CrossRef](#)]

12. Ngo, K.H.; Guillaud, M.; Decurninge, A.; Yang, S.; Schniter, P. Multi-user detection based on expectation propagation for the non-coherent SIMO multiple access channel. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 6145–6161. [[CrossRef](#)]
13. Zhang, Z.; Jiao, C.; Zhong, C. Impact of mobility on the uplink sum rate of MIMO-OFDMA cellular systems. *IEEE Trans. Commun.* **2017**, *65*, 4218–4231. [[CrossRef](#)]
14. Anand, A.; Murthy, C.R. Impact of subcarrier allocation and user mobility on the uplink performance of multiuser massive MIMO-OFDM systems. *IEEE Trans. Commun.* **2022**, *70*, 5285–5299. [[CrossRef](#)]
15. Hong, M.; Hwang, I.; Heo, J.; Hong, D. A new scheduling algorithm for time-varying MIMO channels with a channel aging metric. In Proceedings of the IEEE Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 8–11 January 2023.
16. Jin, L.; Gu, X.; Hu, Z. Low-complexity scheduling strategy for wireless multiuser multiple-input multiple-output downlink system. *IET Commun.* **2023**, *5*, 990–995. [[CrossRef](#)]
17. Shehata, M.; Crussire, M.; Hlard, M.; Pajusco, P. Leakage based users selection for hybrid beamforming in millimeter wave MIMO. In Proceedings of the Annual International Symposium on Personal, Indoor and Mobile Radio Communications, Bologna, Italy, 9–12 September 2018; pp. 1144–1150.
18. Bu, G.; Jiang, J. Reinforcement learning-Based user scheduling and resource allocation for massive MU-MIMO system. In Proceedings of the IEEE/CIC International Conference on Communications in China, Istanbul, Turkey, 11–13 August 2019; pp. 1–6.
19. Thet, N.M.W.; Baykas, T.; Ozdemir, M.K. Performance analysis of user scheduling in massive MIMO with fast moving users. In Proceedings of the 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications, Changchun, China, 8–11 September 2019; pp. 641–646.
20. Yoon, S.; Lee, S.J. Improved Hierarchical Codebook-Based Channel Estimation for mmWave Massive MIMO Systems. *IEEE Wirel. Commun. Lett.* **2022**, *11*, 2095–2099. [[CrossRef](#)]
21. Yang, B.; Yu, Z.; Lan, J.; Zhang, R.; Zhou, J.; Hong, W. Digital beamforming-based massive MIMO transceiver for 5G millimeter-wave communications. *IEEE Trans. Microw. Theory Tech.* **2018**, *66*, 3403–3418. [[CrossRef](#)]
22. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; The MIT Press Cambridge: London, UK, 2018.
23. Song, Y.; Liu, C.; Zhang, W.; Liu, Y.; Zhou, H.; Shen, X. Two stage beamforming in massive MIMO: A combinatorial multi-armed bandit based approach. *IEEE Trans. Veh. Technol.* **2023**, *72*, 6794–6799. [[CrossRef](#)]
24. Vershynin, R. *High-Dimensional Probability—An Introduction with Applications in Data Science*; University of California: Irvine, CA, USA, 2020.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.