Two Stage Beamforming in Massive MIMO: A Combinatorial Multi-Armed Bandit Based Approach

Yunchao Song, *Member, IEEE*, Chen Liu, *Member, IEEE*, Wenyi Zhang, Yiliang Liu, *Member, IEEE*, Haibo Zhou, *Senior Member, IEEE*, and Xuemin Shen, *Fellow, IEEE*

***Abstract*—In frequency division duplex (FDD) massive multi-input multi-output (MIMO), the two-stage beamforming (TSB) using channel covariance matrices (CCM) can significantly reduce the downlink training length (DTL) and channel feedback. However, the overhead to estimate the CCM is large. In this paper, a combinatorial multi-armed bandit (CMAB) based TSB scheme is proposed without requirement of CMM. Particularly, the problem of the pre-beamforming matrix design is transformed into a CMAB problem. We consider the pre-beamforming matrix design in each slot as the arm selection in the CMAB, and convert the problem of the arm selection into a 0-1 integer linear programming problem, which can be solved by the branch-and-bound method. During the training process, the maximum likelihood method is used to detect the power of angle spectrum, and the angle range of each user is determined adaptively. We prove that the regret grows logarithmically with time, such that the proposed scheme converges towards the optimal action. Finally, simulation results demonstrate that the proposed scheme can significantly improve the spectral efficiency.**

***Index Terms*—Massive MIMO, combinatorial multi-armed bandit, upper confidence bound, chi-square distribution.**

I. Introduction

Massive multi-input multi-output (MIMO) systems have shown a great potential in 6 G wireless communications [1], [2]. Since the channel reciprocity does not hold in frequency division duplex (FDD) [3], [4], the downlink training length (DTL) and channel state information (CSI) feedback to obtain the instantaneous CSI at the base station (BS) bring a large burden for FDD systems, which make the FDD massive MIMO challenging. Considering that the statistical CSI remains unchanged over the intervals much longer than the symbol transmission time scale, a two-stage beamforming (TSB) scheme, namely joint spatial division and multiplexing (JSDM), based on the channel covariance matrix (CCM) is proposed [3], [4]. In the first stage, the JSDM groups the users and designs a prebeamformer by the statistical CSI to reduce the dimension of the effective channel matrix, yielding reduced DTL and CSI feedback. In the second stage, a multi-user precoder is designed to reduce the interference. Many works have been dedicated to improving the performance of the TSB scheme. A minimum mean squared error (MMSE) based JSDM scheme designs the prebeamformer and multi-user precoder based on the MMSE and weighted MMSE criterion, respectively [5]. To address the grouping issue, a neighbor-based JSDM (NJSDM) that fully utilizes the signal space was proposed to obtain a higher spectral efficiency (SE) [6]. The TSB scheme considering the multi-cluster channel was proposed in [7]. The aforementioned TSB schemes assume an available CCM. In [8], Khalilsarai et al. estimated the CCM and proposed a TSB scheme called active channel sparsification (ACS), which uses the discrete fourier transform (DFT) matrix to design the prebeamformer. In practice, the overhead to obtain the CCM is large [4], [8], especially when the users move fast such that the CCM changes frequently.

In this paper, we consider a single-cell FDD massive MIMO system, and propose a combinatorial multi-armed bandit (CMAB) based TSB scheme. Different from other TSB schemes, the pre-beamforming matrix is designed by the CMAB which does require the CCM. In the first stage of the TSB scheme, we transform the pre-beamforming design problem into a CMAB problem, where the pre-beamforming matrix is regarded as a super arm. We use a linear upper confidence bound (UCB) strategy to select the super arm, and the problem of selecting the super arm is transformed into a linear integer 0-1 programming problem. The branch-and-bound method is used to solve this programming problem. In the second stage, the multi-user precoder is designed to mitigate the interference. During the CMAB process, we use the maximum likelihood detection method to determine the power angular spectrum (PAS), and the angle range (AR) is reduced iteratively. We prove that the proposed method achieves regret that grows logarithmically with time. Simulation results also validate the good performance of the proposed scheme.

Notations: Bold uppercase letters denote matrices, and bold lowercase letters denote column vectors. Also, $\mathbf{a}_i$ and $\mathbf{a}^i$ denote the i-th column and the i-th row of the matrix $\mathbf{A}$, respectively. The superscripts $(\cdot)^H$, $(\cdot)^T$, and $(\cdot)^\dagger$ indicate the matrix conjugate-transpose, transpose, and pseudo-inverse operations, respectively. Furthermore, $\mathcal{M}$ denotes the set $\{1, 2, \ldots, M\}$. The complex number field is represented by $\mathbb{C}$. $\|\cdot\|_2$ denotes the l_2-norm, and $|\mathcal{A}|$ denotes the cardinal number of the set $\mathcal{A}$.

II. Preliminary

Consider a single-cell FDD massive MIMO system, where the BS serves K single-antenna user equipments (UEs). The BS is equipped with a unitary linear antenna (ULA) array of M elements. In downlink transmission, let $\mathbf{h}_k^H \in \mathbb{C}^{1\times M}$ be the downlink channel vector from BS to user k, and $\mathbf{y} = [y_1, y_2, \ldots, y_K]^T \in \mathbb{C}^{K\times 1}$ be the received signal, where y_k is the received signal of user k. Let $\mathbf{B}$ and $\mathbf{W}$ be the prebeamfoming matrix and multi-user precoding matrix, respectively, in the TSB scheme. Then

$$\mathbf{y} = \mathbf{H}^H\mathbf{B}\mathbf{W}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2 \ldots, \mathbf{h}_K]$, $\mathbf{s} = [s_1, s_2, \ldots, s_K]^T \in \mathbb{C}^{K\times 1}$ is the transmitted signal with $\mathbb{E}(\mathbf{s}\mathbf{s}^H) = \mathbf{I}$, $\mathbb{E}(\cdot)$ is the expectation, and $\mathbf{n} \in \mathbb{C}^{K\times 1} \sim \mathcal{CN}(0, \sigma_n^2\mathbf{I})$. In the first stage of the TSB, $\mathbf{B}$ is designed using the CCM, which make the effective channel matrix $\mathbf{H}^H\mathbf{B}$ having special structure such that the DTL and the channel feedback is small. In the second stage, using the zero-forcing criterion [9], [10], the multi-user precoder $\mathbf{W} = (\mathbf{H}^H\mathbf{B})^\dagger$ is designed to cancel the interference, where $(\cdot)^\dagger$ is the pseudo-inverse.

Manuscript received 11 June 2022; revised 31 October 2022; accepted 19 November 2022. Date of publication 6 January 2023; date of current version 18 May 2023. This work was supported in part by the Natural Science Foundation of China under Grant 62101282 and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20190732. The review of this article was coordinated by Prof. Yue Gao. *(Corresponding author: Chen Liu.)*

Yunchao Song, Chen Liu, and Wenyi Zhang are with the College of Electronic and Optical Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: songyc@njupt.edu.cn; liuch@njupt.edu.cn; 1020020928@njupt.edu.cn).

Yiliang Liu is with the School of Cyber Science and Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China (e-mail: liuyiliang@xjtu.edu.cn).

Haibo Zhou is with the School of Electronic Science and Engineering, Nanjing University, Jiangsu 210023, China (e-mail: haibozhou@nju.edu.cn).

Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L3G1, Canada (e-mail: sshen@uwaterloo.ca).

Digital Object Identifier 10.1109/TVT.2022.3229312

A one-ring channel model [3], [6] is considered, where user k has an azimuth center angle θ_k and an angular spread (AS) Δ_k. By θ_k and Δ_k, the CCM is calculated as

$$\mathbf{C}_k = \eta_k \int_{\theta_k-\Delta_k}^{\theta_k+\Delta_k} \gamma_k(\theta)\,\mathbf{a}(\theta)\,\mathbf{a}(\theta)^H\, d\theta, \tag{2}$$

where $\mathbf{a}(\theta) = \frac{1}{\sqrt{M}}\left[1, e^{j2\pi\frac{d}{\lambda_c}\sin(\theta)}, \ldots, e^{j2\pi\frac{d(M-1)}{\lambda_c}\sin(\theta)}\right]^T$ is the steering vector, λ_c is the wavelength, $d = \frac{\lambda_c}{2}$ is the spacing between two antenna elements. $\gamma_k(\theta)$ is the channel PAS, and η_k is a normalized density function such that $\eta_k \int_{\theta_k-\Delta_k}^{\theta_k+\Delta_k} \gamma_k(\theta)d\theta = 1$. Using the Karhunen-Loeve representation, user k's channel vector is represented as $\mathbf{h}_k = \mathbf{C}_k^{1/2}\mathbf{z}_k$ where $\mathbf{z}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the small-scale fading. In the conventional TSB schemes, it will cost a large overhead to obtain the CCM.

III. The Proposed CMAB Based TSB Scheme

A. Problem Formulation

In the first stage of the proposed TSB scheme, the prebeamforming matrix is designed by the DFT matrix $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_M]$, where $\mathbf{d}_i = \mathbf{a}(\arcsin(-1 + \frac{2(i-1)}{M}))$ [11], [12]. We define the interval of the statistical CSI remaining unchanged as a time block (TB). Suppose that the AR of user k at the $(\bar{t}-1)$-th TB is $\Theta_k^{\bar{t}-1} = [\theta_k^{lo}, \theta_k^{up}]$, where θ_k^{lo} is the lower bound and θ_k^{up} is the upper bound. In the $\bar{t}$-th TB, we set a large constant Δ, such that the AR of user k belongs in $\Theta_k^{\bar{t}} = [\theta_k^{lo} - \Delta, \theta_k^{up} + \Delta]$.

Define $\Phi_k^{\bar{t}} = \{i | \theta_k^{lo} - \Delta \le \arcsin(-1 + \frac{2(i-1)}{M}) \le \theta_k^{up} + \Delta, i \in \mathcal{M}\}$, where $\mathcal{M} = \{1, 2, \ldots, M\}$. Then $\Phi_k^{\bar{t}}$ is a set containing consecutive numbers. Let $\mathbf{h}_k^H(t)$ be the channel vector $\mathbf{h}_k^H$ at the t-th slot in the $\bar{t}$-th TB. As the number of antennas approaches infinity, we have $\mathbf{a}(\theta_i)^H\mathbf{a}(\theta_j) = 0$ for $\theta_i \neq \theta_j$ [13]. It is easy to verify that if $l \notin \Phi_k^{\bar{t}}$, we have $\mathbf{a}(\theta)^H\mathbf{d}_l = 0$ and $\mathbf{d}_l^H\mathbf{a}(\theta) = 0$ for $\theta \in [\theta_k^{lo} - \Delta, \theta_k^{up} + \Delta]$. Thus, for $l \notin \Phi_k^{\bar{t}}$,

$$\mathbf{d}_l^H\mathbf{C}_k\mathbf{d}_l = \eta_k \int_{\theta_k-\Delta_k}^{\theta_k+\Delta_k} \gamma_k(\theta)\,\mathbf{d}_l^H\mathbf{a}(\theta)\,\mathbf{a}(\theta)^H\,\mathbf{d}_l d\theta = 0. \tag{3}$$

Since $\mathbf{C}_k$ is symmetric, we have $\mathbf{C}_k^{1/2} = (\mathbf{C}_k^{1/2})^H$. Then $\mathbf{d}_l^H\mathbf{C}_k\mathbf{d}_l = \mathbf{d}_l^H\mathbf{C}_k^{\frac{1}{2}}\mathbf{C}_k^{\frac{1}{2}}\mathbf{d}_l = \mathbf{d}_l^H(\mathbf{C}_k^{\frac{1}{2}})^H\mathbf{C}_k^{\frac{1}{2}}\mathbf{d}_l = (\mathbf{C}_k^{\frac{1}{2}}\mathbf{d}_l)^H\mathbf{C}_k^{\frac{1}{2}}\mathbf{d}_l = 0$, yielding that $\mathbf{C}_k^{\frac{1}{2}}\mathbf{d}_l = \mathbf{0}$ where all the elements in $\mathbf{0}$ are zeros. As a result,

$$\mathbf{h}_k^H(t)\mathbf{d}_l = \mathbf{z}_k^H(t)\mathbf{C}_k^{1/2}\mathbf{d}_l = 0, l \notin \Phi_k^{\bar{t}}. \tag{4}$$

Denote by $\mathcal{A}_{\mathbf{a}(t)}$ the set of the selected DFT vectors and $\mathbf{D}_{\mathcal{A}_{\mathbf{a}(t)}}$ the pre-beamforming matrix. Then

$$\mathbf{h}_k^H(t)\mathbf{D}_{\mathcal{A}_{\mathbf{a}(t)}} = \left[\cdots, 0, \mathbf{h}_k^H(t)\mathbf{D}_{\Phi_k^{\bar{t}} \cap \mathcal{A}_{\mathbf{a}(t)}}, 0, \cdots\right] \tag{5}$$

is a sparse vector, where $\mathbf{D}_{\Phi_k^{\bar{t}} \cap \mathcal{A}_{\mathbf{a}(t)}}$ is a matrix whose columns are $\mathbf{d}_i, i \in \Phi_k^{\bar{t}} \cap \mathcal{A}_{\mathbf{a}(t)}$. Denote by d_k the number of elements in $\mathbf{h}_k^H(t)\mathbf{D}_{\Phi_k^{\bar{t}} \cap \mathcal{A}_{\mathbf{a}(t)}}$. Let the pilot matrix have the circular structure as $\mathbf{X} \in \mathbb{C}^{L \times |\mathcal{A}_{\mathbf{a}(t)}|} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L, \mathbf{x}_1, \mathbf{x}_2, \cdots]$, where $L \ge \max_k d_k$, $|\mathcal{A}_{\mathbf{a}(t)}|$ is the cardinal number of $\mathcal{A}_{\mathbf{a}(t)}$, and $[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L] \in \mathbb{C}^{L \times L}$ is a unitary matrix. The received signal of user k in the channel estimation is

$$\mathbf{y}_k^H(t) = \mathbf{h}_k^H(t)\mathbf{D}_{\mathcal{A}_{\mathbf{a}(t)}}\mathbf{X} + \mathbf{n}^H(t). \tag{6}$$

Substitute (5) into (6), and removing the zero entries, we have

$$\mathbf{y}_k^H(t) = \mathbf{h}_k^H(t)\mathbf{D}_{\Phi_k^{\bar{t}} \cap \mathcal{A}_{\mathbf{a}(t)}}\tilde{\mathbf{X}} + \mathbf{n}^H(t). \tag{7}$$

where $\tilde{\mathbf{X}}$ is also a tall unitary matrix composed by partial vectors $\mathbf{x}_i$. Then $\bar{\mathbf{h}}_k^H(t) = \mathbf{h}_k^H(t)\mathbf{D}_{\Phi_k^{\bar{t}} \cap \mathcal{A}_{\mathbf{a}(t)}}$ can be estimated by least square (LS) criterion, and the DTL is L. By LS criterion, we multiply (7) with $\tilde{\mathbf{X}}^H$, and then we have

$$\mathbf{y}_k^H(t)\tilde{\mathbf{X}}^H = \bar{\mathbf{h}}_k^H(t) + \mathbf{n}^H(t)\tilde{\mathbf{X}}^H. \tag{8}$$

$\bar{\mathbf{h}}_k^H(t)$ is estimated as $\hat{\bar{\mathbf{h}}}_k^H(t) = \mathbf{y}_k^H(t)\tilde{\mathbf{X}}^H$, which is fed back to the BS. Using $\hat{\bar{\mathbf{h}}}_k^H(t)$ and (5), the BS can obtain the estimation of $\mathbf{H}^H(t)\mathbf{D}_{\mathcal{A}_{\mathbf{a}(t)}}$. In the second stage, the multi-user precoder $\mathbf{W}$ is designed as $(\hat{\bar{\mathbf{H}}}(t))^\dagger$ to mitigate the interference, where $\hat{\bar{\mathbf{H}}}(t)$ is the estimation of $\mathbf{H}^H(t)\mathbf{D}_{\mathcal{A}_{\mathbf{a}(t)}}$.

The main problem is to design the pre-beamforming matrix $\mathbf{D}_{\mathcal{A}_{\mathbf{a}(t)}}$ to maximize the SE, where the DTL is limited. Since maximizing the received energy can help to improve the SE [14], [15], we maximize the received signal energy, i.e., $\sum_k \|\mathbf{h}_k^H(t)\mathbf{D}_{\mathcal{A}_{\mathbf{a}(t)}}\|_2^2$ to obtain a high SE, where $\|\cdot\|_2$ denotes the l_2-norm. Considering (5), to estimate the channel well, the DTL is set to $\kappa \max_k d_k$, where κ is constant. The problem of the pre-beamforming matrix design is

$$\begin{aligned} \mathcal{P}_1: \ &\max_{\mathcal{A}_{\mathbf{a}(t)} \subseteq \mathcal{M}} \sum_t \sum_k \left\|\mathbf{h}_k^H(t)\mathbf{D}_{\mathcal{A}_{\mathbf{a}(t)}}\right\|_2^2, \\ &\text{s.t.} \ \max_k d_k < \tau. \end{aligned} \tag{9}$$

where $\tau = \frac{DTL}{\kappa}$. Without CCM, we use a multi-armed bandit (MAB) [16], [17] based algorithm to obtain $\mathcal{A}_{\mathbf{a}(t)}$. In $\mathcal{P}_1$, multiple DFT vectors are required, and thus $\mathcal{P}_1$ is a CMAB problem [17]. In the CMAB problem, there are some arms and each arm is regarded as a base arm. At each time, a super arm (i.e., multiple base arms) is selected, and all of the selected base arms are observed at that time. A linearly weighted combination of the observations of the selected base arms is the reward of the super arm. The goal is to find a policy that minimizes regret, defined as the difference between the reward obtained by a genie that knows the mean of each base arm, and that obtained by the given policy [17]. Next, we show a method to reduce the ARs of the users, such that the AR becomes accurate and the effective SE (ESE) is improved.

B. Adaptive Angle Range Adjustment

From the previous subsection, the AR $\Theta_k^{\bar{t}} = [\theta_k^{lo} - \Delta, \theta_k^{up} + \Delta]$ is large due to the large constant Δ, yielding a long DTL and a low ESE. To improve the ESE, we propose a method to adaptively adjust $\Theta_k^{\bar{t}}$ such that $\Theta_k^{\bar{t}}$ get closer to the actual AR, which can reduce the DTL.

Considering (4), assume $\Phi_k^{\bar{t}} = [lo^k, lo^k + 1, \ldots, up^k]$. Note that if the PAS corresponding to the edge vector $\mathbf{d}_{up^k}$ is not zero, then the PAS corresponding to $\mathbf{d}_{up^k-1}$ is also not zero. If the PAS corresponding to $\mathbf{d}_{up^k}$ is zero, we cannot determine the PAS corresponding to $\mathbf{d}_{up^k-1}$. Similarly, if the PAS corresponding to the edge vector $\mathbf{d}_{lo^k}$ is not zero, then the PAS corresponding to $\mathbf{d}_{lo^k+1}$ is also not zero. If the PAS corresponding to $\mathbf{d}_{lo^k}$ is zero, we cannot determine the PAS corresponding to $\mathbf{d}_{lo^k+1}$. Thus the PAS corresponding to the edge vectors has a larger probability to be zero than the non-edge vectors, and the edge vectors $\mathbf{d}_{lo^k}$ or $\mathbf{d}_{up^k}$ are considered in each iteration.

The power $p_{k,m} = \mathbb{E}(|\mathbf{h}_k^H(t)\mathbf{d}_m|^2)$ of user k at the vector $\mathbf{d}_m$ is calculated as follows. We assume that if $p_{k,m}$ is smaller than a constant ζ^2, the power is zero. Otherwise, the power is non-zero. In (8), if the vector $\mathbf{d}_m$ is contained in $\mathbf{D}_{\mathcal{A}_{\mathbf{a}(t)}}$, then considering $\mathbf{d}_m$, we have

$$\bar{h}_{k,m}(t) + \bar{n}_{k,m}(t) = \bar{y}_{k,m}(t), \tag{10}$$

where $\bar{h}_{k,m} = \mathbf{h}_k^H(t)\mathbf{d}_m$, and $\bar{n}_{k,m}(t)$ and $\bar{y}_{k,m}(t)$ are the elements in $\mathbf{n}^H(t)\tilde{\mathbf{X}}$ and $\mathbf{y}_k^H(t)\tilde{\mathbf{X}}^H$ corresponding to $\mathbf{d}_m$, respectively. Since $\bar{h}_{k,m}(t) = \mathbf{h}_k^H(t)\mathbf{d}_m = \mathbf{z}_k^H\mathbf{R}_k^{1/2}\mathbf{d}_m$, $\bar{h}_{k,m}(t)$ is a Gaussian variable with zero mean and the variance is $p_{m,k}$. Thus, $\bar{h}_{k,m}(t) + \bar{n}_{k,m}(t)$ is a Gaussian variable with zero mean and the variance is $p_{k,m} + \sigma_n^2$. If the vector $\mathbf{d}_m$ has been picked by n times, we have $\bar{\mathbf{y}}_{k,m}$, where $\bar{\mathbf{y}}_{k,m} = [\bar{y}_{k,m}(1), \ldots, \bar{y}_{k,m}(n)]$. We can use $\bar{\mathbf{y}}_{k,m}$ to detect the variance $p_{m,k} + \sigma_n^2$ and $p_{k,m}$. Considering (10), the likelihood-ratio (LR) is

$$\gamma_{k,m} = \frac{\int_\zeta^\infty P(\bar{\mathbf{y}}_{k,m}|p)\,dp}{\int_{\sigma_n^2}^\zeta P(\bar{\mathbf{y}}_{k,m}|p)\,dp} = \frac{\int_\zeta^\infty \prod_{t=1}^n \frac{1}{\sqrt{2\pi p}} e^{-\frac{\bar{y}_{k,m}^2(t)}{2p^2}}\,dp}{\int_{\sigma_n}^\zeta \prod_{t=1}^n \frac{1}{\sqrt{2\pi p}} e^{-\frac{\bar{y}_{k,m}^2(t)}{2p^2}}\,dp} = \frac{\int_\zeta^\infty \frac{1}{p^n} e^{-\frac{\sum_{t=1}^n \bar{y}_{k,m}^2(t)}{2p^2}}\,dp}{\int_{\sigma_n}^\zeta \frac{1}{p^n} e^{-\frac{\sum_{t=1}^n \bar{y}_{k,m}^2(t)}{2p^2}}\,dp}. \tag{11}$$

Letting $\lambda = \frac{\sum_{t=1}^n \bar{y}_{k,m}^2(t)}{2}$, $\int_0^\infty \frac{1}{p^n} e^{-\lambda p^{-2}}dp = \frac{1}{2}\int_0^\infty q^{\frac{n-3}{2}} e^{-\lambda q}dq = \frac{\Gamma(\frac{n-1}{2})}{2\lambda^{\frac{n-1}{2}}}$, where $\Gamma(\cdot)$ is the Gamma function. Calculate $\int_0^\zeta \frac{1}{p^n} e^{-\lambda p^{-2}}dp$ by the numerical method, and $\int_\zeta^\infty \frac{1}{p^n} e^{-\lambda p^{-2}}dp$ is obtained. If the $\gamma_{k,m} \geq \varepsilon$, the power $p_{k,m}$ is detected as non-zero. If $\gamma_{k,m} \leq \frac{1}{\varepsilon}$, $p_{k,m}$ is detected as 0. In other cases, the power is not determined. ε is a constant used to show the probability that the power detected correctly. The larger ϵ is, the larger probability the power is detected correctly.

C. The CMAB Based Pre-Beamformer Design

Next, we propose a CMAB based pre-beamformer design algorithm to solve $\mathcal{P}_1$. The DFT vectors $\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_M$ are the base arms and are labeled 1 to M. Let $R_{i,t} = \|\mathbf{H}^H(t)\mathbf{d}_i\|_2^2 = \sum_{k=1}^K \bar{y}_{k,i}^2(t)$ be the reward of the i-th base arm. A super arm $\mathcal{A}_{\mathbf{a}(t)} = [a_1, a_2, \cdots]$ is composed by the base arms $a_i, i = 1, 2, \cdots$, where $a_i \in \{1, 2, \ldots, M\}$. Given a super arm $\mathcal{A}_{\mathbf{a}(t)}$, the pre-beamforming matrix is $\mathbf{D}_{\mathcal{A}_{\mathbf{a}(t)}} = [\mathbf{d}_{a_1}, \mathbf{d}_{a_2}, \ldots,]$, and the reward is the power of the received signal given by

$$R_{\mathcal{A}_{\mathbf{a}(t)},t} = \left\|\mathbf{H}^H(t)\mathbf{D}_{\mathcal{A}_{\mathbf{a}(t)}}\right\|_F^2 = \sum_{i\in\mathcal{A}_{\mathbf{a}(t)}} R_{i,t}. \tag{12}$$

Considering that the reward of the super arm $\mathcal{A}_{\mathbf{a}(t)}$ is a linear combination of the rewards of the selected base arms, we use a linear UCB strategy to select the super arm in each slot. In the linear UCB strategy, we set N UCB values for the N base arms. This strategy only requires storage that grows linearly in the number of base arms. Considering the chi-square distribution of the reward, the UCB values of the base arms in the t-th slot are defined as

$$u_i(t) = \begin{cases} \bar{R}_{i,t} + \bar{c}\ \max\left\{\frac{16(M+1)\ln(t)}{m_{i,t}},\right. \\ \left.\sqrt{\frac{2K(M+1)\ln(t)}{m_{i,t}}}\right\} & m_{i,t} \neq 0 \\ \infty & m_{i,t} = 0. \end{cases} \tag{13}$$

where $m_{i,t}$ denotes the number of the action i has been selected, $\bar{c}$ is a arbitrary constant satisfing $\bar{c} \geq \sqrt{\max_k p_{k,i} + \sigma_n^2}$, $\bar{R}_{i,t} = \frac{\bar{R}_{i,t-1}m_{i,t-1} + R_{i,t}}{m_{i,t-1}+1}$. Since the PAS of each user is bounded (i.e., $p_{m,k}$ is bounded), we can find the value $\bar{c}$. Then the UCB value for the super arm $\mathcal{A}_{\mathbf{a}(t)}$ is $\sum_{i\in\mathcal{A}_{\mathbf{a}(t)}} u_i(t)$, which is proved to be converged to the optimal action latter.

In each slot, we use $\mathbf{v}$ to denote whether the base arms are selected in the super arm, where $v_i = 1$ denotes that the i-th base arm is selected and $v_i = 0$ denotes that the i-th base arm is not selected. Let $\mathbf{u} = [u_1(t), u_2(t), \ldots, u_M(t)]$. Define a matrix $\mathbf{A}$, whose k-th row and l-th column element is

$$\mathbf{A}_{k,l} = \begin{cases} 1, & \tilde{\mathbf{H}}(t)_{k,l} \neq 0 \\ 0, & \tilde{\mathbf{H}}(t)_{k,l} = 0. \end{cases} \tag{14}$$

The problem of maximizing the UCB value is

$$\mathcal{P}_2: \max_{x_i\in\{0,1\}} \mathbf{u}^H\mathbf{v} \tag{15}$$

$$\text{s.t. } \mathbf{A}\mathbf{v} \leq \tau\mathbf{e}, \tag{16}$$

where $\mathbf{e}$ is a vector with all elements being 1.

To reduce the DTL such that the ESE can be improved, the exact AR $\Theta_k^{\bar{t}}$ should be found. When $\Theta_k^{\bar{t}}$ of all the users decrease to a convergence, $\Theta_k^{\bar{t}}$ is seen as the actural AR, and the DTL is the smallest, leading to a large SE. Thus, the actions corresponding the edge vectors have higher priorities. Consequently, we divide the actions into two sets. One is the candidate set $\mathcal{S}_C$ which involves the edge vectors, and the other one is the remaining set $\mathcal{S}_R$ which involves the remaining vectors. $\mathbf{v}$ is divided into $\mathbf{v}_1$ and $\mathbf{v}_2$, where $\mathbf{v}_1$ denotes the actions in the candidate set, and $\mathbf{v}_2$ denotes the remaining actions. Correspondingly, $\mathbf{u}$ is divided into $\mathbf{u}_1$ and $\mathbf{u}_2$, where $\mathbf{u}_1$ is composed by $u_i, i \in \mathcal{S}_C$, and $\mathbf{u}_2$ is composed by $u_i, i \in \mathcal{S}_R$. Similarly, $\mathbf{A}$ is divided into $\mathbf{A}_1$ and $\mathbf{A}_2$, where the columns of $\mathbf{A}_1$ are $\mathbf{a}_i, i \in \mathcal{S}_C$, and the columns of $\mathbf{A}_2$ are $\mathbf{a}_i, i \in \mathcal{S}_R$.

Then $\mathcal{P}_2$ is divided into two subproblems. The first one is to choose the base arms in $\mathcal{S}_C$ to maximize the UCB value as

$$\mathcal{P}_{S1}: \max_{v_i\in\{0,1\}} \mathbf{u}_1^H\mathbf{v}_1 \tag{17}$$

$$\text{s.t. } \mathbf{A}_1\mathbf{v}_1 \leq \tau\mathbf{e}, \tag{18}$$

where the arms in $\mathcal{S}_R$ are not considered. Once $\mathbf{v}_1$ is obtained, we substitute $\mathbf{v}_1$ into $\mathcal{P}_2$, and then consider the base arms in $\mathcal{S}_R$. The second subproblem is formulated as

$$\mathcal{P}_{S2}: \max_{v_i\in\{0,1\}} \mathbf{u}_2^H\mathbf{v}_2 \tag{19}$$

$$\text{s.t. } \mathbf{A_2}\mathbf{v}_2 \leq \tau\mathbf{e} - \mathbf{A}_2\mathbf{v}_1. \tag{20}$$

$\mathcal{P}_{S1}$ and $\mathcal{P}_{S2}$ are linear 0-1 integer programming problems, which can be solved by the branch-and-bound method.

We have shown the process of obtaining the super arm in each slot. In each slot, the reward of each base arm can be updated. Then, we use the LR to calculate the power of the arms in the candidate set. If the base arms in the candidate set are determined to have zero or non-zero power, then the edge vectors should be updated. The detail is given in Algorithm 1.

Theorem 1: The expected regret of the CMAB based algorithm is $O(\ln t)$.

Proof: In (11), there exists a constant $\bar{N}$ such that all users' ARs keep unchanged after $\bar{N}$ slots. Next, we consider the latter slots. Define a variable $Z_{i,t}$. In t-th slot, if the optimal action $\mathbf{a}^*$ is selected, $Z_{i,t}$ keeps unchanged. If a non-optimal action $\mathbf{a}$ is selected, $Z_{i,t}$ is added by 1, where $i = \arg\min_{j\in\mathbf{a}} m_{j,t}$ (If there are multiple solutions, we select one of them arbitrarily). The number of non-optimal actions picked is $\sum_{i=1}^M Z_{i,t}$, and $Z_{i,t} \leq m_{i,t}$. Denote by $I_{i,t}$ the indicator function that is 1 if $Z_{i,t}$ is added by one at time t. Let l be an arbitrary positive integer. Then

$$Z_{i,t} = \sum_{p=\bar{N}+1}^t \mathbf{1}\{I_{i,p} = 1\} \leq l + \sum_{p=\bar{N}+1}^t \mathbf{1}\{I_{i,p} = 1, Z_{i,p-1} \geq l\}, \tag{21}$$

Algorithm 1: The CMAB based Pre-beamformer Design.

Data: $\Theta_k^{\bar{l}} = \left[\theta_k^{lo} - \Delta, \theta_k^{up} + \Delta\right]$
Result: $\mathcal{A}_{\mathbf{a}(t)}$
Initialize: $m_{i,0} = 0$;
for $k = 1 : K$ **do**
  $lo^k = \left\lfloor M\cos\left(\theta_k^{lo} - \Delta\right)\right\rfloor$ and $up^k = \left\lceil M\cos\left(\theta_k^{up} + \Delta\right)\right\rceil$;
end
for $t = 1 : T$ **do**
  Use (13) to get $u_i(t)$;
  Solve $\mathcal{P}_{S1}$ and $\mathcal{P}_{S2}$ to get $\mathcal{A}_{\mathbf{a}(t)}$;
  Calculate the LR, and update lo^k, up^k and $m_{i,t}$;
end
Procedure End

where $\mathbf{1}(x)$ is 1 when the event x is true, and 0 when it is false. When $I_{i,p} = 1$, a non-optimal action $\mathbf{a}(p)$ is picked for which $m_{i,t} = \min_{j\in\mathcal{A}_{\mathbf{a}(p)}} m_{j,t}$. Then

$$\begin{aligned} Z_{i,t} &\le l + \sum_{p=\bar{N}+1}^{t} \mathbf{1}\Bigg\{\sum_{j\in\mathcal{A}_{\mathbf{a}^*}} \bar{R}_{j,p-1} + c_{j,p-1} \\ &\le \sum_{j\in\mathcal{A}_{\mathbf{a}(p)}} \bar{R}_{j,p-1} + c_{j,p-1}, Z_{i,p-1} > l\Bigg\} \le l + \sum_{p=1}^{t}\mathbf{1} \\ &\Bigg\{\sum_{j\in\mathcal{A}_{\mathbf{a}^*}} \bar{R}_{j,p} + c_{j,p} \le \sum_{j\in\mathcal{A}_{\mathbf{a}(p+1)}} \bar{R}_{j,p} + c_{j,p}, Z_{i,p} > l\Bigg\}. \end{aligned} \tag{22}$$

Let $Q^* = |\mathcal{A}_{\mathbf{a}^*}|$ and $Q^p = |\mathcal{A}_{\mathbf{a}(p+1)}|$. Since $l \le Z_{i,p} \le m_{i,p}$,

$$\begin{aligned} Z_{i,t} &\le l \\ &+ \sum_{p=1}^{t}\mathbf{1}\Bigg\{\min_{0<m_{h_1,t_1},m_{h_2,t_2},\cdots,m_{h_{Q^*},t_{Q^*}}\le p} \sum_{j=1}^{Q^*} \bar{R}_{h_j,t_j} + c_{h_j,t_j} \\ &\le \max_{l\le m_{g_1,t_1},m_{g_2,t_2},\cdots,m_{g_{Q^p},t_{Q^p}}\le p} \sum_{j=1}^{Q^p} \bar{R}_{g_j,t_j} + c_{g_j,t_j}\Bigg\} \\ &\le l + \sum_{p=1}^{\infty} \sum_{m_{h_1,t_1}=1}^{p} \cdots \sum_{m_{h_{Q^*},t_{Q^*}}=1}^{p} \sum_{m_{g_1,t_1}=l}^{p} \cdots \sum_{m_{g_{Q^p},t_{Q^p}}=l}^{p} \\ &\mathbf{1}\left\{\sum_{j=1}^{Q^*} \bar{R}_{h_j,t_j} + c_{h_j,t_j} \le \sum_{j=1}^{Q^p} \bar{R}_{g_j,t_j} + c_{g_j,t_j}\right\}. \end{aligned} \tag{23}$$

$\sum_{j=1}^{Q^*} \bar{R}_{h_j,t_j} + c_{h_j,t_j} \le \sum_{j=1}^{Q^p} \bar{R}_{g_j,t_j} + c_{g_j,t_j}$ means that at least one of the following events must be true.

$$\mathcal{E}1: \sum_{j=1}^{Q^*} \bar{R}_{h_j,t_j} \le R_{\mathbf{a}^*} - \sum_{j=1}^{Q^*} c_{h_j,t_j}; \tag{24}$$

$$\mathcal{E}2: \sum_{j=1}^{Q^p} \bar{R}_{g_j,t_j} \ge R_{\mathbf{a}(p+1)} + \sum_{j=1}^{Q^p} c_{g_j,t_j}; \tag{25}$$

$$\mathcal{E}3: R_{\mathbf{a}^*} < R_{\mathbf{a}(p+1)} + \sum_{j=1}^{Q^p} c_{g_j,t_j} + \sum_{j=1}^{Q^*} c_{h_j,t_j}. \tag{26}$$

where $R_{\mathbf{a}^*} = \sum_{j=1}^{Q^*} R_{h_j}$ and $R_{\mathbf{a}(p+1)} = \sum_{j=1}^{Q^*} R_{g_j}$. For $\mathcal{E}1$, at least one j satisfies $\bar{R}_{h_j,t_j} \le R_{h_j} - c_{h_j,t_j}$, and we have

$$P\{\mathcal{E}1\} \le \sum_{j=1}^{Q^*} P\left\{\bar{R}_{h_j,t_j} \le R_{h_j} - c_{h_j,t_j}\right\}. \tag{27}$$

In (10), $\bar{y}_{k,i}^2(t)$ follows a chi-square distribution, which is a $(2\tau_{i,k}^2, 4\tau_{i,k})$-sub-exponential function [18], where $\tau_{i,k}^2 = p_{i,k} + \sigma_n^2$. From [18], $\sum_{k=1}^{K} \bar{y}_{k,i}^2(t)$ is a $(2\sum_{k=1}^{K}\tau_{i,k}^2, 4\bar{\tau})$-sub-exponential function, where $\tilde{\tau}^2 = 2\sum_{k=1}^{K}\tau_{i,k}^2$, $\bar{\tau} = \max_k \tau_{i,k}$. Applying the Chernoff-Hoeffding bound for this sub-exponential function [18], each item in (27) satisfies

$$P\left\{\bar{R}_{h_j,t_j} \le R_{h_j} - c_{h_j,t_j}\right\} \le 2e^{-\min\left\{\frac{m_{h_j,t_j} c_{h_j,t_j}^2}{\tilde{\tau}^2}, \frac{m_{h_j,t_j} c_{h_j,t_j}}{8\bar{\tau}}\right\}},$$

From (13), $c_{h_j,t_j} \ge \bar{c}\sqrt{\frac{2K(M+1)\ln t}{m_{h_j,t_j}}}$, $e^{-\frac{m_{h_j,t_j} c_{h_j,t_j}^2}{\tilde{\tau}^2}} \le t^{\frac{-2(M+1)K\bar{c}^2}{\tilde{\tau}^2}}$.

Since $K\bar{c}^2 \ge \tilde{\tau}^2$, then $e^{-\frac{m_{h_j,t_j} c_{h_j,t_j}^2}{\tilde{\tau}^2}} \le t^{-2(M+1)}$. Moreover, from (13), $c_{h_j,t_j} \ge \bar{c}\frac{16(M+1)\ln(t-1)}{m_{i,t-1}}$ $e^{-\frac{m_{h_j,t_j} c_{h_j,t_j}}{8\bar{\tau}}} \le t^{\frac{-2(M+1)\bar{c}}{\bar{\tau}}}$. Since $\bar{c} \ge \bar{\tau}$, then $e^{-\frac{m_{h_j,t_j} c_{h_j,t_j}^2}{\tilde{\tau}^2}} \le t^{-2(M+1)}$. As a result,

$$\mathbb{E}\left(P\{\mathcal{E}1\}\right) \le Mt^{-2(M+1)}. \tag{28}$$

Similarly,

$$\mathbb{E}\left(P\{\mathcal{E}2\}\right) \le Mt^{-2(M+1)}. \tag{29}$$

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

$$\begin{aligned} \sum_{j=1}^{Q^p} \frac{16\bar{c}(M+1)\ln t}{m_{h_j,t_j}} &\le M\frac{16\bar{c}(M+1)\ln t}{l} \\ &\le M\frac{\bar{c}(M+1)\ln t}{2\bar{c}(M+1)\ln t}\frac{\Delta_{\mathbf{a}(t+1)}}{M} \le \frac{\Delta_{\mathbf{a}(t+1)}}{2}, \end{aligned} \tag{30}$$

and

$$\begin{aligned} &\sum_{j=1}^{Q^p}\sqrt{\frac{2\bar{c}(M+1)\ln t}{m_{i,t}}} \le M\sqrt{\frac{2\bar{c}(M+1)\ln t}{l}} \\ &M\sqrt{\frac{2\bar{c}(M+1)\Delta_{\mathbf{a}(t+1)}^2\ln t}{8\bar{c}(M+1)M^2\ln t}} \le M\frac{\Delta_{\mathbf{a}(t+1)}}{2M} \le \frac{\Delta_{\mathbf{a}(t+1)}}{2}. \end{aligned} \tag{31}$$

From (30) and (31), $\sum_{j=1}^{Q^p} c_{g_j,t_j} \le \frac{\Delta_{\mathbf{a}(t+1)}}{2}$. Similarly, we have $\sum_{j=1}^{Q^*} c_{h_j,t_j} \le \frac{\Delta_{\mathbf{a}(t+1)}}{2}$. Hence,

$$R_{\mathbf{a}^*} - R_{\mathbf{a}(p+1)} - \sum_{j=1}^{Q^p} c_{h_j,t_j} - \sum_{j=1}^{Q^*} c_{h_j,t_j} \ge 0, \tag{32}$$

and $\mathcal{E}3$ is always satisfied. Note that

$$\begin{aligned} &\sum_{p=1}^{\infty} \sum_{m_{h_1,t_1}=1}^{p} \cdots \sum_{m_{h_{Q^*},t_{Q^*}}=1}^{p} \sum_{m_{g_1,t_1}=l}^{p} \cdots \sum_{m_{g_{Q^p},t_{Q^p}}=l}^{p} \\ &2Mp^{-2(M+1)} \le \sum_{p=1}^{\infty} 2Mp^{-2} \le 1 + \frac{\pi^2}{3}M. \end{aligned} \tag{33}$$

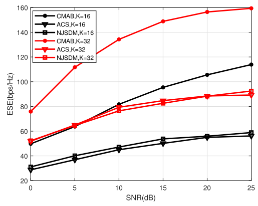

Fig. 1. ESE vs. SNR.

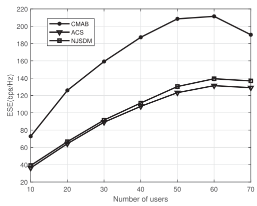

Fig. 2. ESE vs. number of users.

Thus, $\mathbb{E}(Z_{i,t}) \leq l + 1 + \frac{\pi^2}{3}M = O(\ln t)$. As a result,

$$Regret \leq \Delta_{max} \sum_{i=1}^{M} \mathbb{E}\left(Z_{i,t}\right) = O\left(\ln t\right). \tag{34}$$

IV. Simulation Results

In this section, we demonstrate the efficiency of the proposed CMAB based scheme through simulations. The number of antennas is 128. The center angles of the users are uniformly distributed in $[-\frac{\pi}{3}, \frac{\pi}{3}]$. The change of the center angles of all users is uniformly distributed in $[0^\circ, 5^\circ]$. The AS of each user is uniformly distributed in $[5^\circ, 15^\circ]$. In the proposed scheme, ε is set to 0.01, and ζ^2 is set to 0.1. In (13), $\bar{c}$ is affected by the system parameters. In this paper, since $p_{k,m} \leq \mathbb{E}(\|\mathbf{h}_k(t)\|)^2 \leq Trace(\mathbf{C}_k)$ and $Trace(\mathbf{C}_k) = \eta_k \int_{\theta_k-\Delta_k}^{\theta_k+\Delta_k} \gamma_k(\theta)d\theta = 1$, we have $p_{k,m} \leq 1$, and we set $\bar{c} = \sqrt{1+\sigma_n^2}$. The parameter κ in the pilot length is set to 1.5. We compare the CMAB based scheme with the ACS [8] and NJSDM schemes [6]. In the ACS and N-JSDM schemes, the CCM is obtained through 500 slots, and a TB includes 1000 slots, which corresponds to 5 m for the speed of 10 m/s in LTE system. The ESE of user k in a coherence block with T_c symbols is given by

$$R_k = \left(1 - \frac{\text{DTL}}{T_c}\right) \log\left(1 + SINR_k\right) \tag{35}$$

where T_c is set to 100 and $SINR_k$ is the signal-to-interference-and-noise ratio of user k.

Fig. 1 compares the ESEs of different schemes under different signal-to-noise ratios (SNRs). The numbers of users are 16 and 32, respectively. Δ is set to 15°, and the DTL is 30. The CMAB based scheme achieves higher ESEs than that of the NJSDM and ACS schemes. This is because, the ACS and NJSDM spend many slots to estimate the CCM. The CMAB, NJSDM and the ACS schemes for 16 users have lower ESEs than the CMAB, NJSDM and the ACS schemes for 32 users, respectively. Fig. 2 shows the ESEs of different schemes under different numbers of users. The SNR is 20 dB. Δ is 15°, and the DTL is 30. Our proposed scheme has a higher ESE than other schemes. The ESE of the proposed scheme first increases and then decreases by increasing the number of users. The reason is as follows. For a small number of users, the inter-user interference is not large, and the multi-user gain improves the ESE. However, when the number of users is large, the inter-user interference is high which decreases the ESE. Fig. 3 compares the ESEs of different schemes under different DTL. The SNR is 20 dB. Δ is 15°, and the number of user is 32. The CMAB based scheme has a higher ESE than the ACS. The ESEs of the CMAB and ACS schemes first increase and then decrease. This is because, for small DTL, the influence of the DTL is reduced. The number of DFT vectors will increase by increasing DTL, which increases the ESE. Thus, the ESEs are increased. However, when the DTL is large, the DTL will be important. By increasing the DTL, the SE will not be increased, and thus the ESE will be decreased. The ESE of the NJSDM scheme keeps unchanged, since the DTL of the NJSDM scheme is fixed and cannot be designed [6]. Fig. 4 shows the ESEs of the CMAB based scheme in each slots. The SNR is 20 dB, and the DTL is 30. Due to the characteristics of the UCB strategy, the ESEs of the CMAB based scheme first increase and then converge with the increasing of the slots. When Δ is large, it will spend more slots exploiting the actions, and thus the ESE converges slowly.

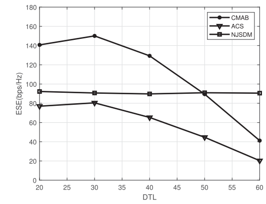

Fig. 3. ESE vs. DTL.

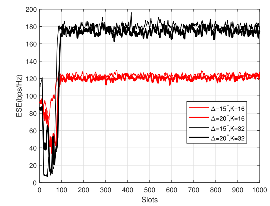

Fig. 4. ESE vs. slots

V. Conclusion

In this paper, we have proposed a CMAB based TSB scheme in FDD massive MIMO. The ESE can be significantly improved since the CCM is not required and the overhead of channel estimation can be reduced. We have also proved that the regret grows logarithmically with time, and the simulation has shown that the proposed TSB scheme can improve the ESE. In our future work, we will consider the multiple scatter channels in massive MIMO.

References

[1] Y. Sun, J. Liu, J. Wang, Y. Cao, and N. Kato, "When machine learning meets privacy in 6G: A survey," *IEEE Commun. Surv. Tut.*, vol. 22, no. 4, pp. 2694–2724, Oct–Dec. 2020.

[2] W. Wang, N. Cheng, K. C. Teh, X. Lin, W. Zhuang, and X. Shen, "On countermeasures of pilot spoofing attack in massive MIMO systems: A double channel training based approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6697–6708, Jul. 2019.

[3] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing-the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[4] J. Nam, A. Adhikary, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.

[5] Y. Jeon, C. Song, S.-R. Lee, S. Maeng, J. Jung, and I. Lee, "New beamforming designs for joint spatial division and multiplexing in large-scale MISO multi-user systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3029–3041, May 2017.

[6] Y. Song, C. Liu, Y. Liu, N. Cheng, Y. Huang, and X. Shen, "Joint spatial division and multiplexing in massive MIMO: A neighbor-based approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7392–7406, Nov. 2020.

[7] Y. Song, C. Liu, W. Wang, and Y. Huang, "Cluster-group-based two-stage beamforming for massive MIMO," *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 1984–1998, Mar. 2022.

[8] M. B. Khalilsarai, S. Haghighatshoar, X. Yi, and G. Caire, "FDD massive MIMO via UL/DL channel covariance extrapolation and active channel sparsification," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 121–135, Jan. 2019.

[9] Z. Yin et al., "UAV-assisted physical layer security in multi-beam satellite-enabled vehicle communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 2739–2751, Mar. 2022.

[10] Z. Yin, N. Cheng, T. H. Luan, Y. Hui, and W. Wang, "Green interference based symbiotic security in integrated satellite-terrestrial communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9962–9973, Nov. 2022.

[11] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[12] L. You, K.-X. Li, J. Wang, X. Gao, X.-G. Xia, and B. Ottersten, "Massive MIMO transmission for LEO satellite communications," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1851–1865, Aug. 2020.

[13] F. Zhu, F. Gao, S. Jin, H. Lin, and M. Yao, "Robust downlink beamforming for BDMA massive MIMO system," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1496–1507, Apr. 2018.

[14] W. Shen, X. Bu, X. Gao, C. Xing, and L. Hanzo, "Beamspace precoding and beam selection for wideband millimeter-wave MIMO relying on lens antenna arrays," *IEEE Trans. Signal Process.*, vol. 67, no. 24, pp. 6301–6313, Dec. 2019.

[15] S. Han, C. L. I., C. Rowell, Z. Xu, S. Wang, and Z. Pan, "Large scale antenna system with hybrid digital and analog beamforming structure," in *Proc. IEEE Int. Conf. Commun.*, 2014, pp. 842–847.

[16] M. Fouda, S. Hashima, S. Sakib, Z. Fadlullah, K. Hatano, and X. Shen, "Optimal channel selection in hybrid RF/VLC networks: A multi-armed bandit approach," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6853–6858, Jun. 2022.

[17] Y. Gai, B. Krishnamachari, and R. Jain, "Combinatorial network optimization with unknown variables: Multi-armed bandits with linear rewards," *IEEE/ACM Trans. Netw.*, vol. 62, no. 17, pp. 4505–4515, Sep. 2014.

[18] R. Vershynin, *High-Dimensional Probability-an Introduction with Applications in Data Science*. Irvine, CA, USA: Univ. California, 2020.