Thompson Sampling-Based Antenna Selection With Partial CSI for TDD Massive MIMO Systems

Zhenran Kuai and Shaowei Wang, *Senior Member, IEEE*

***Abstract*— Antenna selection (AS) is a promising technology that can reduce the implementation complexity and hardware cost of a massive MIMO system, in which a part of the attainable antennas are selected and connected to the radio frequency chains in each time slot. In this article, we propose a low-complexity AS algorithm to maximize the downlink achievable rate in the time-division duplexing massive MIMO system, which is based on online Thompson sampling technique and significantly reduces the pilot overhead required for channel estimation with only partial channel state information. We prove that the distribution-dependent upper bound of the proposed algorithm is sub-linear as a function of time slot by introducing the concept of *regret*. We also develop three discounting factors to accommodate the large-scale variations across antennas. Numerical results show that the downlink achievable rate can be greatly improved with our proposed scheme as compared to other typical ones.**

***Index Terms*— Antenna selection, combinatorial multi-armed bandit, massive MIMO, online learning, Thompson sampling.**

I. Introduction

MASSIVE MIMO has been considered as a leading technology that provides a performance boost for wireless communication systems [1]. By deploying large-scale antenna arrays at base stations (BSs), massive MIMO can serve multiple single-antenna users with the same time-frequency resources, achieving significant improvements in spectral and energy efficiency [2]. However, the radio frequency (RF) chain deployed at the BSs, including power amplifiers, frequency mixers, AD/DA converters, etc., also increases with the number of antennas, which leads to a high implementation complexity and hardware cost [3]. In the extreme scenario such as millimeter wave communications [4], the tight space among antennas and the intensive power consumption result in compact hardware layout and insufficient thermal dissipation, which becomes a big challenge for actualizing massive MIMO systems.

Hybrid signal processing techniques for reducing the complexity and the cost of massive MIMO systems have been extensively studied in the literature [5], among which the antenna selection (AS) adopts simple RF switches to achieve a low hardware cost and power consumption while being benefitted from the diversity gains of antenna arrays [6]. For the conventional MIMO, the AS can be well resolved since there are only few antennas to be dealt with [7]. However, it becomes computationally intractable for massive MIMO systems since the combinatorial complexity increases exponentially over antennas [8]. Therefore, low-complexity AS algorithm is urgently needed for massive MIMO systems and has engrossed attentions in both academia and industry. In [9], a joint transmitting and receiving AS scheme is proposed, where genetic algorithm, which uses a priority mechanism in mutation and crossover processes, is introduced to solve the formulated problem. In [10], AS is formulated as an integer programming problem and is solved by convex relaxation technique. In [11], a heuristic AS method is developed by sorting and shifting techniques on channel matrix, which achieves a higher sum rate compared with using all antennas in a reference system. In [12], channel correlation is exploited, based on which a two-step AS algorithm is developed. In [13], trace-based algorithms are introduced to reduce the complexity of the AS problem for both single-cell and multiple-cell massive MIMO systems. The AS with fixed power allocation is formulated as a maximum entropy sampling problem in [14] by leveraging the submodularity, based on which a greedy algorithm with low complexity is developed. The authors in [15] employ the AS technique to optimize constructive interference at the receivers so as to improve the system energy efficiency. In [16], the AS for capacity maximization is transformed into a sequential decision-making problem and a Monte Carlo tree search method is proposed to select antennas.

As far as the authors have known, these schemes rely heavily on perfect channel state information (CSI), which means the channel coefficients are complete and totally correct. In a massive MIMO system, CSI is obtained by processing the received pilot sequences, which generally generates radio resource consumption proportional to both the numbers of users and antennas. On the other hand, if the massive MIMO system is equipped with a reduced number of RF chains, such as the AS scenario, only partial CSI can be obtained from each pilot transmission. Acquiring full CSI inevitably generates heavy signalling overhead due to multiple times of pilot transmissions. Consider a typical downlink time-division duplexing (TDD) massive MIMO system with 64 antennas and 4 users. An uplink pilot sequence occupies 0.071 ms in each

Manuscript received January 7, 2020; revised April 25, 2020 and August 9, 2020; accepted September 7, 2020. Date of publication September 15, 2020; date of current version December 16, 2020. This work was partially supported by the National Natural Science Foundation of China under Grants 61931023, 61671233 and U1936202. The associate editor coordinating the review of this article and approving it for publication was H. Suraweera. *(Corresponding author: Shaowei Wang.)*

The authors are with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: dz1923021@smail.nju.edu.cn; wangsw@nju.edu.cn).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2020.3024199

time slot with total 0.5 ms duration [17]. If the BS is equipped with only 16 RF chains for the AS, the time of uplink pilot transmission to obtain full CSI is 0.284 ms which is more than half of a time slot. Moreover, CSI imperfection caused by the effect of noise should be taken into account [18].

In this article, we investigate the transmit AS problem in the TDD massive MIMO system. To reduce the pilot overhead of full CSI requirement, we propose a slot structure for acquiring partial CSI. The transmitting AS task is then formulated as a combinatorial multi-armed bandit (CMAB) problem, and an online decision-making technique, referred to as Thompson sampling, is adopted to select a set of active antennas in each time slot. The introduced online learning algorithm achieves a trade-off between the exploration and the exploitation of the subset of active antennas. Analysis show that it can approach the antenna subset with the highest average contribution to channel capacity in hindsight as time slot goes, which is also confirmed by numerical results. We also introduce three discounting factors such that the proposed online method can address the scenarios of the large-scale variations of wireless channels. The contributions of this article are summarized as follows:

- We propose a novel slot structure with partial CSI for AS in TDD massive MIMO systems, by which the pilot overhead of the system for channel estimation is significantly cut down and the savings of the slot can be exploited for data transmission.
- We employ online learning technique, referred to as Thompson sampling, to address the formulated optimization task effectively and efficiently, which distinguishes the contributions of the antennas adaptively and selects the promising subset of them so that the system throughput can be guaranteed with limited RF chains.
- We introduce the discounting factors to deal with dynamic scenarios in practical environment, which can alleviate the effects of large-scale fading variations over antennas. Numerical experiments show that the proposed method can track the contribution variations of the antennas and yield higher throughput as compared to conventional convex relaxation or power-based algorithms.
- We prove that the *regret* of the proposed algorithm is sublinear as a function of time slot, which implies that its asymptotical performance is close to the antenna set with the highest average contribution in hindsight as time slot goes.

The rest of the paper is organized as follows. In Section II, we propose the slot structure with partial CSI, and show that the AS in massive MIMO can be formulated as a CMAB problem. In Section III, we give the details of the online AS algorithms, as well as the theoretical analysis of *regret*. In Section IV, the proposed algorithms are validated by numerical experiments. We conclude our work in Section V.

Notations: We adopt the following notations throughout this article. Non-bold, bold lower-case and bold upper-case letters are used to denote scalars, vectors, and matrices, respectively. $(\cdot)^H$ represents the Hermitian of a matrix, while $\mathbf{I}_N$ represents an $N \times N$ identity matrix. $\mathbf{tr}(\cdot)$ represents the trace of a matrix. $\mathbb{C}^N$ is an $N-$dimensional complex Euclidean space and the Calligraphic font is used to denote finite sets. $\|\cdot\|$ represents Euclidean norm of a vector. $\mathbb{E}[\cdot]$ represents expectation operator on random variable. $O(\cdot)$ represents the asymptotic growth rate of a function. $\mathbf{1}\{\cdot\}$ is indicator function which equals to 1 if an event holds or 0 otherwise. $\lceil\cdot\rceil$ rounds a variable to the nearest integer no less than itself. $[k]$ represents the shorthand form of set $\{1, 2, \ldots, k\}$.

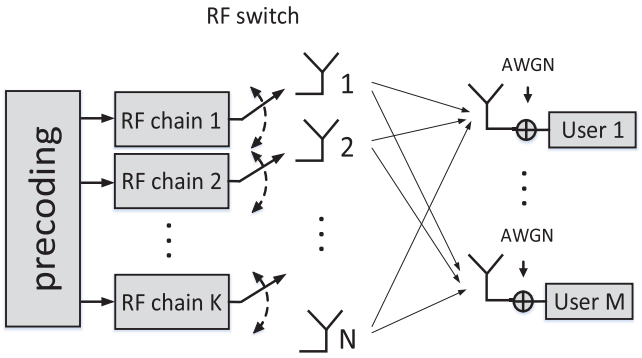


Fig. 1. Transmit antenna selection in massive MIMO systems.

II. System Model and Problem Formulation

A. System Model

Consider the downlink transmission of a single-cell multi-user massive MIMO-OFDM system, where the BS is equipped with K RF chains and N antennas, and M single-antenna mobile users are served simultaneously in TDD mode. As shown in Fig. 1, a subset of K antennas is selected to connect the RF chains in each time slot. The set of available antennas at BS is denoted by $[N] = \{1, 2, \ldots, N\}$, and the set of selected antennas is denoted by $\mathcal{K} \subset [N]$. Denote $\mathcal{S}$ as the set of possible antenna subsets, i.e., $\mathcal{S} = \{s \subset [N] \mid |s| = K\}$, we have $\mathcal{K} \in \mathcal{S}$.

Suppose that the total number of subcarriers in the massive MIMO system is U. Every U/L consecutive subcarriers are grouped into a sub-band, so there are L sub-bands. Before the downlink transmission, the users first transmit the uplink pilot OFDM symbols to the BS, which are received by the BS to estimate channel matrix $\hat{\mathbf{H}}_l^{\mathcal{K}}$. Here we denote the true channel matrix by $\mathbf{H}_l^{\mathcal{K}} = [\mathbf{h}_{1,l}^{\mathcal{K}} \cdots \mathbf{h}_{M,l}^{\mathcal{K}}]^T$, where $\mathbf{H}_l^{\mathcal{K}} \in \mathbb{C}^{M \times K}$ represents the channel matrix for antenna subset $\mathcal{K}$ at the subcarrier in sub-band l [19]. Orthogonal uplink pilot sequences are used for different users in one allocated subcarrier for each sub-band l while the remaining subcarriers on the sub-band are preserved. Once the channel is estimated for user m by the pilot on the allocated subcarrier, it can be used for the other subcarriers in the sub-band. Meanwhile, we consider the effect of noise on CSI imperfection. Each user transmits a pilot signal sequence $\Phi_{m,l} \in \mathbb{C}^{\tau_p}$ to the BS. The pilot sequence has unit-magnitude elements, i.e., $\Phi_{m,l}^H \Phi_{m,l} = \tau_p$. The received signal $\mathbf{Y}_l^{\mathcal{K}} \in \mathbb{C}^{K \times \tau_p}$ is given by

$$\mathbf{Y}_l^{\mathcal{K}} = \sum_{m=1}^{M} \sqrt{\rho_{ul}} \mathbf{h}_{m,l}^{\mathcal{K}} \Phi_{m,l}^H + \mathbf{N}_l, \tag{1}$$

where ρ_{ul} is the transmit power. $\mathbf{N}_l$ is the noise matrix whose entries are i.i.d $\mathcal{CN}(0,1)$. The BS can multiply $\mathbf{Y}_l^{\mathcal{K}}$ with $\Phi_{m,l}/\sqrt{\tau_p}$, leading to the processed pilot signal, given as

$$\mathbf{y}_{m,l}^{\mathcal{K}} = \sqrt{\rho_{ul}\tau_p}\mathbf{h}_{m,l}^{\mathcal{K}} + \mathbf{n}_{m,l}, \tag{2}$$

where the entries of the noise vector, $\mathbf{n}_{m,l}$, are also i.i.d $\mathcal{CN}(0,1)$. The channel vector $\mathbf{h}_{m,l}^{\mathcal{K}}$ of the m-th user is modeled as

$$\mathbf{h}_{m,l}^{\mathcal{K}} = \tilde{\mathbf{R}}_{m,l}^{\mathcal{K}}\mathbf{v}_{m,l}^{\mathcal{K}}, \tag{3}$$

where $\mathbf{R}_{m,l}^{\mathcal{K}} \triangleq \tilde{\mathbf{R}}_{m,l}^{\mathcal{K}}(\tilde{\mathbf{R}}_{m,l}^{\mathcal{K}})^H \in \mathbb{C}^{K\times K}$ is the spatial correlation matrix [20]. $\mathbf{v}_{m,l}^{\mathcal{K}} \sim \mathcal{CN}(0,\mathbf{I}_K)$ represents the independent fast-fading channel vector. Thus the MMSE estimation $\hat{\mathbf{h}}_{m,l}^{\mathcal{K}}$ of $\mathbf{h}_{m,l}^{\mathcal{K}}$ is given by

$$\hat{\mathbf{h}}_{m,l}^{\mathcal{K}} = \sqrt{\rho_{ul}\tau_p}\mathbf{R}_{m,l}^{\mathcal{K}}\mathbf{Q}_{m,l}^{\mathcal{K}}\mathbf{y}_{m,l}^{\mathcal{K}}, \tag{4}$$

and $\hat{\mathbf{h}}_{m,l}^{\mathcal{K}} \sim \mathcal{CN}(0,\mathbf{\Psi}_{m,l}^{\mathcal{K}})$. The matrices $\mathbf{\Psi}_{m,l}^{\mathcal{K}}$ and $\mathbf{Q}_{m,l}^{\mathcal{K}}$ are defined as

$$\begin{aligned}\mathbf{\Psi}_{m,l}^{\mathcal{K}} &= \rho_{ul}\tau_p\mathbf{R}_{m,l}^{\mathcal{K}}\mathbf{Q}_{m,l}^{\mathcal{K}}\mathbf{R}_{m,l}^{\mathcal{K}},\\ \mathbf{Q}_{m,l}^{\mathcal{K}} &= \left(\mathbf{I}_K + \rho_{ul}\tau_p\mathbf{R}_{m,l}^{\mathcal{K}}\right)^{-1}.\end{aligned}$$

$\tilde{\mathbf{h}}_{m,l}^{\mathcal{K}} = \mathbf{h}_{m,l}^{\mathcal{K}} - \hat{\mathbf{h}}_{m,l}^{\mathcal{K}}$ is the estimation error and $\tilde{\mathbf{h}}_{m,l}^{\mathcal{K}} \sim \mathcal{CN}(0,\mathbf{R}_{m,l}^{\mathcal{K}} - \mathbf{\Psi}_{m,l}^{\mathcal{K}})$.

For the allocated subcarrier in sub-band l, the BS transmits Q OFDM symbols in each time slot. Denote $\mathbf{s}_{l,q} \in \mathbb{C}^M$ as the q-th OFDM symbol vector of M users, and normalize it as $\mathbb{E}[\|\mathbf{s}_{l,q}\|^2] = 1$. Denote $\mathbf{W}_{l,q}^{\mathcal{K}} \in \mathbb{C}^{K\times M}$ as the precoding matrix of the BS for any antenna subset $\mathcal{K}$, the transmit signal vector $\mathbf{x}_{l,q} \in \mathbb{C}^K$ is then given by

$$\mathbf{x}_{l,q} = \sqrt{\lambda}\mathbf{W}_{l,q}^{\mathcal{K}}\mathbf{s}_{l,q}. \tag{5}$$

λ is a normalization constant following $\mathbb{E}[\|\mathbf{x}_{l,q}\|^2] = 1$, i.e.,

$$\lambda = \frac{1}{\mathbb{E}\left[\mathbf{tr}\left(\mathbf{W}_{l,q}^{\mathcal{K}}(\mathbf{W}_{l,q}^{\mathcal{K}})^H\right)\right]}. \tag{6}$$

Assume zero-forcing is utilized for precoding, we have

$$\mathbf{W}_{l,q}^{\mathcal{K}} = \left(\hat{\mathbf{H}}_{l,q}^{\mathcal{K}}\right)^H\left(\hat{\mathbf{H}}_{l,q}^{\mathcal{K}}\left(\hat{\mathbf{H}}_{l,q}^{\mathcal{K}}\right)^H\right)^{-1}. \tag{7}$$

The received OFDM symbol vector $\mathbf{y}_{l,q} \in \mathbb{C}^M$ of all K users is given by

$$\mathbf{y}_{l,q} = \sqrt{M\rho_{dl}}\mathbf{H}_{l,q}^{\mathcal{K}}\mathbf{x}_{l,q} + \mathbf{n}_{l,q}, \tag{8}$$

where $\mathbf{n}_{l,q}$ represents the zero-mean circularly symmetric complex Gaussian noise vector with unit variance, i.e., $\mathbf{n}_{l,q} \sim \mathcal{CN}(0,\mathbf{I}_M)$, and ρ_{dl} represents the normalized transmit SNR per user. To fix the transmit power per user, we multiply ρ_{dl} by M such that the total transmit power $M\rho_{dl}$ increases with the number of users, and the average transmit power per user is ρ_{dl}.

Notice that $\mathbf{H}_{l,q}^{\mathcal{K}} = \tilde{\mathbf{H}}_{l,q}^{\mathcal{K}} + \hat{\mathbf{H}}_{l,q}^{\mathcal{K}}$, by substituting (7) into (8), we have the received signal

$$\mathbf{y}_{l,q} = \sqrt{M\rho_{dl}\lambda}\mathbf{s}_{l,q} + \mathbf{n}_{l,q} + \sqrt{M\rho_{dl}}\tilde{\mathbf{H}}_{l,q}^{\mathcal{K}}\mathbf{x}_{l,q}. \tag{9}$$

The effective signal-to-interference-plus-noise ratio (SINR) at the q-th OFDM symbol for the subcarrier in sub-band l received by the m-th user is given by

$$\mathsf{SINR}_{l,q}^{\mathcal{K}}[m] = \frac{M\rho_{dl}\lambda}{M\rho_{dl}\mathbf{tr}\left(\mathbf{R}_{m,l}^{\mathcal{K}} - \mathbf{\Psi}_{m,l}^{\mathcal{K}}\right) + 1}. \tag{10}$$

The capacity bound in the l-th sub-band with the selected antenna subset $\mathcal{K}$ can be calculated by

$$C_l(\mathcal{K}) = \frac{1}{Q}\sum_{q=1}^{Q}\sum_{m=1}^{M}\log_2\left(1 + \mathsf{SINR}_{l,q}^{\mathcal{K}}[m]\right). \tag{11}$$

In each time slot, the system capacity averaged over L sub-bands with the selected antenna subset $\mathcal{K}$ is bounded by

$$C(\mathcal{K}) = \frac{1}{QL}\sum_{q=1}^{Q}\sum_{l=1}^{L}\sum_{m=1}^{M}\log_2\left(1 + \mathsf{SINR}_{l,q}^{\mathcal{K}}[m]\right). \tag{12}$$

B. Full CSI Vs. Partial CSI

In a TDD massive MIMO system, the users transmit pilot sequences in uplink and the BS utilizes the pilot signals to estimate CSI. Then the users transmit uplink data to the BS. Downlink data is precoded and transmitted based on the estimated CSI after a guard period [21]. As shown in Fig. 2(a), when AS is utilized with a reduced number of RF chains, the uplink pilot should be transmitted $\lceil N/K \rceil$ times to achieve full CSI. Denote T_s and T_p as the length of a time slot and a pilot sequence, respectively. T_{UL} represents the length of uplink data transmission, and T_g represents the length of a switch guard. B_l is used to denote the bandwidth of the sub-band l. Given any selected antenna subset $\mathcal{K}_f \in \mathcal{S}$ to acquire full CSI, the average throughput of TDD massive MIMO with the conventional full CSI structure is given by

$$R_f(\mathcal{K}_f) = \frac{T_s - \left\lceil\frac{N}{K}\right\rceil T_p - T_{UL} - 2T_g}{T_s}\sum_l C_l(\mathcal{K}_f)B_l. \tag{13}$$

Obviously, extra pilot overheads waste radio resources that could be utilized for downlink data transmission. Here, we propose a simplified slot structure with partial CSI, as shown in Fig. 2(b), by which channel estimation is performed only for the current active antenna subset $\mathcal{K}_p$ in a slot. Therefore, for any online strategy selecting antenna subset $\mathcal{K}_p \in \mathcal{S}$, the average throughput of the massive MIMO system with partial CSI is given by

$$R_p(\mathcal{K}_p) = \frac{T_s - T_p - T_{UL} - 2T_g}{T_s}\sum_l C_l(\mathcal{K}_p)B_l. \tag{14}$$

Note that the antenna subset $\mathcal{K}_f$ with full CSI usually achieves a higher capacity than the antenna subset $\mathcal{K}_p$ with partial CSI that needs to transmit pilot sequences multiple times. However, by leveraging the outputs of previous selections, the antennas with the highest contribution to channel capacity can be selected to make the capacity with partial CSI $C(\mathcal{K}_p)$ approach that with full CSI $C(\mathcal{K}_f)$. Since the

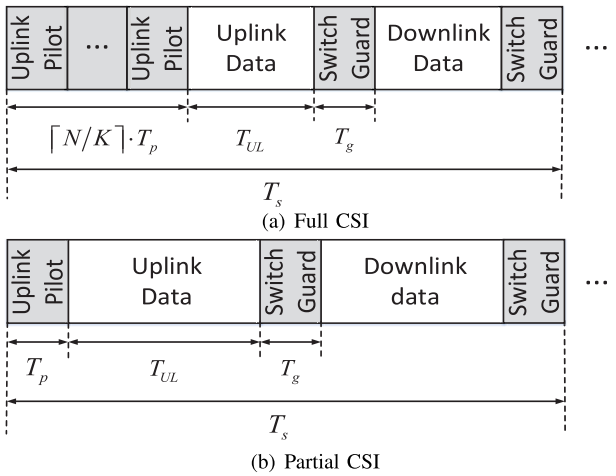


Fig. 2. Slot structure with full and partial CSI AS in TDD massive MIMO.

pilot overhead is highly reduced in the slot structure with partial CSI, the throughput with partial CSI $R_p(\mathcal{K}_p)$ may exceed that with full CSI $R_f(\mathcal{K}_f)$ [22]. Here we fix the length of the uplink data to study the influence of pilot overhead on the downlink throughput of AS. Recall that the uplink transmission can also be affected by pilot overhead, in this case, it can be transformed into studying the influence of AS on uplink rate with finite length of slot.

C. Combinatorial Multi-Armed Bandit

In the AS problem, K out of N antennas are selected for channel estimation and transmission in each time slot. Given selected antennas, channel capacity is related to the outputs of AS which are utilized for the following selections. The procedure of the AS can be formulated as a CMAB problem that can be analyzed conveniently.

In round $t = 1, \ldots, T$ of a multi-armed bandit (MAB) problem, only one arm is chosen to play from N arms, where arm i obeys a distribution of reward r_i with unknown mean μ_i. Take action $i(t)$, the player observes a reward $r_{i(t)}$. The target is to maximize the total expected reward with an online strategy π, where $\pi = \{\pi(t)\}_{t\geq 1}$ is a series of sequential policies, and $\pi(t)$ is the mapping of the previous observations of rewards $r_{i(1)}, \cdots, r_{i(t-1)}$ to the current action $i(t)$. If multiple arms are selected simultaneously in each round, then the MAB is referred to as a CMAB [23]. The action is represented by a subset of arms $\mathcal{K}(t)$ and the corresponding total reward is denoted by $r_{\mathcal{K}(t)}$.

Definition 1: For a CMAB problem with N arms, a subset of K arms is selected in each round and the reward of each arm is associated with a specific probability distribution with mean μ_i. The *regret* of any strategy π is defined as the difference between its cumulative expected rewards and the total expected rewards of the optimal arm subset $\mathcal{K}^* = \arg\max_{\mathcal{K}}\{\sum_{i\in\mathcal{K}} \mu_i\}$ until round T,

$$Reg^{\pi}(T) = T \sum_{i\in\mathcal{K}^*} \mu_i - \sum_{t=1}^{T} \sum_{i\in\mathcal{K}(t)} \mu_i. \tag{15}$$

For the AS problem in massive MIMO systems, the N available antennas can be regarded as N arms. In each slot t, a subset of attainable antennas $\mathcal{K}(t) \in \mathcal{S}$ is selected to maximize its average contribution to the system capacity. To this end, we divide the total capacity $C(\mathcal{K}(t))$ into K equal parts, each of which is defined as the reward of any arm $i \in \mathcal{K}(t)$:

$$r(t) = r_i(t) = \frac{1}{K} C(\mathcal{K}(t)). \tag{16}$$

Assume that $C(\mathcal{K}(t)) \in [C_{min}, C_{max}]$, thus $r_i(t) \in [C_{min}/K, C_{max}/K]$. Denote $r_{max} = C_{max}/K$ and $r_{min} = C_{min}/K$ as the upper bound and the lower bound of $r_i(t)$, respectively. Further, we normalize $r_i(t)$ on $[C_{min}/K, C_{max}/K]$ such that $\hat{r}_i(t) \in [0, 1]$, i.e.,

$$\hat{r}_i(t) = \frac{r_i(t) - r_{min}}{r_{max} - r_{min}}. \tag{17}$$

The normalized antenna-specific reward $\hat{r}_i(t)$ of each antenna $i \in [N] = \{1, 2, \ldots, N\}$ is complied with a fixed but unknown probability distribution $\varphi(\mu_i)$, where $\mu_i \in [0, 1]$ is the mean of the normalized reward. Maximizing the average throughput is equivalent to minimizing the difference between the cumulative expected normalized rewards of the AS strategy and the total expected normalized rewards of the optimal antenna subset $\mathcal{K}^* = \arg\max_{\mathcal{K}}\{\sum_{i\in\mathcal{K}} \mu_i\}$ in hindsight until T slots. The *regret* can be regarded as the cumulative differences of the expected rewards between the antenna set chosen by the AS strategy and the antenna set with the largest average contribution to the system capacity:

$$Reg^{AS}(T) = T \sum_{i\in\mathcal{K}^*} \mu_i - \sum_{t=1}^{T} \sum_{i\in\mathcal{K}(t)} \mu_i. \tag{18}$$

Then the AS in massive MIMO systems can be formulated as a CMAB problem in this way.

III. Online Antenna Selection Based on Thompson Sampling

We develop an online AS algorithm based on Thompson sampling to address the formulated optimization task and provide the upper bound on the *regret*. Discounting factors are also introduced to extend the proposed algorithm to accommodate large-scale variations across antennas in practical radio propagation environment.

A. Online Antenna Selection Using Thompson Sampling

Thompson sampling can achieve a good trade-off between exploration and exploitation for the MAB problems as discussed in [24]. For each arm, the mean value of the reward obeys a prior distribution. Sampling results from these prior distributions are the references for selecting arms, whose rewards are used to update the parameters of the corresponding prior distributions by Bayesian rule for the next round of selection. The random sampling facilitates the process of exploration, and the prior reward distribution for sampling exploits the outputs of selection.

For each antenna, the corresponding reward is a random variable with an unknown distribution. In each slot, the information about reward distribution should be extracted by utilizing the updated reward [25], [26]. Thompson sampling uses Beta distribution to estimate the mean of the reward. Specifically, the mean reward μ_i for antenna i is approximated by $\Theta_i \sim Beta(S_i, F_i)$ with the probability density function

$$p_i(\Theta_i) = \frac{\Gamma(S_i + F_i)}{\Gamma(S_i)\Gamma(F_i)} \Theta_i^{S_i - 1} (1 - \Theta_i)^{F_i - 1}, \quad (19)$$

where Γ is Gamma function, and S_i, F_i are the parameters of Beta distribution. Beta distribution is chosen as the prior distribution due to its conjugacy properties. Either of its parameters can be increased by 1 to show its deviation and convergence to the mean value of the normalized reward. Note that the random variable Θ_i is distributed in $[0, 1]$ with mean $S_i/(S_i+F_i)$ and variance $S_iF_i/(S_i+F_i+1)(S_i+F_i)^2$. S_i and F_i are updated in each time slot such that the mean reward μ_i of antenna i can be approximated accurately as time slot goes.

Given the CSI of selected antenna subset $\mathcal{K}(t)$, we calculate the normalized reward $\hat{r}_i(t)$ by (17). For antenna $i \in \mathcal{K}(t)$, we take an independent sample $b_i(t)$ from Bernoulli distribution with success probability $\hat{r}_i(t)$ and use the samples to update the parameters of Beta distribution [27]

$$\begin{aligned} S_i(t+1) &= S_i(t) + b_i(t), \\ F_i(t+1) &= F_i(t) + 1 - b_i(t). \end{aligned} \quad (20)$$

If $\hat{r}_i(t)$ is close to 1, which implies that $\mathcal{K}(t)$ produces high capacity, there is a high probability that $S_i(t+1)$ increases by 1 while $F_i(t+1)$ keeps unchanged, and the mean of Beta distribution increases accordingly. Otherwise, if $\hat{r}_i(t)$ is close to 0, which implies that $\mathcal{K}(t)$ has poor performance, there is a high probability that $F_i(t+1)$ increases by 1 while $S_i(t+1)$ keeps unchanged, and the mean of Beta distribution decreases accordingly. As the variance decreases with $S_i(t) + F_i(t)$, the sampling value of Beta distribution $Beta(S_i(t), F_i(t))$ approaches μ_i.

In time slot $t+1$, independent sample is drawn from the Beta distribution for each antenna, i.e. $\theta_i(t+1) \sim Beta(S_i(t+1), F_i(t+1))$, which is regarded as an approximation $\hat{r}_i(t)$. Using Beta distribution to estimate the mean value of the reward, the sampling results can represent the quality of the corresponding antennas and the sampling process facilitates the exploration. The set of antennas that maximizes the total expected rewards is selected for downlink transmission:

$$\mathcal{K}(t+1) = \arg\max_{\mathcal{K} \in \mathcal{S}} \sum_{i \in \mathcal{K}} \theta_i(t+1). \quad (21)$$

If the antennas in $\mathcal{K}(t)$ are selected from Beta distributions with a higher mean value, we can exploit it to achieve a higher throughput. If the Beta distribution of the selected antenna has a lower mean value, its parameters can be updated for a better description of the antenna-specific reward, which explores the potential antennas. In brief, the online AS algorithm can achieve a balance between the exploration of antennas and the exploitation of well-performing antennas. The procedure of the proposed online AS is summarized in **Algorithm 1**.

Algorithm 1 Online Antenna Selection

1: Initialize $S_i(0) = 1$, $F_i(0) = 1$ for antenna $i \in [N]$.
2: **for** $t = 1, 2, 3 \cdots$ **do**
3: Sample $\theta_i(t)$ from $Beta(S_i(t-1), F_i(t-1))$ for antenna $i \in [N]$;
4: Select antenna set $\mathcal{K}(t) \in \mathcal{S}$ with (21);
5: Calculate $\hat{r}_i(t)$ using (17) for $i \in \mathcal{K}(t)$;
6: **for** antenna $i \in \mathcal{K}(t)$ **do**
7: Sample $b_i(t)$ from Bernoulli distribution with success probability $\hat{r}_i(t)$;
8: Update $S_i(t)$, $F_i(t)$ using (20);
9: **end for**
10: **end for**

B. Regret Analysis

Without loss of generality, we assume that the means of the rewards for all antennas are distinct from each other and have been sorted, i.e., $\mu_1 > \mu_2 > \cdots > \mu_N$. Recall that the average contribution of each antenna to the system capacity is normalized as the antenna-specific reward. Our target is to select antennas with the largest average contribution to the overall capacity. The *regret* defined by (18) is the cumulative differences of the rewards between the antenna set chosen by the proposed AS scheme and the antenna set with the largest average contribution in hindsight.

Theorem 1: Consider selecting K from N antennas, the expected *regret* of the online AS algorithm is bounded by

$$\mathbb{E}[Reg^{AS}(T)] \leq \sum_{i \in [N] \setminus [K]} \frac{\Delta_{i,K} \log T}{\epsilon_1 d^B(\mu_i, \mu_K)} + J_\mu(T) + C_\mu. \quad (22)$$

where ϵ_1 is a sufficiently minimal positive. C_μ is a constant independent of T and is related to the mean rewards of antennas μ_i's. $J_\mu(T)$ is also related to the mean rewards μ_i's and achieves the growth rate of order $O(\log \log T)$. $d^B(p, q)$ is the Kullback-Leibler divergence between two Bernoulli distributions with means p and q, respectively. The set of the optimal K antennas with the highest average contribution is denoted by $[K]$ and the set of the rest antennas is $[N] \setminus [K]$.

Proof of Theorem 1: For a sufficiently minimal $\delta > 0$, define $\mu_K^{(-)} = \mu_K - \delta$ to satisfy $\mu_K^{(-)} \in (\mu_{K+1}, \mu_K)$. For antenna $i \in [N] \setminus [K]$, define $\mu_i^{(+)} = \mu_i + \delta$ to satisfy $\mu_i^{(+)} \in (\mu_i, \mu_K)$. $L_i(t)$ is the number of times that antenna i is selected before slot t, i.e. $L_i(t) = \sum_{t'=1}^{t-1} \mathbf{1}\{i \in \mathcal{K}(t')\}$. We define $\theta_{\mathcal{S}}^{(K)}(t)$ as the K-th largest sample in $\mathcal{S}$. Let $\nu = (\mu_{K-1} + \mu_K)/2$. Thus, we can define the events as follows:

$$\begin{aligned} \mathcal{A}_i(t) &= \{i \in \mathcal{K}(t)\}, \\ \mathcal{B}(t) &= \{\theta_{[N]}^{(K)}(t) \geq \mu_K^{(-)}\}, \\ \mathcal{C}_i(t) &= \bigcap_{j \in [N] \setminus ([K-1] \cup \{i\})} \left\{\theta_{[N] \setminus \{i,j\}}^{(K-1)}(t) \geq \nu\right\}, \\ \mathcal{D}_i(t) &= \left\{L_i(t) < \frac{\log T}{d^B(\mu_i^{(+)}, \mu_K^{(-)})}\right\}, \end{aligned} \quad (23)$$

where $\mathcal{A}_i(t)$ indicates that antenna i is selected in slot t. $\mathcal{B}(t)$ indicates that the K-th largest sample is not less than $\mu_K^{(-)}$. $\mathcal{C}_i(t)$ indicates that antenna i in the optimal subset $[K]$ is not included in $\mathcal{K}(t)$. $\mathcal{D}_i(t)$ means that the number of selections is not large enough to measure the average contribution for antenna i accurately.

The reward loss between the suboptimal antenna i and the antenna with the highest average contribution in $[K]$ excluded by $\mathcal{K}(t)$ is given by

$$\Delta_i(t) = \begin{cases} x = \left(\max_{j\in[K]\setminus\mathcal{K}(t)} \mu_j\right) - \mu_i & \mathcal{K}(t) \neq [K], \\ 0 & otherwise. \end{cases} \tag{24}$$

The maximum *regret* of antenna $i \in [N] \setminus [K]$ is

$$Reg_i^{max}(T) = \sum_{t=1}^{T} \mathbf{1}\{i \in \mathcal{K}(t)\}\Delta_i(t).$$

Then the *regret* is bounded by the total maximum *regret* of antennas in $[N] \setminus [K]$:

$$Reg^{AS}(T) \leq \sum_{i\in[N]\setminus[K]} Reg_i^{max}(T).$$

We use the events given by (23) to divide $Reg_i^{max}(T)$ into different parts for proof.

The maximum *regret* of suboptimal antennas $i \in [N] \setminus [K]$ can be resolved into:

$$\begin{aligned} &Reg_i^{max}(T) \\ &\leq \underbrace{\sum_{t=1}^{T} \mathbf{1}\{\mathcal{B}^c(t)\}}_{\mathsf{A}} + \underbrace{\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{C}_i^c(t)\}}_{\mathsf{B}} \\ &\quad + \underbrace{\sum_{j\in[N]\setminus([K-1]\cup\{i\})} \sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{C}_i(t), \mathcal{D}_i(t), \mathcal{A}_j(t)\}}_{\mathsf{C}} \\ &\quad + \underbrace{\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{B}(t), \mathcal{D}_i^c(t)\}}_{\mathsf{D}} + \frac{\log T}{d^B(\mu_i^{(+)}, \mu_K^{(-)})}\Delta_{i,K}, \end{aligned} \tag{25}$$

where $\mathcal{A}^c$ is the complement of $\mathcal{A}$ and $\{\mathcal{A}, \mathcal{B}\}$ is the shorthand for $\{\mathcal{A}\cap\mathcal{B}\}$. $\Delta_{i,K}$ means that only antenna K is excluded from $\mathcal{K}(t)$. The proof of (25) can be referred to Lemma 2 in [28].

On the right side of (25), term A indicates that some antennas in $[K]$ are excluded from $\mathcal{K}(t)$ in slot t. Term B indicates the antenna i is in $\mathcal{K}(t)$ but some in $[K-1]$ are not. Term C indicates that antennas $i \in [N] \setminus [K]$ and $j \in [N] \setminus ([K-1] \cup \{i\})$ are simultaneously selected. Term D indicates that the antenna is selected with enough information about the normalized reward.

For sufficiently minimal $\epsilon_2 > 0$ and $\delta > 0$, the terms $\mathsf{A} - \mathsf{D}$ can be bounded as follows:

$$\mathbb{E}[\mathsf{A}] = O\left(\frac{1}{(\mu_k - \mu_K^{(-)})^2}\right) = O\left(\frac{1}{\delta^2}\right), \tag{26}$$

$$\mathbb{E}[\mathsf{B}] = O(\log\log T) \tag{27}$$

$$\mathbb{E}[\mathsf{C}] \leq \sum_{j\in[N]\setminus([K-1]\cup\{i\})} \left(\epsilon_2 + \frac{4T^{-\epsilon_2 d^B(\nu_2,\nu)}}{d^B(\mu_i, \mu_K)}\right)\log T + O(1), \tag{28}$$

$$\mathbb{E}[\mathsf{D}] \leq 2 + \frac{1}{d^B(\mu_i^{(+)}, \mu_i)} = O\left(\frac{1}{d^B(\mu_i^{(+)}, \mu_i)}\right). \tag{29}$$

The proofs of (26)-(29) are referred to Appendix B. We move on to bound the right hand side of (28).

Let $\epsilon_2 = (\log\log T)/(d^B(\nu_2, \nu)\log T)$, we have

$$\begin{aligned} &\inf_{\epsilon_2>0}\left\{\epsilon_2 + \frac{4T^{-\epsilon_2 d^B(\nu_2,\nu)}}{d^B(\mu_i, \mu_K)}\right\} \\ &= \inf_{\epsilon_2>0}\left\{\epsilon_2 + \frac{4e^{-\epsilon_2 d^B(\nu_2,\nu)\log T}}{d^B(\mu_i, \mu_K)}\right\} \\ &\leq \frac{\log\log T}{d^B(\nu_2,\nu)\log T} + \frac{4e^{-\log\log T}}{d^B(\mu_i,\mu_K)} \\ &= \frac{\log\log T}{d^B(\nu_2,\nu)\log T} + \frac{4}{d^B(\mu_i,\mu_K)\log T} = O\left(\frac{\log\log T}{\log T}\right) \end{aligned}$$

$\mathbb{E}[\mathsf{C}]$ can be bounded by $O(\log\log T)$.

The last term in (25) can be bounded as follows. Define a positive c_i such that

$$d^B(\mu_i^{(+)}, \mu_K^{(-)}) \geq (1 - c_i\delta)d^B(\mu_i, \mu_K),$$

the upper bound of *regret* can be given by

$$\begin{aligned} \mathbb{E}[Reg^{AS}(T)] &\leq \sum_{i\in[N]\setminus[K]} \mathbb{E}[Reg_i^{max}(T)] \\ &\leq \sum_{i\in[N]\setminus[K]} \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t)\}\Delta_i(t)\right] \\ &\leq \sum_{i\in[N]\setminus[K]} \{\mathbb{E}[\mathsf{A}+\mathsf{B}+\mathsf{C}+\mathsf{D}] \\ &\quad + \frac{\log T}{d^B(\mu_i^{(+)}, \mu_K^{(-)})}\Delta_{i,K}\} \\ &\leq \sum_{i\in[N]\setminus[K]} \frac{\Delta_{i,K}\log T}{(1-c_i\delta)d^B(\mu_i,\mu_K)} \\ &\quad + O(\log\log T) \\ &\quad + O\left(\frac{1}{\delta^2}\right) + O\left(\frac{1}{d^B(\mu_i^{(+)}, \mu_i)}\right). \end{aligned}$$

If $c_i\delta < 1/2$, $1 - c_i\delta > c_i\delta$. Let $\epsilon_1 < 1/2$ and $\delta = \epsilon_1 / \max_{i\in[N]\setminus[K]} c_i$. We can see $J_\mu(T)$ is on the order of $O(\log\log T)$, and C_μ is a constant on the order of $O(1/\delta^2) + O(1/d^B(\mu_i^{(+)}, \mu_i))$. ■

C. Extension to Dynamic Environment

In the static scenario, the reward of AS is used to update the contribution of each antenna to the system capacity. Due to the user mobility, the large-scale fading across the antenna array changes [29], and the previous rewards of AS are not efficient to measure the Beta distribution accurately. Moreover, as the channel conditions change, the upper and lower bounds of the current average capacity also vary. We focus on addressing

these adverse effects of large-scale fading variations across antennas.

Recall that the antenna-specific reward in (16) is within an interval $[r_{max}, r_{min}]$. To track the large-scale variation of channels, the parameters r_{min} and r_{max} should be updated dynamically. We introduce two positive discounting factors $\alpha_u < 1$ and $\alpha_l > 1$ for updating the interval $[r_{min}, r_{max}]$ after observing the antenna-specific reward $r(t)$.

In slot t, if antenna-specific reward falls in the interval $[r_{min}(t), r_{max}(t)]$, i.e., $r_{min}(t) \leq r(t) \leq r_{max}(t)$, the reward bounds are contracted as

$$\begin{aligned} r_{max}(t+1) &= \alpha_u r_{max}(t), \\ r_{min}(t+1) &= \alpha_l r_{min}(t). \end{aligned} \tag{30}$$

Otherwise, if the antenna-specific reward $r(t)$ is larger than the upper bound, i.e., $r(t) > r_{max}(t)$, keep the lower bound and replace $r_{max}(t)$ with $r(t)$:

$$r_{max}(t+1) = r(t). \tag{31}$$

If $r(t)$ falls below the lower bound, i.e., $r(t) < r_{min}(t)$, keep the upper bound and replace $r_{min}(t)$ with $r(t)$:

$$r_{min}(t+1) = r(t). \tag{32}$$

The large-scale fading variations also render the previous updates on the parameters of Beta distributions ineffective. We introduce a discounting factor $\gamma < 1$ to update S_i and F_i as discussed in [30]:

$$\begin{aligned} S_i(t+1) &= \gamma S_i(t) + b_i(t), \\ F_i(t+1) &= \gamma F_i(t) + 1 - b_i(t). \end{aligned} \tag{33}$$

γ can yield an exponential filtering of S_i and F_i to gradually reduce the influences of previous rewards. Therefore, these discounting factors guide the algorithm to focus on the latest outputs of AS and quantize the contribution of each antenna more precisely. This method can address the issue of large-scale fading variations over the antenna array effectively. The extended online AS algorithm is summarized in **Algorithm 2**.

Algorithm 2 Extended Online Antenna Selection

1: Initialize $S_i(0) = 1$, $F_i(0) = 1$ for antenna $i \in [N]$, and $r_{min}(t)$, $r_{max}(t)$ randomly.
2: **for** $t = 1, 2, 3 \cdots$; **do**
3: Sample $\theta_i(t)$ from $Beta(S_i(t-1), F_i(t-1))$ for antenna $i \in [N]$;
4: Select $\mathcal{K}(t) \in \mathcal{S}$ with (21);
5: Calculate $r(t)$ using (16) and $\hat{r}_i(t)$ using (17) for $i \in \mathcal{K}(t)$;
6: **if** $r_{min}(t) \leq r(t) \leq r_{max}(t)$ **then**
7: Update $[r_{min}(t), r_{max}(t)]$ with (30);
8: **else if** $r(t) > r_{max}(t)$ or $r(t) < r_{min}(t)$ **then**
9: Update $r_{max}(t)$ using (31) or $r_{min}(t)$ using (32);
10: **end if**
11: **for** each antenna $i \in \mathcal{K}(t)$, **do**
12: Sample $b_i(t)$ from Bernoulli distribution with success probability $\hat{r}_i(t)$;
13: Update $S_i(t)$, $F_i(t)$ with (33);
14: **end for**
15: **end for**

IV. Simulation Results

To evaluate our proposed online AS algorithms, we compare it with typical methods: convex relaxation, power-based and full antenna schemes. In the convex relaxation scheme, the active antenna set is obtained by solving a relaxed convex optimization problem. In the power-based scheme, the antennas with the highest received power are selected in each slot. For the full antenna scheme, the BS is equipped with the same number of antennas as the RF chains.

A. Simulation Settings

Consider the downlink transmission with 2.6 GHz central frequency and 5 MHz bandwidth in a single-cell massive MIMO system, where the uniform linear array on the BS is equipped with $N = 128$ antennas. The COST 2100 channel model is used for simulation and the OFDM parameters are set as described in Section II. The bandwidth is divided into 25 sub-bands and each sub-band contains 12 subcarriers. Orthogonal uplink pilot sequences are used for different users in the allocated subcarrier. The estimated channel gains are also used by other subcarriers in the sub-band. The subcarrier space is 15 kHz and the number of total subcarriers is 300. The number of users varies from 2 to 8, and the velocities of users are given by 0 m/s and 1 m/s, respectively, corresponding to the static and dynamic scenarios. The number of RF chains ranges from 64 to 96. Each slot consists of 7 OFDM symbols and each pilot occupies one symbol [17], as shown in Fig. 3. The parameters of the extended online AS algorithm are given by $\alpha_u = 0.95$, $\alpha_l = 1.05$, and $\gamma = 0.999$. The simulation parameters are listed in Table I.

TABLE I
Simulation Parameters

Parameter	Value
Slot length	0.5 ms
Pilot length	0.071 ms
Central frequency	2.6 GHz
Total bandwidth	5 MHz
Subcarrier space	15 kHz
Number of subcarriers	300
Number of sub-bands	25
Interference-free per-user SNR	$\rho_{dl} = [-5, 20]$ dB
Number of RF chains	$K = [64, 96]$
Number of mobile users	$M = [2, 8]$
Velocity of users	$v = \{0, 1\}$ m/s
Number of BS antennas	$N = 128$
Antenna spacing at BS	0.0577 m
Length of BS-side visibility regions	$L_{VR} = 3.2$ m
Intensity parameter	$\xi = 2.9$
Slope of visibility region gain	$\mathcal{N}(0, 0.9)$ dB/m

The effects of the large-scale fading across the antenna arrays in massive MIMO systems are discussed comprehensively in [10]. The standard COST 2100 channel model [31] designed for SISO or MIMO channel cannot be directly used in massive MIMO scenarios. We adopt the massive MIMO extension of the COST 2100 channel model proposed in [32].

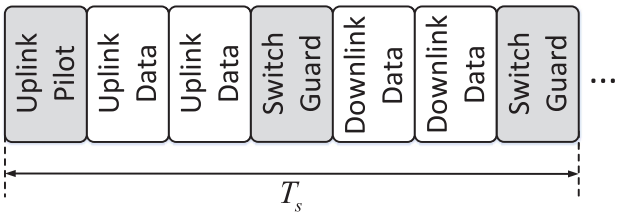


Fig. 3. The adopted slot structure.

The spatial variation across a large array can be modeled by introducing the BS-side visibility regions. The region illustrates the visibility of partial antennas to a particular group of multipath components. An antenna is active if it locates in a region visible to the user. Multiple active visibility regions overlap on the antenna array, resulting in different average channel gains.

The BS-side visibility region can be modeled as arising from a birth-death process [32]. The number of observed visibility regions on the antenna array follows the distribution

$$N_{VR} = \mathbf{Po}(\xi \cdot L_{BS} + \xi \cdot \mathbb{E}(Z)),$$

where $\mathbf{Po}(\cdot)$ denotes the Poisson distribution and ξ denotes the intensity parameter. Z represents the length of visibility region and L_{BS} represents the length of the antenna array. The visibility region center positions are uniformly distributed along the antenna array. It turns out that the true visibility region length in channel measurements can be well approximated by an exponential distribution, which can be expressed by

$$f_Z(z) = \begin{cases} -\dfrac{1}{L_{VR}} e^{-\frac{z}{L_{VR}}} & for\ z \geq 0, \\ 0 & otherwise, \end{cases}$$

where L_{VR} is the mean length of the BS-side visibility regions. Linear slopes in dB are used to fit the power variations within the visibility regions, and the distribution of slopes is approximated by normal distribution.

B. Antenna Selection Performance

Recall the *regret* is the cumulative differences of the rewards between the antenna set chosen by the online AS scheme and the antenna set with the largest average contribution in hindsight. Fig. 4 shows the ratio of *regret* to $\log T$ of the online AS scheme in static scenarios. The *regret* converges to a logarithmic order, which implies that the normalized reward can be well approximated by Beta distribution and the proposed online algorithm tends to select the antennas with the highest average performance as time slot goes.

The capacity performance of AS schemes is compared with full antenna scheme, which uses the same number of RF chains as the proposed online AS schemes. The deployment positions of the antennas are consistent with the antennas in the central interval of the AS schemes. Fig. 5 shows the capacity as a function of time slots in the static ($v = 0$ m/s) and dynamic ($v = 1$ m/s) scenarios, respectively. The numbers of users and RF chains are $M = 4$ and $K = 64$, respectively, and the interference-free per-user SNR is $\rho_{dl} = -5$ dB.

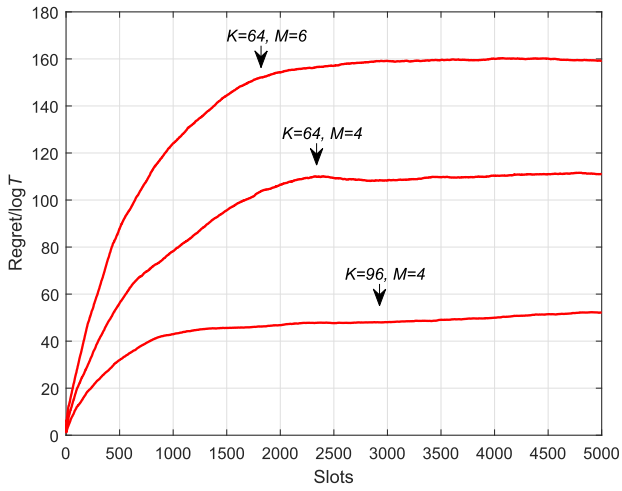
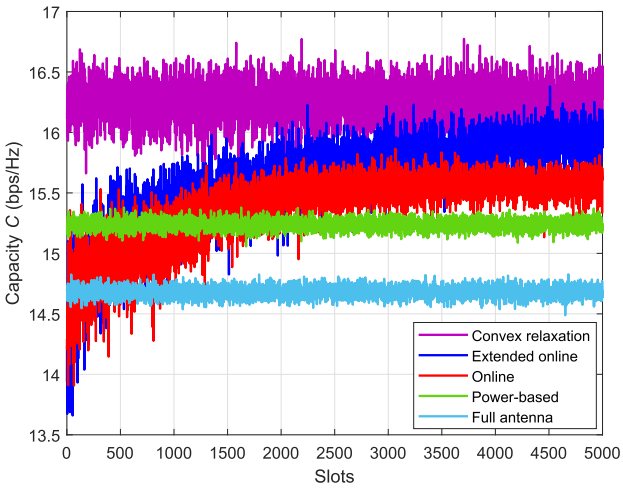
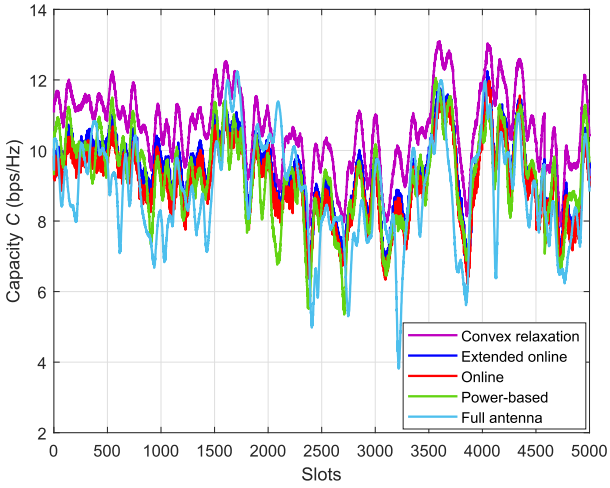


Fig. 4. $Regret/\log T$ as a function of time slots in static scenarios.



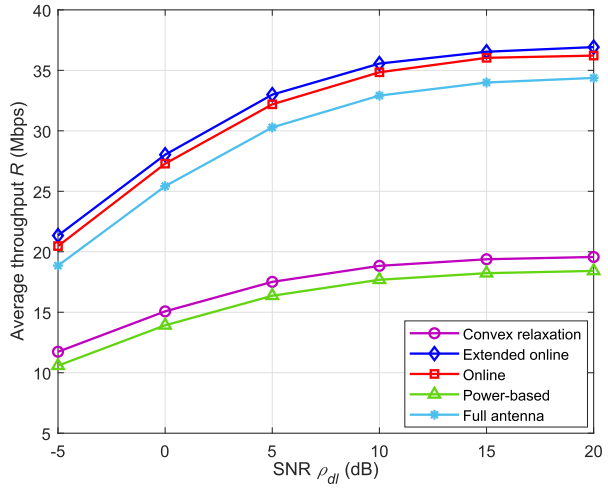
(a) Static scenario



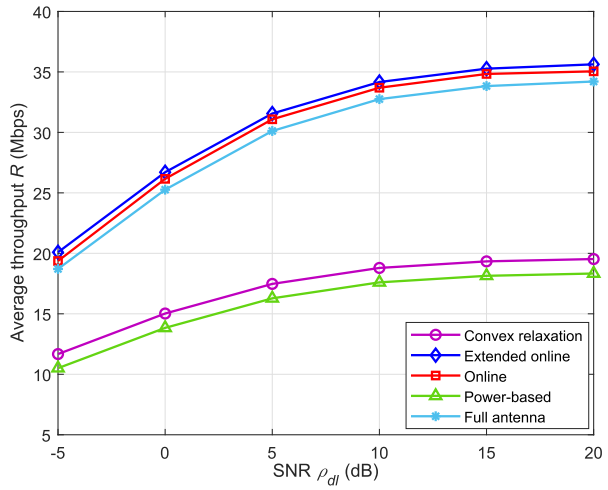
(b) Dynamic scenario

Fig. 5. Capacity as a function of time slots. $M = 4$, $\rho_{dl} = -5$ dB, and $K = 64$.

As we can see in Fig. 5(a), the convex relaxation algorithm achieves the highest capacity. However, the computational complexity is at least $\mathcal{O}(N^3KM)$, while the proposed online

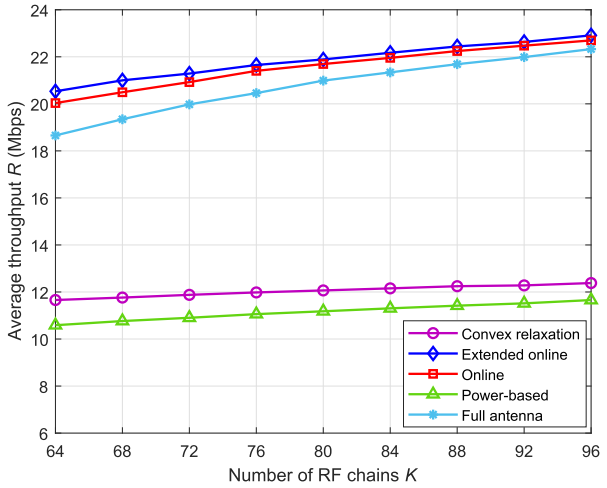


(a) Static scenario

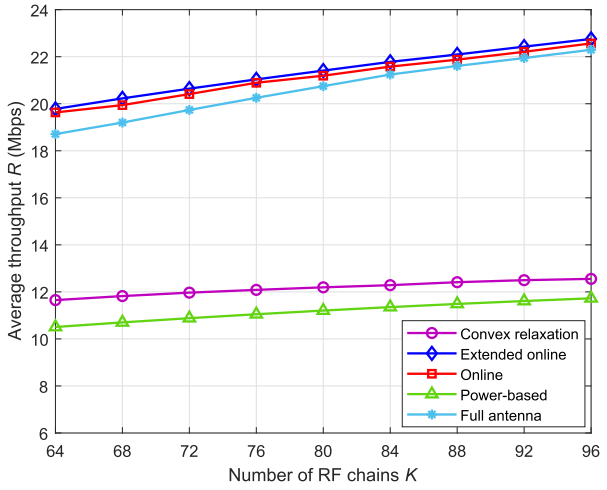


(b) Dynamic scenario

Fig. 6. Average throughput as a function of the interference-free per-user SNR. $M = 4$ and $K = 64$.



(a) Static scenario



(b) Dynamic scenario

Fig. 7. Average throughput as a function of the number RF chains. $M = 4$ and $\rho_{dl} = -5$ dB.

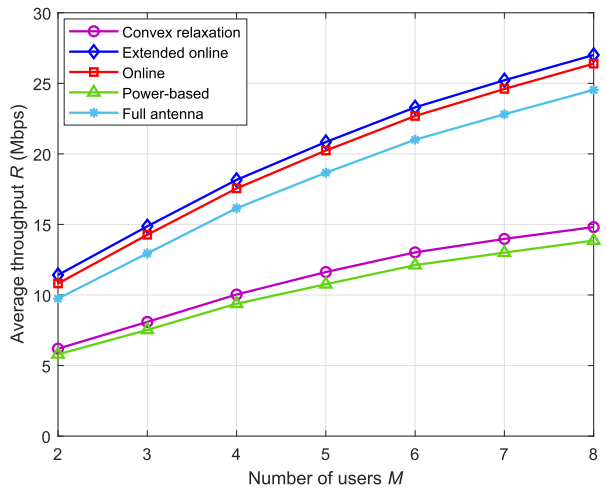
schemes have a much lower computational complexity of $\mathcal{O}(N \log N + K)$. Even if the performance is relatively low at first, the proposed online AS schemes still yield higher capacity than the power-based scheme and the full antenna scheme because the online learning algorithms can exploit the previous outputs of AS to achieve a better decision as time slot goes.

Note that the full antenna scheme has the lowest capacity, but it sometimes may achieve a high performance, as shown in Fig. 5(b). The reason is that the antenna with the largest average channel gain may be concentrated in the deployment location of the full antenna scheme.

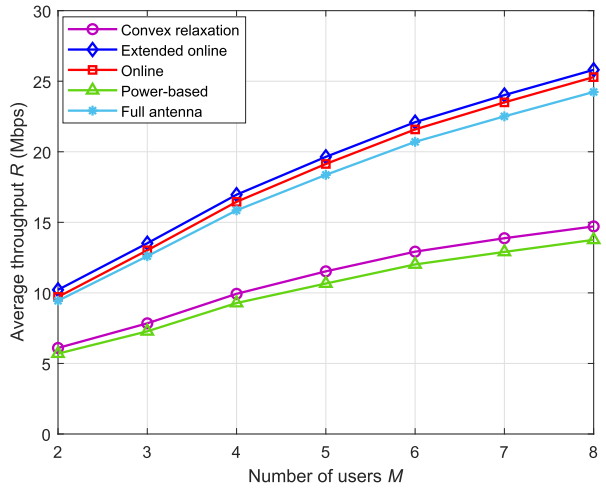
In the dynamic scenario, the convex relaxation algorithm still achieves the highest capacity as compared to others. The extended online AS scheme exhibits a slightly better performance than the power-based and the full antenna schemes. Consider the slot structure shown in Fig. 2, our proposed online AS schemes can exploit the saved pilot overhead for data transmission, which means they could yield higher throughputs than other algorithms, including the convex relaxation.

C. System Throughput

Fig. 6 shows the average throughput as a function of the interference-free per-user SNR. The online AS algorithm outperforms the convex relaxation and the power-based schemes by around 71% and 86% in the static scenario, respectively, while the extended online AS algorithm outperforms the convex relaxation and the power-based schemes by around 79% and 95%, respectively. The reason is that the reduced pilot transmission time in each slot is used for data transmission now. Specifically, since the number of RF chains is 64, $\lceil N/K \rceil = 2$ OFDM symbols are utilized for uplink pilot transmission to obtain full CSI for the convex relaxation and power-based algorithms. One symbol is remained for downlink data with fixed uplink length. In the proposed online AS schemes, only 1 symbol is needed to obtain partial CSI and an extra symbol can be used for downlink data, as shown in Fig. 3. On the other hand, even though the full antenna scheme uses the same symbol length for pilot, the antenna subset selection diversity can produce throughput improvement as can be seen from Fig. 6.



(a) Static scenario



(b) Dynamic scenario

Fig. 8. Average throughput as a function of the number of users. $\rho_{dl} = -5$ dB and $K = 64$.

The extended online AS algorithm outperforms the online one by around 5% and 3% for the static scenario and the dynamic scenario, respectively. The reason is that the discounting factors guide the extended online AS track the large-scale variations across antennas, which make the algorithm distinguish the contributions of the antennas adaptively.

Fig. 7 shows the average throughput as a function of the number of RF chains. The throughput increases with the number of RF chains, and the proposed online algorithms also outperform the other two AS algorithms and the full antenna scheme. The throughput improvements of the proposed online AS schemes over the full antenna one decreases with the increase of RF chains. The reason is that it is more likely for the full antenna scheme to cover the antenna array with a large contribution to the performance. When the number of RF chains increases, the throughput improvements of the online AS schemes increase. This gap growth is because the pilot overhead of CSI acquisition does not decrease in the convex relaxation and power-based algorithms while the active antennas are growing in the online AS schemes.

Fig. 8 shows the average throughput as a function of the number of users. In the static scenario, the online AS algorithm outperforms the convex relaxation and power-based algorithms by around 74% and 83%, respectively, while the extended one achieves a 85% and 94% throughput gain over the convex relaxation and power-based algorithms, respectively. Even in the dynamic scenario, the throughput improvements are only slightly reduced. Meanwhile, the online AS algorithm and the extended one outperforms the full antenna scheme by around 6% and 8% in the static scenario, respectively, while in the dynamic scenario, the throughput improvements of the online AS algorithm and the extended one over the full antenna scheme are around 4% and 6%, respectively.

V. Conclusion

In this article, we investigated the downlink AS problem in the TDD massive MIMO systems with partial CSI acquisition, where a subset of antennas should be selected to maximize the average throughput. We proposed a novel slot structure to reduce the pilot overhead and developed two online learning algorithms based on Thompson sampling technique to address the AS problem with imperfect CSI effectively and efficiently. By incorporating the massive MIMO extensions into the standard COST 2100 model, the large-scale fading across antennas can be modeled accurately and the proposed online AS algorithms yield significant throughput improvements as compared to the convex relaxation and power-based schemes. Our proposed online AS scheme also outperforms the full antenna scheme with the same number of RF chains in a massive MIMO system. Moreover, the extended online AS scheme for dynamic scenarios can alleviate the effects of large-scale fading variations over the antenna arrays on capturing the system capacity. We proved that the online AS algorithm can achieve a sublinear *regret* upper bound by selecting the antenna subset with the largest average contribution. In other words, the performance of the proposed online AS scheme is optimal asymptotically, indicating that it is promising for applications in practical massive MIMO systems.

Appendix A
Facts and Preparatory Lemmas

Fact 1 (Chernof-Hoeffding Bound Extension in [33]): Let $X_1, \ldots, X_n$ be random variables within the interval $[0, 1]$ such that $S_n = \sum_{i=1}^{n} X_i$. Let $\mu = \mathbb{E}[X_i]$. Define $d^B(p, q) = p \log(p/q) + (1-p) \log((1-p)/(1-q))$ as the Kullback-Leibler divergence between two Bernoulli distributions with mean value p and q. Then for $a \in (0, 1-\mu)$,

$$Pr(S_n \geq n(\mu + a)) \leq e^{-nd^B(\mu+a,\mu)}.$$

For $a \in (0, \mu)$,

$$Pr(S_n \leq n(\mu - a)) \leq e^{-nd^B(\mu-a,\mu)}.$$

Fact 2 (Pinsker's Inequality): For $p, q \in (0, 1)$, the Kullback-Leibler divergence between two Bernoulli distributions is bounded by

$$d^B(p, q) \geq 2(p - q)^2.$$

Lemma 1 (Lemma 9 in [28]): Let $k \in [N]$ and define $z < \mu_k$. Define $\hat{\mu}_i$ as the mean of $Beta(S_i(t), F_i(t))$ in slot t. The events $\mathsf{S}(t)$, $\mathsf{T}(t)$ and $\mathsf{U}(t)$ are defined as

1) If $\theta_k(t) \geq z$, $\mathsf{S}(t)$ and $\mathsf{T}(t)$ occur, antenna k is selected at slot t.
2) $\theta_k(t)$, $\mathsf{S}(t)$ and $\mathsf{T}(t)$ are mutually independent given $\{\hat{\mu}_i(t)\}_{i=1}^N$ and $\{L_i(t)\}_{i=1}^N$.
3) $\mathsf{U}(t)$ is deterministic given $\{\hat{\mu}_i(t)\}_{i=1}^N$ and $\{L_i(t)\}_{i=1}^N$.
4) Given $\{\hat{\mu}_i(t)\}_{i=1}^N$ and $\{L_i(t)\}_{i=1}^N$ such that $\mathsf{U}(t)$ holds, then $\mathsf{T}(t)$ has a probability of at least q.

Then

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\theta_k(t) < z, \mathsf{S}(t), \mathsf{U}(t), L_k(t) < L_c\}\right] = O\left(\frac{1}{q(\mu_k - z)^2}\right) + N_c \frac{1-q}{q}.$$

By setting $\mathsf{T}(t)$ and $\mathsf{U}(t)$ the trivial events that always hold, then

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\theta_k(t) < z, \mathsf{S}(t)\}\right] = O\left(\frac{1}{(\mu_k - z)^2}\right). \quad (34)$$

Lemma 2 (Appendix B.1 in [34]): Let $k \in [N]$ and define $z < \mu_k$, we have

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \mathbf{1}\{\mathsf{A}_k(t), \hat{\mu}_k(t) > z\}\right] < 1 + \frac{1}{d^B(z, \mu_k)}.$$

Lemma 3: Let $k \in [N]$, $n > 1$, x_1, $x_2 \in [0, 1]$ satisfying $x_1 > x_2$, then

$$Pr(\theta_k(t) \geq x_1 \mid \hat{\mu}_k(t) \leq x_2, L_k(t) = n) \leq e^{-nd^B(x_2, x_1)}$$

Proof:

$$\begin{aligned} &Pr(\theta_k(t) \geq x_1 \mid \hat{\mu}_k(t) \leq x_2, L_k(t) = n) \\ &= Pr(\theta \sim Beta(\hat{\mu}_j(t)n + 1, (1 - \hat{\mu}_j(t))n + 1), \\ &\quad \theta \geq x_1 \big| \hat{\mu}_j(t) \leq x_2) \\ &= 1 - F^{beta}_{x_2 n+1, (1-x_2)n+1}(x_1) \\ &= F^{B}_{n+1, x_1}(x_2 n) \\ &\leq F^{B}_{n, x_1}(x_2 n) \leq e^{-nd^B(x_2, x_1)}, \end{aligned}$$

where $F^{beta}_{\alpha,\beta}(x)$ represents the cumulative distribution function of Beta distribution with parameters α and β. $F^{B}_{n,p}(x)$ represents the cumulative distribution function of binomial distribution with mean np. The last inequality holds for *Fact 1*. ∎

APPENDIX B

Proof of Term A*:* First we can find $\theta^{(K)}(t) < \mu_K^{(-)}$ in term A indicates that at least one sample of antennas in $[N]$ is less than μ_K, which means

$$\{\theta^{(K)}_{[N]}(t) < \mu_K^{(-)}\} \subset \bigcup_{k \in [K]} \{\theta_k(t) < \mu_K^{(-)}\}.$$

Thus

$$\begin{aligned} \{\theta^{(K)}_{[N]}(t) < \mu_K^{(-)}\} &= \bigcup_{k \in [K]} \{\theta_k(t) < \mu_K^{(-)}, \theta^{(K)}_{[N]}(t) < \mu_K^{(-)}\}, \\ &\subset \bigcup_{k \in [K]} \{\theta_k(t) < \mu_K^{(-)}, \theta^{(K)}_{[K]\setminus\{k\}}(t) < \mu_K^{(-)}\}. \end{aligned}$$

The inequality is given by

$$\begin{aligned} &\mathbf{1}\{\theta^{(K)}_{[N]}(t) < \mu_K^{(-)}\} \\ &\qquad \leq \sum_{k \in [K]} \mathbf{1}\{\theta_k(t) < \mu_K^{(-)}, \theta^{(K)}_{[K]\setminus\{k\}}(t) < \mu_K^{(-)}\}. \end{aligned}$$

Since $\theta^{(K)}_{[K]\setminus\{k\}}(t) < \mu_K^{(-)}$ satisfies the condition for the event $\mathsf{S}(t)$ in *Lemma 1* with $z = \mu_K^{(-)}$, we have

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\theta^{(K)}_{[N]}(t) < \mu_K^{(-)}\}\right] = O\left(\frac{1}{q(\mu_k - \mu_K^{(-)})^2}\right).$$

∎

Proof of Term B*:* For term B,

$$\begin{aligned} &\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{C}_i^c(t)\} \\ &= \sum_{t=1}^{T} \sum_{j \in [N]\setminus([K-1]\cup\{i\})} \big\{\mathbf{1}\{\mathcal{A}_i(t), \hat{\mu}_i(t) > \mu_K\} \\ &\qquad + \mathbf{1}\{\mathcal{A}_i(t), \hat{\mu}_i(t) \leq \mu_K, \theta^{(K-1)}_{[N]\setminus\{i,j\}}(t) < \nu\}\big\}. \quad (35) \end{aligned}$$

The first term in (35) can be bounded with *Lemma 2*:

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \hat{\mu}_i(t) > \mu_K\}\right] \leq 1 + \frac{1}{d^B(\mu_K, \mu_i)} = O(1).$$

The second term in (35) can be decomposed as

$$\begin{aligned} &\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \hat{\mu}_i(t) \leq \mu_K, \theta^{(K-1)}_{[N]\setminus\{i,j\}}(t) < \nu\} \\ &\leq \frac{\log\log T}{d^B(\mu_K, \nu)} + \sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), L_i(t) > \frac{\log\log T}{d^B(\mu_K, \nu)}, \\ &\qquad \hat{\mu}_i(t) \leq \mu_K, \theta^{(K-1)}_{[N]\setminus\{i,j\}}(t) < \nu\} \\ &\leq \frac{\log\log T}{d^B(\mu_K, \nu)} + \sum_{t=1}^{T} \mathbf{1}\{L_i(t) > \frac{\log\log T}{d^B(\mu_K, \nu)}, \\ &\qquad \hat{\mu}_i(t) \leq \mu_K, \theta^{(K-1)}_{[N]\setminus\{i,j\}}(t) < \nu\}. \quad (36) \end{aligned}$$

The inequality $\theta^{(K-1)}_{[N]\setminus\{i,j\}}(t) < \nu$ implies that the sample of at least one antenna in $[K-1]$ is less than ν, which means

$$\{\theta^{(K-1)}_{[N]\setminus\{i,j\}}(t) < \nu\} \subset \bigcup_{k \in [K-1]} \{\theta_k(t) < \nu, \theta^{(K-1)}_{[N]\setminus\{i,j,k\}}(t) < \nu\}$$

The inequality is then given by

$$\mathbf{1}\{L_i(t) > \frac{\log\log T}{d^B(\mu_K, \nu)}, \hat{\mu}_i(t) \leq \mu_K, \theta^{(K-1)}_{[N]\setminus\{i,j\}}(t) < \nu\}$$

$$\leq \sum_{k\in[K-1]} \mathbf{1}\{L_i(t) > \frac{\log\log T}{d^B(\mu_K,\nu)}, \hat{\mu}_i(t) \leq \mu_K, \theta_k(t) < \nu, \theta^{(K-1)}_{[N]\setminus\{i,j,k\}}(t) < \nu\}.$$

Define $\nu_2 = (\nu + \mu_K)/2$ and $L^*(t) = \log T / d^B(\nu_2, \nu)$. For $k \in [K-1]$, we have $\mu_k > \nu > \nu_2 > \mu_K$, and

$$\begin{aligned} &Pr\{\theta_k(t) < \nu, L_k(t) \geq L^*(t)\} \\ &\leq \sum_{n=L^*(t)}^{T} Pr\{\theta_k(t) < \nu, L_k(t) = n\} \\ &\leq \sum_{n=L^*(t)}^{T} Pr\{\theta_k(t) < \nu, \hat{\mu}_k > \nu_2, L_k(t) = n\} \\ &\quad + \sum_{n=L^*(t)}^{T} Pr\{\hat{\mu}_k \leq \nu_2, L_k(t) = n\} \\ &\leq \sum_{n=L^*(t)}^{T} e^{-nd^B(\nu_2,\nu)} + \sum_{n=L^*(t)}^{T} e^{-nd^B(\nu_2,\mu_k)} \end{aligned}$$

The first term of last inequality holds by *Lemma 3* and the second term holds by *Fact 1*. Since $d^B(\nu_2, \mu_k) > d^B(\nu_2, \nu)$ for $\mu_k > \nu > \nu_2$, we have

$$\sum_{n=L^*(t)}^{T} e^{-nd(\nu_2,\nu)} + \sum_{n=L^*(t)}^{T} e^{-nd(\nu_2,\mu_k)} = O(T^{-1}).$$

Then we can derive

$$\begin{aligned} &\sum_{t=1}^{T} \sum_{k\in[K-1]} Pr\Big\{L_i(t) > \frac{\log\log T}{d^B(\mu_K,\nu)}, \hat{\mu}_i(t) \leq \mu_K, \\ &\quad \theta_k(t) < \nu, \theta^{(K-1)}_{[N]\setminus\{i,j,k\}}(t) < \nu\Big\} \\ &\leq \sum_{t=1}^{T} \sum_{k\in[K-1]} Pr\Big\{L_i(t) > \frac{\log\log T}{d^B(\mu_K,\nu)}, L_k(t) < L^*(t), \\ &\quad \hat{\mu}_i(t) \leq \mu_K, \theta_k(t) < \nu, \theta^{(K-1)}_{[N]\setminus\{i,j,k\}}(t) < \nu\Big\} \\ &\quad + \sum_{t=1}^{T} \sum_{k\in[K-1]} Pr\{\theta_k(t) < \nu, L_k(t) \geq L^*(t)\} \\ &\leq \sum_{t=1}^{T} \sum_{k\in[K-1]} Pr\Big\{L_i(t) > \frac{\log\log T}{d^B(\mu_K,\nu)}, L_k(t) < L^*(t), \\ &\quad \hat{\mu}_i(t) \leq \mu_K, \theta_k(t) < \nu, \theta^{(K-1)}_{[N]\setminus\{i,j,k\}}(t) < \nu\Big\} + O(1). \end{aligned} \tag{37}$$

$z = \nu$, $\mathsf{S}(t) = \{\theta^{(K-1)}_{[N]\setminus\{i,j,k\}}(t) < \nu\}$, $\mathsf{T}(t) = \theta_k(t) < \nu$, and $\mathsf{U}(t) = \hat{\mu}_i(t) \leq \mu_K$ satisfy the conditions in *Lemma 1*. Follow *Lemma 3*, $\mathsf{T}(t)$ has a probability of

$$1 - exp\left(-d^B(\mu_K,\nu)\left(\frac{\log\log T}{d^B(\mu_K,\nu)}\right)\right) = 1 - (\log T)^{-1}$$

By using *Lemma 1* with $L_c = L^*(t)$,

$$\begin{aligned} &\mathbb{E}\Big[\sum_{t=1}^{T} \mathbf{1}\{L_i(t) > \frac{\log\log T}{d^B(\mu_K,\nu)}, L_k(t) < L^*(t), \\ &\quad \hat{\mu}_i(t) \leq \mu_K, \theta_k(t) < \nu, \theta^{(K-1)}_{[N]\setminus\{i,j,k\}}(t) < \nu\}\Big] \\ &\leq O\left(\frac{1}{1-(\log T)^{-1}(\mu_k-\nu)^2}\right) \\ &\quad + O\left(\frac{(\log T)^{-1}}{1-(\log T)^{-1}}\frac{\log T}{d^B(\nu_2,\nu)}\right) = O(1). \end{aligned} \tag{38}$$

Combining (35), (36), (37) and (38), term B is bounded by $O(\log\log T)$.

■

Proof of Term C*:* According to Appendix A.4 in [28], the term C can be decompose for each antenna $j \in [N] \setminus ([K-1] \cup \{i\})$ as

$$\begin{aligned} &\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{A}_j(t), \mathcal{C}_i(t), \mathcal{D}_i(t)\} \\ &\leq \epsilon_2 \log T + \sum_{n=0}^{\frac{\log T}{d^B(\mu_i^{(+)},\mu_K^{(-)})}-1} \sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{A}_j(t), \mathcal{C}_{i,j}(t), \\ &\quad L_i(t) = n, \hat{\mu}_j(t) \leq \nu_2, \mathcal{E}_j(t)\} + O(1). \end{aligned} \tag{39}$$

We first prove

$$\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{A}_j(t), \mathcal{C}_{i,j}(t), L_i(t) = n, \hat{\mu}_j(t) \leq \nu_2, \mathcal{E}_j(t)\}. \tag{40}$$

Since the event $\{\mathcal{A}_i(t), L_i(t) = n\}$ occurs at most once, (40) is at most 1. Define τ as the first slot where $\{\mathcal{C}_{i,j}(t), \theta^{(K-1)}_{[N]\setminus\{i,j\}}(t) \leq \theta_i(t), \mathcal{A}_i(t), L_i(t) = n\}$ is satisfied.

$\{\theta_i(\tau) \geq \theta^{(K-1)}_{[N]\setminus\{i,j\}}(\tau)\}$ is necessary for (40) to be 1, because $\theta_i(\tau)$ and $\theta_j(\tau)$ should be larger than $\theta^{(K-1)}_{[N]\setminus\{i,j\}}(\tau)$ for selecting antennas i, j simultaneously. If $\theta_j(\tau) < \theta^{(K-1)}_{[N]\setminus\{i,j\}}(\tau)$, antenna i is selected and $\{L_i(t) = n\}$ is not satisfied when $t > \tau$. By using *Lemma 3*, we have

$$\begin{aligned} &Pr\{\theta_i(\tau) \geq \theta^{(K-1)}_{[N]\setminus\{i,j\}}(\tau), \theta^{(K-1)}_{[N]\setminus\{i,j\}}(\tau) \geq \nu, \hat{\mu}_j(t) \leq \nu_2\} \\ &\qquad \leq e^{-L_j(\tau)d(\nu_2,\nu)} \end{aligned}$$

Then

$$\begin{aligned} &\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{A}_j(t), \mathcal{C}_{i,j}(t), L_i(t) = n, \hat{\mu}_j(t) \leq \nu_2\}\right] \\ &\qquad \leq e^{-d(\nu_2,\nu)\epsilon_2 \log T} = T^{-\epsilon_2 d^B(\nu_2,\nu)}. \end{aligned} \tag{41}$$

The term C for each antenna j in (39) can be bounded as

$$\begin{aligned} &\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{A}_j(t), \mathcal{C}_i(t), \mathcal{D}_i(t)\} \\ &\leq \epsilon_2 \log T + \frac{\log T}{d^B(\mu_i^{(+)},\mu_K^{(-)})} T^{-\epsilon_2 d^B(\nu_2,\nu)} + O(1) \\ &\leq \left(\epsilon_2 + \frac{4T^{-\epsilon_2 d^B(\nu_2,\nu)}}{d^B(\mu_i,\mu_K)}\right)\log T + O(1). \end{aligned}$$

The last inequality holds for *Fact 2* and $(1+\delta)^2 < 4$. Finally, we have

$$\mathbb{E}[\mathsf{C}] \leq \sum_{j\in[N]\setminus([K-1]\cup\{i\})} \left(\epsilon_2 + \frac{4T^{-\epsilon_2 d^B(\nu_2,\nu)}}{d^B(\mu_i,\mu_K)}\right)\log T + O(1).$$

■

Proof of Term D*:* For term D, we can decompose it as

$$\mathbb{E}[\mathsf{D}] \leq \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{B}(t), \hat{\mu}_i(t) > \mu_i^{(+)}, \mathcal{D}_i^c(t)\}\right] + \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{B}(t), \hat{\mu}_i(t) \leq \mu_i^{(+)}, \mathcal{D}_i^c(t)\}\right].$$

Follow *lemma 2*, the first term can be bounded as

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \mathcal{B}(t), \hat{\mu}_i(t) > \mu_i^{(+)}, \mathcal{D}_i^c(t)\}\right] \leq \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{1}\{\mathcal{A}_i(t), \hat{\mu}_i(t) > \mu_i^{(+)}\}\right] \leq 1 + \frac{1}{d^B(\mu_i^{(+)}, \mu_i)}$$

The second term can be bounded as

$$\begin{aligned}
&\mathbb{E}\left[\mathbf{1}\{\mathcal{A}_i(t), \mathcal{B}(t), \hat{\mu}_i(t) \leq \mu_i^{(+)}, \mathcal{D}_i^c(t)\}\right] \\
&\leq \mathbb{E}\left[\mathbf{1}\{\theta_i(t) \geq \mu_K^{(-)}, \hat{\mu}_i(t) \leq \mu_i^{(+)}, \mathcal{D}_i^c(t)\}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{\theta_i(t) \geq \mu_K^{(-)}, \hat{\mu}_i(t) \leq \mu_i^{(+)}, \mathcal{D}_i^c(t)\}|\hat{\mu}_i(t), L_i(t)\right]\right] \\
&\leq \mathbb{E}\Big[\mathbb{E}\Big[\mathbf{1}\{\hat{\mu}_i(t) \leq \mu_i^{(+)}, \mathcal{D}_i^c(t)\} \\
&\quad Pr\{\theta_i(t) \geq \mu_K^{(-)}|\hat{\mu}_i(t), L_i(t)\}|\hat{\mu}_i(t), L_i(t)\Big]\Big] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[e^{-\frac{\log T}{d(\mu_i^{(+)},\mu_K^{(-)})} d(\mu_i^{(+)},\mu_K^{(-)})}\middle|\hat{\mu}_i(t), L_i(t)\right]\right] \\
&= e^{-\frac{\log T}{d(\mu_i^{(+)},\mu_K^{(-)})} d(\mu_i^{(+)},\mu_K^{(-)})} = T^{-1}.
\end{aligned}$$

The penultimate inequality holds for the fact $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$, while the last inequality holds for *Lemma 3*. Finally, we come to the conclusion that

$$\mathbb{E}[\mathsf{D}] \leq 2 + \frac{1}{d^B(\mu_i^{(+)}, \mu_i)}.$$

■

ACKNOWLEDGMENT

The authors would like to thank the editors and the anonymous reviewers, whose invaluable comments helped to improve the presentation of this article substantially.

REFERENCES

[1] V. W. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang, *Key Technologies for 5G Wireless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[2] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[3] S. Sanayei and A. Nosratinia, "Antenna selection in MIMO systems," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, Oct. 2004.

[4] M. Xiao *et al.*, "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[5] R. W. Heath, Jr., N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[6] S. Y. Park and D. J. Love, "Capacity limits of multiple antenna multicasting using antenna subset selection," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2524–2534, Jun. 2008.

[7] A. Mukherjee and A. Hottinen, "Learning algorithms for energy-efficient MIMO antenna subset selection: Multi-armed bandit framework," in *Proc. EUSIPCO*, Aug. 2012, pp. 659–663.

[8] K. T. Truong and R. W. Heath, Jr., "The viability of distributed antennas for massive MIMO systems," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2013, pp. 1318–1323.

[9] H.-Y. Lu and W.-H. Fang, "Joint transmit/receive antenna selection in MIMO systems based on the priority-based genetic algorithm," *IEEE Antennas Wireless Propag. Lett.*, vol. 6, pp. 588–591, 2007.

[10] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Massive MIMO in real propagation environments: Do all antennas contribute equally?" *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 3917–3928, Nov. 2015.

[11] T.-W. Ban and B. C. Jung, "A practical antenna selection technique in multiuser massive MIMO networks," *IEICE Trans. Commun.*, vol. E96.B, no. 11, pp. 2901–2905, 2013.

[12] J. Joung and S. Sun, "Two-step transmit antenna selection algorithms for massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[13] M. Hanif, H.-C. Yang, G. Boudreau, E. Sich, and H. Seyedmehdi, "Antenna subset selection for massive MIMO systems: A trace-based sequential approach for sum rate maximization," *J. Commun. Netw*, vol. 20, no. 2, pp. 144–155, Apr. 2018.

[14] A. Konar and N. D. Sidiropoulos, "A simple and effective approach for transmit antenna selection in multiuser massive MIMO leveraging submodularity," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4869–4883, Sep. 2018.

[15] P. V. Amadori and C. Masouros, "Interference-driven antenna selection for massive multiuser MIMO," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 5944–5958, Aug. 2016.

[16] J. Chen, S. Chen, Y. Qi, and S. Fu, "Intelligent massive MIMO antenna selection using Monte Carlo tree search," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5380–5390, Oct. 2019.

[17] J. Vieira *et al.*, "A flexible 100-antenna testbed for massive MIMO," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2014, pp. 287–293.

[18] O. Raeesi, A. Gokceoglu, Y. Zou, E. Björnson, and M. Valkama, "Performance analysis of multi-user massive MIMO downlink under channel non-reciprocity and imperfect CSI," *IEEE Trans. Commun.*, vol. 66, no. 6, pp. 2456–2471, Jun. 2018.

[19] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[20] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, nos. 3–4, pp. 154–655, 2017.

[21] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Tech. J.*, vol. 20, pp. 11–22, Mar. 2015.

[22] Z. Kuai, T. Wang, and S. Wang, "Transmit antenna selection in massive MIMO systems: An online learning framework," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Aug. 2019, pp. 496–501.

[23] N. Cesa-Bianchi and G. Lugosi, "Combinatorial bandits," *J. Comput. Syst. Sci.*, vol. 78, no. 5, pp. 1404–1422, Sep. 2012.

[24] O. Chapelle and L. Li, "An empirical evaluation of Thompson sampling," in *Proc. NeurIPS*, 2011, pp. 2249–2257.

[25] D. J. Russo, B. Van Roy, A. Kazerouni, I. Osband, and Z. Wen, "A tutorial on Thompson sampling," *Found. Trends Mach. Learn.*, vol. 11, no. 1, pp. 1–96, 2018.

[26] M. Zhou, T. Wang, and S. Wang, "Spectrum sensing across multiple service providers: A discounted Thompson sampling method," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2402–2406, Dec. 2019.

[27] S. Agrawal and N. Goyal, "Analysis of Thompson sampling for the multi-armed bandit problem," in *Proc. COLT*, 2012, pp. 39.1–39.26.

[28] J. Komiyama, J. Honda, and H. Nakagawa, "Optimal regret analysis of Thompson sampling in stochastic multi-armed bandit problem with multiple plays," in *Proc. ICML*, 2015, pp. 1152–1161.

[29] R. Chopra, C. R. Murthy, H. A. Suraweera, and E. G. Larsson, "Performance analysis of FDD massive MIMO systems under channel aging," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1094–1108, Feb. 2018.

[30] N. Gupta, O.-C. Granmo, and A. Agrawala, "Thompson sampling for dynamic multi-armed bandits," in *Proc. 10th Int. Conf. Mach. Learn. Appl. Workshops*, vol. 1, Dec. 2011, pp. 484–489.

[31] L. Liu *et al.*, "The COST 2100 MIMO channel model," *IEEE Wireless Commun.*, vol. 19, no. 6, pp. 92–99, Dec. 2012.

[32] J. Flordelis, X. Li, O. Edfors, and F. Tufvesson, "Massive MIMO extensions to the COST 2100 channel model: Modeling and validation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 380–394, Jan. 2020.

[33] D. P. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[34] S. Agrawal and N. Goyal, "Further optimal regret bounds for Thompson sampling," in *Proc. AISTATS*, 2013, pp. 99–107.

Zhenran Kuai received the B.S. degree from Nanjing University, Nanjing, China, in 2019, where he is currently pursuing the Ph.D. degree with the School of Electronic Science and Engineering. His current research interests include wireless communications and machine learning.

Shaowei Wang (Senior Member, IEEE) received the Ph.D. degree from Wuhan University, Wuhan, China, in 2006. From 2012 to 2013, he was a Visiting Scholar/Professor with Stanford University, Stanford, CA, USA, and The University of British Columbia, Vancouver, BC, Canada. He joined the School of Electronic Science and Engineering, Nanjing University, Nanjing, China, as a Faculty Member, in 2006, where he is currently a Full Professor. His research interests include communications and networking, operations research, and machine learning.