Two-dimensional Beam Selection for Multi-user Massive MIMO System Based on Multi-armed Bandit Algorithm

Maki Arai
Dept. of Electrical Engineering, Faculty of Engineering
Tokyo University of Science
Tokyo, Japan
maki.arai@rs.tus.ac.jp

Mikio Hasegawa
Dept. of Electrical Engineering, Faculty of Engineering
Tokyo University of Science
Tokyo, Japan
hasegawa@ee.kagu.tus.ac.jp

***Abstract*—A massive Multiple-input Multiple-output (MIMO) system for multiple users is important to achieve large capacity for future wireless communications. As the number of beams and users increases, appropriate beam selection becomes difficult because of a large number of searches. In this paper, we propose a simple 2-dimensional beam selection method using the information of the received power for multi-user massive MIMO systems and formulate the 2-dimensional beam selection method with the ϵ-greedy method. The objective is to improve the system performance by independently selecting the beam that maximizes the received power of each user terminal. By the numerical analysis, the validity of the proposed method is confirmed and the relationship between the search rate and the achieved received power is clarified.**

***Index Terms*—massive MIMO, beamforming, beam alignment, MAB**

I. INTRODUCTION

To achieve the large capacity of future wireless communication systems such as Beyond 5G and 6G systems, multi-user massive multiple-input multiple-output (MU massive MIMO) technology is indispensable. In massive MIMO systems, a hybrid beamforming configuration, which combines analog beamforming and digital pre/post-processing, is used at a base station (BS) side to achieve both low cost and high performance [1]–[3]. Multiple high-gain beams are formed by using a large number of antenna elements and analog beamforming weight matrix, which improves propagation gain in a high-frequency band and enables space-division multiplexing transmission to multiple user terminals (UT). However, as the number of beams and UTs increases, the number of searches for appropriate beam selection becomes larger. As a result, it is difficult to complete the appropriate beam selection before the propagation environment and UT location change, so the system performance degrades.

To solve the above trade-off between search time and beam optimality, multiple beam selection in MIMO systems is considered as a multi-armed bandit (MAB) problem, and several optimal beam selection method using the MAB algorithm has been proposed as follows. In [4], the fast beam selection can be achieved by using the correlation between the optimal beam and the neighboring beams, so the delay due to beam selection was significantly reduced compared to the exhaustive search method used in IEEE.802.11ad. [5] presented the antenna state selection without an instantaneous full channel state information (CSI), then a signal-to-noise ratio (SNR) and a throughput can be improved because the overhead of CSI information is reduced. Also, [6] proposed a two-stage beam selection for a millimeter wave communication between a BS, a reconfigurable intelligent surface and a UT, then the authors confirmed the reduction of beam training overhead and improvement of spectral efficiency. However, 2-dimensional high-gain beam selection for multi-user massive MIMO systems with hybrid beamforming has not been studied.

In this paper, we propose a 2-dimensional beam selection method in both horizontal and vertical directions for multi-user environments and hybrid beamforming configurations.

First, we formulate a problem of selecting a beam for each UT independently, which is transformed from a problem for multiple UTs. Then, we apply an ϵ-greedy algorithm to select the optimal beam for each UT by using the information of the received power. Finally, we evaluate the performance of the proposed method by numerical analysis.

II. SYSTEM MODEL

In this section, a channel model, an antenna and a beamforming configuration used in this paper are described.

A. Signal model

Figure 1 shows a massive MIMO system model with one BS and K UTs. The number of antennas at the BS side is N_{BS} and each user has one antenna. In this system, a hybrid beamforming configuration is used, where two types of weight matrix are used for analog beamforming and digital pre/post-processing [3], [7]. The received signal vector $\mathbf{y}$ from BS to all UTs is defined as follows,

$$\begin{aligned}\mathbf{y} &= [y_1, \cdots, y_K]^{\mathrm{T}} \\ &= \mathbf{H}\mathbf{W}_{\mathrm{a}}\mathbf{W}_{\mathrm{d}}\mathbf{s} + \mathbf{n},\end{aligned} \tag{1}$$

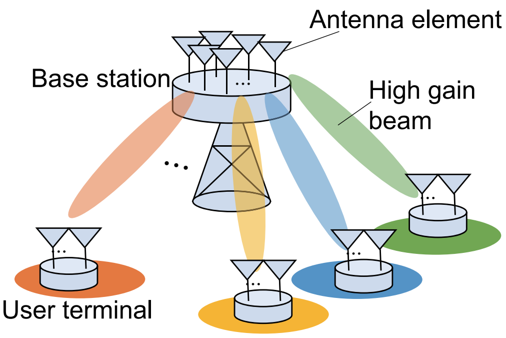


Fig. 1: Multi-user massive MIMO system model.

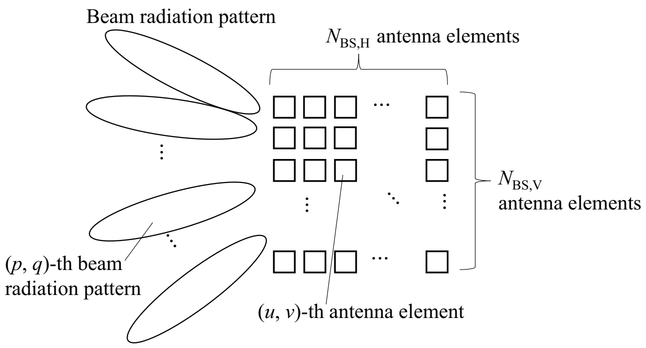


Fig. 2: 2-dimensional planar array configuration.

where y_k is received signal of the k-th UT, $\mathbf{H}$ is a channel matrix between BS and all UTs, $\mathbf{W}_{\mathrm{a}}$ is an analog beamforming weight, $\mathbf{W}_{\mathrm{d}}$ is a digital pre/post-processing weight matrix such as zero-forcing, $\mathbf{s}$ is a vector of transmit signals, and $\mathbf{n}$ is a vector of noise at all UTs.

B. Channel model

In this paper, we assume the line-of-sight (LOS) propagation channel because a higher frequency band such as a millimeter wave band is used. Under the LOS environment, the component of the channel matrix h_{kn} is expressed by a distance between the k-th user and the n-th BS antenna element d_{kn} and a carrier frequency f_0.

$$h_{kn} = \frac{1}{\sqrt{PL(d_{kn}, f_0)}} \exp(-j\frac{2\pi}{\lambda_0} d_{kn}) g_0(\theta_k, \phi_k), \quad (2)$$

where $\sqrt{PL(d_{kn}, f_0)}$ is a pathloss coefficient determined by d_{kn} and f_0. $\exp(-j\frac{2\pi}{\lambda_0} d_{kn})$ is a phase rotation due to LOS path between the k-th user and the n-th BS antenna element and λ_0 means a wavelength at the carrier frequency. g_0 is a radiation pattern of each antenna element at BS side and θ_k, ϕ_k indicate angles of directions from BS to the k-th UT in vertical and horizontal planes, respectively.

C. Antenna configuration at BS side

To communicate with UTs at any location, 2-dimensional beamforming towards horizontal and vertical planes is needed. In this paper, a uniform planar antenna array with a full-array beamforming network is used at the BS side, where all antenna elements and all RF chains are connected with phase shifters [7].

As shown in Fig. 2, the planar antenna array has N_{BS} antenna elements and $N_{\mathrm{BS,H}}$ and $N_{\mathrm{BS,V}}$ antenna elements are placed towards horizontal and vertical directions with identical spacing l. The number of RF chains is defined as the same as the number of UTs K. The relationship among the number of antenna elements, the number of RF chains, and the number of users is $N_{\mathrm{BS}} \geq K$.

The analog beamforming weight matrix is defined by using a discrete Fourier transform (DFT) matrix. The weight for the n-th antenna element and the m-th beam direction is expressed as

$$w_{\mathrm{a},nm} = \frac{1}{\sqrt{N_{\mathrm{BS}}}} \exp(\frac{-j2\pi(u-1)(p-1)}{aN_{\mathrm{BS,V}}}) \cdot \exp(\frac{-j2\pi(v-1)(q-1)}{aN_{\mathrm{BS,H}}}). \quad (3)$$

u, v is the indices of the antenna element in vertical and horizontal directions, and p, q are the indices of the beam direction in vertical and horizontal directions, respectively. a is a coefficient to adjust the angle spacing between beams; the larger a is, the closer the main lobes of the beams are. In this configuration, $a^2 N_{\mathrm{BS}}$ different beams can be formed and K types of beam directions are selected for K UTs. To improve the communication performance, the optimal beam selection for each UT is needed.

Next, the digital pre/post-coding weight matrix is defined by a zero-forcing (ZF) matrix.

$$\mathbf{W}_{\mathrm{d}} = [\mathbf{w}_{\mathrm{d},1}, \cdots, \mathbf{w}_{\mathrm{d},K}] \propto (\mathbf{H}\mathbf{W}_{\mathrm{a}})^{+} \quad (4)$$

$$|\mathbf{w}_{\mathrm{d},k}|^2 = 1, \quad (5)$$

where $\mathbf{X}^{+}$ means the Moore-Penrose inverse matrix of $\mathbf{X}$. By using ZF, an equivalent channel matrix is obtained, where the channels are orthogonalized.

$$\tilde{\mathbf{H}} = \mathbf{H}\mathbf{W}_{\mathrm{a}}\mathbf{W}_{\mathrm{d}} = \mathrm{diag}[\tilde{h}_1, \cdots, \tilde{h}_K]. \quad (6)$$

D. Received power and channel capacity

When the m-th beam is used between BS and the k-th UT, the received power of the k-th user is expressed as follows,

$$P_{\mathrm{r},km} = |\mathbf{h}_k \mathbf{w}_{\mathrm{a},m}|^2 P_{\mathrm{s}} \quad (7)$$

$$\mathbf{h}_k = [h_{k,1}, \cdots, h_{k,N_{\mathrm{BS}}}], \quad (8)$$

where it is assumed that the transmit power per user is equally divided into P_{s}.

Then, the inter-user interference can be canceled and channels are equalized for all UTs. Finally, the total channel capacity of all UTs is derived by using an equalized channel response $\tilde{h}$, given by

$$C_{\mathrm{total}} = B \log_2 \det(\mathbf{I}_K + \mathbf{H}\mathbf{W}_{\mathrm{a}}\mathbf{W}_{\mathrm{d}}(\mathbf{H}\mathbf{W}_{\mathrm{a}}\mathbf{W}_{\mathrm{d}})^{\mathrm{H}} \frac{P_{\mathrm{s}}}{P_{\mathrm{n}}})$$
$$= B \sum_k \log_2(1 + |\tilde{h_k}|^2 \frac{P_{\mathrm{s}}}{P_{\mathrm{n}}}). \quad (9)$$

III. PROPOSED BEAM SELECTION METHOD

In this chapter, we describe the definitions of the arms and the reward in the MAB problem to select optimal 2-dimensional beams to maximize the received power of each UT.

A. *Multi-armed bandit problem*

The MAB problem [8] aims to maximize the final reward by playing multiple slot machines with unknown rewards within a limited number of trials. In this case, there is a trade-off between “exploration” of the best machine and “exploitation” of the information obtained from the search to select the better machine. The method of decision-making on whether “exploration” or “exploitation” is important in the MAB problem.

In this paper, we use the ϵ-greedy method, which is one of the algorithms for solving the MAB problem [9]. The ϵ-greedy method is a simple algorithm that uses a parameter ϵ, which is called as a search rate in this paper, to determine the trade-off between exploration and exploitation. On each trial, $i (i = 1, \cdots, I)$, an arm is randomly selected with probability ϵ and the arm that achieves the highest reward is selected with probability $1 - \epsilon$.

The beam selection towards horizontal and vertical directions for multiple UTs is considered as an MAB problem of selecting a combination of horizontal and vertical beam indices. In this paper, the horizontal and vertical beam indices are replaced by a single index $m = p + (q - 1)aN_{\mathrm{BS,V}}$, then we select the beam index with the highest reward among all $a^2 N_{\mathrm{BS}}$ beam indices. By using the ϵ-greedy method, the index of the beam selected for the k-th UT on the $i+1$-th trial, defined as $\hat{m}_k(i)$, is expressed as follows.

$$\hat{m}(i+1) = \begin{cases} \text{Random} & (\text{for } \epsilon) \\ \arg\max_{m \in \{1, \cdots, a^2 N_{\mathrm{BS,H}} N_{\mathrm{BS,V}}\}} \; Q(i) & (\text{for } 1 - \epsilon), \end{cases} \quad (10)$$

B. *Objective in 2-dimensional beam selection*

When the SNR is large, the total channel capacity is proportional to the total received power from Eq. 9. Since the magnitude of $\mathbf{w}_{\mathrm{d},k}$ is 1, the total power of the desired signal remains the same before and after digital pre/post-processing. In this case, the problem of maximizing the total channel capacity for all UTs can be replaced by the problem of maximizing the received power for all UTs. In addition, when all phase shifters are variable in a full array configuration, the same beam can be used for different RF chains and the beam for each UT is selected independently regardless of the other UTs' beam selection. Therefore, the selection of the beams that maximize the sum of the total channel capacity for all UTs can be considered as a problem of independently selecting the beam that maximizes the received power for each UT as follows.

$$\max_m C_{\mathrm{total}} \approx \max_m \sum_k |\tilde{h_k}|^2$$
$$= \max_m \mathrm{trace}(\mathbf{H}\mathbf{W}_{\mathrm{a}}\mathbf{W}_{\mathrm{d}}(\mathbf{H}\mathbf{W}_{\mathrm{a}}\mathbf{W}_{\mathrm{d}})^{\mathrm{H}})$$
$$= \max_m \mathrm{trace}(\mathbf{H}\mathbf{W}_{\mathrm{a}}(\mathbf{H}\mathbf{W}_{\mathrm{a}})^{\mathrm{H}})$$
$$= \max_m \sum_k P_{\mathrm{r},km_k}. \quad (11)$$

TABLE I: Analysis condition.

Location (BS)	$(x, y, z) = (0, 0, 25)$
Location (UT)	$(x, y, z) = (10, 10, 1.5)$, $(-15, 20, 1.5)$
The number of antenna elements (BS)	$N_{\mathrm{BS,H}} = 16$ $N_{\mathrm{BS,V}} = 16$
Antenna element spacing (BS)	$l = 0.5\lambda_0$
The number of antenna elements (UT)	1
The number of UTs	$K = 2$
The number of RF chains	$M = 2$
Coefficient for beam angle spacing	$a = 1$
Carrier frequency	$f_0 = 25$ [GHz]
Pathloss coefficient	3GPP TR 38.901 Release 17, UMa LOS [10]
Transmit power	$KP_{\mathrm{s}} = 20$ [dBm]
The number of trials	$I = 300$
Repetition Count	$N_{\mathrm{average}} = 500$

C. *Reward definition*

We define the received power of each UT P_{r,km_k} as a reward given by

$$r(i) = P_{\mathrm{r},k\hat{m}(i)}. \quad (12)$$

By using this reward, the beam is selected with not complete CSI information but the information about the received power. The expected reward for the m-th beam on the $i+1$-th trial is calculated by

$$Q_m(i+1) = Q(i) + \frac{1}{N_m(i+1)}(r(i+1) - Q(i)), \quad (13)$$

where $N_m(i)$ is the number that the m-th beam is selected by i trials.

IV. NUMERICAL ANALYSIS

In this section, by numerical analysis, we confirm the validity of the 2-dimensional beam selection method that we proposed.

A. *Analysis condition*

For the fundamental analysis, the MU massive MIMO system with one BS and 2 UTs is considered. Fig. 3 shows a bird's-eye view of locations in meters of the BS and UTs. The planar antenna array of BS is placed in the xz-plane and multiple beams are formed in the positive direction of y-axis. The analysis condition is described in Table I.

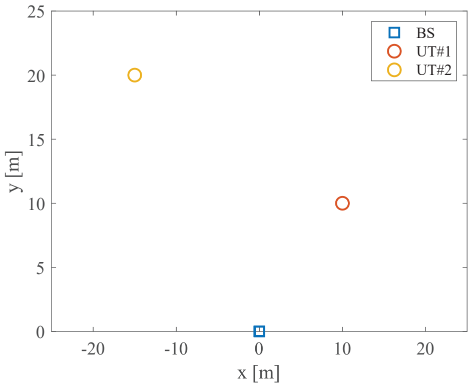


Fig. 3: Locations of BS and UTs in xy-plane.

B. Beam radiation patterns.

The beam radiation patterns formed by the analog beamforming weight matrix are calculated for the front of the planar antenna array in the range of $\phi = 0\ 180$ [deg.] and $\theta = 0\ 180$ [deg.]. From the horizontal and vertical symmetries, we describe the vertical beam radiation patterns. The radiation patterns in the horizontal plane at $\theta = 90$ [deg.] are shown in Fig. 4, where multiple beams are formed in 16 different directions. In the horizontal direction, the beams in 16 different directions are formed as well. Thus, $16 \times 16 = 256$ different types of beams are formed as a whole.

C. Optimal beams derived by exhaustive search.

The optimal beams to maximum received power at each UT are derived by an exhaustive search as shown in Fig. 5, where the circle marker indicates UT location. It can be seen that there is the peak gain of the selected beam towards the direction of each UT. Also, the optimal beams to maximum total channel capacity are derived by exhaustive search as shown in Fig. 6, where the selected beam is the same in the case of maximizing the received power of each UT. From the results, it is confirmed that the maximization of total channel capacity is replaced by the maximization of the received power of each UT, as shown in Eq. (11).

D. Beam selection result for search rate

We describe the results of the beam selection by using the ϵ-greedy method with Eqs. (10), (12) and (13).

Figures (7) and (8) show the mean of the beam index that maximizes the expected reward at the end of I trials as a function of the search rate ϵ. The obtained beam index was averaged by repeating N_{average} times of I trials. The result shows that when ϵ is small, the search for the optimal beam is not completed due to the large number of beams than the number of searches.

On the other hand, as ϵ increases, the mean of the beam index that maximizes the expected reward agrees well with the optimal value because of the sufficient number of searches. Therefore, learning the optimal beam that achieves maximum received power requires a sufficiently large search rate.

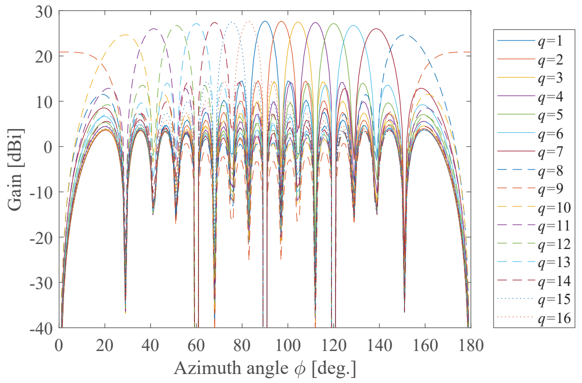
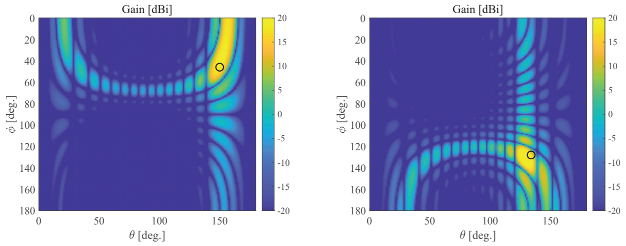
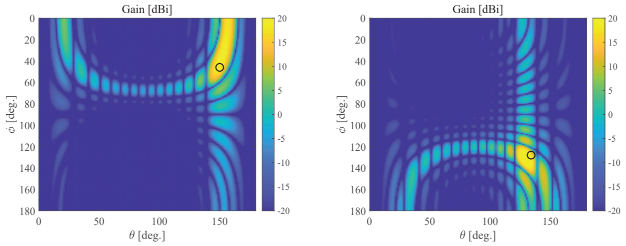


Fig. 4: Radiation patterns in the horizontal plane at $\theta = 90$ [deg.].



(a) The case of UT#1 (b) The case of UT#2

Fig. 5: Gain of the optimal beam to maximize received power.



(a) The case of UT#1 (b) The case of UT#2

Fig. 6: Gain of the optimal beam to maximize total channel capacity.

E. Achieved received power

The achieved received power of UE#1 and #2 in I trials is shown in Figs. 9 and 10. For both UT#1 and #2, the probability of selecting the optimal beam increases as ϵ becomes large, and the achieved received power also increases. However, when ϵ becomes too large, the maximum received power is achieved but the optimal beam is less selected. It can be seen that an appropriate ϵ is needed to obtain larger received power more often.

Figure 11 shows the average received power for I trials and N_{average} repetitions as a function of the search rate ϵ. The average received power is maximized at $\epsilon = 0.5$ in UT#1 and at $\epsilon = 0.35$ in UT#2, respectively.

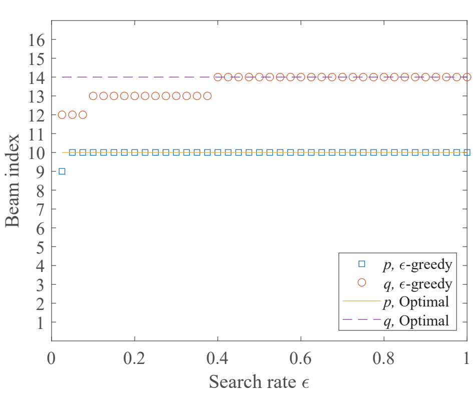


Fig. 7: Search rate versus selected and optimal beam indices in the case of UT#1.

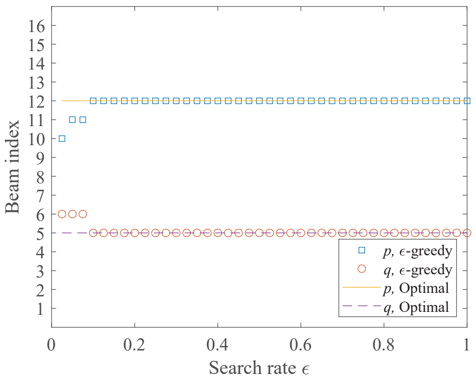


Fig. 8: Search rate versus selected and optimal beam indices in the case of UT#2.

V. CONCLUSION

In this paper, we proposed a method for selecting beams toward horizontal and vertical directions to maximize the received power in the multi-user massive MIMO system. Using the received power information of each UT, we formulated a 2-dimensional beam selection method with the ϵ-greedy method. By the numerical analysis, the validity of the proposed method is confirmed and the relationship between the search rate and the achieved received power is clarified.

The method that we proposed enables to solve the 2-dimensional beam selection problem with not any complete CSI information but the information about the received power. Therefore, the proposal in this paper is expected to contribute to the improvement of the communication performance for future wireless communication systems such as Beyond5G and 6G.

For future works, we will analyze communication characteristics with a larger number of antenna elements and user terminals and apply other MAB algorithms to improve the search efficiency.

ACKNOWLEDGMENT

This work was supported by JST [Moonshot R&D][Grant Number JPMJMS2216].

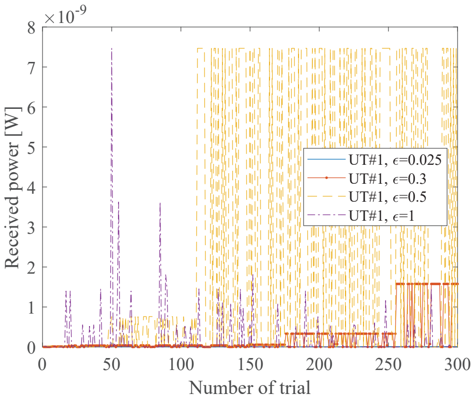


Fig. 9: Received power of each trial in the case of UT#1.

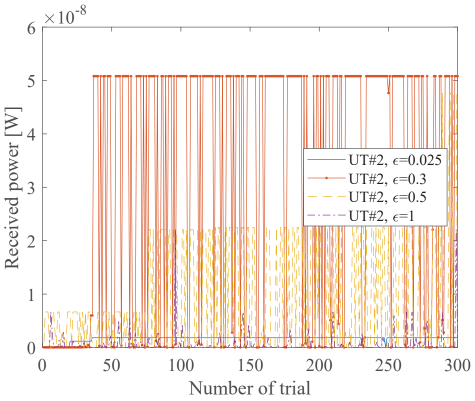


Fig. 10: Received power of each trial in the case of UT#2.

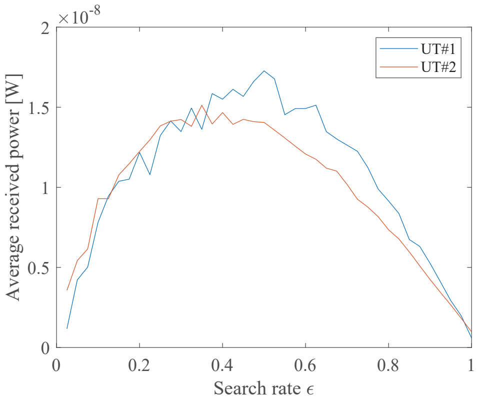


Fig. 11: Search rate versus Average received power.

REFERENCES

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive MIMO for next generation wireless systems,” IEEE Commun. Mag., vol. 52, no. 2, pp. 186-195, Feb. 2014.

[2] A. Gupta and R. K. Jha, “A Survey of 5G Network: Architecture and Emerging Technologies,” IEEE Access, vol. 3, pp. 1206-1232, 2015.

[3] M. Rihan, T. Abed Soliman, C. Xu, L. Huang and M. I. Dessouky, “Taxonomy and Performance Evaluation of Hybrid Beamforming for 5G and Beyond Systems,” IEEE Access, vol. 8, pp. 74605-74626, 2020.

[4] W. Wu, N. Cheng, N. Zhang, P. Yang, W. Zhuang and X. Shen, “Fast mmwave Beam Alignment via Correlated Bandit Learning,” IEEE Transactions on Wireless Communications, vol. 18, no. 12, pp. 5894-5908, 2019.

[5] N. G. A. K. Dandekar, “Learning State Selection for Reconfigurable Antennas: A Multi-Armed Bandit Approach,” IEEE Transactions on Antennas and Propagation, vol. 62, no. 3, pp. 1027-1038, 2014.

[6] E. M. Mohamed, S. Hashima, K. Hatano, and S. A. Aldossari, “Two-Stage Multiarmed Bandit for Reconfigurable Intelligent Surface Aided Millimeter Wave Communications,” Sensors, vol. 22, no. 6, Mar. 2022.

[7] M. Arai, K. Sakaguchi, and K. Araki, “A study on optimal beam patterns for single user massive MIMO transmissions,” IEICE Trans. Commun., vol. E102.B, no. 2, pp. 324-336, 2019.

[8] P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time Analysis of the Multiarmed Bandit Problem,” Mach. Learn., vol. 47, no. 2, pp. 235-256, May 2002.

[9] R. Sutton and A. Barto, “Reinforcement learning: An introduction,” MIT Press, 1998.

[10] 3GPP TR 38.901, “3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on channel model for frequency spectrum above 6 GHz (Release 17),” V17.0.0, 2022-03.