一般社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report

A Joint Interference Coordination based on Graph Coloring Algorithm and Deep Reinforcement Learning for Cluster-wise Distributed MU-MIMO

Chang Ge‡,a)　Sijie Xia‡,b)　Qiang Chen‡,b)　Fumiyuki Adachi†, b)

†Resilient Wireless Communication Research Group
International Research Institute of Disaster Science, Tohoku University

468-1 Aoba, Aramaki, Aoba-ku, Sendai, Miyagi, Japan
‡Department of Communications Engineering, Graduate school of Engineering, Tohoku University

6-6-05 Aramaki Aza Aoba, Aoba-ku, Sendai, Miyagi, 980-8579, Japan
E-mail: a) ge.chang.q2@dc.tohoku.ac.jp, b) {xia-s, adachi, chenq}@eceic.tohoku.ac.jp

Abstract In our previous studies, we proposed a 2-layer IC framework based on O-RAN architecture for a cellular system with cluster-wise distributed MU-MIMO and applied a graph coloring algorithm (GCA)-based interference coordination (IC) to mitigate both the intercell interference and intracell interference. In this paper, knowing that the mobile radio environment is time-varying, we propose a joint IC consisting of a GCA-based intracell IC and a deep reinforcement learning (DRL)-based intercell IC in our 2-layer IC framework. The simulation results of our proposed joint IC have revealed that it could achieve a significant increase in capacity compared to the no IC case.

Keywords Interference Coordination, Deep Reinforcement Learning, Graph Coloring, Distributed MU-MIMO

1. Introduction

In 5G and beyond, massive MU-MIMO has been regarded as a promising technique [1]. Distributed MU-MIMO [2], which exploits distributed antennas (DAs) over the base station coverage area (hereafter, simply called the cell), can relieve the problem of radio link blockage resulting from the utilization of mm-wave band. A large-scale cell-wise MU-MIMO requires a prohibitively large computational complexity. Hence, in our previous study, we proposed a cluster-wise distributed MU-MIMO [3], where users are adaptively divided into non-overlapping sub-groups called user-clusters (hereafter, simply called clusters) based on the user location information to greatly reduce the computational complexity. However, in return, the problem of inter-cluster interference is produced.

In a cellular system with cluster-wise distributed MU-MIMO, the inter-cluster interference can be of two types: intracell interference and intercell interference. Considering of the system scalability, we want to mitigate these two types of interference jointly in a fully decentralized manner, that is, each cell works independently with no information exchange among each other. Under this decentralized scenario, the intracell interference coordination (IC), which aims to mitigate the interference caused by clusters in the own cell, is relatively straightforward because each BS has all the information about its governing clusters. While the intercell IC, which aims to mitigate the interference from clusters belonging to surrounding cells but facing each other along a cell boundary, is much more difficult to realize.

In recent years, with the rise of artificial intelligence (AI) technology, especially the deep reinforcement learning (DRL), some new intercell IC schemes in a cellular system have emerged. In 2020, in order to solve the intercell IC problem in an ultra-dense small-cell network deployed in a residential area, Y. Wang, et al. [4] applied the actor-critic (AC) algorithm to minimize each BS's transmit power so as to reduce the intercell interference to the user equipments (UEs) of the surrounding BSs. In order to realize a fully decentralized scheme without information exchange between BSs, the Mean Field Theory is employed together with AC algorithm. Similarly, in 2021, in order to solve the intercell IC problem in HetNets, M. Yan, et al. [5] applied the Double deep Q network (DQN) to schedule sub-channels to individual users. In order to improve the robustness of Double DQN, Wasserstein Generative Adversarial Networks (W-GANs) was incorporated together in [5].

In our previous study, we proposed a 2-layer IC framework [6] based on O-RAN architecture [7]. Besides that, we also proposed a graph coloring algorithm-based IC (GCA-IC) [6] to be applied in the 2-layer IC framework. However, the mobile radio environment is time-varying,

the static GCA-based IC does not cope very well with such a dynamically varying mobile radio environment. Therefore, in this paper, we propose a joint IC, in which the DRL is applied to our GCA-IC. The joint IC under 2-layer IC framework can be expected to mitigate both the intercell interference and the intracell interference in a totally distributed manner under a dynamically varying environment.

To ensure that the joint IC can adapt quickly to the varying environment, the DQN needs to be trained with the real-time data obtained from interaction with the environment. Therefore, DQN is trained online instead of offline in this paper, which guarantees that our proposed joint IC can adapt to the varying environment and react in real time.

The remainder of this paper is organized as follows. Section 2 gives a brief introduction of the 2-layer IC framework. Section 3 provides the system model and the problem formulation. In Section 4, the proposed joint IC is described. The performance evaluation is conducted by computer simulation in Section 5, and Section 6 concludes this paper.

2. 2-layer IC framework

The proposed 2-layer IC framework is designed based on O-RAN architecture as shown in Fig.1. The key functional components introduced by O-RAN architecture is the near-real-time (near-RT) radio access network intelligent controllers (RICs) with the xAPPs, and the non-RT RIC with rApps. The near-RT RICs, with the control loop of 10ms ~ 1s, are designed to be the specific executor to control one or several cells, while the non-RT RIC, with the control loop of longer than 1s, is to provide guidance for the near-RT RICs with its global optimization and monitoring capability.

The proposed 2-layer IC framework allows the two kinds of RICs to cooperate with each other and fully exploit their respective advantages. The clustering, together with the IC are designed to applied as the xAPPs on each near-RT RICs, respectively. While the non-RT RIC is responsible for the cellular reconstruction and guiding the application of IC of each near-RT RICs.

In our previous study [6], we realized a successful application of applying GCA in the 2-layer IC framework. This paper is a preliminary attempt in the application of DRL in it. Our proposed joint IC consists of a GCA-based intracell IC and a DRL-based intercell IC. Both ICs are designed to work in the fully decentralized manner, which means that both ICs can be applied independently by each near-RT RIC with only the locally observed information. During the communication, each near-RT RIC updates the clustering results based on the users' movement and associates the DAs to each cluster according to the principle of proximity. The updating of the clustering results will trigger the GCA-based intracell IC (described in Sect. 4 (B)) to allocate the different sub-bands to the neighboring clusters to mitigate the intracell interference. After that, the non-RT RIC with its broader system-level view will send guidance information to the near-RT RICs to turn on some of the non-adjacent cells' DRL-based intercell IC (described in Sect.4(C)). Then, the selected cells will work independently to mitigate the intercell interference with only the locally observed information.

The 2-layer IC framework can be classified as a semi-decentralized framework that adds an additional centralized layer on top of the decentralized layer. As a preliminary study of the application of DRL for IC under O-RAN architecture, in this paper we only focus on the near-RT RIC part, that is how to jointly apply the GCA and DRL to mitigate both the intracell interference and the intercell interference, while leave the details about the higher-level control from the non-RT RIC for a future study.

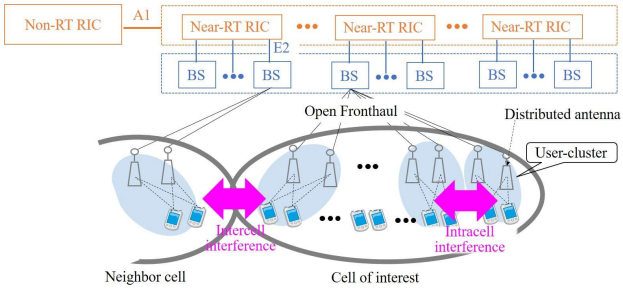

Fig. 1. 2-layer IC framework based O-RAN architecture.

3. System model and the problem formulation

In our proposed joint IC, the entire bandwidth is segmented into M sub-bands, where M is called the bandwidth segmentation factor, and one of the sub-bands is allocated to each cluster. The set of clusters in the entire communication service area and the set of clusters which are allocated the m^{th} sub-band in the entire communication service area are denoted by κ and κ_m, $m \in \{1, \dots, M\}$, respectively. In this paper, the numbers of users, DAs, and clusters in κ are denoted by N_U, N_A, and N_C, respectively. While those in κ_m are denoted by N_U^m, N_A^m, and N_C^m, respectively. The i^{th} user in the k^{th} cluster

in κ_m is denoted by $u_{i,k}^m$. Below, the matrices are represented as bold upper-case letters and the superscripts $(i,:)$ and $(:,i)$ represent the i^{th} row and column vectors of the matrix, respectively. Assuming the zero-forcing (ZF) based cluster-wise MU-MIMO to eliminate the multi-user interference within each cluster and by approximating the sum of inter-cluster interference and noise as a complex Gaussian process, the received signal-to-interference plus noise ratio (SINR) of user $u_{i,k}^m$ is given as

$$SINR_{u_{i,k}^m} = \frac{P_k \left\| \mathbf{H}_k^{(i,:)} \mathbf{W}_k^{(:,i)} \right\|^2}{\sum_{l=1,l\neq k}^{N_C^m} P_l \sum_{j=1}^{N_{U,l}^m} \left\| \mathbf{H}_{k,l}^{(j,:)} \mathbf{W}_l^{(:,j)} \right\|^2 + 1}, \quad (1)$$

where $\mathbf{W}_k$ and $\mathbf{W}_l$ are the ZF precoder matrices, $\mathbf{H}_k$ and $\mathbf{H}_{k,l}$ are respectively the channel matrix of the k^{th} cluster and the interference channel matrix between users in the k^{th} cluster and DAs in the l^{th} cluster in κ_m. $N_{U,k\ or\ l}^m$ denotes the number of users in the k^{th} or l^{th} cluster in κ_m. P_k and P_l are the transmit powers allocated to the k^{th} and l^{th} clusters, respectively and can be expressed as

$$P_{k\ \mathrm{or}\ l} = \frac{N_{U,k\ \mathrm{or}\ l}^m P}{\left\| \mathbf{W}_{k\ or\ l} \right\|_F^2}, \quad (2)$$

where P is the transmit power-to-noise ratio equal to all N_U users. Using the SINR expression in Eq. (1), the user capacity of user $u_{i,k}^m$ can be expressed as

$$C_{u_{i,k}^m} = \frac{1}{M} \log_2 (1 + SINR_{u_{i,k}^m}). \quad (3)$$

Assigning different sub-bands to different clusters is equivalent to dividing the clusters into different cluster subsets $\{\kappa_m; m \in \{1, \ldots, M\}\}$. Therefore, our goal is to select optimal cluster subset $\kappa_m \subseteq \kappa$ which maximizes the sum capacity. We set our optimization objective as follows:

$$\begin{aligned} &\max_{\kappa_m \subseteq \kappa} \sum_{m=1}^{M} C_m \\ &s.t. \quad \forall m \in M, \\ &\quad \bigcup_{m \in M} \kappa_m = \kappa, \text{ and } \kappa_n \bigcap \kappa_m = \varnothing, \ \forall n \neq m, \end{aligned} \quad (4)$$

where

$$C_m = \sum_{k=1}^{N_C^m} \sum_{i=1}^{N_{U,k}^m} C_{u_{i,k}^m}. \quad (5)$$

4. Joint IC based on GCA and DRL

The framework of our proposed joint IC is illustrated in Fig.2. Each near-RT RIC independently forms clusters according to the changes of user locations. Once the clustering result is updated, the GCA-based intracell IC is triggered to reassign sub-bands to the newly formed clusters to mitigate the intracell interference. After that, the non-RT RIC will send commanding signals to the near-RT RICs and then, some of the non-adjacent cells' DRL-based intercell IC will be turned on to work independently to mitigate the intercell interference caused by color collision.

During the implementation process of the DRL-based intercell IC, each near-RT RIC first estimates the current state $s^{(t)}$ in the timeslot t, which is used as the input to the DQN to derive the estimated value of each color-adaptation actions. The action $a^{(t)}$ with the highest value will be selected, which as a consequence, will change the existing coloring results to minimize the occurrence of color collision near the cell boundary. The selected $a^{(t)}$ actually serves the next timeslot t+1, therefore $s^{(t+1)}$ is estimated again and the reward $r^{(t+1)}$ is defined by the near-RT RIC to evaluate the merit of the selected $a^{(t)}$ by comparing $s^{(t)}$ and $s^{(t+1)}$.

Because the online training strategy is adopted in this paper, we assume that the wireless environment at $s^{(t)}$ and $s^{(t+1)}$ are different. Unlike the commonly used offline training [8][9], online training can ensure that the parameters of DQN are constantly updated and thus is able to follow the time-varying environment and provide real-time solutions. As a result, our proposed joint IC based on online training can naturally explore the unknown environment and find solutions with well adaptability to the time-varying environment.

To enable efficient online training, in this paper, we assume that each near-RT RIC is equipped with a fixed size of memory pool, in which the state transition sequence $\Delta^{(t)} = \left(s^{(t)}, a^{(t)}, s^{(t+1)}, r^{(t+1)}\right)$ that happened in latest timeslots are stored. During the online training process, a batch of data D is randomly selected from the memory pool to train the DQN. The application of memory replay and batch selection [10] can effectively eliminate the correlation between training data and improve the data utilization. Meanwhile, it ensures that the training dataset for online training is up-to-date and also, it greatly reduces the size of dataset during each training episode so as to reduce the training overhead.

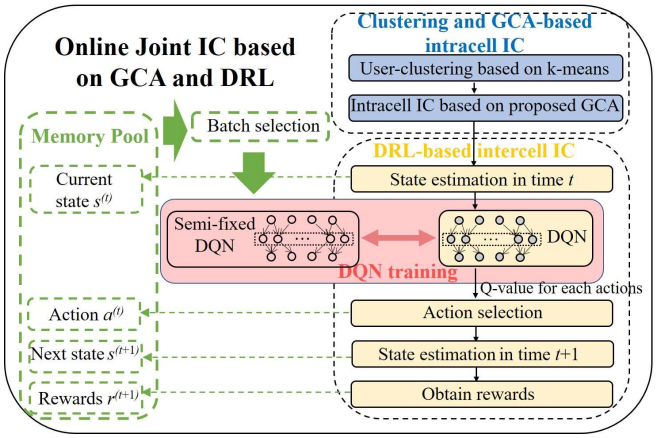

Fig.2. The framework of joint IC.

4.1. GCA-based intracell IC

In our previous study [6], we modelled the IC problem as a graph and applied GCA from graph theory to optimize the sub-bands allocation in order to mitigate the intracell interference. We revealed that there is a tradeoff between the bandwidth segmentation and the interference mitigation and that the maximum capacity is obtained when the bandwidth segmentation factor M=4. We also proposed an GCA in which the value of M is controllable. In this paper, we apply the GCA of [6] for intracell IC.

4.2. DRL-based intercell IC

Since we assume a fully decentralized framework, we suppose that each BS is a single agent and that the IC problem in each cell can be modeled as a Markov decision process (MDP), which can be expressed as a triplet $\{S, A, R\}$, where S represents the state space, A represents the action space, and R is the reward function. They are described below.

- **State space:** At timeslot t, we define the states for each BS agent as the instantaneous sum capacities of the clusters those belong to κ_m in each cell based on the current coloring result, which is expressed as $s^{(t)} = [C_0^{(t)}, C_1^{(t)}, \cdots, C_{M-1}^{(t)}]$.

- **Action space:** The action that each BS can take is designed as $A = \{1, 2, \ldots, M\}$. Let the coloring result for the k^{th} cluster after GCA be $g_k \in \{0, 1, \ldots M-1\}$. In timeslot t+1, after the action $a^{(t)}$ is chosen by the BS, the coloring result of each cluster is adjusted based on the modulo operation as

$$g_k^{(t+1)} = \left(g_k^{(t)} + a^{(t)}\right) \bmod M \,. \tag{6}$$

As for the action selection policy (π), we adopt the well-known ε-greedy policy [10] to balance the exploration and the exploitation.

- **Reward function:** The reward function is defined as the difference in the change of sum capacity after taking $a^{(t)}$ to change the coloring result and is given as

$$r^{(t+1)} = \sum_{m=0}^{M-1} C_m^{(t+1)} - \sum_{m=0}^{M-1} C_m^{(t)} \,. \tag{7}$$

The DQN used in this paper is an extension of the basic Q-learning algorithm [10], which applies the Bellman equation to update the Q value with the learning rate α as

$$Q(s^{(t)}, a^{(t)}) \leftarrow Q(s^{(t)}, a^{(t)}) + \alpha[r^{(t+1)} + \gamma \max_{a \in A} Q'(s^{(t+1)}, a) - Q(s^{(t)}, a^{(t)})] \tag{8}$$

Since the environment are changing in time, the state space S becomes infinite. Therefore, the DQN, in which the tabular-based storage is replaced by a neural network, is our better choice. In order to better train the DQN, we adopt the semi-fixed target network method [10], in which one local DQN and one semi-fixed DQN coexist. The local DQN updates its weight θ during the online training and calculates the estimated Q value $Q(s^{(t)}, a^{(t)}, \theta)$ (denoted by Q_estimated). While the semi-fixed DQN, with weight θ' been copied from θ every T^* timeslots, is to calculate the target Q value $Q(s^{(t+1)}, a^{(t+1)}, \theta')$ (denoted by Q_target). The cooperation of semi-fixed DQN and local DQN can improve the convergence during the DQN training.

In DQN, the Q_target and Q_estimated can be obtained by local DQN and semi-fixed DQN, respectively, as follows

$$\text{Q\_target} = r^{(t+1)} + \gamma \max_{a \in A} Q(s^{(t+1)}, a, \theta') \,, \tag{9}$$

$$\text{Q\_estimated} = Q(s^{(t)}, a^{(t)}, \theta) \,. \tag{10}$$

Therefore, the loss is defined as

$$\text{loss}(\theta) = \sum_{D} (\text{Q\_target} - \text{Q\_estimated})^2 \,. \tag{11}$$

The DQN training process is to minimize the loss by updating the value of θ.

5. Simulation results

We consider a normalized area of 5×5 over which 25 cells are constructed as shown in Fig.3(a). In our

simulation, the user locations are generated randomly 100 times, and for each generation of user locations, the quasi-static channel is realized as follows. The distance dependent pathlosses are computed based on the generated user locations. The log-normally distributed shadowing losses are generated 10 times for each generation of user locations. The Rayleigh fading gains are generated 10 times for each generation of shadowing losses. In this paper, the cell which is located in the center area and receives the intercell interference from every direction is selected as the cell of interest to evaluate the performance of our proposed joint IC. The DQN is made up of a 3-layer fully connected neural network, and the Gradient Descent with Momentum and Adaptive Learning Rate Backpropagation [11] is used as the network training function. Other detailed parameters are shown in Table I.

Figures 3(b) and (c) illustrate the sub-bands allocation results at timeslot $t = 0$ and 100, respectively. At the beginning ($t = 0$) when only GCA-based intracell IC is applied, the neighboring clusters inside each cell have been allocated different sub-bands so as to avoid the intracell interference. But a lot of color collisions are seen along the cell boundary, thereby causing the intercell interference. While when $t = 100$, due to the implementation of DRL-based intercell IC, the BSs can adjust their coloring results and thus, minimize the intercell interference.

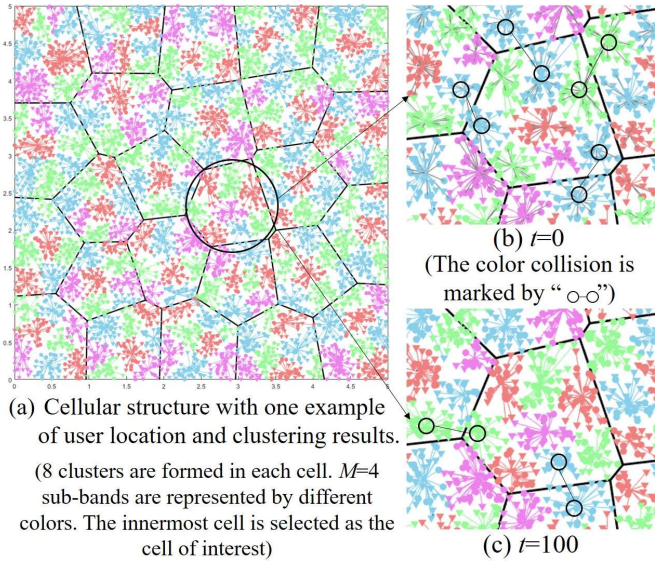

Fig. 3. An example of sub-band allocation by joint IC.

TABLE I

PARAMETER SETTING

Total number of DAs, N_A	3200
Total number of users, N_U	2400
Total number of clusters, N_C	200
The number of sub-bands, M	4
Pathloss exponent	3.5
Shadowing loss standard deviation in dB	8
P in Eqs. (2)	0dB
γ in Eqs. (9)	1
ε for ε-greedy policy	0.8
Memory size	50
Batch size	10
T^*	50
Number of neurons in each layer	[256,256,32]

In Fig.4, we plot the cumulative distribution function (CDF) of the sum capacity to evaluate our proposed joint IC when 8 clusters are formed in each cell. When GCA-based intracell IC is applied alone (indicated as red line), the intracell interference can be mitigated effectively, thus increasing the capacity at CDF=50% by 33% compared to the no IC case (indicated as the black line). While the joint IC, which adds DRL-based intercell IC on top of the GCA-based intracell IC, can mitigate both the intercell and intracell interferences, thus can further increase the sum capacity by 18% (by a total of 51% compared to the no IC case).

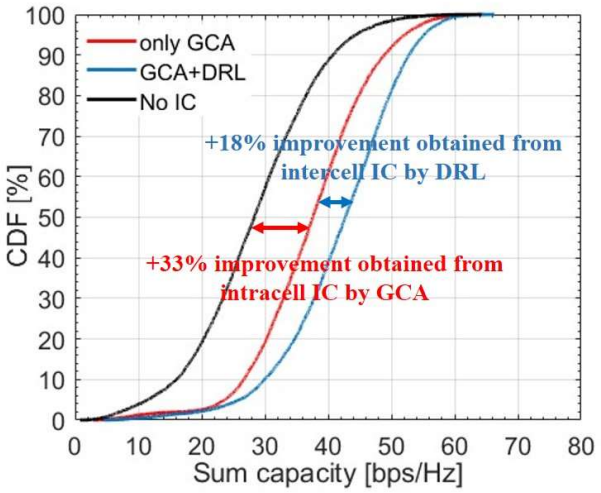

Fig. 4. The CDF of sum capacity.

6. Conclusion

In this paper, we proposed a joint IC consisting of a graph coloring algorithm (GCA)-based intracell IC and a deep reinforcement learning (DRL)-based intercell IC under the 2-layer IC framework to mitigate both the intercell interference and intracell interference in cellular system with cluster-wise distributed MU-MIMO. The simulation results have revealed that the proposed joint IC achieves a significant capacity increase compared to the no IC case in a time varying mobile radio environment.

Acknowledgement

A part of this work was conducted under “R&D for further advancement of the 5th generation mobile communication system” (JPJ000254) commissioned by Research and Development for Expansion of Radio Wave Resources of the Ministry of Internal Affairs and Communications in Japan.

References

[1] J. Gozalvez, “5G Worldwide Developments [Mobile Radio],” *IEEE Vehicular Technology Magazine*, vol. 12, no. 1, pp. 4-11, Mar. 2017. doi: 10.1109/MVT.2016.2641138.

[2] J. Joung, Y. K. Chia and S. Sun, “Energy-Efficient, Large-Scale Distributed-Antenna System (L-DAS) for Multiple Users,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 954-965, Oct. 2014. doi: 10.1109/JSTSP.2014.2309942.

[3] S. Xia, C. Ge, Q. Chen and F. Adachi, “Cellular Structuring and Clustering for Distributed Antenna Systems,” *Proc. 2021 24th International Symposium on Wireless Personal Multimedia Communications (WPMC2021)*, Okayama, Japan, 14-16 Dec. 2021, pp. 1-6. doi: 10.1109/WPMC52694.2021.9700460.

[4] Y. Wang, G. Feng, Y. Sun, S. Qin and Y. -C. Liang, “Decentralized Learning Based Indoor Interference Mitigation for 5G-and-Beyond Systems,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 12124-12135, Oct. 2020. doi: 10.1109/TVT.2020.3012311.

[5] M. Yan, J. Yang, K. Chen, Y. Sun and G. Feng, “Self-Imitation Learning-Based Inter-Cell Interference Coordination in Autonomous HetNets,” *IEEE Transactions on Network and Service Management*, vol. 18, no. 4, pp. 4589-4601, Dec. 2021. doi: 10.1109/TNSM.2021.3088837.

[6] C. Ge, S. Xia, Q. Chen and F. Adachi, "2-layer Interference Coordination Framework Based on Graph Coloring Algorithm for A Cellular System with Distributed MU-MIMO," in *IEEE Transactions on Vehicular Technology*, 2022, doi: 10.1109/TVT.2022.3219411.

[7] W. Jiang, B. Han, M. A. Habibi and H. D. Schotten, "The Road Towards 6G: A Comprehensive Survey," in IEEE Open Journal of the Communications Society, vol. 2, pp. 334-366, 2021, doi: 10.1109/OJCOMS.2021.3057679.

[8] Y. Qian, J. Xu, S. Zhu, W. Xu, L. Fan and G. K. Karagiannidis, "Learning to Optimize Resource Assignment for Task Offloading in Mobile Edge Computing," in IEEE Communications Letters, vol. 26, no. 6, pp. 1303-1307, June 2022, doi: 10.1109/LCOMM.2022.3159742.

[9] J. Xu, P. Zhu, J. Li and X. You, "Deep Learning-Based Pilot Design for Multi-User Distributed Massive MIMO Systems," in IEEE Wireless Communications Letters, vol. 8, no. 4, pp. 1016-1019, Aug. 2019, doi: 10.1109/LWC.2019.2904229.

[10] V. Mnih, K. kavukcuoglu, D. Silver. et al, “Human-level control through deep reinforcement learning,” *Nature,* vol. 518, pp. 529–533 (2015). doi: 10.1038/nature14236

[11] C. Yu and B. Liu, “A Backpropagation Algorithm with Adaptive Learning Rate and Momentum Coefficient,” *Proc. 2002 International Joint Conference on Neural Networks (IJCNN'02), Cat. No.02CH37290*, vol.2, pp. 1218-1223. doi: 10.1109/IJCNN.2002.1007668.