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

Reinforcement Learning-based Graph Coloring Algorithm for Interference Coordination in Distributed MU-MIMO

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

†Research Organization of Electrical Communication, Tohoku University

2-1-1 Katahira, Aoba-ku, Sendai, Miyagi, 980-8577, 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 this paper, we propose a reinforcement learning based graph coloring algorithm (RL-GCA) to solve the interference coordination problem for cluster-wise distributed multi-user multi-input multi-output (MU-MIMO). Compared with other non-intelligent GCAs, such as our previously proposed RCN-GCA and the well-known DSATUR, our newly proposed RL-GCA is able to significantly improve the link capacity of cluster-wise distributed MU-MIMO. Also, the detailed discussion about the chromatic number and convergence analysis are included in this paper.

Keywords Reinforcement learning, graph coloring, interference coordination, distributed MU-MIMO, frequency allocation, machine learning

1. Introduction

The deployment of 5G has begun in many countries to provide new mobile communication services. The mobile data traffic is growing at a compound annual growth rate (CAGR) of 46% [1]. The massive multi-user multi-input multi-output (MU-MIMO) is adopted to solve the ever-growing mobile data traffic using the limited radio bandwidth [2]. There are two architectures in massive MU-MIMO, known as the co-located MU-MIMO and distributed MU-MIMO [3]. The distributed MU-MIMO becomes more and more attractive than the co-located MU-MIMO in recent years because when the mm-wave band is used, the spatially deployed antennas in distributed MU-MIMO is able to relieve the problem of radio link blockage which is caused by the nature of rectilinear propagation [4]. However, the large-scale MU-MIMO requires a prohibitively high computational complexity. To solve this problem, idea of clustering is introduced. With clustering, the large-scale distributed MU-MIMO is divided into several small-scale distributed MU-MIMO [5][6]. However, the introduction of clusters brings a problem of inter-cluster interference [5]. Therefore, effective interference coordination algorithm is required.

The severe inter-cluster-interference comes from the neighboring clusters if the same frequency band is assigned to them. For interference coordination, frequency allocation is a commonly used method [7]. Borrowing the idea from graph theory, we describe the clusters and their mutual relationship as an undirected graph $G=(V,E)$, in which vertices denote the clusters while the edge denotes the neighboring relationship. Based on the graph G, the frequency allocation problem can be abstracted as a vertex graph coloring problem and graph coloring algorithm (GCA) [8] can be applied.

The heuristic GCA can be applied due to its simplicity and reasonable computational complexity. In our previous study [9], we proposed a heuristic algorithm named as restricted color number-based GCA (RCN-GCA), in which a more than 50% increase in the sum capacity can be obtained compared with 1-color case. Besides our application in distributed MU-MIMO, Y. Zhao [10], L. Chen [11], and Q. Zhang [12] also applied heuristic GCA to mitigate the severe co-tier interference in dense small cell or femtocell systems.

However, what heuristic algorithm can obtain is only a sub-optimal result. The machine learning, which has been attracting attention in the communications field in recent years [14]~[17], has a potential for further improvement. Especially, the reinforcement learning (RL) [13], is attracting much attention. Among the RL, the Q-learning is probably the most well-known RL algorithm, which has been successfully applied in some related research areas [14]~[17]. In [14], M. Simsek proposed a Q-learning based time-domain inter-cell interference coordination in a heterogeneous network (HetNet). In [15], K. Nakashima proposed a deep learning-based channel allocation scheme

This article is a technical report without peer review, and its polished and/or extended version may be published elsewhere.

for densely deployed wireless local area networks (WLANs). In [16][17], the multi-agent Q-learning-based RL was applied. In [16], in order to optimize the joint subcarrier and power resource allocation, Y. Hu firstly applied non-intelligent algorithm to get the sub-optimal results, and then applied the multi-agent Q-learning to further improve the spectral efficiency in multi-cell OFDM system. While in [17], G. Bu solved the user scheduling and resource allocation in massive MU-MIMO system by multi-agent Q-learning based scheme. Compared with the single-agent Q-learning (such as applied in [14][15]), the multi-agent Q-learning used in [16][17] is able to handle more complex situation with sequential decisions.

In this paper, we propose a multi-agent Q-learning based interference coordination algorithm to mitigate the inter-cluster-interference in ultra-dense RAN with distributed MU-MIMO. Since we abstract our problem as graph coloring problem, we name our algorithm as RL-GCA. We will compare it with our previously proposed RCN-GCA [9] and also the well-known DSATUR [18]. We will show by the computer simulation that our proposed RL-GCA overwhelms those two non-intelligent GCA.

The rest of paper is organized as follows. In Sect. II, cluster-wise distributed MU-MIMO system model is presented. In Sect. III, a quick review of RCN-GCA and DSATUR is presented. In Sect. IV, the proposed RL-GCA is described. The link capacity evaluation by computer simulation is presented together with the convergence analysis in Sect. V. Finally, Sect. VI offers the conclusion and future research plan.

2. Cluster-wise Distributed MU-MIMO System Model

In order to make a fair comparison with our previous studies, the same simulation model is adopted. In our simulation, the base station coverage area (or cell) is supposed to be a 1 by 1 square shaped area where 128 antennas are randomly distributed. For MU-MIMO, the total number of users in the area is set between the number of clusters (at least each cluster has one user) and the number of antennas (to meet the requirement of Zero-Forcing (ZF) algorithm). In this paper we adopt the user-based clustering method (based on k-means algorithm) [9]. Fig. 1 illustrates the system model with 8 clusters.

3. Review of Previous Work

In this section, we quickly review our previously proposed RCN-GCA and DSATUR in order to better illustrate the difference between the intelligent GCA (such as the RL-GCA) and non-intelligent GCA (such as the RCN-GCA and DSATUR). For non-intelligent GCA, graph $G = (V, E)$ should be defined first. Based on the graph G, the heuristic method can be applied. The commonly used heuristic method is to firstly set all the vertex in a specific order, and then each time, assign the smallest color index to each vertex on the premise that this color is not used by its neighbors. In RCN-GCA, we set the clusters in descending order according to their degrees, while in DSATUR, the ordering is based on the degree of saturation. The different ordering decides the final coloring results, which in turns decides the chromatic number $\chi(G)$. The coloring results based on RCN-GCA and DSATUR are shown in Fig.2, where $\chi(G) = 4$ for RCN-GCA while $\chi(G) = 3$ for DSATUR. The $\chi(G)$ is of great importance because it decides how many parts the entire bandwidth will be divided into. Thus, in non-intelligent GCA, we try to restrict the chromatic number $\chi(G)$ as less as possible so as not to divide the entire bandwidth into too many narrow parts.

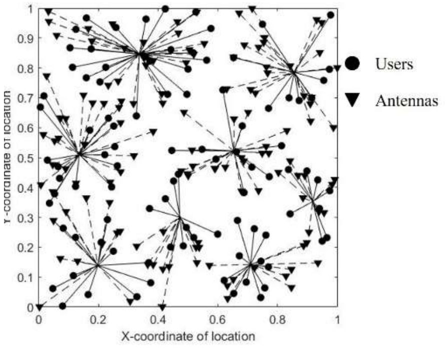
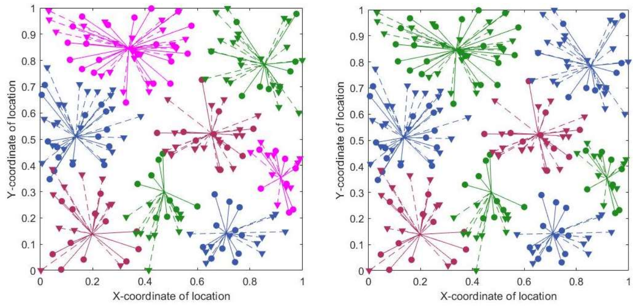


Fig. 1. System model with 8 clusters.



(a) RCN-GCA (b) DSATUR

Fig. 2. Coloring results of RCN-GCA and DSATUR.

4. Proposed RL-GCA

In this section, we will explain how our proposed RL-GCA works. RL can be interpreted as a Markov Decision Process (MDP), which in general, contains *states(S)*, *actions(A)* and *rewards(R)*. Fig.3 illustrates the framework of our proposed RL-GCA. We suppose that each cluster works as an *agent* and makes its own decision to choose the appropriate color for itself. The entire single-cell ultra-dense MU-MIMO system is regarded as the *environment*. Each *agent* observes its *state(s)* and based on the *state*, it selects the *action(a)* with the help of *action selection policy (π)*. The performed *action* reacts on the *environment*. Therefore, the *state* which the agent-1 observed is different from the *state* which the agent-2 observed. Meantime, the *environment* will evaluate this *action* by offering *reward(r)*. After a set of training, the *agents* are able to make the best decisions (*action*) at the moment (*state*).

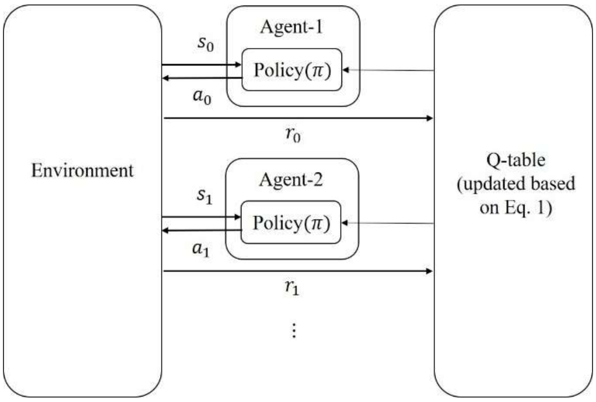


Fig. 3. Framework of RL-GCA.

The core of the Q-learning algorithm is to iteratively update the Q-value through Bellman equation. The Q-learning algorithm can be expressed as

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_t + \lambda \max_{a\in A} Q(s_{t+1},a) - Q(s_t,a_t)], (1)$$

where α denotes the learning rate and λ denotes the discount factor. The Q-value's function is to evaluate the quality of the action been selected, store it in Q-table, and update it in each step t iteratively. The discount factor λ is set to be 1 in this paper because in our situation each cluster's coloring result has influence to the remaining clusters until each training episode end.

Our objective of utilizing RL-GCA is to mitigate the inter-cluster interference. Therefore, our optimization objective is to minimize the sum of inter-cluster interference in each color group. The set of total clusters is denoted by κ. For the case when N colors are available, κ_n denotes the cluster subset of n^{th} color group. $\mathbf{D}_{\text{n}} = (d_{ij}) \in \mathbb{R}^{N_{k_n} \times N_{k_n}}$ denotes the cluster-centroids' distance matrix of κ_n in which N_{k_n} denotes the number of clusters in κ_n . If we simply use pathloss to represent the inter-cluster interference level in κ_n, the interference matrix can be denoted as $\mathbf{ICI}_{\text{n}} = (ic_{ij}) \in \mathbb{R}^{N_{k_n} \times N_{k_n}}$, in which $ici_{ij} = d_{ij}{}^{-\gamma}$with γ being the pathloss exponent. The optimization objective can be expressed as follows

$$\begin{aligned} &\min_{\kappa_n \subseteq \kappa} \sum_{n=1}^{N} \|\mathbf{ICI}_n\|_F \\ &s.t.\ \forall n \in N \qquad\qquad\qquad , \qquad (2) \\ &\quad \bigcup_{n\in N} \kappa_n = \kappa, \text{ and } \kappa_n \bigcap \kappa_m = \varnothing,\ \forall n \neq m \end{aligned}$$

where $\|\cdot\|_F$ represents the Frobenius norm.

The three elements of RL-GCA (actions, states, rewards) are defined as follows:

- **Actions**: In our case, the *action* is defined as each cluster picks one color from the color pool (all the colors available) $a_t \in \{1,2,\ldots,N\}$. As for the policy (π) , we adopt the widely used ε-greedy [16] as follows.

  $$a_t = \begin{cases} \arg\max_{a\in A} Q(s_t,a), & \text{with probability } \varepsilon \\ \text{Choose a random color,} & \text{with probability 1-}\varepsilon \end{cases} . (3)$$

  The ε-greedy policy is able to deal with the tradeoff between the exploration and exploitation. The discussion about the value of ε is shown in Sect. V.

- **States:** The state at step t can be denoted as $s_t = [s_1\, s_2 \cdots s_{N_\kappa}]$, in which N_κ represents the number of clusters in total set κ. Accordingly, $s_i \in \{0,1,2,\ldots,N\}$, which denotes the color index for each cluster, while "0" indicates the clusters not been colored.

- **Rewards**: The rewards works as a way to judge whether the action is good or not. Therefore, how to define the rewards is to be carefully considered. In this paper, the reward at step t is defined as

  $$r_t = -(\sum_{n=1}^{N} \|\mathbf{IC}_n\|_F)_t - (\sum_{n=1}^{N} \|\mathbf{IC}_n\|_F)_{t-1}, \qquad (4)$$

  where $(\|\mathbf{IC}_n\|_F)_t$ and $(\|\mathbf{IC}_n\|_F)_{t-1}$ represent the total inter-cluster interference in each κ_n at step t and $t-1$.

The algorithm of our proposed RL-GCA [19] is described in **Algorithm**. In RL-GCA, each episode is a training session, while each step represents each agent's Q-learning process. In order to perform our proposed RL-GCA, only the information about users' location is required, which can be easily got via Global Positioning System (GPS) or other methods.

```
Algorithm: RL-GCA [19]
  Input: $\alpha,\varepsilon,\lambda,\boldsymbol{D}$.
```

```
Initialize Q-table.
for each episode
Initialize current-state.
for each step $t$ do
    Select action (for the scheduled cluster,
     select one color based on polity ($\pi$)).
     Update next-state.
     Calculate reward.
     Update Q-table.
     current-state ← next-state.
 end do
 end for
```

The final coloring results obtained by RL-GCA is shown in Fig. 4.The chromatic number $\chi(G) = 4$.

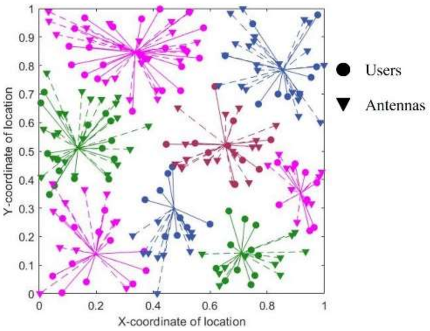


Fig. 4. RL-GCA coloring results.

5. Monte Carlo Simulation

5.1 Link capacity analysis

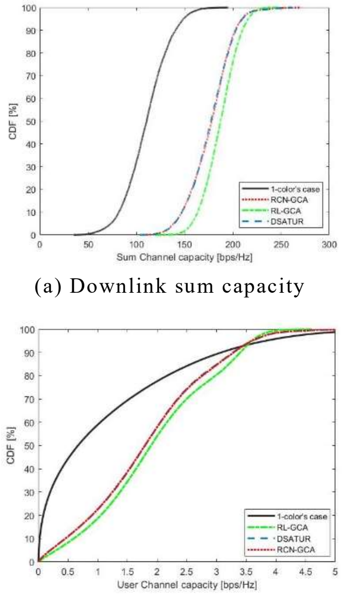
In this section, we evaluate the downlink sum capacity and user capacity to compare the performance of our newly proposed RL-GCA with other non-intelligent GCAs. For simplicity, a square-shaped BS area is assumed. The simulation setting is shown in Table I. In this paper, we adopt equal power allocation, the normalized transmit signal power-to-noise ratio for each user is set to 0dB, which means that the transmit power for each user is set so as to the received signal-to-noise ratio becomes 0dB when the distance between the transmitter and receiver is equal to the side length of square-shaped BS area.

TABLE I. SIMULATION SETTINGS

Parameter	Value/State
Number of antennas	128
Number of users	96
Number of clusters	6,8
Pathloss exponent	3.5
Shadowing standard deviation	8 [dB]
Fading type	Rayleigh fading
Number of different user location patterns	100
Number of shadowing generation per user location pattern	10
Number of fading generation per shadowing generation	10
Discount factor λ	1

In our simulation we consider the quasi-static environment, which means that user location does not change during the communication duration. This environment is simulated by changing shadowing loss and fading several times for each realization of users' locations. In the simulation, the locations of users are randomly generated 100 times. For each generation of user locations, the shadowing loss and Rayleigh fading are generated 100 times. After a total of 10,000 channel realizations including pathloss, shadowing loss, and Rayleigh fading, the cumulative distribution functions (CDFs) of the sum capacity and the user capacity are obtained.

Fig. 5 plots the CDFs of the sum capacity and the user capacity for the case of 96 users, 128 antennas, and 6 clusters. It can be seen from Fig. 5(a) that our proposed RL-GCA provides the highest capacity and improves the sum capacity at CDF=50% by 69% compared with 1-color case (i.e., no interference coordination). On the other hand, the non-intelligent GCAs (RCN-GCA and DSATUR) achieve around 60% improvement. From the results of user capacity in Fig. 5(b), the RL-GCA achieves 28 times higher user capacity at CDF=10% than the 1-color case while the non-intelligent GCA achieves 20 times higher than the 1-color case.



(a) Downlink sum capacity

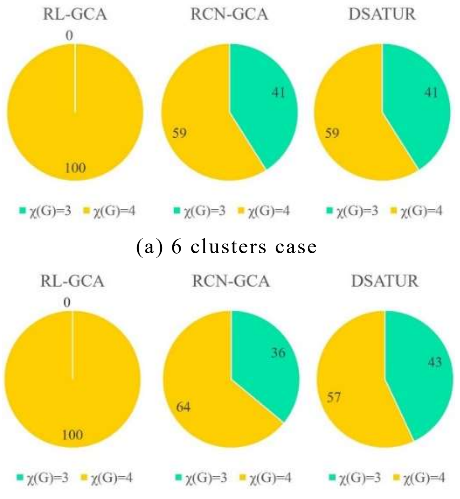
(b) Downlink user capacity

Fig. 5. Link capacity comparison between RL-GCA, RCN-GCA, and DSATUR.

5.2 Chromatic number $\chi(G)$ analysis

Besides the link capacity discussion, we also find out some interesting conclusion through the analysis of chromatic number $\chi(G)$. As we mentioned earlier, a smaller number of colors was considered to be desirable. Therefore, the 3-color solution was considered better than the 4-color solution [8]. However, according to the statistic results of the chromatic number distribution over 100 user location patterns shown in Fig. 6, it can be seen that the RL-GCA chooses 4 colors always while the non-intelligent GCAs choose less chromatic number. Note that the RL-GCA provides higher capacity than the non-intelligent GCAs. Therefore, excessively pursuing the less chromatic number $\chi(G)$ is not necessary.

The reason that the RL-GCA can provide better solution is that it can provide end-to-end global optimization. On the other hand, for the case of non-intelligent GCAs, the tradeoff between the mitigation of interference and the segmentation of bandwidth must be carefully considered. Therefore, it is an indirect process, and a degree of assumptions and simplifications exists.



(a) 6 clusters case

(b) 8 clusters case

Fig. 6. $\chi(G)$ distribution over 100 user location patterns.

5.3 The comparison about the exploitation rate ε

In this paper, we adopted the ε-greedy as the *action selection policy* (π). However, the different value of exploitation rate ε will affect the convergence speed. $\varepsilon\,\epsilon[0,1]$, if $\varepsilon=0.9$, which means 90% episodes are used for exploitation based on the Q-table, while 10% episodes are used for exploration based on randomly searching. Actually, a certain level of random search is to avoid to fall in local optimization due to the lack of exploration. Fig.7 compares the convergence speed for different values of ε. When $\varepsilon=0.9$, even though the capacity increases very fast at the beginning, but it soon gets trapped in a local optimization between episode 500 and episode 3000. On the other hand, lack of exploitation will have difficulties in obtaining the optimized results. When $\varepsilon=0.3$ the converge is slow because a lot of episodes are "wasted" due to random search. In this case of 96 users, 128 antennas and 6 clusters, $\varepsilon=0.5$ or 0.7 is a better choice.

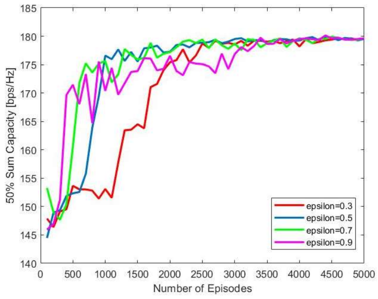


Fig. 7. The comparison of the exploitation rate ε.

5.4 The comparison about the learning rate α

The learning rate α (or the step size) also influences how quickly the final results can be obtained. Too large α may results in divergence while too small α will lower the convergence speed. In Fig. 8, when $\alpha=0.1$, the training converges until around 4500 episodes, and with smaller value of $\alpha=0.01$, the training even cannot converge within 5000 episodes. But if α is increased to 1, less than 1000 episodes is enough for training. However, when α increased to be as large as 5, the training does not work well.

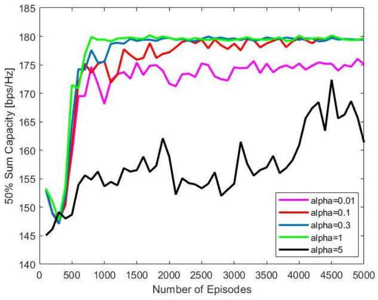


Fig. 8. The comparison of the learning rate α

6. Conclusion

In this paper, we proposed a reinforcement learning-based graph coloring algorithm (RL-GCA) for inter-cluster interference coordination for cluster-wise distributed MU-MIMO. It was confirmed that the RL-GCA achieve higher link capacity compared with our previously proposed RCN-

GCA and the well-known DSATUR. Besides that, an interesting conclusion that smaller chromatic number does not necessarily achieve higher link capacity was obtained . The best chromatic number which maximizes the achievable link capacity was shown to be 4. Also based on our convergence analysis, the parameter setting, such as α and ε, has a great influence on the convergence speed. Too big α will result in diverge but too small value is time-wasting. Similarly, too big ε will be easily trapped in sub-optimal value, while too small ε also lower the converge speed, therefore should be carefully designed.

In our future study, we will introduce the neural network into our algorithm to form a deep-learning based architecture to further reduce the computational complexity.

Acknowledgment

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] V. Cisco, "Cisco visual networking index: Forecast and trends, 2017-2020," White paper, vol. 1, 2018.

[2] X. Ge, S. Tu, G. Mao,C. Wang, "5G Ultra-Dense Cellular Networks," IEEE Wireless Communications, Vol. 23, Issue:1, pp. 72–79, March 2016.

[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," IEEE Communications Magazine, vol. 52, no.2, pp. 186-195, February 2014.

[4] J. Joung, Y. K. Chia, and S. Sun, "Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users," IEEE J. Selected Topics in Signal Processing, Vol. 8, No. 5, pp.954965, Oct. 2014.

[5] S. Xia, C. Ge, Q. Chen, and F. Adachi, "A Study on User-antenna Cluster Formation for Cluster-wise MU-MIMO," 2020 23rd International Symposium on Wireless Personal Multimedia Communications (WPMC),19-26 Oct. 2020.

[6] T. D. Tsvetkov and I. G. Iliev, "Computational Complexity Analysis of Cognitive Radio using PCA with Various Clustering Methods," 2020 28th National Conference with International Participation (TELECOM), Sofia, Bulgaria, 2020, pp. 145-148, doi: 10.1109/TELECOM50385.2020.9299558.

[7] A.S.Hamza, S.S.Khalifa, H.S.Hamza, and K. Elsayed, "A Survey on Inter-Cell Interference Coordination Techniques in OFDMA-Based Cellular Networks," IEEE Communications Survey & Tutorials, Vol. 15, No. 4, Fourth Quarter, 2013.

[8] M. Garey, D. Johnson, "Computers and Intractability: A Guide to the Theory of NP-Completeness," W.H. Freeman and Co., USA.

[9] C. Ge, S. Xia, Q. Chen, and F. Adachi, "2-Step Graph Coloring Algorithm for cluster-wise Distributed MU-MIMO in ultra-dense RAN," 2020 23rd International Symposium on Wireless Personal Multimedia Communications (WPMC),19-26 Oct. 2020.

[10] Y. Zhao, H. Xia, Z. Zeng and S. Wu, "Joint clustering-based resourse alllocation andpower control in dense small cell networks," Proc. IEEE/CIC ICCC 2015 Symposium on Wireless Communications Systems, Shenzhen, China, 2-4 Nov. 2015. DOI: 10.1109/ICCChina.2015.7448694.

[11] L. Chen, H. Xia, C. Feng and S.WU, "Clustering-based co-tier interference coordination in dense small cell networks," 2015 IEEE 26th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Hong Kong, China, 30 Aug.- 2 Sept. 2015. DOI: 10.1109/PIMRC.2015.7343605.

[12] Q. Zhang, X. Zhu, and L. Wu, K. Sandrasegaran, "A coloring-based resource allocation for OFDMA femtocell networks," Proc. 2013 IEEE Wireless Communications and networking Conference (WCNC), Shanghai, China, 7-10 Apr. 2013. DOI: 10.1109/WCNC.2013.6554644.

[13] A. Gosavi, "Reinforcement Learning: A Tutorial Survey and recent Advances," Informs Journal on Computing 21(2): 178-192, DOI:10.1287/ijoc.1080.0305.

[14] M. Simsek, M. Bennis, İ. Güvenç, "Learning Based Frequency- and Time-Domain Inter-Cell Interference Coordination in HetNets," IEEE Transactions on Vehicular Technology, Vol. 64, No. 10, October 2015.

[15] K. Nakashima, S. Kamiya, K. Ohtsu, K. Yamamoto, "Deep Reinforcement Learning-Based Channel Allocation for Wireless LANs With Graph Convolutional Networks," IEEE Access, Vol.8, pp. 31823-31834, Feb. 2020, DOI: 10.1109/ACCESS.2020.2973140.

[16] Y. Hu, M. Chen, Z. Yang, M. Chen, G. Jia, "Optimization of Resource Allocation in Multi-Cell OFDM Systems: A Distributed Reinforcement Learning Approach," 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, 31 Aug.-3 Sept. 2020, London UK, DOI: 10.1109/PIMRC48278.2020.9217276.

[17] G. Bu, J. Jiang, "Reinforcement Learning-Based User Scheduling and Resource Allocation for Massive MU-MIMO System," 2019 IEEE/CIC International Conference on Communications in China (ICCC), 11-13 Aug. 2019, Changchun China, DOI: 10.1109/ICCChina.2019.8855949.

[18] D. Brélaz, "New methods to color the vertices of a graph," Communications of the ACM, Vol.22, Issue 4, pp. 251-256, Apr. 1979.

[19] C. Ge, S. Xia, Q. Chen, and F. Adachi, "Reinforcement Learning-based Interference Coordination for Distributed MU-MIMO," submitted to The 24th International Symposium on Wireless Personal Multimedia Communications (WPMC), 12-16 Dec. 2021.