Learning Based on Graph: A Joint Interference Coordination for Cluster-Wise Distributed MU-MIMO

Chang Ge, *Graduate Student Member, IEEE*, Sijie Xia, *Graduate Student Member, IEEE*, Qiang Chen, *Senior Member, IEEE*, and Fumiyuki Adachi, *Life Fellow, IEEE*

***Abstract*— In a cellular system with distributed MU-MIMO, an application of cluster-wise distributed MU-MIMO reduces the computational complexity. However, both the intracell interference and the intercell interference are produced. Considering the scalability of the system, in this letter, we propose a fully decentralized interference coordination (IC) which jointly applies the graph coloring algorithm (GCA) and the deep reinforcement learning (DRL). Based on online training with consideration of the time-varying wireless environment, our proposed joint IC can adapt quickly to the changing environment. The simulation reveals that our proposed joint IC can significantly improve the capacity compared to the no IC case.**

***Index Terms*— Interference coordination, deep reinforcement learning, graph coloring algorithm, distributed MU-MIMO.**

I. Introduction

In 5G and beyond, massive MU-MIMO has been regarded as a promising technique [1]. In particular, distributed MU-MIMO [2], which exploits distributed antennas (DAs) over the base station coverage area (or BS cell), can relieve the problem of radio link blockage resulting from the utilization of mm-wave band. In our previous research, we proposed a cluster-wise distributed MU-MIMO [3], where users are dynamically divided into non-overlapping sub-groups called user-clusters (hereafter, simply called clusters) based on the user location information. Then, a large-scale cell-based MU-MIMO can be replaced with performing small-scale cluster-based MU-MIMOs in parallel, so that the computational complexity required for signal processing can be greatly reduced. 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 from the same cell, is relative straightforward because each base station (BS) has all the information about its governing clusters. While the intercell IC, which aims to mitigate the interference exists between clusters that belong to different cells but face each other along a cell boundary, is much more difficult to realize.

In recent years, with the rise of artificial intelligence (AI), especially the reinforcement learning (RL), some new progresses for intercell IC in cellular system have emerged. As early as in 2015, Simsek et al. [4] have tried to apply the Q-learning algorithm to solve the intercell IC among macrocells and picocells in a Heterogeneous network (HetNet). In order to overcome the memory and computational limitation problems that come with tabular-based Q-learning algorithm, the authors proposed to store the probability distribution over all actions instead of the state-action combination in Q table.

In more recent years, with the development of deep learning technology, deep reinforcement learning (DRL) embedded with updatable neural networks has been able to solve large-scale problems more efficiently than tabular-based RL. In 2020, in order to solve the intercell IC problem in an ultra-dense network with small-cell BSs deployed in a residential area, Wang et al. [5] 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, Yan et al. [6] applied the Double DQN to schedule sub-channels to individual users. In order to improve the robustness of Double DQN, Wasserstein Generative Adversarial Networks (W-GANs) is incorporated together.

Inspired by the above-mentioned contributions, we also want to explore the application of DRL for intercell IC in the cellular system with cluster-wise distributed MU-MIMO. However, the intercell IC problem we face is even more challenging than the above-mentioned scenarios, which requires the consideration of intracell IC constraints while doing intercell IC. In our previous research [7], we have explored the application of graph coloring algorithm (GCA) for intracell IC. The GCA-based intracell IC is able to divide the entire available bandwidth into several sub-bands, and assign the different sub-bands to neighboring clusters inside each cell.

Manuscript received 25 November 2022; revised 25 December 2022; accepted 20 January 2023. Date of publication 25 January 2023; date of current version 10 March 2023. 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. The associate editor coordinating the review of this letter and approving it for publication was W. Jiang. *(Corresponding author: Chang Ge.)*

Chang Ge, Sijie Xia, and Qiang Chen are with the School of Engineering, Tohoku University, Sendai 980-8579, Japan (e-mail: ge.chang.q2@dc.tohoku.ac.jp; xia.sijie.p2@dc.tohoku.ac.jp; qiang.chen.a5@tohoku.ac.jp).

Fumiyuki Adachi is with the International Research Institute of Disaster Science, Tohoku University, Sendai 980-8572, Japan (e-mail: fumiyuki.adachi.b4@tohoku.ac.jp).

Digital Object Identifier 10.1109/LCOMM.2023.3239605

Because the neighboring clusters do not share the same sub-bands, the serious intracell interference can be successfully mitigated. However, since each cell applies the GCA-based intracell IC independently, and the coloring result will not be shared with the surrounding cells, the color collision may happen at clusters facing each other along the cell boundary, thus the intercell interference is unavoidable. Besides that, in cellular system with cluster-wise distributed MU-MIMO, the clustering results will change according to the changes in the user's location, and the changes of the clustering results will directly lead to the changes in the GCA results. Therefore, in a dynamic environment considering the users' mobility, where the color collision occurs and among which color it occurs is time-varying and totally unpredictable. Under this scenario, the DRL which springs up recently overwhelms the traditional methods in both the flexibility and adaptability in dynamic environment. By interacting with the unknown environment, the DRL is able to figure out the solution on its own with limited number of trial and errors. Therefore, we try to apply the Deep Q network (DQN) from DRL to dynamically eliminate the color collision, and the corresponding intercell interference, by self-learning with only locally observed information.

In this letter, we propose a joint IC solution, aiming at maximizing the capacity of a cellular system with cluster-wise distributed MU-MIMO in a fully decentralized manner by combining the GCA and the DRL. The GCA-based intracell IC is applied first in each cell independently to allocate different sub-bands to the neighboring clusters. Then, under the constraint of the existing GCA results, based on only the locally observed information, the selected cells turn on the DRL-based intercell IC to dynamically choose one color-adaptation scheme to adjust the existing coloring result in order to minimize the occurrence of the intercell interference from surrounding cells.

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

The remainder of this letter is organized as follows. Section II provides the system model and the problem formulation. In Section III, the proposed joint IC based on GCA and DRL is described. The simulation analysis is conducted in Section IV, and Section V concludes this letter.

II. System Model and Problem Formulation

The structure of a cellular system with cluster-wise distributed MU-MIMO is illustrated in Fig.1. Our proposed joint IC is planned to be applied based on O-RAN architecture [8], in which the near-real-time (near-RT) radio access network intelligent controllers (RICs) with the xAPPs, and the non-RT RIC with rApps are introduced. The near-RT RICs are designed to be the specific executor to control one or several cells, while the non-RT RIC is to provide guidance for the near-RT RICs with its global optimization and monitoring capability.

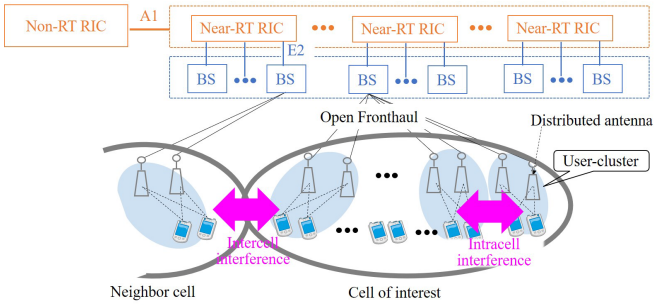

Fig. 1. System model of cellular system with cluster-wise distributed MU-MIMO in O-RAN architecture.

In our proposed joint IC, the entire bandwidth is divided into M sub-bands and one of the sub-bands is assigned to each cluster. The set of entire clusters and the set of clusters which are assigned to the m^{th} sub-band in the service area are denoted by κ and κ_m, $m \in \{1, \cdots, M\}$, respectively. In this letter, the numbers of users, DAs, and clusters in κ are denoted by N_U, N_A, and N_C, respectively. While those in the 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 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

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

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 $\{\kappa_m; m \in \{1, \cdots, M\}\}$. 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:

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

where

$$C_m = \sum_{k=1}^{N_C^m} \sum_{i=1}^{N_{U,k}^m} C_{u_{i,k}^m} \tag{5}$$

III. Joint IC Based on GCA and DRL

A. The Framework of Joint IC

The framework of our proposed joint IC is illustrated in Fig.2. The clustering, together with the joint IC (including the GCA-based intracell IC and the DRL-based intercell IC) are designed to be applied as the xAPPs on each near-RT RICs, respectively. During the communication, each near-RT RIC updates the clustering results based on the users' movement and associate the DAs according to the principle of proximity. The updating of the clustering results will trigger the GCA-based intracell IC (described in Sect. III-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. III-C). Then, the selected cells will work independently to mitigate the intercell interference with only the locally observed information.

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 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 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 the change in $s^{(t)}$ and $s^{(t+1)}$.

Because the online training strategy is adopted in this letter, we assume that the wireless environment at $s^{(t)}$ and $s^{(t+1)}$ are different, with future information completely unknown. Unlike the commonly used offline training [9], [10], online training can ensure that the parameters of DQN been constantly updated during the real communication process and thus is able to follow the changing 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 dynamic environment.

To enable efficient online training, in this letter, we assume that each cell is equipped with a fixed size of memory pool, in which the state transition sequence $\Delta^{(t)} = (s^{(t)}, a^{(t)}, s^{(t+1)}, r^{(t+1)})$ 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.

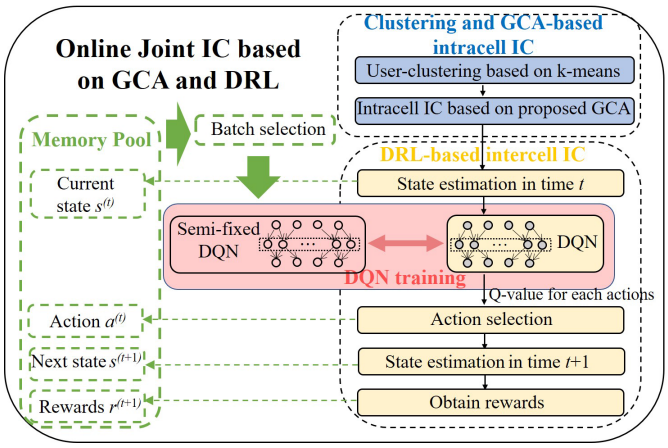

Fig. 2. The framework of joint IC.

The application of memory replay and batch selection [11] 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.

As a preliminary study of the application of DRL for IC under O-RAN architecture, in this letter 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 future researches.

B. GCA-Based Intracell IC

In our previous study [7], we explored about how to model the problem of IC as a graph and apply GCA from graph theory to optimize the sub-bands allocation in order to mitigate 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 M =4. We also proposed an GCA in which the value of M is controllable. In this letter, we apply the GCA of [7] for intracell IC.

C. DRL-Based Intercell IC

Since we assume a fully decentralized framework, we suppose each BS is a single agent, and 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 noted 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 $A = \{1, 2, \cdots, M\}$. Let the coloring result for the k^{th} cluster after GCA be $g_k \in \{0, 1, \cdots, M-1\}$. 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

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

As for the action selection policy (π), we adopt the well-known ε-greedy policy [11] 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 letter is an extension of the basic Q-learning algorithm [11], 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 been replaced by a neural network, is our better choice. The computational complexity of DQN during implementation process only depends on the complexity of matrix multiplication, therefore it is $\mathrm{O}(\sum_{l=1}^{L} n_l n_{l-1})$ [10], in which $\mathscr{L} = \{0, \cdots, L\}$ represent the set of layers, $l = 0$ and $l = L$ denote the input layer and output layer respectively, n_l denote the number of neurons of each layer $l \in \mathscr{L}$. In order to better train the DQN, we adopt the semi-fixed target network method [11], 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 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.

IV. Simulation Results

We consider a normalized area of 5 × 5 over which 25 cells are constructed as shown in Fig.3(a). Because the user movement in a small range will not affect the clustering results, in this letter, we adopted the quasi-static simulation.

TABLE I
PARAMETER SETTINGS

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
DQN	3-layer fully connected artificial neural network
Number of neurons in each layer	[256,256,32]

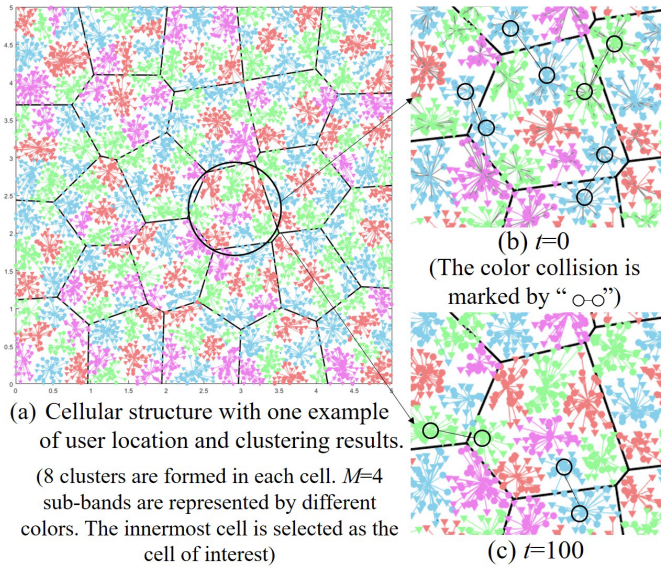

Fig. 3. An example of sub-band allocation results by joint IC.

The users' locations are generated randomly for 100 times, and for each generation of user locations, the channels are 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 letter, we suppose each cell applies the GCA-based intracell IC, while the cell which locates in the center area and receives the intercell interference from every direction is selected as the cell of interest and turns on the DRL-based intercell IC. In this way, by comparing the performance of the no IC case, only GCA case and the GCA+DRL case in the cell of interest, the effectiveness of GCA-based intracell IC and DRL-based intercell IC of our proposed joint IC can be evaluated, respectively. 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. 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 that the intracell interference can be mitigated. But a lot of color collisions are seen along the cell boundary,

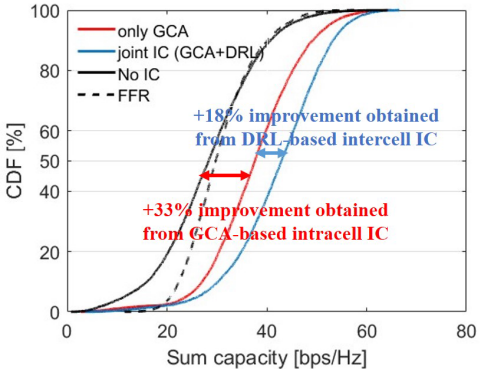

Fig. 4. The CDF of sum capacity.

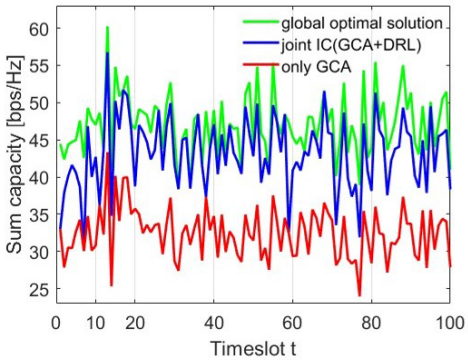

Fig. 5. The convergence of DQN.

thereby causing the intercell interference. While when $t = 100$, due to the implementation of DRL-based intercell IC, the BSs can adjust the coloring result and thus, minimize the intercell interference.

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 for 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 increases the sum capacity by 18% (by a total of 51% compared to the no IC case). Compared with the well-known fractional frequency reuse scheme (FFR) [12], which can only achieve 5% improvement, our proposed joint IC has significant advantages.

To achieve the adaptability of the DRL-based intercell IC to the dynamic environment via online training, the convergence speed of DQN is of vital importance. Therefore in Fig.5, we illustrate the sum capacity variation during the beginning 100 timeslots for a new updating of clusters when DRL-based intercell IC is applied. For comparison, the only GCA case and the global optimal case are also provided. (Note that the global optimal solution is obtained by applying the GCA in the non-RT RIC in a fully centralized manner.) It is clearly seen from Fig. 5 that after a dozen of timeslots for training, the DQN can adapt to the environment, thus providing a sub-optimal solution which has higher capacity than the IC with GCA only. The convergence speed of DQN can accommodate the requirements of online training, therefore, convinced that our proposed joint IC based on online training can quickly keep up with the changes in the dynamic environment.

V. Conclusion

In this letter, we proposed a joint interference coordination (IC) which combines the advantages of graph coloring algorithm (GCA) and the deep reinforcement learning (DRL) so as to realize the fully decentralized intracell IC and intercell IC in a cellular system with cluster-wise distributed MU-MIMO. The GCA is firstly applied to mitigate the intracell interference which is produced between user-clusters inside each cell, then DRL is applied based on only locally observed information to adjust the existing coloring result to mitigate the intercell interference from user-clusters in surrounding cells. Based on online training with consideration of the time-varying wireless environment, our proposed joint IC can adapt quickly to the changing environment. The simulation confirmed that our proposed joint IC can approximates the optimal solution and significantly improve the sum capacity compared to the no IC case and FFR.

In the letter, the perfect CSI is assumed. The capacity improvement achievable with our proposed joint IC under the imperfect CSI is left as our future study.

References

[1] J. Gozalvez, "5G worldwide developments [mobile radio]," *IEEE Veh. Technol. Mag.*, vol. 12, no. 1, pp. 4–11, Mar. 2017.

[2] J. Joung, Y. K. Chia, and S. Sun, "Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 954–965, Oct. 2014.

[3] S. Xia, C. Ge, Q. Chen, and F. Adachi, "Cellular structuring and clustering for distributed antenna systems," in *Proc. 24th Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Okayama, Japan, Dec. 2021, pp. 1–6.

[4] M. Simsek, M. Bennis, and I. Güvenç, "Learning based frequency- and time-domain inter-cell interference coordination in HetNets," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4589–4602, Oct. 2015.

[5] Y. Wang, G. Feng, Y. Sun, S. Qin, and Y.-C. Liang, "Decentralized learning based indoor interference mitigation for 5G-and-beyond systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12124–12135, Oct. 2020.

[6] M. Yan, J. Yang, K. Chen, Y. Sun, and G. Feng, "Self-imitation learning-based inter-cell interference coordination in autonomous HetNets," *IEEE Trans. Netw. Service Manage.*, vol. 18, no. 4, pp. 4589–4601, Dec. 2021.

[7] 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," *IEEE Trans. Veh. Technol.*, early access, Nov. 4, 2022, doi: 10.1109/TVT.2022.3219411.

[8] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021, doi: 10.1109/OJCOMS.2021.3057679.

[9] 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," *IEEE Commun. Lett.*, vol. 26, no. 6, pp. 1303–1307, Jun. 2022, doi: 10.1109/LCOMM.2022.3159742.

[10] J. Xu, P. Zhu, J. Li, and X. You, "Deep learning-based pilot design for multi-user distributed massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1016–1019, Aug. 2019, doi: 10.1109/LWC.2019.2904229.

[11] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb. 2015, doi: 10.1038/nature14236.

[12] *Inter-Cell Interference Handling for E-UTRA*, document 3GPP R1-050764, Ericsson, Aug. 2005.