DQN and Heuristic Precoding-aware Scheduling in Cell-free mMIMO Networks

Adam Girycki*†, Md Arifur Rahman*, and Sofie Pollin†
* IS-Wireless, Pulawska Plaza, ul. Pulawska 45b, 05-500 Piaseczno, Poland
† Department of Electrical Engineering, KU Leuven, Belgium
a.girycki@is-wireless.com; a.rahman@is-wireless.com; sofie.pollin@kuleuven.be

***Abstract*—We propose a deep Q-Learning (DQN) and heuristic-based radio resource scheduling (RRS) algorithms, which determine serving access points (APs), for maximum ratio transmission (MRT) precoder in the cell-free massive multiple-input multiple-output (CF mMIMO) networks. We show that, DQN-based algorithm yields up to 6% higher sum spectral efficiency (SE) than the heuristic and the reinforcement learning based algorithms for MRT precoder, but the heuristic algorithm yields 1000 times shorter computation time than the DQN-based algorithm. Thus, the low complexity MRT heuristic algorithms are the trade off between the performance and the cost.**

***Index Terms*—Spectral efficiency, radio resource scheduling algorithm, deep Q-Learning, and computational time.**

I. Introduction

The multiple-input-multiple-output (MIMO) technique appears to be one of the most effective disruptive technologies in recent wireless (mobile) network [1] resulting in a significant spectral efficiency improvement. Similarly, the concept of multi-user MIMO provides high capacity by utilizing the benefits of space-division multiple access [2]. To further enhance the SE, the massive MIMO solution has been proposed in [3] where a number of antennas are simultaneously serving several user equipments (UEs). In MIMO systems, the large number of antennas can be co-located in a single array or distributed geographically in a cell [4]. The cell-free mMIMO is a network paradigm shift to an architecture where a number of UEs in a geographic area is served by multiple access points (APs) [5], [6]. Recently, the CF mMIMO system has received a lot of interest due to its high SE [7]. Therefore, in this work, we consider the CF mMIMO network concept and address a DQN-based precoding-aware resource scheduling method to improve the downlink (DL) sum spectral efficiency performance of the network.

Several studies have already been carried out from various aspects such as precoding optimization [8], [10], power allocation [11]–[14], and clustering [15] to enhance the performance of the CF mMIMO networks. In particular, the authors in [8] revealed the proposed scheme to optimize the precoding vector at the APs for both centralized and decentralized fashions. It has been shown, that centralized precoding, i.e., computing the precoding vector in the central processor unit (CPU) instead of the AP, can effectively improve the performance over decentralized design with reasonably low complexity as there is more information available in the centralised CPU to determine the optimal precoding vector. Mapping of the AP and CPU to the O-RAN terminology has been done in [9], where the AP has been identified as the O-RU and the CPU as the O-DU. To reduce the complexity of the precoding vector computation of user-centric CF mMIMO, the authors in [10] collect partial channel state information from the APs at the CPU. The power allocation (PA) and QoS-aware resource scheduling problems are addressed in [11], [12], where a distributed deep neural network and deep reinforcement learning algorithms are used to maximize SE per user and system throughput of CF mMIMO networks. Similarly, a deep Q-learning-based DL power transmission method for CF mMIMO is proposed in [13] to maximize the network SE. Also, a deep neural network (DNN) based unsupervised learning method for PA is studied by Rajapaksha *et.al* in [14] to maximize the minimum rate of the user in CF mMIMO networks.

A. Contributions and Organization of the paper

In summary, existing works on ML algorithms for CF mMIMO networks have generally studied precoding, PA, and clustering problems either separately or jointly, to maximize the system SE, per user SE, system throughput, a minimum rate of the users [11]–[15]. The low-complexity precoding aware radio resource scheduling scheme to maximize the sum SE is still an open problem for CF mMIMO networks. Therefore, in this work:

- Firstly, we propose a DQN-based states crossover accelerated PRB allocation algorithm for CF mMIMO. Inherited from the genetic algorithm, a random crossover of good states rather then a random state is used for DQN start point, yielding 100 times shorter computation time than the DQN alone.
- Secondly, we propose a low-complexity heuristic PRB scheduling algorithm for MRT precoder, which yields 3 ranges of magnitude shorter computation time than the DQN with crossover.
- Finally, we illustrate that, depending on the noise level, MRT precoder results in 10-100 times lower fronthaul load compared to zero forcing (ZF) and optimized ZF (OZF) precoders, which makes MRT an attractive solution even though it yields lower sum SE compared to the ZF and OZF precoders.

In section II of the paper, we discuss the system model and formulate the optimisation problem. In section III, the proposed DQN and heuristic-based PRB scheduling algorithms are described. In section IV, we present and discuss the simulation results and we conclude in section V.

II. System model and problem formulation

A. System model

Let us consider a CF mMIMO network with N APs jointly transmitting to single-antenna UEs. The received signal y_{ub} at UE u on physical resource block (PRB) b consisting of 12 consecutive subcarriers, can be written as:

$$y_{ub} = \sum_{n}^{N} \sum_{u'}^{U} h_{unb} w_{nu'b} s_{u'b} + n_{ub}, \tag{1}$$

where h_{unb} is the complex-valued channel (propagation) coefficient between UE u and transmit AP n in PRB b, w_{nub} denotes the complex precoding value that AP n applies to transmit the symbol s_{ub} on PRB b and n_{ub} is the noise at UE u in a PRB b. The UE received signal-to-interference and noise (SINR) in PRB b is the total received power divided by the total interference power in the PRB, where interference is caused by transmission to other users u' :

$$\text{SINR}_{ub} = \frac{\left| \sum_{n}^{N} h_{unb} w_{nub} \right|^2}{\left| \sum_{n}^{N} \sum_{u' \neq u}^{U} h_{unb} w_{nu'b} \right|^2 + |n_{ub}|^2}. \tag{2}$$

The system achievable downlink SE, referred to as the *sum SE*, denoted by SE_{DL}, can be expressed as:

$$\text{SE}_{\text{DL}} = \frac{1}{B} \sum_{b}^{B} \sum_{u}^{U} \log_2(1 + \text{SINR}_{ub}), \tag{3}$$

where B is the number of PRBs in the carrier bandwidth. The notation used in this paper is listed in Table I.

B. Problem formulation

The target of this work is to maximize the sum SE by optimizing the precoders w_{nub} with AP power constrain:

$$\max \text{SE}_{\text{DL}} \tag{4}$$

$$\sum_{u}^{U} |w_{nub}|^2 \leq P_{\text{max}}, \tag{5}$$

where P_{max} is the maximum AP transmit power per PRB.

C. Search space of the problem

Let q be the quantizied precoder search space, e.g., q = 2^{16} for precoder coded with 16 bits. Considering that an AP serves only one UE per PRB, then the search space per PRB of the 2-dimensional AP-UE optimisation problem is $\omega = (qU)^N$ and the overall search space of the 3-dimensional AP-UE-PRB optimisation problem is $\Omega = \omega^B$. For $q = 2^{16}$, U = 10, N = 64 and B = 100, the search space $\omega = (2^{16} * 10)^{64} \approx 1.8 * 10^{372}$ and $\Omega \sim 10^{37225}$. The search space will be greater if an AP can serve simultaneously more than one UE per PRB.

TABLE I: Notation

Symbols	Meaning
N	Number of APs
Index n	AP index
U	Number of UEs
Index u	UE index
B	Number of PRBs in the carrier bandwidth
Index b	PRB index
s	Transmitted symbol
y	Received signal
h	Channel
w	Precoding
SINR	Signal-to-interference and noise ration
n	Noise
SE	Spectral efficiency
P	AP transmit power per PRB
ω	Search space per PRB
Ω	Overall search space
c	RRS algorithm complexity per PRB
C	Overall complexity of the RRS algorithm
S	State
a	Action
r	Reward
Q	Q-function, i.e., cumulative future reward
γ	Discount factor
k	Signal to jamming ratio

III. Precoding-aware RRS

A significant contributor to the huge overall search space is the quantizied precoder search space, q, therefore, we use the *precoding aware* RRS approach where the precoder is determined based on a well known precoder type, e.g., MRT or ZF, but the APs serving each UE are derived by scheduling algorithm optimised for the precoder type. Thus, the overall radio resource allocation is divided into two steps:

- **Precoding-aware RRS**, where APs serving UEs are selected independently for each PRB knowing the precoder type. Not selected AP-UE-PRB resources will have the precoder $w_{nub} = 0$ indicating no transmission from the AP n to the UE u in the PRB b. We propose two RRS algorithms:
  - The DQN-based for MRT precoding, which uses DQN to explore relations between APs.
  - The heuristic algorithm for MRT precoding, which selects APs following a predefined search path.
- **Precoding**, where the exact value of the precoder w_{nub} is determined following the MRT, ZF or OZF scheme.

A. Assumptions reducing algorithm complexity

We take an advantage of the interference independence between PRBs and carry out RRS independently for each PRB. Furthermore, we reduce the algorithms search space allowing an AP to serve one UE per PRB. The search space, where an AP can simultaneously serve two UEs yields 3.7% higher sum SE compared to the search space where an AP can serve one UE [16], however, this sum SE improvement is at the cost of several ranges of magnitude greater search space, while we search for *efficient practically applicable algorithms*.

Therefore, we consider the following two search spaces where an AP can serve one UE:

- *Search space 1*, where an AP can select to serve the UE with the highest power on that PRB.
- *Search space 2*, where an AP can serve either the highest or second-highest power UE on that PRB.

In the case of the search space 1, the number of possible schedule outputs per PRB is $\omega_1 = 2^N$ and in the case of the search space 2, $\omega_2 = 3^N$. With 64 APs, $\omega_1 = 1.84 \times 10^{19}$ and $\omega_2 = 3.4 \times 10^{30}$. Because we carry out scheduling for each PRB independently, therefore, the overall search space, Ω, is derived as $\Omega = B\omega$. In the following sections we propose DQN and heuristic RRS algorithms for the two search spaces.

B. The proposed DQN-based RRS for MRT precoder

The DQN is the deep learning based reinforcement learning (RL) algorithm. It explores the environment and stores information about the predicted cumulative future reward of each action in a given state in a Q-function that is represented by a neural network. The state, the action and the reward definitions are the essential aspects of a RL algorithm and we define them as follows:

- The **state** indicates the AP-UE association. If an AP serves the UE then the AP-UE associations is denoted by $+1$ (or '+') otherwise it is denoted by -1 (or '$-$'). The DQN algorithm search space 1 for MRT precoder is denoted as MRT-DQN1 and its state is coded by length-N binary sequence, as illustrated in Fig. 1. The algorithm for search space 2 is denoted as MRT-DQN2 and its state is coded by length-$2N$ binary sequence.
- The **action** is a change of the status of one AP. The action can be an addition action, when the AP-UE association becomes serving, as shown in Fig. 1, or a removal action, when a serving AP stops serving the UE. The number of possible actions is equal to the number of states.
- The **reward** is the sum SE change and it can be positive, if an action improves sum SE, or negative, if an action yields decreased sum SE.

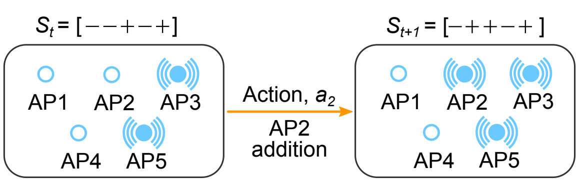

Fig. 1: An illustration of the RL state and action for the search space 1, i.e., when an AP can serve the strongest power UE only, in an example network consisting of 5 APs. In the state $S_t = [- - + - +]$, AP3 and AP5 are serving, thus, denoted by '+'. Not serving APs are denoted by '$-$'. The action a_2 adds the AP2 to the serving APs and transitions the network to the next state $S_{t+1} = [- + + - +]$.

For the DQN, we use a fully connected 3-layer neural network with the input layer encoding states, the output layer encoding Q-values for actions, one hidden layer with the same number of nodes as the input and output layers, and sigmoid activation function. Let $Q(S_t, a_t)$ be the output of the neural network for a given state s_t and action a_t, where the output is the *predicted* cumulative future reward of the action a_t in state s_t. In the learning phase, the DQN weights and biases are updated in order to minimize the $Loss$ defined as the square difference between the *target* and the *prediction* value, where the target value is calculated as a sum of the reward r_t of the action a_t and the cumulative future reward of the best action in the state S_{t+1} multiplied by the discount factor $\gamma = 0.1$:

$$Loss = \left[\overbrace{r_t + \gamma \max_{a_{t+1}} Q(S_{t+1}, a_{t+1})}^{\text{target}} - \overbrace{Q(S_t, a_t)}^{\text{prediction}} \right]^2. \quad (6)$$

When channels h_{unb} change, a new single-use on-line search for optimized RRS is necessary. We use the DQN to find the optimized RRS, rather than to learn the DQN. We accelerate the DQN search by starting a new search batch from a crossover of good states, where AP states from two different states are selected randomly. Furthermore, we transfer the optimized RRS from MRT-DQN1 to MRT-DQN2 for further optimization. During the search, we explore the search space separately for each PRB running 3000 or 5000 batches per PRB for MRT-DQN1 and MRT-DQN2, respectively. Each batch consists of N or $2N$ sequential actions for MRT-DQN1 and MRT-DQN2, respectively. Thus, for $N = 64$, the algorithms complexity per PRB can be estimated as $c_1 = 1.92 \times 10^5$ and $c_2 = 8.32 \times 10^5$ actions for MRT-DQN1 and MRT-DQN2, respectively. The complexity of the algorithm compared with the corresponding search space indicates the algorithm efficiency.

Algorithm 1 The heuristic resource scheduling algorithm for MRT precoder

Start;
For each AP n in the network calculate the metric $k_n = \text{abs}(h_{nu_{n,1}})^2/\text{abs}(h_{nu_{n,2}})^2$, defined as the ratio between the received power by the strongest power UE, $u_{n,1}$, and the received power by the second strongest power UE, $u_{n,2}$, in the AP.
Sort the APs in the descending order of k_n. Denote the sorted APs as $m_1, .., m_N$, the strongest power UEs in the sorted AP as $u_{m_1,1}, ..., u_{m_N,1}$ and the second strongest power UEs in the sorted AP as $u_{m_1,2}, ..., u_{m_N,2}$.
For each AP m_i on the $m_1, .., m_N$ list, select an UE $u_{m_i,1}$ or $u_{m_i,2}$, whichever provides higher sum SE improvement.
Return the list of selected AP-UE associations.
End;

C. The proposed heuristic RRS for MRT precoder

In addition, we propose the heuristic MRT precoder-aware RRS algorithm for the *search space 2* denoted as MRT-H2 and described in Algorithm 1, as an extension of the MRT-H1 denoted in [16] as 'MRT with heuristic'. The MRT-H2 algorithm runs independently for each PRB and its key aspect is the sequence order in which the AP-UE associations are considered for selection. The APs with highest signal to jamming ratio k, defined as the power received by the strongest power UE to the power received by the second strongest power UE in the AP in the PRB, are considered first. This ensures that APs are selected for the transmission in an order which causes little interference for other UEs. With this algorithm, all relations between APs are not searched. Only signal to jamming dependencies between already selected APs and the AP being considered for selection are checked.

The MRT-H2 algorithm calculates the sum SE twice per AP, thus, the algorithm complexity per PRB is $c_2 = 128$ for $N = 64$, which is three ranges of magnitude less than the MRT-DQN2 algorithm. The comparison of the algorithm complexity per PRB, $c_2 = 128$, with the corresponding search space per PRB, $\omega_2 = 3.4 \times 10^{30}$, shows that the heuristic scheduling algorithm for MRT precoder attempts to solve a very complex problem in an extremely simple way.

IV. Simulation Results and Discussions

The simulated network of size 320 m $\times$ 277 m consists of 64 APs distributed in 8 rows and spaced by 40 m. Each AP is equipped with 2.15 dBi gain omni-directional antenna at 10 m above the ground. The 4 GHz carrier with 10 PRBs and 15 kHz carrier spacing transmits maximum 20 mW per PRB. In 100 simulation iterations 10 stationary single antenna UEs have been randomly distributed in the network and the channel has been modeled according to [16].

A. Convergence of the DQN algorithm

Fig. 2 shows the MRT-DQN1 algorithm convergence at 57.9 kbit/s/Hz during 3000 batches search and Fig. 3 shows, that after this short learning the neural network does not reach convergence yet. Fig. 4 shows 300000 batches search after which the DQN provides the same maximum sum SE of 57.9 kbit/s/Hz as during short search. However, after long learning, the DQN provides repetitive converged search attempts regardless of the starting state, as presented in Fig. 5. The example shows, that long training leads to the neural network convergence, but the neural network convergence is not required for the purpose of single-use finding an optimized state if DQN is combined with states crossover. The DQN combined with stages crossover accelerates the search 100 times.

B. Reference algorithms

Table II lists the proposed MRT-DQN1, MRT-DQN2 and MRT-H2 RRS algorithms together with reference CF SISO and CF mMIMO algorithms, the latter utilizing MRT, ZF and OZF precoders. In the CF SISO, in each PRB, the UE is served by a single AP, the one which provides the strongest DL signal power in the PRB. Because the CF SISO does not rely on CF mMIMO precoding therefore it can be consider as a reference for the CF mMIMO algorithms. We also compare the proposed algorithms with the MRT reinforcement learning based, MRT-RL1, and MRT heuristic of search space 1, MRT-H1, in [16].

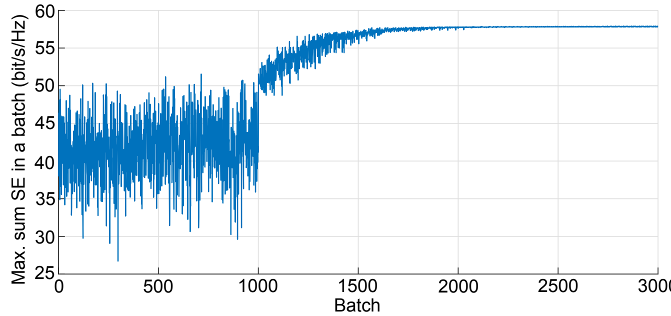

Fig. 2: DQN convergence during short learning. The first 1000 batches start from a random state and the remaining 2000 batches start from a crossover of two out of 40 best states.

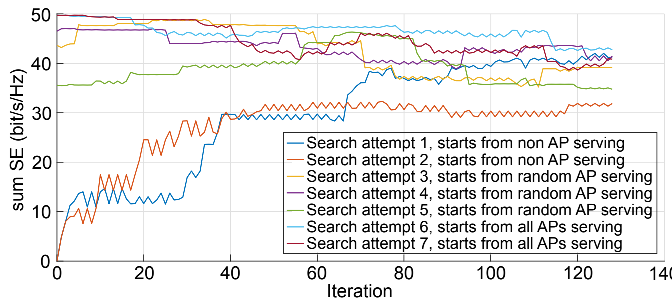

Fig. 3: DQN performance during exploitation mode after learning in Fig. 2. The search attempts do not converge indicating that the DQN was not trained.

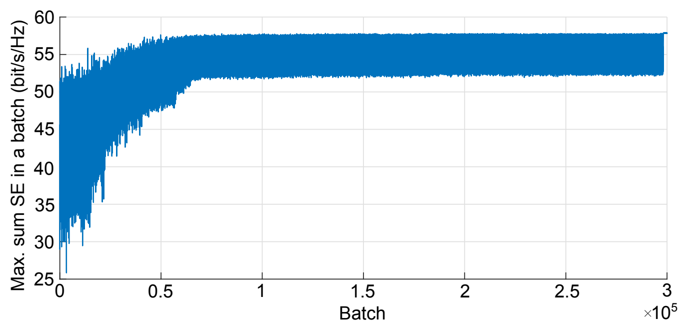

Fig. 4: DQN convergence during long learning. The first 298000 batches start from a random state and the last 2000 batches start from a crossover of two out of 40 best states.

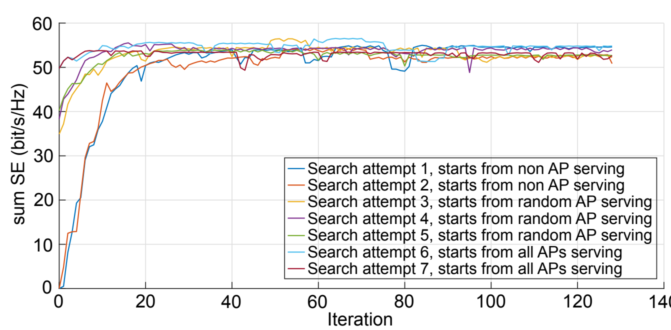

Fig. 5: DQN exploitation mode convergence after learning in Fig. 4. The search attempts converge indicating that the DQN was trained.

TABLE II: Radio resource allocation algorithms

Algorithm	Scheduling[1]	Precoding	Time[2] (s)
CF SISO	Best AP-PRB	–	0.0002
MRT-DQN2	DQN (2)	MRT	$15 + 85$
MRT-DQN1	DQN (1)	MRT	15
MRT-RL1	RL (1)	MRT	1.0
MRT-H2	Heuristic (2)	MRT	0.004
MRT-H1	Heuristic (1)	MRT	0.0016
ZF-all	All APs	ZF	0.002
ZF-RL	RL	ZF	3.6
OZF-all	All APs	OZF	0.063

[1]In the bracket, the searched number of best power UEs per AP per PRB.
[2]The computation time measured in Matlab R2016b on the MacBook Air M1 2020 with macOS 13.4 for one PRB, 64 APs and 10 UEs.

In ZF-all and OZF-all, all APs serve all UEs. In ZF-RL [16], the RL algorithm determines the APs which serve all UEs.

C. DQN and heuristic algorithms computation time

Table II also shows the computation time of the algorithms. The MRT-DQN1 computation time is 3 ranges of magnitude greater than the MRT-H1, which searches the same area. Also the MRT-DQN2 computation time is 3 ranges of magnitude greater than the MRT-H2, which searches the same area.

D. DQN and heuristic algorithms sum SE

It can be seen in Fig. 6, that at each noise power, the CF mMIMO RRS provides significantly higher sum SE compared to the CF SISO scheduler. At low noise of -90 dBm per PRB, the CF mMIMO with ZF or OZF precoder yield the sum SE of 130 bit/s/Hz, which is 2.7 times higher compared to 48.4 bit/s/Hz sum SE of the CF SISO scheduler. MRT-DQN2 yields the sum SE of 74.5 bit/s/Hz and performs 6% better than its corresponding heuristic algorithm MRT-H2, which yields the sum SE of 70.1 bit/s/Hz. The DQN is able to search for dependencies between the APs, while the heuristic algorithm only checks if the AP considered for selection provides sum SE improvement when serving together with the already selected APs. Interestingly, the MRT-DQN1 performs only 1% worse than the MRT-DQN2 indicating that the search space 1, where the AP is allowed to serve only the highest power UE on the PRB, provides satisfactory sum SE performance. At high noise of -50 dBm per PRB, all MRT algorithms show similar sum SE performance and they outperform the CF SISO scheduler providing 2.4 times higher sum SE, since the MRT-DQN2 yield 22 bit/s/Hz and CF SISO 9.3 bit/s/Hz.

E. DQN and heuristic algorithms power utilization

Fig. 7 shows the power utilization defined as the ratio of the mean AP transmit power, resulted from the applied RRS algorithm, to the maximum AP power. For example, if each UE is served by one AP transmitting on its maximum power, then U out of N APs transmit with full power and the power utilization is $\frac{U}{N} = \frac{10}{64} = 0.156$. For CF SISO, the power utilization is 0.142, which is less than 0.156, because more than one UE may be served by an AP, thus, less than 10

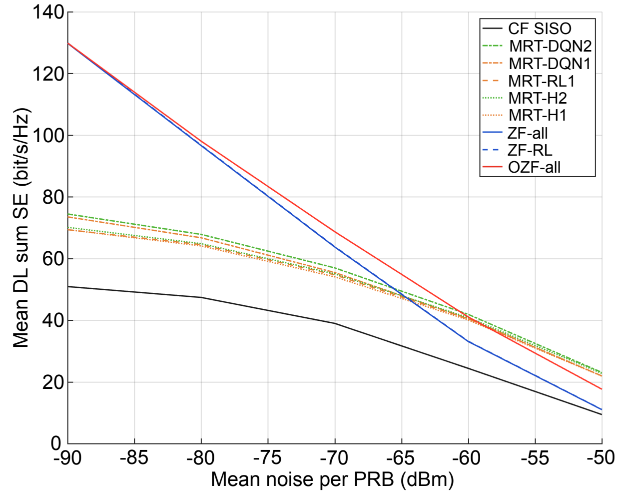

Fig. 6: Mean sum SE performance of the radio resource allocation algorithms.

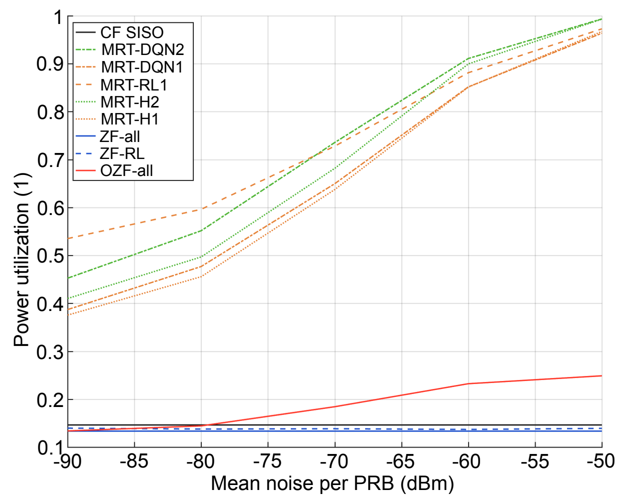

Fig. 7: APs power utilization for the radio resource allocation algorithms.

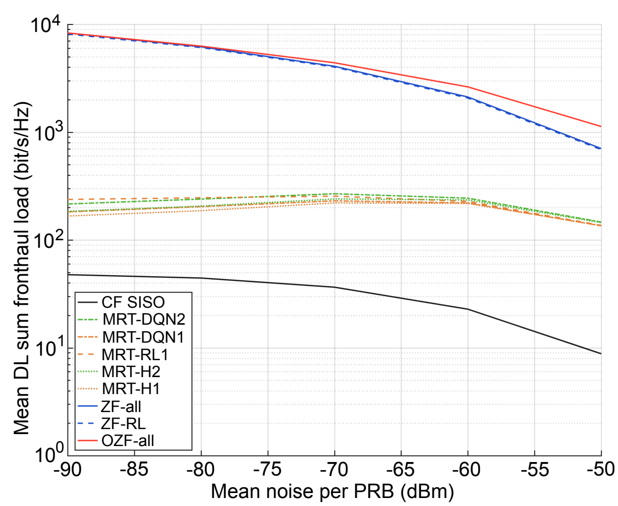

Fig. 8: Mean sum fronthaul load for radio resource allocation algorithms.

APs transmit. The RRS algorithms with MRT precoder choose more than one serving AP for each UE, therefore, the power utilization is significantly higher compared to the CF SISO. For the MRT precoder, the mean number of serving APs per UE grows with the noise level, thus also the power utilization grows. At -90 dBm noise, the MRT-DQN2 uses in average 2.90 AP/UE resulting in power utilization of 0.45, but, at the noise level of -50 dBm, 6.35 AP/UE results in the power utilization of 0.99, i.e., almost all APs transmit. The MRT-DQN1 selects less APs compared to the MRT-DQN2 and the heuristic algorithms tent to selct less serving APs than the DQN-based algorithms. The ZF-all yields the lowest power utilization of 0.13 across all noise levels, since the ZF does not use the full AP power. At low noise, the OZF precoder yields similar power utilization to the ZF precoder, but increases the transmit power at higher noise.

F. DQN and heuristic algorithms fronthaul load

Fig. 8, shows the mean sum fronthaul load defined as the mean sum SE multiplied by the mean number of APs serving the UE, which is applicable when the beamforming is done at the RU as in the O-RAN 7-2 split, and fronthaul modulation compression is used. For -90 dBm noise level, the MRT-DQN2 yields the sum SE of 74.5 bit/s/Hz with mean number of 2.90 APs serving an UE, yielding the mean sum fronthaul load of $74.46 \text{ bit/s/Hz} \times 2.9 = 216 \text{ bit/s/Hz}$. With the CF SISO an UE is served by one AP per PRB thus the mean sum fronthaul load is equal to the sum SE. With the ZF-all and OZF-all each UE is served by all APs, thus, the mean sum fronthaul load is N times greater than the sum SE, i.e., 8312 bit/s/Hz for -90 dBm noise. At this noise, ZF-RL selects 62.5 serving APs yielding the same sum SE as the ZF-all but with 2.3% lower fronthaul load. Thus, scheduling can reduce fronthaul load. Fig. 8 also shows that the 7-2 split fronthaul load for MRT RRS algorithms does not change much across different noise levels because, with higher noise, the mean number of serving AP increases, but the sum SE decreases.

V. Conclusions

We show that the MRT precoded transmission, even though yields lower sum SE, is an interesting solution due to 10-100 times lower 7-2 split O-RAN fronthaul load compare to ZF and OZF precoders. The DQN-based algorithm, when used for optimized radio resource scheduling search, can be accelerated 100 times when combined with states crossover, however, the crossover does not reduce the DQN learning time. The DQN-based algorithm for the MRT precoder, at low interference of -90 dBm per PRB, yields 6% higher sum SE compared to the heuristic algorithm, but at the high noise of -60 dBm both algorithms yield similar sum SE. Due to 3 ranges of magnitude lower computational time of the heuristic algorithm compared to the DQN algorithm, the heuristic algorithm seems to be an attractive real deployment solution. We also investigate the transmit power of the proposed algorithms and observe that the ZF precoded transmissions dissipate less power than the MRT precoded. As future work we will seek for DQN-based scheduling algorithm which allows for off-line learning.

Acknowledgment

6G-BRICKS project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grand Agreement No 101096954.

References

[1] A. J. Paulraj, D. A. Gore, R. U. Nabar and H. Bolcskei, "An overview of MIMO communications-a key to gigabit wireless," *Proceedings of the IEEE*, vol. 92, pp. 198-218, Feb. 2004.

[2] P. Krishna, T. A. Kumar and K. K. Rao, "Multiuser MIMO systems: Spectral and energy efficiencies, estimations and capacity limits," in *Proc. 2015 Twelfth International Conference on Wireless and Optical Communications Networks (WOCN)*, Bangalore, India, 2015, pp. 1-6, doi: 10.1109/WOCN.2015.8064514.

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

[4] J. Xu, P. Zhu, J. Li, X. Wang and X. You, "Secrecy Energy Efficiency Optimization for Multi-User Distributed Massive MIMO Systems," *IEEE Trans. on Commun.*, vol. 68, no. 2, pp. 915-929, Feb. 2020.

[5] J. Zhang, L. Hanzo, "Cell-Free Massive MIMO: A New Next-Generation Paradigm," *IEEE Access*, 2019.

[6] G. Interdonato, E. Björnson, "Ubiquitous cell-free Massive MIMO communications," *EURASIP Journal on Wireless Communications and Networking*, 2019:197.

[7] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau and K. V. Srinivas, "User-Centric Cell-Free Massive MIMO Networks: A Survey of Opportunities, Challenges and Solutions," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 611-652.

[8] E. Shi, J. Zhang, J. Zhang, D. W. K. Ng and B. Ai, "Decentralized Coordinated Precoding Design in Cell-Free Massive MIMO Systems for URLLC," *IEEE Transactions on Vehicular Technology*, 2022, doi: 10.1109/TVT.2022.3210253.

[9] V. Ranjbar, A. Girycki, M. A. Rahman, S. Pollin, M. Moonen and E. Vinogradov, "Cell-Free mMIMO Support in the O-RAN Architecture: A PHY Layer Perspective for 5G and Beyond Networks," *IEEE Communications Standards Magazine* vol. 6, no. 1, pp. 28-34, March 2022, doi: 10.1109/MCOMSTD.0001.2100067.

[10] X. Gao, Y. Li, W. Cheng, L. Dong and P. Liu, "Secure Optimal Precoding for User-Centric Cell-Free Massive MIMO System," *IEEE Wireless Communications Letters*, vol. 12, no. 1, pp. 31-35, Jan. 2023.

[11] M. Zaher, Ö. T. Demir, E. Björnson and M. Petrova, "Learning-Based Downlink Power Allocation in Cell-Free Massive MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 174-188, Jan. 2023.

[12] Huang, Chih-Wei, et al., "Joint QoS-Aware Scheduling and Precoding for Massive MIMO Systems via Deep Reinforcement Learning," arXiv preprint arXiv:2104.04492 (2021).

[13] Y. Zhao, I. G. Niemegeers and S. H. De Groot, "Deep Q-network based dynamic power allocation for cell-free massive MIMO," in *Proc. 2021 IEEE 26th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Porto, Portugal, 2021, pp. 1-7, doi: 10.1109/CAMAD52502.2021.9617768.

[14] N. Rajapaksha, K. B. Shashika Manosha, N. Rajatheva and M. Latva-Aho, "Deep Learning-based Power Control for Cell-Free Massive MIMO Networks," in *Proc. ICC 2021 - IEEE International Conference on Communications*, Montreal, Canada, 2021, pp. 1-7.

[15] N. Athreya, V. Raj and S. Kalyani, "Beyond 5G: Leveraging Cell Free TDD Massive MIMO Using Cascaded Deep Learning," *IEEE Wireless Communications Letters*, vol. 9, no. 9, pp. 1533-1537, Sept. 2020

[16] A. Girycki, M. A. Rahman, E. Vinogradov and S. Pollin, "Learning-based Precoding-aware Radio Resource Scheduling for Cell-free mMIMO Networks," in IEEE Transactions on Wireless Communications, doi: 10.1109/TWC.2023.3323152.