IEICE TRANS. COMMUN., VOL.E106–B, NO.7 JULY 2023

PAPER

UE Set Selection for RR Scheduling in Distributed Antenna Transmission with Reinforcement Learning

Go OTSURU[†a)], ***Member and*** **Yukitoshi SANADA**[†b)], ***Fellow***

SUMMARY In this paper, user set selection in the allocation sequences of round-robin (RR) scheduling for distributed antenna transmission with block diagonalization (BD) pre-coding is proposed. In prior research, the initial phase selection of user equipment allocation sequences in RR scheduling has been investigated. The performance of the proposed RR scheduling is inferior to that of proportional fair (PF) scheduling under severe intra-cell interference. In this paper, the multi-input multi-output technology with BD pre-coding is applied. Furthermore, the user equipment (UE) sets in the allocation sequences are eliminated with reinforcement learning. After the modification of a RR allocation sequence, no estimated throughput calculation for UE set selection is required. Numerical results obtained through computer simulation show that the maximum selection, one of the criteria for initial phase selection, outperforms the weighted PF scheduling in a restricted realm in terms of the computational complexity, fairness, and throughput.

key words: *distributed antenna transmission, round-robin scheduling, resource allocation, reinforcement learning*

1. Introduction

Recently, many Internet of Things (IoT) applications have been launched and the amount of mobile traffic has increased explosively [1]. Distributed antenna transmission (DAT) has been studied as one form of fifth-generation (5G) mobile communication deployment and can resolve the problem of larger path loss in high-frequency bands. Radio-resource scheduling in DAT among multiple transmission points (TPs) with lower computational complexity is a challenge to solve under a trade-off relationship between system throughput and fairness among users [2].

To achieve higher system throughput and mitigate co-channel interference (CCI), especially for cell-edge users, cooperative DAT (CDAT) with multi-user spatial multiplexing has been proposed [3]–[7]. The combination of CDAT, user equipment (UE) classification, UE clustering, and cluster-antenna association has been investigated as an evolved CDAT in [4]. Fractional frequency reuse (FFR) has been adopted in [4] to mitigate the inter-cell interference of cell edge UEs. In [5], Max C/I scheduling, proportional fair (PF) scheduling, and round-robin (RR) scheduling in CDAT are compared. In spite of the low complexity of RR scheduling, the system throughput and the fairness are close to those of PF scheduling. On the other hand, coordinated radio-resource scheduling with a global scheduler has been introduced in [6], [7]. The global scheduler computes a PF metric for each combination of UEs while local schedulers determine the association between TPs and UEs.

However, the above mentioned research studies for UE allocation apply no scheduling criterion for RR scheduling in DAT. The authors have investigated RR scheduling in DAT and proposed the initial phase selection of UE allocation sequences [8], [9]. The problem of the scheme in [8], [9] is that no fairness among UEs was taken into account even through the system throughput was improved. It is necessary that the system throughput is increased and user fairness is maintained. The authors have introduced the user set elimination in allocation sequences of RR scheduling [10]. However, the proposed scheme in [10] does not include multi-user multi-input multi-output (MIMO) technology in order to mitigate the CCI. Therefore, the fairness of the proposed RR scheduling is inferior to that of PF scheduling.

The research on resource allocation with reinforcement learning has been flourishing because it can be applied to a system that is difficult to model. For example, in [11], to utilize the limited backhaul capacity of millimeter-wave communication, the blockage patterns of channel states can be captured and predicted with deep reinforcement learning (DRL). In [12], [13], the authors have applied DRL to high mobility transport systems, such as trains or unmanned aerial vehicles (UAVs), whereby unpredicted and fluctuating links are generated. They support the time division duplex (TDD) configuration in real time and adaptively change the TDD uplink/downlink ratio. As the integrated access and the backhaul architecture can be huge and time-varying, DRL has been introduced [14] to the problem in which the optimal solution that maximizes the sum rate of all UEs is intractable to find. In [15], the aggregate network capacity employing beamforming and non-orthogonal multiple access (NOMA) have been maximized by utilizing three reinforcement learning (RL) methods. The RL on resource allocation is promising and has the possibility of spectrum efficiency improvement. However, to the best of our knowledge, the RL has not been adapted to the RR scheduling in DAT. In this paper, RR scheduling with RL is proposed to realize the efficient UE allocation sequence of the RR scheduling for DAT with a block diagonalization (BD) algorithm [16]. The proposed RR scheduling is compared to the weighted PF scheduling in terms of computational complexity, fairness, and system throughput [17].

Manuscript received August 21, 2022.
Manuscript revised November 6, 2022.
Manuscript publicized January 13, 2023.

[†]The authors are with the Dept. of Electronics and Electrical Engineering, Keio University, Yokohama-shi, 223-8522 Japan.
a) E-mail: gootsuru@snd.elec.keio.ac.jp
b) E-mail: sanada@elec.keio.ac.jp

DOI: 10.1587/transcom.2022EBP3136

This paper is organized as follows. Section 2 describes a system model and scheduling schemes. Section 3 explains simulation conditions. Numerical results obtained through computer simulation are then presented. Section 4 gives our conclusions.

2. System Description

2.1 Cell Model

The cell model shown in Fig. 1 is assumed. One macro cell consists of seven hexagonal micro cells. A distributed antenna called a TP is located at the center of each micro cell. The number of TPs in each macro cell is $N_A = 7$. All TPs are controlled by the same central unit (CU). CCI is caused by reusing the same frequency channel in the other macro cells. The colored macro cells in Fig. 1 exchange UE allocation information, as shown in Fig. 2, and the system throughput is evaluated over the seven colored macro cells in this paper. Moreover, radio-resource scheduling for the allocation of UEs over resource blocks (RBs) is adopted for orthogonal frequency division multiplexing (OFDM) signal transmission. Multiple UEs can be assigned to each RB and served by TPs within a macro cell. The number of UEs in a macro cell is N_U and the maximum number of UEs allocated to each RB is N_S.

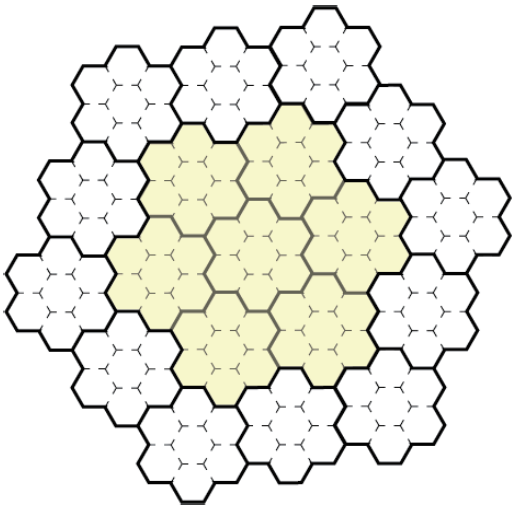


Fig. 1 Cell model.

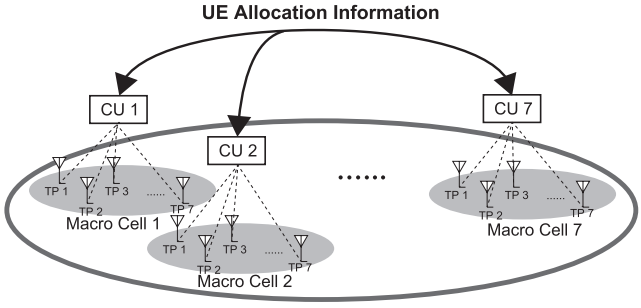


Fig. 2 Overview of cell model.

2.2 Antenna Selection

In the assumed system model, multi-user MIMO with the BD algorithm is introduced to the DAT [16]. As shown in Fig. 3, N_S TPs are selected from N_A TPs and signals for N_S single antenna UEs are spatially multiplexed. The TPs should be selected at the initial stage of radio-resource scheduling so as to maximize the estimated throughput of those UEs. The estimated throughput takes no inter-cell interference into account because which TPs in adjacent macro cells cause CCI to UEs may change all the time. The TP is selected exclusively so that N_S TPs are connected to N_S UEs. Suppose that m^{rc} is the TP set index of the TPs associated in the r-th RB at the c-th macro cell as shown in Fig. 3. The number of TP set indexes is ${}_{N_A}\mathrm{C}_{N_S}$. The signals for the n-th UE in the r-th RB are transmitted only from the TPs of the m^{rc}-th TP. The transmit signals are pre-coded with the BD algorithm so that only the desired signal reaches a specific UE. The transmit signal to the n-th UE on the l-th subcarrier in the r-th RB is represented by x_n^{rl}. The received signal for the n-th UE at the c-th macro cell is given by

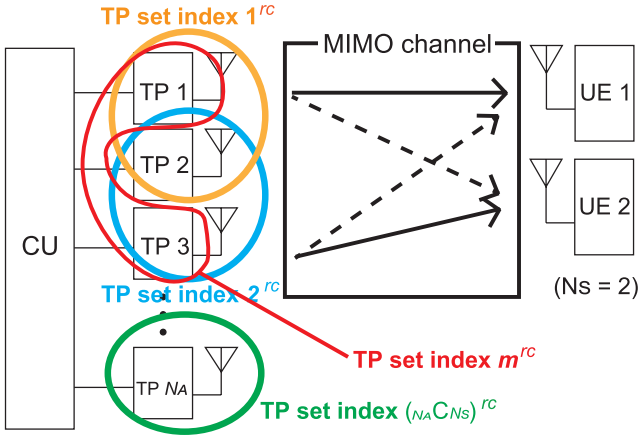


Fig. 3 Multi-user MIMO systems in DAT ($N_S = 2$).

$$y_n^{rlc} = \mathbf{H}_{nm^{rc}}^{rlc}\mathbf{W}_{nm^{rc}}^{rlc}x_n^{rlc} + \sum_{\nu\in\{\mu^{rc}\}}\mathbf{H}_{nm^{rc}}^{rlc}\mathbf{W}_{\nu m^{rc}}^{rlc}x_\nu^{rlc} + z_n^{rlc} \quad (1)$$

where $\mathbf{H}_{nm^{rc}}^{rlc}$ is the channel response vector with a size of $1 \times N_S$ between the TPs of the m^{rc}-th TP set and the n-th UE, $\mathbf{W}_{nm^{rc}}^{rlc}$ is the pre-coding vector with a size of $N_S \times 1$ between the TPs in the m^{rc}-th TP set index and the n-th UE, ν is the index of a UE that causes interference to the n-th UE, z_n^{rlc} is the additive white Gaussian noise (AWGN) with a mean of zero and a variance of σ^2 on the l-th subcarrier in the r-th RB, and $\{\mu^{rc}\}$ is the set of N_S UE indexes allocated to the r-th RB at the c-th macro cell. The number of UE index sets, $\{\mu^r\}$, is ${}_{N_U}\mathrm{C}_{N_S}$.

The tentative throughput for the n-th UE on the l-th subcarrier in the r-th RB at the c-th macro cell is calculated

as

$$\hat{T}_n^{r_l c}(m^{r_l c}) = \log_2\left(1 + \frac{P_{nm^{rc}}^{r_l c}}{\sum_{\nu\in\{\mu^{rc}\}} P_{\nu m^{rc}}^{r_l c} + \sigma^2}\right) \tag{2}$$

where the received signal power is represented as $P_{nm^{r_l c}}^{r_l c} = \left|\mathbf{H}_{nm^{rc}}^{r_l c}\mathbf{W}_{nm^{rc}}^{r_l c}\right|^2$ for the ν-th UE from the TPs of the m^{rc}-th TP set on the l-th subcarrier in the r-th RB at the c-th macro cell. This is the tentative throughput for TP association without taking inter-cell interference into account as it is determined after the association of TPs to UEs in the adjacent cells.

2.3 PF Scheduling

In this paper, one subframe consists of multiple timeslots and one timeslot consists of 14 OFDM symbols. Because of TDD, half of the symbols are allocated to downlink communication.

The total sum of the tentative throughputs to the n-th UE over the subcarriers in the r-th RB in the t-th timeslot at the c-th macro cell, $\hat{T}^{rc}$, is given by

$$\hat{T}^{rc}(n,t) = \sum_{l\in\{l^r\}} \hat{T}_n^{r_l c}(m^{r_l c}). \tag{3}$$

Weighted PF scheduling is applied in this paper as a reference [17]–[19]. In the r-th RB at the c-th macro cell, it calculates the following metric for the set of N_S UE indexes:

$$f^{rc}(\mu^{rc}) = \prod_{n\in\{\mu^r\}}\left(1 + \frac{(\hat{T}^{rc}(n,t))^{\beta-\gamma}}{(C_n(t))^{\gamma}}\right) \tag{4}$$

where $\hat{T}^{rc}(n,t)$ is the tentative throughput derived from Eq. (3) for the n-th UE set over the subcarriers in the r-th RB in the t-th timeslot at the c-th macro cell and β and γ are the weights for the weighted PF scheduling. A larger weight, β, as well as a smaller weight, γ, tends to allocate a RB to UEs with larger tentative throughputs. $C_n(t)$ is the average user throughput for the n-th UE at the t-th timeslot. The interference from outer macro cells is not taken into account because the resource allocation in the macro cells controlled by the same CU is simultaneously conducted in the PF scheduling. The PF scheduling selects the N_S UEs with the largest PF metric. The PF metric is calculated at every subframe. The throughput estimation is conducted $N_{RB}\cdot {}_{N_U}\mathrm{C}_{N_S}$ times at each RB allocation.

2.4 Proposed RR Scheduling

Unlike the tentative throughput of the PF scheduling, the interference from the other macro cells is included in the estimated throughput calculation of the proposed RR scheduling. That is because the resource allocation in the macro cells controlled by the same CU is sequentially conducted with initial phase selection [8]–[10] so that the TPs in the outer macro cells can be identified. The RR scheduling takes no fairness among UEs into account. However, the fairness among UEs may be maintained when the RR scheduling is applied to the DAT [5].

In the proposed RR scheduling, a Q-table is updated in accordance with the current action a_t^r and the UE set with the lowest Q-value is eliminated from the RR sequence. On the other hand, the Q-value for each UE set varies in accordance with the frequency selectivity of a multipath channel. Thus, the fairness is not greatly deteriorated even though it is not counted in the scheduling algorithm directly.

2.4.1 Throughput Estimation

Suppose that the estimated throughput for the n-th UE set on the l-th subcarrier in the r-th RB corresponding to the initial phase δ_c for the c-th macro cell is represented as $\bar{T}_{n_{\delta_c}}^{r_l c}(m^{rc})$ and it is given by

$$\bar{T}_{n_{\delta_c}}^{r_l c}(m^{r_l c}) = \log_2\left(1 + \frac{P_{n_{\delta_c} m^{r_l c}}^{r_l c}}{\sum_{\nu\in\{\mu^{rc}\}} P_{\nu m^{r_l c}}^{r_l c} + \eta_{n_{\delta_c}}^{r_l c\ 2}}\right) \tag{5}$$

where $\eta_{n_{\delta_c}}^{r_l c}$ is the sum of the noise and the interference from the outer macro cells to the n_{δ_c}-th UE on the l-th subcarrier in the r-th RB when the initial phase for the c-th macro cell, δ_c, is selected.

The total sum of the estimated throughput over all the UEs and the subcarriers of the RBs for the c-th macro cell is calculated from Eq. (5) and is given by

$$\bar{T}^c(\delta_c) = \sum_r \sum_{l\in\{l^r\}} \sum_{n_{\delta_c}\in\{\mu_{\delta_c}\}} \bar{T}_{n_{\delta_c}}^{r_l c}(m^{r_l c}). \tag{6}$$

The expected system throughput corresponding to the set of initial phases, $\{\delta_c\}$, over the macro cells is then given by

$$\bar{T}(\delta_1,\cdots,\delta_7) = \sum_{c=1}^{7} \bar{T}^c(\delta_c). \tag{7}$$

2.4.2 Algorithms for Initial Phase Selection

In this paper, two different initial phase selection algorithms based on the estimated throughputs are applied in the initial phase selection [8], [9]. The initial phases in the macro cells controlled by the same CU are sequentially optimized with a maximum selection scheme to reduce the inter-cell interference.

(1) Random Selection

Random selection selects the initial phases in all the macro cells randomly and sequentially. Therefore, no throughput estimation is carried out over the entire RB allocation.

(2) Maximum Selection

Maximum selection selects the initial phases of the UE allocation sequences sequentially over multiple macro cells and it is repeated iteratively. Suppose that t is the time index and $\hat{\delta}_c^{(t)}$ is the candidate of the initial phase selected in the c-th macro cell at the t-th time index, the sum of the tentative

throughputs given by the selected initial phases at the c-th macro cell, $\bar{T}(\hat{\delta}_1^{(t)}, \cdots, \hat{\delta}_{c-1}^{(t)}, \hat{\delta}_c, \hat{\delta}_{c+1}^{(t-1)}, \cdots, \hat{\delta}_7^{(t-1)})$, is calculated from Eq. (7) for all of $\delta_c (0 \le \delta_c \le (1-x)_{N_U}\mathrm{C}_{N_S} - 1)$. The maximum selection selects the phase with the largest estimated throughput. The maximum selection is presented as

$$\hat{\delta}_c^{(t)} = \arg\max_{\hat{\delta}_c} \bar{T}(\hat{\delta}_1^{(t)}, \cdots, \hat{\delta}_{c-1}^{(t)}, \hat{\delta}_c, \hat{\delta}_{c+1}^{(t-1)}, \cdots, \hat{\delta}_7^{(t-1)}). \tag{8}$$

Since this criterion selects the initial phases sequentially, the system throughput may fall into a local optimum. The throughput estimation is conducted $N_{RB} \cdot (1-x) \cdot {}_{N_U}\mathrm{C}_{N_S}$ times at each RB allocation.

2.4.3 Effective UE Set Selection in RR Allocation Sequence with RL

In the proposed scheduling, the RR allocation sequence is modified for each RB because the channel response in each RB differs depending on the frequency selectivity of a fading channel. The estimated throughput for each UE set in each RB is calculated from Eq. (5) with the maximum selection because the initial phases are exhaustively searched in the maximum selection. In the proposed RR scheduling, the RL is applied to the RR allocation sequence in order to eliminate UE sets that suffer from CCI. The CUs select the action a_t^r with the largest Q-value in the r-th RB. In other words, the CUs eliminate the UE sets under severe interference at every timeslot after the initial phase selection.

The RR allocation sequence in the r-th RB at the t-th timeslot is expressed as the state s_t^r. The elimination of the specific UE set in the r-th RB at the t-th timeslot is the possible action of the CU, and it is denoted as a_t^r. The predicted Q-value for the next timeslot in the r-th RB at the t-th timeslot is given as $Q_t(s_{t+1}^r, a_t^r)$. The Q-value of the action a_t^r for the state s_t^r is renewed as

$$\begin{aligned} Q_t(s_t^r, a_t^r) \leftarrow & (1-\alpha) Q_t(s_t^r, a_t^r) \\ & + \alpha[R_{t+1}^r + \varepsilon \max_{a^r} Q_{t+1}(s_{t+1}^r, a^r)] \end{aligned} \tag{9}$$

where α is the learning rate that indicates the impact of the current and past learning, ε is the discount rate, and R_{t+1}^r is the reward value for the transition to the state s_{t+1}^r. R_{t+1}^r is calculated from the estimation throughput averaged over UE sets in the allocation sequence except for the UE set that is eliminated in the action a_t^r.

The possible transition state at the t-th timeslot is shown in Fig. 4. Suppose that the length of the allocation sequence for the initial state s_t^r is L, the CU takes an action a_t^r which is the elimination of a UE set. The next state after the initial action is expressed as s_{t+1}^r, in which the length of an allocation sequence is $L-1$ and the reward for the initial action, R_{t+1}^r, is calculated from the remaining UE sets.

The UE set that results in the smallest system throughput is excluded by the action a_t^r and the reward R_{t+1}^r is derived from the state as s_{t+1}^r, where R_{t+1}^r is the reward for the transition to the state as s_{t+1}^r at the t-th timeslot and corresponds to the average estimation throughput when the UE set that realizes the smallest system throughput is excluded. The CU then takes the action with the largest Q-value.

It is necessary to halt the UE set selection because the excessive elimination of UE sets leads to the deterioration of the fairness among UEs. It is determined from the minimum fairness targeted by the CU.

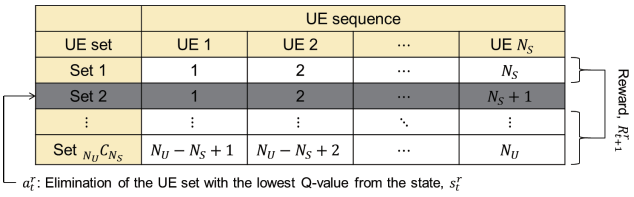


Fig. 4 UE set elimination with RL.

Table 1 Simulation conditions.

Inter-antenna distance	50, 100, 150, 200 m
Minimum distance between UE and TP	5 m
Height of TP	10 m
Height of UE	1.5 m
Carrier frequency	4.65 GHz
System bandwidth	72 MHz
RB bandwidth	720 kHz
No. of RBs	100
No. of subcarriers per RB	12
Transmit power	30 dBm
LOS probability	$P_{LOS} = \min(\frac{18}{d}, 1)\{1 - \exp(-\frac{d}{d_{LOS}})\} + \exp(-\frac{d}{d_{LOS}})$ d: distance from UE to TP
Path loss	$L_{LOS} = 22.0\log_{10}(d) + 28.0 + 20\log_{10}(f_c)$ dB $L_{NLOS} = 36.7\log_{10}(d) + 22.7 + 26\log_{10}(f_c)$ dB f_c: carrier frequency
Shadowing standard deviation	4 dB
Channel model	LOS: Rician path + 15-path uniform Rayleigh (K-factor:10) NLOS: 16-path uniform Rayleigh
Receiver noise density	−174 dB/Hz
Noise figure	9 dB
Allocation	2-user allocation
No. of UEs per macro cell	5, 10
Learning rate	0.1
Weight for PF $\beta - \gamma$	$0 \le \beta - \gamma \le 1.0$
Weight for PF γ	1.0

3. Numerical Results

3.1 Simulation Conditions

The inter-antenna distance is selected from 50, 100, 150, and 200 meters. The height of the TPs is 10 meters and that of

the UEs is 1.5 meters. The carrier frequency is 4.65 GHz, the system bandwidth is 72 MHz, and the RB bandwidth is 720 kHz. The number of RBs is 100 and the number of subacarriers per RB is 12. The transmit power per antenna is set to 30 dBm. The amounts of average propagation loss, L_{LOS} and L_{NLOS}, are different between line-of-sight (LOS) and non-line-of-sight (NLOS) conditions. The LOS probability and path loss models are the same as those in [20], [21]. The shadowing deviation is 4 dB. The first Rician fading path and the following 15 uniform Rayleigh paths are assumed in the LOS condition while a 16-path uniform Rayleigh fading channel is assumed in the NLOS condition. The K-factor in the LOS model is 10. The receiver noise density is set to -174 dB/Hz and the noise figure is 9 dB. Two user allocation ($N_S = 2$) is assumed. The number of UEs per macro cell is 5 or 10 and the uniform user distribution is applied. The average system throughput per subcarrier per cell is evaluated for different phase selection criteria unless it is specified. The number of timeslots in which UE sets are selected with RL is set from 0 to 21 when the number of UEs is 5, while that is set from 0 to 140 when the number of UEs is 10. The weight for PF scheduling, γ, is set to 1.0 while the weight $\beta - \gamma$ are varied from 0.0 to 1.0.

The throughput for the n-th UE on the l-th subcarrier in the r-th RB at the c-th macro cell, $T_n^{r_l c}(m^{r_l c})$, is given by

$$T_n^{r_l c}(m^{r_l c}) = \log_2\left(1 + \frac{P_{nm^{r_l c}}^{r_l c}}{\sum_{\nu\in\{\mu^{rc}\}} P_{\nu m^{r_l c}}^{r_l c} + \eta_n^{r_l c 2}}\right) \tag{10}$$

where $\eta_n^{r_l c}$ is the sum of the noise and the interference from the outer macro cells to the n-th UE on the l-th subcarrier in the r-th RB. The total sum of the throughputs to the n-th UE over the subcarriers in the r-th RB at the c-th macro cell, T^{rc}, is given by

$$T^{rc}(n) = \sum_{l\in\{l^r\}} T_n^{r_l c}(m^{r_l c}). \tag{11}$$

The system throughput over seven macro cells, RBs, and allocated UEs normalized by the number of the macro cells is given as

$$T = \frac{1}{7}\sum_{c=1}^{7}\sum_{r}\sum_{n\in\{\mu^{rc}\}} T^{rc}(n) \tag{12}$$

where $\{\mu^{rc}\}$ is the set of UE indexes on the l-th subcarrier in the r-th RB.

The total sum of the throughputs for the n-th UE in the t-th timeslot at the c-th macro cell over the subcarriers and RBs, $T^c(n,t)$, is calculated from Eq. (11) and given by

$$T^c(n,t) = \sum_{r=1}^{N_{RB}}\sum_{l\in\{l^r\}} T_n^{r_l c}(m^{r_l c}). \tag{13}$$

To evaluate the fairness among UEs, the fairness index (FI) is calculated as [18]

$$\mathrm{FI} = \frac{\sum_{c=1}^{7}\left(\sum_{n=1}^{N_U}\frac{1}{T_{ave}}\sum_{t=1}^{T_{ave}} T^c(n,t)\right)^2}{7\sum_{c=1}^{7} N_U \sum_{n=1}^{N_U}\left(\frac{1}{T_{ave}}\sum_{t=1}^{T_{ave}} T^c(n,t)\right)^2} \tag{14}$$

where T_{ave} is the period for averaging the radio-resource scheduling.

The system throughput given in Eq. (12) and the FI given in Eq. (14) are used for the evaluation of the conventional and proposed schemes.

3.2 Parameter Dependence of Proposed Scheme

The system throughput versus the learning rate α is shown in Fig. 5. The number of UEs is 5, the inter-antenna distance is 100 meters, and the number of timeslots in UE selection is 7. The highest system throughput is achieved when the learning rate is 0.1. The Q-value is less affected by the reward, as given in Eq. (9), when the learning rate α is small. Thus, the reward is underestimated and the UE set that achieves a lower throughput on average tends to survive. Moreover, the Q-value is strongly affected by the reward, as given in Eq. (9), when the learning rate α is large. Thus, the reward is overestimated and the UE set that suffers from the instantaneous inter-cell interference is tends to be eliminated. We apply the learning rate α of 0.1 in the following numerical results.

The transient response of the algorithm is shown in Fig. 6. The number of UEs is 5 and the number of timeslots in which UE sets are selected with RL is 21. The convergence of the system throughput is delayed until the RL is completed.

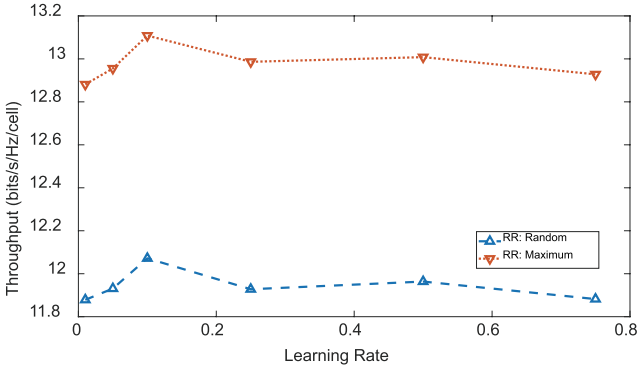


Fig. 5 System throughput vs. learning rate (N_U=5, inter-antenna distance 100 m, timeslot in UE selection 7).

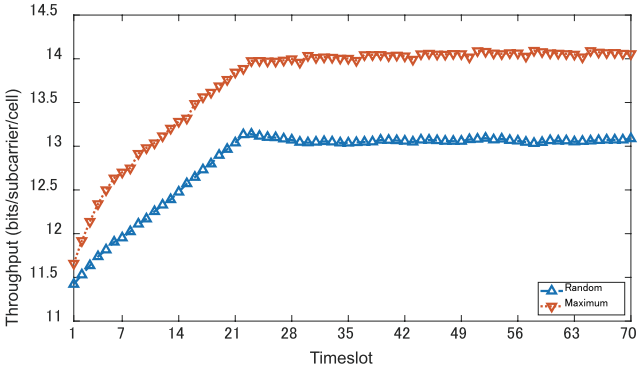


Fig. 6 System throughput vs. timeslot (N_U=5, inter-antenna distance 100 m, no. of timeslots for RL=21).

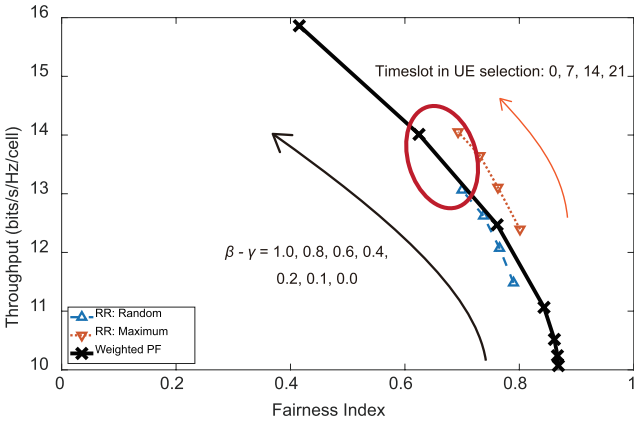


Fig. 7 System throughput vs. FI (N_U=5, inter-antenna distance 100 m).

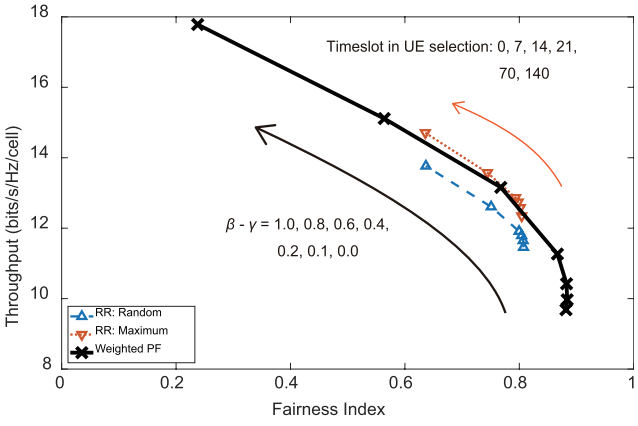
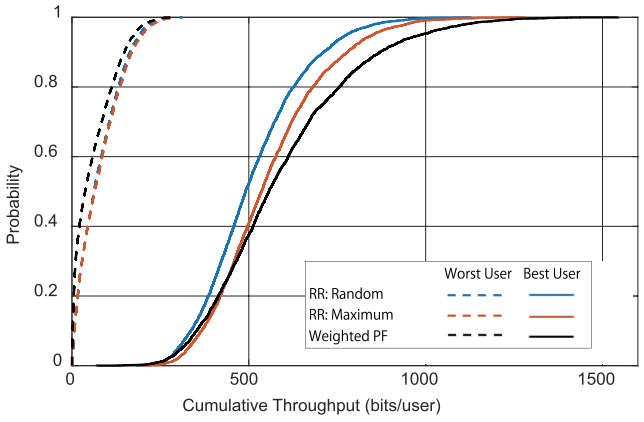
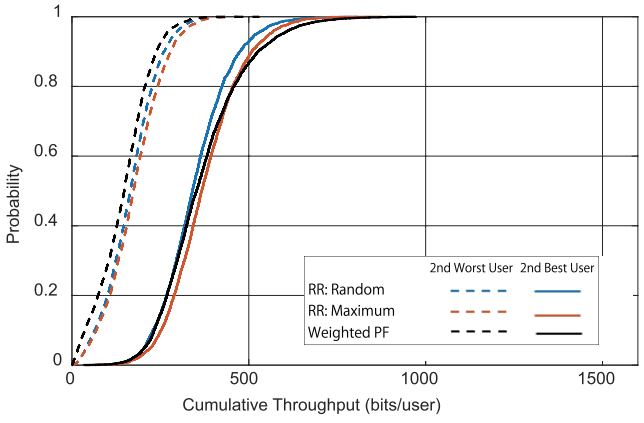


Fig. 8 System throughput vs. FI (N_U=10, inter-antenna distance 100 m).



(a) Best and worst user throughput (no. of timeslots for RL=21, $\beta - \gamma = 0.1$, inter-antenna distance 100 m).



(b) 2nd best and 2nd worst user throughput (no. of timeslots for RL=21, $\beta - \gamma = 0.1$, inter-antenna distance 100 m).

Fig. 9 CDF of user throughputs (N_U=5).

3.3 Effect of Number of UEs

The system throughput versus FI is shown in Figs. 7 and 8. The numbers of UEs are 5 and 10. The inter-antenna distance is 100 meters.

As the number of timeslots in which UE sets are selected with RL increases, the system throughput improves and the fairness among UEs deteriorates in the proposed RR scheduling as shown in Figs. 7 and 8. The CU selects an action with the maximum Q-value to modify the RR sequences and the system throughput increases as the reward is given by the average system throughput after the action. This leads to the deterioration of the fairness among UEs because the UE sets with lower throughputs are eliminated. In the proposed RR scheduling, the performance of the random selection is inferior to that of the maximum selection because the maximum selection chooses UE sets with the highest estimated throughput. The curves in Fig. 7 show that the fairness among UEs with the weighted PF becomes the highest when the weight $\beta - \gamma$ is set to 1.0. However, the computational complexity of the PF scheduling is much larger than that of the propped RR scheduling after UE sets are selected with RL. The computational complexity of the weighted PF scheduling exponentially increases with the number of UEs. On the other hand, in a steady state after RL, no estimated throughput calculation is required, except for the initial phase selection [8]. Moreover, the fairness among UEs and the system throughput for the maximum selection in the proposed RR scheduling are superior to those for the weighted PF when the weighted PF scheduling puts more weight on the system throughput (i.e., the weight for the PF scheduling, $\beta - \gamma$, is from 0.1 to 0.4). The performance for the maximum selection in the proposed RR scheduling is equivalent to that for the weighted PF scheduling when the number of UEs is 10 as shown in Fig. 8.

The numerical results in Figs. 7 and 8 indicate that the range of the system throughput increases with the number of UEs. This is because the number of candidates of UE sets increases and the UE sets with lower inter-cell interference can be allocated.

The cumulative distribution function (CDF) of user throughputs for $N_U = 5$ is shown in Fig. 9, in which the points surrounded by the red circle in Fig. 7 are compared. The inter-antenna distance is 100 meters. The number of timeslots in which UE sets are selected with RL is 21 and the weight for PF scheduling, $\beta - \gamma$, is 0.1. The CDF curves

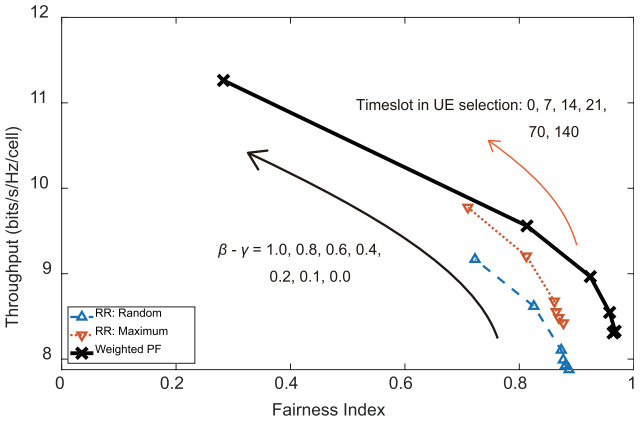


Fig. 10 System throughput vs. FI (N_U=10, inter-antenna distance 50 m).

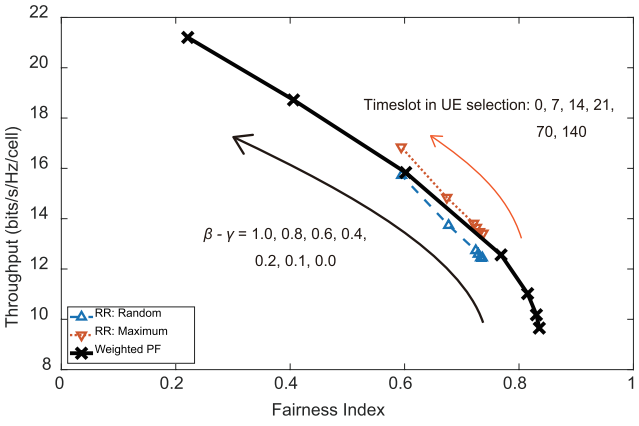


Fig. 12 System throughput vs. FI (N_U=10, inter-antenna distance 150 m).

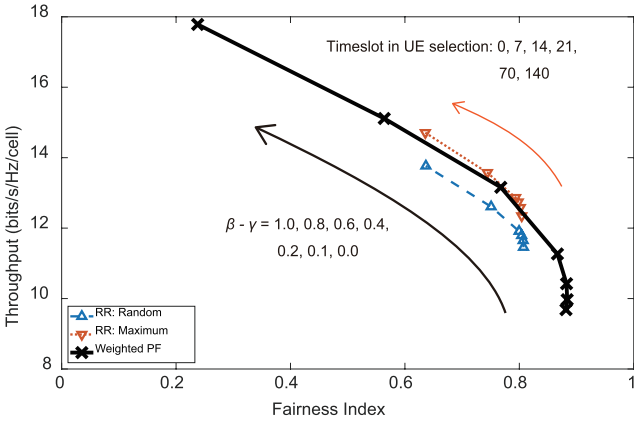


Fig. 11 System throughput vs. FI (N_U=10, inter-antenna distance 100 m).

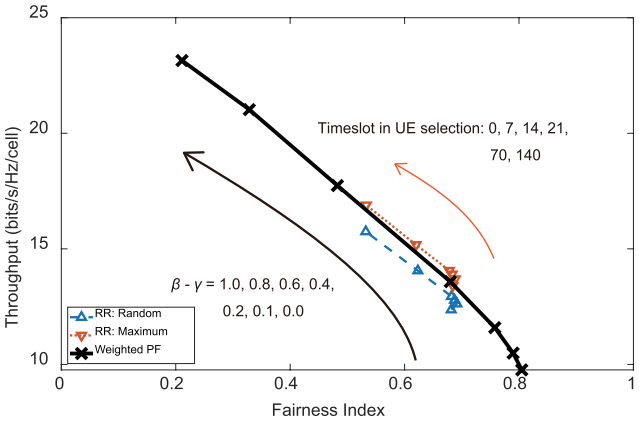


Fig. 13 System throughput vs. FI (N_U=10, inter-antenna distance 200 m).

of the worst and best user throughputs are shown in Fig. 9(a) and those of the 2nd best and 2nd worst user throughputs are shown in Fig. 9(b). It is shown in Fig. 9(a) that the user throughput for the worst user of the maximum selection in the proposed RR scheduling is larger than that in the weighted PF scheduling while the user throughput for the best user of the maximum selection in the proposed RR scheduling is less than that in the weighted PF scheduling. It is also shown in Fig. 9(b) that the user throughput for the 2nd worst user of the maximum selection in the proposed RR scheduling is larger than that in the weighted PF scheduling. In addition, the same tendency as that of the best user can be observed for the 2nd best user. The range of throughput values with the maximum selection in the proposed RR scheduling is smaller than that in the weighted PF scheduling. This is the reason why the maximum selection shows better fairness for the same system throughput as indicated by the plots surrounded by the red circle in Fig. 7.

3.4 Effect of Inter-Antenna Distance

The system throughputs versus FI for different inter-antenna distances are shown in Figs. 10, 11, 12, and 13. The inter-antenna distances are 50, 100, 150, and 200 meters, respectively. The number of UEs is 10 and the number of timeslots in which UE sets are selected with RL is set from 0 to 140. Regardless of the inter-antenna distance, as UE set selection with RL progresses, the system throughput improves and the fairness deteriorates. Larger inter-antenna distance leads to longer distance from TPs to cell-edge UEs. It is then difficult to balance the fairness among UEs because the larger inter-antenna distance leads to a larger range of throughput among UEs. The performance of the weighted PF scheduling is superior to that of the proposed RR scheduling when the inter-antenna distance is 50 meters as shown in Fig. 10. The performance for the maximum selection of the proposed RR scheduling is comparable to that for the weighted PF scheduling in terms of both the fairness and the throughput when the weight $\beta-\gamma$ is from 0.4 to 0.1 and the inter-antenna distance is 100 meters as shown in Fig. 11. The performance for the random selection of the proposed RR scheduling is also closer to that of the weighted PF scheduling when the inter-antenna distance increases to 150 and 200 meters as shown in Figs. 12 and 13. This is because the weighted PF scheduling tends to allocate the worse throughput users in order to improve the fairness among UEs while the proposed RR scheduling allocates the selected UE sets with the bet-

ter system throughput. In the proposed RR scheduling, the estimated throughput which is calculated for the reward values takes no inter-cell interference into account and is not accurate if the inter-antenna distance is small. This is why the weighted PF scheduling is superior to the proposed RR scheduling in Fig. 10.

4. Conclusions

In this paper, UE set selection in the allocation sequences of RR for DAT with BD pre-coding is proposed. In prior research, the initial phase selection of user equipment allocation sequences in RR scheduling has been proposed and the computational complexity has been indicated. In the proposed RR scheduling, the CU takes an action with the highest Q-value calculated from the estimated throughput and the UE sets in the allocation sequences are selected. No throughput estimation calculation is required except for the initial phase selection in the RR scheduling in the steady state after RL. Numerical results obtained through computer simulation show that the maximum selection, one of the criteria for initial phase selection, is comparable to or outperforms the weighted PF scheduling in terms of the computational complexity, fairness, and throughput when the weight $\beta - \gamma$ is from 0.4 to 0.1.

Acknowledgments

This work is supported in part by a Grant-in-Aid for Scientific Research (C) under Grant No. 16K06366 from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

[1] "IMT Vision - Framework and Overall Objectives of the Future Development of IMT for 2020 and beyond," ITU-R M.2083-0, Sept. 2015.
[2] E. Dahlman, S. Parkvall, and J. Sköld, 4G:LTE/LTE-Advanced for Mobile Broadband Edition, Academic Press, 2011.
[3] F. Adachi, A. Boonkajay, T. Saito, S. Kumagai, and H. Miyazaki, "Cooperative distributed antenna transmission for 5G mobile communications network," IEICE Trans. Commun., vol.E100-B, no.8, pp.1190–1240, Aug. 2017.
[4] F. Adachi, R. Takahashi, and H. Matsuo, "Enhanced interference coordination and radio resource management for 5G advanced ultra-dense RAN," IEEE 91st Vehicular Technology Conference, May 2020.
[5] Y. Seki and F. Adachi, "Downlink capacity comparison of MMSE-SVD and BD-SVD for cooperative distributed antenna transmission using multi-user scheduling," IEEE 86th Vehicular Technology Conference, Sept. 2017.
[6] Y. Arikawa, T. Sakamoto, and S. Shigematsu, "Hierarchical scheduling with FPGA-based accelerator for flexible 5G mobile network," IEEE 91st Vehicular Technology Conference, May 2020.
[7] Y. Arikawa, T. Sakamoto, and S. Kimura "User throughput analysis of coordinated radio-resource scheduler with hardware accelerator for 5G mobile systems," IEICE Communications Express, vol.7, no.6, pp.183–188, 2018.
[8] G. Otsuru and Y. Sanada, "User allocation with round-robin scheduling sequence for distributed antenna system," IEEE 90th Vehicular Technology Conference, Sept. 2019.
[9] G. Otsuru and Y. Sanada, "Phase selection in round-robin scheduling sequence for distributed antenna system," IEICE Trans. Commun., vol.E103-B, no.10, pp.1155–1163, Oct. 2020.
[10] G. Otsuru and Y. Sanada, "User set elimination in allocation sequences of RR scheduling for distributed antenna transmission," IEEE 93rd Vehicular Technology Conference, April 2021.
[11] M. Feng and S. Mao, "Dealing with limited backhaul capacity in millimeter-wave systems: A deep reinforcement learning approach," IEEE Commun. Mag., vol.57, no.3, pp.50–55, March 2019.
[12] F. Tang, Y. Zhou, and N. Kato, "Deep reinforcement learning for dynamic uplink/downlink resource allocation in high mobility 5G HetNet," IEEE J. Sel. Areas Commun., vol.38, no.12, pp.2773–2782, June 2020.
[13] Y. Zhou, F. Tang, Y. Kawamoto, and N. Kato, "Reinforcement learning-based radio resource control in 5G vehicular network," IEEE Wireless Commun. Lett., vol.9, no.5, pp.611–614, Dec. 2019.
[14] W. Lei, Y. Ye, and M. Xiao, "Deep reinforcement learning-based spectrum allocation in integrated access and backhaul networks," IEEE Trans. Cogn. Commun. Netw., vol.6, no.3, pp.970–979, Sept. 2020.
[15] M. Elsayed, M. Erol-Kantarci, and H. Yanikomeroglu, "Transfer reinforcement learning for 5G new radio mmWave networks," IEEE Trans. Wireless Commun., vol.20, no.5, pp.2838–2849, May 2021.
[16] Q.H. Spencer, A.L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO systems," IEEE Trans. Signal Process., vol.52, no.2, pp.461–471, Feb. 2004.
[17] R. Agrawal, A. Bedekar, R.J. La, and V. Subramanian, "Class and channel condition based weighted proportional fair scheduler," Proc. ITC2001, Sept. 2001.
[18] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," DEC Research Report, TR-301, Sept. 1984.
[19] X. Chen, A. Benjebbour, A. Li, and A. Harada, "Multi-user proportional fair scheduling for uplink non-orthogonal multiple access (NOMA)," IEEE 79th Vehicular Technology Conference, May 2014.
[20] "Study on Channel Model for Frequencies from 0.5 to 100 GHz," 3GPP, TR 38.901, V14.3.0, Jan. 2018.
[21] T. Okuyama, S. Suyama, J. Mashino, and Y. Okumura, "5G distributed massive MIMO with ultra-high density antenna deployment in low SHF bands," IEICE Trans. Commun., vol.E100-B, no.10, pp.1921–1927, Oct. 2017.

Go Otsuru was born in Nagasaki, Japan, in 1995. He received his B.E. degree and M.E. degree in electronics engineering from Keio University, Japan, in 2019 and 2021, respectively. Since April 2019, he has been a graduate student in the School of Integrated Design Engineering, Graduate School of Science and Technology, Keio University. His research interests are mainly in resource allocation.

Yukitoshi Sanada was born in Tokyo in 1969. He received his B.E. degree in electrical engineering from Keio University, Yokohama, Japan, his M.A.Sc. degree in electrical engineering from the University of Victoria, B.C., Canada, and his Ph.D. degree in electrical engineering from Keio University, Yokohama, Japan, in 1992, 1995, and 1997, respectively. In 1997 he joined the Faculty of Engineering, Tokyo Institute of Technology as a Research Associate. In 2000, he joined Advanced Telecommunication Laboratory, Sony Computer Science Laboratories, Inc, as an associate researcher. In 2001, he joined the Faculty of Science and Engineering, Keio University, where he is now a professor. He received the Young Engineer Award from IEICE Japan in 1997. His current research interests are in software defined radio, cognitive radio, and OFDM systems.