

AI/ML-Based TRP Selection Method for System Throughput Maximization in Multi-UE Environments Using a 6G Simulator

Yuta Hayashi, Hiromasa Terashi, Dan Mohri, Satoshi Suyama, Yuyuan Chang, and Huiling Jiang
NTT DOCOMO, INC., 3-6 Hikarino-oka, Yokosuka-shi, Kanagawa, 239-8536 Japan
E-mail: yuuta.hayashi.fb@nttdocomo.com

Abstract—The sixth-generation mobile communication system (6G) sets forth a wide range of requirements, including extreme-high-speed and high-capacity communications. To realize 6G, the advancement of wireless communication technologies and the exploration of new network topologies are essential. Furthermore, as a newly added element in 6G, the use of artificial intelligence (AI) and machine learning (ML) is being investigated to control and operate wireless communication systems more intelligently. Therefore, the authors have been developing a 6G system-level simulator (6G simulator) to evaluate and visualize 6G-oriented technologies. Our previous works include the implementation and evaluation of new candidate frequency bands such as the sub-terahertz and centimeter-wave bands, the integration of advanced distributed network technologies like reconfigurable intelligent surfaces (RIS), and the deployment of AI/ML-driven predictive control techniques for communication quality. In this paper, we further advance the AI/ML scenarios implemented in the 6G simulator by introducing a transmission and reception point (TRP) selection scenario aimed at maximizing system throughput. We conduct visualization and performance evaluation, compare the performance using different learning methods, and clarify an effective TRP selection approach to maximize system throughput.

Keywords—6G, System level simulation, AI, Machine learning, TRP selection, System throughput maximization

I. INTRODUCTION

Sixth-generation mobile communication system (6G) aims to further enhance the key features of fifth-generation mobile communication system (5G), namely high data rate and large capacity, low latency, and massive connectivity. Specifically, 6G targets extreme-high-speed and large-capacity communications exceeding 100 Gbps, extreme-low-latency communications with latency as low as 0.1 ms, and extreme massive connectivity supporting up to 10 million devices/km² [1], [2]. In addition, 6G is expected to provide ubiquitous coverage that enables network access from virtually anywhere in the world. Thus, 6G has set ambitious goals across a wide range of areas. To achieve these goals, extensive studies are being conducted on sub-terahertz and centimeter-wave bands, which are expected to provide wide bandwidths [3], as well as on advanced distributed networking technologies, referred to as new radio network topology (NRNT), such as reconfigurable intelligent surfaces (RISs), meta-surfaces, and distributed multiple-input and multiple-output (MIMO) systems [4]. Furthermore, in addition to the evolution of high-capacity, low-latency, and massive-connectivity communications introduced in 5G, 6G is

expected to incorporate new features such as wireless sensing, non-terrestrial networks (NTNs), and the utilization of artificial intelligence (AI) [5]. Among these, the use of AI has attracted particularly strong interest, as more intelligent control and operation of wireless communication systems are expected to enable more efficient and higher-quality communications than conventional approaches.

In 6G, the aforementioned elements and technologies are expected to create new values. For example, advances in communications will enable the fusion of cyberspace and physical space. Information obtained from internet of things (IoT) devices and wireless sensing can be processed in cyberspace to predict conditions several seconds ahead, and the predicted information can then be fed back into physical space, thereby proactively preventing or mitigating communication-quality degradation, sudden traffic surges, and increases in power consumption. In addition, since anything can be connected to the network from virtually any location, the use of robots is also being considered. Rather than being limited to localized applications in factories and facilities as in conventional scenarios, a world is envisioned in which robots play active roles in a wide range of daily-life situations, including urban infrastructure inspection, cleaning, and tourist guidance [6].

To satisfy these advanced requirements and create new value, various companies, universities, and research institutions are actively conducting research and development toward 6G. In this context, the authors have been developing a 6G system-level simulator (6G simulator) to evaluate and visualize these individual technologies from a system-wide perspective and to clarify the potential of 6G [7], [8]. Fig. 1 shows the overview of the 6G simulator. To date, sub-terahertz bands, centimeter-wave bands, and NRNT have been incorporated into the 6G simulator, and evaluations have been conducted in a virtual outdoor urban environment. In addition, a machine learning (ML) algorithm, as an AI technology, has been implemented in the 6G simulator. Based on the results of pre-training, transmission and reception point (TRP) selection is performed using predictive control to avoid blockage caused by unexpected obstacles, thereby demonstrating a use case of AI in wireless communication systems [9]. The results in [9] confirmed that AI/ML-based predictive control enables TRP selection that mitigates throughput degradation caused by sudden blockage, and clarified that this is an important approach for use cases, in



Fig. 1 Overview of the 6G simulator

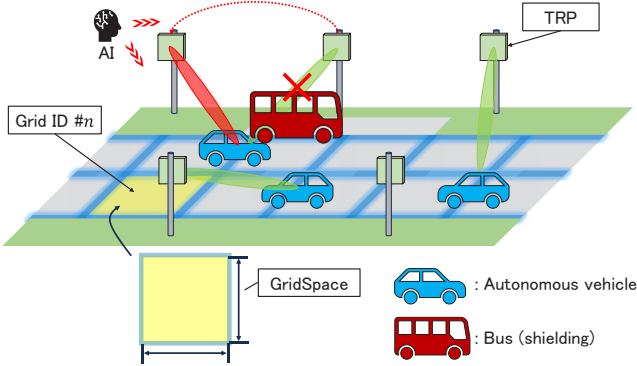


Fig. 2 Overview of the AI/ML scenario

which throughput degradation must be avoided, such as autonomous driving. On the other hand, the previous study considered only single user equipment (UE), and therefore its effectiveness in multi-UE scenarios needed to be verified. In this paper, we enhance the AI/ML scenario implemented in the 6G simulator and evaluate whether the benefits of AI/ML-based control observed in the single-UE case can also be obtained in multi-UE scenarios. Furthermore, in multi-UE environments, if the objective is only to maximize the throughput of each UE, severe mutual interference may occur. Therefore, we examine three learning methods, including one that incorporates interference into the reward design for AI/ML training, compare their performance, and clarify an effective TRP selection method for maximizing the system throughput. While previous studies [10] and [11] focused on optimizing AI-based handover and target cell selection for mobile users, this paper considers a more complex scenario that incorporates dynamic environmental changes, such as periodic signal blockages, in addition to user mobility.

II. AI/ML-BASED TRP SELECTION FOR MAXIMIZING SYSTEM THROUGHPUT

A. Overview

In 5G and 6G, high-frequency bands characterized by strong directionality and large propagation losses are used to realize high-data-rate, high-capacity, and low-latency communications. Therefore, it is important to maintain communication over as short a distance as possible and under line-of-sight (LoS) conditions. Instead of deploying new base stations (BSs), techniques such as RIS and drone-based TRPs are being considered to intentionally establish communication paths.

However, since the radio environment is continuously changing, throughput degradation is highly likely to occur due to unexpected blockages and other factors. In addition, because of the severe propagation loss in high-frequency bands, more TRPs are expected to be deployed than in conventional systems to ensure coverage. As described above, 6G will connect a greater number of devices to the network, which also increases the likelihood of throughput degradation caused by mutual interference. Therefore, intelligent control becomes essential, including continuously monitoring the surrounding environment, selecting TRPs to avoid throughput degradation caused by blockage when necessary, and switching the serving TRP to suppress interference when appropriate.

Accordingly, in this paper, we consider a scenario in which autonomous vehicles travel in an outdoor urban environment, as shown in Fig. 2. For communications between the autonomous vehicles and the TRPs deployed along the roadside, AI/ML-based TRP selection is performed with the objective of maximizing the system throughput. Autonomous vehicles are selected as the target UEs because, in 6G use cases, it is particularly important to provide high-quality and consistent throughput, making them an appropriate use case for investigating one application of AI/ML. In addition, to create a more complex simulation environment, the scenario includes an event in which a bus periodically stops at the roadside and acts as a blockage between the TRP and the UE.

B. Deep Reinforcement Learning

Real world communication environments change continuously, and various factors, such as weather conditions, the number of people in the area, and the presence or absence of blockages, affect the communication environment depending on the time of day. Therefore, in this simulation, deep reinforcement learning (DRL) is employed for TRP selection. DRL is a technique that combines deep learning and reinforcement learning, and this combination improves decision-making capability in more complex environments.

To apply DRL, it is necessary to define the *environment*, *state*, *action*, and *reward*. In this simulation, the environment is defined as “communications between TRPs and autonomous vehicles on a straight road,” and the state is defined as “the UE location information and the presence or absence of a bus acting as an obstacle.” Note that, as shown in Fig. 2, the UE location information is represented by dividing the AI/ML application area into grid cells and using the corresponding grid index. In this paper, the length of one side of a grid cell is defined as “GridSpace.” The action is defined as “TRP selection.” For the reward, different learning models are generated using the three equations given in (1), (2), and (4), which are defined below.

A) Maximization of System Throughput (Reward₁):

$$RW = TP_{sum} = \sum_t \left(\sum_u TP_{u,t} \right), \quad (1)$$

where t denotes the TRP selection period, u denotes the UE index, TP_{sum} denotes the cumulative system throughput, and $TP_{u,t}$ denotes the throughput of each UE. In this reward design, the reward is defined as the sum of the individual throughput obtained by all UEs from the time they connect to a given TRP



Fig. 3 Screenshot of the AI/ML scenario

until the next TRP selection. Note that the same neural network (NN) is used commonly for all UEs.

B) Interference-Aware System Throughput Maximization (Reward₂):

$$RW = |TP_{sum} - \alpha I_{sum}|, \quad (2)$$

$$I_{sum} = \sum_t \left(\sum_u I_{u,t} \right), \quad (3)$$

where I_{sum} denotes the cumulative interference power, $I_{u,t}$ denotes the interference power received by UE, and α is a hyperparameter that determines the extent to which interference is taken into account in the reward. In this simulation, α is set to a fixed value. In this reward design, the total interference in the system is incorporated into the reward in addition to (1). Note that the same NN is also shared among all UEs in this reward design.

C) User Throughput Maximization Considering Generated Interference (Reward₃):

$$RW = \begin{cases} TP - \alpha I_{inf} & (TP \geq \alpha I_{inf}) \\ 0 & (TP < \alpha I_{inf}) \end{cases}, \quad (4)$$

$$TP = \sum_t TP_{u,t}, \quad (5)$$

$$I_{inf} = \sum_t I_{inf-u,t}. \quad (6)$$

In (4)-(6), TP denotes the cumulative throughput of each UE, $I_{inf-u,t}$ denotes the interference caused by UE to the other UEs, and I_{inf} denotes the cumulative caused interference. Unlike (1) and (2), this reward design constructs a different NN for each UE, and each UE optimizes its own throughput while suppressing the interference it causes to other UEs. Note that the reward is made condition-dependent to facilitate convergence of the learning process.

In all of these equations, the throughput and interference are defined as the values obtained from the time a TRP is selected until the next TRP selection. In this paper, these reward functions are referred to as “Reward₁”, “Reward₂”, and “Reward₃”, respectively. In this simulation, an ϵ -greedy policy is employed as the action-selection policy for TRP selection. This policy enables a good balance between “exploration,” in which actions are selected randomly by varying ϵ , and “exploitation,” in which actions are selected based on the rewards obtained through previous exploration, thereby allowing a policy better suited to the environment to be learned. The update rule for ϵ is defined as

$$\epsilon_{t+1} = 0.999 - \{0.001 \times (n - n_{down})\}, \quad (7)$$

where n denotes the episode number, and n_{down} is a correction term, which is set to 1 in this simulation. When (7) is used, ϵ_{t+1} becomes negative after the number of episodes exceeds a certain threshold. To prevent this, a lower bound of 0.001 is imposed on ϵ_{t+1} . Adaptive moment estimation (Adam) [12] is used to train the NN. The NN is updated using the results obtained under the ϵ -greedy policy, and the update timing is synchronized with the TRP selection period. For example, suppose that TRP selection is performed every four slots. If the current time is slot 8, the learning process uses the UE position, the presence or absence of the bus, and the selected TRP at slot 4, together with the throughput and interference obtained up to slot 8 as a result of connecting to the selected TRP. By repeating this process at every TRP selection period, TRP selection that maximizes the system throughput becomes possible.

III. SIMULATION

A. Simulation Environment

Fig. 3 shows a schematic screenshot of the AI/ML scenario implemented in the 6G simulator, while Tables 1 and 2 summarize the learning parameters and the simulation

Table 1 Learning parameters

Learning method	Deep reinforcement learning
Discount factor	0.5
Learning rate	0.003
Size of mini-batch	64
GridSpace	20 m
ε	$0.001 \leq \varepsilon \leq 1.0$
α	2.5×10^{20}

Table 2 Simulation parameters

Center frequency	15 GHz
Bandwidth	400 MHz
Number of antennas of TRP	576
Tx power of TRP	23 dBm
Height of TRP	10 m
Period of TRP selection	10 ms
Number of antennas of UE	144
Tx power of UE	23 dBm
Height of UE	1.5 m
Speed of UE	60 km/h
Period of beam selection	1 ms

respectively. The simulation procedure is as follows. First, a learning model is generated using the parameter values listed in Table 1. Then, the generated model is used to perform TRP selection so as to maximize the system throughput.

In this simulation, a 15 GHz centimeter-wave band is used. Two-dimensional planar array antennas are employed at both the TRPs and the UEs. Each TRP is equipped with a total of 576 antenna elements in an $8H \times 8V \times 9$ panels configuration, and beamforming is performed by analog beamforming at each TRP. Each UE is equipped with a total of 144 antenna elements in a $4H \times 4V \times 9$ panels configuration and also performs analog beamforming in the same manner as the TRPs. In the absence of AI/ML-based control, each UE selects the TRP that provides the highest received power. The number of deployed TRPs is set to eight, the number of target UEs is set to three, and the speed of UE is set to 60 km/h. α used in (2) and (4) is set to the value shown in Table 1. It should be noted that this value was determined empirically through several trials to achieve the best performance.

In this simulation, data transmission is performed using time division duplex (TDD) with a downlink-to-uplink ratio of 10:0. Therefore, this paper evaluates the downlink throughput. In addition, the bus shown in Fig. 3 serves as a blockage between the TRP and the UE by stopping at the roadside in this scenario. For this blockage bus, a model based on the blockage model defined in TR 38.901 [13] is applied. This model places a plate-shaped obstacle and assigns attenuation according to the difference between the direct distance from the transmitter to the receiver and the distance via the upper, lower, left, or right edge of the plate. The simulator also implements a proportional fairness (PF) scheduler, which is applied in this scenario as well.



Fig. 4 Temporal variation of system throughput



Fig. 5 Temporal variation of user throughput for UE #0

B. Simulation Result

Fig. 4 shows the temporal variation in the system throughput. The results compare the case without AI/ML, the case using the Reward_1, the case using the Reward_2, and the case using the Reward_3. The results confirm that, for all reward designs, the use of AI/ML improves the system throughput compared with the case without AI/ML. In particular, significant improvements are observed during the periods when the bus stops, indicating that throughput degradation caused by blockage can be mitigated. A comparison among the reward designs shows no significant difference in the achieved system throughput. The average system throughput is 27.5 Gbps for Reward_1, 27.4 Gbps for Reward_2, and 26.7 Gbps for Reward_3.

Fig. 5 shows the user throughput of UE #0, which travels in the same lane as the bus. As shown in Fig. 5, the use of AI/ML improves the user throughput. In particular, the best performance is achieved with Reward_2, which attains an average user throughput of 9.18 Gbps. In contrast, Reward_3 results in lower user throughput than the other reward designs.

The above results can first be attributed to the change in TRP selection introduced by AI/ML. Fig. 6 shows the TRP selection behavior with and without AI. The timing shown in Fig. 6 corresponds to the period when the bus acting as a blockage is stopped at the roadside. Without AI/ML, each UE selects the TRP that provides the highest received power at each TRP selection instant. As a result, two UEs connect to TRP ID #5, while one UE connects to TRP ID #2, which is located near the stopped bus. In contrast, when AI/ML is used, each UE connects to a different TRP to maximize the system throughput, and

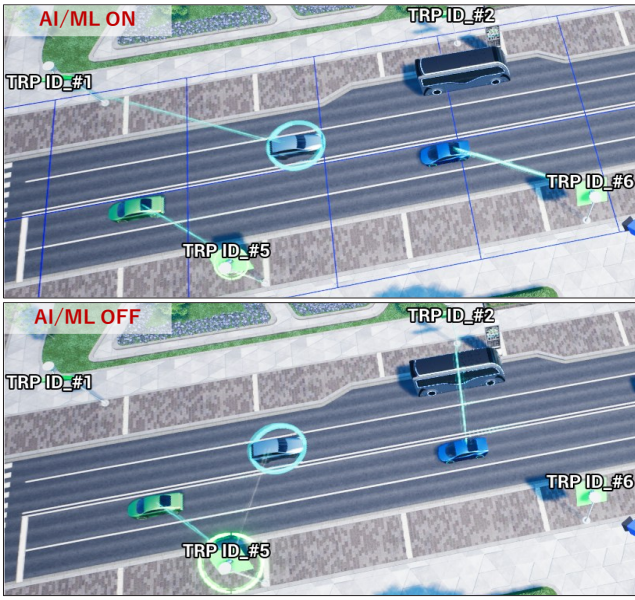


Fig. 6 TRP Selection with and without AI/ML

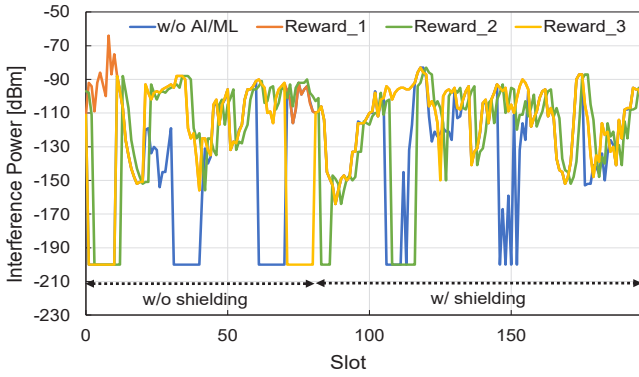


Fig. 7 Temporal variation of interference power for UE #0

connection to TRP ID #2 near the stopped bus is also avoided. When two UEs are connected to the same TRP, the PF scheduler allocates radio resources to each UE, and therefore the UE with better reception quality tends to achieve higher throughput, whereas the UE with poorer reception quality tends to achieve lower throughput. Although this corresponds to the intended operation of the PF scheduler, depending on the surrounding conditions, selecting a TRP other than the one with the highest received power can instead improve both the user throughput and the system throughput. In addition, for communication with TRP ID #2, which is blocked by the bus, the reference signal received power (RSRP) decreases by approximately 30 dB due to the blockage, which becomes a major cause of throughput degradation. By contrast, when AI/ML is used, the UE is instead connected to TRP ID #6, making it possible to maintain high throughput. For these reasons, we consider that the use of AI/ML leads to an improvement in system throughput.

On the other hand, when each UE is connected to a different TRP, mutual interference arises. Fig. 7 shows the interference power for UE #0. In terms of the average interference power, the case without AI/ML achieves the lowest interference level, at -98 dB. When AI/ML is used, the average interference power is -86 dB for Reward_1, -97 dB for Reward_2, and -96 dB

for Reward_3. These results indicate that, if the learning objective is solely to maximize the system throughput, the interference power increases by more than 10 dB. In contrast, for the two methods that take interference into account, the interference level can be suppressed to a level comparable to that without AI/ML. In 6G, massive device connectivity and base station deployments will make networks highly interference-dominated. Thus, interference-aware model training, as proposed here, becomes essential. Future work must verify its effectiveness in highly dense environments.

IV. CONCLUSION

In this paper, we implemented in the 6G simulator an AI/ML-based TRP selection scenario aimed at maximizing the system throughput and compared three different reward designs to clarify their characteristics and effectiveness. In future work, we plan to investigate the robustness of each reward design and evaluate their performance under different environments.

REFERENCES

- [1] XGMF, *Beyond 5G White Paper: Message to the 2030s (Version 4.0)*, Dec. 2024. [online]. Available: https://xgmf.jp/wp-content/uploads/2025/04/Beyond_5G_White_Paper_v4_0.pdf
- [2] NTT DOCOMO, INC., *5G Evolution and 6G White Paper (Version 5.0)*, Jan. 2023. https://www.ntt.com/english/binary/pdf/corporate/technology/whitepaper_6g/DOCOMO_6G_White_PaperEN_v5.0.pdf
- [3] Y. Hayashi, T. Tomie, S. Suyama, D. Mohri, Y. Chang, and H. Jiang, "A Study on Transmission Characteristics of Mid-band Massive MIMO in Yokosuka City Using Color Image Method and Link-level-simulation," *IEICE Tech. Report*, RCS2025-132, Oct. 2025.
- [4] D. Mohri, S. Suyama, Y. Chang, and H. Jiang, "Hybrid Beamforming Algorithm with Null Steering for RIS-empowered 6G System," *IEEE VTC-Spring 2025*, June 2025.
- [5] ITU-R, *Framework and overall objectives of the future development of IMT for 2030 and beyond*, Nov. 2023. [online]. https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2160-0-202311-I!!PDF-E.pdf
- [6] NTT DOCOMO, INC., *AI Centric NW*, [online]. https://www.ntt.com/english/corporate/technology/rd/6g/001/?icid=CRP_CORP_technology_rd_6g_001_to_CRP_en_CORP_technology_rd_6g_001
- [7] K. Tateishi, K. Kitao, S. Suyama, and T. Yamada, "Advancement of 6G System-Level Simulator," *NTT DOCOMO Technical Journal*, vol.31, no.2, Jul. 2023.
- [8] K. Tateishi, Y. Hayashi, S. Suyama, Y. Chang, and H. Jiang, "6G System Level Simulator for Evaluation Introduction of Mid-Band and Future Predictive Control Technology," *IEICE Tech. Report*, SR2024-75, Jan. 2025.
- [9] Y. Hayashi, K. Tateishi, S. Suyama, Y. Chang, and H. Jiang, "Shielding Avoidance Evaluation and Visualization through Future Predictive Control Using AI/ML on 6G Simulator," *IEEE VTS APWCS2025*, Aug. 2025.
- [10] S. Schneider, H. Karl, R. Khalili, and A. Hecker, "Multi-Agent Deep Reinforcement Learning for Coordinated Multipoint in Mobile Networks," *IEEE Transactions on Network and Service Management*, Feb. 2024.
- [11] R. Ou, Y. Liu, Y. Xie, X. Liu, P. Wu, and T. Qiu, "Optimizing Multi-Cell Selection Handover in Cellular Networks: A Deep Reinforcement Learning Approach," *MSN 2024*, Dec. 2024.
- [12] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," *ICLR 2015*, May 2015.
- [13] 3GPP TR38.901 V16.1.0, Rel-16, 5G: *Study on channel Model for Frequencies from 0.5 to 100 GHz*, Nov. 2020.