IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 December 2023, accepted 14 January 2024, date of publication 25 January 2024, date of current version 7 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3358409

RESEARCH ARTICLE

Ultrafast Resource Allocation by Parallel Bandit Architecture Using Chaotic Lasers for Downlink NOMA Systems

MASAKI SUGIYAMA[1], TAKATOMO MIHANA[2], AOHAN LI[1,3], (Member, IEEE), MAKOTO NARUSE[2], AND MIKIO HASEGAWA[1], (Member, IEEE)

[1]Department of Electrical Engineering, Graduate School of Engineering, Tokyo University of Science, Tokyo 125-8585, Japan
[2]Department of Information Physics and Computing, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan
[3]Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan

Corresponding author: Aohan Li (aohanli@ieee.org)

This work was supported in part by the Japan Society for the Promotion of Science through the Grant-in-Aid for Transformative Research Areas (A) under Grant JP22H05197.

ABSTRACT The effectiveness of laser chaos decision-makers in facilitating ultrafast decision-making makes it possible for a real-time process of channel allocation scheme for the non-orthogonal multiple access (NOMA) technology in the next generation of wireless communications. However, managing the increasing number of users is challenging as the complexity of channel allocation increases significantly. To resolve this challenge, this paper proposes a novel approach to address scalability problems by introducing a parallel bandit architecture using an array of laser chaos decision-makers. In the proposed method, each user is allocated a specific channel by a dedicated laser chaos decision-maker, thereby reducing the number of available options compared with the conventional approach. This parallel bandit architecture enables the system to efficiently manage increasing users while maintaining scalability and ultrafast channel allocation in NOMA. Additionally, fairness is considered by incorporating a logarithmic utility function for design compensation. Numerical simulation results demonstrate that the proposed method achieves higher data rate and enhanced fairness than conventional NOMA approaches such as minimum distance-NOMA (MD-NOMA), conventional-NOMA (C-NOMA), and uniformed channel gain difference-NOMA (UCGD-NOMA). Moreover, the system performance is evaluated on a larger scale, accommodating a significant number of users, with the study considering up to 64 users, surpassing the limitations of the conventional approach of one laser chaos decision-maker, which is constrained to 10 users.

INDEX TERMS Non-orthogonal multiple access (NOMA), channel allocation, logarithmic utility function, laser chaos decision-maker, multi-armed bandit problem.

I. INTRODUCTION

With the rapid development of wireless networks, the number of mobile devices is increasing rapidly. Thus, the technology to process the massive mobile data traffic for wireless communication systems should be continually developed [1], [2]. Non-orthogonal multiple access (NOMA) is one of the promising technologies to satisfy the demands of next-generation wireless communication systems [3], [4]. Compared with orthogonal multiple access (OMA), NOMA has various advantages, such as superior spectral efficiency and scale connectivity [5], [6], [7]. The core concept of NOMA is to serve multiple users concurrently in the same resource block (i.e., the same time and channel) [8], [9]. For power-domain NOMA systems, superposition coding is used at the transmitters, and the receivers use successive interference cancellation (SIC) to distinguish the desired signal from the multiplexed signal [10], [11]. In a NOMA system, resource allocation, such as channels and power allocation, is essential to maximizing the benefits of the

The associate editor coordinating the review of this manuscript and approving it for publication was Hailong Sun.

NOMA. Channel allocation is extremely important as it can improve performance, for instance, a high signal-to-interference-plus-noise ratio (SINR) and throughput [12].

Several studies have been conducted on channel allocation schemes for NOMA systems [13], [14], [15], [16], [17], [18]. In [13] and [14], conventional-NOMA (C-NOMA) and uniformed channel gain difference-NOMA (UCGD-NOMA) are proposed. In C-NOMA and UCGD-NOMA, up to two users are multiplexed in the same channel. The common concept of C-NOMA and UCGD-NOMA is to pair far and near users. Here, the definition of far and near users is based on the distance between the users and the base station (BS), and pairing means multiplexing the signals of multiple users in the same channel. In these two methods, users are first classified into two groups, ''near area'' and ''far area'', according to their distance from the BS. In C-NOMA, the user closest to the BS in the near area is paired with the user farthest from the BS in the far area. The second closest user in the near area is paired with the second farthest user in the far area. Other users are paired in order according to this rule. In UCGD-NOMA, the user that is nearest to the BS in the near area is paired with the user that is nearest to the BS in the far area. The second nearest user in the near area is paired with the second nearest user in the far area. Other users are paired in order according to this rule. Although these two methods are easy to implement, the prior location information of the users should be known. Hence, fast channel allocation may be difficult because location estimation is necessary. In addition, these two methods assume that up to two users are multiplexed in the same channel, which may lead to the loss of the benefit of the NOMA system.

Deep reinforcement learning (DRL) has been studied in various fields of wireless communication because of its superior performance, such as Reconfigurable Intelligent Surfaces (RIS), Millimeter Wave (mmWave), Multiple-Input and Multiple-Output (MIMO), Unmanned Aerial Vehicle (UAV), Internet of Things (IoT), vehicular systems, etc., [19], [20], [21], and [22]. NOMA is no exception for it. To achieve the full benefit of NOMA, scholars have recently studied DRL based channel allocation approaches [15], [16], [17], [18]. More than two users can be multiplexed in the same channel with DRL approaches. In [16], channel assignment was based on DRL with the framework called actor-critic. In [17], a deep learning method using a long short-term memory (LSTM) network was applied to NOMA systems. In [18], DRL was applied with an attention-based neural network (ANN). In [15], a DRL-based method was proposed to optimize channel allocation. Specifically, the study adopted a paradigm of DRL called actor-critic, and a recurrent neural network (RNN) module was incorporated into an agent's decision-making process to capture relationships by leveraging the RNN's ability to integrate contextual information through time. This helps the agent make decisions by extracting additional information from past time slots. Simulation results indicated that the proposed method obtained a high total data rate. In [16], a method that conducts power and channel allocations using the approximate solution method and Deep Q Learning (DQL), respectively, was proposed. This method first solves power allocation using Lagrange multipliers, considering various constraints of the NOMA system and satisfying the Karush-Kuhn-Tucker (KKT) optimality conditions. After power allocation, channel allocation, which is priority-based, is conducted using DQL. Simulation results revealed a high performance of the proposed method in terms of sum data rate and energy efficiency. Reference [17] proposed a framework that uses deep learning to primarily address energy-efficient user association, subchannel allocation, and power allocation in NOMA mmWave heterogeneous networks. Simulation results showed that the proposed scheme improves the energy efficiency of systems. In [18], a method using Deep Q-learning network (DQN) and back propagation neural network (BPNN) for user clustering and power allocation, respectively, was proposed. User clustering is conducted using the DQN to maximize the total data rate. Power allocation is conducted using BPNN, which learns the correlation between the power allocation coefficients and users' channel state information (CSI) in each cluster. Simulation results revealed that the proposed approach achieves a high system spectrum efficiency that closely matches that of an exhaustive search.

However, these proposed schemes are complex for channel allocation. This means that these methods have computational complexity for training the model. These methods require a long training time to obtain proper results. Therefore, high-speed, real-time processing for future dynamic environments would be difficult to achieve with these methods.

Recently, the interest in optical technology as a key component of artificial intelligence, particularly machine learning, has increased [23]. Photonic-based machine learning methods have been reported using laser chaos, which provides an ultrafast chaotic laser output [24], [25], [26]. In [24], the authors solved two-armed bandit problems in the GHz order using the laser chaos time series. This paper refers to such a system as a laser chaos decision-maker. In addition, the time-domain correlations of laser chaos can accelerate the solution of the bandit problem [27]. Reference [25] introduced time-division multiplexing of chaotically oscillating ultrafast time series and demonstrated a scalable, pipelined, multi-armed bandit problem (MAB)-solving principle. The MAB algorithm is in which an agent ''explores'' an optimal action and ''exploits'' the result of the selected action to maximize its reward [28], [29].

To achieve ultrafast decisions in wireless communications to adapt to dynamic environments, the applications of the laser chaos decision-maker to wireless communications have been studied [12], [30], [31], [32]. In [30], the MAB algorithm based on laser chaotic time series was used for dynamic and autonomous channel selection in

TABLE 1. Comparison of the proposed method and previous studies.

	[15]–[18]	[12], [32]	this paper
Channel allocation method	DRL	Single laser chaos decision-maker	Parallel bandit architecture using chaotic laser
Scalability	60 users	10 users	64 users
Variables being optimized	Channel allocation variables	Pairing variables	Channel allocation variables
State space	Necessary, Multidimensional	0 (unnecessary)	0 (unnecessary)
Processing computation time	$10^{-3} \sim 10^{-1}$ s [36]	$10^{-9} \sim 10^{-8}$ s [26]	$10^{-9} \sim 10^{-8}$ s [26]
Hardware requirement for implementation	CPU or GPU	Optical equipment	Optical equipment
Memory	State space, Actions, Rewards	Threshold value and reward probability	Threshold value and reward probability
Implementation complexity	CPU or GPU	Single laser chaos decision-maker	X laser chaos decision-makers
Number of options for laser chaos decision maker	—	$(X-1)!!$	Number of channels ($< X$)

an IEEE802.11a-based four-channel wireless local area network. The algorithm successfully solved the channel selection problem in wireless communication systems with dynamically changing communications environments due to traffic, interference, and fading. In [31], the laser chaos decision-maker solved channel selection problems, including channel bonding in IEEE802.11ac networks. Autonomous and adaptive channel bonding was experimentally demonstrated, and better performance was achieved regarding throughput than other MAB-based methods such as ϵ-greedy and upper confidence bound (UCB) 1-tuned algorithms.

In [12] and [32], the user pairing problem in the NOMA systems was considered. These studies demonstrated that the laser chaos decision-maker can accommodate user pairing and achieve optimal/near-optimal pairing solutions. However, with the increase in users, these schemes cannot easily achieve optimal pairing because the total number of possible pairings is large. In other words, the total number of possible pairings is $(X-1)!!$ when up to two users are multiplexed in the same channel, where X is the total number of users and $(X-1)!!$ means $(X-1) \times (X-3) \times \cdots \times 3 \times 1$. Furthermore, when more than two users are multiplexed in the same channel, achieving optimal/near-optimal solutions becomes even more difficult. Owing to the high complexity of the designed methods based on a scalable laser chaos decision-maker in [12] and [32], the performance was evaluated under the NOMA system with up to 10 users, which is difficult to apply to the next-generation wireless systems, where the number of users will expand rapidly. In addition, these studies did not consider "fairness" regarding the equal distribution of wireless resources among users. In NOMA systems, considering fairness is essential to actualizing efficient communications.

To address the abovementioned problems, we propose a scalable architecture for managing numerous users in the next generation of NOMA systems using multiple laser chaos decision-makers. Specifically, each laser chaos decision-maker is used by the BS to allocate channels for each user. Because the laser chaos decision-maker can make decisions and perform channel selection at ultrahigh-speed without prior information, it is implemented for channel assignment in the NOMA system. In addition, recent advances in integrated and silicon photonics [34], [35], including on-chip photonic decision-makers, suggest using photonic accelerators to enhance electronic systems [23]. Hence, the proposed architecture can be used in the BS to allocate channels for the users in future wireless systems. To highlight the motivation of selecting channels based on the parallel bandit architecture laser chaos proposed in this paper, we compare and analyze the performance, e.g., variables being optimized, state space, implementation complexity, etc. of the proposed method with the DRL methods [15], [16], [17], [18], and our previous work on single laser chaos decision maker [12], [32]. The comparison of the proposed and previous studies is summarized in Table 1.

As shown in Table 1, the channel allocation method using laser chaos decision-maker is much faster in operation than DRL. That is, the DRL methods need to be implemented on Central Processing Unit (CPU) or Graphics Processing Unit (GPU), the processing computation time of which is around $10^{-3} \sim 10^{-1}$ s [36]. The laser chaos methods are implemented using the laser chaos decision-maker, the processing time is $10^{-9} \sim 10^{-8}$ s [26]. This is because the generation of laser chaos is very fast. In addition, the parallel bandit architecture can expand the scalability compared to the channel allocation method using the single laser chaos decision-maker proposed in our previous study. This is because the number of options in laser chaos decision-maker is reduced for the parallel bandit architecture. Specifically, the number of options for single laser chaos decision-maker is $(X-1)!!$ while that for the parallel bandit architecture equals the number of channels, which is less than X, where X is the number of users. Regarding the state space, it is necessary to work DRL because DRL utilizes a neural network. Regarding the state space, DRL-based methods have a multidimensional state-space as it is a neural net-based technique, while the laser-based methods work on a different principle; hence, state-space is not applicable for such. Regarding hardware requirements for implementation, the laser chaos decision-maker method needs optical equipment such as a laser, optical fiber, coupler, and so on. Due to the low implementation complexity of the parallel bandit architecture

TABLE 2. The notations used in this paper.

Symbol	Description	Symbol	Description
$\mathbf{U}$	Set of users	U	Number of users
$\mathbf{C}$	Set of channels	C	Number of channels
d_u	Distance between the u-th user and BS	B	Total bandwidth
B_c	Bandwidth of the c-th channel	$\chi_u^c \in \{0,1\}$	Binary variable which indicates if the u-th user is allocated to the c-th channel
M_c	Number of users multiplexed in the c-th channel	x_c	NOMA symbol on the c-th channel
b_u	Transmission symbols of the u-th user	p_u	Transmission power of the u-th user
h_u	Channel response between the BS and the u-th user	z_c	Noise (AWGN in this paper)
$\sigma_{z_c}^2$	Variance of AWGN	λ	Pathloss exponent
R_u^{NOMA}	Data rate of the u-th user when performing NOMA	R_u^{OMA}	Data rate of the u-th user when performing OMA
$T(t)$	Threshold value at step t	$TA(t)$	Threshold adjuster value at step t
Z	Natural number which decides the total number of possible threshold levels	k	Fixed value determining the range of the resultant $T(t)$
α	Forgetting rate	Δ	Constant increment
Ω	Increment parameter	J_i	Number of times that slot machine i was selected
L_i	Number of wins by playing the selected slot machine i	$\hat{P}_i$	Estimated reward probability of the slot machine i
$\mathbf{L}$	Set of laser chaos decision-makers	L_u	The u-th laser chaos decision-maker
S_j	Binary variable to represent channel	K	The K-th MSB in binary code to represent channel
τ	The number of iterations that the data rate are considered	ξ_u	The order of the long distance between the u-th user and BS among that of all users assigned to the c-th channel.
P_c	Transmit power assigned to the c-th channel	a_c	Power allocation factor of the c-th channel

chaotic laser method for more users in NOMA systems as described above, [33] proposes channel allocation using multiple laser chaos decision-makers. Compared to [33], this paper considers user fairness and the integration of OMA and NOMA. The following are the main contributions of this paper:

- We design a fast channel allocation scheme based on a parallel structure with multiple laser chaos decision-makers that can perform real-time decisions for downlink NOMA systems with a larger number of users.
- To consider fairness for the users in the NOMA system, we improve the proposed scheme by inducing a logarithmic utility function.
- To confirm the effectiveness of the proposed method, we evaluate the performance of the proposed scheme and compare it with other methods. The simulation results show that the proposed scheme has superior performance in terms of total throughput and fairness.

The remainder of this paper is organized as follows. In Section II, we describe the system model of this study. Section III introduces the laser chaos decision-maker. Section IV presents the proposed method. Section V presents numerical simulations of the proposed method. Section VI concludes the paper. The notations used in this paper and their descriptions are summarized in Table 2.

II. SYSTEM MODEL

A downlink NOMA system consisting of one BS and multiple users is considered. As shown in Fig. 1, multiple users can be multiplexed on one channel simultaneously, which

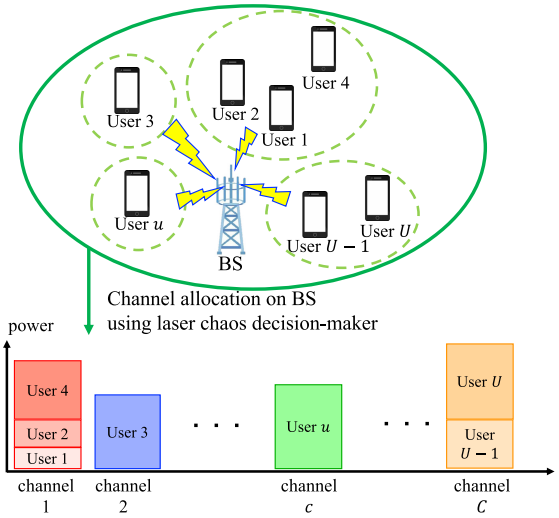


FIGURE 1. System model.

can improve spectrum efficiency. We assume that all the transmitters and receivers have one antenna each. Let the set of users in the NOMA system be $\mathbf{U} = \{1, 2, \cdots, u, \cdots, U\}$ and the index of each channel be $\mathbf{C} = \{1, 2, \cdots, c, \cdots, C\}$, where U and C denote the total number of users and channels, respectively. Assuming that the distances between the BS and users are ordered as $d_1 \geq d_2 \geq \cdots \geq d_u \geq \cdots \geq d_U$, where d_u represents the distance between the u-th user and the BS. We assume that the total bandwidth is B. In addition, we assume that the bandwidth for each channel is equal.

Hence, the bandwidth for the c-th channel can be represented as $B_c = B/C$. Let a binary variable $\chi_u^c \in \{0, 1\}$ indicate if the u-th user is allocated at the c-th channel. If the u-th user is allocated at the c-th channel, $\chi_u^c = 1$; otherwise $\chi_u^c = 0$. Accordingly, $\sum_{u=1}^{U}\sum_{c=1}^{C}\chi_u^c = U$, $\sum_{u=1}^{U}\chi_u^c = M_c$, and $\sum_{c=1}^{C}\chi_u^c = 1$, where M_c represents the number of users multiplexed in the c-th channel.

The NOMA symbol x_c on the c-th channel is defined as follows [37]:

$$x_c = \sum_{u=1}^{U}\chi_u^c(\sqrt{p_u}b_u), \tag{1}$$

where b_u denotes the transmission symbols of the u-th user, and p_u is the transmission power of the u-th user. At the receiver side, the multiplexed signals of the u-th user assigned in the c-th channel can be expressed as follows:

$$y_u = \sum_{c=1}^{C}\chi_u^c(h_u x_c + z_c), \tag{2}$$

where h_u represents the channel response between the BS and the u-th user, and z_c is the noise. This paper assumes that z_c is the additive white Gaussian noise (AWGN) with variance $\sigma_{z_c}^2$ and zero mean. h_u is calculated as follows:

$$|h_u|^2 = (\hbar_u)^2 d_u^{-\lambda}, \tag{3}$$

where $\hbar_u$ is a coefficient that follows a Rayleigh distribution, $d_u^{-\lambda}$ represents the pathloss between the u-th user and BS, and λ denotes the pathloss exponent.

According to the SIC principle, when the SIC is operating exhaustively, the signals of the user with a stronger power are decoded perfectly by considering the signals from the other users as noise. In addition, the signals of the user with a weaker power are completely detected by discarding the signals with a stronger power than themselves using its replica. Therefore, when performing NOMA, the data rate of the u-th user can be expressed as

$$R_u^{NOMA} = \sum_{c=1}^{C}\chi_u^c B_c \log_2\left(1 + \frac{p_u|h_u|^2}{\sum_{i=u+1}^{U}\chi_i^c p_i|h_u|^2 + \sigma_{z_c}^2}\right). \tag{4}$$

If the data rate obtained by each user in OMA is greater than that obtained in NOMA, the better option is to use OMA for the user. OMA divides the channels used by NOMA using the number of users. Thus, the data rate for the u-th user obtained using OMA is expressed as follows:

$$R_u^{OMA} = \sum_{c=1}^{C}\chi_u^c \frac{B_c}{M_c}\log_2\left(1 + \frac{p_u|h_u|^2}{\sigma_{z_c}^2}\right). \tag{5}$$

Because switching between NOMA and OMA is performed according to the data rate obtained, the data rate that can be obtained by the u-th user, from (4) and (5), is obtained as follows:

$$R_u = \begin{cases} \sum_{c=1}^{C}\chi_u^c B_c \log_2\left(1 + \frac{p_u|h_u|^2}{\sum_{i=u+1}^{U}\chi_i^c p_i|h_u|^2 + \sigma_{z_c}^2}\right), \\ \quad \text{if } \sum_{c=1}^{C}\chi_u^c R_u^{NOMA} \geq \sum_{c=1}^{C}\chi_u^c R_u^{OMA} \\ \sum_{c=1}^{C}\chi_u^c \frac{B_c}{M_c}\log_2\left(1 + \frac{p_u|h_u|^2}{\sigma_{z_c}^2}\right), \text{otherwise} \end{cases}. \tag{6}$$

In this paper, the objective is to maximize the total throughput of the NOMA system. The total throughput of the NOMA system can be expressed as $\sum_{u=1}^{U} R_u$; hence, the objective function of this paper is expressed as follows:

$$\max_{\chi_u^c}\sum_{u=1}^{U} R_u. \tag{7}$$

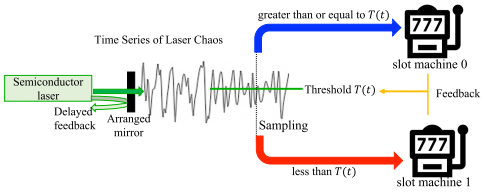


FIGURE 2. Principle of laser chaos decision-maker for the two-armed bandit problem [24].

III. PRINCIPLE OF LASER CHAOS DECISION-MAKER

In this section, we explain the principle of the laser chaos decision-maker, an ultrafast reinforcement learning scheme based on the chaotic oscillatory dynamics of lasers called laser chaos [24], [25].

Laser chaos is the phenomenon when the output of a semiconductor laser exhibits chaotic behavior. When a portion of the laser's optical output is delayed for a certain time via an externally placed mirror and returned to its cavity, instability is induced in the laser, causing it to exhibit chaotic oscillation behavior [24]. Figure 2 shows the decision-making based on a time series of laser chaos produced by a semiconductor laser in the two-armed bandit problem [26]. The principle of laser chaos decision-maker is summarized as follows. The slot machine selection is decided by comparing the sampled amplitude level of the laser chaos time series with the threshold value. Here, we consider the two-armed bandit problem, where the challenge is to quickly and accurately determine the slot machine of two with the higher reward probability. The two slot machines are called machines 0 and 1. If the sampled amplitude of the laser chaos time series is less than the threshold value, slot machine 1 is selected. Otherwise, the decision is to select slot machine 0. The threshold is adjusted depending on the results of playing the selected slot machine (in other words, win or lose); thus, the slot machine with a higher reward probability is selected in the subsequent trials.

More precisely, the threshold value $T(t)$ at step t, which is used in comparison with the amplitude of laser chaos signal, is given as follows:

$$T(t) = k \times \lfloor TA(t) \rfloor, \tag{8}$$

where $TA(t)$ is the threshold adjuster value at step t, $\lfloor TA(t) \rfloor$ is the nearest integer to $TA(t)$ rounded to zero, and k is a fixed value determining the range of the resultant $T(t)$. The value of $\lfloor TA(t) \rfloor$ can be one of the values in $-Z \leq \lfloor TA(t) \rfloor \leq Z$, where Z is a natural number. Thus, the total number of possible threshold levels is $2Z + 1$. Possible values for $T(t)$ are in the range of $-kZ \leq T(t) \leq kZ$ because of setting $\lfloor TA(t) \rfloor = Z$ when $\lfloor TA(t) \rfloor$ is greater than Z, as well as $\lfloor TA(t) \rfloor = -Z$ when $\lfloor TA(t) \rfloor$ is less than $-Z$. If the selected slot machine yields a reward or not reward at cycle t (in other words, wins or loses the slot machine play), the TA value is updated at cycle $t + 1$ based on the following [24] and [25]:

$$TA(t+1) = -\Delta + \alpha TA(t), \text{ if selected slot machine 0 wins,} \tag{9}$$

$$TA(t+1) = +\Delta + \alpha TA(t), \text{ if selected slot machine 1 wins,} \tag{10}$$

$$TA(t+1) = +\Omega + \alpha TA(t), \text{ if selected slot machine 0 loses,} \tag{11}$$

$$TA(t+1) = -\Omega + \alpha TA(t), \text{ if selected slot machine 1 loses} \tag{12}$$

where α is a forgetting rate to reduce the influence of the past results of decisions, which is in the range of $(0 \leq \alpha \leq 1)$, Δ is the constant increment, and Ω is the increment parameter. If the selected slot machine wins, i.e., if the result of playing the selected slot machine is a reward, the threshold adjustment value $TA(t)$ is varied based on $\mp\Delta + \alpha TA(t)$. Otherwise, the selected slot machine loses, i.e., the result of playing the selected slot machine is not a reward, the threshold adjustment value $TA(t)$ is updated by $\pm\Omega + \alpha TA(t)$, where Ω is the increment parameter based on the history of choices and rewards and penalties. We define J_i and L_i as the number of times that slot machine i was selected and that of wins by playing the selected slot machine i until step t, respectively. Here, let the estimated reward probability, or winning probability of the slot machine i be $\hat{P}_i$, and it is given by

$$\hat{P}_i = \frac{L_i}{J_i}. \tag{13}$$

Considering the two-armed bandit problem, using the estimated winning probability of (13), we donate Ω as follows:

$$\Omega = \frac{\hat{P}_0 + \hat{P}_1}{2 - (\hat{P}_0 + \hat{P}_1)}, \tag{14}$$

where $\hat{P}_0$ and $\hat{P}_1$ are the estimated reward probabilities of the two slot machines, respectively.

Scalability of the laser chaos decision-maker is achieved by cascading the above algorithms [25]. By lining up the thresholds in the pipeline, the laser chaos decision-maker can select one of several slot machines. Figure 3 shows the scalable decision-making using laser chaos time series, e.g., the eight-armed bandit problem. Specifically, scalable decision-making is conducted according to the following rules. At step t_1, the sampled chaotic signal $s(t_1)$ compares with the threshold TH_1. At step t_2, the sampled chaotic signal $s(t_2)$ compares with the threshold $TH_{2,0}$, if $s(t_1)$ is greater than or equal to TH_1. Otherwise, $s(t_2)$ compares with the threshold $TH_{2,1}$. Similarly, at step t_3, $s(t_3)$ compares with $TH_{3,0,0}$, $TH_{3,0,1}$, $TH_{3,1,0}$, or $TH_{3,1,1}$. In this rule, the comparison with the threshold value and laser chaos is terminated when it selects one slot machine. Thus, the laser chaos decision-maker can make a scalable decision.

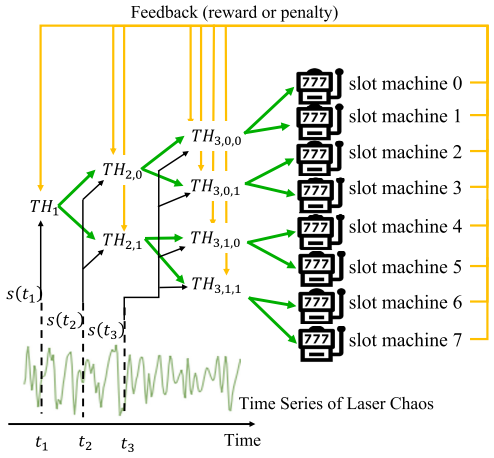


FIGURE 3. Scalable decision-making by a hierarchical architecture in the case of the 8-armed bandit problem [25].

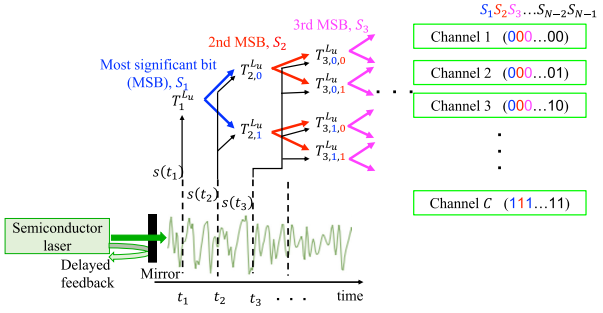


FIGURE 4. Selection of channel to allocate u-th user in L_u.

IV. PARALLEL BANDIT ARCHITECTURE FOR SCALABLE CHANNEL ALLOCATION IN NOMA SYSTEMS

This section describes the proposed scalable channel allocation scheme based on a parallel array of laser chaos decision-makers. The number of laser chaos decision-makers in the BS equals the number of users U. For simplicity, we assume that U is given by 2^N, where N is a natural number. In addition, we assume that the number of channels is half the number of users, i.e., the number of channels C is $C = 2^{N-1}$. Each laser chaos decision-maker selects a channel for a distinct user.

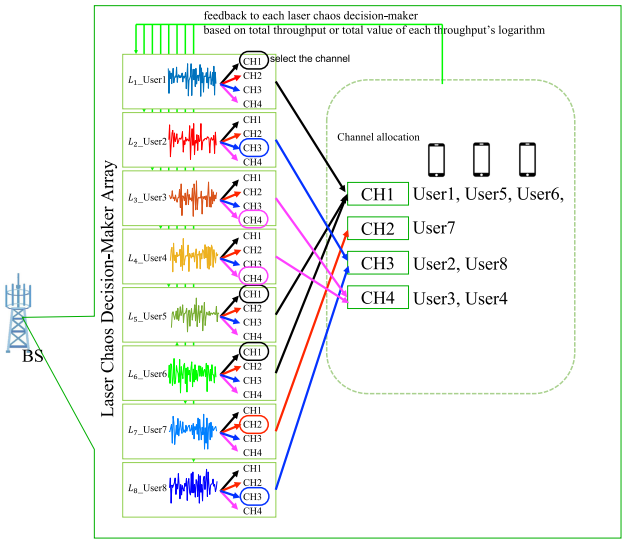


FIGURE 5. Channel allocation by parallel array of laser chaos decision-maker in the case of 8 users.

The proposed method operates with a reinforcement learning approach that maximizes the total data rate of all users by repeating the following process.

1) Select the channel to be assigned to each user by the corresponding laser chaos decision-maker.
2) Evaluate the total data rate of all users based on (7), which corresponds to the reward of reinforcement learning.
3) Adjust the threshold of each laser chaos decision-maker based on the calculated results or reward.

Let the set of laser chaos decision-makers be $\mathbf{L} = \{L_1, L_2, \cdots, L_u, \cdots, L_U\}$, and each can solve the C $(= 2^{N-1})$-armed bandit problem. L_u is associated with the u-th user. Each arm symbolizes the channel used for communication. To elaborate, if L_u selects the c-th channel, the u-th user is allocated in the c-th channel (or $\chi_u^c = 1$).

To describe the behavior of each laser chaos decision-maker, we focus on the u-th laser chaos decision-maker L_u shown in Fig. 4. L_u selects one channel from $\mathbf{C}$ based on the following principle. The C channels are distinguished by the index given by ranging from 0 to $C-1$, which are also represented in an $N-1$-bit binary code given by $S_1S_2\cdots S_{N-1}$, where S_j $(j = 1, 2, \cdots, N-1)$ is 0 or 1. For example, when $C = 8$ (or $2^N = 16, N = 4$), the channels are numbered as $S_1S_2S_3 = 000, 001, 010, \cdots, 111$. S_1 is the most significant bit (MSB), whereas S_{N-1} is the least significant bit (LSB). Thus, the identity of the channel to be selected is determined bit by bit from S_1 (MSB) to S_{N-1} (LSB) in a pipelined manner.

The decision is made for each of the bits based on a comparison between the measured signal level of laser chaos and the designed threshold value. First, the laser chaos signal $s(t_1)$ measured at $t = t_1$ is compared with a threshold value denoted as $T_1^{L_u}$. The output of this comparison is the decision of the MSB concerning the channel to select. If the time series is less than the threshold value, the output bit of the decision is 1, which is donated as $S_1 = 1$ (MSB $= 1$). Otherwise, the output will be 0 ($S_1 = 0$) [24], [25]. Subsequently, we determine the second MSB based on the determination of the first MSB. When the MSB is 1 ($S_1 = 1$), the laser chaos signal $s(t_2)$ is subject to the threshold value $T_{2,1}^{L_u}$. The "2" to the left of the subscript indicates that this threshold relates to the second MSB, and the "1" to the right in the subscript represents that the one previous determination is 1 ($S_1 = 1$). If $s(t_2)$ is less than the threshold $T_{2,1}^{L_u}$, then S_2, which is the second MSB, is 1; otherwise, it is 0 ($S_2 = 0$). Finally, we should decide on the LSB. Based on the aforementioned manners, the comparison between thresholds and laser chaos signals is terminated when all $N-1$ bits are designated, or the selected channel is determined. When $U = 16\ (= 2^4)$, $C = 8\ (= 2^3)$, and $S_1S_2S_3 = 101$, then L_u has selected the fifth channel, i.e., the ten hex of 101, and the u-th user is allocated to the fifth channel.

The formula for updating the threshold adjuster value (TA) of L_u is given as follows:

$$TA^{L_u}_{K,S_1,S_2,\ldots,S_{(K-1)}}(t+1) = -\Delta + \alpha TA^{L_u}_{K,S_1,S_2,\ldots,S_{(K-1)}}(t), \quad (15)$$

$$TA^{L_u}_{K,S_1,S_2,\ldots,S_{(K-1)}}(t+1) = +\Delta + \alpha TA^{L_u}_{K,S_1,S_2,\ldots,S_{(K-1)}}(t), \quad (16)$$

$$TA^{L_u}_{K,S_1,S_2,\ldots,S_{(K-1)}}(t+1) = -\Omega_{K,S_1,S_2,\ldots,S_{(K-1)}} + \alpha TA^{L_u}_{K,S_1,S_2,\ldots,S_{(K-1)}}(t), \quad (17)$$

$$TA^{L_u}_{K,S_1,S_2,\ldots,S_{(K-1)}}(t+1) = +\Omega_{K,S_1,S_2,\ldots,S_{(K-1)}} + \alpha TA^{L_u}_{K,S_1,S_2,\ldots,S_{(K-1)}}(t), \quad (18)$$

where Δ is a constant, and $\Omega_{K,S_1,S_2,\ldots,S_{(K-1)}}$ can be calculated based on (14). K corresponds to the K-th MSB. K is $1 \le K \le N-1$, and K_0 is not described. $TA^{L_u}_{K,S_1,S_2,\ldots,S_{(K-1)}}(t+1)$ represents the threshold adjustment value, which is used to set the threshold for comparison with the laser chaos amplitude to confirm the K-th MSB with the decision $S_1S_2\ldots S_{(K-1)}$. Equations (15), (16), (17), and (18) are formulated based on (9), (10), (11), and (12), respectively. The update of the $TA^{L_u}_{K,S_1,S_2,\ldots,S_{(K-1)}}(t+1)$ is rewarded or non-rewarded based on the output of the L_u decision maker, i.e., the results of the channel allocation.

Here, we define the conditions of reward in the proposed method. At each step t, the total data rate of users is calculated, which decides whether to reward or punish based on it (equivalent to a win or loss on a slot machine). Specifically, the average of the total data rate of the most recent τ times is compared with the total data rate in step t to determine if a reward has been earned. At step t, the average of the total data rate in the last τ times can be

expressed as follows:

$$R^{\tau}_{ave}(t) = \sum_{\theta=t-\tau}^{t-1} \frac{\sum_{u=1}^{U} R_u(\theta)}{\tau}, \tag{19}$$

Accordingly, based on (19), if $\sum_{u=1}^{U} R_u(t) \geq R^{\tau}_{ave}(t)$, all laser chaos decision-makers are rewarded; otherwise, they do not obtain any reward. Here, $R_u(t)$ denotes the data rate of the u-th user at step t.

In addition, to account for fairness, the total value of the data rate of the logarithm of each user is compared one with the last τ times. It is expressed as follows:

$$R^{\tau}_{log,ave}(t) = \sum_{\theta=t-\tau}^{t-1} \frac{\sum_{u=1}^{U} \log_{\xi} R_u(\theta)}{\tau}, \tag{20}$$

where ξ is the base of the logarithm.

Without the logarithmic utility function, i.e., (19), the total data rate will be large if the data rate of some users is very large, even if the data rate of some users is small. With the logarithmic utility function, i.e., (20), from the nature of the logarithm function, the difference between users with large data rate and those with small data rate becomes smaller than when the logarithm is not taken. Moreover, we consider that if users with a small data rate exist, the total logged data rate will not be large, even if some users have a very large data rate. Therefore, to increase the sum of logged data rates, we must increase the data rate of users whose data rate is small. Thus, we believe that taking the logarithm of each user's data rate enables us to consider fairness among users.

For clarity, Fig. 5 provides an example of channel allocation on BS in a NOMA system with eight users. Four channels are available. In Fig. 5, User1 is allocated to CH1 because L_1 selects CH1. Similarly, other users are assigned to the channel selected by their respective laser chaos decision-makers, which are associated with each user. After completing the channel allocation of all users, the BS calculates the achieved data rate of each user. If the calculation is based on (19), the results of the calculations are summed to determine the reward. If the calculation is based on (20), the logarithm of the calculation results is taken and then summed to determine the reward. After feedback is provided to each laser chaos decision-maker, and when each decision-maker has updated its threshold value, the channel assignment process is repeated. This procedure is repeated for a certain number of iterations.

Moreover, a fixed power allocation approach is used because the BS lacks detailed user information. In our fixed power allocation method, users are assigned a fixed power level in a particular channel according to their distance, with the closest user receiving the highest power. The power allocation assignments can be determined as follows:

$$p_u = \begin{cases} \sum_{c=1}^{C} \chi_u^c (1-a_c) P_c \times 10^{-(\xi_u - 1)}, \text{ if } \xi_u < M_c \\ \sum_{c=1}^{C} \chi_u^c a_c P_c \times 10^{-(M_c - 1)}, \text{ if } \xi_u = M_c, \end{cases} \tag{21}$$

where $\xi_u = \sum_{n=1}^{u} \chi_n^c$, which indicates the order of the long distance between the u-th user and BS among that of all users assigned to the c-th channel. P_c and a_c donate the transmit power assigned to and the power allocation factor of the c-th channel, respectively.

V. SIMULATION RESULTS

This section evaluates the performance of our proposed method in terms of total data rate and fairness using MATLAB R2022a. The sampled laser chaos time series data generated by the practical semiconductor laser chaos decision-maker are used to make decisions in the simulation. The laser chaos time series used in this simulation are sampled at 100 Gsample/s (10 ps sampling interval) with 8-bit resolution, and signal levels are integer values from -127 to 128 [25]. The laser chaos time series data is sampled at suitable interval, i.e., 50 ps [25]. The number of the generated laser chaos time series equals to that of the users. In summary, the implementation steps are described as follows. The laser chaos time series for each user are generated by the practical semiconductor laser that is the same one detailed in the schematic diagram Fig. 1-c of reference [25]. The details about the practical semiconductor laser also can be found in [24], [26], [35], [38], [39], and [40]. Then, performance is evaluated based on the generated laser chaos time series using MATLAB R2022a. In the simulation, the access channel for each user is selected by comparing the threshold and the value of the sampled laser chaos time series.

We compare the performance of our method to three existing techniques: MD-NOMA [41], C-NOMA [13], and UCGD-NOMA [14]. The procedure of MD-NOMA is summarized as follows: BS computes the pairing distance threshold $d_{Threshold}$ according to $d_U(\frac{u}{v(\beta_1-\beta_2 u)-u})^{-\lambda}$, where d_U is the distance between the BS and the user that is farthest from the BS. β_1 and β_2 are the power allocation coefficients for the paired users. $u = 2^{\frac{R^1_{min}}{B}} - 1$ while $v = 2^{\frac{R^2_{min}}{B}} - 1$. R^1_{min} and R^2_{min} are the minimum required throughput of the paired users. According to the distance threshold, BS distinguishes users farther than $d_{Threshold}$ into the far user group and others into the near user group. The nearest user from the BS in the far user group is paired with the user in the near user group that is nearest from BS. Paired users perform NOMA. When all users in the far user group or in the near group have been paired, the remaining users in the other group perform OMA. In the simulation, we considered a circular cell with $U = 2^N$ users, where U represents the total number of users and N is a natural number. In addition, we set the number of channels to $C = 2^{N-1}$, as defined in Section II. Thus, the number of users was twice the number of channels. Furthermore, we set the minimum distance between a user and BS to 10 m. The pathloss exponent λ was set to 3. The noise power spectral density and total bandwidth were set as -170 dBm and 20 MHz, respectively. The power allocation is fixed. The total transmission power for the users allocated to the same channel was set to 20 dBm regardless of the number of users

TABLE 3. Parameter settings.

Parameter	Value
Number of users	$U = 2^N$
Number of channels	$C = 2^{N-1}$
Cell shape	Circular
Minimal distance between user and BS	10 m
Transmission power allocated to the c-th channel	$P_c = 20$ dBm, $\forall c$
Power allocation factor for the c-th channel	$a_c = 0.1, \forall c$
Pathloss exponent	$\lambda = 3.0$
Noise power spectral density	-170 dBm
Total bandwidth	$B = 20$ MHz
Constant increment Δ	$\Delta = 1.0$
Base of the logarithm	$\xi = 10$
Number of runs	100

in each channel. The power allocation factor was $a_c = 0.1$ for all c ($c = 1, 2, \cdots, 2^{N-1}$). The Δ related to laser chaos decision-maker was set to 1.0. The base of the logarithm ξ was set to 10. The simulation results of the total data rate and fairness described in this section were the average values of the 100 runs. The common parameter settings in the performance evaluation are presented in Table 3.

In this study, we used the Jain's index J to evaluate the fairness of the users which is defined by [42]

$$J = \frac{(\sum_{u=1}^{U} R_u)^2}{U \cdot \sum_{u=1}^{U} (R_u)^2}. \tag{22}$$

In addition, note that we denote the simulation results based on (19) and (20) as *Proposed (1)* and *Proposed (2)*, respectively.

1) CONVERGENCE OF THE PROPOSED METHOD

To assess the convergence of the proposed method, Figs. 6 and 7 present the simulation results changing with the number of iterations. Figs. 6a and 7a show the channels selected for each iteration using Proposed (1) and Proposed (2), respectively. The blue points in Fig. 6b represent the total data rate obtained using Proposed (1), whereas the orange points in Fig. 7b represent the total value of logarithmic data rate achieved using Proposed (2). The numbers of users and channels in the simulation were 8 and 4, respectively. The cell radius and pathloss exponent were set to 500 m and 3.0, respectively. The thresholds Z and the forgetting rate α were set to $Z = 128$ and $\alpha = 0.999$, respectively. Fig. 6a shows the convergence of channel allocation using Proposed (1) and Proposed (2), and Fig. 6b shows that the obtained total data rate, which also converges as a result of the convergence of the channel allocation. Proposed (1) is based on (19), which determines rewards or non-rewards based on the value of the total data rate. Therefore, Fig. 6b showcases the evolution of the total data rate achieved through channel allocation. As observed in the figure, the channel allocation is carried out to maximize the reward, leading to an increase in the total data rate.

Similarly, Fig. 7a shows the convergence of channel allocation using Proposed (2), and Fig. 7b shows that the total value of logarithmic data rate also converges as a result of the convergence of the channel allocation. Proposed (2) is based on (20), which determines rewards or non-rewards based on the total value of the logarithmic data rate. Thus, Fig. 7b depicts the evolution of the total value of the logarithmic data rate achieved through channel assignment. Fig. 7b shows that the channel allocation is performed to maximize the reward, leading to an increase in the total value of the logarithmic data rate. From the aforementioned observations, we concluded that Proposed (1) and Proposed (2) have been functioning properly.

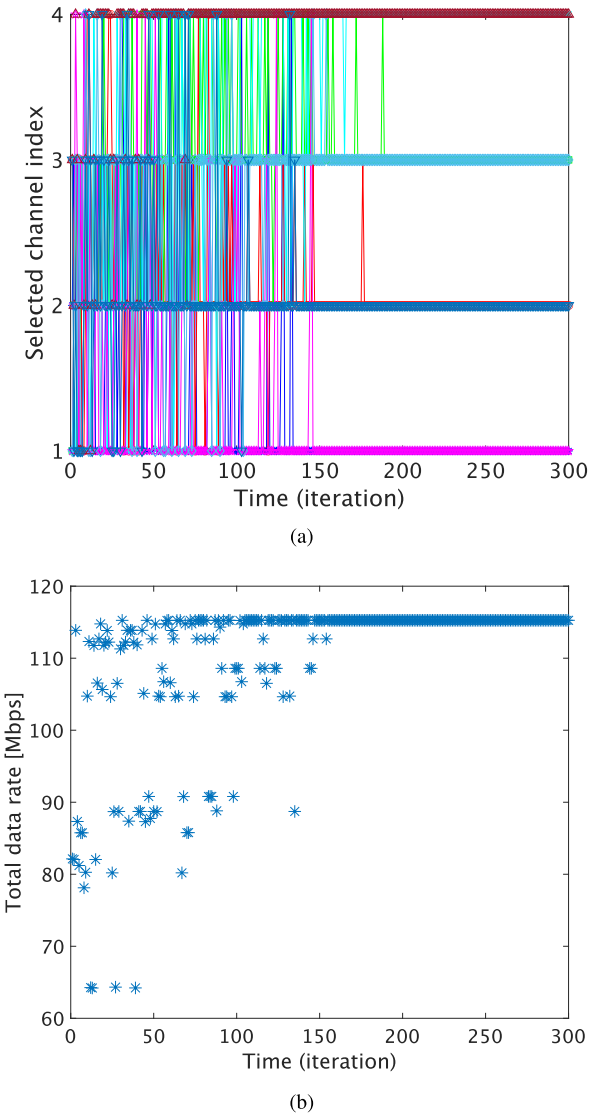


FIGURE 6. Convergence of Proposed (1). (a) Selected channel index in each iteration of Proposed (1). (b) Total data rate in each iteration using Proposed (1).

2) Z DEPENDENCIES

Next, we assess the total data rate of Proposed (2) while varying the parameter Z, which determines the number of thresholds in the laser-chaos decision maker. In the simulation, we set Z to 1, 2, 4, 8, 16, 32, 64, and 128 to

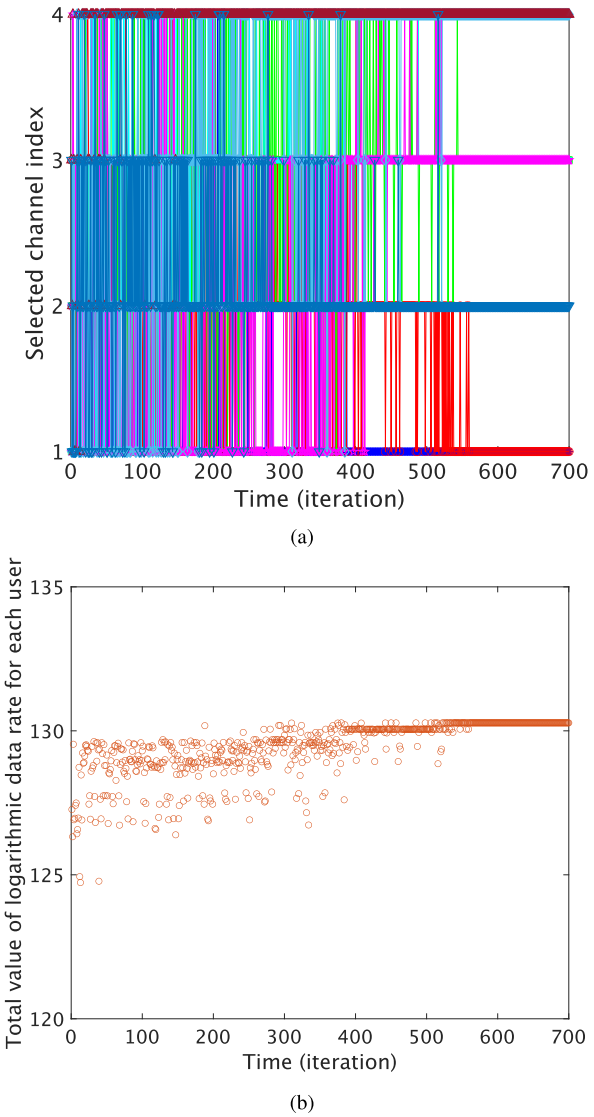


FIGURE 7. Convergence of Proposed (2). (a) Selected channel index in each iteration of Proposed (2). (b) Total data rate in each iteration using Proposed (2).

analyze its impact on the total data rate. The number of users, cell radius, and the number of iterations were 64, 1000 m, and 60000, respectively. The forgetting rate α was set to $\alpha = 0.999$. Fig. 8 presents the simulation results. The blue bars represent the total data rate achieved using Proposed (2), and the orange line represents Jain's index. These metrics were used to assess the performance of the system and analyze the effectiveness of Proposed (2).

Fig. 8 shows that the best fairness was achieved when Z was 128 and the number of iterations was sufficient. In the laser chaos decision-maker, when the number of available thresholds is small, the thresholds can reach the upper or lower limits more quickly. Thus, selection convergence is generally faster, but accurate selection becomes more difficult. Conversely, if the number of thresholds is large, a threshold may reach its upper or lower limit after a sufficient search. Thus, convergence takes a longer time, but correct

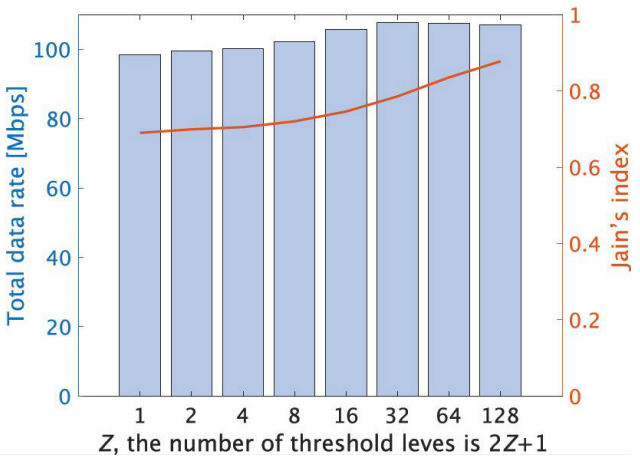


FIGURE 8. Total data rate and Jain's index when the number of threshold levels increased using Proposed (2).

selection may be highly achievable [26]. In this simulation, we set the same number of iterations regardless of the Z value; thus, the performance was considered to be better when Z was 128 (maximum). Fig. 8 shows that the optimal performance in terms of both total data rate and fairness was attained at $Z = 128$. Hence, the figures after Fig. 8 show results for a setting of $Z = 128$.

3) FORGETTING PARAMETER DEPENDENCIES

Fig. 9 depicted the evolution of total data rate and fairness when the number of users, cell radius, and number of iterations were 32, 500 m, and 50000, respectively, using Proposed (2). From (15)(16)(17)(18), if α is set smaller than 0.999, the fluctuation of threshold update becomes smaller, learning speed becomes slower, and the speed of solution cannot match the scale of the problem, increasing number of iterations required for learning and leading to delays. For the case of $\alpha = 0.99,\ 0.9$, the 50000 iterations we set did not converge, and performance deteriorated rapidly, as shown in Fig. 9 because of the small width of the threshold variation. In addition, Fig. 9 shows that the case of $\alpha = 0.999$ was the best in both total data rate and fairness. Therefore, the setting of $\alpha = 0.999$ was adopted when compared with other methods.

4) SMALL AREA, SMALL PATHLOSS EXPONENT

Fig. 10 shows the evolution of total data rate and fairness with the increasing number of users when the cell radius and pathloss exponent λ were 500 m and 3.0, respectively. For Proposed (1), the number of iterations was set to 1000, 2000, 5000, and 20000 when the number of users was 8, 16, 32, and 64, respectively. For Proposed (2), the number of iterations was set to 2000, 5000, 20000, and 50000 when the number of users was 8, 16, 32, and 64, respectively. As shown in Fig. 10, Proposed (1) had the highest total data rate for any number of users, but its fairness is not the best. In contrast, Proposed (2) has a reasonable total data rate, and its fairness

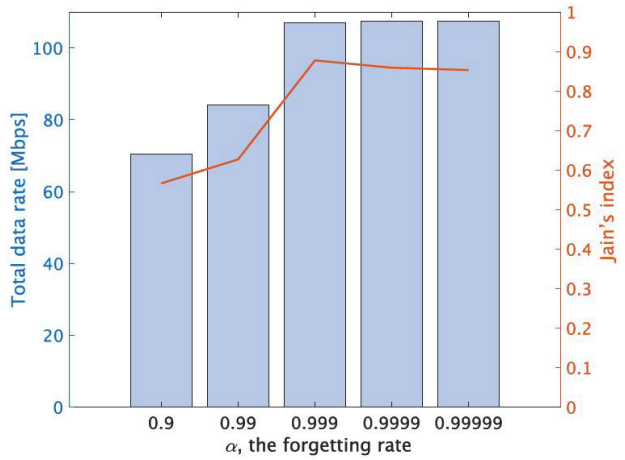


FIGURE 9. Total data rate and Jain's index when the forgetting rate α increased using Proposed (2).

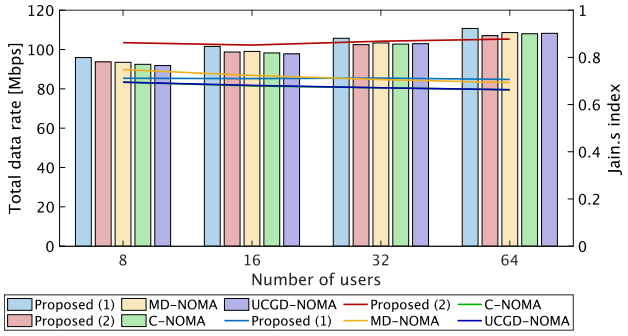


FIGURE 10. Total data rate and Jain's index obtained for the number of users when the cell radius and pathloss exponent λ were 500 m and 3.0, respectively.

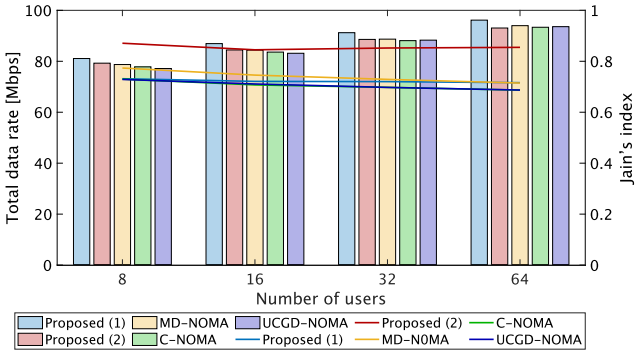


FIGURE 11. Total data rate and Jain's index obtained for the number of users when the cell radius and pathloss exponent λ were 1000 m and 3.0, respectively.

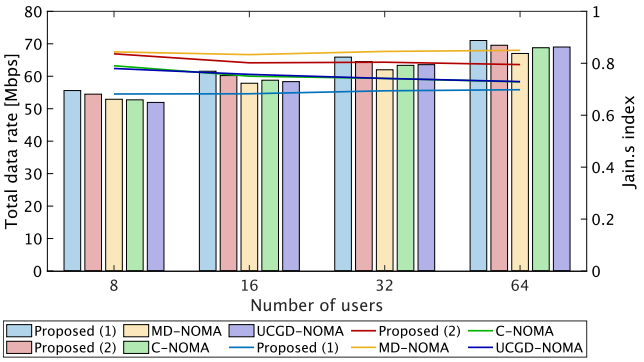


FIGURE 12. Total data rate and Jain's index obtained for the number of users when the cell radius and pathloss exponent λ were 500 m and 4.0, respectively.

was better than those of the other methods for any number of users. This is because the log-taken value was larger under the situation where all users had a similar overall data rate than the situation where some users had a low data rate while some had a very high data rate.

5) LARGE AREA, SMALL PATHLOSS EXPONENT

Fig. 11 shows the evolution of total data rate and fairness as the number of users increased, considering a cell radius of 1000 m and a pathloss exponent of 3.0. The number of iterations was set the same as the case of a cell radius of 500 m and pathloss exponent of 3.0. Figs. 10 and 11 show that Proposed (1) had the highest total data rate for any number of users, but its fairness is not the highest. Conversely, Proposed (2) had the best fairness and a reasonably high total data rate.

6) SMALL AREA, LARGE PATHLOSS EXPONENT

Fig. 12 shows the evolution of the total data rate and fairness as the number of users increased, considering a cell radius of 500 m and pathloss exponent of 4.0. The number of iterations was set the same as the case of a cell radius of 500 m and pathloss exponent of 3.0. Fig. 12 shows that Proposed (1) had the highest total data rate for any number of users, but its fairness was poor because the pathloss was more influential than before, probably because it was difficult to maintain fairness while increasing the total data rate with the power allocation of this study. The fairness of Proposed (2) also became poor, but it was greater than that of Proposed (1). The total data rate of Proposed (2) was higher than that of MD-NOMA.

7) LARGE AREA, LARGE PATHLOSS EXPONENT

Fig. 13 shows the evolution of the total data rate and fairness as the number of users increased, considering a cell radius of 1000 m and pathloss exponent of 4.0. The number of iterations was set the same as the case of a cell radius of 500 m and pathloss exponent of 3.0. Fig. 13 shows the same conclusions as in Fig. 12.

8) PATHLOSS DEPENDENCIES

Fig. 14 shows the evaluation of the total data rate and Jain's index as the pathloss exponent increased while maintaining the number of users and cell radius at 64 and 500 m, respectively. The number of iterations in Proposed (1) and Proposed (2) was set to 5000 and 20000, respectively. Proposed (1) outperformed other schemes in terms of total data rate. However, Proposed (1) was not as effective in achieving fairness. In contrast, Proposed (2) exhibited

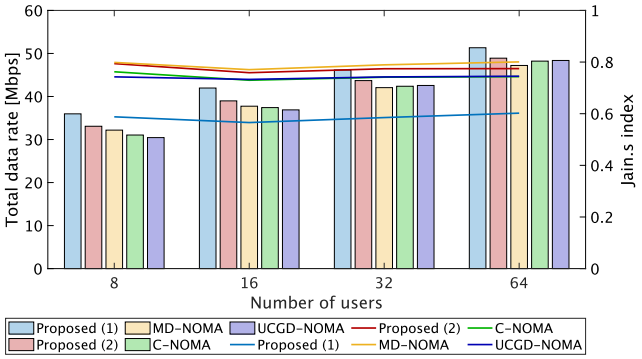


FIGURE 13. Total data rate and Jain's index obtained for the number of users when the cell radius and pathloss exponent λ were 1000 m and 4.0, respectively.

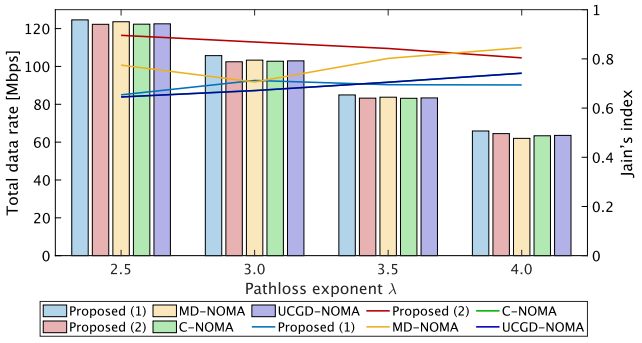


FIGURE 14. Total data rate and Jain's index obtained for the pathloss exponent when the number of users and cell radius were 32 and 500 m.

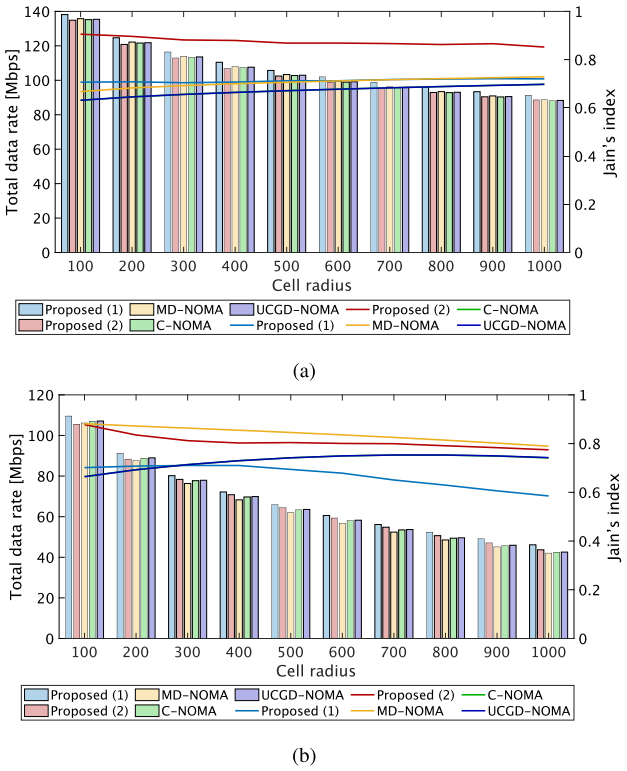


FIGURE 15. Total data rate and Jain's index obtained for the cell radius. (a) Number of users and pathloss exponent λ were 32 and 3.0. (b) Number of users and pathloss exponent λ were 32 and 4.0.

promising results in terms of fairness, particularly when the pathloss exponent was not large. For larger pathloss exponents, specifically when the pathloss exponent was 4.0, the fairness performance of Proposed (2) was inferior to that of MD-NOMA. However, the total data rate performance of Proposed (2) surpassed that of MD-NOMA. When the pathloss exponent was large, the power attenuation increased, which posed challenges in maintaining fairness. As a result, the expected improvement in fairness with Proposed (2) was poorer than that with a small pathloss exponent.

9) CELL RADIUS DEPENDENCIES

Fig. 15a shows the evaluation of the total data rate and Jain's index while maintaining the number of users and pathloss exponent at 32 and 3.0, respectively, as the cell radius expanded. The number of iterations in Proposed (1) and Proposed (2) was set to 5000 and 20000, respectively. Proposed (1) had the highest total data rate for any cell radius, but its fairness was poorer than MD-NOMA when the cell radius was larger than 600 m. In contrast, Proposed (2) had the highest fairness for any cell radius.

Fig. 15b shows the evaluation of the total data rate and fairness when the number of users and pahtloss exponent were at 32 and 4.0, respectively, while the cell radius expanded. The number of iterations in Proposed (1) and Proposed (2) was set to 5000 and 20000, respectively. Proposed (1) had the best performance in terms of the data rate at all cell radii. However, the fairness of Proposed (1) decreased as the cell radius expanded. In contrast, Proposed (2) attained a reasonable total data rate and fairness. Specifically, the total data rate of Proposed (2) surpassed that of MD-NOMA, and its fairness was inferior to MD-NOMA but superior to those of other methods. These were caused by the difficulty in increasing the total data rate while maintaining fairness owing to the increased impact of path loss.

Based on the aforementioned findings, we can conclude that Proposed (1) is effective when prioritizing the total data rate over fairness, even if it means compromising fairness to some extent. This approach is suitable when the total data rate must be maximized regardless of fairness considerations. In contrast, Proposed (2) is effective in scenarios where increasing the total data rate is important, but fairness must be considered. This approach balances between total data rate improvements and maintaining a certain level of fairness among users.

VI. CONCLUSION

In this paper, we proposed and demonstrated a scalable channel allocation principle for NOMA systems based on the parallel array of laser chaos decision-makers. The computational complexity associated with the combinatorial explosion in channel allocation is reduced significantly.

In other words, by assigning a dedicated laser chaos decision-maker to each user and linking the reward to the total data rate or aggregated logarithmic data rate of all users, the scalability problem of managing an increasing number of users, as observed in previous studies, is effectively resolved. In addition, the use of logarithms in terms of compensation was considered to improve fairness. Simulation results showed that the proposed algorithm achieves greater total throughput and better fairness than conventional NOMA algorithms. In future research, we plan to evaluate performance in several environments where users and terminals are moving and have multiple base stations. In addition, we will also incorporate new technologies to further enhance the performance of the NOMA system, such as the Stacked Intelligent Metasurface (SIM) technique [43]. The SIM offers substantial signal processing capabilities by stacking an array of multiple programmable nearly passive metasurfaces. The signal propagation in an SIM is at the speed of light. As such, SIM can accomplish advanced computation and signal processing tasks, such as MIMO, as the electromagnetic (EM) wave propagates through the multiple layers of the metasurface, which may further effectively reduce the processing delay and improve the total data rate for the NOMA system.

REFERENCES

[1] H. Yahya, A. Ahmed, E. Alsusa, A. Al-Dweik, and Z. Ding, "Error rate analysis of NOMA: Principles, survey and future directions," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1682–1727, 2023.

[2] X. Mu and Y. Liu, "Exploiting semantic communication for non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2563–2576, Aug. 2023.

[3] V. Andiappan and V. Ponnusamy, "Hierarchical decoding mechanism for interference cancellation in full-duplex intelligent reflecting surfaces aided non-orthogonal multiple access system," *IEEE Access*, vol. 11, pp. 74312–74323, 2023.

[4] L. Xu, H. Xing, Y. Deng, A. Nallanathan, and C. Zhuansun, "Fairness-aware throughput maximization for underlying cognitive NOMA networks," *Syst. J.*, vol. 15, no. 2, pp. 1881–1892, Jun. 2021.

[5] R. Liu, K. Guo, K. An, Y. Huang, F. Zhou, and S. Zhu, "Resource allocation for cognitive satellite-HAP-terrestrial networks with non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 9659–9663, Mar. 2023.

[6] M. Zhu, Y. Wang, X. Liu, S. Ma, X. Zhang, and Y. Fu, "Performance analysis for DF relay-aided visible light communication system with NOMA," *IEEE Photon. J.*, vol. 14, no. 5, pp. 1–9, Oct. 2022.

[7] Y. Wang, J. Wang, V. W. S. Wong, and X. You, "Effective throughput maximization of NOMA with practical modulations," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 4, pp. 1084–1100, Apr. 2022.

[8] P. Wang, N. Ye, J. Li, B. Di, and A. Wang, "Asynchronous multi-user detection for code-domain NOMA: Expectation propagation over 3D factor-graph," *IEEE Trans. Veh. Technol.*, vol. 71, no. 10, pp. 10770–10781, Oct. 2022.

[9] M. Salehi, H. Tabassum, and E. Hossain, "Meta distribution of SIR in large-scale uplink and downlink NOMA networks," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 3009–3025, Apr. 2019.

[10] W. U. Khan, M. A. Javed, T. N. Nguyen, S. Khan, and B. M. Elhalawany, "Energy-efficient resource allocation for 6G backscatter-enabled NOMA IoV networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 9775–9785, Jul. 2022.

[11] H. Al-Obiedollah, K. Cumanan, H. B. Salameh, G. Chen, Z. Ding, and O. A. Dobre, "Downlink multi-carrier NOMA with opportunistic bandwidth allocations," *IEEE Wireless Commun. Lett.*, vol. 10, no. 11, pp. 2426–2429, Nov. 2021.

[12] Z. Duan, A. Li, N. Okada, Y. Ito, N. Chauvet, M. Naruse, and M. Hasegawa, "User pairing using laser chaos decision maker for NOMA systems," *Nonlinear Theory Appl.*, vol. 13, no. 1, pp. 72–83, 2022.

[13] L. Zhu, J. Zhang, Z. Xiao, X. Cao, and D. O. Wu, "Optimal user pairing for downlink non-orthogonal multiple access (NOMA)," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 328–331, Apr. 2019.

[14] M. B. Shahab, M. Irfan, M. F. Kader, and S. Y. Shin, "User pairing schemes for capacity maximization in non-orthogonal multiple access systems," *Wireless Commun. Mobile Comput.*, vol. 16, no. 17, pp. 2884–2894, Dec. 2016.

[15] J. Zheng, X. Tang, X. Wei, H. Shen, and L. Zhao, "Channel assignment for hybrid NOMA systems with deep reinforcement learning," *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1370–1374, Jul. 2021.

[16] M. Abd-Elnaby, G. G. Sedhom, and M. Elwekeil, "Priority-based joint resource allocation with deep Q-learning for heterogeneous NOMA systems," *IEEE Trans. Commun.*, vol. 19, no. 12, pp. 41468–41481, Jun. 2021.

[17] H. Zhang, H. Zhang, K. Long, and G. K. Karagiannidis, "Deep learning based radio resource management in NOMA networks: User association, subchannel and power allocation," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 4, pp. 2406–2415, Oct. 2020.

[18] Y. Cao, G. Zhang, G. Li, and J. Zhang, "A deep Q-network based-resource allocation scheme for massive MIMO-NOMA," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1544–1548, May 2021.

[19] W. Xu, J. An, C. Huang, L. Gan, and C. Yuen, "Deep reinforcement learning based on location-aware imitation environment for RIS-aided mmWave MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 7, pp. 1493–1497, Jul. 2022.

[20] J.-H. Lee, J. Park, M. Bennis, and Y.-C. Ko, "Integrating LEO satellites and multi-UAV reinforcement learning for hybrid FSO/RF non-terrestrial networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 3647–3662, Mar. 2023.

[21] A. Li, "Deep reinforcement learning based resource allocation for LoRaWAN," in *Proc. IEEE 96th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2022, pp. 1–4.

[22] A. S. Kumar, L. Zhao, and X. Fernando, "Multi-agent deep reinforcement learning-empowered channel allocation in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1726–1736, Feb. 2022.

[23] K. Kitayama, "Novel frontier of photonics for data processing—Photonic accelerator," *APL Photonics.*, vol. 4, no. 19, pp. 1–24, Sep. 2019.

[24] M. Naruse, Y. Terashima, A. Uchida, and S.-J. Kim, "Ultrafast photonic reinforcement learning based on laser chaos," *Sci. Rep.*, vol. 7, no. 1, p. 8772, Aug. 2017.

[25] M. Naruse, T. Mihana, H. Hori, H. Saigo, K. Okamura, M. Hasegawa, and A. Uchida, "Scalable photonic reinforcement learning by time-division multiplexing of laser chaos," *Sci. Rep.*, vol. 8, no. 1, p. 10890, Jul. 2018.

[26] M. Naruse, N. Chauvet, A. Uchida, A. Drezet, G. Bachelier, S. Huant, and H. Hori, "Decision making photonics: Solving bandit problems using photons," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, pp. 1–10, Jan. 2020.

[27] N. Okada, M. Hasegawa, N. Chauvet, A. Li, and M. Naruse, "Theory of acceleration of decision-making by correlated time sequences," *Complexity*, vol. 19, no. 12, Jun. 2021, Art. no. 5205580.

[28] A. Li, M. Fujisawa, I. Urabe, R. Kitagawa, S.-J. Kim, and M. Hasegawa, "A lightweight decentralized reinforcement learning based channel selection approach for high-density LoRaWAN," in *Proc. IEEE Int. Symp. Dynamic Spectr. Access Netw. (DySPAN)* 2021, pp. 9–14.

[29] S. Hasegawa, R. Kitagawa, A. Li, S.-J. Kim, Y. Watanabe, Y. Shoji, and M. Hasegawa, "Multi-armed-bandit based channel selection algorithm for massive heterogeneous Internet of Things networks," *Appl. Sci.*, vol. 12, no. 15, p. 7424, Jul. 2022.

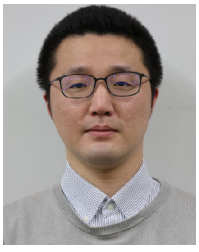
[30] S. Takeuchi, M. Hasegawa, K. Kanno, A. Uchida, N. Chauvet, and M. Naruse, "Dynamic channel selection in wireless communications via a multi-armed bandit algorithm using laser chaos time series," *Sci. Rep.*, vol. 10, no. 1, p. 1574, Jan. 2020.

[31] H. Kanemasa, A. Li, Y. Ito, N. Chauvet, M. Naruse, and M. Hasegawa, "Dynamic channel bonding in WLANs by hierarchical laser chaos decision maker," *Nonlinear Theory Appl.*, vol. 13, no. 1, pp. 84–100, 2022.

[32] M. Sugiyama, A. Li, Z. Duan, M. Naruse, and M. Hasegawa, "BER minimization by user pairing in downlink NOMA using laser chaos decision-maker," *Electronics*, vol. 11, no. 9, p. 1452, Apr. 2022.

[33] M. Sugiyama, A. Li, M. Naruse, and M. Hasegawa, "Scalable channel allocation in downlink NOMA using parallel array of laser chaos decision-maker," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Jan. 2023, pp. 461–466.
[34] T. Harayama, S. Sunada, and S. Shinohara, "Universal single-mode lasing in fully chaotic two-dimensional microcavity lasers under continuous-wave operation with large pumping power," *Photon. Res.*, vol. 5, no. 6, p. B39, Dec. 2017.
[35] R. Homma, S. Kochi, T. Niiyama, T. Mihana, Y. Mitsui, K. Kanno, A. Uchida, M. Naruse, and S. Sunada, "On-chip photonic decision maker using spontaneous mode switching in a ring laser," *Sci. Rep.*, vol. 9, no. 1, p. 9429, Jul. 2019.
[36] T. Otsuka, A. Li, H. Takesue, K. Inaba, K. Aihara, and M. Hasegawa, "High-speed resource allocation algorithm using a coherent Ising machine for NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 73, no. 1, pp. 707–723, Jan. 2024.
[37] H. Yahya, E. Alsusa, and A. Al-Dweik, "Exact BER analysis of NOMA with arbitrary number of users and modulation orders," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6330–6344, Sep. 2021.
[38] T. Mihana, K. Kanno, M. Naruse, and A. Uchida, "Photonic decision making for solving competitive multi-armed bandit problem using semiconductor laser networks," *Nonlinear Theory Appl.*, vol. 13, no. 3, pp. 582–597, 2022.
[39] K. Morijiri, T. Mihana, K. Kanno, M. Naruse, and A. Uchida, "Decision making for large-scale multi-armed bandit problems using bias control of chaotic temporal waveforms in semiconductor lasers," *Sci. Rep.*, vol. 12, no. 1, p. 8073, May 2022.
[40] R. Iwami, T. Mihana, K. Kanno, S. Sunada, M. Naruse, and A. Uchida, "Controlling chaotic itinerancy in laser dynamics for reinforcement learning," *Sci. Adv.*, vol. 8, no. 49, Dec. 2022, Art. no. 125225.
[41] S. Mounchili and S. Hamouda, "Pairing distance resolution and power control for massive connectivity improvement in NOMA systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4093–4103, Apr. 2020.
[42] N. Chauvet, G. Bachelier, S. Huant, H. Saigo, H. Hori, and M. Naruse, "Entangled N-photon states for fair and optimal social decision making," *Sci. Rep.*, vol. 10, no. 1, p. 20420, Nov. 2020.
[43] J. An, C. Yuen, C. Xu, H. Li, D. Wing Kwan Ng, M. D. Renzo, M. Debbah, and L. Hanzo, "Stacked intelligent metasurface-aided MIMO transceiver design," 2023, *arXiv:2311.09814*.

MASAKI SUGIYAMA received the B.E. degree from the Department of Electrical Engineering, Tokyo University of Science, Japan, in 2022, where he is currently pursuing the master's degree. His current research interests include NOMA, reinforcement learning, and laser chaos.



TAKATOMO MIHANA received the B.E., M.E., and Ph.D. degrees in engineering from Saitama University, in 2017, 2019, and 2022, respectively. Since 2022, he has been an Assistant Professor with the Department of Information Physics and Computing, Graduate School of Information Science and Technology, The University of Tokyo.

AOHAN LI (Member, IEEE) received the Ph.D. degree from Keio University, Tokyo, Japan, in 2020. From 2020 to 2022, she was an Assistant Professor with the Tokyo University of Science, Tokyo, where she is currently a Visiting Researcher. She is also an Assistant Professor with The University of Electro-Communications, Tokyo. She has published over 70 peer-reviewed journals and international conference papers. Her current research interests include resource management, quantum annealing, machine learning, and the Internet of Things. She is a member of IEICE. She was a recipient of the Ninth International Conference on Communications and Networking in China (CHINACOM'14) Best Paper Award, in 2014; the Third International Conference on Artificial Intelligence in Information and Communication (ICAIIC'21) Excellent Paper Award; and the Telecom System Technology Student Excellent Paper Award of the Telecommunications Advancement Foundation, Japan, in 2021.

MAKOTO NARUSE received the B.E., M.E., and Ph.D. degrees in engineering from The University of Tokyo, in 1994, 1996, and 1999, respectively. After working as a Research Associate and an Assistant Professor with The University of Tokyo, from 1999 to 2002, he joined the National Institute of Information and Communications Technology, Ministry of Internal Affairs and Communications, Tokyo, in 2002. In 2017, he was an Invited Processor with Université Grenoble Alpes, Grenoble, France. Since 2019, he has been a Professor with the Department of Information Physics and Computing, Graduate School of Information Science and Technology, The University of Tokyo.

MIKIO HASEGAWA (Member, IEEE) received the B.Eng., M.Eng., and Dr.Eng. degrees from the Tokyo University of Science, Japan, in 1995, 1997, and 2000, respectively. From 1997 to 2000, he was a Research Fellow with the Japan Society for the Promotion of Science (JSPS). From 2000 to 2007, he was with the Communications Research Laboratory (CRL), Ministry of Posts and Telecommunications, which was reorganized as the National Institute of Information and Communications Technology (NICT), in 2004. Currently, he is a Professor with the Department of Electrical Engineering, Faculty of Engineering, Tokyo University of Science. His research interests include mobile networks, cognitive radio, neural networks, machine learning, and optimization techniques. He is a Senior Member of IEICE.