JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 38, 859-875 (2022)
DOI: 10.6688/JISE.202207_38(4).0010

Examination of User Pairing NOMA System Considering the DQN Scheme Over Time-Varying Fading Channel Conditions

RAVI SHANKAR, AYAZ AHMAD, SAMINATHAN VEERAPPAN,
MANOJ KUMAR BEURIA, PATTETI KRISHNA,
SUDHANSU SEKHAR SINGH AND V. GOKULA KRISHNAN
ECE Department
Madanapalle Institute of Technology and Science
Madanapalle, 517325 India
E-mail: {ravishankar.nitp1}@gmail.com

In this paper, we employ the deep Q-network (DQN) algorithm to study user pairing downlink (D/L) non-orthogonal multiple access (NOMA) system considering the multiple user equipment (UEs). In this work, the independent and identically distributed (i.i.d.) fading links are considered. The channel is expected to become time-varying as a result of node mobility. User pairing and optimum power distribution algorithms based on reinforcement learning (RL) are investigated initially in NOMA systems. To make the analysis simpler and for reducing the computational complexity the Q-learning-based scheme is used jointly to investigate the user pairing and optimal power allocation problems. In real-time propagation conditions with numerous users, DQN was employed to conduct user pairing and power allocation at the same time. When the learning rate is 0.2000, the DQN method, on the other hand, converges quicker but does not reach the maximum throughput (or average sum rate). In our simulation, we have eight UEs, and it has been observed that employing near-user far-user (N-F) results in a better sum rate. The symbol error rate (SER) performance falls dramatically as the node velocity increases, according to simulation curves. This is because, when node mobility increases, the channel will change extremely quickly. The simulation results confirm the derived analytical expressions.

Keywords: NOMA, channel state information (CSI), successive interference cancellation (SIC), deep learning (DL), deep neural network (DNN), SER, time selective fading, DQN

1. INTRODUCTION

The internet of things (IoT), big data analytics, and rapid improvements in the production of fifth generation (5G) cellular and portable wireless devices have all led to a considerable rise in network traffic [1]. The objective of today's wireless networks is to provide extremely high signal transmission rates, lower latency, and massive connectivity, which can only be achieved by integrating a range of 5G technologies [2]. The integration of various 5G schemes can serve many users from a single resource block to increase energy efficiency (EE) and spectral efficiency (SE). However, handling asynchronous data created by machines to offer end-users massive IoT connectivity and diverse quality of

Received September 29, 2021; revised December 22, 2021; accepted March 14, 2022.
Communicated by Changqia Xu.

service (QoS) is difficult. NOMA techniques are used to overcome the issues in the context of 5G in the third-generation partnership project's new radio (NR) release-14 for D/L and NR release-15 for uplink (U/L) standards. Because of its high SE, massive connectivity, and potential as a 5G enabler, NOMA has been an active research topic in recent years. It has been explored by several researchers. Previous studies [3,4] by the researcher mostly focused on the basic schemes, waveform design, outage probability performance, benefits and limits, comparison with orthogonal multiple access (OMA), applications, and future network possibilities. Another system, power domain NOMA, uses the superposition techniqueat the base station (BS) and the SIC scheme at the receiver, allowing different UEs to access resources (time and frequency) based on the power allocation factor value [5]. In the work [6], the authors investigated the user pairing NOMA and resource allocation algorithms under frequency-flat Rayleigh fading links without considering the node mobility and time-selective fading channel conditions. In the work [7], the authors have investigated the combined resource allocation and user pairing scheme for a virtual multiple-input multiple-output (MIMO) network. First, a multiple-level water-filling-based resource allocation strategy is employed to address resource allocation with known matched user groups. The next stage is to execute joint user pairing and power allocation using an iterative algorithm based on the analysis from the previous phase. In [7], the authors introduced user pairing and scheduling methods for massive MIMO–NOMA systems, with the purpose of lowering inter-pair interference and therefore boosting the sum rate.

In the work [8], the authors have investigated the 2 cellular user D/L MIMO-NOMA system considering the optimal resource allocation. Further, the non-convex optimization framework has been developed for the MIMO-NOMA system and suboptimal and optimal solutions are provided. Furthermore, in the work [9], an optimum power allocation system for maximum fairness is developed; based on the max-min rate criterion power allocation scheme, all UEs have the same data rate. In the work [10], it has been observed that the user pairing D/L NOMA with optimal resource allocation provides a much better SE than the conventional NOMA system. To improve SE, the UEs were sorted by fading channel strength and then given the best resource allocation. In [11], the user pairing NOMA system is investigated for a maximum of two users. In the works [12–14], the optimal resource allocation scheme is developed for the NOMA system. The basic MIMO technique considers the convex optimization-based resource allocation problem and seeks to distribute power by addressing the Karush Kuhn Tucker (KKT) conditions analytically. The RL, on the other hand, is used in the MIMO-NOMA system to determine the power of the UEs in each pair. Furthermore, even though previous schemes for evaluating user pairing and resource allocation in a MIMO-NOMA strategy required a high level of computational complexity, we identify user pairing and power allocation jointly with a low level of computational complexity.

DL has been applied in several research in 5G networks [15, 16], including approaches like supervised, unsupervised, and RL. Adapting to time and frequency selective fading channel conditions may be challenging since supervised learning requires a large number of datasets. Data classification, statistical distributions, user matching, and resource allocation are all demanding challenges in unsupervised learning. One of the most well-known RL systems is Q-learning, a famous model-free method for RL. Q-learning can solve the problem of user pairing and power allocation by taking "action". The CSI

between the UE and the BS changes continuously at every time slot as a result of node mobility between the UEs and large-scale fading. As a result, Q-learning, which uses CSI to determine the best "reward" without utilizing a dataset, may be more suitable for 5G than other supervised learning algorithms that require a huge number of datasets. In the literature [17–26], the DL algorithm was applied in the NOMA scheme. In work [17], the authors proposed a DL-based sparse code multiple access (SCMA) method in which data is mapped to the resource and incoming signals are analyzed using a deep neural network (DNN). The data used for training is unstructured and loud. The authors developed a dual DNN-based deep RL (DRL)-based power allocation to overcome this problem [18]. Furthermore, in [19], a recurrent neural network (RNN) was utilized to calculate the NOMA channel, which is used to learn the NOMA system's CSI via offline and online training.

The authors of [20, 21] used the greedy algorithm to investigate the rapid RL technique based on the DQN in the context of jamming assaults. The authors of [22] used multiple agent RL to pair users in multiple carrier NOMA systems. For NOMA systems, the authors recommended a DQN-based combined power and channel assignment [23]. For the channel assignment problem, the authors developed an approach for identifying the best power allocation parameters using an attention-based DQN. The authors of [24] create a partially observable Markov decision process for dynamic channel access difficulty, and DQN is used to discover the best access policy via online learning. The authors in the work [25] have explored a multiple agent DNN strategy to estimate the spectrum occupancy of unknown adjacent networks in slotted 5G systems, in which they trained the DNN in real-time using both RL and supervised learning. To improve the sum rate, the authors [25] proposed using a DQN-based resource allocation for a multiple cell system. The authors used the combined precoding and SIC decoding approach for the MIMO-NOMA network in [26], considering the incorrect SIC decoding situation. Non-convex optimization, resource allocation, antenna beamforming, SIC ordering, and user pairing are the primary concerns for MIMO-NOMA systems. These issues have been investigated in tandem or in part, using specific performance measurements. MIMO-NOMA is a technique that can improve SE in 5G, however, it has a significant computational complexity restriction. In the work [27], the authors have solved the sum-rate maximization problem and consider the Marcum Q-function with first-order and log-concavity qualities to identify the best transmit power and subchannel allocation techniques while keeping QoS in mind for users. In addition, to optimize the sum rate of the NOMA system, in the work [28], the authors have proposed spectrum resource and power distribution using an adaptive proportional fair user matching mechanism. Many of the recommended resource allocation and channel assignment approaches in the literature [29,30] are nondeterministic polynomial-hard [29] and solving these issues directly is extremely challenging. Furthermore, classic optimization procedures have not been properly investigated in terms of computing efficiency, even though they frequently comprise many stages and complicated computer processes [30]. In the work [20,31], the authors have used an RL-based system called Q-learning to allocate power resources to improve the SE and EE of the NOMA transmission. The DQN algorithm, like a traditional DRL approach, has the advantage of increased learning efficiency and more policy selection capacity due to its neural network, which outperforms typical RL algorithms. The experience relay buffer, a device in the DQN approach, can reduce the interdependence of the input data and make the training process more trustworthy. *The contributions of this paper are as follows:*

- For NOMA systems, the RL-based user pairing and optimal power distribution strategies are researched first. Previous studies looked into user pairing and power allocation problems separately or looked into them using mathematical methodologies like convex optimization in a simpler system with a few users.
- To the authors' knowledge, this is the first time DQN has been used to execute user pairing and power allocation simultaneously in a real-world system with multiple users.
- The main aim of this paper is to enhance the SER performance of the NOMA system while decreasing computational complexity by applying DQN based optimal power allocation and user pairing.
- The DQN method is used to solve the user pairing problem, allowing the overall sum rate of all UEs to be maximized.

The rest of the paper is structured as follows: Section 2 investigates the signal and fading channel models, and formulations for the time selective fading channels are produced. In addition, the user pairing NOMA system has been described, and several types of user pairing systems have been examined. Section 3 investigates the DRL-based user pairing NOMA scheme and provides optimal power allocation and user pairing based on the Q-Learning algorithm. Section 4 contains the simulation findings, while Section 5 concludes the work.

2. SIGNAL AND CHANNEL MODEL

2.1 Channel Model

It is considered in this work that node mobility causes the channel to transition from i.i.d. frequency flat fading channel to a time-varying fading channel. It has been assumed that the UEs move in relation to each other. The first-order autoregressive process [32] may be used to simulate a time-varying fading channel. $z(\tau) = \rho z(\tau-1) + \sqrt{1-\rho^2}e(\tau)$, where τ and $\tau-1$ represents the two adjacent time instants. The term $e(\tau)$ is a random process and it can be modeled as the CN(0, σ^2), ρ is the fading channel correlation coefficients arises due to the Doppler spread, expressed as $\rho = J_0\{2\pi f_c v/R_S c\}$. Where f_c represents the frequency, v represents the relative velocity between two communicating UEs, c denotes the velocity of light, $J_0(.)$ denotes the Bessel function of zeroth order and first kind, and R_S represents the symbol rate.

2.2 System Description

In this work, we explore the D/L NOMA system in a microcell with a radius of 600 m, demonstrated in Fig. 1 [5]. The power transmitted from the BS is represented as P_{BS}. The transmitted power is considered to be distributed evenly among all antennas. As a result, the selection combining technique is employed at the BS. The BS transmits the superimposed signal to all the UEs, and it considers the features of NOMA. All M UEs are randomly placed in a cell to produce a NOMA-applicable situation. The transmitted power at each beam can be expressed as $P_n = P_{BS}/N$. The instantaneous fading channel gain is considered in the following order [5–9],

$$|z_{n,i}(\tau)|^2 \leq \left|z_{n,j}(\tau)\right|^2, \quad for\ i \leq j. \tag{1}$$

In the NOMA system under consideration, the SIC technique is used by the closest user (with excellent channel strength) to cancel the interference signal, which might be the signal transmitted to the UE with poor channel conditions. In this situation, the SIC should function with few or no errors. The BS is also in charge of pairing UEs and calculating each UE's transmits power. Each UE undergoes from frequency flat Rayleigh fading links and additive white Gaussian noise (AWGN) with zero average value and noise variance $\sigma_{n,k}$. After performing the section combining scheme the transmitted signal from the BS is expressed as [4–6],

$$x_n = \sum_{k=1}^{K} \sqrt{\alpha_{n,k}} P_n s_{n,k}, \tag{2}$$

Where $\alpha_{n,k}$ represents the power allocation factor, the signal transmitted from the BS is represented as $s_{n,k}$ and P_n denote the beam power. The signal obtained at the $UE_{n,k}$ is expressed as,

$$y_{n,k} = \mathrm{z}_{n,k}(\tau) \sum_{n=1}^{N} \mathrm{w}_n x_n + \eta_{n,k}, \tag{3}$$

Where $\mathrm{z}_{n,k}(\tau)$ is channel vector consists of the Rayleigh fading channel coefficients from the BS to the $UE_{n,k}$. Precoding matrix consists of precoding vectors w_n for each beam, expressed as, $\mathrm{W} = [\mathrm{w}_1, \mathrm{w}_2, \mathrm{w}_3, ..., \mathrm{w}_n]$, $\mathrm{w}_n \in {}^{1\times N}$, and $\eta_{n,k}$ is modeled as AWGN noise. The channel vector $\mathrm{z}_{n,k}(\tau)$ consists of Rayleigh fading channel coefficients expressed as,

$$\mathrm{z}_{n,k}(\tau) = z_{n,k}(\tau) \sqrt{d_{n,k}^{-\eta}}. \tag{4}$$

Furthermore, $d_{n,k}$ represents the distance between the BS and $UE_{n,k}$, η represents the route loss exponent, and $z_{n,k}$ represents the state of the RL. Rewrite Eq. (3) as follows,

$$y_{n,k} = \mathrm{z}_{n,k}(\tau) \sqrt{P_n \alpha_{n,k}} s_{n,k} + \underbrace{\mathrm{z}_{n,k}(\tau) \mathrm{w}_n \sum_{k'=k+1}^{K} \sqrt{P_n \alpha_{n,k'}} s_{n,k'}}_{\text{intra-beam interference}} + \underbrace{\mathrm{z}_{n,k}(\tau) \sum_{n'=1, n\neq n}^{N} \mathrm{w}_{n'} x_{n'}}_{\text{inter-beam interference}} + \eta_{n,k}. \tag{5}$$

After SIC, the above expression may be expressed as,

$$y_{n,k} = \begin{cases} \mathrm{z}_{n,k}(\tau) \sqrt{P_n \alpha_{n,k}} s_{n,k} + \mathrm{z}_{n,k}(\tau) \sum\limits_{n'=1, n\neq n}^{N} \mathrm{w}_{n'} x_{n'} + \eta_{n,k}, & if\ k = K \\ \mathrm{z}_{n,k}(\tau) \sqrt{P_n \alpha_{n,k}} s_{n,k} + \mathrm{z}_{n,k}(\tau) \mathrm{w}_n \sum\limits_{k'=k+1}^{K} \sqrt{P_n \alpha_{n,k'}} s_{n,k'} \\ \quad + \mathrm{z}_{n,k}(\tau) \sum\limits_{n'=1, n\neq n}^{N} \mathrm{w}_{n'} x_{n'} + \eta_{n,k}, & if\ 1 \leq k \leq K, k \neq K. \end{cases} \tag{6}$$

According to the NOMA concept, the power allocation coefficient $\alpha_{n,k}$ of each UE is stated as follows,

$$0 \leq \alpha_{n,k} \leq 1, \sum_{k=1}^{K} \alpha_{n,k} = 1, \;\; \alpha_{n,k} \in \Omega, \tag{7}$$

where Ω represents the space of the feasible power allocation factors.

2.3 User Pairing for 8 UE NOMA Scheme

Fig. 2 demonstrates the D/L NOMA system represented as UE1-UE8, considering 8 UEs. The distance between UE1 and UE8 from the BS is represented by $d_1 - d_8$. Fig. 2 shows that UE1 is the user that is closest to the BS, whereas UE8 is the user who is furthest away. UE1 has the best channel condition, whereas UE8 has the worst, based on the assumption. The channel conditions are assumed to be listed in decreasing order, *i.e.*, $|z_1(\tau)|^2 > |z_2(\tau)|^2 > |z_3(\tau)|^2 > |z_4(\tau)|^2 > |z_5(\tau)|^2 > |z_6(\tau)|^2 > |z_7(\tau)|^2 > |z_8(\tau)|^2$. There are four users in each of the two orthogonal blocks or resources (in this case, time). In this study, two simple schemes for user pairing-based just on distance are provided.

2.3.1 Paring of near user with the far users (N-F pair)

The user who is closest to the BS is paired with the user who is farthest away from the BS in this form of user pairing. UE1 and UE8 are the closest and furthest users, respectively, in our case. As a result, UE1 and UE8 will couple up and share the same resource block as UE2 and UE7, the following N-F pair. Another N-F pair option is for UE1 to be paired with UE7 and UE2 to be partnered with UE8, in which case they can share resource blocks. Only the user closest to the BS is paired with the furthest in this article, while the second closest to the BS is paired with the second last farthest user, as shown in Fig. 3. Now, because the UE1 is nearest to the BS and the UE8 is farthest away, we assume that the UE8 will have a larger power allocation coefficient, *i.e.*, $\alpha_1 < \alpha_8$. UE8 may decode its data from the D/L signal directly using this method. UE1, on the other hand, must decode its data via SIC. The remainder of the pairings are assumed to be the same, therefore $\alpha_2 < \alpha_7, \alpha_3 < \alpha_6, \alpha_4 < \alpha_5$. As a result, UE2, UE3, and UE4 will need SIC to decode their data, whereas UE7, UE6, and UE5 will be able to decode their data directly in relation to UE2, UE3, and UE4. For the user N-F pairing the achievable rates for first N-F pair,

$$R_{1,nf} = (1/2) \times \log_2(1 + P\alpha_1 |z_1(\tau)|^2/\sigma^2) \;\; (\text{After Performing SIC}) \tag{8}$$

$$R_{8,nf} = (1/2) \times \log_2\{1 + P\alpha_8 |z_8(\tau)|^2/(P\alpha_1 |z_8(\tau)|^2 + \sigma^2)\} \;\; (\text{After DirectDecoding}) \tag{9}$$

Similarly, for 2nd N-F Pair,

$$R_{2,nf} = (1/2) \times \log_2(1 + P\alpha_2 |z_2(\tau)|^2/\sigma^2) \;\; (\text{After Performing SIC}) \tag{10}$$

$$R_{7,nf} = (1/2) \times \log_2\{1 + P\alpha_7 |z_7(\tau)|^2/(P\alpha_2 |z_7(\tau)|^2 + \sigma^2)\} \;\; (\text{After DirectDecoding}) \tag{11}$$

For 3rd N-F pair,

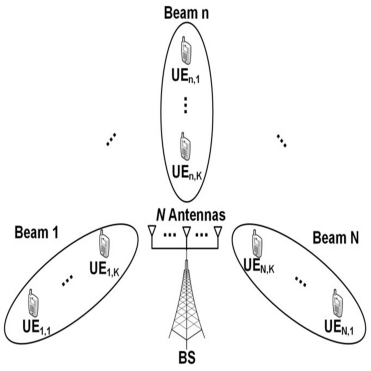

Fig. 1. Schematic representation of the D/L NOMA with multiple antennas.

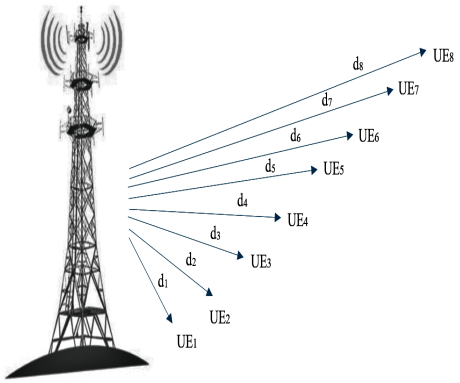

Fig. 2. System model for hybrid NOMA.

$$R_{3,nf} = (1/2) \times \log_2(1 + P\alpha_3 |z_3(\tau)|^2/\sigma^2) \text{ (After Performing SIC)} \tag{12}$$

$$R_{6,nf} = (1/2) \times \log_2\{1 + P\alpha_6 |z_6(\tau)|^2/(P\alpha_3 |z_6(\tau)|^2 + \sigma^2)\} \text{ (After DirectDecoding)} \tag{13}$$

For 4th N-F pair,

$$R_{4,nf} = (1/2) \times \log_2(1 + P\alpha_4 |z_4(\tau)|^2/\sigma^2) \text{ (After Performing SIC)} \tag{14}$$

$$R_{5,nf} = (1/2) \times \log_2\{1 + P\alpha_5 |z_5(\tau)|^2/(P\alpha_4 |z_5(\tau)|^2 + \sigma^2)\} \text{ (After Direct Decoding)} \tag{15}$$

The average achievable sum rate for N-F pairing scheme is expressed as,

$$R_{nf} = \sum_{i=1}^{8} R_{i,nf}. \tag{16}$$

2.3.2 Paring of near user with the near users and pairing of far users with far users (N-N pair, F-F pair)

As seen in Fig. 4, close users are paired with nearby users, whereas distant users are paired with far users. In this system, as illustrated in Fig. 4, the closest user (closest to the BS) will be paired with the next closest user going away from the BS, and the furthest user will be paired with the next farthest user moving toward the BS. We anticipated that UE1-UE4 are users who are closer to the BS and would engage in N-N pairing, whereas UE5-UE8 are users who are further away from the BS and will participate in F-F pairing. UE1 will be paired with UE2, and the resource block will be shared. We anticipated that the closer a user pair is to the BS, the smaller the power allocation coefficient, *i.e.*, $\alpha_1 < \alpha_2, \alpha_3 < \alpha_4, \alpha_5 < \alpha_6$ & $\alpha_7 < \alpha_8$. Hence, in every pairing, the user closest to the BS will perform SIC to decode its signal while the second in the pair can directly decode its signal. For the user N-N pairing the achievable rates for 1st N-N pair,

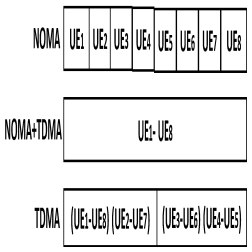

Fig. 3. Schematic representation of the N-F pairing scheme.

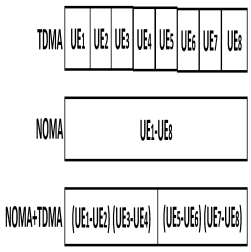

Fig. 4. Schematic representation of the N-N pair & F-F pair scheme.

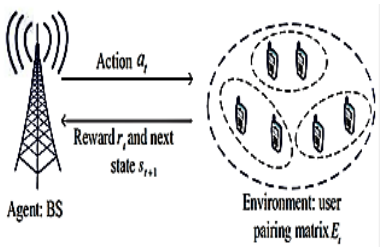

Fig. 5. DRL based scenario.

$$R_{1,nn} = (1/2) \times \log_2(1 + P\alpha_1 |z_1(\tau)|^2/\sigma^2) \quad \text{(After Performing SIC)} \tag{17}$$

$$R_{2,nn} = (1/2) \times \log_2\{1 + P\alpha_2 |z_2(\tau)|^2/(P\alpha_1 |z_2(\tau)|^2 + \sigma^2)\} \quad \text{(After Direct Decoding)} \tag{18}$$

Similarly for 2nd N-N pair

$$R_{3,nn} = (1/2) \times \log_2(1 + P\alpha_3 |z_3(\tau)|^2/\sigma^2) \quad \text{(After Performing SIC)} \tag{19}$$

$$R_{4,nn} = (1/2) \times \log_2\{1 + P\alpha_4 |z_4(\tau)|^2/(P\alpha_3 |z_4(\tau)|^2 + \sigma^2)\} \quad \text{(After Direct Decoding)} \tag{20}$$

for 1st F-F pair,

$$R_{5,ff} = (1/2) \times \log_2(1 + P\alpha_5 |z_5(\tau)|^2/\sigma^2) \quad \text{(After Performing SIC)} \tag{21}$$

$$R_{6,ff} = (1/2) \times \log_2\{1 + P\alpha_6 |z_6(\tau)|^2/(P\alpha_5 |z_6(\tau)|^2 + \sigma^2)\} \quad \text{(After Direct Decoding)} \tag{22}$$

Similarly for 2nd F-F pair

$$R_{7,ff} = (1/2) \times \log_2(1 + P\alpha_7 |z_7(\tau)|^2/\sigma^2) \quad \text{(After Performing SIC)} \tag{23}$$

$$R_{8,ff} = (1/2) \times \log_2\{1 + P\alpha_7 |z_8(\tau)|^2/(P\alpha_7 |z_8(\tau)|^2 + \sigma^2)\} \quad \text{(After Direct Decoding)} \tag{24}$$

The overall achievable sum rate for N-N, F-F pairing scheme will be

$$R_{nf} = \sum_{i=1}^{4} R_{i,nn} + \sum_{i=5}^{8} R_{i,ff}. \tag{25}$$

3. DRL BASED USER PAIRING NOMA SCHEME

The authors developed a methodology for getting the best power allocation factors using the DRL scheme in their paper [33]. This research also uses the user pairing approach to determine the best power allocation. The DRL environment is then used to transform the user pairing scheme. The DQN technique is used to study average sum rate performance as well as optimal power allocation. Finally, the algorithm for describing the entire procedure is presented.

3.1 Formulation of DRL Algorithm

In this sub-section the user pairing NOMA framework is transformed to the DRL environment [33]. The DRL scenario is demonstrated in Fig. 5 [5, 32, 33]. Agent, Environment and state, action, reward and target, policy are the main components of the DRL algorithm [33], expressed below:
(1) In a real-time propagation context, RL seeks to train an agent how to do a job. The agent is a policy maker as well as a learner. The agent responds by sending actions to the environment after receiving observations and a reward from the environment. The agent incorporates both a policy and a learning algorithm. BS serves as an agent in DRL. (2) In DRL algorithm the "agent" will interact with the environment (acts as "object") after each interaction and the change results are specified as "state". In the meantime, given the DRL scenario and proposed NOMA user pairing problem, the DRL environment can only represent the user pairing matrix E_t, where each element in G_1 is a 2×2 line matrix 1. and each element in G_2 is a column, and the agent can only take one "action" per step when it interacts with the environment. Because the process of user pairing takes place inside time slots, the NOMA system is also utilised to designate a time slot as a training session. During each step of a training period, the agent chooses an action to engage with the environment, and the status is therefore changed from current state s_t to the next state s_{t+1}. We regard every stage of each formation to be a limited Markov decision-making process due to the limited number of users (MDP). (3) The agent must choose an appropriate action based on unique strategies in the current state s_t since the various actions have different environmental implications. The action space of A_t may be described by the NOMA system as $A_t = \{u_t^{N+1}, u_t^{N+2}, ..., u_t^N\}$, with the action t representing $u_t^p (p \in \{N+1, N+2, ..., N\})$ at step t. . Since the user pairing matrix row represents wireless users in G_1, we assume that the transmission partner is created when user t selects user , which is interpreted as up $u_t^p = 1$, in every step t; otherwise, $u_t^p = 0$. (4) After completing a task, the agent is rewarded with an immediate positive or negative reward. The agent's goal is to find and detect a policy that will maximise the cumulative discount reward, which is calculated by multiplying the current immediate reward by a discount factor at each training session. In our NOMA system, the immediate reward may be calculated as $r_t = r_t^{\pi}(s_t, a_t)$, where s_t represents the state and a_t represents the action taken in step t. The average sum rate of the t^{th} user is r_t, whereas the sum rate of users who are moved on the same user pair is r_t. If there are more than two users in a user pair, the immediate reward is set to zero, and the current training period is ended. As a result, the purpose of the DQN algorithm is to maximize the discount cumulative reward, which may be thought of as the aggregate sum rate of all users. (5) The process is selected by

the agent in the same way that the policy is selected by the agent. The DQN approach uses the $\varepsilon - Greedy$ policy to select an action. That is, the action is randomly selected with a probability of ε, and the action produces the highest action state value $Q(s_t, a_t)$ with a probability of $1-\varepsilon$. To prevent the selected algorithm from being tuned locally to the optimal solution, the agent can use processes to investigate unknown actions and conditions.

3.2 Optimal Power Allocation and User Pairing based on Q-Learning Algorithm

The average sum-rate of D/L NOMA UEs is the "reward", which represents the reward at time t, expressed below [5, 20, 31–33],

$$\hat{R}^{all} = \sum_{n=1}^{N} \sum_{k=1}^{K} \log_2 \left(1 + \frac{\alpha_{n.k} p_n \left| \hat{z}_{n,k}(\tau) w_n \right|^2}{I_{n,k}^U + \sigma_n^2} \right), \tag{26}$$

The sum-rate computed using $\hat{z}_{n,k}(\tau)$ is represented as $\hat{R}^{all}$. In Q-learning, the Q-function updates $\hat{R}^{all}$ on a regular basis, whereas $\hat{R}^{all}$ calculates $\hat{z}_{n,k}(\tau)$. The user pairing index and the power allocation factors are both determined simultaneously using Q-learning. Furthermore, given system state s and action a, $Q(s^t, \theta^t)$, denotes the BS's Q-function [20, 31, 33],

$$Q(s^t, \theta^t) \leftarrow (1-\beta) Q(s^t, \theta^t) + \beta [r(s^t, \theta^t) + \delta \max_{\theta'} Q(s^{t+1}, \theta^t). \tag{27}$$

Where $\beta \in (0, 1]$ denotes the importance of recent learning experiences. The discount factor $\delta \in [0, 1]$ determines the importance of present and future benefits [30, 31, 33–35].

4. SIMULATION RESULTS

This section shows the simulation results for analysing the DL-based NOMA system's end-to-end performance. We also simulate the end-to-end system performance of the other three approaches, namely the DQN algorithm, DRL algorithm, and conventional NOMA scheme considering the optimal power allocation, to make a comparison.

4.1 Simulation Setting

It has been assumed that the BS is situated at the cell's center, with the M UEs dispersed randomly about the cell at distances ranging from 50 to 250 metres. Key simulation parameters are listed in Table 1. Open source tool developed from the Google library TensorFlow Core v2.8.0 primarily for DL applications, which runs on Python 3.9.0, implements the DQN algorithm. Table 2 shows a few of the parameters used in the DQN algorithm.

4.2 Simulation Analysis by Considering Frequency Flat Fading Conditions Without Node Mobility

In the analysis, the average sum rate is represented as the "average cumulative reward," and the BS acts as an "agent." Fig. 6 demonstrates the learning graphs, which de-

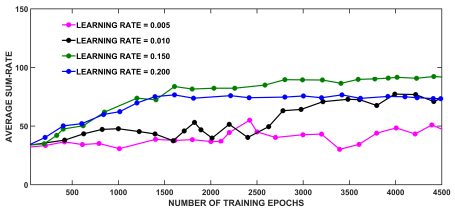

Fig. 6. Convergence variation during the training process.

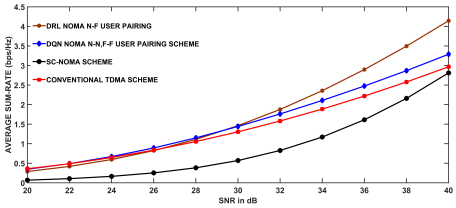

Fig. 7. Average Sum rate versus SNR in dB for various user pairing schemes.

monstrate the average sum rate acquired by the BS in each training period with various learning rates. In the simulation, we consider 30 UEs. When the learning rate is set to 0.1500, the average sum rate, *i.e.*, cumulative reward, converges to roughly 100 Mbps after 3900 epochs, resulting in the best end-to-end system performance. The DQN scheme, on the other hand, converges faster when the learning rate is 0.2000, but does not attain the maximum throughput (or average sum rate). This problem may be explained by the fact that when the learning rate is high, the gradient update speed is too quick, causing the optimal solution to slide. Meanwhile, the DQN algorithm fails to converge when the learning rate is set to 0.010 or 0.005. This is since when the learning rate is low, gradient updating is likewise sluggish. As a result, the training process will not be able to converge in less than 4500 epochs.

Figs. 7 and 8 provide a comparison of several DL algorithms. The following simulation settings are used: $d_1 = 15,\ d_2 = 10,\ d_3 = 8,\ d_4 = 7,\ d_5 = 6,\ d_6 = 5,\ d_7 = 4,\ d_8 = 3$; path loss exponent = 4; and number of iterations = 10^5. The N-F and N-N, F-F techniques were expressed analytically in the preceding section. Figs. 7 and 8 provide a comparison of these two user pairing procedures. Additionally, the single carrier NOMA (SC-NOMA) approach has been studied, in which all users are multiplexed on the same carrier without user pairing. We assume eight UEs in our simulation, and it has been discovered that using N-F results in a higher sum rate. This supports our analytical outcomes that NOMA operates better when the channel conditions between the two users differ [34] [34]. NOMA still outperforms TDMA when NN, FF pairing is employed, but the difference is not substantial. When compared to TDMA, the performance of SC-NOMA is inferior. Because jamming occurs when everyone is on the same carrier. This also verifies our suspicion that increasing the number of users sharing the same carrier without a cost is impossible.

It is seen in Fig. 8 that the DQN scheme has overtaken the DRL, traditional TDMA and SC-NOMA schemes. It is also easy to see that SC-NOMA performance is lower than TDMA performance. Due to the interference experienced by all users on the same carrier, it turns out that it is impossible to increase the number of users on the same subcarrier without paying for it. In simulation the learning rate is fixed at 0.20. The DQN NOMA algorithm outperforms the DRL NOMA scheme and can demonstrate that the achievable data rate is much greater than the Nyquist Shannon rate. The DQN NOMA algorithm, in contrast to the DRL NOMA method, uses the DL approach to estimate the Q value of the action state. It is easy to see that the DQN NOMA algorithm can extract features from input data symbols using DL symbol training. In addition, the large amount of data symbols complicates the storage and retrieval of Q values in the Q-learning process.

Table 1. Simulation parameters.

Parameter	value
$d_{\min}$(minimum distance) between two users	15m
The net transmission bandwidth	20 MHz
Sub-carrier	2 MHz
Power transmitted from the BS	45 dBm
Power allocated to each subcarrier	30 dBm
Path loss exponent	3
AWGN noise spectral density	–170 dBm/Hz
Quality of Service threshold limit	2.50 bps/Hz

Table 2. The parameters of the DQN algorithm.

Parameter	value
Total number of fully connected (FC) hidden layers in DNN	4
Number of Neurons in 1st hidden layer	130
Number of Neurons in 2nd hidden layer	135
Discount factor	0.90
Batch size	258
ε	from 0.07 to 0.02
The number of FC hidden layers	2
The number of neurons in the 1st hidden layer	128

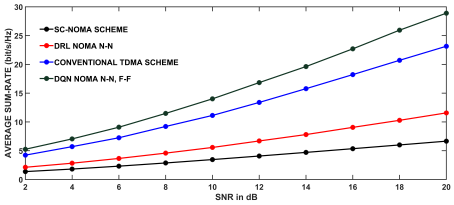

Fig. 8. Performance comparison between DQN NOMA, DRL NOMA and SC NOMA schemes.

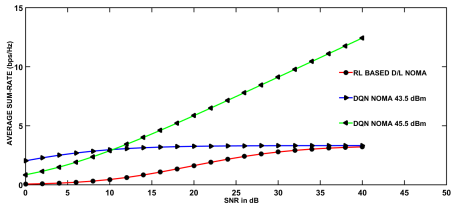

Fig. 9. Achievable sum rate for 45.5 and 43.5 dBm.

Therefore, the DQN NOMA algorithm is superior to the DRL NOMA method in terms of overall rate performance and SE. In addition, the DQN-NOMA method is superior to both random user pairing and TDMA. This is because the best pairing for transmission on the evaluated NOMA system is 8 UEs. The NOMA approach uses more spectra than the TDMA method because only one UE is transmitted on a subchannel of the OMA system. The NOMA approach can significantly improve data transfer rates by using resources in both the frequency and power domains.

As shown in Fig. 9, if BS adopts different transmit power recommendations such as 45.5 dBm and 43.5 dBm, the overall data transmission rate of the proposed optimal power allocation scheme is the overall fixed power allocation system and OFDMA scheme. It will be larger than the target rate. This demonstrates the benefits of an optimal power allocation scheme. It also shows that the data transmission rate increases as the BS transmission power increases. This is because as BS transmit power increases, so does the transmit power limit for each subchannel, resulting in more transmit power allocated to each user. This will significantly improve the data rate for each user. Fig. 10 shows that as the number of UEs increases, the data transfer rate increases. As the number of UEs increases, net throughput increases significantly. As the number of UEs increases, the performance difference of the NF method, which distributes the same power as the TDMA

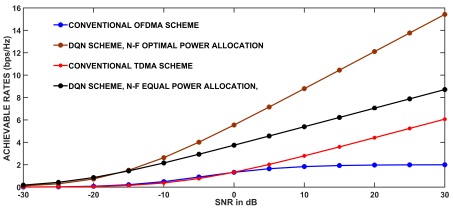

Fig. 10. Performance comparison between the equal and optimal power allocation factors.

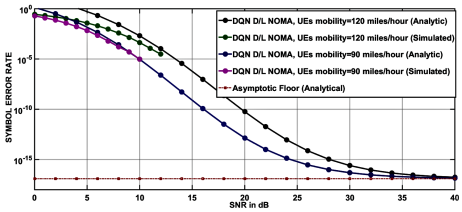

Fig. 11. Impact of the node mobility on D/L NOMA system.

system, decreases. Due to the growing number of "states", the performance gap between 10 UEs is about 6.49 percent. The proposed approach increases the overall rate by 14.30 percent and 49.70 percent, respectively, compared to OFDMA and random user selection schemes. On the other hand, both the DQN algorithm and the phased DRL scheme work the same.

4.3 Simulation Analysis by Considering Time Selective Fading Conditions With Node Mobility

In this sub-section we consider the node mobility scenario, considering the channel estimation error. Fig. 11 shows the SER performance of the D/L NOMA system considering the i.i.d. time varying fading channel conditions. Simulation curves show that with increase in the node velocity the SER performance decreases significantly. This is since with increase in the node mobility the channel will change very rapidly. Due to change in channel coefficients instantly it is not possible to estimate the channel perfectly. Due the imperfect CSI the SER performance decreases significantly, and the net throughput decreases significantly.

5. CONCLUSION

In this work DQN scheme is compared with the conventional NOMA schemes. The joint resource allocation and user pairing schemes are jointly investigated. If the BS follows different transmit power recommendations such as 45.5 dBm and 43.5 dBm, the overall rate of the proposed optimal power allocation scheme would be higher than that of fixed power allocation systems and OFDMA schemes, as shown in the simulation section. The SER performance falls dramatically as the node velocity increases, according to simulation curves. This is because, when node mobility increases, the channel will change extremely quickly. It is impossible to estimate the channel correctly due to instantaneous changes in channel coefficients. The SER performance suffers as a result of the faulty CSI, and the net throughput suffers as a result.

REFERENCES

1. M. H. C. Garcia, A. Molina-Galan, M. Boban, J. Gozalvez, B. Coll-Perales, T. Sahin, and A. Kousaridas, "A tutorial on 5G NR V2X communications," *IEEE Communications Surveys Tutorials*, Vol. 23, 2021, pp. 1972-2026.

2. Y. Siriwardhana, P. Porambage, M. Liyanage, and M. Ylianttila, "A survey on mobile augmented reality with 5G mobile edge computing: Architectures, applications, and technical aspects," *IEEE Communications Surveys Tutorials*, Vol. 23, 2021, pp. 1160-1192.
3. B. P. Chaudhary, R. Shankar, and R. K Mishra, "A tutorial on cooperative non-orthogonal multiple access networks," *The Journal of Defense Modeling and Simulation*, 2021, No. 1548512920986627.
4. R. Shankar, S. Nandi, and A. Rupani, "Channel capacity analysis of non-orthogonal multiple access and massive multiple-input multiple-output wireless communication networks considering perfect and imperfect channel state information," *The Journal of Defense Modeling and Simulation*, 2021, No. 15485129211000139.
5. J. Lee and J. So, "Reinforcement learning-based joint user pairing and power allocation in MIMO-NOMA systems," *Sensors*, Vol. 20, 2020, No. 7094.
6. B. Jia, H. Hu, Y. Zeng, T. Xu, and H. H. Chen, "Joint user pairing and power allocation in virtual MIMO systems," *IEEE Transactions on Wireless Communications*, Vol. 17, 2018, pp. 3697-3708.
7. X. Chen, F. K. Gong, G. Li, H. Zhang, and P. Song, "User pairing and pair scheduling in massive MIMO-NOMA systems," *IEEE Communications Letters*, 2018, Vol. 22, pp. 788-791.
8. O. Abuajwa, M. B. Roslee, and Z. B. Yusoff, "Simulated annealing for resource allocation in downlink NOMA systems in 5G networks," *Applied Sciences*, Vol. 11, 2021, No. 4592.
9. S. Ravi, G. R. Kulkarni, S. Ray, M. Ravisankar, V. G. Krishnan, and D. S. K. Chakravarthy, "Analysis of user pairing non-orthogonal multiple access network using deep Q-network algorithm for defense applications," *The Journal of Defense Modeling and Simulation*, 2022, No. 15485129211072548.
10. J. Guo, X. Wang, J. Yang, J. Zheng, and B. Zhao, "User pairing and power allocation for downlink non-orthogonal multiple access," in *Proceedings of IEEE Globecom Workshops*, 2016, pp. 1-6.
11. F. Liu, P. Mahonen, and M. Petrova, "Proportional fairness-based user pairing and power allocation for non-orthogonal multiple access," in *Proceedings of IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications*, 2015, pp. 1127-1131.
12. H. Sun, Y. Xu, and R. Q. Hu, "A NOMA and MU-MIMO supported cellular network with underlaid D2D communications," in *Proceedings of IEEE 83rd Vehicular Technology Conference*, 2016, pp. 1-5.
13. Q. Sun, S. Han, C.-L. I, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Communications Letters*, Vol. 4, 2015, pp. 405-408.
14. S. Timotheou and I. Krikidis, "Fairness for non-orthogonal multiple access in 5G systems," *IEEE Signal Processing Letters*, Vol. 22, 2015, pp. 1647-1651.
15. C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications Surveys Tutorials*, Vol. 21, 2019, pp. 2224-2287.
16. N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys Tutorials*, Vol. 21, 2019, pp. 3133-3174.

17. M. Kim, N.-I. Kim, W. Lee, and D.-H. Cho, "Deep learning-aided SCMA," *IEEE Communications Letters*, Vol. 22, 2018, pp. 720-723.
18. K. N. Doan, M. Vaezi, W. Shin, H. V. Poor, H. Shin, and T. Q.-S. Quek, "Power allocation in cache-aided NOMA systems: Optimization and deep reinforcement learning approaches," *IEEE Transactions on Communications*, Vol. 68, 2020, pp. 630-644.
19. G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Transactions on Vehicular Technology*, Vol. 67, 2018, p. 844.
20. L. Xiao, Y. Li, C. Dai, H. Dai, and H. V. Poor, "Reinforcement learning-based NOMA power allocation in the presence of smart jamming," *IEEE Transactions on Vehicular Technology*, Vol. 67, 2018, pp. 3377-3389.
21. P.-G. Ye, Y.-G. Wang, J. Li, and L. Xiao, "Fast reinforcement learning for anti-jamming communications," *arXiv Preprint*, 2020, arXiv:2002.05364.
22. S. Wang, T. Lv, and X. Zhang, "Multi-agent reinforcement learning-based user pairing in multi-carrier NOMA systems," in *Proceedings of IEEE International Conference on Communications Workshops*, 2019, pp. 1-6.
23. C. He, Chaofan, Y. Hu, Y. Chen, and B. Zeng, "Joint power allocation and channel assignment for NOMA with deep reinforcement learning," *IEEE Journal on Selected Areas in Communications*, Vol. 37, 2019, pp. 2200-2210.
24. S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Transactions on Cognitive Communications and Networking*, Vol. 4, 2018, pp. 257-265.
25. R. Mennes, D. Figueiredo, A. P. Felipe, and S. Latr, "Multi-agent deep learning for multi-channel access in slotted wireless networks," *IEEE Access*, Vol. 8, 2020, pp. 95032-95045.
26. K. I. Ahmed and E. Hossain, "A deep Q-learning method for downlink power allocation in multi-cell networks," *arXiv Preprint*, 2019, arXiv:1904.13032.
27. Z. Li, S. Wang, P. Mu, and Y.-C. Wu, "Sum rate maximization of secure NOMA transmission with imperfect CSI," in *Proceedings of IEEE International Conference on Communications*, 2020, pp. 1-6.
28. K. Long, P. Wang, W. Li, and D. Chen, "Spectrum resource and power allocation with adaptive proportional fair user pairing for NOMA systems," *IEEE Access*, Vol. 7, 2019, pp. 80043-80057.
29. G. Liang, Q. Zhu, J. Xin, Y. Feng, and T. Zhang, "Joint user-channel assignment and power allocation for non-orthogonal multiple access relaying networks," *IEEE Access*, Vol. 7, 2019, pp. 30361-30372.
30. N. Yang, H. Zhang, K. Long, H.-Y. Hsieh, and J. Liu, "Deep neural network for resource management in NOMA networks," *IEEE Transactions on Vehicular Technology*, Vol. 69, 2020, pp. 876-886.
31. S. Pandya, P. Krishna, R. Shankar, and A. S. Bist, "Examination of the fifth-generation vehicular simultaneous wireless information and power transfer cooperative non-orthogonal multiple access network in military scenarios considering time-varying and imperfect channel state information conditions," *The Journal of Defense Modeling and Simulation*, 2021, No. 15485129211033040.
32. R. Shankar and R. K. Mishra, "PEP and OP examination of relaying network over time-selective fading channel," *SN Applied Sciences*, Vol. 2, 2020, pp. 1-3.

33. F. Jiang, Z. Gu, C. Sun, and R. Ma, "Dynamic user pairing and power allocation for NOMA with deep reinforcement learning," in *Proceedings of IEEE Wireless Communications and Networking Conference*, 2021, pp. 1-6.
34. M. Aldababsa, M. Toka, S. Gökçeli, G. K. Kurt, and O. Kucur, "A tutorial on non- orthogonal multiple access for 5G and beyond," *Wireless Communications and Mobile Computing*, Vol. 2018, 2018, pp. 1-24.
35. R. Shankar and R. K. Mishra, "S-DF cooperative communication system over time selective fading channels," *Journal of Information Science and Engineering*, Vol. 35, 2019, pp. 1223-1248

Ravi Shankar is working as an Associate Professor at Madanapalle Institute of Technology and Science, Madanapalle, India.

Ayaz Ahmad is an Assistant Professor of Department of Mathematics, National Institute of Technology, Patna.

Saminathan Veerappan is working as an Associate Professir at Karpagam College of Engineering, Coimbatore, India. His current research interests cover cooperative communication, D2D communication, IoT/M2M networks and network protocols.

Manoj Kumar Beuria is pursuing Ph.D. from KIIT University, India. His current research interests include wireless communication, information propagation analysis, NOMA, deep learning, machine learning.

Patteti Krishna is working as a faculty at Netaji Subhas University of Technology East Campus, New Delhi, India. His current research interests cover cooperative communication, D2D communication, IoT/M2M networks and network protocols.

Sudhansu Sekhar Singh is working as a Professor in School of Electronics Enginering, KIIT University, Bhubaneswar, India.

V.G. Krishnan is working as a Professor at CVR College of Engineering, Hyderabad, Telangana, India, 501510.