This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2019.2933762, IEEE Journal on Selected Areas in Communications

Joint Power Allocation and Channel Assignment for NOMA with Deep Reinforcement Learning

Chaofan He, *Student Member, IEEE,* Yang Hu, Yan Chen*, *Senior Member, IEEE,* Bing Zeng, *Fellow, IEEE*

***Abstract*—Non-orthogonal multiple access (NOMA) has been considered as a significant candidate technique for the next generation wireless communication to support high throughput and massive connectivity. It allows different users to be multiplexed on one channel through applying superposition coding at the transmitter and successive interference cancellation (SIC) at the receiver. To fully utilize the benefit of the NOMA technique, the key problem is how to optimally allocate resources, such as power and channels, to users to maximize the system performance. There have been some existing works on the power allocation for the single-carrier NOMA system. However, how to optimally assign channels in the multi-carrier NOMA system is still unclear. In this paper, we propose a deep reinforcement learning framework to allocate resources to users in a near optimal way. Specifically, we exploit an attention-based neural network (ANN) to perform the channel assignment. Simulation results show that the proposed framework can achieve better system performance, compared with the state-of-the-art approaches.**

***Index Terms*—Non-orthogonal multiple access (NOMA), channel assignment, power allocation, deep reinforcement learning, attention-based neural network.**

I. Introduction

With the rapid growth in the number of mobile devices and the volume of mobile data [1]–[6], there are dramatic demands to improve the capacity and connectivity of mobile communications. To satisfy such demands, the fifth generation (5G) communication is being standardized in recent years, where one key technique is the non-orthogonal multiple access (NOMA), a novel multiple access scheme promised to significantly improve the system throughput [7]–[9].

With traditional multiple access schemes, multiple users are allocated with orthogonal resources, such as time, frequency and codes, to avoid inter-user interference. However, with more and more mobile devices accessing to the wireless communication system, the benefit from resources exploited by the orthogonal multiple access schemes will saturate. To solve this problem, the NOMA technique introduces an extra power domain, which enables multiple users to be multiplexed on the same channel. Specifically, it uses superposition coding at the transmitter, and applies successive interference cancellation (SIC) at the receivers to differentiate signals from multiple users in the power domain. Therefore, the NOMA technique is able to impel the 5G communication system to achieve high data rate and massive connectivity.

To fully utilize the benefit of the NOMA technique, the key issue is to optimally perform joint channel assignment and power allocation with limited resources. Such joint channel assignment and power allocation problem has been proven to be NP-hard [10], [11], i.e., to derive the optimal solution, all possible combinations of channel assignment should be evaluated, which is computationally expensive if not infeasible. Therefore, researchers have proposed many suboptimal or heuristic approaches [10], [12]–[21] to resolve this optimization problem.

Since channel characteristics among different users can be complicated, conventional approaches may not be able to capture the underlying relationship among users. Meanwhile, the solution space to the optimization problem is huge, nonlinear searching procedures are ineluctable. Therefore, conventional approaches are not efficient and reliable enough to obtain good channel assignment, due to which the performance of NOMA system can still be quite limited. In recent years, machine learning as a promising technique has been incorporated in the wireless communication communication system design [22]–[25]. It improves the system performance by exploiting the nonlinear relationship in the training data.

However, how to derive the optimal solution to the channel assignment problem in the multi-carrier NOMA system is still unclear. Inspired from the power of machine learning, we consider utilizing machine learning techniques to resolve the channel assignment problem. The channel assignment problem can be re-formulated as a sequential decision-making process. At each step, one channel is assigned to a corresponding user according to the decision-making process. The process will be terminated until there is no available channel resource. The objective is to find the optimal process, i.e., channel assignment that maximizes the system performance. However, due to the absence of the optimal "labels" in the channel assignment, the supervised machine learning techniques are not applicable to the channel assignment problem. To revolve the problem, in this paper we propose a deep reinforcement learning based resource allocation scheme to maximize the performance of the multi-carrier NOMA system under the MSR and MMR metrics. With the benefit from deep reinforcement learning, the proposed scheme can explore different channel assignment processes, observe their corresponding rewards and discover

Manuscript received December 6, 2018; revised April 2, 2019; accepted May 20, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61672137 and Grant 61602090, in part by the 111 Project No. B17008, and in part by the Thousand Youth Talents Program of China (to Yan Chen).(*Corresponding author: Yan Chen.*)

C. He and Y. Chen are with the Center for Intelligent Networking and Communications (CINC), and the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, China, 611731 (emails: eecfhe@std.uestc.edu.cn, eecyan@uestc.edu.cn).

Y. Hu and B. Zeng are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, China, 611731 (emails:{yanghu, eezeng}@uestc.edu.cn).

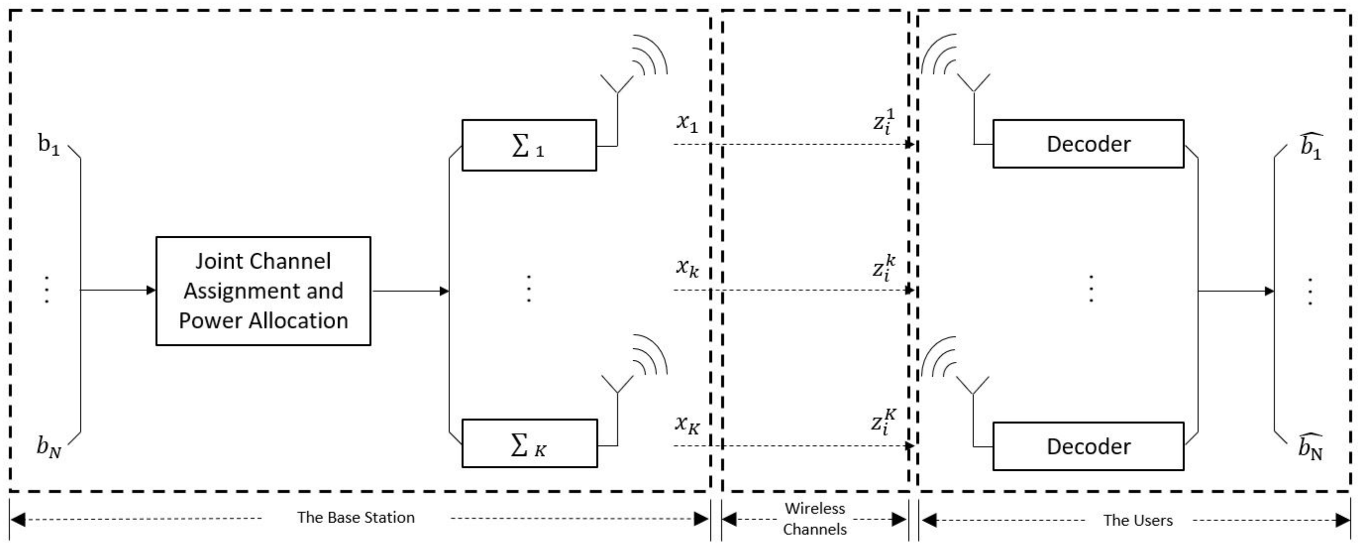


Fig. 1. The block diagram of the transmission and reception procedures of the NOMA system.

the behind heuristics.

Specifically, we formulate the joint channel assignment and power allocation as an optimization problem. Similar to JRA, we derive an optimal power allocation given channel assignment. Then, with the optimal power allocation, we propose to utilize the deep reinforcement learning to learn an optimal channel assignment policy. To capture the sequential relations between input and output of channel assignment problem, the policy network in the proposed framework applies a Transformer architecture [26]. It only uses an attention mechanism, instead of recurrence or convolution, to draw sequential dependencies in transduction works, which has been demonstrated to be able to achieve good performance with reasonable training time.

Finally, we compare the proposed framework with two other approaches, i.e., the joint resource allocation (JRA) in [18] and the exhaustive search (ES) method. Simulation results show that the proposed framework can achieve better system performance than JRA. Compared with ES, the proposed framework can achieve similar performance with much lower computational complexity.

The rest of paper is organized as follows. Section II briefly reviews the related works on resource allocation and machine learning for NOMA system design.The system model is introduced in detail in section III. The problem formulation on the joint channel assignment and power allocation is discussed in section IV. The optimal power allocation is derived in section V, while the deep reinforcement learning based channel assignment is discussed in section VI. The simulation results are shown in section VII and conclusions are drawn in section VIII.

II. Related works

A. *Resource Allocation for NOMA Systems*

How to optimally allocate resources, such as channel and power to users is an important topic in the NOMA system design. Therefore, researchers have proposed many approaches to resolve this optimization problem under different objective functions, such as maximizing sum rate (MSR) and maximizing minimal rate (MMR). The MSR dedicates to improve the overall data rate of the wireless communication system. In [12], an efficient power allocation and precoding design scheme is proposed for the single-carrier NOMA system to maximize the sum rate. It addresses the resource allocation problem through a two-step approach, where the sum rate maximization problem is first transformed into an equivalent form and then the suboptimal power allocation and complex precoding vectors are derived by iteratively using the minimization-maximization algorithm [13]–[16].

However, the overall data rate can be further improved by applying NOMA in the multi-carrier communication system with benefit from frequency diversity. The authors in [10] propose a suboptimal joint power and channel allocation algorithm to maximize the throughput performance of the multi-carrier NOMA system. The sum rate maximization problem is first converted to an equivalent problem by relaxing the power constraints, and then the power weights and assigned channels for users are derived through dynamic programming. A joint resource allocation algorithm is proposed in [17] to maximize the weighted system data rate through an iterative approach, where the authors first reformulate the maximization problem into a class of difference of convex function programming and then obtain the local optimal solution with successive convex approximation method [27]. In [18], a NOMA system that maximizes the weighted sum rate with individual user quality of service (QoS) constraints is considered. Assuming the number of users multiplexed on each channel is no more than two, the optimal power allocation is first obtained with the given channel assignment. Then by iteratively performing channel assignment and power allocation, a near optimal solution to the maximization problem is derived.

Different from MSR which neglects the users with bad channel condition, MMR can ensure the fairness among users.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSAC.2019.2933762, IEEE Journal on Selected Areas in Communications

The authors in [19] provide a low-complexity power allocation algorithm to maximize the fairness among users of the NOMA system. The maximization problem is first decomposed into a sequence of subproblems via a bisection iterative algorithm. For each subproblem, the solution with closed and semi-closed form is obtained. A proportional fairness scheduling scheme is proposed in [20] to maximize the minimal user rate, where the maximization problem is approximated by a convex optimization. In [21], the authors propose to maximize the minimal user rate with outage probabilities and power constraints. The decoding order for users is first derived, and then the power allocation is obtained by performing the Newton's method.

B. Machine Learning for Wireless Communication Systems

To enhance the system performance, machine learning has been incorporated in the wireless communication system design [22]–[25].

A robust and efficient deep learning scheme is proposed in [22] to learn the unknown channel state information (CSI) of the NOMA system to achieve better performance in terms of block error rate and sum rate. The scheme is composed of a pretraining network and a long short-term memory (LSTM) network. The pretraining network exploits restricted boltzmann machines to reduce the dimension of the input data and enhance the generalization ability, while the LSTM is employed to learn the CSI of the NOMA system via offline training and online training. The authors in [23] propose a fast reinforcement learning based power allocation scheme to improve the spectral efficiency of multiple-input and multiple-output (MIMO) NOMA system with interference from smart jammers. An anti-jamming MIMO NOMA transmission game is first formulated and the Stackelberg equilibrium is derived to reveal the impact of multiple antennas and channel state information. Then, to resolve the dynamic game, a Q-learning based power allocation scheme is exploited to allocate power to users against jamming attack. The authors in [25] propose a reinforcement learning scheme to learn the optimal feedback allocation in LTE network to maximize the system performance. They first build the feedback model and analyze the impact of feedback. Then a Q learning based reinforcement learning scheme is exploited to optimally allocate feedback. An actor-critic reinforcement learning approach is proposed in [24] to learn an optimal policy for user scheduling and resource allocation in HetNets to maximize the energy efficiency. The actor part uses the Gaussian distribution as the parameterized policy to generate continuous stochastic actions, while the critic part evaluates the value function with compatible function and helps the actor learn the gradient of the policy.

III. System Model

In this paper, we consider a downlink multi-carrier NOMA system, where the base station transmits data to multiple users over wireless channels. In such a system, the signals of different users are multiplexed on the wireless channels through channel assignment and power allocation. Then at the decoder, users recover the specific signals from each channel via SIC. The key problem here is how to allocate limited resources, such as power and channels, to multiple users to maximize the system performance.

The block diagram of the transmission and reception processes for the NOMA system is shown in Fig. 1, where we assume that there are N users and K channels. The total bandwidth is B_{tot}, and thus each channel is with bandwidth $B_c = B_{tot}/K$. Suppose there are N_k users multiplexed on the k^{th} channel. Let b_n denote the transmission symbols for the n^{th} user. Then, the base station multiplexes symbols of different users on each channel through channel assignment and power allocation module, and transmits them over wireless channels. The multiplexed signal on the k^{th} channel can be written as

$$x_k = \sum_{i=1}^{N_k} \sqrt{p_i^k} b_i, \tag{1}$$

where p_i^k is the power allocated to the i^{th} user transmitted on the k^{th} channel.

At the receiver, the distorted signals of users are received from each channel. For the n^{th} user, the received signal from the k^{th} channel is represented as

$$y_n^k = \sqrt{p_n^k} h_n^k b_n + \sum_{i=1, i \neq n}^{N_k} \sqrt{p_i^k} h_i^k b_i + z_n^k, \tag{2}$$

where h_n^k is the channel response between base station and the n^{th} user that considers both the path loss and shadowing effect, z_n^k denotes the additive white gaussian noise (AWGN) with zero mean and variance $\sigma_{z_k}^2$.

To reconstruct the signals for users on each channel, the decoder applies the SIC technique. Specifically, let $\Gamma_n^k = |h_n^k|^2/\sigma_{z_k}^2$ represent the channel-to-noise-ratio (CNR) of the n^{th} user on the k^{th} channel. Without loss of generality, let us assume the CNRs of users multiplexed on the k^{th} channel are ordered as $\Gamma_1^k > ... > \Gamma_n^k > ... > \Gamma_{N_k}^k$. According to the NOMA protocol, the users with lower CNR are assigned with more power, i.e., $p_1^k < ... < p_n^k < ... < p_{N_k}^k$. Hence, on the k^{th} channel, the n^{th} user is able to decode signals of users who are allocated with more power ($p_i^k > p_n^k$), while treating signals of users who are allocated with less power ($p_i^k < p_n^k$) as interference. Therefore, the signal to interference plus noise ratio (SINR) of the n^{th} user on the k^{th} channel could be written as

$$\gamma_n^k = \frac{p_n^k \Gamma_n^k}{1 + \sum_{i=1}^{n-1} p_i^k \Gamma_n^k}, \tag{3}$$

and the corresponding data rate is

$$R_n^k(\Gamma_n^k, p_1^k, ..., p_n^k) = B_c log_2(1 + \frac{p_n^k \Gamma_n^k}{1 + \sum_{i=1}^{n-1} p_i^k \Gamma_n^k}). \tag{4}$$

By performing SIC on each channel, a user has to decode the signals of other users in addition to its own signal. Thus, the hardware complexity and processing delay increase with the number of users on each channel. Practically, each channel is constrained to be allocated to two users [28], [29]. In this paper, we assume there are two users multiplexed on each channel, i.e., $N_k = 2, \forall k = 1, ..., K$. Let $\Gamma_1^k > \Gamma_2^k$, and then we can respectively derive the data rate for the two users on the k^{th} channel as

$$\begin{aligned} R_1^k(\Gamma_1^k, p_1^k, p_2^k) &= B_c log_2(1 + p_1^k \Gamma_1^k), \\ R_2^k(\Gamma_2^k, p_1^k, p_2^k) &= B_c log_2(1 + \frac{p_2^k \Gamma_2^k}{1 + p_1^k \Gamma_2^k}). \end{aligned} \tag{5}$$

IV. Problem Formulation

In this section, we discuss how to optimize the joint channel assignment and power allocation module of the NOMA system in Fig. 1. Let $\Gamma = \left\{\Gamma_1^1, \Gamma_2^1, ..., \Gamma_1^K, \Gamma_2^K\right\}$ and $P = \left\{p_1^1, p_2^1, ..., p_1^K, p_2^K\right\}$ be the channel assignment and power allocation set, respectively. In this paper, we focus on the following two system performance, i.e., maximizing the sum rate (MSR) and maximizing the minimal rate (MMR).

For the MSR performance metric, the corresponding objective function dedicates to improve the overall data rate of all users. Here, we consider the sum rate with QoS constraints as follows

$$\begin{aligned} \max_{\Gamma, P} \quad & \sum_{k=1}^{K} \left[R_1^k(\Gamma_1^k, p_1^k, p_2^k) + R_2^k(\Gamma_2^k, p_1^k, p_2^k)\right], \\ s.t. \quad & R_n^k \geq (R_n^k)_{min}, \ n = 1, 2, \ \forall k = 1, ..., K, \end{aligned} \tag{6}$$

where $(R_n^k)_{min}$ denotes the minimal data rate requirement for the n^{th} user on the k^{th} channel.

The objective function for the MMR performance metric aims to achieve the fairness among users, which can be written as

$$\max_{\Gamma, P} \quad \underset{k=1,...,K}{min} \left\{R_1^k(\Gamma_1^k, p_1^k, p_2^k), R_2^k(\Gamma_2^k, p_1^k, p_2^k)\right\}. \tag{7}$$

Suppose that the total power provided for the base station is P_T, and thus there is a power constraint for all users as follows

$$\sum_{k=1}^{K} (p_1^k + p_2^k) \leq P_T. \tag{8}$$

Due to the coupling between the power weights and assigned channel in (5), it is generally difficult to derive the optimal solution to the optimization problem on joint channel assignment and power allocation. To resolve this problem, we first derive the optimal power weights given channel assignment. Then, we propose a deep reinforcement learning framework to solve the channel assignment problem and thus jointly obtain a near optimal assigned channels and power weights for users.

V. Optimal power allocation

In this section, we illustrate how to conduct power allocation given the channel assignment for two performance metrics, MSR and MMR. The power allocation method we employ in this paper is similar to that in [18]. Therefore, in the following, we will skip the detailed derivation and directly illustrate the optimal power allocation solution.

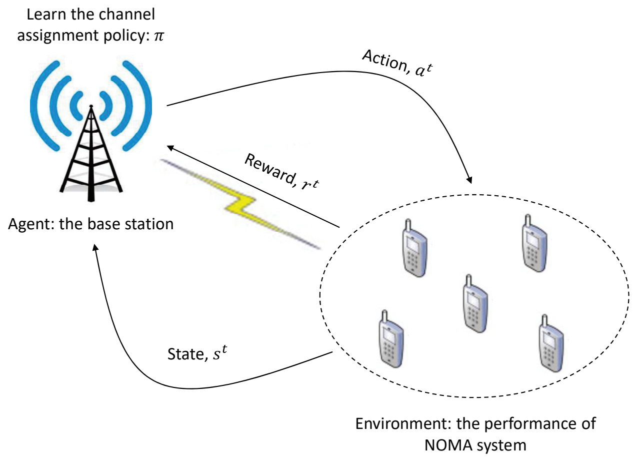

Learn the channel assignment policy: π
Action, a^t
Reward, r^t
Agent: the base station
State, s^t
Environment: the performance of NOMA system

Fig. 2. Reinforcement learning formulation for channel assignment of the NOMA system.

For MSR, the power allocation problem can be written as

$$\begin{aligned} \max_{P} \quad & \sum_{k=1}^{K} \left[R_1^k(p_1^k, p_2^k) + R_2^k(p_1^k, p_2^k)\right], \\ s.t. \quad & R_n^k \geq (R_n^k)_{min}, \ n = 1, 2, \ \forall k = 1, ..., K, \\ & \sum_{k=1}^{K} (p_1^k + p_2^k) \leq P_T, \\ & 0 \leq p_1^k \leq p_2^k, \ \forall k = 1, ..., K. \end{aligned} \tag{9}$$

Let $A_n^k = 2^{\frac{(R_n^k)_{min}}{B_c}}$ and assume $A_2^k \geq 2$, the solution to (9) can be written as follows

$$\begin{aligned} p_1^k &= \frac{\Gamma_2^k q_k - A_2^k + 1}{A_2^k \Gamma_2^k}, \\ p_2^k &= q^k - p_1^k, \end{aligned} \tag{10}$$

where q^k and γ_k are given as

$$\begin{aligned} q^k &= \left[\frac{B_c}{\lambda} - \frac{A_2^k}{\Gamma_1^k} + \frac{A_2^k}{\Gamma_2^k} - \frac{1}{\Gamma_2^k}\right]_{\gamma_k}^{\infty}, \\ \gamma_k &= \frac{A_2^k(A_1^k - 1)}{\Gamma_1^k} + \frac{A_2^k - 1}{\Gamma_2^k}, \end{aligned} \tag{11}$$

with λ satisfying $\sum_{k=1}^{K} q_k = P_T$

For MMR, the optimization on power allocation is formulated as follows

$$\begin{aligned} \max_{P} \quad & \underset{k=1,...,K}{min} \left\{R_1^k(p_1^k, p_2^k), R_2^k(p_1^k, p_2^k)\right\}, \\ s.t. \quad & \sum_{k=1}^{K} (p_1^k + p_2^k) \leq P_T, \\ & 0 \leq p_1^k \leq p_2^k, \ \forall k = 1, ..., K. \end{aligned} \tag{12}$$

The solution to (12) can be written as

$$\begin{aligned} p_1^k &= \frac{-(\Gamma_1^k + \Gamma_2^k) + \sqrt{(\Gamma_1^k + \Gamma_2^k)^2 + 4\Gamma_1^k (\Gamma_2^k)^2 q_k}}{2\Gamma_1^k \Gamma_2^k}, \\ p_2^k &= q^k - p_1^k, \end{aligned} \tag{13}$$

where $q^k = \frac{(Z(\lambda)\Gamma_2^k + \Gamma_1^k)(Z(\lambda)-1)}{\Gamma_1^k\Gamma_2^k}$, $Z(\lambda) = X + \sqrt{X^2 + \frac{B_c}{2\lambda\sum_{k=1}^{K} 1/\Gamma_1^k}}$, and $X = \frac{\sum_{k=1}^{K}(\Gamma_2^k - \Gamma_1^k)/(\Gamma_1^k\Gamma_2^k)}{4\sum_{k=1}^{K} 1/\Gamma_1^k}$.

VI. Deep Reinforcement Learning Framework for Channel Assignment

With the optimal power allocation derived in the previous section, in this section, we introduce a deep reinforcement learning framework to solve the channel assignment problem. Specifically, we first formulate the channel assignment problem into an optimization and describe how it can be represented as a reinforcement learning task. Then, we introduce the network designed for deep reinforcement learning framework, discuss the corresponding training algorithm in detail and analyze the computational complexity.

A. Channel Assignment Formulation

In this subsection, we formulate the channel assignment problem under two performance metrics into optimization problems, respectively. The optimization on channel assignment with the MSR objective can be written as

$$\max_{\Gamma} \sum_{k=1}^{K} \left[R_1^k(\Gamma_1^k) + R_2^k(\Gamma_2^k)\right], \tag{14}$$

and the channel assignment optimization that applies the MMR objective can be represented as

$$\max_{\Gamma} \min_{k=1,\ldots,K} \left\{R_1^k(\Gamma_1^k), R_2^k(\Gamma_2^k)\right\}. \tag{15}$$

The optimization problems in (14) and (15) is computationally expensive since all possible combinations on channel assignment have to be evaluated. To resolve this challenge, in the following, we propose a deep reinforcement learning framework to optimize channel assignment for the NOMA system.

B. Deep Reinforcement Learning Formulation

In this subsection, the optimization on channel assignment is modeled as a reinforcement learning task, which consists of an agent and environment interacting with each other, as shown in Fig 2. Specifically, the base station is treated as the agent and the performance of NOMA system is the environment. The action taken by the agent is based on the collective information on channel condition from users. Then at each step, based on the observed state s^t of the environment, the agent picks an action a^t from the action space A to assign channels to users according to the channel assignment policy, π, where the policy is learned by an attention-based neural network (ANN). With the action, the environment evolves into a new state, s^{t+1}. The channel assignment process terminates when there is no extra channel resources. Then, with the obtained channel assignment, the optimal power allocation is conducted and the step reward, r^t can be computed and fed back to the agent. This reward is the objective data rate of the NOMA system.

1) State Space: The state of the environment is characterized by the channel information. Here, we represent the channel information as user-channel pairs. Let $s^t = (U_{e_t}, \Gamma^{c_t})$ $(1 \leq e_t \leq N,\ 1 \leq c_t \leq K)$ be the user-channel pair, i.e., the $(c_t)^{th}$ channel assigned to the $(e_t)^{th}$ user at step t. Hence, the state space contains NK states, denoted as $S = \left\{(U_1, \Gamma^1), ..., (U_1, \Gamma^K), ..., (U_N, \Gamma^1), ..., (U_N, \Gamma^K)\right\}$.

2) Action Space: At each step, the agent takes an action, $a^t \in A$, which selects a channel for a user for transmission. However, to satisfy the requirements of channel assignment for the NOMA system, such action is constrained, i.e., the two users chosen by an action for each channel should be different. After N actions, the channel assignment process is completed.

3) Reward Function: In the NOMA system, the user's data rate is an important metric to evaluate the system performance. At the end of each epoch, the power allocation is conducted with given channel assignment. Then, the reward for step l that selects the state $s^l \in S$ is defined as

$$r^l = \begin{cases} R_1^{c_l}(s^l), \text{if user of } s^l \text{ is first assigned to the channel of } s^l, \\ R_2^{c_l}(s^l), \text{otherwise.} \end{cases} \tag{16}$$

To evaluate the overall system performance at the end of each epoch, the reward in previous states should also be taken into consideration. Therefore, we respectively define the returned reward of each epoch for MSR and MMR as follows

$$\begin{aligned} G_N^{msr} &= \sum_{l=1}^{N} r^l, \\ G_N^{mmr} &= \min_{l=1,\ldots,N} \left\{r^l\right\}. \end{aligned} \tag{17}$$

The objective of the NOMA system under two performance metrics is to maximize the returned award in (17), respectively. Through solving the maximization problem, we can derive the optimal channel assignment policy.

C. Attention-based Neural Network

In this subsection, we model the channel assignment policy by an ANN and introduce the design of ANN in detail.

After the agent takes action a^{t-1} based on the learned policy, the state of the environment transits from s^{t-1} to s^t, which can be characterized by the state transition probability $\pi : p(s_i|S, s_{i-1})$. To derive the state transition probability, we apply the proposed ANN to parameterize it as $p_\theta(s_i|S, s_{i-1})$. Motivated by the sequence transduction model [26], the proposed ANN employs an encoder-decoder structure, as shown in Fig. 3. The encoder computes the embedding of state space as E^s, and the decoder outputs probability distribution over all states at each step.

1) Encoder: The encoder applies the structure proposed in [26] without the positional encoding. For each d_{in}-dimensional input ($d_{in} = 2$ in this paper), the encoder linearly projects it into an initial d_e-dimensional output. Then, through L identical layers, the result embedding with dimension d_e of each state is computed, where each layer is composed

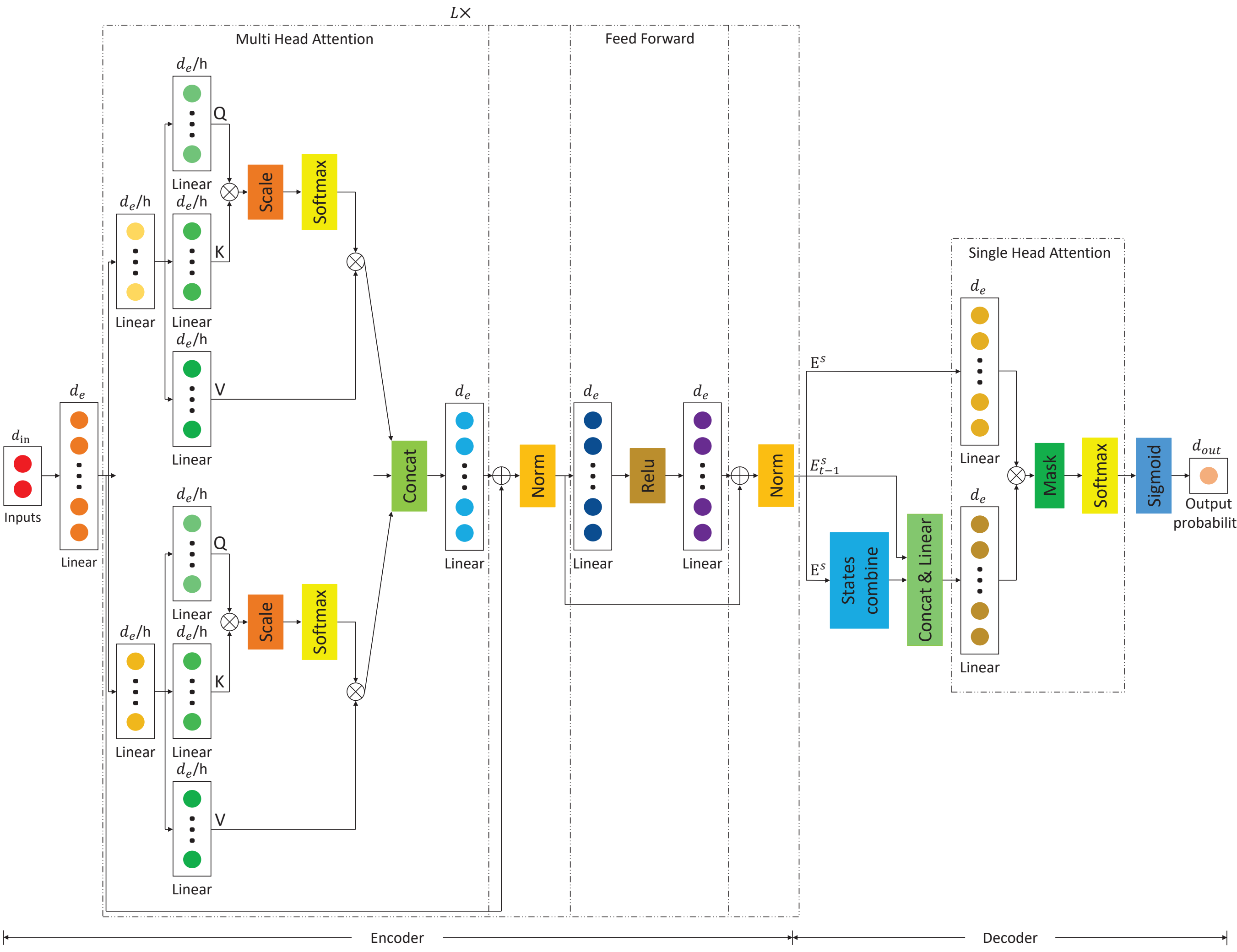


Fig. 3. The structure of the proposed attention-based neural network (ANN).

of two sub-layers. The first sub-layer is the Multi Head Attention layer with heads $h = 8$. It concatenates the h heads derived from the [26]. By performing linear transformation, the concatenated heads are mapped back to a d_e-dimensional output. The second sub-layer is the Feed Forward layer, which is composed of two linear transformations and a RELU activation in between. Followed by each sub-layer, the residual connection and batch normalization are appended sequentially.

2) Decoder: The decoder applies the single attention mechanism to compute the output probability, $p_\theta(s_t|S, s_{t-1})$ at each step. For the embedding of state space derived from the encoder, it linearly combines all states into one d_e-dimensional general state, e^g. Then, at each step of the decoding process, the embedding of the last state and the general state are concatenated and linearly projected to one decoding state, denoted as e^d. At the Single Head Attention layer, it computes the compatibility of each state with the decoding state and masks the states that satisfies the following conditions: (1) the states have already been selected ; (2) the states are not reachable due to the constraints of the action space. Specifically, the key (K) and the query (Q) come from the last state and the decoding state, respectively, which are represented as

$$\begin{aligned} K &= E^s W^K, \\ Q &= e^d W^Q, \end{aligned} \tag{18}$$

where $W^K \in R^{d^e \times d^e}$ and $W^Q \in R^{d^e \times d^e}$.

After masking the states specified above, it calculates the compatibility for residual state as

$$Compatibility = softmax(mask(KQ^T)), \tag{19}$$

where $K \in R^{NK \times d^e}$ and $Q \in R^{1 \times d^e}$.

Finally, through the sigmoid layer, the state transition probability, $p_\theta(s^t|S, s^{t-1})$ is obtained.

D. Training Algorithm

The proposed ANN is trained in an epochal setting. In each epoch, the state, i.e., the user-channel pair, is selected step by step using bootstrap sampling method according to the output probability derived from the ANN. The epoch terminates until all channels are assigned. Therefore, the solution to the channel assignment problem can be represented as $\zeta = \{s^1, ..., s^N\}$, which is a combination of states. The conditional probability of solution ζ given state space, S, can be written as follows

$$p_\theta(\zeta|S) = \prod_{i=1}^{N} p_\theta(s^t|S, s^{t-1}). \tag{20}$$

The loss of the proposed ANN for two performance metrics is defined as the averaged reward over multiple channel assignment solutions on the base of (17), which can be written as

$$Loss(\boldsymbol{\zeta}|S) = E_{\boldsymbol{\zeta}}[G_N^{msr}(\boldsymbol{\zeta})] \ or \ E_{\boldsymbol{\zeta}}[G_N^{mmr}(\boldsymbol{\zeta})]. \tag{21}$$

To derive the gradient of ANN, we use a variation of the reinforcement estimator [30], which introduces a baseline to reduce the gradient variance and reinforce the policy towards better direction. The baseline here is defined as a neural network with the same structure of the ANN, and initialized by the parameter of the ANN, denoted as $\theta^{bl} \leftarrow \theta$. According to [30], the gradient of the ANN is computed as

$$\nabla_\theta Loss(\boldsymbol{\zeta}|S) = E_{\boldsymbol{\zeta}}\left[(Loss(\boldsymbol{\zeta}|S) - Loss(\boldsymbol{\zeta}^{bl}|S)\nabla log p_\theta(\boldsymbol{\zeta}|S)\right], \tag{22}$$

where $\boldsymbol{\zeta}^{bl}$ is the solution derived from baseline, which greedily chooses the state with maximal probability at each step of an epoch.

Therefore, at the end of each epoch, the ANN updates its parameter according to the derived gradient in (22). The optimizer here we used is Adam [31]. Then we evaluate the system performance of the proposed ANN and the baseline respectively on the validation dataset. If ANN can achieve better performance than the baseline, the baseline replaces its parameters by that of ANN. Otherwise, the baseline keeps its parameters.

The training algorithm stops when the ANN cannot successively outperform the baseline over T_s epochs. Then the learned ANN model, i.e., the baseline is saved for test. The training algorithm is described in Algorithm 1. The dimension of vector in Algorithm 1 is equal to the batch size. Finally, with the learned channel assignment policy, i.e., the ANN model, we can derive a near optimal solution to the optimization on channel assignment in (14) and (15).

Algorithm 1 Training algorithm for channel assignment

```
Input: State space, $S$; The initialized ANN; Batch size, $B$
Output: The learned ANN; Channel Assignment; Power Allocation,
for each epoch do
    for each step do
        $p_\theta(\boldsymbol{s}^t|S, \boldsymbol{s}^{t-1}) \leftarrow$ output probability of ANN
        $\boldsymbol{\zeta}_t \leftarrow$ bootstrap sampling based on $p_\theta(\boldsymbol{s}^t|S, \boldsymbol{s}^{t-1})$
        $\boldsymbol{\zeta}_t^{bl} \leftarrow$ greedy sampling based on $p_\theta(\boldsymbol{s}^t|S, \boldsymbol{s}^{t-1})$
    end for
    if stopping criteria is not satisfied then
        $\theta \leftarrow Adam(\theta, \nabla_\theta Loss(\boldsymbol{\zeta}|S))$
        if $Loss(\boldsymbol{\zeta}|S) < Loss(\boldsymbol{\zeta}^{bl}|S)$ then
            $\theta^{bl} \leftarrow \theta$
        end if
    else
        save the learned ANN
        break
    end if
end for
```

E. Complexity Analysis

From the above discussions, we can see that the training algorithm is composed of an attention-based neural network, a baseline neural network and state space. In the following, we will derive the time and space complexity of the training algorithm.

The state space needs some space to be stored, hence the corresponding space complexity is $O(NK)$. As the ANN and baseline share the same encoder-decoder structure, the space and time complexity for them are identical.

The encoder consists of an embedding layer and L identical layers. The space complexity for the embedding layer is $O(d_{in}d_e)$. Each identical layer is composed of two sub-layers, i.e., a Multi Head Attention layer and a Feed Forward layer. According to [26], the space complexity for the Single Head Attention layer and the Multi Head Attention layer are similar, which is $O((NK)^2 d_e)$, and the Feed Forward layer has a space complexity of $O(d_e^2)$. Therefore, the space complexity for the encoder is written as

$$O(L(NK)^2 d_e + Ld_e^2). \tag{23}$$

For the decoder, since the Single Head Attention layer restricts attention to only last state, the corresponding space complexity is reduced to $O(NKd_e)$. The complexity for the linear layer and sigmoid layer are calculated as $O(d_e^2)$ and $O(d_e)$, respectively. Hence, the space complexity of the decoder is represented as

$$O(d_e^2 + NKd_e). \tag{24}$$

Therefore, the space complexity of ANN and baseline is represented as

$$O(L(NK)^2 d_e + Ld_e^2) + O(d_e^2 + NKd_e). \tag{25}$$

The time complexity of the ANN and baseline can be written as

$$O(TNL(NK)^2 d_e + TNLd_e^2), \tag{26}$$

where T is the number of training epoches and N is the number of channel assignment steps in an epoch, i.e, the number of users in the NOMA system.

Hence, the the overall space complexity of our training algorithm is

$$O(2L(NK)^2 d_e + 2Ld_e^2) + O(2d_e^2 + 2NKd_e), \tag{27}$$

and the overall time complexity of our training algorithm is

$$O(2TNL(NK)^2 d_e + 2TNLd_e^2). \tag{28}$$

Note that the complexity of exhaustive search in channel assignment problem can be calculated as $O(C_{NK}^N)$, which denotes all qualified combinations that chooses N states from state space containing NK states. Therefore, we can find the complexity of the proposed training algorithm is much lower.

VII. Simulation Results

In this section, we conduct multiple simulations to evaluate the performance of the proposed deep reinforcement learning framework, by comparing with two other approaches: the joint resource allocation (JRA) method proposed in [18] and the exhaustive search (ES) method. Also, we conduct the parameter analysis to show the influence of the parameters in the proposed ANN on the training process and results.

A. Simulation Settings

In the simulations, we assume the base station is located at the center of cell and the N users are randomly distributed around it ranging form $50m$ to $300m$. The minimal distance between users is set to $30m$. The total power provided for the base station is $P_T = 2 \sim 12 Watt$. The total bandwidth offered to the NOMA system is $B_{tot} = 5MHz$. The channel response of the k^{th} channel between the n^{th} user and the base station is specified as

$$h_n^k = g_n^k d_n^{-\alpha}, \tag{29}$$

where g_n^k follows the Rayleigh distribution, d_n is the distance between the n^{th} user and the base station, and $\alpha = 2$ is the path loss coefficient. The variance of channel noise is defined as $\sigma_{z_k}^2 = B_{tot}N_0/K$ for $\forall k = 1, ..., K$, with $N_0 = -170dbm$. The minimal user rate requirement in (6) is set to $R_n^k = 2\ bps/Hz,\ \forall k = 1, ..., K,\ \ n = 1, 2$.

The ANN for policy learning in the proposed framework is set up as follows. The parameters, such as weights and biases of the ANN are initialized to be uniformly distributed within $(-1/\sqrt{d_{in}}, 1/\sqrt{d_{in}})$, where $d_{in} = 2$ is the input dimension. For a trade-off between the quality and the complexity, we use $L = 3$ identical layers at the encoder, and the embedding dimension d_e is set to 32. The T_s for stopping criteria is set to 5.

The training and validation datasets are randomly generated at each epoch consisting of 40000 and 2000 instances, respectively. The test dataset contains 2000 instances. Each instance is composed of NK states, i.e., the user-channel pairs. The batch size applied for training is set to 400.

B. Performance Comparison

In this subsection, we compare the proposed framework with two other approaches, JRA and ES. In the following, the proposed Joint Resource Allocation Framework using Deep Reinforcement Learning is called as “JRA-DRL”. In JRA, the solution to the optimization on power allocation given channel assignment is first derived. Then with optimal power allocation, JRA conducts channel assignment and power allocation iteratively to find the power and channels allocated to each user. The ES method is a direct extension of JRA. It carries out the power allocation in the same way as that in JRA. For channel assignment, instead of the iterative optimization used by JRA, it exhaustively searches all combinations of channel assignment and finds the one that can maximize the objective data rate of the NOMA system.

The sum rate performance comparisons versus the power of base station with $N = 10$ users in the NOMA system are shown in Fig. 4. We can see that the sum rate achieved by the JRA-DRL is higher than that by JRA, which means that compared with JRA, the JRA-DRL is able to find better channel assignment for the NOMA system. From Fig. 4, we can also observe that the achieved sum rate by both the JRA-DRL and JRA methods increases with the power supplied to base station, but the increment becomes smaller when the power supply is larger. Such a phenomenon is because that according to (5), the channel data rate increases as power allocated to users multiplexed on that channel increases, but the benefit will saturate with the power supply is large.

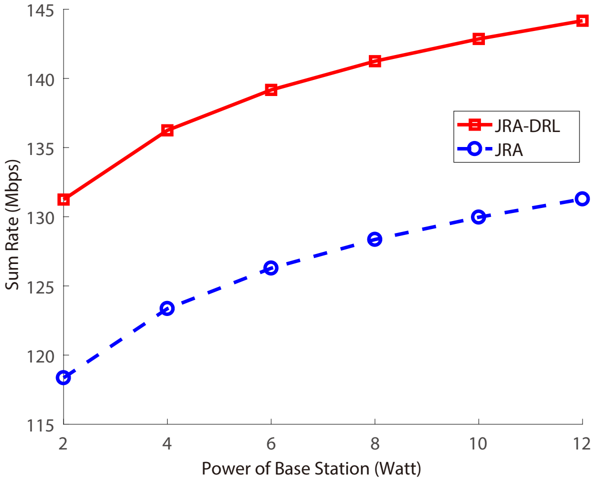


Fig. 4. The system sum rate performance comparisons versus the power of base station with $N = 10$ users in the NOMA system.

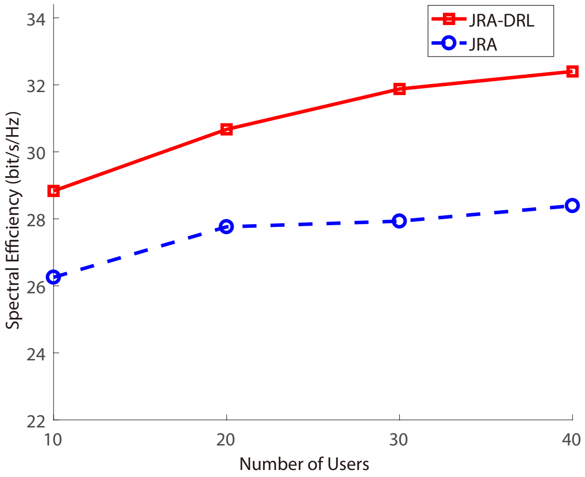


Fig. 5. The spectral efficiency versus different number of users with $P_T = 12Watt$.

The spectral efficiency of the NOMA system under the MSR performance metric versus different number of users when the power supplied to the base station is fixed as $P_T = 12Watt$ is shown in Fig 5. We can see that with the increase of users in the NOMA system, the spectral efficiency realized by the JRA-DRL and JRA methods increases, and JRA-DRL achieves much higher spectral efficiency than that of JRA.

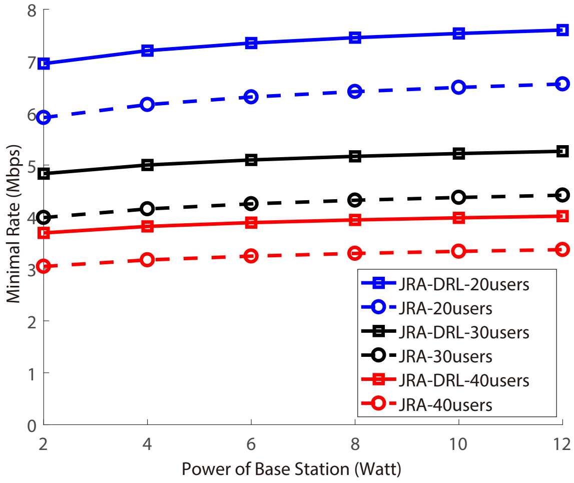


Fig. 6. The minimal rate performance via different methods versus the power offered to base station when different number of users are deployed in the NOMA system.

This is mainly because that the JRA-DRL can find better combination of channel assignment than JRA.

We also evaluate the minimal rate performance via different methods versus the power offered to base station when different number of users are deployed in the NOMA system, and the result is shown in Fig. 6. We can see that the minimal rate increases with the power of base station. This is reasonable since the rate will be larger when more power is available. We can find that at any fixed power supplied to base station, the minimal rate achieved by the JRA-DRL is always larger than that by JRA no matter how many users are deployed in the system. This is because the JRA-DRL is able to better assign channels than JRA under the MMF performance metric. We can also observe that the minimal rate of system reduces with the increase of the deployed users when the power supplied to base station is fixed. This is because that to realize MMF that assures the fairness among users, the minimal use rate would decrease as the number of users in the NOMA system increases when the total power supply is limited.

Due to the high computational complexity in conducting ES to find optimal combination of channel assignment, we set $N = 6$ to compare the objective rate performance with ES at different power configuration of base station under both MSR and MMF performance metrics. The results are shown in Fig. 7. We can see that under both performance metrics, the JRA performs the worst. By exploiting better channel assignment, the JRA-DRL achieves higher objective data rate, which is almost the same as that realizes by ES.

C. Parameter Analysis

In this subsection, we conduct parameter analysis to the proposed ANN. To facilitate comparison, we in the following assume there are $N = 40$ users and $K = 20$ channels in the NOMA system.

The influence of the batch size is shown in Fig. 8, where the learning rate is set to 0.01. We can see that the sum rate of NOMA system increases with the batch size. This is partly because with a larger batch size, the ANN can explore more channel assignment processes at the same time, which is helpful to discover the deep relations and output a better channel assignment policy network. We can also observe that the results will converge quickly when a large batch size is applied.

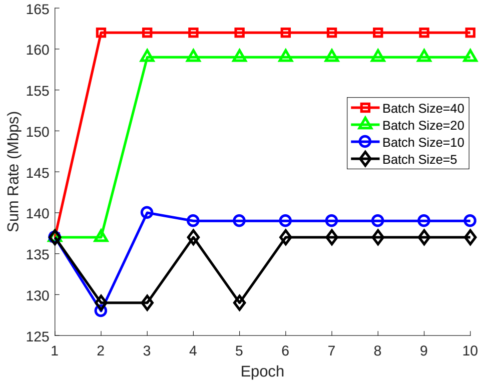


Fig. 8. Convergence of the channel assignment policy network with different batch size.

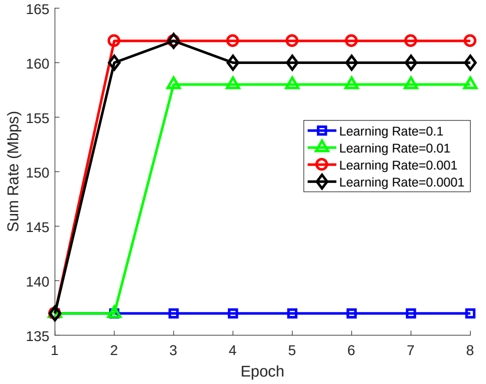
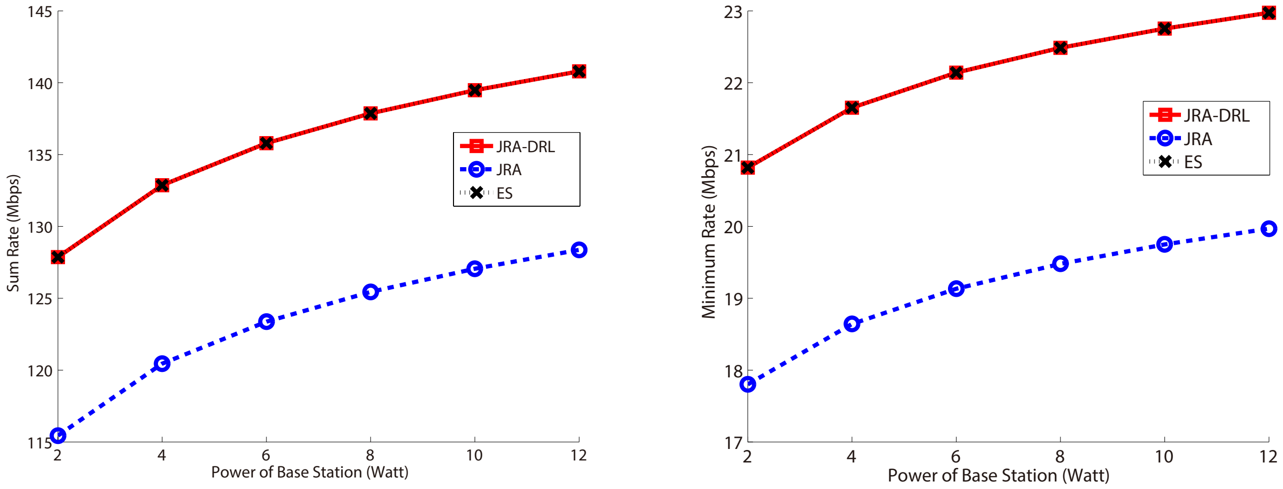


Fig. 9. Convergence of the channel assignment policy network with different learning rate.

In Fig. 9 shows the influence of the learning rate, where the batch size is set to 20. We can see that when the learning rate is 0.1 and 0.01, the ANN cannot learn a good channel assignment policy. With a learning rate of 0.001, the results will converge quickly to a good channel assignment policy. If the learning rate goes even smaller, e.g., 0.0001, the ANN converges quickly but with a slightly worse performance.

VIII. Conclusion

In this paper, we propose a deep reinforcement learning based resource allocation scheme to maximize the performance of the multi-carrier NOMA system. The joint channel assignment and power allocation problem is first formulated into an optimization problem. To resolve the optimization problem, we first derive a closed-form solution to the power allocation problem given channel assignment. Then, with the optimal power weights, a deep reinforcement learning framework, which utilizes an attention-based neural network,



(a)

(b)

Fig. 7. Comparisons on objective rate performance among three approaches at different power configuration of base station under different performance measures with $N = 6$: (a) under the MSR performance metric ; (b) under the MMF performance metric.

is proposed to address the channel assignment problem. The attention-based neural network exploits an encoder-decoder structure, where the encoder computes an embedding of state space and the decoder outputs probability distribution over all states at each step. In simulation results, we compare the proposed framework with two other approaches and show that the proposed framework can achieve better system performance under different performance metrics.

References

[1] J. G. Andrews *et al.*, “What will 5G be?” *IEEE J. Sel. Areas in Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[2] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, “5G network capacity: Key elements and technologies,” *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 71–78, 2014.

[3] Z. Chang, Y. Hu, Y. Chen, and B. Zeng, “Cluster-oriented device-to-device multimedia communications: Joint power, bandwidth and link selection optimization,” *IEEE Trans. on Veh. Technol.*, vol. 67, no. 2, pp. 1570–1581, 2018.

[4] Y. Chen, B. Wang, Y. Han, H. Lai, Z. Safar, and K. J. R. Liu, “Why Time Reversal for future 5G wireless?” *IEEE Signal Processing Mag.*, vol. 33, no. 2, pp. 17–26, March 2016.

[5] C. He, H. Wang, Y. Hu, Y. Chen, X. Fan, H. Li, and B. Zeng, “Mcast: High-quality linear video transmission with time and frequency diversities,” *IEEE Trans. on Image Processing*, vol. 27, no. 7, pp. 3599–3610, 2018.

[6] N. Li, Y. Hu, Y. Chen, and B. Zeng, “Lyapunov optimized resource management for multiuser mobile video streaming,” *to appear in IEEE Trans. on Circuits and Syst. for Video Technol.: 10.1109/TCSVT.2018.2850445*.

[7] P. Wang, J. Xiao, and P. Li, “Comparison of orthogonal and non-orthogonal approaches to future wireless cellular systems,” *IEEE Veh. Technol. Mag.*, vol. 1, no. 3, pp. 4–11, 2006.

[8] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, and T. Nakamura, “Concept and practical considerations of non-orthogonal multiple access (noma) for future radio access,” in *Intelligent Signal Processing and Communications Systems (ISPACS)*, 2013, pp. 770–774.

[9] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, “Non-orthogonal multiple access (NOMA) for cellular future radio access,” in *Proceedings of the 77th IEEE Vehicular Technology Conference, VTC Spring 2013, Dresden, Germany, June 2-5, 2013*. IEEE, 2013, pp. 1–5.

[10] L. Lei, D. Yuan, C. K. Ho, and S. Sun, “Joint optimization of power and channel allocation with non-orthogonal multiple access for 5g cellular systems,” in *IEEE Global Communications Conference (GLOBECOM), San Diego, CA, USA*, Dec. 2015, pp. 1–6.

[11] Y. Liu and Y. Dai, “On the complexity of joint subcarrier and power allocation for multi-user OFDMA systems,” *IEEE Trans. Signal Processing*, vol. 62, no. 3, pp. 583–596, 2014.

[12] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, “A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems,” *IEEE Trans. Signal Processing*, vol. 64, no. 1, pp. 76–88, 2016.

[13] D. R. Hunter and K. Lange, “A tutorial on mm algorithms,” *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[14] D. R. Hunter and R. Li, “Variable selection using mm algorithms,” *Annals of Statistics*, vol. 33, no. 4, pp. 1617–1642, 2005.

[15] A. J. Smola, S. V. N. Vishwanathan, and T. Hofmann, “Kernel methods for missing variables,” in *Proc. of the 10th International Workshop on Artificial Intelligence and Statistics*, 2005, pp. 325–332.

[16] P. Stoica and Y. Selen, “Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher,” *IEEE Signal Processing Mag.*, vol. 21, no. 1, pp. 112–114, Jan. 2004.

[17] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, “Optimal joint power and subcarrier allocation for MC-NOMA systems,” in *IEEE Global Communications Conference (GLOBECOM), Washington, DC, USA*, Dec. 2016, pp. 1–6.

[18] J. Zhu, J. Wang, Y. Huang, S. He, X. You, and L. Yang, “On optimal power allocation for downlink non-orthogonal multiple access systems,” *IEEE J. on Sel. Areas in Commun.*, vol. 35, no. 12, pp. 2744–2757, 2017.

[19] S. Timotheou and I. Krikidis, “Fairness for non-orthogonal multiple access in 5G systems,” *IEEE Signal Processing Lett.*, vol. 22, no. 10, pp. 1647–1651, 2015.

[20] J. Choi, “Power allocation for max-sum rate and max-min rate proportional fairness in NOMA,” *IEEE Commun. Lett.*, vol. 20, no. 10, pp. 2055–2058, 2016.

[21] J. Cui, Z. Ding, and P. Fan, “A novel power allocation scheme under outage constraints in NOMA systems,” *IEEE Signal Processing Lett.*, vol. 23, no. 9, pp. 1226–1230, 2016.

[22] G. Gui, H. Huang, Y. Song, and H. Sari, “Deep learning for an effective nonorthogonal multiple access scheme,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, 2018.

[23] 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 Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3377–3389, 2018.

[24] Y. Wei, F. R. Yu, M. Song, and Z. Han, “User scheduling and resource allocation in hetnets with hybrid energy supply: An actor-critic reinforcement learning approach,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 1, pp. 680–692, 2018.

[25] A. Chiumento, C. Desset, S. Pollin, L. Van der Perre, and R. Lauwereins, "Impact of csi feedback strategies on lte downlink and reinforcement learning solutions for optimal allocation," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 550–562, Jan 2017.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *CoRR*, vol. abs/1706.03762, 2017. [Online]. Available: http://arxiv.org/abs/1706.03762

[27] Q. T. Dinh and M. Diehl, "Local convergence of sequential convex programming for nonconvex optimization," in *Recent Advances in Optimization and its Applications in Engineering*, 2010, pp. 93–102.

[28] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5g nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, 2016.

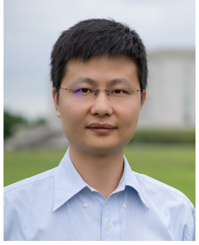
[29] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5g systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, 2015.

[30] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," in *Proc. of the 12th International Conference on Neural Information Processing Systems (NIPS)*, Cambridge, MA, USA, 1999, pp. 1057–1063. [Online]. Available: http://dl.acm.org/citation.cfm?id=3009657.3009806

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

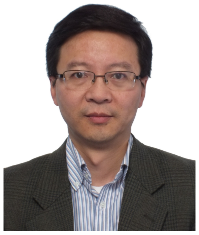
Chaofan He received B.S. degree from University of Electronic Science and Technology of China, Chengdu, in 2016. Now he is currently pursuing the Ph.D. degree in the School of Information and Communication Engineering at University of Electronic Science and Technology of China. His research interests include multimedia signal processing, wireless multimedia communication and networking, and multimedia social network.

Yang Hu received the B.S. and Ph.D. degrees in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2004 and 2009 respectively. She was with the University of Maryland Institute for Advanced Computer Studies as a research associate from 2010 to 2015. She is currently an associate researcher with the School of Information and Communication Engineering at the University of Electronic Science and Technology of China, Chengdu, China. Her current research interests include computer vision, machine learning and multimedia signal processing.



Yan Chen (SM'14) received the bachelor's degree from the University of Science and Technology of China in 2004, the M.Phil. degree from the Hong Kong University of Science and Technology in 2007, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2011. He was with Origin Wireless Inc. as a Founding Principal Technologist. Since Sept. 2015, he has been a full Professor with the School of Information and Communication Engineering at the University of Electronic Science and Technology of China. His research interests include multimedia, signal processing, game theory, and wireless communications.

He was the recipient of multiple honors and awards, including the best student paper award at the PCM in 2017, best student paper award at the IEEE ICASSP in 2016, the best paper award at the IEEE GLOBECOM in 2013, the Future Faculty Fellowship and Distinguished Dissertation Fellowship Honorable Mention from the Department of Electrical and Computer Engineering in 2010 and 2011, the Finalist of the Dean's Doctoral Research Award from the A. James Clark School of Engineering, the University of Maryland in 2011, and the Chinese Government Award for outstanding students abroad in 2010.



Bing Zeng (M'91-SM'13-F'16) received his BEng and MEng degrees in electronic engineering from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1983 and 1986, respectively, and his PhD degree in electrical engineering from Tampere University of Technology, Tampere, Finland, in 1991.

He worked as a postdoctoral fellow at University of Toronto from September 1991 to July 1992 and as a Researcher at Concordia University from August 1992 to January 1993. He then joined the Hong Kong University of Science and Technology (HKUST). After 20 years of service at HKUST, he returned to UESTC in the summer of 2013, through Chinas 1000-Talent-Scheme. At UESTC, he leads the Institute of Image Processing to work on image and video processing, 3D and multi-view video technology, and visual big data.

During his tenure at HKUST and UESTC, he graduated more than 30 Master and PhD students, received about 20 research grants, filed 8 international patents, and published more than 260 papers. Three representing works are as follows: one paper on fast block motion estimation, published in IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) in 1994, has so far been SCI-cited more than 1000 times (Google-cited more than 2200 times) and currently stands at the 8th position among all papers published in this Transactions; one paper on smart padding for arbitrarily-shaped image blocks, published in IEEE TCSVT in 2001, leads to a patent that has been successfully licensed to companies; and one paper on directional discrete cosine transform (DDCT), published in IEEE TCSVT in 2008, receives the 2011 IEEE CSVT Transactions Best Paper Award. He also received the best paper award at ChinaCom three times (2009 Xian, 2010 Beijing, and 2012 Kunming).

He served as an Associate Editor for IEEE TCSVT for 8 years and received the Best Associate Editor Award in 2011. He was General Co-Chair of VCIP-2016 and PCM-2017. He received a 2nd Class Natural Science Award (the first recipient) from Chinese Ministry of Education in 2014 and was elected as an IEEE Fellow in 2016 for contributions to image and video coding.