Online Power Control for 5G Wireless Communications: A Deep Q-network Approach

Changqing Luo, Jinlong Ji, Qianlong Wang, Lixing Yu, and Pan Li
Department of Electrical Engineering and Computer Science
Case Western Reserve University, Cleveland, OH 44106
Emails: {cxl881, jxj405, qxw204,lxy257, lipan}@case.edu.

***Abstract*—The popularity of smart mobile devices has resulted in the surged growth of mobile data traffic, which makes current cellular communication systems overloaded. To accommodate the data, the current wireless communication system is evolving to a 5G wireless communication system that employs multiple technologies to boost its system capacity. We notice that non-line-of-sight (NLOS) transmission is ubiquitous in wireless communication systems, and is even more common in 5G wireless communication systems due to using millimeter-Wave (mmWave) communications. Previous works employ beamforming techniques to enhance NLOS transmission performance but suffer from the high cost for controlling antennas. In this paper, we propose a dynamic transmission power control scheme for improving NLOS transmission performance. Particularly, we explore the control of UE association with MBS/SBSs and power allocation to maximize UEs' sum-rate under the constraints of transmission power and UEs' quality of service (QoS). To solve this maximization problem, we propose a deep Q-network (DQN) scheme, in which we apply a convolutional neural network (CNN) to estimate the Q-function offline and conduct a deep Q-learning online to find the control strategy. We offer simulation results to show the efficacy of the proposed scheme.**

***Index Terms*—5G wireless communications, deep learning, non-line-of-sight (NLOS) transmission, power control**

I. Introduction

The exploding growth and popularity of mobile devices like smartphones and tablets have resulted in sudden surges of various mobile applications [1]–[4], such as anywhere anytime online social networking, Internet of Things (IoT), vehicular communications, and smart health. These mobile applications daily create the massive amount of data. According to Cisco Visual Networking Index [5], the volume of mobile data traffic will experience a 1000-fold growth by the year 2020.

However, current cellular communication systems, even the newly-developed LTE/LTE-advanced ones, have been nearly or already overloaded. As a consequence, accommodating the newly-increased mobile data traffic inevitably incurs severe degradation of communication performance in these current cellular communication systems. For example, with the ever-growing mobile devices and mobile data traffic, mobile users will be gradually suffering from data traffic congestion and low per-user throughput. Therefore, current cellular communication systems need to evolve to the next-generation (5G) wireless communication systems for providing higher system capacity and per-user data rate.

To boost system capacity, academia and industry have proposed various wireless technologies like massive MIMO (multiple-input-multiple-output) and millimeter-wave (mmWave) communications, and consider such technologies as ideal candidates for 5G wireless communications. For massive MIMO, by using large-scale antenna arrays, a base station (BS) can transmit high-speed data streams to multiple user equipment (UEs) with different spatial patterns concurrently. mmWave communications allow 5G wireless communication systems to exploit high-frequency band, i.e., 30 - 300 GHz, for providing UEs with high available bandwidth. Due to employing these technologies for supporting a great number of mobile devices, a 5G wireless communication system is a dense and large-scale system.

Due to utilizing high-frequency spectrum band, 5G wireless communications easily suffer from performance degradation incurred by obstacles. Basically, in a 5G wireless communication system, there exist two types of information transmissions: the line-of-sight (LOS) and the non-line-of-sight (NLOS). LOS transmission often happens when a transmitter is close to its corresponding receiver and NLOS transmission is very common since obstacles (e.g., buildings, trees, and hills) can be easily seen everywhere. NLOS transmission performance degrades sharply since radio signals are attenuated a lot after they penetrate large obstacles. Due to adopting mmWave communications for 5G wireless communications, the signal attenuation loss becomes even worse. As a consequence, NLOS transmission in 5G wireless communication systems is much more vulnerable to degrading communication performance.

To tackle this issue, researchers have proposed to employ beamforming techniques. Lin and Akyildiz [6] consider employing the beamforming technique to improve NLOS transmission performance. They allocate beamforming weights for UEs at each remote radio head and maximize UEs' sum-rate of the considered 5G wireless communication system. Renzo and Lu [7] try to enhance system performance by jointly considering a realistic channel model, cell association criteria, and directional beamforming. Turgut and Gursoy [8] offer an analytical framework for improving NLOS transmission performance in heterogeneous downlink mmWave cellular networks by considering directional beamforming and derive signal-to-interference-plus-noise ratio (SINR) coverage probability. Some other researchers also consider coordinating multiple antennas of multiple BSs to enhance NLOS transmission performance [9]. However, it is not cost-effective to control antennas for achieving the beamforming in wireless communication systems. Particularly, due to using large-scale antenna arrays, it is more difficult to achieve the beamforming in 5G wireless communication systems. We notice that so far the improvement of NLOS transmission performance with low cost in 5G wireless communication systems is still an open and challenging problem.

In this paper, we propose to dynamically control transmission power for improving NLOS transmission performance in 5G wireless communication systems. In particular, we aim to maximize the total data rate (i.e., sum-rate) achieved by all UEs in a 5G wireless communication system. Specifically, we explore UE association with MBSs/SBSs and power allocation and formulate the maximization problem of UEs' sum-rate under the constraints of transmission power and UEs' quality of service (QoS) requirements. We notice that the formulated optimization problem is a large-scale mixed integer nonlinear programming (MINLP) problem that is generally NP-hard [10]. To efficiently solve the formulated problem, we propose a deep Q-network (DQN) that is based on the reinforcement learning but applies a deep neural network to estimate the Q-function. In particular, the Q-function is estimated offline and the deep Q-learning is conducted online to derive the control strategy (i.e., UE association with MBSs/SBSs and power allocation), hence leading to the significant computation time reduction.

The rest of this paper is organized as follows. We briefly depict the considered system model in Section II. We then present the problem formulation in Section III. Afterwards, we describe Q-network based solution to the formulated problem in IV. Finally, we present simulation results in Section V, and conclude this paper in Section VI.

II. System Description

To simplify the presentation, we consider a simple scenario for 5G wireless communications, as shown in Fig. 1. In this scenario, we have a single MBS presented by m. Within the coverage area of this MBS, there are a set of SBSs denoted by $\mathcal{S} = \{1, 2, \cdots, S\}$ and a large number of UEs presented by $\mathcal{U} = \{1, 2, \cdots, U\}$, where $s \in \mathcal{S}$ and $u \in \mathcal{U}$ denote an SBS and a UE, respectively. We consider that all SBSs are carefully located to offer services and all UEs are randomly distributed within this coverage area. Moreover, we have a cloud server that connects with the MBS and SBSs via wired links like optical fiber cables. The learning operations are performed at this server.

We consider downlink information transmission in this scenario. In particular, we employ the CoMP transmission/reception technique [11] to coordinate multiple-path information transmission. Hence, a UE can receive (or transmit) signals from (or to) the MBS and multiple SBSs at the same time. Specifically, a UE has the following three information transmission cases: 1) a UE only receives information from the MBS; 2) a UE only receives information from one or several SBSs; and 3) a UE receives information from the MBS and one or several SBSs.

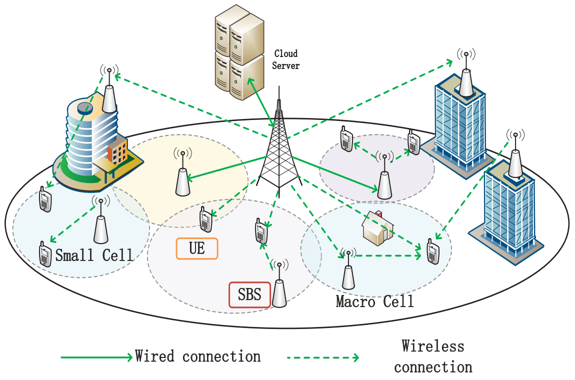

Fig. 1. The considered scenario for 5G wireless communications.

We evaluate link quality by channel gain. Let $\{h_{u,m}|u \in \mathcal{U}\}$ and $\{h_{u,s}|u \in \mathcal{U}, s \in \mathcal{S}\}$ denote the channel gain of UE/MBS (U-M) link between UE u and MBS m and UE/SBS (U-S) link between UE u and SBS s, respectively. In particular, we consider a block-fading channel model [12]. The channel gain is divided into discrete levels, each of which is related to a state. The state set is denoted by $\mathcal{C} = \{1, 2, \cdots, C\}$, where $|\mathcal{C}|$ is the number of states.

A UE receives signals transmitted over links from MBS and SBSs. When MBS m transmits signals $x_{u,m}$ to UE u over the U-M link between MBS m and UE u, UE u receives signals $y_{u,m}$ that can be expressed as follows:

$$y_{u,m} = \sqrt{P_{u,m} h_{u,m}} x_{u,m} + n_{u,m}, \tag{1}$$

where $P_{u,m}$ is the transmission power at MBS m and $n_{u,m}$ is the additive Gaussian white noise (AGWN) received by UE u. Likewise, when SBS s ($\forall s \in \mathcal{S}$) transmits signals $x_{u,s}$ to UE u over the U-S link between SBS s and UE u, UE u receives signals $y_{u,s}$ that can be expressed as follows:

$$y_{u,s} = \sqrt{P_{u,s} h_{u,s}} x_{u,s} + n_{u,s}, \tag{2}$$

where $P_{u,s}$ is the transmission power at SBS s and $n_{u,s}$ is the AGWN received by the UE. We consider that the noise power, denoted by σ^2, is identical for all links.

III. Problem Formulation

A. The transmission power constraints

In 5G wireless communication systems, the transmission power is always limited due to devices' capabilities. To achieve efficient 5G wireless communications, we need to allocate transmission power to active links. Let $x_{u,m}$ and $x_{u,s}$, binary variables, denote whether the U-M link between UE u and MBS m and the U-S links between UE u and SBS s (for $u \in \mathcal{U}$ and $s \in \mathcal{S}$) are active or not, and we have

$$x_{u,m} = \begin{cases} 1, & \text{if the U-M link between } u \text{ and } m \text{ is active} \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

for $u \in \mathcal{U}$, and

$$x_{u,s} = \begin{cases} 1, & \text{if the U-S link between } u \text{ and } s \text{ is active} \\ 0, & \text{otherwise} \end{cases}, \tag{4}$$

for $u \in \mathcal{U}$ and $s \in \mathcal{S}$.

Transmission power at MBS and SBSs is naturally limited due to the capability of their equipment. We denote by P_m^{mx} and P_s^{mx} the maximum transmission power an MBS and an SBS can provide, respectively. Besides, for each active link, no matter U-M and U-S links, its transmission power is also constrained. Let $P_{u,m}^{mx}$ and $P_{u,s}^{mx}$ denote the maximum transmission power of U-M or U-S links, respectively. Thus, for U-M and U-S links, we have

$$0 \leq x_{u,m}P_{u,m} \leq P_{u,m}^{mx}, \text{ and } 0 \leq x_{u,s}P_{u,s} \leq P_{u,s}^{mx}, \tag{5}$$

for $u \in \mathcal{U}$ and $s \in \mathcal{S}$. Furthermore, for all U-M and U-S links, we have the constraints as follows:

$$0 \leq \sum_{u=1}^{U} x_{u,m}P_{u,m} \leq P_m^{mx}, \text{ and } 0 \leq \sum_{u=1}^{U} x_{u,s}P_{u,s} \leq P_s^{mx}, \tag{6}$$

for $s \in \mathcal{S}$. Note, we consider that all SBSs have identical capabilities, i.e., P_s^{mx}'s (for $s \in \mathcal{S}$) are the same for all SBSs.

B. The UEs' QoS Requirements

UEs usually have QoS requirements for their information transmission. Thus, we consider to provide them with satisfactory services. In this paper, we use SINR to capture the QoS requirement, and hence formulate the SINR received by a UE. In practical 5G wireless communications, when MBS m transmits information to UE u, other UEs $u' \in \mathcal{U}$ nearby can be interfered since they can also receive the information at the same time. The received interference power at UE u' can be expressed by $|h_{u',m}|^2 P_{u,m}$. Similarly, when SBS s sends information to UE u, other UEs $u' \in \mathcal{U}$ nearby can also be interfered. We can find the received interference power at UE u' as $|h_{u',s}|^2 P_{u,s}$. Let $\gamma_{u,m}$ and $\gamma_{u,s}$ denote the received SINR for the U-M link between UE u and MBS m and the U-S link between UE u and SBS s, respectively. We can obtain $\gamma_{u,m}$ and $\gamma_{u,s}$ by using (7).

Let γ_u and γ_{min} denote the received SINR and the minimum SINR requirement of the UE u, respectively. Based on the study in [13], we can formulate the SINR received by UE u as follows:

$$\gamma_u = \gamma_{u,m} + \sum_{s=1}^{S} \gamma_{u,s}. \tag{8}$$

We consider that all UEs in the considered scenario have the identical minimum SINR requirement. To satisfy the QoS requirement, the SINR received by each UE needs to satisfy as follows:

$$\gamma_u \geq \gamma_{min}, \tag{9}$$

for $u \in \mathcal{U}$.

From the above description, we can notice that the SINR received by a UE largely depends on transmission power. When we increase transmission power for transmitting information to a UE, the power at the receiver of a UE will be grown up. However, this will result in the increasing interference power at other UEs' receiver in the neighborhood. As a consequence, the SINR received by other UEs' SINR will be getting low. Thus, we need to well control transmission power for all active links.

C. The Data Rate Achieved by a UE

Due to employing the CoMP transmission technique, the data rate achieved by a UE is the sum of the data rate achieved each active link owned by this UE. In the following, we first find the data rate that an active link is able to achieve, and then the data rate achieved by a UE.

When MBS m transmits information to UE u with transmission power $P_{u,m}$, we can find the data rate of the link as follows:

$$C_{u,m} = B_{u,m} \log_2(1 + \gamma_{u,m}), \tag{10}$$

where $B_{u,m}$ is the transmission bandwidth of a U-M link. Similarly, when SBS s transmits information to UE u, we can obtain the data rate of the link as follows:

$$C_{u,s} = B_{u,s} \log_2(1 + \gamma_{u,s}), \tag{11}$$

where $B_{u,s}$ is the transmission bandwidth of a U-S link. Thus, the sum-rate achieved by a UE u can be expressed as follows:

$$C_u = C_{u,m} + \sum_{s=1}^{S} C_{u,s}. \tag{12}$$

D. The Formulated Optimization Problem

So far, we have successfully characterized transmission power constraints, UEs' QoS requirements, and data rate achieved by a UE. We notice that these together affect the sum-rate achieved by all UEs. To maximize the sum-rate, we need to control them. Thus, we formulate the sum-rate maximization problem as follows:

$$\begin{aligned} \textbf{P-NLOS:} \quad & \textbf{max} \quad \sum_{u=1}^{U} C_u, \\ & \textbf{s.t.} \quad (5), (6), \text{and } (9). \end{aligned}$$

Through solving **P-NLOS**, we can find an optimal control strategy about UE association with MBS/SBSs and transmission power, i.e., $x_{u,m}$, $x_{u,s}$, $P_{u,m}$, and $P_{u,s}$, for $u \in \mathcal{U}$ and $s \in \mathcal{S}$. In particular, we can meet UEs' QoS requirements and enhance NLOS transmission performance by finding the solution to this formulated optimization problem.

IV. Dynamic Power Control via Deep Reinforcement Learning

We can notice that **P-NLOS** is a large-scale mixed integer nonlinear programming (MINLP) problem that is generally NP-hard. To efficiently solve **P-NLOS**, we propose a deep Q-network (DQN) that is based on reinforcement learning but applies a deep neural network to estimate the Q-function values. Specifically, the proposed approach combines two

$$\gamma_{u,*} = \frac{|h_{u,*}|^2 x_{u,*} P_{u,*}}{\sigma^2 + \sum_{u'=1,u'\neq u}^{U} \sum_{s=1}^{S} (|h_{u,m}|^2 x_{u',m} P_{u',m} + |h_{u,s}|^2 x_{u',s} P_{u',s})}, \text{ for } * \in \{m, s\}. \quad (7)$$

parts: one is a CNN, a variant of the deep neural network, which is performed offline for estimating Q-function values and the other is the deep Q-learning for online control of UE association with MBS/SBSs and power allocation. In what follows, we describe the reinforcement learning based dynamic control and DQN based dynamic control, respectively.

A. Reinforcement Learning based Dynamic Control

The DQN approach is evolved from the reinforcement learning. The difference between them is the deep neural network used to estimate Q-function values. Therefore, in this part, we first describe the reinforcement learning based dynamic control scheme for solving **P-NLOS**.

In our considered scenario, there is an agent located within the cloud. This agent is responsible for conducting the required learning process and outputting best strategies for online control of UE association with MBS/SBSs and power allocation. Specifically, this agent continually interacts with the considered 5G wireless communication system for a long term. It observes system states and finds its accumulated rewards it can obtain. At a decision epoch, it derives a control strategy, based on which an action is designated. After an action is taken, the system evolves into a new state that will be later presented to the agent.

To derive the best control strategy, we identify system state space, action space, and reward functions to construct the reinforcement learning process. In the following, we describe them one by one.

1) System state space: In 5G wireless communication systems, the state is characterized by all UEs' states. A UE's state is characterized by the channel state of all U-M and U-S links. Thus, at epoch k, we can characterize the state of UE u by $s_u^k = (s_{u,m}^k, s_{u,1}^k, s_{u,2}^k, \cdots, s_{u,S}^k)$, where $s_{u,*}^k \in \mathcal{C}$ is the state of the link between UE u and MBS (or SBS) $*$. As a result, we can have the state of the considered system by $s^k = (s_1^k, s_2^k, \cdots, s_U^k)$. Recall that each link has an identical channel state set $\mathcal{C}$. Note that, we can see that the system state space is very large in our consider 5G wireless communication system.

2) Action space: In our formulated problem, we need to control UE association with MBS/SBSs and transmission power allocation, which leads to a composite action. For UE u, its composite action at epoch k is $a_u^k = (x_{u,m}^k P_{u,m}^k, x_{u,1}^k P_{u,1}^k, \cdots, x_{u,S}^k P_{u,S}^k, x_{u,m}^k)$. Here, if $x_{u,*} \ \triangle_{u,*}= 0$, where $*$ presents symbols m and s, and $\triangle$ presents symbols P, we say $x_{u,*} = 0$ and $_{u,*} = 0$, otherwise $x_{u,*} = 1$ and $\triangle_{u,*}= x_{u,*} \ \triangle_{u,*}> 0$. For all UEs, the action space is $a^k = (a_1^k, a_2^k, \cdots, a_U^k)$.

3) Reward function: In a 5G wireless communication system, UEs' sum-rate is one of important parameters used to evaluate UEs' communication performance. Thus, we consider the sum-rate as the system reward that is originally defined as follows:

$$R(s^k, a^k) = \sum_{u=1}^{U} C_u^k, \quad (13)$$

where $R(a^k)$ is the reward function, showing that the reward highly depends on the system state and the action to be taken and C_u^k is the data rate received by UE u at epoch k.

Due to natural dynamics, the system state of a 5G wireless communication system transits over time. Thus, the agent interacts with this system continually. At epoch k, the agent knows s^k and derives an optimal strategy π by using a learning process. Then, the agent maps s^k to a^k based on the strategy π. As a result, the system takes an action a^k, and afterwards obtains reward $R(s^k, a^k)$. Finally, the system state transits to a new one s^{k+1} and the agent continues the above operations until it reaches a final epoch. The returned reward R_k is the accumulated and discounted reward that is calculated as follows:

$$R_k = \sum_{t=0}^{T} \alpha R(s^{k+t}, a^{k+t}), \quad (14)$$

where t is the epoch, T is the maximum epoch, and $\alpha \in (0, 1]$ is the discount factor.

The objective of the agent is to maximize the expected accumulated reward, i.e., $\max E[R_k|s^k]$. Through solving the maximization problem, it can derive the optimal strategy π. For solving this maximization problem, researchers have already developed two types of widely used methods: the value function based [14] and the policy based [15]. We adopt the value function based method in this paper.

Value function is a fundamental notion in reinforcement learning and Q-learning is a popular algorithm for learning state-action value functions. To find a state-action value function, we first define a state value function, denoted by $V_\pi(s^k)$. $V_\pi(s^k)$ is the accumulated reward for the strategy π. Thus, we have $V_\pi(s^k) = E[R_k|s^k]$. Due to the feature that the channel state transits independently in 5G wireless communication systems, we can rewrite this state value function as follows:

$$V_\pi(s^k) = r(s^k, \pi_k) + \alpha \sum_{s^{k'} \in s} P_{s^k s^{k'}}(\pi_k) V_\pi(s^{k'}), \quad (15)$$

where $P_{s^k s^{k'}}(\pi_k)$ is the state transition probability when strategy π_k is taken.

Furthermore, we define a Q-function $Q_\pi(s^k, a^k)$, a state-action value function, to characterize the expected reward for choosing action a^k at system state s^k by following the strategy π. We have $Q_\pi(s^k, a^k) = E[R_k|s^k, a^k]$. Due to the special feature, which the system state changes independently of the control of UE association and power allocation, we have $Q_\pi(s^k, a^k) = V_\pi(s^k)$, i.e.,

$$Q_\pi(s^k, a^k) = r(s^k, a_k) + \alpha \sum_{s^{k'} \in s} P_{s^k s^{k'}}(a_k) V_\pi(s^{k'}, a^{k'}). \tag{16}$$

At this moment, the agent needs to maximize the value of Q-function $Q_\pi(s^k, a^k)$. Through solving this optimization problem, this agent is able to find the best strategy.

We can notice that $r(s^k, a_k)$ and $P_{s^k s^{k'}}(a_k)$ are unknown in this optimization problem. To solve such types of optimization problems, Q-learning is one of most commonly used algorithms. For Q-learning based algorithms, Q-function is usually found in a recursive fashion using the available information $(s^k, a^k, r(s^k, a^k), s^{k+1}, a^{k+1})$. Thus, we can update $Q(s^k, a^k)$ as follows:

$$\begin{aligned} Q(s^{k+1}, a^{k+1}) = Q(s^k, a^k) + \beta(r(s^k, a^k) + \\ \alpha[\max_{a'^k} Q(s'^k, a'^k)] - Q(s^k, a^k), \end{aligned} \tag{17}$$

where β is the learning rate, and s'^k and a'^k are states and actions in system state space and action space at epoch k, respectively.

We can see that updating $Q(s^{k+1}, a^{k+1})$ needs to search over the whole system state space and action space. Due to the very large-scale system state space, the so-called curse of dimensionality, it needs to take a very long time to update the Q-function value. As a consequence, this reinforcement learning method does not work well in solving large-scale optimization problems.

B. DQN based Dynamic Control

To accelerate the learning process of the Q-learning method, we propose a deep reinforcement learning method, i.e., DQN. In this method, DNN is used to estimate the values of the Q-function $Q_\pi(s^k, a^k)$ offline and the deep Q-learning based online control is conducted based on the estimated values. In particular, since DNN based Q-learning causes the instability, we integrate the experience replay technique into the deep Q-learning.

To estimate the Q-function values, the current system state and historical system states are input into a CNN network and the output is $Q(ss^k, a^k|\theta^k)$ for a given weight vector θ^k, where $ss^k = (s^{k-N}, s^{k-N+1}, \cdots, s^k)$. Based on the experience replay, θ^k is updated for each time. The memory is used to store the experiences. We denote the experience at epoch k by $e^k = (ss^k, a^k, r(s^k, a^k), ss^{k+1})$ and the memory by $\mathcal{D} = (e^1, e^2, \cdots, e^k)$. Based on the experience replay technique, the agent selects at random an experience from $\mathcal{D}$ used to update the weight vector θ^k of the CNN.

We input a Q-function value $Q_\pi(s^k, a^k)$ into a deep Q-learning process to derive an optimal online control strategy for UE association with MBS/SBSs and power allocation. We employ a stochastic gradient descent method to update weight parameters. The learning process is as follows. First, we replace the value function $Q_\pi(s^k, a^k)$ by a deep Q-network with parameters θ^k, i.e., $Q_\pi(s^k, a^k) \approx Q(s^k, a^k, \theta^k)$. This approximation is then used to define the objective function as follows:

$$\begin{aligned} L(\theta^k) &= E[(r(s^k, a^k) + \beta \max_{a^{k+1}} Q(s^{k+1}, a^{k+1}, \theta^k) \\ &\quad -Q(s^k, a^k, \theta^k))^2]. \end{aligned} \tag{18}$$

Subsequently, we can find its gradients as follows:

$$\begin{aligned} \frac{\partial L(\theta^k)}{\partial \theta^k} &= -E[(r(s^k, a^k) + \beta \max_{a^{k+1}} Q(s^{k+1}, a^{k+1}, \theta^k) \\ &\quad -Q(s^k, a^k, \theta^k)) \frac{\partial Q(s^k, a^k, \theta^k)}{\partial \theta^k}]. \end{aligned} \tag{19}$$

At each epoch, we continue this process until it reaches T times. For each updating, we select θ^k according to the randomly chosen experiences from $\mathcal{D}$. The best control strategies are derived online. As a result, our proposed scheme is so efficient that it can adapt to the highly dynamic 5G wireless communications environment.

V. Simulation Results

A. Simulation Settings

In the simulations, we consider a 5G wireless communication system. In this system, we have 1 MBS, 10 SBSs, and 100 UEs. The 100 UEs are randomly dispersed over the MBS's coverage area that we set as a 200 m × 200 m square. Besides, the coverage area of a SBS is set as a 50 m × 50 m square and these SBSs are also randomly distributed in the considered area. To characterize the dynamics of the 5G wireless communications, we adopt the Manhattan mobility model [16] in this simulations and all UEs move at a low speed around the MBS's coverage area. The maximum transmission power of the MBS and SBS are $P_m^{mx} = 150\ mw$ and $P_s^{mx} = 50\ mw$, respectively. The maximum transmission power of the active U-M link and U-S link are $P_{um}^{mx} = 20\ mw$ and $P_{us}^{mx} = 15\ mw$, respectively. The noise power is set as $\sigma^2 = 10\ mw$. Similarly, the wireless links' bandwidth is set as 5 MHz for the U-M link and 25 MHz for the U-S link, respectively. Moreover, to characterize the QoS requirement, we set $\gamma_{min} = 6\ dBw$. Recall that the channel state is produced by divided the channel gain. We set the number of channel state to be 5 for each wireless link.

We also set up a CNN used for estimating the Q-function. There are three convolutional layers, two pooling layer, and two fully connected layers. The values of the initial weight vector are set as random values. The number of previous states is set as $N = 50$. We have $\epsilon = 0.2$. The learning rate used for the Q-learning is $\beta = 0.2$. The discount is $\alpha = 0.7$. The number of time slots to be traversed is set to be $T = 40$. The memory size is $K = 5000$.

B. The Convergence Performance

To evaluate the convergence performance of the proposed scheme, we offer Fig. 2 to show this. We have the following observations from this figure. At the beginning of the learning process, the sum-rate achieved by all UEs is very low. As the time elapses, the sum-rate is increasing but fluctuant. This is

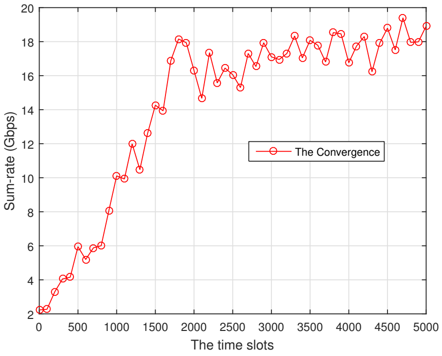

Fig. 2. The convergence of the proposed scheme.

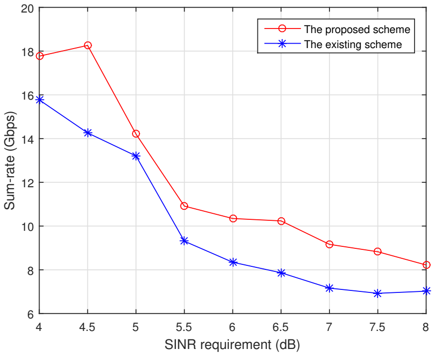

Fig. 3. The comparison of the sum-rate achieved by all UEs.

because the agent is still learning the parameters of the system and does not find the best parameters. After 2200 around time slots, the fluctuation of the sum-rate is very small, which implies the sum-rate converges. Recall that the estimation of the Q-function is conducted offline by using a CNN. Thus, the convergence time has a little influence on the performance of the proposed scheme.

C. The Improvement of the Sum-Rate

To explore the performance of the proposed scheme, we measure the sum-rate achieved by all UEs and compare that achieved by the proposed scheme to the existing scheme. The measured sum-rate is shown in Fig. 3. Specifically, we show the sum-rate as the minimum SINR requirement increases. We can find that the sum-rate is improved a lot by the proposed scheme. This gives us an insight that we need to maximize the sum-rate for 5G wireless communications. Moreover, the sum-rate is decreasing with the increase in the minimum SINR requirement. Since the higher minimum SINR is required, some links need to be allocated higher transmission power and some other links may have to be allocated lower transmission power. As a result, the sum-rate is decreasing.

VI. Conclusions

In this paper, we have studied the problem of enhancing NLOS transmission performance for 5G wireless communications. To improve NLOS transmission performance, we propose to dynamically control the transmission power in 5G wireless communication systems. In particular, we explore the control of UE association with MBSs/SBSs and power allocation and formulate the maximization problem of UEs' sum-rate under the constraints of transmission power and UEs' QoS requirements. To solve the formulated problem, we first apply a CNN to perform the Q-function estimation offline, and based on the estimation results, then conduct the deep Q-learning for online control of UE association with MBSs/SBSs and power allocation. The simulation results show the significant performance improvement achieved by our proposed scheme.

References

[1] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Comm. Survey & Tutorials*, vol. 18, 3rd Quater 2016.

[2] C. Luo, G. Min, F. R. Yu, M. Chen, L. T. Yang, and V. C. M. Leung, "Energy-efficient distributed relay and power control in cognitive radio cooperative communications," *IEEE J. Select. Areas Commun.*, vol. 31, pp. 2442–2452, Nov. 2013.

[3] W. Liao, M. Li, S. Salinas, P. Li, and M. Pan, "Energy-source-aware cost optimization for green cellular networks with strong stability," *IEEE Transactions on Emerging Topics in Computing*, vol. 4, no. 4, pp. 541–555, 2016.

[4] W. Liao, M. Li, S. Salinas, P. Li, and M. Pan, "Optimal energy cost for strongly stable multi-hop green cellular networks," in *Distributed Computing Systems (ICDCS), 2014 IEEE 34th International Conference on*, pp. 62–72, IEEE, 2014.

[5] CISCO, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast (2015 - 2020)," *http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html.*

[6] S. Lin and I. F. Akyildiz, "Dynamic base station formulation for solving NLOS problem in 5g millimeter-wave communication," in *Proc. IEEE INFOCOM'17*, (Atlanta, GA, USA), pp. 2556–2564, 1 - 4 May 2017.

[7] M. D. Renzo and W. Lu, "System-level analysis and optimization of cellular networks with simultaneous wireless information and power transfer: Stochastic geometry modeling," *IEEE Trans. Veh. Technol.*, vol. 66, pp. 2251–2275, Mar. 2017.

[8] E. Turgut and M. C. Gursoy, "Coverage in heterogeneous downlink millimeter wave cellular networks," vol. PP, pp. 1–1, May 2017.

[9] O. Tervo, L. N. Tran, and M. Juntti, "Decentralized coordinated beamforming for weighted sum energy efficiency maximization in multi-cell miso downlink," in *Proc. IEEE GlobalSIP'15*, (Orlando, FL, USA), pp. 1387–1391, 1 - 4 May 2015.

[10] C. Luo, S. Guo, S. Guo, L. T. Yang, and G. Min, "Green communication in energy renewable wireless mesh networks: Routing, rate control, and power allocation," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, pp. 3211–3220, Dec. 2014.

[11] J. Jia, Y. Deng, J. Chen, A. Aghvami, and A. Nallanathan, "Availability analysis and optimization in CoMP and CA-enabled HetNets," *IEEE Trans. Comm.*, vol. 65, pp. 2438–2450, Jun. 2017.

[12] Q. Zhang and S. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, pp. 1688–1692, Nov. 1999.

[13] A. H. Muqaibel and A. N. Jadallah, "SINR evaluation for improved practical coordinated multi-point clustering," *Wireless Pers. Commun., Springer*, vol. 83, pp. 3091–3102, Aug. 2015.

[14] C. Luo, F. R. Yu, H. Ji, and V. C. M. Leung, "Cross-layer design for TCP performance improvement in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 59, pp. 2485–2495, Jun. 2010.

[15] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning, Springer*, vol. 8, pp. 229–256, May 1992.

[16] Y. Lu, H. Lin, Y. Gu, and A. Helmy, "Towards mobility-rich analysis in Ad Hoc networks: Using contraction, expansion and hybrid models," in *Proc. IEEE ICC'04*, (Paris, France), pp. 4346–4351, 20 - 24 Jul. 2004.