2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) | 978-1-6654-6483-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/PIMRC56721.2023.10293819

Multi-Connectivity Mobility Management in Downlink FD-RAN: A Learning Based Approach

Tianqi Zhang*, Jianzhe Xue*, Jiwei Zhao*, Jiacheng Chen†, Haibo Zhou* and Xuemin (Sherman) Shen‡
*School of Electronic Science and Engineering, Nanjing University, Nanjing, China, 210023
†Department of Mathematics and Theories, Peng Cheng Laboratory, Shenzhen, China, 518000
‡Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada, N2L 3G1
Email: tianqizhang@smail.nju.edu.cn, jianzhexue@smail.nju.edu.cn, jw_zhao@smail.nju.edu.cn,
chenjch02@pcl.ac.cn, haibozhou@nju.edu.cn, sshen@uwaterloo.ca

***Abstract*—We consider a fully-decoupled radio access network (FD-RAN), where base stations (BSs) are physically decoupled into control BSs, uplink BSs and downlink BSs, and multi-connectivity becomes the default user equipment (UE) association mode. Specifically, we study the inter-frequency multi-connectivity in downlink of FD-RAN and present a deep reinforcement learning based online multi-connectivity mobility management scheme. We formulate a UE dynamic multiple access problem and transform it into a handover decision problem, then apply the double deep Q-network (DDQN) algorithm to make real time mobility management decisions. Simulation results show that the proposed scheme outperforms benchmarks in terms of handover frequency and quality of service, while ensuring real-time performance.**

***Index Terms*—FD-RAN, multi-connectivity, mobility management, handover decision, deep reinforcement learning**

I. Introduction

The fifth-generation mobile communication networks (5G) have brought unprecedented opportunities and challenges for the radio access network (RAN) design [1]. 5G aims to achieve faster transmission rate and lower air interface delay. However, due to higher frequency bands, smaller base station (BS) coverage and denser BS deployment, 5G mobile user equipments (UEs) suffer from more frequent service interruption brought by handovers [2]. In addition, the inflexible BS association mode and fixed resource scheduling method of 5G networks make it difficult to meet diverse and dynamic quality of service (QoS) requirements of UEs. Hence, in the next generation mobile communication networks (6G), how to provide continuous and seamless on-demand services for mobile UEs remains a significant challenge.

Recently, an original RAN architecture, namely fully-decoupled RAN (FD-RAN) [3], has been proposed for 6G. FD-RAN decouples the traditional communication BS into control BS and data BS, and further decouples the data BS as into uplink BS and downlink BS (DBS). The complete decoupling of BSs facilitates elastic UE-BS association and resource cooperation, which can achieve more flexible networking and higher spectral efficiency [4, 5]. Moreover, multi-connectivity becomes the default association mode in FD-RAN where multiple BSs collaborate to serve a UE. However, FD-RAN brings new challenges for mobility management especially for downlink. Since each UE have multiple serving DBSs, it is necessary to coordinate the handover decisions among different DBSs to avoid unnecessary handovers and adapt to UE's dynamic QoS requirements. Thus, typical event-based handover policy in 3GPP 5G protocols [6] is no longer applicable, so it is of great significance to develop a novel online multi-connectivity mobility management mechanism.

To address challenges of multi-connectivity access and online mobility management, many studies have been carried out [5, 7–9]. For multi-connectivity access problem, matching theory-based algorithms were proposed to obtain the optimal multi-BS association in [5, 7]. However, both papers considered static scenarios and their real-time performance were not satisfactory enough. For online mobility management problem, deep reinforcement learning (DRL) based methods were applied to dynamically select optimal serving BS in each step, where different BS association of adjacent steps was treated as a handover [8, 9]. Both methods allowed for real-time requirements, but they were only suitable for single-connectivity instead of multi-connectivity scenario.

In this paper, inspired by previous works and considering potential challenges, we propose a DRL based approach for online multi-connectivity mobility management in downlink FD-RAN to seamlessly satisfy UEs' QoS requirements while decreasing handover numbers. Note that we consider the inter-frequency multi-connectivity form where multiple DBSs send signals to a UE over distinct frequency bands. In that way, data are split on the edge cloud (EC), then multiple data flows are delivered to DBSs and sent to UE independently and simultaneously. The main contributions are three-fold:

- We develop an inter-frequency multi-connectivity system model of downlink FD-RAN and formulate a long-term mathematical problem about mobile UEs' QoS satisfying.
- We transform the original UE multiple access problem into a difference form handover decision problem, where handover refers to the updating of UE's serving BS set, and apply DDQN algorithm to obtain real-time decisions.
- We carry out extensive simulations, which indicate that our scheme can provide seamless on-demand services with low handover numbers and real-time performance.

The rest of this paper is organized as follows. We first present system model and problem formulation in Section II. To solve it, we develop a DRL framework and resort to

978-1-6654-6483-3/23/$31.00 ©2023 IEEE

DDQN algorithm in Section III. Simulation results are shown in Section IV. At last, conclusions are drawn in Section V.

II. System Model and Problem Formulation

Notations: We use letters to denote scalars (e.g., k and K), while bold letters are used to denote vectors and matrices, (e.g., $\mathbf{h}$ and $\mathbf{H}$), respectively. $(\bullet)^{\dagger}$ stands for the conjugate transpose operation. The notation $||\bullet||_n$ is the n-norm operator.

A. Network System Model

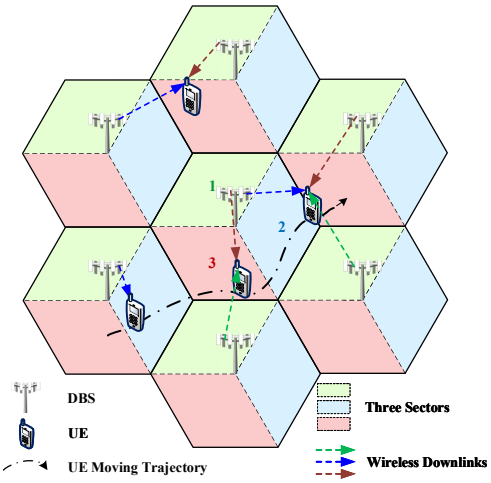

Fig. 1. Network system model in downlink FD-RAN

As shown in Fig. 1, we consider a single-tier cellular network consisting of D hexagonally deployed DBSs denoted as $\mathcal{D} = \{1, 2, ..., D\}$. All the DBSs share a common resource pool, and each DBS consists of three sectors using different resource trisected from the pool. Specifically, sectors with the same color in Fig. 1 reuse the same resources. We assume that each sector has 120-degree coverage and is equipped N_T directional antennas. We call that each DBS has three cells, and the cell set is denoted as $\mathcal{C} = \{1, 2, ..., C\}$, where $C = 3 \times D$. We define each DBS has N (can be divided by 3) sub-channels (SCs) in total denoted as $\mathcal{N} = \{1, 2, ..., N\}$, so each cell has $\frac{N}{3}$ SCs. UE set is denoted as $\mathcal{K} = \{1, 2, ..., K\}$, each is equipped with one omnidirectional antenna.

We define that the system runs in a time-slot mode with the slot index $t \in \{1, 2, ..., T\}$, and the length of the time slot is t_T. At time slot t, The Euclidean distance (in meters) between UE k and cell c is denoted by $d^t_{k,c}$, and the azimuth angle (in radians) of the UE k relative to the sector centreline of cell c is denoted by $\theta^t_{k,c}$. Each UE has its own QoS requirements, the transmission rate q^t_k, specifically. The QoS requirements level is discretized into three levels: $q^t_k \in \{q_a, q_b, q_c\}$.

B. Channel Model

We consider an inter-frequency multi-connectivity system leveraging OFDMA technique, where each UE can get multiple data flows from multiple cells using different SCs . We define the maximum number of serving cell is 3, and each UE is allocated with 1 SC from each serving cell. Especially, if a UE connects to multiple cells, each link has its own signal-to-noise-plus-interference-ratio (SINR) and achievable rate.

For the three-sector DBS, the sector antenna gain is maximum at the sector centreline direction and decays to both sides. The horizontal sector antenna gain pattern (in decibels) is specified as 3GPP TR 38.901 [10]:

$$A(\theta)_{\mathrm{dB}} = A_g - \min\left\{12\left(\frac{\theta}{\theta_{3\mathrm{dB}}}\right)^2, A_m\right\}, \ \theta \in [-\pi, \pi], \tag{1}$$

where $\theta_{3\mathrm{dB}}$ is 3dB beam width, A_g is the maximum antenna gain and A_m is the maximum directional attenuation.

We model each link as a multi-input single-output (MISO) system. The channel response $\mathbf{h}^t_{k,c,n} \in \mathbb{C}^{N_T \times 1}$ is

$$\mathbf{h}^t_{k,c,n} = \boldsymbol{\alpha}^t_{k,c,n}\sqrt{A(\theta^t_{k,c}) l^t_{k,c} \beta^t_{k,c}}, \tag{2}$$

where the fast Rayleigh fading $\boldsymbol{\alpha}^t_{k,c,n} \in \mathbb{C}^{N_T \times 1}$ is an N_T-dimensional vector whose entries are i.d.d random variables following complex Gaussian distribution with zero-mean and one-variance. The fractional antenna gain $A(\theta^t_{k,c})$ is converted from $A(\theta^t_{k,c})_{\mathrm{dB}}$. The pass loss $l^t_{k,c}$ is a function of $d^t_{k,c}$. $\beta^t_{k,c}$ is the log-normal shadow fading. Considering the autocorrelation of adjacent shadow fading [10, 11], the shadow fading in decibel form is modeled as a first-order autoregressive process:

$$\beta^t_{\mathrm{dB}} = \beta^{t-1}_{\mathrm{dB}} e^{-\Delta x / d_{cor}} + \chi(1 - e^{-\Delta x / d_{cor}}), \tag{3}$$

where Δx is the displacement distance of UE between adjacent time slots, d_{cor} is correlation length and χ follows Gaussian distribution with mean 0 and standard deviation σ_{SF}.

The useful received signal at UE k from cell c on SC n is:

$$\delta^t_{k,c,n} = {\mathbf{h}^t_{k,c,n}}^{\dagger} \mathbf{w}^t_{k,c,n} b^t_{k,c,n} f^t_{k,c,n}, \tag{4}$$

where $b^t_{k,c,n}$ is the transmitted symbol. $f^t_{k,c,n}$ is the binary UE-cell-SC allocation indicator with $f^t_{k,c,n} = 1$ implying the UE k is served by cell c on SC n at time slot t and $f^t_{k,c,n} = 0$ otherwise. Note that if cell c does not have the SC n in its resource pool, then for any k and t, $f^t_{k,c,n} \triangleq 0$. Similarly, we use $g^t_{k,c} = \sum_{n \in \mathcal{N}} f^t_{k,c,n}$ to denote the binary UE-cell association indicator. $\mathbf{w}^t_{k,c,n} \in \mathbb{C}^{N_T \times 1}$ is the beamforming vector leveraging maximum ratio transmission (MRT) technique:

$$\mathbf{w}^t_{k,c,n} = \sqrt{P}\frac{\mathbf{h}^t_{k,c,n}}{||\mathbf{h}^t_{k,c,n}||_2}, \tag{5}$$

where P is the constant transmitting power. The SINR at UE k from cell c on SC n at time slot t can be given by:

$$\gamma^t_{k,c,n} = \frac{|{\mathbf{h}^t_{k,c,n}}^{\dagger} \mathbf{w}^t_{k,c,n} f^t_{k,c,n}|^2}{\sum\limits_{k' \in \mathcal{K} \setminus k} \sum\limits_{c' \in \mathcal{C} \setminus c} |{\mathbf{h}^t_{k,c',n}}^{\dagger} \mathbf{w}^t_{k',c',n} f^t_{k',c',n}|^2 + \sigma^2}, \tag{6}$$

where σ^2 denotes the noise power. Note that for each UE, only SINRs on the serving cell and serving SC have valid non-zero values. UEs' SINRs equal to zero on the unused SCs.

We adopt Shannon theorem to indicate the achievable rate on each SC, where B denote the bandwidth of SC:

$$\varphi^t_{k,c} = \sum_{n \in \mathcal{N}} B \log_2(1 + \gamma^t_{k,c,n}). \tag{7}$$

In our inter-frequency multi-connectivity system, UE can be served by multiple links from multiple cells. The total rate for UE k at time slot t is the sum achievable rate from all cells:

$$r_k^t = \sum_{c \in \mathcal{C}} g_{k,c}^t \varphi_{k,c}^t. \tag{8}$$

C. UE Measurement and Reporting Model

Let cells periodically broadcast the channel state information reference signal (CSI-RS) at constant power. UEs are configured to periodically measure CSI-RS of all the cells and get the CSI reference signal received power (CSI-RSRP).

We define that CSI-RS transmitting and measuring are done synchronously at each time slot, and measurement reporting is done every λ time slots. $j \in \{1, 2, ..., J\}$ is used to denote the reporting slot. So the reporting slot length $j_J = \lambda t_T$.

Let $\mathrm{RSRP_{k,c}}[t]$ (in Watts) be the measured CSI-RSRP for UE k from cell c at time slot t. To reduce the channel measuring randomness brought by fast fading, we average the measured CSI-RSRPs of the past λ time slots and use $\overline{M}_{k,c}^j$ to denote the averaged CSI-RSRP at reporting slot j:

$$\overline{M}_{k,c}^j = \frac{1}{\lambda} \sum_{i=0}^{\lambda-1} \mathrm{RSRP_{k,c}}[j\lambda - i]. \tag{9}$$

During a reporting slot, the UE-cell connections remain the same, so for UE k at reporting slot j, we use $G_{k,c}^j = g_{k,c}^{j\lambda}$ to denote its cell access indicator, use $\mathrm{Q}_k^j = q_k^{j\lambda}$ to denote its QoS requirement, and use $\overline{r}_k^j$ to denote its averaged sum rate:

$$\overline{r}_k^j = \frac{1}{\lambda} \sum_{i=0}^{\lambda-1} r_k^{j\lambda - i}. \tag{10}$$

D. Problem Formulation and Transformation

The objective of this work is to provide continuous and seamless on-demand services for UEs. We denote $\boldsymbol{F} = [f_{k,c,n}^t, \forall k, c, n, t]$ as all UE-cell-SC allocation configurations. The problem is formulated as:

$$\min_{\boldsymbol{F}} \quad \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{k \in \mathcal{K}} |r_k^t - q_k^t| \tag{11a}$$

$$\text{s.t.} \quad q_k^t \in \{q_a, q_b, q_c\}, \quad \forall k \in \mathcal{K} \tag{11b}$$

$$f_{k,c,n}^t \in \{0, 1\}, \quad \forall (k, c, n) \in \mathcal{K} \times \mathcal{C} \times \mathcal{N} \tag{11c}$$

$$\sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} f_{k,c,n}^t \leq \tfrac{N}{3}, \quad \forall c \in \mathcal{C} \tag{11d}$$

$$\sum_{n \in \mathcal{N}} f_{k,c,n}^t \leq 1, \quad \forall (k, c) \in \mathcal{K} \times \mathcal{C} \tag{11e}$$

$$1 \leq \sum_{c \in \mathcal{C}} g_{k,c}^t \leq 3, \quad \forall k \in \mathcal{K} \tag{11f}$$

$$f_{k,c,n}^t = 0, \quad \forall (k, c) \in \mathcal{K} \times \mathcal{C}$$
$$\forall n \in \mathcal{N} \backslash \{\tfrac{N}{3}(c \bmod 3 - 1) + 1, ..., \tfrac{N}{3}(c \bmod 3)\}, \tag{11g}$$

where the objective (11a) is to minimize the difference between UEs' received rate and QoS requirements, thus provide on-demand services. Constraint (11b) represents the optional QoS levels. Constraint (11d) is the SC number constraint for each cell. Constraint (11e) means that each UE is permitted to access at most 1 SC from each cell. Constraint (11f) means that each UE can connect 1 to 3 cells. Equation (11g) ensures that UE cannot use SCs that are not owned by its serving cell.

Optimization problem (11) is considered intractable due to its integer programming nature and cannot be effectively solved using convex optimization. Moreover, in highly dynamic mobile scenario, mobility management namely dynamic UE access management, needs to be done in real time. Hence, we resort to DRL to obtain online mobility management policy.

In single-connectivity scenarios, DRL is widely used for access and handover control. However, in our multi-connectivity system, UE can connect to 1, 2 or 3 cells, so there are $\mathbf{C}_C^1 + \mathbf{C}_C^2 + \mathbf{C}_C^3$ UE-cell access cases in total, resulting in too large DRL action space and difficult to converge.

Therefore, we transform the original UE access problem into a difference form. That is, we use DRL to output handover decision for each UE, instead of explicit UE access configurations. The details of our DRL framework are described below.

III. Deep Reinforcement Learning Framework

The framework of DRL is composed of agents and environment. In our scenario, agents representing each UE are created at the EC, and the environment simulates a FD-RAN downlink communication system. We define that DRL agents and environment interact in a discrete step manner, and the step j is equivalent to the reporting slot j. The UEs' QoS requirements are slightly dynamic, that is, a certain proportion of UEs' QoS level will be regenerated every λ_Q steps.

For every step j, each agent receives the report from its corresponding UE and obtain the state $\boldsymbol{s}^j$. Based on the state and DRL algorithm, the agent takes an action a^j. After interacting with the environment using a^j, agent receives the reward R^j from environment, and gets the next state $\boldsymbol{s}^{j+1}$. Here are some important elements of our DRL framework.

A. Initialization

At first, each UE are randomly distributed in circular area and randomly generate QoS level from $\{q_a, q_b, q_c\}$. Let UEs with low and medium QoS levels access the nearest cell, and let UEs with high QoS level access the two nearest cells.

B. State

We define that the state $\boldsymbol{s}_k^j$ of UE k at step j is composed of three components. The first component is a vector of cell access 0-1 indicators, $\boldsymbol{G}_k^j = [G_{k,1}^j, G_{k,2}^j, ..., G_{k,C}^j]$. The second component is a vector of averaged measured CSI-RSRPs, $\boldsymbol{M}_k^j = [\overline{M}_{k,1}^j, \overline{M}_{k,2}^j, ..., \overline{M}_{k,C}^j]$. The third component is the rate achieving radio, $Z_k^j = \bar{r}_k^j / \mathrm{Q}_k^j$. So $\boldsymbol{s}_k^j$ is given by

$$\boldsymbol{s}_k^j \triangleq \left[\boldsymbol{G}_k^j, \boldsymbol{M}_k^j, Z_k^j\right]. \tag{12}$$

C. Action

1) Action Space: The discrete action space $\mathcal{A}$ consists of four kinds of handover decisions, and can be expressed as:

$$\mathcal{A} = \{Remain, Add, Remove, Change\}. \tag{13}$$

2) Invalid Action Masking [12]: Note that in some cases, not all actions are valid due to the serving cells number constraint. For UE k at step j, let $\kappa_k^j = ||\boldsymbol{G}_k^j||_1$ denote its serving cell number, then its invalid action $\hat{a}_k^j$ can be obtained:

$$\hat{a}_k^j = \begin{cases} Remove, & \text{if } \kappa_k^j = 1 \\ Add \quad , & \text{if } \kappa_k^j = 3 \\ \emptyset \quad\ , & \text{if } \kappa_k^j = 2 \end{cases}. \tag{14}$$

So the valid action space is $\mathcal{A}\backslash\hat{a}_k^j$. In the action selection stage of the DRL, UE k can only select action from $\mathcal{A}\backslash\hat{a}_k^j$.

3) Handover Performing: The action $a_k^j \in \mathcal{A}\backslash\hat{a}_k^j$ is the handover decision of UE k at step j. For the UE k, after an action a_k^j is chosen, the corresponding handover decision is performed and its serving cell set is updated. The specific four kinds of handover performing and serving cell set updating processes are shown as follows:

- *Remain*: UE k keeps its cell association unchanged. So its serving cell set remains the same: $\boldsymbol{G}_k^{j+1} \leftarrow \boldsymbol{G}_k^j$.
- *Add*: UE k adds the cell c_{add} to its serving cell set: $\boldsymbol{G}_k^{j+1} \leftarrow \boldsymbol{G}_k^j,\ with\ \boldsymbol{G}_k^j[c_{add}] = 1$.
- *Remove*: UE k deletes cell c_{sub} from its serving cell set: $\boldsymbol{G}_k^{j+1} \leftarrow \boldsymbol{G}_k^j,\ with\ \boldsymbol{G}_k^j[c_{sub}] = 0$.
- *Change*: UE k replaces the cell c_{sub} with the cell c_{add}: $\boldsymbol{G}_k^{j+1} \leftarrow \boldsymbol{G}_k^j,\ with\ \boldsymbol{G}_k^j[c_{add}] = 1\ and\ \boldsymbol{G}_k^j[c_{sub}] = 0$.

Here c_{add} is the index of cell that has the largest CSI-RSRP in the unconnected cell set, and c_{sub} is the index of cell that offers the lowest rate φ in the serving cell set.

4) SC Allocation: We apply a simple SC allocation scheme. When a UE is newly connected to a cell, it will be randomly allocated an idle SC. When a UE is disconnected from a cell, the allocated SC will be released and becomes idle. Assume that UEs will not cluster in large numbers in a single cell, so cell overloading is not considered herein.

D. Reward

In our multi-connectivity mobility management framework, the objectives are threefold: satisfying UEs' QoS requirements, avoiding frequent handovers, and saving network resources. Hence, we define that the reward function $R_k^j(\boldsymbol{s}_k^j, a_k^j)$ of UE k after taking action a_k^j at state $\boldsymbol{s}_k^j$ consists of three components: rate bonus, handover cost and multi-connectivity penalty.

The rate bonus RR_k^j is designed for satisfying UE's QoS without wasting network resources:

$$RR_k^j = \begin{cases} 1 - 2.0|Z_k^{j+1} - 0.95|, \text{if } Z_k^{j+1} < 0.95 \\ 1 \qquad\qquad\quad , \text{if } 0.95 \le Z_k^{j+1} < 1.25 \\ 1 - 0.4|Z_k^{j+1} - 1.25|, \text{if } Z_k^{j+1} \ge 1.25 \end{cases} \tag{15}$$

The handover cost RH_k^j is designed for avoiding frequent handovers. The values of RH_k^j is selected as:

$$RH_k^j = \begin{cases} 0, & \text{if } a_k^j = Remain \\ -0.8, & \text{if } a_k^j = Add \\ -0.6, & \text{if } a_k^j = Remove \\ -1.0, & \text{if } a_k^j = Change \end{cases}. \tag{16}$$

The multi-connectivity penalty RM_k^j is designed for encouraging UE to access fewer cells in order to save network resources:

$$RM_k^j = -0.5(\kappa_k^{j+1} - 1). \tag{17}$$

The total reward function is the weighted sum of the above three components, where τ_R, τ_H and τ_M are the weights:

$$R_k^j(\boldsymbol{s}_k^j, a_k^j) = \tau_R RR_k^j + \tau_H RH_k^j + \tau_M RM_k^j. \tag{18}$$

E. Double Deep Q-Network (DDQN)

In this paper, we apply DDQN algorithm which combines Q-learning and deep learning. Q-learning is a reinforcement learning algorithm that learns how to choose the best action in each state based on the expected future reward. Q-learning uses action-value Q-function to represent the Q-value of taking an action in a state, uses Q-table to store Q-values for all possible state-action pairs. However, Q-learning cannot handle large state spaces because Q-table cannot be impractically large. Therefore, deep learning is used to deal with the drawbacks of Q-learning, then we get DDQN algorithm, which uses deep neural network to approximate the Q-function. In DDQN algorithm, two identical neural networks, training network $Q(\boldsymbol{\theta})$ and target network $Q(\boldsymbol{\theta'})$, are created initially, where $\boldsymbol{\theta}$ and $\boldsymbol{\theta'}$ are the parameters of two networks, respectively.

At each step, for each agent, the training network takes the state $\boldsymbol{s}$ as input and outputs the Q-values $Q(\boldsymbol{s}, a; \boldsymbol{\theta})$ for all possible actions $a \in \mathcal{A}$. Then, we use ε-greedy policy for action selection, which is a simple and common way to balance exploration and exploitation in DDQN:

$$a = \begin{cases} \arg\max\limits_{a \in \mathcal{A}\backslash\hat{a}} Q(\boldsymbol{s}, a; \boldsymbol{\theta}) & ,\ with\ prob.\ 1 - \varepsilon \\ random\ \ a \in \mathcal{A}\backslash\hat{a} & ,\ with\ prob.\ \varepsilon \end{cases}, \tag{19}$$

where ε decays exponentially over step from ε_{init} to ε_{final} and fixes at ε_{final} thereafter. After interacting with environment using selected action a, the agent receives reward $R(\boldsymbol{s}, a)$ and state of next step $\hat{\boldsymbol{s}}$, and this experience $\{\boldsymbol{s}, a, R, \hat{\boldsymbol{s}}\}$ is stored in experience replay memory $\mathcal{D}$. Note that all the agents interact with the environment but share the same neural network and the same replay memory.

Subsequently, DDQN algorithm randomly samples a mini-batch of experience $\mathcal{D}_m$ from $\mathcal{D}$ to train the training network. Training network $Q(\boldsymbol{\theta})$ is trained using gradient descent (GD) algorithm by minimizing a loss function, which is the mean square error (MSE) between the estimated Q-values and the estimation target values over mini-batch $\mathcal{D}_m$:

$$L(\boldsymbol{\theta}) = \mathop{\mathbb{E}}_{\{\boldsymbol{s}, a, R, \hat{\boldsymbol{s}}\} \in \mathcal{D}_m} \left[\left(y - Q(\boldsymbol{s}, a; \boldsymbol{\theta}) \right)^2 \right], \tag{20}$$

where the estimation target value y is denoted as

$$y = R + \gamma Q(\hat{\boldsymbol{s}}, \tilde{a}; \boldsymbol{\theta'}), \tag{21}$$

where $\gamma \in (0, 1]$ is the discounting factor and $\tilde{a}$ is given by

$$\tilde{a} = \arg\max_{a \in \mathcal{A}\backslash\hat{a}} Q(\hat{\boldsymbol{s}}, a; \boldsymbol{\theta}). \tag{22}$$

Note that in (21), DDQN uses training network $Q(\boldsymbol{\theta})$ to select action in (22) and uses target network $Q(\boldsymbol{\theta'})$ to calculate

target Q-value in (21). Thus, compared to the traditional DQN that uses the same target network to select action and calculate target value, Q-value overestimation errors can be reduced and network can be trained more effectively.

In our DDQN algorithm, the target network $Q(\boldsymbol{\theta'})$ is only used to obtain the target of estimation in (21), and does not carry out gradient descent calculation. However, every λ_U steps, the target network $Q(\boldsymbol{\theta'})$ is updated with its parameters $\boldsymbol{\theta'}$ copied from training network parameters $\boldsymbol{\theta}$.

IV. Simulation Results and Analysis

A. Simulation Settings

We consider a cellular network consisting of $D = 7$ hexagonally deployed DBSs (that is, $C = 21$ cells, as shown in Fig. 1) and default $K = 30$ randomly distributed UEs in a 600-meter-radius circular area. All the DBSs reuse a common resource pool of $N = 30$ SCs and each cell has $\frac{N}{3} = 10$ SCs. The number of antennas in each cell is $N_T = 8$. UEs are evenly divided into two groups, one half keeps stationary, the other half keeps moving following Gauss-Markov mobility model [13]. The channels between UEs and cells are generated based on 3GPP TR 38.901 UMa scenario [10]. Our DDQN network is a four-layer fully connected neural network with two hidden layers and ReLU activation function. The numbers of neurons in the two hidden layers are 128 and 64, respectively. The simulation is implemented in Python 3.9 and PyTorch 1.12.0, with other main parameters shown in Table I.

TABLE I
Simulation Parameters

Parameters	Value
Time length of one time slot t_T	10 ms
QoS levels $\{q_a, q_b, q_c\}$	$\{15, 30, 45\}$ Mbps
Antenna gain parameter $\theta_{3\text{dB}}$, A_g, A_m	0.361π, 8 dBi, 30 dB
Shadow fading parameter d_{cor}, σ_{SF}	50 m, 6 dB
Path loss model (in decibels)	$13.54 + 39.08\log_{10}(d)$
SC bandwidth B	2 MHz
Transmitting power P	30 dBm
Noise power density σ^2/B	-174 dBm/Hz
UE Reporting period (slots) λ	10
QoS updating period (steps) λ_Q	500
Mean speed of moving UEs	Default 10 m/s
Time length of one step j_J	100 ms
Gradient descent Optimizer	Adam
Exploration rate ε_{init}, ε_{final}	0.5, 0.05
Replay memory size $\lvert\mathcal{D}\rvert$	2^{17}
Mini-batch size $\lvert\mathcal{D}_m\rvert$	128
Discounting factor γ	0.94
Target network updating period (steps) λ_U	200
Training network learning rate	0.0005

B. Results and Analysis

The DRL training process consists of 6 episodes, where each episode has 5000 steps. We use DQN to compare with the applied DDQN, and Fig. 2 shows that as the training process goes on, the agents average reward of DDQN can eventually converge around 0.87, bringing a little advantage over DQN.

After the training process, the training network is fixed, which will be used by each agent to take optimal action according to the obtained state. Before analysing the testing

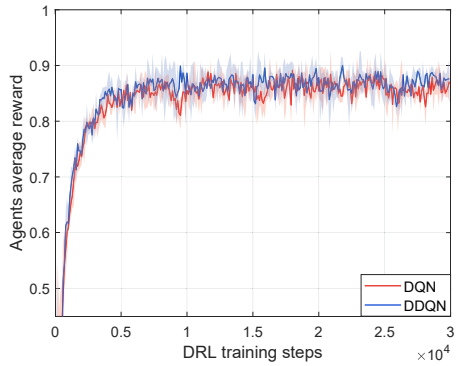

Fig. 2. Reward variation during training process.

results, we first define two performance indexes: handover number and QoS substandard time proportion. A handover means a UE takes an action other than $Remain$. QoS substandard time proportion means the time proportion when 10-step-averaged QoS achieving ratio of the UE is less than 1.

Fig. 3 illustrates the impact of UE moving speeds on communication performance using our scheme in 5000 testing steps. We consider five cases, where the case 0 represents stationary UEs and other cases represent moving UEs with different mean speeds. We can see that as UE speed grows, each UE experiences more handovers. Note that stationary UEs will still perform handover due to the dynamic QoS requirements updating. and channel condition changing. However, regardless of the UEs' speed, their QoS can be well satisfied, with only about 2% to 3% time proportion that their serving rate is substandard.

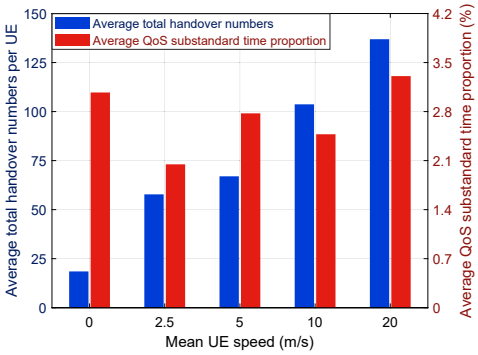

Fig. 3. Average communication performance versus UE speed.

Two greedy methods are executed as benchmarks. The first is a single-connectivity handover scheme called 1-cell benchmark, where each UE connects to the cell with the highest RSRP value in each step. The other is a triple-connectivity handover scheme called 3-cell benchmark, where each UE connects to three cells with the three highest RSRP values in each step. In these two benchmarks, a change in the UE's serving cell or serving cell set is considered as a handover. The 5000-step performance comparisons among our scheme and two benchmarks are shown in Fig. 4 and Fig. 5.

Fig. 4 shows the comparisons of average handover number per moving UE among the three methods under different total UE numbers. We find that UE number has no obvious impact on handover numbers for these methods. Moreover, compared

with the 3-cell benchmark, our proposed scheme can reduce UE averaged handover numbers by 68.8%, 54.6%, 60.1% and 53.2%, under the UE number of 10, 20, 30 and 40, respectively. As for the 1-cell benchmark, our proposed scheme has similar performance in terms of handover numbers.

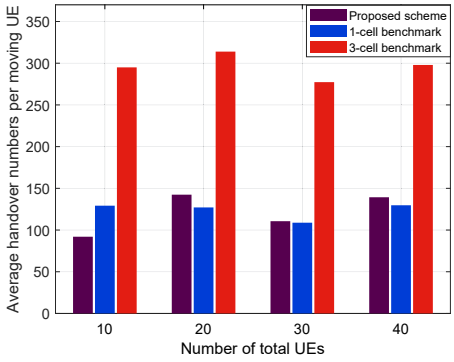

Fig. 4. Moving UE average handover numbers of different UE numbers.

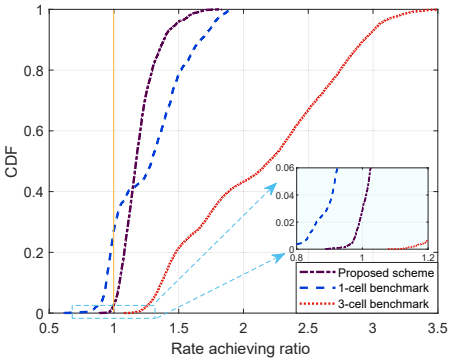

Fig. 5. Rate achieving ratio CDF of proposed scheme and two benchmarks.

Fig. 5 shows the cumulative distribution function (CDF) of the rate achieving ratio of the three methods under the 30 UEs case. Note that in our scenario, we hope the UE's rate achieving ratio is not less than 1, but not too large. That is because that an excessively high rate achieving ratio usually means the UE has too many serving cells, which wastes the network resources. The figure shows that when rate achieving ratio equals to 1, the CDF of the proposed scheme is much lower than 1-cell benchmark, meaning that proposed scheme is better than 1-cell benchmark in guaranteeing UEs' QoS. In the part of QoS achieving ratio greater than 1, the CDF of 3-cell benchmark is much lower than proposed scheme, meaning that 3-cell benchmark occupies unnecessary network resources to provide too much unnecessary services for UEs.

In combination with Fig. 4 and Fig. 5, we can find that compared with the 1-cell benchmark, our proposed scheme significantly improves UEs' QoS without increasing handover numbers. Compared with the 3-cell benchmark, our scheme largely reduces UEs' handover numbers and saves network resources while meeting the rate demands almost as much. Furthermore, under the 30 UEs case, the average total action selection time of each testing step in proposed scheme is 0.585 ms, which indicates that our learning-based mobility management scheme has good real-time performance, and can adapt to the highly dynamic mobile communication scenario.

V. Conclusions

In this paper, we have investigated inter-frequency multi-connectivity in downlink FD-RAN, and developed a DRL based mobility management scheme to provide seamless on-demand services for mobile UEs. The handover decision problem has been transformed from original UE multiple access problem, and DDQN has been leveraged to make real-time decisions. Simulation results have shown that our scheme can adapt to different UE speeds and provide more balanced and superior QoS compared to the benchmarks. In the future, we will consider an intelligent SC allocation and power control method to further enhance our mobility management scheme.

Acknowledgment

This work is supported in part by the National Natural Science Foundation Original Exploration Project of China under Grant 62250004, the National Natural Science Foundation of China under Grant 62271244, the Natural Science Fund for Distinguished Young Scholars of Jiangsu Province under Grant BK20220067.

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[2] T. Bilen, B. Canberk, and K. R. Chowdhury, "Handover management in software-defined ultra-dense 5G networks," *IEEE Network*, vol. 31, no. 4, pp. 49–55, 2017.

[3] Q. Yu, H. Zhou, J. Chen, Y. Li, J. Jing, J. J. Zhao, B. Qian, and J. Wang, "A fully-decoupled RAN architecture for 6G inspired by neurotransmission," *Journal of Communications and Information Networks*, vol. 4, no. 4, pp. 15–23, 2019.

[4] J. Zhao, Q. Yu, B. Qian, K. Yu, Y. Xu, H. Zhou, and X. Shen, "Fully-decoupled radio access networks: A resilient uplink base stations cooperative reception framework," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[5] K. Yu, Q. Yu, Z. Tang, J. Zhao, B. Qian, Y. Xu, H. Zhou, and X. Shen, "Fully-decoupled radio access networks: A flexible downlink multi-connectivity and dynamic resource cooperation framework," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2022.

[6] 3rd Generation Partnership Project, "NR; Requirements for support of radio resource management," 3GPP, TS 38.133, 2023, version 18.0.0.

[7] M. Simsek, T. Hobler, E. Jorswieck, H. Klessig, and G. Fettweis, "Multiconnectivity in multicellular, multiuser systems: A matching-based approach," *Proceedings of the IEEE*, vol. 107, no. 2, pp. 394–413, 2019.

[8] C. Lee, J. Jung, and J. M. Chung, "Intelligent dual active protocol stack handover based on double DQN deep reinforcement learning for 5g mmwave networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7572–7584, 2022.

[9] D. Guo, L. Tang, X. Zhang, and Y.-C. Liang, "Joint optimization of handover control and power allocation based on multi-agent deep reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 13 124–13 138, 2020.

[10] 3rd Generation Partnership Project, "Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP, TR 38.901, 2022, version 17.0.0.

[11] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[12] S. Huang and S. Ontan, "A closer look at invalid action masking in policy gradient algorithms," *The International FLAIRS Conference Proceedings*, vol. 35, 2022.

[13] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," *Wireless Communications and Mobile Computing*, vol. 2, no. 5, pp. 483–502, 2002.