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.2020.3005495, IEEE Journal on Selected Areas in Communications

Deep Reinforcement Learning for Dynamic Uplink/Downlink Resource Allocation in High Mobility 5G HetNet

Fengxiao Tang, *Member, IEEE,* Yibo Zhou, *Student Member, IEEE,* and Nei Kato, *Fellow, IEEE.*

***Abstract*—Recently, the 5G is widely deployed for supporting communications of high mobility nodes including train, vehicular and unmanned aerial vehicles (UAVs) largely emerged as the main components for constructing the wireless heterogeneous network (HetNet). To further improve the radio utilization, the Time Division Duplex (TDD) is considered to be the potential full-duplex communication technology in the high mobility 5G network. However, the high mobility of users leads to the high dynamic network traffic and unpredicted link state change. A new method to predict the dynamic traffic and channel condition and schedule the TDD configuration in real-time is essential for the high mobility environment. In this paper, we investigate the channel model in the high mobility and heterogeneous network and proposed a novel deep reinforcement learning based intelligent TDD configuration algorithm to dynamically allocate radio resources in an online manner. In the proposal, the deep neural network is employed to extract the features of the complex network information, and the dynamic Q-value iteration based reinforcement learning with experience replay memory mechanism is proposed to adaptively change TDD Up/Down-link ratio by evaluated rewards. The simulation results show that the proposal achieves significant network performance improvement in terms of both network throughput and packet loss rate, comparing with conventional TDD resource allocation algorithms.**

***Index Terms*—5G, High Mobility, Resource Allocation, Unmanned Aerial Vehicle (UAV), Time Division Duplex (TDD), Reinforcement Learning (RL), Q-learning, Deep Learning, Deep Belief Network, Heterogeneous network (HetNet).**

I. Introduction

The fifth-generation (5G) network is being fast deployed throughout the world, which enables ultra-reliable and low-latency communications (uRLLC) of users. The key promising technology of 5G is the usage of millimeter-wave (mmWave) with high frequency and wide bandwidth. The mmWave brings high data transmission rates but also suffers high attenuation leading to short transmission distance. Therefore, the 5G network is constructed with high density and heterogeneous structure of enormous macro and small cells [1]. Nevertheless, such a dense and heterogeneous structure brings high channel interference that requires sufficient spectral resources. As both the channel and space are limited in the high-density 5G heterogeneous network (HetNet), for better utilize the power and channel resources, the Orthogonal Frequency Division Multiple Access (OFDMA) is proposed to enable multiple users to access to the internet synchronously [2], [3]. Meanwhile, the beamforming (BF) and Massive Multi-input Multi-output (MIMO) are widely deployed in 5G secenario for improving the spectral utilization [4], [5]. On the other hand, 5G enables massive machine-type communications (mMTC) bringing unfair network traffic demand for the Up-link and Down-link. For balancing the Up/Down traffic and adaptively allocate channel resources, the time-division-duplexing (TDD) is widely researched as the promising full-duplex technology for 5G communications [6]. The new releases (e.g., 13, 14) of Long-Term Evolution (LTE) for 5G all embrace the TDD mode.

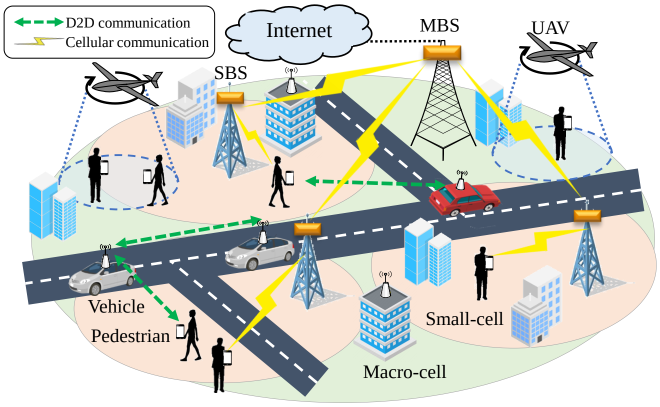


Fig. 1: Considered HetNet with high mobility users.

The TDD tailing the channel resources of frequency carriers to many sub-frames, which are dynamically allocated to Up-link and Down-link depending on different traffic demands. In the LTE-TDD mode, the TDD is considered to be tailed with 7 configurations and dynamically scheduled on a per-frame basis of 5/10 ms [7]. Many dynamic TDD resource allocation algorithms for optimizing both power efficiency and transmission performance are proposed in recent years, and various TDD based network applications are investigated in different scenarios [8]–[10]. However, all of those researches are lack of considering the high mobility of 5G HetNet with high mobility nodes such as vehicles, trains, and unmanned aerial vehicle (UAV)s [11]. The high mobility nodes create and broke connections frequently which causes frequent fluctuation of links and unexpected bursty network traffic. The conventional TDD algorithms only focus on the current state of links and traffic load is not suitable for the high mobility

Fengxiao Tang, Yibo Zhou, and Nei Kato are with the Graduate School of Information Sciences (GSIS), Tohoku University, Sendai, Japan. Emails: {fengxiao.tang, yibo.zhou, kato}@it.is.tohoku.ac.jp

5G scenarios [12]. Besides, the heterogeneous structure of the 5G HetNet leads to the high complexity of the TDD resource allocation. As shown in Fig. 1, there are two types of base station (BS). The small-cell base station (SBS) collected data through Up-links from various user equipments (UEs), and relay those data to the macro-cell base station (MBS) through the second layer Up-links for accessing to the core network, and vice versa. In such a heterogeneous structure, the MBS act as both the integrator of UEs and relay node for SBSs. The TDD configuration of MBS should consider both the UE to MBS connections and the SBS to MBS connections. For the HetNet, the network traffic demands for both UP-link and Down-link are highly dynamic and hard to be estimated by conventional methods.

Recently, deep learning (DL) as a promising artificial intelligence tool, is widely discussed for network optimizations including intelligent radio access and resource allocation and traffic control these days [13]–[16]. In this paper, we exploit the usage of the DL to the adaptive TDD Up/Down-link configuration problem. The reinforcement learning (RL) is one kind of machine learning that imitates human behaviors, which trains the agents (human) to do actions based on the cumulative reward (reflection from environment) of previous actions (temptations) [16]. Combined with the deep structure, deep reinforcement learning (DRL) is recently proposed as a powerful state estimation and function approximation tool in various network applications [17], [18]. Naturally, in this research, we consider using the DRL to learn the dynamic and heterogeneous environment of high mobility 5G HetNet and propose an intelligent DRL based TDD intelligent configuration algorithm to adaptively allocate Up/Down-links resource.

In summary, our contributions are summarized as follows:

1) We construct the high mobility scenario of the heterogeneous 5G network, considering both the inner-cell UE to BS connections and inter-cell BS to BS connections.
2) We employ the DRL to extract the features of the considered high mobility HetNet, and design a DRL based intelligent TDD Up/Down-link resource allocation algorithm.
3) We compare the performance in terms of packet loss rate and network throughput between the proposal and both the conventional algorithm and shallow RL based algorithm.

The remainder of this paper is organized as follows. We firstly introduce the related work and research flow of TDD configuration in Section III. Then, Section III detailedly illustrates the network, channel, and learning model, and the problem is also formulated in this section. In Section IV, the proposed DRL based intelligent TDD Up/Down-link resource allocation algorithm is introduced. Section V simulates our proposal and evaluates the performance of the proposal compared with conventional algorithms. Finally, the conclusion is summarized in Section VI.

II. Related works

As a promising full-duplex technology, the TDD resource allocation algorithms are widely researched in many years. In [19], C. Chiang et al. firstly consider the TDD configuration can be dynamic scheduled based on the cross-layer detected traffic loads. In [9], M. Ding et al. proposed a TDD dynamic configuration by considering the channel interference between wireless links. In this paper, the authors analyze the relation between dynamic channel condition and TDD configuration, a cell clustering and interference cancellation technology is employed for TDD interference control. However, the considered homogeneous network is simple and the traffic load variation is not fully considered. Instead of the homogeneous network, the authors in [8] first explore the TDD resource allocation problem in the LTE HetNet. To improve the capacity of TDD communications in LTE system, the work in [20] derives important theoretical bounds for interference-limited networks, and proposes several novel coordinated interference optimization frameworks. Besides, T. Ding et al. analyze the performance of employing dynamic TDD for dense small cell networks towards 5G [21]. However, the strong assumption that the arrival traffic follows Poisson distribution for previous works is not reasonable in the practical high mobility 5G network. Besides, in 2016, the same authors in [9] further proposed an enhanced dynamic TDD resource allocation algorithm for the HetNet [12]. Their proposal chooses the TDD configuration depending on the incoming UL/DL traffic load from the user equipment in the last time slot, which is not suitable for the high mobility environment with high dynamic traffic.

Besides considering the TDD channel resource allocation, some researches focus on joint resource allocation such as joint power and channel allocation. For the joint resource allocation problem, some classical methods such as Monte Carlo method [22], graph coloring, and game theory are employed. T. Yang et al. considering a HetNet with both vehicle to vehicle (V2V) and vehicle to infrastructure (V2I) communications and propose a graph coloring based joint power and channel resource allocation algorithm [23]. Besides, the [24] proposes a game theory based TDD resource allocation in the high mobility environment. However, both the graph coloring and game theory based algorithms are time-consuming due to the slow convergence of the coloring and game process [25], [26].

In recent years, machine learning is widely used for solving high dynamic and complexity problems in various networks [27], [28]. Both the supervised learning and reinforcement learning are used for the intelligent TDD resource allocation. In our early research in [29], the Long Short-Term Memory (LSTM) which is one kind of supervised learning is proposed in TDD to predict the traffic and link-state and dynamically change the TDD configuration. Besides, the machine learning is widely researched for various kinds of adaptive configuration and resource allocation in physical-Layer 5G communications [30]. However, the supervised learning method highly depends on the labeled big dataset which is expensive and limited by the benchmarked algorithm. To further improve it, in our previous work [31], the RL is proposed to intelligent allocate the channel resources to the Up/Down-link based on the refined Q-matrix and online collected rewards. However, the conventional RL with shallow structure is hard to handle the complex environment of the high

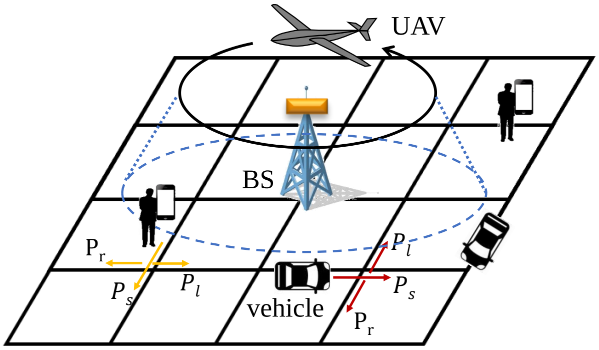


Fig. 2: Considered mobility model of various users.

mobility and heterogeneous structure of 5G HetNet.

III. System Model and Problem Statement

A. Network model

As depicted in Fig. 1, the considered HetNet is heterogeneously constructed with macro-cell and small-cells. In such network, there are N_{sbs} SBSs with corresponding small cells deployed in the network and denoted as $\mathbf{SBS} = \{sbs_1, sbs_2, \ldots, sbs_{N_{sbs}}\}$, and one MBS denoted as mbs. For each small cell, there are N^i_{user} users connect to the sbs_i for the instant. The users within the small cell directly transmit data to the SBS, and the SBS relay those data to the MBS. Besides, the users inner the macro cell but outside small cells can direct connect to the MBS, and there are D2D communications re-using the same radio frequency between users, which highly interfere with the cellular Up/Down-links. In such a scenario, two kinds of connections namely inner-cell user to BS connection and inter-cell BS to BS connection simultaneously coexist. For the simultaneous inter-cell and inner-cell connections, the OFDMA is employed for enabling multiple access. With OFDMA, the links connect to the same SBS are orthogonal with each others and no self-interferences exist. For each pair of direct connection, the TDD is used for two-way full-duplex transmission. For the dynamic TDD, we assume the TDD Up/Down-link configuration can be dynamically changed between two time slots. The TDD modulation divides each time slot t to N_{sub} sub-frames, and there are N_{tdd} types of configurations for the Up/Down-link sub-frames allocation. N_{tup}, N_{tdown} and N_{spec} denote the allocated number of Up-link, Down-link and special sub-frames respectively, $N_{tdd} = N_{tup} + N_{tdown} + N_{spec}$. In TDD configuration, the N_{spec} is always a constant value and the ratio of N_{tup}/N_{tdown} can dynamically change. The illustration of all 6 types (two configurations with the same Up/Donw link ratio are treated as the same type) of TDD configurations with different N_{tup}/N_{tdown} ratio are as shown in Fig. 3. In the considered high mobility 5G HetNet, both the links and network traffic demands are dynamically changing with users moving. As the channel state of the HetNet is complex, we model the channel with inner-cell and inter-cell respectively. For links inner each small cell, we consider in the beginning of time slot t_f, there are M_{up} UP-links denoted as $L_{up_i} (i = 1, 2, \ldots, M_{up})$, M_{down} Down-links denoted as $L_{down_j} (j = 1, 2, \ldots, M_{down})$, and totally M^i_{link} links connect to the sbs_i. Besides, considering L_{d2d_k} represents the k^{th} D2D link, and there are M_{d2d} D2D connections between the high mobility users. The corresponding transmit power and the channel gain of Up/Down-link are denoted with P_{up_i}, P_{down_j}, G_{up_i} and G_{down_j} respectively. Besides, the P_{d2d_j} and G_{d2d_j} denote the transmit power and channel gain of D2D link L_{d2d_j} respectively. The thermal noise at each UE is given by σ^2, which is assumed to be constant value. The channel capacity C_i of link L_i is calculated with Shannon theory as follow:

$$C_i = B \cdot log_2(1 + \gamma_i) \tag{1}$$

Here, the B stands for the bandwidth of link L_i, which is constant for all links in the considered TDD-OFDMA network. For each small cell, the Signal to Interference plus Noise Ratio (SINR) of $\gamma_{up_i}/\gamma_{down_j}$ Up/Down-link L_{up_i}/L_{down_j} can be given as

$$\gamma_{up_i} = \frac{P_{up_i} \cdot G_{up_i}}{\sigma^2 + \Sigma_{j=1}^{M_{d2d}} P_{d2d_j} G_{d2d_j,up_i}} \tag{2}$$

where G_{d2d_j,up_i} stands for the channel gain between the D2D link L_{d2d_j} and up-link L_{up_i} [23]. This dynamic SINR is caused by the dynamic interference between the Up-links and D2D links which allocated with the same sub-channel by the OFDMA algorithm inner the small cell. And the SINR of Down-links are the same as the Up-links and denoted as follow:

$$\gamma_{down_j} = \frac{P_{down_j} \cdot G_{down_j}}{\sigma^2 + \Sigma_{n=1}^{M_{d2d}} P_{d2d_n} G_{d2d_n,down_j}} \tag{3}$$

However, as mentioned above, the interference in the considered HetNet not only exists inner the small cell but inter cells. We use $L_{up^{ma}_i}/L_{down^{ma}_j}$ to denote the i^{th} and j^{th} inter-cell Up/Down-links between SBS to MBS, and there are totally M^{ma}_{up} and M^{ma}_{down} Up/Down-links inter cells. Thus, due to the inter-cell interference (ICI), the SINR of the inter-cell link is formulated as:

$$\begin{aligned}
\gamma_{down^{ma}_j} &= \frac{P_{down^{ma}_j} \cdot G_{down^{ma}_j}}{\sigma^2 + \Sigma_{k=1}^{N_{sbs}} \theta_k + \Sigma_{k=1}^{N_{sbs}} \nu_k + \Sigma_{k=1}^{N_{sbs}+1} \xi_k} \\
\theta_k &= \Sigma_{i=1}^{M_{up}} P_{up_i} G_{up_i,down^{ma}_j} \\
\nu_k &= \Sigma_{i=1}^{M_{down}} P_{down_i} G_{down_i,down^{ma}_j} \\
\xi_k &= \Sigma_{n=1}^{M_{d2d}} P_{d2d_n} G_{d2d_n,down^{ma}_j}
\end{aligned} \tag{3}$$

Here, the $\theta_k/\nu_k/\xi_k$ denote interference factor between inner-cell Up/Down/D2D-links in small cell k and the inter-cell Down-links respectively. And the D2D connections near the MBS should also be considered. Because the users are dynamic moving in the scenario, the channel gain G should be dynamically calculated based on the instantaneous distance as follow:

$$G = \sqrt{\frac{P_r}{P_t \cdot A \cdot d^{-\alpha} \cdot 10^{-\delta/10}}} \tag{4}$$

$$P_r = \frac{P_t}{4\pi^2 \cdot (d/\lambda)^r} \left\{1+\beta^2+2\beta cos\left(4\pi h^2/d\lambda\right)\right\}$$

where A, $d^{-\alpha}$ and $10^{-\delta/10}$ stand for constants that depend on transmit/receive antenna gains, path loss, and shadowing loss respectively. α is a path loss exponent, and δ is a standard deviation. Moreover, P_r is the received power, d is the distance between antennas in transmitter and receiver, h is the antenna height, λ is the wavelength of the propagation signal, β is the reflection rate, and γ is the path-loss factor [32].

From the equations, we know that the channel capacity of all links can be calculated based on the global information such as locations, transmit power of all base stations and users. However, collecting such global information in real-time is difficult in the real world. Thus, by only collecting the traffic patterns and historical links conditions of SBSs, immediately predicting the future traffics and channel states to intelligent configure the TDD modulation is our main research goal. In the next part, we first introduce the high mobility model of our scenario and then illustrate the TDD Up/Down-link resource allocation problem with mathematical models.

B. Mobility Model

In contrast to traditional wireless networks, the considered HetNet exhibits high flexibility, and nodes can continuously move and be added/removed dynamically. In this research, we consider the high mobility users with different speeds randomly go through the grid road. Consider there are $\mathbf{A_{old}} = \{a_1, a_2, \ldots, a_N\}$ users exist in the network. If M nodes are added to the network in an instant, they are denoted by $\mathbf{B} = \{b_1, b_2, \ldots, b_M\}$. Therefore, the total nodes, in the considered system, are represented by $\mathbf{A} = \mathbf{A_{old}} \cup \mathbf{B} = \{a_1, a_2, \ldots, a_N, a_{N+1}, a_{N+2}, \ldots, a_{N+M}\}$. Each node in the considered network is represented by its specific features, i.e., speed, latitude, longitude, and height. As shown in Fig. 2, for simplifying the research, we consider there are three kinds of mobility users. One is the high-speed vehicles ($\geq 10m/s$) go through the grid-like road which is constructed with many same-size squares covered by SBSs. The second kind of user is the pedestrians with relatively slow speed ($\leq 3m/s$). Those vehicles and pedestrians move straight along the road until the next intersections. At the intersection, the mobility node chooses whether to move straight, turn right or left with fixed probability $p_{\rm s}$, $p_{\rm r}$, and $p_{\rm l}$, and $p_{\rm s} + p_{\rm r} + p_{\rm l} = 1$. Besides, there are UAVs move in circles around the SBSs with ultra high speed ($\geq 50m/s$). For such a scenario, the moving orbits and the channel condition of corresponding connections become complex and hard to be tracked.

C. Problem Statement

The TDD is a promising full-duplex technology that tailors the channel resources into N_{tdd} sub-frames, in which N_{tup} sub-frames are for Up-link and N_{tdown} sub-frames are for Down-link. For each time slot, the configuration of N_{tup} and N_{tdown} can be correspondingly changed by TDD algorithm depending on the changed channel condition and requirement. The incorrectly allocated sub-frames for Up/Down-link cause link congestion and channel waste, which significantly affect the transmission efficiency and decrease network throughput. Therefore, the research objective of this paper is to adaptively change TDD configuration to minimize the gap between the required channel capacity and allocated channel capacity for Up/Down link. In other words, the objective of our proposal is to keep a high transmission data rate of all up-links and down-links in the whole network. Thus, the problem can be formulated as follow:

$$D_{total} = \underset{(R_{up}, R_{down}) \in W}{\arg\max} \; (D_{up} + D_{down})$$
$$\text{s.t.} \begin{cases} D_{up} = min(\Sigma_{i=1}^{N_{sbs}} \Sigma_{j=1}^{M_{up}^i} \lambda_{up}^i, C_{sum} \cdot R_{up}) \\ D_{down} = min(\Sigma_{i=1}^{N_{sbs}} \Sigma_{j=1}^{M_{down}^i} \lambda_{down}^i, C_{sum} \cdot R_{down}) \\ W = \{(1,9),(2,8),(3,7),(4,6),(5,5),(6,4)\} \\ R_{up} = \frac{N_{up}}{N_{up}+N_{down}}, R_{down} = 1 - R_{up} \end{cases} \tag{5}$$

Here the D_{total} denotes the total transmission rate of the network. The $D_{total} = D_{up} + D_{down}$, where the D_{up}/D_{down} represent the total transmission rate of Up/Down links respectively. The demanded data rate of Up-link i and Down-link j are donated as λ_{up}^i and λ_{down}^j. Moreover, C_{sum} means the total channel capacity calculated with the equation. 2. And the allocated ratio of up/down-link is represented with pair $<R_{up}$, $R_{down}>$. Set W stands for all TDD configurations [10], which contains 6 pairs of $<N_{up}, N_{down}>$ ratio.

IV. Proposed deep reinforcement learning based Up/Down-link resource allocation

With equations mentioned in Section. III, we can calculate the optimal allocation of TDD Up/Down-resources with instant information of users transmit power, location and data requirement. However, in the high mobility and heterogeneous structure of the considered 5G network, the channel conditions and data requirements are impossible to be correctly collected in real-time. Especially with various types of users, the data requirements of users are changing continuously. Thus, instead of the conventional method only utilizes the current (i.e., last time slot) condition for TDD configuration [12], our proposal considers the historical information including traffic patterns and past channel conditions as input to predict future sate and intelligently change TDD configuration to fit for the dynamic changing environment. Without the requirement of users' information, the signaling overhead can be decreased. Furthermore, with the dynamic learning of deep reinforcement learning (DRL), the algorithm in base stations can change the TDD configuration in advance to adapt to the predicted future network traffic and channel condition, which avoids potential network congestion and significantly improves the channel utilization.

In this section, we introduce the proposed DRL based intelligent TDD scheduling algorithm in two parts. In the first part, we simply introduce the employed deep learning model and deep-Q learning flow. Then, the proposed adaptively learning and scheduling algorithm is illustrated in terms of online Q-learning and offline memory experience replay.

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.2020.3005495, IEEE Journal on Selected Areas in Communications

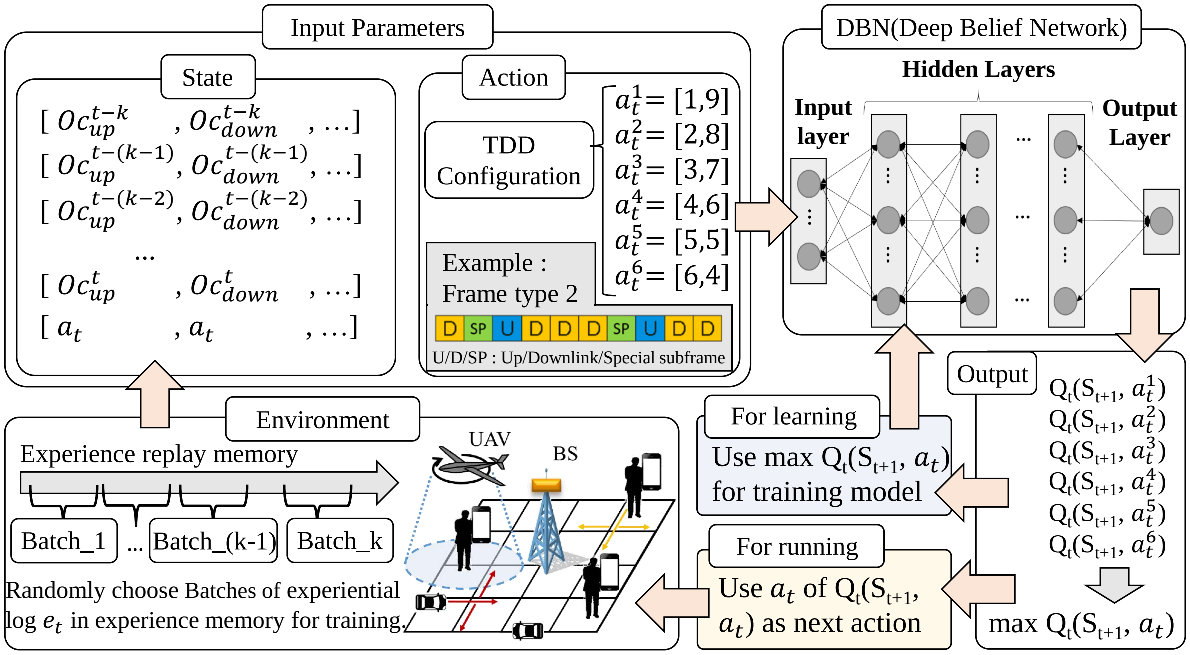


Fig. 3: Overview of the deep reinforcement learning based Up/Down-link resource allocation.

A. Deep Belief Neural Network for Feature Extraction

To extract the features of the complex network environment, we employ the deep belief architecture (DBA) as the encoder in our proposal. As shown in Fig. 3, the chosen DBA is constructed with L layers, the input and output are formatted as matrix x_{input} and y_{output}. Consider the q^{th} layer l_q constructed with N_{unit} neural units, b_j is the bias of unit j and w_{ij} denotes the weight of link between units i in l_{q-1} and unit j in (l_q). The training of the DBA consists of two steps, namely forward propagation and back-propagation processes. The forward propagation is used to construct the neural network structure and activate output, while the back-propagation is used to adapt the structure and fine-tune the values in matrix W and B, which constructed with all wights and bias of between layers. For forward propagation, there are two operations of linear activation and non-linear activation between two layers in DBA as shown in Eqs. 6 and 7.

$$u_i^{(l_q)} = \sum_j a_j^{(l_q-1)} w_{ji}^{(l_q)} + b_i^{(l_q)}, \tag{6}$$

$$a_j^{(l_q)} = f(u_i^{(l_q)}). \tag{7}$$

Here, the $a_j^{(l_q-1)}$ is the output of unit i in l_{q-1}^{th} layer. $u_i^{(l_q)}$ and $a_j^{(l_q)}$ denote the output of linear activation and non-linear activation of layer l_q. Then, the Mean Square Error (MSE) is used as the loss function in the final layer as follow:

$$loss(W, B, x_{input}, y_{output}) = \frac{1}{\hat{m}} \sum_{j=1}^{\hat{m}} (\hat{y}_{output} - y_{output})^2, \tag{8}$$

Here, y_{output} is the calculated output in each iteration with forward propagation. $\hat{y}_{output}$ stands for the benchmark Q-value collected from the learning process, $\hat{m}$ is the set number of the training data.

With the continuous online training process, the final purpose is to minimize the value of $loss(w, b, x_{input}, y_{output})$. For each iteration, the gradient descent method is adopted to fine-tune the matrix W and B. Consider the η is the used learning rate, the back-propagation process to fine-tune can be represented as:

$$W = W + \eta \frac{\partial loss(W, B, x_{input}, y_{output})}{\partial W}, \tag{9}$$

$$B = B + \eta \frac{\partial loss(W, B, x_{input}, y_{output})}{\partial B}, \tag{10}$$

B. Deep Q-learning Model

In reinforcement learning, the agents continuously collect states from the environment and choose action base on calculated rewards from feedback. As shown in fig. 3, in our considered scenario, the agents are the base stations, and the actions are the candidate TDD configurations. After all base stations choose the configuration, the network running for the next time slot and the feedback of network performance such as throughput, packet loss rate and end-to-end delay are collected to calculate the rewards of corresponding actions. In such a process, the state and actions of all agents are used as input x_{input} in the learning system.

As the x_{input} including state and actions, we format the two parts separately.

1) State: We format the state s_t^k of base station k at time t to a vector, the element is a different kind of features of the base station including up/down-link occupancy Oc_{up}^t and Oc_{down}^t, buffer occupancy and so on.

$$s_t^k = (Oc_{up}^t, Oc_{down}^t, \zeta_{up}, \zeta_{down}, ...) \tag{11}$$

$$Oc_{up}^t = \Sigma_{q=1}^{N_{up}^i} \lambda_{up}^q / C_q^t \cdot R_{up} \times \%$$

$$Oc_{down}^t = \Sigma_{q=1}^{N_{down}^i} \lambda_{down}^q / C_q^t \cdot R_{down} \times \%$$

$$\zeta_{up} = \Sigma_{q=1}^{N_{up}^i} C_q^t / C_q(max) \cdot R_{up} / N_{up}^i \times \%$$

$$\zeta_{down} = \Sigma_{q=1}^{N_{down}^i} C_q^t / C_q(max) \cdot R_{down} / N_{down}^i \times \%$$

where, C_q^t represents the average channel capacity of link L_q during time slot t. The ζ_{up} and ζ_{down} denote the channel condition of all Up/Down-links to the base station, which can be calculated with static channel state information (CSI) of each time slot [33]. The $C_q(max)$ stands for the ideal channel capacity of link L_q when no interference at all.

We regard all of the state parameters as discrete values and normalize all values. Instead of conventional Q-learning used in our previous work in [31] that only considers the state of one time slot, states of historical time sequence with N_t time slots are considered for representing the state S_t in time slot t. The S_t can be formulated as a vector as

$$S_t = (s_{t-N_t}, s_{t-N_t+1}, ..., s_{t-1}) \tag{12}$$

2) Action: In our research, the action space of each station is the configurations of TDD sub-frame allocation. We consider there are 6 different types of TDD configuration which are employed in LTE-release 11 and beyond [10]. Thus, the action a collected from set action space W is defined as

$$\begin{gathered} a = (N_{up}, N_{down}) \in W \\ W = \{(1,9),(2,8),(3,7),(4,6),(5,5),(6,4)\} \end{gathered} \tag{13}$$

With the formatted state and action, the input x_{input} is constructed with a tuple $< S_t, a_t >$. The structure of the formatted input including state and action is shown as in fig. 3

3) Reward: The agent at first chooses action following ϵ-greedy policy, the reward $r(S_t, a)$ is calculated to evaluate the feedback of chosen action a in the state S_t. As mentioned in the equation. 5, our goal is to maximize the data transmission rate which is related to the channel occupancy. Therefore, the reward function can be defined as

$$r(S_t, a) = r_{up} + r_{down} \tag{14}$$

$$r_{up} = \begin{cases} c_a \cdot Oc_{up}^{t+1} & (Oc_{up}^{t+1} \leq 1.0) \\ -c_a \cdot Oc_{up}^{t+1} + d_a & (Oc_{up}^{t+1} > 1.0) \end{cases}$$

$$r_{down} = \begin{cases} c_b \cdot Oc_{down}^{t+1} & (Oc_{down}^{t+1} \leq 1.0) \\ -c_b \cdot Oc_{down}^{t+1} + d_b & (Oc_{down}^{t+1} > 1.0) \end{cases}$$

where, r_up/r_down denote the up-down channel utility. The c_a,c_b,d_a, and d_b are constant coefficient. The total rewards mainly depend on the channel occupancy of the next time slot with chosen TDD configuration.

Algorithm 1 Algorithm of experience replay

```
Input: Batch size $N_{batch}$, training epoch $N_{epoch}$, learning rate $\vartheta$.
Randomly choose $e$ from $DBA_k$ with batch size $N_{batch}$.
for $i = 1$ to $N_{batch}$ do
    Format the $x_{input}$ from $e_i$.
    for $a_t \in W$ do
        List all possible next action $a_{t+1}$.
    end for
    update the $y_{output}$ by Eq. (14).
end for
Train $DBA_k$ with $(x_{input}, y_{output}, N_{epoch}, \vartheta)$ by Eqs. (6-10) .
```

In the following part, we introduce the DRL based Up/down-link resource allocation algorithm in terms of the online learning process and offline memory experience replay process.

C. Q-value Iteration based Online Training and Experience Replay Mechanism

The Online training process includes two parts, the information collection and online training. In the considered heterogeneous network, the control control is hard to be deployed. Thus, the base station need to collect the state and reward information continuously in distribute manner. The transmit demand packets λ^q of all up and down links are recorded in each time slot t. Besides, the nodes in the small cell report the static CSI of transmitter and receiver to the base station in every time slot. Thus, the channel occupancy Oc^t of all up and down links can be calculated with equation.11, and the state s_t^k of base station k in time slot t can be formatted correspondingly. Meanwhile, the reward r_{t-1} of last time slot is calculated with equation.14.

For training the agents to learn the policy, the Q-value Q_t is used as $\hat{y}_{output}$ for the DBA model. In time t, The agent format the state S_t with equation.12 and do action a_t. After one time slot, the state of environment correspondingly change with the chosen action a_t of agent, and the new reward $r(S_t, a)$ are collected by the agent. Thus, the $\hat{y}_{output}$ updating operation can be defined as

$$\begin{aligned} \hat{y}_{output} = Q(s_{t+1}, a_t) = (1-\vartheta)Q(s_t, a_t) + \\ \vartheta \left\{ r(S_t, a_t) + \gamma \max_{a_{t+1}} r(S_{t+1}, a_{t+1}) \right\}. \end{aligned} \tag{15}$$

Here, ϑ is the learning rate and γ is the discount rate which is to adjust the weight of reward in the next state. The agent decides action based on ϵ greedy algorithm to cover all possible actions. Consider i is a random decimal between 0 and 1, ϵ-greedy algorithm is defined as

$$a_t = \begin{cases} Random \quad a \in W & (i \leq \epsilon) \\ \arg\max_{a \in W}(Q(S_{t+1}, a_t) & (i > \epsilon) \end{cases} \tag{16}$$

To further improve the learning efficiency, an experience replay memory mechanism is also used in our algorithm for better convergence when the rewards are of low variance [34]. Let DBA_k denotes the DBA weight matrix stored in the sbs_k,

Algorithm 2 Algorithm of deep reinforcement learning based intelligent Up/Down-link resource allocation

```
Input: learning count $N_{learn}$, learning time range $N_{range}$, action change
    interval $N_{change}$, experience replay count $N_{replay}$.
for $i = 1$ to $N_{learn}$ do
    $Oc_{up}^{t}$,$Oc_{down}^{t}$ randomly initialized.
    for $j = 1$ to $N_{range}$ do
        $t = 0$
        Network runs following system model. (3) and (4).
        if $j$ % $N_{change}$ = 0 then
            $t = t + 1$
            for $k = 1$ to $N_{sbs}$ do
                Collect $Oc_{up}^{t}$,$Oc_{down}^{t}$ by Eq. (11).
                Calculate $r_{up}$, $r_{down}$ by Eq. (14).
                Choose action by the $\epsilon$-greedy algorithm Eq. (16).
                The $DBA_k$ store experience memory $e_t$.
            end for
        end if
    end for
    for $j = 1$ to $N_{replay}$ do
        for $k = 1$ to $num - BS$ do
            $DBA_k$ randomly do Alg. 1.
        end for
    end for
end for
```

and the experiential log of each time slot is denoted as $e_t =< S_t, a_t, r_t, S_{t+1} >$. In the learning phase, the base stations store the e_t continuously until the packets forwarding process of the whole network is finished. For each training interval, the sbs_k randomly select training batch from the experience replay memory and train the DBA_k. The algorithm of experience replay is defined as Alg. 1, and the whole learning algorithm is as shown in Alg. 2.

TABLE I: The simulation parameters

Name	Value	Name	Value
Learning Count N_{learn}	30	Learning rate ϑ	0.1
Training epoch N_{epoch}	10	Discount rate γ	0.9
Experience count N_{replay}	10	Block Height	100m
Batch size N_{batch}	32	Block Width	100m
Learning time N_{range}	20000	Speed of pedestrian	3 m/s
Action interval N_{change}	30	Speed of vehicle	20 m/s
Hidden layers	4	Speed of UAV	50 m/s
Turning probability p_r,p_l	0.1	Straight Probability p_s	0.8

V. Performance Evaluation

In this section, we evaluate the performance of our proposal in terms of learning efficiency and network performance aspects comparing with conventional algorithms. The simulation is constructed with Python and deep learning framework of TensorFlow/Keras.

To simplify the simulation, we modeled the heterogeneous with 3 SBS with 3 small-cell and 1 MBS with the corresponding macro-cell as shown in Fig. 1. Consider there are totally 16 Blocks constructed as 4×4 grids, and each block is bounded by four streets. The pedestrians and vehicles move along streets, and the UAVs is hovering around the BSs. The ratio of pedestrians and users is set to 1:1. The number of UAVs is equal to the number of BSs, and all users move following the mobility model illustrated in Section III. The user leaves the network when it arrives at the bound of grids, and a new one joins the network immediately.

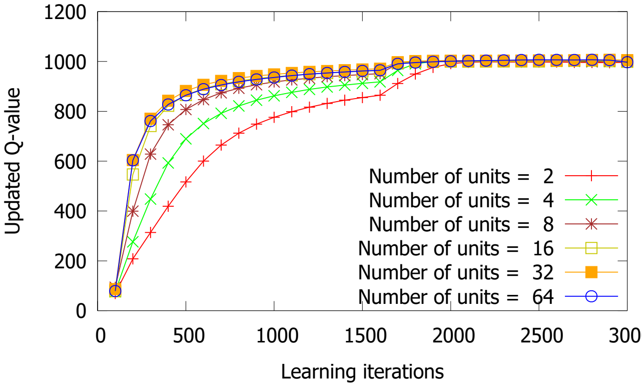


Fig. 4: The Q-value convergence speed comparison in terms of different number of units in each hidden layer.

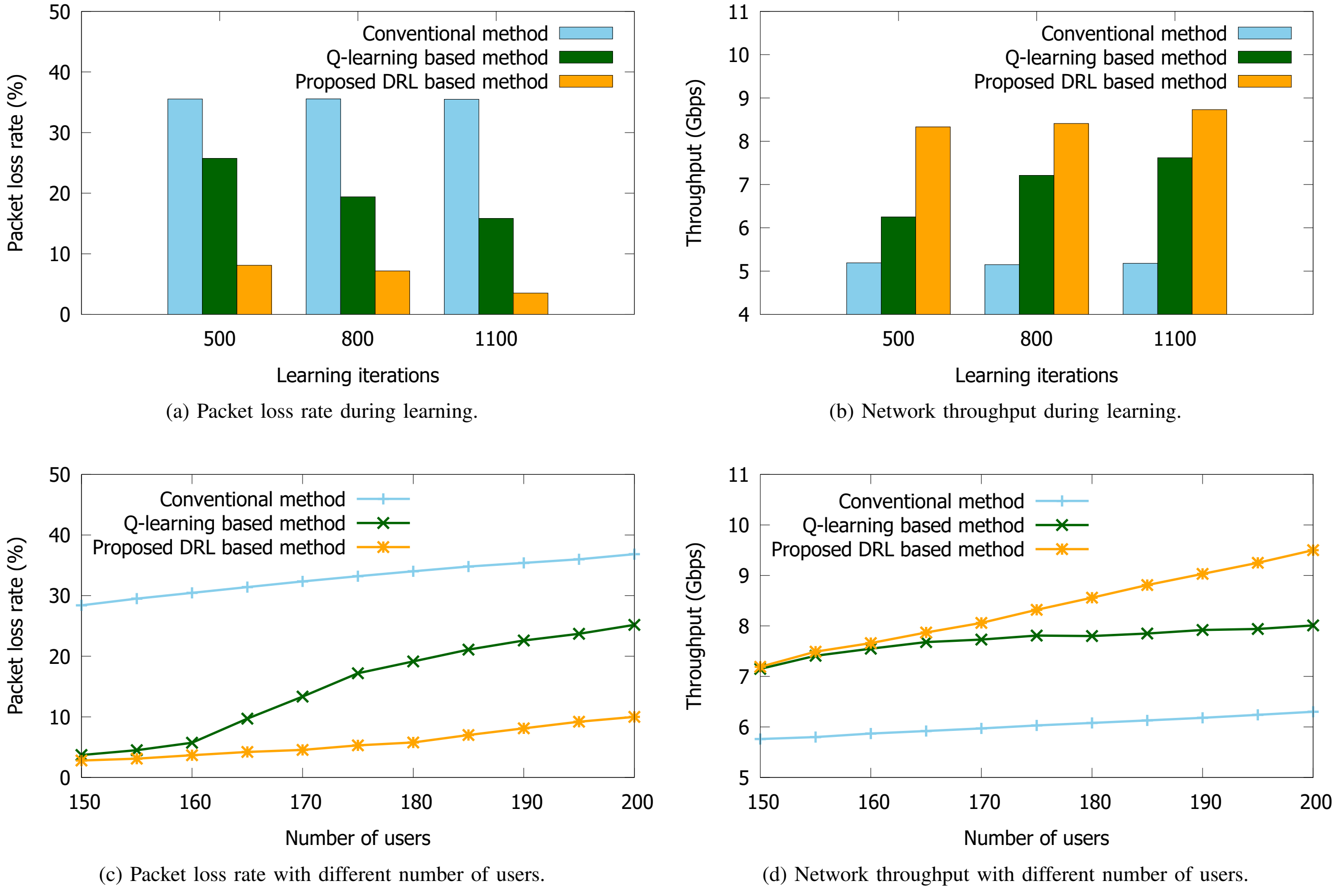
The patterns of user's traffic demand follows the data generation setting in our previous work in [26]. The average traffic demand ratio of Up-link and Down-link is set to 1:4, and the average traffic demand including both Uploading and downloading for each node is 50Mbps. The 5G beam coverage of SBS is considered to be fixed, and the maximum coverage distance is 200m in our simulation. We assume the MBS can cover the whole network to connect with all SBSs. The bandwidths of all sub-channels are considered to be the same with a fixed value of 100MHz, 10 orthogonal sub-channels are used for each link between MBS and SBS, and the sub-channels are enough for all users inner the same small cell. We also assume packets in D2D communications not go though BS and not calculated in the network throughput. Some other parameters are listed in Table. I.

We firstly run different structures of deep neural networks to analyze the training efficiency as shown in fig. 4. There are 4 hidden layers in the DBA, and the number of neural units in each layer is set from 2 to 64. In the figure, the x-axis represents the learning iteration and the y-axis is the updated average Q-values with continuous learning which is smoothed to show the trends. The result demonstrates that the deep neural network with different neural units all convergent after 2000 times of training. However, the structure with 32 neural units in each layer has the best convergence performance which is also used as the training structure in the next network performance evaluation part.

In the network performance comparison part, we compare both the network throughput and average packet loss rate between our proposal and conventional dynamic TDD resource allocation focus on the past channel conditions [12] and shallow Q-learning based TDD radio resource allocation algorithm proposed in our previous work [31]. For simplify the simulation, we assume the past channel conditions are know in the simulation.

We at first analyze the network performance in terms of different learning iterations. As shown in Fig. 5(a) and Fig. 5(b), with the learning process going on, the throughput and packet loss rate of conventional method which is scheduled just based on current network state is almost constant. However, the intelligent Q-learning based method and our

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.2020.3005495, IEEE Journal on Selected Areas in Communications



(a) Packet loss rate during learning.

(b) Network throughput during learning.

(c) Packet loss rate with different number of users.

(d) Network throughput with different number of users.

Fig. 5: The network performance of the proposal compares with the conventional dynamic and Q-learning based TDD resource allocation algorithm.

proposal continuously improve, and the proposal shows better performance than the shallow Q-learning based method. After 1100 times of iteration, the throughput of our proposal almost touches the upper-bound of total network capacity.

Then, we further illustrate the network performance with the different number of users in the network. We collect the trained DBA weight matrix which is trained with 1100 learning iterations to run the network and collect throughput and packet loss rate. As shown in Fig. 5(a) and Fig. 5(b), the throughput and packet loss rate of both the conventional dynamic TDD resource allocation and shallow Q-learning based algorithm are worse than our proposal. This is because of the wrongly allocated channel resource of the conventional algorithm mismatches the future traffic demands of users.

VI. Conclusion

Deep learning is a promising artificial intelligence (AI) tool to enable intelligence for various network functions. In this paper, we explored employing deep reinforcement learning to adaptively allocate TDD Up/Down-link resources in the high mobility 5G HetNet. With the deep neural network, the proposed reinforcement learning algorithm can reflect the complex states of HetNet, potential TDD configuration and corresponding feedback with continuous Q-value iterations. The simulation shows that the proposal can dynamically change TDD configuration to optimize Up/Down-link radio resource allocation with low overhead, which significantly improves the network performance including network throughput and packet loss rate. This research is the initial attempt to involve DRL into dynamic TDD configuration in HetNet, there are still many works can be further explored. For our future work, we will continue to consider the more realistic scenario with varying movement of nodes and improve the computation overhead of the deep reinforcement learning algorithm.

References

[1] X. Ge, S. Tu, G. Mao, C. Wang, and T. Han, “5g ultra-dense cellular networks,” *IEEE Wireless Communications*, vol. 23, pp. 72–79, February 2016.

[2] K. Hamdi, W. Zhang, and K. B. Letaief, “Opportunistic spectrum sharing in cognitive mimo wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 8, pp. 4098–4109, August 2009.

[3] G. Gui, H. Huang, Y. Song, and H. Sari, “Deep learning for an effective nonorthogonal multiple access scheme,” *IEEE Transactions on Vehicular Technology*, vol. 67, pp. 8440–8450, Sep. 2018.

[4] B. D. Van Veen and K. M. Buckley, “Beamforming: a versatile approach to spatial filtering,” *IEEE ASSP Magazine*, vol. 5, pp. 4–24, April 1988.

[5] H. Huang, Y. Peng, J. Yang, W. Xia, and G. Gui, “Fast beamforming design via deep learning,” *IEEE Transactions on Vehicular Technology*, vol. 69, pp. 1065–1069, Jan 2020.

[6] J. Flordelis, F. Rusek, F. Tufvesson, E. G. Larsson, and O. Edfors, "Massive mimo performance—tdd versus fdd: What do measurements say?," *IEEE Transactions on Wireless Communications*, vol. 17, pp. 2247–2261, April 2018.

[7] 3GPP, TR 22.885 v14.0.0, "http://www.3gpp.org/DynaReport/22885.htm," *Study on LTE Support for V2X Services*, May, accessed 2019.

[8] Z. Shen, A. Khoryaev, E. Eriksson, and X. Pan, "Dynamic uplink-downlink configuration and interference management in td-lte," *IEEE Communications Magazine*, vol. 50, pp. 51–59, November 2012.

[9] Ming Ding, D. L. Pérez, A. V. Vasilakos, and Wen Chen, "Dynamic tdd transmissions in homogeneous small cell networks," in *2014 IEEE International Conference on Communications Workshops (ICC)*, pp. 616–621, Sydney, Australia, June 2014.

[10] A. Khoryaev, A. Chervyakov, M. Shilov, S. Panteleev, and A. Lomayev, "Performance analysis of dynamic adjustment of tdd uplink-downlink configurations in outdoor picocell lte networks," in *2012 IV International Congress on Ultra Modern Telecommunications and Control Systems*, pp. 914–921, Oct 2012.

[11] X. Ge, H. Cheng, G. Mao, Y. Yang, and S. Tu, "Vehicular communications for 5g cooperative small-cell networks," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 7882–7894, Oct 2016.

[12] M. Ding, D. López-Pérez, R. Xue, A. V. Vasilakos, and W. Chen, "On dynamic time-division-duplex transmissions for small-cell networks," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 8933–8951, Nov 2016.

[13] N. Kato, Z. M. Fadlullah, B. Mao, F. Tang, O. Akashi, T. Inoue, and K. Mizutani, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Wireless Communications*, vol. 24, pp. 146–153, June 2017.

[14] B. Mao, Z. M. Fadlullah, F. Tang, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "Routing or computing? the paradigm shift towards intelligent computer network packet transmission based on deep learning," *IEEE Transactions on Computers*, vol. 66, pp. 1946–1960, Nov 2017.

[15] F. Tang, B. Mao, Z. M. Fadlullah, and N. Kato, "On a novel deep-learning-based intelligent partially overlapping channel assignment in sdn-iot," *IEEE Communications Magazine*, vol. 56, pp. 80–86, Sep. 2018.

[16] F. Tang, Y. Kawamoto, N. Kato, and J. Liu, "Future intelligent and secure vehicular network toward 6g: Machine-learning approaches," *Proceedings of the IEEE*, vol. 108, pp. 292–307, Feb 2020.

[17] F. A. Aoudia and J. Hoydis, "Model-free training of end-to-end communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 37, pp. 2503–2516, Nov 2019.

[18] J. L. Carrera Villacrés, Z. Zhao, T. Braun, and Z. Li, "A particle filter-based reinforcement learning approach for reliable wireless indoor positioning," *IEEE Journal on Selected Areas in Communications*, vol. 37, pp. 2457–2473, Nov 2019.

[19] C. Chiang, W. Liao, T. Liu, I. K. Chan, and H. Chao, "Adaptive downlink and uplink channel split ratio determination for tcp-based best effort traffic in tdd-based wimax networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 182–190, February 2009.

[20] S. Sun, Q. Gao, Y. Peng, Y. Wang, and L. Song, "Interference management through comp in 3gpp lte-advanced networks," *IEEE Wireless Communications*, vol. 20, pp. 59–66, February 2013.

[21] T. Ding, M. Ding, G. Mao, Z. Lin, A. Y. Zomaya, and D. López-Pérez, "Performance analysis of dense small cell networks with dynamic tdd," *IEEE Transactions on Vehicular Technology*, vol. 67, pp. 9816–9830, Oct 2018.

[22] Sida Song, Yongyu Chang, Hao Xu, Daixu Zheng, and Dacheng Yang, "Energy efficiency model based on stochastic geometry in dynamic tdd cellular networks," in *2014 IEEE International Conference on Communications Workshops (ICC)*, pp. 889–894, Sydney, Australia, June 2014.

[23] T. Yang, R. Zhang, X. Cheng, and L. Yang, "A graph coloring resource sharing scheme for full-duplex cellular-vanet heterogeneous networks," in *2016 International Conference on Computing, Networking and Communications (ICNC)*, pp. 1–5, Feb 2016.

[24] M. Al-Imari, M. Ghoraishi, P. Xiao, and R. Tafazolli, "Game theory based radio resource allocation for full-duplex systems," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp. 1–5, May 2015.

[25] F. Tang, Z. M. Fadlullah, N. Kato, F. Ono, and R. Miura, "Ac-poca: Anticoordination game based partially overlapping channels assignment in combined uav and d2d-based networks," *IEEE Transactions on Vehicular Technology*, vol. 67, pp. 1672–1683, Feb 2018.

[26] F. Tang, Z. M. Fadlullah, B. Mao, and N. Kato, "An intelligent traffic load prediction-based adaptive channel assignment algorithm in sdn-iot: A deep learning approach," *IEEE Internet of Things Journal*, vol. 5, pp. 5141–5154, Dec 2018.

[27] F. Tang, B. Mao, Z. M. Fadlullah, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "On removing routing protocol from future wireless networks: A real-time deep learning approach for intelligent traffic control," *IEEE Wireless Communications*, vol. 25, pp. 154–160, February 2018.

[28] Z. M. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Communications Surveys Tutorials*, vol. 19, pp. 2432–2455, Fourthquarter 2017.

[29] Y. Zhou, Z. M. Fadlullah, B. Mao, and N. Kato, "A deep-learning-based radio resource assignment technique for 5g ultra dense networks," *IEEE Network*, vol. 32, pp. 28–34, November 2018.

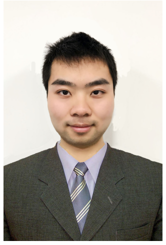
[30] H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, H. Sari, and F. Adachi, "Deep learning for physical-layer 5g wireless techniques: Opportunities, challenges and solutions," *IEEE Wireless Communications*, vol. 27, pp. 214–222, February 2020.

[31] Y. Zhou, F. Tang, Y. Kawamoto, and N. Kato, "Reinforcement learning based radio resource control in 5g vehicular network," *IEEE Wireless Communications Letters*, pp. 1–1, 2019.

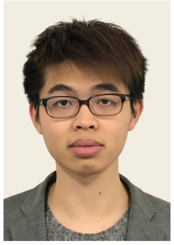
[32] P. Fazio, F. De Rango, C. Sottile, and C. Calafate, "A new channel assignment scheme for interference-aware routing in vehicular networks," in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Budapest, Hungary, May 2011.

[33] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Transactions on Information Theory*, vol. 45, pp. 2007–2019, Sep. 1999.

[34] J. Foerster, N. Nardelli, G. Farquhar, T. Afouras, P. H. S. Torr, P. Kohli, and S. Whiteson, "Stabilising experience replay for deep multi-agent reinforcement learning," in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pp. 1146–1155, Sydney, NSW, Australia, 2017.



Fengxiao Tang (S'15-M'19) received the B.E. degree in measurement and control technology and instrument from the Wuhan University of Technology, Wuhan, China, in 2012 and the M.S. degree in software engineering from the Central South University, Changsha, China, in 2015. He received the Ph.D. degrees from the Graduate School of Information Science, Tohoku University, Japan. Currently, He is an Associate Professor at the Graduate School of Information Sciences (GSIS) of Tohoku University. His research interests are unmanned aerial vehicles system, IoT security, game theory optimization, network traffic control and machine learning algorithm. He was a recipient of the prestigious Dean's and President's Awards from Tohoku University in 2019, and several best paper awards at conferences including IC-NIDC 2018, GLOBECOM 2017/2018. He was also a recipient of the prestigious Funai Research Award in 2020.



Yibo Zhou (S'17) received his B.E. and M.S. degree from the Graduate School of Information Sciences (GSIS) at Tohoku University in 2018 & 2020, Japan. His research interest is machine learning, TDD optimization and network traffic control.

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.2020.3005495, IEEE Journal on Selected Areas in Communications

Nei Kato (M'04-SM'05-F'13) is a full professor (Deputy Dean) with Graduate School of Information Sciences(GSIS) and the Director of Research Organization of Electrical Communication(ROEC), Tohoku University, Japan. He has been engaged in research on computer networking, wireless mobile communications, satellite communications, ad hoc & sensor & mesh networks, smart grid, AI, IoT, Big Data, and pattern recognition. He has published more than 400 papers in prestigious peer-reviewed journals and conferences. He is the Vice-President (Member & Global Activities) of IEEE Communications Society(2018-), the Editor-in-Chief of IEEE Transactions on Vehicular Technology(2017-), and the Chair of IEEE Communications Society Sendai Chapter. He served as the Editor-in-Chief of IEEE Network Magazine (2015-2017), a Member-at-Large on the Board of Governors, IEEE Communications Society(2014-2016), a Vice Chair of Fellow Committee of IEEE Computer Society(2016), and a member of IEEE Communications Society Award Committee (2015-2017). He has also served as the Chair of Satellite and Space Communications Technical Committee (2010-2012) and Ad Hoc & Sensor Networks Technical Committee (2014-2015) of IEEE Communications Society. His awards include Minoru Ishida Foundation Research Encouragement Prize(2003), Distinguished Contributions to Satellite Communications Award from the IEEE Communications Society, Satellite and Space Communications Technical Committee(2005), the FUNAI information Science Award(2007), the TELCOM System Technology Award from Foundation for Electrical Communications Diffusion(2008), the IEICE Network System Research Award(2009), the IEICE Satellite Communications Research Award(2011), the KDDI Foundation Excellent Research Award(2012), IEICE Communications Society Distinguished Service Award(2012), IEICE Communications Society Best Paper Award(2012), Distinguished Contributions to Disaster-resilient Networks R&D Award from Ministry of Internal Affairs and Communications, Japan(2014), Outstanding Service and Leadership Recognition Award 2016 from IEEE Communications Society Ad Hoc & Sensor Networks Technical Committee, Radio Achievements Award from Ministry of Internal Affairs and Communications, Japan (2016), IEEE Communications Society Asia-Pacific Outstanding Paper Award(2017), Prize for Science and Technology from the Minister of Education, Culture, Sports, Science and Technology, Japan(2018), Award from Tohoku Bureau of Telecommunications, Ministry of Internal Affairs and Communications, Japan(2018), and Best Paper Awards from IEEE ICC/GLOBECOM/WCNC/VTC. Nei Kato is a Distinguished Lecturer of IEEE Communications Society and Vehicular Technology Society. He is a fellow of The Engineering Academy of Japan and IEICE.