Spatial Deep Learning-Based Dynamic TDD Control for UAV-Assisted 6G Hotspot Networks

Van Dat Tuong, Wonjong Noh, and Sungrae Cho

***Abstract*—Compared to static time-division duplexing (TDD), dynamic TDD (D-TDD) has significantly increased the spectral efficiency of cellular networks. However, conventional systems operate based on exact channel state information (CSI), resulting in high communication overhead and delay. Spatial deep learning refers to using spatial geographical information as the training data. This study investigates a spatial deep learning-based D-TDD scheme for 6G hotspot networks. First, we represent geographical location information in forms of traffic demand density grid matrices. Second, we use spatial convolution filters to extract discriminative features of uplink and downlink service gains and harms, taking the traffic demand density grid matrices as the input. Subsequently, extracted feature matrices are processed with sparse convolution blocks to reduce computation cost for the classification. Finally, we develop novel deep dueling neural networks, leveraging the extracted features to efficiently learn the near-optimal radio slot configurations for all BSs. Numerical results show that the proposed approach improves average rate per user by 2.5%, 6%, and 523.3% over those achieved in state-of-the-art centralized D-TDD, the competitive reinforcement learning (RL), and greedy approaches, respectively. Additionally, the proposed approach achieves up to 98.7% of the data rate performance of the optimum scheme with an exhaustive search algorithm.**

***Index Terms*—Dynamic time division duplexing (D-TDD), geographic location information, hot-spot networks, spatial deep learning, unmanned aerial vehicle (UAV).**

I. Introduction

TIME-division duplexing (TDD), a promising multiplexing technique in 5G and 6G, can support asymmetric and coupled transmissions by allocating non-overlapping time slots for the uplink (UL) and downlink (DL) channels. In general, TDD schemes can be categorized as static or dynamic. In static TDD, time-slot configuration is pre-determined and remains fixed regardless of any arising interference and dynamic traffic demand requests. As a result, static TDD schemes cannot adapt to rapid traffic ranges, leading to the inefficient utilization of radio resources. Moreover, static TDD requires all base stations (BSs) of a small-cell network to synchronize their operations in both the UL and DL modes with variations in the traffic demand across network cells, which results in sub-optimal quality of service (QoS) for all user equipments (UEs) and overall network performance [1].

To address the limitations of static TDD schemes, dynamic TDD (D-TDD) schemes have been developed. In D-TDD, each time slot can be dynamically allocated for UL or DL transmissions, allowing the network to flexibly and promptly adapt to rapid changes in the traffic demands. Also, D-TDD allows network cells to operate at different time-slot configurations. Those can achieve improved QoS for all users, such as lower delay and higher throughput. Because of the benefits of D-TDD schemes, several researchers have explored their use for long-term evolution, 5G, and 6G network design [2]–[4]. However, migrating to a D-TDD system poses several challenges, such as severe cross-link interference control and complex signaling control [5]. Therefore, it is necessary to determine appropriate interference mitigation strategies and efficient signaling control methods to improve the network performance for D-TDD systems [6].

Unmanned aerial vehicles (UAVs) plays an important role in the upcoming 6G and future wireless networks that focus on spatial data transmissions. To be more specific, the modernest UAVs are equipped with radio components to establish stable line-of-sight air-to-ground data transmissions for ground user devices. With great benefits of a flying object, UAV can flexibly adjust its trajectory to serve ground users more efficiently. Moreover, the UAVs can be equipped with high-resolution cameras and three-dimensional (3D) lidar sensors that effectively collect the geographical map information of a large ground area. This geographical map information is useful for evaluating not only the strength of transmissions but also the interference between adjacent links. In particular, UAV-assisted wireless communication has been applied in industrial Internet of Things systems to effectively process diverse service requests [7]–[10].

A. Related Work

The existing D-TDD schemes can be categorized into centralized [11]–[14] and distributed schemes [15]–[17].

Razlighi *et al.* [11] developed centralized D-TDD schemes that maximized the rate region for a full-duplex wireless network by selecting the scheduling mode per time slot, i.e., reception, transmission, simultaneous reception and transmission, or silence (corresponding to DL, UL, combined DL and UL, or flexible time slot allocation), for every network node. Ghermezcheshmeh *et al.* [12] extended the work of Razlighi *et al.* [11] and integrated the centralized D-TDD

This research was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-RS-2022-00156353) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation) and in part by the National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIT) (RS-2023-00209125 and NRF-2023R1A2C1003003). *(Corresponding authors: Wonjong Noh and Sungrae Cho.)*

V. D. Tuong and S. Cho are with the School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, Republic of Korea. *(email: vdtuong@uclab.re.kr, srcho@cau.ac.kr)*

W. Noh is with the School of Software, Hallym University, Chuncheon 24252, Republic of Korea. *(email: wonjong.noh@hallym.ac.kr)*

TABLE I
COMPARISON OF THE STATE-OF-THE-ART STUDIES

Schemes	Objectives	Approaches	Outcomes/Features	Shortcomings
[11]	Maximizing the weighted sum-rate considering half-duplex and full-duplex transmission nodes	Centralized D-TDD	Achieved a near-optimal sum-rate performance	High computation cost and required full CSI
[12]	Maximizing the rate region of the network with interference alignment	Centralized D-TDD	Interference alignment helped improve the performance	High computation cost
[15]	Minimizing the inter-cell interference	Fully distributed multi-agent deep Q-network	Jointly optimized subframe configuration, channel assignment, and computation offloading	Did not simultaneously guarantee optimal performance for all network cells
[16]	Improving spectral efficiency for cellular mMIMO systems	Greedy algorithm	Polynomial time	Non-optimized spectral efficiency performance
[18]	Mitigating cross-link interference	Machine learning with lightweight feedforward neural network	Reduced computational complexity	Required channel estimation with radio frequency chain model
[19]	Maximizing data rate for dense wireless and mobile networks	Stackelberg game and DDPG-based deep learning	Reduced communication costs by estimating interference penalty	The communication overhead could not be thoroughly mitigated
[20]	Quickly adapting traffic pattern for 5G BS	Deep reinforcement learning based	Timely and efficient radio configuration	Only counting UL and DL buffers of the BSs, non-applicable
[21]	Improving radio utilization	Deep reinforcement learning based	Considered mobility and heterogeneous networks	Time complexity was not provided to validate with mobility
Proposed	Maximizing the sum rate for UAV-assisted industrial hotspot networks	RL-based deep dueling algorithm using geographical location information	Completely eliminates communication overhead by using only geographical information data	The offline training time is large owing to geographical information feature extraction

scheme [11] with interference alignment. Pedersen *et al.* [13] introduced a novel coordination scheme for TDD radio frame configurations between neighboring cells, which could help enhance the D-TDD system performance. Nwalozie *et al.* [14] investigated reconfigurable intelligent surfaces-aided D-TDD schemes, aiming to maximize system spectral efficiency while minimizing cross-link interference.

Unlike the centralized approaches, Song *et al.* [15] investigated the joint optimization problem of subframe configuration, channel assignment, and computation offloading in multicell D-TDD networks. Chowdhury *et al.* [16] investigated D-TDD for distributed antenna array massive MIMO systems. They proved that the performance of D-TDD depends significantly on the scheduling of UL and DL transmissions. A greedy algorithm is developed to solve the scheduling problem in polynomial time. Cavalcante *et al.* [17] studied bidirectional sum-power minimization beamforming to reduce cross-link interference for MIMO D-TDD networks. The authors proposed an alternating direction method of multipliers to obtain the optimal beamforming in a distributed manner.

Recently, several machine learning-based D-TDD approaches [18]–[21] have been developed. Tan *et al.* [18] studied general flexible duplexing techniques for 5G and 6G mobile communication networks. They aimed to minimize cross-link interference. Two machine learning algorithms are developed that used lightweight feedforward neural network to improve the performance while reducing computational complexity. Tuong *et al.* [19] aimed to maximize data rate for dense wireless and mobile networks. A novel framework combining Stackelberg game and deep deterministic policy gradient is developed to solve the formulated problem while reducing communication cost needed to obtain global channel state information. Bagaa *et al.* [20] focused on optimizing UL/DL pattern for 5G New Radio. They proposed deploying deep reinforcement learning (DRL) at the base station to quickly adapt to time-varying traffic pattern. Tang *et al.* [21] also addressed the TDD configuration problem in 5G vehicular networks with highly mobile user devices. A novel DRL-based algorithm was proposed to dynamically allocate radio resources.

It is highlighted that despite of the great outcomes, state-of-the-art schemes with centralized, distributed, machine learning, and game theoretical approaches have a common drawback of raising communication overhead for using exact CSI data. The proposed scheme completely eliminates this overhead by using only geographical information data. In other words, the proposed scheme does not depend on exact CSI data update. Therefore, it is considered a low-overhead solution.

B. Main Contributions

We aim to develop a high-performance D-TDD scheme that integrates the merits of the centralized and distributed approaches. Regarding centralized approach, the proposed D-TDD scheme considers UAV deployment to collect the geographical map of the network coverage area and obtain the channel conditions of all data transmissions. The proposed D-TDD scheme also regards the distributed approach by enabling each BS to observe its own features of data rate gain for its subscribed users and inter-cell interference harm toward its non-subscribed users. Subsequently, the near-optimal radio configurations are derived distributively for all BSs. The key novelty and features of the proposed scheme compared to

the existing approaches are summarized in Table I. The main contributions of this work can be summarized as follows:

- We formulated an non-convex D-TDD problem that maximizes the achievable system rate for the 6G hot-spot networks while using the spatial geographical location information instead of explicit channel state information (CSI), which can be efficiently obtained by UAVs.
- To this end, first, we represented the geographical location and load information for each network cell based on traffic demand density grid matrices. Different from the prior art study [22], the traffic demand density grid matrices considered summing product of relative distance between UE and its associated BS with the quantized traffic demand state. Second, we used spatial convolution filters to extract discriminative features of UL and DL service gains and harms. Furthermore, to reduce the computation cost for further classification, we proposed to continue processing the feature matrices with sparse convolutional blocks, e.g., speed-accuracy balancing and deep feature processing blocks. Third, we developed a novel deep dueling neural network to efficiently learn the near-optimal slot configurations for all BSs.
- Through simulations, we confirmed that the proposed D-TDD framework stably converges under various settings. The proposed D-TDD scheme also improves average rate per user by 2.5%, 6%, and 523.3% when it is compared to the state-of-the-art centralized D-TDD scheme [11], the competitive reinforcement learning (RL) such as deep Q-network (DQN), deep double Q-network (DDQN), and deep double policy gradient (DDPG), and greedy approaches, respectively. Additionally, the proposed approach achieves up to 98.7% of the data rate performance of the optimum scheme with the ES algorithm. Therefore, the proposed scheme is considered a near-optimal scheme.

The remaining paper is organized as follows: Section II introduces the system model and describes the problem formulation. Section III introduces the proposed algorithm based on deep dueling neural networks. The simulation results are discussed in Section IV. Section V presents the concluding remarks and highlights the future research directions.

II. System Model and Problem Formulation

A. Network Model

As illustrated in Fig. 1, we consider a cellular wireless network, in which multiple hot-spots (small-cells) are distributed in a macro cell coverage area. The BSs are expressed as by $k \in \mathcal{K} = \{0, 1, \ldots, K\}$, where $k = 0$ indicates the macro BS (MBS) and K is the total number of hot-spot small-BSs (SBSs). It is worth noting that practical complex antenna settings for the BSs only change the expression of computing channel gain and achievable data rate and they do not affect to the solution. Therefore, for the sake of simplicity, we assume that each BS operates with a single antenna. Specifically, the macro base station (MBS) employs 3D antenna pattern while each small base stations (SBSs) is equipped with omnidirectional antenna pattern. UEs are distributed uniformly and independently in each cell. We denote the UEs as $u \in \mathcal{U} = \{1, \ldots, U\}$, where U is the number of UEs. Based on the reference signal received power (RSRP), each UE selects the BS providing the maximal RSRP as the service BS and sets the corresponding association indicator value to 1, denoted by $\kappa_{u,k} = 1$. All BSs are assumed to reuse the same frequency resource. However, each BS serves its subscribed UEs using orthogonal frequency-division multiplexing transmissions. We denote the orthogonal sub-channels of each BS as $s \in \mathcal{S} = \{1, \ldots, S\}$, where S is the number of orthogonal sub-channels. Therefore, the transmissions of UEs that are associated with the same BS exhibit orthogonality, which eliminates intra-BS interference. Nevertheless, inter-BS interference exists between UEs associated with different BSs. We assume that several UAVs are employed to cooperatively and sufficiently cover the network area. They collect geographical location information that is processed to build a grid of location information of all BSs and UEs.

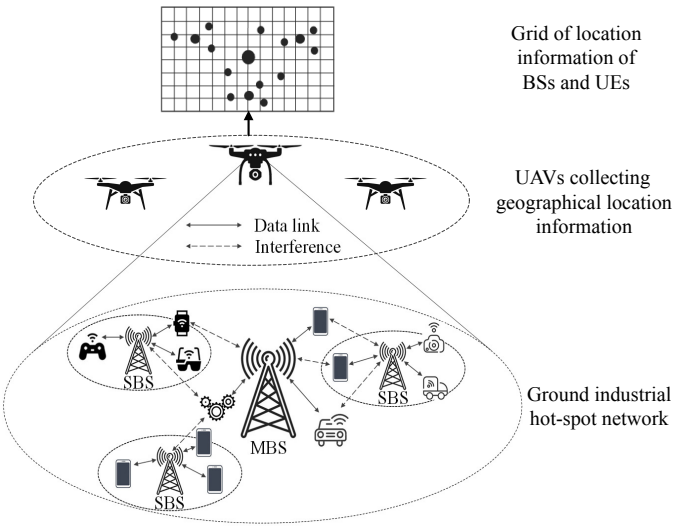


Fig. 1. Illustration of UAV-assisted industrial hot-spot networks.

Each network cell is assumed to operate in TDD mode with a fully dynamic and flexible frame configuration. In other words, all network cells can independently and dynamically switch time slot configurations for UL or DL transmissions. We consider the radio frame configurations, as indicated in Release 16 of 3GPP's Technical Specification (TS) 38.211 [23]. As per this standard, there exist five settings of the sub-carrier spacing and slot duration, determined using a numerology parameter $\mu \in \{0, \ldots, 4\}$, corresponding to $15 \times 2^{\mu}$ kHz and $1/2^{\mu}$ ms, respectively. Fig. 2 illustrates the radio frame configurations when $\mu = 0$, sub-carrier spacing is 15 kHz, and slot duration is 1 ms. The number of symbols in each slot and number of sub-frames in each radio frame are set as 14 and 10, respectively. As 5G NR allows symbol-level configurations, we can change the slot patterns for each network cell to adapt to changes in the UL and DL traffic demand.

B. Problem Formulation

We consider a static power scheme, which assumes the same UL transmission power for all UEs, and the same DL

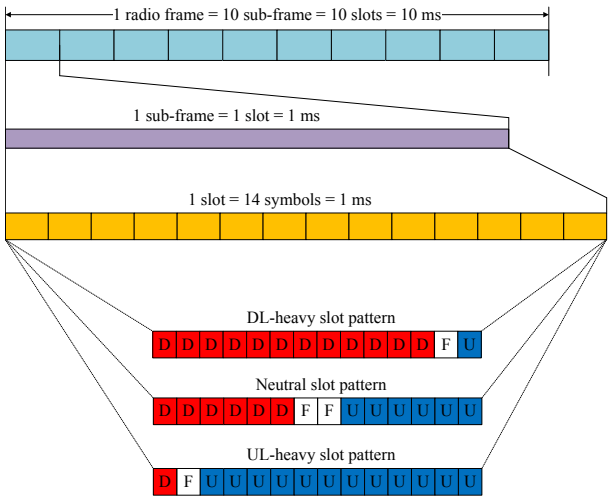


Fig. 2. Radio frame configurations: $\mu = 0$.

transmission power for all BSs. The sub-channel resources are uniformly and independently allocated to UEs. Let $p^{UL}_{u,k,s}$ and $p^{DL}_{u,k,s}$ denote the transmission power in the UL and DL between UE u and BS k on each sub-channel s, respectively. If sub-channel s is allocated to UE u, the assignment indicator $\eta_{u,k,s}$ is set as 1; otherwise, it is set as 0. Considering the significant heterogeneity in the network model, the channel power gain of the transmission between UE u and BS k over sub-channel s may be affected by nearby transmissions. Path-loss and fading models are applied to reflect the transmissions between MBS and MBS, MBS and SBS, MBS and UE, SBS and SBS/UE, and UE and UE. Specifically, the channel power gain is computed as follows:

$$g^s_{u,k} = |h^s_{u,k}|^2 \alpha_{u,k}, \tag{1}$$

where $h^s_{u,k}$ is the small-scale fading component, and $\alpha_{u,k}$ is the large-scale path-loss component.

According to 5G New Radio settings [24], each slot format is within a number of consecutive time slots denoted by N, which corresponds to a transmission time interval. Let $x_{k,n}|n \in \mathcal{N} = \{1, \ldots, N\}$ denote the slot configuration at time n of BS k, $x_{k,n}$ can take one of the three values $\{-1, 0, 1\}$, which represent UL, flexible, and DL configurations, respectively. The combined variable for the slot configurations of all BSs is defined as $\mathbf{x} = \{x_{k,n}|k \in \mathcal{K}, n \in \mathcal{N}\}$. The UL and DL signal to interference plus noise ratios at time n of UE u considering BS k can be defined as in (2) and (3), respectively:

$$\Upsilon^{UL}_{u,k,n} = \frac{\mathbf{1}_{(x_{k,n}=-1)} \kappa_{u,k} \eta_{u,k,s} g^s_{u,k} p^{UL}_{u,k,s}}{\sum\limits_{v \in \mathcal{U} \setminus \{u\}} \left[\left(\mathbf{1}_{(x_{l_v,n}=-1)} I_{v,k,s} + \mathbf{1}_{(x_{l_v,n}=1)} I_{l_v,k,s} \right] + \sigma^2 \right.}, \tag{2}$$

and

$$\Upsilon^{DL}_{u,k,n} = \frac{\mathbf{1}_{(x_{k,n}=1)} \kappa_{u,k} \eta_{u,k,s} g^s_{u,k} p^{DL}_{u,k,s}}{\sum\limits_{v \in \mathcal{U} \setminus \{u\}} \left[\left(\mathbf{1}_{(x_{l_v,n}=-1)} I_{v,u,s} + \mathbf{1}_{(x_{l_v,n}=1)} I_{l_v,u,s} \right] + \sigma^2 \right.}, \tag{3}$$

where $\mathbf{1}_{(*)}$ is a binary function such that $\mathbf{1}_{(*)} = 1$ if $*$ is true, otherwise $\mathbf{1}_{(*)} = 0$, l_v is the BS serving UE v, $I_{v,k,s}$ and $I_{v,u,s}$ denote the interferences from UE v to BS k and UE u, respectively, $I_{l_v,k,s}$ and $I_{l_v,u,s}$ denote the interferences from BS l_v to BS k and UE u, respectively, and σ^2 is the additive white Gaussian noise power spectral density. The interferences are computed as follows:

$$I_{v,k,s} = \kappa_{v,l_v} \eta_{v,l_v,s} p^{UL}_{v,l_v,s} g^s_{v,k}, \tag{4}$$

$$I_{l_v,k,s} = \kappa_{v,l_v} \eta_{v,l_v,s} p^{DL}_{v,l_v,s} g^s_{k,l_v}, \tag{5}$$

$$I_{v,u,s} = \kappa_{v,l_v} \eta_{v,l_v,s} p^{UL}_{v,l_v,s} g^s_{u,v}, \tag{6}$$

and

$$I_{l_v,u,s} = \kappa_{v,l_v} \eta_{v,l_v,s} p^{DL}_{v,l_v,s} g^s_{u,l_v}, \tag{7}$$

where g^s_{k,l_v} and $g^s_{u,v}$ indicate the direct channel power gain from BS l_v to BS k and from UE v to UE u, respectively. The UL and DL transmission rate at time n of UE u considering BS k can be computed as follows:

$$R^{UL}_{u,k,n} = W \times \log_2 \left(1 + \Upsilon^{UL}_{u,k,n} \right), \tag{8}$$

and

$$R^{DL}_{u,k,n} = W \times \log_2 \left(1 + \Upsilon^{DL}_{u,k,n} \right), \tag{9}$$

where W is the sub-channel bandwidth.

We aim to maximize the achievable sum rate of the network by optimizing the slot configurations of all BSs. The corresponding optimization problem is mathematically formulated as follows:

$$\max_{\mathbf{x}} \sum_{u=1}^{U} \sum_{k=0}^{K} \sum_{n=1}^{N} \left(R^{UL}_{u,k,n} + R^{DL}_{u,k,n} \right), \tag{10}$$

$$\text{s.t. } x_{k,n} \in \{-1, 0, 1\}, \forall k \in \mathcal{K}; \forall n \in \mathcal{N}, \tag{10a}$$

where $\mathbf{x} = \{x_{k,n}|k \in \mathcal{K}, n \in \mathcal{N}\}$ is the target variable as slot configurations of all BSs. To be more specific, the slot configuration of each BS does not only affect its own achievable downlink and uplink rate but also make interference, especially severe cross-link interference, to its adjacent BSs and UEs. Therefore, the sum rate objective function in formula (10) is nonlinear in terms of the integer variable $\mathbf{x}$ [25]. According to [26] and [27], the formulated problem is a nonlinear combinatorial integer programming optimization problem, which is nonconvex and a subclass of NP-hard.

III. Proposed Solution

To simplify the formulated multicell problem, we assume that users maintain their locations for a duration without any handovers, and the scheduling decisions for UL and DL are governed by the generated packet. Notably, the upcoming 6G networks can provide geographical information of the transmission links based on topology captured from high altitudes using UAVs. In practice, it is more convenient and cost-effective to collect geographical location information than the explicit CSI. Also, the geographical location information can be efficiently exploited to estimate the channel environment, which is typically characterized by signal and interference strength. Specifically, the geographical information can include various factors affecting signal and interference strength, such as the link distance, blockage, line-of-sight conditions, and

beamforming trajectory. Therefore, leveraging the geographical location information can derive a globally near-optimal solution for the formulated problem.

A. Traffic Demand Density Grid Input

Assuming that the location of users does not significantly change during a transmission period T, the UE–BS association is maintained during the transmission period, and no handover occurs across BSs. Notably, the coverage area of the BSs may overlap, and the coverage area for each BS may include both its own subscribed users and the subscribed users of adjacent BSs. Therefore, the slot configuration and corresponding scheduling decisions made by a BS to serve DL or UL requests not only cater to its subscribed users but also induce inter-cell interferences for the subscribed users of adjacent BSs. This interference effect can be learnt from geographical location information by using convolution neural networks.

The geographical location information can be represented in a matrix form as $\{\{l_k|k = 0, \ldots, K\}, \{l_u|u = 1, \ldots, U\}\}$, where $l_k \in \mathbb{R}^2$ and $l_u \in \mathbb{R}^2$ are the locations of BSs k and UEs u, respectively. For each BS, the location information of the users located in its coverage area are quantized based on the original continuous location information. Without loss of generality, we assume that the coverage area of each BS k is a circle of radius r_k can be divided into multiple Z smaller circles indexed as $1, \ldots, Z$ in increasing order, i.e., a smaller circle is assigned a larger index. Users located in smaller circles experience higher received signal power than those located in larger circles. We quantize the relative signal strength of the users located in each circle $z \in \{1, \ldots, Z\}$ by its index z. Furthermore, we construct grid matrices of each network cell, in which the BS is located at the origin, and the subscribed and non-subscribed users are distributed in the coverage area. The geographical location information of each network cell is represented using the tuple $\langle (l_1^U, \ldots, l_{\Omega_k}^U), (l_1^K, \ldots, l_{\Xi_k}^K), (l_1^{NU}, \ldots, l_{\Phi_k})^{NU} \rangle$, where Ω_k is the number of subscribed users of BS k, Ξ_k is the number of adjacent BSs of BS k, and Φ_k is the number of non-subscribed users of BS k that are located in the coverage area of BS k. Moreover, we quantize the traffic request of the users located in the coverage area of each BS. For example, DL traffic requests are quantized into low, normal, and high levels. Similarly, UL traffic requests are quantized into low, normal, and high levels. The objective of such quantization is to ensure fair QoS for users in terms of DL and UL services. According to 5G and 6G mobile communication network specifications [28], [29], the peak data rate per device is at least 10-fold over that of LTE and 5G networks, respectively, corresponding to 10 Gbps and 1 Tbps. Additionally, assuming each BS serves traffic requests of only one device, we define the low, normal, and high traffic request states according to Table II below.

We decompose the geographical location information of each network cell into four grid matrices, as illustrated in Fig. 3, considering the traffic requests of the subscribed users and non-subscribed users. This decomposition aims to extract both the positive (performance gain in terms of DL and UL rate achieved by serving subscribed users) and negative effects (interferences arising for the non-subscribed users and adjacent BSs). The first density grid matrix captures DL requests of the subscribed users and is constructed by counting the quantized DL requests at each subscribed-user node. The second density grid matrix pertains to UL requests of the subscribed users and is constructed by counting the quantized UL requests at each subscribed-user node. The third grid matrix pertains to the DL requests at the non-subscribed-user nodes inside the network cell coverage area. The fourth grid matrix pertains to the UL requests toward adjacent BSs of the network cell. Assuming that the network coverage area is in a rectangular form and located in a specified coordinate system, each grid matrix is constructed from multiple grid-cells with the unit size of 1 m × 1 m, divided by rows and columns which are determined based on the origin's location, directions of $x-$ and $y-$axis, and the unit size. Moreover, the density grid values are calculated as the sum product of the quantized traffic demand requests with the index of the largest coverage circle containing the corresponding subscribed and non-subscribed users. For example, the traffic demand quantization values are set as 1, 2, and 3 to indicate low, normal, and high traffic demand, respectively. Thus, for a BS k and a user located in a coverage circle with index z having a high UL request, the corresponding UL density grid value is calculated as $3 * z$.

TABLE II
TRAFFIC REQUEST STATES

Traffic request state	5G network	6G network
Low	< 10 Mbps	$< 10^2$ Mbps
Normal	$10 - 10^3$ Mbps	$10^2 - 10^5$ Mbps
High	$10^3 - 10^4$ Mbps	$10^5 - 10^6$ Mbps

B. Efficient Feature Extraction

Using the traffic demand density grid matrices as the input, we extract the features pertaining to the DL and UL service gains and harms using convolution neural networks (CNNs). Spatial convolution filters are used to screen the traffic demand density grid matrices and compute the corresponding service gains and harms. The weights of the convolution filters are optimized during the training process. The obtained features are independently computed in parallel on the traffic demand density grid matrices to minimize the training time. For each feature extraction, the convolution filter is represented as a square matrix with a predefined size and trainable weighting coefficients. As in the research of [22], the values of the convolution filters are determined based on the channel coefficients received at the corresponding position relative to the serving BS. Throughout the training process, the weights of the spatial convolution filters are adjusted to achieve a circular symmetric pattern with radial decay.

Following the processing with spatial convolution filters, we obtain grid matrices with the traffic demand gains (for the subscribed users) and harms (for the non-subscribed users and adjacent BSs). Owing to the large size and numerous variations in these matrices, significant computation capabilities are required for further classification. To address this problem,

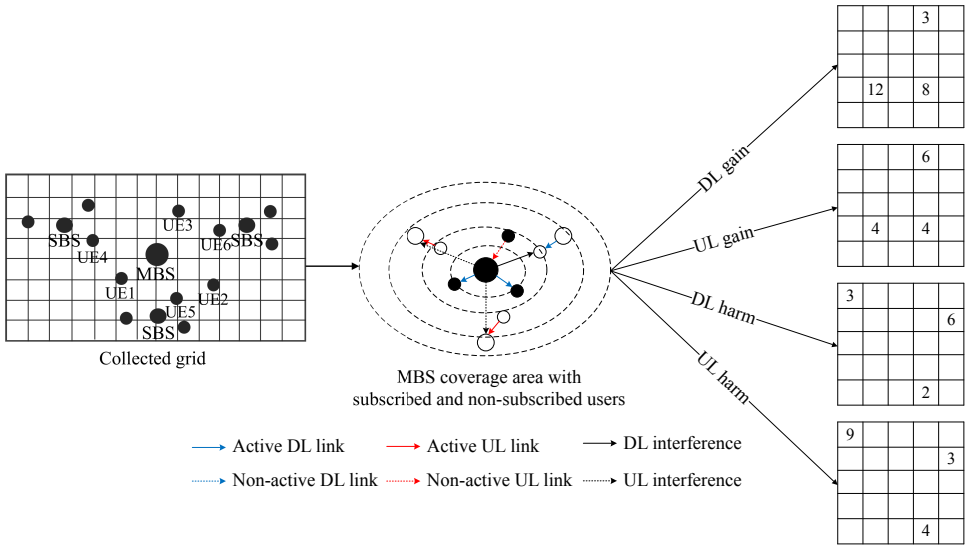


Fig. 3. Traffic demand density grid input for the macro base station: UE 1 (DL request high, UL request low), UE 2 (DL request normal, UL request low), UE 3 (DL request low, UL request normal), UE 4 (DL request low, UL request high), UE 5 (DL request low, UL request normal), and UE 6 (DL request normal, UL request low)

.

we use sparse convolution blocks, including speed–accuracy balancing (SAB) and deep feature processing (DFP) blocks [30], to effectively extract the traffic demand gain and harm features. Both the SAB and DFP blocks are equipped with depthwise and pointwise convolution filters to promote feature extraction while reducing the computation cost for resource-constrained devices. The SAB block helps balance the speed and accuracy. However, it may lead to suboptimal performance when used in deep layers with numerous free parameters. The DFP block is employed to extract discriminative features more effectively in deep layers. To prevent inter-block feature loss, skip connections are employed to enable feature sharing between the subblocks of the SAB and DFP blocks.

C. Reinforcement Learning Task Formulation

This work considers classification of the near-optimal radio configuration output for each collected geographical location map of the network coverage area. To do this, the dedicated features regarding the achievable data rate gain and inter-cell interference harm are extracted first. Subsequently, we develop a RL-based deep dueling algorithm to promote the classification job, in which the extracted features are the input and the near-optimal radio configuration output is learnable as the expected class of the collected geographical map. In this subsection, we present a novel RL-based model that deploys classification for mapping a correct slot format with the derived features of DL and UL service gains and harms. Notably, the problem formulated in (10) is challenging to solve in a straightforward manner. Therefore, we transform this problem into a long-term determination of slot configuration variables that maximize the sum rate for all users. We consider the time to be slotted such that a specific duration $\mathcal{T}$ is divided into T time slots indexed by $t \in \{1, \ldots, T\}$, resulting in $\lfloor T/N \rfloor + 1$ consecutive slot configurations of each BS. The corresponding problem can be formulated as follows:

$$\max_{\mathbf{x}} \mathbb{E}\left[\sum_{t=1}^{T} \gamma^{t-1} \sum_{u=1}^{U} \sum_{k=0}^{K} \left(R^{UL}_{u,k,n} + R^{DL}_{u,k,n}\right)_{n=t\%N}\right], \quad (11)$$

$$\text{s.t. } x_{k,n} \in \{-1, 0, 1\}, \forall k \in \mathcal{K}; \forall n \in \mathcal{N}, \quad (11a)$$

where γ is a discounting factor that ensures γ^{t-1} approaches 0 when t is large. $R^{UL}_{u,k,n}$ and $R^{DL}_{u,k,n}$ are updated corresponding to time t. Specifically, if $t > N$, n is reset by $n = t\%N$ and a new slot configuration is applied for each BS with updated network information, which is assumed to be maintained no change periodically during each N time slots.

An RL-based training process is developed to solve the problem defined in (11). Specifically, an RL agent is employed at a high altitude with a UAV to control the optimization of the slot format variable $\mathbf{x}$ of all BSs, with the goal of maximizing the achievable sum rate in the long term. In RL terminology, the RL agent iteratively interacts with the environment by observing the state of the environment denoted by $S(t)$, providing instantaneous action $a(t)$, observing changes in the environment evolved state $S(t+1)$, and computing its step reward $U(t)$. The objective of this interaction process is to learn a policy of action selection that maximizes the achievable reward in the long term. By applying this concept to solve the problem formulated as (11), we define the following RL terms.

1) State space: The system state is determined based on the spatial geographical information of each network cell, characterized by the obtained traffic demand density feature matrices representing the DL and UL service gains and harms. In general, changing the service from DL to UL or vice versa requires a time gap with no service, which can degrade the system performance in terms of data rate and delay. Therefore, we also consider the previous selected action to formulate the system state. The overall system state is defined as follows:

$$S(t) \leftarrow [\{\chi_k(\varphi_{\text{DL}}, \varphi_{\text{UL}}, \psi_{\text{DL}}, \psi_{\text{UL}}) | \forall k \in \mathcal{K}\}, a(t-1)], \quad (12)$$

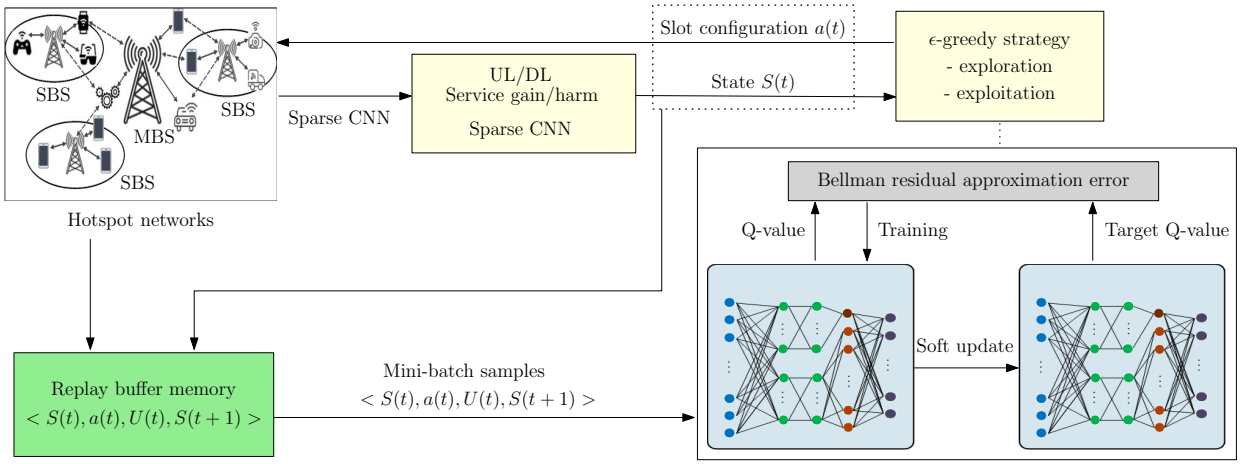


Fig. 4. Architecture of the proposed deep dueling algorithm.

where $\chi_k(\varphi_{\mathrm{DL}}, \varphi_{\mathrm{UL}}, \psi_{\mathrm{DL}}, \psi_{\mathrm{UL}})$ denotes the output traffic demand density features corresponding to network cell k.

2) Action space: The action includes the slot format of each BS, which can take three values, i.e., $x_k(t) \in \{-1, 0, 1\}, \forall k \in \mathcal{K}$, corresponding to UL, flexible, and DL service states. Notably, in flexible slot format, neither UL nor DL services are provided, resulting in no service gain or interference generation.

3) Reward function: The sum rate of all users is considered the reward for the RL agent, defined as follows:

$$U(t) = \sum_{u=1}^{U} \sum_{k=0}^{K} \left(R^{UL}_{u,k,n}(t) + R^{DL}_{u,k,n}(t) \right)_{n=t\%N}. \tag{13}$$

4) State transition: After each iteration, the action is known to the agent and remains as an element of the next state. Additionally, based on the evolved graphical location information and traffic demand requests of all users, we use convolution filters to extract the evolved traffic demand density features for processing in the next training iteration.

D. Deep Dueling Algorithm

We employ the deep dueling-based learning framework, which is an off-policy maximum entropy algorithm that offers both stability and sample-efficient learning. The concept of dueling neural network is based on Generative Adversarial Network, which takes two neural networks as the simplified mathematical models of the human brain. The dueling neural network structure significantly improves the learning by allowing the networks to differentiate actions from one another during the learning process [31]. Adopting dueling neural network to reinforcement learning with Deep Q-learning algorithm, two streams of estimating the state-value and advantages for each action are separated. This algorithm is especially suitable for high-dimensional tasks with complex state and action spaces and can effectively address overestimation problems in large-scale systems. The core concept of the deep dueling algorithm is to separately estimate the values of the states and advantages of actions using two streams of neural network layers, and combine them at the final output layer. For many samples, in certain states, the action selection has no effect on the outcome. In such circumstances, the deep dueling algorithm can avoid the unnecessary estimation of the action value and focus more on estimating the state value to improve the convergence and stability. For a given stochastic policy π, the value functions of the state–action pair (S, a) and state S can be expressed as follows:

$$Q_\pi(S, a) = \mathbb{E}\left[U(t) | S(t) = S, a(t) = a, \pi\right], \tag{14}$$

and

$$V_\pi(S) = \mathbb{E}\left[Q_\pi(S, a)\right]. \tag{15}$$

Here, the Q-value function $Q(S, a)$ is the state–action pair function that evaluates the value of selecting action a in state S. Moreover, state value function $V(S)$ estimates the effectiveness of the action in a specific state S. Assuming action $\mathbf{a}$ is taken under policy π with state S, its advantage value function is expressed as

$$G_\pi(S, \mathbf{a}) = Q_\pi(S, \mathbf{a}) - V_\pi(S). \tag{16}$$

This function obtains a relative value of each action by decoupling the state value from the state–action pair Q-value function. Accordingly, $\mathbb{E}\left[G(S, \mathbf{a})\right] = 0$, $a^* = \mathrm{argmax}_a Q(S, a)$, and $Q(S, a^*) = V(S)$.

The value functions of the state S and corresponding action a can be evaluated using a dueling neural network, which has two output streams of the fully connected layers: one for $V(S)$ and the other for $G(S, a)$. Thereafter, the Q-value function can be estimated when combining outputs of the two streams as

$$Q(S, a) = V(S) + G(S, a). \tag{17}$$

To specify the unique state and action value functions of a given Q-value, we can implement the dueling neural network that combines the state and action value outputs following a specific mapping as

$$Q(S, a) = V(S) + \left(G(S, a) - \max_a G(S, a) \right). \tag{18}$$

In state S, given the near-optimal action $a^* = \mathrm{argmax}\ Q(S, a) = \mathrm{argmax}\ G(S, a)$, we obtain $Q(S, a^*) = V(S)$. Therefore, the Q-value function can

be simplified by replacing the max operator with an average operator as follows:

$$Q(S,a) = V(S) + \left(G(S,a) = \frac{1}{\Pi} \sum_{a} G(S,a) \right), \tag{19}$$

where Π indicates the size of the action space. Accordingly, we directly obtain the Q-value by estimating the value functions $V(S)$ and $G(S,a)$ of state S and action a using a deep dueling neural network, without any extra modifications.

The architecture of the proposed deep dueling algorithm is shown in Fig. 4. There are four main components: (i) the small-cell network environment, (ii) the sparse convolution filters, (iii) the deep dueling neural network model, and (iv) a replay buffer memory. The observed current density grid matrices of all BSs are processed to extract the UL and DL service gains and harms features using spatial and sparse convolution filters. Based on the extracted features, state $S(t)$ is formulated and a slot configuration action $a(t)$ is selected using the ϵ-greedy strategy. Subsequently, the small-cell network environment is returned with an evolved state $S(t+1)$ and step reward $U(t)$. Tuple $\langle S(t), a(t), U(t), S(t+1) \rangle$ is saved in the replay buffer memory as a historical experience. Notably, the store capacity of the replay buffer memory is limited, and in the absence of any more space, the least recent experiences are replaced by new ones. To train the near-optimal slot configuration policy, random experiences are sampled from the replay memory and processed using the deep dueling neural networks. The learning process iterates over several training steps to minimize the Bellman residual approximation error of the Q-values by using the gradient descent technique.

$$\mathcal{L}(\theta) = \mathbb{E}_{t=1}^{\Psi} \left[(y(t) - Q(S(t), a(t); \theta))^2 \right], \tag{20}$$

where Ψ is the batch size; $Q(S(t), a(t); \theta)$ denotes the output of the main neural network Q for the state–action pair $(S(t), a(t))$; and $y(t) = U(t) + \gamma Q(S(t+1), a(t); \theta)$, $t \in \{1, \ldots, \Psi\}$, is a target Q-value with $a(t) = \text{argmax}_a Q_t(S(t+1), a; \theta)$. The training process is terminated once the updated amount for θ becomes significantly small.

IV. Performance Evaluation

A. Simulation Settings

The parameter settings for networks are summarized in Table III. We construct a heterogeneous network consisting of a one-tier macro-cell with outdoor small cells. The SBSs are uniformly distributed, whereas the UEs are randomly distributed in the MBS coverage area. Here, each UE selects its serving cell based on the maximum RSRP in the DL, with 20 dB range expansion bias for each SBS. We use Pytorch to build the training framework, in which deep neural networks are trained using the generated network data. The parameter settings for deep learning are selected based on common settings in [31]. The deep dueling mechanism takes two streams of neural networks, each of them consists of a hidden layer with size of 128, input and output layers. The mini-batch size is 64. The replay buffer size is 50,000. Discounting factor is valued 0.7. A learning rate of 0.9 is applied. Moreover, to avoid timing issues associated with data collection for supervised learning, the neural network model is trained offline in a model-free DRL framework, using generated data of the interaction between an RL agent and the network environment.

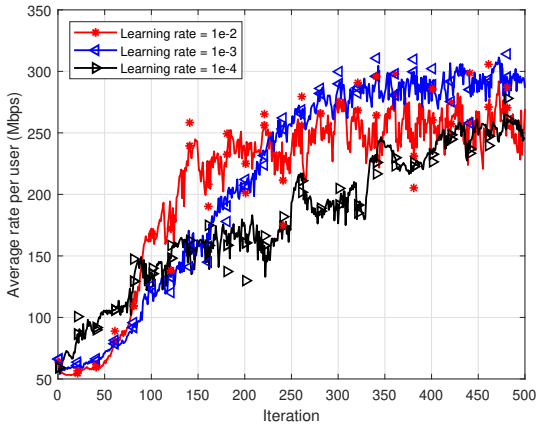


Fig. 5. Achievable rate per user with different learning rate for training, $U = 10$ and $K = 4$.

Extensive simulations are conducted to evaluate the performance of the proposed duplexing framework under different parameter settings. The results are compared to those achieved in existing duplexing schemes:

- *Optimum scheme with ES algorithm*, which involves a central controller that collects all the information regarding the CSI and traffic demand requests per slot in the network. Based on this information, the algorithm determines the optimal slot configurations for all BSs.
- *Greedy approach*, in which each BS selects a slot format based on the traffic demand requests of only its subscribed users. Specifically, by comparing the total traffic demand requests in the UL and DL transmissions of all subscribed users, the BS decides to serve either the UL or DL requests based on the transmission with higher demand.
- *Random approach*, in which each BS randomly selects its radio slot configuration.
- *Centralized D-TDD scheme [11]*, in which all BSs and UEs are iteratively select their UL or DL transmission until convergence is achieved for the best slot configurations of all nodes. The iteration is based on a binary matrix that models the relative information of the desired and undesired links for each node.

B. Results

Fig. 5 illustrates the convergence of the proposed training process in terms of the achievable rate per user for different learning rates, specifically, $1e-2$, $1e-3$, and $1e-4$. As shown in the figure, convergence is achieved after approximately 300 iterations for the learning rates of $1e-2$ and $1e-3$. For the smaller learning rate of $1e-4$, the achievable rate exhibits significant fluctuations. Furthermore, the achievable rate is higher with a learning rate of $1e-3$ than those with the learning rates $1e-2$ and $1e-4$. This can be attributed

TABLE III
SYSTEM-LEVEL SIMULATION PARAMETERS FOR NETWORKS

Parameter	Value	Parameter	Value
Deployment	3GPP case 1 set (MBS-to-MBS ISD 1 km) (CF 2 GHz, Speed 3 km/h)	Traffic model	FTP Model 3
Number of MBS Sectors	3	Channel bandwidth	10 MHz (single channel)
Number of SBSs	12	Fading model	Rayleigh
Number of UEs	50	Path-Loss: MBS and MBS	$100.7 + 23.5 \log_{10}(d(\text{km}))$[dB]
Scheduling	Proportional fair	Path-Loss: MBS and SBS	$125.2 + 36.3 \log_{10}(d(\text{km}))$[dB]
MBS Height	32 m	Path-Loss: MBS and UE	$128.1 + 37.6 \log_{10}(d(\text{km}))$[dB]
SBS Height	5 m	Path-Loss: SBS and SBS/UE	$140.7 + 36.7 \log_{10}(d(\text{km}))$[dB]
UE Height	1.5 m	Path-Loss: UE and UE	$140.7 + 36.7 \log_{10}(d(\text{km}))$[dB]
Noise Figure	9 dB	Range Expansion	20 dB
MBS Transmit Power	46 dBm	Min. Dist. MBS and SBS	$\geq$ 35 m
SBS Transmit Power	30 dBm	Min. Dist. MBS and UE	$\geq$ 35 m
UE Transmit power	20 dBm	Min. Dist. SBS and UE	$\geq$ 10 m
SBS Antenna Pattern	0 dB (omnidirectional)	Min. Dist. SBS and SBS	$\geq$ 30 m
MBS Antenna Pattern	3D pattern, in Table A.2.1.1-2 in [32]	Placing SBSs and UEs	Configuration 1 in Table A.2.1.1.2-3 in [32]

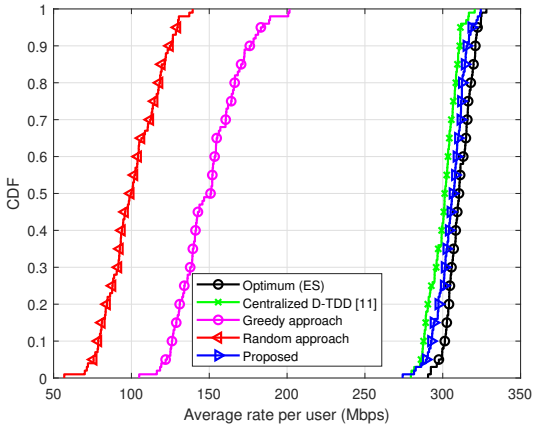


Fig. 6. Cumulative distribution functions of the achievable data rates.

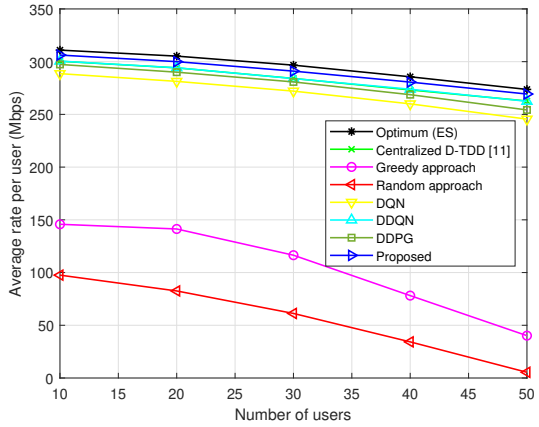


Fig. 7. Achievable rates with different numbers of users.

to the non-linear relationship of the achievable rate with the slot configurations and geographical location information. Furthermore, with a learning rate of $1e-3$, the fluctuation of the achievable rate is considerably smaller, indicating its higher potential for providing the optimal performance. Based on these observations, we select the learning rate of $1e-3$ that gives stable and high performance of the achievable rate.

Fig. 6 shows the cumulative distributions of the achievable data rates for the proposed scheme, random approach, greedy approach, state-of-the-art centralized D-TDD scheme [11], and optimal scheme with ES algorithm. The achievable rate for the proposed scheme is significantly higher than those of the random and greedy approaches. Moreover, the proposed scheme outperforms the centralized D-TDD scheme, while its performance is close to that of the optimal scheme with the ES algorithm.

Fig. 7 compares the achievable rates of the proposed scheme and other schemes for different numbers of users. From the figure, the achievable rate per user decreases as the number of users increases. Moreover, the proposed scheme exhibits greater achievable rates compared to the random and greedy approaches and centralized D-TDD scheme [11]. Especially, the rate performance of proposed scheme is comparable with that of the optimum scheme using ES algorithm.

Fig. 8 compares the achievable rates of the proposed scheme and other schemes for different numbers of small cells. The achievable rate per user increases as the number of small cells increases. The proposed scheme significantly outperforms the random and greedy approaches in terms of average rate per user. It also has slightly higher rate compared to that achieved in the centralized D-TDD [11] and achieves a performance comparable with that of the optimum scheme based on the ES algorithm.

The superiority of the proposed scheme over conventional deep learning frameworks such as deep Q-network (DQN), deep double Q-network (DDQN), and deep double policy gradient (DDPG) are reflected in Figs. 7 and 8. Specifically, for various settings of number of users and small cells, the proposed scheme improves the average rate per user by up to 6% over that achieved using DQN, DDQN, and DDPG frameworks.

Fig. 9 compares the achievable DL and UL rate of the MBS and SBS users. In Fig. (a), for the MBS users, the DL achievable rate with the proposed scheme is higher than that

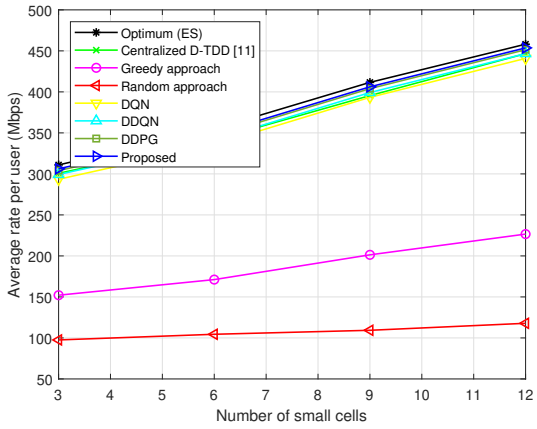


Fig. 8. Achievable rate with different numbers of small cells.

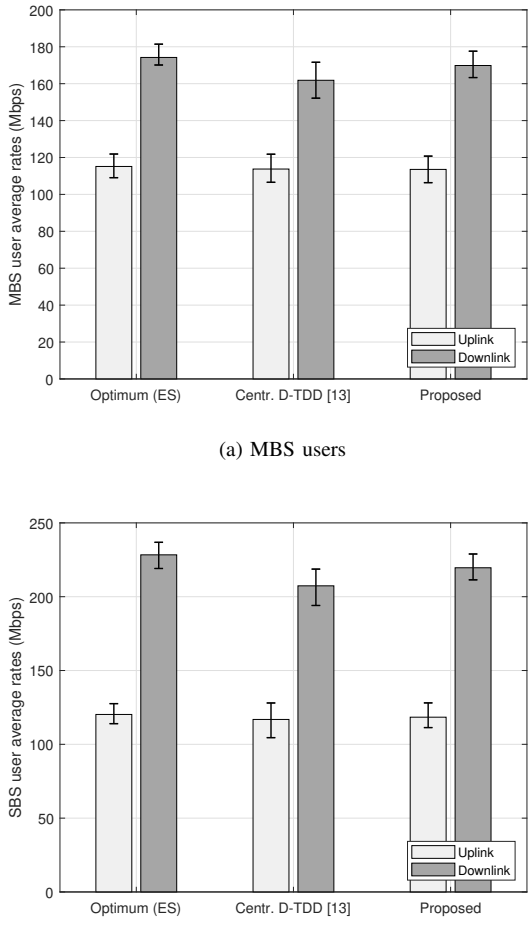
with the centralized D-TDD scheme [11], and similar to that with the optimum scheme. The UL rate of the MBS users are approximately same with all three schemes. In Fig. (b), for the SBS users, the DL and UL achievable rate for the comparing schemes have the same tendency as in the case of MBS users above, but the achievable rate of the SBS users are higher than those of the MBS users with all three schemes.

C. Optimality Analysis

According to [33], a Q-learning-based algorithm can achieve an optimal solution when its training process is sufficiently implemented with stationary state transition probabilities and bounded step rewards. In this work, the conditions can be satisfied. First, the state transition of the DL and UL traffic demand requests at each time slot is stationary when we consider several specified industrial network systems such as factorial and smart-farm networks [19]. Second, the step reward $U(t)$ is bounded because the achievable data rate of all users cannot exceed a threshold value depending on the transmission power of the BSs and UEs and channel attenuation. Therefore, the proposed scheme that adopts dueling neural network with Deep Q-learning algorithm can achieve optimal solution of the proposed RL problem, which is near-optimal slot configurations of the BSs.

V. Conclusions

This paper develops a near-optimal and low-overhead D-TDD framework that can support various traffic demands in UAV-assisted industrial hotspot networks. First, the geographical location and load information for each network cell is represented using traffic demand density grid matrices. Second, taking the density grid matrices as the input, spatial convolution filters and sparse convolution blocks are employed to extract UL and DL service gains and harms while reducing the computation cost for further classification. Third, we develop novel deep dueling neural networks, which can efficiently learn the optimal slot configurations for all BSs. The simulation results demonstrated that the proposed D-TDD framework stably converges, and it achieves near-optimal data rate for all users as the traffic demand dynamically evolves.



(a) MBS users

(b) SBS users

Fig. 9. Downlink and uplink achievable rate of MBS and SBS users.

References

[1] J. M. B. da Silva, G. Wikström, R. K. Mungara, and C. Fischione, "Full Duplex and Dynamic TDD: Pushing the Limits of Spectrum Reuse in Multi-Cell Communications," *IEEE Wireless Commun.*, vol. 28, no. 1, pp. 44–50, 2021.

[2] H. Iimori, J. Huschke, and J. Vieira, "Radio Unit Configuration for Dynamic Time Division Duplex in Distributed MIMO Systems," in *IEEE Global Communications Conference*. IEEE, 2023, pp. 2542–2547.

[3] A. Chowdhury and C. R. Murthy, "Half-Duplex APs with Dynamic TDD vs. Full-Duplex APs in Cell-Free Systems," *IEEE Trans. Commun.*, 2024.

[4] M. N. Alam, R. Jäntti, and Z. Uykan, "Hopfield Neural Network Based Uplink/Downlink Transmission Order Optimization for Dynamic Indoor TDD Femtocells," *IEEE Access*, vol. 11, pp. 85 414–85 425, 2023.

[5] H. Kim, J. Kim, and D. Hong, "Dynamic TDD Systems for 5G and Beyond: A Survey of Cross-Link Interference Mitigation," *IEEE Commun. Surv. Tut.*, vol. 22, no. 4, pp. 2315–2348, 2020.

[6] F. S. Samidi *et al.*, "5G New Radio: Dynamic Time Division Duplex Radio Resource Management Approaches," *IEEE Access*, vol. 9, pp. 113 850–113 865, 2021.

[7] Y. Zhang *et al.*, "Packet-Level Throughput Analysis and Energy Efficiency Optimization for UAV-Assisted IAB Heterogeneous Cellular Networks," *IEEE Trans. Veh. Technol.*, vol. 22, no. 7, pp. 9511–9526, 2023.

[8] Z. Su *et al.*, "Energy-Efficiency Optimization for D2D Communications Underlaying UAV-Assisted Industrial IoT Networks With SWIPT," *IEEE Internet Things J.*, vol. 10, no. 3, pp. 1990–2002, 2023.

[9] S. Khisa and S. Moh, "Priority-Aware Fast MAC Protocol for UAV-Assisted Industrial IoT Systems," *IEEE Access*, vol. 9, pp. 57 089–57 106, 2021.

[10] D. K. Jain *et al.*, "Enabling Unmanned Aerial Vehicle Borne Secure Communication With Classification Framework for Industry 5.0," *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5477–5484, 2022.

[11] M. M. Razlighi, N. Zlatanov, S. R. Pokhrel, and P. Popovski, "Optimal Centralized Dynamic-Time-Division-Duplex," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 28–39, 2021.

[12] M. Ghermezcheshmeh, M. M. Razlighi, V. Shah-Mansouri, and N. Zlatanov, "Centralized Dynamic-Time Division Duplex Utilizing Interference Alignment," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6852–6866, 2021.

[13] K. Pedersen *et al.*, "Advancements in 5G New Radio TDD Cross Link Interference Mitigation," *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 106–112, 2021.

[14] G. C. Nwalozie, K. Ardah, and M. Haardt, "Reflection Design Methods for Reconfigurable Intelligent Surfaces-Aided Dynamic TDD Systems," in *IEEE 12th Sensor Array and Multichannel Signal Processing Workshop*. IEEE, 2022, pp. 36–40.

[15] J. Song *et al.*, "QoE-Driven Distributed Resource Optimization for Mixed Reality in Dynamic TDD Systems," *IEEE Trans. Commun.*, vol. 70, no. 11, pp. 7294–7306, 2022.

[16] A. Chowdhury, C. R. Murthy, and R. Chopra, "Dynamic TDD Enabled Distributed Antenna Array Massive MIMO System," in *IEEE 12th Sensor Array Multichannel Signal Processing*. IEEE, 2022, pp. 131–135.

[17] E. de O. Cavalcante, G. Fodor, Y. C. Silva, and W. C. Freitas, "Bidirectional Sum-Power Minimization Beamforming in Dynamic TDD MIMO Networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 10, pp. 9988–10 002, 2019.

[18] J.-S. Tan *et al.*, "Lightweight Machine Learning for Digital Cross-Link Interference Cancellation With RF Chain Characteristics in Flexible Duplex MIMO Systems," *IEEE Wireless Commun. Lett.*, vol. 12, no. 7, pp. 1269–1273, 2023.

[19] V. D. Tuong, N.-N. Dao, W. Noh, and S. Cho, "Deep Reinforcement Learning-Based Hierarchical Time Division Duplexing Control for Dense Wireless and Mobile Networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7135–7150, 2021.

[20] M. Bagaa, K. Boutiba, and A. Ksentini, "On using Deep Reinforcement Learning to dynamically derive 5G New Radio TDD pattern," in *IEEE Global Communications Conference*. IEEE, 2021, pp. 1–6.

[21] F. Tang, Y. Zhou, and N. Kato, "Deep Reinforcement Learning for Dynamic Uplink/Downlink Resource Allocation in High Mobility 5G HetNet," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 12, pp. 2773–2782, 2020.

[22] W. Cui, K. Shen, and W. Yu, "Spatial Deep Learning for Wireless Scheduling," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1248–1261, 2019.

[23] "3GPP TS 38.211: NR; Physical channels and modulation (Version 16.2. 0 Release 16)," 3GPP, European Telecommunications Standards Institute, 2022.

[24] "3GPP TS 38.300: NR; NR and NG-RAN Overall description," 3GPP, European Telecommunications Standards Institute, 2017.

[25] L. Liberti, "Undecidability and hardness in mixed-integer nonlinear programming," *RAIRO-Operations Research*, vol. 53, no. 1, pp. 81–109, 2019.

[26] D. Du, P. M. Pardalos, X. Hu, and W. Wu, *Introduction to Combinatorial Optimization*. Springer, 2022.

[27] L. A. Wolsey, *Integer Programming*. John Wiley & Sons, 2020.

[28] A. A. Ateya, A. Muthanna, M. Makolkina, and A. Koucheryavy, "Study of 5G Services Standardization: Specifications and Requirements," in *10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*. IEEE, 2018, pp. 1–6.

[29] M. El-Moghazi and J. Whalley, "The ITU IMT-2020 Standardization: Lessons from 5G and Future Perspectives for 6G," *Journal of Information Policy*, vol. 12, pp. 281–320, 2022.

[30] V. D. Tuong, W. Noh, and S. Cho, "Sparse CNN and Deep Reinforcement Learning-Based D2D Scheduling in UAV-Assisted Industrial IoT Networks," *IEEE Trans. Ind. Informat.*, vol. 20, no. 1, pp. 213–223, 2024.

[31] Z. Wang *et al.*, "Dueling Network Architectures for Deep Reinforcement Learning," in *International Conference on Machine Learning*. PMLR, 2016, pp. 1995–2003.

[32] "3GPP TR 36.814: Further advancements for E-UTRA physical layer aspects (Release 9)," EUTR Access, European Telecommunications Standards Institute, 2010.

[33] C. J. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, pp. 279–292, 1992.

Van Dat Tuong received the B.S. degree in mechatronics from Hanoi University of Science and Technology, Hanoi, Vietnam, in 2012, and the M.S. degree in system software, from Chung-Ang University, Seoul, South Korea, in 2021, where he is currently pursuing the Ph.D. degree in big data. From 2012 to 2018, he was a Software Engineer with the Viettel Software Center, Viettel Telecom and the Samsung Vietnam Mobile Center (SVMC), Samsung Electronics Vietnam, Hanoi, Vietnam. He was a recipient of the Global Korea Scholarship sponsored by the Korean Government, from 2018 to 2021. His research interests include machine learning, game theory, convex optimization, and their applications in wireless networking and ubiquitous computing.

Wonjong Noh received the B.S., M.S., and Ph.D. degrees from the Department of Electronics Engineering, Korea University, Seoul, South Korea, in 1998, 2000, and 2005, respectively. From 2005 to 2007, he conducted the postdoctoral research with Purdue University, West Lafayette, IN, USA, and University of California at Irvine, Irvine, CA, USA. From 2008 to 2015, he was a Principal Research Engineer with the Samsung Advanced Institute of Technology (SAIT), Samsung Electronics, South Korea. After that, he worked as an Assistant Professor with the Department of Electronics and Communication Engineering, Gyeonggi University of Science and Technology, South Korea, and since 2019, he has been worked as a Professor with the School of Software, Hallym University, Chuncheon, South Korea. He received the Government Postdoctoral Fellowship from the Ministry of Information and Communication, South Korea, in 2005. He was also a recipient of the Samsung Best Paper Gold Award in 2010, the Samsung Patent Bronze Award in 2011, and the Samsung Technology Award in 2013. His current research interests include fundamental analysis and evaluations on machine learning-based 5G and 6G wireless communications and networks.

Sungrae Cho is a professor with the school of computer sciences and engineering, Chung-Ang University (CAU), Seoul. Prior to joining CAU, he was an assistant professor with the department of computer sciences, Georgia Southern University, Statesboro, GA, USA, from 2003 to 2006, and a senior member of technical staff with the Samsung Advanced Institute of Technology (SAIT), Kiheung, South Korea, in 2003. From 1994 to 1996, he was a research staff member with electronics and telecommunications research institute (ETRI), Daejeon, South Korea. From 2012 to 2013, he held a visiting professorship with the national institute of standards and technology (NIST), Gaithersburg, MD, USA. He received the B.S. and M.S. degrees in electronics engineering from Korea University, Seoul, South Korea, in 1992 and 1994, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2002. He has been a KICS fellow since 2021. He received numerous awards including Haedong Best Researcher of 2022 in Telecommunications and Award of Korean Ministry of Science and ICT in 2021. His current research interests include wireless networking, network intelligence, and network optimization. He has been an editor-in-chief (EIC) of ICT Express (Elsevier) since 2022, a subject editor of IET Electronics Letter since 2018, an executive editor of Wiley Transactions on Emerging Telecommunications Technologies since 2023, and was an area editor of Ad Hoc Networks Journal (Elsevier) from 2012 to 2017. He has served numerous international conferences as a general chair, TPC chair, or an organizing committee chair, such as IEEE ICC, IEEE SECON, IEEE ICCE, ICOIN, ICTC, ICUFN, APCC, TridentCom, and the IEEE MASS.