Federated Deep Reinforcement Learning for THz-Beam Search with Limited CSI

Po-Chun Hsu, Li-Hsiang Shen, Chun-Hung Liu*, and Kai-Ten Feng

Department of Electrical and Computer Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan
Department of Electrical and Computer Engineering, Mississippi State University, MS, USA*
e-mail: {g309513013.c, ktfeng, gp3xu4vu6.cm04g}@nycu.edu.tw and chliu@ece.msstate.edu*

***Abstract*—Terahertz (THz) is a promising technique which provides a ultra-wide frequency band for high requirement of data rates and low-latency services. Nevertheless, the disadvantage of THz networks is the severe propagation attenuation even under short transmission distance. Thus, it becomes compellingly imperative to employ a larger scale of antenna arrays. For a network deployed with multiple THz base station (BS), we propose a federated deep reinforcement learning (FDRL) scheme to coordinate THz beamforming. Each BS conducts deep deterministic policy gradient (DDPG) based DRL to obtain THz beamforming policy with limited channel state information (CSI). While, multi-BSs are controlled by a single federated edge learning (FEL) server to exchange DDPG model with hidden information capable of mitigating inter-cell interference. The simulation results demonstrate the throughput convergence of each BS. We can observe that higher throughput can be achieved with a larger antenna arrays for more THz CSI and hidden neurons of DDPG. Compared to full-model upload of FEL, it requires lower operational overhead using partial-model upload. Moreover, the proposed FDRL outperforms the existing benchmarks using non-FEL and conventional non-learning based on optimization methods.**

***Index Terms*—Terahertz, federated learning, deep reinforcement learning, beamforming, edge computing.**

I. Introduction

For the purpose of meeting the increasing ultra-high data requirement, such as virtual reality (VR), augmented reality (AR) and hologram technologies in the future sixth generation (6G) communication network, the prospect of terahertz communication is considered to be able to provide higher frequency bands ranging from 0.1 to 10 THz. However, compared to conventional millimeter wave (mmWave) at GHz-bands, the most different and important challenges for THz are severe power attenuations, blockages and additional molecular absorption that brings about a much shorter propagation distance and corresponding limited. Therefore, beamforming techniques are utilized to enhance the transmit direction towards the desired receiving user equipment (UE) rather than omni-directional transmissions, which requires a large-scale antenna arrays deployed to obtain high beam gain and spatial diversity. Although there exist potentially rich channel paths in a multi-antenna system, only a single THz path can be utilized with line-of-sight (LoS) condition in most cases. The related beamforming designs in [1], [2] are proposed for the THz system. However, when associating with enormous UEs in different cells, there will still emerge serious interfered beamforming under short transmission distance of THz network which should be coordinated among different BSs appropriately in order to improve overall system performance. In addition, [3]–[5] apply full CSI for beamforming requiring high-complexity channel estimation which is somehow non-implementable and impractical due to numerous and dynamic antenna arrays. However, It is difficult to estimate full CSI around the environment, which results in a dynamic and uncertain wireless communication network. Accordingly, the traditional optimization method cannot perfectly tackle THz system deployed with large-scale antennas under the constantly-changing and limited channels.

Recently, the deep learning techniques are widely applied in the different fields in wireless communication systems. As a prospect, the deep reinforcement learning (DRL) enables the agent, which may be BS or UE to adjust its wireless state and action, i.e., policy output according to the changed environment. Different from model-free reinforcement learning, the deep Q network (DQN) architecture is implemented via deep neural network (DNN) to decide the Q-value instead of the Q-table which benefits the problems with non-countable or near-continuous variables with infinite solution sets. However, when the action is high-dimensional and continuous, it is inefficient and unuseful to apply basic DQN to quantize the decision space. Deep deterministic policy gradient (DDPG) using a two-layered DNN as actor-critic (AC) network is conceived to deal with this problem. Federated edge learning (FEL) has been considerably studied for improving the training progress via learning model exchange with less information uploaded to edge server [6]. The main concept of FEL is to cooperate the local training model in order to acquire a more complete global model, which includes certain hidden information in different BSs or UEs ended with greatly reduced data overhead. In [7]–[9], they have studied in improving the learning speed to deal with the tasks of minimizing computing delay and energy consumption through FEL. In the meanwhile, FEL server integrates the local model to the global model to enhance privacy of every participating client [10]. However, they consider that the BS has full-CSI which is difficult to be estimated in practice, whereas interference generated from different BSs has not been mentioned. Therefore, it becomes promisingly imperative to design an FEL-enabled interference

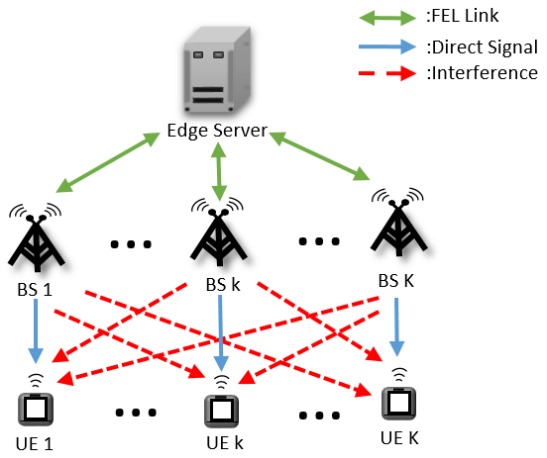


Fig. 1. A schematic diagram of the cellular network with edge computing considered in this paper. In the network, there and there are K base stations, each of them equipped with N_t transmit antennas. A downlink communication scenario is considered and BS k serves UE k equipped with a single antenna in the THz frequency band. All the BSs are connected to an edge server through a high-speed optical link.

mitigation scheme in THz beamforming using state-of-the-art deep learning techniques. The contribution of this paper is summarized as follows.

- We have conceived a federated deep reinforcement learning (FDRL) leveraging the benefits of both FEL and DRL architectures. FEL aims at model exchange from neural networks extracting hidden information of partially estimated CSI, which potentially alleviates interference from other BSs. While, AC-based DDPG is designed to search candidate THz beams to maximize the total throughput performance.
- We characterize the performance in terms of complexity and throughput. We can infer that higher throughput can be achieved with more antennas, exchanged data, and more neurons of FDRL under a compromised computational complexity of deep learning. The proposed FDRL scheme outperforms the baseline using pure deep Q-learning and conventional non-deep learning based beamforming methods.

The rest of the paper is organized as follows: Section II describes the system model and formulates the THz beamforming problem. Section III elaborates our proposed FDRL algorithm for coordinating THz beamforming under a multi-BS network. Section IV shows simulation results, whlist conclusions are drawn in Section V.

II. System Model

In this paper, we consider a cellular network in which each UE is equipped with a single antenna and there are K base stations (BSs) operating in the THz (frequency) band, each of which equipped with N_t antennas. In the downlink, each BS is assumed to adopt different resource blocks to serves different UEs in its cell and thereby there is no intra-cell interference in the network. We also assume that the frequency reuse factor is one in the network so that the K BSs interferes each other in the downlink so that UE k receives interference from the other $K-1$ BSs when it is served by BS k. All the BSs are connected to a edge server through a high-speed optical link where edge computing can be conducted. A schematic diagram of the cellular network with edge computing considered in this paper is shown in Fig. I. In the following, we will first introduce the channel model in the THz band and then specify the signal model transmitted over a THz channel.

A. THz Channel Model

Due to THz signals' nature of extremely high frequency, transmitting them significantly suffers from two serious environmental impairments, i.e., severe attenuation and molecular absorption [11]–[14]. As such, THz signals undergo much higher path loss than mmWave as well as UHF signals. For a THz channel with frequency f, its channel response for transmitting a signal over distance d, denoted by complex vector $\mathbf{h} \in \mathbb{C}^{N_t}$, can be modeled as

$$\mathbf{h} = G\left[1 + \sum_{l=1}^{L} \Lambda_l(f)\right] a_L(f,d)\mathbf{a}_t(\theta_t), \tag{1}$$

where G is called the integrated antenna gain consisting of the transmitted and received antenna gains of the antenna array, L denotes the number of non-line-of-sight (NLoS) paths, $\Lambda_l(f)$ is a frequency-dependent constant consisting of the reflection factor and roughness coefficient of NLoS paths affected by the reflective interfaces and material impedance. Moreover, $a_L(f,d)$ is defined as

$$a_L(f,d) = \frac{c}{4\pi f d} e^{-\frac{1}{2}\rho(f)d}$$

by considering a uniform linear array in which c is the speed of light and $\rho(f)$ is the medium absorption factor of frequency f, $\mathbf{a}_t(\theta_t)$ is defined as

$$\mathbf{a}_t(\theta_t) = \frac{1}{N_t}\left[1, e^{j\frac{2\pi}{\lambda}d_a \sin(\theta_t)}, \cdots, e^{j\frac{2\pi}{\lambda}d_a(N_t-1)\sin(\theta_t)}\right]^T, \tag{2}$$

where $\theta_t \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ is the angle of departure, d_a is the distance between two antennas, and T denotes the transpose operation of a vector. Note that $\mathbf{h}$ in (1) consists of line-of-sight (LoS) and NLoS components, that is, $Ga_L(f,d)\mathbf{a}(t)(\theta_t)$ is the LoS component whereas the other term is the NLoS component. Moreover,

B. THz Signal Model

Let $\mathbf{w}_k \in \mathbb{C}^{N_t}$ be the beamforming vector for BS k and $x_k \in \mathbb{C}$ be the signal with unit power transmitted to the kth UE by BS k. Since there are K BSs in the network, we consider the worst scenario that all the BSs interferes each other when they serve their UE. As a result, the signal received by UE k can be specifically written as

$$y_k = \underbrace{\sqrt{P}\mathbf{h}_{kk}^{\mathcal{H}}\mathbf{w}_k x_k}_{\text{desired signal}} + \underbrace{\sum_{j=1, j\neq i}^{K} \sqrt{P}\mathbf{h}_{jk}^{\mathcal{H}}\mathbf{w}_j x_j}_{\text{interference signal}} + \underbrace{n_k}_{\text{noise}}, \tag{3}$$

where $k \in \{1, \ldots, K\}$, P is the transmit power of each BS, and superscript $\mathcal{H}$ stands for the Hermitian operation of a complex vector, $n_k \in \mathbb{C}$ denotes the Gaussian noise, and $\mathbf{h}_{kk} \in \mathbb{C}^{N_t}$ and $\mathbf{h}_{jk} \in \mathbb{C}^{N_t}$ are the channel vectors from BS k to UE k and from (interfering) BS j to UE k, respectively. Note that $\mathbf{h}_{kk}$ and $\mathbf{h}_{jk}$ adopt the channel model defined in (1). As such, the signal-to-noise-plus-interference ratio (SINR) received at the kth UE can be defined as

$$\Gamma_k = \frac{P\,|\mathbf{h}_{kk}^{\mathcal{H}}\mathbf{w}_k|^2}{\sum_{j=1,j\neq i}^{K} P\,|\mathbf{h}_{jk}^{\mathcal{H}}\mathbf{w}_j|^2+\sigma_n^2}, \tag{4}$$

where $|\cdot|$ represents the operator of absolute value and σ_n^2 is the power of the Gaussian noise n_k for all $k \in \{1, \ldots, K\}$. According to (4), the downlink achievable rate (spectral efficiency) of BS k can be written as

$$C_k = \log_2\left(1 + \Gamma_k\right), \quad \text{(bits/sec/Hz)} \tag{5}$$

for all $k \in \{1, \ldots, K\}$. In the following, we will use C_k to formulate an optimization problem of beam search that is able to maximize the sum rate of all the BSs in the scenario that only limited CSI is available at each BS.

C. Sum-Rate Optimization with Limited CSI

In the multi-BS multi-user network, the attenuation of THz channel and beamforming interference contribute to the uncertain system performance. The purpose is to coordinately strengthen the beamforming signal for each user while mitigating the interference. In other words, our objective is to maximize the sum rate of the network by optimizing the beamforming vector $\mathbf{w}_k$ with limited CSI $\mathbf{h}_{kk}^{lim} \in \mathbb{C}^{N_t}$ that is estimated at BS k. The optimization problem can be formulated as

$$\begin{aligned} &\max_{\mathbf{w}_k} \sum_{k=1}^{K} C_k \\ &\text{s.t. } \mathrm{tr}\left\{\mathbf{w}_k\mathbf{w}_k^{\mathcal{H}}\right\} \leq 1,\ \mathbf{h}_{kk} = \mathbf{h}_{kk}^{lim},\ \forall k \in \{1, \ldots, K\}. \end{aligned} \tag{6}$$

However, problem (6) cannot be readily solved via conventional optimization methods with respect to the digital beamforming and partially-attainable CSI. Furthermore, due to high computational complexity of global optimum and high-overhead of CSI exchange, the traditional method is difficult to analyze the sophisticated and unpredictable communication network. As a consequence, we design a deep learning based scheme by leveraging the DRL architecture and federated edge learning architecture to resolve the complex problem.

III. Proposed Federated Deep Reinforcement Learning (FDRL) for THz beam Search

In this section, due to the non-analytic optimization problem and limited attainable CSI of THz network, we propose an FEL based DDPG learning scheme iteratively to coordinate to attain the appropriate THz beamforming policy. We consider that each BS conducts a DDPG to obtain THz beamforming policy with limited CSI. While, multi-BSs are controlled by a single FEL server to exchange training model with hidden information mitigating THz interference.

A. DRL-based DDPG Network

As a brief concept, DRL framework contains state $\mathcal{S}$, action $\mathcal{A}$ and reward function $\mathcal{R}$, where an agent (may be BS or UE) conducts a certain action to obtain the corresponding reward while updating the current status. Therefore, the action will be reinforced iteratively to obtain better rewards under the changing environment. However, our THz beamforming problem exists a large state-action space, which is inappropriate to employ conventional DRL algorithm due to its slow convergence and huge storage of table-mapping. Moreover, traditional reinforcement learning deep neural network (DNN) is adopted in DRL to become a Q-table generator instead of directly accessing and computing Q values in a table-mapping manner. Additionally, since THz beamforming vector is deemed to be continuous variables with substantially-high quantization levels, we adopt the DDPG to establish a two-layered actor-critic network to resolve the problem with continuous solutions. For beamforming policy in the THz network, we define the state, action and rewards as follows.

1) *State space $\mathcal{S}$*: In THz, the state space is situation of each BS under current THz channels, denoted by $\mathcal{S} = \{s_i|\forall i = 1, 2, ..., K\}$, which consists of the serving CSI $\mathbf{h}_{ii}$ linked to the i-th UE, and SINR Γ_i fedback from the UE i. Note that $\mathbf{h}_{ii}$ may be partially attainable due to limited measured CSI under a large-scale THz antenna array. Therefore, state of each BS should be $s_i = \left\{\mathbf{h}_{ii}^{lim}, \Gamma_i|\forall i = 1, 2, ..., K\right\}$.

2) *Action space $\mathcal{A}$*: The action set represents the decision-making of THz beamfoming vector defined as $\mathcal{A} = \{a_i = \{\mathbf{w}_i\}|\forall i = 1, 2, ..., K\}$. Note that each BS will only determine its own action, i.e., $\mathbf{w}_i$ according to current input state and reward.

3) *Reward function $\mathcal{R}$*: We define the overall reward as $\mathcal{R} = \{r_i|\forall i = 1, 2, ..., K\}$. Since we aim at maximizing the sum throughput in (6), we consider the reward function as individual throughput of each BS, i.e., $r_i = C_i$.

As shown in Fig. 2, DDPG architecture contains the main and target networks which individually consist of actor and critic sub-networks, which θ_i^{μ}, θ_i^{Q} denote DNN-enabled actor/critic weights in the main network, and $\theta_i^{\mu'}$ and $\theta_i^{Q'}$ denote the actor/critic weights in the target network, respectively. The main network determines the beamforming action of the i-th THz BS as $a_{i,t} = \mu(s_{i,t}|\theta_i^{\mu}) + N_G$, where $\mu(s_{i,t}|\theta_i^{\mu})$ is the output layer of the DNN-based actor network. To promote the exploration of the environment, the deterministic policy will obtain the probabilistic action by adding the perturbation N_G as Gaussian noise. On the other hand, the target network input is fed by the output action of actor network, which provides the Q-value outcome $Q\left(s_{i,t}, a_{i,t}|\theta_i^{Q}\right)$ via hidden DNN layers at t-th epoch to evaluate the selected action, which is written as

$$Q\left(s_{i,t}, a_{i,t}|\theta_i^{Q}\right) = \mathbb{E}\left[r_i + \gamma Q\left(s_{i,t+1}, a_{i,t+1}|\theta_i^{Q}\right)\right], \tag{7}$$

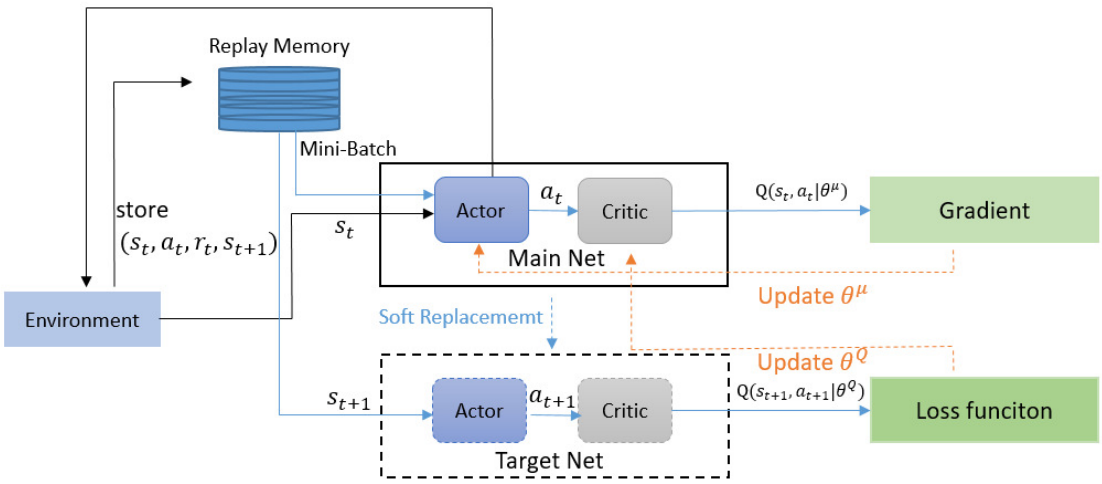


Fig. 2. The DRL-based DDPG algorithm. The main/target networks contain their actor-critic sub-networks established by DNN.

where γ is a discount factor, and $\mathbb{E}[.]$ is the expectation. In order to update the DDPG network, the gradient of actor network is acquired as

$$\begin{aligned}\nabla_{\theta_i^\mu} J_i =& \nabla_{\theta_i^\mu} \mathbb{E}\left[Q\left(s_{i,t}, a_{i,t}|\theta_i^Q\right)\right] \\ =& \mathbb{E}\left[\nabla_{a_{i,t}} Q\left(s_{i,t}, a_{i,t}|\theta_i^Q\right) \cdot \nabla_{\theta_i^\mu} \mu\left(s_{i,t}|\theta_i^\mu\right)\right],\end{aligned} \tag{8}$$

where critic loss function can be given by

$$L_i = \mathbb{E}\left[y_i - Q\left(s_i, a_i|\theta_i^Q\right)\right]^2, \tag{9}$$

where $y_i = r_i + \gamma Q\left(s_{i,t+1}, a_{i,t+1}|\theta_i^{Q'}\right)$. The target network will periodically update the network weights from the main network based on the soft update [15] for both actor-critic sub-networks which is represented by

$$\theta_i^{Q'} = \tau_a \theta_i^Q + (1-\tau_a)\theta_i^{Q'}, \tag{10}$$

$$\theta_i^{\mu'} = \tau_c \theta_i^\mu + (1-\tau_c)\theta_i^{\mu'}, \tag{11}$$

where τ_a, τ_c are constants meaning the significance of parameters in the target and main networks.

B. Federated Edge Learning for Interference Mitigation

After DDPG learning for THz beamforming adjustment at each BS, each BS trains their local DDPG training model of the actor and critic network's weights will be sent from BSs to the edge server every T iteration. The edge server aggregates local training model to exchange the hidden information in neurons of interference information rather than directly upload compellingly-high overhead of full-channel dataset. Based on the weighted FEL method [15], the model aggregation can be presented as

$$\theta^{global} = \frac{1}{K}\sum_{i=1}^{K} \xi_i \theta_i^{local}, \tag{12}$$

where ξ_i is a ratio indicating the importance of each training model depending on certain property of dataset in each BS.

Algorithm 1: Proposed FDRL Algorithm

1: Input: $\mathbf{h}_{ii}, \Gamma_i, \forall i$
2: Output: $a_i = \mathbf{w}_i, \forall i$
3: Initialize: θ_i^μ, $\theta_i^{\mu'}$, θ_i^Q, $\theta_i^{Q'}$ $\forall i$, θ^{global}, replay memory M_r
4: **for** $t = 1, 2..., E$ **do**
5: **for** each BS i **do**
6: Decide the action $a_{i,t} = \mu(s_{i,t}|\theta_i^\mu) + N_G$
7: Interact with the environment and save result of $(s_{i,t}, a_{i,t}, r_i, s_{i,t+1})$ to replay memory M_r
8: Off-line actor/critic model training by mini-batching data with a size of B
9: Soft update $\theta_i^{\mu'}, \theta_i^{Q'}$
10: **end for**
11: **if** $\mod(t, T) = 0$ **then**
12: FEL model aggregation: $\theta^{global} = \frac{1}{K}\sum_{i=1}^{K} \xi_i \theta_i^{local}$
13: Model update after aggregation: $\theta_i^{local} = \theta^{global}$
14: **end if**
15: **end for**

In (12), $\theta_i^{local} = \left\{\theta_i^\mu, \theta_i^Q\right\}$ consists of neural weights of main actor/critic network. Note that in this case, we consider $\xi_i = 1$ as equivalent importance of each beamforming model since the THz BS could provide potentially useful information of limited estimated THz channels. After finishing the model aggregation, the edge server returns the global parameters to each BS, which is repeatedly performed until convergence. Thus, the candidate beamforming of each THz BS $\mathbf{w}_i$ will converge to near optimum by searching for the higher DDPG reward through the iterative training. The concrete algorithm of proposed FDRL is demonstrated in Algorithm 1.

IV. Simulation Results

In this section, we have performed simulations of proposed FDRL-enabled THz beamforming with maximization of sys-

TABLE I
PARAMETERS OF FDRL-ENABLED THZ NETWORK

Definition	Symbol	Value
Discount factor	γ	0.9
Significance of actor network	τ_a	0.01
Significance of critic network	τ_c	0.01
Number of epoch	E	300
Buffer of the memory size	M_r	10
Batch size of DDPG training	B	5
The cycle of FEL	T	20
Number of antenna	N_t	$\{8, 16, 32, 64, 128, 256\}$
Number of BS	K	$\{2, 3, 6\}$
Operating frequency	f	0.3 THz
Distance between BS and UE	d	$[10, 100]$ m
The medium absorption factor	$\rho(f)$	0.1
Bandwidth	W	10 GHz
Number of NLOS paths	N_{NL}	5
Transmit Antenna gains	G_t	10 dB
Receiving Antenna gains	G_r	10 dB
Noise power spectral density	N_0	-174 dBm/Hz
Transmit power	P	10 dBm

tem throughput while mitigating network interference. The THz BS and serving UEs are uniformly-randomly distributed in the radius from 10 to 100 meters. We consider $N_{NL} = 5$ NLOS paths and THz frequency is set to be 0.3 THz. The parameter setting of the THz network is listed in Table I.

In Fig. 3, the convergence of the throughput with $K = 3$, clients served by three BS equipped with $N_t = 8$ antennas. Each BS have a two-layer actor-critic neural network with $\{100, 70\}$ neurons. The actor network decides the beamforming vector $\mathbf{w}_i$, while the critic network using the Q-learning network evaluates the decision of actor network. Initially, the beamforming vector is randomly selected with the perturbation N_G with variance equal to 3 leveraging exploitation and exploration. However, it will gradually decay to 0.99 as deterministic decision. At about 100-th epoch, the client 1 tends to be stable, but the others are still looking for the potential solution from DDPG network. At around 150-th epoch, the performance of throughput is almost converged. Note that the result only shows a run of a certain channel condition in Fig. 3; however, we will conduct more than 100 Monte Carlo runs in the following comparisons.

In Fig. 4, we illustrate the throughput that is affected by the number of the actor and critic neurons. The number of the BSs is $K = 4$ equipped with $N_t = 128$ antennas. As the actor-critic neuron is set as $(20, 20)$, the performance of full FEL upload achieves higher throughput than that of partial upload with around 2 and 0.5 Gbps for 10% and 50% upload of FEL parameters. However, the operational overhead, i.e., operation for training in local network and FEL server is quite higher using full upload than that of 10% upload. That is, the overhead of 10% upload is half the overhead compared to that of full upload while sustaining sufficiently high throughput performance, which also implies that the

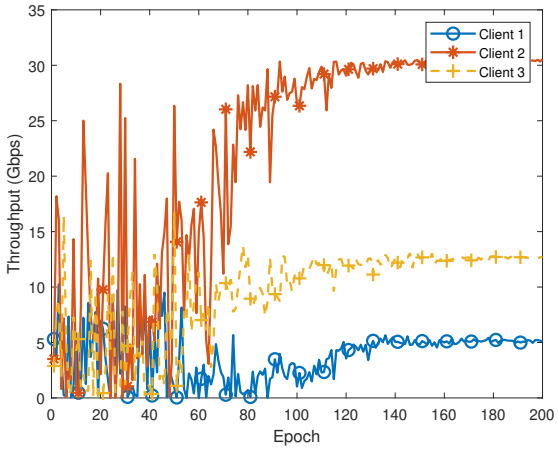
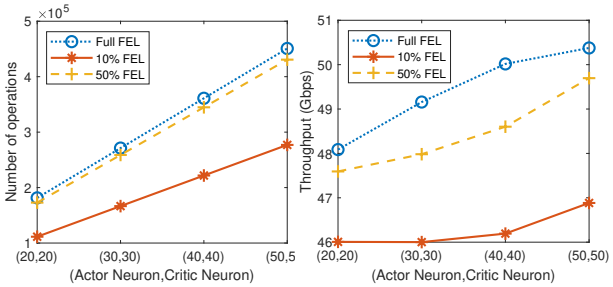


Fig. 3. Covergence of system throughput of the proposed FDRL with $K = 3$ clients.



(a) Complexity of FDRL (b) Throughput of FDRL

Fig. 4. The proposed FDRL compared to different numbers of actor and critic neurons with $\{(20, 20), (30, 30), (40, 40), (50, 50)\}$.

exchanged hidden information is enough featured to alleviate induced interference. Moreover, when the number of actor-critic neurons is $(30, 30)$, there provokes higher computational overhead but reaches high throughput performance.

Fig. 5 demonstrates the throughput considering input states of CSI and only SINR feedback with different number of antennas and THz BS deployment with $K = \{2, 6\}$ and $N_t = \{8, 16, 32, 64, 128, 256\}$ antennas. We can observe the result of $K = 2$ and $K = 6$ that higher performance can be obtained due to advantageous FDRL of interference mitigation. Furthermore, the overall throughput performance is proportional to the number of antennas due to higher spatial diversity. In addition, throughput difference becomes increasingly larger when $N_t = 256$ antennas are equipped by comparing the mechanism of THz CSI and of only SINR feedback. This is because that more hidden information from the estimated CSI is extracted and exchanged by FEL server, which is compellingly advantageous to interference cancellation.

Fig. 6 demonstrates the throughput considering distances between THz BS-UE. The throughput of all algorithms decreases to near zero due to limitation of severe intrinsic pathloss from the THz channel. The proposed FDRL algorithm with

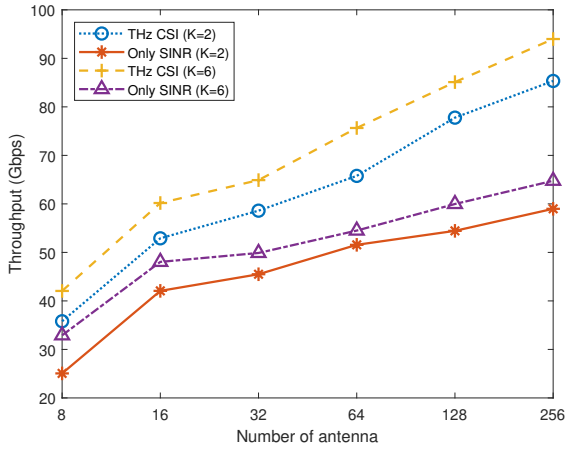


Fig. 5. Throughput of proposed FDRL with different number of antennas considering limited CSI and only SINR feedback with $K = \{2, 6\}$ BS.

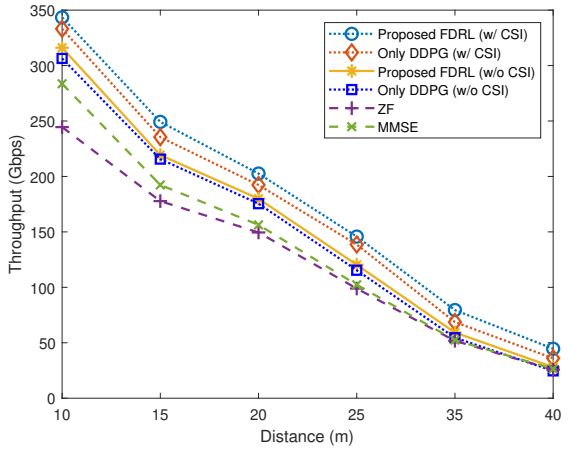


Fig. 6. The proposed FDRL algorithm compared to non-FEL benchmarks and conventional optimization methods.

estimated CSI can exchange more hidden information than that without CSI, which achieves higher throughput performance. In addition, DDPG lacks training model exchange benefited by FEL, i.e., each BS has training without CSI exchange has lower throughput than our FDRL algorithm. Moreover, the proposed FDRL is capable of exchanging sufficient hidden training models from powerful deep learning based DDPG, which outperforms the based conventional beamforming of zero forcing (ZF) and minimum mean square error (MMSE).

V. Conclusions

THz communication is a highly well-known research in the future 6G communication network. However, signal suffers from the serious power attenuation resulting in short propagation distance. We have an proposed FDRL algorithm to maximize the sum rate to intelligently adjust beamformer under limited THz CSI. The performance shows that with more available CSI, FDRL is capable of exchanging representative features among THz BSs to achieve higher throughput. With more deployed antenna arrays, it reaches higher system throughput because of higher spatial diversity. Moreover, it becomes a compelling tradeoff between overhead of exchange information and throughput, i.e., high-throughput can be reached at a cost of high upload overhead. The proposed FDRL scheme outperforms the baseline using pure deep Q-learning, which is beneficial for interference mitigation from information exchange through FEL. Also, FDRL triumphs over the existing beamforming mechanisms using non-learning methods.

References

[1] B. Ning, Z. Chen, W. Chen, Y. Du, and J. Fang, "Terahertz multi-user massive MIMO with intelligent reflecting surface: Beam training and hybrid beamforming," *IEEE Trans. Veh. Technol.*, vol. 70, no. 2, pp. 1376–1393, Feb. 2021.

[2] C. Huang, Z. Yang, G. C. Alexandropoulos, K. Xiong, L. Wei, C. Yuen, and Z. Zhang, "Hybrid beamforming for RIS-empowered multi-hop terahertz communications: A DRL-based method," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2020, pp. 1–6.

[3] P. V. Tuan, T. Trung Duy, and I. Koo, "Multiuser MISO beamforming design for balancing the received powers in secure cognitive radio networks," in *Proc. IEEE Seventh International Conference on Communications and Electronics (ICCE)*, 2018, pp. 39–43.

[4] G. Taricco, "On the beamforming capacity of MISO channels," *IEEE Wireless Commun. Lett.*, vol. 1, no. 2, pp. 141–144, 2012.

[5] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmwave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, Jul. 2017.

[6] S. Yang and Y. Liu, "Training efficiency of federated learning: A wireless communication perspective," in *Proc. International Conference on Wireless Communications and Signal Processing (WCSP)*, 2020, pp. 922–926.

[7] X. Mo and J. Xu, "Energy-efficient federated edge learning with joint communication and computation design," *Journal of Communications and Information Networks*, vol. 6, no. 2, pp. 110–124, 2021.

[8] K.-H. Liu, Y.-H. Hsu, W.-N. Lin, and W. Liao, "Fine-grained offloading for multi-access edge computing with actor-critic federated learning," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, 2021, pp. 1–6.

[9] S. Zarandi and H. Tabassum, "Federated double deep q-learning for joint delay and energy minimization in IoT networks," in *Proc. IEEE International Conference on Communications Workshops (ICC Workshops)*, 2021, pp. 1–6.

[10] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 2031–2063, 2020.

[11] C. Lin and G. Y. Li, "Adaptive beamforming with resource allocation for distance-aware multi-user indoor terahertz communications," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2985–2995, Aug. 2015.

[12] S. Priebe and T. Kurner, "Stochastic modeling of THz indoor radio channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4445–4455, Sep. 2013.

[13] C. Han, A. O. Bicen, and I. F. Akyildiz, "Multi-ray channel modeling and wideband characterization for wireless communications in the terahertz band," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2402–2412, May 2015.

[14] X. Gao, L. Dai, Y. Zhang, T. Xie, X. Dai, and Z. Wang, "Fast channel tracking for terahertz beamspace massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5689–5696, Jul. 2017.

[15] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *Computer Science*, vol. 8, no. 6, 2015.