Orthogonalized RSMA-based Flexible Multiple Access in Digital Twin Edge Networks

Thanh Phung Truong, Hieu V. Nguyen, Nhu-Ngoc Dao, Wonjong Noh, and Sungrae Cho

***Abstract*—This paper proposes a flexible and efficient access control scheme that combines the orthogonal frequency division multiple access and rate-splitting multiple-access techniques for enhancing the uplink transmission in a digital twin edge network system. We formulate a non-convex mixed integer optimization problem that minimizes the energy consumption of all Internet of Things devices (IoTDs) and maximizes the number of successful IoTD tasks. To this end, we propose a deep reinforcement learning (DRL) framework by normalizing a DRL training algorithm named deep deterministic policy gradient for efficiently designing the variables while ensuring the problem constraints. However, in the inference stage, the proposed DRL method may encounter different devices and services. Therefore, we design an exhaustive-improved DRL method that can improve the proposed DRL effectively using information from a digital-twin module. We also propose a mathematical approximation-based solution employing two convexification approach: Dinkelbach's method and relaxed Linear Matrix Inequality (LMI). Through extensive simulations over different parameters and scenarios, we identify the polynomial complexity, stable convergence, and operating regime of the proposed solutions. It is also confirmed that the proposed approaches work well even with digital twin defects and provide improved performance in terms of energy consumption and number of successful tasks in comparison with benchmark schemes.**

***Index Terms*—Rate-splitting multiple access, orthogonal frequency division multiple access, digital twin edge networks**

I. Introduction

THE next era of wireless communication has imposed numerous demands on networks, such as high spectral and energy efficiency, low latency, and ultra-dense connections. To support them, new network architecture, such as mobile edge computing (MEC), is being actively developed [1]. Here, mobile users can offload complex tasks to the MEC server, enabling the execution of intricate applications on smart devices with limited computational capacity and power. Furthermore, the application of digital twin (DT) technologies to advanced networks has recently been widely considered. By constructing a digital world that replicates the physical world, it can provide advanced network systems with robust infrastructure control and strong intelligence [2]. The combination of MEC and DT systems, known as the digital twin edge network (DITEN) or digital twin-assisted mobile edge network, can improve system performance regarding reliability and latency, intelligent decision-making, and technological application [3].

In addition to the DITEN architecture, the rate-splitting multiple access (RSMA) technique [4] has recently been developed as a promising multiple-access strategy [5]. It brings many benefits to networks' spectral and energy efficiency by splitting a transmitted message into multiple parts that one or more receivers can decode flexibly [6], demonstrating its suitability for the next era of communication. In uplink systems, RSMA allows users to split their messages into multiple streams, each with appropriately allocated transmit power. The base station (BS) employs successive interference cancellation (SIC) at the receiver to recover and reconstruct the original messages. This approach dynamically handles inter-user interference by splitting messages [7]. Interestingly, non-orthogonal multiple access (NOMA), a critical multiple access technique for 5G and beyond, is a specific case of RSMA where user messages are not split. As a result, RSMA offers significant advantages over NOMA. Firstly, RSMA provides more decoding order options, making it a more robust and flexible scheme. Secondly, in RSMA, each user's rate is the sum of sub-messages, whereas NOMA depends on decoding a single message. These features make RSMA a more effective solution for managing interference and enhancing performance in uplink communication systems [8], [9].

On the other hand, the evolution of the Internet of Things (IoT) has led to its widespread applications and a new class of devices named NR-RedCap (Reduced Capability New Radio) [10]. These applications and devices have diverse quality of service (QoS) requirements in environments with high interference levels or limited bandwidth [11]. However, the legacy and emerging random access approaches in the DITEN could be more efficient in dealing with the services and devices having widely different QoS requirements. Inspired by this observation, this paper explores a flexible and efficient multiple-access scheme that combines RSMA and orthogonal frequency division multiple access (OFDMA) in the DITEN system. Here, the OFDMA technique is applied to slice the communication channel into orthogonal sub-channels with different amounts of resources suitable for various service demands [12], [13]. Per sub-channel, RSMA is applied to

This work 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), and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2023R1A2C1003003).

Corresponding authors: Wonjong Noh and Sungrae Cho.

Thanh Phung Truong and Sungrae Cho are with the School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, Republic of Korea (e-mail: tptruong@uclab.re.kr; srcho@cau.ac.kr).

Hieu V. Nguyen is with the Faculty of Electronic and Telecommunication Engineering, The University of Danang, University of Science and Technology, Da Nang 50000, Vietnam (email: nvhieu@dut.udn.vn).

Nhu-Ngoc Dao is with the Department of Computer Science and Engineering, Sejong University, Seoul 05006, Republic of Korea (e-mail: nndao@sejong.ac.kr).

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

enhance the communication efficiency of multiuser access [8], [9]. We refer to this scheme as orthogonalized RSMA-based flexible multiple access. In this context, IoTDs that have similar requirements can be grouped and served by an appropriate sub-channel. As each group is transmitted via an orthogonal sub-channel, inter-subchannel interference can be avoided while exploiting RSMA.

A. Related Works

As a promising multiple-access technique for next-gen communication networks, RSMA has recently been studied in broad research [8], [9], [14]. In particular, the authors in [8] aimed to optimize power allocation for a two-user uplink transmission system, in which each transmitted message is split into two sub-streams using the RSMA scheme. In this work, they also demonstrated the outperformance of RSMA to NOMA in a minimum rate maximization problem. The authors in [9] considered decoding order and transmit power as optimization variables in a sum rate maximization problem for an uplink RSMA system. Because finding the decoding order by exhaustive search is highly complex, the authors demonstrated the optimal decoding order in a two-user scenario and established a sub-optimal decoding order for a multi-user case. In this work, the authors also proved the outperformance of RSMA compared with NOMA, TDMA (time-division multiple access), and FDMA (frequency-division multiple access) schemes regarding the system sum rate. The author in [14] investigated a system sum rate maximization problem in a multiuser uplink RSMA system. They proposed a deep reinforcement learning (DRL) algorithm to optimize the users' precoding matrices and a graph-based method to search for the decoding order. Many recent works have also studied the DRL-based optimization schemes for RSMA [15]–[17]. Huang *et al*. [15] proposed a DRL framework for designing power control and resource allocation in RSMA-based low-earth orbit satellite-terrestrial networks, intending to maximize the system sum rate. In [16], Tran *et al*. examined the improvement of RSMA and an intelligent reflecting surface in a streaming system. They applied a DRL algorithm named proximal policy optimization to optimize the resource allocation, RSMA parameters, phase shift, and bitrate adaption to maximize the quality of experience in the system. The authors in [17] formulated an energy efficiency maximization problem in a two-user RSMA network, where they proposed a multi-agent DRL algorithm to optimize the rate and power allocations at the BS. Observably, RSMA studies have focused on several aspects with different objectives. However, reports on combining RSMA with orthogonal multiple-access techniques such as OFDMA for further improvement remain limited.

Considering the efficiency of multiple-access techniques in next-generation communication systems, their application to MEC systems has been widely explored [18]–[20]. For example, OFDMA has been applied to a collaborative MEC network in [18] to resolve the joint problem of task offloading and resource allocation. Chen *et al*. [19] proposed an RSMA-aided MEC system with randomly deployed users, demonstrating the excellent performance attainable using the RSMA-MEC scheme and the superiority of RSMA over NOMA in terms of average throughput and successful computation probability. Diamanti *et al*. [20] investigated the effect of RSMA in a multi-server MEC system, considering a system delay minimization problem. The outcomes highlighted the efficiency of RSMA and its superiority over other multiple-access schemes, such as NOMA and OFDMA, in terms of users' maximum experienced delay.

Additionally, research on DITEN systems has been considered recently [21]–[25]. Liu *et al*. [21] investigated the cooperation of MEC servers in a DITEN system, aiming to minimize the network delay and system power overhead by optimizing the task portion allocated to MEC servers and the task offloading ratio by proposing a DRL-based algorithm. Hao *et al*. [22] focused on minimizing latency and energy consumption in a DITEN system by jointly optimizing resource allocation, task offloading, and power management using a learning-based algorithm. Similar to [21], [22], DRL algorithms have been widely used in DITEN systems, demonstrating their effectiveness in solving DITEN problems [23]–[25]. Moreover, some researchers have explored the effect of multiple-access techniques in DITEN systems [26], [27], demonstrating the effectiveness of NOMA and TDMA in improving the spectral efficiency and energy consumption of DITEN systems. Given the advantages of RSMA over other multiple-access techniques, exploring RSMA and its combination with DITEN systems remains an appealing topic.

Obviously, RSMA has been studied extensively in a variety of research contexts. As a result, RSMA demonstrated its efficiency compared to NOMA, TDMA, and FDMA because of the high diversity order and the flexibility of transmit power and decoding order. It supports robust interference management that improves the spectral efficiency and thus enhances the transmission rate, i.e., improving the system latency [8], [9], [15]–[17]. Besides, previous research has demonstrated the efficiency of dividing the communication bandwidth into multiple sub-channels and applying the NOMA technique at each sub-channel [26], [28]. Therefore, with the effectiveness of RSMA, a multiple access scheme that divides the communication bandwidth into multiple sub-channels and applies the RSMA technique at each sub-channel is an engaging multiple access scheme worth considering. Also, with the above advantages, applying the proposed RSMA-based multiple access scheme to improve MEC systems, especially DITEN systems, opens an attractive research aspect. To highlight the novel of this study, we summarize the related works and our contributions in Table I.

B. Contribution and Organization

To the best of our knowledge, the flexible multiple access scheme based on orthogonalized RSMA proposed in this paper is still in its early development stages. Motivated by the above observations, our study delves into the potential of the proposed orthogonalized RSMA-based flexible multiple access scheme to enhance task execution performance in DITEN systems. We summarize the main contributions of this study as follows.

TABLE I
COMPARISON OF THE PROPOSED STUDY WITH RELATED WORKS

Research	MA Scheme	MEC	DT	Objective	Optimization Method
[8]	RSMA	✗	✗	Maximize users' minimum rate.	SCA and GP-based algorithms.[a]
[9]	RSMA	✗	✗	Maximize system sum-rate.	Alternating-based optimization.
[14], [15]	RSMA	✗	✗	Maximize system sum-rate.	DRL-based optimization.
[16]	RSMA	✗	✗	Maximize quality of experience.	DRL-based optimization.
[17]	RSMA	✗	✗	Maximize system energy effciency.	DRL-based optimization.
[18]	OFDMA	✓	✗	Minimize the total energy consumption.	Alternating-based and DRL-based algorithms.
[19]	RSMA	✓	✗	Maximize successful computation probability.	Closed-form-based solutions.
[20]	RSMA	✓	✗	Minimize users' maximum delay.	Alternating-based algorithm.
[21]	✗	✓	✓	Minimize power consumption and system delay.	DRL-based algorithm.
[22]	✗	✓	✓	Minimize latency and energy consumption.	Learning-based algorithm.
[23]	✗	✓	✓	Maximize energy efficiency.	DRL-based algorithm.
[24]	✗	✓	✓	Minimize long-term energy consumption.	DRL-based optimization.
[25]	✗	✓	✓	Minimize system cost.	Federated DRL-based algorithm.
[26]	NOMA	✓	✓	Minimize tasks' completion delay.	Alternating optimization.
[27]	TDMA	✓	✓	Minimize energy consumption.	DRL-based optimization.
Proposed	Orthogonalized RSMA	✓	✓	Maximize the number of completed tasks while minimizing energy consumption.	Two approaches: LMI-based optimization and DRL-based optimization.

[a]SCA: successive convex approximation, GP: geometric programming

- We explore a flexible and efficient access control scheme that combines RSMA and OFDMA techniques for enhancing uplink transmission in a DITEN system. Here, we formulate a non-convex mixed integer optimization problem aimed at minimizing the energy consumption of all IoT devices (IoTDs) and maximizing the number of successful tasks of IoTDs, considering sub-channel allocation, offloading ratio control, decoding order control, splitting ratio control, and transmit power allocation of the MEC server and service users.
- We propose a deep reinforcement learning (DRL) framework by normalizing the deep deterministic policy gradient (DDPG) algorithm to design variables while adhering to problem constraints. Here, we present new normalization functions to handle discrete variables and develop a constraint processing stage to satisfy all problem constraints. However, in the inference stage, the proposed DRL method may encounter devices other than those used in the training stage. Therefore, we suggest an exhaustive-improved DRL method that can improve the proposed DRL effectively by using information from the digital-twin module.
- We develop a mathematical approximation-based optimal solution to assess the proposed DRL framework. It deals with non-convex combinatorial variables by leveraging power allocation in the successive interference cancellation (SIC)-enabled broadcasting system. Also, it employs the Dinkelbach framework and relaxed linear matrix inequality (LMI) framework for problem convexification.
- Through extensive simulations over different parameters and scenarios, we identify the polynomial complexity, stable convergence, and operating regime of the proposed solutions. It is also confirmed that the proposed approaches work well even with DT defects and provide improved performance in terms of energy consumption and number of successful tasks compared with benchmark schemes. In particular, the proposed framework enhances performance by about 9.7% and 30.7% on average compared to the competitive RSMA and NOMA schemes.

The remaining paper is organized as follows : Section II describes the system model and problem formulation. Sections III and IV outline the deep reinforcement learning-based solution and the mathematical approximation-based solution, respectively. Sections V and VI present the simulation results and conclusion, respectively.

II. System Model and Problem Formulation

Fig. 1 shows the considered system model, which consists of two layers: a physical layer and a digital twin layer. First, in the physical layer, a BS equipped with an edge server plays the role of the MEC, assisting users in a specific range. There exist U IoTDs within the BS range, expressed as $\mathcal{U} = \{1, 2, \ldots, u, \ldots, U\}$, connected with the BS via wireless communication links. With reference to existing studies [29]–[31], all nodes are assumed to be equipped with a single-antenna. Each IoTD u is considered to be assigned a computational task that needs to be executed at each time slot t, expressed as $\tau_u[t] = \{z_u[t], c_u[t], d_u[t]\}$, where $z_u[t], c_u[t]$, and $d_u[t]$ represent the size in bits, required computing resource, and maximum delay of each task, respectively. The duration Δt is used to divide time into separate discrete time slots. The tasks are assumed to be bit-wise independent and can be arbitrarily segmented into sub-tasks. Therefore, each IoTD can process its task locally or offload it to the MEC using a partial offloading scheme with an offloading ratio of $o_u[t] \in [0, 1]$. At time slot t, the IoTD u offloads $o_u[t]$ percent of task $\tau_u[t]$ to the MEC for processing, and the remaining $(1 - o_u[t])$ percent of the task is executed locally. Second, a digital twin layer is established to facilitate the execution of tasks in the physical space. This entity contains a digital mirror that virtually replicates the physical space in real-time and digital twin services that enable monitoring, system simulation, and feeding of the optimal control back into the physical space to

TABLE II
NOTATION

Notation	Definition
$U\backslash\mathcal{U}$	Number\set of IoTDs.
$S\backslash\mathcal{S}$	Number\set of sub-channels.
$U_s\backslash\mathcal{U}_s$	Number\set of IoTDs allocated on sub-channel s.
$K\backslash\mathcal{K}$	Number\set of splitting sub-messages.
$\tau_u = \{z_u, c_u, d_u\}$	Task of IoTD u.
Δ_t	Time slot duration.
o_u	Offloading ratio of IoTD u.
$\alpha_{u,s}$	Channel allocation variable.
$p_{u,k}$	Transmit power of sub-message k of IoTD u.
$h_{u,s}$	Channel response coefficient between IoTD u and BS on sub-channel s.
$\pi^s_{u,k}$	Decoding order of sub-message k of IoTD u on sub-channel s.
π^s	Set of decoding order on sub-channel s.
$x_{u,s}$	Transmit message from IoTD u on sub-channel s.
$x_{u,s,k}$	Sub-message k of transmit message $x_{u,s}$.
n_0, σ_0^2	Additive white Gaussian noise and its variance.
$r_{u,s,k}$	Offloading rate of sub-message $x_{u,s,k}$.
δ_u	Splitting ratio of IoTD u.
$z^o_{u,k}$	Size in bits of sub-offloaded part k of task τ_u.
$T^o_{u,k}\backslash E^o_{u,k}$	Time\energy consumption for offloading part k of task τ_u.
$\hat{f}^l_u\backslash\hat{f}^l_u$	Estimated\deviation computing resource of IoTD u.
$T^l_u\backslash E^l_u$	Time\energy consumption for local execution of task τ_u.
$\hat{f}^e_u\backslash\hat{f}^e_u$	Estimated\deviation computing resource at MEC allocated for IoTD u.
T^{oM}_u	MEC-offloaded time of task τ_u.
T_u	Total latency of task τ_u.
E_u	Total energy consumption of task τ_u.
P_{u-max}	Maximum transmit power of IoTD u.

improve network performance in real-time. To construct the digital mirror model, the MEC gathers raw data, such as the network and user information, from the physical space in real time and sends it to the DT layer [3], [21]. The notation used in this work is summarized in Table II.

A. Communication Model

We divide the bandwidth of the BS into S sub-channels, each with a set of orthogonal subcarriers, numbered based on ascending order of the number of subcarriers in the set. The set of sub-channels is expressed as $\mathcal{S} = \{1, \ldots, s, \ldots, S\}$, where sub-channel S (1) provide the highest (lowest) communication resources. Without loss of generality, the numbers of sub-channels and subcarriers in each sub-channel are assumed to be predefined by the system, considering on the type, requirement, and purpose of specific applications. The connectivity of IoTD u on sub-channel s at time slot t is represented by the channel allocation variable $\alpha_{u,s}[t] \in \{0, 1\}$, where $\alpha_{u,s}[t] = 1$ if IoTD u connects to sub-channel s and $\alpha_{u,s}[t] = 0$ otherwise. Each IoTD can utilize at most one sub-channel at each time slot. We assume that the number of IoTDs is greater than the number of sub-channels. Hence, we formulate the following sub-channel allocation constraint:

$$\sum_{s\in\mathcal{S}} \alpha_{u,s}[t] \leq 1, u \in \mathcal{U}. \tag{1}$$

Then, to ensure that all offloaded tasks are successfully offloaded to the MEC server, the offloading event of IoTD u is defined to occur only if this IoTD is accessed through the communication channel. To this end, the offloading ratios must satisfy the following channel allocation constraint:

$$o_u[t] \in \left[0, \sum_{s\in\mathcal{S}} \alpha_{u,s}[t]\right], u \in \mathcal{U}. \tag{2}$$

Here, we let $U_s[t]$ and $\mathcal{U}_s[t]$ represent the number and set of IoTDs transmitting signals on sub-channel s in time slot t, respectively, with $U_s[t] = \sum_{u\in\mathcal{U}} \alpha_{u,s}[t], s \in \mathcal{S}$ and $\mathcal{U}_s[t] \triangleq \{u \in \mathcal{U} \mid \alpha_{u,s}[t] = 1\},\ s \in \mathcal{S}$.

In uplink RSMA, each transmitted message of IoTD u on sub-channel s is split into K sub-messages and transmitted to the BS via the same frequency slot (communication channel). Without loss of generality, we assume that each IoTD can split its transmitted message into two sub-messages ($K = 2$)[1] and define $\mathcal{K} \triangleq \{1, \cdots, K\}$. Then, the transmitted message from IoTD u on sub-channel s in time slot t is expressed as

$$x_{u,s}[t] = \sum_{k=1}^{2} \sqrt{p_{u,k}[t]} x_{u,s,k}[t], \tag{3}$$

where $p_{u,k}[t] \geq 0$ is the transmitted power that the IoTD u spends in transmitting sub-message $x_{u,s,k}[t]$. At the BS, the messages received in time slot t contain all messages received at all sub-channels, expressed as

$$\begin{aligned} x_t[t] &= \sum_{s\in\mathcal{S}} \sum_{u\in\mathcal{U}_s[t]} \alpha_{u,s}[t] h_{u,s}[t] x_{u,s}[t] + n_0 \\ &= \sum_{s\in\mathcal{S}} \sum_{u\in\mathcal{U}_s[t]} \alpha_{u,s}[t] h_{u,s}[t] \sum_{k=1}^{2} \sqrt{p_{u,k}[t]} x_{u,s,k}[t] + n_0, \end{aligned} \tag{4}$$

where $h_{u,s}[t] \in \mathbb{C}$ denotes the channel response coefficient between IoTD u and the BS on sub-channel s in time slot t, while $n_0 \sim \mathbb{C}(0, \sigma_0^2)$ represents additive white Gaussian noise. Subsequently, in each sub-channel, the BS applies the SIC technique to decode the sub-messages received within time slot t, in accordance with a decoding order. We define $\pi^s_{u,k}[t],\ (u, k) \in \mathcal{U}_s[t] \times \mathcal{K}$, as the decoding order of sub-message k of IoTD u connected via channel s and $\pi^{\mathbf{s}}[t] \triangleq \{\pi^s_{u,k}[t] \mid \bar{U}_s[t] \mid \forall (u, k) \in \mathcal{U}_s[t] \times \mathcal{K}\}$ as the set of decoding orders on channel s, where $\bar{U}_s[t] \triangleq \{1, 2, \cdots, 2|\mathcal{U}_s[t]|\}$. By setting $\mathcal{C}^s_{u,k}[t] \triangleq \{(u', k') \in \mathcal{U}_s[t] \times \mathcal{K} \mid \pi^s_{u',k'}[t] > \pi^s_{u,k}[t]\}$ and using the ascending decoding order, the offloading rate of sub-message $x_{u,s,k}$ in time slot t can be calculated as

$$r_{u,s,k}[t] = B_s \log_2\Big(1 + \frac{\alpha_{u,s}[t]|h_{u,s}[t]|^2 p_{u,k}[t]}{\mathrm{IN}_{u,s,k} + \sigma_0^2}\Big), \tag{5}$$

where $\mathrm{IN}_{u,s,k} \triangleq \sum_{(u',k')\in\mathcal{C}^s_{u,k}[t]} \alpha_{u',s}[t]|h_{u',s}|^2[t] p_{u',k'}[t]$, and B_s is the communication bandwidth of sub-channel s.

B. RSMA Offloading Model

For the RSMA scheme, we define a splitting ratio for each IoTD u within time slot t, $\delta_u[t] \in [0, 1]$, which splits each offloaded part of a task into two sub-offloaded parts. Then,

[1]Many previous studies have set $K = 2$ [9], [32], [33].

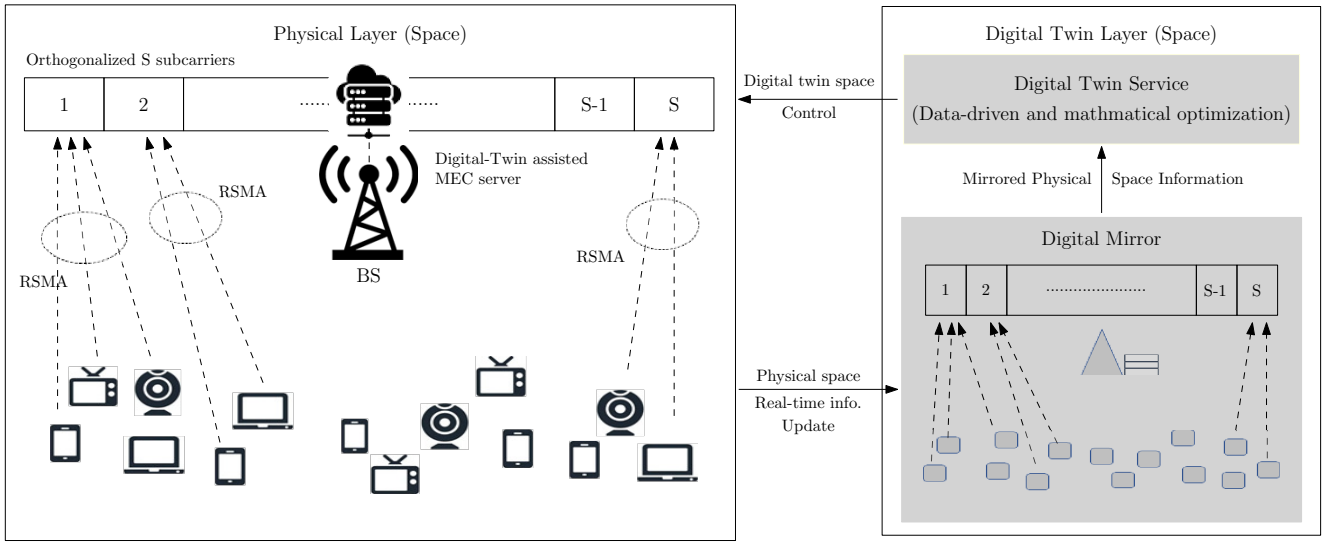


Fig. 1. System model: RSMA-enabled multiple access for DITEN systems.

the size (in bits) of the sub-offloaded parts of task $\tau_u[t]$ in time slot t, $z^o_{u,k}[t], k \in \{1,2\}$, is calculated as

$$\begin{aligned} z^o_{u,1}[t] &= \delta_u[t] o_u[t] z_u[t], \\ z^o_{u,2}[t] &= (1-\delta_u[t]) o_u[t] z_u[t]. \end{aligned} \tag{6}$$

To ensure that all sub-offloaded parts are successfully offloaded to the BS, the transmitted powers must be greater than zero during offloading. Therefore, we introduce the following constraint for the RSMA offloading power:

$$\begin{cases} p_{u,k}[t] > 0 & \text{, if } z^o_{u,k}[t] > 0, \\ p_{u,k}[t] = 0 & \text{, otherwise.} \end{cases} \tag{7}$$

Before deriving the energy consumption, we introduce the following proposition.

Proposition 1. When $z^o_{u,k}[t] > 0$, there is always one value of the offloading rate that satisfies $\sum_{s\in\mathcal{S}} r_{u,s,k}[t] > 0$, $s \in \mathcal{S}$.

Proof. According to (6), $o_u[t] > 0$ when $z^o_{u,k}[t] > 0$. Then, from (2), $\sum_{s\in\mathcal{S}} \alpha_{u,s}[t] > 0$, that is, there exists at least one value of $\alpha_{u,s}[t]$ greater than zero, i.e., $\exists \alpha_{u,s}[t] = 1$. In accordance with (1), there exists only one non-zero value of $\alpha_{u,s}[t]$, $s \in \mathcal{S}$. In addition, based on (7), $p_{u,k} > 0$ when $z^o_{u,k}[t] > 0$. Therefore, only one non-zero value of $\alpha_{u,s}[t] p_{u,k}[t]$ exists, resulting in $\sum_{s\in\mathcal{S}} r_{u,s,k}[t] > 0$ (5), $s \in \mathcal{S}$. This proves **Proposition 1**. □

According to Proposition 1, the following statement holds: If each sub-offloaded part $z^o_{u,k}[t] > 0$, during which information is offloaded from IoTD u to the MEC through all subcarriers, it is necessary and sufficient to calculate the corresponding offloading time as

$$T^o_{u,k}[t] = \frac{z^o_{u,k}[t]}{\sum_{s\in\mathcal{S}} r_{u,s,k}[t]}, \tag{8}$$

and thus, the energy consumption of IoTD u for offloading $z^o_{u,k}[t]$ can be obtained by

$$E^o_{u,k}[t] = p_{u,k}[t] T^o_{u,k}[t] = \frac{p_{u,k}[t] z^o_{u,k}[t]}{\sum_{s\in\mathcal{S}} r_{u,s,k}[t]}. \tag{9}$$

Remark 1. When $z^o_{u,k}[t] = 0$, the offloading power $p_{u,k}[t]$ in (7) and offloading rate $r_{u,s,k}[t]$ in (5) become zero, and thus, the values in (8) and (9) are $\frac{0}{0}$. To avoid these cases, we define $T^o_{u,k}[t] = E^o_{u,k}[t] = 0$ if $z^o_{u,k}[t] = 0$. In other words, the offloading time and energy consumption for offloading are set as zero when no offloading occurs.

C. DT Computation Model

In the DT layer, the digital mirror replicates the computation resource model of the IoTDs and MEC server in the physical layer by estimating the computing resource for each entity. However, DT cannot fully reflect the state and can exhibit a deviation in the computing resource between the physical and digital layers [34], [35]. First, the computation resource model of each IoTD u in the DT layer can be expressed as

$$DT^{IoTD}_u = \{\tilde{f}^l_u, \hat{f}^l_u\}, \tag{10}$$

where $\tilde{f}^l_u$ and $\hat{f}^l_u$ are the estimated computing resources of the u-th IoTD and deviation in the computing resource between the physical IoTD u and its DT counterpart, respectively. Assuming that the values of $\tilde{f}^l_u$ and $\hat{f}^l_u$ can be acquired in advance [21], [34], the estimated local execution time at IoTD u for task $\tau_u[t]$ can be calculated as

$$\tilde{T}^l_u[t] = \frac{(1-o_u[t]) z_u[t] c_u[t]}{\tilde{f}^l_u}, \tag{11}$$

and the local execution time gap between the real time and DT estimated time is computed as

$$\Delta T^l_u[t] = -\frac{(1-o_u[t]) z_u[t] c_u[t] \hat{f}^l_u}{\tilde{f}^l_u (\tilde{f}^l_u + \hat{f}^l_u)}. \tag{12}$$

Then, the actual local execution time can be computed as

$$T^l_u[t] = \tilde{T}^l_u[t] + \Delta T^l_u[t] = \frac{(1-o_u[t]) z_u[t] c_u[t]}{(\tilde{f}^l_u + \hat{f}^l_u)}. \tag{13}$$

Accordingly, the local execution energy consumption of the m-th IoTD can be computed as

$$E_u^l[t] = \kappa_u(1 - o_u[t])z_u[t]c_u[t](\tilde{f}_u^l + \hat{f}_u^l)^2, \quad (14)$$

where κ_u is a hardware architecture-related energy coefficient of IoTD u. Second, similarly, the computation resource model of the MEC server in the DT layer can be expressed in terms of the estimated computing resource allocated to each IoTD u, $\tilde{f}_u^e$, and deviation in the computing resource allocated to IoTD u between the physical MEC and its DT counterpart, $\hat{f}_u^e$, acquired in advance. Therefore, the actual MEC execution time of task $\tau_u[t]$ for the offloaded part is computed as

$$T_u^M[t] = \tilde{T}_u^M[t] + \Delta T_u^M[t] = \frac{o_u[t]z_u[t]c_u[t]}{(\tilde{f}_u^e + \hat{f}_u^e)}, \quad (15)$$

where $\tilde{T}_u^M[t]$ and $\Delta T_u^M[t]$ denote the estimated execution time at MEC of task $\tau_u[t]$ for the offloaded part and gap between the real and estimated execution time values, respectively.

Remark 2. The predefined estimated and deviation values ($\tilde{f}_u^l$, $\hat{f}_u^l$, $\tilde{f}_u^e$, $\hat{f}_u^e$) determine the DT system's performance, with zero deviation indicating an ideal digital mirror of the DT system. In general, a well-functioning DT system should exhibit small deviation values. Therefore, to ensure the effectiveness of the predefined DT system, the deviation values are considered to be smaller than the estimated values, that is,

$$(\tilde{f}_u^l + \hat{f}_u^l) > 0;\ (\tilde{f}_u^e + \hat{f}_u^e) > 0. \quad (16)$$

D. Problem Formulation

This research seeks to reduce the energy consumption of all IoTDs while increasing the number of successful tasks completed by IoTDs. A task $\tau_u[t]$ is considered successful if its execution time meets the maximum delay requirement $d_u[t]$, assumed to be completed within time slot Δt. Based on the RSMA offloading scheme, the execution time of each task $\tau_u[t]$ is defined as the larger value between the MEC-offloaded time and the local execution time. Here, the MEC-offloaded time of the offloaded part, $T_u^{oM}[t]$, is calculated as the sum of the offloading time and MEC execution time:

$$T_u^{oM}[t] = \max\{T_{u,1}^o[t], T_{u,2}^o[t]\} + T_u^M[t]. \quad (17)$$

Then, the execution time of task $\tau_u[t]$, $T_u[t]$, is

$$T_u[t] = \max\{T_u^l[t], T_u^{oM}[t]\}. \quad (18)$$

The energy consumption of IoTD u required for executing task $\tau_u[t]$, $E_u[t]$, is computed as the sum of the local execution energy consumption and total energy consumption for offloading its sub-offloaded parts:

$$E_u[t] = E_u^l[t] + \sum_{k\in\{1,2\}} E_{u,k}^o[t]. \quad (19)$$

Accordingly, we establish a utility function for the system, which is proportional to the number of successful tasks and inversely proportional to the energy consumption of the IoTDs, formulated as

$$\mathcal{V}[t] = \frac{\sum_{u\in\mathcal{U}} \mathcal{H}(d_u[t] - T_u[t])}{\sum_{u\in\mathcal{U}} E_u[t]} \quad (20)$$

where $\mathcal{H}(x)$ denotes a step function, i.e., $\mathcal{H}(x) = 1$, if $x \geq 0$, otherwise $\mathcal{H}(x) = 0$, implying that the function has a value of 1 if the task is successful and 0 otherwise.

Proposition 2. The energy consumption of the u-th IoTD is always greater than zero when it processes a valuable task $\tau_u[t]$, i.e., $E_u[t] > 0, \forall t$, if $z_u[t], c_u[t], d_u[t] > 0$.

Proof. We analyze the value of $E_u[t]$ for three cases of offloading ratios, $o_u[t] \in [0, 1]$:

- $o_u[t] = 0$: Based on (6), $z_{u,k}^o[t] = 0$, $\forall k \in \{1, 2\}$. Then, the energy consumption is determined through (19) as

$$\begin{aligned} E_u[t] = E_u^l[t] &= \kappa_u(1 - o_u[t])z_u[t]c_u[t](\tilde{f}_u^l + \hat{f}_u^l)^2 \\ &= \kappa_u z_u[t]c_u[t](\tilde{f}_u^l + \hat{f}_u^l)^2. \end{aligned} \quad (21)$$

  Given that $\left(\tilde{f}_u^l - \hat{f}_u^l\right)^2 > 0$ (from *Remark* 2), $E_u[t] > 0$.
- $o_u[t] = 1$: In this case, the energy consumption can be determined through (19) as

$$E_u[t] = \sum_{k\in\{1,2\}} E_{u,k}^o[t] = \sum_{k\in\{1,2\}} p_{u,k}[t]\frac{z_{u,k}^o[t]}{r_{u,s,k}[t]}. \quad (22)$$

  According to (6), there exists at least one sub-offloaded part that satisfies $z_{u,k}^o[t] > 0$. Then, according to (7) and **Proposition 1**, there exists at least one value of $p_{u,k}[t] > 0$ and $r_{u,s,k}[t] > 0$, i.e., $E_u[t] > 0$.
- $0 < o_u[t] < 1$, i.e., $(1 - o_u[t]) > 0$: Based on *Remark* 2, the local execution energy consumption is computed as

$$E_u^l[t] = \kappa_u\left(1 - o_u[t]\right) z_u[t]c_u[t]\left(\tilde{f}_u^l + \hat{f}_u^l\right)^2 > 0. \quad (23)$$

  According to (6), (7), and **Proposition 1**, there exists at least one set of values that satisfies $z_{u,k}^o[t] > 0$, $p_{u,k}[t] > 0$, and $r_{u,s,k}[t] > 0$. Then, the total energy consumption for offloading is computed as

$$\sum_{k\in\{1,2\}} E_{u,k}^o[t] = \sum_{k\in\{1,2\}} p_{u,k}[t]\frac{z_{u,k}^o[t]}{r_{u,s,k}[t]} > 0. \quad (24)$$

  Therefore, the energy consumption in (19) can be expressed as

$$E_u[t] = E_u^l[t] + \sum_{k\in\{1,2\}} E_{u,k}^o[t] > 0. \quad (25)$$

Consequently, the value of $E_u[t]$ is always greater than zero, which proves **Proposition 2**. □

It follows from Proposition 2 that the positive denominator, i.e., $\sum_{u\in\mathcal{U}} E_u[t] > 0, \forall t$, in (20) numerically validates the utility rational function within all time slots.

To attain the multiple objective of minimizing the energy consumption of all IoTDs and maximizing the number of successful tasks, we formulate a utility maximization problem subject to constraints of the offloading operation, splitting ratios, channel allocation, IoTD transmit powers, and decoding orders. Defining $\mathbf{o}[t] = \{o_u[t], u \in \mathcal{U}\}$, $\delta[t] = \{\delta_u[t], u \in \mathcal{U}\}$, $\alpha[t] = \{\alpha_{u,s}[t], u \in \mathcal{U}, s \in \mathcal{S}\}$, $\mathbf{p}[t] = \{p_{u,o,1}[t], p_{u,o,2}[t]\}$,

and $\pi[t] = \{\pi^s[t], s \in \mathcal{S}\}$, the target problem can be formulated as

$$\text{(P1):} \quad \max_{\mathcal{F}[t]} \quad \mathcal{V}[t] \tag{26a}$$

$$\text{s.t.} \quad o_u[t] \in \left[0, \sum_{s\in\mathcal{S}} \alpha_{u,s}[t]\right], u \in \mathcal{U}, \tag{26b}$$

$$\delta_u[t] \in [0,1], u \in \mathcal{U}, \tag{26c}$$

$$\alpha_{u,s}[t] \in \{0,1\}, u \in \mathcal{U}, s \in \mathcal{S}, \tag{26d}$$

$$\sum_{s\in\mathcal{S}} \alpha_{u,s}[t] \le 1, u \in \mathcal{U}, \tag{26e}$$

$$\begin{cases} p_{u,k}[t] > 0 & \text{, if } z^o_{u,k}[t] > 0, \\ p_{u,k}[t] = 0 & \text{, otherwise.} \end{cases}, \forall k, \ \forall u, \tag{26f}$$

$$\sum_{k\in\{1,2\}} p_{u,k}[t] \le P_{u-\max}, u \in \mathcal{U}, \tag{26g}$$

where $\mathcal{F}[t] \triangleq \{\mathbf{o}[t], \delta[t], \alpha[t], \mathbf{p}[t], \pi[t]\}$. Constraints (26b), (26c), and (26d) define the value ranges of the variables. Constraint (26e) ensures that each IoTD can access up to one subcarrier in a certain time slot. Constraint (26f) is to guarantee that all sub-offloaded parts of tasks are successfully offloaded. Notably, $z^o_{u,k}[t]$ represents a product function associated with two variables ($\delta_u[t]$ and $o_u[t]$), as indicated in (6). Finally, the power constraint is specified by (26g), with $P_{u-\max}$ being the maximum transmit power of IoTD u.

In (P1), the objective function is non-convex. In addition, the decoding order and channel allocation variables render the considered problem a mixed-integer problem. Therefore, problem (P1) is a mixed-integer non-convex problem, for which it is challenging to determine an optimization solution. A realistic strategy to solve this problem is to develop an approximation approach to identify the optimal solution. To this end, we propose two approaches based on DRL and mathematical approximation.

III. Proposed DRL Framework

A. Reinforcement learning-based Problem

We formulate (P1) as a reinforcement learning problem, leveraging concepts from the Markov decision process (MDP), a stochastic control mechanism that models decision-making problems in which the decision-maker learns from interactions to achieve a goal. The decision-maker (the agent) decides actions and interacts with the environment, including everything outside of the agent, by observing the environmental state. The environment then presents the agent with a new state and a reward defined according to the goal [36]. We define the state space, action space, and reward function as follows:

Definition 1. State space: The state space contains the IoTD task information and channel gain between the IoTDs and communication channels, expressed as

$$s[t] = \{z_u[t], c_u[t], d_u[t], h_{u,s}[t], u \in \mathcal{U}, s \in \mathcal{S}\} \tag{27}$$

Definition 2. Action space: The action space contains optimization variables, given as

$$a[t] = \mathcal{F}[t] \triangleq \{\mathbf{o}[t], \delta[t], \alpha[t], \mathbf{p}[t], \pi[t]\}. \tag{28}$$

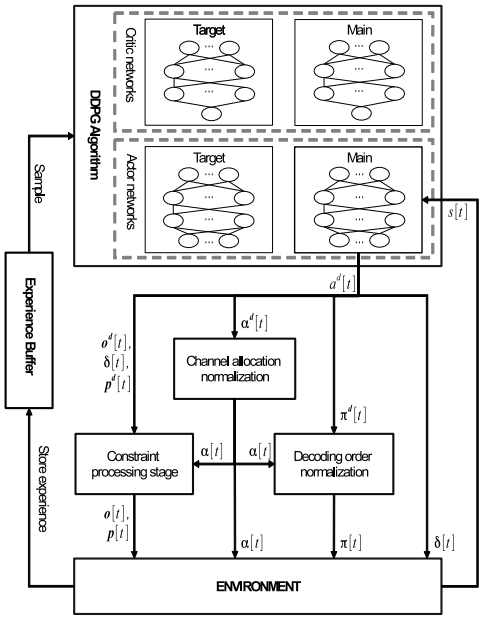


Fig. 2. Proposed deep reinforcement learning framework.

Algorithm 1 Decoding order determination.

Input:
  Channel allocation variables: $\alpha[t]$.
  Decoding priority variables: $\pi^{\mathbf{d}}[t]$.
for $s \in \mathcal{S}$ **do**
  Set of IoTDs in channel s: $\mathcal{U}_s = \{u \in \mathcal{U} | \alpha_{u,s} = 1\}$.
  Sort $\mathcal{U}_s$ by $\pi^d_{u,k}[t]$ in descending order.
  for $(u,k) \in \mathcal{U}_s[t] \times \mathcal{K}$ **do**
    $\pi^s_{u,k}[t] \leftarrow$ index of $\pi^d_{u,k}$ in $\mathcal{U}_s[t]$.
  end for
end for
return Decoding order $\pi[t]$.

Definition 3. Reward function: Because the objective is to maximize the value functions of all IoTDs, the reward function within each time slot is calculated by the sum of the IoTDs' utility functions:

$$r[t] = \sum_{u\in\mathcal{U}} \mathcal{V}_u[t]. \tag{29}$$

According to Proposition 2, the value function $\mathcal{V}_u[t]$ exists in each time slot, rendering the reward function always valuable.

B. Proposed Normalized DDPG Algorithm

To train the agent, we incorporate an actor-critic algorithm named DDPG into the proposed DRL framework. The algorithm is composed of actor and critic networks (ANs and CNs), each with its own set of primary and target networks. The primary AN, presented by a parameterized policy function $\mu(s|\theta^\mu)$, acts as a decision-maker, deciding an action based on the observed state. The primary CN, presented by a parameterized action-value function $Q(s,a|\theta^Q)$, evaluates the chosen action a at each state s by estimating the expected cumulative return. The target networks help enhance the training performance [37]. Furthermore, the algorithm explores

samples for training the networks by adding noise to generate actions, expressed as:

$$a = \mu\left(s|\theta^{\mu}\right) + \mathcal{N}, \tag{30}$$

where $\mathcal{N}$ denotes noise generated by the Ornstein–Uhlenbeck process [38]. Accordingly, the AN's parameter is updated by the policy gradient of the expected return based on the action, expressed as

$$\nabla_{\theta^{\mu}} J = \frac{1}{D}\sum_{i=1}^{D}\left(\nabla_a Q(s,a|\theta^{Q})|_{s=s_i,a=\mu(s_i)}\nabla_{\theta^{\mu}}\mu(s_i|\theta^{\mu})\right), \tag{31}$$

where D is the training samples batch size. The CN's parameter is updated by minimizing the loss function expressed as

$$L = \frac{1}{D}\sum_{i=1}^{D}\left(Q(s_i,a_i|\theta^{Q}) - y_i\right)^2, \tag{32}$$

where y_i denotes the estimated return value. Denoting the target CN ans AN as $Q'\left(s,a|\theta^{Q'}\right)$ and $\mu'\left(s|\theta^{\mu'}\right)$, respectively, the estimated return value can be calculated as

$$y_i = r_i + \gamma Q'\left(s_i', \mu'(s_i'|\theta^{\mu'})|\theta^{Q'}\right), \tag{33}$$

where s_i' is the next state of the current sample state s_i. Subsequently, the parameters of the target networks are updated from the primary networks through a soft update with coefficient τ:

$$\begin{aligned}\theta^{\mu'} &\leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'},\\ \theta^{Q'} &\leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'}.\end{aligned} \tag{34}$$

The networks are trained off-policy, with the training samples randomly selected from an experience replay buffer, which stores the interaction experiences.

By using the AN, the action value range can be designed by selecting the activation function. Considering the constraints in (26), we scale actions to the range between 0 and 1. Accordingly, we define $o_u^d[t], \delta_u^d[t], \alpha_{u,s}^d[t], p_{u,k}^d[t], \pi_{u,k}^d \in [0,1], u \in \mathcal{U}, k \in \{1,2\}, s \in \mathcal{S}$ as the variables that the DDPG decides, i.e., the offloading ratios, splitting ratios, channel allocation, transmit power, and decoding orders, respectively. Consequently, the constraint of splitting ratio variables in (26c) is satisfied. Therefore, the splitting ratio of IoTD u can be directly obtained from the AN, i.e., $\delta_u[t] = \delta_u^d[t]$, $u \in \mathcal{U}$. However, the AN produces continuous variables that do not correspond to the discrete form of the channel allocation and decoding order. Therefore, we propose the following normalization functions to handle the variable form.

1) Channel Allocation Normalization: By letting $\alpha_{u,s}^d[t]$ be the probability that IoTD u connects to the BS on channel s in time slot t, the u-th IoTD only connects to the channel with the highest probability, otherwise it does not establish any connections. Denoting $\alpha_u^{\mathbf{d}}[t] = \{\alpha_{u,s}^d[t], s \in \mathcal{S}\}$ as the set of probabilities that IoTD u connects to the channels, the channel allocation variables of IoTD u can be expressed as

$$\alpha_{u,s}[t] = \begin{cases} 1 & \text{, if } s = \operatorname{argmax}\{\alpha_u^{\mathbf{d}}[t]\} \\ 0 & \text{, otherwise.}\end{cases} \tag{35}$$

Thus, by applying the channel allocation probability variables, the constraints (26d) and (26e) are satisfied.

Algorithm 2 Proposed DRL-based training algorithm

```
Initialize μ(s|θ^μ), Q(s,a|θ^Q), μ'(s|θ^μ'), and Q'(s,a|θ^Q').
for e = 1, 2, ..., E do
    Obtain initial state s[1].
    for t = 1, 2, ..., T do
        Actor network decides action a^d[t] by (30).
        Calculate channel allocation variables, α[t], using (35).
        Determine the decoding order, π[t], as in Algorithm 1.
        Apply the constraint processing stage: equations (36)–(39)
        → obtain o[t], p[t].
        Apply a[t] to the environment → obtain r[t], s[t+1].
        Store {s[t], a[t], r[t], s[t+1]} into buffer.
        Update state: s[t] ← s[t+1].
        Randomly sample a batch (size D) from the replay buffer.
        Update neural network parameters using (31), (32), (34).
    end for
end for
return Trained actor network.
```

2) Decoding Order Normalization: By letting $\pi_{u,k}^d[t]$ represent the decoding priority of sub-message k of IoTD u, the decoding order of the sub-messages in each sub-channel is designed based on the decoding priority in descending order, determined using Algorithm 1, where $\pi^{\mathbf{d}}[t] \triangleq \{\pi_{u,k}^d[t], u \in \mathcal{U}, k \in \{1,2\}\}$. In each sub-channel, the sub-message with the highest decoding priority is first decoded, and the same principle is followed for the remaining sub-messages. After considering all sub-channels, the decoding order of all sub-messages can be identified.

3) Constraint Processing Stage: By using the normalization functions, three constraints remain in problem P1: (26b), (26f), and (26g). To manage the constraints, we introduce a constraint processing stage as follows. First, let $o_u^m[t]$ denote the offloading ratio that satisfies constraint (26b), it can be calculated as

$$o_u^m[t] = o_u^d[t]\sum_{s\in\mathcal{S}}\alpha_{u,s}[t]. \tag{36}$$

By applying this calculation, the offloading ratio is set as zero when no communication channel exists, i.e., $\sum_{s\in\mathcal{S}}\alpha_{u,s}[t] = 0$, thereby satisfying constraint (26b).

Furthermore, to adhere to the constraint in (26f), we denote $p_{u,k}^m[t]$ as the mapped transmit power variable of sub-message k of IoTD u in time slot t, calculated as

$$\begin{aligned}p_{u,1}^m[t] &= p_{u,1}^d[t]\lceil\delta_{u,1}[t]\rceil,\\ p_{u,2}^m[t] &= p_{u,2}^d[t]\lceil\delta_{u,2}[t]\rceil,\end{aligned} \tag{37}$$

where $\delta_{u,1}[t] \triangleq \delta_u[t]o_u^m[t]$, $\delta_{u,2}[t] \triangleq (1-\delta_u[t])\,o_u^m[t]$, and $\lceil . \rceil$ is the ceiling function, which returns the smallest integer greater or equal to the input number. Accordingly, $\lceil\delta_{u,1}[t]\rceil$ and $\lceil\delta_{u,2}[t]\rceil$, indicating the corresponding offloading events of sub-messages 1 and 2, have two possible values (0 or 1). If the offloading event is flagged as zero, i.e., $z_{u,k}^o[t] = 0$ (according to (6)), the corresponding transmit power is also zero, i.e., $p_{u,k}^m[t] = 0$; otherwise, it is $p_{u,k}^m[t] = p_{u,k}^d[t]$. Through this mapping function, the second condition in (26f) is satisfied. To avoid violating the first condition in (26f), i.e., $z_{u,k}^o[t] > 0$ and $p_{u,k}[t] = 0$, we force the offloading ratio to zero, preventing the IoTDs from offloading tasks with insufficient transmit

power. Hence, the offloading ratio of IoTD u at time slot t that satisfies all its constraints is

$$o_u[t] = \begin{cases} 0, & \text{if } \exists_{k\in\mathcal{K}} \lceil p^m_{u,k}[t] \rceil < \lceil \delta_{u,k}[t] \rceil, \\ o^m_u[t], & \text{otherwise.} \end{cases} \tag{38}$$

With the mapped transmit power and valid offloading ratio, the only remaining constraint is (26g). Therefore, we propose the following proposition to adjust the transmit power variables to appropriate values.

Proposition 3. To satisfy constraint (26g), $p_{u,1}[t]$ and $p_{u,2}[t]$ can be calculated using the following function:

$$p_{u,1}[t] = p^m_{u,1}[t] P_{u-max}[t], \tag{39a}$$

$$p_{u,2}[t] = p^m_{u,2}[t] \left(P_{u-max}[t] - p_{u,1}[t]\right). \tag{39b}$$

Proof. By substituting $\delta_u[t], p^d_{u,k}[t] \in [0,1]$ into (37), we obtain $p^m_{u,k}[t] \in [0,1]$:

- According to (39a), the value range of $p_{u,1}[t]$ is $p_{u,1}[t] \in [0, P_{u-max}]$, i.e., $p_{u,1}[t] \geq 0$, and $p_{u,1}[t] \leq P_{u-max}$.
- According to (39b), we have $p_{u,2}[t] \geq 0$ and $p_{u,2}[t] \leq (P_{u-max} - p_{u,1}[t]) \Rightarrow p_{u,1}[t] + p_{u,2}[t] \leq P_{u-max}$, i.e., $\sum_{k\in\mathcal{K}} p_{u,k}[t] \leq P_{u-max}$.

Consequently, constraint (26g) is satisfied, which proves **Proposition 3** □

Consequently, the constraint processing stage ensures that all constraints in the proposed problem are satisfied.

Fig. 2 illustrates the proposed DRL framework, and the pseudocode of the algorithm is presented in Algorithm 2. The agent is trained in E episodes, each with T time steps. In each time step t, the training process is divided into interaction and network update processes. In the interaction process, the actor network chooses action according to the observed state $s[t]$ by exploration, as indicated in (30), where $\mathbf{a^d}[t] \triangleq \{o^d_u[t], \delta^d_u[t], \alpha^d_{u,s}[t], p^d_{u,k}[t], \pi^d_{u,k} \in [0,1], u \in \mathcal{U}, k \in \{1,2\}, s \in \mathcal{S}\}$. Accordingly, the channel allocation and decoding order variables are calculated. Then, we execute the constraint processing stage to the action, redesigning the offloading ratio and transmit power variables that satisfy the problem constraints. Subsequently, action $a[t]$ is performed onto the environment, reward $r[t]$ is obtained based on (29), and the environment updates the next state as $s[t+1]$. Subsequently, the experience sample (including $s[t]$, $a[t]$, $r[t]$, and $s[t+1]$) is stored in the replay buffer for training the networks. In the network update process, a batch of samples (D samples) is randomly selected from the replay buffer: s^D, a^D, r^D, s'^D to update the neural network parameters as described in (31), (32), and (34). Finally, the trained actor network is obtained for interacting with the environment.

C. Exhaustive-improved DRL

According to (39), the designed transmit powers are fractions of their maximum transmit power. However, during training, the DRL model considers a fixed transmit power, P_{u-max}, based on the setting environment parameters. However, unlike during model training, during the inference process in real operating situations, various devices may have different maximum transmit powers. That is, in the inference process, ignoring the change of P_{u-max} may degrade the system performance due to the lack of optimal power values. Therefore, by leveraging DT features, the device status can be obtained, enabling us to establish an exhaustive-improved DRL phase based on the exhaustive search strategy to determine suitable transmit powers in each state, described in Algorithm 3. In each interaction step, actions $\delta[t], \pi[t], \alpha[t]$ are designed by the trained DRL model as indicated in lines $5-7$ in Algorithm 2. Next, we search for the suitable value of the transmit power from P_{min} to P_{u-max} with a step of Δp, and implement the constraint processing stage. Here, the maximum transmit power value is replaced by the searched power value p_{es}, i.e., the transmit powers in (39) are recalculated as: $p_{u,1}[t] = p^m_{u,1}[t] p_{es}[t]$; $p_{u,2}[t] = p^m_{u,2}[t] \left(p_{es}[t] - p_{u,1}[t]\right)$. Subsequently, the DITEN system estimates the reward $r[t]$ with the decided action. After searching for all transmit power values, the optimal action $a^*[t]$, which yields the best reward, is obtained to interact with the physical environment.

Algorithm 3 Exhaustive-improved DRL phase

Input: Trained DRL model.
while interaction **do**
  Observe environment state $s[t]$.
  Set the best reward: $r_{best} = 0$.
  Design $\delta[t], \pi[t], \alpha[t]$ according to line $5-7$ in algorithm 2.
  for $p_{es} \leftarrow P_{min}$ to P_{u-max}, $p_{es} : p_{es} + \Delta p$ **do**
    Apply the constraint processing stage (equations (36)–(39)) with $P_{u-max} = p_{es} \rightarrow$ obtain $\mathbf{o}[t], \mathbf{p}[t]$.
    Estimate reward $r[t]$.
    if $r[t] \geq r_{best}$ **then**
      $a^*[t] \leftarrow a[t]$, $r_{best} \leftarrow r[t]$.
    end if
  end for
  return Optimal action $a^*[t]$.
end while

Remark 3. This exhaustive-improved DRL phase is applied to only the inference process and thus does not affect training performance.

IV. Proposed Mathematical Approximation based Solution: Relaxed LMI Framework

Owing to the non-linear objective function and mixed-integer constraints, problem (26) belongs to a class of mixed-integer nonlinear programming (MINLP). Moreover, (26) represents a combinatorial problem owing to the presence of variable $\pi[t]$. To overcome the challenges associated with these problem types, we first develop an optimization method involving two steps. The first step is to tackle the combinatorial variables, exploiting the power allocation in the SIC-enabled broadcasting system. The second step is to introduce the LMI framework with relaxation for integer variables $\{\mathbf{o}[t], \boldsymbol{\delta}[t], \boldsymbol{\alpha}[t]\}$.

A. Solution for combinatorial variables $\boldsymbol{\pi}[t]$

As mentioned earlier, the sub-message is decoded using the SIC technique, which requires a predefined decoding order. As the strongest signal must be prioritized in decoding, the

optimal solution for the decoding order is one in which the sub-messages transmitted through higher-gain channels are decoded and removed from the received signal before the other sub-messages are decoded through weaker-gain channels [39]–[41]. The optimal value is obtained considering the power allocation to be real-number variables.

Let $\boldsymbol{\pi}^*[t] \triangleq \{\bar{\boldsymbol{\pi}}^{\mathrm{s}}[t]\}_{s\in\mathcal{S}}$ be the solution of problem (26). Based on [39], we can determine element $\bar{\boldsymbol{\pi}}^{\mathrm{s}}[t]$ of $\boldsymbol{\pi}^*[t]$ as

$$\begin{aligned}\bar{\boldsymbol{\pi}}^{\mathrm{s}}[t] =\Big\{&\pi^{\mathrm{s}}_{u,k}[t] \in \{1,\cdots,|\mathcal{U}_s[t]|\times K\}| \\ &\pi^{\mathrm{s}}_{u,k}[t] < \pi^{\mathrm{s}}_{u',k'}[t] \ \wedge\ |h_{u,s}[t]|^2 \geq |h_{u',s}[t]|^2, \\ &\forall (u,k),(u',k') \in \mathcal{U}_s[t]\times\mathcal{K}\Big\}. \end{aligned} \tag{40}$$

Recall that the optimal decoding order can be obtained by sorting the channel gains in the broadcast system. Then, with the given decoding order, the resource allocation can be flexibly optimized by solving the following problem:

$$\begin{aligned}
\text{(P2):}\quad \max_{\hat{\mathcal{F}}[t]}\quad & \hat{\mathcal{V}}[t] \triangleq \mathcal{V}[t]|_{\boldsymbol{\pi}[t]=\boldsymbol{\pi}^*[t]} && \text{(41a)}\\
\text{s.t.}\quad & o_u[t] \in \left[0, \sum_{s\in\mathcal{S}} \alpha_{u,s}[t]\right], u\in\mathcal{U}, && \text{(41b)}\\
& \delta_u[t] \in [0,1], u\in\mathcal{U}, && \text{(41c)}\\
& \alpha_{u,s}[t] \in [0,1], u\in\mathcal{U}, s\in\mathcal{S}, && \text{(41d)}\\
& \sum_{s\in\mathcal{S}} \alpha_{u,s}[t] \leq 1, u\in\mathcal{U}, && \text{(41e)}\\
& \begin{cases} p_{u,k}[t] > 0 & \text{, if } z^o_{u,k}[t] > 0,\\ p_{u,k}[t] = 0 & \text{, otherwise.}\end{cases}, \forall k,\ \forall u, && \text{(41f)}\\
& \sum_{k\in\{1,2\}} p_{u,k}[t] \leq P_{u-\max}, u\in\mathcal{U}, && \text{(41g)}
\end{aligned}$$

where $\hat{\mathcal{F}}[t] \triangleq \big\{\mathcal{F}[t] \mid \boldsymbol{\pi}[t] = \boldsymbol{\pi}^*[t]\big\}$. Here, variable $\boldsymbol{\alpha}[t]$ in constraint (41d) is relaxed from binary values $\{0,1\}$ to the continuous interval $[0,1]$. Clearly, problem (41) is still non-convex owing to the non-convexity of the objective function and non-linear boundary condition of constraint (41f). Therefore, the analysis described in the remaining section is focused on addressing the difficulty in solving problem (41).

B. Solution for resource allocation variables $\{\mathbf{o}[t], \boldsymbol{\delta}[t], \boldsymbol{\alpha}[t], \boldsymbol{p}[t]\}$

1) Treating the objective function: First, we apply the Dinkelbach framework to transform (41) into a non-fractional form, such that the optimal value is obtained as

$$\hat{\mathcal{V}}^*[t] = \max_{\hat{\mathcal{F}}[t]} \Big\{\sum_{u\in\mathcal{U}} \mathcal{H}(d_u[t]-T_u[t]) - \mu^*[t]\sum_{u\in\mathcal{U}} E_u[t]\Big\} = 0, \tag{42}$$

if and only if the following condition is satisfied:

$$\mu^*[t] = \frac{\sum_{u\in\mathcal{U}} \mathcal{H}(d_u[t]-T^*_u[t])}{\sum_{u\in\mathcal{U}} E^*_u[t]} = \max_{\hat{\mathcal{F}}[t]}\quad \hat{\mathcal{V}}[t]. \tag{43}$$

To numerically determine $\mu^*[t]$, a finite number of iterations is executed, equivalent to solving the following problem:

$$\begin{aligned}\mu^{(n+1)}[t] =& \max_{\hat{\mathcal{F}}[t]} \Big\{\sum_{u\in\mathcal{U}} \mathcal{H}(d_u[t]-T_u[t]) \\ &- \mu^{(n)}[t]\sum_{u\in\mathcal{U}} E_u[t]\Big\}, \end{aligned} \tag{44}$$

a) Addressing $T_u[t]$: First, $T_u[t]$ is rewritten as

$$T_u[t] = \max\{T^l_u[t], T^o_{u,1}[t] + T^M_u[t], T^o_{u,2}[t] + T^M_u[t]\}. \tag{45}$$

Because $T^l_u[t]$ and $T^M_u[t]$ are linear, it is necessary to consider only the offloading term $T^o_{u,k}[t],\ k\in\mathcal{K}$. Clearly, $o_u[t]$, $\delta_u[t]$, and $1-\delta_u[t]$ are positive. Therefore, the variable can be modified as

$$\rho^2_{u,1}[t] = \delta_u[t] o_u[t] z_u[t], \tag{46}$$
$$\rho^2_{u,2}[t] = (1-\delta_u[t]) o_u[t] z_u[t]. \tag{47}$$

Then, $T^o_{u,k}[t]$ in (8) can be represented as linear variable $\beta_{u,k}[t]$ with coupled constraints:

$$\frac{\rho^2_{u,k}[t]}{\bar{r}_{u,k}[t]} \leq \beta_{u,k}[t], \tag{48a}$$
$$\bar{r}_{u,k}[t] \leq \sum_{s\in\mathcal{S}} r_{u,s,k}[t]. \tag{48b}$$

By applying the Schur complement, constraint (48) can be expressed as

$$\begin{bmatrix} \bar{r}_{u,k}[t] & \rho_{u,k}[t] \\ \rho_{u,k}[t] & \beta_{u,k}[t]\end{bmatrix} \succeq \mathbf{0} \tag{49a}$$
$$\sum\nolimits_{s\in\mathcal{S}} r_{u,s,k}[t] \geq \bar{r}_{u,k}[t]. \tag{49b}$$

Evidently, this constraint is convex, and $T_u[t]$ is completely addressed.

b) Addressing $E_u[t]$: Consider $E_u[t]$ in (19). The energy required for local computation (first term) is linear, whereas the energy required for offloading operation (second term) is non-convex. To convexify $E_u[t]$, the expression under the summation can be rewritten as

$$p_{u,k}[t]\frac{z^o_{u,k}[t]}{\sum_{s\in\mathcal{S}} r_{u,s,k}[t]} = p_{u,k}[t]\frac{\rho^2_{u,k}[t]}{\bar{r}_{u,k}[t]} \overset{(i)}{\leq} \varphi_{u,k}[t]. \tag{50}$$

Then, the inequality *(i)* in (50) can be treated as

$$\frac{p_{u,k}[t]}{\bar{\beta}_{u,k}[t]} \leq \varphi_{u,k}[t], \tag{51}$$

where the smooth variable $\bar{\beta}_{u,k}[t] = \frac{\bar{r}_{u,k}[t]}{\rho^2_{u,k}[t]}$. Subsequently, $E_u[t]$ can be transformed into the linear form, proportional to variable $\varphi_{u,k}[t]$ under additional convex constraints:

$$\begin{bmatrix} \bar{\beta}_{u,k}[t] & \bar{p}_{u,k}[t] \\ \bar{p}_{u,k}[t] & \varphi_{u,k}[t]\end{bmatrix} \succeq \mathbf{0}, \tag{52}$$

with the power allocation being reconstructed as $p_{u,k}[t] = \bar{p}^2_{u,k}[t]$.

c) Convex objective function: By substituting (49) and (52) into (44), we can obtain the convex form of the objective function using the Dinkelbach algorithm, i.e.,

$$\mu^{(n+1)}[t] = \max_{\mathcal{D}[t]} \Big\{ \sum_{u\in\mathcal{U}} \mathcal{H}(d_u[t]-\omega_u[t]) - \mu^{(n)}[t] \sum_{u\in\mathcal{U}}\sum_{k\in\mathcal{K}} \varphi_{u,k}[t] \Big\}, \quad (53)$$

where the feasible region $\mathcal{D}[t]$ is determined as

$$\mathcal{D}[t] \triangleq \left\{ \hat{\mathcal{F}}[t] \cup \tilde{\mathcal{F}}[t] \middle| \text{Constraints } (\mathcal{C}) \text{ are satisfied.} \right\}, \quad (54)$$

with $\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2$, $\mathcal{C}_1 = \{(41b)-(41g)\}$, and

$$\mathcal{C}_2 = \begin{Bmatrix} T_u^l[t] \le \omega_u[t] \\ T_{u,1}^o[t] + T_u^M[t] \le \omega_u[t] \\ T_{u,2}^o[t] + T_u^M[t] \le \omega_u[t] \\ (49), (52) \end{Bmatrix}. \quad (55)$$

The auxiliary-variable set for convexifying the objective function is

$$\tilde{\mathcal{F}}[t] \triangleq \Big\{ \{\omega_u[t]\}, \{\varphi_{u,k}[t]\}, \{\rho_{u,k}\}, \{\bar{r}_{u,k}\}, \{\beta_{u,k}\}, \{\bar{\beta}_{u,k}[t]\}, \{\bar{p}_{u,k}\} \Big\}, \forall u \in \mathcal{U},\ \forall k \in \mathcal{K}.$$

Although the objective function is now linear, the feasible set $(\mathcal{C})$ is non-convex owing to the non-convexity constraint (41f) in $(\mathcal{C}_1)$. Hence, the final part of the analysis is to address this constraint.

2) Convexifying the feasible set ($\mathcal{C}_1$): We outline the steps for transforming $(\mathcal{C}_1)$ into the LMI form. To this end, the following vector variable is defined:

$$\boldsymbol{\theta}[t] \triangleq \left[(\mathbf{o}[t])^T\ (\boldsymbol{\delta}[t])^T\ \mathrm{vec}(\boldsymbol{\alpha}[t])^T \right]^T \in [0,1]^{(S+2)U\times 1}. \quad (56)$$

Next, the constraints of MEC-resource and channel allocation (41b)– (41e) in $(\mathcal{C}_1)$ are expressed as

$$o_u[t] \ge 0, u \in \mathcal{U}, \quad (57a)$$
$$\delta_u[t] \ge 0,\ u \in \mathcal{U}, \quad (57b)$$
$$\alpha_{u,s}[t] \ge 0,\ u \in \mathcal{U}, s \in \mathcal{S}, \quad (57c)$$
$$o_u[t] - \sum_{s\in\mathcal{S}} \alpha_{u,s}[t] \le 0, u \in \mathcal{U}, \quad (57d)$$
$$\delta_u[t] \le 1,\ u \in \mathcal{U}, \quad (57e)$$
$$\alpha_{u,s}[t] \le 1,\ u \in \mathcal{U}, s \in \mathcal{S}, \quad (57f)$$
$$\sum_{s\in\mathcal{S}} \alpha_{u,s}[t] \le 1, u \in \mathcal{U}. \quad (57g)$$

In a straightforward manner, we can rewrite (57) in the LMI form as follows:

$$\mathbf{A}_1.\boldsymbol{\theta}[t] \preceq \mathbf{b}_1, \quad (58)$$

where matrix $\mathbf{A}_1$ and vector $\mathbf{b}_1$ are constructed as

$$\mathbf{A}_1 = \begin{bmatrix} -\mathbf{I}_U & \mathbf{0}_{U\times U} & \mathbf{0}_{U\times S.U} \\ \mathbf{0}_{U\times U} & -\mathbf{I}_U & \mathbf{0}_{U\times S.U} \\ \mathbf{0}_{S.U\times U} & \mathbf{0}_{S.U\times U} & -\mathbf{I}_{S.U} \\ \mathbf{I}_U & \mathbf{0}_{U\times U} & -\mathbf{1}_{1\times S}\otimes\mathbf{I}_U \\ \mathbf{0}_{U\times U} & \mathbf{I}_U & \mathbf{0}_{U\times S.U} \\ \mathbf{0}_{S.U\times U} & \mathbf{0}_{S.U\times U} & \mathbf{I}_{S.U} \\ \mathbf{0}_{U\times U} & \mathbf{0}_{U\times U} & \mathbf{1}_{1\times S}\otimes\mathbf{I}_U \end{bmatrix}, \quad (59)$$

$$\mathbf{b}_1 = \begin{bmatrix} \mathbf{0}_{1\times(S+3)U} & \mathbf{1}_{1\times(S+2)U} \end{bmatrix}^T. \quad (60)$$

The remaining challenge is to address constraint (41f). In this context, the limit of power coefficient variables depends on the value of $z_{u,k}^o[t]$, specified as the compound of two other variables $\boldsymbol{\delta}[t]$ and $\mathbf{o}[t]$. To overcome this issue, we transform constraints for power allocation (41f)–(41g) (for $K=2$) as follows:

$$p_{u,s,1}[t] \le \delta_u[t] P_{u-\max},\ \forall u \in \mathcal{U},\ \forall s \in \mathcal{S}, \quad (61a)$$
$$p_{u,s,2}[t] \le (1-\delta_u[t]) P_{u-\max},\ \forall u \in \mathcal{U},\ \forall s \in \mathcal{S}, \quad (61b)$$
$$\sum_k p_{u,s,k}[t] \le \alpha_{u,s}[t] P_{u-\max},\ \forall u \in \mathcal{U},\ \forall s \in \mathcal{S}, \quad (61c)$$
$$\sum_{s\in\mathcal{S}}\sum_{k\in\mathcal{K}} p_{u,s,k} \le o_u[t] P_{u-\max},\ \forall u \in \mathcal{U}, \quad (61d)$$

which is equivalent to the following LMI expression:

$$\begin{bmatrix} \mathbf{A}_2 & \mathbf{C} \end{bmatrix} . \begin{bmatrix} \boldsymbol{\theta}[t] \\ \boldsymbol{\varrho}[t] \end{bmatrix} \preceq \mathbf{b}_2, \quad (62)$$

with $\boldsymbol{\varrho}[t] \triangleq \mathrm{vec}(\mathbf{p}[t])$, and

$$\mathbf{A}_2 = P_{u-\max} \begin{bmatrix} \mathbf{0}_{S.U\times U} & -\big(\mathbf{1}_{S\times 1}\otimes\mathbf{I}_U\big) & \mathbf{0}_{S.U\times S.U} \\ \mathbf{0}_{S.U\times U} & \big(\mathbf{1}_{S\times 1}\otimes\mathbf{I}_U\big) & \mathbf{0}_{S.U\times S.U} \\ \mathbf{0}_{S.U\times U} & \mathbf{0}_{S.U\times U} & -\big(\mathbf{I}_S\otimes\mathbf{I}_U\big) \\ -\mathbf{I}_U & \mathbf{0}_{U\times U} & \mathbf{0}_{U\times S.U} \end{bmatrix} \quad (63)$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{I}_S\otimes\big([1\ 0]\otimes\mathbf{I}_U\big) \\ \mathbf{I}_S\otimes\big([0\ 1]\otimes\mathbf{I}_U\big) \\ \mathbf{I}_S\otimes\big([1\ 1]\otimes\mathbf{I}_U\big) \\ \mathbf{1}_{1\times S.K}\otimes\mathbf{I}_U \end{bmatrix} \in \{0,1\}^{(3S+1)U\times S.K.U}, \quad (64)$$

$$\mathbf{b}_2 = P_{u-\max}\begin{bmatrix} \mathbf{0}_{1\times S.U} & \mathbf{1}_{1\times S.U} & \mathbf{0}_{1\times(S+1).U} \end{bmatrix}^T. \quad (65)$$

From (58) and (62), the constraints in $(\mathcal{C}_1)$ are equivalent to the following LMI form (denoted by $\hat{\mathcal{C}}_1 \subseteq \mathcal{C}_1$):

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{0}_{(S+5)U\times S.K.U} \\ \mathbf{A}_2 & \mathbf{C} \end{bmatrix} . \begin{bmatrix} \boldsymbol{\theta}[t] \\ \boldsymbol{\varrho}[t] \end{bmatrix} \preceq \mathbf{b}, \quad (66)$$

where $\mathbf{b} = [\mathbf{b}_1^T\ \mathbf{b}_2^T]^T$.

Finally, the equivalent convex optimization problem for solving (41) is formulated as

$$\mu^{(n+1)}[t] = \max_{\hat{\mathcal{D}}[t]} \Big\{ \sum_{u\in\mathcal{U}} \mathcal{H}(d_u[t]-\omega_u[t]) - \mu^{(n)}[t] \sum_{u\in\mathcal{U}}\sum_{k\in\mathcal{K}} \varphi_{u,k}[t] | \Big\}, \quad (67)$$

where

$$\hat{\mathcal{D}}[t] \triangleq \left\{ \hat{\mathcal{F}}[t] \cup \tilde{\mathcal{F}}[t] \middle| \hat{\mathcal{C}}_1 \cap \mathcal{C}_2 \text{ is satisfied.} \right\}. \quad (68)$$

Problem (67) can be conveniently solved using convex optimization tools. Notably, $\hat{\mathcal{C}}_1$ is a convex subset of $\mathcal{C}_1$, and then, the convex subset $(\hat{\mathcal{C}}_1 \cap \mathcal{C}_2) \subseteq (\mathcal{C}_1 \cap \mathcal{C}_2)$. It follows that every feasible point in $\hat{\mathcal{D}}[t]$ for problem (67) is feasible in $\mathcal{D}[t]$ for the Dinkelbach-form problem (53). Moreover, $(\hat{\mathcal{C}}_1 \cap \mathcal{C}_2) \subseteq (\mathcal{C}_1 \cap \mathcal{C}_2) \subseteq \mathcal{C}_1$, and according to the iterative Dinkelbach algorithm for the combinatorial problem, the optimal solution for problem (53) is a sub-optimal solution for the original problem (26).

Remark 4. Building upon the reported results [39]–[41], the proposed solution for solving problem (26) within the scope of this work may be more efficient than state-of-the-art methods, i.e., SCA or block successive upper minimization (BSUM). This can be attributed to two reasons: (a) The LMI-based solution has a convex form, where the sub-problems following Dinkelbach's framework can be linearly handled by the available convex packages; and (b) SCA methods require continuous convex feasible subsets. Thus, smooth variables and approximation constraints are frequently introduced, resulting in high complexity, especially for combinatorial problems.

TABLE III
ENVIRONMENTAL PARAMETERS

Parameter	Value
U	20
z_u	$100 - 500$ Kbits
c_u	$700 - 1000$ cycles/bit
d_u	0.35 s
S	3
B_s	$(B_1, B_2, B_3) = (10, 30, 60)$ MHz
NF	5 dB
σ^2	$-174 + \text{NF} + 10\log_{10}(B_s)$ dBm
$\tilde{f}_u^l$	10^9 cycles/s
$\tilde{f}_u^e$	5.10^9 cycle/s
$\hat{f}_u^l, \hat{f}_u^e$	10 % of estimated values
κ_u	10^{-28}

V. PERFORMANCE EVALUATION

We simulate scenarios in which the BS serves IoTDs randomly distributed within a 10-to-200-m radius of the BS. The channel response coefficients between the IoTDs and BS are generated as [42]

$$h_{u,s} = \hat{h}_{u,s}\sqrt{h_{u,s}^L}, \quad (69)$$

where $\hat{h}_{u,s}$ is the small-scale fading, with the distribution $\mathcal{CN}(0,1)$; and $h_{u,s}^L$ represents the large-scale fading between the BS and IoTD u, specified as

$$h_{u,s}^L = 10^{\frac{L(d)+n_{sh}\sigma_{sh}}{10}}, \quad (70)$$

where $\sigma_{sh} = 8$ (dB), $n_{sh} \sim \mathcal{N}(0,1)$, and $L(d)$ is the path loss (in dB) between BS and IoTD u, calculated as [42]

$$\begin{aligned} L(d) = &- 140.7 - 35\log_{10}(d) \\ &+ 20c_0\log_{10}\left(\frac{d}{d_0}\right) + 15c_1\log_{10}\left(\frac{d}{d_1}\right). \end{aligned} \quad (71)$$

Other environmental simulation parameters are summarized in Table III, where the parameters are chosen from [43]–[45]. The neural networks in the DRL algorithm have two hidden layers, each has 512 nodes. The replay buffer has size of 10^5, other parameters are set as $\gamma = 0.99$, $\tau = 0.01$, and batch size $D = 16$. The following schemes are considered in this comparative analysis:

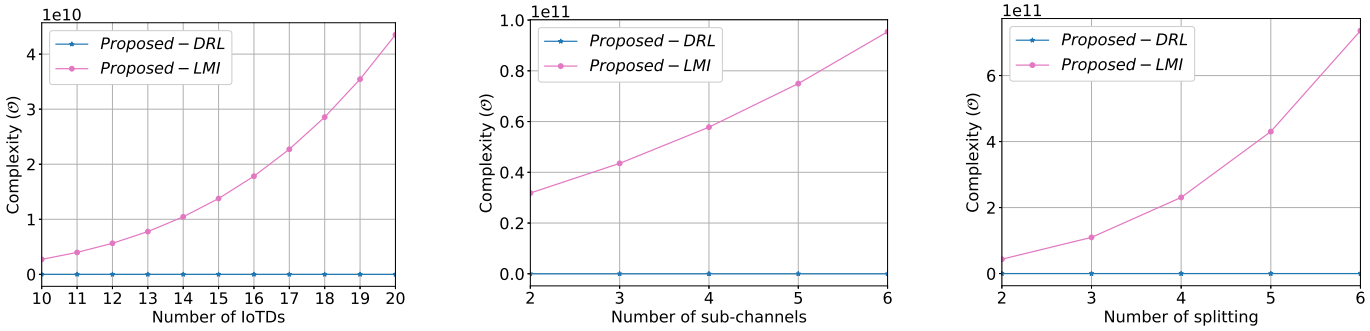
- *Proposed-DRL*: Proposed approach using the DRL framework, as outlined in Section III. This approach does not contain the exhaustive-improved DRL phase.
- *Proposed-EID*: Proposed DRL scheme with the exhaustive-improved DRL phase. We set $P_{min} = -30$ (dBm) and $\Delta p = 1$ (dBm) for searching the suitable transmit power in each step.
- *Proposed-LMI*: Proposed approach using the sup-optimal solution, defined in Section IV.
- *RSMA*: Only the RSMA technique is used for uplink communication between IoTDs and the BS. This multiple-access technique has been considered in numerous existing studies [9], [32].
- *NOMA*: According to [26], this scheme applies NOMA as the multiple access technique at each sub-channel.
- *FDMA*: The FDMA technique is applied to communication between IoTDs and BS, with the communication channel divided into ten sub-channels with the same resources, each serving two IoTDs at a time.
- *FMA-GS (Proposed orthogonalized RSMA-based flexible multiple access scheme with greedy searching)*: Proposed multiple-access scheme. To solve the optimization problem, we discretize the variables and use a greedy-based local search algorithm, described in [46].

A. Complexity

1) Complexity of the proposed DRL approach: We evaluate the computational complexity of the proposed DRL algorithm, factoring in the neural networks and supplementary processing functions. Given that training occurs at DITEN, we equate the complexity of the neural networks with that of the primary actor network utilized during inference. According to [47], [48], the computational complexity of the primary actor network is calculated as $\mathcal{O}_\mu = \mathcal{O}\left(\sum_{l=1}^{L}\psi_{l-1}\psi_l\right)$, where ψ_l denotes the number of nodes in layer l of the neural network, with $l = 0$ and $l = L$ being the input and output layers, respectively. According to (27) and (28), the number of nodes in the input and output layers are calculated as

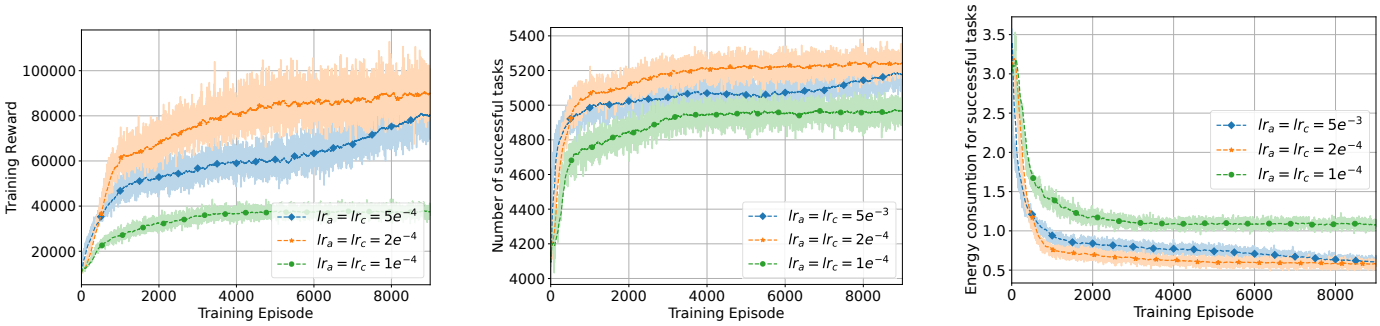
$$\psi_0 = U * (3 + 2S); \quad \psi_L = U * (6 + S). \quad (72)$$

According to (35), we normalize the channel allocation variables by calculating argmax over S entries for $U*S$ variables. Then, its computational complexity is $\mathcal{O}_\alpha = \mathcal{O}(U * S^2)$. For determining the decoding order indicated in Algorithm 1, the sorting operation has a complexity of $\mathcal{O}(U_s \log U_s)$ [49]. Accordingly, the computational complexity for determining the decoding order is $\mathcal{O}_\pi = \mathcal{O}(\sum_{s\in\mathcal{S}}(U_s \log U_s + U_s * K))$. Additionally, the computational complexity for the constraint processing stage can be calculated as $\mathcal{O}_c = \mathcal{O}(U * S)$, where



(a) Number of IoTDs. (b) Number of sub-channels. (c) Number of splittings.

Fig. 3. Complexity of the algorithms according to environment scale.



(a) Training reward. (b) Number of successful task. (c) Energy consumption for successful tasks.

Fig. 4. Training results with different learning rates.

the operations in (37), (38), and (39) have constant complexity. Thus, the proposed DRL algorithm has a computational complexity of $\mathcal{O}_{DRL} = \mathcal{O}_{\mu} + \mathcal{O}_{\alpha} + \mathcal{O}_{\pi} + \mathcal{O}_{c}$.

2) Complexity of the proposed mathematical approximation approach: The complexity of the algorithm for solving problem (26) is determined by the number of variables and constraints in $\hat{\mathcal{D}}[t]$. The number of variables is $\hat{\mathcal{F}}[t] \cup \tilde{\mathcal{F}}[t]$, with the following specifications:

- Variable set $\hat{\mathcal{F}}$: This set includes $U \times (S+4)$ variables, related to the variable set $\mathbf{o}[t], \boldsymbol{\delta}[t], \boldsymbol{\alpha}[t]$, and $\mathbf{p}[t]$.
- Variable set $\tilde{\mathcal{F}}[t]$: This set requires $U \times (6K+1)$ new variables.

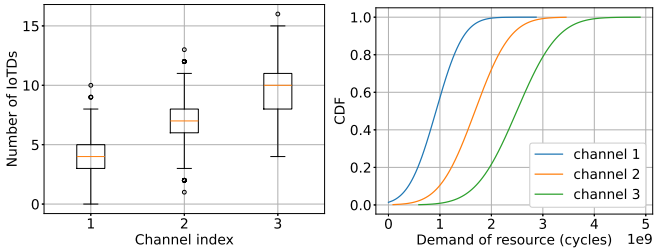
In total, the number of variables is $\mathtt{V} = U \times (S + 6K + 5)$.

The number of constraints is determined by $\hat{\mathcal{C}}_1 \cap \mathcal{C}_2$, which is equivalent to the summation of the number of constraints in (55) and total number of rows of matrices $\mathbf{A}_1$ and $\mathbf{A}_2$. From (55) and (66), the total number of constraints is calculated to be $\mathtt{C} = U \times (5S + 5K + 9)$. According to [50], the LMI-based algorithm has a computational complexity of $\mathcal{O}(\mathtt{V}^3\mathtt{C})$.

3) Complexity comparison: The complexity values of the *Proposed-DRL* and the *Proposed-LMI* with environment growth, i.e., the number of IoTDs, are $U \log U$ and U^4, respectively. As shown in Fig. 3, the increase in the complexity of the *Proposed-DRL* scheme is minor with the rise in the scale of the environment, i.e., the number of IoTDs, sub-channels, and splittings. In contrast, the complexity of the *Proposed-LMI* scheme rapidly increases. That is, we can say that the proposed DRL approach can be applied more practically in realistic environments, whereas applying the mathematical approach requires the use of a hyper-powerful computer. However, all the proposed algorithms have polynomial complexity, i.e., scalable algorithms. In Table IV, we summarize the complexity of the algorithms according to environment parameters.

TABLE IV
COMPLEXITY COMPARISON

	Number of IoTDs	Number of sub-channels	Number of splitting
Proposed-DRL	$\mathcal{O}(U \log U)$	$\mathcal{O}(S^2)$	$\mathcal{O}(K)$
Proposed-LMI	$\mathcal{O}(U^4)$	$\mathcal{O}(S^4)$	$\mathcal{O}(K^4)$



(a) Distribution of num. of IoTDs. (b) CDF of computational resource demand.

Fig. 5. Statistics on traffic usage of sub-channels.

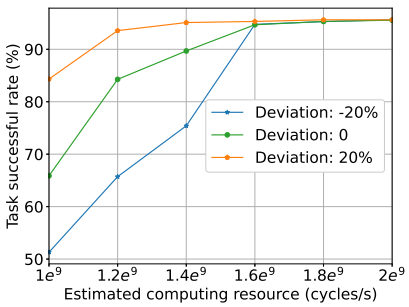


Fig. 6. Task success rate for different DT deviations.

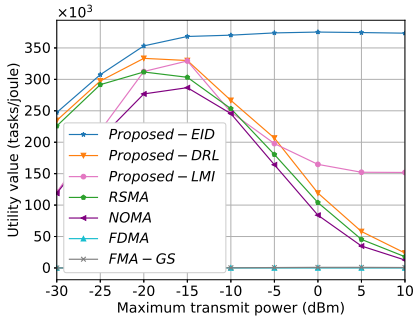


Fig. 7. Performance against different maximum transmit power values.

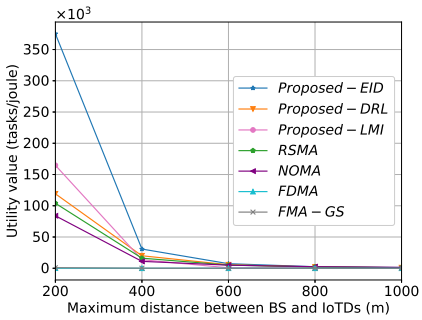


Fig. 8. Performance in different maximum distance values.

B. Numerical Experiments

First, we evaluate the convergence of the DRL training algorithm by training the model using different learning rates. As mentioned in Remark 3, the exhaustive-improved DRL does not affect the training process. Thus, we skip this phase in this analysis. Experiments are conducted with three actor learning rates (lr_a) and critic learning rates (lr_c): $lr_a = lr_c = \{5e^{-4}, 2e^{-4}, 1e^{-4}\}$. As illustrated in Fig. 4, case $lr_a = lr_c = 2e^{-4}$ yields the best result, where the model converges after approximately 8000 episodes, and the reward is approximately 18.75 % and 140 % superior than those when the rates are $5e^{-4}$ and $1e^{-4}$, respectively. Additionally, the reward increases with the number of successful tasks and energy consumption to complete the tasks reduces, as shown in Figs. 4b and 4c, respectively.

Second, we estimate the user allocation on the sub-channels, as illustrated in Fig. 5a. In this scenario, we model three sub-channels indexed 1, 2, and 3, with the communication resource increased from 1 to 3 and evaluate the number of users access to each channel. The results indicate that channel 3 serves the highest number of users in each step, approximately 2.6 times and 1.4 times higher than the values for channels 1 and 2, respectively. This suits the system property where the higher-capacity channel can serve more users. Furthermore, we evaluate the distribution of users' resource demand on each channel by calculating the cumulative distribution function (CDF) in 3000 testing steps. As shown in Fig. 5b, channels with higher communication bandwidth serve users with higher demand. Specifically, the mean resource demands of users on channels 1, 2, and 3 are approximately $0.92e^9$, $1.68e^9$, and $2.49e^9$ cycles per step, respectively. The results show that our goal of dividing the communication bandwidth is reasonable, as users with high service demand use the channel with more communication resources.

The performance of the proposed framework is assessed in different DT systems, in terms of the task success rate corresponding to three values of DT deviations: -20 %, 0 %, and 20 % with different MEC capacities. As illustrated in Fig. 6, the deviation of 20 % yields the best result, with the most inferior result obtained when -20 %. The deviation of 20 % indicates that the computing resources in the physical devices are 20 % higher than the estimated values in the DT layer. Therefore, with positive deviations, the physical layer performs better than expected from the DT layer, owing to the higher computing resources. In the case of negative deviations, it performs worse than expected from the DT layer. In addition, the system performs well in all three cases of deviation when the computing resource at the MEC are adequate (higher than $1.6e^9$ cycles/s), demonstrating the stability of the system. Therefore, in the considered scenario, with $\tilde{f}_u^e = 5e^9$ cycles/s (from the Table 1), the system appears to be effective across different DT systems, demonstrating the excellent system performance.

Furthermore, we assess the system performance by the utility value. As illustrated in Fig. 7, we measure the utility value over different IoTD transmit powers, varying from -30 to 10 dBm. The utility value has a quasi-concave function of P_{u-max}, where the peak lies in the interval $[-20, -15]$ (dBm). The proposed approaches outperform other benchmark schemes in all cases, with the value for *Proposed-DRL* being approximately 9.7% and 30.7% higher than those for *RSMA* and *NOMA*, respectively, and it significantly outperforms *FDMA*. The *FMA-GS* scheme exhibits inferior performance because the high complexity of the problem causes the greedy-based search algorithm to become stuck in an inferior state. Among the proposed algorithms, notably, the *Proposed-EID* scheme outperforms the *Proposed-DRL* scheme. That is, the proposed exhaustive-improved DRL phase significantly improves the performance of the *Proposed-DRL* algorithm. Its utility value is stable when the maximum transmit power value is greater than -10 dBm. This is because *Proposed-EID* uses the appropriate power values for offloading the messages by searching the transmit power values even as the maximum transmit power increases. In contrast, the *Proposed-DRL* scheme uses a fraction of the maximum transmit power for offloading, resulting in increased energy consumption as the maximum transmit power increases, reducing the utility value. The *Proposed-LMI* approach exhibits a stable performance, retaining the highest output compared with other schemes when the maximum transmit power is higher than 0 dBm. As a result, the proposed orthogonalized RSMA scheme shows its effectiveness. Here, dividing the communication bandwidth can reduce inter-user interference by using different orthogonal sub-channels. Besides, the RSMA-based schemes demonstrated outperformance in transmission compared to the NOMA-based scheme, which is similar to the results in

previous studies [8], [9].

Next, we evaluate the system performance according to the change in distance between BS and IoTDs, with $P_{u-max} = 0$ (dBm). The maximum distance increases from 200 m to 1000 m. As shown in Fig. 8, the utility value decreases with the rise in distance, as increasing the distance reduces the large-scale fading, as indicated in (70), thereby decreasing the gain of the communication channel. The *Proposed-LMI* approach outperforms the *Proposed-DRL* approach in the first case, i.e., 200 m, and significantly decreases with higher distance because it has to keep a convex feasible set during iteration. As a result, the *Proposed-LMI* approach performs excellently in permissive environments, such as allowing high transmit power and channel gain. On the other hand, the *Proposed-DRL* approach performs better in strict environments. Significantly, the *Proposed-EID* demonstrates its outperformance when performing best in all cases, indicating the effectiveness of the proposed exhaustive-improved DRL phase. Notably, increasing the maximum distance between the BS and IoTDs briskly reduces the utility value. The increase in path loss according to the distance leads to a stricter environment, which reduces the number of successful tasks. Therefore, deploying the BS or access points' position in practical environments should be carefully considered.

Remark 5. Although the numerical results proved our proposed approach's effectiveness, this study opens up some points for further discussion in future research. First, the proposed RSMA-based scheme splits the transmitted message into two sub-messages. However, splitting more sub-messages may improve the spectral efficiency due to the higher decoding order diversity [51], [52]. Thus, splitting more sub-messages in the proposed scheme is worth considering. Second, fixing bandwidth may reduce the flexibility of the proposed orthogonalized RSMA scheme. Therefore, more flexible communication sub-channels can be studied to further improve the system's performance.

VI. Conclusion

This study proposed a flexible and efficient access control scheme combining RSMA and OFDMA techniques in a DITEN system. We formulated a mixed-integer non-convex optimization problem that jointly maximizes the number of successful tasks and minimizes the energy consumption of IoTDs under the constraints of the offloading operation, splitting ratios, channel allocation, IoTD transmit powers, and decoding orders. To solve this problem, we first proposed a DRL framework by developing a normalized DDPG algorithm. Further, we improved it by introducing an exhaustive-improved DRL phase. We also proposed an established mathematical optimization-based solution by applying the relaxed LMI framework. We evaluated the proposed algorithms under different environments and parameters and confirmed that the proposed controls converge stably and provide much more enhanced performance than the benchmark schemes.

In future work, more advanced flexible multiple access schemes with flexible sub-message and sub-channel configurations and new components, such as intelligent reflecting surfaces in DITEN, can be studied further.

References

[1] C.-X. Wang, X. You, X. Gao, X. Zhu, Z. Li, C. Zhang, H. Wang, Y. Huang, Y. Chen, H. Haas, J. S. Thompson, E. G. Larsson, M. D. Renzo, W. Tong, P. Zhu, X. Shen, H. V. Poor, and L. Hanzo, "On the road to 6G: Visions, requirements, key technologies, and testbeds," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 905–974, 2023.

[2] S. Mihai, M. Yaqoob, D. V. Hung, W. Davis, P. Towakel, M. Raza, M. Karamanoglu, B. Barn, D. Shetve, R. V. Prasad, H. Venkataraman, R. Trestian, and H. X. Nguyen, "Digital twins: A survey on enabling technologies, challenges, trends and future prospects," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 4, pp. 2255–2291, 2022.

[3] F. Tang, X. Chen, T. K. Rodrigues, M. Zhao, and N. Kato, "Survey on digital twin edge networks (DITEN) toward 6G," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1360–1381, 2022.

[4] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the gaussian multiple-access channel," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 364–375, 1996.

[5] Y. Mao, O. Dizdar, B. Clerckx, R. Schober, P. Popovski, and H. V. Poor, "Rate-splitting multiple access: Fundamentals, survey, and future research trends," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 4, pp. 2073–2126, 2022.

[6] G. Zhou, Y. Mao, and B. Clerckx, "Rate-splitting multiple access for multi-antenna downlink communication systems: Spectral and energy efficiency tradeoff," *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 4816–4828, 2022.

[7] B. Clerckx, Y. Mao, E. A. Jorswieck, J. Yuan, D. J. Love, E. Erkip, and D. Niyato, "A primer on rate-splitting multiple access: Tutorial, myths, and frequently asked questions," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1265–1308, 2023.

[8] O. Abbasi and H. Yanikomeroglu, "Transmission scheme, detection and power allocation for uplink user cooperation with NOMA and RSMA," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 471–485, 2023 .

[9] Z. Yang, M. Chen, W. Saad, W. Xu, and M. Shikh-Bahaei, "Sum-rate maximization of uplink rate splitting multiple access (RSMA) communication," *IEEE Transactions on Mobile Computing*, vol. 21, no. 7, pp. 2596–2609, 2022.

[10] 3GPP, "Study on support of reduced capability NR devices," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.875, March 2021.

[11] M. Vaezi, A. Azari, S. R. Khosravirad, M. Shirvanimoghaddam, M. M. Azari, D. Chasaki, and P. Popovski, "Cellular, wide-area, and non-terrestrial iot: A survey on 5G advances and the road toward 6G," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1117–1174, 2022.

[12] D. Vargas and Y. J. D. Kim, "Two-layered superposition of broadcast/multicast and unicast signals in multiuser OFDMA systems," *IEEE Transactions on Wireless Communications*, 2020, volume=19, number=2, pages=979-994.

[13] Z. Liu and S. Feng, "Joint subcarrier assignment and power allocation for ofdma full duplex distributed antenna systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 11 554–11 564, 2021.

[14] T. P. Truong, T. M. Tuyen Nguyen, T.-V. Nguyen, N.-N. Dao, and S. Cho, "RSMA for uplink MIMO systems: DRL-based achievable system sum rate maximization," in *2023 IEEE Globecom Workshops (GC Wkshps)*, 2023, pp. 878–883.

[15] J. Huang, Y. Yang, J. Lee, D. He, and Y. Li, "Deep reinforcement learning-based resource allocation for RSMA in LEO satellite-terrestrial networks," *IEEE Transactions on Communications*, vol. 72, no. 3, pp. 1341–1354, 2024.

[16] N.-P. Tran, T. P. Truong, Q. T. Do, N.-N. Dao, and S. Cho, "Joint wireless resource allocation and bitrate adaptation for QoE improvement in IRS-aided RSMA-enabled IoMT streaming systems," *Internet of Things*, p. 101145, 2024.

[17] M. Diamanti, G. Kapsalis, E. E. Tsiropoulou, and S. Papavassiliou, "Energy-efficient rate-splitting multiple access: A deep reinforcement learning-based framework," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 2397–2409, 2023 .

[18] L. Tan, Z. Kuang, L. Zhao, and A. Liu, "Energy-efficient joint task offloading and resource allocation in OFDMA-based collaborative edge computing," *IEEE Transactions on Wireless Communications*, vol. 21, no. 3, pp. 1960–1972, 2022.

[19] P. Chen, H. Liu, Y. Ye, L. Yang, K. J. Kim, and T. A. Tsiftsis, "Rate-splitting multiple access aided mobile edge computing with randomly deployed users," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1549–1565, 2023.

[20] M. Diamanti, C. Pelekis, E. E. Tsiropoulou, and S. Papavassiliou, "Delay minimization for rate-splitting multiple access-based multi-server MEC offloading," *IEEE/ACM Transactions on Networking*, pp. 1–13, 2023.

[21] T. Liu, L. Tang, W. Wang, Q. Chen, and X. Zeng, "Digital-twin-assisted task offloading based on edge collaboration in the digital twin edge network," *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1427–1444, 2022.

[22] Y. Hao, J. Wang, D. Huo, N. Guizani, L. Hu, and M. Chen, "Digital twin-assisted URLLC-enabled task offloading in mobile edge network via robust combinatorial optimization," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 10, pp. 3022–3033, 2023.

[23] Y. Chen, W. Gu, J. Xu, Y. Zhang, and G. Min, "Dynamic task offloading for digital twin-empowered mobile edge computing via deep reinforcement learning," *China Communications*, vol. 20, no. 11, pp. 164–175, 2023.

[24] Y. Dai, K. Zhang, S. Maharjan, and Y. Zhang, "Deep reinforcement learning for stochastic computation offloading in digital twin networks," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 4968–4977, 2021.

[25] Y. Dai, J. Zhao, J. Zhang, Y. Zhang, and T. Jiang, "Federated deep reinforcement learning for task offloading in digital twin edge networks," *IEEE Transactions on Network Science and Engineering*, vol. 11, no. 3, pp. 2849–2863, 2024.

[26] L. Zhang, H. Wang, H. Xue, H. Zhang, Q. Liu, D. Niyato, and Z. Han, "Digital twin-assisted edge computation offloading in industrial internet of things with NOMA," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 9, pp. 11 935–11 950, 2023.

[27] B. Li, Y. Liu, L. Tan, H. Pan, and Y. Zhang, "Digital twin assisted task offloading for aerial edge computing and networks," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 10, pp. 10 863–10 877, 2022.

[28] J. Tang, J. Luo, J. Ou, X. Zhang, N. Zhao, D. K. C. So, and K.-K. Wong, "Decoupling or learning: Joint power splitting and allocation in MC-NOMA with SWIPT," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5834–5848, 2020.

[29] Z. Ding, D. Xu, R. Schober, and H. V. Poor, "Hybrid NOMA offloading in multi-user MEC networks," *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 5377–5391, 2022.

[30] K. Wang, H. Li, Z. Ding, and P. Xiao, "Reinforcement learning based latency minimization in secure NOMA-MEC systems with hybrid SIC," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 408–422, 2023.

[31] Y.-C. Wu, C. Lin, and T. Q. S. Quek, "A robust distributed hierarchical online learning approach for dynamic MEC networks," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 2, pp. 641–656, 2022.

[32] T. P. Truong, N.-N. Dao, and S. Cho, "HAMEC-RSMA: Enhanced aerial computing systems with rate splitting multiple access," *IEEE Access*, vol. 10, pp. 52 398–52 409, 2022.

[33] M. Katwe, K. Singh, B. Clerckx, and C.-P. Li, "Rate splitting multiple access for sum-rate maximization in IRS aided uplink communications," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2022.

[34] W. Sun, H. Zhang, R. Wang, and Y. Zhang, "Reducing offloading latency for digital twin edge networks in 6G," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 12 240–12 251, 2020.

[35] W. Sun, S. Lei, L. Wang, Z. Liu, and Y. Zhang, "Adaptive federated learning and digital twin for industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5605–5614, 2021.

[36] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* The MIT Press, 2018.

[37] 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." in *ICLR (Poster)*, 2016.

[38] G. E. Uhlenbeck and L. S. Ornstein, "On the theory of the brownian motion," *Phys. Rev.*, vol. 36, pp. 823–841, Sep 1930. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.36.823

[39] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels .I. ergodic capacity," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 1083–1102, 2001.

[40] D. Tse and S. Hanly, "Multiaccess fading channels. I. polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2796–2815, 1998.

[41] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, D. N. Nguyen, E. Dutkiewicz, and O.-S. Shin, "Joint power control and user association for NOMA-based full-duplex systems," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 8037–8055, 2019.

[42] H. V. Nguyen, V.-D. Nguyen, O. A. Dobre, S. K. Sharma, S. Chatzinotas, B. Ottersten, and O.-S. Shin, "On the spectral and energy efficiencies of full-duplex cell-free massive MIMO," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1698–1718, 2020.

[43] H. Adhami, M. Alja'afreh, M. Hoda, J. Zhao, Y. Zhou, and A. El Saddik, "Suitability of SDN and MEC to facilitate digital twin communication over LTE-A," *Digital Communications and Networks*, vol. 10, no. 2, pp. 347–354, 2024.

[44] N. M. Rekoputra, C.-W. Tseng, J.-T. Wang, S.-H. Liang, R.-G. Cheng, Y.-F. Li, and W.-H. Yang, "Implementation and evaluation of 5G MEC-enabled smart factory," *Electronics*, vol. 12, no. 6, p. 1310, 2023.

[45] F. Morselli, S. Modarres Razavi, M. Z. Win, and A. Conti, "Soft information-based localization for 5G networks and beyond," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 9923–9938, 2023 .

[46] X. Ma, Z. Chen, W. Chen, Z. Li, Y. Chi, C. Han, and S. Li, "Joint channel estimation and data rate maximization for intelligent reflecting surface assisted terahertz MIMO communication systems," *IEEE Access*, vol. 8, pp. 99 565–99 581, 2020.

[47] T. P. Truong, V. D. Tuong, N.-N. Dao, and S. Cho, "FlyReflect: Joint flying IRS trajectory and phase shift design using deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 4605–4620, 2023.

[48] R. Zhong, Y. Liu, X. Mu, Y. Chen, X. Wang, and L. Hanzo, "Hybrid reinforcement learning for STAR-RISs: A coupled phase-shift model based beamformer," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2556–2569, 2022.

[49] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press, 2022.

[50] P. Gahinet, *LMI Control Toolbox: For Use with MATLAB ; User's Guide ; Version 1*, ser. Computation, visualization, programming. MathWorks, 1995. [Online]. Available: https://books.google.com.vn/books?id=GdWlXwAACAAJ

[51] H. Jiang, L. You, A. Elzanaty, J. Wang, W. Wang, X. Gao, and M.-S. Alouini, "Rate-splitting multiple access for uplink massive MIMO with electromagnetic exposure constraints," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1383–1397, 2023.

[52] M. Katwe, K. Singh, B. Clerckx, and C.-P. Li, "Improved spectral efficiency in STAR-RIS aided uplink communication using rate splitting multiple access," *IEEE Transactions on Wireless Communications*, vol. 22, no. 8, pp. 5365–5382, 2023.