Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.063206

ARTICLE

URLLC Service in UAV Rate-Splitting Multiple Access: Adapting Deep Learning Techniques for Wireless Network

Reem Alkanhel[1,#], Abuzar B. M. Adam[2,#], Samia Allaoua Chelloug[1], Dina S. M. Hassan[1,*], Mohammed Saleh Ali Muthanna[3] and Ammar Muthanna[4]

[1]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, 11671, Saudi Arabia

[2]Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg City, 1855, Luxembourg

[3]China-Korea Belt and Road Joint Laboratory on Industrial Internet of Things, Key Laboratory of Industrial Internet of Things and Networked Control, Ministry of Education, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China

[4]Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), Moscow, 117198, Russia

*Corresponding Author: Dina S. M. Hassan. Email: dshassan@pnu.edu.sa

[#]These authors contributed equally to this work

Received: 08 January 2025; Accepted: 27 April 2025; Published: 09 June 2025

ABSTRACT: The 3GPP standard defines the requirements for next-generation wireless networks, with particular attention to Ultra-Reliable Low-Latency Communications (URLLC), critical for applications such as Unmanned Aerial Vehicles (UAVs). In this context, Non-Orthogonal Multiple Access (NOMA) has emerged as a promising technique to improve spectrum efficiency and user fairness by allowing multiple users to share the same frequency resources. However, optimizing key parameters–such as beamforming, rate allocation, and UAV trajectory–presents significant challenges due to the nonconvex nature of the problem, especially under stringent URLLC constraints. This paper proposes an advanced deep learning-driven approach to address the resulting complex optimization challenges. We formulate a downlink multiuser UAV, Rate-Splitting Multiple Access (RSMA), and Multiple Input Multiple Output (MIMO) system aimed at maximizing the achievable rate under stringent constraints, including URLLC quality-of-service (QoS), power budgets, rate allocations, and UAV trajectory limitations. Due to the highly nonconvex nature of the optimization problem, we introduce a novel distributed deep reinforcement learning (DRL) framework based on dual-agent deep deterministic policy gradient (DA-DDPG). The proposed framework leverages inception-inspired and deep unfolding architectures to improve feature extraction and convergence in beamforming and rate allocation. For UAV trajectory optimization, we design a dedicated actor-critic agent using a fully connected deep neural network (DNN), further enhanced through incremental learning. Simulation results validate the effectiveness of our approach, demonstrating significant performance gains over existing methods and confirming its potential for real-time URLLC in next-generation UAV communication networks.

KEYWORDS: Deep learning; quality-of-service (QoS); rate-splitting multiple access (RSMA); unmanned aerial vehicle (UAV); ultra-reliable low-latency communication (URLLC)

1 Introduction

As 5G has inspired widespread research efforts, ultra-reliable and low-latency communications (URLLC) will be crucial in driving the future of wireless industrial automation and are designed to support critical applications that require highly reliable communication [1,2]. By the 3rd Generation Partnership

Project (3GPP) standard, the implementation of URLLC needs to ensure that data packets are delivered in a few milliseconds with the target reliability of 99.999% or higher [3]. However, this result can be obtained through a combination of advanced network technologies like non-orthogonal multiple access (NOMA). When NOMA is applied, data is decoded sequentially based on its power levels. This approach facilitates improved interference management and ensures that users with high reliability demands receive more consistent and reliable data [4]. Moreover, NOMA-based beamforming can meet the specific needs of URLLC communications by adapting networks to constantly changing conditions and varied application requirements [5,6]. Although NOMA has several benefits for URLLC communications, it comes up against problems of great complexity and unsatisfactory effect when multiple antennas are introduced for transmission [7]. Consequently, Rate-Splitting Multiple Access (RSMA) [8–10] has recently been suggested as an appropriate approach to achieve a more general and robust transmission framework than NOMA Systems [11–13], particularly in environments where users have different expectations when it comes to transmission rates. The main idea is to split user messages into two parts, a common part and a private part. Thus, the common parts and the private parts are encoded separately in a common data stream and private data streams, respectively [9]. Noting that interference generated by other users is partially decoded and treated as noise [14]. During reception, the signals are successively decoded. First, the user subtracts the common data stream and decodes their own private data stream, followed by the decoding of their own private data stream [15]. Applied in this context, RSMA could offer great spectral efficiency and reduce interference between users. RSMA flexibility can better control the problems inherent in NOMA by adapting the undercoats to suit channel conditions and user requirements. In the URLLC, the time and reliability requirements can hardly be met by a single communication link [16]. To overcome this challenge, a new support system architecture should be designed.

Joint optimization of beamforming vectors, rate allocation, and UAV trajectory under URLLC constraints forms a highly non-convex problem that cannot be solved efficiently using conventional optimization techniques [17]. Existing approaches typically address these components separately, leading to suboptimal solutions that fail to meet URLLC's real-time processing requirements. For instance, the alternating optimization method proposed in [18] requires iterative updates that introduce unacceptable latency ($\geq$2.5 ms) for true URLLC applications. While Deep Reinforcement Learning (DRL) has shown promise for wireless resource allocation [19], current single-agent frameworks lack the capability to simultaneously handle the continuous action space of beamforming/rate allocation and the discrete nature of trajectory planning. The work in [20] demonstrates this limitation, where a 23% performance degradation was observed when applying single-agent DRL to similar problems. There exists a significant gap in incorporating domain-specific knowledge from communication theory into DRL frameworks. Most existing solutions treat the optimization problem as a black box [21], failing to leverage known mathematical structures that could accelerate convergence and improve interpretability. This is particularly problematic for URLLC applications where performance guarantees are essential. To this end, we study the URLLC in multiuser UAV rate splitting multiple access to maximize the average data rate.

1.1 Motivations and Contributions

Motivated by recent advancements and the need for real-time URLLC services in multiuser UAV-aided RSMA networks, this paper leverages state-of-the-art deep learning methodologies to address the inherent complexities and optimization challenges. Our contributions are detailed as follows.

- First, we investigate a downlink multiuser UAV RSMA MIMO network targeting the maximization of the achievable rate, subject to stringent URLLC quality-of-service (QoS), power budget constraints, rate allocation, and UAV trajectory limitations. The resulting problem formulation leads to a nonconvex

optimization challenge that is difficult to solve directly within the latency constraints characteristic of next-generation networks.

- To effectively address the intractability of jointly optimizing beamforming vectors, rate allocations, and UAV trajectory parameters, we propose a novel distributed deep reinforcement learning (DRL) framework utilizing dual-agent deep deterministic policy gradient (DA-DDPG). This framework introduces a sophisticated neural network architecture for beamforming and rate allocation. Specifically, we integrate inception-like and deep unfolding mechanisms to construct the network layers, allowing multi-scale feature extraction and accelerated convergence.
- To apply the above mechanisms, we develop a successive pseudo-convex approximation (SPCA) and numerical solutions for rate allocation and the beamforming subproblem. To the best of our knowledge, this work represents the first attempt at incorporating these recent deep learning advances into solving complex optimization problems in wireless communications. Additionally, a fully connected deep neural network (DNN) is employed for the critic network.
- Regarding UAV trajectory optimization, we utilize a fully connected DNN to build the corresponding actor-critic-based agent, ensuring efficient learning and robust decision-making.
- We also apply an incremental learning training procedure [1] to improve model accuracy, reduce the complexity of the deep unfolding-based model, and significantly lower the volume of required training sequences.
- Simulation experiments conducted validate the efficacy of the proposed DA-DDPG framework. Numerical results illustrate that our proposed method consistently outperforms existing state-of-the-art techniques, underscoring its suitability and effectiveness for real-time URLLC-enabled UAV communications.

2 Related Work

Over the past few years, the research community has made significant efforts to investigate the complex and new challenges of URLLC-UAV communication based on the NOMA or the RSMA approaches [22–24]. Notably, the study in [22] proposed a model to deploy UAVs as relays between the base station and remote devices. The aim was to overcome the poor connectivity due to the presence of obstacles. In the proposed model, the authors targeted the maximization of the transmission rate on the backward link while satisfying the requirement of URLLC on the forward link. More recently, the authors in [23] considered the joint optimization problem for UAV-enabled URLLC-based mobile edge computing, which is divided into three subproblems. The first two subproblems optimize the UAV's horizontal and vertical locations, while the third one optimizes the offloading bandwidths and processing frequencies. Within this framework, they minimized the system's computation latency under an overall resource constraint.

In [24], a global iterative algorithm based on the alternating direction method of multipliers and random perturbation was proposed to minimize total error probability and jointly optimize blocklength allocation and UAV deployment. These contributions have made significant efforts in the field of URLLC-assisted UAV communication. However, the studies referred to above were carried out in the NOMA scenario, and the RSMA has received relatively little attention in the context of URLLC-assisted UAV networks. NOMA uses successive interference cancellation (SIC), which increases the complexity of processing in UAVs. This additional complexity can result in delays, causing problems in URLLC scenarios where extremely low latency is essential. In the current technical literature, only a few studies have explored RSMA-based URLLC-assisted wireless networks, as evidenced by [25]. In [25], RSMA was introduced to achieve energy-efficient (EE) URLLC in cell-free massive MIMO systems. In particular, the power allocation problem was formulated to enhance the EE.

Huang et al. introduced an artificial intelligence named RSMA-Deep Reinforcement Learning (DRL)-based approach. The authors demonstrated that, compared to the space-division multiple access (SDMA)-DRL protocol, the proposed method can achieve higher energy efficiency. Similarly, to maximize transmission power at every moment under energy harvesting, the authors in [26] presented a DRL method called the soft actor-critic algorithm. In particular, they used the Han–Powell quasi-Newton approach in sequential least squares programming (SLSQP) to optimize the sum-rate for the specified transmission power through DRL. The work in [27] presents a framework that integrates the Reconfigurable Intelligent Surface (RIS) with an intelligent satellite UAV-terrestrial network. The multiple access technique used in the communication architecture was NOMA. The authors formulated a multi-objective optimization problem that optimizes UAV trajectory, RIS phase shift, and transmit beamforming while minimizing UAV energy consumption and maximizing the sum rate. To solve the optimization problem, the multi-objective deep deterministic policy gradient (MO-DDPG) technique was suggested. In simulations, better data rates, energy consumption, and throughput were achieved when compared to systems without integrated RIS or random phase shift systems. The authors in [28] proposed a deep learning (DL)-based RSMA system for RIS-assisted terahertz massive multiple-input multiple-output (MIMO) systems. They presented a low-complexity approximate weighted minimum mean square error (AWMMSE) digital precoding scheme. Then, by combining AWMMSE with deep learning (DL), the deep unfolding [29] active precoding network (DFAPN) scheme was performed at the Base Station (BS), while passive precoding was performed using a transformer-based data-driven RIS reflecting network (RRN) at the RIS. Results showed that channel state information has a better achievable rate for the worst user device when compared to corresponding SDMA-based systems. Authors in [30] introduced the deep deterministic policy gradient (DDPG) approach to maximize the sum rate in downlink UAV-aided RSMA systems, where the trajectory and UAV beamforming matrix are jointly optimized. In this context, the UAV's uniform rectangular array (URA) beamforming design and mobility functions are considered. Unfortunately, there is currently a lack of research on the combination of URLLC and RSMA in a wireless UAV network. Hence, it is a mathematically challenging task to study the downlink achievable rate in UAV communication based on URLLC-RSMA.

3 System Model and Problem Formulation

In this work, a downlink multiuser UAV RSMA-URLLC system is considered, as depicted in Fig. 1, where the UAV is equipped with M antennas and serving N single-antenna users via short-packet communication (SPC), where the latency is always considered less than 1 ms, which indicates that the channel conditions can be viewed as quasi-static since this latency is shorter than the channel coherence time. The flying period of the UAV is divided into T time slots, with δ_t denoting the set of users as $N = \{1, 2, \cdots, N\}$ and the locations of the UAV and the users are respectively given as $L_n = (x_n, y_n)^T$, $n \in N$ and $q_n[t] = (x_u[t], y_u[t], z_u[t])^T$. The UAV movement can be defined as

$$\begin{aligned} &\|q_u(t+1) - q_u(t)\|^2 \leq D^2, \quad t = 1, \cdots, T-1 \\ &q_u(0) = q_0, \quad t = 1, \cdots, T-1 \\ &\|q_u(T) - q_u^F\|^2 \leq D^2, \end{aligned} \tag{1}$$

where $D = \delta_t v_{max}$ with v_{max} represents the maximum velocity of the UAV.

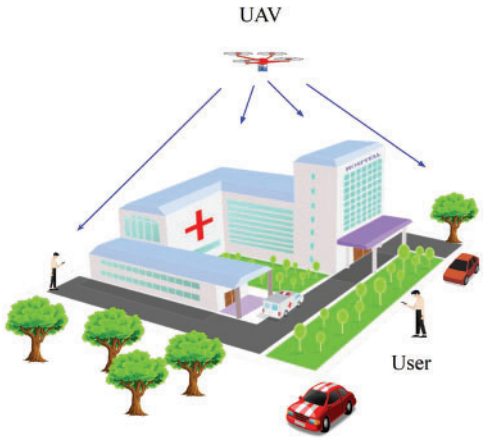


Figure 1: Multiuser RSMA-URLLC UAV network

Since RSMA is employed, the transmitted message from the UAV to the user n is divided into a common message w_c and a private message w_n. Hence, two data streams are transmitted, and performing the precoding on these data streams, the transmitted signal at the time slot t is given as

$$x[t] = v_c[t]s_s[t] + \sum_{n=1}^{N} v_n[t]s_n[t], \tag{2}$$

where $v_c[t] \in \mathbb{C}^{M\times 1}$ and $v_n[t] \in \mathbb{C}^{M\times 1}$ are the beamforming vectors with $P_c[t] = v_c v_c^H[t]$ and $P_n[t] = v_n v_n^H[t]$ denote the transmit power of the common and private streams, respectively; s_c and s_n are the data streams of the common and private messages with $\mathbb{E}ss^H = I$, and s is the total transmit data stream. The received signal at the user n at the time slot t is given as

$$y_n[t] = h_{u,n}^H[t]x[t] + \eta_n[t], \tag{3}$$

where $h_{u,n} \in \mathbb{C}^{M\times 1}$ is the channel vector between the UAV and the user n and $\eta_n \sim \mathbb{CN}(0, \sigma_n^2)$ is the additive white Gaussian noise (AWGN). Based on RSMA protocol, each user first decodes the common message using successive interference cancellation (SIC) and then decodes its own private signal. The signal-to-interference-plus-noise ratio (SINR) of the user n decoding the common message at the time slot t is given as

$$\gamma_{n,c}[t] = \frac{|h_{u,n}^H[t]v_c[t]|^2}{\sum_{l=1}^{K}|h_{u,n}^H[t]v_l[t]|^2 + \sigma_n^2}. \tag{4}$$

Similarly, the SINR of the user n decoding its private message at the time slot t is given as follows:

$$\gamma_{n,p}[t] = \frac{|h_{u,n}^H[t]v_n[t]|^2}{\sum_{l=1,l\neq n}^{K\,\Sigma_n^2}|h_{u,n}^H[t]v_l[t]|^2}. \tag{5}$$

To formulate the quality-of-service (QoS) constraints of URLLC, we follow the derivation in Eq. (5), where the approximate maximum achievable rate of the user n at the time slot t and for a multi-antenna with quasi-static flat fading channel can be expressed as follows:

$$r_{u,n,c}^{(\varepsilon)}[t] = \log_2(1+\gamma_{n,c}[t]) - \sqrt{\frac{V(\gamma_{n,c}[t])}{B}}\frac{\mathbb{Q}^{-1}(\epsilon_{n,c})}{\ln_2}, \tag{6}$$

where $V(\gamma_{n,c}[t]) = 1 - (1 - \gamma_{n,c}[t])^{-2}$ is the channel dispersion, and B is the transmission channel block-length, $\mathbb{Q}^{-1}(\varepsilon_{n,c})$ is the inverse of the Gaussian Q-function, and $\varepsilon_{n,c}[t] = \mathbb{Q}(f(\gamma_{n,c}[t], B, r_{u,n,c}[t]))$ denotes the decoding error probability with $f(\gamma_{n,c}[t], B, r_{u,n,c}[t])$ defined as follows:

$$f(\gamma_{n,c}[t], B, r_{u,n,c}[t]) = \ln_2 \sqrt{\frac{B}{v}}(\log_2)(1 + \gamma_{n,c}[t]) - r^{(\varepsilon)}_{u,n,c}[t]. \tag{7}$$

To guarantee the decoding of the common message by the user n, the constraints $\sum_{n=1}^{N} C^{(\varepsilon)}_{u,n,c}[t] \leq \min_{n\in\mathbb{N}}\{r^{(\varepsilon)}_{u,n,c}[t]\}$ is imposed.

Similarly, the achievable data rate of the user n to decode the private stream is given as

$$r^{(\varepsilon)}_{u,n,c}[t] = \log_2(1 + \gamma_{n,p}[t]) - \sqrt{\frac{V(\gamma_{n,p}[t])}{B}}\frac{\mathbb{Q}^{-1}(\varepsilon_{n,p})}{\ln_2}. \tag{8}$$

The achievable rate of the user n at the time slot t is defined as

$$R^{(\varepsilon)}_{u,n}[t] = C^{(\varepsilon)}_{u,n,c}[t] + r^{(\varepsilon)}_{u,n,p}[t]. \tag{9}$$

Aiming at maximizing the total achievable rate of the system, we formulate our optimization problem as follows:

$$\max_{v,C,Q} \sum_{n=1}^{N} R^{(\varepsilon)}_{u,n}[t],$$

s.t.

$$\begin{aligned}
&\mathbf{C1}: \sum_{n=1}^{N} C^{(\varepsilon)}_{u,n,c}[t] \leq \min_{n\in\mathbb{N}}\{r^{(\varepsilon)}_{u,n,c}[t]\} && \mathbf{C2}: |v_c[t]|^2 + \sum_{n=1}^{N} |v_n[t]|^2 \leq P_{max} \mathbf{C3}: C^{(\varepsilon)}_{u,n,c}[t] \geq 0,\\
&\mathbf{C4}: r^{(\varepsilon)}_{u,n,c}[t] \geq 0, && \mathbf{C5}: \varepsilon_{n,p} \leq \varepsilon^{th}, \varepsilon_{n,c} \leq \varepsilon^{th} \mathbf{C6}: R^{(\varepsilon)}_{u,n}[t] \geq R^{(\min)}_{u,n}\\
&\mathbf{C7}: v_n[t] \geq 0, v_c[t] \geq 0,
\end{aligned} \tag{10}$$

where $v = [v_c, v_1, ..., v_N]^T \in \mathbb{C}^{(N+1)\times M}$ and $C = \left[r^{(\varepsilon)}_{u,n,c}, C^{(\varepsilon)}_{u,1,c}, C^{(\varepsilon)}_{u,2,c}, \cdots, C^{(\varepsilon)}_{u,N,c}, C^{(\varepsilon)}_{u,1,p}, C^{(\varepsilon)}_{u,2,p}, \cdots, C^{(\varepsilon)}_{u,N,p}\right]^T \in \mathbb{R}^{(2N+1)\times 1}$. The constraint **C1** ensures the successful decoding of the common stream by all users. The constraint **C2** is the power budget. The constraints **C5** and **C6** are the QoS of URLLC with $R^{min}_{u,n}$ representing the minimum rate requirements.

Problem (10) is nonconvex due to the nonconvexity of the objective function and the tight coupling of v, C, and Q. Joint optimization of the beamforming, rate allocation, and UAV trajectory is intractable. Next, instead of directly solving the problem mathematically, we design a DRL framework to jointly optimize the above three parameters.

In the following sections, we provide a novel deep reinforcement framework to handle the problem in (10). The proposed framework combines inception-like mechanisms, deep unfolding, and dual-agent DRL within the context of URLLC and RSMA-assisted UAV networks. It offers a practical and scalable DRL framework tailored specifically for real-time wireless optimization scenarios involving multiple antennas and stringent QoS demands.

4 Deep Unfolding-Based Inception-like Model

The proposed model architecture incorporates an inception-like block, where the inception mechanism is employed to enhance the design of the beamforming and rate allocation actor. The inception mechanism, inspired by convolutional neural networks, allows the model to process different features at multiple scales by applying several operations (e.g., different filter sizes or layers) in parallel. This increases the flexibility and expressiveness of the model, allowing it to better capture the diverse patterns and correlations between beamforming and rate allocation decisions. Additionally, the concept of deep unfolding is integrated into the actor's design [31,32]. Deep unfolding involves converting iterative optimization algorithms into neural networks by mapping each iteration to a neural network layer. This technique allows the model to mimic traditional optimization algorithms while leveraging the learning capability of deep neural networks. By using deep unfolding, the beamforming and rate allocation actor is able to emulate efficient optimization steps, improving both accuracy and convergence speed.

To build the proposed deep unfolding-based inception-like model, first, we solve the problem using a traditional optimization method, namely, successive pseudo-convex optimization for the beamforming and the interior point for the rate allocation. The beamforming and rate allocation subproblem is given as (we removed $[t]$ for simplicity)

$$\max_{v,\mathbf{C}} \quad \sum_{n=1}^{N} R_{u,n}^{(\varepsilon)}, \tag{11}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} C_{u,n,c}^{(\varepsilon)} \leq \min_{n\in\mathbb{N}}\{r_{u,n,c}^{(\varepsilon)}\} \tag{11a}$$

$$|v_c|^2 + \sum_{n=1}^{N} |v_n|^2 \leq P_{max} \tag{11b}$$

$$C_{u,n,c}^{(\varepsilon)} \geq 0, \tag{11c}$$

$$r_{u,n,c}^{(\varepsilon)} \geq 0, \tag{11d}$$

$$\varepsilon_{n,p} \leq \varepsilon^{th}, \varepsilon_{n,c} \leq \varepsilon^{th} \tag{11e}$$

$$R_{u,n}^{(\varepsilon)} \geq R_{u,n}^{(\min)} \tag{11f}$$

$$v_n \geq 0, v_c \geq 0. \tag{11g}$$

The problem in (11) can be further decoupled into rate allocation and beamforming. The rate allocation subproblem is given as

$$\max_{v,\mathbf{C}} \quad \sum_{n=1}^{N} R_{u,n}^{(\varepsilon)} \tag{12}$$

$$\text{s.t.} \quad (11a),\ (11c)\ ,(11d),(11e),(11f).$$

Following the analysis in [2], constraint (11f) can be transformed into $C_{u,n,c}^{(\varepsilon)} \geq R_{u,n}^{(\min)} - r_{u,n,p}^{(\varepsilon)^*}$, assuming that the solution $r_{u,n,p}^{(\varepsilon)^*}$ is given. Then, the problem is decoupled into two subproblems as follows:

$$\mathbf{C}_c^* = \underset{\mathbf{C}_c}{\operatorname{argmax}} \sum_{n=1}^{N} C_{u,n,c}^{(\varepsilon)} \tag{13}$$

$$\text{s.t.} \quad C_{u,n,c}^{(\varepsilon)} \geq R_{u,n}^{(\min)} - r_{u,n,p}^{(\varepsilon*)} \forall n \in \mathcal{N}, \tag{13a}$$

$$\mathbf{C}_{n,p}^{*} = \arg\max_{\mathbf{C}_{n,p}} \sum_{n=1}^{N} r_{u,n,p}^{(\varepsilon)} \tag{14}$$

$$\text{s.t.} \quad 0 \le r_{u,n,p}^{(\varepsilon)} \le \log_2\left(1+\gamma_{n,p}\right) - \sqrt{\frac{V\left(\gamma_{n,p}\right)}{B}} \frac{\mathcal{Q}^{-1}\left(\varepsilon_{n,p}\right)}{\ln 2}, \forall_n \in \mathcal{N}, \tag{14a}$$

where $\mathbf{C}_c^{*}, \mathbf{C}_{n,p}^{*} \in \mathbf{C} = \left[r_{u,n,c}^{(\varepsilon)}, \mathbf{C}_{u,1,c}^{(\varepsilon)}, \mathrm{C}_{u,2,c}^{(\varepsilon)}, \ldots, \mathbf{C}_{u,N,c}^{(\varepsilon)}, \mathbf{C}_{u,1,p}^{(\varepsilon)}, \ldots, \mathbf{C}_{u,N,p}^{(\varepsilon)}\right]^T \in \mathbb{R}^{(2N+1)\times 1}$. The problem in (13) is convex and can be solved by standard optimization methods such as the interior point. For $\mathbf{C}_c^{*}$, the objective function is linear, and the optimal solution can be expressed as follows:

$$\mathbf{C}_c^{*} = \begin{cases} C_{n,c}^{max} & \text{if } \gamma_{n,c} = \max\limits_{l\in N}\{\gamma_{l,c}\} \\ C_{n,c}^{min} & \text{otherwise} \end{cases}, \tag{15}$$

where $C_{n,c}^{max} = max\left\{r_{u,n,c}^{(\varepsilon)} - \sum_{l=1,l\neq n}^{N} C_{u,n,c}^{(\varepsilon)}, 0\right\}$ and $C_{n,c}^{min} = max\left\{R_{u,n}^{min} - r_{u,n,c}^{(\varepsilon^*)}, 0\right\}$. The structure of the proposed solution in (15) is very helpful in designing deep unfolding layers for our proposed rate allocation network. However, the solution for (14) is numerical. Therefore, we use linear layers to design the corresponding neural network of (14).

The successive pseudo-convex approximation (SPCA) has emerged as a pivotal technique in wireless communications for transforming nonconvex optimization problems into tractable convex forms through iterative approximations [3,33]. This approach is particularly effective for beamforming design, where the core subproblem is formulated as

$$\max_{v} \quad \sum_{n=1}^{N} R_{u,n}^{(\varepsilon)} \tag{16}$$

$$\text{s.t.} \quad |v_c[t]|^2 + \sum_{n=1}^{N} |v_n[t]|^2 \le P_{max}, \tag{16a}$$

$$\varepsilon_{n,p} \le \varepsilon^{th}, \varepsilon_{n,c} \le \varepsilon^{th}, \tag{16b}$$

$$v_n[t] \ge 0, v_c[t] \ge 0. \tag{16c}$$

Problem (16) is nonconvex due to the nonconvexity of the objective function and constraint (16b). Using the conventional Software Communications Architecture (SCA), a surrogate function $\hat{R}_{u,n}^{(\varepsilon)}$ is approximated at the points $v_n^{(\tau)}$ of $R_{u,n}^{(\varepsilon)}$ is defined. Assuming that problem (10) has a closed and convex solution set V and an optimal point $V^{(\tau)}$, where $v_n^{(\tau)} \in V$, $\hat{R}_{u,n}^{(\varepsilon)}$ should satisfy the following technical conditions:

1. $\hat{R}_{u,n}^{(\varepsilon)}$ is pseudo-concave for any $v_n^{(\tau)} \in \mathbf{V}$.
2. $\hat{R}_{u,n}^{(\varepsilon)}$ is continuously differentiable in $v_n \in \mathbf{V}$ and for any given $v_n^{(\tau)} \in \mathbf{V}$. Moreover, $\hat{R}_{u,n}^{(\varepsilon)}$ is continuous in $v_n \in \mathbf{V}$ for $v_n^{(\tau)} \in \mathbf{V}$.
3. The gradients of $\hat{R}_{u,n}^{(\varepsilon)}$ is identical to that of $R_{u,n}^{(\varepsilon)}$ for $v_n^{(\tau)} \in v_n$.
4. The set V is nonempty for $\tau = 1, \cdots,$
5. The sequence $\left\{\mathcal{V}^{(\tau)}\right\}_{\tau=1}^{T}$ is convergent for any convergent sequence $\left\{v^{(\tau)}\right\}_{\tau=1}^{T}$.

On the highlights of these technical conditions, the surrogate $\hat{R}_{u,n}^{(\varepsilon)}$ function is defined as

$$\hat{R}_{u,n}^{(\varepsilon)} = \hat{R}_{u,-n}^{(\varepsilon)}\left(v; v^{(\tau)}\right) + \left(v - v^{(\tau)}\right) \sum_{l=1,l\neq n}^{N} \nabla_v R_{u,n}^{(\varepsilon)}\left(v_n^{(\tau)}\right). \tag{17}$$

Using the objective function, problem (16) can be rewritten as

$$\max_{\mathrm{v}} \quad \sum_{n=1}^{N} \hat{R}_{u,n}^{(\varepsilon)} \tag{18}$$

$$\text{s.t.} \quad (16a), (16c),$$

$$f\left(\gamma_{-n,c}\left(\mathrm{v}_c;\mathrm{v}_c^{(\tau)}\right), B, r_{u,-n,c}^{(\varepsilon)}\right) + \left(\mathrm{v}_c - \mathrm{v}_c^{(\tau)}\right) \sum_{l=1}^{N} \nabla_{\mathrm{v}_c} f\left(\gamma_{n,c}\left(\mathrm{v}_n^{(\tau)}\right), B, r_{u,n,c}^{(\varepsilon)}\right) \leq Q^{-1}\left(\varepsilon^{th}\right), \tag{18a}$$

$$f\left(\gamma_{-n,p}\left(\mathrm{v}_n;\mathrm{v}_n^{(\tau)}\right), B, r_{u,-n,p}^{(\varepsilon)}\right) + \left(\mathrm{v}_n - \mathrm{v}_n^{(\tau)}\right) \sum_{l=1,l\neq n}^{N} \nabla_{\mathrm{v}_n} f\left(\gamma_{n,p}\left(\mathrm{v}_n^{(\tau)}\right), B, r_{u,n,p}^{(\varepsilon)}\right) \leq Q^{-1}\left(\varepsilon^{th}\right), \tag{18b}$$

where constraints (18a) and (18b) are transformations of constraint (16b). Applying Lagrange relaxation and Karush-Kuhn-Tucker (KKT) conditions, the beamforming can be expressed in the following closed-form expressions:

$$\mathbf{v}_n^{(\tau)} = \sqrt{\frac{\left(\beta_{n,p}^{(\tau)} + \zeta_{n,p}^{(\tau)}\right)\Psi_{u,n,1}^2\left(2\beta_{n,p}^{(\tau)} - 2\right)}{\left(\mathrm{Tr}\left(\mathbf{h}_{u,n}\mathbf{h}_{u,n}^H\right)\right)^6\left(2 - \left(\Theta_{u,n}^{(\varepsilon)}\right)^2\right)} - \Omega_{u,n}^{(\varepsilon)} \frac{\Psi_{u,n,2}^6}{\mathrm{Tr}\left(\mathbf{h}_{u,n}\mathbf{h}_{u,n}^H\right)\Psi_{u,n,3}^2\beta_{n,p}^{(\tau)}}}, \tag{19}$$

where $\Theta_{u,n}^{(\varepsilon)} = \frac{\sqrt{2}}{2\ln 2} e^{2(erfc^{-1}(\varepsilon))^2}$, $\Omega_{u,n}^{(\varepsilon)} = \frac{Q^{-1}(\varepsilon)}{\ln 2}\left(\frac{\beta_{n,p}^{(}\tau)-1}{\left(\beta_{n,p}^{(}\tau)\right)^2}\right)$ with $\beta_{n,p}^{(\tau)}$ and $\zeta_{n,p}^{(\tau)}$ denote Lagrange multipliers associated with constraint (15). $\psi_{u,n,1} = \sum_{l=1,l\neq n}^{K\Sigma_2} \mathbf{Tr}\left(h_{u,l}h_{u,l}^H\right)$, $\psi_{u,n,2} = \sum_{n=1}^{K} \mathbf{Tr}\left(h_{u,n}h_{u,n}^H\right) + \sigma_n^2$ and $\psi_{u,n,3}^2 = \sum_{l=1}^{K} \mathbf{Tr}\left(h_{u,l}h_{u,n}^H\right) + \sigma_n^2$. The proposed deep unfolding-based solution is designed using the formulas in (15) and (19), and the solution of (13). Fig. 2 illustrates a block structure of the proposed deep-unfolding-based neural network.

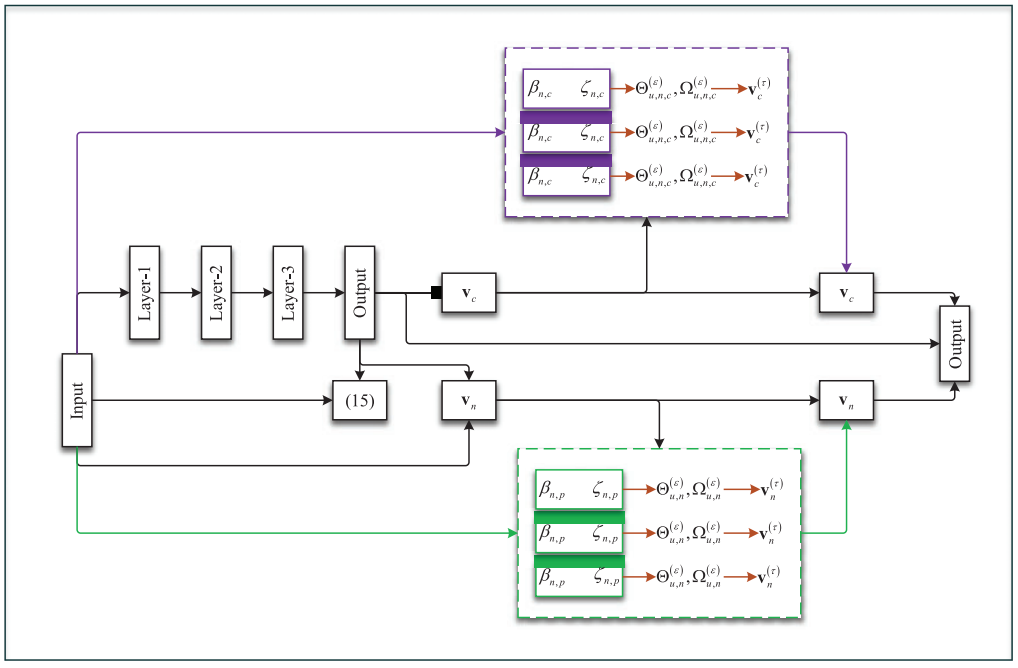


Figure 2: The structural design of a block of the proposed deep-unfolding-based neural network

The deep unfolding-based layer in 2 is built to resemble an inception block by accommodating neurons and layers in a mixed sequential and parallel structure to avoid the vanishing gradient problem. The block in 2 will be used in the actor network in the proposed DA-DDPG framework in Section 4.

5 Proposed Deep Unfolding-Based DDPG Framework

This section presents a novel design of a DRL framework using the dual-agent deep deterministic policy gradient (DA-DDPG) algorithm, and reformulates the problem as a Markov decision process (MDP). DA-DDPG is an extension of the standard DDPG method, which incorporates two agents working in parallel. One agent focuses on optimizing the continuous action space for beamforming and rate allocation, while the other agent handles the UAV trajectory optimization. By using dual agents, the framework can effectively decouple these two tasks, allowing for more specialized learning and faster convergence during training.

5.1 MDP Formulation

The traditional MDP consists of the following six-tuples: $\langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{T}, \ v\rangle$, which respectively stand for the state space, action space, reward function, transition policies, and discount factor. During the learning process, the state value $Q_\pi(s)$ function is maximized to find the optimal policy $\pi^*(s,a)$. The transition policy $\pi\left(a^{(t)}|s^{(t)}\right)$ is mapping the state $s^{(t)}$ and the probability of choosing each possible action. The state-value function $Q_\pi(s)$ is defined as

$$Q_\pi\left(s\right) = \mathbb{E}_\pi\left[A^{(t)}|s^{(t)}\right] = \mathbb{E}_\pi\left[\sum_{k=1}^{\infty} v^{(k)} r^{(t+k+1)}|s^{(t)}\right] \forall s \in \mathcal{S}, \tag{20}$$

where $\mathbb{E}_\pi$ is the expectation of the random variable when the agent follows the transition policy. $A^{(t)} = \sum_{k=1}^{\infty} v^{(k)} r^{(t+k+1)}$ can be considered the action that maximizes the sum discount reward obtained during the process. The discount parameter $v \in [0,1]$ is used to weight the short-term and long-term rewards.

The state-value function $Q_\pi(s|a)$ can be written for the policy $\pi\left(s^{(t)}|a^{(t)}\right)$ when the state value is known to obtain the expected reward. Thus, we have

$$Q_\pi(s|a) = \mathbb{E}_\pi\left[A^{(t)}|s^{(t)}, a^{(t)}\right] = \mathbb{E}_\pi\left[\sum_{k=1}^{\infty} v^{(k)} r^{(t+k+1)}|s^{(t)}, a^{(t)}\right] \forall a \in \mathcal{A}. \tag{21}$$

And the Bellman expectation equation can be expressed

$$\begin{aligned} V_\pi\left(s\right) &= \mathbb{E}_\pi\left[A^{(t)}|s^{(t)}\right] = Q_\pi\left(s|a\right) \\ &= \mathbb{E}_\pi\left[A^{(t)}|s^{(t)}, a^{(t)}\right] = \mathbb{E}_\pi\left[\sum_{k=1}^{\infty} v^{(k)} r^{(t+k+1)}|s^{(t)}\right] \\ &= \sum_{a\in\mathcal{A}} \pi\left(a^{(t)}|s^{(t)}\right)\left(\mathcal{R}^{(s,a)} + v\sum_{s'\in\mathcal{S}} \Gamma^{(s,s',a)} Q_\pi\left(s'\right)\right), \end{aligned} \tag{22}$$

where $\mathcal{R}^{(s,a)} = \mathbb{E}_\pi\left[r^{(t+1)}|s^{(t)}, a^{(t)}\right]$ represents the reward function, and $\Gamma^{(s,s',a)} = \mathbb{P}\left[s^{(t+1)}|s^{(t)}, a^{(t)}\right]$ is the matrix probability of the state transition.

By interacting with the environment through behaviors, the agent in DRL learns the policy and modifies its behavior in response to rewards. The following provides a thorough explanation of the state space, action space, and reward:

1. **State space $\mathcal{S}$:** The state space in our case includes two state subspaces $\mathcal{S}_1^{(t)}$ for the case of the beamforming and rate allocation agent, and $\mathcal{S}_2^{(t)}$ for the UAV trajectory. $\mathcal{S}_1^{(t)}$ and $\mathcal{S}_2^{(t)}$ are given as follows:

$$\mathcal{S}_1^{(t)} = \{\{\mathbf{h}_{u,n}[t]\}_{n=1}^{N}, P_{max}, a_1^{(t-1)}, A^{(t)}, M\}, \tag{23}$$

$$\mathcal{S}_2^{(t)} = \{\mathbf{q}_u[t], a_2^{(t-1)}, A^{(t)}, \{\mathbf{L}_n\}_{n=1}^{N}\}, \tag{24}$$

where $a_1^{(t-1)}$ and $a_2^{(t-1)}$ denote the actions from the last time slot for the beamforming and rate allocation network and the UAV trajectory network, respectively.

2. **Action space** $\mathcal{A}$: Similar to the case of the state space, the action space includes the actions from both agents. Two action subspaces can be defined as follows:

$$\mathcal{A}_1^{(t)} = \{\mathbf{v}[t], \mathbf{C}[t]\}, \tag{25}$$

$$\mathcal{A}_2^{(t)} = \{\mathbf{q}_u[t]\}, \tag{26}$$

3. **Reward:** Following the structure of problem (10), the reward function should be built to maximize the total achievable rate for the given constraints. Therefore, penalties are set to force the satisfaction of the constraints while maximizing the objective function. Hence, the reward can be defined as follows:

$$\mathcal{R}^{(t)} = \begin{cases} -\rho & \text{if } s \text{ is negative} \\ \sum_{n=1}^{N} R_{u,n}^{(\varepsilon)}[t] & \text{otherwise} \end{cases}. \tag{27}$$

5.2 Algorithm Description

The DA-DDPG algorithm is a reinforcement learning framework tailored specifically for continuous optimization problems, incorporating two separate agents. The actor network creates actions depending on the current state of each agent's actor-critic structure, and the critic network assesses these actions to guide policy improvements.

Specifically, each agent consists of two neural networks: a training network and a slowly updated target network. The target networks are periodically updated using parameters from the training networks, significantly contributing to stable and robust training by mitigating potential divergence issues common in reinforcement learning.

The critic network is responsible for estimating the expected cumulative reward (Q-value) associated with performing a specific action given a particular state. During training, the critic network parameters θ_{critic} are adjusted by minimizing the temporal difference loss, which quantifies the discrepancy between the predicted Q-value from the critic training network and the target network's estimated Q-value. This update is mathematically represented as

$$L(\theta_{critic}^{train}) = \left(r^{(t)} + \nu Q(\theta_{critic}^{target} | s^{(t+1)}, a') - Q(\theta_{critic}^{train} | s^{(t+1)}, a')\right)^2, \tag{28}$$

where $r^{(t)}$ is the immediate reward received after executing action a' in state $s^{(t)}$, ν is the discount factor balancing immediate vs. future rewards, $Q(\theta_{critic}^{target} | s^{(t+1)}, a')$ is the target critic network's Q-value prediction, and $Q(\theta_{critic}^{train} | s^{(t+1)}, a')$ is the critic training network's current Q-value estimation. The actor network aims to find the optimal policy that maximizes expected rewards by following the policy gradient derived from

the critic network. Its parameters are updated using the critic's gradients concerning the action, thus directly influencing the actions produced by the actor network to optimize the policy.

$$\theta_{actor}^{(t+1)} = \theta_{actor}^{(t)} - i_{actor} \nabla_a Q(\theta_{critic}^{target} | s^{(t)}, a') \nabla_{\theta_{actor}^{train}} \pi(\theta_{actor}^{train} | s^{(t)}). \tag{29}$$

The parameters of the target networks for both the actor and critic are updated through a soft update procedure to slowly track the training networks, helping stabilize the training process.

$$\theta^{target} \leftarrow \tau \theta^{train} + (1 - \tau) \theta^{target}, \tag{30}$$

where τ is typically set to a small value such as 0.001. Algorithm 1 illustrates the training procedure for the proposed DA-DDPG.

Algorithm 1: DA-DDPG training procedure

1: Initialize actor and critic networks $\theta_{actor}, \theta_{critic}$ and their target networks $\theta_{actor}^{target}, \theta_{critic}^{target}$
2: Initialize replay buffer $\mathcal{B}$
3: **for** each episode **do**
4: Observe initial state s
5: **for** each timestep t **do**
6: Select action $a = \pi(s|\theta_{actor}) + \text{noise}$
7: Execute action a, observe reward r and next state s'
8: Store transition (s, a, r, s') in replay buffer $\mathcal{B}$
9: Sample random minibatch of transitions from $\mathcal{B}$
10: Update critic by minimizing loss:

$$L = \left(r + \nu Q(\theta_{critic}^{target} | s', a') - Q(\theta_{critic}^{train} | s, a)\right)^2$$

11: Update actor network using sampled policy gradient:

$$\nabla_{\theta_{actor}} \approx \frac{1}{N} \sum \nabla_a Q(\theta_{critic} | s, a) \nabla_{\theta_{actor}} \pi(s | \theta_{actor})$$

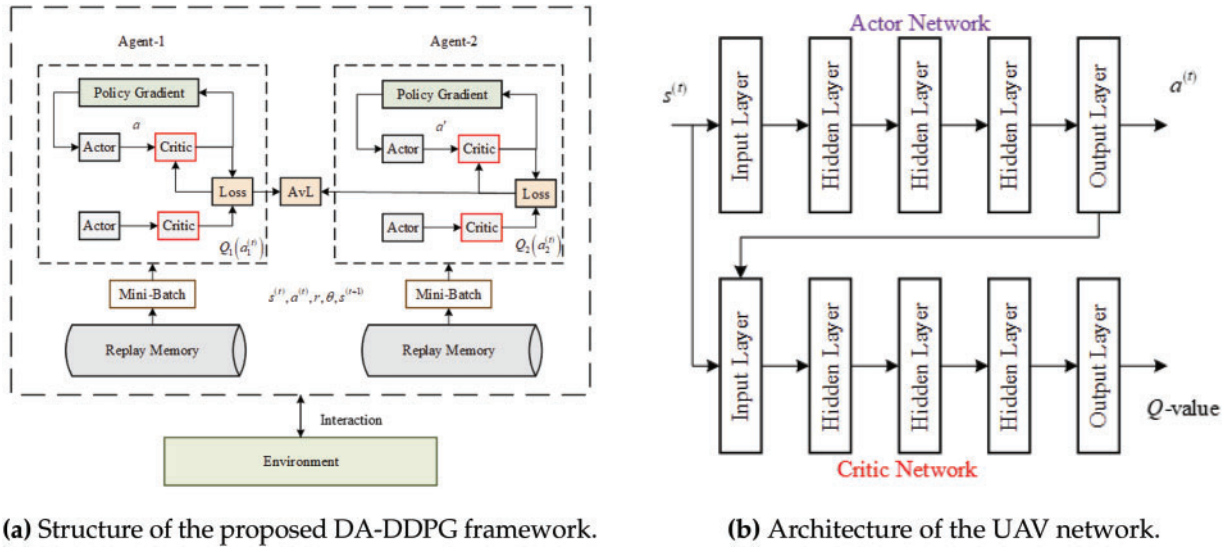
12: Soft-update target networks:

$$\theta^{target} \leftarrow \tau \theta^{train} + (1 - \tau) \theta^{target}$$

13: Update state $s = s'$
14: **end for**
15: **end for**

The overall structure of the proposed DRL framework is given as in Fig. 3a. The network consists of two agents, each agent has actor-critic structure for both training and target networks.

The beamforming and rate allocation actor is constructed using an inception-like deep unfolding-based model from Section 4, while the critic is built using a fully connected deep neural network (DNN). On the other hand, a fully connected DNN-based actor-critic architecture is used to design the UAV trajectory network. The structure of the UAV trajectory is a DNN, as in Fig. 3b. In this architecture, the actor predicts the optimal UAV trajectory, while the critic evaluates the predicted trajectory by calculating the value function. The actor-critic structure ensures that the UAV trajectory network continuously improves its trajectory decisions based on feedback from the critic, leading to more efficient trajectory optimization over time.



(a) Structure of the proposed DA-DDPG framework. **(b)** Architecture of the UAV network.

Figure 3: Architectural diagrams of the proposed system: **(a)** The DA-DDPG framework structure and **(b)** The UAV network architecture

5.3 Inference Steps and Complexity Analysis

During inference, the trained DA-DDPG algorithm executes actions based on learned policies without further training. Initially, given the current state $s^{(t)}$, the beamforming and rate allocation actor network generates optimized beamforming vectors $\mathbf{v}[t]$ and rate allocation $\mathbf{C}[t]$, processing the state information, which includes channel conditions $\mathbf{h}_{u,n}[t]n = 1^N$, previous actions, available power $Pmax$, and antenna configurations M. Concurrently, the UAV trajectory actor receives its state comprising UAV location $\mathbf{q}_u[t]$, previously taken actions $a_2^{(t-1)}$, and user locations $L_{n\,n=1}^{N}$, then computes the next UAV position $q_u[t+1]$, adhering to trajectory constraints.

The computational complexity during inference primarily arises from the neural network architectures. For the beamforming and rate allocation agent, employing inception-like and deep unfolding structures, the complexity can be represented as $\mathcal{O}(L_u N_{inc}^2)$, where L_u denotes the number of unfolding layers and N_{inc} indicates the number of neurons per inception-like layer. For the UAV trajectory agent, using a fully connected deep neural network, the complexity is $\mathcal{O}(L_{traj} N_{traj}^2)$, with L_{traj} being the number of layers and N_{traj} the number of neurons per layer. Consequently, the total inference complexity of DA-DDPG is expressed as

$$\mathcal{O}(L_u N_{inc}^2 + L_{traj} N_{traj}^2), \tag{31}$$

making the framework suitable for real-time applications required by URLLC scenarios.

6 Simulation Results

This section presents simulation results to judge the performance of the proposed DA-DDPG framework for the multiuser UAV URLLC network. We consider a network architecture as outlined in Section II, where a total of N users are uniformly and randomly distributed within a circular area with a radius of R = 500 m. To support the ground Internet of Things (IoT) units in URLLC applications, the UAV is deployed over the region, with altitudes of 150 m. Unless stated otherwise, other simulation parameters are summarized in Table 1. The simulation parameters for the DA-DDPG framework are illustrated in Table 2. Fig. 4 shows

the performance evaluation of the proposed framework through four different metrics. Fig. 4a gives the total achievable rate for different values of P_{max}. We fix the minimum rate at $R_{u,n}^{min}$ bps/Hz, and the number of users is set as 10. The performance of the proposed DA-DDPG is closer to that of SPCA and surpasses the performance of DRL. The behavior of the proposed DA-DDPG is similar to that of SPCA, and that is because the structure of the proposed neural network is mostly based on the highlights of the SPCA solution. The performance gap between DA-DDPG and SPCA is about 1.86%. The gap is mostly due to the data-driven parts of the proposed agents. In the mean, the performance of DA-DDPG is 10.77% better than that of DRL for M = 2 and 8.09% for M = 4.

Table 1: Simulation parameters

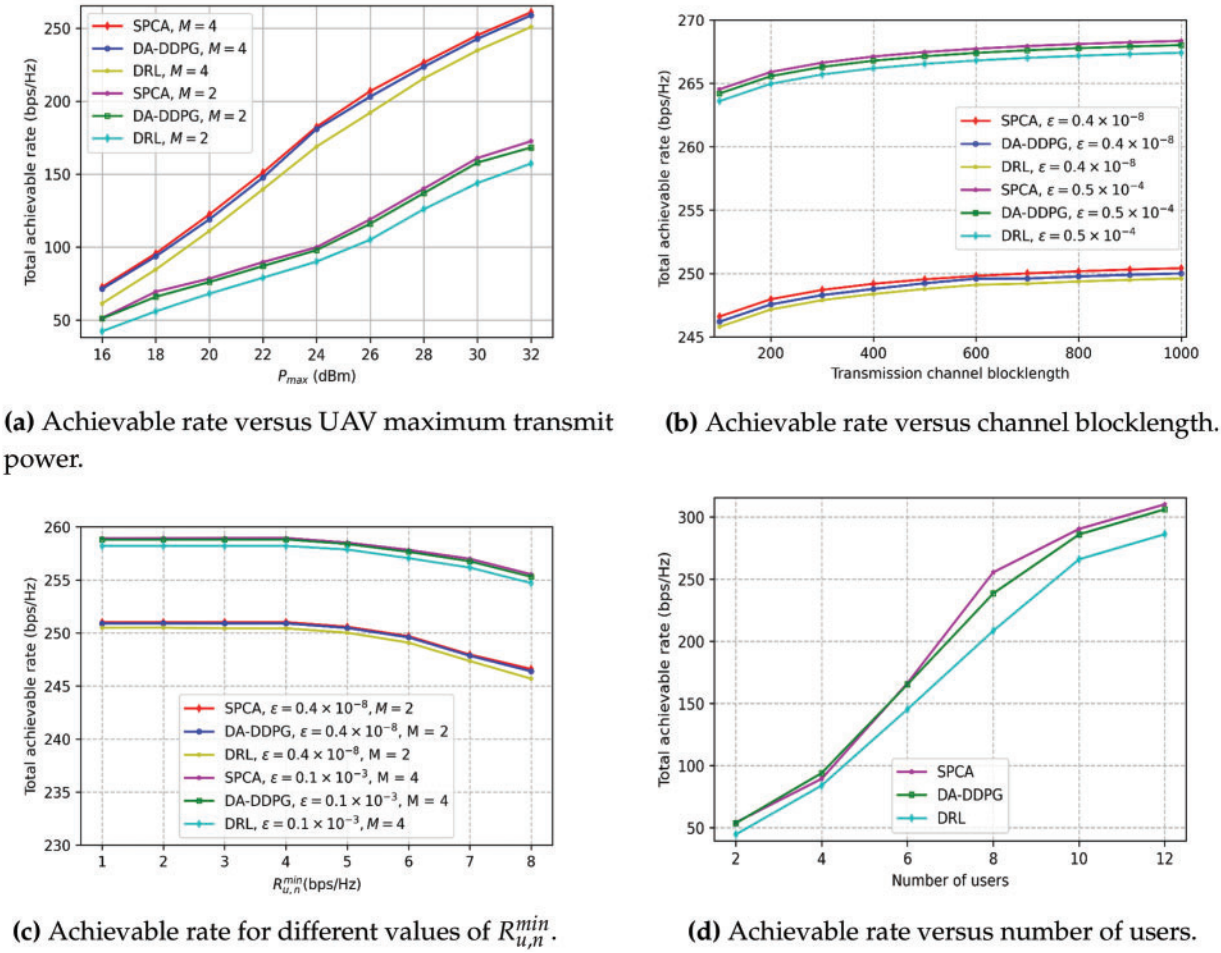
Description	Value
Number of users	50
Carrier frequency	6 GHz
Maximum available bandwidth of each UAV	20 MHz
Length of transmission packet	32 bytes
Single-side noise spectral density	−80 dBm/Hz
Maximum transmit power of each device	120 mW
Speed of light	3×10^8 m/s
Packet re-transmission delay	0.7 ms
Packet decoding and processing delay	0.2 ms
Threshold of decoding error	10^{-5}
Excessive path loss of Line-of-Sight (LoS)	1 dB
Excessive path loss of Non-Line-of-Sight (NLoS)	20 dB

Table 2: DA-DDPG framework parameters

Parameters	Values
Number of layers for actor	5
Number of training episodes	1500
Learning rate for actor	10^{-3}
Learning rate for critic	10^{-3}
Replay memory size	10,000 bytes
Batch size	64
Initial exploration variance	2.0
Final exploration variance	0.1
Soft target updates parameter	0.001

The performance in terms of achievable rate vs. transmission channel blocklength is depicted in Fig. 4b. Different values of the error threshold are set to examine the impact. We set $R_{u,n}^{min}$ bps/Hz, and the number of users is set to 6, while the number of antennas is set to $M = 2$ and $M = 4$. As can be observed from the figure, the total achievable rates of all three frameworks increase with the increase of the blocklength. Moreover, it is obvious that DA-DDPG has closer performance to that of SPCA; however, it requires a longer blocklength

to achieve similar performance. Similarly, DRL requires a longer blocklength than that in the case of DA-DDPG to achieve similar performance to that of SPCA. The increase of the error threshold leads to better performance, and the performance gap between DA-DDPG and SPCA slightly decreases as such. However, the performance gap between the DRL and the other two frameworks increases, i.e., for the case of error threshold 0.5×10^{-4}.



(a) Achievable rate versus UAV maximum transmit power.

(b) Achievable rate versus channel blocklength.

(c) Achievable rate for different values of $R_{u,n}^{min}$.

(d) Achievable rate versus number of users.

Figure 4: Performance evaluation of the proposed DA-DDPG framework. **(a)** Achievable rate vs. UAV maximum transmit power, **(b)** Achievable rate vs. channel blocklength, **(c)** Achievable rate for different values of $R_{u,n}^{min}$, and **(d)** Achievable rate vs. number of users

Fig. 4c shows the total achievable rate for different values of $R_{u,n}^{min}$. Different values of the error threshold and number of antennas are considered. The total achievable rate is decreasing with the increasing of $R_{u,n}^{min}$. The performance gap between SPCA and DA-DDPG is small, while it is relatively larger for the case of DRL. The decrease in the total achievable rate becomes more severe for higher $R_{u,n}^{min}$ in the case of DRL. For instance, when $R_{u,n}^{min}$ bps/Hz and $M = 2$, the performance gap between DRL and DA-DDPG is 6.92%, and when $R_{u,n}^{min}$ bps/Hz, the gap is 7.83%. In general, similar conclusions to those in Fig. 4b can be drawn from Fig. 4c.

The relationship between the total achievable rate and the number of users is depicted in Fig. 4d. The number of antennas is fixed at $M = 4$, the error threshold is $\varepsilon^{th} = 5 \times 10^{-3}$, and $R_{u,n}^{min}$ bps/Hz. The performance

in terms of achievable rate increases dramatically with the increase of the number of users. Even with the fluctuation in the performance gap between DA-DDPG and SPCA, on average the gap is smaller compared to that between SPCA and DRL or DA-DDPG and DRL. On the other hand, Table 3 clearly shows the superiority of the proposed DA-DDPG over other frameworks in terms of inference speed. The proposed DA-DDPG proves to be a better choice for large-scale communication systems compared with DRL.

Table 3: Computational time comparison

Approach	$N = 50$		$N = 100$	
	CPU (ms)	GPU (ms)	CPU (ms)	GPU (ms)
SPCA	25.652	10.401	41.112	28.432
DRL	13.753	9.008	16.095	11.265
DA-DDPG	10.104	6.990	11.455	8.998

7 Conclusion

In this work, we proposed a design of a distributed DRL framework with dual-agent deep deterministic policy gradient (DA-DDPG). A novel structure of the beamforming and rate allocation is proposed, where an inception-like mechanism and deep unfolding are employed to build the layers. To apply these two mechanisms, we designed a successive pseudo-convex approximation (SPCA) to handle both beamforming and rate allocation. A fully connected deep neural network is used for the critic network. Successive pseudo-convex approximation (SPCA) lays the ground for applying the above two techniques. The second agent is the UAV trajectory, in which both the actor and critic are based on DNN. Simulation results demonstrated the effectiveness of the proposed framework. Moreover, the proposed model outperforms other well-known methods in the literature. Additionally, developing joint resource allocation frameworks for multiple UAVs and RISs will improve scalability and load balancing, particularly in dense urban areas. Addressing security challenges in UAV-RIS networks will also ensure robust and secure communication in mission-critical applications.

Acknowledgement: The authors extend their appreciation to the Deputyship of Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number RI-44-0291.

Funding Statement: The study was supported by the Deputyship of Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number RI-44-0291.

Author Contributions: The authors confirm their contribution to the paper as follows: Conceptualization, Samia Allaoua Chelloug and Ammar Muthanna; methodology, Dina S. M. Hassan; software, Mohammed Saleh Ali Muthanna; validation, Reem Alkanhel and Samia Allaoua Chelloug; formal analysis, Abuzar B. M. Adam and Dina S. M. Hassan; investigation, Samia Allaoua Chelloug; resources, Ammar Muthanna and Abuzar B. M. Adam; data curation, Mohammed Saleh Ali Muthanna and Abuzar B. M. Adam; writing—original draft preparation, Reem Alkanhel and Abuzar B. M. Adam; writing—review and editing, Samia Allaoua Chelloug; visualization, Dina S. M. Hassan.; supervision, Ammar Muthanna; project administration, Reem Alkanhel; funding acquisition, Samia Allaoua Chelloug. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Chen K, Wang Y, Zhao J, Wang X, Fei Z. URLLC-oriented joint power control and resource allocation in UAV-assisted networks. IEEE Int Things J. 2021;8(12):10103–16. doi:10.1109/jiot.2021.3051322.
2. Zeng C, Wang JB, Xiao M, Ding C, Chen Y, Yu H, et al. Task-oriented semantic communication over rate splitting enabled wireless control systems for URLLC services. IEEE Transact Commun. 2023;72(2):722–39. doi:10.1109/tcomm.2023.3325901.
3. 3GPP. Study on scenarios and requirements for next generation access technologies. 3rd generation partnership project (3GPP), TR 38913; 2022 [Internet]. [cited 2025 Apr 26]. Available from: https://www.etsi.org/deliver/etsi_tr/138900_138999/138913/18.00.00_60/tr_138913v180000p.pdf.
4. Bing L, Gu Y, Aulin T, Wang J. Design of autoconfigurable random access NOMA for URLLC industrial IoT networking. IEEE Transact Indust Informat. 2023;20(1):190–200. doi:10.1109/tii.2023.3257841.
5. Soleymani M, Santamaria I, Jorswieck E, Clerckx B. Optimization of rate-splitting multiple access in beyond diagonal RIS-assisted URLLC systems. IEEE Transact Wireless Commun. 2023;23(5):5063–78. doi:10.1109/twc.2023.3324190.
6. Ou X, Xie X, Lu H, Yang H. Channel blocklength minimization in MU-MISO nonorthogonal multiple access for URLLC services. IEEE Systems J. 2023;18(1):36–9. doi:10.1109/jsyst.2023.3329721.
7. Sun Y, Ding Z, Dai X. On the performance of downlink NOMA in multi-cell mmWave networks. IEEE Commun Lett. 2018;22(11):2366–9. doi:10.1109/lcomm.2018.2870442.
8. Mao Y, Dizdar O, Clerckx B, Schober R, Popovski P, Poor HV. Rate-splitting multiple access: fundamentals, survey, and future research trends. IEEE Commun Surv Tutor. 2022;24(4):2073–126. doi:10.1109/COMST.2022.3191937.
9. Kong C, Lu H. Cooperative rate-splitting multiple access in heterogeneous networks. IEEE Commun Lett. 2023;27(10):2807–11. doi:10.1109/lcomm.2023.3309818.
10. Lyu X, Aditya S, Kim J, Clerckx B. Rate-splitting multiple access: the first prototype and experimental validation of its superiority over SDMA and NOMA. IEEE Transact Wireless Commun. 2024;23(8):9986–10000. doi:10.1109/twc.2024.3367891.
11. Li X, Li J, Liu Y, Ding Z, Nallanathan A. Residual transceiver hardware impairments on cooperative NOMA networks. IEEE Transact Wireless Commun. 2019;19(1):680–95. doi:10.1109/twc.2019.2947670.
12. Pivoto DGS, de Figueiredo FAP, Cavdar C, Tejerina GRDL, Mendes LL. A comprehensive survey of machine learning applied to resource allocation in wireless communications. IEEE Commun Surv Tutor. 2025;1. doi:10.1109/COMST.2025.3552370.
13. Huang Y, Jiang Y, Zheng FC, Zhu P, Wang D, You X. Energy-efficient optimization in user-centric cell-free massive MIMO systems for URLLC with finite blocklength communications. IEEE Transact Vehic Technol. 2024;73(9):12801–14. doi:10.1109/tvt.2024.3382341.
14. Nguyen DC, Cheng P, Ding M, Lopez-Perez D, Pathirana PN, Li J, et al. Enabling AI in future wireless networks: a data life cycle perspective. IEEE Commun Surv Tutor. 2021;23(1):553–95. doi:10.1109/COMST.2020.3024783.
15. Wang D, Liu Y, Yu H, Hou Y. Three-dimensional trajectory and resource allocation optimization in multi-unmanned aerial vehicle multicast system: a multi-agent reinforcement learning method. Drones. 2023;7(10):641. doi:10.3390/drones7100641.
16. Robaglia BM, Coupechoux M, Tsilimantos D. Deep reinforcement learning for uplink scheduling in NOMA-URLLC networks. IEEE Transact Mach Learn Commun Netw. 2024;2(12):1142–58. doi:10.1109/tmlcn.2024.3437351.
17. Ouamri MA, Barb G, Singh D, Adam AB, Muthanna MSA, Li X. Nonlinear energy-harvesting for D2D networks underlaying UAV with SWIPT using MADQN. IEEE Commun Lett. 2023;27(7):1804–8. doi:10.1109/lcomm.2023.3275989.
18. Alkama D, Azni M, Ouamri MA, Li X. Modeling and performance analysis of vertical heterogeneous networks under 3D blockage effects and multiuser MIMO systems. IEEE Transact Vehic Technol. 2024;37(7):10090–105. doi:10.1109/tvt.2024.3366655.
19. Ranjha A, Kaddoum G. Quasi-optimization of distance and blocklength in URLLC aided multi-hop UAV relay links. IEEE Wireless Commun Lett. 2019;9(3):306–10. doi:10.1109/lwc.2019.2953165.

20. Ranjha A, Javed MA, Piran MJ, Asif M, Hussien M, Zeadally S, et al. Towards facilitating power efficient URLLC systems in UAV networks under jittering. IEEE Transact Consumer Elect. 2023;70(1):3031–41. doi:10.1109/tce.2023.3305550.
21. Liu CF, Wickramasinghe ND, Suraweera HA, Bennis M, Debbah M. URLLC-aware proactive UAV placement in internet of vehicles. IEEE Transact Intell Transport Syst. 2024;25(8):10446–51. doi:10.1109/tits.2024.3352971.
22. Cai Y, Jiang X, Liu M, Zhao N, Chen Y, Wang X. Resource allocation for URLLC-oriented two-way UAV relaying. IEEE Transact Vehic Technol. 2022;71(3):3344–9. doi:10.1109/tvt.2022.3143174.
23. Wu Q, Cui M, Zhang G, Wang F, Wu Q, Chu X. Latency minimization for UAV-enabled URLLC-based mobile edge computing systems. IEEE Transact Wireless Commun. 2023;23(4):3298–311. doi:10.1109/TWC.2023.3307154.
24. Feng R, Li Z, Wang Q, Huang J. An ADMM-based optimization method for URLLC-enabled UAV relay system. IEEE Wirel Commun Lett. 2022;11(6):1123–7. doi:10.1109/lwc.2022.3153142.
25. Huang Y, Jiang Y, Zheng FC, Zhu P, Wang D, You X. Enhancing energy-efficient URLLC in cell-free mMIMO systems with transceiver impairments: an RSMA-DRL based approach. IEEE Wirel Commun Lett. 2024;13(5):1443–7. doi:10.1109/lwc.2024.3373826.
26. Seong J, Toka M, Shin W. Sum-rate maximization of RSMA-based aerial communications with energy harvesting: a reinforcement learning approach. IEEE Wirel Commun Lett. 2023;12(10):1741–5. doi:10.1109/lwc.2023.3290372.
27. Guo K, Wu M, Li X, Song H, Kumar N. Deep reinforcement learning and NOMA-based multi-objective RIS-assisted IS-UAV-TNs: trajectory optimization and beamforming design. IEEE Transact Intell Transport Syst. 2023;24(9):10197–210. doi:10.1109/tits.2023.3267607.
28. Wu M, Gao Z, Huang Y, Xiao Z, Ng DWK, Zhang Z. Deep learning-based rate-splitting multiple access for reconfigurable intelligent surface-aided tera-hertz massive MIMO. IEEE J Selec Areas Commun. 2023;41(5):1431–51. doi:10.1109/jsac.2023.3240781.
29. Khalili A, Monfared EM, Zargari S, Javan MR, Mokari N, Jorswieck EA. Resource management for transmit power minimization in UAV-assisted RIS HetNets supported by dual connectivity. IEEE Transact Wirel Commun. 2021;21(3):1806–22. doi:10.1109/twc.2021.3107306.
30. Hua DT, Do QT, Dao NN, Cho S. On sum-rate maximization in downlink UAV-aided RSMA systems. ICT Express. 2024;10(1):15–21. doi:10.1016/j.icte.2023.03.001.
31. Adam ABM, Elhassan MAM. Enhancing Secrecy in UAV RSMA Networks: Deep Unfolding Meets Deep Reinforcement Learning. In: 2023 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI). China: Chongqing: 2015. p. 1–5. doi:10.1109/CCCI58712.2023.10290852.
32. Adam ABM, Wan X, Elhassan MAM, Muthanna MSA, Muthanna A, Kumar N, et al. Intelligent and robust UAV-aided multiuser RIS communication technique with jittering UAV and imperfect hardware constraints. IEEE Trans Veh Technol. 2023;72(8):10737–53. doi:10.1109/tvt.2023.3255309.
33. Adam ABM, Wang Z, Wan X, Xu Y, Duo B. Energy-efficient power allocation in downlink multi-cell multi-carrier NOMA: special deep neural network framework. IEEE Transact Cognit Commun Netw. 2022;8(4):1770–83. doi:10.1109/tccn.2022.3198652.