NOMA-Enhanced Quantized Uplink Multi-user MIMO Communications

Thanh Phung Truong*, Anh-Tien Tran*, Van Dat Tuong*, Nhu-Ngoc Dao†, and Sungrae Cho*

*School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, South Korea.
†Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea.
Email: {tptruong, attran, vdtuong}@uclab.re.kr, nndao@sejong.ac.kr, srcho@cau.ac.kr.

***Abstract*—This research examines quantized uplink multi-user MIMO communication systems with low-resolution quantizers at users and base stations (BS). In such a system, we employ the non-orthogonal multiple access (NOMA) technique for communication between users and the BS to enhance communication performance. To maximize the number of users that satisfy the quality of service (QoS) requirement while minimizing the user's transmit power, we jointly optimize the transmit power and precoding matrices at the users and the digital beamforming matrix at the BS. Owing to the non-convexity of the objective function, we transform the problem into a reinforcement learning-based problem and propose a deep reinforcement learning (DRL) framework named `QNOMA-DRLPA` to overcome the challenge. Because the nature of the action decided by the DRL algorithm may not satisfy the problem constraints, we propose a post-actor process to redesign the actions to meet all the problem constraints. In the simulation, we assess the proposed framework's performance in training convergence and demonstrate its superior performance under various environmental parameters compared with other benchmark schemes.**

***Index Terms*—low-resolution quantizers, multiple-input multiple-output, non-orthogonal multiple access**

I. Introduction

Wireless communication has evolved significantly, and there is a growing demand for low-power, high-speed wireless communication. The Internet of Things (IoT) produces applications that rely on devices with limited battery life and computing capabilities but demand a high spectral efficiency. Therefore, it is crucial to develop efficient and effective wireless communication technologies [1]. Fortunately, low-power hardware components, such as low-resolution digital-to-analog converters (DACs) and analog-to-digital converters (ADCs), are available for adoption. By reducing the number of quantization bits, the power consumption of the quantizers decreases exponentially. This procedure offers a practical way to alleviate power constraints and improve the system's overall energy efficiency without significantly compromising performance [2].

The need for power-efficient wireless communication has led to the widespread use of low-resolution DACs and ADCs in communication systems, particularly in multiple-input multiple-output (MIMO) systems. In [3], the use of low-resolution ADCs has been proposed as a potential solution to effectively decrease power consumption in MIMO systems. This study carefully investigates the influence of ADC resolution, the Rician factor, and the number of antennas on the uplink spectral efficiency, using rigorous mathematical formulations. The findings indicate that using inexpensive and low-resolution techniques may nonetheless provide satisfactory spectral efficiency in huge MIMO systems. The additive quantization noise model (AQNM) was used to linearly approximate the quantized signal in several works as in [4]–[6]. The antenna selection methods were established by the authors of [4] using AQNM to simulate the signal. The researchers focused their attention on enhancing the efficiency of transmit beamforming with the objective of maximizing energy efficiency while minimizing the impact on spectral efficiency. The investigation performed by [5] focused on examining the issue of antenna selection for downlink transmission and uplink reception in a scenario including a multi-antenna base station (BS) and single-antenna mobile stations equipped with low-resolution ADCs. The theoretical study is applicable to the scenario of a wideband Orthogonal Frequency Division Multiplexing (OFDM) system, where all subcarriers use a shared subset of antennas. Additionally, a novel antenna selection approach is proposed in order to optimize capacity while considering the impact of quantization effects. The study conducted by [6] integrates the impact of low-resolution ADCs and limited blocklength channel coding into the optimization of non-orthogonal multiple access (NOMA) in the downlink. This research focuses on a situation where a multiple-antenna access point is responsible for serving numerous single-antenna IoT devices. The attainable rate may be described as a mathematical function that depends on many factors, including the amount of quantization bits, the precoding vectors, the blocklength, and the likelihood of error. The study conducted by [7] examines the problem of cost-effective design for enabling widespread connectivity in cellular IoT applications, specifically focusing on the challenges posed by spatially linked Rician fading channels. The authors obtained analytical formulas for the spectral efficiency of uplink and downlink transmissions, assuming the presence of a large number of IoT devices and the use of low-complexity successive interference

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) grants funded by the Korea government (MSIT) (RS-2023-00209125).

cancellation (SIC) receivers. In addition, they appropriately allocate time for channel estimation to mitigate the decline in performance resulting from the use of low-resolution ADC, as well as the transmission of uplink and downlink data inside a data frame.

Although there has been extensive research on quantized MIMO systems, the application of multiple access techniques in such systems, particularly in uplink communication, has been relatively limited. In this study, we investigate a novel system where the NOMA technique is applied to enhance uplink transmission in a quantized MIMO system. We summarize the main contributions of this study as follows:

- We investigate a novel NOMA-enhanced quantized uplink multi-user MIMO communication system. Here, the multiple-antenna BS and users are adopted with low-resolution quantizers. The NOMA technique is applied to the communication between users and the BS to enhance communication efficiency. Accordingly, we formulate an optimization problem to maximize the number of users meeting the QoS requirement while minimizing the user's transmit power by optimizing the transmit power and precoding matrix at users and digitally received beamforming vector at the BS.
- We transform the problem into a reinforcement learning (RL)-based problem due to the non-convexity in the objective function. Then, we propose a deep reinforcement learning framework, which employs a DRL algorithm to solve the problem. Because the action determined from the DRL algorithm cannot satisfy the problem constraints, we propose a post-actor process that modifies the decided action to meet all the requirements. We name the proposed framework as `QNOMA-DRLPA`.
- We demonstrate the effectiveness of our proposed framework through numerical simulations. We evaluate the convergence of the training algorithm by presenting the training reward and policy loss outcomes. In addition, we prove the effectiveness of `QNOMA-DRLPA` by showing its outperformance compared with other benchmark schemes under different environmental parameters. Besides, we analyze the system performance regarding the change in the number of quantization bits to assess the effect of the device's resolution on communication performance.

We organize the rest of this study as follows. Section II introduces the proposed NOMA-enhanced quantized uplink multi-user MIMO communication system, where we present the problem formulation. In section III, we describe the proposed solution with the detail of `QNOMA-DRLPA` framework. Accordingly, we demonstrate our proposed framework's performance in Section IV. Finally, we conclude the work in Section V.

Notation: Some major specific symbols are utilized to present this article as follows: $\mathcal{CN}(\mu, \sigma^2)$ denotes the circularly symmetric complex Gaussian distribution with variance σ^2 and mean μ; $\mathbf{0}_{r\times c}$ and $\mathbf{1}_{r\times c}$ denote the matrix with r rows and c columns with all elements values are 0 and 1, respectively; $\text{diag}(\mathbf{A})$ denotes the diagonal matrix of $\mathbf{A}$; $\text{Tr}(.)$ and $\mathbb{E}[.]$ denote the trace and expectation operation of matrix, respectively; $(.)^{-1}$, $(.)^H$, and $(.)^T$ denote the inverse, Hermitian, and transpose of matrix, respectively.

TABLE I: Signal-to-quantized-noise ratio

$b_{(T,k,n)/(R,m)}$	1	2	3	4	5
$\rho_{(T,k,n)/(R,m)}$	0.3634	0.1175	0.03454	0.009497	0.002499

II. Problem Statement

A. Quantized Uplink Multi-user MIMO Communications

We examine a single-cell uplink multiuser MIMO system, where a set of K users, $\mathcal{K} \triangleq \{1, 2, \ldots, K\}$, each equipped with N antennas, transmit their signal to the BS equipped with M antennas. The digital baseband signal at user k, $\mathbf{x}_k \in \mathbb{C}^{N\times 1}$, is expressed as

$$\mathbf{x}_k = \sqrt{p_k}\mathbf{f}_k s_k, \tag{1}$$

where p_k, $\mathbf{f}_k \triangleq [f_1, \ldots, f_N]^T \in \mathbb{C}^{N\times 1}$, and s_k denote the transmit power, precoding matrix, and transmit signal, respectively.

At each transmitter, pairs of DACs are employed, each pair with $b_{(T,k,n)}-$bit resolution includes DACs for real and imaginary parts, where $b_{(T,k,n)}$ is the number of DAC's quantization bits at the antenna n of user k. Accordingly, we utilize the AQNM method [8], which approximates the quantization process using a linear representation, the uplink quantized signal from user k, $\mathbf{x}_k^q \in \mathbb{C}^{N\times 1}$, is expressed as [9]

$$\begin{aligned}\mathbf{x}_k^q &= \mathcal{Q}(\mathbf{x}_k) \approx \mathbf{\Theta}_{(\alpha,T,k)}\mathbf{x}_k + \mathbf{n}_{(T,k)}^q \\ &= \sqrt{p_k}\mathbf{\Theta}_{(\alpha,T,k)}\mathbf{f}_k s_k + \mathbf{n}_{(T,k)}^q,\end{aligned} \tag{2}$$

where $\mathcal{Q}(.)$ denotes the quantizer function, $\mathbf{\Theta}_{(\alpha,T,k)} \triangleq \text{diag}(\alpha_{(T,k,1)}, \alpha_{(T,k,2)}, \ldots, \alpha_{(T,k,N)}) \in \mathbb{C}^{N\times N}$ is the quantization loss matrix, and $\mathbf{n}_{(T,k)}^q$ is the additive Gaussian quantization noise vector at user k. The quantization loss element $\alpha_{(T,k,n)}$ is calculated founded on the inverse of the signal-to-quantized-noise ratio (SQNT), $\rho_{(T,k,n)}$, expressed as

$$\alpha_{(T,k,n)} = 1 - \rho_{(T,k,n)}, \tag{3}$$

where $\rho_{(T,k,n)}$ is determined based on $b_{(T,k,n)}$ [10], which is specified in Table I if $b_{(T,k,n)} \leq 5$ and $\rho_{(T,k,n)} = \frac{\pi\sqrt{3}}{2}2^{-2b_{(T,k,n)}}$ otherwise. According to [9], the quantization noise follows $\mathbf{n}_{(T,k)}^q \sim \mathcal{CN}(\mathbf{0}_{N\times 1}, \mathbf{R}_{(T,k)})$, where $\mathbf{R}_{(T,k)} \triangleq \mathbf{\Theta}_{(\alpha,T,k)}\mathbf{\Theta}_{(\rho,T,k)}\text{diag}\left(\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^H\right]\right)$ is the covariance matrix, with $\mathbf{\Theta}_{(\rho,T,k)} \triangleq \text{diag}(\rho_{(T,k,1)}, \rho_{(T,k,2)}, \ldots, \rho_{(T,k,N)}) \in \mathbb{C}^{N\times N}$. Let $P_{(k,max)}$ denote the maximum transmit power of user k, the uplink quantized signal at each user should follow the constraint

$$\text{Tr}\left(\mathbb{E}\left[\mathbf{x}_k^q(\mathbf{x}_k^q)^H\right]\right) \leq P_{(k,max)}. \tag{4}$$

Accordingly, the received signal vector at the BS, $\mathbf{y} \in \mathbb{C}^{M\times 1}$, is expressed as

$$\mathbf{y} = \sum_{k=1}^{K}\mathbf{H}_k\mathbf{x}_k^q + \mathbf{n}, \tag{5}$$

where $\mathbf{H}_k \in \mathbb{C}^{M\times N}$, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_{M\times 1}, \sigma^2\mathbf{1}_{M\times 1})$ denote the uplink channel matrix between $k-$th user and BS, and the additive white Gaussian noise vector, respectively.

Similarly, the received signal at the BS is quantized by the ADCs with $b_{R,m}-$bit resolution. Consequently, the received baseband signal, $\mathbf{y}^q \in \mathbb{C}^{M\times 1}$, is expressed as

$$\begin{aligned}\mathbf{y}^q &= \mathcal{Q}(\mathbf{y}) \approx \mathbf{\Theta}_{(\alpha,R)}\mathbf{y} + \mathbf{n}_R^q \\ &= \mathbf{\Theta}_{(\alpha,R)} \sum_{k=1}^{K} \mathbf{H}_k \left(\sqrt{p_k}\mathbf{\Theta}_{(\alpha,T,k)}\mathbf{f}_k s_k + \mathbf{n}_{(T,k)}^q\right) \\ &\quad + \mathbf{\Theta}_{(\alpha,R)}\mathbf{n} + \mathbf{n}_R^q \\ &= \mathbf{\Theta}_{(\alpha,R)} \sum_{k=1}^{K} \mathbf{H}_k \sqrt{p_k}\mathbf{\Theta}_{(\alpha,T,k)}\mathbf{f}_k s_k \\ &\quad + \mathbf{\Theta}_{(\alpha,R)} \sum_{k=1}^{K} \mathbf{H}_k \mathbf{n}_{(T,k)}^q + \mathbf{\Theta}_{(\alpha,R)}\mathbf{n} + \mathbf{n}_R^q,\end{aligned} \tag{6}$$

where $\mathbf{\Theta}_{(\alpha,R)} \triangleq \text{diag}(\alpha_{(R,1)}, \alpha_{(R,2)}, \ldots, \alpha_{(R,M)}) \in \mathbb{C}^{M\times M}$ is the quantization loss matrix, and $\mathbf{n}_R^q$ is the additive Gaussian quantization noise vector at the BS. The quantization loss element $\alpha_{(R,m)}$ is calculated as $\alpha_{(R,m)} = 1 - \rho_{(R,m)}$, where $\rho_{(R,m)}$ is specified in Table I if $b_{(R,m)} \leq 5$ and $\rho_{(R,m)} = \frac{\pi\sqrt{3}}{2}2^{-2b_{(R,m)}}$ otherwise. The quantization noise $\mathbf{n}_R^q$ follows $\mathcal{CN}(\mathbf{0}_{M\times 1}, \mathbf{R}_R)$, where $\mathbf{R}_R \triangleq \mathbf{\Theta}_{(\alpha,R)}\mathbf{\Theta}_{(\rho,R)}\text{diag}\left(\mathbb{E}\left[\mathbf{y}\mathbf{y}^H\right]\right)$ denotes the covariance matrix, with $\mathbf{\Theta}_{(\rho,R)} \triangleq \text{diag}(\rho_{(R,1)}, \rho_{(R,2)}, \ldots, \rho_{(R,M)}) \in \mathbb{C}^{M\times M}$.

In baseband processing, a received digital beamforming matrix, $\mathbf{W} \triangleq [\mathbf{w}_1, \ldots, \mathbf{w}_K]^T \in \mathbb{C}^{K\times M}$, is applied to detect the uplink signal [11]–[13], where $\mathbf{w}_k \in \mathbb{C}^{1\times M}$ is a normalized beamforming vector to detect the signal of $k-$th user, which satisfies $||\mathbf{w}_k||_2 = 1$. The detected signal is then represented as

$$\begin{aligned}\hat{\mathbf{y}} &= \mathbf{W}\mathbf{y}^q \\ &= \mathbf{W}\mathbf{\Theta}_{(\alpha,R)} \sum_{k=1}^{K} \mathbf{H}_k \sqrt{p_k}\mathbf{\Theta}_{(\alpha,T,k)}\mathbf{f}_k s_k \\ &\quad + \mathbf{W}\mathbf{\Theta}_{(\alpha,R)} \sum_{k=1}^{K} \mathbf{H}_k \mathbf{n}_{(T,k)}^q + \mathbf{W}\mathbf{\Theta}_{(\alpha,R)}\mathbf{n} + \mathbf{W}\mathbf{n}_R^q.\end{aligned} \tag{7}$$

Here, $\hat{\mathbf{y}} \triangleq [\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_K] \in \mathbb{C}^{K\times 1}$, where $\hat{y}_k$ is the detected signal of user k.

B. Problem Formulation

The system uses the NOMA technique to facilitate communication between the users and the BS. Then, the BS applies the SIC technique to decode the individual signals from the detected signal, where the decoding order is ranked based on the effective channels between the users and the BS. Here, the signal from the user having the strongest effective channel is decoded first [14]. Accordingly, the achievable rate of each user k can be calculated as

$$r_k = \mathcal{B}\log\left(1 + \text{SU}_k\left(\text{IU}_k + \text{QE}_k + \text{NU}_k\right)^{-1}\right), \tag{8}$$

Algorithm 1 Achievable rate calculation

1: **Input:**
2: Calculate R_k^{su}, $k \in \mathcal{K}$.
3: **for** $k \in \mathcal{K}$ **do**
4: Set $\text{IU}_k = 0$.
5: **for** $j \in \mathcal{K}\backslash k$ **do**
6: **if** $R_j^{su} < R_k^{su}$ **then**
7: Calculate SU_j as (9a).
8: Update interference: IU_k += SU_j.
9: **end if**
10: **end for**
11: Calculate r_k as (8).
12: **end for**
13: **return** The achievable rate r_k, $k \in \mathcal{K}$.

where B is the communication bandwidth, and

$$\text{SU}_k = p_k\left|\mathbf{w}_k\mathbf{\Theta}_{(\alpha,R)}\mathbf{H}_k\mathbf{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\right|^2, \tag{9a}$$

$$\text{IU}_k = \sum_{j\in\mathcal{IU}_k} p_j\left|\mathbf{w}_k\mathbf{\Theta}_{(\alpha,R)}\mathbf{H}_j\mathbf{\Theta}_{(\alpha,T,j)}\mathbf{f}_j\right|^2, \tag{9b}$$

$$\text{QE}_k = \sum_{i=1}^{K} \mathbf{w}_k\mathbf{\Theta}_{(\alpha,R)}\mathbf{H}_i\mathbf{H}_i^H\mathbf{\Theta}_{(\alpha,R)}^H\mathbf{w}_k^H\mathbf{1}_{1\times N}\mathbf{R}_{(T,i)}, \tag{9c}$$

$$\text{NU}_k = \sigma^2\mathbf{w}_k\mathbf{\Theta}_{(\alpha,R)}\mathbf{\Theta}_{(\alpha,R)}^H\mathbf{w}_k^H + \mathbf{w}_k\mathbf{w}_k^H\mathbf{1}_{1\times M}\mathbf{R}_R, \tag{9d}$$

where $\mathcal{IU}_k$ is the set of users having weaker effective channels than the user k. Without loss of generality, we assume the effective channels of users are determined according to their achievable rate in a single-user system [15]. Therefore, let R_k^{su} denote the achievable rate of user k in the system without any interference from other users, the users are ranked as $R_1^{\text{su}} > R_2^{\text{su}} > \cdots > R_K^{\text{su}}$. R_k^{su} is calculated as

$$R_k^{\text{su}} = \mathcal{B}\log\left(1 + p_k\text{SU}_k\left(\text{QE}_k^{(i=k)} + \text{NU}\right)^{-1}\right), \tag{10}$$

where SU_k, and NU are defined in (9a), and (9d), respectively, $\text{QE}_k^{(i=k)} = \mathbf{w}_k\mathbf{\Theta}_{(\alpha,R)}\mathbf{H}_k\mathbf{H}_k^H\mathbf{\Theta}_{(\alpha,R)}^H\mathbf{w}_k^H\mathbf{1}_{1\times N}\mathbf{R}_{(T,k)}$. Accordingly, the achievable rate of users can be calculated as Algorithm 1. First, the achievable rate of each user in the single-user system is estimated using (10). Then, for the considered user k, the lower-rate users are counted as the interference users (lines 7-12). Consequently, the achievable rate r_k is calculated using (8).

Besides, each user's achievable rate must satisfy the minimum QoS requirement, which is determined according to an achievable rate threshold, r_{th}. Accordingly, we formulate an optimization problem that maximizes the number of users meeting the QoS requirement, i.e., satisfied users, while minimizing the users' transmit power. To do so, we establish a value function as follows:

$$\mathcal{C} = \sum_{k=1}^{K} \epsilon\lambda(r_k - r_{th}) - p_k, \tag{11}$$

where ϵ is an auxiliary variable determining the priority of the number of satisfied users over the transmit power, and $\lambda(x)$

denotes a satisfaction function, expressed as

$$\lambda(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ \eta, & \text{otherwise,} \end{cases} \tag{12}$$

where η is a negative harm value to penalize users breaking the QoS requirement.

Consequently, to maximize the number of satisfied users while minimizing the users' transmit power, we formulate an optimization problem of maximizing the value function by optimizing the transmit power, precoding matrix, and the received digital beamforming matrix. By denoting $\mathbf{p} \triangleq \{p_1, p_2, \ldots, p_K\}$, $\mathbf{F} \triangleq \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_K\}$, the problem is formulated as

$$(P1): \max_{\mathbf{p},\mathbf{F},\mathbf{W}} \ \mathcal{C} \tag{13a}$$

$$\text{s.t. } \mathrm{Tr}\left(\mathbb{E}\left[\mathbf{x}_k^q (\mathbf{x}_k^q)^H\right]\right) \leq P_{(k,max)}, k \in \mathcal{K} \tag{13b}$$

$$||\mathbf{w}_k||_2 = 1, k \in \mathcal{K}, \tag{13c}$$

where (13b) is the constraint of transmit power, and (13c) is the receive beamforming matrix constraint.

The objective function exhibits non-convexity because of the satisfaction function, which poses challenges in finding the global optimum solution and formulating efficient optimization algorithms to solve the problem. In light of the non-convex nature of the problem, developing an approximation approach to find an approximate solution is a practical and feasible choice. Fortunately, RL has gained recognition as a valuable tool for implementing such an approach. By leveraging RL algorithms, it is possible to effectively navigate the complexities of the problem and discover satisfactory solutions. Considering this observation, we address the issue by transforming the problem (P1) into a problem based on the RL model and proposing a DRL framework to tackle the transformed problem.

III. Proposed Deep Reinforcement Learning Framework

In this section, we first transform *(P1)* into an RL-based problem. Then, we propose a DRL framework that leverages a DRL algorithm and a proposed post-actor process to solve it. We illustrate the proposed framework in Fig. 1, which is detailed as follows.

A. RL-based Problem

We formulate the problem *(P1)* as an RL model [16], where the BS assumes the role of the RL agent, while the entire system functions as the environment in which the agent operates. At each time slot t, the state space, action space, and reward function are defined as follows:

Definition 1: The state space contains environment observations, which includes channel matrices from users to the BS. Accordingly, the state space at time slot t is expressed as

$$s[t] = \{\mathbf{H}_1[t], \mathbf{H}_2[t], \ldots, \mathbf{H}_K[t]\}. \tag{14}$$

Definition 2: The action space specifies the actions that the agent needs to decide, including the transmit powers, precoding matrices, and the received digital beamforming matrices. Then, the action space at time slot t is expressed as

$$a[t] = \{\mathbf{p}[t], \mathbf{F}[t], \mathbf{W}[t], k \in \mathcal{K}\}. \tag{15}$$

Definition 3: The reward function is calculated according to the objective function, with a focus on maximizing value function $\mathcal{C}$. To achieve this, we formulate it founded on the value function, which is calculated as

$$r[t] = \sum_{k=1}^{K} \epsilon \lambda(r_k - r_{th}) - p_k \tag{16}$$

Accordingly, the RL-based problem is formulated to optimize long-term return by determining the best action at each state in the RL model while considering the constraints in (13). Then, we formulate the RL-based problem as

$$(P2): \max_{a[t]} \ \sum_{l=t}^{T} r[t]\gamma^{l-t} \tag{17a}$$

$$\text{s.t. (13b), (13c),} \tag{17b}$$

where T is the number of examined time slots and γ is the discount factor.

B. DRL Training Algorithm

To decide the suitable RL action, we employ a widely-used DRL algorithm named Deep Deterministic Policy Gradient (DDPG) in our framework. DDPG utilizes critic and actor networks, each comprising both main and target networks, to facilitate the decision-making process [17]. Each network in the DRL framework is implemented using a neural network. The main actor network, $\mu(s|\theta^\mu)$, where θ^μ is the network's parameter, plays the decision-maker role, trained to map observation state $s[t]$ to corresponding action $a[t]$. To measure the chosen action, DDPG uses the main critic network, $Q(s, a|\theta^Q)$, where θ^Q is the network's parameter, to estimate action-value function for the chosen action $a[t]$ at state $s[t]$. Accordingly, the main actor network's parameter is updated based on the policy gradient ascent function as

$$\nabla_{\theta^\mu} J = \frac{1}{B} \sum_{b=1}^{B} \left(\nabla_a Q(s, a|\theta^Q)|_{s=s_b, a=\mu(s_b)} \nabla_{\theta^\mu} \mu(s_b|\theta^\mu)\right), \tag{18}$$

where B is the mini-batch size of training samples. The parameter of the main critic network is then updated by using gradient descent on the loss function as

$$L = \frac{1}{B} \sum_{b=1}^{B} \left(Q(s_b, a_b|\theta^Q) - y_b\right)^2, \tag{19}$$

where $y_b = r_b + \gamma Q'\left(s'_b, \mu'(s'_b|\theta^{\mu'})|\theta^{Q'}\right)$ with s'_b is the next state of $b-$th sample, $\mu'(s|\theta^{\mu'})$ and $Q'(s, a|\theta^{Q'})$ are the target actor and critic networks, respectively. The corresponding

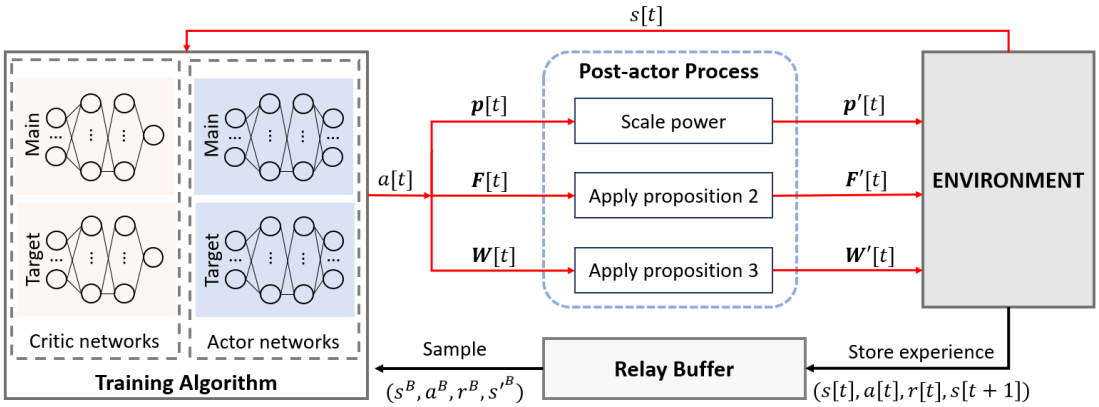


Fig. 1: Proposed DRL framework.

parameters $\theta^{\mu'}$ and $\theta^{Q'}$ are updated by a soft-update with a parameter $\tau \in [0,1]$, given as

$$\begin{aligned} \theta^{\mu'} &\leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'}, \\ \theta^{Q'} &\leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'}. \end{aligned} \tag{20}$$

The exploration in the training process is ensured by adding noise to the action. In DDPG, the noise is generated founded on the Ornstein-Uhlenbeck process [18]. Let $\mathcal{ON}[t]$ denote the noise at time slot t, the action decided by the DDPG algorithm in training is expressed as

$$a[t] = \mu(s[t]|\theta^{\mu}) + \mathcal{ON}[t]. \tag{21}$$

C. Post-actor Process

By employing the DDPG algorithm, the actor network can decide the appropriate actions $\mathbf{p}$, $\mathbf{F}$, and $\mathbf{W}$. However, these actions may not comply with the constraints stated in (17b), which violates the environmental requirements. To address this issue, we propose a post-actor process to guarantee that all constraints are maintained throughout the system.

The action value range in the DRL algorithm can be decided by selecting the activation function in the actor network. To determine a suitable range of values for actions, we first reconsider the problem constraints by proposing the following proposition:

Proposition 1: With the quantization loss matrix $\mathbf{\Theta}_{(\alpha,T,k)} \triangleq \text{diag}(\alpha_{(T,k,1)}, \alpha_{(T,k,2)}, \ldots, \alpha_{(T,k,N)})$, where $\alpha_{(T,k,n)} = 1 - \rho_{(T,k,n)}$, $n \in \{1,\ldots,N\}$ are real numbers, constraint of transmit power in (13b) can be reformulated as

$$p_k \le P_{(k,max)}, k \in \mathcal{K}, \tag{22a}$$

$$\text{Tr}\big(\mathbf{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\mathbf{f}_k^H\big) = 1, k \in \mathcal{K}. \tag{22b}$$

Proof: Please see Appendix A. ■

Accordingly, we normalize the action value to the range of $[0,1]$ and introduce a scaled value of p_k as the new transmit power of user k that satisfies constraint (22a), denoted as p'_k, which is calculated as

$$p'_k = p_k P_{(k,max)}, \tag{23}$$

where $p_k \in [0,1]$, i.e., $p'_k \in [0, P_{(k,max)}]$. Let $\mathbf{p}' \triangleq \{p'_1, \ldots, p'_K\}$ represent the new transmit power set. By applying this set, the constraint (22a) is satisfied. Then, the remaining constraints are (22b) and (13c). To ensure constraint (22b) is satisfied, we propose a new precoding matrix $\mathbf{f}'_k$ for each user k, which is calculated according to $\mathbf{f}_k$. In particular, we introduce the following proposition:

Proposition 2: Let $\mathbf{f}'_k \triangleq [f'_1, \ldots, f'_N]^T \in \mathbb{C}^{N\times 1}$ represent the precoding matrix of user k that satisfies constraint (22b), i.e., $\text{Tr}\big(\mathbf{\Theta}_{(\alpha,T,k)}\mathbf{f}'_k(\mathbf{f}'_k)^H\big) = 1$. Its $n-$th element, f'_n, is calculated as

$$f'_n = \frac{f_n}{\sqrt{\sum_{n=1}^{N} \alpha_{(T,k,n)}|f_n|^2}} \tag{24}$$

Proof: Please see Appendix B. ■

By applying Proposition 2, the new precoding matrix, $\mathbf{F}' \triangleq \{\mathbf{f}'_1, \ldots, \mathbf{f}'_K\}$, is obtained that meets the constraint (22b). Then, the only remaining constraint is (13c). To achieve this requirement, we introduce the following proposition, which defines the function for recalculating the received digital beamforming matrix to the appropriate values:

Proposition 3: Let $\mathbf{w}'_k \triangleq [w'_{k,1}, \ldots, w'_{k,M}] \in \mathbb{C}^M$ denote the normalized beamforming vector that satisfies constraint (13c), i.e., $||\mathbf{w}'_k||_2 = 1$. The $m-$th element of $\mathbf{w}'_k$ is calculated as

$$w'_{k,m}, = \frac{w_{k,m}}{\sqrt{\sum_{m=1}^{M}|w_{k,m}|^2}}. \tag{25}$$

Proof: The square norm of $\mathbf{w}'_k$ is expressed as

$$||\mathbf{w}'_k||_2 = \sqrt{\sum_{m=1}^{M}|w'_{k,m}|^2} \tag{26}$$

Replacing $w'_{k,m}$ by (25), we obtain

$$\begin{aligned} ||\mathbf{w}'_k||_2 &= \sqrt{\sum_{m=1}^{M}\left|\frac{w_{k,m}}{\sqrt{\sum_{m=1}^{M}|w_{k,m}|^2}}\right|^2} \\ &= \sqrt{\sum_{m=1}^{M}\frac{|w_{k,m}|^2}{\sum_{m=1}^{M}|w_{k,m}|^2}} = 1. \end{aligned} \tag{27}$$

Algorithm 2 Proposed DRL-based algorithm

1: Set up algorithm parameters.
2: **while** $e < E$ **do**
3: **for** t from 1 to T **do**
4: Observe environment state $s[t]$.
5: Select $a[t]$ as in (21).
6: Normalize action into the range of $[0, 1]$.
7: **for** $k \in \mathcal{K}$ **do**
8: Scale transmit power $p_k[t] \leftarrow p'_k[t]$ as (23).
9: Calculate new precoding matrix $\mathbf{f}'_k[t]$ as Proposition 2.
10: Calculate new received beamforming vector $\mathbf{w}'_k[t]$ as Proposition 3.
11: **end for**
12: Perform $\mathbf{p}'[t]$, $\mathbf{F}'[t]$, $\mathbf{W}'[t]$, get state-next $s[t+1]$, reward $r[t]$.
13: Store $(s[t], a[t], r[t], s[t+1])$ in buffer.
14: Update new state $s[t+1] \rightarrow s[t]$.
15: Randomly choose a batch of samples from buffer, (s^B, a^B, r^B, s'^B)
16: Training neural networks as sub-section III-B.
17: **end for**
18: **end while**
19: **return** the trained main actor network, $\mu^*(s|\theta^{\mu^*})$.

This completes the proof. ∎

By applying Proposition 3, a new received beamforming matrix $\mathbf{W}' \triangleq \{\mathbf{w}'_1, \ldots, \mathbf{w}'_M\}$ is obtained, which satisfies constraint (13c). Consequently, all constraints in the problem $(P2)$ are satisfied.

D. Framework Formulation

Our proposed framework utilizes the DDPG algorithm in conjunction with the proposed post-actor process, referred to as QNOMA-DRLPA, to address the optimization problem. The entire proposed algorithm is shown in Algorithm 2. The algorithm takes place in E episodes, each with T time steps. In each time step t, the agent decides action $a[t]$ according to the observed state $s[t]$. Then, the proposed post-actor process is applied (lines 6-11) to deal with the problem constraints. Accordingly, the new actions, including $\mathbf{p}'[t]$, $\mathbf{F}'[t]$, and $\mathbf{W}'[t]$, are performed to the environment. Here, the users' achievable rates are calculated according to Algorithm 1, the reward $r[t]$ is then calculated as (16), and the environment state is updated to the next state $s[t+1]$. Consequently, an experienced sample combined from $s[t]$, $a[t]$, $r[t]$, and $s[t+1]$ is pushed into the buffer for training. To train the neural networks, a batch of samples is randomly taken out from the replay buffer, and the training process is executed based on the DDPG algorithm as introduced in sub-section III-B.

IV. Simulation Results

A. Simulation Setting

To assess the proposed framework's performance, we conduct simulations in an environment where the BS serves 10 users randomly distributed within a range of 10 to 200 meters from the BS. The channel matrices between the BS and users, $\mathbf{H}_k$, $k \in \mathcal{K}$, are generated as [19]

$$\mathbf{H}_k = \hat{\mathbf{H}}_k \sqrt{10^{-L_k/10}}, \tag{28}$$

where $L_k = 103.8 + 20.9 \log_{10}(d_k)$ denote the path loss in dB between the BS and user k at the distance d_k (Kilometer); the matrix $\hat{\mathbf{H}}_k$ represents the small-scale fading, where each element is independently distributed as $\mathcal{CN}(0, 1)$. The users are equipped with 4 antennas, while the number of antennas in BS is 8. We set the communication bandwidth $\mathcal{B} = 10$ MHz, $\sigma^2 = -174$ dBm/Hz, $\epsilon = 10$, and $\eta = 2$. The time step duration is 0.1 (s). The neural networks in the DRL algorithm are deployed with two hidden layers, each has 512 nodes. The batch size, buffer size, and discount factor in the algorithm are set to 16, 10^5, and 0.999, respectively. The training is executed in 3000 episodes, each with 300 time steps. In this simulation, we consider a homogeneous quantization resolution system, where the antennas in each node have the same DAC/ADC. The maximum transmit power, $P_{(k,max)}$, is from 0 to 10 (dBm). The achievable rate threshold, r_{th}, is from 2 to 10 (Mbits/s). And the number of quantization bits in each node varies from 2 to 10 bits. To assess performance, we rely on two primary measures. The performance value represents the reward received during each time step, and the percentage of satisfied users reflects the proportion of users meeting the QoS constraint about the total number of users being evaluated.

Besides, we evaluate the following schemes to compare with the proposed QNOMA-DRLPA framework's performance:

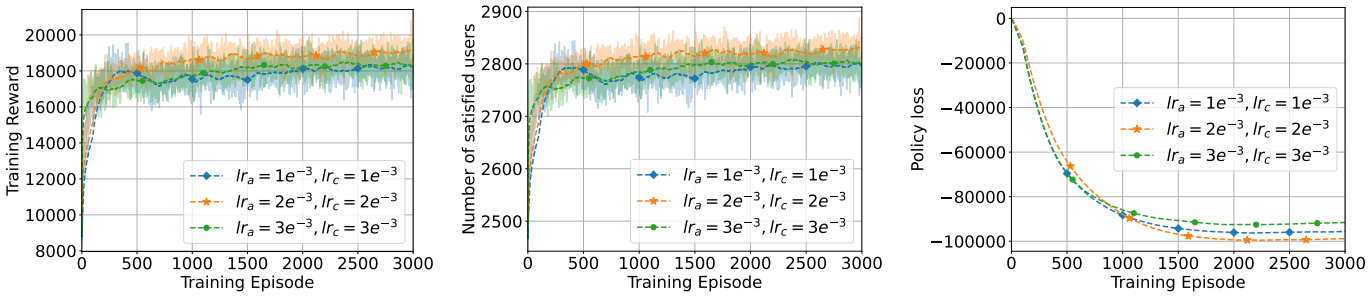
- Quantization with OMA (QOMA): We compare the performance of the NOMA and orthogonal multiple access (OMA) schemes in the quantized uplink multi-user MIMO system. Based on [20], we simulate an OMA system using the FDMA technique, where each user is allocated a dedicated part of the bandwidth, and interference is set to zero.
- Quantization without multiple access technique (QWoMA): In this scheme, multiple access techniques are not employed in the communications. Consequently, signals from other users are fully treated as interference when considering the transmission of user k. According to [13] (sub-section 4.1), the value of IU_k in (9b) is calculated as:

$$\text{IU}_k = \sum_{j \in \mathcal{K} \setminus k} p_j \left| \mathbf{w}_k \mathbf{\Theta}_{(\alpha,R)} \mathbf{H}_j \mathbf{\Theta}_{(\alpha,T,j)} \mathbf{f}_j \right|^2. \tag{29}$$

- Discrete searching algorithm (DSA): We transform continuous actions into discrete spaces and select the optimal action at each time step. However, when there are numerous options, it is unfeasible to explore all potential actions. Therefore, we implement a low-complexity search strategy based on a greedy algorithm to overcome this challenge and determine the action yielding the highest reward. This approach can be considered a local-optimal method.

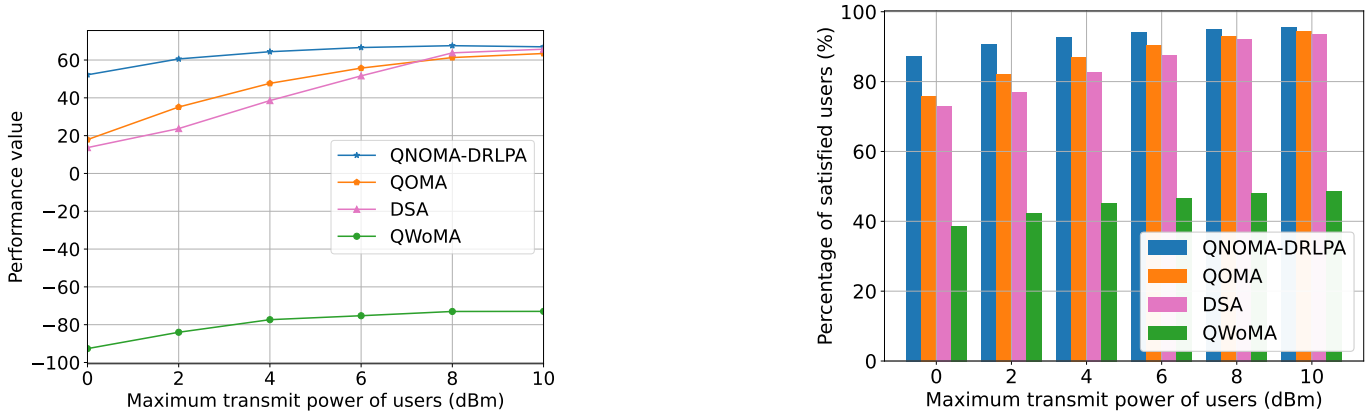
B. Performance Evaluation

First, we estimate the convergence of the proposed QNOMA-DRLPA framework by training it with different learning rate values. The actor learning rate (lr_a) and critic learning rate (lr_c) are chosen from three values of 0.001, 0.002, and



(a) Training reward. (b) Number of satisfied users. (c) Policy loss.

Fig. 2: Training results with different learning rate.



(a) Performance of value function. (b) Percentage of satisfied users.

Fig. 3: Performance with different value of maximum transmit powers.

TABLE II: Percentage of satisfied users with different resolutions.

(a) DAC resolution

Number of DAC's quantization bits	2	3	4	5	6	7	8	9	10
Percentage of satisfied users (%)	56.34	85.59	95.43	98.04	98.72	98.84	98.99	98.92	99.05

(b) ADC resolution

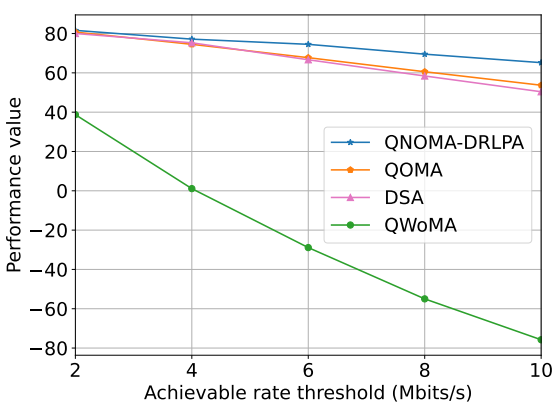
Number of ADC's quantization bits	2	3	4	5	6	7	8	9	10
Percentage of satisfied users (%)	36.86	74.89	91.02	96.17	97.74	97.91	98.07	98.07	98.05

0.003. As shown in Fig. 2a, the case when $lr_a = lr_c = 0.002$ gives the best reward, where its value increases in accordance with the increase in the number of satisfied users in Fig. 2b. Besides, we examine a policy loss to see how the actor network's parameter is updated. According to the gradient ascent in (18), the policy loss at each training step, denoted as GL_μ, is measured by

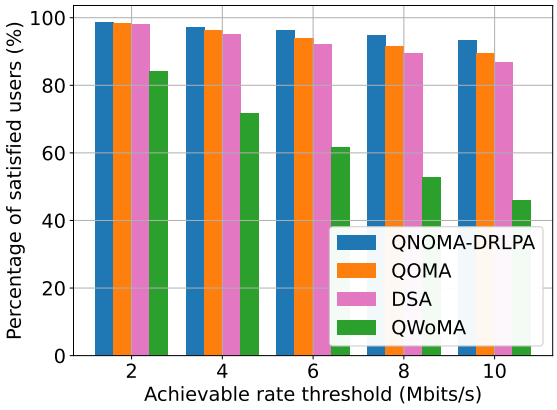
$$GL_\mu = -\frac{1}{B}\sum_{b=1}^{B}\left(Q(s_b, \mu(s_b|\theta^\mu))\right). \tag{30}$$

As illustrated in Fig. 2c, the policy loss fades during the training process, where it finally reaches convergence after approximately 2000 training episodes. Consequently, we select the trained model in the case of $lr_a = lr_c = 0.002$ after training in 3000 episodes to evaluate the framework's performance.

Second, we evaluate the framework's performance with different environmental parameters by compared with other benchmark schemes. In Fig. 3, we assess the system's performance by varying the maximum transmit power of users from 0 to 10 (dBm), the numbers of quantization bits are set to 4 and 10 at the DACs and ADC, respectively, and the achievable rate threshold is 10 (Mbits/s). As a result, our proposed `QNOMA-DRLPA` framework outperforms `QOMA`, `DSA`, and `QWoMA`, where the percentage of satisfied users is always higher than 85 %. Besides, the `QWoMA` scheme yields the worst result because, in the absence of multiple access techniques, the interference experienced by other users becomes high in a multi-user system. Combined with the quantization loss, it reduces transmission quality, adversely affecting overall system's performance. In Fig. 4, we evaluate

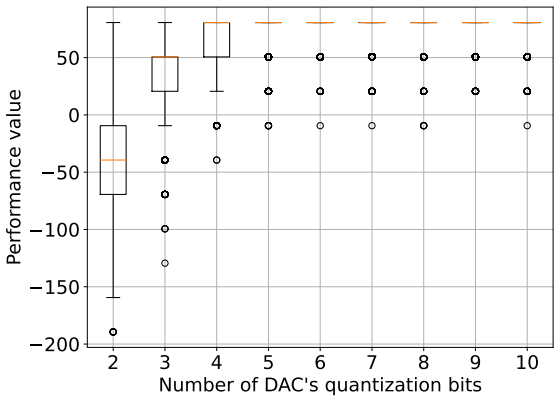


(a) Performance of value function.

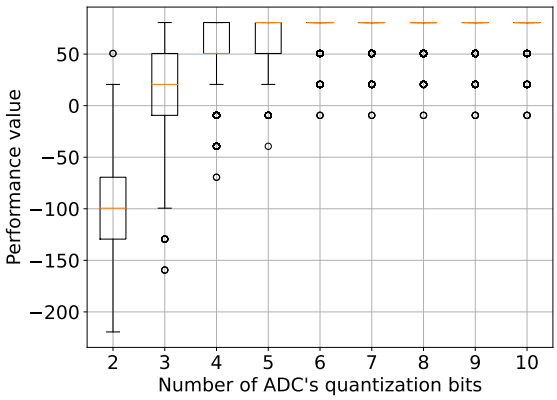


(b) Percentage of satisfied users.

Fig. 4: Performance with different value of achievable rate thresholds.



(a) DAC resolutions.



(b) ADC resolutions.

Fig. 5: Performance with different resolutions.

our proposed framework under different application requirements by varying the achievable rate threshold value from 2 to 10 (Mbits/s). The result shows that our proposed framework performs exceptionally well in all cases, achieving a high percentage of satisfied users, up to 99 % in the case of $r_{th} = 2$ (Mbits/s). In addition, increasing the threshold requirement poses a significant challenge for systems employing the OMA or no multiple access techniques. As the threshold expands, the system's performance gradually declines, with reductions of approximately 7.7 %, 9.1 %, 10.6 %, and 11.3 % when the threshold increases from 2 to 4, 4 to 6, 6 to 8, and 8 to 10 (Mbits/s), respectively, in QOMA. Also, a local-optimal method like DSA faces the same issue as QOMA, while QNOMA−DRLPA only reduces about 5-6 % after each increase of r_{th}. Consequently, our proposed scheme consistently outperforms other benchmark schemes in all simulation cases, demonstrating its superiority and effectiveness in handling varying threshold requirements and achieving high system's performance.

Next, we analyze the framework's performance in different numbers of quantization bits. In Fig. 5b, we evaluate the framework when modifying the number of DAC's quantization bits from 2 to 10 bits, where the number of ADC's quantization bits is 10, the maximum powers of users are 10 (dBm), and $r_{th} = 10$ (Mbits/s). In the worst case (2 bits), the performance value varies a lot, and its percentage of satisfied users is only 56.34 %, shown in Table IIa, due to the high value of quantization loss. When increasing the resolution, the performance gets better, and the framework performs excellently when the number of DAC's quantization bits is greater than 4, where its percentage of satisfied users is approximately 99 %. Then, we fixed the resolution of DACs to 5 bits and changed the number of ADC's quantization bits from 2 to 10 to observe results in Fig. 5a and Table IIb. Similarly, the lowest resolution gives the worst performance, where the performance value ranges from -220 to 20, and the percentage of satisfied users is about 36.86 %. The framework performs well and is stable when the resolution is greater than 5 bits, where its percentage of satisfied users is about 98 %. As a result, the resolution of communication devices significantly impacts communication performance. While developing efficient and effective wireless communication technologies is crucial, it is equally important to consider the trade-off between network requirements and

device cost during the design and manufacturing of telecommunications equipment.

V. Conclusion

Our research focused on improving the performance of a quantized uplink multi-user MIMO communication system by applying the NOMA technique. Specifically, we aimed to optimize the transmit power and precoding matrix at users and the received beamforming matrix at the BS. The objective was to maximize the number of users meeting the QoS requirement while minimizing the user's transmit power. We faced a challenge with the objective function not being convex. To tackle this, we transformed the problem into an RL-based problem and proposed a DRL framework named `QNOMA-DRLPA`, which employs a well-known DRL algorithm named DDPG, to resolve it. However, the DDPG cannot handle constraints in the problem, so we proposed a post-actor process that recalculates the value of actions decided by the DDPG to meet all the problem constraints. In the simulation, we demonstrated the convergence of the DRL training algorithm by examining the training reward and the policy loss values. In addition, we proved the superior performance of the `QNOMA-DRLPA` framework compared to benchmark schemes in different environmental parameters. In addition, we evaluated the performance under different resolutions, where we concluded the impact of resolution on the communication system. Besides, the suggested system holds appeal for numerous research endeavors, such as exploring the system's energy efficiency, integrating it with other next-gen communication techniques like mobile edge computing and reconfigurable intelligent surfaces, and considering alternative multiple access schemes like rate splitting multiple access.

Appendix A
Proof of Proposition 1

The left side of constraint (13b) can be reformulated as

$$\begin{aligned}\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{x}_k^q(\mathbf{x}_k^q)^H\right]\right) = \mathrm{Tr}\left(p_k\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\mathbf{f}_k^H\boldsymbol{\Theta}_{(\alpha,T,k)}^H\right) \\ + ||\mathbf{R}_{(T,k)}||_1,\end{aligned} \tag{31}$$

where $\mathbf{R}_{(T,k)}$ is calculated as

$$\begin{aligned}\mathbf{R}_{(T,k)} &= \boldsymbol{\Theta}_{(\alpha,T,k)}\boldsymbol{\Theta}_{(\rho,T,k)}\mathrm{diag}\left(\mathbb{E}\left[\mathbf{x}_k\mathbf{x}_k^H\right]\right) \\ &= \boldsymbol{\Theta}_{(\alpha,T,k)}\boldsymbol{\Theta}_{(\rho,T,k)}\mathrm{diag}\left(p_k\mathbf{f}_k\mathbf{f}_k^H\right).\end{aligned} \tag{32}$$

Besides, from (3), we can obtain:

$$\boldsymbol{\Theta}_{(\rho,T,k)} = \mathbf{I}_N - \boldsymbol{\Theta}_{(\alpha,T,k)}. \tag{33}$$

Therefore, the left side of constraint (13b) becomes:

$$\begin{aligned}&\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{x}_k^q(\mathbf{x}_k^q)^H\right]\right) \\ &= \mathrm{Tr}\big(p_k\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\mathbf{f}_k^H\boldsymbol{\Theta}_{(\alpha,T,k)}^H \\ &\quad + p_k\boldsymbol{\Theta}_{(\alpha,T,k)}(\mathbf{I}_N - \boldsymbol{\Theta}_{(\alpha,T,k)})\mathbf{f}_k\mathbf{f}_k^H\big) \\ &= p_k\mathrm{Tr}\big(\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\mathbf{f}_k^H\boldsymbol{\Theta}_{(\alpha,T,k)}^H\big) + p_k\mathrm{Tr}\big(\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\mathbf{f}_k^H\big) \\ &\quad - p_k\mathrm{Tr}\big(\boldsymbol{\Theta}_{(\alpha,T,k)}\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\mathbf{f}_k^H\big)\end{aligned} \tag{34}$$

Due to the diagonal matrix $\boldsymbol{\Theta}_{(\alpha,T,k)} \triangleq \mathrm{diag}(\alpha_{(T,k,1)}, \alpha_{(T,k,2)}, \ldots, \alpha_{(T,k,N)})$, with $\alpha_{(T,k,n)}$, $n \in \{1, \ldots, N\}$ are real numbers, we can perform:

$$\begin{aligned}&\mathrm{Tr}\big(\boldsymbol{\Theta}_{(\alpha,T,k)}\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\mathbf{f}_k^H\big) = \mathrm{Tr}\big(\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\mathbf{f}_k^H\boldsymbol{\Theta}_{(\alpha,T,k)}\big) \\ &\quad = \mathrm{Tr}\big(\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\mathbf{f}_k^H\boldsymbol{\Theta}_{(\alpha,T,k)}^H\big).\end{aligned} \tag{35}$$

Accordingly. the equation (34) is simplified as

$$\mathrm{Tr}\left(\mathbb{E}\left[\mathbf{x}_k^q(\mathbf{x}_k^q)^H\right]\right) = p_k\mathrm{Tr}\big(\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\mathbf{f}_k^H\big). \tag{36}$$

Consequently, constraint (13b) is rewritten as

$$p_k\mathrm{Tr}\big(\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\mathbf{f}_k^H\big) \le P_{(k,max)}, \tag{37}$$

where we can split it into two sub-constraints as

$$\begin{cases} p_k \le P_{(k,max)}, \\ \mathrm{Tr}\big(\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k\mathbf{f}_k^H\big) = 1. \end{cases} \tag{38}$$

This completes the proof.

Appendix B
Proof of Proposition 2

To prove $\mathrm{Tr}\big(\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k'(\mathbf{f}_k')^H\big) = 1$, we first perform the matrix multiplication in the trace operator of the left side, which is represented as

$$\begin{aligned}&\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k'(\mathbf{f}_k')^H \\ &= \mathrm{diag}(\alpha_{(T,k,1)}, \ldots, \alpha_{(T,k,N)}) \begin{bmatrix} f_1' \\ \cdots \\ f_N' \end{bmatrix} [\bar{f_1'}, \ldots, \bar{f_N'}] \\ &= \begin{bmatrix} \alpha_{(T,k,1)}f_1' \\ \cdots \\ \alpha_{(T,k,N)}f_N' \end{bmatrix} [\bar{f_1'}, \ldots, \bar{f_N'}] \\ &= \begin{bmatrix} \alpha_{(T,k,1)}|f_1'|^2 & \alpha_{(T,k,1)}f_1'\bar{f_2'} & \cdots & \cdots \\ \alpha_{(T,k,2)}f_2'f_1' & \alpha_{(T,k,2)}|f_2'|^2 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{(T,k,N)}f_N'\bar{f_1'} & \alpha_{(T,k,N)}f_N'\bar{f_2'} & \cdots & \alpha_{(T,k,N)}|f_N'|^2 \end{bmatrix}.\end{aligned} \tag{39}$$

Then, the trace operator is calculated as

$$\begin{aligned}\mathrm{Tr}\big(\boldsymbol{\Theta}_{(\alpha,T,k)}\mathbf{f}_k'(\mathbf{f}_k')^H\big) &= \sum_{n=1}^{N}\alpha_{(T,k,n)}|f_n'|^2 \\ &= \sum_{n=1}^{N}\alpha_{(T,k,n)}\left|\frac{f_n}{\sqrt{\sum_{n=1}^{N}\alpha_{(T,k,n)}|f_n|^2}}\right|^2 \\ &= \frac{\sum_{n=1}^{N}\alpha_{(T,k,n)}\,|f_n|^2}{\left|\sqrt{\sum_{n=1}^{N}\alpha_{(T,k,n)}|f_n|^2}\right|^2} = 1.\end{aligned} \tag{40}$$

This completes the proof.

REFERENCES

[1] S. Pattar, R. Buyya, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, "Searching for the IoT resources: Fundamentals, requirements, comprehensive review, and future directions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2101–2132, 2018.

[2] J. Choi, J. Park, and N. Lee, "Energy efficiency maximization precoding for quantized massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 9, pp. 6803–6817, 2022.

[3] J. Zhang, L. Dai, S. Sun, and Z. Wang, "On the spectral efficiency of massive MIMO systems with low-resolution ADCs," *IEEE Communications Letters*, vol. 20, no. 5, pp. 842–845, 2016.

[4] E. Vlachos and J. Thompson, "Energy-efficiency maximization of hybrid massive MIMO precoding with random-resolution DACs via RF selection," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1093–1104, 2020.

[5] J. Choi, J. Sung, N. Prasad, X.-F. Qi, B. L. Evans, and A. Gatherer, "Base station antenna selection for low-resolution ADC systems," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1951–1965, 2019.

[6] S. Kim, J. Choi, and J. Park, "Downlink NOMA for short-packet internet of things communications with low-resolution ADCs," *IEEE Internet of Things Journal*, vol. 10, no. 7, pp. 6126–6139, 2022.

[7] G. Yu, X. Chen, and D. W. K. Ng, "Low-cost design of massive access for cellular internet of things," *IEEE Transactions on Communications*, vol. 67, no. 11, pp. 8008–8020, 2019.

[8] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, "Robust predictive quantization: Analysis and design via convex optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 618–632, 2007.

[9] S. Park, J. Choi, J. Park, W. Shin, and B. Clerckx, "Rate-splitting multiple access for quantized multiuser MIMO communications," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[10] L. Fan, S. Jin, C.-K. Wen, and H. Zhang, "Uplink achievable rate for massive MIMO systems with low-resolution ADC," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2186–2189, 2015.

[11] G. Li, M. Zeng, D. Mishra, L. Hao, Z. Ma, and O. A. Dobre, "Energy-efficient design for IRS-empowered uplink MIMO-NOMA systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 9490–9500, 2022.

[12] Y. Ma, S. Ren, Z. Quan, and Z. Feng, "Data-driven hybrid beamforming for uplink multi-user MIMO in mobile millimeter-wave systems," *IEEE Transactions on Wireless Communications*, vol. 21, no. 11, pp. 9341–9350, 2022.

[13] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017. [Online]. Available: http://dx.doi.org/10.1561/2000000093

[14] W. Hao, M. Zeng, G. Sun, O. Muta, O. A. Dobre, S. Yang, and H. Gacanin, "Codebook-based max–min energy-efficient resource allocation for uplink mmwave MIMO-NOMA systems," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8303–8314, 2019.

[15] 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.

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

[17] 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.

[18] 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

[19] 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.

[20] Z. Wei, L. Yang, D. W. K. Ng, J. Yuan, and L. Hanzo, "On the performance gain of NOMA over OMA in uplink communication systems," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 536–568, 2020.