Energy Efficiency in Quantized MIMO-NOMA Communication Systems

Thanh Phung Truong, Nam-Phuong Tran, Junsuk Oh, Donghyun Lee, Van Dat Tuong, Nhu-Ngoc Dao, and Sungrae Cho

***Abstract*—Reducing power consumption while maintaining high spectral efficiency is crucial in modern communication systems, especially given the rapid expansion of network services and devices. To catch up on this challenge, we explore an energy-efficiency problem in an uplink quantized MIMO (Multiple-Input Multiple-Output) system empowered by NOMA (Non-Orthogonal Multiple Access) in this work. In this context, employing low-resolution quantizers can significantly reduce power consumption, while NOMA enhances spectral efficiency in MIMO systems. Here, we propose a post-actor-added deep deterministic policy gradient (P-DDPG) algorithm that incorporates a novel post-actor process into the DDPG training algorithm to optimize the users' precoding and base station's detection matrices while considering their constraints. By leveraging the P-DDPG algorithm, we aim to enhance the system's energy efficiency, boosting spectral efficiency while minimizing power consumption. Additionally, we design a power-exhaustive searching function added to the trained model in the inference process, ensuring device diversity. Numerical results confirm the algorithm's convergence and demonstrate its superior performance under various environmental conditions compared to benchmark schemes. Furthermore, we thoroughly analyze the impact of low-resolution quantizers on system performance.**

***Index Terms*—low-resolution quantizers, multiple-input multiple-output systems, non-orthogonal multiple access, quantized networks.**

I. Introduction

MIMO (Multiple-Input Multiple-Output) networks are a key technology in modern wireless communication systems. They significantly enhance performance and capacity compared to traditional single-antenna systems by utilizing multiple antennas at both the transmitter and receiver. This enhancement is achieved through spatial diversity and multiplexing, which improve the transmission rate and spectral efficiency [1]–[3]. However, a significant challenge facing MIMO systems, particularly as they scale up, is their high power consumption. The multiple radio frequency (RF) chains required for each antenna, including power amplifiers, analog-to-digital converters (ADCs), and digital-to-analog converters (DACs), contribute to increased energy usage [4]. The high power consumption not only impacts the operational costs but also poses challenges for battery life in mobile devices and for deploying MIMO in energy-constrained scenarios.

Thanh Phung Truong, Nam-Phuong Tran, Junsuk Oh, Donghyun Lee, Van Dat Tuong, and Sungrae Cho are with the School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, Republic of Korea (e-mail: {tptruong, tnphuong, jsoh, dhlee, vdtuong}@uclab.re.kr; srcho@cau.ac.kr).

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

Moreover, the demand for energy-efficient and rapid wireless communication has become increasingly crucial in today's technological landscape, particularly in emerging fields like the Internet of Things (IoT). Devices in these fields often face limited battery life or modest energy resources. Despite these limitations, achieving high spectral efficiency remains essential for their operation [5]. To address this challenge, integrating energy-efficient hardware components, such as low-resolution quantizers, into transceivers presents a viable strategy for reducing power consumption. Employing low-power hardware solutions balances operational longevity and performance, especially in scenarios where energy conservation is paramount [6], [7]. Low-resolution quantizers effectively reduce power consumption in transceivers by lowering the quantization levels in DACs and ADCs. This reduction is particularly impactful in multiple-antenna systems, where low-resolution quantizers can significantly decrease overall power usage [8]. Nevertheless, quantized networks employing low-resolution quantizers in transmitting and receiving processes introduce significant quantization errors. These errors distort both transmitted and received signals due to the limitations of the DACs and ADCs. These distortions result in substantial inter-user interference, posing a fundamental constraint on the spectral efficiency gains [9], [10]. As a result, the demand for improving low-resolution quantizer systems while considering quantized distortion opens an attractive avenue for future research.

Meanwhile, the number of devices used in communication networks has dramatically increased recently. This explosion requires efficient multiple access techniques to handle user interference and improve transmission efficiency. In this context, non-orthogonal multiple access (NOMA) appears to be a significant technique for 5G and beyond [11], [12]. By superimposing users' signals at different power levels, NOMA enables simultaneous transmission, allowing multiple users to share the same time-frequency resources. At the receiver, the Successive Interference Cancellation (SIC) technique is applied to decode each user's signal sequentially, effectively mitigating users' interference [13], [14]. Previous research has investigated integrating NOMA into MIMO networks, referred to as MIMO-NOMA networks [15]. These networks can improve transmission efficiency by leveraging MIMO's spatial diversity and NOMA's efficient multiple access capabilities, which can significantly boost system performance by increasing spectral efficiency and lowering the transmit power demand [16], [17]. While low-resolution quantizers reduce circuit power consumption at the expense of increased distortion

and user interference, NOMA's power-domain superposition coding and the SIC technique effectively mitigate such interference [18]. Therefore, integrating NOMA into quantized MIMO networks not only enables efficient user multiplexing but also helps counteract the spectral-efficiency degradation introduced by quantization, making quantized MIMO-NOMA a compelling design choice for energy-efficient future wireless systems. This integration offers a multifaceted solution to several concurrent wireless communication network challenges: *(a)* efficient MIMO network design, *(b)* the need for a multiple access technique in networks with numerous devices, and *(c)* the application of low-resolution quantizers for MIMO networks with resource-constrained devices.

To the best of our knowledge, research on low-resolution quantizers in MIMO-NOMA systems is in its early stages. Accordingly, this work explores the energy efficiency problem in a quantized uplink multi-user MIMO-NOMA system, and the major contributions are summarized below.

- We explore an energy efficiency problem in a quantized uplink multi-user MIMO-NOMA system. Here, the transceivers are adopted with multiple low-resolution quantizer antennas, and NOMA is applied to enhance the transmission between users and the BS. We analyze the spectral efficiency and system power consumption under the distortion due to low-resolution quantizers. We then formulate an energy efficiency maximization problem by optimizing the users' precoding and BS detection matrix.
- The non-convexity of the objective function and the coupled variables pose a significant challenge in solving the problems using traditional optimization methods. As a result, we propose a post-actor-added deep deterministic policy gradient (P-DDPG) algorithm to form a DRL framework to solve the problem. Unlike standard DDPG or constrained DRL methods that rely on penalty functions, the proposed P-DDPG employs an analytical post-actor normalization. By applying closed-form formulas to normalize the action decided from the neural network, our method guarantees exact constraint satisfaction with low overhead, which is crucial for stable operation in quantized MIMO-NOMA systems. Moreover, we propose a power-exhaustive searching function to effectively decide actions in the inference process to meet the diversity of user devices.
- Numerical experiment proves the convergence of our proposed algorithm. Moreover, the results demonstrate the proposed scheme's outperformance compared to conventional benchmarks. In addition, we analyze the effect of the quantization level on the system's spectral efficiency, power consumption, and energy efficiency.

The remainder of this work is organized as follows. In Section II, we summarize the state-of-the-art related works. Section III introduces the proposed system and formulates the problem. The proposed solution is detailed in Section IV. Section V shows the numerical results. And Section VI concludes the work.

II. Related work

MIMO networks have been widely considered in research because of their effectiveness [19]–[21]. For instance, Song *et al.* [19] proposed a learning method that employs a learnable unitary matrix to optimize the beamforming matrix from the BS to the users. This method showed its effectiveness in simulation, outperforming several existing algorithms in maximizing the system's sum rate. The authors in [20] focused on maximizing the system sum rate in an uplink rate-splitting MIMO network. They integrated graph searching and deep reinforcement learning methods to design the users' precoding matrices and the decoding order at the BS. Kim *et al.* [21] proposed a semi-exhaustive search optimization and an alternating optimization method to design beam forming, task assignment, channel selection, and CPU allocation for minimizing overhead in a MIMO device-to-device edge network. In addition, with the demand for low-power hardware in MIMO networks for energy-efficient transmission, researchers have gradually considered quantized MIMO networks [22]–[24]. In [22], the authors considered a quantized MIMO system with low-resolution DACs at the transmitter. They proposed an alternating-based method to optimize the analog and digital precoding strategies and analyzed the spectral and energy efficiency under quantized distortion. Paper [23] considered low-resolution ADCs at the BS. The authors proposed a tracking algorithm for a mobile user localization issue in quantized MIMO systems in this work. Cho *et al.* [24] aimed to minimize the maximum transmit power at the quantized transmitter antennas. For this purpose, they proposed a Lagrangian-based optimization method to jointly design the beamforming vectors and the peak transmit power under signal-to-quantization-plus-interference-and-noise ratio constraints. Although quantized MIMO networks have been increasingly considered in recent studies because of their ability to reduce power consumption, the distortion noise from low-resolution quantizers negatively impacts spectral efficiency. Therefore, exploring advanced technologies to enhance spectral efficiency in these networks is a compelling study area. Multiple access techniques present a promising solution, especially as device density increases.

Meanwhile, the NOMA technique has demonstrated its efficiency in enhancing MIMO networks [25]–[27]. For instance, the authors in [25] considered an energy efficiency maximization problem in terahertz MIMO-NOMA systems. They proposed an approach combining machine learning and distributed alternating direction methods to optimize hybrid precoding, user clustering, and resource management. Besides, the results in this paper showed the effectiveness of NOMA in MIMO networks, especially when compared with other multiple access schemes. In [26], the authors studied the maximization of system sum rate and energy efficiency in incorporating NOMA and MIMO networks. They proposed a deep neural network-based method to optimize the BS's precoder and power allocation variables. The simulation results highlighted the enhancement of NOMA to the MIMO system, especially compared with other multiple access schemes such as OMA (orthogonal multiple access). Mamat *et al.* [27] applied a geometric program approach for optimizing channel

TABLE I
QUANTIZATION DISTORTION FACTOR

$b_{k,n}^T \backslash b_m^R$	1	2	3	4	5
$\beta_{k,n} \backslash \beta_m$	0.3634	0.1175	0.03454	0.009497	0.002499

direction information bits and transmit power in MIMO-OMA and MIMO-NOMA networks. Aiming to maximize the minimum SINR (signal-to-interference-plus-noise ratio) fairness, the authors proved the outperformance of NOMA over OMA in low-to-moderate CDI (channel direction information) rate regimes, showing that NOMA can achieve up to a 4 dB gain in some scenarios. Their study further demonstrated that user grouping and progressive-filling allocation schemes can significantly reduce computational complexity while maintaining near-optimal fairness. Obviously, NOMA has demonstrated its effectiveness in enhancing MIMO networks. As the demand for improved spectral efficiency in low-power-consumption quantized MIMO networks grows, particularly in multi-user scenarios, investigating quantized MIMO-NOMA networks becomes crucial for advancing wireless communication systems.

III. PROBLEM STATEMENT

A. Quantized Uplink Multi-user MIMO-NOMA Communications Networks

Fig. 1 illustrates the considered quantized uplink multi-user MIMO-NOMA communication system. Here, a set of users, $\mathcal{K} \triangleq \{1, 2, ..., K\}$, each is equipped with N_k antennas, $k \in \mathcal{K}$, transmits signal to an M-antenna BS via wireless channel. The transmission between users and BS is enhanced by applying the NOMA technique. Without loss of generality, we adopt a per-antenna quantization model that supports heterogeneous resolutions across antennas. In this work, the effects of low-resolution DACs and ADCs are modeled using a linear gain plus additive noise approximation, which is a standard and well-established approach in quantization analysis. Here, the quantization error is uncorrelated with the input signal and can be accurately characterized by its second-order statistics [22], [28], [29].

1) Quantized Transceivers: At each user k, the transmit symbol $\mathbf{s}_k$ is precoded by using a linear precoder $\mathbf{P}_k \in \mathbb{C}^{N_k \times 1}$. Without loss of generality, we assume that the transmit symbols have unit power, i.e., $\mathbb{E}(\mathbf{s}_k \mathbf{s}_k^H) = 1$ [30]. The precoded signal at k-user can be expressed as

$$\mathbf{x}_k = \mathbf{P}_k \mathbf{s}_k. \tag{1}$$

Then, each user employs pairs of DACs to convert the digital signal to an analog signal. The DAC pair at antenna n of user k (for real and imaginary parts) is designed with $b_{k,n}^T$-bit resolution. According to [28], we apply a linear representation to approximate the quantization process. The quantized signal at user k is expressed as

$$\mathbf{x}_k^q = \mathcal{Q}_k(\mathbf{x}_k) \approx \mathbf{\Theta}_k^\alpha \mathbf{x}_k + \mathbf{n}_k^q = \mathbf{\Theta}_k^\alpha \mathbf{P}_k \mathbf{s}_k + \mathbf{n}_k^q, \tag{2}$$

where $\mathcal{Q}_k(\mathbf{x}_k)$ is the quantizer function, $\mathbf{\Theta}_k^\alpha$ and $\mathbf{n}_k^q$ are the quantization loss matrix and the additive Gaussian quantization noise at the user k, respectively. The quantization loss matrix, $\mathbf{\Theta}_k^\alpha \triangleq diag\big(\alpha_{k,1}^T, ..., \alpha_{k,N_k}^T\big) \in \mathbb{C}^{N_k \times N_k}$, is estimated based on the quantization distortion factor, $\beta_{k,n}^T, n \in \mathcal{N}_k \triangleq \{1, ..., N_k\}$, where its elements are calculated as

$$\alpha_{k,n}^T = 1 - \beta_{k,n}^T, n \in \mathcal{N}. \tag{3}$$

The value of $\beta_{k,n}^T$ is determined according to the quantization level [31], which is specified as in Table I if $b_{k,n}^T \leq 5$; otherwise, $\beta_{k,n}^T = \frac{\pi\sqrt{3}}{2} 2^{-2b_{k,n}^T}$. The quantization noise follows $\mathbf{n}_k^q \sim \mathcal{CN}(\mathbf{0}_{N_k \times 1}, \mathbf{R}_k^T)$, where $\mathbf{R}_k^T$ denotes the covariance matrix of $\mathbf{n}_k^q$, which can be calculated as:

$$\mathbf{R}_k^T = \mathbf{\Theta}_k^\alpha \mathbf{\Theta}_k^\beta diag(\mathbb{E}[\mathbf{x}_k \mathbf{x}_k^H]), \tag{4}$$

where $\mathbf{\Theta}_k^\beta \triangleq diag\big(\beta_{k,1}^T, ..., \beta_{k,N_k}^T\big) \in \mathbb{C}^{N_k \times N_k}$. Accordingly, the transmitted signal at each user has a power constraint [22], expressed as

$$P_k^T \triangleq tr\left(\mathbb{E}[\mathbf{x}_k^q (\mathbf{x}_k^q)^H]\right) \leq P_{kmax}, k \in \mathcal{K}, \tag{5}$$

where P_{kmax} denotes the maximum transmit power.

By letting $\mathbf{H}_k \in \mathcal{C}^{M \times N_k}$ denote the channel matrix between user k and the BS, the received signal at the BS can be represented as

$$\mathbf{y} = \sum_{k \in \mathcal{K}} \mathbf{H}_k \mathbf{x}_k^q + \mathbf{n_0}, \tag{6}$$

where $\mathbf{n_0}$ is the additive white Gaussian noise with zero mean and variance σ^2. Accordingly, the received signal at each antenna, $m \in \mathcal{M} \triangleq \{1, ..., M\}$, is quantized by a pair of ADCs with b_m^R-bit resolution. Similarly, the quantized received signal is approximated as

$$\mathbf{y}^q = \mathcal{Q}_B(\mathbf{y}) \approx \mathbf{\Theta}_B^\alpha \mathbf{y} + \mathbf{n}_B^q, \tag{7}$$

where $\mathcal{Q}_B(\mathbf{y})$ is the quantizer function, $\mathbf{\Theta}_B^\alpha$ and $\mathbf{n}_B^q$ are the quantization loss matrix and the additive Gaussian quantization noise at the BS, respectively. Similarly, the quantization loss matrix at the BS, $\mathbf{\Theta}_B^\alpha \triangleq diag(\alpha_1^R, ..., \alpha_M^R)$, is calculated based on the resolution of the quantizer, where its elements are calculated as

$$\alpha_m^R = 1 - \beta_m^R, m \in \mathcal{M}. \tag{8}$$

The value of β_m^R is determined according to the quantization level [31], which is specified as in Table I if $b_m^R \leq 5$; otherwise, $\beta_m^R = \frac{\pi\sqrt{3}}{2} 2^{-2b_m^R}$. The quantization noise follows $\mathbf{n}_B^q \sim \mathcal{CN}(\mathbf{0}_{M \times 1}, \mathbf{R}^R)$, where $\mathbf{R}^R$ denotes the covariance matrix of $\mathbf{n}_B^q$, which can be calculated as:

$$\mathbf{R}^R = \mathbf{\Theta}_B^\alpha \mathbf{\Theta}_B^\beta diag(\mathbb{E}[\mathbf{y}\mathbf{y}^H]), \tag{9}$$

where $\mathbf{\Theta}_B^\beta \triangleq diag\big(\beta_1^R, ..., \beta_M^R\big) \in \mathbb{C}^{M \times M}$.

Prior to decoding the signal, the quantized signal is performed with a unitary detection matrix, expressed as [32]–[34]

$$\begin{aligned} \hat{\mathbf{y}} = \mathbf{W}\mathbf{y}^q = \mathbf{W}\mathbf{\Theta}_B^\alpha \sum_{k \in \mathcal{K}} \mathbf{H}_k \mathbf{\Theta}_k^\alpha \mathbf{P}_k \mathbf{s}_k \\ + \mathbf{W}\mathbf{\Theta}_B^\alpha \sum_{k \in \mathcal{K}} \mathbf{H}_k \mathbf{n}_k^q + \mathbf{W}\mathbf{\Theta}_B^\alpha \mathbf{n_0} + \mathbf{W}\mathbf{n}_B^q, \end{aligned} \tag{10}$$

where $\mathbf{W} \triangleq [\mathbf{w}_1, ..., \mathbf{w}_K]^T \in \mathbb{C}^{K \times M}$, with $\mathbf{w}_k \in \mathbb{C}^{1 \times M}, k \in \mathcal{K}$ being the detection vector to detect signal of user k; and $\hat{\mathbf{y}} \triangleq [\hat{y}_1, ..., \hat{y}_K]$, with $\hat{y}_k, k \in \mathcal{K}$, being the detected signal of user k.

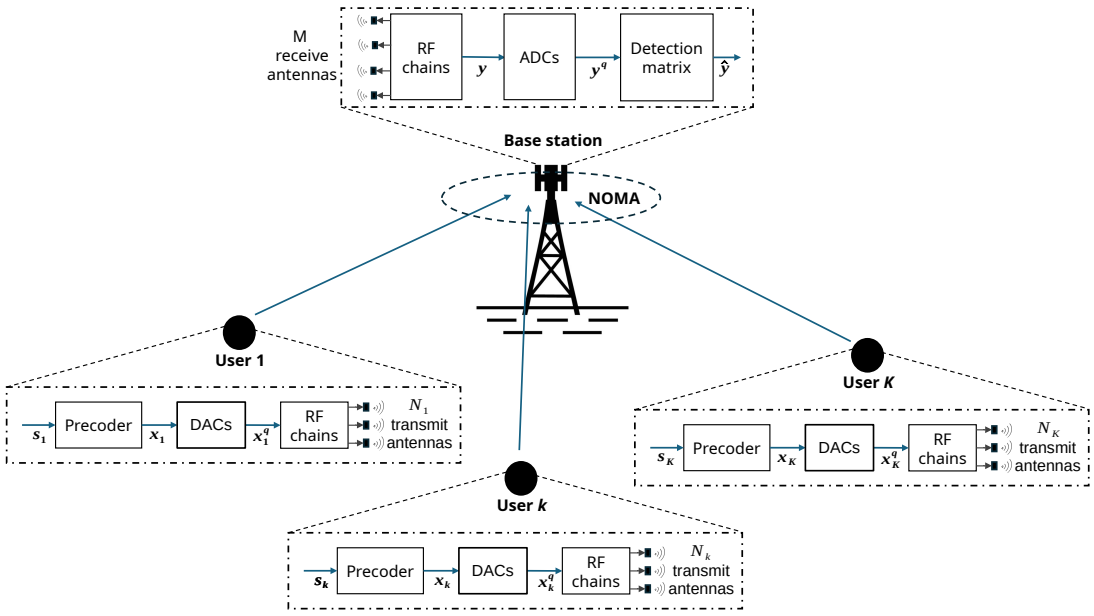


Fig. 1. Quantized MIMO-NOMA systems.

2) Transmission with NOMA: By involving the NOMA technique in the transmission, the BS applies the SIC technique to decode the users' signals from the detected signals. The decoding order is ranked according to the channel effectiveness between the users and the BS, with the user having the highest channel effectiveness decoded first [35]. Without loss of generality, we assume that users' channel effectiveness is determined based on their spectral efficiency in a single-user system, since this metric serves as a practical proxy for post-detection reliability and has been shown to yield effective and low-complexity SIC ordering [36]. Here, the spectral efficiency of user k in a single-user system can be calculated as

$$r_k^{SU} = log\Big(1+\frac{\mathbf{w}_k\mathbf{\Theta}_B^\alpha\mathbf{H}_k\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{P}_k^H(\mathbf{\Theta}_k^\alpha)^H\mathbf{H}_k^H(\mathbf{\Theta}_B^\alpha)^H\mathbf{w}_k^H}{\mathrm{NS}_k}\Big), \tag{11}$$

where the total noise in the single-user system is $\mathrm{NS}_k \triangleq \mathbf{w}_k\mathbf{\Theta}_B^\alpha\mathbf{H}_k\mathbf{R}_k^T(\mathbf{w}_k\mathbf{\Theta}_B^\alpha\mathbf{H}_k)^H + \sigma^2\mathbf{w}_k\mathbf{\Theta}_B^\alpha(\mathbf{w}_k\mathbf{\Theta}_B^\alpha)^H + \mathbf{w}_k\mathbf{R}^R\mathbf{w}_k^H$. Accordingly, the decoding order is ranked as $r_1^{SU} > ... > r_K^{SU}$.

B. Performance Metric

1) Spectral Efficiency: Given the decoding order, the spectral efficiency of user k can be calculated as

$$r_k = log\Big(1+\frac{\mathbf{w}_k\mathbf{\Theta}_B^\alpha\mathbf{H}_k\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{P}_k^H(\mathbf{\Theta}_k^\alpha)^H\mathbf{H}_k^H(\mathbf{\Theta}_B^\alpha)^H\mathbf{w}_k^H}{\mathrm{IU}_k+\mathrm{NU}_k}\Big), \tag{12}$$

where IU_k and NU_k are the interference signals and the total noise from quantization and transmission, respectively, which are calculated as

$$\begin{aligned}\mathrm{IU}_k &= \sum_{j\in\mathcal{IU}_k}\mathbf{w}_k\mathbf{\Theta}_B^\alpha\mathbf{H}_j\mathbf{\Theta}_j^\alpha\mathbf{P}_j\mathbf{P}_j^H(\mathbf{\Theta}_j^\alpha)^H\mathbf{H}_j^H(\mathbf{\Theta}_B^\alpha)^H\mathbf{w}_k^H,\\ \mathrm{NU}_k &= \sum_{i\in\mathcal{K}}\mathbf{w}_k\mathbf{\Theta}_B^\alpha\mathbf{H}_i\mathbf{R}_i^T\mathbf{H}_i^H(\mathbf{\Theta}_B^\alpha)^H\mathbf{w}_k^H\\ &\quad+\sigma^2\mathbf{w}_k\mathbf{\Theta}_B^\alpha(\mathbf{\Theta}_B^\alpha)^H\mathbf{w}_k^H+\mathbf{w}_k\mathbf{R}^R\mathbf{w}_k^H,\end{aligned} \tag{13}$$

where $\mathcal{IU}_k$ is the set of the users decoded after user k.

2) Power Consumption: The power consumption at each user is combined from the amplifier and analog circuit powers [9]. Here, the amplifier power consumption can be calculated as

$$P_k^{AP} = \eta^{-1}P_k^T, \tag{14}$$

where η is a power efficiency coefficient. The analog circuit power consumption of user k is calculated according to the local oscillator (LO), DACs, and radio frequency (RF) chain, which is calculated as [9], [22]

$$P_k^{AC} = P_k^{LO} + \sum_{n\in\mathcal{N}_k}(2P_{k,n}^{DAC}+P_k^{RF}). \tag{15}$$

Here, the power consumption of each DAC at antenna n of user k, $n\in\mathcal{N}_k$, can be calculated as

$$P_{k,n}^{DAC} = 1.5\times10^{-5}\cdot2^{b_{k,n}^T}+9\times10^{-12}\cdot b_{k,n}^T\cdot F_s, \tag{16}$$

where F_s is the sampling frequency [37]. The power consumption of the RF chain at user k is calculated as

$$P_k^{RF} = 2(P_k^{LP}+P_k^M)+P_k^H, \tag{17}$$

where P_k^{LP}, P_k^M, and P_k^H are the power consumption of the low-pass filter, mixer, and hybrid buffer at user k, respectively. We summarize some parameters for power consumption model in Table II, which are compiled from [9], [22].

TABLE II
PARAMETERS FOR POWER CONSUMPTION.

Parameters	Values
η	27%
F_s	1 GHz
$\{P_k^{LO}, P_k^{LP}, P_k^{M}, P_k^{H}, k \in \mathcal{K}\}$	$\{22.5, 14, 0.3, 3\}$mW

3) Energy Efficiency: To maximize the system spectral efficiency while minimizing the power consumption, we formulate an energy efficiency metric, which can be calculated by dividing the sum of the users' spectral efficiency by the sum of the users' power consumption, calculated as

$$\mathcal{E} = \frac{\sum_{k\in\mathcal{K}} r_k}{\sum_{k\in\mathcal{K}} P_k^U}, \tag{18}$$

where $P_k^U = P_k^{AP} + P_k^{AC}$ is the power consumption of user k.

C. Problem Formulation

In this study, we address the dual objectives of enhancing spectral efficiency while reducing the transmitters' power consumption under low-resolution quantizer effects by formulating an optimization problem focusing on maximizing the energy efficiency by optimizing the precoding matrices at the users and the detection matrix at the BS. Accordingly, the formulated problem can be expressed as

$$\begin{aligned} \text{(P1):} \quad & \max_{\mathbf{W},\mathbf{P}} \quad \mathcal{E} && \text{(19a)} \\ & \text{s.t.} \quad P_k^T \le P_{kmax}, k \in \mathcal{K}, && \text{(19b)} \\ & \qquad \mathbf{w}_k \mathbf{w}_k^H = 1, k \in \mathcal{K}, && \text{(19c)} \end{aligned}$$

where $\mathbf{P} \triangleq \{\mathbf{P}_k, k \in \mathcal{K}\}$ is the set of precoding matrices of all users, (19b) denotes the users' power constraint, and (19c) is the constraint of the detection matrix [38].

Problem (P1) poses significant challenges due to the non-convexity of the objective function and the strong interdependence among the optimization variables. These characteristics make it difficult to apply standard optimization methods directly, necessitating more advanced or specialized solution approaches. In particular, conventional deterministic optimization methods are difficult to apply to Problem (P1) due to the lack of convexity and tractable analytical gradients, as well as the strong coupling among optimization variables. Moreover, conventional supervised deep learning approaches are unsuitable because they require labeled training data generated from optimal solutions, which are unavailable due to the lack of an efficient benchmark solver. Fortunately, deep reinforcement learning (DRL) has emerged as an efficient approach to solving this class of problems, as demonstrated by many recent works. In particular, the deep deterministic policy gradient (DDPG) algorithm has been widely adopted due to its effectiveness in addressing wireless communication optimization problems with continuous action spaces [39]–[42]. Hence, in this work, we propose a DRL-based framework and develop a novel DDPG-based algorithm, referred to as post-actor-added DDPG (P-DDPG).

IV. PROPOSED P-DDPG-BASED DRL FRAMEWORK

A. Reinforcement Learning Model

To apply the DRL framework to solve (P1), we first transform the problem into the reinforcement learning (RL) model [43], where the BS plays the role of the RL agent, and the RL environment is the whole system. Throughout this work, the BS is assumed to have perfect channel state information (CSI) to focus on the joint optimization of users' precoders and the BS detection matrix under quantization constraints. In practice, CSI can be obtained via standard uplink pilot-based training, where users transmit known pilot sequences, and the BS estimates the channels from the received signals [44]. When low-resolution ADCs are employed, quantization during the pilot phase attenuates the received pilots. It introduces nonlinear distortion that degrades estimation accuracy, which can be mitigated by quantization-aware channel estimation techniques [45], [46]. At each time slot t, we define the state space, action space, and reward function as

- *State space:* This space represents the set of all possible configurations or conditions that the system can perceive or experience, which are the channel matrices between users to the BS, expressed as

$$s[t] = \{\mathbf{H}_1[t], \mathbf{H}_2[t], ..., \mathbf{H}_K[t]\}. \tag{20}$$

- *Action space:* The action space defines the set of decisions of the agent, which are the precoding matrices and the detection matrix, expressed as

$$a[t] = \{\mathbf{P}[t], \mathbf{W}[t]\}. \tag{21}$$

- *Reward function:* Aiming to maximize energy efficiency, we calculate the reward function at each time slot using energy efficiency, expressed as

$$z[t] = \mathcal{E}[t]. \tag{22}$$

The RL problem is structured to maximize long-term rewards by identifying the optimal action for each state. This optimization process takes into account the constraints specified in (19), ensuring that the chosen actions adhere to the defined requirements of the system.

B. Proposed P-DDPG Algorithm

To train the agent to decide on the action, we apply the DDPG algorithm, which determines action using a neural network named actor network. At each time slot t, the actor network decides the action, denoted as $a^d[t]$, according to the observed state $s[t]$, expressed as

$$a^d[t] = \mu_{\theta_\mu}(s[t]) + \mathcal{OU}[t], \tag{23}$$

where $\mathcal{OU}[t]$ is the noise for training exploration [47], the noise is set to zero after training, $\mu_{\theta_\mu}(s)$ denotes the actor network with the parameter θ_μ, which takes state s as the input. The chosen action is evaluated by the critic network, $Q_{\theta_Q}(s, a)$, with the parameter θ_Q, which takes a pair of action

and state (s, a) as input. Accordingly, the parameter of the actor network is updated by a policy gradient expressed as

$$\nabla_{\theta_\mu} J = \frac{1}{S}\sum_{i=1}^{S}\left(\nabla_a Q_{\theta_Q}(s,a)|_{s=s_i,a=\mu_{\theta_\mu}(s_i)}\nabla_{\theta_\mu}\mu_{\theta_\mu}(s_i)\right), \tag{24}$$

where S is the size of training sample batch, s_i denotes the state in the sample i. Then, the critic network parameter is trained by minimizing a loss function expressed as

$$L = \frac{1}{S}\sum_{i=1}^{S}(Q_{\theta_Q}(s_i,a_i) - y_i)^2, \tag{25}$$

where y_i denotes the target value, and a_i is the action in the sample i. Here, the target value is estimated as

$$y_i = z_i + \gamma Q'_{\theta_{Q'}}(s'_i, \mu'_{\theta_{\mu'}}(s'_i)), \tag{26}$$

where z_i and s'_i denote the reward and the next state in sample i, $Q'_{\theta_{Q'}}(s,a)$ and $\mu'_{\theta_{\mu'}}(s)$ denote the target critic and actor networks, respectively, with the corresponding parameters $\theta_{Q'}$ and $\theta_{\mu'}$, which enhance the training performance [48]. The target network parameters are updated based on the primary networks by a soft update function with a coefficient τ, expressed as

$$\begin{aligned}\theta_{\mu'} &\leftarrow \tau\theta_\mu + (1-\tau)\theta_{\mu'},\\ \theta_{Q'} &\leftarrow \tau\theta_Q + (1-\tau)\theta_{Q'}.\end{aligned} \tag{27}$$

The networks are trained off-policy using training samples randomly drawn from an experience replay buffer, which holds the interaction experiences. Accordingly, the actor network decides the optimal action to maximize the expected reward at each time step. However, it does not consider the problem constraints when determining the action. Therefore, the actions determined by the actor network may not satisfy the problem constraints. Hence, we propose a post-actor process that integrates with the DDPG algorithm to ensure all the problem constraints, resulting in the P-DDPG algorithm.

To ensure compliance with constraint (19b), we examine it closely to understand how the chosen action influences this constraint. Here, users' transmit power can be expressed by the following proposition.

Proposition 1. *The transmit power of user k in* (5) *can be expressed as follows*

$$P_k^T = tr\left(\mathbf{\Theta}_k^\alpha \mathbf{P}_k \mathbf{P}_k^H\right). \tag{28}$$

Proof. According to (5), the transmit power at user k can be calculated as

$$\begin{aligned} &tr\left(\mathbb{E}\left[\mathbf{x}_k^q(\mathbf{x}_k^q)^H\right]\right)\\ &= tr\left(\mathbb{E}\left[\left(\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{s}_k + \mathbf{n}_k^q\right)\left(\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{s}_k + \mathbf{n}_k^q\right)^H\right]\right)\\ &\overset{(a)}{=} tr\left(\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{P}_k^H(\mathbf{\Theta}_k^\alpha)^H + \mathbf{R}_k^T\right), \end{aligned} \tag{29}$$

where (a) is obtained by the fact that $\mathbb{E}\left(s_k s_k^H\right) = 1$, and the noise is uncorrelated with the transmitted signal [49]. According to (1) and (4), the covariance matrix of quantization noise at user k can be calculated as

$$\mathbf{R}_k^T = \mathbf{\Theta}_k^\alpha\mathbf{\Theta}_k^\beta diag\left(\mathbf{P}_k\mathbf{P}_k^H\right). \tag{30}$$

Based on the quantization loss and distortion factors calculated by (3) and Table I, the corresponding matrices $\mathbf{\Theta}_k^\alpha$ and $\mathbf{\Theta}_k^\beta$ are the real diagonal matrices. Accordingly, the transmit power in (29) can be rewritten as

$$\begin{aligned} &tr\left(\left[\mathbb{E}(\mathbf{x}_k(\mathbf{x}_k)^H)\right]\right) = tr\left(\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{P}_k^H(\mathbf{\Theta}_k^\alpha)^H + \mathbf{R}_k^T\right)\\ &= tr\left(\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{P}_k^H(\mathbf{\Theta}_k^\alpha)^H\right) + tr\left(\mathbf{\Theta}_k^\alpha\mathbf{\Theta}_k^\beta diag\left(\mathbf{P}_k\mathbf{P}_k^H\right)\right)\\ &\overset{(b)}{=} tr\left(\mathbf{\Theta}_k^\alpha\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{P}_k^H\right)\\ &\quad + tr\left(\mathbf{\Theta}_k^\alpha\mathbf{I}_{N_k}diag(\mathbf{P}_k\mathbf{P}_k^H)\right) - tr\left(\mathbf{\Theta}_k^\alpha\mathbf{\Theta}_k^\alpha diag(\mathbf{P}_k\mathbf{P}_k^H)\right)\\ &= tr\left(\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{P}_k^H\right), \end{aligned} \tag{31}$$

where (b) is due to $\mathbf{\Theta}_k^\beta = \mathbf{I}_{N_k} - \mathbf{\Theta}_k^\alpha$ with $\mathbf{I}_{N_k}$ being the identity matrix, and $\mathbf{\Theta}_k^\alpha$ and $\mathbf{\Theta}_k^\beta$ are the real diagonal matrices. As a result, the transmit power of user k is expressed as

$$P_k^T = tr\left(\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{P}_k^H\right), \tag{32}$$

which completes the proof. □

By applying Proposition 1, constraint (19b) is rewritten as

$$tr\left(\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{P}_k^H\right) \le P_{kmax}. \tag{33}$$

To clarify this constraint, we declare a new variable, $f_k \in [0,1]$, as the fraction of usage power at user k. Accordingly, the constraint in (33) is equivalently written as

$$f_k \in [0,1], \tag{34a}$$

$$tr\left(\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{P}_k^H\right) = f_k P_{kmax} \tag{34b}$$

Here, constraint (34a) can be naturally satisfied by adapting the activation function to decide the range of action values when using the neural network [50]. Let $\mathbf{V}_k \in \mathbb{C}^{N_k\times 1}$ denote the temporary precoding matrix at user k decided by the actor network, we propose the following proposition to ensure constraint (34b).

Proposition 2. *The precoding matrix at user k, $\mathbf{P}_k$, can satisfy the constraint* (34b) *by normalizing from $\mathbf{V}_k$, where its element at n-th element, $p_{k,n}$, is calculated as*

$$p_{k,n} = v_{k,n}\sqrt{\frac{f_k P_{kmax}}{\sum_{n=1}^{N_k}\alpha_{k,n}^T|v_{k,n}|^2}}, \tag{35}$$

where $v_{k,n}$ is the n-th element of $\mathbf{V}_k$.

Proof. To prove that constraint (34b) is true with the precoding matrix $\mathbf{P}_k$, we first explore the trace operator on the left-hand side, which can be expressed as

$$tr\left(\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{P}_k^H\right) \overset{(c)}{=} \sum_{n=1}^{N_k}\alpha_{k,n}^T|p_{k,n}|^2, \tag{36}$$

where (c) is obtained from (37) at the top of the next page with $\mathbf{\Theta}_k^\alpha$ being the real diagonal matrix. By replacing $p_{k,n}$ in (35) to (36), we obtain

$$\begin{aligned} tr\left(\mathbf{\Theta}_k^\alpha\mathbf{P}_k\mathbf{P}_k^H\right) &= \sum_{n=1}^{N_k}\alpha_{k,n}^T\frac{|v_{k,n}|^2 f_k P_{kmax}}{\sum_{n=1}^{N_k}\alpha_{k,n}^T|v_{k,n}|^2}\\ &= f_k P_{kmax}\frac{\sum_{n=1}^{N_k}\alpha_{k,n}^T|v_{k,n}|^2}{\sum_{n=1}^{N_k}\alpha_{k,n}^T|v_{k,n}|^2}\\ &= f_k P_{kmax}, \end{aligned} \tag{38}$$

$$tr\left(\mathbf{\Theta}_k^{\alpha}\mathbf{P}_k\mathbf{P}_k^H\right) = tr\left(\text{diag}\left(\alpha_{k,1}^T,\ldots,\alpha_{k,N_k}^T\right)\begin{bmatrix}p_{k,1}\\ \cdots\\ p_{k,N_k}\end{bmatrix}\begin{bmatrix}p_{k,1}\\ \cdots\\ p_{k,N_k}\end{bmatrix}^H\right) = tr\left(\begin{bmatrix}\alpha_{k,1}^T|p_{k,1}|^2 & \cdots & \cdots & \cdots\\ \cdots & \alpha_{k,2}^T|p_{k,2}|^2 & \cdots & \cdots\\ \cdots & \cdots & \cdots & \alpha_{k,N_k}^T|p_{k,N_k}|^2\end{bmatrix}\right) = \sum_{n=1}^{N_k}\alpha_{k,n}^T|p_{k,n}|^2. \tag{37}$$

which establishes the satisfaction of constraint (34b), thereby completing the proof. □

Upon implementing the results from Proposition 2, the problem constraints have been simplified. The only constraint that persists in the problem is (19c). Let $\mathbf{w}_k^d \in \mathbb{C}^{1\times M}, k \in \mathcal{K}$ denote the temporary detection vector for user k decided by the actor network, we propose the following proposition to ensure constraint (19c)

Proposition 3. *The detection vector for user k, $\mathbf{w}_k \triangleq [w_{k,1},...,w_{k,M}]$, can satisfy constraint* (19c) *by normalizing from $\mathbf{w}_k^d$, where its elements can be calculated as*

$$w_{k,m} = \frac{w_{k,m}^d}{||\mathbf{w}_k^d||_2}, k \in \mathcal{K}, m \in \mathcal{M}, \tag{39}$$

where $w_{k,m}^d$ denotes the m-th element of $\mathbf{w}_k^d$.

Proof. To demonstrate that the detection vector calculated by (39) can satisfy constraint (19c), we explore the left-hand side of (19c) as

$$\mathbf{w}_k\mathbf{w}_k^H = \sum_{m=1}^{M}|w_{k,m}|^2. \tag{40}$$

By replacing (39) to (40), we obtain

$$\mathbf{w}_k\mathbf{w}_k^H = \sum_{m=1}^{M}\left|\frac{w_{k,m}^d}{||\mathbf{w}_k^d||_2}\right|^2 = \frac{\sum_{m=1}^{M}|w_{k,m}^d|^2}{||\mathbf{w}_k^d||_2^2} = 1, \tag{41}$$

which completes the proof. □

By applying proposition 3, constraint (19c) is satisfied. As a result, all constraints in the problem are ensured while deciding the action.

C. Framework Formulation

The proposed P-DDPG algorithm is detailed in Algorithm 1, where the P-DDPG-based DRL framework is illustrated in Fig. 2. The framework includes the P-DDPG algorithm, which decides the action to interact with the environment. The algorithm is trained in E episodes, each has T time steps. At each time step, the action a^d is determined by the actor network based on the observed state s, which is comprised of $f_k, \mathbf{V}_k, \mathbf{w}_k^d$, i.e.,

$$\{f_k[t], \mathbf{V}_k[t], \mathbf{w}_k^d[t], k \in \mathcal{K}\} = a^d[t]. \tag{42}$$

Then, the post-actor process is applied to modify the action to meet the problem constraints. Here, the precoding and detection matrices are calculated according to Propositions 2 and 3, respectively, shown in lines 6–13 in Algorithm 1. Accordingly, the obtained matrices $\mathbf{P}[t]$ and $\mathbf{W}[t]$ are performed to the environment. The environment changes to the next state, $s'[t]$, and the reward, $z[t]$, is estimated. After that, a tuple of experience, including $s[t], a[t], z[t], s'[t]$ is stored in the replay buffer for training. At each training step, a batch of S samples is randomly drawn from the replay buffer to train the neural network, where the training process is performed by the DDPG algorithm as described in (24)-(27). After training, the trained actor network is obtained to interact with the environment in the inference process.

Algorithm 1 Proposed P-DDPG algorithm

1: Set up algorithm parameters.
2: **while** $e < E$ **do**
3: **for** $t = 1 : T$ **do**
4: Observe $s[t]$.
5: Decide $a^d[t]$ using $\mu_{\theta_\mu}(s)$.
6: **for** $k \in \mathcal{K}$ **do**
7: **for** $n = 1 : N_k$ **do**
8: Calculate precoding element as Proposition 2.
9: **end for**
10: **for** $m = 1 : M$ **do**
11: Calculate detection vector element as Proposition 3.
12: **end for**
13: **end for**
14: Perform $\mathbf{P}[t]$, $\mathbf{W}[t]$, get state-next $s'[t]$, reward $z[t]$.
15: Store $(s[t], a[t], z[t], s'[t])$ in buffer.
16: Update state $s'[t] \rightarrow s[t]$.
17: Draw sample batch from buffer, (s^S, a^S, z^S, s'^S)
18: Update networks parameter as (24)-(27).
19: **end for**
20: **end while**
21: **return** the trained actor network, $\mu_{\theta_{\mu^*}^*}(s)$.

D. Power-exhaustive Searching Function

The transmit powers derived from equation (34) are proportional to the maximum available transmit power. During the training phase, the Deep Reinforcement Learning (DRL) model assumes a constant maximum transmit power, P_{kmax}, as defined by the environment parameters. However, this assumption may not hold in real-world scenarios during the inference stage, where different devices often have varying maximum transmit power capabilities. Consequently, applying the model without accounting for these variations in P_{kmax} could lack optimal power allocation value, potentially compromising overall system performance. Therefore, we introduce a power-exhaustive searching function founded on the exhaustive search method, adding to the trained model to decide appropriate transmit powers in the inference stage. The inference flow is illustrated in Fig. 3, which is detailed in Algorithm 2.

For each inference step, the trained actor network designs $a^d[t]$ (42) (line 5). An exhaustive search is then conducted to find the most suitable power level. The maximum power

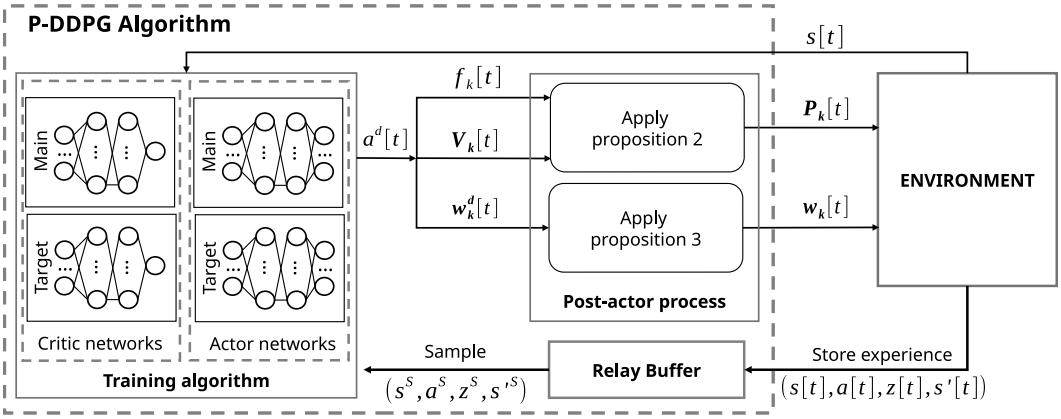


Fig. 2. Proposed P-DDPG-based DRL Framework

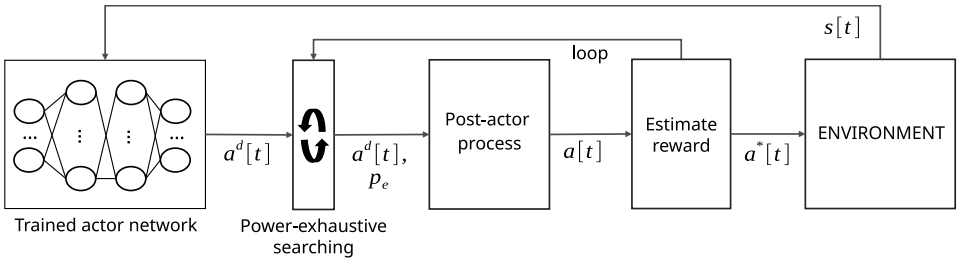


Fig. 3. Inference process flow.

Algorithm 2 Power-exhaustive searching function

1: **Input:** Trained actor network
2: **while** inference **do**
3: Observe state $s[t]$.
4: Initialize best reward: $z_{best}[t] = 0$.
5: Decide $a^d[t] = \mu^*_{\theta_{\mu^*}}(s[t])$.
6: **for** $p_e \leftarrow P_{kmin}$ to P_{kmax}, $p_e : p_e + \Delta p$ **do**
7: Calculate $\mathbf{P}[t], \mathbf{W}[t]$ based on post-actor process (Algorithm 1, lines 6-13) with $P_{kmax} = p_e$.
8: Estimate temporary reward $z_t[t]$.
9: **if** $z_t[t] \geq z_{best}$ **then**
10: $a^*[t] \leftarrow a[t]$, $z_{best} \leftarrow z_t[t]$.
11: **end if**
12: **end for**
13: **return** Optimal action $a^*[t]$.
14: **end while**

level for each user ranges from P_{kmin} to P_{kmax} with a step size of Δ_p. For each power level value p_e, the precoding and detection matrices are calculated using the post-actor process described in Algorithm 1 (lines 6-13), where P_{kmax} is set to p_e. The reward is then estimated. After searching through all power levels, the optimal action $a^*[t] \triangleq \{\mathbf{P}^*[t], \mathbf{W}^*[t]\}$, which yields the best reward, is selected to interact with the environment.

Remark 1. *The power-exhaustive search function is employed during the inference process, using the trained model to determine actions at each interaction step. Therefore, integrating this function into the model does not impact the training process's performance.*

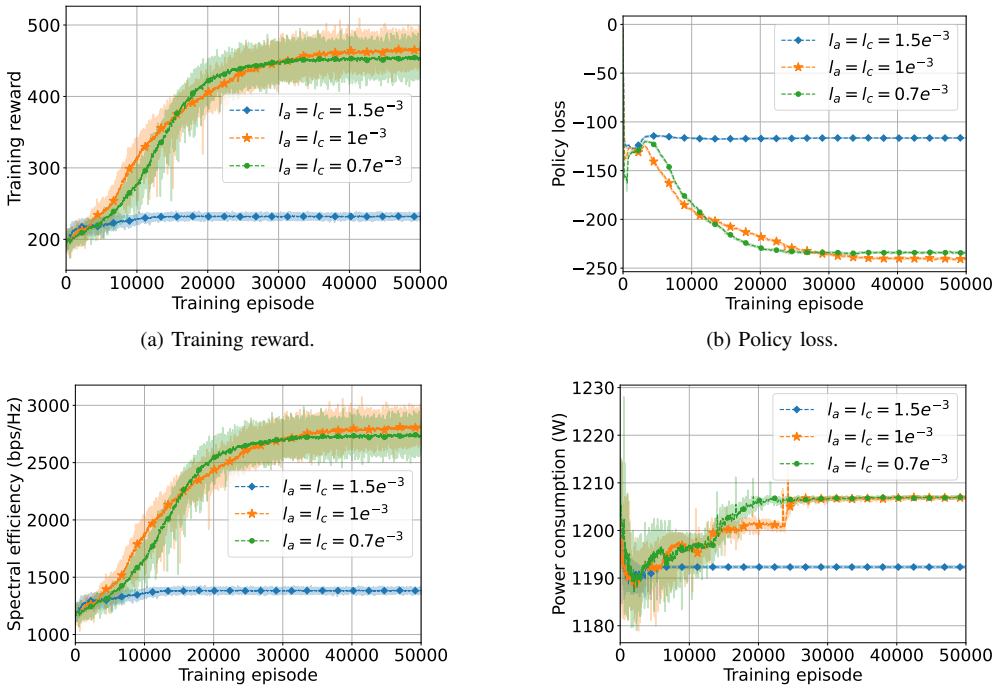
E. Complexity Analysis

As the training process is performed only once and the trained model is subsequently used to interact with the environment, we evaluate the complexity of the proposed framework based on the inference stage. The computational complexity of the inference flow consists of two main components: *(i)* the action decision by the trained actor network inference and *(ii)* the power-exhaustive searching function.

According to [30, Sec. VII-A], the computational complexity of the trained actor network during inference can be determined from the network architecture, yielding $\mathcal{O}\left(\sum_{k=1}^{K} MN_k + \sum_{k=1}^{K} N_k + KM\right)$, where $\sum_{k=1}^{K} MN_k$ corresponds to the number of entries in the state space, and $\sum_{k=1}^{K} N_k + KM$ represents the number of entries in the action space. Consequently, the upper-bound complexity of the actor network inference can be expressed as $\mathcal{O}(KMN)$, where $N = \max(N_k,\, k \in \mathcal{K})$.

For the power-exhaustive searching function, the computational complexity depends on the number of search iterations, defined as $L \triangleq \left\lfloor \frac{P_k^{\max} - P_k^{\min}}{\Delta p} \right\rfloor + 1$. In each iteration, the post-actor process is executed, which involves applying Propositions 2 and 3. The resulting complexity per iteration is $\mathcal{O}(KN + KM)$. Therefore, the overall complexity of the power-exhaustive searching function is $\mathcal{O}\big(LK(N+M)\big)$.

As a result, the total computational complexity of the proposed inference process is $\mathcal{O}\big(KMN + LKN + LKM\big)$, which scales linearly with the environment parameters, making the framework feasible for large-scale deployment.



(a) Training reward.

(b) Policy loss.

(c) Spectral efficiency.

(d) Power consumption.

Fig. 4. Training results of the P-DDPG algorithm under different learning rates.

TABLE III
PARAMETERS FOR POWER CONSUMPTION.

Environmental Parameters	Values
K	10
M	8
$N_k, k \in \mathcal{K}$	4
σ^2	-104 dBm
Bandwidth	10 MHz
P_{kmin}	-10 dBm
P_{kmax}	20 dBm
Δ_p	1
$b^T_{k,n}$	[1,12] bits
b^R_m	[1,12] bits

V. NUMERICAL RESULTS

A. Simulation Setting

We evaluate the system's performance by simulating an environment where the BS serves ten users randomly distributed over a 100-meter radius. The path-loss (in dB) between user k and the BS is calculated as [51]

$$PL_k = 103.8 + 20.9log_{10}(d_k), \tag{43}$$

where d_k (in Km) denotes the link distance. Then, the corresponding channel matrix is calculated as

$$\mathbf{H}_k = \sqrt{10^{-PL_k/10}}\hat{\mathbf{H}}_k, \tag{44}$$

where $\hat{\mathbf{H}}_k \sim \mathcal{CN}(0,1)$ denotes the small scale-fading. Other environmental parameters are summarized in Table III.

We evaluate the proposed framework's performance by comparing it with the following benchmark schemes:

- ***WoPEF (P-DDPG without power-exhaustive searching function):*** In this scheme, the model trained by the P-DDPG algorithm is applied to design the actions in the inference process without adding the power-exhaustive searching function.
- ***PPO-based (PPO algorithm with the proposed post-actor process):*** In this scheme, we train the agent using an online policy algorithm named proximal policy optimization (PPO), which is also widely used in network optimization.
- ***QOMA (quantization with orthogonal multiple access):*** We simulate this scheme to compare the effectiveness of NOMA and OMA in the considered quantized system. According to [52], the FDMA is applied to the communication between users and the BS, where the bandwidth is divided between users, and the interference from other users is zero.
- ***QWoMA (quantization without using multiple access technique):*** To further analyze the effectiveness of multiple access techniques, in this scheme, users share the entire communication bandwidth, and signals from other users are fully treated as noise during signal decoding. According to [53, sub-section 4.1] , the interference signals of decoding user k can be calculated as $\text{IU}_k = \sum_{j\neq k} \mathbf{w}_k \mathbf{\Theta}_B^{\alpha} \mathbf{H}_j \mathbf{\Theta}_j^{\alpha} \mathbf{P}_j \mathbf{P}_j^H (\mathbf{\Theta}_j^{\alpha})^H \mathbf{H}_j^H (\mathbf{\Theta}_B^{\alpha})^H \mathbf{w}_k^H$.
- ***RPaD (Random precoding and detection matrices):*** This scheme randomly chooses the precoding and detection matrices according to their value ranges.

B. Convergence Analysis

The neural networks applied in this experiment are combined from two hidden layers, where the first layer has 2048 nodes and the second has 1024 nodes. Other training parameters are set as: $\gamma = 0.99$, $S = 16$, $\tau = 0.01$, the buffer size is $1e^5$. The training is conducted over 50,000

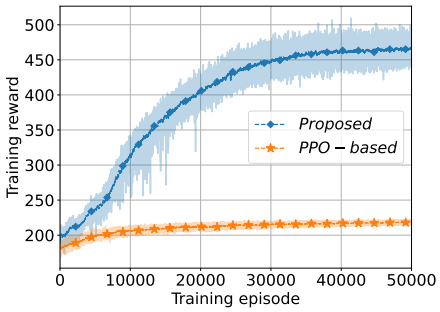


Fig. 5. Training reward performance under different DRL-based approaches.

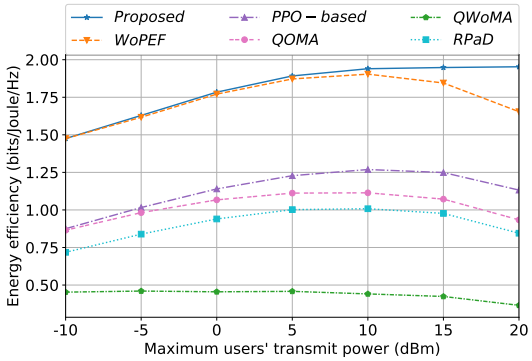


Fig. 6. Energy efficiency under different power levels.

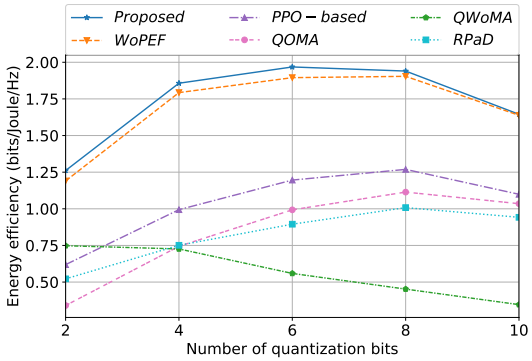


Fig. 7. Energy efficiency under different quantization levels.

episodes, each consisting of 200 time steps. Accordingly, the values plotted in Fig. 4 represent the accumulated results per episode, rather than instantaneous per-step values. We analyze the convergence of training the model under three cases of the learning rate, where the actor and critic learning rates (l_a and l_c, respectively) are selected from $l_a = l_c = \{1.5e^{-3}, 1e^{-3}, 0.7e^{-3}\}$. As shown in Fig. 4a, the reward increases during training and converges after about 40,000 episodes at around 470 per episode. This increase in reward, calculated based on system energy efficiency, aligns with the growth in spectral efficiency and stabilization of power consumption. As illustrated in Fig. 4c, the system's spectral efficiency per episode rises from about 1100 to 2800 after training and stabilizes after about 40,000 episodes. At the beginning of the training process, the model focuses on increasing the reward by reducing system power consumption, as seen in Fig. 4d. However, during this stage, the spectral efficiency increases slowly due to insufficient transmit power, leading to a slower reward increase. After approximately 3000 episodes, the model learns to increase the power, boosting spectral efficiency and, accordingly, the reward. The power consumption then stabilizes after about 25,000 episodes.

Moreover, we analyze the training loss to guarantee learning convergence. This experiment assesses the policy loss obtained from training the actor network. According to (24), we estimate the policy loss, L_μ, as

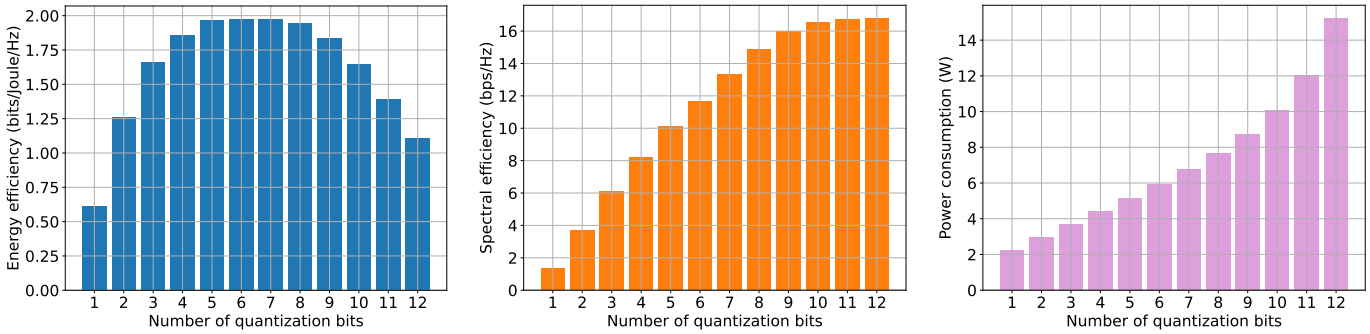
$$L_\mu = \frac{1}{S}\sum_{i=1}^{S}\left(Q_{\theta_Q}(s_i, \mu_{\theta_\mu}(s_i))\right). \tag{45}$$

As illustrated in Fig. 4b, the policy loss gradually decreases during training and stabilizes after approximately 40,000 episodes. This decline implies an increase in the designed action Q-value, demonstrating the improvement of the actor network during training and, accordingly, the effectiveness of the learning process. Additionally, the results in Fig. 4 indicate that the optimal learning rate for this model is $l_a = l_c = 1e^{-3}$, as it provides the highest reward and lowest policy loss. Conversely, a larger learning rate ($l_a = l_c = 1.5e^{-3}$) leads to a stuck finding of optimal network parameters due to oversized optimization steps, causing the training process to stall. Therefore, we use the best-trained model to evaluate the results.

Beyond these convergence curves, the stability of P-DDPG is further supported by the post-actor normalization in Propositions 2 and 3, which ensures that all actions remain feasible and bounded during training. This complements the observed results in Fig. 4, where the reward, spectral efficiency, and power consumption stabilize after sufficient episodes, confirming reliable convergence of the learned policy in practice. Also, we compare the performance of training the proposed algorithm with that of the PPO-based algorithm. As shown in Fig. 5, the proposed P-DDPG converges faster and achieves a significantly higher training reward than the PPO-based approach. This behavior is explained by the structure of the considered problem, which involves continuous actions under hard constraints, which aligns it well with deterministic policy gradient methods. Although the same post-actor normalization is applied to both algorithms, PPO is inherently disadvantaged by its stochastic on-policy learning mechanism. In particular, action normalization weakens the correspondence between sampled actions and their practical impact on the reward, increasing gradient variance in PPO's likelihood-ratio updates. Combined with conservative clipped policy updates and the inability to reuse past samples, PPO converges to a suboptimal solution. In contrast, P-DDPG exploits off-policy learning and experience replay to efficiently refine strategies, resulting in superior convergence and higher steady-state reward.

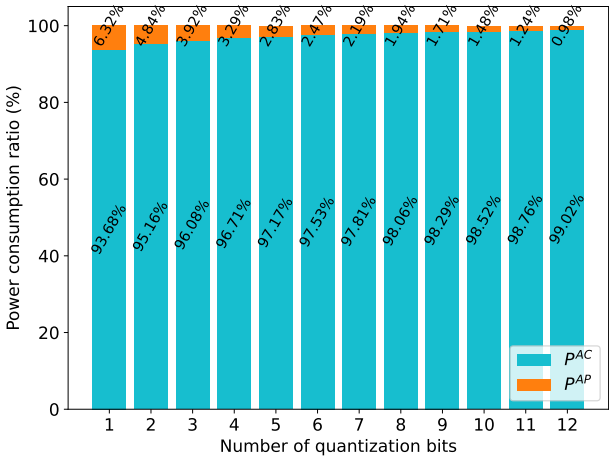
C. Performance Evaluation

We begin by evaluating the system's energy efficiency across different transmit power levels, with the maximum power of users varying from -10 to 20 dBm. As shown

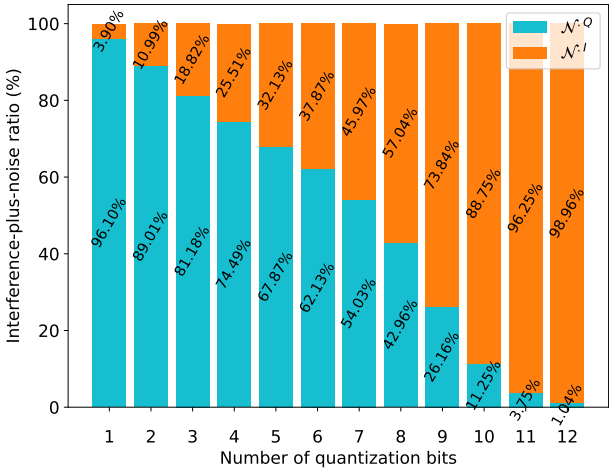


(a) Energy efficiency. (b) Spectral efficiency. (c) Power consumption.

Fig. 8. System performance under different quantization levels.



(a) Power consumption ratio.



(b) Interference-plus-noise ratio.

Fig. 9. Power and interference-plus-noise ratios under different quantization levels.

in Fig. 6, the *Proposed* scheme exhibits increasing energy efficiency as the transmit power increases and maintains stable performance at higher power levels. In contrast, several benchmark schemes experience noticeable efficiency degradation in the high-power regime due to unregulated power usage. These results confirm the effectiveness of the power-exhaustive searching function, which avoids unnecessary power consumption when the transmit power budget increases. In particular, the *WoPEF* scheme performs comparably to the *Proposed* scheme at low-to-moderate power levels but declines as the power budget increases over the peak value of 10 dBm. Overall, the *Proposed* approach consistently outperforms the benchmark schemes, improving energy efficiency on average by approximately 59.6% over the *PPO-based* scheme, 76.6% over *QOMA*, 99.4% over *RPaD*, and more than three times over *QWoMA*. This highlights the effectiveness of NOMA in quantized MIMO networks, especially when compared to orthogonal multiple access or scenarios without multiple access techniques.

Next, we evaluate the energy efficiency under different quantization levels, assuming identical resolution for ADCs and DACs. As shown in Fig. 7, the energy efficiency of the *Proposed* scheme increases with the quantization level. It reaches its maximum at moderate resolutions, after which it gradually decreases as the quantization level increases. This behavior results from the trade-off between reduced quantization distortion and increased circuit power consumption. While higher quantization levels improve spectral efficiency by mitigating quantization noise, the associated increase in power consumption eventually offsets these gains, reducing overall energy efficiency. Averaged across all quantization levels, the *Proposed* scheme improves energy efficiency by approximately 67.5% over the *PPO-based* scheme, 105.2% over *QOMA*, 110.4% over *RPaD*, and more than three times over QWoMA.

For a more in-depth understanding, we provide a detailed analysis of the impact of quantization levels on system performance in Fig. 8. As shown in Fig. 8a, the energy efficiency increases according to the quantization levels and reaches peak values at 5, 6, and 7 bits quantization, then decreases afterward. Figs. 8b and 8c illustrate the reasons for this effect. When the quantization levels increase to 9 bits, the spectral efficiency stabilizes and does not increase further. At the same time, the power consumption continues to increase with the number of quantization bits. Furthermore, the spectral efficiency increases from about 13.3 (bps/Hz) to 16.7 (bps/Hz), an improvement of approximately 25.6%, when increasing the quantization level from 7 to 12 bits. Meanwhile, the power consumption increases from about 6.7 (W) to about 15.2 (W),

which is more than double. Hence, the stability of the spectral efficiency and the continued increase in power consumption cause a reduction in the energy efficiency beyond certain quantization levels.

Finally, we assess the impact of varying quantization levels on the overall system power consumption and the resulting distortion, often referred to as quantization noise, as depicted in Fig. 9. Specifically, we analyze the contribution of the amplifier power, denoted by $P^{AP} \triangleq \sum_{k\in\mathcal{K}} P_k^{AP}$, and the analog circuit power, denoted as $P^{AC} \triangleq \sum_{k\in\mathcal{K}} P_k^{AC}$, to the total system power consumption, as defined in (14) and (15), respectively. As illustrated in Fig. 9a, the analog circuit power, P^{AC}, constitutes the predominant portion of the system's power consumption, exceeding 93% across all scenarios. Consequently, increasing the quantization level leads to a substantial rise in analog circuit power, thereby significantly escalating the overall system power consumption. Notably, when the quantization level is increased to 12 bits, P^{AC} dominates the power consumption, accounting for over 99% of the total power consumption. Given that the amplifier power is constrained by the maximum transmit power limit, the continuous increase in quantization levels directly amplifies the analog circuit power, resulting in the outcomes observed in Fig. 8c. Next, we analyze the quantization noise and evaluate its impact on the interference-plus-noise value, which, in turn, significantly influences the spectral efficiency. As outlined in (13), we decompose the interference signals and total noise into two components: quantization noise, denoted by $\mathcal{N}_k^{Q}$, and interference plus environmental noise, denoted by $\mathcal{N}_k^{I}$, quantified as

$$\begin{aligned}\mathcal{N}_k^{Q} &= \sum_{i\in\mathcal{K}} \mathbf{w}_k \mathbf{\Theta}_B^{\alpha} \mathbf{H}_i \mathbf{R}_i^T \mathbf{H}_i^H (\mathbf{\Theta}_B^{\alpha})^H \mathbf{w}_k^H + \mathbf{w}_k \mathbf{R}^R \mathbf{w}_k^H, \\ \mathcal{N}_k^{I} &= \mathrm{IU}_k + \sigma^2 \mathbf{w}_k \mathbf{\Theta}_B^{\alpha} (\mathbf{\Theta}_B^{\alpha})^H \mathbf{w}_k^H.\end{aligned} \tag{46}$$

Accordingly, we measure the mean of quantization noise, $\mathcal{N}^{\cdot Q} \triangleq \frac{1}{K}\sum_{k\in\mathcal{K}} \mathcal{N}_k^{Q}$, and the mean of interference plus environmental noise, $\mathcal{N}^{\cdot I} \triangleq \frac{1}{K}\sum_{k\in\mathcal{K}} \mathcal{N}_k^{I}$, under different quantization levels. As depicted in Fig. 9b, the quantization noise decreases as the quantization levels increase. This is because higher quantization levels reduce the distortion factor, as defined in Table I, which, in turn, lowers the quantization noise described in (4) and (9). Notably, $\mathcal{N}^{Q}$ constitutes over 96% of the total interference plus noise when using a 1-bit quantizer, but this value dramatically drops to just 1% with a 12-bit quantizer. Consequently, higher quantization levels significantly reduce quantization noise, thereby improving SINR. However, this improvement comes at the cost of increased power consumption, which raises system costs and reduces energy efficiency. Therefore, a careful trade-off between energy and spectral efficiency is essential when designing practical quantized MIMO networks.

VI. Conclusion

In this work, we addressed the challenge of maximizing energy efficiency in quantized uplink multi-user MIMO-NOMA networks. Low-resolution quantizers were employed at the user devices and BS to reduce power consumption, while the NOMA technique was integrated to enhance transmission efficiency. We formulated the system model by analyzing the characteristics of quantized networks and defined the energy efficiency maximization problem with users' precoding matrices and the BS's detection matrix as optimization variables. Given the nonconvex nature of the objective function, we developed a DRL framework and introduced a post-actor process into the DDPG algorithm, resulting in the P-DDPG method that determines optimization variables while ensuring constraint satisfaction. Additionally, we proposed a power-exhaustive search function to augment the trained model, enabling robust power adaptation across diverse devices during inference. Numerical results confirmed the convergence of the proposed algorithm and demonstrated its superiority over benchmark schemes under various scenarios. We further analyzed the effect of quantization resolution, showing that moderate bit levels offer the best trade-off between efficiency and complexity. These findings provide practical insights for designing quantized MIMO-NOMA systems and contribute to the advancement of next-generation wireless networks.

While this work employed P-DDPG as the training algorithm, the proposed DRL framework is flexible and can readily incorporate other continuous-action DRL algorithms, such as A2C or SAC, making the comparative evaluation of different DRL approaches in quantized MIMO-NOMA systems an interesting direction for future research. Also, future extensions of this work may consider robustness under dynamic conditions, including varying user loads, heterogeneous quantization resolutions, and fast-fading channels, which are critical for practical deployment of quantized MIMO-NOMA systems.

References

[1] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "User-centric cell-free massive MIMO networks: A survey of opportunities, challenges and solutions," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 611–652, 2022.

[2] D. López-Pérez, A. De Domenico, N. Piovesan, G. Xinli, H. Bao, S. Qitao, and M. Debbah, "A survey on 5G radio access network energy efficiency: Massive MIMO, lean carrier design, sleep modes, and machine learning," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 653–697, 2022.

[3] K. Min, T. Kim, and M. Jung, "Performance analysis of multiuser massive MIMO with multi-antenna users: Asymptotic data rate and its application," *ICT Express*, vol. 9, no. 5, pp. 821–826, 2023.

[4] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems," *IEEE Communications Magazine*, vol. 56, no. 7, pp. 205–211, 2018.

[5] W. Mao, Z. Zhao, Z. Chang, G. Min, and W. Gao, "Energy-efficient industrial internet of things: Overview and open issues," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 11, pp. 7225–7237, 2021.

[6] M. M. Kiasaraei, K. Nikitopoulos, and R. Tafazolli, "Toward ultra-power-efficient, Tbps wireless systems via analogue processing: Existing approaches, challenges and way forward," *IEEE Communications Surveys & Tutorials*, vol. 26, no. 2, pp. 747–780, 2024.

[7] A.-T. Le, D.-T. Do, N.-N. Dao, N. D. Nguyen, and A. Silva, "New look at secure performance of massive MIMO with low-resolution DACs," *ICT Express*, vol. 9, no. 4, pp. 608–613, 2023.

[8] J. Choi, G. Lee, A. Alkhateeb, A. Gatherer, N. Al-Dhahir, and B. L. Evans, "Advanced receiver architectures for millimeter-wave communications with low-resolution ADCs," *IEEE Communications Magazine*, vol. 58, no. 8, pp. 42–48, 2020.

[9] 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.

[10] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized precoding for massive MU-MIMO," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 4670–4684, 2017.

[11] A. Akbar, S. Jangsher, and F. A. Bhatti, "NOMA and 5G emerging technologies: A survey on issues and solution techniques," *Computer Networks*, vol. 190, p. 107950, 2021.

[12] M. C. Ho, A. T. Tran, D. Lee, J. Paek, W. Noh, and S. Cho, "A DDPG-based energy efficient federated learning algorithm with SWIPT and MC-NOMA," *ICT Express*, vol. 10, no. 3, pp. 600–607, 2024.

[13] V. D. Tuong, T. P. Truong, T.-V. Nguyen, W. Noh, and S. Cho, "Partial computation offloading in NOMA-assisted mobile-edge computing systems using deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 8, no. 17, pp. 13 196–13 208, 2021.

[14] S. Pakravan, J.-Y. Chouinard, X. Li, M. Zeng, W. Hao, Q.-V. Pham, and O. A. Dobre, "Physical layer security for NOMA systems: Requirements, issues, and recommendations," *IEEE Internet of Things Journal*, vol. 10, no. 24, pp. 21 721–21 737, 2023.

[15] O. Maraqa, A. S. Rajasekaran, S. Al-Ahmadi, H. Yanikomeroglu, and S. M. Sait, "A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2192–2235, 2020.

[16] M. Wang, S. Shi, D. Zhang, C. Wu, and Y. Wang, "Joint computation offloading and resource allocation for MIMO-NOMA assisted multi-user MEC systems," *IEEE Transactions on Communications*, vol. 71, no. 7, pp. 4360–4376, 2023.

[17] S. Kiani, M. Dong, S. ShahbazPanahi, G. Boudreau, and M. Bavand, "Learning-based user clustering in NOMA-aided MIMO networks with spatially correlated channels," *IEEE Transactions on Communications*, vol. 70, no. 7, pp. 4807–4821, 2022.

[18] X. Xu, Y. Liu, X. Mu, Q. Chen, H. Jiang, and Z. Ding, "Artificial intelligence enabled NOMA toward next generation multiple access," *IEEE Wireless Communications*, vol. 30, no. 1, pp. 86–94, 2023.

[19] Q. Song, J. Wang, J. Li, G. Liu, and H. Xu, "A learning-only method for multi-cell multi-user MIMO sum rate maximization," in *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications*, 2024, pp. 291–300.

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

[21] J. Kim, T. Kim, M. Hashemi, D. J. Love, and C. G. Brinton, "Minimum overhead beamforming and resource allocation in D2D edge networks," *IEEE/ACM Transactions on Networking*, vol. 30, no. 4, pp. 1454–1468, 2022.

[22] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, "Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 2, pp. 298–312, 2018.

[23] W. Fan, X. Li, G. Yang, C. Li, and Y. Huang, "Low-complexity mobile user tracking in quantized mmWave MIMO systems," *IEEE Transactions on Mobile Computing*, vol. 23, no. 9, pp. 8569–8581, 2024.

[24] Y. Cho, J. Choi, and B. L. Evans, "Coordinated per-antenna power minimization for multicell massive MIMO systems with low-resolution data converters," *IEEE Transactions on Communications*, vol. 72, no. 2, pp. 1119–1134, 2024.

[25] H. Zhang, H. Zhang, W. Liu, K. Long, J. Dong, and V. C. M. Leung, "Energy efficient user clustering, hybrid precoding and power optimization in terahertz MIMO-NOMA systems," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 9, pp. 2074–2085, 2020.

[26] H. Huang, Y. Yang, Z. Ding, H. Wang, H. Sari, and F. Adachi, "Deep learning-based sum data rate and energy efficiency optimization for MIMO-NOMA systems," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5373–5388, 2020.

[27] K. Mamat and W. Santipach, "Optimal transmit power and channel-information bit allocation with zeroforcing beamforming in MIMO-NOMA and MIMO-OMA downlinks," *IEEE Transactions on Communications*, vol. 71, no. 4, pp. 2028–2041, 2023.

[28] 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.

[29] J. Choi, Y. Cho, and B. L. Evans, "Quantized massive MIMO systems with multicell coordinated beamforming and power control," *IEEE Transactions on Communications*, vol. 69, no. 2, pp. 946–961, 2021.

[30] T. Phung Truong, T. My Tuyen Nguyen, T. Vi Nguyen, N.-N. Dao, and S. Cho, "Energy efficiency in rsma-enhanced active ris-aided quantized downlink systems," *IEEE Journal on Selected Areas in Communications*, vol. 43, no. 3, pp. 834–850, 2025.

[31] 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.

[32] S. Chen, S. X. Ng, E. F. Khalaf, A. Morfeq, and N. D. Alotaibi, "Multiuser detection for nonlinear MIMO uplink," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 207–219, 2020.

[33] A. Krishnamoorthy and R. Schober, "Uplink and downlink MIMO-NOMA with simultaneous triangularization," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3381–3396, 2021.

[34] 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.

[35] 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.

[36] 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.

[37] S. Cui, A. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2349–2360, 2005.

[38] 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.

[39] W. J. Yun, S. Park, J. Kim, M. Shin, S. Jung, D. A. Mohaisen, and J.-H. Kim, "Cooperative multiagent deep reinforcement learning for reliable surveillance via autonomous multi-UAV control," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 7086–7096, 2022.

[40] D. Kwon, J. Jeon, S. Park, J. Kim, and S. Cho, "Multiagent DDPG-based deep learning for smart ocean federated learning IoT networks," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9895–9903, 2020.

[41] S. Park, C. Park, and J. Kim, "Learning-based cooperative mobility control for autonomous drone-delivery," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 4, pp. 4870–4885, 2024.

[42] T. T. H. Pham, W. Noh, and S. Cho, "Multi-agent reinforcement learning based optimal energy sensing threshold control in distributed cognitive radio networks with directional antenna," *ICT Express*, 2024.

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

[44] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.

[45] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 4075–4089, 2017.

[46] B. Fesl, N. Turan, B. Böck, and W. Utschick, "Channel estimation for quantized systems based on conditionally gaussian latent models," *IEEE Transactions on Signal Processing*, vol. 72, pp. 1475–1490, 2024.

[47] 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

[48] T. Lillicrap, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[49] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[50] J. Han and C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *International workshop on artificial neural networks*. Springer, 1995, pp. 195–201.

[51] T. P. Truong, A.-T. Tran, V. D. Tuong, N.-N. Dao, and S. Cho, "NOMA-enhanced quantized uplink multi-user MIMO communications," in *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications*, 2024, pp. 281–290.

[52] 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.

[53] 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