A DRL Framework to Optimize Energy Efficiency for Quadrature Rate-Splitting Multiple Access

Anh-Tien Tran
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
attran@uclab.re.kr

Thanh Phung Truong
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
tptruong@uclab.re.kr

Dang-Huy Mac
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
hdmac@uclab.re.kr

Dongwook Won
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
dwwon@uclab.re.kr

Jaemin Kim
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
jmkim@uclab.re.kr

Nhu-Ngoc Dao
Department of Computer Science and Engineering
Sejong University
Seoul, Korea (South)
nndao@sejong.ac.kr

Sungrae Cho
School of Computer Science and Engineering
Chung-Ang University
Seoul, Korea (South)
srcho@cau.ac.kr

***Abstract*—Maximizing energy efficiency (EE) in Multiple-Input Single-Output (MISO) downlink networks employing Quadrature Rate-Splitting Multiple Access (Q-RSMA) is a challenging task due to the non-convex optimization problem involving power allocations, beamforming vectors, and rate allocations under multiple constraints. In this paper, we propose a Deep Reinforcement Learning (DRL) framework based on the Deep Deterministic Policy Gradient (DDPG) algorithm to maximize EE. We handle the minimum rate constraints by formulating the rate allocation as a linear programming (LP) problem, which allows for a computationally efficient solution. The beamforming vector normalization is clarified to ensure the unit norm constraints are satisfied. Simulation results demonstrate the effectiveness of the proposed approach in achieving high EE while satisfying all system constraints.**

***Index Terms*—Energy efficiency, Quadrature Rate-Splitting Multiple Access, Deep Reinforcement Learning, DDPG, MISO downlink networks.**

I. Introduction

The demand for high data rates and energy-efficient communication has driven the exploration of advanced multiple access schemes to address the challenges of spectral scarcity and increasing user density in modern wireless networks [1]. Among these schemes, Quadrature Rate-Splitting Multiple Access (Q-RSMA) has emerged as a cutting-edge technique that capitalizes on the quadrature components of the signal to enable the simultaneous transmission of both common and private messages [2]. By exploiting this dual signal representation, Q-RSMA achieves improved spectrum utilization and enhanced interference management, making it a strong candidate for next-generation wireless communication systems, including 6G networks.

In the context of MISO (multiple-input single-output) downlink networks, the adoption of Q-RSMA necessitates sophisticated resource allocation strategies to maximize energy efficiency while satisfying key operational constraints. These include a stringent total power budget, minimum rate guarantees for each user, and effective interference suppression. The optimization of power allocations, beamforming vectors, and rate allocations collectively forms a highly coupled and non-convex problem, which is further exacerbated by the dynamic and heterogeneous nature of wireless environments. [3].

Traditional optimization techniques, such as iterative algorithms or convex approximation methods, often struggle to deliver real-time solutions due to their computational complexity and reliance on accurate channel state information (CSI). This motivates the adoption of machine learning-based approaches, particularly Deep Reinforcement Learning (DRL), which has demonstrated remarkable potential in solving complex, non-convex optimization problems in wireless networks [4], [5].

In this work, we introduce a DRL-based framework leveraging the Deep Deterministic Policy Gradient (DDPG) algorithm to optimize energy efficiency in Q-RSMA-enabled MISO downlink networks. The DDPG algorithm, an actor-critic method tailored for continuous action spaces, is particularly well-suited for the joint optimization of beamforming and power allocation. By training the DRL agent in a simu-

lated environment, we enable it to learn an efficient policy that dynamically adapts to varying channel conditions and user requirements [6], [7]. This approach not only reduces computational overhead but also enhances the scalability and adaptability of Q-RSMA systems in practical deployments.

Our proposed framework is validated through extensive simulations that evaluate its performance under realistic scenarios, including imperfect CSI, user mobility, and interference-limited conditions. The results demonstrate significant gains in energy efficiency compared to conventional optimization techniques and other multiple access schemes, highlighting the viability of integrating DRL with Q-RSMA to meet the demands of future wireless networks.

II. System Model

We consider a downlink communication system where a base station (BS) equipped with M antennas serves N single-antenna users using Q-RSMA.

A. Signal Model

The transmitted signal $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is given by:

$$\mathbf{x} = \sqrt{p_0}\Re\left\{\mathbf{w}_c s_c\right\} + j\sum_{n=1}^{N}\sqrt{p_n}\Im\left\{\mathbf{w}_n s_n\right\}, \tag{1}$$

where s_c is the common message symbol with $\mathbb{E}[|s_c|^2] = 1$, s_n is the private message symbol for user n with $\mathbb{E}[|s_n|^2] = 1$, $\mathbf{w}_c \in \mathbb{C}^{M\times 1}$ is the beamforming vector for the common message, $\mathbf{w}_n \in \mathbb{C}^{M\times 1}$ is the beamforming vector for the private message of user n. Beside, p_0 is the power allocated to the common message and p_n is the power allocated to the private message of user n.

B. Received Signal

The received signal at user n is:

$$y_n = \sqrt{p_0}\Re\left\{\mathbf{h}_n^H\mathbf{w}_c s_c\right\} + j\sqrt{p_n}\Im\left\{\mathbf{h}_n^H\mathbf{w}_n s_n\right\} + z_n, \tag{2}$$

where $\mathbf{h}_n \in \mathbb{C}^{M\times 1}$ is the channel vector between the BS and user n, and $z_n \sim \mathcal{CN}(0,\sigma^2)$ is the additive white Gaussian noise. Due to the orthogonality of the in-phase and quadrature components, when decoding the common message, there is no interference from the private messages.

III. Problem Formulation

Our goal is to maximize the energy efficiency (EE) defined as the ratio of the total achievable rate to the total power consumption.

A. Achievable Rates

1) Common Message Rate: The Signal-to-Noise Ratio (SNR) for the common message at user n is:

$$\gamma_{0,n} = \frac{p_0|\mathbf{h}_n^H\mathbf{w}_c|^2}{\sigma^2/2}. \tag{3}$$

The achievable rate for the common message at user n is:

$$R_{0,n} = \frac{B}{2}\log_2\left(1+\gamma_{0,n}\right). \tag{4}$$

The common message rate is determined by the worst-case user:

$$R_0 = \min_n R_{0,n}. \tag{5}$$

2) Private Message Rates: The Signal-to-Interference-plus-Noise Ratio (SINR) for the private message at user n is:

$$\gamma_n = \frac{p_n|\mathbf{h}_n^H\mathbf{w}_n|^2}{\sum_{k\neq n}p_k|\mathbf{h}_n^H\mathbf{w}_k|^2 + \sigma^2/2}. \tag{6}$$

The achievable rate for the private message at user n is:

$$R_n = \frac{B}{2}\log_2\left(1+\gamma_n\right). \tag{7}$$

3) Total Rate for User n: Each user's total rate is:

$$R_n^{\text{total}} = R_{0,n}^{\text{alloc}} + R_n, \tag{8}$$

where $R_{0,n}^{\text{alloc}}$ is the portion of the common message rate allocated to user n.

B. Optimization Problem

We formulate the optimization problem as:

$$\max_{\substack{p_0,\{p_n\},\\ \mathbf{w}_c,\{\mathbf{w}_n\},\\ \{R_{0,n}^{\text{alloc}}\}}} \quad \eta = \frac{\sum_{n=1}^{N}\left(R_{0,n}^{\text{alloc}} + R_n\right)}{p_0 + \sum_{n=1}^{N}p_n + P_{\text{circuit}}} \tag{9}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} R_{0,n}^{\text{alloc}} \le R_0 \tag{10}$$

$$R_{0,n}^{\text{alloc}} + R_n \ge R_n^{\min}, \quad \forall n \tag{11}$$

$$p_0 + \sum_{n=1}^{N} p_n \le P_{\max} \tag{12}$$

$$\|\mathbf{w}_c\| \le 1, \quad \|\mathbf{w}_n\| \le 1, \quad \forall n \tag{13}$$

$$p_0 \ge 0, \quad p_n \ge 0, \quad R_{0,n}^{\text{alloc}} \ge 0, \quad \forall n \tag{14}$$

IV. Proposed DRL Solution

To solve the optimization problem in (9)–(14), we propose a DRL framework using the DDPG algorithm. The DRL agent learns to make decisions on power allocation, beamforming design, and rate allocation.

A. DDPG Algorithm

DDPG is suitable for high-dimensional continuous action spaces. It consists of actor and critic networks that are updated using sampled policy gradients.

1) State and Action Spaces:

- **State Space $\mathcal{S}$:** The state s_t includes the real and imaginary parts of the channel vectors:

$$s_t = \left[\text{Re}(\mathbf{H}), \text{Im}(\mathbf{H})\right], \tag{15}$$

  where $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N]$.
- **Action Space $\mathcal{A}$:** The action a_t includes:

$$a_t = \left[p_0, \{p_n\}, \text{vec}(\mathbf{w}_c), \{\text{vec}(\mathbf{w}_n)\}, \{R_{0,n}^{\text{alloc}}\}\right]. \tag{16}$$

2) Reward Function: The reward is designed to maximize energy efficiency:

$$r_t = \eta_t = \frac{\sum_{n=1}^{N} \left(R_{0,n}^{\text{alloc}} + R_n\right)}{p_0 + \sum_{n=1}^{N} p_n + P_{\text{circuit}}}. \tag{17}$$

B. Action Post-Processing for Constraint Satisfaction

After the agent outputs an action a_t, we apply post-processing to ensure all constraints are satisfied.

Proposition IV.1. *Given any beamforming vector* $\mathbf{w} \in \mathbb{C}^{M\times 1}$, *the normalized beamforming vector* $\mathbf{w}_{norm} = \frac{\mathbf{w}}{\max\{\|\mathbf{w}\|, 1\}}$ *satisfies the unit norm constraint* $\|\mathbf{w}_{norm}\| \leq 1$.

Proof. If $\|\mathbf{w}\| \leq 1$, then $\mathbf{w}_{\text{norm}} = \mathbf{w}$, so $\|\mathbf{w}_{\text{norm}}\| = \|\mathbf{w}\| \leq 1$.
If $\|\mathbf{w}\| > 1$, then $\mathbf{w}_{\text{norm}} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$, so $\|\mathbf{w}_{\text{norm}}\| = 1$.
Therefore, in both cases, $\|\mathbf{w}_{\text{norm}}\| \leq 1$. ■

Proposition IV.2. *To satisfy the minimum rate constraints, the common rate allocations* $\{R_{0,n}^{alloc}\}$ *can be obtained by solving the following linear programming (LP) problem:*

$$\max_{\{R_{0,n}^{alloc}\}} \quad \sum_{n=1}^{N} R_{0,n}^{alloc} \tag{18}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} R_{0,n}^{alloc} \leq R_0 \tag{19}$$

$$R_{0,n}^{alloc} \geq R_n^{min} - R_n, \quad \forall n \tag{20}$$

$$R_{0,n}^{alloc} \geq 0, \quad \forall n \tag{21}$$

Proof. The LP aims to maximize the total allocated common rate while ensuring *(i)* the total allocated common rate does not exceed R_0; *(ii)* each user meets the minimum rate requirement $R_n^{\min}$; and *(iii)* allocated rates are non-negative.
Since R_n and $R_n^{\min}$ are constants in this LP, the problem is linear and can be efficiently solved. ■

We propose to use traditional water-filling algorithm for rate allocation due to its simplicity and effectiveness. The water-filling method is well-known for allocating power or rates optimally under certain conditions. We initially set $R_{0,n}^{\text{alloc}}$ to 0 for all users as starting values. For each user n, we compute the deficit between the minimum required rate and the private rate:

$$D_n = R_n^{\min} - R_n \tag{22}$$

Note that users with $D_n \leq 0$ do not need additional common rate allocation to meet the minimum rate requirement. To allocate the common rate R_0 efficiently among users using the water-filling principle, we set water level λ such that:

$$\sum_{n\in\mathcal{N}_+} \max\left(0, \lambda - D_n\right) = R_0 \tag{23}$$

where $\mathcal{N}_+$ is the set of users with $D_n > 0$. Then, for each user $n \in \mathcal{N}_+$, specify the amount of allocated rate as:

$$R_{0,n}^{\text{alloc}} = \max\left(0, \lambda - D_n\right) \tag{24}$$

and adjust λ accordingly if the sum of allocated rates exceeds R_0.

Algorithm 1: DDPG Algorithm with Action Post-Processing

Initialize actor network $\mu(s|\theta^\mu)$ and critic network $Q(s,a|\theta^Q)$ with random weights
Initialize target networks $\theta^{\mu'} \leftarrow \theta^\mu$, $\theta^{Q'} \leftarrow \theta^Q$
Initialize replay buffer $\mathcal{D}$
for *episode* $= 1$ *to* M **do**
  Initialize a random process $\mathcal{N}$ for action exploration
  Receive initial state s_1
  for $t = 1$ *to* T **do**
    Select action $a_t = \mu(s_t|\theta^\mu) + \mathcal{N}_t$
    Adjust power allocations to satisfy total power constraint (12)
    Normalize beamforming vectors using Proposition IV.1
    Compute achievable rates R_n, R_0
    Use water-filling algorithm to obtain $R_{0,n}^{\text{alloc}}$
    Apply action a_t to the environment
    Observe reward r_t and next state s_{t+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in $\mathcal{D}$
    Sample a minibatch from $\mathcal{D}$
    Update critic by minimizing loss:
      $L = \frac{1}{D}\sum_{i=1}^{D}\left(y_i - Q(s_i, a_i|\theta^Q)\right)^2$
      where $y_i = r_i + \gamma Q(s_{i+1}, \mu(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$
    Update actor using policy gradient:
      $\nabla_{\theta^\mu} J \approx$
      $\frac{1}{D}\sum_{i=1}^{D} \nabla_a Q(s,a|\theta^Q)|_{s_i, a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s_i|\theta^\mu)$
    Update target networks:
      $\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$
      $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$
    Update state $s_t \leftarrow s_{t+1}$
  end
end

C. Algorithm

The overall algorithm is summarized in Algorithm 1. The presented algorithm is a modified Deep Deterministic Policy Gradient (DDPG) framework tailored for environments requiring power and beamforming optimizations. It begins by initializing the actor and critic networks, as well as their target counterparts, and employs a replay buffer for experience storage. During each episode, actions are selected using the actor network with added noise for exploration, followed by post-processing steps to meet power constraints and normalize beamforming vectors. Achievable rates are computed, and a water-filling algorithm is used for resource allocation. The algorithm stores transitions, updates the critic using a temporal-difference loss, and refines the actor via policy gradient. Target networks are updated using soft updates to ensure stable learning. These enhancements enable effective policy learning in scenarios with strict constraints on resource allocation.

V. Conclusion

We proposed a DRL framework using the DDPG algorithm to maximize energy efficiency in Q-RSMA MISO downlink networks. By formulating the rate allocation as a linear programming problem and normalizing the beamforming vectors, we ensure all constraints are satisfied without iterative adjustments. The proposed method efficiently solves the original optimization problem and is suitable for practical implementations.

Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2025-RS-2022-00156353) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

References

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

[2] G. Arora and A. Jaiswal, "Zero sic based rate splitting multiple access technique," *IEEE Communications Letters*, vol. 26, no. 10, pp. 2430–2434, 2022.

[3] A. Mishra, Y. Mao, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access for downlink multiuser mimo: Precoder optimization and phy-layer design," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 874–890, 2021.

[4] A. Alwarafy, M. Abdallah, B. S. Ciftler, A. Al-Fuqaha, and M. Hamdi, "Deep reinforcement learning for radio resource allocation and management in next generation heterogeneous wireless networks: A survey," *arXiv preprint arXiv:2106.00574*, 2021.

[5] A. M. Nagib, H. Abou-zeid, and H. S. Hassanein, "Toward safe and accelerated deep reinforcement learning for next-generation wireless networks," *IEEE Network*, vol. 37, no. 2, pp. 182–189, 2022.

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

[7] T. P. Truong, A.-T. Tran, T. M. T. Nguyen, T.-V. Nguyen, A. Masood, and S. Cho, "Mec-enhanced aerial serving networks via hap: A deep reinforcement learning approach," in *2022 international conference on information networking (ICOIN)*. IEEE, 2022, pp. 319–323.