Weighted Sum-Rate Maximization in Rate-Splitting MISO Downlink Systems

Anh-Tien Tran, Thanh Phung Truong, Dongwook Won, Nhu-Ngoc Dao, and Sungrae Cho

Weighted Sum-Rate Maximization in Rate-Splitting MISO Downlink Systems

***Abstract*—Rate-splitting multiple access (RSMA) and successive interference cancellation (SIC) are essential approaches in the next-generation communication systems that boost spectrum efficiency by effectively managing and mitigating interference between multiple signals. However, a challenge arises in ensuring that users can distinguish the common message from the remaining non-decoded private messages without considering a separate SIC constraint per user. This imperfection cancellation leads to residual interference from the common stream that remains in the received signal. This work investigates the maximization of the weighted sum-rate (WSR) in single-layer RSMA multiple input single output (MISO) downlink network by proposing explicit SIC constraints. In particular, we suggest an approach that initially addresses the critical problem of allocating power and precoding vectors for streams using a deep reinforcement learning (DRL) method, and then determines the user-specific allocations within the common rate to meet the criteria of users' minimum rate by solving a linear programming problem. Simulation results exhibit the supremacy of the proposed DRL framework over SDMA and other DRL approaches in terms of spectral efficiency leading to an improvement of approximately 30% of WSR in several scenarios.**

***Index Terms*—Rate-splitting multiple access, deep deterministic policy gradient, successive interference cancellation, weighted sum-rate, deep reinforcement learning, multiple-input-single-output downlink network.**

I. Introduction

THE rapid evolution of wireless communication technologies has led to an unprecedented surge in demand for high-speed, reliable, and efficient data transmission. As we transition into the era of 5G and beyond, the need for advanced multiple access techniques has become increasingly critical to address the growing challenges of spectrum scarcity, network capacity, and diverse user requirements [1]. In current wireless communication systems, multiple access schemes play a pivotal role in enabling simultaneous connectivity for a multitude of users while optimizing the utilization of limited spectral resources. The explosive growth of Internet of Things (IoT) devices, the proliferation of multimedia applications, and the emergence of new use cases such as ultra-reliable low-latency communications (URLLC) and massive machine-type communications (mMTC) have further intensified the demand for innovative multiple access solutions [2]. Traditional multiple access techniques, such as Time Division Multiple Access (TDMA), Frequency Division Multiple Access (FDMA), and Code Division Multiple Access (CDMA), have served as the backbone of wireless networks for decades [3]. However, these conventional approaches are increasingly challenged by the exponential growth in connected devices and modern applications' ever-increasing data rate requirements.

In response to these challenges, researchers and industry stakeholders have been actively developing and implementing advanced multiple access schemes. Rate-splitting multiple access (RSMA) is a communication technique developed to boost the efficiency and dependability of wireless networks [4]. In contrast to traditional multiple access techniques, RSMA leverages non-orthogonal resource sharing by combining shared and private messages, instead of allocating fixed or orthogonal resources to each user. The system's versatility allows for efficient interference management and adaptation to diverse user requirements, resulting in improved spectrum efficiency, faster data rates, and enhanced quality of service (QoS). RSMA's importance grows as wireless networks evolve toward large-scale connectivity and diverse applications. This technology aligns well with 5G and beyond, effectively addressing interference and spectrum scarcity challenges for future generations. RSMA facilitates concurrent information delivery to multiple users in a shared frequency spectrum by dividing user messages into separate sections and applying individual encoding. In [5], a performance comparison of RSMA, SDMA, and NOMA was conducted in multiple-input single-output (MISO) networks, specifically focusing on sum-rate (WSR). RSMA outperformed the other methods in terms of WSR. Although RSMA has several exceptional benefits, maximizing its WSR presents a significant challenge. In contrast to conventional techniques, RSMA requires the transmitter to precisely designate transmission power to each message, ensuring compliance with power constraints and QoS requirements. In another study, [6] concurrently optimized the precoder design and sum-rate maximization, considering that the channel state may fluctuate during transmission in accordance with certain known stationary distributions. Numerous articles on RSMA downlink transmission strategies have focused on developing suboptimal algorithms with controllable and affordable complexity, such as [7]–[13], to reduce com-

This work was supported in part by the IITP (Institute of Information & Communications Technology Planning & Evaluation) - ITRC (Information Technology Research Center) (IITP-2025-RS-2022-00156353, 50%) grant funded by the Korea government (Ministry of Science and ICT), in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00209125), and in part by the Chung-Ang University Young Scientist Scholarship in 2018.

A.-T. Tran, T.P. Truong, D. Won and S. Cho are with Chung-Ang University, School of Computer Science and Engineering, Seoul, Republic of Korea (emails: {attran, tptruong, dwwon }@uclab.re.kr, and srcho@cau.ac.kr)

N.-N. Dao is with Department of Computer Science and Engineering, Sejong University, Seoul, Republic of Korea (email: nndao@sejong.ac.kr).

Corresponding authors: Nhu-Ngoc Dao and Sungrae Cho

putational complexity. The applications of RSMA have been extended to various areas, including uplink networks [14], caching networks [15], and aerial networks [16].

When comparing RSMA with Non Orthogonal Multiple Access (NOMA), it is evident that RSMA offers several advantages making it a superior option in certain situations [11]. RSMA employs a distinctive method that divides a user's message into shared and individual components, using interference to improve the efficiency of the frequency spectrum without stringent orthogonality requirements. In contrast to NOMA, RSMA leverages interference exploitation to enhance spectral efficiency and potentially reduce decoding complexity, rather than relying on power domain multiplexing and successive interference cancellation (SIC). An important distinction in single-layer RSMA downlink networks is that SIC is solely used to extract the common message from private messages, without the need to consider SIC decoding order as in NOMA and other RSMA schemes. Therefore, the rate calculation does not depend on SIC decoding order. This simplifies the receiver design and reduces complexity, making RSMA a more practical choice in certain scenarios. However, introducing SIC constraints poses challenges in ensuring that users can successfully decode the common message without interference from the remaining non-decoded private messages. This necessitates careful power allocation and beamforming design to satisfy the SIC constraints, which can be complex in multi-user scenarios. Balancing the power allocation between common and private messages while meeting users' quality of service (QoS) requirements and adhering to system constraints becomes a significant challenge.

A. Contributions

In this study, we optimize the power allocation coefficient vectors using a DRL framework, focusing on users' common rate share. A state is defined by observed channel gains between the BS and its associated users, and the minimum transmission rate requirement of each user. Employing the Deep Deterministic Policy Gradient (DDPG) [17] method, actions are generated, and a linear LP problem is subsequently solved. The crucial contributions of this paper are as follows:

- We investigate a MISO downlink system in which a multi-antenna base station (BS) serves multiple single-antenna users. In this system, we implement the RSMA technique to enhance transmission efficiency from the BS to users, all operating within the same transmission bandwidth. We aim to maximize the system's weighted sum rate by optimizing the BS's beamforming vectors, power allocation, and common rate allocation. To ensure the practicality of the RSMA system, we also consider the constraints of the SIC process. These constraints guarantee that the minimum signal power received by each user meets the threshold required for successful decoding, thereby maintaining the system's reliability and performance.
- Expressing rate allocation and power control as an optimization issue with the objective of maximizing the WSR in downlink MISO systems. To address this challenge, we propose a procedure for carefully determining beamforming vectors and power share ratios for both common and private messages. Subsequently, the data rate allocation to each user is split to meet diverse minimum data rate requirements.
- Leveraging the DDPG framework to simultaneously generate temporary values of beamforming vectors and power allocation ratios. In the suggested framework, the user precoding matrix verification and compliance with the power budget limitation of the BS are supported by two novel propositions. Subsequently, using the transmission rates derived from the first phase, we find the closed-form expressions of rate allocation such that all individual rate requirements are satisfied.
- Employing simulations to demonstrate that the proposed DRL framework achieves significant gains in terms of spectral efficiency compared with other learning benchmarks and the SDMA case. We compare the performance with and without SIC constraints and analyze the impact of SIC constraints on system performance.

The remaining sections of this paper are organized as follows. Section III describes the system model and problem formulation. Section IV details the state and action spaces, and reward functions designed for the DDPG framework. In Section V-A, detailed explanations and proofs of propositions aiming at refining the action of neural networks are provided. The training process for the DDPG is described in Section IV-C. Simulation results are interpreted in Section VI, and conclusions are written in Section VII.

II. Related Works

The rapid evolution of multiple access technologies has brought RSMA to the forefront as a highly promising candidate for future wireless networks. From a broader network science and engineering perspective, RSMA-based beamforming and rate allocation can be interpreted as a networked control policy acting on a multi-user interference graph, where the base station adaptively steers power and common/private streams to regulate interference and satisfy user quality-of-service (QoS) constraints. Recent research has systematically demonstrated the unique flexibility and performance gains of RSMA compared to traditional schemes, while parallel works on radio resource management and deep learning-based control highlight the importance of scalable optimization and learning frameworks in large networked systems.

Park et al. [18] developed a generalized power-iteration-based linear precoder for downlink MIMO RSMA with imperfect CSIT. By approximating the minimum common-rate term with a LogSumExp function and recasting the sum spectral efficiency maximization into a nonlinear eigenvalue problem, their algorithm attains comparable or higher spectral efficiency than WMMSE-based designs at reduced complexity. However, this framework remains deterministic and does not provide real-time, learning-based adaptation or explicit SIC-gap control, which are central to our DRL-based approach. Building on RSMA's robust performance, Zhou et al. [13] benchmarked RSMA, SDMA, and NOMA in multi-antenna

TABLE I
COMPARISON OF RELATED WORKS WITH OUR PROPOSED METHOD (PART I)

Reference	Focus	Main Contribution	Key Difference from Our Work
RSMA precoding / performance analysis			
[18]	Downlink MIMO RSMA with incomplete CSIT.	Proposed a GPI-based linear precoder for downlink MIMO RSMA with imperfect CSIT, outperforming WMMSE in sum spectral efficiency.	Offline deterministic precoding without learning or explicit SIC-gap tuning; we learn an online DDPG policy that jointly sets beams, powers, and SIC gaps in single-cell MISO RSMA.
[13]	Spectral/energy efficiency tradeoff in multi-antenna RSMA.	Jointly optimized RSMA precoders and power via WMMSE and fractional programming to characterize SE-EE tradeoffs versus SDMA/NOMA.	Uses snapshot-based iterative solvers with no DRL or SIC-gap actions; we replace them by a DDPG controller that outputs beamforming and power under explicit SIC-gap constraints.
[15]	RSMA-enabled wireless caching networks with latency focus.	Minimized user-perceived latency by jointly optimizing caching placement and RSMA resource allocation.	Targets cache/latency design in content-centric networks, whereas we solve PHY-layer WSR maximization with SIC-gap enforcement in a generic one-layer MISO RSMA downlink.
[16]	Survey and analysis of OMA, NOMA, and RSMA for aerial/UAV networks.	Surveyed architectures, performance trends, and challenges of OMA/NOMA/RSMA for aerial systems and provided design guidelines.	Offers system-level guidelines but no concrete joint beamforming-power-SIC control algorithm; our work provides a fully specified DRL-based control policy for downlink MISO RSMA.
[19]	RSMA precoding for aerial computing networks using deep CNNs.	Trained a CNN (DCLPNN) to emulate AO/WMMSE-based RSMA precoders in aerial computing networks with much lower online complexity.	Emulates offline AO/WMMSE precoders without explicit SIC-gap variables or online policy learning; we directly learn beamforming and power actions with SIC-gap control in terrestrial MISO RSMA.
DRL-based RSMA control			
[20]	DRL for RSMA power allocation in downlink MU-MISO.	Applied PPO to learn RSMA power allocation under QoS and power constraints, improving sum-rate over heuristic baselines.	Optimizes only power with fixed beamforming and no explicit SIC-gap variable; our DDPG agent jointly optimizes beamforming and power with SIC-gap-aware action refinement and LP-based common-rate allocation.
[21]	DRL-based RSMA resource allocation in LEO satellite-terrestrial networks.	Designed a hybrid DQN/PPO framework for joint resource and power allocation in LEO satellite-terrestrial RSMA systems.	Targets mixed satellite-terrestrial topology with hybrid discrete/continuous actions and no explicit SIC-gap modeling; we consider a terrestrial single-cell MISO RSMA with fully continuous DDPG control and SIC-gap constraints.
[22]	Delay analysis and DRL-based power control in downlink RSMA.	Derived SNC-based delay-violation bounds and defined a delay/energy utility optimized via DDPG-based RSMA power and resource allocation.	Uses DDPG to tune RSMA power and resource fractions under SNC-based delay and energy efficiency, without beamforming design or explicit SIC-gap margins; we optimize beamforming and power for a WSR objective with explicit SIC-gap modeling and no SNC delay component.
Deterministic RSMA rate / power / QoS optimization			
[23]	Uplink RSMA sum-rate maximization.	Analyzed uplink RSMA and provided closed-form and exhaustive-search solutions for power allocation and decoding order to maximize sum-rate.	Addresses uplink optimization via analytical/exhaustive solvers rather than a downlink DRL controller; our framework is for downlink MISO RSMA and learns a continuous beamforming-power-SIC policy.
[24]	RSMA precoding design and complexity reduction.	Proposed interference-nulling and hierarchical stream design to reduce RSMA precoding complexity while benchmarking rate performance.	Uses deterministic low-complexity precoders with no learning or SIC-gap variables; our method keeps general beamforming and delegates complexity and SIC-gap handling to a DRL controller.
[25]	Rate allocation and power control in SISO/MISO RSMA.	Formulated joint rate and power control with user-specific SIC constraints and solved it via sequential convex approximation.	Relies on SCA-based snapshot optimization of rate/power with SIC constraints; our approach learns a DDPG policy that outputs beamforming and power online, embedding SIC-gap constraints into action refinement rather than repeatedly solving SCA problems.
[26]	Resource allocation for MU-MISO RSMA under URLLC.	Proposed block coordinate descent and SCA algorithms to maximize effective throughput under URLLC constraints and imperfect SIC.	Focuses on URLLC-oriented effective throughput using deterministic solvers; we address generic WSR optimization via DRL with SIC-gap constraints, without explicit URLLC latency modeling.
HARQ and reliability with RSMA			
[27]	HARQ protocol design assisted by RSMA.	Proposed two flexible HARQ-RSMA policies that mix RSMA/NOMA/SDMA with IR/CC combining and analyzed throughput, delay, and EE under perfect and imperfect CSI.	Operates at the HARQ/link layer by adapting retransmission mode and combining while keeping PHY signaling fixed; our work keeps standard HARQ and instead adapts PHY beamforming, power, and SIC gaps via DRL.

downlink systems, optimizing spectral and energy efficiency via fractional programming and WMMSE-based iterative algorithms. Their methodology thoroughly characterizes RSMA's SE/EE gains but does not incorporate explicit SIC power-gap constraints or adaptive, data-driven optimization, which our unified DRL framework provides. Beyond general system performance, recent research has also explored the integration of RSMA with new network functionalities. For instance, Hua et al. [15] focused on latency reduction in RSMA-enabled wireless caching networks by jointly optimizing caching decisions and RSMA transmission parameters. They treat long-term caching via dynamic programming and solve the short-term resource allocation with an alternating optimization framework and one-dimensional search to minimize average user-perceived latency. While effective for content-driven latency optimization, this design is not aimed at generic sum-rate maximization or explicit SIC-gap control as in our MISO RSMA setting. Extending the scope, Jaafar et al. [16] surveyed OMA, NOMA, and RSMA for aerial and UAV networks, detailing their respective rate-splitting strategies, SIC mechanisms, and comparative performance, and outlining key research challenges for RSMA adoption. Their work is primarily conceptual and does not propose a concrete resource-allocation or learning-based optimization framework, a gap our study addresses. In the related context of aerial computing networks, Wang et al. [19] designed a deep convolutional linear precoder neural network (DCLPNN) that maps CSI and user weights to RSMA linear precoders, closely approximating the sum-rate performance of AO-WMMSE-based designs while markedly reducing online computation and feedback overhead. Their beamformer exemplifies offline data-driven RSMA precoding but does not provide an online DRL controller that jointly adapts beamforming and SIC margins as in our approach. These RSMA precoding and performance-analysis works are summarized in the first block of Table I.

More recently, Hieu et al. [20] applied a proximal policy optimization (PPO)-based DRL scheme for downlink RSMA, learning a power allocation policy that maximizes long-term sum-rate under a total power constraint and varying QoS settings. Their agent operates on scalar power variables and improves sum-rate over heuristic and optimization-based baselines, but it does not model SIC-gap variables explicitly or jointly optimize beamforming and SIC control as in our design. Similarly, Huang et al. [21] developed a hybrid DQN/PPO-based DRL framework for RSMA in LEO satellite-terrestrial networks, where a DQN agent allocates discrete resources (e.g., subchannels) and a PPO agent controls continuous transmit powers to maximize system utility. Their focus is on hybrid action-space control in LEO satellite environments, whereas our framework targets terrestrial MISO RSMA with a single continuous DDPG agent and explicit SIC-gap enforcement. Complementing these efforts, Chen et al. [22] combined stochastic network calculus with a DDPG-based power control policy for downlink RSMA, deriving analytical delay-violation probability bounds and training a DRL agent to optimize a utility that balances energy efficiency and delay. Their scheme adapts scalar RSMA power allocation to channel dynamics under strict latency constraints but does not optimize beamforming or treat SIC-gap parameters as explicit control variables as in our framework.

Other algorithmic alternatives exist: Yang et al. [23] studied sum-rate maximization for uplink RSMA, deriving a closed-form optimal rate-splitting vector and employing exhaustive search (with low-complexity heuristics) over decoding orders and power allocations under SIC decoding constraints. In contrast, our approach targets downlink MISO RSMA and replaces such offline combinatorial search with DRL-based continuous control and explicit SIC-gap modeling; Sadeghabadi and Blostein [24] proposed an RSMA precoding design based on interference nulling and a sum-rate upper bound, where hierarchical streams are used to transmit only a subset of possible streams, greatly reducing precoding complexity compared with exhaustive RSMA. Their approach is fully deterministic and does not address real-time RSMA control via RL or explicit SIC-gap enforcement as in our work. These DRL-based RSMA controllers, together with deterministic rate/power/QoS optimization and HARQ-enhanced RSMA reliability mechanisms, are grouped in the remaining blocks of Table I, which contrasts them with our unified DDPG-based RSMA control policy.

Beyond RSMA-specific designs, a rich body of work has investigated radio resource management and DRL-driven network control in interference-limited wireless systems. Zhang et al. [28] proposed a deep learning-based radio resource management framework for mmWave NOMA heterogeneous networks, where user association is handled via Lagrange dual decomposition and semi-supervised DNNs learn subchannel and power allocation to maximize energy efficiency under QoS, power, and interference constraints. Their scheme targets NOMA and offline-trained DNNs, whereas we focus on RSMA and online DDPG-based beamforming/power control with explicit SIC-gap constraints. At a larger network scale, Alwarafy et al. [29] introduced DeepRAT, a hierarchical multi-agent DRL framework for joint multi-RAT assignment and power allocation in heterogeneous networks. A single-agent DQN selects the radio access technology for each edge device, while a multi-agent DDPG stage allocates downlink transmit powers to maximize a cost-aware sum-rate objective. Their results show that DRL scales to high-dimensional heterogeneous resource-allocation problems, reinforcing the feasibility of our DDPG-based RSMA control design, but they do not exploit RSMA stream structures or SIC-gap control. From the perspective of joint communication and economic incentives, Liu et al. [30] studied joint incentive and resource allocation design in user-provided networks under 5G integrated access/backhaul constraints, aligning user contributions with system utility through optimization and pricing mechanisms. Their framework focuses on economic incentives and backhaul-aware sharing rather than RSMA signaling or DRL-based beamforming as in our work, while Huang et al. [31] formulated joint request offloading and resource allocation for collaborative edge inference with time-coupled communication and computing resources, aiming to maximize a long-term utility metric. This work addresses temporal coupling and long-term optimization at the edge-application layer, whereas we optimize per-slot RSMA weighted sum-rate with

TABLE II
COMPARISON OF RELATED WORKS WITH OUR PROPOSED METHOD (PART II)

Reference	Focus	Main Contribution	Key Difference from Our Work
DRL / resource control in non-RSMA systems			
[28]	Deep learning-based RRM in mmWave NOMA HetNets.	Learned user association, subchannel allocation, and power control in mmWave NOMA HetNets to maximize energy efficiency under QoS.	Focuses on NOMA-based HetNets with offline DNN-based RRM and no RSMA or SIC-gap design; our work targets RSMA PHY-layer beamforming and power with online DDPG and explicit SIC-gap enforcement.
[29]	Hierarchical multi-agent DRL for multi-RAT heterogeneous networks.	Proposed DeepRAT, combining DQN for multi-RAT selection and multi-agent DDPG for downlink power allocation.	Operates at multi-RAT and cell-power layers with a generic PHY and no RSMA streams or SIC-gap modeling; our single-agent DDPG specializes in MISO RSMA and directly outputs beamformers, power, and SIC-gap-feasible actions.
[30]	Joint incentive and resource allocation in user-provided IAB networks.	Designed joint incentive mechanisms and resource allocation to align user participation with system utility in user-provided IAB.	Works at the network/economic layer without RSMA structure or DRL-based PHY control; our method operates at the PHY layer with explicit RSMA streams and SIC constraints.
[31]	Long-term utility optimization for collaborative edge inference.	Formulated joint request offloading and resource allocation with time-coupled computing/communication queues to maximize long-term edge-inference utility.	Optimizes long-term edge-inference utility and queuing dynamics, whereas our controller makes per-slot PHY-layer RSMA decisions to maximize WSR under SIC-gap and rate constraints.
[32]	DRL-based joint computing/communication allocation in NOMA-aided IIoT.	Developed a dual-agent PPO framework for joint NOMA communication and edge-computing resource allocation to maximize energy productivity in IIoT.	Uses dual-agent PPO for NOMA-based communication and MEC with SIC for NOMA but no RSMA or beamforming variables; our single-agent DDPG instead targets RSMA PHY beamforming and power in MISO downlink.
Non-RSMA physical-layer optimization (cell-free / UAV / CoMP)			
[33]	RB-granularity precoding and compression in cell-free mobile networks.	Jointly optimized RB-level precoding and fronthaul compression to trade off fronthaul load and user rates in cell-free networks.	Considers cell-free architectures and fronthaul constraints with no RSMA or learning; our work deals with single-cell RSMA beamforming and power via DRL under SIC-gap constraints.
[34]	User-centric clustering and beamforming for satellite-assisted cell-free systems.	Proposed user-centric clustering and robust beamforming schemes for satellite-assisted cell-free systems to maximize the minimum user rate.	Designs robust clustering and beamforming via convex optimization in satellite-assisted cell-free networks, without RSMA streams or DRL; our method targets a conventional single-cell MISO RSMA setting and replaces convex solvers with a learned DDPG controller.
[35]	Outage analysis of MISO–NOMA UAV-assisted Agri-IoT with SWIPT and TAS.	Derived closed-form and asymptotic outage probabilities and diversity orders for MISO-NOMA UAV-assisted Agri-IoT with SWIPT and transmit-antenna selection.	Provides analytical outage/diversity results for NOMA-based UAV Agri-IoT without RSMA or learning-based beamforming; our focus is RSMA downlink beamforming and power control via DRL with explicit SIC gaps.
[36]	Power/subcarrier allocation in multi-cell NOMA (CoMP).	Developed centralized and distributed algorithms for power and subcarrier allocation in CoMP NOMA to maximize throughput under SIC constraints.	Formulates CoMP-NOMA resource allocation with SIC constraints and solves it via convex optimization, remaining NOMA-based and non-learning; our method targets RSMA and replaces convex programs with a learned DDPG beamforming-power-SIC control policy.

explicit PHY-layer SIC-gap enforcement. In NOMA-aided industrial IoT systems, Lotfolahi and Ferng [32] addressed joint computing-communication resource allocation and proposed a dual-decentralized multi-agent PPO framework, where one MAPPO agent allocates edge computing resources and another manages NOMA transmission resources to maximize an energy productivity metric under task-queue and NOMA-specific constraints. This further illustrates the scalability of DRL-based resource management in interference-limited networks, albeit without RSMA streams or SIC-gap control.

In ultra-dense and cell-free architectures, Chen et al. [33] optimized resource-block-level precoding and fronthaul compression strategies in cell-free mobile networks, using an alternating optimization framework to trade off fronthaul load and user rates. Their design targets cell-free architectures without RSMA or DRL, whereas we study single-cell MISO RSMA with DDPG-based beamforming and SIC-gap-aware power control, whereas Lee et al. [34] designed user-centric clustering and coordinated beamforming strategies for satellite-assisted cell-free networks, formulating robust beamforming problems to enhance coverage and minimum user rates under power constraints. Their work addresses clustering and convex beamforming in integrated satellite-cell-free architectures but does not involve RSMA streams or DRL-based online control as in our terrestrial MISO RSMA design. Finally, He et al. [35] analyzed outage probability and diversity performance for MISO-NOMA UAV-assisted agricultural IoT systems with SWIPT and transmit-antenna selection, deriving closed-form and asymptotic outage expressions for edge sensors and showing that properly designed TAS significantly improves reliability. This work provides analytical reliability guarantees for NOMA-based UAV systems, whereas we study DRL-based RSMA beamforming with explicit SIC-gap constraints in terrestrial downlink networks. These DRL and resource-

control methods in non-RSMA systems, together with non-RSMA physical-layer optimization in cell-free, UAV, and CoMP settings, are systematically contrasted with our RSMA design in Table II.

Emerging multiple access techniques such as RSMA and NOMA fundamentally rely on SIC to enhance throughput and user fairness. As a result, an important thread in recent research has been the explicit consideration of SIC power constraints in resource allocation. Yang et al. [25] directly address this aspect by formulating the joint rate allocation and power control problem for both SISO and MISO RSMA networks, with a particular focus on practical user-specific rate and SIC constraints. Leveraging a sequential convex approximation (SCA) framework, they transform the inherently nonconvex optimization problem into a series of tractable convex subproblems, ensuring convergence to a locally optimal solution. Their work not only demonstrates the critical impact of explicit SIC constraints on achievable rates but also provides valuable analytical insights into optimal rate-splitting strategies. However, despite these strengths, their iterative SCA-based approach incurs significant computational complexity, limiting its applicability in rapidly changing or large-scale network environments, a limitation that motivates the adoption of DRL techniques in our present study for scalable and adaptive resource allocation. In a complementary direction, Ou et al. [26] investigate resource allocation in multi-user MISO RSMA systems for ultra-reliable and low-latency communications (URLLC) scenarios, where the accuracy of SIC cannot always be guaranteed. Their contribution includes both single-carrier and multi-carrier RSMA schemes, with a rigorous formulation of the effective throughput maximization problem under stringent latency and reliability constraints. To address this challenging nonconvex problem, they develop an iterative solution that combines block coordinate descent and successive convex approximation, along with a low-complexity heuristic tailored for the multi-carrier case.

Building on RSMA's flexibility, Abidrabbu and Arslan [27] proposed a flexible HARQ protocol assisted by RSMA, in which the base station dynamically switches between RSMA, SDMA, and NOMA frame structures and adjusts message splitting and power allocation across retransmissions; their throughput, delay, and energy-efficiency analysis under imperfect CSI feedback shows that the HARQ-RSMA scheme reduces the average number of transmissions and sojourn time compared with conventional HARQ-RSMA baselines, but it does not provide a DRL-based controller for continuous-time beamforming and SIC-gap tuning as in our work. Parallel research into power allocation under SIC constraints can also be found in the context of NOMA. For example, Ali et al. [36] examine dynamic power allocation in downlink multi-cell NOMA systems employing coordinated multi-point (CoMP) transmission. Their study formulates a joint optimization over subcarrier and power allocation, explicitly incorporating both rate and SIC constraints to maximize system throughput and efficiency. They propose centralized convex optimization strategies as well as distributed algorithms to enable scalability in large networks. Although their results demonstrate substantial spectral and energy efficiency gains for CoMP-NOMA, especially for cell-edge users, their mathematical formulation and algorithmic approach are based entirely on convex analysis. In contrast, our work departs from deterministic optimization, advancing a learning-based approach that is better suited to highly dynamic, interference-limited RSMA environments. The roles of these SIC-conscious RSMA and NOMA designs within the broader landscape are also reflected in the "Deterministic RSMA rate/power/QoS optimization," "HARQ and reliability with RSMA," and "Non-RSMA physical-layer optimization" blocks of Tables I and II.

Although previous studies have explored SIC constraints, reliability mechanisms, and learning-based resource allocation in both RSMA and NOMA settings [20]–[27], their integration within RSMA downlink systems, especially in dynamic, large-scale environments, remains underdeveloped. Likewise, existing works on deep learning and DRL-based resource management and long-term network utility optimization [19], [28]–[35] primarily focus on NOMA, multi-RAT heterogeneous networks, cell-free or satellite-assisted architectures, edge inference, aerial computing, or industrial IoT systems rather than RSMA downlink with explicit SIC-gap modeling. The unique advantages of DDPG-based, data-driven resource allocation for RSMA with explicit SIC enforcement and real-time adaptability have not been fully exploited in the literature. This gap directly motivates our investigation, where we propose a unified DRL-based optimization framework for downlink RSMA that treats beamforming vectors, power allocation ratios, and SIC-gap constraints as continuous control actions. Tables I and II provide a two-part technical comparison of these representative RSMA-, NOMA-, and resource-management works, grouping them by problem class (RSMA precoding and DRL-based RSMA control, deterministic RSMA/SIC optimization, HARQ and reliability, DRL and resource control in non-RSMA systems, and non-RSMA physical-layer optimization) and highlighting how our study advances the state of the art in terms of joint RSMA design, explicit SIC-gap enforcement, and DRL-based network control.

III. System Model

In our single-cell MISO downlink network, a multi-antenna base station serves multiple single-antenna users using a one-layer RSMA scheme. Each user's message is split into a common part and a private part, where the common parts are combined into one stream while private streams are sent separately. The transmitted signal is formed by beamforming vectors and power allocation ratios that meet the total power and SIC constraints. The following subsections detail the channel model and the rest of the system framework. Table. III summarizes the notations adopted in this study.

A. Channel Model

In this work, we examine a single cell downlink network in which the BS transmits with a consistent power of P_{tot}. Users, each with a single antenna, are randomly spread inside the BS's coverage radius. Also, we consider that all wireless signals are influenced by both large-scale path loss and small-scale block Rayleigh fading. Besides. the channel between the

TABLE III
LIST OF NOTATIONS

Notation	Description
M	Number of antennas
K	Number of users
$\sigma^2_{n,k}$	Variance of AWGN
B	Bandwidth
θ	Detection threshold of SIC constraints
P_{tot}	Power budget of BS
μ_k	Power allocation ratio for private streams of the k-th User
μ_0	Power allocation ratio for common stream
$\mathbf{w}_k$	Precoding vector for private message of the k-th user
$\mathbf{w}_0$	Precoding vector for common message
C_k	Common rate allocated to the k-th user
Γ_k	Private message for the k-th user
Γ_0	Common message
γ^0_k	SINR of common message decoded at the k-th user
R_0	Attainable rate of common message
R_k	Attainable rate of private message of the k-th user
R^{tot}_k	Attainable rate of the k-th user

BS and the k-th user, $k \in \mathcal{K} = \{1, \ldots, k, \ldots, K\}$ where $K = |\mathcal{K}|$, is represented as $128.1 + 37.6 \log_{10}(d_k)$ dB, where d_k is the distance in kilometers [13].The system runs on a time-slotted basis, with each time slot having identical duration and non-overlapping. Additionally, we assume a block fading channel, where the channel remain stable in a distinct time slot but oscillate separately across different time periods [37].

B. Rate Splitting Multiple Access

We apply the 1-layer RS framework [4], [38], where each message Γ_k planned for the k-th user is divided into a common part Γ^0_k and a private part Γ^p_k. The common parts of all users $\Gamma^0_k, \ldots, \Gamma^0_K$ are merged into a unified common message Γ_0 and encoded into a common stream s_0 applying a codebook accessible to everyone. Each user's private part $\Gamma^p_k, \ldots, \Gamma^p_K$ is individually encoded into private streams $s_1, \ldots, s_K$ which are decodable only by the respective associated users. In this study, we divided the power action into two independent variables, i.e. $\mathbf{p_k} = \sqrt{\mu_k P_{tot}}\mathbf{w_k}, k \in \{0, 1, 2, \ldots, K\}$ such that the transmission power of the BS is fully utilized,

$$||\mathbf{w}_k||^2 = 1, k \in \{0, 1, 2, \ldots, K\} \quad \text{AND} \quad \sum_{i=0}^{K} \mu_i \leq 1 \quad (1)$$

The transmitted signal of the BS is specified as follows:

$$\boldsymbol{x} = \sqrt{\mu_0 P_{tot}}\boldsymbol{w}_0 s_0 + \sum_{k=1}^{K} \sqrt{\mu_k P_{tot}}\boldsymbol{w}_k s_k \quad (2)$$

where $\mathbf{w}_0 \in \mathbb{C}^{M\times 1}$ and $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$ represent the precoding vectors of the common and private messages, respectively. μ_0 and μ_k denote the power share ratios for the common and private messages, respectively.

The perceived signal at the k-th user is

$$\begin{aligned} y_k = {} & \boldsymbol{h}_k^H \sqrt{\mu_0 P_{tot}}\boldsymbol{w}_0 s_0 + \boldsymbol{h}_k^H \sqrt{\mu_k P_{tot}}\boldsymbol{w}_k s_k \\ & + \boldsymbol{h}_k^H \left(\sum_{k' \neq k}^{K} \sqrt{\mu_{k'} P_{tot}}\boldsymbol{w}_{k'} s_{k'} \right) + n_k \end{aligned} \quad (3)$$

where n_k denotes the additive white Gaussian noise (AWGN) following the distribution $\mathcal{CN}\left(0, \sigma^2_{n,k}\right)$ and $\boldsymbol{h}_k \in \mathbb{C}^{M\times 1}$ is the channel gain between the BS and the k-th user. In (3), the first term describes the received common stream, the second term describes the designated private stream, and the last term sums the interference from co-scheduled users. Each k-th user decodes the common stream s_0 into $\bar{\Gamma}_0$ by considering the interference from all private streams as noise. The SINRs of common messages are determined as follows:

$$\gamma^0_k = \frac{\mu_0 P_{tot} |\boldsymbol{h}_k^H \boldsymbol{w}_0|^2}{\sum_{j=1}^{K} \mu_j P_{tot} |\boldsymbol{h}_k^H \boldsymbol{w}_j|^2 + \sigma^2_{n,k}} \quad (4)$$

Leveraging SIC, $\bar{\Gamma}_0$ is then re-encoded, precoded, and deducted from the perceived signal. To guarantee effective decoding for all users, the attainable rate of the common message is computed as follows:

$$R_0 = \min_{k\in\mathcal{K}} \left\{ B \log_2 \left(1 + \gamma^0_k\right) \right\} \quad (5)$$

where B represents the total downlink bandwidth. To successfully achieve the SIC process at the receiver, the transmit power of each user must comply with the following constraint [39]:

$$\mu_0 P_{tot} |\boldsymbol{h}_k^H \boldsymbol{w}_0|^2 - \sum_{j=1}^{K} \mu_j P_{tot} |\boldsymbol{h}_k^H \boldsymbol{w}_j|^2 - \sigma^2_{n,k} \geq \theta, \forall k \in \mathcal{K} \quad (6)$$

where θ represents the minimum disparity between the decoding signal power and the combined undecoded inter-user interference signal power and noise power [40]. Given that R_0 is distributed among all users, we denote C_k to represent the quantity of the rate R^c distributed to the k-th user as follows:

$$\sum_{k=1}^{K} C_k \leq R_0. \quad (7)$$

After removing $\bar{\Gamma}_0$, each user interprets its private stream s_k into $\bar{\Gamma}^p_k$ by considering multi-user interference from the private streams of other co-scheduled users as noise, similar to the principle of space division multiple access (SDMA). Next, the k-th user recovers the original message by deriving $\bar{\Gamma}^0_k$ from $\bar{\Gamma}_0$, and concatenating it with $\bar{\Gamma}^p_k$ to form $\bar{\Gamma}_k$. Given the bandwidth of B, the spontaneous transmission rates for decoding the common and private streams at the k-th user are calculated as follows:

$$R_k = B \log_2 \left(1 + \frac{\mu_k P_{tot} |\boldsymbol{h}_k^H \boldsymbol{w}_k|^2}{\sum_{k'=1, k' \neq k}^{K} \mu_{k'} P_{tot} |\boldsymbol{h}_k^H \boldsymbol{w}_{k'}|^2 + \sigma^2_{n,k}} \right) \quad (8)$$

The obtainable rate of the k-th user is presented by

$$R^{tot}_k = C_k + R_k \quad (9)$$

C. Optimization Problem

This study optimizes beamforming vectors and users' portion of the common rate to maximize the WSR of all users.

$$(\mathcal{P}): \max_{\boldsymbol{w},\boldsymbol{\mu},\boldsymbol{c}} \sum_{k\in\mathcal{K}} \epsilon_k R_k^{tot}, \tag{10}$$

$$\text{s.t.} \quad \mu_0 P_{tot}|\boldsymbol{h}_k^H \boldsymbol{w}_0|^2 - \sum_{j=1}^{K} \mu_j P_{tot}|\boldsymbol{h}_k^H \boldsymbol{w}_j|^2 - \sigma_{n,k}^2 \geq \theta, \forall k \in \mathcal{K}, \tag{11}$$

$$\|\mathbf{w}_i\|^2 = 1, i \in \{0, 1, \ldots, K\}, \tag{12}$$

$$\sum_{k=0}^{K} \mu_k \leq 1, \tag{13}$$

$$C_k \geq \max\left(R_k^{min} - R_k, 0\right), \forall k \in \mathcal{K}, \tag{14}$$

where $\boldsymbol{c} = [C_1, \ldots, C_k, \ldots, C_K] \in \mathbb{R}^{K\times 1}$ denotes the common rate allocation vector; $\boldsymbol{w}$ denotes the beamforming matrix and $\boldsymbol{\mu}$ denotes the power allocation ratio for streams. Constraint (14) ensures that each user's rate exceeds the minimum transmission rate requirement. $\epsilon_k > 0$ are user weights used in the WSR objective and reward shaping. Moreover, the constraints of total power budget, SIC requirements for reliable common stream decoding, and minimum rate guarantees for each user, reflect practical limitations that must be respected in real-world deployments.

In this work, we present a unified framework that leverages DRL to maximize the WSR as in problem $(\mathcal{P})$. Specifically, we employ DDPG method to jointly optimize power allocation and beamforming vectors (denoted as μ and $\boldsymbol{w}$), along side the allocation of common rates to satisfy minimum user rate constraints.

IV. Deep Reinforcement Learning-Resolvable Transformation

The original optimization problem is formulated as a Markov decision process-based (MDP) problem. In an MDP, an agent interplay with its environment in discrete-time steps. During each iteration, the agent observes the current state of the environment, selects an action based on its predetermined strategy, and transitions to a new state dependent on the environment's dynamics and the selected action. The state transitions are determined by a Markovian property, indicating that the future state is only influenced by the present state and the action made, without considering the past history of states and actions. An MDP is detailed as $\mathcal{M} = \{X, A, P, R, \tau\}$, where X is the state space, A denotes the set of plausible actions, P signifies the transition probability matrix, R is the reward function assessing the action with respect to the current observation, and $\tau \in (0, 1)$ is the positive discount factor. We employed DDPG [17] to deal with continuous power allocation variables. From the standpoint of the BS, the following subsections define the MDP components.

A. State and Action Spaces

This study describes the state consisting of channel gains between the BS, weights of users, and its users and the margin rate requirement of each user.

$$X = \left\{h_k^t, \epsilon_k^t, R_{k,min}^t\right\}_{k\in\mathcal{K}}, \tag{15}$$

After observing the current state, the agent outputs a temporary action which is defined as: $A = \{\boldsymbol{w}, \boldsymbol{\mu}\}$ where $\boldsymbol{w}$ includes the imaginary and real parts of the complex beamforming vectors and $\boldsymbol{\mu}$ is the ratio of the power allocation normalized coefficient vector such that $\mu_k, \mu_0 \in [0, 1]$ and $\mu_0 + \sum_{k\in\mathcal{K}} \mu_k \leq 1$. The temporary values of the imaginary and real parts of the beamforming vectors are then refined through Proposition 1 to ensure valid beamforming vectors, i.e. $\text{tr}\left(\mathbf{w}^H\mathbf{w}\right) = 1$. Next, the refined complex beamforming vectors and temporary power allocation $\boldsymbol{\mu}$ are inserted into Proposition 2 to ensure a valid power allocation matrix. The details and proofs of the propositions are clearly explained in the next section.

B. Reward Function

State-specific actions are assessed using the reward function. Assuming that the criteria are met, the reward value is the sum WSR of all users, as follows:

$$r[t] = (1 - \chi(\boldsymbol{a}[t])) \sum_{k\in\mathcal{K}} \epsilon_k R_k^{tot} \tag{16}$$

where $\chi(\boldsymbol{a}[t])$ represents the penalty for action $a[t]$. In particular, the penalty $\chi(\boldsymbol{a}[t])$ is applied, i.e. $\chi(\boldsymbol{a}[t]) = 1$, if the optimization problem $(\mathcal{P})_3$ is infeasible. This indicates that no $\boldsymbol{C}$ can be found that simultaneously satisfies the QoS constraint (14) and the total common rate constraint (7). Otherwise, the immediate reward is the objective value obtained by solving $(\mathcal{P}_1)$.

The objective of the BS is to boost the expected discounted long-term return by determining the optimal power allocation variables and user share of the common rate.

$$\mathcal{R} = \max_{a[t]} \mathbb{E}\left[\sum_{t=1}^{\infty} \tau^t r[t]\right], \tag{17}$$

Algorithm 1: DDPG-WSR-RSMA Algorithm.

```
Initialize MA, MC, TA, and TC networks parameters λ, Ψ, λ' = λ, Ψ' = Ψ;
while t ≤ EP do
    Obtain the current state x[t] and decide the corresponding action a[t] via MA network.;
    Add noise to the outputted action for better exploration. ;
    Perform action refinement procedure. ;
    Find corresponding c regard to p by solving (P_3). ;
    Perform the concatenated action and obtain immediate reward according to (16). ;
    Obtain the subsequent state x[t + 1];
    if t ≥ E then
        Discard the most outdated transition in the replay buffer and add the most recent transition ; Randomly select a E-sized mini-batch of transitions ;
        Update MA network according to (20);
        Update MC network by reducing the loss (18);
        Update TA network according to (21);
        Update TC network according to (22);
    else
        Add the most recent transition;
    t++
```

C. DDPG Algorithm

The learning agent comprises four components: the main actor (MA), target actor (TA), main critic (MC), and target critic (TC) networks. These networks are employed as non-linear function approximators. At the beginning, the TA and TC networks are clones of the MA and MC networks, and they are periodically upgraded during training. According to a policy θ parameterized by λ, the MA network deterministically derives a specific action $a[t]$ based on the given state $x[t]$, i.e. $a[t] = \theta\left(x[t]|\lambda\right)$. We denote θ' as the policy of the TA network parameterized by λ'. The process of the embedded DDPG framework is described as in Algorithm. 1.

The DDPG-WSR-RSMA algorithm operates in a sequence of episodes, as detailed in Algorithm 1. At each episode, the main actor network generates candidate beamforming vectors and power allocation ratios based on the observed state. To ensure the feasibility of these actions with respect to system constraints, the outputs are refined by applying Propositions 1 and 2, yielding valid beamforming matrices and power ratios. Next, the optimal allocation of the common rate ensuring all users' minimum rate requirements is computed and appended to form the final action. This action is executed in the environment, which in turn provides the resulting state and reward, allowing the experience to be stored in a replay buffer. During learning, random samples from this buffer are used to update the main and critic networks of the DDPG framework, following the actor-critic paradigm with soft target updates and experience replay.

We maintain a replay buffer to accumulate transition $\langle x[t], a[t], r[t], x[t+1]\rangle$ for training. In particular, a mini-batch of E transitions is randomly extracted from the replay buffer to disrupt any potential link among the sampled data. To prevent needless confusion, we denote e as the recorded time index of the transition used for training during the t-th episode. In response to the e-th transition, the agent's MA network generates the appropriate action based on the stored state based on current parameterized policy. The TA network of the agent outputs the subsequent action regarding the stored subsequent state $x[e+1]$. The MC network is updated by reducing the loss.

$$\mathcal{L} = \frac{1}{E}\sum_{e}\left(y[e] - Q\left(x[e], a[e]\middle|\Psi\right)\right)^2 \tag{18}$$

where $y[e]$ is the target value for the e-th transition, which is the amount of the discounted state-action value collected from the TC network and the recorded recent reward,

$$y[e] = r[e] + \tau Q'\left(x[e+1], \theta'\left(x[e+1]|\lambda'\right)\middle|\Psi'\right) \tag{19}$$

The MA network revises the policy with the assistance of the MC network in accordance with the updating gradient of the policy.

$$\nabla_\lambda J \approx \frac{1}{E}\sum_{e}\nabla_a Q\left(x, a|\Psi\right)|_{x=x[e],a=\theta(x[e])}\nabla_\lambda\theta\left(x|\lambda\right)|_{x[e]} \tag{20}$$

The weights of TA and TC are then modified through gradual monitoring in accordance with the soft-update rule [17],

$$\lambda' \leftarrow \rho_\lambda\lambda + \left(1-\rho_\lambda\right)\lambda', \tag{21}$$

$$\Psi' \leftarrow \rho_\Psi\Psi + \left(1-\rho_\Psi\right)\Psi', \tag{22}$$

where $\rho_\lambda \ll 1$ and $\rho_\Psi \ll 1$ are pre-determined step sizes. The DDPG algorithm also recommends the use of Ornstein-Uhlenbeck process to introduce exploration noise to obtain continuous action as output from the MA network, which enables the learning behavior escape undesirable overfitting.

V. Proposed DDPG-RSMA-SIC algorithm

A. Action Refinement Procedure

By employing the DDPG algorithm, the actor-network can choose the appropriate actions $\boldsymbol{w}$ and $\boldsymbol{\mu}$. However, these actions may not always satisfy the problem's constraints, which violates environmental requirements. To address this issue, we propose a post-actor processing step. This step ensures that all constraints are satisfied before the system solves the problem of finding the optimal common rate allocations.

The process begins by verifying the user precoding matrix against the specified requirement. If the constraints are not satisfied, the precoding matrix is modified using the following proposition.

Proposition 1. *Let* $\mathbf{W}' \triangleq [\mathbf{w}'_1, \mathbf{w}'_2, \ldots, \mathbf{w}'_M]^T \in \mathbb{C}^{M\times(K+1)}$ *represent the* ***refined*** *precoding matrix of the BS that satisfies constraint* (12) *, i.e.* $||\mathbf{w_k}|| = 1, k \in \{0, 1, \ldots, K\}$. *Its element at row* m *and column* k, $w'_{mk}, m \in \mathcal{M} \triangleq \{1, 2, \ldots, M\}, k \in \mathcal{K} \triangleq \{0, 1, 2, \ldots, K\}$, *is measured as follows:*

$$w'_{mk} = \frac{w_{mk}}{\parallel \mathbf{w}_k \parallel}, \tag{23}$$

where $w_{m,k}$ denotes the element in row m and column k of $\mathbf{W}$.

Proof: According to constraint (23), the precoding constraint is satisfied using the refined precoding matrix $\mathbf{W}'$, i.e., $||\mathbf{w}'_k|| = 1, k \in \{0, 1, \ldots, K\}$. The vector-norm can be calculated as

$$||\mathbf{w}'_k||^2 = \left(\sqrt{\sum_{m=1}^{M}|w'_{mk}|^2}\right)^2 = \sum_{m=1}^{M}\frac{|w_{mk}|^2}{||\mathbf{w_k}||^2} = \frac{\sum_{m=1}^{M}|w_{mk}|^2}{||\mathbf{w_k}||^2} = 1. \tag{24}$$

This completes the proof. ■

Suppose that we have $\boldsymbol{\mu}$ as the output result from the actor network. The action will be invalid if either the SIC requirements or the power budget are violated. To address this problem, we compute a new set of values for all power ratios using the following proposition:

Proposition 2. *Let* $\boldsymbol{\mu}' \triangleq [\mu'_0, \mu'_1, \ldots, \mu'_K]^T \in \mathbb{C}^{(K+1)\times 1}$ *represent the power ratio vector that satisfies the SIC constraints of each user and the power budget of the BS. The calculations of its elements are as follows:*

$$\mu'_0 = \frac{\hat{\mu}_0}{\hat{\mu}_0 + \sum_{k=1}^{K}\mu_k}(1 - u(\mu)) + \hat{\mu}_0 u(\mu), \tag{25}$$

and

$$\mu_k' = \left[\frac{\mu_k}{\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k} - \frac{\mu_k(\theta + \sigma_{n,k}^2)}{\sum_{j=1}^{K} \mu_j P_{tot} \left| \mathbf{h}_k^H \mathbf{w}_j' \right|^2} \right] (1 - u(\mu)) + \mu_k u(\mu), \tag{26}$$

where

$$\hat{\mu}_0 = \begin{cases} \mu_0, & \text{if (11) is true,} \\ \max_k \left(\frac{\theta + \sigma_{n,k}^2 + \sum_{j=1}^{K} \mu_j P_{tot} \left| \mathbf{h}_k^H \mathbf{w}_j' \right|^2}{P_{tot} \left| \mathbf{h}_k^H \mathbf{w}_0' \right|^2} \right), & \text{otherwise,} \end{cases} \tag{27}$$

and $u(\mu)$ denotes the unit function that equals 1 when constraint (13) *is satisfied with $\hat{\mu}_0$, i.e., $\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k \leq 1$, and equals 0 in otherwise.*

Proof: Please see Appendix A. ∎

By applying Propositions 1 and 2, the designed precoding matrix and power allocation ratios satisfy all the related constraints in problem $(\mathcal{P}_1)$, which ensures the accuracy of the actions.

After this step, we can reliably calculate all rates including R_0 and R_k. However, the optimal value of $\boldsymbol{c}$ cannot be straightforwardly computed. Instead, its closed-form expressions can be resulted with the calculated value of R_0 and R_k as in following Proposition.

Proposition 3. *Consider the linear program*

$$\begin{aligned} \max_{\{\boldsymbol{c}\}} \quad & \sum_{k \in \mathcal{K}} \epsilon_k C_k, \\ \text{s.t.} \quad & C_k \geq \max\{R_k^{\min} - R_k, 0\}, \quad \forall\, k \in \mathcal{K}, \\ & \sum_{k \in \mathcal{K}} C_k \leq R_0, \end{aligned} \tag{28}$$

where ϵ_k, R_k, $R_k^{\min}$, and R_0 are nonnegative. Assume that

$$\sum_{k \in \mathcal{K}} \max\{R_k^{\min} - R_k, 0\} \leq R_0. \tag{29}$$

Define $\epsilon^* \triangleq \max_{k \in \mathcal{K}} \epsilon_k$ and $\mathcal{K}^* \triangleq \{k \in \mathcal{K} : \epsilon_k = \epsilon^*\}$. Then an optimal solution $\{C_k^*\}_{k \in \mathcal{K}}$ is given by

$$C_k^* = \begin{cases} \max\{R_k^{\min} - R_k, 0\}, & \text{if } \epsilon_k < \epsilon^*, \\ \max\{R_k^{\min} - R_k, 0\} + \delta_k, & \text{if } k \in \mathcal{K}^*, \end{cases} \tag{30}$$

with $\delta_k \geq 0$ for $k \in \mathcal{K}^*$ satisfying $\sum_{k \in \mathcal{K}^*} \delta_k = R_0 - \sum_{k \in \mathcal{K}} \max\{R_k^{\min} - R_k, 0\}$.

Proof: For each $k \in \mathcal{K}$, the constraint $C_k \geq \max\{R_k^{\min} - R_k, 0\}$ forces a minimum allocation of $\max\{R_k^{\min} - R_k, 0\}$. Express C_k as

$$C_k = \max\{R_k^{\min} - R_k, 0\} + d_k, \quad d_k \geq 0. \tag{31}$$

Then the total budget constraint becomes

$$\sum_{k \in \mathcal{K}} \max\{R_k^{\min} - R_k, 0\} + \sum_{k \in \mathcal{K}} d_k \leq R_0, \tag{32}$$

or equivalently,

$$\sum_{k \in \mathcal{K}} d_k \leq R_0 - \sum_{k \in \mathcal{K}} \max\{R_k^{\min} - R_k, 0\}. \tag{33}$$

Since the objective function can be written as

$$\sum_{k \in \mathcal{K}} \epsilon_k \left[\max\{R_k^{\min} - R_k, 0\} + d_k \right], \tag{34}$$

and the term $\sum_{k \in \mathcal{K}} \epsilon_k \max\{R_k^{\min} - R_k, 0\}$ is constant, maximizing the objective reduces to maximizing $\sum_{k \in \mathcal{K}} \epsilon_k d_k$ subject to $\sum_{k \in \mathcal{K}} d_k \leq R_0 - \sum_{k \in \mathcal{K}} \max\{R_k^{\min} - R_k, 0\}$ and $d_k \geq 0$. The optimal strategy is to assign the entire surplus to the indices with the highest weight ϵ^*. Hence, for any k with $\epsilon_k < \epsilon^*$, set $d_k^* = 0$, and for $k \in \mathcal{K}^*$ choose $d_k^* = \delta_k$ with

$$\sum_{k \in \mathcal{K}^*} \delta_k = R_0 - \sum_{k \in \mathcal{K}} \max\{R_k^{\min} - R_k, 0\}.$$

This establishes the stated optimal solution. ∎

B. DDPG-WSR-RSMA Algorithm

As described in Algorithm 1, the training algorithm proceeds in EP episodes. During each episode, the DDPG algorithm generates precoding vectors and the power allocation ratio vector through the actor network. Subsequently, the action procedure is applied by recalculating their values using Propositions 1 and 2 to ensure satisfaction of the problem constraints. The state of the environment is updated and the interaction process iterates. At each update step, a mini-batch of samples is chosen by chance from the replay buffer to train the neural networks using the DDPG algorithm.

The full procedure of DDPG-WSR-RSMA Algorithm can be described as in Fig. 1. Beside traditional learning procedure of DDPG, we introduce the Procedures following the proved Propositions in previous subsection. The temporal value of action outputted from the neural networks of DDPG are processed in proper manners and yield a refined action. The common rate allocation is then solved and contributes to the final action. This action would be executed by the environment, assessing the current quality of solution of DDPG-WSR-RSMA. The experience is then stored cumulatively at the experience buffer. The random samples from experience buffer triggers the learning process of original DDPG framework.

The complete workflow consists of state observation, temporary action generation, two-stage refinement for constraint satisfaction, common rate allocation, action execution, experience storage, and neural network updates driven by replayed experiences. This integrated process enables the agent to learn high-quality, constraint-satisfying resource allocation strategies for RSMA in a continuous, data-driven manner.

C. Complexity Analysis of DDPG-WSR-RSMA Algorithm

In this subsection, we investigate the complexity of DDPG-WSR-RSMA in execution stage after finishing the training procedure. In execution stage, the optimal MA is utilized to decide the action interacting with the environment. The back-propagation forwards are skipped. Therefore, the computational complexity is determined according to the feed-forward passes starting from the input layer to output layer. The numbers of neurons in input and output layers are respectively the number of entries of state and action spaces. As a result, the time complexity of the proposed DDPG-WSR-RSMA

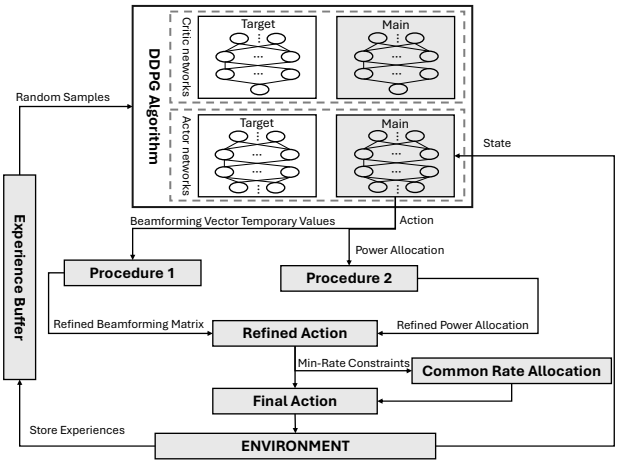


Fig. 1. Overview of Framework of DDPG-WSR-RSMA Algorithm.

algorithm sums up the matrix multiplication of layers [41], which is computed as follows:

$$O(MA) = O\Big((2 \times K \times (M+1)) \times n_{a_1} + n_{a_1} \times n_{a_2} + n_{a_2} \times ((2M+1) \times (K+1) \times M) \Big) \tag{35}$$

where n_{a_1} and n_{a_2} are the numbers of hidden neurons of first and second layer of MA, respectively.

In addition to the matrix multiplications associated with the neural network feed-forward pass, our algorithm incorporates two refinement steps (Propositions 1 and 2) to ensure that the actions satisfy all constraints. The normalization of each beamforming vector (Proposition 1) requires $\mathcal{O}((K+1) \times M)$ operations, and the power allocation refinement (Proposition 2), which involves verifying the power budget and enforcing SIC constraints, incurs a computational overhead of $\mathcal{O}\left(K^2\right)$. Since the number of users K and the number of antennas M are typically moderate, these additional computations are negligible compared to the overall $\mathcal{O}(MA)$ complexity of the feed-forward passes. Consequently, the refinement steps enhance the reliability of the DRL output without imposing a significant computational burden, thereby preserving the efficiency of the DDPG-WSR-RSMA algorithm.

VI. Performance Evaluation

A. Simulation Scenario

The simulation scenario considers a single-cell MISO downlink network in which a multi-antenna base station (BS) serves multiple single-antenna users randomly distributed within its coverage area. Unless otherwise specified, the BS is equipped with $M = 4$ antennas and communicates with $K = 10$ users. Large-scale path loss and small-scale block Rayleigh fading jointly model the channel, while the system operates with a total bandwidth of $B = 1$ MHz and a transmit power budget $P_{tot} = 20$ dBm. The noise power is set to $\sigma^2_{n,k} = -84$ dBm and the SIC detection threshold to $\theta = -104$ dBm, following standard RSMA evaluation settings [13]. Key simulation and DDPG hyperparameters are summarized in Table IV. All powers in (3)-(8) and (11) are in linear watts; dBm values in Table IV are converted via $P[\mathrm{W}] = 10^{(P_{\mathrm{dBm}}-30)/10}$.

TABLE IV
Environmental Parameters and DDPG Hyperparameters

Parameter	Value
P_{tot}	20 dBm
$\sigma^2_{n,k}$	-84 dBm
θ	-104 dBm
B	1 MHz
DDPG Hyperparameters	
Critic Network Architecture	2 hidden layers (1024, 512 neurons)
Actor Network Architecture	2 hidden layers (512, 256 neurons)
Replay Buffer Size	10^5
Batch Size	16
Noise Decay Schedule	1 to 10^{-5} over 1000 episodes

All algorithms are executed on a PC with a 2.1 GHz Intel i7-12700 CPU and 32 GB RAM. The DDPG agent employs an actor network with two hidden layers (512 and 256 neurons) and a critic network with two hidden layers (1024 and 512 neurons). A replay buffer of size 10^5 is maintained, and training uses mini-batches of 16 experiences. Exploration noise decays from 1 to 10^{-5} over 1000 episodes to balance exploration and exploitation. Unless stated otherwise, each performance result is averaged over 10,000 independent channel realizations to ensure statistical significance.

B. Comparison Schemes

To thoroughly evaluate our proposed DDPG-WSR-RSMA algorithm, we compare it against several baseline schemes that reflect both alternative system designs and different learning frameworks. The benchmark schemes are detailed as follows:

- **DDPG-WSR-RSMA without SIC constraints (DDPG-NoSIC):** This variant uses the same DDPG framework as our proposed scheme, but it omits the refinement provided by Proposition 2. In other words, while the actor network still outputs temporary beamforming vectors and power allocation ratios, these values are not adjusted to enforce the SIC constraints before the environment executes the action. This benchmark isolates the effect of incorporating explicit SIC constraints on system performance.
- **DDPG-WSR-SDMA:** This baseline applies the DDPG algorithm to a conventional space division multiple access (SDMA) system. Here, the RSMA strategy is replaced by a standard SDMA approach, meaning that each user is served without message splitting. This comparison highlights the performance benefits that RSMA provides over traditional SDMA when optimized via DRL.

To further demonstrate the superiority of our continuous-action DDPG approach, we also consider benchmarks based on branching dueling Q networks (BDQN) [42], which are action-value methods designed for discrete action spaces. Because BDQN is not inherently suited for continuous actions, we discretize the power allocation space using a uniform quantization mechanism [20]. In particular, let Φ denote the

number of discretization levels and Ξ the number of actions; while a standard DDQN would require exploring an action space of size Φ^{Ξ}, BDQN reduces this to $\Phi \times \Xi$, making the problem more tractable. Within the BDQN framework, we introduce two variants:

- **BDQN-WSR-RSMA:** This scheme employs both Proposition 1 and Proposition 2 for refining the temporary outputs, similar to our proposed method. However, the underlying learning structure and update mechanisms differ due to the BDQN framework.
- **BDQN-WSR-RSMA-noSIC:** This variant is identical to BDQN-WSR-RSMA except that it does not apply the SIC-related refinement (i.e., Proposition 2). This serves to further isolate the impact of explicitly enforcing SIC constraints.

For consistency and fairness, Proposition 1 is applied across all schemes to ensure that the beamforming vectors are properly normalized. In addition, a softmax function is used on the temporary power allocation outputs to guarantee that the total power budget is fully utilized. Simulation results are reported as averages over 10,000 channel realizations. This detailed breakdown of benchmark schemes allows for a clear comparison of the benefits offered by our proposed DRL-based RSMA approach relative to both alternative system designs and different learning frameworks.

C. Non-RL Baseline: WMMSE-SCA-SIC and WMMSE-SCA-noSIC

We benchmark the proposed DRL against a deterministic alternating-optimization (AO) solver using the rate-WMMSE equivalence for 1-layer RSMA [5] under an explicit SIC-gap constraint. Let the transmit vectors be $\mathbf{p}_i \triangleq \sqrt{\mu_i P_{\mathrm{tot}}}\,\mathbf{w}_i \in \mathbb{C}^M$ for $i \in \{0,1,\ldots,K\}$. The total-power constraint is $\sum_{i=0}^{K} \|\mathbf{p}_i\|_2^2 \le P_{\mathrm{tot}}$ (unit-norm $\mathbf{w}_i$ absorbed). Introduce scalar equalizers $g_{c,k}, g_k \in \mathbb{C}$ and weights $u_{c,k}, u_k > 0$. With $\mathbf{H}_k \triangleq \mathbf{h}_k\mathbf{h}_k^H$, the MSEs at user k are

$$e_{c,k} = |g_{c,k}|^2\Big(|\mathbf{h}_k^H\mathbf{p}_0|^2 + \textstyle\sum_{j=1}^{K}|\mathbf{h}_k^H\mathbf{p}_j|^2 + \sigma_{n,k}^2\Big) - 2\Re\{g_{c,k}\mathbf{h}_k^H\mathbf{p}_0\} + 1, \tag{36}$$

$$e_k = |g_k|^2\Big(\textstyle\sum_{j\neq k}|\mathbf{h}_k^H\mathbf{p}_j|^2 + \sigma_{n,k}^2\Big) - 2\Re\{g_k\mathbf{h}_k^H\mathbf{p}_k\} + 1. \tag{37}$$

Define augmented WMSEs $\xi_{c,k} = u_{c,k}e_{c,k} - \log u_{c,k} - 1$ and $\xi_k = u_k e_k - \log u_k - 1$. Using $\min_{u_{c,k},g_{c,k}} \xi_{c,k} = 1 - R_{c,k}$ and $\min_{u_k,g_k} \xi_k = 1 - R_k$, maximizing $\sum_k \epsilon_k(C_k + R_k)$ is equivalent to minimizing $\sum_k(\xi_{c,k}+\xi_k) - \sum_k \epsilon_k C_k$ subject to (i) $\sum_{i=0}^{K}\|\mathbf{p}_i\|_2^2 \le P_{\mathrm{tot}}$, (ii) the common-rate epigraph $R_0 \le R_{c,k}\ \forall k$ and $\sum_k C_k \le R_0$, and (iii) the explicit SIC-gap

$$|\mathbf{h}_k^H\mathbf{p}_0|^2 - \textstyle\sum_{j=1}^{K}|\mathbf{h}_k^H\mathbf{p}_j|^2 - \sigma_{n,k}^2 \;\ge\; \theta, \quad \forall k \tag{38}$$

For given $\{\mathbf{p}_i\}$,

$$g_{c,k}^{\mathrm{mmse}} = \frac{\mathbf{h}_k^H\mathbf{p}_0}{\sum_{i=0}^{K}|\mathbf{h}_k^H\mathbf{p}_i|^2 + \sigma_{n,k}^2}, \qquad u_{c,k}^{\mathrm{mmse}} = e_{c,k}^{-1}, \tag{39}$$

$$g_k^{\mathrm{mmse}} = \frac{\mathbf{h}_k^H\mathbf{p}_k}{\sum_{j\neq k}|\mathbf{h}_k^H\mathbf{p}_j|^2 + \sigma_{n,k}^2}, \qquad u_k^{\mathrm{mmse}} = e_k^{-1}. \tag{40}$$

TABLE V
ASYMPTOTIC COMPLEXITY OF COMPARED SCHEMES

Scheme	Dominant time complexity (per decision)
DDPG-WSR-RSMA-SIC (proposed)	$\mathcal{O}\big(C_{\mathrm{A}} + KM + K^2\big)$
DDPG-WSR-RSMA (no SIC)	$\mathcal{O}\big(C_{\mathrm{A}} + KM\big)$
DDPG-WSR-SDMA	$\mathcal{O}\big(C_{\mathrm{A}} + KM\big)$
BDQN-WSR-RSMA-SIC	$\mathcal{O}\big(C_{\mathrm{B}} + \Xi\Phi + KM + K^2\big)$
BDQN-WSR-RSMA (no SIC)	$\mathcal{O}\big(C_{\mathrm{B}} + \Xi\Phi + KM\big)$
WMMSE-SCA-SIC	$\mathcal{O}\big(I_{\mathrm{AO}}\, I_{\mathrm{SCA}}\, ((K+1)M)^3\big)$
WMMSE-SCA-noSIC	$\mathcal{O}\big(I_{\mathrm{AO}}\, ((K+1)M)^3\big)$

For fixed (u, g), update $\{\mathbf{p}_i\}, R_0, \{C_k\}$ by solving a convex subproblem. The only nonconvex part is (38). At AO iterate n, lower-bound $|\mathbf{h}_k^H\mathbf{p}_0|^2 = \mathbf{p}_0^H\mathbf{H}_k\mathbf{p}_0$ at $\mathbf{p}_0^{(n)}$ via

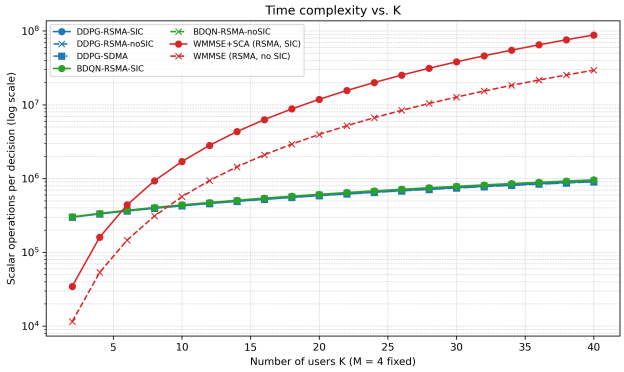
$$\ell_k^{(n)}(\mathbf{p}_0) \;\triangleq\; 2\,\Re\big\{(\mathbf{p}_0^{(n)})^H\mathbf{H}_k\mathbf{p}_0\big\} - |\mathbf{h}_k^H\mathbf{p}_0^{(n)}|^2, \tag{41}$$

and enforce the convex surrogate $\ell_k^{(n)}(\mathbf{p}_0) - \sum_{j=1}^{K}|\mathbf{h}_k^H\mathbf{p}_j|^2 - \sigma_{n,k}^2 \ge \theta$. Together with the affine common-rate epigraph and $\sum_i \|\mathbf{p}_i\|_2^2 \le P_{\mathrm{tot}}$, this yields a convex QCQP solvable by interior-point methods [43]. A few SCA inner iterations per AO step ensure monotone WSR ascent. After each precoder update set $R_0 = \min_k R_{c,k}$ and solve $\max_{\{C_k\}} \sum_k \epsilon_k C_k$ s.t. $\sum_k C_k \le R_0$ and $C_k \ge \max\{R_k^{\min} - R_k, 0\}$. The optimal policy meets all lower bounds and assigns the leftover entirely to the user(s) with the largest ϵ_k. Dropping (38) (or setting $\theta = 0$) removes the SCA inner loop; the precoder update reduces to a single convex QCQP per AO iteration [5]. This WMMSE-SCA baseline (and its WMMSE-SCA-noSIC variant) attains strong local optima per snapshot, at the cost of solving a convex program (with SCA when SIC is enforced), whereas our DRL performs a single forward pass at test time. Generalized power-iteration (GPI) is a related non-RL alternative for RSMA precoding under different objectives [12].

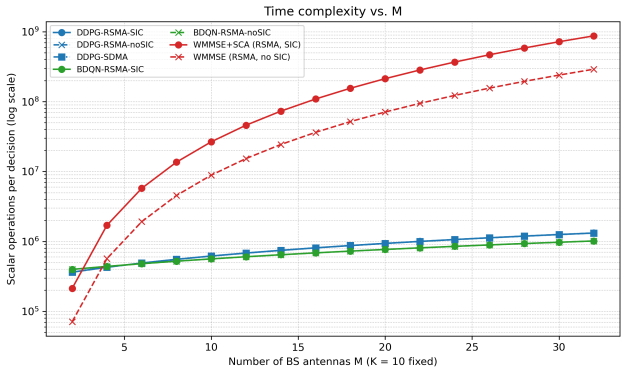
D. Time Complexity and Inference Cost

Table VI-D and Fig. 2 jointly show three key insights. Here, "per decision" means the test-time cost to produce one action for a single channel snapshot. We use K for the number of users and M for BS antennas; $((K+1)M)$ is the decision dimension counting the common stream plus K private streams. For DRL, C_{A} denotes the DDPG actor forward-pass cost and C_{B} the BDQN trunk+branch forward cost; BDQN discretizes Ξ action branches (here $\Xi{=}K{-}1$ due to the simplex), each with Φ discrete levels. For deterministic solvers, I_{AO} is the number of AO outer iterations and I_{SCA} the number of SCA inner iterations used only when SIC is enforced.

- The deterministic non-RL baselines (WMMSE-SCA) scale cubically in the decision dimension $\approx ((K{+}1)M)$, since each AO step solves a QCQP via interior-point Newton factorization; with explicit SIC, additional SCA inner loops (counted by I_{SCA}) further multiply the cost, leading to steep (orders-of-magnitude) growth as K or M increase, which is impractical for real-time control [43].



(a) *Ops* vs. number of users K (M=4).



(b) *Ops* vs. number of antennas M (K=10).

Fig. 2. Explicit operation counts per decision (log scale). Same color denotes the same family (DDPG, BDQN, WMMSE). Markers distinguish SIC (•) and noSIC (×); □ denotes SDMA.

- DDPG inference adds only one MLP forward pass (cost C_{A}) plus lightweight post-processing, $\mathcal{O}(KM)$, for beamformer normalization and $\mathcal{O}(K^2)$ for the closed-form SIC refinement, so operation counts increase smoothly with K/M.
- BDQN scales linearly with the discretization budget through branching ($\Xi\Phi$ logits and an argmax per branch, on top of C_{B}), avoiding exponential blow-up from joint grids; in practice its cost lies between DDPG and WMMSE [42].

Overall, these results make the deployment implication clear: deterministic solvers can achieve strong stationary points per snapshot, but their per-decision complexity becomes prohibitive in large systems, whereas RL inference offers predictable and lightweight scalability.

E. Simulation Results

1) Convergence of DDPG-WSR-RSMA Algorithm: We begin by evaluating the convergence of the DDPG-WSR-RSMA algorithm using different hyperparameter settings for the Adam optimizer. Figure 3 demonstrates the robustness of the DDPG-WSR-RSMA-SIC framework's convergence across various of neural network learning rates. Overall, the DDPG-WSR-RSMA-SIC framework exhibit robust convergence toward stationary values. However, careful learning rate selection remains crucial to balance convergence speed and

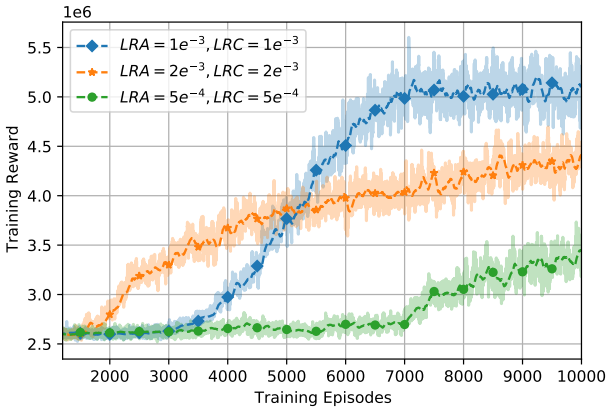


Fig. 3. Convergence of DDPG-WSR-RSMA-SIC Framework.

the risk of overshooting local minima. In this figure, the hyperparameter setting $LRA = 1e^{-3}, LRC = 1e^{-3}$ achieves the highest reward, and its training reward begins to increase earlier (at 1500 episode) compared with the other learning rate configurations. Consequently, we used this hyperparameter setting for all DDPG-based comparison models.

2) Varying Number of Users and Antennas: Figure 4 depicts a decreasing trend in WSR as the number of users rises. This is intuitive for multiple access schemes, where a big number of users leads to intensified interference levels for all users. In SDMA, each user undergoes interference from all other users simultaneously. This results in significant WSR losses as individual user rates diminish because of rising interference from other users. RSMA schemes generally outperform SDMA schemes because of their flexible power allocation capabilities. The performance gap widens significantly with 10 users. Our proposed scheme DDPG-RSMA-SIC and DDPG-RSMA-noSIC still exhibit clear gains over DDPG-SDMA and BDQN-based counterparts; additionally, the two deterministic baselines, WMMSE-SCA-SIC and WMMSE-SCA-noSIC, now appear at the top of the figure with a significant lead. This is consistent with the rate-WMMSE equivalence and SCA literature, where alternating optimization around MMSE equalizers and convexified subproblems often attains strong stationary points for WSR in interference-limited regimes.

Figure 5 clearly demonstrates how the WSR increases as the number of antennas at the BS grows. Generally, the multiuser gain of RSMA is more pronounced than that of SDMA. A higher number of antennas enhances the spatial diversity of communication channels, granting the learning agent more flexibility and decision-making power in action selection. It also implies that in addition to the power allocation ratios, the transmit beamforming vectors also contributes significantly to the increases in WSR. In line with Fig. 4, the two non-RL baselines remain the best across M, while the margins over RL schemes are moderate when array gain is abundant and become more visible in smaller arrays. This trend agrees with deterministic AO designs based on weighted-MMSE updates and convex inner approximations, which are known to be

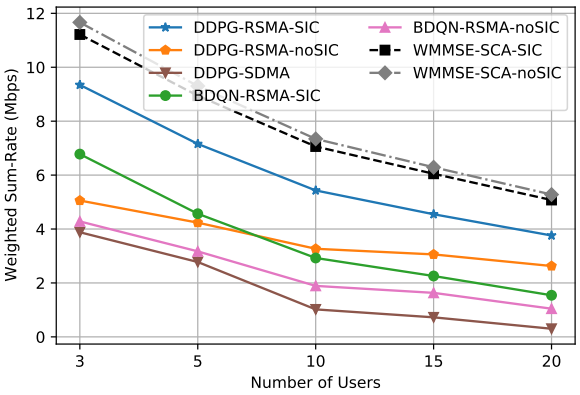


Fig. 4. WSR versus varying number of users

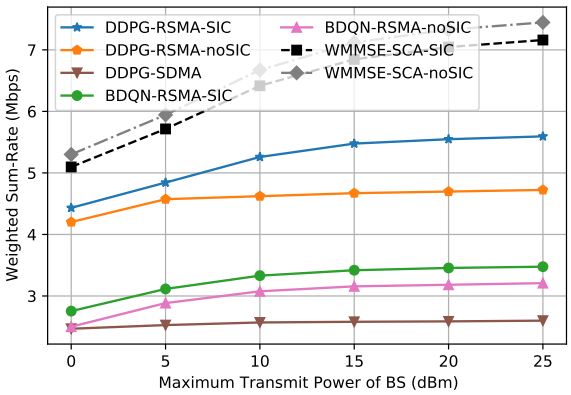


Fig. 6. WSR versus maximum transmit power of the BS

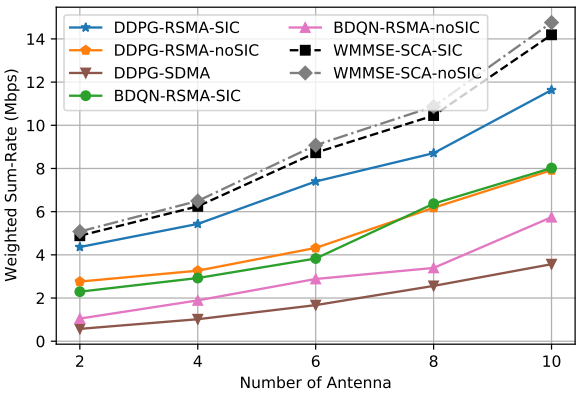


Fig. 5. WSR versus varying number of antennas

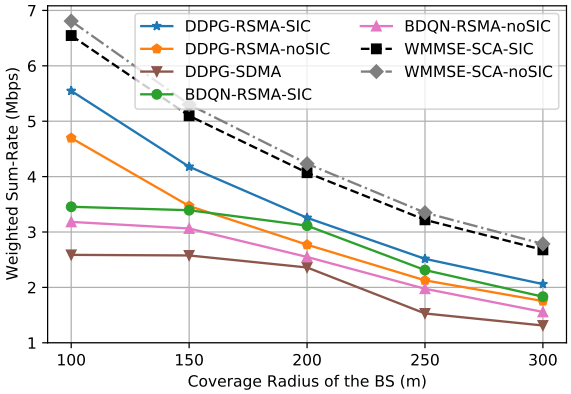


Fig. 7. WSR versus coverage radius of BS

especially effective when the feasible set is well captured by successive convex surrogates.

3) Power Budget, Channel Strengths, and Detection Threshold: Figure 6 shows a linear increase with the logarithmic maximum transmit power of the BS. Notably, the performance gap between the DDPG-based and BDQN-based schemes widens as P_{tot} increases. However, the WSR eventually plateaus as the maximum transmit power of the BS increases. The deterministic WMMSE-SCA-SIC and WMMSE-SCA-noSIC curves stay consistently above all RL curves across P_{tot}, with a noticeable margin in the mid-to-high power region before all methods saturate. This behavior reflects the ability of AO/WMMSE updates to directly optimize a tight surrogate of WSR at each step while respecting constraints via convexified inner loops

The results in Figure 7 confirm the effectiveness of combining the RSMA and SIC constraints. The DDPG-based strategies outperform their BDQN-based counterparts in terms of throughput across varying distances to the BS. In addition, both non-RL baselines preserve a clearly higher WSR level throughout the cell, and the advantage becomes more evident as the coverage radius increases and channels weaken, again matching prior evidence that WMMSE-driven AO with SCA handles strong inter-user coupling efficiently.

Figure 8 illustrates the reduction in WSR as the SIC detection threshold rises. A higher SIC detection threshold requires the BS to give more power ratio to the common message. The marginal reductions in WSR indicate that RSMA is advisable for detecting high SIC thresholds. In the case of SDMA, the WSR remains constant regardless of any changes in value of θ simply because the SIC makes no contribution to the calculation of rates in the SDMA scheme.

VII. Conclusions

This study explores the potential of employing DRL to enhance RSMA-based networks. We propose a practical DDPG-based approach to address the joint problem of power allocation and user-specific common rate allocation. This approach maximizes the WSR of all users while adhering to constraints such as the BS's limited transmission power and each user's minimum rate requirement. The final action combines the power allocation coefficient vectors received from DDPG agent's actor network with the corresponding optimal shared rates of users, determined by solving an LP problem. Simulations demonstrate that the DDPG-RSMA-SIC framework operate reliably and achieves a higher WSR in most

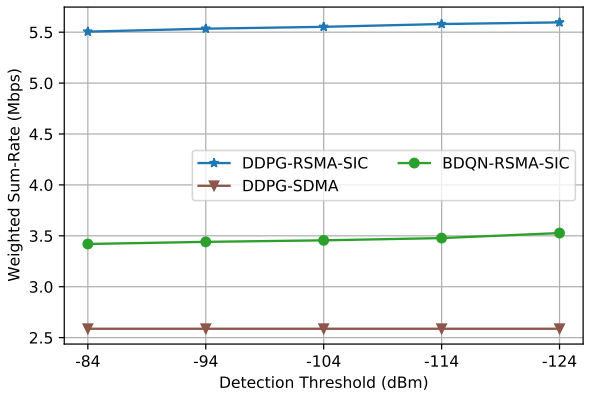


Fig. 8. WSR versus the detection threshold θ

cases than other DRL benchmarks; additionally, deterministic WMMSE-based baselines (with/without explicit SIC via SCA) consistently attain the top WSR across scenarios, in line with the established rate-WMMSE and SCA literature.

Appendix A
Proof of Proposition 2

In Proposition 2, we normalize the power allocation vector to ensure the fulfillment of the SIC and power budget constraints. Based on Equation (11), for any k-th user, the minimum value of μ_0 that complies with the SIC constraint can be calculated as

$$\mu_0 = \frac{\theta + \sigma_{n,k}^2 + \sum_{j=1}^{K} \mu_k P_{tot} \left|\mathbf{h}_k^H \mathbf{w}_j'\right|^2}{P_{tot} \left|\mathbf{h}_k^H \mathbf{w}_0'\right|^2}. \tag{42}$$

Considering K users, the minimum value of μ_0 is selected as

$$\mu_0 = \max_k \left(\frac{\theta + \sigma_{n,k}^2 + \sum_{j=1}^{K} \mu_k P_{tot} \left|\mathbf{h}_k^H \mathbf{w}_j'\right|^2}{P_{tot} \left|\mathbf{h}_k^H \mathbf{w}_0'\right|^2} \right). \tag{43}$$

Therefore, if μ_0 violates the constraint (11), we force it to the minimum value expressed in Equation (43). Consequently, the new value of μ_0 that satisfies the SIC constraint in Equation (11) can be calculated as

$$\hat{\mu}_0 = \begin{cases} \mu_0, & \text{if (11) is true,} \\ \max_k \left(\frac{\theta + \sigma_{n,k}^2 + \sum_{j=1}^{K} \mu_j P_{tot} \left|\mathbf{h}_k^H \mathbf{w}_j'\right|^2}{P_{tot} \left|\mathbf{h}_k^H \mathbf{w}_0'\right|^2} \right), & \text{otherwise.} \end{cases} \tag{44}$$

With the new value of μ_0, we then prove that the normalized values of power allocation variables defined in Equation (25) and Equation (26) satisfy the power budget constraint in (13), i.e., $\mu_0' + \sum_{k=1}^{K} \mu_k' \leq 1$. To achive this, we consider two cases:

- $u(\mu) = 1$ ($\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k \leq 1$): in this case, the normalized power allocation variables are unchanged and hold the constraint, i.e.,

$$\begin{aligned} \mu_0' &= \hat{\mu}_0, \\ \mu_k' &= \mu_k, k \in \mathcal{K}. \end{aligned} \tag{45}$$

- $u(\mu) = 0$ ($\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k > 1$): in this case, the sum of the power allocation variables is calculated as

$$\begin{aligned} \mu_0' + \sum_{k=1}^{K} \mu_k' &= \frac{\hat{\mu}_0}{\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k} + \frac{\sum_{k=1}^{K} \mu_k}{\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k} \\ &\quad - \sum_{k=1}^{K} \frac{\mu_k (\theta + \sigma_{n,k}^2)}{\sum_{j=1}^{K} \mu_k P_{tot} \left|\mathbf{h}_k^H \mathbf{w}_j'\right|^2} \\ &= 1 - \sum_{k=1}^{K} \frac{\theta + \sigma_{n,k}^2}{\sum_{j=1}^{K} P_{tot} \left|\mathbf{h}_k^H \mathbf{w}_j'\right|^2} \leq 1, \end{aligned} \tag{46}$$

which proves the power budget constraint.

Next, we prove that the normalized power allocation vector $\boldsymbol{\mu'}$ satisfies the SIC constraint in Equation (11). To achieve this, we first calculate the left side of Equation (11), expressed as

$$\begin{aligned} &\mu_0' P_{tot} |\mathbf{h}_k^H \mathbf{w}_0'|^2 - \sum_{j=1}^{K} \mu_j' P_{tot} |\mathbf{h}_k^H \mathbf{w}_j'|^2 - \sigma_{n,k}^2 \\ &= \frac{\hat{\mu}_0}{\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k} P_{tot} |\mathbf{h}_k^H \mathbf{w}_0'|^2 \\ &\quad - \sum_{j=1}^{K} \frac{\mu_j}{\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k} P_{tot} |\mathbf{h}_k^H \mathbf{w}_j'|^2 \\ &\quad + \sum_{j=1}^{K} \frac{\mu_j (\theta + \sigma_{n,k}^2)}{\sum_{j=1}^{K} \mu_j P_{tot} \left|\mathbf{h}_k^H \mathbf{w}_j'\right|^2} P_{tot} |\mathbf{h}_k^H \mathbf{w}_j'|^2 - \sigma_{n,k}^2 \\ &= \frac{\hat{\mu}_0}{\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k} P_{tot} |\mathbf{h}_k^H \mathbf{w}_0'|^2 \\ &\quad - \sum_{j=1}^{K} \frac{\mu_j}{\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k} P_{tot} |\mathbf{h}_k^H \mathbf{w}_j'|^2 \\ &\quad + \frac{(\theta + \sigma_{n,k}^2)}{\sum_{j=1}^{K} \mu_j P_{tot} \left|\mathbf{h}_k^H \mathbf{w}_j'\right|^2} \sum_{j=1}^{K} \mu_j P_{tot} \left|\mathbf{h}_k^H \mathbf{w}_j'\right|^2 - \sigma_{n,k}^2 \\ &= \frac{\hat{\mu}_0}{\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k} P_{tot} |\mathbf{h}_k^H \mathbf{w}_0'|^2 \\ &\quad - \sum_{j=1}^{K} \frac{\mu_j}{\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k} P_{tot} |\mathbf{h}_k^H \mathbf{w}_j'|^2 + \theta + \sigma_{n,k}^2 - \sigma_{n,k}^2 \\ &= \frac{\hat{\mu}_0}{\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k} P_{tot} |\mathbf{h}_k^H \mathbf{w}_0'|^2 \\ &\quad - \sum_{j=1}^{K} \frac{\mu_j}{\hat{\mu}_0 + \sum_{k=1}^{K} \mu_k} P_{tot} |\mathbf{h}_k^H \mathbf{w}_j'|^2 + \theta \\ &\overset{(a)}{\geq} \theta, \end{aligned} \tag{47}$$

where (a) occurs due to $\hat{\mu}_0 P_{tot} |\mathbf{h}_k^H \mathbf{w}_0'|^2 \geq \sum_{j=1}^{K} \mu_j P_{tot} |\mathbf{h}_k^H \mathbf{w}_j'|^2$, which is obtained from Equation (44). Consequently, the SIC constraint in Equation (11) is hold. Therefore, the normalized power allocation vector $\boldsymbol{\mu'}$ satisfies the SIC and power budget constraints, which proves Proposition 2.

REFERENCES

[1] J. Li, S. Dang, M. Wen, Q. Li, Y. Chen, Y. Huang, and W. Shang, "Index modulation multiple access for 6g communications: Principles, applications, and challenges," *IEEE Network*, vol. 37, no. 1, pp. 52–60, 2023.

[2] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for iot: A survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1805–1838, 2020.

[3] X. Chen, D. W. K. Ng, W. Yu, E. G. Larsson, N. Al-Dhahir, and R. Schober, "Massive access for 5g and beyond," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 3, pp. 615–637, 2020.

[4] 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*, 2022.

[5] Y. Mao, B. Clerckx, and V. O. Li, "Rate-splitting multiple access for downlink communication systems: Bridging, generalizing, and outperforming sdma and noma," *EURASIP journal on wireless communications and networking*, vol. 2018, pp. 1–54, 2018.

[6] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser miso systems with partial csit: A rate-splitting approach," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4847–4861, 2016.

[7] Y. Mao, E. Piovano, and B. Clerckx, "Rate-splitting multiple access for overloaded cellular internet of things," *IEEE Transactions on Communications*, vol. 69, no. 7, pp. 4504–4519, 2021.

[8] 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, feb 2022. [Online]. Available: http://dx.doi.org/10.1109/tcomm.2021.3138437

[9] Z. Li, C. Ye, Y. Cui, S. Yang, and S. Shamai, "Rate splitting for multi-antenna downlink: Precoder design and practical implementation," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1910–1924, 2020.

[10] A. A. Ahmad, H. Dahrouj, A. Chaaban, A. Sezgin, and M.-S. Alouini, "Interference mitigation via rate-splitting and common message decoding in cloud radio access networks," *IEEE Access*, vol. 7, pp. 80 350–80 365, 2019.

[11] S. K. Singh, K. Agrawal, K. Singh, and C.-P. Li, "Ergodic capacity and placement optimization for rsma-enabled uav-assisted communication," *IEEE Systems Journal*, vol. 17, no. 2, pp. 2586–2589, jun 2023. [Online]. Available: http://dx.doi.org/10.1109/jsyst.2022.3220249

[12] J. Park, J. Choi, N. Lee, W. Shin, and H. V. Poor, "Rate-splitting multiple access for downlink mimo: A generalized power iteration approach," *IEEE Transactions on Wireless Communications*, vol. 22, no. 3, pp. 1588–1603, mar 2023. [Online]. Available: http://dx.doi.org/10.1109/TWC.2022.3205480

[13] G. Zhou, Y. Mao, and B. Clerckx, "Rate-splitting multiple access for multi-antenna downlink communication systems: Spectral and energy efficiency tradeoff," *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 4816–4828, jul 2022. [Online]. Available: http://dx.doi.org/10.1109/twc.2021.3133433

[14] T. P. Truong, T. M. Tuyen Nguyen, T.-V. Nguyen, N.-N. Dao, and S. Cho, "Rsma for uplink mimo systems: Drl-based achievable system sum rate maximization," in *2023 IEEE Globecom Workshops (GC Wkshps)*. Institute of Electrical and Electronics Engineers (IEEE), dec 2023, pp. 878–883. [Online]. Available: http://dx.doi.org/10.1109/gcwkshps58843.2023.10464584

[15] Y. Hua, Y. Fu, and Q. Zhu, "Latency minimization for advanced rsma-enabled wireless caching networks," *IEEE Wireless Communications Letters*, vol. 12, no. 8, pp. 1329–1333, aug 2023. [Online]. Available: http://dx.doi.org/10.1109/lwc.2023.3273225

[16] W. Jaafar, S. Naser, S. Muhaidat, P. C. Sofotasios, and H. Yanikomeroglu, "Multiple access in aerial networks: From orthogonal and non-orthogonal to rate-splitting," *IEEE Open Journal of Vehicular Technology*, vol. 1, pp. 372–392, 2020. [Online]. Available: http://dx.doi.org/10.1109/ojvt.2020.3032844

[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," *arXiv preprint arXiv:1509.02971*, 2015.

[18] J. Park, J. Choi, N. Lee, W. Shin, and H. V. Poor, "Rate-splitting multiple access for downlink MIMO: A generalized power iteration approach," *IEEE Transactions on Wireless Communications*, vol. 22, no. 3, pp. 1588–1603, mar 2023. [Online]. Available: http://dx.doi.org/10.1109/TWC.2022.3205480

[19] Z. Wang, R. Ma, H. Shi, Z. Cai, L. Lin, and H. Guan, "Deep convolutional linear precoder neural network for rate splitting strategy of aerial computing networks," *IEEE Transactions on Network Science and Engineering*, vol. 11, no. 6, pp. 5228–5243, 2024.

[20] N. Q. Hieu, D. T. Hoang, D. Niyato, and D. I. Kim, "Optimal power allocation for rate splitting communications with deep reinforcement learning," *IEEE Wireless Communications Letters*, vol. 10, no. 12, pp. 2820–2823, dec 2021. [Online]. Available: http://dx.doi.org/10.1109/lwc.2021.3118441

[21] J. Huang, Y. Yang, J. Lee, D. He, and Y. Li, "Deep reinforcement learning-based resource allocation for rsma in leo satellite-terrestrial networks," *IEEE Transactions on Communications*, vol. 72, no. 3, pp. 1341–1354, mar 2024. [Online]. Available: http://dx.doi.org/10.1109/tcomm.2023.3331021

[22] Y. Chen, M. Sun, X. Xu, S. Han, H. Gao, X. Cheng, P. Cui, and P. Zhang, "Snc-enabled delay analysis and drl-based power control for rsma systems," *IEEE Transactions on Network Science and Engineering*, 2025.

[23] Z. Yang, M. Chen, W. Saad, W. Xu, and M. Shikh-Bahaei, "Sum-rate maximization of uplink rate splitting multiple access (rsma) communication," *IEEE Transactions on Mobile Computing*, vol. 21, no. 7, pp. 2596–2609, oct 2020. [Online]. Available: http://dx.doi.org/10.1109/tmc.2020.3037374

[24] E. Sadeghabadi and S. Blostein, "Rsma precoding design based on interference nulling and sum rate upper bound," *IEEE Transactions on Communications*, vol. 71, no. 7, pp. 4091–4104, jul 2023. [Online]. Available: http://dx.doi.org/10.1109/tcomm.2023.3277033

[25] Z. Yang, M. Chen, W. Saad, and M. Shikh-Bahaei, "Optimization of rate allocation and power control for rate splitting multiple access (RSMA)," *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 5988–6002, sep 2021. [Online]. Available: http://dx.doi.org/10.1109/TCOMM.2021.3091133

[26] X. Ou, X. Xie, H. Lu, and H. Yang, "Resource allocation in mu-miso rate-splitting multiple access with sic errors for urllc services," *IEEE Transactions on Communications*, vol. 71, no. 1, pp. 229–243, jan 2023. [Online]. Available: http://dx.doi.org/10.1109/tcomm.2022.3224388

[27] S. Abidrabbu and H. Arslan, "A new flexible harq assisted by rate splitting multiple access," *IEEE Transactions on Network Science and Engineering*, 2025.

[28] H. Zhang, H. Zhang, K. Long, and G. K. Karagiannidis, "Deep learning based radio resource management in NOMA networks: User association, subchannel and power allocation," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 2406–2415, oct 2020. [Online]. Available: http://dx.doi.org/10.1109/TNSE.2020.3004333

[29] A. Alwarafy, B. S. Ciftler, M. Abdallah, M. Hamdi, and N. Al-Dhahir, "Hierarchical multi-agent DRL-based framework for joint multi-RAT assignment and dynamic resource allocation in next-generation hetnets," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 4, pp. 2481–2494, jul 2022. [Online]. Available: http://dx.doi.org/10.1109/TNSE.2022.3164648

[30] Y. Liu, A. Tang, and X. Wang, "Joint incentive and resource allocation design for user-provided network under 5G integrated access and backhaul networks," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 673–685, apr 2020. [Online]. Available: http://dx.doi.org/10.1109/TNSE.2019.2910867

[31] J. Huang, J. Wu, Y. Wu, and J. Wu, "Joint request offloading and resource allocation for long-term utility optimization in collaborative edge inference with time-coupled resources," *IEEE Transactions on Network Science and Engineering*, vol. 12, no. 4, pp. 2622–2639, jul 2025. [Online]. Available: http://dx.doi.org/10.1109/tnse.2025.3551148

[32] A. Lotfolahi and H.-W. Ferng, "Drl-based resource allocation in noma-aided industrial iot towards energy productivity maximization," *IEEE Transactions on Network Science and Engineering*, 2025.

[33] Z. Chen, K. Cai, J. Ye, Q. Li, and X. Ge, "Resource block-granularity precoding optimization and compression for cell-free mobile networks," *IEEE Transactions on Network Science and Engineering*, vol. 12, no. 5, pp. 3641–3655, sep 2025. [Online]. Available: http://dx.doi.org/10.1109/tnse.2025.3563370

[34] Y. Lee, C. Song, D. Lee, W. Noh, and S. Cho, "User-centric clustering and beamforming design for satellite-assisted cell-free networks," *IEEE Transactions on Network Science and Engineering*, vol. 12, no. 5, pp. 3467–3479, sep 2025. [Online]. Available: http://dx.doi.org/10.1109/tnse.2025.3561505

[35] Y. He, F. Huang, D. Wang, and R. Zhang, "Outage probability analysis of MISO-NOMA downlink communications in UAV-assisted agri-IoT with SWIPT and TAS enhancement," *IEEE Transactions on Network*

Science and Engineering, vol. 12, no. 3, pp. 2151–2164, may 2025. [Online]. Available: http://dx.doi.org/10.1109/TNSE.2025.3545148

[36] M. S. Ali, E. Hossain, A. Al-Dweik, and D. I. Kim, "Downlink power allocation for comp-noma in multi-cell networks," *IEEE Transactions on Communications*, vol. 66, no. 9, pp. 3982–3998, sep 2018. [Online]. Available: http://dx.doi.org/10.1109/tcomm.2018.2831206

[37] M. Z. Hassan, M. J. Hossain, J. Cheng, and V. C. M. Leung, "Energy-spectrum efficient content distribution in fog-ran using rate-splitting, common message decoding, and 3d-resource matching," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 4929–4946, aug 2021. [Online]. Available: http://dx.doi.org/10.1109/TWC.2021.3063283

[38] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, "Rate splitting for mimo wireless networks: a promising phy-layer strategy for lte evolution," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 98–105, may 2016. [Online]. Available: http://dx.doi.org/10.1109/mcom.2016.7470942

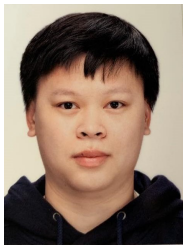
[39] Z. Yang, M. Chen, W. Saad, and M. Shikh-Bahaei, "Optimization of rate allocation and power control for rate splitting multiple access (rsma)," *IEEE Transactions on Communications*, vol. 69, no. 9, pp. 5988–6002, sep 2021. [Online]. Available: http://dx.doi.org/10.1109/TCOMM.2021.3091133

[40] X. Liu, J. Feng, F. Li, and V. C. M. Leung, "Downlink energy efficiency maximization for rsma-uav assisted communications," *IEEE Wireless Communications Letters*, vol. 13, no. 1, pp. 98–102, jan 2024. [Online]. Available: http://dx.doi.org/10.1109/lwc.2023.3321765

[41] T. P. Truong, N.-N. Dao, S. Cho *et al.*, "Flyreflect: Joint flying irs trajectory and phase shift design using deep reinforcement learning," *IEEE Internet of Things Journal*, vol. 10, no. 5, pp. 4605–4620, 2022.

[42] A. Tavakoli, F. Pardo, and P. Kormushev, "Action branching architectures for deep reinforcement learning," in *Proceedings of the aaai conference on artificial intelligence*, vol. 32, no. 1, 2018.

[43] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.



Anh Tien Tran received the B.S. degree in electronics and telecommunications from the Da Nang University of Science and Technology, Da Nang, Vietnam, in 2018. He is currently working toward the Ph.D. degree in computer science and engineering with Chung-Ang University, Seoul, South Korea. His research interests include wireless network communication, video streaming, and machine learning.

Thanh Phung Truong (Graduate Student Member, IEEE) received the B.E. degree in electronics and telecommunications from Ho Chi Minh City University of Technology, Vietnam National University Ho Chi Minh City, Vietnam, in 2018, and the M.E. degree in computer science and engineering from Chung-Ang University, South Korea, in 2022, where he is currently pursuing the Ph.D. degree in computer science and engineering. From 2019 to 2020, he was a FPGA engineer with VinSmart Research and Manufacture Joint Stock Company, Vingroup, Vietnam. From 2022 to 2023, he was a Data Scientist with Viettel Networks, Viettel Group, Vietnam. He has been with the Ultra-Intelligent Computing/Communication Laboratory, Chung-Ang University, South Korea, since 2020. His research interests include wireless communication, applied machine learning, and network architectures.

Dongwook Won received a B.S. and M.S. degree in Electronics and Communications Engineering from Kwangwoon University, Seoul, Korea in 2020, 2022. He is currently pursuing a Ph.D. in the School of Computer Science and Engineering at Chung-Ang University, Seoul, South Korea. His research interests include Semantic Communications, 5G/6G Networks, and Satellite Communication.

Nhu-Ngoc Dao is an Assistant Professor at the Department of Computer Science and Engineering, Sejong University, Seoul, Korea. He received his M.S. and Ph.D. degrees in computer science at the School of Computer Science and Engineering, Chung-Ang University, Seoul, Korea, in 2016 and 2019, respectively. He received the B.S. degree in electronics and telecommunications from the Posts and Telecommunications Institute of Technology, Hanoi, Viet Nam, in 2009. Prior to joining Sejong University, he was a Visiting Researcher with the University of Newcastle, Callaghan, NSW, Australia, in 2019 and a Postdoc Researcher with the Institute of Computer Science, University of Bern, Switzerland, from 2019 to 2020. He was a Visiting Professor at Chung-Ang University from 2023 to 2024. He has been selected to receive the 2025 Springer Nature Editorial Contribution Award. He is currently an Editor of the ICT Express, Scientific Reports, and PLOS ONE journals. Dr. Dao is a Senior Member of IEEE and a Member of ACM.

Sungrae Cho is a professor with the school of computer sciences and engineering, Chung-Ang University (CAU), Seoul. Prior to joining CAU, he was an assistant professor with the department of computer sciences, Georgia Southern University, Statesboro, GA, USA, from 2003 to 2006, and a senior member of technical staff with the Samsung Advanced Institute of Technology (SAIT), Kiheung, South Korea, in 2003. From 1994 to 1996, he was a research staff member with electronics and telecommunications research institute (ETRI), Daejeon, South Korea. From 2012 to 2013, he held a visiting professorship with the national institute of standards and technology (NIST), Gaithersburg, MD, USA. He received the B.S. and M.S. degrees in electronics engineering from Korea University, Seoul, South Korea, in 1992 and 1994, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2002. He has been an associate member of National Academy of Engineering of Korea (NAEK) since 2025 and a KICS fellow since 2021. He received numerous awards including the Haedong Best Researcher of the Year in Telecommunications in 2022, the Medal from Korean Prime Mister in 2025, and the Award from Korean Ministry of Science and ICT in 2021.

His current research interests include wireless networking, network intelligence, and network optimization. He has been an editor-in-chief (EIC) of ICT Express (Elsevier) since 2024, a subject editor of IET Electronics Letter since 2018, an executive editor of Wiley Transactions on Emerging Telecommunications Technologies since 2023, and was an area editor of Ad Hoc Networks Journal (Elsevier) from 2012 to 2017. He has served numerous international conferences as a general chair, TPC chair, or an organizing committee chair, such as IEEE ICC, IEEE SECON, IEEE ICCE, ICOIN, ICTC, ICUFN, APCC, TridentCom, and the IEEE MASS.