# Enhancing Energy Efficiency in STAR-RIS-Assisted Massive MIMO-RSMA Networks

Ridho Hendra Yoga Perdana*, Thong-Nhat Tran*, Wonseok Lee*, Young Jeon*, and Taejoon Kim*
*School of Information and Communication Engineering, Chungbuk National University, Cheongju 28644, South Korea
Emails: *perdanarhy@ieee.org, *trantnhat@gmail.com, *ws_lee@chungbuk.ac.kr,
*jeony9672@chungbuk.ac.kr, *ktjcc@chungbuk.ac.kr

***Abstract*—This paper investigates energy-efficiency maximization (EEM) for downlink simultaneously transmitting and reflecting (STAR-RIS)–assisted massive MIMO (mMIMO) networks employing rate-splitting multiple access (RSMA). The energy efficiency formulation problem is a non-convex mixed-integer program, driven by (i) an upper bound on the allocable common SE, (ii) the transmit-power budget at the base station (BS), and (iii) the discrete phase shifts at the STAR-RIS, which together preclude direct optimal solutions. To address this, we first relax the problem and decouple it into two subproblems—phase-shift design and beamforming—that are solved in an alternating fashion. The phase-shift subproblem is handled via a bisection search, while the beamforming subproblem is tackled using an inner-approximation (IA) procedure that converts the original non-convex constraints into tractable surrogates. For real-time operation, we design a deep-learning (DL) framework that infers optimal phase shifts and precoding matrices under varying system parameters. Simulations demonstrate that the proposed scheme yields notable EE gains, and we further quantify the impact of the number of STAR-RIS and STAR-RIS elements on the overall system performance.**

***Index Terms*—Discrete phase shift, energy efficiency, mMIMO, non-convex optimization, RSMA, STAR-RIS.**

## I. Introduction

High spectral efficiency (SE) and energy efficiency (EE) are critical performance objectives for next-generation wireless networks [1], [2]. To boost SE, manage interference, and sustain high capacity in dense deployments, massive multiple-input multiple-output (mMIMO), non-orthogonal multiple access (NOMA), and rate-splitting multiple access (RSMA) have been widely explored and adopted [3]–[5]. Yet, in the presence of obstacles between the base station (BS) and users, received-signal quality can degrade markedly due to blockage and non-line-of-sight conditions. A promising remedy is the use of simultaneously transmitting and reflecting reconfigurable intelligent surfaces (STAR-RIS) [6], [7]. In contrast to conventional RIS, which only reflects, STAR-RIS supports both transmission and reflection, extending coverage and improving link quality for users on both sides of the surface [8]. This dual functionality is particularly effective for mitigating blockage and enhancing reception in complex propagation scenarios. In parallel, RSMA can further increase EE by managing multiuser interference through splitting data into a common stream and user-specific private streams [9]. However, as mMIMO arrays and STAR-RIS apertures scale up, the associated resource-allocation and configuration problems become increasingly complex, introducing new challenges for practical system design and optimization. To address this challenge, a deep learning (DL)-based framework has been proposed to obtain an optimal solution in mMIMO–RSMA networks [10].

Prior work has typically examined RSMA and STAR-RIS in mMIMO systems separately. However, beyond-5G networks call for their integrated deployment, which—while promising substantial gains—introduces pronounced challenges in interference suppression, resource allocation, and overall system complexity. In particular, jointly optimizing the EE of user equipment becomes difficult due to the large numbers of BS antennas and user antennas, and STAR-RIS elements. Integrating DL with their integrated deployment system can boost system performance while keeping complexity and latency low. However, the impact of jointly using DL and STAR-RIS to address the EE problem in mMIMO–RSMA networks remains underexplored. Motivated by this gap, we study a downlink multiuser in a STAR-RIS-assisted mMIMO–RSMA network to improve EE performance. The main contributions of the paper are as follows:

- A downlink multi-user mMIMO-RSMA network is considered, where multiple STAR-RISs are deployed to improve the received quality signal of users. The energy efficiency maximization (EEM) problem is formulated under constraints on each user's minimum data rate, the total transmit-power budget, and the STAR-RIS phase shifts. The resulting formulation is highly challenging, as it leads to a non-linear mixed-integer program.
- To tackle the EEM problem, we decompose it into two subproblems—phase-shift design and beamforming—and solve them sequentially: the phase-shift subproblem via a bisection search and the beamforming subproblem via a low-complexity IA-based iterative algorithm.
- To support real-time operation, we design a DL framework with a CNN that predicts the optimal solution produced by the proposed algorithm.
- Simulation results demonstrate EE gains for the proposed system. The DL-based framework matches the optimality of the conventional EEM solver while achieving substantially shorter run time. We also conduct a thorough sensitivity analysis of key parameters.

*Notation*: $a$, $\mathbf{a}$, $\mathbf{A}$ denote the scalar, vectors, and matrices, respectively. $\mathrm{diag}(\mathbf{A})$, $|.|$, $(.)^*$ is the diagonal matrix, absolute

of values, complex conjugate, respectively. $\Re$ and $\mathbb{C}$ represent real part and complex numbers, respectively.

## II. System Model

We consider a downlink multi-user mMIMO-RSMA system assisted by a set of AAV-attached STAR-RIS $\mathcal{R} = \{\mathrm{R}_r | r = 1, ..., R\}$, as illustrated in Fig. A flying-based station (BS)

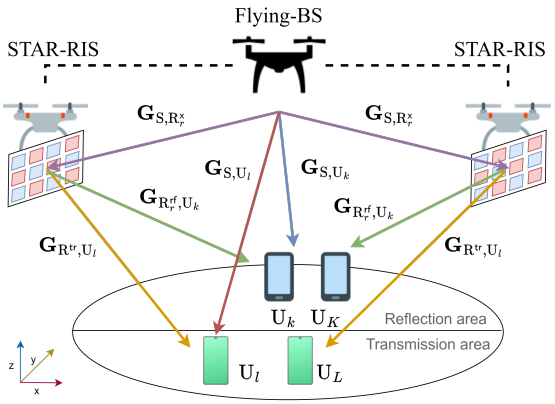



Fig. 1. The proposed system model STAR-RIS-assisted mMIMO-RSMA networks.

simultaneously serves a set of users $\mathcal{K} = \{\mathrm{U}_k | k = 1, ..., K\}$ and users $\mathcal{L} = \{\mathrm{U}_l | l = 1, ..., L\}$, and the STAR-RISs assist by enhancing their received signals. The base station, $\mathrm{U}_K$, and $\mathrm{U}_L$ are equipped with $M_{\mathrm{S}} > 1$, $M_{\mathrm{K}} > 1$, and $M_{\mathrm{L}} > 1$ antennas, respectively. We consider a mode-switching (MS) STAR-RIS with $M_{\mathrm{R}} > 1$ passive elements split into a transmit subset $M_R^{\mathsf{tr}}$ and a reflect subset $M_R^{\mathsf{rf}}$, such that $M_{\mathrm{R}}^{\mathsf{tr}} + M_{\mathrm{R}}^{\mathsf{rf}} = M_{\mathrm{R}}$ [8]. $\mathrm{U}_k$ and $\mathrm{U}_l$ are randomly distributed across reflection and transmission areas, respectively. The STAR-RIS is randomly deployed at the perimeter of the reflection area. We denote the relative position vector from $\mathsf{x} \in \{\mathrm{BS}, \mathrm{R}_r\}$ to $\mathsf{y} \in \{\mathrm{R}_r, \mathrm{U}_k, \mathrm{U}_l\}$ in 3D Cartesian coordinates by $\mathbf{c}_{\mathsf{xy}} = [x_{\mathsf{xy}}, y_{\mathsf{xy}}, z_{\mathsf{xy}}]$. The BS is assumed to possess perfect channel state information (CSI) for every link [10].

### A. STAR-RIS-assisted mMIMO-RSMA

Under single-layer rate-splitting (RS), the BS sends a common stream for $\mathrm{U}_k$ and $\mathrm{U}_l$ and private streams per user. Let $\mathbf{s}^{\mathsf{c}} \in \mathbb{C}^{L\times 1}$ denote the common stream and $\mathbf{s}_k^{\mathsf{p}}, \mathbf{s}_l^{\mathsf{p}} \in \mathbb{C}^{L\times 1}$ the private streams, with $1 < L \leq \min\{M_{\mathrm{S}}, M_{\mathrm{K}}, M_{\mathrm{L}}\}$. With $\mathbf{s} = [\mathbf{s}^{\mathsf{c}}, \mathbf{s}_k^{\mathsf{p}}, \mathbf{s}_l^{\mathsf{p}}]^{\mathsf{T}}$ and linear precoding uses $\mathbf{W} \in \mathbb{C}^{M_{\mathrm{S}}\times L} = [\mathbf{W}^{\mathsf{c}}, \mathbf{W}_k^{\mathsf{p}}, \mathbf{W}_l^{\mathsf{p}}]$. Then the transmitted signal can be expressed as

$$\mathbf{X} = \mathbf{W}^{\mathsf{c}}\mathbf{s}^{\mathsf{c}} + \sum_{k\in\mathcal{K}} \mathbf{W}_k^{\mathsf{p}}\mathbf{s}_k^{\mathsf{p}} + \sum_{l\in\mathcal{L}} \mathbf{W}_l^{\mathsf{p}}\mathbf{s}_l^{\mathsf{p}}. \tag{1}$$

For compactness, we set $\mathbf{W}_1 = \mathbf{W}^{\mathsf{c}}$, $\mathbf{W}_2 = [\mathbf{W}_k^{\mathsf{p}}]_{k\in\mathcal{K}}$, and $\mathbf{W}_3 = [\mathbf{W}_l^{\mathsf{p}}]_{l\in\mathcal{L}}$.

Let $\mathbf{G}_{\mathrm{S,R}_r^{\mathsf{tr}}}$ and $\mathbf{G}_{\mathrm{S,R}_r^{\mathsf{rf}}}$ denote the $\mathrm{BS}-\mathrm{R}_r$ channel matrices associated with the reflecting and transmitting element sets, respectively. While $\mathbf{G}_{\mathrm{S,U}_k}$, and $\mathbf{G}_{\mathrm{S,U}_l}$ denote the $\mathrm{BS}-\mathrm{U}_k$ and $\mathrm{BS}-\mathrm{U}_k$ channel matrices, respectively. Likewise, $\mathbf{G}_{\mathrm{R}_r^{\mathsf{tr}},\mathrm{U}_l}$ and $\mathbf{G}_{\mathrm{R}_r^{\mathsf{rf}},\mathrm{U}_k}$ denote the $\mathrm{R}_r^{\mathsf{tr}} - \mathrm{U}_l$ (reflection mode) and $\mathrm{R}_r^{\mathsf{rf}} - \mathrm{U}_l$ (transmission mode) channel matrices. We model $\mathbf{G} = \sqrt{\rho_g}[\tilde{\mathbf{g}}_1, ..., \tilde{\mathbf{g}}_i]$, where $\mathbf{G} \in \{\mathbf{G}_{\mathrm{S,R}_r^{\mathsf{tr}}}, \mathbf{G}_{\mathrm{S,R}_r^{\mathsf{rf}}}, \mathbf{G}_{\mathrm{S,U}_k}, \mathbf{G}_{\mathrm{S,U}_l}\}$, where $\rho_g$ is the large-scale fading and each $\tilde{\mathbf{g}}_i$ is a small-scale fading. The large-scale fading $\rho_g$ can be modeled as $\rho_g = \mathcal{A}(d_{\mathsf{xy}}/d_0)^{-\sigma_{\mathsf{PL}}}$, where $d_0$, $\mathcal{A}$, $d_{\mathsf{xy}}$, and $\sigma_{\mathsf{PL}}$ denote the reference distance, the measured pathloss at $d_0$, the distance between $\mathsf{x}$ and $\mathsf{y}$ (in meter), and the pathloss exponent, respectively. Furthermore, $d_{\mathsf{xy}}$ can be calculated as $d_{\mathsf{xy}} = \sqrt{(x_{\mathsf{x}} - x_{\mathsf{y}})^2 + (y_{\mathsf{x}} - y_{\mathsf{y}})^2 + (z_{\mathsf{x}} - z_{\mathsf{y}})^2}$. While the small-scale fading has i.i.d. $\mathcal{CN} \sim (0, 1)$ elements. Thus, the equivalent channel from $\mathrm{BS} - \mathrm{U}_k$ denoted by $\hat{\mathbf{G}}_{\mathrm{S,U}_k}(\mathbf{\Phi}) \in \mathbb{C}^{M_{\mathrm{K}}\times M_{\mathrm{S}}}$ and from $\mathrm{BS}-\mathrm{U}_l$ denoted by $\hat{\mathbf{G}}_{\mathrm{S,U}_l}(\mathbf{\Psi}) \in \mathbb{C}^{M_{\mathrm{L}}\times M_{\mathrm{S}}}$ can be expressed, respectively, as

$$\hat{\mathbf{G}}_{\mathrm{S,U}_k}(\mathbf{\Phi}) = \mathbf{G}_{\mathrm{S,U}_k} + \mathbf{G}_{\mathrm{S,R}_r^{\mathsf{rf}}}\mathbf{\Phi}_r\mathbf{G}_{\mathrm{R}_r^{\mathsf{rf}},\mathrm{U}_k}, \tag{2}$$

$$\hat{\mathbf{G}}_{\mathrm{S,U}_l}(\mathbf{\Psi}) = \mathbf{G}_{\mathrm{S,U}_l} + \mathbf{G}_{\mathrm{S,R}_r^{\mathsf{tr}}}\mathbf{\Psi}_r\mathbf{G}_{\mathrm{R}_r^{\mathsf{tr}},\mathrm{U}_l}, \tag{3}$$

where $\mathbf{\Phi}_r$ and $\mathbf{\Psi}_r$ denote the reflection-mode and transmission-modes phase shift of $\mathrm{R}_r$, respectively, which can be determine, respectively, by

$$\mathbf{\Phi} = \mathrm{diag}\left(\lambda_1^{\mathsf{rf}} e^{j\phi_1}, ..., \lambda_{M_{\mathrm{R}}^{\mathsf{rf}}}^{\mathsf{rf}} e^{j\phi_{M_R^{\mathsf{rf}}}}\right), \tag{4}$$

$$\mathbf{\Psi} = \mathrm{diag}\left(\lambda_1^{\mathsf{tr}} e^{j\psi_1}, ..., \lambda_{M_{\mathrm{R}}^{\mathsf{rf}}}^{\mathsf{tr}} e^{j\psi_{M_R^{\mathsf{tr}}}}\right), \tag{5}$$

where $\phi_{M_R^{\mathsf{rf}}} \in (0, 2\pi]$ and $\psi_{M_R^{\mathsf{tr}}} \in (0, 2\pi]$ denote the reflection and transmission phase shifts of the $M_R^{\mathsf{rf}}$-th and $M_R^{\mathsf{tr}}$-th element of STAR-RIS, respectively. Each element uses discrete phase values with resolution $\varepsilon = 2^b$, where $b$ is the number of quantization bits [8]. Hence, the admissible phase sets can be expressed, respectively, as

$$\mathcal{D}_\phi = \left\{0, \frac{2\pi}{\varepsilon}, ..., \frac{2\pi(\varepsilon-1)}{\varepsilon}\right\}, \mathcal{D}_\psi = \left\{0, \frac{2\pi}{\varepsilon}, ..., \frac{2\pi(\varepsilon-1)}{\varepsilon}\right\}, \tag{6}$$

The received signals at $\mathrm{U}_k$ and $\mathrm{U}_l$ are, respectively, given by

$$y_k = \hat{\mathbf{G}}_{\mathrm{S,U}_k}\mathbf{W}^{\mathsf{c}}\mathbf{s}^c + \sum_{k\in\mathcal{K}} \hat{\mathbf{G}}_{\mathrm{S,U}_k}\mathbf{W}_k^{\mathsf{p}}\mathbf{s}_k^{\mathsf{p}} + \sum_{l\in\mathcal{L}} \hat{\mathbf{g}}_{\mathrm{S,U}_k}\mathbf{W}_l^{\mathsf{p}}\mathbf{s}_l^p + \mathbf{n}_k, \tag{7}$$

$$y_l = \hat{\mathbf{G}}_{\mathrm{S,U}_l}\mathbf{W}^{\mathsf{c}}\mathbf{s}^c + \sum_{l\in\mathcal{L}} \hat{\mathbf{G}}_{\mathrm{S,U}_l}\mathbf{W}_l^{\mathsf{p}}\mathbf{s}_l^{\mathsf{p}} + \sum_{k\in\mathcal{K}} \hat{\mathbf{g}}_{\mathrm{S,U}_l}\mathbf{W}_k^{\mathsf{p}}\mathbf{s}_{kl}^p + \mathbf{n}_l, \tag{8}$$

where $\mathbf{n}_k \sim \mathcal{CN}(0, \mathbf{I}\sigma_k^2)$ and $\mathbf{n}_l \sim \mathcal{CN}(0, \mathbf{I}\sigma_l^2)$ denote the additive white noises at $\mathrm{U}_k$ and $\mathrm{U}_l$, respectively.

According to RSMA, the SINR for common-message decoding at $\mathrm{U}_k$ and $\mathrm{U}_l$ can, respectively, be written as

$$\gamma_k^{\mathsf{c}}(\mathbf{W}, \mathbf{\Phi}) = \frac{|\hat{\mathbf{G}}_{\mathrm{S,U}_k}^{\mathsf{H}}\mathbf{W}^{\mathsf{c}}|^2}{\varsigma_k^{\mathsf{c}}(\mathbf{W}, \mathbf{\Phi})}, \gamma_l^{c}(\mathbf{W}, \mathbf{\Psi}) = \frac{|\hat{\mathbf{G}}_{\mathrm{S,U}_l}^{\mathsf{H}}\mathbf{W}^{\mathsf{c}}|^2}{\varsigma_l^{\mathsf{c}}(\mathbf{W}, \mathbf{\Psi})}, \tag{9}$$

where

$$\varsigma_k^{\mathsf{c}}(\mathbf{W}, \mathbf{\Phi}) = \sum_{k\in\mathcal{K}} |\hat{\mathbf{G}}_{\mathrm{S,U}_k}^{\mathsf{H}}\mathbf{W}_k^{\mathsf{p}}|^2 + \sum_{l\in\mathcal{L}} |\hat{\mathbf{G}}_{\mathrm{S,U}_k}^{\mathsf{H}}\mathbf{W}_l^{\mathsf{p}}|^2 + \sigma_k^2,$$

$$\varsigma_l^{\mathsf{c}}(\mathbf{W}, \mathbf{\Psi}) = \sum_{l\in\mathcal{L}} |\hat{\mathbf{G}}_{\mathrm{S,U}_l}^{\mathsf{H}}\mathbf{W}_l^{\mathsf{p}}|^2 + \sum_{k\in\mathcal{K}} |\hat{\mathbf{G}}_{\mathrm{S,U}_l}^{\mathsf{H}}\mathbf{W}_k^{\mathsf{p}}|^2 + \sigma_l^2.$$

Whereas the SINR to decode the private message at $\mathrm{U}_k$ and

$\mathrm{U}_l$ can, respectively, be written as

$$\gamma_k^{\mathsf{p}}(\mathbf{W},\mathbf{\Phi})=\frac{|\hat{\mathbf{G}}_{\mathrm{S},\mathrm{U}_k}\mathbf{W}_k^{\mathsf{p}}|^2}{\varsigma_k^{\mathsf{p}}(\mathbf{W},\mathbf{\Phi})},\gamma_l^{\mathsf{p}}(\mathbf{W},\mathbf{\Psi})=\frac{|\hat{\mathbf{G}}_{\mathrm{S},\mathrm{U}_l}\mathbf{W}_l^{\mathsf{p}}|^2}{\varsigma_l^{\mathsf{p}}(\mathbf{W},\mathbf{\Psi})}, \tag{10}$$

where

$$\varsigma_k^{\mathsf{p}}(\mathbf{W},\mathbf{\Phi})=\sum_{k'\in\mathcal{K}\setminus\{k\}}|\hat{\mathbf{G}}_{\mathrm{S},\mathrm{U}_{k'}}\mathbf{W}_{k'}^{\mathsf{p}}|^2+\sum_{l\in\mathcal{L}}|\hat{\mathbf{G}}_{\mathrm{S},\mathrm{U}_k}^{\mathsf{H}}\mathbf{W}_l^{\mathsf{p}}|^2+\sigma_k^2,$$

$$\varsigma_l^{\mathsf{p}}(\mathbf{W},\mathbf{\Psi})=\sum_{l'\in\mathcal{L}\setminus\{l\}}|\hat{\mathbf{G}}_{\mathrm{S},\mathrm{U}_{l'}}\mathbf{W}_{l'}^{\mathsf{p}}|^2+\sum_{k\in\mathcal{K}}|\hat{\mathbf{G}}_{\mathrm{S},\mathrm{U}_l}^{\mathsf{H}}\mathbf{W}_k^{\mathsf{p}}|^2+\sigma_l^2.$$

The achievable SE (nat/s/Hz) for the common and private messages at $\mathrm{U}_k$ and $\mathrm{U}_l$ can, respectively, be expressed as

$$\mathsf{SE}_k^{\mathsf{c}}(\mathbf{W},\mathbf{\Phi})=\ln(1+\gamma_k^{\mathsf{c}}),\ \mathsf{SE}_k^{\mathsf{p}}(\mathbf{W},\mathbf{\Phi})=\ln(1+\gamma_k^{\mathsf{p}}), \tag{11}$$

$$\mathsf{SE}_l^{\mathsf{c}}(\mathbf{W},\mathbf{\Psi})=\ln(1+\gamma_l^{\mathsf{c}}),\ \mathsf{SE}_l^{\mathsf{p}}(\mathbf{W},\mathbf{\Psi})=\ln(1+\gamma_l^{\mathsf{p}}). \tag{12}$$

As the common SE is shared across users, the sum allocated to them satisfies $\sum_{k\in\mathcal{K}}\mathsf{SE}_k^{\mathsf{c}}+\sum_{l\in\mathcal{L}}\mathsf{SE}_l^{\mathsf{c}}\le\mathsf{SE}^{\mathsf{c}}$, where $\mathsf{SE}^{\mathsf{c}}=\min\{\mathsf{SE}_1^{\mathsf{c}},...,\mathsf{SE}_K^{\mathsf{c}},\mathsf{SE}_1^{\mathsf{c}},...,\mathsf{SE}_L^{\mathsf{c}}\}$. Hence, the total achievable SEs of $\mathrm{U}_k$ and $\mathrm{U}_l$ can, respectively, be expressed as

$$\mathsf{SE}_k(\mathbf{W},\mathbf{\Phi})=\mathsf{SE}_k^{\mathsf{c}}(\mathbf{W},\mathbf{\Phi})+\mathsf{SE}_k^{\mathsf{p}}(\mathbf{W},\mathbf{\Phi}), \tag{13}$$

$$\mathsf{SE}_l(\mathbf{W},\mathbf{\Psi})=\mathsf{SE}_l^{\mathsf{c}}(\mathbf{W},\mathbf{\Psi})+\mathsf{SE}_l^{\mathsf{p}}(\mathbf{W},\mathbf{\Psi}). \tag{14}$$

### *B. Problem Formulation*

We assume the STAR-RISs mounted on AAVs remain static within a coherence interval, ensuring channel stability for effective joint optimization of the precoders and phase shifts [**?**]. Based on (13) and (14), our goal is to maximize energy efficiency subject to the BS power budget, discrete STAR-RIS phase constraints, and QoS requirements, by optimizing $\mathbf{W}^{\mathsf{c}},\mathbf{W}_k^{\mathsf{p}},\mathbf{W}_l^{\mathsf{p}},\mathbf{\Phi}_r$, and $\mathbf{\Psi}_r$. The EE metric incorporates total hardware power consumption, modeled as

$$\begin{aligned}\check{P}=&(\|\mathbf{W}_1\|^2+\|\mathbf{W}_2\|^2+\|\mathbf{W}_3\|^2)/\xi+M_{\mathrm{S}}P_{\mathrm{S}}^{\mathsf{dy}}+P_{\mathrm{S}}^{\mathsf{st}}\\&+\sum_{r\in\mathcal{R}}M_{\mathrm{R}}P_r^{\mathsf{st}}+\sum_{k\in\mathcal{K}}M_{\mathrm{K}}P_k^{\mathsf{st}}+\sum_{l\in\mathcal{L}}M_{\mathrm{L}}P_l^{\mathsf{st}},\end{aligned} \tag{15}$$

where $\xi\in(0,1]$ denotes the transmit-power efficiency, $P_{\mathrm{S}}^{\mathsf{dy}}$ denotes the BS dynamic power, and $P_{\mathrm{S}}^{\mathsf{st}}$ denotes the BS static power. Likewise, $P_k^{\mathsf{st}}$, $P_k^{\mathsf{st}}$, and $P_l^{\mathsf{st}}$ are the static hardware powers of $\mathrm{R}_r$, $\mathrm{U}_k$, and $\mathrm{U}_l$, respectively. The total circuit power is $\check{P}^{\mathsf{CP}}=P_{\mathrm{S}}^{\mathsf{st}}+\sum_{r\in\mathcal{R}}M_{\mathrm{K}}P_k^{\mathsf{st}}+\sum_{k\in\mathcal{K}}M_{\mathrm{K}}P_k^{\mathsf{st}}+\sum_{l\in\mathcal{L}}M_{\mathrm{L}}P_l^{\mathsf{st}}$. Accordingly, the EEM problem can be formulated as

P1 : Formulated Problem

$$\max_{\mathbf{W},\mathbf{\Phi},\mathbf{\Psi}}\ \mathcal{E}\triangleq\frac{\sum_{k\in\mathcal{K}}\mathsf{SE}_k(\mathbf{W},\mathbf{\Phi})+\sum_{l\in\mathcal{L}}\mathsf{SE}_l(\mathbf{W},\mathbf{\Psi})}{(\|\mathbf{W}_1\|^2+\|\mathbf{W}_2\|^2+\|\mathbf{W}_3\|^2)/\xi+\check{P}^{\mathsf{CP}}} \tag{16a}$$

$$\text{s.t.}\quad \mathsf{SE}_k(\mathbf{W},\mathbf{\Phi})\ge\bar{\mathsf{SE}}_k,\quad\forall k\in\mathcal{K}, \tag{16b}$$

$$\mathsf{SE}_l(\mathbf{W},\mathbf{\Phi})\ge\bar{\mathsf{SE}}_l,\quad\forall l\in\mathcal{L}, \tag{16c}$$

$$||\mathbf{W}_1||^2+||\mathbf{W}_2||^2+||\mathbf{W}_3||^2\le\bar{P}_{\mathrm{BS}}, \tag{16d}$$

$$\phi_{M_r^{\mathsf{rf}}}\in\mathcal{D}_\phi,\quad\psi_{M_r^{\mathsf{tr}}}\in\mathcal{D}_\psi, \tag{16e}$$

$$\lambda_{M_{\mathrm{R}}^{\mathsf{rf}}}^{\mathsf{rf}}+\lambda_{M_{\mathrm{R}}^{\mathsf{tr}}}^{\mathsf{tr}}=1, \tag{16f}$$

$$\lambda_{M_{\mathrm{R}}^{\mathsf{rf}}}^{\mathsf{rf}}\in\{0,1\},\quad\lambda_{M_{\mathrm{R}}^{\mathsf{tr}}}^{\mathsf{tr}}\in\{0,1\}, \tag{16g}$$

where Constraints (16b) and (16c) enforce per-user QoS, requiring the SEs of $\mathrm{U}_k$ and $\mathrm{U}_l$ to exceed the preset thresholds $\bar{\mathsf{SE}}_k>0$ and $\bar{\mathsf{SE}}_l>0$, respectively. Constraint (16d) limits the aggregate transmit power to the BS budget. Constraint (16e) enforces discretized STAR-RIS phase shifts. Constraints (16f) and (16g) capture the MS operation via binary variables. Consequently, the objective in (16a) is non-convex in $\mathbf{W},\mathbf{\Phi},\mathbf{\Psi}$, and the problem is a mixed-integer non-convex program, making global optimization highly challenging.

## III. The Proposed EEM Algorithm

Solving problem (16) is inherently more challenging than the SE problem in [10], since the EEM is a mixed-integer, non-convex fractional program, and entails combinatorial (exponential) complexity for global optimality. Nevertheless, we will show—via suitable transformations—that an IA-based algorithm can efficiently handle the EEM. A key step is to relax the discrete variables to continuous ones. The resulting relaxed problem can be approximated as

P2 : Relaxed Problem

$$\max_{\mathbf{W},\mathbf{\Phi},\mathbf{\Psi}}\ \mathcal{E}\triangleq\frac{\sum_{k\in\mathcal{K}}\mathsf{SE}_k(\mathbf{W},\mathbf{\Phi})+\sum_{l\in\mathcal{L}}\mathsf{SE}_l(\mathbf{W},\mathbf{\Psi})}{(\|\mathbf{W}_1\|^2+\|\mathbf{W}_2\|^2+\|\mathbf{W}_3\|^2)/\xi+\check{P}^{\mathsf{CP}}} \tag{17a}$$

$$\text{s.t. }\phi_{M_r^{\mathsf{rf}}}\in[0,2\pi],\quad\psi_{M_r^{\mathsf{tr}}}\in[0,2\pi], \tag{17b}$$

$$(16\text{b}),(16\text{c}),(16\text{d}),(16\text{f}),(16\text{g}). \tag{17c}$$

We observe that (17) is a non-convex fractional program. Our approach is to split it into phase-shift and beamforming subproblems and handle them via alternating optimization.

### *A. Phase Shift Subproblem*

We first tackle the phase-shift subproblem by fixing the beamforming variables. Consequently, (17) can be reformulated as

P3 : Phase Shift Subproblem

$$\max_{\mathbf{\Phi},\mathbf{\Psi}}\ \mathcal{E}\triangleq\frac{\sum_{k\in\mathcal{K}}\mathsf{SE}_k(\mathbf{\Phi})}{(1+K+L)/\xi+\check{P}^{\mathsf{CP}}}+\frac{\sum_{l\in\mathcal{L}}\mathsf{SE}_l(\mathbf{\Psi})}{(1+K+L)/\xi+\check{P}^{\mathsf{CP}}} \tag{18a}$$

$$\text{s.t.}\quad(16\text{b}),(16\text{c}),(16\text{f}),(16\text{g}),(17\text{b}). \tag{18b}$$

Noting that (18a) is concave and (17b) is linear within (18), we design an efficient solver using a Bisection Search Algorithm, presented in Alg. 1.

**Algorithm 1** Proposed Bisection-Search Algorithm for Solving Phase Shift Subproblem (18)

**Input:** $K$, $L$, $R$, $M_{\mathrm{R}}$.
**Output:** $\mathbf{\Phi}_r^\star$, $\mathbf{\Psi}_r^\star$.
Initialize the upper ($\hat{\phi}$, $\hat{\psi}$) and lower ($\check{\phi}$,$\check{\psi}$) bounds;
**repeat**
  Calculate $\phi^\star=(\check{\phi}+\hat{\phi})/2$; $\psi^\star=(\check{\psi}+\hat{\psi})/2$
  Update $\phi_{m_r^{\mathrm{rf}}}(\phi^\star)$; $\psi_{m_r^{\mathrm{rf}}}(\psi^\star)$
  Solve the problem (18);
**until** Convergence

Note that the optimal phase-shift solution is continuous and therefore cannot be used directly in the original problem. To

bridge this gap, we apply a rounding operation to (18) after obtaining the continuous optimum, i.e.,

$$\mathbf{\Lambda}^{\star} = \lceil \mathbf{\Lambda}^{(\star)} + \delta \rceil, \ \forall r \in \mathcal{R}, \ \forall m_{\mathrm{R}} \in M_{\mathrm{R}}, \tag{19}$$

where $\mathbf{\Lambda}^{(\star)} \in \{\mathbf{\Phi}_r^{\star}, \mathbf{\Psi}_r^{\star}\}$ and $\delta = (360/\varepsilon)/2$ denotes the rounding step size. After obtaining the optimal phase shifts, we proceed to the beamforming subproblem, which is solved in alternation and described in the next subsection.

### *B. Beamforming Subproblem*

We now turn to the beamforming optimization. To iteratively approximate the non-convex terms, we introduce an auxiliary variable $\Xi > 0$ that satisfies

$$(\|\mathbf{W}_1\|^2 + \|\mathbf{W}_2 + \|\mathbf{W}_3\|^2)/\xi + P^{\mathsf{CP}} \leq \Xi. \tag{20}$$

Plugging the optimal phase-shift values from Alg. 1 into (17), we obtain

$$\begin{aligned} \mathrm{P4}: \quad & \text{Beamforming Subproblem} \\ \max_{\mathbf{W},\Xi} \quad & \mathcal{E} \triangleq \sum_{k\in\mathcal{K}} \mathsf{SE}_k(\mathbf{W})/\Xi + \sum_{l\in\mathcal{L}} \mathsf{SE}_l(\mathbf{W})/\Xi && (21a) \\ \text{s.t.} \quad & (16b),\ (16c),\ (16d). && (21b) \end{aligned}$$

We define $\mathbf{e} \triangleq \{e_k, e_l\}_{k\in\mathcal{K}, l\in\mathcal{L}}$ to represent soft energy efficiencies, $\boldsymbol{\gamma} \triangleq \{\gamma_k^{\mathsf{c}}, \gamma_k^{\mathsf{p}}, \gamma_l^{\mathsf{c}}, \gamma_l^{\mathsf{p}}\}_{k\in\mathcal{K}, l\in\mathcal{L}}$ represent the users' SINRs ($\mathrm{U}_k, \mathrm{U}_l$), $\bar{\mathsf{SE}} \triangleq \{\bar{\mathsf{SE}}_k, \bar{\mathsf{SE}}_l\}_{k\in\mathcal{K}, l\in\mathcal{L}}$ represent the minimum SE threshold of users, and $\boldsymbol{\varsigma} \triangleq \{\varsigma_k^{\mathsf{c}}, \varsigma_k^{\mathsf{p}}, \varsigma_l^{\mathsf{c}}, \varsigma_l^{\mathsf{p}}\}$ represent the interference of users. With these variables, (21) can be rewritten as

$$\begin{aligned} \max_{\mathsf{W},\Xi,\boldsymbol{\gamma},\boldsymbol{e}} \quad & \mathcal{E} \triangleq \sum_{k\in\mathcal{K}} e_k + \sum_{l\in\mathcal{L}} e_l && (22a) \\ \text{s.t.} \quad & \boldsymbol{\gamma}(\mathbf{W}) \geq 1/\boldsymbol{\gamma}, && (22b) \\ & \ln(1+1/\gamma_k^{\mathsf{c}})/\Xi + \ln(1+1/\gamma_k^{\mathsf{p}})/\Xi \geq e_k, && (22c) \\ & \ln(1+1/\gamma_l^{\mathsf{c}})/\Xi + \ln(1+1/\gamma_l^{\mathsf{p}})/\Xi \geq e_l, && (22d) \\ & \ln(1+1/\boldsymbol{\gamma}) \geq \bar{\mathsf{SE}}, && (22e) \\ & (16d),\ (20). && (22f) \end{aligned}$$

With $x^{(\kappa)}$ the feasible point at iteration $(\kappa)$, and following Lemmas 1 and 2 in [11, Lemma 1, Lemma 2], we approximate constraints (22b) at the $\kappa$ iteration by

$$\boldsymbol{\varsigma}(\mathbf{W})/\boldsymbol{\gamma} \leq f^{(\kappa)}(\mathbf{W}), \tag{23}$$

where $f^{(\kappa)}(\mathbf{W}) \triangleq \|\hat{\mathbf{G}}\mathbf{W}\|^2 \geq 2\Re\{(\hat{\mathbf{G}}\mathbf{W}^{(\kappa)})^*(\hat{\mathbf{G}}\mathbf{W})\} - \|\hat{\mathbf{G}}\mathbf{W}^{(\kappa)}\|^2$.

The term $\ln(1+1/\gamma_k^{\mathsf{c}})/\Xi + \ln(1+1/\gamma_k^{\mathsf{p}})/\Xi$ and $\ln(1+1/\gamma_l^{\mathsf{c}})/\Xi + \ln(1+1/\gamma_l^{\mathsf{p}})/\Xi$ on the left-hand side of (22c)–(22d) is convex in $(\boldsymbol{\gamma}, \Xi)$. Its first-order approximation at $(\boldsymbol{\gamma}^{(\kappa)}, \Xi^{(\kappa)})$ is given by [12, Eq. (18)] $\ln(1+1/\gamma^c)/\Xi + \ln(1+1/\gamma^p)/\Xi \geq (2\ln(1+1/\gamma^{c,(\kappa)})/\Xi^{(\kappa)} + 1/(\Xi^{(\kappa)}(\gamma^{c(\kappa)}+1)) - \gamma^c/(\Xi^{(\kappa)}\ \gamma^{c,(\kappa)}(\gamma^{c(\kappa)}+1)) - \ln(1+1/\gamma^{c(\kappa)})\Xi/(\Xi^{(\kappa)})^2) + (2\ln(1+1/\gamma^{p,(\kappa)})/\Xi^{(\kappa)} + 1/(\Xi^{(\kappa)}(\gamma^{p(\kappa)}+1)) - \gamma^p/(\Xi^{(\kappa)}\ \gamma^{p,(\kappa)}(\gamma^{p(\kappa)}+1)) - \ln(1+1/\gamma^{c(\kappa)})\Xi/(\Xi^{(\kappa)})^2) \triangleq \mathcal{B}^{(\kappa)}(\gamma, \Xi), \ \forall \Xi^{(\kappa)} > 0, \gamma^{c(\kappa)} > 0, \gamma^{p(\kappa)} > 0$.

Using [10, Eq. (34)], constraint (22e) is approximated by $\mathcal{C}^{(\kappa)}(\boldsymbol{\gamma}) \triangleq \ln(1+1/\boldsymbol{\gamma}) \geq \ln(1+(\boldsymbol{\gamma}^{(\kappa)})^{-1}) + (\boldsymbol{\gamma}^{(\kappa)}+1)^{-1} - \boldsymbol{\gamma}[\boldsymbol{\gamma}^{(\kappa)}(\boldsymbol{\gamma}^{(\kappa)}+1)]^{-1}$.

Consequently, (22) is approximated by the following convex formulation at iteration $(\kappa+1)$:

$$\begin{aligned} \mathrm{P5}: \quad & \text{Convex Problem} \\ \max_{\mathbf{W},\boldsymbol{\Xi},\boldsymbol{\gamma},\mathbf{e}} \quad & \mathcal{E} \triangleq \sum_{k\in\mathcal{K}} e_k + \sum_{l\in\mathcal{L}} e_l && (24a) \\ \text{s.t.} \quad & \mathcal{B}_k^{(\kappa)}(\boldsymbol{\gamma}, \Xi) \geq e_k, \ \mathcal{B}_l^{(\kappa)}(\boldsymbol{\gamma}, \Xi) \geq e_f, && (24b) \\ & \mathcal{C}^{(\kappa)}(\boldsymbol{\gamma}) \geq \bar{\mathsf{SE}} && (24c) \\ & (16d),\ (20),\ (23). && (24d) \end{aligned}$$

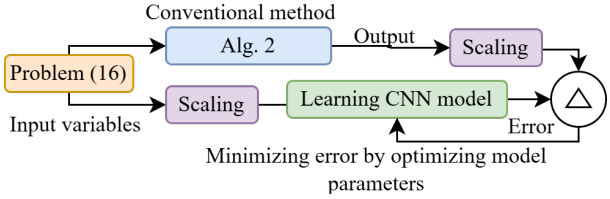
To wrap up, Alg. 2 presents the proposed low-complexity iterative algorithm, including the bisection module in Alg. 1.

**Algorithm 2** : Proposed Inner-Approximation Framework for Problem (16)

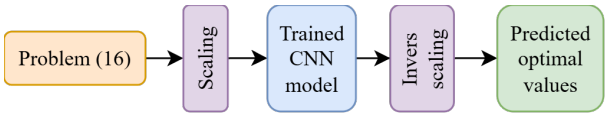
1: **Initialization:** $(\mathbf{W}, \mathbf{\Phi}, \mathbf{\Psi}) \leftarrow 0$, and generate an initial feasible point $(\mathbf{W}^{(0)}, \Xi^{(0)}, \boldsymbol{\gamma}^{(0)})$ randomly;
2: **Output:** $\mathcal{E}$ and $(\mathbf{W}^{\star}, \mathbf{\Phi}^{\star}, \mathbf{\Psi}^{\star})$.
3: **repeat**
4: Solve problem (18) by running Alg. 1 to achieve $(\mathbf{\Phi}^{\star}, \mathbf{\Psi}^{\star})$;
5: Round $\mathbf{\Phi}^{\star}$ and $\mathbf{\Psi}^{\star}$ by using (19);
6: Solve problem (24) to achieve $(\mathbf{W}^{\star}, \Xi^{\star}, \boldsymbol{\gamma}^{\star})$;
7: **until** Convergence
8: Calculate $\mathcal{E}$ in (16) based on $(\mathbf{W}^{\star}, \mathbf{\Phi}^{\star}, \mathbf{\Psi}^{\star})$;

## IV. Deep Learning Framework Design

In this section, we present a CNN-based deep learning framework for the EEM problem in multi-STAR-RIS–assisted mMIMO–RSMA networks, as illustrated in Fig. 2. As shown




(a) Offline Learning




(b) Online Predicting

Fig. 2. The proposed DL-based CNN framework to the EE problem

in Fig. 2(a), the CNN is trained offline to learn the mapping from the input parameters of problem (16) to the optimal solutions (targets) produced by Alg. 2. A scaling function normalizes all inputs to $[0, 1]$, which stabilizes gradients and accelerates convergence. After training, the resulting deep

CNN (weights and biases) can be used for real-time inference to predict the optimal precoding matrices and STAR-RIS phase shifts with high accuracy, as illustrated in Fig. 2(b). In contrast to the conventional pipeline—where Alg. 2 must be run to obtain the optimal transmission/reflection phases at the STAR-RISs and the precoding matrix—the trained CNN outputs these directly for new inputs. In the proposed DL framework, the CNN inputs include the channel matrices from BS to STAR-RIS $\{\mathbf{G}_{\mathrm{S,R}_r^{\mathrm{rf}}}, \mathbf{G}_{\mathrm{S,R}_r^{\mathrm{tf}}}\}$, BS to users, and STAR-RIS to users $\{\mathbf{G}_{\mathrm{R}_r^{\mathrm{rf}},\mathrm{U}_k}, \mathbf{G}_{\mathrm{R}_r^{\mathrm{tr}},\mathrm{U}_l}\}$; the positions of all users $\{\mathrm{U}_k^{\mathsf{Pss}}, \mathrm{U}_l^{\mathsf{Pss}}\}$; the position of BS $\mathrm{BS}^{\mathsf{Pss}}$; the position of all STAR-RISs $\mathrm{R}_r^{\mathsf{Pss}}$; antenna counts at the BS and users $(M_{\mathrm{S}}, \{M_{\mathrm{K}}, M_{\mathrm{L}}\})$; the number of elements per STAR-RIS $M_{\mathrm{R}}$; and the numbers of users and STAR-RISs $(R, K, L)$. The outputs are the precoding matrices $\mathbf{W} \in \{\mathbf{W}^{\mathsf{c}}, \mathbf{W}_k^{\mathsf{p}}, \mathbf{W}_l^{\mathsf{p}}\}$ and the STAR-RIS transmission/reflection phase shifts $\{\mathbf{\Phi}_r, \mathbf{\Psi}_r\}$. Accordingly, the input dimension is $M_{\mathrm{S}}((2M_{\mathrm{R}}+2)+M_{\mathrm{L}}+M_{\mathrm{K}})+M_{\mathrm{R}}(M_{\mathrm{L}}+M_{\mathrm{K}})R\mathrm{R}_r^{\mathsf{Pss}}+K\mathrm{U}_k^{\mathsf{Pss}}+L\mathrm{U}_l^{\mathsf{Pss}}+8$, while the output dimension is $(K+L)(M_{\mathrm{BS}}\times L)+(M_{\mathrm{BS}}\times L)+RM_{\mathrm{R}}$. During inference, we apply the same input scaling before prediction and perform inverse scaling on the CNN outputs to recover values in the original physical units.

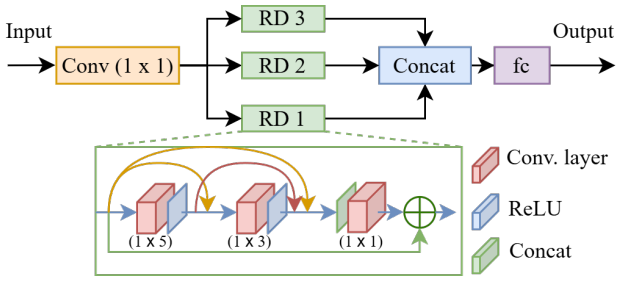



Fig. 3. The architecture of the Deep CNN design.

Fig. 3 depicts the proposed deep-CNN architecture, built around three parallel residual-dense (RD) blocks. These blocks capture rich representations by forming multi-level connections among feature maps and stacking multiple convolutional layers for stronger feature extraction and stable gradient flow. Within each RD block, a $1\times1$ convolution reduces dimensionality while preserving salient information, and outputs from intermediate layers are fused via concatenation to aggregate features across processing stages. A fully connected (fc) layer then maps the fused features to the prediction space. The dense skip connections mitigate vanishing gradients, speed up convergence, and improve learning effectiveness; using multiple RD blocks also helps curb overfitting and supports better generalization. This design enables accurate, efficient inference—even in real time—by combining residual-dense learning with feature aggregation. ReLU is used as the activation function, and all convolutional and fc layers employ 64 kernels/neurons.

## V. Simulation Results

To evaluate the proposed algorithms for the formulated problem in STAR-RIS–assisted mMIMO–RSMA networks, we use the following simulation settings: cell size 500 m $\times$ 500 m $\times$ 50 m, $\bar{\mathsf{SE}}_k = \bar{\mathsf{SE}}_l$ = 1 bps/Hz, $M_{\mathrm{K}} = M_{\mathrm{L}}$ = 10, $P_{\mathrm{S}}^{\mathsf{dy}}$ = 10 dBm, $P_{\mathrm{S}}^{\mathsf{st}}$ = 15 dBm, and $P_k^{\mathsf{st}} = P_l^{\mathsf{st}}$ = 5 dBm, $P_r^{\mathsf{st}}$ = 0 dBm. The convex optimization is carried out in MATLAB via YALMIP (SDPT3 backend) [10]. The dataset is divided into 90% for training and 10% for testing.

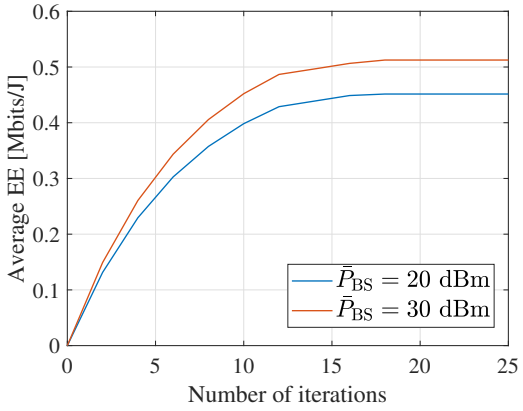


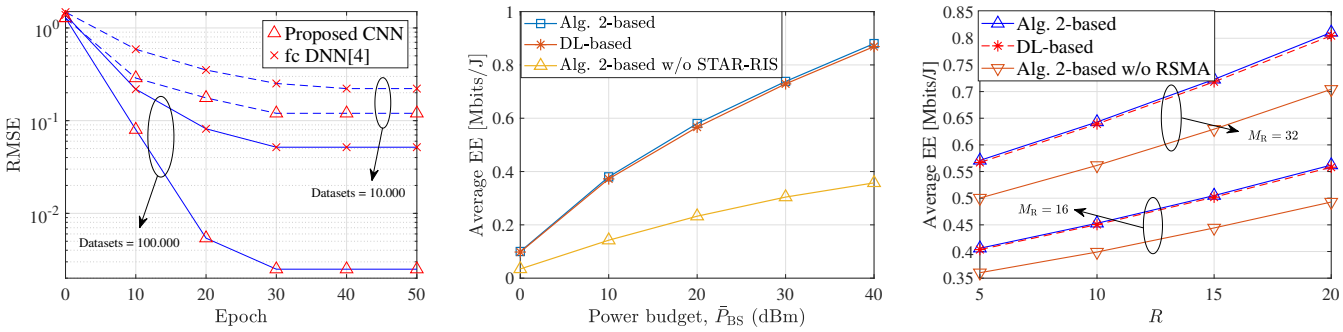

Fig. 4. Convergence behaviour of the proposed Alg. 2.

As depicted in Fig. 4, Algorithm 2 solves problem (16) efficiently across a range of BS power budgets, reaching the optimal value in roughly 13 iterations via monotonic objective improvement. In addition, a larger power budget leads to higher average EE, owing to the corresponding increase in achievable capacity, in line with constraint (16d).

As illustrated in Fig. 5(a), RMSE drops with increasing epochs, reflecting progressive refinement of network parameters. The proposed CNN yields the lowest RMSE, outperforming the fc DNN benchmark [4] and confirming its advantage on high-dimensional data. Additionally, both DNN and CNN benefit from larger training sets, which provide broader feature coverage and lead to lower errors.

Fig. 5(b) illustrates how the maximum power budget at the BS affects the average EE. As $\bar{P}_{\mathrm{BS}}$ increases, the average SE rises because higher transmit power improves the received signal quality at ground users. The Alg. 2–based method attains the highest SE, as it explores the feasible set to globally optimize the objective, yielding superior solutions. By contrast, the Alg. 2 variant without (w/o) STAR-RIS performs worst since users rely solely on the direct BS to user links with no additional STAR-RIS gain. Finally, the DL-based approach closely tracks the benchmark Alg. 2–based curve, indicating that the DL model predicts the output parameters with high accuracy.

Fig. 5(c) plots average EE versus the number of STAR-RISs $R$ for different element counts. EE increases with $R$ as additional panels improve users' effective channels. Likewise, more elements per STAR-RIS further raise EE by enabling tighter signal focusing. The conventional scheme without RSMA performs worst because it allocates orthogonal time–frequency resources to each user. By contrast, the DL-




(a) Epoch vs. RMSE (b) Impact of the power budget on the average EE (c) $R$ vs. Average EE

Fig. 5. Epoch vs. RMSE and the impact of the power budget and the number of STAR-RISs $R$ on the average EE

based method accurately predicts the precoders and STAR-RIS phase shifts, achieving strong EE.

TABLE I
THE EXECUTION TIME OF ALG. 2-BASED VS. DL-BASED APPROACHES

| Number of users | Alg. 2-based | DL-based |
|---|---|---|
| 8 | 54.3 s | 0.0137 s |
| 12 | 80.6 s | 0.0134 s |
| 16 | 105.3 s | 0.0153 s |

Finally, we assess the execution time of the DL-based method for obtaining the optimal solution (Table I). The simulation was run on hardware specifications of an AMD Ryzen 7 9700X 8-core processor, a single NVIDIA GeForce GTX 1050 Ti graphic processing unit, and 16 GB of RAM. The results show that the DL approach attains the optimum with lower runtime, even as the number of users grows, whereas the Alg. 2-based solver slows significantly with scale. This is because the DL model directly maps inputs to near-optimal outputs, while Alg. 2 requires multiple iterations, increasing computation time.

## VI. CONCLUSIONS

This paper investigated a DL-based EEM framework for STAR-RIS–assisted mMIMO–RSMA networks. We formulated joint optimization of the BS precoding matrix and STAR-RIS phase shifts under BS power, discrete phase, and minimum-QoS constraints, yielding a mixed-integer non-convex problem. To solve it, we decoupled the task into phase-shift and beamforming subproblems: the former handled via a bisection search, the latter transformed into a tractable surrogate and solved by an IA method. For real-time operation, we further designed a DL framework that predicts near-optimal phase shifts and precoders across varying system parameters. Simulations show that the DL approach attains optimal-quality solutions with markedly lower runtime than conventional solvers, and we quantified the influence of key parameters on overall performance.

## ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2023-00244014). This work was supported by Chungbuk National University BK21 program (2025).

## REFERENCES

[1] M. Na, J. Lee, G. Choi, T. Yu, J. Choi, J. Lee, and S. Bahk, "Operator's Perspective on 6G: 6G Services, Vision, and Spectrum," *IEEE Commun. Mag.*, vol. 62, no. 8, pp. 178–184, 2024.

[2] R. H. Y. Perdana, T.-V. Nguyen, Y. Pramitarini, and B. An, "Deep Learning-Based Energy Efficiency Maximization in Massive MIMO-NOMA Networks With Multiple RISs," in *2024 Int. Conf. Artif. Intell. Inf. Commun.* Osaka, Japan: IEEE, 2024, pp. 382–387.

[3] M. Soleymani, I. Santamaria, and E. Jorswieck, "Energy-efficient Rate Splitting for MIMO STAR-RIS-assisted Broadcast Channels with I/Q Imbalance," in *Eur. Signal Process. Conf.* Helsinki, Finland: EURASIP, 2023, pp. 1504–1508.

[4] R. H. Y. Perdana, T.-V. Nguyen, and B. An, "A Deep Learning-Based Spectral Efficiency Maximization in Multiple Users Multiple STAR-RISs Massive MIMO-NOMA Networks," in *2023 Twelfth Int. Conf. Ubiquitous Futur. Networks*. Paris, France: IEEE, 2023, pp. 675–680.

[5] J. Zheng, J. Zhang, H. Du, D. Niyato, D. I. Kim, and B. Ai, "Rate-Splitting for CF Massive MIMO Systems with Channel Aging," *IEEE Trans. Veh. Technol.*, vol. 73, no. 1, pp. 1485–1490, 2024.

[6] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously Transmitting and Reflecting (STAR) RIS Aided Wireless Communications," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 5, pp. 3083–3098, 2022.

[7] Y. Pramitarini, R. H. Y. Perdana, K. Shim, and B. An, "Secure Multicast Routing Against Collaborative Attacks in FANETs with CF-mMIMO and STAR-RIS: Blockchain and Federated Learning Design," *IEEE Internet Things J.*, vol. 12, no. 12, pp. 22 404–22 426, 2025.

[8] R. H. Y. Perdana, T. V. Nguyen, Y. Pramitarini, D. H. Nguyen, and B. An, "Enhancing Spectral Efficiency of Short-Packet Communications in STAR-RIS-Assisted SWIPT MIMO-NOMA Systems with Deep Learning," *IEEE Trans. Wirel. Commun.*, vol. 24, no. 1, pp. 842–859, 2025.

[9] D. Galappaththige and C. Tellambura, "Sum Rate Maximization for RSMA-Assisted CF mMIMO Networks With SWIPT Users," *IEEE Wirel. Commun. Lett.*, vol. 13, no. 5, pp. 1300–1304, 2024.

[10] R. H. Y. Perdana, T.-V. Nguyen, and B. An, "Adaptive User Pairing in Multi-IRS-aided Massive MIMO-NOMA Networks: Spectral Efficiency Maximization and Deep Learning Design," *IEEE Trans. Commun.*, vol. 71, no. 7, pp. 4377–4390, 2023.

[11] T. V. Nguyen, V. D. Nguyen, D. B. Da Costa, and B. An, "Hybrid User Pairing for Spectral and Energy Efficiencies in Multiuser MISO-NOMA Networks with SWIPT," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4874–4890, Aug. 2020.

[12] V. D. Nguyen, H. V. Nguyen, O. A. Dobre, and O. S. Shin, "A New Design Paradigm for Secure Full-Duplex Multiuser Systems," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 7, pp. 1480–1498, 2018.