# A Deep Learning-Based Spectral Efficiency Maximization in Multiple Users Multiple STAR-RISs Massive MIMO-NOMA Networks

Ridho Hendra Yoga Perdana* Toan-Van Nguyen†, and Beongku An‡
*Dept. of Software and Communications Engineering in Graduate School, Hongik University, Republic of Korea
†School of Computer Science and Engineering, International University - Vietnam National University, HCM City, Vietnam.
‡Dept. of Software and Communications Engineering, Hongik University, Republic of Korea
Emails: mail.rhyp@gmail.com, vannguyentoan@gmail.com, beongku@hongik.ac.kr

***Abstract*—In this paper, we investigate a deep learning-based spectral efficiency maximization in multiple users multiple simultaneously transmitting and reflecting reconfigurable intelligent surfaces (STAR-RISs) massive multiple-input multiple-output (MIMO)-non-orthogonal multiple access (NOMA) networks, where multiple STAR-RISs are deployed to assist a base station (BS) in communicating with multiple near and far users simultaneously. To improve spectral efficiency (SE), we formulate a joint optimization of the precoding matrix, the phase shift of reflection and transmission of the STAR-RIS problem subject to the individual quality of services (QoS) of each user, maximum power budget at BS and the discrete phase shift. The formulated problem belongs to the class of mixed-integer non-convex optimization problem, which is difficult to solve optimally. We first relax the formulated problem and then transform it into an equivalent non-convex one but with a more tractable form. We then apply the inner approximation method to approximate the non-convex parts iteratively. Towards real-time optimization, we propose a novel deep learning framework with a convolutional neural network (CNN) model to predict the downlink SE based on users' position information and channel gains. Numerical results show that the proposed approach improves the SE compared to the massive MIMO systems with conventional RIS. Besides, the DL framework for the proposed scheme predicts the optimal SE within a short time compared to the conventional approach.**

***Index Terms*—Deep convolutional neural networks, inner approximation, massive MIMO, NOMA, non-convex optimization, phase shift, spectral efficiency, STAR-RIS.**

## I. Introduction

Spectral efficiency (SE) is one of the key performance metrics in the fifth generation (5G) and beyond 5G [1]. In massive multiple-input multiple-output (MIMO) systems, a base station (BS) equipped large antenna array can improve the SE performance [2], [3]. To provide high SE and large connectivity in the network, massive MIMO combined with non-orthogonal multiple access (NOMA) has been realized in [4], [5]. In [4], massive MIMO-NOMA was investigated for the uplink channels to improve the number of connected devices and overall SE while beamforming selection and user grouping for massive MIMO-NOMA to improve the downlink SE in [6].

Simultaneously transmitting and reflecting reconfigurable intelligent surface (STAR-RIS) has recently proposed to achieve $360^0$ wireless coverage and improve network performance by using both sides of the surface [7]. However, in the context of the SE improvement, the analysis has been restricted to a single STAR-RIS. Moreover, recent works have not investigated the impact of multiple users multiple STAR-RISs on the massive MIMO-NOMA networks, which brings several technical challenges related to the system's complexity.

On the other hand, convolutional neural network (CNN) has gained significant attention recently since it can effectively solve various problem in wireless communication systems [8], [9]. Due to its accuracy in predicting high non-linear functions at low complexity, it has been studied for a wide range of applications in wireless networks such as beamforming optimization [10], power allocation [11] and channel estimation in cellular massive MIMO systems [12]. The CNN can help to predict the optimal solution in real-time settings in wireless communication networks since the deep CNN with multiple convolutional blocks effectively gathers the highly relevant features of each block at multiple signal input resolutions [13]. In this paper, we investigate a spectral efficiency problem in multiple users multiple STAR-RISs massive MIMO-NOMA networks, where multiple STAR-RISs aim to enhance the system performance, accompanied by an efficient deep convolutional neural network evaluation approach. The benefits of implementing a CNN, STAR-RIS, and massive MIMO-NOMA networks due to a large number of STAR-RIS and users can reduce the processing time and complexity, also promising to achieve optimal solutions with high SE. At the same time, other works have not studied the effect of CNN and multiple STAR-RISs in massive MIMO-NOMA networks. We briefly outline the main contributions of this paper as follows:

- We propose a joint optimization problem of a linear precoding matrix and phase shift of transmission and reflection for multiple users multiple STAR-RISs-aided massive MIMO-NOMA networks to achieve the maximum downlink SE. The problem belongs to the class of mixed-integer non-convex program problem, which is very challenging to solve globally.
- We propose an iterative algorithm with low complexity to

solve the formulated problem effectively. Firstly, we relax the phase shift constraint to be continuous and transform the formulated problem into a non-convex equivalent form yet with a more tractable form. We then apply the inner approximation (IA) method to approximate the non-convex parts iteratively.

- Towards real-time optimization, we propose a novel deep learning (DL) framework with a convolutional neural network model to predict the optimal solution of the SE maximization problem based on the designed IA method.
- Numerical results show the SE performance improvement of the proposed scheme over its counterpart with conventional RIS. Furthermore, the proposed DL framework can predict the optimal solution of the SE maximization problem within a short time compared to the conventional approach. Besides, the impact of multiple STAR-RISs on massive MIMO-NOMA networks is discussed and evaluated thoroughly.

*Mathematical Notations*: Small face, boldface lower and upper-case letters denote scalars, vectors and matrices, respectively. $\|.\|$ denotes the Euclidean norm of matrices. The complex numbers denote as $\mathbb{C}$, and $\Re$ symbolizes the real part. The circularly symmetric complex Gaussian distribution denotes as $\mathcal{CN}(\upsilon, \mathbf{A})$ with covariance matrix $\mathbf{A}$ and mean $\upsilon$.

## II. System Model and Problem Formulation

### A. System Model

Let us consider a downlink of multiple users in multiple STAR-RIS massive MIMO-NOMA networks, where a base station (BS) with $M_{\mathsf{BS}} > 1$ antennas transmits its message to $N$ near users ($\mathsf{U}_n$) and $F$ far users ($\mathsf{U}_f$) with $M_{\mathsf{N}}$ and $M_{\mathsf{F}} > 1$ antennas assisted by $K$ STAR-RISs (R) using NOMA, as shown in Fig. 1. The considered cell is assumed to be

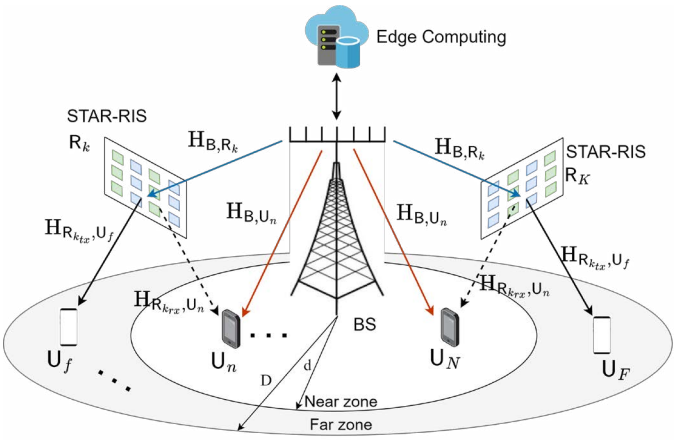



Fig. 1. The proposed multiple users massive MIMO-NOMA systems with multiple STAR-RISs.

divided virtually into two zones, the near zone and the far zone. The BS is located in the center of the cell and connected to edge computing for offline training deep learning model while STAR-RIS is deployed randomly at the edge of the near zone. A set $\mathcal{N} = \{\mathsf{U}_n, n = 1, \cdots, N\}$ of near users are deployed randomly in the near zone or front of the STAR-RIS, where a set $\mathcal{F} = \{\mathsf{U}_f, f = 1, \cdots, F\}$ of far users are deployed randomly in the far zone or behind STAR-RIS. Moreover, only $\mathsf{U}_n$ has direct links to BS, whereas no direct link between BS and $\mathsf{U}_f$ due to human or natural-made obstacles. It is important to note that the channel condition of near users in the near zone and far users in the far zone is different due to the radius of near zone d and far zone D. The STAR-RIS works in mode switching (MS) protocol, allowing each element to operate in either full reflection or transmission mode [14]. Thus, the STAR-RIS is equipped with two main parts: the first part is transmitting $L_t$, and the second part is reflecting $L_r$ elements to help the users in far and near zone, respectively. Besides that, we also consider the phase shift amplitude of transmission and reflection to be ideal (i.e., $\kappa_t = \kappa_r = 1$).

The signal from BS to $\mathsf{U}_n$ and $\mathsf{U}_f$ can be expressed as

$$\mathbf{x} = \sum_{n}^{N} \mathbf{W}_n \varsigma_n + \sum_{f}^{F} \mathbf{W}_f \varsigma_f, \tag{1}$$

where $\mathbf{W}_n$ and $\mathbf{W}_f \in \mathbb{C}^{M_{\mathrm{BS}} \times d}$ denote the linear precoding matrices from BS to $n$-th $\mathsf{U}_n$ and $f$-th $\mathsf{U}_f$, respectively; $\varsigma_n \in \mathbb{C}^{d \times 1}$ and $\varsigma_f \in \mathbb{C}^{d \times 1}$ denote the $(d \times 1)$-element downlink data signal for $n$-th $\mathsf{U}_n$ and $f$-th $\mathsf{U}_f$, respectively; and $d$ is the number of data streams transmitted for $\mathsf{U}_n$ and $\mathsf{U}_f$, with $1 \leq d \leq \min\{M_{\mathsf{BS}},\ M_N,\ M_F\}$.

The equivalent channel gains from BS to $\mathsf{U}_n$ and $\mathsf{U}_f$ ($\hat{\mathbf{H}}_{\mathsf{B},n}$, $\hat{\mathbf{H}}_{\mathsf{B},f}$) can be expressed, respectively, as

$$\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta) = \mathbf{H}_{\mathsf{B},\mathsf{U}_n} + \sum_{k}^{K} \mathbf{H}_{\mathsf{B},\mathsf{R}_k} \Theta_k \mathbf{H}_{\mathsf{R}_{k_r},\mathsf{U}_n}, \tag{2}$$

$$\hat{\mathbf{H}}_{\mathsf{B},f}(\Psi) = \sum_{k}^{K} \mathbf{H}_{\mathsf{B},\mathsf{R}_k} \Psi_k \mathbf{H}_{\mathsf{R}_{k_t},\mathsf{U}_f}, \tag{3}$$

where $\mathbf{H}_{\mathsf{R}_{k_r},\mathsf{U}_n} \in \mathbb{C}^{L_r \times M_{\mathsf{N}}}$, $\mathbf{H}_{\mathsf{B},\mathsf{R}_k} \in \mathbb{C}^{L_{\mathsf{X}} \times M_{\mathsf{BS}}}$, $\mathsf{X} \triangleq \{r, t\}$, $\mathbf{H}_{\mathsf{B},\mathsf{U}_n} \in \mathbb{C}^{M_{\mathsf{N}} \times M_{\mathsf{BS}}}$, and $\mathbf{H}_{\mathsf{R}_{k_t},\mathsf{U}_f} \in \mathbb{C}^{L_t \times M_{\mathsf{F}}}$ denote channel gain from $k$-th STAR-RIS with reflection elements to $\mathsf{U}_n$, BS to $k$-th STAR-RIS, BS to $\mathsf{U}_n$, $k$-th STAR-RIS with transmission elements to $\mathsf{U}_f$, respectively. $\Theta_k$ denotes the phase shift matrix of $k$-th STAR-RIS with reflection elements with the diagonal reflecting matrix being $\Theta_k = \mathrm{diag}(\kappa_1 \phi_{k,1},\ \kappa_2 \phi_{k,2}, \cdots,\ \kappa_r \phi_{k,L_r})$, where $\phi_{k,L_r} = e^{j\theta_{k,L_r}}$ with $\theta_{k,L_r} \in (0, 2\pi]$ denotes the phase shift of reflection elements on the $k$-th STAR-RIS. $\Psi_k$ denotes the phase shift matrix of $k$-th STAR-RIS with transmission elements with the diagonal transmission matrix being $\Psi_k = \mathrm{diag}(\kappa_1, \delta_{k,1},\ \kappa_2 \delta_{k,2}, \cdots,\ \kappa_t \delta_{k,L_t})$, where $\delta_{k,L_t} = e^{j\psi_{k,L_t}}$ with $\psi_{k,L_t} \in (0, 2\pi]$ denote the phase shift of transmission elements on the $k$-th STAR-RIS. The channel gains $\mathbf{H} = \sqrt{\rho_h}\tilde{\mathbf{H}}$, $\mathbf{H} \in \{\mathbf{H}_{\mathsf{B},\mathsf{U}_n}, \mathbf{H}_{\mathsf{B},\mathsf{R}_k}, \mathbf{H}_{\mathsf{R}_{k_r},\mathsf{U}_n}, \mathbf{H}_{\mathsf{R}_{k_t},\mathsf{U}_f}\}$, where $\rho_h$ denotes the large scale fading and $\tilde{\mathbf{H}}$ denotes the small scale fading matrix both generated with distribution $\mathcal{CN}(0, 1)$ [15].

The received signal at $\mathsf{U}_n$ and $\mathsf{U}_f$ can be expressed, respectively, as

$$\mathbf{y}_n = \hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_n\varsigma_n + \sum_{n'\in N\setminus\{n\}} \hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_{n'}\varsigma_{n'} + \sum_{f\in F} \hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_f\varsigma_f + \mathbf{n}_n, \quad (4)$$

$$\mathbf{y}_f = \hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_f\varsigma_f + \sum_{f'\in F\setminus\{f\}} \hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_{f'}\varsigma_{f'} + \sum_{n\in N} \hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_n\varsigma_n + \mathbf{n}_f, \quad (5)$$

where $\mathbf{n}_n$ and $\mathbf{n}_f$ are noise vector at $\mathsf{U}_n$ and $\mathsf{U}_f$ with variance $\sigma_n^2$ and $\sigma_f^2$ and zero-mean, respectively. By following the principle of successive interference cancellation (SIC), the SINR at $\mathsf{U}_n$ to eliminate the $\mathsf{U}_f$'s message can be expressed as

$$\gamma_n^{\varsigma_f}(\mathbf{W},\Theta) = \frac{\|\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_f\|^2}{\Phi_{n,f}(\mathbf{W},\Theta)}, \quad (6)$$

where

$$\Phi_{n,f}(\mathbf{W},\Theta) = \sum_{n\in N} \|\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_n\|^2 + \sum_{f'\in F\setminus\{f\}} \|\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_{f'}\|^2 + \sigma_n^2. \quad (7)$$

Furthermore, the $\mathsf{U}_n$ by using SINR can decode its own message, which can be expressed as

$$\gamma_n^{\varsigma_n}(\mathbf{W},\Theta) = \frac{\|\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_n\|^2}{\Xi_n(\mathbf{W},\Theta)}, \quad (8)$$

where

$$\Xi_n(\mathbf{W},\Theta) = \sum_{n'\in N\setminus\{n\}} \|\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_n\|^2 + \sum_{f\in F} \|\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_f\|^2 + \sigma_n^2. \quad (9)$$

The $\mathsf{U}_f$ by using SINR can decode its own message directly, which can be written as

$$\gamma_f^{\varsigma_f}(\mathbf{W},\Psi) = \frac{\|\hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_f\|^2}{\Phi_f(\mathbf{W},\Psi)}, \quad (10)$$

where

$$\Phi_f(\mathbf{W},\Psi) = \sum_{n\in N} \|\hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_n\|^2 + \sum_{f'\in F\setminus\{f\}} \|\hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_{f'}\|^2 + \sigma_f^2. \quad (11)$$

The downlink spectral efficiency of $\mathsf{U}_n$ and $\mathsf{U}_f$ in nat/sec/Hz can be written, respectively, as

$$\mathcal{R}_n(\mathbf{W},\Theta) = \ln\big(1+\gamma_n^{\varsigma_n}(\mathbf{W},\Theta)\big), \quad \forall n\in\mathcal{N}, \quad (12)$$

$$\mathcal{R}_f(\mathbf{W},\Psi) = \ln\big(1+\gamma_f^{\varsigma_f}(\mathbf{W},\Psi)\big), \quad \forall f\in\mathcal{F}. \quad (13)$$

### *B. Problem Formulation*

Let us consider the joint design of linear precoding matrix $\mathbf{W}$ and phase shift of transmission $\Theta_k$ and reflection $\Psi_k$ at STAR-RIS, which will be combined to maximize spectral efficiency. The goal is to maximize the spectral efficiency in the network subject to individual quality of services (QoS) and maximum power budget at $\mathsf{BS}$ by jointly optimizing the linear precoding matrix, phase shift reflection and transmission of IRSs, which can be formulated as

$$\max_{\mathbf{W},\Theta,\Psi} \ \mathcal{R}_\sum \triangleq \sum_{n\in\mathcal{N}} \mathcal{R}_n(\mathbf{W},\Theta) + \sum_{f\in\mathcal{F}} \mathcal{R}_f(\mathbf{W},\Psi), \quad (14a)$$

$$\text{s.t. } \mathcal{R}_n(\mathbf{W},\Theta) \geq \bar{\mathsf{R}}_n, \quad \forall n\in\mathsf{N}, \quad (14b)$$

$$\mathcal{R}_f(\mathbf{W},\Psi) \geq \bar{\mathsf{R}}_f, \quad \forall f\in\mathsf{F}, \quad (14c)$$

$$||\mathbf{W}_n||^2 + ||\mathbf{W}_f||^2 \leq P_{\mathrm{BS}}^{\max}, \quad (14d)$$

$$(\theta_{k,l_r}|0\leq\theta_{k,l_r}\leq 2\pi], \forall k\in K,\ l_r\in L_r, \quad (14e)$$

$$(\psi_{k,l_t}|0\leq\psi_{k,l_t}\leq 2\pi], \forall k\in K,\ l_t\in L_t, \quad (14f)$$

where (14b) and (14c) indicate the QoS of each user guaranteed by predefined thresholds for $\mathsf{U}_n$ $\bar{\mathsf{R}}_n \geq 0$ and $\mathsf{U}_f$ $\bar{\mathsf{R}}_f \geq 0$. Constraint (14d) indicates the total power of all UEs, which is constrained by the maximum power budget at the BS. Constraints (14e) and (14f) denote the phase shift of reflection and transmission elements of STAR-RIS that have discrete values. Clearly, the objective function (14) is non-convex subject to mixed-integer constraints. Since the precoding matrix, the phase shift of transmission and reflection at STAR-RIS are coupled, the formulated problem in (14) is difficult to solve optimally. Thus, we propose a low-complexity iterative algorithm to solve the formulated problem (14), which will be explained in the next section.

## III. Spectral Efficiency Maximization Algorithm

It is very challenging to solve problem (14) optimally since $\mathbf{W}$, $\Theta$ and $\Psi$ have a strong couple so that it is only possible to use exhaustive search to find the optimal solution for the problem (14) which has very high computational complexity. In fact, when the number of BS antennas, the number of STAR-RIS with the number of elements and the number of users with multiple antennas increase, the computational complexity value of exhaustive search will increase exponentially. Based on mentioned above, we are motivated to propose an algorithm to solve the formulated problem (14) with low computational complexity.

Firstly, the standard way to tackle the discrete properties in problem (14) to be more tractable is by relaxing the discrete properties into continuous properties. Thus, the relaxed problem of (14) can be rewrite as

$$\max_{\mathbf{W},\Theta,\Psi} \ \bar{\mathcal{R}}_\sum \triangleq \sum_{n\in\mathcal{N}} \mathcal{R}_n(\mathbf{W},\Theta) + \sum_{f\in\mathcal{F}} \mathcal{R}_f(\mathbf{W},\Psi), \quad (15a)$$

$$\theta_{k,l_r}\in[0,2\pi], \forall k\in K,\ l_r\in L_r, \quad (15b)$$

$$\psi_{k,l_t}\in[0,2\pi], \forall k\in K,\ l_t\in L_t, \quad (15c)$$

$$(14b), (14c), (14d). \quad (15d)$$

We introduce new variables $\Xi \triangleq \{\Xi_n, \Xi_f\}_{n\in\mathcal{N}, f\in\mathcal{F}}$ and $\mathbf{r} \triangleq \{r_n, r_f\}_{n\in\mathcal{N}, f\in\mathcal{F}}$, where $\Xi_n$ and $\Xi_f$, and $r_n$ and $r_f$ denote soft data rates and SINRs of $\mathsf{U}_n$ and $\mathsf{U}_f$, respectively. Next, we transform problem (15) into an equivalent non-convex one with a more tractable form. Thus, the equivalent of the problem (15) can be rewritten as

$$\max_{\mathbf{W},\Theta,\Psi,\Xi,\mathbf{r}} \ \bar{\mathcal{R}}_{\sum} \triangleq \sum_{n\in\mathcal{N}} r_n + \sum_{f\in\mathcal{F}} r_f, \tag{16a}$$

$$\gamma_n^{\varsigma_n}(\mathbf{W},\Theta) \geq 1/\Xi_n, \ \forall n \in \mathcal{N}, \tag{16b}$$

$$\gamma_f^{\varsigma_f}(\mathbf{W},\Psi) \geq 1/\Xi_f, \ \forall f \in \mathcal{F}, \tag{16c}$$

$$\ln(1+1/\Xi_n) \geq r_n, \ \forall n \in \mathcal{N}, \tag{16d}$$

$$\ln(1+1/\Xi_f) \geq r_f, \ \forall f \in \mathcal{F}, \tag{16e}$$

$$r_n \geq \bar{\mathcal{R}}_n, \ \forall n \in \mathcal{N}, \tag{16f}$$

$$r_f \geq \bar{\mathcal{R}}_f, \ \forall f \in \mathcal{F}, \tag{16g}$$

$$(14d), (15b), (15c). \tag{16h}$$

We introduce new variables $\zeta \in \{\zeta_n, \zeta_f\}$, $\zeta \triangleq \{\zeta \geq 0\}$ which satisfy the convex constraint $\|\hat{\mathbf{H}}_{\mathsf{B},n}(\Psi)\mathbf{W}_n\|^2 < \zeta_n$ and $\|\hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_f\|^2 < \zeta_f$. By following the IA method, the problem (16) can be approximated at $(\kappa+1)$ iteration as

$$\max_{\mathbf{W},\Theta,\Psi,\Xi,\mathbf{r},\zeta} \ \bar{\mathcal{R}}_{\sum}^{(\kappa)} \triangleq \sum_{n\in\mathcal{N}} r_n + \sum_{f\in\mathcal{F}} r_f, \tag{17a}$$

$$\|\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_n\|^2 < \zeta_n, \ \forall n \in \mathcal{N}, \tag{17b}$$

$$\|\hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_f\|^2 < \zeta_f, \ \forall f \in \mathcal{F}, \tag{17c}$$

$$\frac{\hat{\xi}_n^{(\kappa)}(\mathbf{W},\Theta,\zeta)}{\Xi_n} \leq f_n^{(\kappa)}(\mathbf{W},\Theta), \tag{17d}$$

$$\frac{\hat{\xi}_f^{(\kappa)}(\mathbf{W},\Theta,\zeta)}{\Xi_f} \leq f_f^{(\kappa)}(\mathbf{W},\Psi), \tag{17e}$$

$$\ln\left(1+\frac{1}{\Xi_n}\right) + \frac{1}{1+\Xi_n} - \frac{\Xi_n}{\Xi_n^{(\kappa)}(1+\Xi_n^{(\kappa)})} \geq r_n, \tag{17f}$$

$$\ln\left(1+\frac{1}{\Xi_f}\right) + \frac{1}{1+\Xi_f} - \frac{\Xi_f}{\Xi_f^{(\kappa)}(1+\Xi_f^{(\kappa)})} \geq r_f, \tag{17g}$$

$$(14d), (15b), (15c), (16f), (16g), \tag{17h}$$

where

$$2\Re\left\{(\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_n^{(\kappa)})^*(\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_n)\right\} - \|\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_n^{(\kappa)}\|^2 \leq \|\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_n\|^2 \triangleq f_n^{(\kappa)}(\mathbf{W},\Theta), \tag{18}$$

$$2\Re\left\{(\hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_n^{(\kappa)})^*(\hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_f)\right\} - \|\hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_f^{(\kappa)}\|^2 \leq \|\hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_f\|^2 \triangleq f_f^{(\kappa)}(\mathbf{W},\Psi), \tag{19}$$

$$\sum_{n'\in\mathcal{N}\setminus\{n\}} \|\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_{n'}\|^2 + \sigma_{\mathsf{n}}^2 + \sum_{f'\in\mathcal{F}} \zeta_{n,f'} \leq \|\hat{\mathbf{H}}_{\mathsf{B},n}(\Theta)\mathbf{W}_n\|^2 \triangleq \hat{\xi}_n(\mathbf{W},\Theta,\zeta), \tag{20}$$

$$\sum_{f'\in\mathcal{F}\setminus\{f\}} \|\hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_{f'}\|^2 + \sigma_f^2 + \sum_{n'\in\mathcal{N}} \zeta_{f,n'} \leq \|\hat{\mathbf{H}}_{\mathsf{B},f}(\Psi)\mathbf{W}_f\|^2 \triangleq \hat{\xi}_f(\mathbf{W},\Psi,\zeta). \tag{21}$$

When solving problem (17), we achieve the optimal solution of $\theta_{k,l_r}$ and $\psi_{k,l_t}$ which have non-integer values. So, the optimal solution obtained from problem (17) is not feasible for the original problem (14). To overcome this problem, we propose a rounding function after obtaining the optimal solution to the problem (17) as

$$\mathsf{Y}^\star = \lceil \mathsf{Y}^{(\kappa)} - 1/2 \rceil, \ \mathsf{Y} \in \{\Theta_k, \Psi_k\}, \ k \in \mathcal{K}. \tag{22}$$

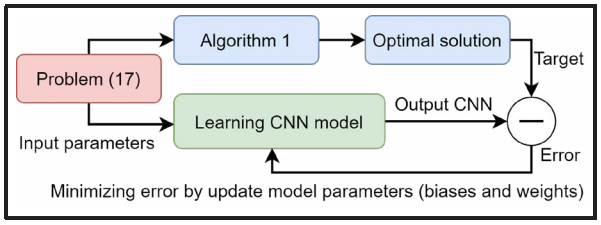
Finally, we summarize the proposed low-complexity iterative algorithm to solve the SEM problem (14) in Algorithm 1.

**Algorithm 1** Proposed IA Algorithm for solving spectral efficiency Maximization (14)

1: **Initialization:**
Set $(\mathbf{W}^\star, \Theta^\star, \Psi^\star) \leftarrow 0$, and generate initial feasible point with a random value $(\mathbf{W}^{(0)}, \Theta^{(0)}, \Psi^{(0)}\Xi^{(0)}, \zeta^{(0)})$ for (17).
2: **for** $\kappa = 0, 1, 2, \cdots$ **do**
3: Solve the convex problem (17), to find $(\mathbf{W}^\star, \Theta^\star, \Psi^\star \Xi^{(\star)}, \zeta^{(\star)})$;
4: Update $(\mathbf{W}^{(\kappa+1)}, \Theta^{(\kappa+1)}, \Psi^{(\kappa+1)}, \Xi^{(\kappa+1)}, \zeta^{(\kappa+1)})$ $\leftarrow (\mathbf{W}^\star, \Theta^\star, \Psi^\star, \Xi^{(\star)}, \zeta^{(\star)})$;
5: **if** Convergence **then**
6: Break;
7: **end if**
8: **end for**
9: Round up the optimal solution $(\Theta^\star, \Psi^\star)$ based on (22);
10: $(\mathbf{W}^\star, \Theta^\star, \Psi^\star) \leftarrow (\mathbf{W}^{(\kappa)}, \Theta^{(\kappa)}, \Psi^{(\kappa)})$;
11: Based on $(\mathbf{W}^\star, \Theta^\star, \Psi^\star)$ calculate $\mathcal{R}_{\sum}$ in (14a).
12: **Output:** Optimal solution $(\mathbf{W}^\star, \Theta^\star, \Psi^\star)$ and $\mathcal{R}_{\sum}$.

## IV. Deep Learning Design for the SEM Prediction

In this section, we introduce the proposed DL framework to predict the optimal solution of SEM consisting of offline training and online predicting, as shown in Fig 5. The proposed DL framework takes the CNN model as the underlying technology since the parameter sharing adopted in the CNN can reduce the number of the learned parameters compared to a fully connected DNN. Furthermore, CNN is well known to be effective for extracting features, which will benefit the predicting optimal solution of SEM problem by using the users'




(a) Training


(b) Predicting

Fig. 2. The proposed DL framework with CNN model for SEM problem

position information and channel gains. In Fig. 5(a), the DL learns to map between input parameters from problem (14) and target (optimal solution) obtained after Algorithm 1 solving the problem (14). During offline training, DL keeps optimizing the CNN model by updating model parameters (weights and biases) to minimize the error between the target and output CNN model. The target of the CNN model is generated from Algorithm 1. The trained CNN model is utilized as a mapping function to predict the optimal solution online via a quick inference process whenever the new parameters are available at the input layer of the CNN model, as shown in Fig. 5(b). Thus, the computational complexity is transferred to offline computational complexity for training. After offline training completed, the computational complexity is transferred to online training, which facilitates the lightweight online execution of the proposed optimization algorithm.

We consider the input parameters of the DL framework consisting of users' position information (geographic and zone), $\mathsf{U}_n^{\mathsf{p}}$, $\mathsf{U}_f^{\mathsf{p}}$, $\mathsf{U}_n^z$, $\mathsf{U}_f^z$, and the channel gains from BS to near users, $\mathbf{H}_{\mathsf{B},\mathsf{U}_n}$, from BS to $k$-th STAR-RIS, $\mathbf{H}_{\mathsf{B},\mathsf{R}_k}$ and from STAR-RIS to users, $\mathbf{H}_{\mathsf{R}_k,\mathsf{U}_n}$, $\mathbf{H}_{\mathsf{R}_k,\mathsf{U}_f}$. Based on these settings, the deep CNN model can predict the optimal solution of the SEM problem in real-time.

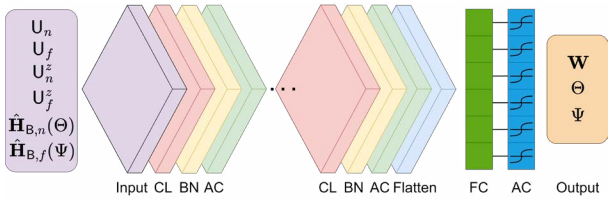



Fig. 3. The proposed CNN model to predict SEM problem.

Fig. 3 illustrates the CNN model to approximate the mapping function from users' position information and channel gains to predict the optimal solution of the SEM problem. The proposed CNN model includes an input layer, batch normalization (BN) layers, convolutional (CL) layers, a flatten layer, activation (AC) layers, and a fully connected (FC) layer. In detail, every CL layer comprises several filters that apply convolutional operations on the input layer, capture the input patterns, and pass the output to the next layer. A variety of channel coefficients shares the filter's parameters. The primary purpose of the BN layers is to use two trainable parameters to normalize the output of the CL layers (i.e., mean and standard deviation). Additionally, the BN layers can enable a higher learning rate also decreasing the probability of over-fitting while AC layers assist the neural networks in extracting useful information and suppressing the input data's insignificant points. In our proposed CNN model, since the predicted variables are continuous and positive values, we consider the rectified linear unit (ReLU) is appropriate for the AC layer. In order for the FC layer to understand the input, the flatten layer must change the input's shape into a vector.

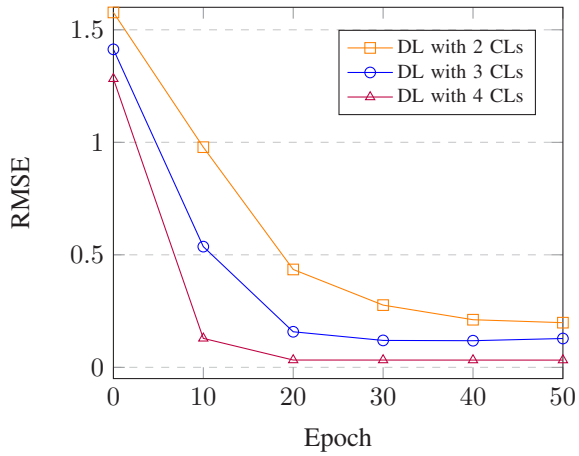


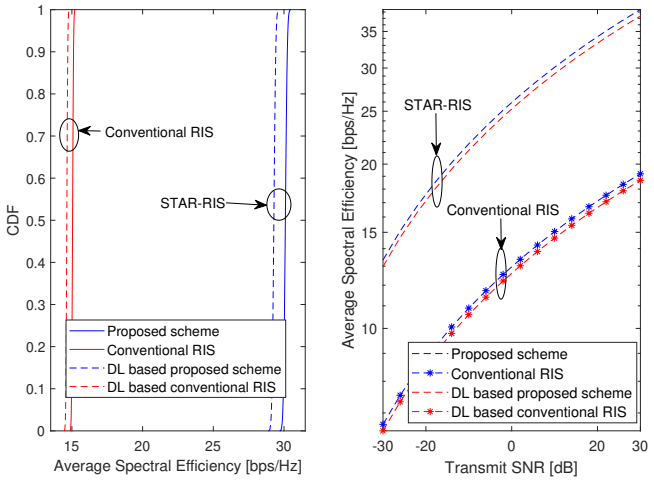

Fig. 4. Impact of the epoch on the DL with a variation of CLs for training.

## V. Numerical Results and Discussions

This section evaluates the numerical results of the proposed scheme and deep learning approach. Simulation parameters are set as follows: D = 100 m, d = 50 m, $L_t = L_r = 32$, $\bar{\mathsf{R}}_n = \bar{\mathsf{R}}_f = 1$ bps/Hz, $M_{\mathsf{BS}} = 100$, $M_{\mathsf{N}} = M_{\mathsf{F}} = 10$, $\mathcal{N} = \mathcal{F} = 4$, $K = 5$ and SNR = [-30:10:30] dB. The result of SE is obtained in nats/sec/Hz, so to get bps/Hz we divide it by ln(2). To solve the convex problem (17), we consider CVX optimization toolbox and SDPT3 convex solver. We generate a 20K dataset and divide it into the training and test sets with a ratio of 90:10. For performance comparison, we consider conventional RIS at the same location as STAR-RIS, with each RIS consisting of $L_r/2$ elements to have the same coverage and a fair comparison with STAR-RIS. The proposed deep CNN model consists of one input, output, FC, and flatten layer, two BN layers, three AC layer, and multiple CL layers. Besides, every CL layer has eight kernels with the size of $3\times3$ while each AC layer uses the ReLU function.

In Fig. 4, we analyze the impact of the epoch on the DL training loss function with a variation of CLs. The training value's root means square error (RMSE) will decrease from 1.57 to 0.19, 1.4 to 0.12, and 1.28 to 0.03 of DL with 2, 3 and 4 CLs when the epoch value increases from 0 to 50, respectively. The reason is that the DL can update the weights and biases with each additional epoch [16].

Fig. 5(a) shows the cumulative distribution function (CDF) of the average SE of the system. As can be observed, the proposed scheme outperforms the conventional RIS scheme since it has twice the number of elements to serve the users compared to the conventional RIS. Besides, the DL can predict the SE with an accuracy of up to 96.7%. In Fig. 5(b), we show the impact of the transmit SNR on the average SE. As can be observed, the average SE will increase when the transmit SNR increases. The proposed scheme has a better SE than the conventional RIS; one of the reasons is that the proposed scheme effectively has a larger element to improve the signal quality received by users. Again, the DL can predict SE with good accuracy.




(a) (b)

Fig. 5. CDF of the average SE and average SE versus SNR

TABLE I
AVERAGE EXECUTION TIME OF CONVENTIONAL APPROACH VERSUS DL APPROACH UNDER VARIATION NUMBER OF USERS

| Number of UE | 8 | 12 | 16 | 20 |
|---|---|---|---|---|
| Conventional [minutes] | 3 | 9 | 22 | 81 |
| DL [seconds] | 1 | 1 | 1 | 1 |

Finally, the average execution time of the conventional approach comparing with the DL approach is shown in Table I. These results show that the DL approach requires a very short execution time even the number of users increases. On the other hand, the conventional approach requires a longer execution time under the same condition. The reason is that the DL approach uses a mapping function to predict the optimal solution to be able to predict the optimal solution quickly. However, in the conventional approach, when running Algorithm 1 requires multiple iterations to achieve optimal solution, so it spends much time compared to the DL approach.

## VI. CONCLUSIONS

This paper investigated deep learning-based multiple users multiple STAR-RISs massive MIMO-NOMA networks to improve SE in the system. We considered a joint design optimization of the precoding matrix, phase shift matrix of reflection and transmission element at the STAR-RIS subject to the individual QoS of each user, maximum power budget at BS and the discrete phase shift. Moreover, the low-complexity iterative algorithm was proposed to solve this problem with guaranteed convergence at a relatively optimal level. To support real-time optimization, a DL framework with deep CNN was proposed to predict the optimal value of SE according to users' position information and the channel gains every link. Furthermore, the proposed approach improved SE over massive MIMO with conventional RIS. Additionally, the DL approach provided a better capability to predict the optimal value under the suggested approach.

## ACKNOWLEDGMENT

This work was supported by National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2022R1A2B5B01001190).
Prof. Beongku An is the corresponding author.

## REFERENCES

[1] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjoland, and F. Tufvesson, "6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities," *Proc. IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021.

[2] R. H. Y. Perdana, T.-V. Nguyen, and B. An, "Deep Learning-based Power Allocation in Massive MIMO Systems with SLNR and SINR Criterions," in *2021 Twelfth Int. Conf. Ubiquitous Futur. Networks*. Jeju Island, Republic of Korea: IEEE, 2021, pp. 87–92.

[3] R. H. Y. Perdana, T. V. Nguyen, and B. An, "Deep neural network design with SLNR and SINR criterions for downlink power allocation in multi-cell multi-user massive MIMO systems," *ICT Express*, 2022.

[4] W. A. Al-Hussaibi and F. H. Ali, "Efficient User Clustering, Receive Antenna Selection, and Power Allocation Algorithms for Massive MIMO-NOMA Systems," *IEEE Access*, vol. 7, pp. 31 865–31 882, 2019.

[5] R. H. Y. Perdana, T. V. Nguyen, Y. Pramitarini, K. Shim, and B. An, "Deep Learning-based Spectral Efficiency Maximization in Massive MIMO-NOMA Systems with STAR-RIS," in *2023 Int. Conf. Artif. Intell. Inf. Commun.* Bali, Indonesia: IEEE, 2023, pp. 644–649.

[6] L. Bai, Z. Huang, Y. Li, and X. Cheng, "A 3D Cluster-Based Channel Model for 5G and beyond Vehicle-to-Vehicle Massive MIMO Channels," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 8401–8414, 2021.

[7] Y. Wang, P. Guan, H. Yu, and Y. Zhao, "Transmit Power Optimization of Simultaneous Transmission and Reflection RIS Assisted Full-Duplex Communications," *IEEE Access*, vol. 10, pp. 61 192–61 200, 2022.

[8] K. Shim, T. N. Do, T. V. Nguyen, D. B. da Costa, and B. An, "Enhancing PHY-Security of FD-Enabled NOMA Systems Using Jamming and User Selection: Performance Analysis and DNN Evaluation," *IEEE Internet Things J.*, vol. 8, no. 24, pp. 17 476–17 494, 2021.

[9] T. V. Nguyen, V. D. Nguyen, D. B. Da Costa, and B. An, "Short-Packet Communications in Multi-Hop WPINs: Performance Analysis and Deep Learning Design," in *2021 IEEE Glob. Commun. Conf. GLOBECOM 2021 - Proc.* IEEE, 2021, pp. 7–12.

[10] J. Zhang, W. Xia, M. You, G. Zheng, S. Lambotharan, and K. K. Wong, "Deep Learning Enabled Optimization of Downlink Beamforming under Per-Antenna Power Constraints: Algorithms and Experimental Demonstration," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 6, pp. 3738–3752, 2020.

[11] T. Van Chien, T. Nguyen Canh, E. Bjornson, and E. G. Larsson, "Power Control in Cellular Massive MIMO with Varying User Activity: A Deep Learning Solution," *IEEE Trans. Wirel. Commun.*, vol. 19, no. 9, pp. 5732–5748, 2020.

[12] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep Learning-Based Channel Estimation," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 652–655, 2019.

[13] T.-V. Nguyen, V.-D. Nguyen, D. B. D. Costa, H.-T. Thien, R. Q. Hu, and B. An, "Short-Packet Communications in Multi-Hop Networks with WET: Performance Analysis and Deep Learning-Aided Optimization," *IEEE Trans. Wirel. Commun.*, pp. 1–27, 2022.

[14] T. Wang, M. A. Badiu, G. Chen, and J. P. Coon, "Outage Probability Analysis of STAR-RIS Assisted NOMA Network with Correlated Channels," *IEEE Commun. Lett.*, vol. 7798, 2022.

[15] 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.

[16] T.-V. Nguyen, T.-N. Tran, K. Shim, T. Huynh-The, and B. An, "A Deep-Neural-Network-Based Relay Selection Scheme in Wireless-Powered Cognitive IoT Networks," *IEEE Internet Things J.*, vol. 8, no. 9, pp. 7423–7436, 2021.