

RIS-Assisted XL-MIMO for Near-Field and Far-Field Communications

Xiaomin Cao, Mohammadali Mohammadi, *Senior Member, IEEE*, Hien Quoc Ngo, *Fellow, IEEE*, Hyundong Shin, *Fellow, IEEE*, and Michail Matthaiou, *Fellow, IEEE*

Abstract—We consider a reconfigurable intelligent surface (RIS)-assisted extremely large-scale multiple-input multiple-output (XL-MIMO) downlink system, where an XL-MIMO array serves two groups of single-antennas users, namely near-field users (NFUEs) and far-field users (FFUEs). FFUEs are subject to blockage, and their communication is facilitated through the RIS. We consider three precoding schemes at the XL-MIMO array, namely central zero-forcing (CZF), local zero-forcing (LZF) and maximum ratio transmission (MRT). Closed-form expressions for the spectral efficiency (SE) of all users are derived for MRT precoding, while statistical-form expressions are obtained for CZF and LZF processing. A heuristic visibility region (VR) selection algorithm is also introduced to help reduce the computational complexity of the precoding scheme. Furthermore, we devise a two-stage phase shifts design and power control algorithm to maximize the sum of weighted minimum SE of two groups of users with CZF, LZF and MRT precoding schemes. The simulation results indicate that, when equal priority is given to NFUEs and FFUEs, the proposed design improves the sum of the weighted minimum SE by 31.9%, 37.8%, and 119.2% with CZF, LZF, and MRT, respectively, compared to the case with equal power allocation and random phase shifts design. CZF achieves the best performance, while LZF offers comparable results with lower complexity. When prioritizing NFUEs or FFUEs, LZF achieves strong performance for the prioritized group, whereas CZF ensures balanced performance between NFUEs and FFUEs.

Index Terms—Extremely large-scale multiple-input multiple-output (XL-MIMO), near-field communications, power control, reconfigurable intelligent surface (RIS), visibility region (VR).

I. INTRODUCTION

To meet the growing demands for high data rates and massive connectivity beyond fifth generation (5G) and sixth generation (6G) networks, extremely large-scale multiple input multiple output (XL-MIMO), as one of the key enablers, has

received significant attention [1]. By increasing the number of antennas by at least an order of magnitude — ranging from several hundred to thousands—XL-MIMO represents a major advance over massive MIMO, enabling high spectral efficiency (SE), energy efficiency (EE) and reliable massive connectivity [1], [2]. Moreover, significant strides have been made in XL-MIMO hardware implementations; for instance, according to [3], a reconfigurable holographic surface-aided platform is capable of supporting the real-time transmission of high-definition video. In [4], the authors concluded that modular XL-MIMO can provide better spatial resolution by availing of its larger array aperture, paving the way for practical large-scale deployments. Despite these impressive advances in XL-MIMO, the signal received by some users may be blocked by the presence of large objects remains unsolved.

We address this problem by introducing a reconfigurable intelligent surface (RIS) which is a planar array equipped with a large number of low-cost passive elements, and each element is able to induce a certain phase shift [5]. Recent breakthroughs have been made in RIS hardware advancements. For example, in [6], the authors proposed the concept of active RISs and focused on signal verification through experimental measurements based on a fabricated active RIS element, which revealed that active RISs can achieve a significant sum-rate gain compared to passive RISs. The paper [7] studied a simultaneously transmitting and reflecting RIS-aided downlink communication system. In [8], the authors explored novel modeling, architecture designs, and optimization schemes for beyond diagonal RIS based on graph theory. Leveraging the capability of RISs to customize the wireless propagation environment, a promising approach is to integrate RISs with XL-MIMO systems. RIS-assisted XL-MIMO systems are expected to harness the combined benefits of both technologies, especially when direct base station (BS)-user links are obstructed [9]–[11]. Specifically, the authors of [9] considered the presence of a RIS to address the challenge of non-line-of-sight (NLoS) scenarios in XL-MIMO systems, while the authors of [10] studied the channel estimation problem for extremely large-scale RIS (XL-RIS) assisted multi-user XL-MIMO systems with hybrid beamforming structures.

Although RIS-assisted XL-MIMO systems provide a promising solution to solve the blockage problem, the increased dimensionality of XL-MIMO will notably alter the electromagnetic (EM) characteristics. Specifically, near-field (NF) propagation dominates as the array size becomes comparable to or larger than the user distance. The distinction between NF and far-field (FF) users (FFUEs) affects the

This work was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) grant (EP/X04047X/2) for TITAN Telecoms Hub. The work of X. Cao was supported in part by the China Scholarship Council. The work of H. Q. Ngo was supported by the U.K. Research and Innovation Future Leaders Fellowships under Grant MR/X010635/1, and a research grant from the Department for the Economy Northern Ireland under the US-Ireland R&D Partnership Programme. The work of M. Matthaiou was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101001331). An earlier version of this paper was presented in part at the 2024 IEEE WCNC [DOI: 10.1109/WCNC57260.2024.10570871]. The associate editor coordinating the review of this article and approving it for publication was Qingqing Wu. (*Corresponding authors: Michail Matthaiou; Hyundong Shin.*)

X. Cao, M. Mohammadi, H. Q. Ngo, and M. Matthaiou are with the Centre for Wireless Innovation (CWI), Queen's University Belfast, BT3 9DT Belfast, U.K. (email: {xcao13, m.mohammadi, hien.ngo, m.matthaiou} @qub.ac.uk.)

H. Q. Ngo, M. Matthaiou, and H. Shin are affiliated with the Department of Electronic Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, Republic of Korea (e-mail: hshin@khu.ac.kr).

choice of the channel model, which is determined based on the Rayleigh distance [12]. Users closer than the Rayleigh distance are modeled using NF (spherical wave) channels, whereas others are modeled using FF (planar wave) models. The precoding design in our work is based on the channel matrix and does not explicitly depend on whether a user is in the NF or FF region. Once the channel is defined, the precoding scheme is applied accordingly.

Motivated by the advantages of XL-MIMO, studies on NF behaviors [13]–[15] highlight the limitations of conventional models, such as the uniform plane wave (UPW) model, advocating for a uniform spherical wave (USW) model [13]. Three fundamental properties should be taken into account when the receiver is in the NF region of the array, which are the distances to the elements, the effective areas of the array, and the losses due to polarization mismatch [14], [15]. However, the extra-large aperture causes each user/scatterer to be observable within a limited visibility region (VR) [16]. Consequently, channel statistics vary along the array, leading to the non-stationarity of XL-MIMO. Nevertheless, the above works have not accounted for this non-stationarity. The main causes of spatial non-stationarity in NF channels are as follows: (i) *The inherent nonlinearity* in the signal space of NF line-of-sight (LoS) paths [17], [18]; (ii) *Potential partial blockage* of NF LoS components [19]; (iii) *Spatially varying scattering environments* across different subarrays [20]. In this study, our consideration of spatial non-stationarity is primarily motivated by the nonlinear phase and amplitude variations across the array, as well as the variation in wave incident angles [13]. To effectively capture and mitigate these effects, we propose a practical and lightweight VR selection algorithm that is well suited for addressing the challenges posed by spatial non-stationarity in NF XL-MIMO systems.

The impact of VR introduced by XL-MIMO arrays was considered in [4], [17], [20]–[26]. In [22], drawing inspiration from coded random access, the authors proposed a decentralized receiver with very low complexity, where processing is executed locally per subarray. The authors introduced a VR detection algorithm and utilized the acquired VR information to design a low-complexity symbol detection scheme in an uplink multi-user scenario in [17]. The authors of [23] proposed four antenna selection approaches in XL-MIMO systems to maximize the total EE. Two novel precoding schemes were proposed to reduce the complexity, namely mean angle-based zero forcing and tensor zero forcing in [24]. In addition, in [25], the authors developed a greedy joint user and subarray scheduling algorithm based on statistical channel state information (CSI) to maximize the achievable SE in the system. In addition, the channel characteristics vary along with the array, leading to non-stationarity. Despite the non-stationarity problem that can be addressed by adopting VRs, the fair resource allocation in the RIS-assisted XL-MIMO system is still unsolved. To the authors' best knowledge, no existing system tackles the non-stationarity and power allocation problem simultaneously.

To fill this gap, we perform a comprehensive analysis of the RIS-aided XL-MIMO system, considering VR selection, phase shift design, and power control with three precoding designs.

Although power allocation algorithms have been extensively studied in massive MIMO [27], [28], little research has focused on XL-MIMO systems. In [29], joint antenna selection and power allocation were optimized to maximize the downlink SE under different loading conditions. In [30], a multi-agent reinforcement learning-based power control algorithm was proposed for cell-free XL-MIMO.

Due to the large aperture size of XL-MIMO arrays and the use of high-frequency bands in future wireless networks, users are likely to reside in either the NF or FF regions. This calls for the adoption of appropriate channel models tailored to different user locations. Consequently, XL-MIMO systems differ significantly from conventional massive MIMO in terms of both system modeling and channel characteristics. In particular, the presence of NF users and the spatially non-stationary nature of XL-MIMO channels can have a pronounced impact on system design. Under such conditions, the extent to which the performance can be improved through optimized system design, such as RIS phase shifts optimization and adaptive power control, remains uncertain. To address this challenge, we adopt a comprehensive channel model that captures both NF and FF propagation effects. This gap motivates our study, with key contributions summarized as follows:

- We consider three linear precoding schemes—maximum ratio transmission (MRT), local zero-forcing (LZF), and central zero-forcing (CZF)—for a RIS-assisted XL-MIMO downlink communication system with NF users (NFUEs) and FFUEs. Rigorous closed-form SE lower bounds are derived for the MRT precoding scheme under VR selection for both NFUEs and FFUEs. Additionally, statistical lower bounds on the SE are obtained for the LZF and CZF schemes with selected VRs. In addition, we also consider the interference within two groups, which is practical and innovative.
- We formulate a joint optimization problem that integrates VR selection, phase shifts design, and power control for MRT, LZF, and CZF precoding schemes to provide a uniformly good service within each group of users. Our objective is to maximize the sum of the weighted minimum SE of both NFUEs and FFUEs, while ensuring that all users meet their quality-of-service (QoS) requirements.
- The highly coupled non-convex optimization problem is transformed into a tractable form with decoupled optimization variables. To achieve this, we propose a heuristic algorithm for VR selection, which selects effective VRs for each user. Then, we devise a two-stage low-complexity algorithm by decoupling the non-convex problem into two subproblems. In the first stage, we propose an algorithm for optimized phase shifts determination, leveraging the penalty method and successive convex approximation (SCA). In the second stage, we introduce a SCA-based power control algorithm.
- Our numerical results reveal that 1) The optimized phase shifts and power allocation coefficients lead to a substantial improvement in the sum of the weighted minimum SE of NFUEs and FFUEs; 2) CZF is generally preferable

TABLE I: Contrasting our contributions to the XL-MIMO literature

Contributions	This paper	[21]	[9]	[10]	[11]	[17]	[20]	[22]	[23]	[24]	[25]	[29]	[30]
VR design	✓	✓	×	×	×	✓	✓	✓	✓	✓	✓	✓	×
Power allocation	✓	×	×	×	×	×	×	×	×	×	×	✓	✓
Mixed NF and FF users	✓	✓	×	×	×	×	×	×	×	×	×	×	×
RIS	✓	✓	✓	✓	✓	×	×	×	×	×	×	×	×

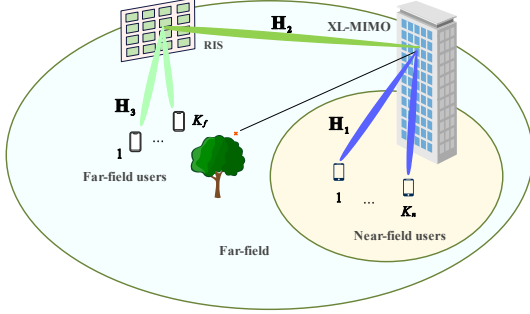


Fig. 1: Illustration of RIS-assisted XL-MIMO system.

while LZF achieves nearly equivalent performance with low complexity when giving equal priority to the NFUEs and FFUEs; 3) When different priorities are assigned to NFUEs and FFUEs, LZF delivers satisfactory performance for the prioritized users, while CZF achieves a balanced performance between NFUEs and FFUEs.

A comparison of our contributions with the state of the art in the field of XL-MIMO is tabulated in Table I.

Notation: We use bold lowercase letters (uppercase) to denote vectors (matrices). The superscripts $(\cdot)^T$ and $(\cdot)^H$ stand for the transpose and Hermitian, respectively; $\lfloor x \rfloor$ denotes the largest integer less than or equal to x ; $[\mathbf{A}]_{(:,j)}$ denotes the j th column of matrix $\mathbf{A}$; $\mathbf{0}_M$ and $\mathbf{0}_{M \times M}$ denote the all-zero vector of size M and all-zero matrix of size $M \times M$, respectively; $\text{diag}(\mathbf{a})$ denotes a diagonal matrix with each diagonal element being the corresponding element in $\mathbf{a}$ and $\text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_n)$ denotes a block-diagonal matrix with the square matrices $\mathbf{A}_1, \dots, \mathbf{A}_n$ on the diagonal. The trace and Kronecker product operations are denoted by $\text{tr}(\cdot)$ and $\otimes$, respectively; $\|\mathbf{A}\|_* = \sum_i \sigma_i(\mathbf{A})$ and $\|\mathbf{A}\|_2 = \sigma_1(\mathbf{A})$ are the nuclear norm and spectral norm, respectively, while $\sigma_i(\mathbf{A})$ denotes the i th largest singular value of matrix $\mathbf{A}$; $\mathbf{u}(\mathbf{A})$ is the eigenvector of $\mathbf{A}$ corresponding to the largest eigenvalue. Moreover, $\mathcal{CN}(0, \sigma^2)$ denotes a complex circularly symmetric Gaussian random variable (RV) with variance σ^2 ; $\mathbb{E}\{\cdot\}$ denotes the statistical expectation; $\delta_{ss'}$ denotes the Kronecker delta, with $\delta_{ss'} = 1$ if $s = s'$, and 0 otherwise. We use $|\mathcal{A}|$ to denote the cardinality of the set $\mathcal{A}$.

II. SYSTEM MODEL

We consider a downlink communication system in which an XL-MIMO array with M antennas serves $K = K_n + K_f$ users, where K_n users are in the NF region, referred to as NFUEs, and K_f users are in the FF region, referred to as FFUEs. For simplicity of notation, we define the sets $\mathcal{K}_N \triangleq \{1, \dots, K_n\}$, $\mathcal{K}_F \triangleq \{1, \dots, K_f\}$ to represent the sets of NFUEs, FFUEs respectively. Transmission to FFUEs is blocked by obstacles, and then a passive RIS is deployed in the system, whereas paths involving multiple interactions with the RIS—such as

signals reflecting off the RIS, then another surface, and back to the RIS—are ignored. The number of reflecting units at the RIS is $N = N_1 N_2$, where N_1 and N_2 are the numbers of elements in rows and columns, respectively. For the XL-MIMO array, we consider a uniform planar array (UPA) with $M = M_x M_y$ antennas, while M_x and M_y denote the number of antennas along the x - and y -axis, respectively. Furthermore, to address NF spatial non-stationary in XL-MIMO systems, the array is divided into S subarrays, where each subarray contains M^* antenna elements, that is, $M^* = M/S$. We set $M^* = M_x^* M_y^*$, where M_x^* and M_y^* are the number of antennas along the x and y directions, respectively. We assume all users are equipped with a single antenna and, to quantify the theoretical upper bound on system performance, we consider perfect knowledge of CSI for all involved channels [13], [14], [17].

A. Near-Field Channel Model of XL-MIMO

We consider a discrete aperture array model with hypothetical isotropic antenna elements. The effective area of each element is given by $\sqrt{A} \times \sqrt{A}$, where $A = e_a \frac{\lambda^2}{4\pi}$ [13], λ is the wavelength and $0 < e_a \leq 1$ denotes the aperture efficiency of each antenna element. Without loss of generality and for simplicity, we assume $e_a = 1$. The inter-element distance is $d = \frac{\lambda}{2}$, while the occupation ratio of the array is defined as $\beta = \frac{A}{d^2} \leq 1$, representing the fraction of the total UPA area occupied by the antenna elements. For convenience, we assume that M_x and M_y are odd numbers, and the location of central point of the (m_x, m_y) th element is $\mathbf{s}_{m_x, m_y} = [m_x d, m_y d, 0]^T$, where $m_x = 0, \pm 1, \dots, \pm \frac{M_x-1}{2}$, $m_y = 0, \pm 1, \dots, \pm \frac{M_y-1}{2}$, respectively. We assume that $\mathbf{u}_k = [u_{k,x}, u_{k,y}, u_{k,z}]^T$ is the location of the k th NFUE. Then, the distance between the k th NFUE and the center of the (m_x, m_y) th element is

$$r_{m_x, m_y, k} = \sqrt{(u_{k,x} - m_x d)^2 + (u_{k,y} - m_y d)^2 + u_{k,z}^2}. \quad (1)$$

The channel from the (m_x, m_y) th element to the k th NFUE, which considers free-space path loss, effective aperture, and polarization mismatch, can be expressed as [31, Eq. (11)], [17]

$$\bar{g}_{m_x, m_y, k} = \sqrt{\eta_{m_x, m_y, k}} e^{-j \frac{2\pi}{\lambda} r_{m_x, m_y, k}} + \sum_{l=1}^{L_k} \tilde{\beta}_{k,l} \sqrt{\eta_{m_x, m_y, l}} e^{-j \frac{2\pi}{\lambda} d_{m_x, m_y, l}}, \quad \forall m_x, m_y, \quad (2)$$

where $\eta_{m_x, m_y, k}$ is the power gain of the channel, which can be written as [17, Eq. (9)]

$$\eta_{m_x, m_y, k} \approx \frac{A}{4\pi} \frac{u_{k,z} \left((u_{k,x} - m_x d)^2 + u_{k,z}^2 \right)}{\left[(u_{k,x} - m_x d)^2 + (u_{k,y} - m_y d)^2 + u_{k,z}^2 \right]^{\frac{5}{2}}}, \quad (3)$$

and $\eta_{m_x, m_y, l}$ can be obtained as $\eta_{m_x, m_y, k}$ in a similar manner. We assume that the scatterer l is located at $\mathbf{p}_{k,l} = [p_{k,l,x}, p_{k,l,y}, p_{k,l,z}]^T$, and $d_{m,l} = \|\mathbf{p}_{k,l} - \mathbf{s}_{m_x, m_y}\|$. Moreover,

TABLE II: List of main symbols and the physical meanings

M	Number of antennas at the XL-MIMO	N	Number of elements at
K_n	Number of NFUEs	K_f	Number of FFUEs
P_n	Maximum transmit power for NFUEs	P_f	Maximum transmit power for FFUEs
$\mathbf{H}_1$	Channel between XL-MIMO and all NFUEs	$\mathbf{\Theta}$	Phase shifts matrix of RIS
$\mathbf{g}_k$	Channel between XL-MIMO and NFUE k	$\mathbf{g}_{sk}$	Channel between s th subarray and NFUE k
$\mathbf{G}$	The channel between RIS and all K_f FFUEs	$\mathbf{h}_k$	Channel between RIS and FFUE k
$\mathbf{H}_2$	Channel from XL-MIMO to RIS	$\mathbf{H}_{2,s}$	Channel from s th subarray of XL-MIMO to RIS
$\mathbf{w}_{sk}$	Precoding vector at subarray s to NFUE k	$\mathbf{w}_{sj}$	Precoding vector at subarray s to FFUE j
$\mathbf{D}_{sk}$	Indicator of subarray s 's active antennas for NFUE k	$\mathbf{D}_{sj}$	Indicator of subarray s 's active antennas for FFUE j
$\tilde{\eta}_k$	Power coefficients allocated to NFUE k	$\tilde{\eta}_j$	Power coefficients allocated to FFUE j
$\tilde{\eta}_{sk}$	Power coefficients at subarray s allocated to NFUE k	$\tilde{\eta}_{sj}$	Power coefficients at subarray s for FFUE j

in (2), L_k is the number of dominant single bounce scatterers visible from NFUE k , $\tilde{\beta}_{k,l}$ includes the impact of the random reflection coefficients of the l th scatterer and channel gain from the scatterer l to NFUE k . It is assumed that the random phase of $\tilde{\beta}_{k,l}$ is independent and identically distributed and uniformly distributed in $(-\pi, \pi]$ [31].

Note that we consider that the channels between XL-MIMO and NFUEs are LoS channels in this paper, since they have more impact on the received signal of NFUEs.

Let $\mathbf{g}_k = [\mathbf{g}_{1k}, \dots, \mathbf{g}_{S_k}]^T \in \mathbb{C}^{M \times 1}$ denote the overall channel vector between the XL-MIMO array and NFUE k , where $\mathbf{g}_{sk} = [\bar{g}_{(s-1)M^*+1,k}, \dots, \bar{g}_{sM^*,k}]^T \in \mathbb{C}^{M^* \times 1}$ represents the channel between the s th subarray of the XL-MIMO system and NFUE k , while $\bar{g}_{m,k}$ can be obtained from (2) by calculating the coordinates (m_x, m_y) of the m th element via the following expressions:

$$m_x = -0.5(M_x - 1) + \text{mod}(m - 1, M_x), \quad (4a)$$

$$m_y = 0.5(M_y - 1) - \left\lfloor \frac{m - 1}{M_x} \right\rfloor. \quad (4b)$$

Moreover, the channel between the XL-MIMO array and all NFUEs is expressed as $\mathbf{H}_1^H = [\mathbf{g}_1, \dots, \mathbf{g}_{K_n}]^H \in \mathbb{C}^{K_n \times M}$.

B. Far-Field Channel Model of XL-MIMO

The channels from the XL-MIMO to the RIS¹ and from the RIS to FFUE k are denoted by $\mathbf{H}_2 \in \mathbb{C}^{N \times M}$ and $\mathbf{h}_k^H \in \mathbb{C}^{1 \times N}$, respectively. Furthermore, the cascaded channel from the XL-MIMO to FFUE k through the RIS is expressed as $\tilde{\mathbf{g}}_k^H = \mathbf{h}_k^H \mathbf{\Theta} \mathbf{H}_2 \in \mathbb{C}^{1 \times M}$, where $\mathbf{\Theta} = \text{diag}(e^{j\theta_1}, \dots, e^{j\theta_N})$, with $\theta_n \in [0, 2\pi)$ representing the phase shift of n th reflecting element, and we set $\boldsymbol{\theta} = [\theta_1, \dots, \theta_N]$. The cascaded channels of all K_f FFUEs are given by $\mathbf{G}^H = [(\tilde{\mathbf{g}}_1^H)^T, \dots, (\tilde{\mathbf{g}}_{K_f}^H)^T]^H \in \mathbb{C}^{K_f \times M}$. Since the RIS is often installed on the facade of a tall building, both LoS and NLoS transmission paths would exist in the XL-MIMO-to-RIS link. Thus, we employ the Ricean fading model for the link from the XL-MIMO to the RIS as

$$\mathbf{H}_2 = \alpha_2 \bar{\mathbf{H}}_2 + \beta_2 \tilde{\mathbf{H}}_2, \quad (5)$$

where $\alpha_2 = \sqrt{\frac{\iota\zeta}{\iota+1}}$, $\beta_2 = \sqrt{\frac{\zeta}{\iota+1}}$, while ζ is the path-loss coefficient corresponding to the link from XL-MIMO array

to the RIS; ι denotes the corresponding Ricean factor; $\bar{\mathbf{H}}_2$ denotes the LoS channel matrix which can be expressed as

$$\bar{\mathbf{H}}_2 = \mathbf{b}_N(\varphi^a, \varphi^e) \mathbf{a}_M^H(\phi^a, \phi^e), \quad (6)$$

where φ^a and φ^e represent the azimuth and elevation angles of arrival (AoAs) at the RIS from the XL-MIMO, while ϕ^a and ϕ^e are the corresponding azimuth and elevation angles of departure (AoDs) from the XL-MIMO to the RIS. Furthermore, $\mathbf{a}_M(\phi^a, \phi^e)$ and $\mathbf{b}_N(\varphi^a, \varphi^e)$ are the array response vectors from the XL-MIMO to the RIS, where

$$\begin{aligned} \mathbf{a}_M(\phi^a, \phi^e) &= [e^{j2\pi m_x \frac{d \cos(\phi^e) \sin(\phi^a)}{\lambda}}]_{m_x \in \mathcal{L}(M_x)}^T \\ &\otimes [e^{j2\pi m_y \frac{d \sin(\phi^e)}{\lambda}}]_{m_y \in \mathcal{L}(M_y)}^T, \end{aligned} \quad (7a)$$

$$\begin{aligned} \mathbf{b}_N(\varphi^a, \varphi^e) &= [e^{j2\pi n_1 \frac{d_R \cos(\varphi^e) \sin(\varphi^a)}{\lambda}}]_{n_1 \in \mathcal{L}(N_1)}^T \\ &\otimes [e^{j2\pi n_2 \frac{d_R \sin(\varphi^e)}{\lambda}}]_{n_2 \in \mathcal{L}(N_2)}^T, \end{aligned} \quad (7b)$$

where $d_R = \frac{\lambda}{2}$ is the RIS element spacing and $\mathcal{L}(n) = \{0, 1, \dots, n-1\}$. Moreover, $\tilde{\mathbf{H}}_2$ is the NLoS channel matrix whose components are independent and identically distributed (i.i.d.) complex Gaussian RVs with zero mean and unit variance. To facilitate notation, let $\mathbf{H}_{2,s} \in \mathbb{C}^{N \times M^*}$ denote the channel between the s th subarray of XL-MIMO and the RIS, which is a submatrix of $\mathbf{H}_2$, given by $\mathbf{H}_{2,s} = \mathbf{H}_2(:, r_s)$, where r_s is the antenna index set corresponding to s th subarray, defined as $r_s \triangleq \{(s-1)M^* + 1, sM^*\}$.

We assume that the links between the RIS and FFUEs are LoS dominant.² Thus, the channel between the RIS and FFUE k can be written as [33]

$$\mathbf{h}_k^H = \sqrt{\varsigma_k} \mathbf{b}_N^H(\vartheta_k^a, \vartheta_k^e), \quad \forall k \in \mathcal{K}_f, \quad (8)$$

where ς_k is the path-loss coefficient; ϑ_k^a and ϑ_k^e represent the azimuth and elevation AoDs reflected by the RIS to the k th FFUE, while the array response vector $\mathbf{b}_N(\vartheta_k^a, \vartheta_k^e)$ is given by (7b), where φ^a and φ^e are replaced with ϑ_k^a and ϑ_k^e , respectively.

It is worth mentioning that a LoS or near-LoS connection between the RIS and the XL-MIMO BS is essential to ensure a strong and stable BS-RIS link. This condition allows the

²In practice, RISs are intentionally placed on building facades/walls to ensure favorable LoS propagation between the RIS and blocked FFUEs by maximizing reflection gain. We note that the gap between the data rate achieved by a purely LoS channel and that achieved by a LoS-dominant channel is negligible, as shown in [32], [33].

¹We consider a RIS that is not extremely large. Therefore, it can be modeled as a single transmitter, and spatial non-stationarities do not arise, unlike in the case of XL-RISs, where such issues must be considered [10].

maximum number of RIS elements to effectively receive and reflect the incident signals with high efficiency. Additionally, FFUEs should be positioned within the RIS reflection region to avail of the beamforming gain. To enhance system performance, orientation strategies should aim to introduce spatial separation—or angular diversity—along the RIS–user paths. This reduces channel correlation and increases degrees of freedom, enabling more effective multi-user transmission.

C. Received Signal

The transmit signal from the s th subarray is given by

$$\mathbf{x}_s = \sum_{k \in \mathcal{K}_N} \sqrt{P_n \tilde{\eta}_{sk}} \tilde{\mathbf{D}}_{sk} \tilde{\mathbf{w}}_{sk} \tilde{s}_k + \sum_{j \in \mathcal{K}_F} \sqrt{P_f \tilde{\eta}_{sj}} \tilde{\mathbf{D}}_{sj} \tilde{\mathbf{w}}_{sj} \tilde{s}_j, \quad (9)$$

where P_n and P_f denote the transmit power for NFUEs and FFUEs; $\tilde{\eta}_{sk}$, $\tilde{\eta}_{sj}$ are the power control coefficients at the s th subarray allocated to the k th NFUE and the j th FFUE, respectively; $\tilde{\mathbf{w}}_{sk} \in \mathbb{C}^{M^* \times 1}$ and $\tilde{\mathbf{w}}_{sj} \in \mathbb{C}^{M^* \times 1}$ are the precoding vectors at the s th subarray associated with NFUE k and FFUE j , respectively; $\tilde{s}_k$ and $\tilde{s}_j$ are the uncorrelated symbols intended for NFUE k and FFUE j , with $\mathbb{E}\{|\tilde{s}_k|^2\} = \mathbb{E}\{|\tilde{s}_j|^2\} = 1$. Moreover, for the sake of generalization, we define a diagonal matrix $\tilde{\mathbf{D}}_{sk}$ with binary entries on its main diagonal to indicate the active antenna elements in subarray s assigned to user k . Since we consider subarray-level activation, $\tilde{\mathbf{D}}_{sk} = \mathbf{I}_{M^*}$ ($\tilde{\mathbf{D}}_{sj} = \mathbf{I}_{M^*}$) implies that NFUE k (FFUE j) is served by subarray s , otherwise, $\tilde{\mathbf{D}}_{sk} = \mathbf{0}_{M^*}$ ($\tilde{\mathbf{D}}_{sj} = \mathbf{0}_{M^*}$).³

The transmitted power satisfies the power constraint $\sum_{s \in \mathcal{S}} \mathbb{E}\{\|\mathbf{x}_s\|^2\} \leq P$, where P is the maximum total transmit power at the XL-MIMO array. Using (9), we can express the power constraint as

$$\sum_{s \in \mathcal{S}} \left(P_n \sum_{k \in \mathcal{K}_N} \tilde{\eta}_{sk} \tilde{c}_{sk} + P_f \sum_{j \in \mathcal{K}_F} \tilde{\eta}_{sj} \tilde{c}_{sj} \right) \leq P, \quad (10)$$

where $\tilde{c}_{sk} = \text{tr}(\tilde{\mathbf{D}}_{sk} \tilde{\mathbf{w}}_{sk} (\tilde{\mathbf{D}}_{sk} \tilde{\mathbf{w}}_{sk})^H)$, $\tilde{c}_{sj} = \mathbb{E}\{\text{tr}(\tilde{\mathbf{D}}_{sj} \tilde{\mathbf{w}}_{sj} (\tilde{\mathbf{D}}_{sj} \tilde{\mathbf{w}}_{sj})^H)\}$.

The signal received at NFUE k can be expressed as:

$$\tilde{y}_k = \sum_{s \in \mathcal{S}} \left(\sqrt{P_n} \sum_{k \in \mathcal{K}_N} \sqrt{\tilde{\eta}_{sk}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{sk} \tilde{\mathbf{w}}_{sk} \tilde{s}_k + \sqrt{P_f} \sum_{j \in \mathcal{K}_F} \sqrt{\tilde{\eta}_{sj}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{sj} \tilde{\mathbf{w}}_{sj} \tilde{s}_j \right) + n_k, \quad (11)$$

where $\mathcal{S}$ is the set of all subarrays, $n_k \sim \mathcal{CN}(0, 1)$ is the additive white Gaussian noise (AWGN) at the k th NFUE.

For FFUE k , the received signal can be expressed as

$$\tilde{y}_k = \sum_{s \in \mathcal{S}} \left(\sqrt{P_f} \sum_{k \in \mathcal{K}_F} \sqrt{\tilde{\eta}_{sk}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{sk} \tilde{\mathbf{w}}_{sk} \tilde{s}_k + \sqrt{P_n} \sum_{i \in \mathcal{K}_N} \sqrt{\tilde{\eta}_{si}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{si} \tilde{\mathbf{w}}_{si} \tilde{s}_i \right) + n_k, \quad (12)$$

where $\tilde{\mathbf{g}}_{sk}^H = \mathbf{h}_k^H \mathbf{\Theta} \mathbf{H}_{2,s}$ denotes the channel between the s th subarray and FFUE k .

III. PERFORMANCE ANALYSIS

For NFUE k , the received signal in (11) can be written as

$$\tilde{y}_k = \overline{\mathbf{D}}_{sk} \tilde{s}_k + \sum_{i \in \mathcal{K}_N \setminus k} \overline{\mathbf{U}}_i \tilde{s}_i + \sum_{j \in \mathcal{K}_F} \overline{\mathbf{U}}_j \tilde{s}_j + n_k, \quad (13)$$

³To simplify the subsequent analysis, we do not include the signal propagation through sidelobes.

where

$$\overline{\mathbf{D}}_{sk} \triangleq \sqrt{P_n} \sum_{s \in \mathcal{S}} \sqrt{\tilde{\eta}_{sk}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{sk} \tilde{\mathbf{w}}_{sk}, \quad (14a)$$

$$\overline{\mathbf{U}}_i \triangleq \sqrt{P_n} \sum_{s \in \mathcal{S}} \sqrt{\tilde{\eta}_{si}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{si} \tilde{\mathbf{w}}_{si}, \quad (14b)$$

$$\overline{\mathbf{U}}_j \triangleq \sqrt{P_f} \sum_{s \in \mathcal{S}} \sqrt{\tilde{\eta}_{sj}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{sj} \tilde{\mathbf{w}}_{sj}, \quad (14c)$$

correspond to the desired signal ($\overline{\mathbf{D}}_{sk}$), the intra-group interference caused by NFUEs ($\overline{\mathbf{U}}_i$) and the inter-group interference caused by FFUEs ($\overline{\mathbf{U}}_j$), respectively. Since $\tilde{s}_k$ is independent of $\tilde{s}_i$ for any $i \neq k$, the first term and the second term of (13) are uncorrelated. Similarly, the third term of (13) is uncorrelated with the first term of (13). Using the arguments from [27] that uncorrelated Gaussian noise represents the worst case, an achievable downlink SE of NFUE k can be obtained as $\overline{\text{SE}}_k = \log_2(1 + \overline{\text{SINR}}_k)$, where

$$\overline{\text{SINR}}_k = \frac{|\overline{\mathbf{D}}_{sk}|^2}{\sum_{i \in \mathcal{K}_N \setminus k} |\overline{\mathbf{U}}_i|^2 + \sum_{j \in \mathcal{K}_F} \mathbb{E}\{|\overline{\mathbf{U}}_j|^2\} + \sigma_k^2}. \quad (15)$$

Note that the SE of the NFUEs is influenced by the interference caused by signals intended for FFUEs, thereby being a function of the RIS phase shift matrix. For FFUE k , we can rewrite the received signal $\tilde{y}_k$ in (12) as

$$\tilde{y}_k = \widetilde{\mathbf{D}}_{sk} \tilde{s}_k + \widetilde{\mathbf{B}}_{\mathbf{U}} \tilde{s}_k + \sum_{j \in \mathcal{K}_F \setminus k} \widetilde{\mathbf{U}}_j \tilde{s}_j + \sum_{i \in \mathcal{K}_N} \widetilde{\mathbf{U}}_i \tilde{s}_i + n_k, \quad (16)$$

where

$$\widetilde{\mathbf{D}}_{sk} \triangleq \sqrt{P_f} \mathbb{E} \left\{ \sum_{s \in \mathcal{S}} \sqrt{\tilde{\eta}_{sk}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{sk} \tilde{\mathbf{w}}_{sk} \right\}, \quad (17a)$$

$$\widetilde{\mathbf{B}}_{\mathbf{U}} \triangleq \sqrt{P_f} \left(\sum_{s \in \mathcal{S}} \sqrt{\tilde{\eta}_{sk}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{sk} \tilde{\mathbf{w}}_{sk} - \mathbb{E} \left\{ \sum_{s \in \mathcal{S}} \sqrt{\tilde{\eta}_{sk}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{sk} \tilde{\mathbf{w}}_{sk} \right\} \right), \quad (17b)$$

$$\widetilde{\mathbf{U}}_j \triangleq \sqrt{P_f} \sum_{s \in \mathcal{S}} \sqrt{\tilde{\eta}_{sj}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{sj} \tilde{\mathbf{w}}_{sj}, \quad (17c)$$

$$\widetilde{\mathbf{U}}_i \triangleq \sqrt{P_n} \sum_{s \in \mathcal{S}} \sqrt{\tilde{\eta}_{si}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{si} \tilde{\mathbf{w}}_{si}, \quad (17d)$$

correspond to the desired signal ($\widetilde{\mathbf{D}}_{sk}$), the beamforming gain uncertainty ($\widetilde{\mathbf{B}}_{\mathbf{U}}$), the intra-group interference caused by the other FFUEs ($\widetilde{\mathbf{U}}_j$) and the inter-group interference caused by NFUEs ($\widetilde{\mathbf{U}}_i$), respectively. Since $\tilde{s}_k$ is independent of $\widetilde{\mathbf{D}}_{sk}$ and $\widetilde{\mathbf{B}}_{\mathbf{U}}$, the first term and the second term of (16) are uncorrelated. Similarly, the third and fourth terms are uncorrelated with the first term of (16). According to [27], the achievable SE at FFUE k is obtained as $\widetilde{\text{SE}}_k = \log_2(1 + \widetilde{\text{SINR}}_k)$, where

$$\widetilde{\text{SINR}}_k = \frac{|\widetilde{\mathbf{D}}_{sk}|^2}{\mathbb{E}\{|\widetilde{\mathbf{B}}_{\mathbf{U}}|^2\} + \sum_{j \in \mathcal{K}_F \setminus k} \mathbb{E}\{|\widetilde{\mathbf{U}}_j|^2\} + \sum_{i \in \mathcal{K}_N} \mathbb{E}\{|\widetilde{\mathbf{U}}_i|^2\} + \sigma_k^2}. \quad (18)$$

Next, we derive the VR-based SE expressions for different precoding schemes, which are MRT, CZF, and LZF and assume that VR selection for NFUEs and FFUEs has already been performed on the XL-MIMO array.⁴ Consequently, each VR-

⁴In the context of massive MIMO, it is well known that linear precoding designs can achieve fairly close results to optimal precoding. Therefore, we apply linear precoding to avoid high computational complexity [27], [34].

based precoding scheme assigns a specific VR to each NFUE and FFUE. In Subsection IV-A, we introduce our proposed VR selection algorithm. From this point on, we will use the superscripts MRT, CZF, and LZF to denote the corresponding precoding schemes and their associated parameters.

A. VR-Based MRT Precoding Scheme

As the channels between the s th subarray and the k th NFUE (FFUE) are denoted as $\bar{\mathbf{g}}_{sk}^H \in \mathbb{C}^{1 \times M^*}$ ($\tilde{\mathbf{g}}_{sk}^H \in \mathbb{C}^{1 \times M^*}$), the VR-based MRT precoding vectors in the subarray s for the k th NFUE and the k th FFUE are given by

$$\bar{\mathbf{w}}_{sk}^{\text{MRT}} = \bar{\mathbf{g}}_{sk}, \quad \tilde{\mathbf{w}}_{sk}^{\text{MRT}} = \tilde{\mathbf{g}}_{sk}. \quad (19)$$

Proposition 1. *The SE of the k th NFUE with VR-based MRT precoding scheme can be represented in closed form as $\overline{SE}_k^{\text{MRT}} = \log_2(1 + \overline{SINR}_k^{\text{MRT}})$, where*

$$\overline{SINR}_k^{\text{MRT}} = \frac{P_n \left(\sum_{s \in \tilde{\mathcal{S}}_k^{\text{MRT}}} \sqrt{\tilde{\eta}_{sk}^{\text{MRT}}} \tilde{\Psi}_{kk}^s \right)^2}{\tilde{\Delta}^{\text{MRT}} + \sigma_k^2}, \quad (20)$$

with $\tilde{\Delta}^{\text{MRT}} \triangleq P_n \sum_{i \in \mathcal{K}_{\mathcal{N}} \setminus k} \left| \sum_{s \in \tilde{\mathcal{S}}_i^{\text{MRT}}} \sqrt{\tilde{\eta}_{si}^{\text{MRT}}} \tilde{\Psi}_{ki}^s \right|^2 + P_f \sum_{j \in \mathcal{K}_{\mathcal{F}}} \sum_{s \in \tilde{\mathcal{S}}_j^{\text{MRT}}} \sum_{s' \in \tilde{\mathcal{S}}_j^{\text{MRT}}} \sqrt{\tilde{\eta}_{sj}^{\text{MRT}}} \sqrt{\tilde{\eta}_{s'j}^{\text{MRT}}} \tilde{\Xi}_{kj}^{ss'}$, where $\tilde{\Psi}_{kk}^s \triangleq \bar{\mathbf{g}}_{sk}^H \bar{\mathbf{g}}_{sk}$, $\tilde{\Psi}_{ki}^s \triangleq \bar{\mathbf{g}}_{sk}^H \bar{\mathbf{g}}_{si}$, $\forall i \neq k$; $\tilde{\mathcal{S}}_k^{\text{MRT}}$, $\tilde{\mathcal{S}}_i^{\text{MRT}}$ and $\tilde{\mathcal{S}}_j^{\text{MRT}}$ are the VRs chosen by NFUEs k , i and FFUE j with MRT precoding scheme, respectively; $\tilde{\Xi}_{kj}^{ss'} \triangleq \Re \left\{ \bar{\mathbf{g}}_{sk}^H (\alpha_2^2 \bar{\mathbf{H}}_{2,s}^H \mathbf{M}_{jj} \bar{\mathbf{H}}_{2,s'} + \delta_{ss'} \zeta_k \beta_2^2 N \mathbf{I}_{M^*}) \bar{\mathbf{g}}_{s'k} \right\}$, where $\mathbf{M}_{jj} \triangleq \Theta^H \mathbf{h}_j \mathbf{h}_j^H \Theta$, while $\mathbf{H}_{2,s'} \in \mathbb{C}^{N \times M^*}$ denotes the channel between the s' th subarray of XL-MIMO and the RIS, which is a submatrix of $\mathbf{H}_2 \in \mathbb{C}^{N \times M}$, and we have $\mathbf{H}_{2,s'} = \mathbf{H}_2(:, r_{s'})$, where $r_{s'} = \{(s' - 1)M^* + 1 : s'M^*\}$.

Proof. See Appendix A. $\square$

Proposition 2. *The SE of FFUE k with VR-based MRT precoding scheme is given by $\widetilde{SE}_k^{\text{MRT}} = \log_2(1 + \widetilde{SINR}_k^{\text{MRT}})$, where*

$$\widetilde{SINR}_k^{\text{MRT}} = \frac{P_f \left(\sum_{s \in \tilde{\mathcal{S}}_k^{\text{MRT}}} \sqrt{\tilde{\eta}_{sk}^{\text{MRT}}} \tilde{\Psi}_k^s \right)^2}{\tilde{\Delta}^{\text{MRT}} + \sigma_k^2}, \quad (21)$$

where $\tilde{\mathcal{S}}_k^{\text{MRT}}$ is the VR chosen by FFUE k , $\tilde{\Psi}_k^s \triangleq \mathbf{h}_k^H \Theta (\alpha_2^2 \bar{\mathbf{H}}_{2,s}^H \bar{\mathbf{H}}_{2,s} + \beta_2^2 M^* \mathbf{I}_N) \Theta^H \mathbf{h}_k$, and

$$\begin{aligned} \tilde{\Delta}^{\text{MRT}} &\triangleq P_f \sum_{s \in \tilde{\mathcal{S}}_k^{\text{MRT}}} \tilde{\eta}_{sk}^{\text{MRT}} \tilde{a}_k^s \\ &+ P_f \sum_{j \in \mathcal{K}_{\mathcal{F}} \setminus k} \sum_{s \in \tilde{\mathcal{S}}_j^{\text{MRT}}} \sum_{s' \in \tilde{\mathcal{S}}_j^{\text{MRT}}} \sqrt{\tilde{\eta}_{sj}^{\text{MRT}}} \sqrt{\tilde{\eta}_{s'j}^{\text{MRT}}} \tilde{b}_{kj}^{ss'} \\ &+ P_n \sum_{i \in \mathcal{K}_{\mathcal{N}}} \sum_{s \in \tilde{\mathcal{S}}_i^{\text{MRT}}} \sum_{s' \in \tilde{\mathcal{S}}_i^{\text{MRT}}} \sqrt{\tilde{\eta}_{si}^{\text{MRT}}} \sqrt{\tilde{\eta}_{s'i}^{\text{MRT}}} \tilde{\Xi}_{ki}^{ss'}, \end{aligned} \quad (22)$$

where $\tilde{\Xi}_{ki}^{ss'} \triangleq \Re \{ \bar{\mathbf{g}}_{si}^T \mathbf{B}_{ki}^{ss'} \bar{\mathbf{g}}_{si}^* \}$, $\mathbf{B}_{ki}^{ss'} = \alpha_2^2 \bar{\mathbf{H}}_{2,s'}^H \mathbf{M}_{kk} \bar{\mathbf{H}}_{2,s} + \delta_{ss'} \zeta_k \beta_2^2 N \mathbf{I}_{M^*}$, with $\mathbf{M}_{kk} \triangleq \Theta^H \mathbf{h}_k \mathbf{h}_k^H \Theta$, and

$$\begin{aligned} \tilde{a}_k^s &\triangleq \Re \left\{ \mathbf{h}_k^H \Theta (\zeta_k \alpha_2^2 \beta_2^2 N \bar{\mathbf{H}}_{2,s}^H \bar{\mathbf{H}}_{2,s} + (\alpha_2^2 \beta_2^2 M^* \right. \\ &\quad \times \mathbf{b}_N^H(\varphi^a, \varphi^e) \mathbf{M}_{kk} \mathbf{b}_N(\varphi^a, \varphi^e) + \zeta_k \beta_2^2 N M^*) \mathbf{I}_N) \Theta^H \mathbf{h}_k \Big\}, \\ \tilde{b}_{kj}^{ss'} &\triangleq \Re \left\{ \mathbf{h}_k^H \Theta (\delta_{ss'} \mathbf{C}_{sj} + (1 - \delta_{ss'}) \mathbf{D}_j^{ss'}) \Theta^H \mathbf{h}_k \right\}, \end{aligned}$$

where $\mathbf{D}_j^{ss'} \triangleq (\alpha_2^2 \bar{\mathbf{H}}_{2,s}^H \bar{\mathbf{H}}_{2,s} + \beta_2^2 M^* \mathbf{I}_N) \mathbf{M}_{jj} (\alpha_2^2 \bar{\mathbf{H}}_{2,s'}^H \bar{\mathbf{H}}_{2,s'} + \beta_2^2 M^* \mathbf{I}_N)$ and

$$\begin{aligned} \mathbf{C}_{sj} &\triangleq \alpha_2^4 \bar{\mathbf{H}}_{2,s}^H \bar{\mathbf{H}}_{2,s} \mathbf{M}_{jj} \bar{\mathbf{H}}_{2,s}^H + \beta_2^2 \left(\alpha_2^2 M^* \bar{\mathbf{H}}_{2,s}^H \bar{\mathbf{H}}_{2,s} \mathbf{M}_{jj} \right. \\ &\quad + \alpha_2^2 \text{tr}(\mathbf{M}_{jj}) \bar{\mathbf{H}}_{2,s}^H \bar{\mathbf{H}}_{2,s} + \alpha_2^2 \text{tr}(\bar{\mathbf{H}}_{2,s}^H \mathbf{M}_{jj} \bar{\mathbf{H}}_{2,s}) \\ &\quad \left. + \alpha_2^2 M^* \mathbf{M}_{jj} \bar{\mathbf{H}}_{2,s}^H \bar{\mathbf{H}}_{2,s} + \beta_2^2 (M^*)^2 \mathbf{M}_{jj} + \beta_2^2 M^* \text{tr}(\mathbf{M}_{jj}) \mathbf{I}_N \right). \end{aligned} \quad (23)$$

Proof. See Appendix B. $\square$

B. VR-Based CZF Precoding Scheme

In order to mitigate the intra-group interference caused by signals transmitted to other NFUEs (FFUEs), the CZF precoding method is employed.⁵ For NFUE k , the precoding vector of the subarray s is given by

$$\bar{\mathbf{w}}_{sk}^{\text{CZF}} = [\bar{\mathbf{H}}_1 (\bar{\mathbf{H}}_1^H \bar{\mathbf{H}}_1)^{-1}]_{(r_s, k)} \in \mathbb{C}^{M^* \times 1}. \quad (24)$$

For FFUEs, we set the precoding vector of the subarray s towards FFUE k as

$$\tilde{\mathbf{w}}_{sk}^{\text{CZF}} = [\tilde{\mathbf{G}} (\tilde{\mathbf{G}}^H \tilde{\mathbf{G}})^{-1}]_{(r_s, k)} \in \mathbb{C}^{M^* \times 1}. \quad (25)$$

Since the VR-based CZF scheme shares similarities with centralized cell-free massive MIMO systems, to eliminate intra-group interference, it is sensible to set the power control coefficients as $\bar{\eta}_k^{\text{CZF}} \triangleq \bar{\eta}_{1k}^{\text{CZF}} = \dots = \bar{\eta}_{Nk}^{\text{CZF}}$ for any NFUE k [35]. For FFUEs, it follows similarly.

Proposition 3. *The SE of NFUE k with VR-based CZF precoding scheme is given by $\overline{SE}_k^{\text{CZF}} = \log_2(1 + \overline{SINR}_k^{\text{CZF}})$, where*

$$\overline{SINR}_k^{\text{CZF}} = \frac{P_n \bar{\eta}_k^{\text{CZF}} \left(\sum_{s \in \tilde{\mathcal{S}}_k^{\text{CZF}}} \bar{\mathbf{g}}_{sk}^H \bar{\mathbf{w}}_{sk}^{\text{CZF}} \right)^2}{\tilde{\Delta}^{\text{CZF}} + \sigma_k^2}, \quad (26)$$

where

$$\begin{aligned} \tilde{\Delta}^{\text{CZF}} &\triangleq P_n \sum_{i \in \mathcal{K}_{\mathcal{N}} \setminus k} \bar{\eta}_i^{\text{CZF}} \left| \sum_{s \in \tilde{\mathcal{S}}_i^{\text{CZF}}} \bar{\mathbf{g}}_{si}^H \bar{\mathbf{w}}_{si}^{\text{CZF}} \right|^2 \\ &+ P_f \sum_{j \in \mathcal{K}_{\mathcal{F}}} \bar{\eta}_j^{\text{CZF}} r_{kj}, \end{aligned} \quad (27)$$

where $r_{kj} \triangleq \Re \left\{ \sum_{s \in \tilde{\mathcal{S}}_j^{\text{CZF}}} \sum_{s' \in \tilde{\mathcal{S}}_j^{\text{CZF}}} \bar{\mathbf{g}}_{sj}^H \mathbb{E} \{ \tilde{\mathbf{w}}_{sj}^{\text{CZF}} (\tilde{\mathbf{w}}_{s'j}^{\text{CZF}})^H \} \bar{\mathbf{g}}_{s'k} \right\}$, $\tilde{\mathcal{S}}_k^{\text{CZF}}$, $\tilde{\mathcal{S}}_j^{\text{CZF}}$ and $\tilde{\mathcal{S}}_i^{\text{CZF}}$ are the VRs selected by FFUEs k , j and NFUE i under CZF precoding scheme, respectively.

Proof. The derivation of r_{kj} follows the proof of $\mathbb{E}\{|\bar{\mathbf{U}}_j|^2\}$ in (75) with MRT and is therefore omitted. $\square$

Proposition 4. *The SE of FFUE k with VR-based CZF design is given by $\widetilde{SE}_k^{\text{CZF}} = \log_2(1 + \widetilde{SINR}_k^{\text{CZF}})$, where*

$$\widetilde{SINR}_k^{\text{CZF}} = \frac{P_f \bar{\eta}_k^{\text{CZF}} \left(\sum_{s \in \tilde{\mathcal{S}}_k^{\text{CZF}}} \mathbb{E} \{ \bar{\mathbf{g}}_{sk}^H \tilde{\mathbf{w}}_{sk}^{\text{CZF}} \} \right)^2}{\tilde{\Delta}^{\text{CZF}} + \sigma_k^2}, \quad (28)$$

⁵While CZF and LZF effectively handle intra-group interference, they overlook inter-group interference. Although optimal precoding can address this issue, it comes with high computational cost and limited gains. To reduce complexity and manage inter-group interference, we propose using protective MRT (PMRT) or protective ZF (PZF) techniques to shield each user group from interference by others.

where

$$\begin{aligned} \tilde{\Delta}^{\text{CZF}} &\triangleq P_f \tilde{\eta}_k^{\text{CZF}} \varepsilon_k + P_n \sum_{i \in \mathcal{K}_N} \tilde{\eta}_i^{\text{CZF}} t_{ki} \\ &+ P_f \sum_{j \in \mathcal{K}_F \setminus k} \tilde{\eta}_j^{\text{CZF}} \mathbb{E} \left\{ \left| \sum_{s \in \tilde{\mathcal{S}}_j^{\text{CZF}}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{w}}_{sj}^{\text{CZF}} \right|^2 \right\}, \end{aligned} \quad (29)$$

with $\varepsilon_k \triangleq \sum_{s \in \tilde{\mathcal{S}}_k^{\text{CZF}}} \mathbb{E} \left\{ \left| \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{w}}_{sk}^{\text{CZF}} - \mathbb{E} \{ \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{w}}_{sk}^{\text{CZF}} \} \right|^2 \right\}$ and $t_{ki} \triangleq \Re \left\{ \sum_{s \in \tilde{\mathcal{S}}_i^{\text{CZF}}} \sum_{s' \in \tilde{\mathcal{S}}_i^{\text{CZF}}} (\tilde{\mathbf{w}}_{si}^{\text{CZF}})^H \mathbf{B}_{ki}^{ss'} \tilde{\mathbf{w}}_{si}^{\text{CZF}} \right\}$.

Proof. The derivation of t_{ki} follows the proof of $\mathbb{E} \{ |\tilde{\mathbf{U}}_i|^2 \}$ in (86) with MRT and is therefore omitted. $\square$

C. VR-Based LZF Precoding Scheme

To reduce the computational complexity of the CZF precoding scheme associated with channel inversion, we propose a low-complexity scheme, referred to as LZF, which performs zero-forcing locally. Specifically, instead of using all VRs, each subarray s independently applies precoding to nullify intra-group interference within its own subarray. However, it cannot eliminate intra-group interference caused by other subarrays. We notice that LZF employs the same VRs as CZF for simplicity.

For NFUE k served by subarray s , the precoding vector with the LZF scheme is given by

$$\tilde{\mathbf{w}}_{sk}^{\text{LZF}} = \left[\tilde{\mathbf{G}}_s (\tilde{\mathbf{G}}_s^H \tilde{\mathbf{G}}_s)^{-1} \right]_{(:,k)} \in \mathbb{C}^{M^* \times 1}, \quad (30)$$

where $\tilde{\mathbf{G}}_s^H = [\tilde{\mathbf{g}}_{s1}^H, \dots, \tilde{\mathbf{g}}_{sK_f}^H]^H \in \mathbb{C}^{K_N \times M^*}$.

Moreover, the LZF precoding vector for FFUE k served by subarray s is

$$\tilde{\mathbf{w}}_{sk}^{\text{LZF}} = \left[\tilde{\mathbf{G}}_s (\tilde{\mathbf{G}}_s^H \tilde{\mathbf{G}}_s)^{-1} \right]_{(:,k)} \in \mathbb{C}^{M^* \times 1}, \quad (31)$$

where $\tilde{\mathbf{G}}_s^H = [(\tilde{\mathbf{g}}_{s1}^H)^T, \dots, (\tilde{\mathbf{g}}_{sK_f}^H)^T]^H \in \mathbb{C}^{K_f \times M^*}$. Note that if the s th subarray serves a single user k , LZF precoder at the s th subarray reduces to MRT, that is, $\tilde{\mathbf{w}}_{sk}^{\text{LZF}} = \tilde{\mathbf{w}}_{sk}^{\text{MRT}}$ and $\tilde{\mathbf{w}}_{sk}^{\text{LZF}} = \tilde{\mathbf{w}}_{sk}^{\text{MRT}}$.

Proposition 5. The SE of NFUE k with VR-based LZF precoding scheme is given by $\overline{\text{SE}}_k^{\text{LZF}} = \log_2 \left(1 + \overline{\text{SINR}}_k^{\text{LZF}} \right)$, where

$$\overline{\text{SINR}}_k^{\text{LZF}} = \frac{P_n \left(\sum_{s \in \tilde{\mathcal{S}}_k^{\text{LZF}}} \sqrt{\tilde{\eta}_{sk}^{\text{LZF}}} \right)^2}{\tilde{\Delta}^{\text{LZF}} + \sigma_k^2}, \quad (32)$$

with $\tilde{\Delta}^{\text{LZF}} \triangleq P_n \sum_{i \in \mathcal{K}_N \setminus k} \left| \sum_{s \in \tilde{\mathcal{S}}_i^{\text{LZF}}} \sqrt{\tilde{\eta}_{si}^{\text{LZF}}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{w}}_{si}^{\text{LZF}} \right|^2 + P_f \sum_{j \in \mathcal{K}_F} \sum_{s \in \tilde{\mathcal{S}}_j^{\text{LZF}}} \sum_{s' \in \tilde{\mathcal{S}}_j^{\text{LZF}}} \sqrt{\tilde{\eta}_{sj}^{\text{LZF}}} \sqrt{\tilde{\eta}_{s'j}^{\text{LZF}}} \tilde{\tau}_{kj}^{ss'}$, with $\tilde{\tau}_{kj}^{ss'} \triangleq \Re \left\{ \tilde{\mathbf{g}}_{sk}^H \mathbb{E} \left\{ \tilde{\mathbf{w}}_{sj}^{\text{LZF}} (\tilde{\mathbf{w}}_{s'j}^{\text{LZF}})^H \right\} \tilde{\mathbf{g}}_{s'k}^H \right\}$.

Proof. The derivation of $\tilde{\tau}_{kj}^{ss'}$ follows the proof of $\mathbb{E} \{ |\tilde{\mathbf{U}}_j|^2 \}$ in (75) with MRT and is therefore omitted. $\square$

Proposition 6. The SE of FFUE k with VR-based LZF design is given by $\widetilde{\text{SE}}_k^{\text{LZF}} = \log_2 \left(1 + \widetilde{\text{SINR}}_k^{\text{LZF}} \right)$, where

$$\widetilde{\text{SINR}}_k^{\text{LZF}} = \frac{P_f \left(\sum_{s \in \tilde{\mathcal{S}}_k^{\text{LZF}}} \sqrt{\tilde{\eta}_{sk}^{\text{LZF}}} \right)^2}{\tilde{\Delta}^{\text{LZF}} + \sigma_k^2}, \quad (33)$$

where $\tilde{\Delta}^{\text{LZF}} \triangleq P_f \sum_{j \in \mathcal{K}_F \setminus k} \mathbb{E} \left\{ \left| \sum_{s \in \tilde{\mathcal{S}}_j^{\text{LZF}}} \sqrt{\tilde{\eta}_{sj}^{\text{LZF}}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{w}}_{sj}^{\text{LZF}} \right|^2 \right\} + P_n \sum_{i \in \mathcal{K}_N} \sum_{s \in \tilde{\mathcal{S}}_i^{\text{LZF}}} \sum_{s' \in \tilde{\mathcal{S}}_i^{\text{LZF}}} \sqrt{\tilde{\eta}_{si}^{\text{LZF}}} \sqrt{\tilde{\eta}_{s'i}^{\text{LZF}}} \tilde{\tau}_{ki}^{ss'}$, with $\tilde{\tau}_{ki}^{ss'} \triangleq \Re \left\{ (\tilde{\mathbf{w}}_{s'i}^{\text{LZF}})^H \mathbf{B}_{ki}^{ss'} \tilde{\mathbf{w}}_{si}^{\text{LZF}} \right\}$.

Proof. The derivation of $\tilde{\tau}_{ki}^{ss'}$ follows the proof of $\mathbb{E} \{ |\tilde{\mathbf{U}}_i|^2 \}$ in (86) with MRT and is therefore omitted. $\square$

Remark 1. It is worth noting that, since the majority of the received power is concentrated within each user's VR, our approach avoids involving the entire antenna array in the processing for every user. This design choice ensures that the computational burden does not scale proportionally with the total number of antennas. Moreover, in the proposed LZF precoding scheme, instead of jointly processing all VRs, each subarray s performs precoding independently to suppress intra-group interference within its own region. This localized and decoupled processing structure eliminates the need for costly large-scale matrix inversions. Consequently, the computational complexity of the precoding stage is reduced from $\mathcal{O}(M \times K_f^2)$ to $\mathcal{O}(M^* \times K_f^2)$ in LZF, where M^* is the number of antennas of a subarray and $M^* \ll M$.

IV. OPTIMIZATION PROBLEM

In this section, we optimize the phase shifts at the RIS and the power control coefficients at the XL-MIMO array to maximize the sum of the weighted minimum SE for both NFUEs and FFUEs, while ensuring that individual QoS requirements are met for all users. At the optimized solution, NFUEs and FFUEs achieve the same SE within their respective groups. This fairness-oriented optimization framework guarantees seamless connectivity for all users, aligning with the core principles of next-generation wireless networks. The problem can be formulated as

$$(P1) : \max_{\{\tilde{\eta}^i, \tilde{\eta}^i, \mathcal{U}_s^i, \mathcal{I}_s^i, \theta\}} w_n \min_{k \in \mathcal{K}_N} \overline{\text{SE}}_k^i + w_f \min_{k \in \mathcal{K}_F} \widetilde{\text{SE}}_k^i \quad (34a)$$

$$\text{s.t.} \quad \sum_{s \in \mathcal{S}} \left(P_n \sum_{k \in \mathcal{U}_s} \tilde{\eta}_{sk}^i \tilde{c}_{sk}^i + P_f \sum_{k \in \mathcal{I}_s} \tilde{\eta}_{sk}^i \tilde{c}_{sk}^i \right) \leq P, \quad (34b)$$

$$\overline{\text{SE}}_k^i \geq \bar{\mathcal{R}}_k, \forall k \in \mathcal{K}_N, \quad (34c)$$

$$\widetilde{\text{SE}}_k^i \geq \tilde{\mathcal{R}}_k, \forall k \in \mathcal{K}_F, \quad (34d)$$

$$\tilde{\eta}_{sk}^i \geq 0, \forall s, \forall k \in \mathcal{U}_s, \quad (34e)$$

$$\tilde{\eta}_{sk}^i \geq 0, \forall s, \forall k \in \mathcal{I}_s, \quad (34f)$$

$$0 \leq \theta_n \leq 2\pi, n = 1, \dots, N, \quad (34g)$$

where $\tilde{\eta}^i = [(\tilde{\eta}_1^i)^T, \dots, (\tilde{\eta}_{K_N}^i)^T]^T$, with $\tilde{\eta}_k^i = [\tilde{\eta}_{1k}^i, \dots, \tilde{\eta}_{K_f}^i]^T$, is a vector of power control coefficients for the NFUEs; $\tilde{\eta}^i = [(\tilde{\eta}_1^i)^T, \dots, (\tilde{\eta}_{K_f}^i)^T]^T$, with $\tilde{\eta}_k^i = [\tilde{\eta}_{1k}^i, \dots, \tilde{\eta}_{K_f}^i]^T$, represents power control coefficients for the FFUEs, where $i \in \{\text{CZF}, \text{LZF}, \text{MRT}\}$, respectively; $\mathcal{U}_s^i$ and $\mathcal{I}_s^i$ denote the sets of NFUEs and FFUEs served by the subarray s ; w_n and w_f are the positive weighting factors of the NF and FF parts, respectively. Constraints (34c) and (34d) denote the QoS requirements, where $\bar{\mathcal{R}}_k$ and $\tilde{\mathcal{R}}_k$ are the minimum SE requirements for NFUEs and FFUEs, respectively.

The optimization problem (P1) is highly intractable due to its non-convex objective function with respect to θ , $\bar{\eta}^i$, $\tilde{\eta}^i$, $\mathcal{U}_s^i$, $\mathcal{I}_s^i$. Finding the global maximizer of the weighted sum SE, subject to SINR and RIS phase shifts constraints, is extremely challenging. To address this, we decompose the problem into three stages. First, we employ a heuristic algorithm to determine suitable VR assignments for each user, resulting in the sets $\mathcal{U}_s^i$ and $\mathcal{I}_s^i$, while satisfying individual QoS requirements. Given these sets, we then propose a two-stage algorithm that decouples the original problem into two more tractable subproblems: RIS phase shifts design and power allocation. This decomposition significantly reduces the computational complexity by breaking the coupling between variables. More specifically, we adopt the penalty method to obtain an optimized phase-shift design and SCA to design an optimized power control algorithm.

A. Heuristic VR Selection Algorithm

We propose a heuristic VR selection algorithm for linear precoding that ensures the QoS of NFUEs and FFUEs with minimum number of subarrays. Without loss of generality, we focus on the CZF scheme. For NFUE k , we first initialize a set $\tilde{\mathcal{D}}_k^{\text{CZF}} = \text{diag}(\tilde{\mathbf{D}}_{1k}, \dots, \tilde{\mathbf{D}}_{Sk})$, with entries $\tilde{\mathbf{D}}_{sk} = \mathbf{I}_{M^*}$, indicating that NFUE k initially utilizes all subarrays. Next, for each subarray s , we calculate the SINR of NFUE k without employing subarray s as $\widehat{\text{SINR}}_k^s$ to assess whether the SINR constraint $\bar{\gamma}_k$ for that user is met. We define $\bar{\gamma}_k = \delta \widehat{\text{SINR}}_k$, where the VR selection ratio $\delta \in [0, 1]$, and $\widehat{\text{SINR}}_k$ can be derived from (15). If $\widehat{\text{SINR}}_k^s$ is greater than $\bar{\gamma}_k$, it implies that the remaining subarrays are sufficient for NFUE k . Then, we update the set $\tilde{\mathcal{D}}_k^{\text{CZF}}$ by setting $\tilde{\mathbf{D}}_{sk} = \mathbf{0}_{M^*}$. Finally, we obtain the matrix $\tilde{\mathcal{D}}_k^{\text{CZF}}$ which represents a set of minimum subarrays that meet the SINR requirement of NFUE k with CZF.

The proposed heuristic VR selection algorithm for NFUEs is summarized in **Algorithm 1**. The algorithm for choosing VRs for FFUEs is similar. The difference is that we compare $\widehat{\text{SINR}}_k^s$, which denotes the SINR without incorporating the subarray s for FFUE k , with the SINR constraint $\bar{\gamma}_k = \delta \widehat{\text{SINR}}_k$, where $\widehat{\text{SINR}}_k$ can be obtained from (18). Consequently, we obtain $\tilde{\mathcal{D}}_k^{\text{CZF}}$ with the CZF precoding scheme. Note that we assume that LZF utilizes the same VRs as CZF for tractability, while the corresponding VRs for MRT can be derived similarly.

The main differences between our proposed VR design and the method in [17] are twofold. First, we use SINR as our VR selection criterion instead of SNR, which is more practical as it accounts for interference from signals intended for other NFUEs. Second, the complexity of the method in [17] is $\mathcal{O}(KS \log(S))$, whereas the complexity of our proposed VR design is $\mathcal{O}(KS)$. This is because, for each user, we sequentially evaluate each subarray by removing it and checking whether the resulting SINR still meets the QoS requirement, resulting in $\mathcal{O}(S)$ complexity per user.

B. Phase-Shift Design at the RIS

By analyzing the SE results for different precoding schemes, we observe that these expressions are highly complex func-

tions of the phase shift matrix. Moreover, the objective function and the constraints in (34b)-(34d) are intricately coupled with respect to this parameter. To address this challenge, we follow the same approach as in [36], [7] and maximize the minimum channel gain of the FFUEs to design the optimized phase shift matrix Θ at the RIS with MRT, and CZF. The optimization problem can be formulated as

$$(P2) : \max_{\Theta} \min_k f_k(\Theta) \quad (35a)$$

$$\text{s.t.} \quad 0 \leq \theta_n \leq 2\pi, \forall n, \quad (35b)$$

where $f_k(\Theta) \triangleq \mathbb{E}\{\|\tilde{\mathbf{g}}_k^H\|^2\}$, which can be expressed as

$$\begin{aligned} f_k(\Theta) &= \mathbb{E}\{\mathbf{h}_k^H \Theta \mathbf{H}_2 \tilde{\mathcal{D}}_k^i (\tilde{\mathcal{D}}_k^i)^H \mathbf{H}_2^H \Theta^H \mathbf{h}_k\} \\ &= \mathbf{v}^H \mathbf{B}_k \mathbf{A} \mathbf{B}_k^H \mathbf{v}, \end{aligned} \quad (36)$$

where $\mathbf{v} \triangleq [e^{j\theta_1}, \dots, e^{j\theta_N}]^H$, $\mathbf{B}_k \triangleq \text{diag}(\mathbf{h}_k^H)$ and $\mathbf{A}_k \triangleq \alpha_2^2 \bar{\mathbf{H}}_2 \tilde{\mathcal{D}}_k^i (\tilde{\mathcal{D}}_k^i)^H \bar{\mathbf{H}}_2^H + \beta_2^2 \text{tr}(\tilde{\mathcal{D}}_k^i (\tilde{\mathcal{D}}_k^i)^H) \mathbf{I}_N \in \mathbb{C}^{N \times N}$. We employ SDP to reformulate problem (P2). To this end, we define $\mathbf{V} \triangleq \mathbf{v} \mathbf{v}^H \in \mathbb{C}^{N \times N}$, and $\mathbf{R}_k \triangleq \mathbf{B}_k \mathbf{A}_k \mathbf{B}_k^H \in \mathbb{C}^{N \times N}$. Then, (P2) can be rewritten as

$$(P2.1) : \max_{\mathbf{V}, t} t \quad (37a)$$

$$\text{s.t.} \quad t \leq \text{tr}(\mathbf{R}_k \mathbf{V}), k \in \mathcal{K}_{\mathcal{F}}, \quad (37b)$$

$$\mathbf{V}_{n,n} = 1, \forall n, \quad (37c)$$

$$\text{rank}(\mathbf{V}) = 1, \quad (37d)$$

$$\mathbf{V} \succeq \mathbf{0}. \quad (37e)$$

Note that Problem (P2.1) is non-convex due to the rank-one constraint. To address this issue, we replace the non-convex rank-one constraint with a difference-of-convex (DC) function formulation, expressed as follows [37], [38]:

$$\text{rank}(\mathbf{V}) = 1 \Leftrightarrow \|\mathbf{V}\|_* - \|\mathbf{V}\|_2 = 0. \quad (38)$$

In general, $\|\mathbf{V}\|_* - \|\mathbf{V}\|_2 \geq 0$, and equality holds when we have the rank-one solution. Then, we add the DC function in (38) into the objective function as a penalty component, yielding

$$(P2.2) : \max_{\mathbf{V}, t} t - \varrho(\|\mathbf{V}\|_* - \|\mathbf{V}\|_2) \quad (39a)$$

$$\text{s.t.} \quad (37b), (37c), (37e), \quad (39b)$$

where $\varrho > 0$ is the penalty factor and penalizes the objective function if $\mathbf{V}$ is not a rank-one matrix. Similarly to [7], we gradually increase ϱ to obtain a rank-one matrix $\mathbf{V}$.

Note that problem (P2.2) is still non-convex because of the convex term $\varrho\|\mathbf{V}\|_2$ in the objective function. To overcome this, we use SCA.⁶ To this end, we find the lower bound for $\|\mathbf{V}\|_2$ as $\|\mathbf{V}\|_2 \geq \bar{\mathbf{V}}^{(n)}$, where $\bar{\mathbf{V}}^{(n)} = (\|\mathbf{V}^{(n)}\|_2 + \text{tr}(\mathbf{u}(\mathbf{V}^{(n)})\mathbf{u}(\mathbf{V}^{(n)})^H(\mathbf{V} - \mathbf{V}^{(n)})))$ [7], [37]. Accordingly, we

⁶The SCA is an iterative optimization technique used to tackle non-convex problems by solving a sequence of convex subproblems. The core idea is to transform problem (39) into a series of tractable subproblems by linearizing the convex term $\varrho\|\mathbf{V}\|_2$ in the objective function. In each iteration, a local convex approximation is solved, improving upon the previous solution. This iterative process continues until convergence to a near-optimal solution [39].

Algorithm 1 Heuristic VR Selection Algorithm

```

1: for  $k = 1, \dots, K_n$  do
2:   set  $\bar{\mathbf{D}}_k^{\text{CZF}} = \text{diag}(\bar{\mathbf{D}}_{1k}, \dots, \bar{\mathbf{D}}_{S_k})$ , where  $\bar{\mathbf{D}}_{sk} = \mathbf{I}_{M^*}$ ;
3:   for  $s = 1, \dots, S$  do
4:     Calculate the SINR without employing subarray
        $s$  as  $\overline{\text{SINR}}_k^s$ ;
5:     if  $\overline{\text{SINR}}_k^s \geq \bar{\gamma}_k$  then
6:       Set  $\bar{\mathbf{D}}_{sk} = \mathbf{0}_{M^*}$ ;
7:     end if
8:   end for
9:   Get set of subarrays  $\bar{\mathbf{D}}_k^{\text{CZF}}$  chosen by NFUE  $k$ .
10: end for

```

obtain the following optimization problem:

$$\begin{aligned} \text{(P2.3)} : \quad & \max_{\mathbf{V}, t \geq 0} \quad t - \varrho(\|\mathbf{V}\|_* - \bar{\mathbf{V}}^{(n)}) \quad (40a) \\ \text{s.t.} \quad & (37b), (37c), (37e). \quad (40b) \end{aligned}$$

It is clear that problem (40) is a standard SDP which can be solved efficiently using existing solvers such as CVX [40]. The penalty-based algorithm for solving (37) consists of two nested loops and is executed iteratively until convergence. In the outer loop, we gradually increase the penalty factor ϱ . Specifically, we increase $\varrho = l\varrho$ until $\|\mathbf{V}\|_* - \|\mathbf{V}\|_2 \leq \epsilon$, where $l > 1$ denotes the scaling factor, and ϵ is a predefined threshold. It means that as ϱ increases, the solutions satisfy $\|\mathbf{V}\|_* - \|\mathbf{V}\|_2 \rightarrow 0$, ensuring near rank-one $\mathbf{V}$. In the inner loop, with the fixed penalty factor, we solve the problem (40) to obtain the optimized solution $\mathbf{V}^*$. The proposed algorithm is outlined in **Algorithm 2**.

C. Power Control Design at the XL-MIMO Array

With $\mathcal{U}_s^i$, $\mathcal{I}_s^i$ obtained from the heuristic VR selection algorithm and the optimized phase shifts $\boldsymbol{\theta}^*$, problem (P1) is reduced to

$$\text{(P3)} : \quad \max_{\{\bar{\eta}^i, \bar{\eta}^i, \bar{t}^i, \bar{t}^i\}} \quad w_n \bar{t}^i + w_f \bar{t}^i \quad (41a)$$

$$\text{s.t.} \quad \overline{\text{SE}}_k^i \geq \bar{t}^i, \forall k \in \mathcal{K}_N, \quad (41b)$$

$$\widetilde{\text{SE}}_k^i \geq \bar{t}^i, \forall k \in \mathcal{K}_F, \quad (41c)$$

$$\bar{t}^i \geq 0, \bar{t}^i \geq 0, \quad (41d)$$

$$(34b), (34e), (34f), \quad (41e)$$

where $\bar{t}^i \triangleq \min_{k \in \mathcal{K}_N} \overline{\text{SE}}_k^i$, $\bar{t}^i \triangleq \min_{k \in \mathcal{K}_F} \widetilde{\text{SE}}_k^i$. Since $\overline{\text{SE}}_k^i = \log_2(1 + \overline{\text{SINR}}_k^i)$, we can replace (41b) with the following two constraints

$$\overline{\text{SINR}}_k^i \geq \bar{T}^i, \forall k \in \mathcal{K}_N, \quad (42a)$$

$$\bar{T}^i \geq 2^{\bar{t}^i} - 1. \quad (42b)$$

Similarly, (41c) can be equivalently replaced by constraints

$$\widetilde{\text{SINR}}_k^i \geq \bar{T}^i, \forall k \in \mathcal{K}_F, \quad (43a)$$

$$\bar{T}^i \geq 2^{\bar{t}^i} - 1. \quad (43b)$$

Then, (P3) can be recast as

Algorithm 2 Penalty-Based Iterative Algorithm for (37)

```

1: Initialization: set initial point  $\mathbf{V}^{(0)}$ , the penalty factor  $\varrho$ ,
   maximum number of iterations  $I_1$ ,  $I_2$  for outer loop and
   inner loop.
2: repeat
3:   Set iteration index  $n = 0$  for inner loop;
4:   repeat
5:     Solve problem (40), get the solution  $(\mathbf{V}^*, t^*)$ ;
6:     Set  $n = n + 1$ , update  $\mathbf{V}^{(n)} = \mathbf{V}^*$ ;
7:   until  $\|\mathbf{V}\|_* - \|\mathbf{V}\|_2 \leq \epsilon$  or  $n = I_2$ .
8:   Set  $\varrho = l\varrho$ ;
9:   Update  $\mathbf{V}^{(0)}$  with  $\mathbf{V}^{(n)}$ .
10: until  $\|\mathbf{V}\|_* - \|\mathbf{V}\|_2 \leq \epsilon$  or  $n = I_1$ .

```

$$\text{(P3.1)} : \quad \max_{\{\bar{\eta}^i, \bar{\eta}^i, \bar{t}^i, \bar{t}^i, \bar{T}^i\}} \quad w_n \bar{t}^i + w_f \bar{t}^i \quad (44a)$$

$$\text{s.t.} \quad (41d), (41e), (42b), (43b), \quad (44b)$$

$$(42a), (43a). \quad (44c)$$

We observe that all the constraints listed in (44b) are convex and the difficulty in solving (P3.1) is due to the non-convexity of the constraints in (42a) and (43a). To this end, we propose a SCA-based design that approximates (42a), (43a) by convex approximation functions at each iteration. Note that the optimization problem varies for different precoding schemes due to differences in SINR expressions and sets of power control coefficients.

1) Power Control Design with MRT: Since (42a) and (43a) follow the same structure, we only present the iterative approximation for (42a). Then, following the same approach, we approximate (43a). To this end, we first define $\bar{\xi}_{sk} \triangleq \sqrt{\bar{\eta}_{sk}^{\text{MRT}}}$, $\forall k \in \mathcal{K}_N$, $\tilde{\xi}_{sk} \triangleq \sqrt{\tilde{\eta}_{sk}^{\text{MRT}}}$, $\forall k \in \mathcal{K}_F$. Then, we set $\bar{\boldsymbol{\xi}}_k \triangleq [\bar{\xi}_{1k}, \dots, \bar{\xi}_{S_k}]^T$, and $\tilde{\boldsymbol{\xi}} \triangleq [\tilde{\xi}_1^T, \dots, \tilde{\xi}_{K_N}^T]^T$ for NFUEs and $\bar{\boldsymbol{\xi}}_k \triangleq [\bar{\xi}_{1k}, \dots, \bar{\xi}_{S_k}]^T$, and $\tilde{\boldsymbol{\xi}} \triangleq [\tilde{\xi}_1^T, \dots, \tilde{\xi}_{K_F}^T]^T$ for FFUEs. Consequently, the optimization problem (P3.1) with MRT can be reformulated as

$$\text{(P3.2)} : \quad \max_{\mathbf{q}^{\text{MRT}}} \quad w_n \bar{t}^{\text{MRT}} + w_f \bar{t}^{\text{MRT}} \quad (45a)$$

$$\text{s.t.} \quad \sum_{s \in \mathcal{S}} \left(P_n \sum_{k \in \mathcal{U}_s} \bar{\xi}_{sk}^2 \bar{c}_{sk}^{\text{MRT}} + P_f \sum_{k \in \mathcal{I}_s} \tilde{\xi}_{sk}^2 \tilde{c}_{sk}^{\text{MRT}} \right) \leq P, \quad (45b)$$

$$\overline{\text{SINR}}_k^{\text{MRT}}(\bar{\boldsymbol{\xi}}, \tilde{\boldsymbol{\xi}}) \geq \bar{T}^{\text{MRT}}, \forall k \in \mathcal{K}_N, \quad (45c)$$

$$\widetilde{\text{SINR}}_k^{\text{MRT}}(\bar{\boldsymbol{\xi}}, \tilde{\boldsymbol{\xi}}) \geq \bar{T}^{\text{MRT}}, \forall k \in \mathcal{K}_F, \quad (45d)$$

$$\bar{\xi}_{sk} \geq 0, \tilde{\xi}_{sk} \geq 0, \forall s, \forall k, \quad (45e)$$

$$(41d), (42b), (43b), \quad (45f)$$

where $\mathbf{q}^{\text{MRT}} \triangleq \{\bar{\xi}^{\text{MRT}}, \tilde{\xi}^{\text{MRT}}, \bar{t}^{\text{MRT}}, \tilde{t}^{\text{MRT}}, \bar{T}^{\text{MRT}}, \tilde{T}^{\text{MRT}}\}$. Now, the remaining non-convexity of (P3.2) is the non-convex constraints (45c) and (45d). To address this problem, we apply SCA to approximate (45c). Note that (45c) can be written as

$$\frac{P_n \left(\sum_{s \in \mathcal{S}_k^{\text{MRT}}} \bar{\xi}_{sk} \bar{\Psi}_{kk}^s \right)^2}{\bar{T}^{\text{MRT}}} \geq \bar{F}_k^{\text{MRT}}(\bar{\boldsymbol{\xi}}, \tilde{\boldsymbol{\xi}}), \forall k \in \mathcal{K}_N, \quad (46)$$

$$\begin{aligned} \bar{F}_k^{\text{MRT}}(\bar{\xi}, \tilde{\xi}) &\leq \bar{F}_k^{\text{MRT},ub}(\bar{\xi}, \tilde{\xi}) \triangleq \frac{\rho_n}{4} \sum_{i \in \mathcal{K}_N \setminus k} \left(\sum_{s \in \bar{\mathcal{S}}_i^{\text{MRT}}} 4\bar{\xi}_{si}^2 |\bar{\Psi}_{ki}^s|^2 + 2 \sum_{s \in \bar{\mathcal{S}}_i^{\text{MRT}}} \sum_{s' \in \bar{\mathcal{S}}_i^{\text{MRT}}, s' > s} [(\bar{\xi}_{si} + \bar{\xi}_{s'i})^2 - 2(\bar{\xi}_{si}^{(n)} - \bar{\xi}_{s'i}^{(n)}) \right. \\ &\quad \times (\bar{\xi}_{si} - \bar{\xi}_{s'i}) + (\bar{\xi}_{si}^{(n)} - \bar{\xi}_{s'i}^{(n)})^2] \Re\{\bar{\Psi}_{ki}^s (\bar{\Psi}_{ki}^{s'})^*\} \Big) + \frac{\rho_f}{4} \sum_{j \in \mathcal{K}_F} \sum_{s \in \bar{\mathcal{S}}_j^{\text{MRT}}} \sum_{s' \in \bar{\mathcal{S}}_j^{\text{MRT}}} [(\tilde{\xi}_{sj} + \tilde{\xi}_{s'j})^2 \\ &\quad - 2(\tilde{\xi}_{sj}^{(n)} - \tilde{\xi}_{s'j}^{(n)}) (\tilde{\xi}_{sj} - \tilde{\xi}_{s'j}) + (\tilde{\xi}_{sj}^{(n)} - \tilde{\xi}_{s'j}^{(n)})^2] \tilde{\Xi}_{kj}^{ss'} + 1, \end{aligned} \quad (49)$$

$$\begin{aligned} \bar{F}_k^{\text{MRT}}(\bar{\xi}, \tilde{\xi}) &\leq \bar{F}_k^{\text{MRT},ub}(\bar{\xi}, \tilde{\xi}) \triangleq \frac{\rho_f}{4} \sum_{j \in \mathcal{K}_F} \sum_{s \in \bar{\mathcal{S}}_j^{\text{MRT}}} \sum_{s' \in \bar{\mathcal{S}}_j^{\text{MRT}}} [(\tilde{\xi}_{sj} + \tilde{\xi}_{s'j})^2 - 2(\tilde{\xi}_{sj}^{(n)} - \tilde{\xi}_{s'j}^{(n)}) (\tilde{\xi}_{sj} - \tilde{\xi}_{s'j}) + (\tilde{\xi}_{sj}^{(n)} - \tilde{\xi}_{s'j}^{(n)})^2] \tilde{d}_{kj}^{ss'} \\ &\quad + \frac{\rho_n}{4} \sum_{i \in \mathcal{K}_N} \sum_{s \in \bar{\mathcal{S}}_i^{\text{MRT}}} \sum_{s' \in \bar{\mathcal{S}}_i^{\text{MRT}}} [(\bar{\xi}_{si} + \bar{\xi}_{s'i})^2 - 2(\bar{\xi}_{si}^{(n)} - \bar{\xi}_{s'i}^{(n)}) (\bar{\xi}_{si} - \bar{\xi}_{s'i}) + (\bar{\xi}_{si}^{(n)} - \bar{\xi}_{s'i}^{(n)})^2] \tilde{\Xi}_{ki}^{ss'} + 1, \end{aligned} \quad (52)$$

where $\bar{F}_k^{\text{MRT}}(\bar{\xi}, \tilde{\xi}) \triangleq P_n \sum_{i \in \mathcal{K}_N \setminus k} \left(\sum_{s \in \bar{\mathcal{S}}_i^{\text{MRT}}} \bar{\xi}_{si} \bar{\Psi}_{ki}^s \right)^2 + P_f \sum_{j \in \mathcal{K}_F} \sum_{s \in \bar{\mathcal{S}}_j^{\text{MRT}}} \sum_{s' \in \bar{\mathcal{S}}_j^{\text{MRT}}} \tilde{\xi}_{sj} \tilde{\xi}_{s'j} \tilde{\Xi}_{kj}^{ss'} + \sigma_k^2$. The left-hand side (LHS) of (46) is a quadratic-over-linear function that is not jointly concave with respect to ξ_{sk} and T^{MRT} on the domain $T^{\text{MRT}} > 0$. In light of SCA, the LHS of (46) can be approximated by the following inequality [41]

$$\frac{x^2}{y} \geq \frac{x^{(n)}}{y^{(n)}} \left(2x - \frac{x^{(n)}}{y^{(n)}} y \right). \quad (47)$$

Moreover, on the right-hand side (RHS) of (46), $\bar{F}_k^{\text{MRT}}(\bar{\xi}, \tilde{\xi})$ is non-convex due to the presence of the term $\xi_{sj} \xi_{s'j}$. To address this issue, we replace the non-convex term by the first Taylor expansion around a point $(x^{(n)}, y^{(n)})$ as follows [41]

$$4xy \leq (x+y)^2 - 2(x^{(n)} - y^{(n)})(x-y) + (x^{(n)} - y^{(n)})^2. \quad (48)$$

Consequently, we obtain the convex upper bound of $\bar{F}_k^{\text{MRT}}(\bar{\xi}, \tilde{\xi})$ as $\bar{F}_k^{\text{MRT},ub}(\bar{\xi}, \tilde{\xi})$, shown in (49), where $\rho_n = \frac{P_n}{\sigma_k^2}$ and $\rho_f = \frac{P_f}{\sigma_k^2}$. Regarding (45d), we have

$$\frac{P_f \left(\sum_{s \in \bar{\mathcal{S}}_k^{\text{MRT}}} \tilde{\xi}_{sk} \tilde{\Psi}_k^s \right)^2}{\tilde{T}^{\text{MRT}}} \geq \tilde{F}_k^{\text{MRT}}(\bar{\xi}, \tilde{\xi}), \forall k \in \mathcal{K}_F, \quad (50)$$

where

$$\begin{aligned} \tilde{F}_k^{\text{MRT}}(\bar{\xi}, \tilde{\xi}) &\triangleq P_f \sum_{s \in \bar{\mathcal{S}}_k^{\text{MRT}}} \tilde{\xi}_{sk}^2 \tilde{a}_k^s \\ &\quad + P_f \sum_{j \in \mathcal{K}_F \setminus k} \sum_{s \in \bar{\mathcal{S}}_j^{\text{MRT}}} \sum_{s' \in \bar{\mathcal{S}}_j^{\text{MRT}}} \tilde{\xi}_{sj} \tilde{\xi}_{s'j} \tilde{b}_{kj}^{ss'} \\ &\quad + P_n \sum_{i \in \mathcal{K}_N} \sum_{s \in \bar{\mathcal{S}}_i^{\text{MRT}}} \sum_{s' \in \bar{\mathcal{S}}_i^{\text{MRT}}} \bar{\xi}_{si} \bar{\xi}_{s'i} \tilde{\Xi}_{ki}^{ss'} + \sigma_k^2. \end{aligned} \quad (51)$$

The LHS of (50) is handled in a similar way to that of (46). For the RHS of (50), the upper bound of $\tilde{F}_k^{\text{MRT}}(\bar{\xi}, \tilde{\xi})$ is given as $\tilde{F}_k^{\text{MRT},ub}(\bar{\xi}, \tilde{\xi})$ in (52), where

$$\tilde{d}_{kj}^{ss'} = \begin{cases} \tilde{a}_k^s & k = j, s = s', \\ 0 & k = j, s \neq s', \\ \tilde{b}_{kj}^{ss'} & k \neq j. \end{cases} \quad (53)$$

As a result, (P3.2) is transformed into the following optimization problem:

$$(P3.3) : \max_{\mathbf{q}^{\text{MRT}}} w_n \bar{t}^{\text{MRT}} + w_f \bar{t}^{\text{MRT}} \quad (54a)$$

$$\text{s.t. } v_k^{(n)} \left(2 \sum_{s \in \bar{\mathcal{S}}_k^{\text{MRT}}} \sqrt{\rho_n} \bar{\xi}_{sk} \bar{\Psi}_{kk}^s - v_k^{(n)} \bar{T}^{\text{MRT}} \right)$$

$$\geq \bar{F}_k^{\text{MRT},ub}(\bar{\xi}, \tilde{\xi}), \quad \forall k \in \mathcal{K}_N, \quad (54b)$$

$$\begin{aligned} &u_k^{(n)} \left(2 \sum_{s \in \bar{\mathcal{S}}_k^{\text{MRT}}} \sqrt{\rho_f} \tilde{\xi}_{sk} \tilde{\Psi}_k^s - u_k^{(n)} \bar{T}^{\text{MRT}} \right) \\ &\geq \tilde{F}_k^{\text{MRT},ub}(\bar{\xi}, \tilde{\xi}), \quad \forall k \in \mathcal{K}_F, \end{aligned} \quad (54c)$$

$$(41d), (42b), (43b), (45b), (45e), \quad (54d)$$

where

$$v_k^{(n)} = \frac{\sqrt{\rho_n} \sum_{s \in \bar{\mathcal{S}}_k^{\text{MRT}}} \bar{\xi}_{sk}^{(n)} \bar{\Psi}_{kk}^s}{(\bar{T}^{\text{MRT}})^{(n)}},$$

$$u_k^{(n)} = \frac{\sqrt{\rho_f} \sum_{s \in \bar{\mathcal{S}}_k^{\text{MRT}}} \tilde{\xi}_{sk}^{(n)} \tilde{\Psi}_k^s}{(\bar{T}^{\text{MRT}})^{(n)}}.$$

In addition, if $\tilde{d}_{kj}^{ss'}$ is negative, a first-order Taylor expansion should be applied to $(\tilde{\xi}_{sj} + \tilde{\xi}_{s'j})^2$ in (52). A similar procedure can also be applied to the other terms in (52) and (49), where the details are omitted for brevity. Note that (54) is a convex optimization problem. Modern convex solvers, such as MOSEK, are capable of efficiently solving (54) for a relatively large problem size, which is sufficient to characterize the performance of XL-MIMO systems. The SCA-based algorithm proposed to solve the problem (45) can be summarized in **Algorithm 3**.

2) *Power Control Design with CZF*: Recall that to eliminate interference with the CZF design, we set the power control coefficients $\bar{\eta}_k^{\text{CZF}} \triangleq \bar{\eta}_{1k}^{\text{CZF}} = \dots = \bar{\eta}_{Sk}^{\text{CZF}}$ ($\bar{\eta}_k^{\text{CZF}} \triangleq \bar{\eta}_{1k}^{\text{CZF}} = \dots = \bar{\eta}_{Sk}^{\text{CZF}}$) for any NFUE (FFUE) k . Then, we have the following optimization problem:

$$(P4) : \max_{\mathbf{q}^{\text{CZF}}} w_n \bar{t}^{\text{CZF}} + w_f \bar{t}^{\text{CZF}} \quad (55a)$$

$$\begin{aligned} \text{s.t. } &P_n \sum_{k \in \mathcal{K}_N} \bar{\eta}_k^{\text{CZF}} \sum_{s \in \mathcal{S}} \bar{c}_{sk} \\ &+ P_f \sum_{k \in \mathcal{K}_F} \bar{\eta}_k^{\text{CZF}} \sum_{s \in \mathcal{S}} \bar{c}_{sk} \leq P, \end{aligned} \quad (55b)$$

$$\overline{\text{SINR}}_k^{\text{CZF}}(\bar{\eta}^{\text{CZF}}, \bar{\eta}^{\text{CZF}}) \geq \bar{T}^{\text{CZF}}, \forall k \in \mathcal{K}_N, \quad (55c)$$

$$\widetilde{\text{SINR}}_k^{\text{CZF}}(\bar{\eta}^{\text{CZF}}, \bar{\eta}^{\text{CZF}}) \geq \bar{T}^{\text{CZF}}, \forall k \in \mathcal{K}_F, \quad (55d)$$

$$\bar{\eta}_k^{\text{CZF}} \geq 0, \bar{\eta}_k^{\text{CZF}} \geq 0, \forall k, \quad (55e)$$

$$(41d), (42b), (43b), \quad (55f)$$

where $\mathbf{q}^{\text{CZF}} \triangleq \{\bar{\eta}^{\text{CZF}}, \bar{\eta}^{\text{CZF}}, \bar{t}^{\text{CZF}}, \bar{t}^{\text{CZF}}, \bar{T}^{\text{CZF}}, \bar{T}^{\text{CZF}}\}$. Since $\bar{T}^{\text{CZF}} > 0$, we can rewrite (55c), for $\forall k \in \mathcal{K}_N$ as

$$P_n \bar{\eta}_k^{\text{CZF}} \left(\sum_{s \in \bar{\mathcal{S}}_k^{\text{CZF}}} \bar{\mathbf{g}}_{sk}^H \bar{\mathbf{w}}_{sk}^{\text{CZF}} \right)^2 \geq \bar{F}_k^{\text{CZF}}(\bar{\eta}^{\text{CZF}}, \bar{\eta}^{\text{CZF}}) \bar{T}^{\text{CZF}}, \quad (56)$$

Algorithm 3 Proposed SCA-based Algorithm for (45)

- 1: Initialization: set $n = 1$, and generate initial points $(\tilde{\xi}^{(1)}, \tilde{\xi}^{(1)}, (\bar{T}^{\text{MRT}})^{(1)}, (\tilde{T}^{\text{MRT}})^{(1)})$. Define a tolerance ϵ_1 and the maximum number of iterations I_3 .
 - 2: Iteration n : solve (54). Let $(\tilde{\xi}^*, \tilde{\xi}^*, (\bar{T}^{\text{MRT}})^*, (\tilde{T}^{\text{MRT}})^*)$ be the solution.
 - 3: If the fractional increase of the objective function value is below a threshold $\epsilon_1 > 0$ or $n = I_3$, stop. Otherwise, go to step 4.
 - 4: Set $n = n + 1$, update $(\tilde{\xi}^{(n)}, \tilde{\xi}^{(n)}, (\bar{T}^{\text{MRT}})^{(n)}, (\tilde{T}^{\text{MRT}})^{(n)}) = (\tilde{\xi}^*, \tilde{\xi}^*, (\bar{T}^{\text{MRT}})^*, (\tilde{T}^{\text{MRT}})^*)$, go to step 2.
-

where

$$\begin{aligned} \bar{F}_k^{\text{CZF}}(\bar{\eta}^{\text{CZF}}, \tilde{\eta}^{\text{CZF}}) &= P_n \sum_{i \in \mathcal{K}_{\mathcal{N}} \setminus k} \bar{\eta}_i^{\text{CZF}} \left| \sum_{s \in \tilde{\mathcal{S}}_i^{\text{CZF}}} \bar{\mathbf{g}}_{sk}^H \bar{\mathbf{w}}_{si}^{\text{CZF}} \right|^2 \\ &+ P_f \sum_{j \in \mathcal{K}_{\mathcal{F}}} \tilde{\eta}_j^{\text{CZF}} r_{kj} + \sigma_k^2. \end{aligned}$$

We note that the LHS of (56) is linear with respect to $\bar{\eta}^{\text{CZF}}$, while the RHS is not convex, e.g., the existence of the term $\bar{\eta}_i^{\text{CZF}} \bar{T}^{\text{CZF}}$. Using SCA, we derive the convex upper bound of $\bar{F}_k^{\text{CZF}}(\bar{\eta}^{\text{CZF}}, \tilde{\eta}^{\text{CZF}}) \bar{T}^{\text{CZF}}$ as $\bar{U}_k^{\text{CZF}, ub}(\bar{\eta}^{\text{CZF}}, \tilde{\eta}^{\text{CZF}}, \bar{T}^{\text{CZF}})$ given in (57). Similarly, the FF SINR constraint (55d) can be expressed as

$$P_f \tilde{\eta}_k^{\text{CZF}} \left(\sum_{s \in \tilde{\mathcal{S}}_k^{\text{CZF}}} \mathbb{E}\{\tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{w}}_{sk}^{\text{CZF}}\} \right)^2 \geq \tilde{F}_k^{\text{CZF}}(\bar{\eta}^{\text{CZF}}, \tilde{\eta}^{\text{CZF}}) \tilde{T}^{\text{CZF}}, \quad (58)$$

where

$$\begin{aligned} \tilde{F}_k^{\text{CZF}}(\bar{\eta}^{\text{CZF}}, \tilde{\eta}^{\text{CZF}}) &= P_f \sum_{j \in \mathcal{K}_{\mathcal{F}} \setminus k} \tilde{\eta}_j^{\text{CZF}} \mathbb{E}\left\{ \left| \sum_{s \in \tilde{\mathcal{S}}_j^{\text{CZF}}} \tilde{\mathbf{g}}_{sj}^H \tilde{\mathbf{w}}_{sj}^{\text{CZF}} \right|^2 \right\} \\ &+ P_f \tilde{\eta}_k^{\text{CZF}} \varepsilon_k + P_n \sum_{i \in \mathcal{K}_{\mathcal{N}}} \bar{\eta}_i^{\text{CZF}} t_{ki} + \sigma_k^2. \end{aligned} \quad (59)$$

The RHS of (58) is non-convex due to the presence of multiplicative terms involving optimization parameters. To address this, we use SCA and derive a convex upper-bound for the RHS of (58) as $\tilde{U}_k^{\text{CZF}, ub}(\bar{\eta}^{\text{CZF}}, \tilde{\eta}^{\text{CZF}}, \tilde{T}^{\text{CZF}})$, shown in (60), where $\tilde{r}_{kj}$ is given by

$$\tilde{r}_{kj} = \begin{cases} \varepsilon_k & k = j, \\ \mathbb{E}\left\{ \left| \sum_{s \in \tilde{\mathcal{S}}_j^{\text{CZF}}} \tilde{\mathbf{g}}_{sj}^H \tilde{\mathbf{w}}_{sj}^{\text{CZF}} \right|^2 \right\} & k \neq j. \end{cases} \quad (61)$$

As a result, the optimization problem (P4) is recast as the

following convex optimization problem:

$$(P4.1) : \max_{\mathbf{q}^{\text{CZF}}} w_n \bar{t}^{\text{CZF}} + w_f \tilde{t}^{\text{CZF}} \quad (62a)$$

$$\begin{aligned} \text{s.t. } & \rho_n \bar{\eta}_k^{\text{CZF}} \left(\sum_{s \in \tilde{\mathcal{S}}_k^{\text{CZF}}} \bar{\mathbf{g}}_{sk}^H \bar{\mathbf{w}}_{sk}^{\text{CZF}} \right)^2 \\ & \geq \bar{U}_k^{\text{CZF}, ub}(\bar{\eta}^{\text{CZF}}, \tilde{\eta}^{\text{CZF}}, \bar{T}^{\text{CZF}}), \forall k \in \mathcal{K}_{\mathcal{N}}, \end{aligned} \quad (62b)$$

$$\begin{aligned} & \rho_f \tilde{\eta}_k^{\text{CZF}} \left(\sum_{s \in \tilde{\mathcal{S}}_k^{\text{CZF}}} \mathbb{E}\{\tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{w}}_{sk}^{\text{CZF}}\} \right)^2 \\ & \geq \tilde{U}_k^{\text{CZF}, ub}(\bar{\eta}^{\text{CZF}}, \tilde{\eta}^{\text{CZF}}, \tilde{T}^{\text{CZF}}), \forall k \in \mathcal{K}_{\mathcal{F}}, \end{aligned} \quad (62c)$$

$$(41d), (42b), (43b), (55b), (55e). \quad (62d)$$

The problem (P4.1) can be effectively solved by modern convex solvers such as CVX [40]. We can adapt **Algorithm 3** with the appropriate modifications to solve (62). The details are omitted for the sake of brevity.

3) *Power Control Design with LZLF*: With LZLF, the optimization problem (P3) can be expressed as

$$(P5) : \max_{\mathbf{q}^{\text{LZF}}} w_n \bar{t}^{\text{LZF}} + w_f \tilde{t}^{\text{LZF}} \quad (63a)$$

$$\begin{aligned} \text{s.t. } & \sum_{s \in \mathcal{S}} \left(P_n \sum_{k \in \mathcal{U}_s} \bar{\psi}_{sk}^2 \bar{c}_{sk}^{\text{LZF}} \right. \\ & \left. + P_f \sum_{k \in \mathcal{I}_s} \tilde{\psi}_{sk}^2 \tilde{c}_{sk}^{\text{LZF}} \right) \leq P, \end{aligned} \quad (63b)$$

$$\overline{\text{SINR}}_k^{\text{LZF}}(\bar{\psi}, \tilde{\psi}) \geq \bar{T}^{\text{LZF}}, \forall k \in \mathcal{K}_{\mathcal{N}}, \quad (63c)$$

$$\widetilde{\text{SINR}}_k^{\text{LZF}}(\bar{\psi}, \tilde{\psi}) \geq \tilde{T}^{\text{LZF}}, \forall k \in \mathcal{K}_{\mathcal{F}}, \quad (63d)$$

$$\bar{\psi}_{sk} \geq 0, \tilde{\psi}_{sk} \geq 0, \forall s, \forall k, \quad (63e)$$

$$(41d), (42b), (43b). \quad (63f)$$

where $\mathbf{q}^{\text{LZF}} \triangleq \{\bar{\psi}, \tilde{\psi}, \bar{t}^{\text{LZF}}, \tilde{t}^{\text{LZF}}, \bar{T}^{\text{LZF}}, \tilde{T}^{\text{LZF}}\}$, and $\bar{\psi} \triangleq [\bar{\psi}_1, \dots, \bar{\psi}_{K_{\mathcal{N}}}]$, where $\bar{\psi}_k \triangleq [\bar{\psi}_{1k}, \dots, \bar{\psi}_{S_k}]^T$ with $\bar{\psi}_{sk} \triangleq \sqrt{\bar{\eta}_{sk}^{\text{LZF}}}$, $\forall s, k \in \mathcal{K}_{\mathcal{N}}$; $\tilde{\psi} \triangleq [\tilde{\psi}_1, \dots, \tilde{\psi}_{K_{\mathcal{F}}}]$, where $\tilde{\psi}_k \triangleq [\tilde{\psi}_{1k}, \dots, \tilde{\psi}_{S_k}]^T$ with $\tilde{\psi}_{sk} \triangleq \sqrt{\tilde{\eta}_{sk}^{\text{LZF}}}$, $\forall s, k \in \mathcal{K}_{\mathcal{F}}$. Constraints (63c) and (63d) can be recast as

$$\frac{P_n \left(\sum_{s \in \tilde{\mathcal{S}}_k^{\text{LZF}}} \bar{\psi}_{sk} \right)^2}{\bar{T}^{\text{LZF}}} \geq \bar{F}_k^{\text{LZF}}(\bar{\psi}, \tilde{\psi}), \forall k \in \mathcal{K}_{\mathcal{N}}, \quad (64a)$$

$$\frac{P_f \left(\sum_{s \in \tilde{\mathcal{S}}_k^{\text{LZF}}} \tilde{\psi}_{sk} \right)^2}{\tilde{T}^{\text{LZF}}} \geq \tilde{F}_k^{\text{LZF}}(\bar{\psi}, \tilde{\psi}), \forall k \in \mathcal{K}_{\mathcal{F}}, \quad (64b)$$

where $\bar{F}_k^{\text{LZF}}(\bar{\psi}, \tilde{\psi})$ and $\tilde{F}_k^{\text{LZF}}(\bar{\psi}, \tilde{\psi})$ are given by

$$\begin{aligned} \bar{F}_k^{\text{CZF}}(\bar{\eta}^{\text{CZF}}, \tilde{\eta}^{\text{CZF}}) \bar{T}^{\text{CZF}} &\leq \bar{U}_k^{\text{CZF}, ub}(\bar{\eta}^{\text{CZF}}, \tilde{\eta}^{\text{CZF}}, \bar{T}^{\text{CZF}}) \triangleq \frac{\rho_n}{4} \sum_{i \in \mathcal{K}_{\mathcal{N}} \setminus k} \left[(\bar{\eta}_i^{\text{CZF}} + \bar{T}^{\text{CZF}})^2 - 2 \left((\bar{\eta}_i^{\text{CZF}})^{(n)} - (\bar{T}^{\text{CZF}})^{(n)} \right) (\bar{\eta}_i^{\text{CZF}} - \bar{T}^{\text{CZF}}) \right. \\ &+ \left. \left((\bar{\eta}_i^{\text{CZF}})^{(n)} - (\bar{T}^{\text{CZF}})^{(n)} \right)^2 \right] \left| \sum_{s \in \tilde{\mathcal{S}}_i^{\text{CZF}}} \bar{\mathbf{g}}_{sk}^H \bar{\mathbf{w}}_{si}^{\text{CZF}} \right|^2 + \frac{\rho_f}{4} \sum_{j \in \mathcal{K}_{\mathcal{F}}} \left[(\tilde{\eta}_j^{\text{CZF}} + \tilde{T}^{\text{CZF}})^2 - 2 \left((\tilde{\eta}_j^{\text{CZF}})^{(n)} - (\tilde{T}^{\text{CZF}})^{(n)} \right) (\tilde{\eta}_j^{\text{CZF}} - \tilde{T}^{\text{CZF}}) \right. \\ &\times \left. \left(\tilde{\eta}_j^{\text{CZF}} - \tilde{T}^{\text{CZF}} \right) + \left. \left((\tilde{\eta}_j^{\text{CZF}})^{(n)} - (\tilde{T}^{\text{CZF}})^{(n)} \right)^2 \right] r_{kj} + \bar{T}^{\text{CZF}}, \end{aligned} \quad (57)$$

$$\begin{aligned} \tilde{F}_k^{\text{CZF}}(\bar{\eta}^{\text{CZF}}, \tilde{\eta}^{\text{CZF}}) \tilde{T}^{\text{CZF}} &\leq \tilde{U}_k^{\text{CZF}, ub}(\bar{\eta}^{\text{CZF}}, \tilde{\eta}^{\text{CZF}}, \tilde{T}^{\text{CZF}}) \triangleq \frac{\rho_f}{4} \sum_{j \in \mathcal{K}_{\mathcal{F}}} \left[(\tilde{\eta}_j^{\text{CZF}} + \tilde{T}^{\text{CZF}})^2 - 2 \left((\tilde{\eta}_j^{\text{CZF}})^{(n)} - (\tilde{T}^{\text{CZF}})^{(n)} \right) (\tilde{\eta}_j^{\text{CZF}} - \tilde{T}^{\text{CZF}}) \right. \\ &+ \left. \left((\tilde{\eta}_j^{\text{CZF}})^{(n)} - (\tilde{T}^{\text{CZF}})^{(n)} \right)^2 \right] \tilde{r}_{kj} + \frac{\rho_n}{4} \sum_{i \in \mathcal{K}_{\mathcal{N}}} \left[(\bar{\eta}_i^{\text{CZF}} + \bar{T}^{\text{CZF}})^2 - 2 \left((\bar{\eta}_i^{\text{CZF}})^{(n)} - (\bar{T}^{\text{CZF}})^{(n)} \right) (\bar{\eta}_i^{\text{CZF}} - \bar{T}^{\text{CZF}}) \right. \\ &+ \left. \left((\bar{\eta}_i^{\text{CZF}})^{(n)} - (\bar{T}^{\text{CZF}})^{(n)} \right)^2 \right] t_{ki} + \tilde{T}^{\text{CZF}}, \end{aligned} \quad (60)$$

$$\begin{aligned} \bar{F}_k^{\text{LZF}}(\bar{\psi}, \tilde{\psi}) &\leq \bar{F}_k^{\text{LZF},ub}(\bar{\psi}, \tilde{\psi}) \triangleq \frac{\rho_n}{4} \sum_{i \in \mathcal{K}_{\mathcal{N}} \setminus k} \left(\sum_{s \in \tilde{\mathcal{S}}_i^{\text{LZF}}} 4\tilde{\psi}_{si}^2 |\bar{\mathbf{g}}_{sk}^H \bar{\mathbf{w}}_{si}^{\text{LZF}}|^2 + 2 \sum_{s \in \tilde{\mathcal{S}}_i^{\text{LZF}}} \sum_{s' \in \tilde{\mathcal{S}}_i^{\text{LZF}}, s' > s} [(\bar{\psi}_{si} + \bar{\psi}_{s'i})^2 \right. \\ &\quad \left. - 2(\bar{\psi}_{si}^{(n)} - \bar{\psi}_{s'i}^{(n)})(\bar{\psi}_{si} - \bar{\psi}_{s'i}) + (\bar{\psi}_{si}^{(n)} - \bar{\psi}_{s'i}^{(n)})^2] \Re\{\bar{\mathbf{g}}_{sk}^H \bar{\mathbf{w}}_{si}^{\text{LZF}} (\bar{\mathbf{g}}_{s'k}^T \bar{\mathbf{w}}_{s'i}^{\text{LZF}})^*\} \right) + \frac{\rho_f}{4} \\ &\quad \times \sum_{j \in \mathcal{K}_{\mathcal{F}}} \sum_{s \in \tilde{\mathcal{S}}_j^{\text{LZF}}} \sum_{s' \in \tilde{\mathcal{S}}_j^{\text{LZF}}} [(\tilde{\psi}_{sj} + \tilde{\psi}_{s'j})^2 - 2(\tilde{\psi}_{sj}^{(n)} - \tilde{\psi}_{s'j}^{(n)})(\tilde{\psi}_{sj} - \tilde{\psi}_{s'j}) + (\tilde{\psi}_{sj}^{(n)} - \tilde{\psi}_{s'j}^{(n)})^2] \tilde{\tau}_{kj}^{ss'} + 1, \end{aligned} \quad (66a)$$

$$\begin{aligned} \tilde{F}_k^{\text{LZF}}(\bar{\psi}, \tilde{\psi}) &\leq \tilde{F}_k^{\text{LZF},ub}(\bar{\psi}, \tilde{\psi}) \triangleq \frac{\rho_f}{4} \sum_{j \in \mathcal{K}_{\mathcal{F}} \setminus k} \left(\sum_{s \in \tilde{\mathcal{S}}_j^{\text{LZF}}} 4\tilde{\psi}_{sj}^2 \mathbb{E}\{|\bar{\mathbf{g}}_{sk}^H \tilde{\mathbf{w}}_{sj}^{\text{LZF}}|^2\} + 2 \sum_{s \in \tilde{\mathcal{S}}_j^{\text{LZF}}} \sum_{s' \in \tilde{\mathcal{S}}_j^{\text{LZF}}, s' > s} [(\tilde{\psi}_{sj} + \tilde{\psi}_{s'j})^2 \right. \\ &\quad \left. - 2(\tilde{\psi}_{sj}^{(n)} - \tilde{\psi}_{s'j}^{(n)})(\tilde{\psi}_{sj} - \tilde{\psi}_{s'j}) + (\tilde{\psi}_{sj}^{(n)} - \tilde{\psi}_{s'j}^{(n)})^2] \Re\{\mathbb{E}\{\bar{\mathbf{g}}_{sk}^H \tilde{\mathbf{w}}_{sj}^{\text{LZF}} (\bar{\mathbf{g}}_{s'k}^T \tilde{\mathbf{w}}_{s'j}^{\text{LZF}})^*\}\} \right) + \frac{\rho_n}{4} \\ &\quad \times \sum_{i \in \mathcal{K}_{\mathcal{N}}} \sum_{s \in \tilde{\mathcal{S}}_i^{\text{LZF}}} \sum_{s' \in \tilde{\mathcal{S}}_i^{\text{LZF}}} [(\bar{\psi}_{si} + \bar{\psi}_{s'i})^2 - 2(\bar{\psi}_{si}^{(n)} - \bar{\psi}_{s'i}^{(n)})(\bar{\psi}_{si} - \bar{\psi}_{s'i}) + (\bar{\psi}_{si}^{(n)} - \bar{\psi}_{s'i}^{(n)})^2] \tilde{\tau}_{ki}^{ss'} + 1, \end{aligned} \quad (66b)$$

$$\begin{aligned} \bar{F}_k^{\text{LZF}}(\bar{\psi}, \tilde{\psi}) &\triangleq P_n \sum_{i \in \mathcal{K}_{\mathcal{N}} \setminus k} \left| \sum_{s \in \tilde{\mathcal{S}}_i^{\text{LZF}}} \bar{\psi}_{si} \bar{\mathbf{g}}_{sk}^H \bar{\mathbf{w}}_{si}^{\text{LZF}} \right|^2 \\ &\quad + P_f \sum_{j \in \mathcal{K}_{\mathcal{F}}} \sum_{s \in \tilde{\mathcal{S}}_j^{\text{LZF}}} \sum_{s' \in \tilde{\mathcal{S}}_j^{\text{LZF}}} \tilde{\psi}_{sj} \tilde{\psi}_{s'j} \tilde{\tau}_{kj}^{ss'} + \sigma_k^2, \end{aligned} \quad (65a)$$

$$\begin{aligned} \tilde{F}_k^{\text{LZF}}(\bar{\psi}, \tilde{\psi}) &\triangleq P_f \sum_{j \in \mathcal{K}_{\mathcal{F}} \setminus k} \mathbb{E}\left\{ \left| \sum_{s \in \tilde{\mathcal{S}}_j^{\text{LZF}}} \tilde{\psi}_{sj} \bar{\mathbf{g}}_{sk}^H \tilde{\mathbf{w}}_{sj}^{\text{LZF}} \right|^2 \right\} \\ &\quad + P_n \sum_{i \in \mathcal{K}_{\mathcal{N}}} \sum_{s \in \tilde{\mathcal{S}}_i^{\text{LZF}}} \sum_{s' \in \tilde{\mathcal{S}}_i^{\text{LZF}}} \bar{\psi}_{si} \bar{\psi}_{s'i} \tilde{\tau}_{ki}^{ss'} + \sigma_k^2. \end{aligned} \quad (65b)$$

The two constraints (64a) and (64b) are non-convex. To address this issue, we apply SCA and replace $\bar{F}_k^{\text{LZF}}(\bar{\psi}, \tilde{\psi})$ and $\tilde{F}_k^{\text{LZF}}(\bar{\psi}, \tilde{\psi})$ by their convex upper bounds $\bar{F}_k^{\text{LZF},ub}(\bar{\psi}, \tilde{\psi})$, $\tilde{F}_k^{\text{LZF},ub}(\bar{\psi}, \tilde{\psi})$ given in (66a) and (66b), respectively, at the top of the next page. As a result, the final convex optimization problem for LZLF is given by

$$(P5.1) : \max_{\mathbf{q}^{\text{LZF}}} w_n \bar{t}^{\text{LZF}} + w_f \tilde{t}^{\text{LZF}} \quad (67a)$$

$$\begin{aligned} \text{s.t.} \quad (u_k^{\text{LZF}})^{(n)} &\left(2 \sum_{s \in \tilde{\mathcal{S}}_k^{\text{LZF}}} \sqrt{\rho_f} \tilde{\psi}_{sk} - (u_k^{\text{LZF}})^{(n)} \right. \\ &\quad \left. \times \tilde{T}^{\text{LZF}} \right) \geq \bar{F}_k^{\text{LZF},ub}(\bar{\psi}, \tilde{\psi}), \forall k \in \mathcal{K}_{\mathcal{F}}, \end{aligned} \quad (67b)$$

$$\begin{aligned} (v_k^{\text{LZF}})^{(n)} &\left(2 \sum_{s \in \tilde{\mathcal{S}}_k^{\text{LZF}}} \sqrt{\rho_n} \bar{\psi}_{sk} - (v_k^{\text{LZF}})^{(n)} \tilde{T}^{\text{LZF}} \right) \\ &\geq \tilde{F}_k^{\text{LZF},ub}(\bar{\psi}, \tilde{\psi}), \forall k \in \mathcal{K}_{\mathcal{N}}, \end{aligned} \quad (67c)$$

$$(41d), (42b), (43b), (63b), (63e), \quad (67d)$$

where

$$\begin{aligned} (u_k^{\text{LZF}})^{(n)} &= \frac{\sqrt{\rho_f} \sum_{s \in \tilde{\mathcal{S}}_k^{\text{LZF}}} (\tilde{\xi}_{sk}^{\text{LZF}})^{(n)}}{(\tilde{T}^{\text{LZF}})^{(n)}}, \\ (v_k^{\text{LZF}})^{(n)} &= \frac{\sqrt{\rho_n} \sum_{s \in \tilde{\mathcal{S}}_k^{\text{LZF}}} (\bar{\xi}_{sk}^{\text{LZF}})^{(n)}}{(\tilde{T}^{\text{LZF}})^{(n)}}. \end{aligned}$$

The problem (P5.1) can be effectively solved by CVX [40]. We can adapt **Algorithm 3** with appropriate modifications to solve (67). The details are omitted for brevity.

D. Complexity and Convergence Analysis

Computational complexity and convergence analysis of Algorithm 2: Since the main complexity comes from solving (P2.3) in the inner loop, the complexity of solving this problem is $\mathcal{O}(N^{3.5})$ if the solver is based on the interior point

method [42]. Then, the overall complexity of **Algorithm 2** is $\mathcal{O}(I_{out} I_{in} N^{3.5})$, where I_{out} , I_{in} represent the number of outer and inner iterations for convergence, respectively.

We now study the convergence of **Algorithm 2**, where we adopt SCA to solve problem (P2) to deal with the non-concave penalty term $\varrho(\|\mathbf{V}\|_2)$ in the objective function. Let $f_{org}(t, \mathbf{V}) = t - \varrho(\|\mathbf{V}\|_* - \|\mathbf{V}\|_2)$ denote the original objective. In iteration (n), we linearize the convex term $\|\mathbf{V}\|_2$ as $\bar{\mathbf{V}}^{(n)}$, and obtain the lower bound of the original objective as $f_{lb}^{(n)}(t, \mathbf{V}) = t - \varrho(\|\mathbf{V}\|_* - \bar{\mathbf{V}}^{(n)})$. Since $f_{lb}^{(n)}(t, \mathbf{V})$ is a global lower bound, we have

$$f_{org}(t, \mathbf{V}) \geq f_{lb}^{(n)}(t, \mathbf{V}), \forall (t, \mathbf{V}), \quad (69)$$

and

$$f_{org}(t^{(n)}, \mathbf{V}^{(n)}) = f_{lb}^{(n)}(t^{(n)}, \mathbf{V}^{(n)}), \quad (70)$$

holds at the current point $(t^{(n)}, \mathbf{V}^{(n)})$ since $f_{lb}^{(n)}(t, \mathbf{V})$ is built based on point $(t^{(n)}, \mathbf{V}^{(n)})$.

In addition, we maximize $f_{lb}^{(n)}(t, \mathbf{V})$, which implies that

$$f_{lb}^{(n)}(t^{(n+1)}, \mathbf{V}^{(n+1)}) \geq f_{lb}^{(n)}(t^{(n)}, \mathbf{V}^{(n)}), \quad (71)$$

Combining (69), (70) and (71), we obtain the following inequality chain [43]:

$$\begin{aligned} f_{org}(t^{(n+1)}, \mathbf{V}^{(n+1)}) &\geq f_{lb}^{(n)}(t^{(n+1)}, \mathbf{V}^{(n+1)}) \\ &\geq f_{lb}^{(n)}(t^{(n)}, \mathbf{V}^{(n)}) \\ &= f_{org}(t^{(n)}, \mathbf{V}^{(n)}). \end{aligned} \quad (72)$$

The first inequality holds because $f_{lb}^{(n)}(t, \mathbf{V})$ is a global lower bound on $f_{org}(t, \mathbf{V})$ for all $(t, \mathbf{V})$, by construction of the linear approximation of the convex term $\|\mathbf{V}\|_2$. The second inequality follows from the fact that we maximize $f_{lb}^{(n)}(t, \mathbf{V})$ in each iteration, and the final equality holds since the approximation is exact at iteration (n). Since $t \leq \text{tr}(\mathbf{R}_k \mathbf{V})$ is bounded above, so the objective of **Algorithm 2** is non-decreasing and guarantees to converge to a limit.

Computational complexity and convergence analysis of Algorithm 3: The relaxed optimization problems (P3.3), (P4.1) and (P5.1) each entail a computational complexity of $\mathcal{O}(\sqrt{A_l^i + A_q^i + 3A_{\text{exp}}^i}(A_v^i + A_l^i + A_q^i + A_{\text{exp}}^i)(A_v^i)^2)$ [41], [43],

TABLE III: Summary of optimization problems' parameters for computational complexity analysis

Problem	A_v^i	A_l^i
P3.3 (MRT)	$ \mathcal{S}^{\text{MRT}} K_n + \mathcal{S}^{\text{MRT}} K_f + 4$	$ \mathcal{S}^{\text{MRT}} K_n + \mathcal{S}^{\text{MRT}} K_f + 2$
P4.1 (CZF)	$K + 4$	$K + 3$
P5.1 (LZF)	$ \mathcal{S}^{\text{LZF}} K_n + \mathcal{S}^{\text{LZF}} K_f + 4$	$ \mathcal{S}^{\text{LZF}} K_n + \mathcal{S}^{\text{LZF}} K_f + 2$

where $i \in \{\text{MRT}, \text{CZF}, \text{LZF}\}$. Here, A_v^i represents the number of real-valued scalar variables, while A_l^i , A_q^i and $A_{\text{exp}}^i = 2$ denote the number of linear, quadratic and exponential cone constraints, respectively. For tractability, we consider an ideal scenario in which the VR selected by users within each group includes the same number of subarrays, e.g. $|\mathcal{S}^{\text{MRT}}| = |\mathcal{S}^{\text{MRT}}|$ for FFUEs with MRT. Table III summarizes the values of A_v^i and A_l^i for various optimization problems, while $A_q^i = K + 1$ for MRT and LZF and $A_q^i = K$ for CZF.

Let $(t^i)^{(n)} = w_n(\bar{t}^i)^{(n)} + w_f(\tilde{t}^i)^{(n)}$ denote the optimized objective and $(\mathbf{q}^i)^{(n)}$ be the set of optimized solutions at iteration (n) . Due to the linear approximation in (47) and (48), the updating rule in **Algorithm 3** ensures that $(\mathbf{q}^i)^{(n)}$ is feasible for the optimization problem at iteration $(n + 1)$. More specifically, at iteration (n) , we use $(\mathbf{q}^i)^{(n)}$ as a feasible point to solve the optimization problem and obtain $(\mathbf{q}^i)^{(n+1)}$. Since we maximize the sum of weighted SE, the inequality $(t^i)^{(n+1)} \geq (t^i)^{(n)}$ holds, which implies that **Algorithm 3** yields a non-decreasing sequence. We also note that due to the total power constraint (34b), the objective of (41) is bounded from above. Thus, **Algorithm 3** guarantees the convergence of the objective function in (P3) for different precoding schemes.

V. SIMULATION RESULTS

A. Simulation Setup

We assume that the XL-MIMO array is divided into S subarrays, each consisting of $M^* = 50$ elements. Unless otherwise stated, we set $M_y = 10$ and increase M_x linearly. The RIS is equipped with $N = 10 \times 10$ elements. The distance between the XL-MIMO and the RIS is $d_{\text{MR}} = 100$ m. We consider $K_f = 5$ FFUEs located in a semicircle and the radius is $d_{\text{RU}} = 20$ m. There are $K_n = 5$ NFUEs randomly distributed in an area with $x \in [0, 20]$, $y \in [0, 20]$ and $z \in [2, 20]$, the locations of NFUEs with respect to XL-MIMO is illustrated in Fig. 2. We assume that the Ricean factor is $\iota = 2$, while the path-loss coefficients are $\varsigma_k = 10^{-3}d_{\text{RU}}^{-2}$ and $\zeta = 10^{-3}d_{\text{MR}}^{-2.5}$, respectively [44]. We set the maximum number of iterations $I_1 = I_2 = 30$ for **Algorithm 2** and $I_3 = 30$ for **Algorithm 3**. The thresholds are $\epsilon = 10^{-4}$, $\epsilon_1 = 10^{-3}$ and we set the initial penalty factor $\varrho = 10^{-6}$, the scaling factor $l = 10$ for **Algorithm 2**. The other parameters are set as follows: the carrier frequency of the system is 5 GHz, the transmit power is $P = 30$ dBm, and the noise power is -104 dBm [44].

To evaluate the performance of the proposed design, namely optimized phase shifts design and power control (**OPS-OPC**), we consider a benchmark scheme, random phase shifts and equal power control (**RPS-EPC**) design, where the transmit power P is equally divided between all users with random phase shift matrix at RIS. We also include the heuristic phase

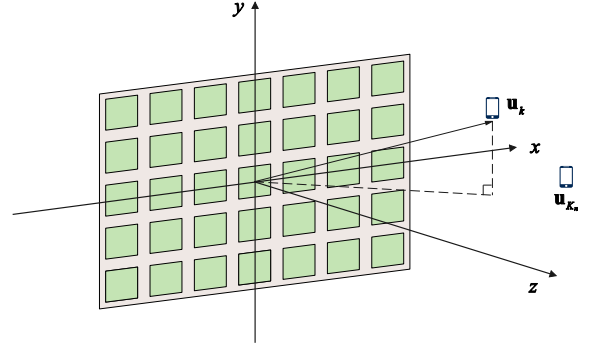


Fig. 2: Illustration of the locations of NFUEs with respect to XL-MIMO.

shifts proposed in [32] and equal power control (**HPS-EPC**), as a benchmark for our proposed design. The main idea of the heuristic phase shifts design is to align the phase shifts of the RIS with the phase of the RIS-related channels of a specific FFUE k , that is, $\mathbf{h}_k^H \mathbf{\Theta} \mathbf{b}_N(\varphi^a, \varphi^e) = N$. To this end, we set the entries of $\mathbf{\Theta}$ as $\theta_n = -\angle\{[\mathbf{h}_k^H]_n[\mathbf{b}_N(\varphi^a, \varphi^e)]_n\}$, $\forall n$ [32]. To have a fair comparison between all precoding designs, we set $\bar{\mathcal{R}}_k = \min\{\bar{\mathcal{S}}_k^{\text{CZF}}, \bar{\mathcal{S}}_k^{\text{LZF}}, \bar{\mathcal{S}}_k^{\text{MRT}}\}$, $\forall k \in \mathcal{K}_N$, $\tilde{\mathcal{R}}_k = \min\{\tilde{\mathcal{S}}_k^{\text{CZF}}, \tilde{\mathcal{S}}_k^{\text{LZF}}, \tilde{\mathcal{S}}_k^{\text{MRT}}\}$, $\forall k \in \mathcal{K}_F$ according to the **RPS-EPC** design.

B. Results and Discussion

1) *Convergence of Algorithms 2 and 3*: We study the convergence behavior of the proposed algorithm with $M = 400$ and random initial values for the phase shifts, that is, $\theta_n \in [0, 2\pi)$, $\forall n$. The convergence analysis consists of two parts: the convergence of the phase shifts design, shown in Figs. 3(a) and 3(b), and the convergence of the power control algorithm, illustrated in Fig. 3(c). As observed in Figs. 3(a) and 3(b), **Algorithm 2** converges to a stationary point after only a few iterations—specifically, after 5 iterations with MRT and 7 iterations with CZF. We also observe that with different initial points, the optimized minimum channel gain values are very close, confirming that the algorithm converges irrespective of the initial value. Note that the small transition in Fig. 3(b) with CZF is because at first the penalty factor ρ is relatively small, $\mathbf{V}$ is not a rank-one matrix. Regarding the convergence of **Algorithm 3**, we can see that the objective values of increase monotonically, as shown in Fig. 3(c), which aligns with our convergence behavior in Section IV. We also observe that MRT requires more iterations to converge—16 iterations—compared to CZF and LZF, which converge in 6 and 7 iterations, respectively. This difference arises because CZF and LZF effectively eliminate intra-group interference, enabling faster convergence to a stationary point.

2) *Performance of Proposed VR Selection*: Figure 4 shows the average of VR efficiency, which is defined as

$$\bar{\omega}^i = \frac{1}{KS} \left(\sum_{k \in \mathcal{K}_N} \text{tr}(\bar{\mathcal{D}}_k^i) + \sum_{k \in \mathcal{K}_F} \text{tr}(\tilde{\mathcal{D}}_k^i) \right),$$

where $i \in \{\text{CZF}, \text{MRT}\}$ versus the number of antennas M . This parameter represents the ratio of the average number of antennas used in the VR-based design relative to the total

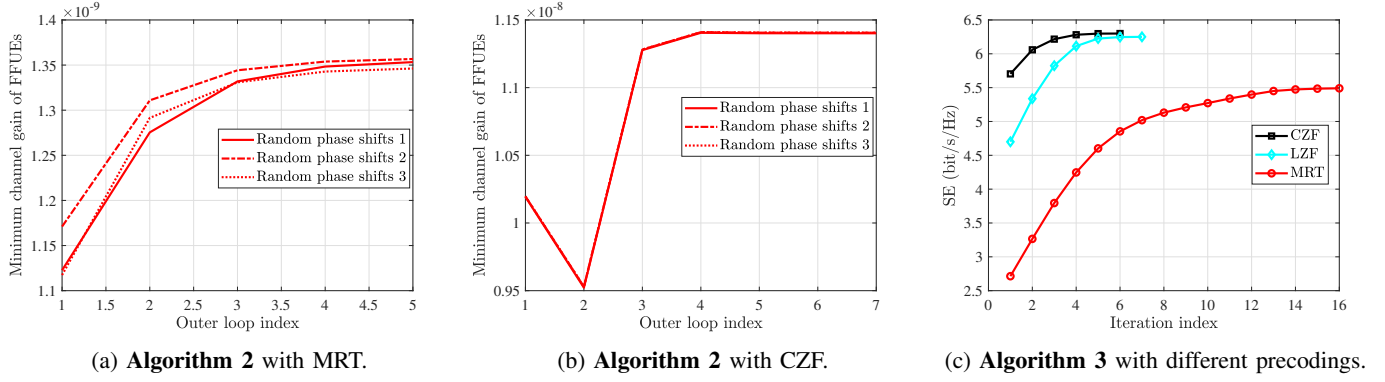


Fig. 3: Convergence behavior of proposed Algorithms.

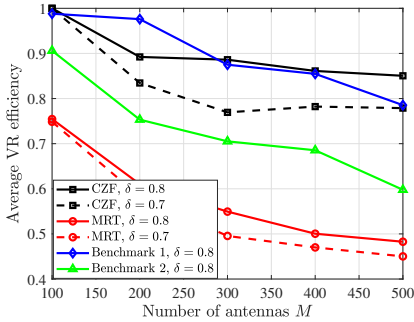


Fig. 4: Average VR efficiency versus M ($w_n = w_f = 0.5$).

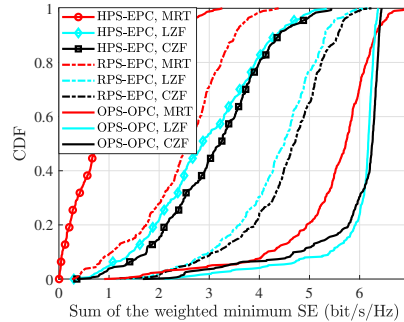


Fig. 5: Comparison between precoding schemes ($M = 400$, $w_n = w_f = 0.5$).

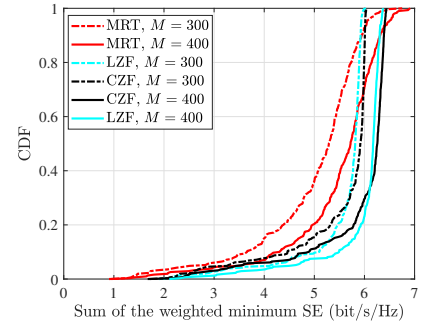


Fig. 6: CDF of the objective function for different antenna numbers ($w_n = w_f = 0.5$).

array size. We consider two VR selection ratios, $\delta = 0.7$ and $\delta = 0.8$. To enable a fair comparison, we define two benchmarks: **Benchmark 1** corresponds to the VR selection scheme in [17], and **Benchmark 2** refers to our proposed scheme applied only to NFUEs. As the number of antennas increases, the proportion used in VR-based design decreases since the received power is limited to a portion of the array, making VR-based design essential for large arrays. CZF requires more antennas than MRT, e.g., for $M = 500$ and $\delta = 0.7$, 78% of the antennas are used with CZF, while only 45% with MRT. This is because CZF suppresses interference, requiring more antennas for better performance. Lowering δ from 0.8 to 0.7 with CZF reduces $\bar{\omega}^{\text{CZF}}$ from 0.85 to 0.78, as fewer antennas are selected to meet the SINR requirements. Similar trends are observed for MRT. Another observation is that **Benchmark 2** selects fewer antennas compared to **Benchmark 1**. This is primarily because we use SINR as the VR selection criterion, which tends to be smaller than the SNR used in **Benchmark 1**. Note that LZF uses the same VR obtained for CZF, thus, the results for LZF are not shown in this figure.

3) *Comparison of Different Precoding Schemes:* Figure 5 presents the cumulative distribution function (CDF) of the weighted minimum SE sum for NFUEs and FFUEs using CZF, LZF, and MRT precoding. The proposed **OPS-OPC** scheme significantly outperforms **RPS-EPC** across all precoding methods. With CZF, **OPS-OPC** achieves a median SE of 6.2 bit/s/Hz, a 31.9% improvement over **RPS-EPC** (4.7 bit/s/Hz). For LZF, the gain is 37.8%, while for MRT, the

median SE increases by 119.2% (from 2.6 to 5.7 bit/s/Hz). Notably, LZF performs comparably to CZF in both schemes, making it an efficient alternative with lower computational complexity, especially for large arrays. Interestingly, the performance of **HPS-EPC** is inferior to that of **RPS-EPC**. The **HPS-EPC** design enhances the system's sum-SE compared to **RPS-EPC** under MRT. However, under CZF and LZF precoding, it fails to improve both the sum-SE and the weighted minimum SE across NFUEs and FFUEs. (Sum-SE results are omitted from the paper due to space limitations.) Specifically, under MRT, configuring all RIS elements to coherently enhance the signal for a particular FFUE k maximizes the reflected power gain for that user. However, this RIS phase configuration causes incoherent—or even destructive—combining at the RIS for other users. Therefore, it degrades the minimum SE across the FFUEs compared to the **RPS-EPC**. In the case of CZF and LZF, customizing the RIS to favor FFUE k results in highly correlated effective channels among the FFUEs. This correlation undermines the ability of CZF and LZF to suppress inter-group interference, leading to a notable degradation in SE—particularly for FFUE k , as interference from other FFUEs cannot be effectively mitigated. As a result, the weighted minimum SE across NFUEs and FFUEs under **HPS-EPC** becomes inferior to that achieved with **RPS-EPC**.

4) *The Effect of Number of Antennas:* In Fig. 6, we study the CDF of the sum of the weighted minimum SE of the XL-MIMO systems for two different number of antennas, i.e.

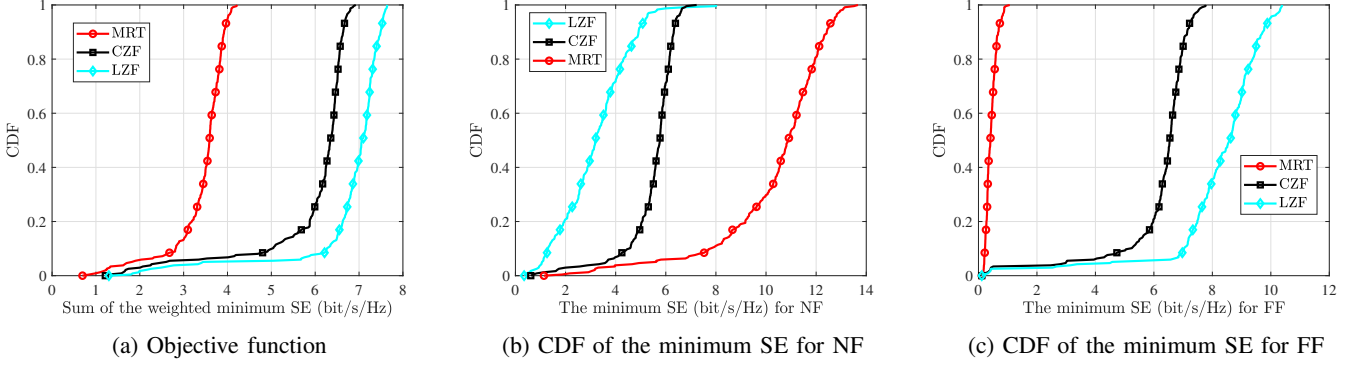


Fig. 7: CDFs of the SE for different precoding schemes and for **OPS-OPC** case ($w_n = 0.3$, $w_f = 0.7$, $M = 400$).

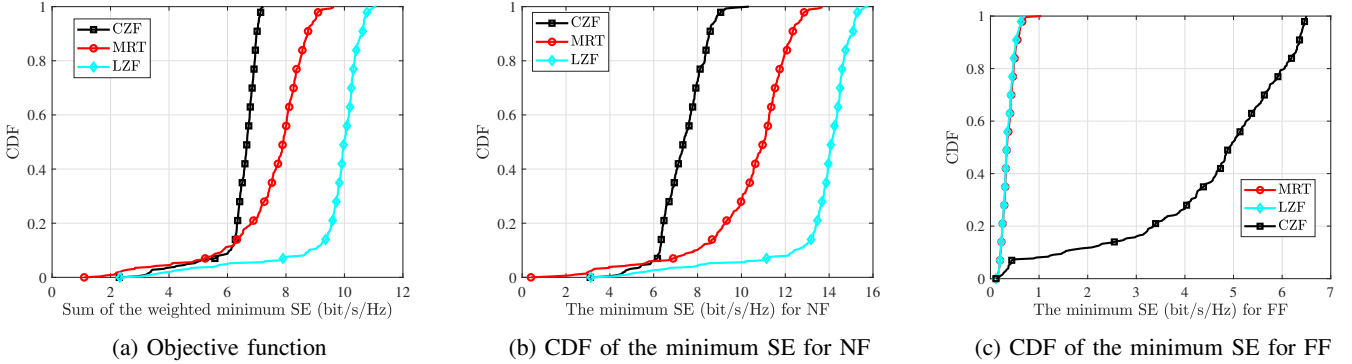


Fig. 8: CDFs of the SE for different precoding schemes and for **OPS-OPC** case ($w_n = 0.7$, $w_f = 0.3$, $M = 400$).

$M = 300$, $M = 400$ and with $w_n = w_f = 0.5$. We can see that by increasing the number of antennas from $M = 300$ to $M = 400$, the median performance of CZF, LZF and MRT can be improved by 5.76%, 6.1% and 8.1%, respectively.

5) *Impact of Priority Factors*: Next, we compare the CDFs of different precoding schemes under the **OPS-OPC** design, prioritizing FFUEs with weights $w_n = 0.3$, $w_f = 0.7$ and $M = 400$. From Fig. 7(a), we observe that LZF achieves the highest median sum of the weighted minimum SE among the three precoding schemes. Specifically, LZF improves the median sum of the weighted minimum SE by a factor of 1.11 compared to CZF and by a factor of 1.97 compared to MRT. This is expected as LZF can mitigate certain interference components within the same subarrays and offers more flexibility in power control design than CZF, leading to better performance. Furthermore, since MRT cannot suppress interference, the median performance with CZF and LZF is significantly higher than that with MRT. The same insights can be obtained for the CDF of minimum SE for FF, see Fig. 7(c). In Fig. 7(b), we observe that MRT achieves the highest median minimum SE for NFUEs, followed by CZF, with LZF performing the worst. This is because, with MRT, nearly all transmit power is allocated to maximize the minimum SE of NFUEs, as the signals transmitted to the FFUEs are weak and MRT lacks interference mitigation capability. As a result, MRT is more effective in enhancing the performance of NFUE. In contrast, CZF provides balanced performance for both NFUEs and FFUEs by effectively eliminating intra-group interference. Meanwhile, with LZF, the priority is given to FFUEs, i.e., less

power is allocated to optimize the performance of NFUEs.

Figure 8 evaluates the performance with CZF, LZF and MRT when prioritizing NFUEs in the **OPS-OPC** scenario with $w_n = 0.7$, $w_f = 0.3$, and $M = 400$. In Fig. 8(b), LZF demonstrates the best performance among the three precoding schemes, with MRT outperforming CZF. This outcome is primarily attributed to the objective function of the optimization problem, which aims to maximize the weighted sum SE of both NFUEs and FFUEs, rather than prioritizing a specific user group. MRT outperforms CZF since, in aiming to maximize the objective function, it allocates more power to enhance the performance of NFUEs, which consequently results in reduced power allocation for FFUEs. This trade-off is evident in Fig. 8(c), where the median of the minimum SE for FFUEs drops to 0.33 bit/s/Hz under MRT, compared to 4.92 bit/s/Hz with CZF. On the other hand, LZF achieves the best overall performance due to its flexible power control design—it allows each subarray to independently adjust its transmit power to each user. In contrast, CZF applies a uniform power control coefficient across all subarrays for a given user. Nonetheless, CZF provides a more balanced performance between NFUEs and FFUEs, as reflected in Figs. 8(b) and 8(c).

VI. CONCLUSION

We investigated the performance of a RIS-assisted downlink XL-MIMO system for both NFUEs and FFUEs with CZF, LZF and MRT precoding schemes, taking into account the effects of VR selection, phase shifts design at the RIS, and power control design at the BS. We proposed a heuristic VR selection design,

a two-stage phase shifts and power control design relying on SCA to maximize the sum of the weighted minimum SE of the system under different priority requirements. Our simulation results indicated that phase shifts design and power control can significantly improve the performance of the XL-MIMO assisted by the RIS with three linear precoding schemes. In particular, when giving equal priority to NFUEs and FFUEs, CZF is the best option, while LZF could achieve comparable performance. If we give priority to NFUEs or FFUEs, LZF can obtain the best performance, and CZF strikes a balance between the performance of NFUEs and FFUEs.

CSI acquisition errors or delays can significantly impact the accuracy of beamforming in XL-MIMO systems, leading to degradation in the achievable rate. Our proposed framework relies on SINR expressions that assume perfect channel knowledge; therefore, any inaccuracy in CSI may result in underestimation or overestimation of SINR values. Furthermore, errors in CSI acquisition can lead to incorrect VR selection, which in turn will compromise the subsequent design of RIS phase shifts and power control strategies. Such imperfections may also affect the convergence and effectiveness of the proposed penalty-based iterative algorithm used for optimization of RIS phase shifts. As such, developing efficient CSI acquisition techniques for RIS-assisted XL-MIMO systems remains a critical and promising direction for future research. To support real-time implementation, hardware acceleration using platforms, such as Field-Programmable Gate Arrays (FPGAs) or Graphics Processing Units (GPUs), can be employed to expedite key computations, while machine learning-based methods can also be explored to further reduce latency.

APPENDIX A

PROOF OF PROPOSITION 1

In order to calculate $\overline{\text{DS}}_k$ for the MRT precoding scheme, substituting the precoding vector $\tilde{\mathbf{w}}_k^{\text{MRT}}$ in (19) into (14), we have the following results:

$$\overline{\text{DS}}_k = \sqrt{P_n} \sum_{s \in \tilde{\mathcal{S}}_k^{\text{MRT}}} \sqrt{\tilde{\eta}_{sk}^{\text{MRT}}} \tilde{\Psi}_{kk}^s, \quad (73)$$

with the fact that $\tilde{\mathbf{D}}_{sk} \tilde{\mathbf{w}}_{sk}^{\text{MRT}} = \tilde{\mathbf{g}}_{sk}$ if $s \in \tilde{\mathcal{S}}_k^{\text{MRT}}$ and $\mathbf{0}_{M^*}$ otherwise.

Similarly, we obtain $|\overline{\text{UI}}_i|^2$ as follows:

$$|\overline{\text{UI}}_i|^2 = P_n \left| \sum_{s \in \tilde{\mathcal{S}}_i^{\text{MRT}}} \sqrt{\tilde{\eta}_{si}^{\text{MRT}}} \tilde{\Psi}_{ki}^s \right|^2. \quad (74)$$

Now, we derive $\mathbb{E}\{|\overline{\text{UI}}_j|^2\}$, which can be expressed as

$$\begin{aligned} \mathbb{E}\{|\overline{\text{UI}}_j|^2\} &= P_f \mathbb{E}\left\{ \left| \sum_{s \in \mathcal{S}} \sqrt{\tilde{\eta}_{sj}^{\text{MRT}}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{sj} \tilde{\mathbf{w}}_{sj}^{\text{MRT}} \right|^2 \right\} \\ &= P_f \mathbb{E}\left\{ \sum_{s \in \tilde{\mathcal{S}}_j^{\text{MRT}}} \sum_{s' \in \tilde{\mathcal{S}}_j^{\text{MRT}}} \sqrt{\tilde{\eta}_{sj}^{\text{MRT}}} \sqrt{\tilde{\eta}_{s'j}^{\text{MRT}}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{g}}_{s'j} \tilde{\mathbf{g}}_{s'j}^H \tilde{\mathbf{g}}_{sk} \right\} \\ &= P_f \sum_{s \in \tilde{\mathcal{S}}_j^{\text{MRT}}} \sum_{s' \in \tilde{\mathcal{S}}_j^{\text{MRT}}} \sqrt{\tilde{\eta}_{sj}^{\text{MRT}}} \sqrt{\tilde{\eta}_{s'j}^{\text{MRT}}} \tilde{\mathbf{g}}_{sk}^H \mathbb{E}\{\tilde{\mathbf{g}}_{sj} \tilde{\mathbf{g}}_{s'j}^H\} \tilde{\mathbf{g}}_{s'k}, \end{aligned} \quad (75)$$

where $\tilde{\mathbf{g}}_{sj}^H = \mathbf{h}_j^H \boldsymbol{\Theta} \mathbf{H}_{2,s}$, and we have

$$\begin{aligned} \mathbb{E}\{\tilde{\mathbf{g}}_{sj} \tilde{\mathbf{g}}_{s'j}^H\} &= \mathbb{E}\{(\mathbf{h}_j^H \boldsymbol{\Theta} \mathbf{H}_{2,s})(\mathbf{h}_j^H \boldsymbol{\Theta} \mathbf{H}_{2,s'})^H\} \\ &= \mathbb{E}\{(\alpha_2 \tilde{\mathbf{H}}_{2,s}^H + \beta_2 \tilde{\mathbf{H}}_{2,s}^H) \mathbf{M}_{jj} (\alpha_2 \tilde{\mathbf{H}}_{2,s'} + \beta_2 \tilde{\mathbf{H}}_{2,s'})^H\} \\ &= \alpha_2^2 \tilde{\mathbf{H}}_{2,s}^H \mathbf{M}_{jj} \tilde{\mathbf{H}}_{2,s'} + \beta_2^2 \mathbb{E}\{\tilde{\mathbf{H}}_{2,s}^H \mathbf{M}_{jj} \tilde{\mathbf{H}}_{2,s'}\}. \end{aligned} \quad (76)$$

When $s = s'$, we have $\mathbb{E}\{\tilde{\mathbf{H}}_{2,s}^H \mathbf{M}_{jj} \tilde{\mathbf{H}}_{2,s'}\} = \text{tr}(\mathbf{M}_{jj}) \mathbf{I}_{M^*} = \varsigma_k N \mathbf{I}_{M^*}$. Otherwise, $\mathbb{E}\{\tilde{\mathbf{H}}_{2,s}^H \mathbf{M}_{jj} \tilde{\mathbf{H}}_{2,s'}\} = \mathbf{0}_{M^* \times M^*}$. Therefore, substituting (76) into (75), we have

$$\mathbb{E}\{|\overline{\text{UI}}_j|^2\} = P_f \sum_{s \in \tilde{\mathcal{S}}_j^{\text{MRT}}} \sum_{s' \in \tilde{\mathcal{S}}_j^{\text{MRT}}} \sqrt{\tilde{\eta}_{sj}^{\text{MRT}}} \sqrt{\tilde{\eta}_{s'j}^{\text{MRT}}} \tilde{\Xi}_{kj}^{ss'}. \quad (77)$$

Substituting (73), (74) and (77) into (15), we obtain the final result given in (20).

APPENDIX B

PROOF OF PROPOSITION 2

For FFUE k with the MRT precoding scheme, we substitute the precoding vector $\tilde{\mathbf{w}}_k^{\text{MRT}}$ in (19) into (17) to derive $\overline{\text{DS}}_k$, $\mathbb{E}\{|\overline{\text{BU}}_k|^2\}$, $\mathbb{E}\{|\overline{\text{UI}}_j|^2\}$ and $\mathbb{E}\{|\overline{\text{UI}}_i|^2\}$ as follows:

$$\begin{aligned} \overline{\text{DS}}_k &= \sqrt{P_f} \mathbb{E}\left\{ \sum_{s \in \mathcal{S}} \sqrt{\tilde{\eta}_{sk}^{\text{MRT}}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{sk} \tilde{\mathbf{w}}_{sk}^{\text{MRT}} \right\} \\ &= \sqrt{P_f} \sum_{s \in \tilde{\mathcal{S}}_k^{\text{MRT}}} \sqrt{\tilde{\eta}_{sk}^{\text{MRT}}} \tilde{\Psi}_k^s, \end{aligned} \quad (78)$$

with the fact that $\tilde{\mathbf{D}}_{sk} \tilde{\mathbf{w}}_{sk} = \tilde{\mathbf{g}}_{sk} \in \mathbb{C}^{M^* \times 1}$ if $s \in \tilde{\mathcal{S}}_k^{\text{MRT}}$, and

$$\begin{aligned} \tilde{\Psi}_k^s &\triangleq \mathbb{E}\{\tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{g}}_{sk}\} = \mathbb{E}\{\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{H}_{2,s} (\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{H}_{2,s})^H\} \\ &= \mathbf{h}_k^H \boldsymbol{\Theta} \mathbb{E}\{\mathbf{H}_{2,s} \mathbf{H}_{2,s}^H\} \boldsymbol{\Theta}^H \mathbf{h}_k \\ &= \mathbf{h}_k^H \boldsymbol{\Theta} (\alpha_2^2 \tilde{\mathbf{H}}_{2,s}^H \tilde{\mathbf{H}}_{2,s} + \beta_2^2 M^* \mathbf{I}_N) \boldsymbol{\Theta}^H \mathbf{h}_k. \end{aligned} \quad (79)$$

To derive $\overline{\text{BU}}_k$, we use the fact that the variance of a sum of independent RVs is equal to the sum of the variances. Therefore, $\mathbb{E}\{|\overline{\text{BU}}_k|^2\}$ can be expressed as

$$\begin{aligned} \mathbb{E}\{|\overline{\text{BU}}_k|^2\} &= P_f \sum_{s \in \tilde{\mathcal{S}}_k^{\text{MRT}}} \tilde{\eta}_{sk}^{\text{MRT}} \mathbb{E}\{|\tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{g}}_{sk}|^2\} - |\mathbb{E}\{\tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{g}}_{sk}\}|^2 \\ &= P_f \sum_{s \in \tilde{\mathcal{S}}_k^{\text{MRT}}} \tilde{\eta}_{sk}^{\text{MRT}} (T_k^s - (\tilde{\Psi}_k^s)^2), \end{aligned} \quad (80)$$

where $T_k^s \triangleq \mathbb{E}\{|\tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{g}}_{sk}|^2\}$, which can be obtained as

$$\begin{aligned} T_k^s &= \mathbb{E}\{|\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{H}_{2,s} (\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{H}_{2,s})^H|^2\} \\ &= \mathbb{E}\{(\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{H}_{2,s} \mathbf{H}_{2,s}^H \boldsymbol{\Theta}^H \mathbf{h}_k)(\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{H}_{2,s} \mathbf{H}_{2,s}^H \boldsymbol{\Theta}^H \mathbf{h}_k)^H\} \\ &= \mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{C}_{sk} \boldsymbol{\Theta}^H \mathbf{h}_k, \end{aligned} \quad (81)$$

where $\mathbf{C}_{sk} \triangleq \mathbb{E}\{\mathbf{H}_{2,s} \mathbf{H}_{2,s}^H \mathbf{M}_{kk} \mathbf{H}_{2,s} \mathbf{H}_{2,s}^H\}$ and $\mathbf{C}_{sk}$ can be obtained by replacing $\mathbf{M}_{jj}$ with $\mathbf{M}_{kk}$ in (23).

Substituting (81) and $\tilde{\Psi}_k^s$ into (80), we have

$$\mathbb{E}\{|\overline{\text{BU}}_k|^2\} = P_f \sum_{s \in \tilde{\mathcal{S}}_k^{\text{MRT}}} \tilde{\eta}_{sk}^{\text{MRT}} \tilde{a}_k^s. \quad (82)$$

Now, we focus on $\mathbb{E}\{|\overline{\text{UI}}_j|^2\}$, which can be expressed as

$$\begin{aligned} \mathbb{E}\{|\overline{\text{UI}}_j|^2\} &= P_f \mathbb{E}\left\{ \left| \sum_{s \in \mathcal{S}} \sqrt{\tilde{\eta}_{sj}^{\text{MRT}}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{D}}_{sj} \tilde{\mathbf{w}}_{sj}^{\text{MRT}} \right|^2 \right\} \\ &= P_f \sum_{s \in \tilde{\mathcal{S}}_j^{\text{MRT}}} \sum_{s' \in \tilde{\mathcal{S}}_j^{\text{MRT}}} \sqrt{\tilde{\eta}_{sj}^{\text{MRT}}} \sqrt{\tilde{\eta}_{s'j}^{\text{MRT}}} \mathbb{E}\{\tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{g}}_{sj} \tilde{\mathbf{g}}_{s'j}^H \tilde{\mathbf{g}}_{s'k}\}. \end{aligned} \quad (83)$$

The expectation term in (83) can be rewritten as

$$\begin{aligned} \mathbb{E}\{\tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{g}}_{sj} \tilde{\mathbf{g}}_{s'j}^H \tilde{\mathbf{g}}_{s'k}\} &= \mathbb{E}\{\mathbf{h}_k^H \boldsymbol{\Theta} \mathbf{H}_{2,s} \mathbf{H}_{2,s}^H \boldsymbol{\Theta}^H \mathbf{h}_j \mathbf{h}_j^H \boldsymbol{\Theta} \mathbf{H}_{2,s'} \mathbf{H}_{2,s'}^H \boldsymbol{\Theta}^H \mathbf{h}_k\} \\ &= \mathbf{h}_k^H \boldsymbol{\Theta} \mathbb{E}\{\mathbf{H}_{2,s} \mathbf{H}_{2,s}^H \mathbf{M}_{jj} \mathbf{H}_{2,s'} \mathbf{H}_{2,s'}^H\} \boldsymbol{\Theta}^H \mathbf{h}_k. \end{aligned} \quad (84)$$

Note that we have $\mathbb{E}\{\tilde{\mathbf{g}}_{sj}^H \tilde{\mathbf{g}}_{sj} \tilde{\mathbf{g}}_{sj}^H \tilde{\mathbf{g}}_{sk}\} = \mathbf{h}_k^H \mathbf{C}_{sj} \mathbf{\Theta}^H \mathbf{h}_k$ when $s = s'$ based on (81). Otherwise, the expectation in (84) can be derived as

$$\begin{aligned} \mathbf{D}_j^{ss'} &\triangleq \mathbb{E}\{\mathbf{H}_{2,s} \mathbf{H}_{2,s}^H\} \mathbf{M}_{jj} \mathbb{E}\{\mathbf{H}_{2,s'} \mathbf{H}_{2,s'}^H\} \\ &= (\alpha_2^2 \tilde{\mathbf{H}}_{2,s} \tilde{\mathbf{H}}_{2,s}^H + \beta_2^2 M^* \mathbf{I}_N) \mathbf{M}_{jj} (\alpha_2^2 \tilde{\mathbf{H}}_{2,s'} \tilde{\mathbf{H}}_{2,s'}^H + \beta_2^2 M^* \mathbf{I}_N). \end{aligned}$$

Substituting (84) into (83), we obtain

$$\mathbb{E}\{|\tilde{\mathbf{U}}_j|^2\} = P_f \sum_{s \in \tilde{\mathcal{S}}_{\text{MRT}}} \sum_{s' \in \tilde{\mathcal{S}}_{\text{MRT}}} \sqrt{\tilde{\eta}_{sj}^{\text{MRT}} \tilde{\eta}_{s'j}^{\text{MRT}}} \tilde{b}_{kj}^{ss'}. \quad (85)$$

Finally, we compute $\mathbb{E}\{|\tilde{\mathbf{U}}_i|^2\}$, which can be written as

$$\begin{aligned} \mathbb{E}\{|\tilde{\mathbf{U}}_i|^2\} &= P_n \mathbb{E}\left\{\left|\sum_{s \in \tilde{\mathcal{S}}_{\text{MRT}}} \sqrt{\tilde{\eta}_{si}^{\text{MRT}}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{w}}_{si}^{\text{MRT}}\right|^2\right\} \\ &= P_n \mathbb{E}\left\{\sum_{s \in \tilde{\mathcal{S}}_{\text{MRT}}} \sum_{s' \in \tilde{\mathcal{S}}_{\text{MRT}}} \sqrt{\tilde{\eta}_{si}^{\text{MRT}} \tilde{\eta}_{s'i}^{\text{MRT}}} \tilde{\mathbf{g}}_{sk}^H \tilde{\mathbf{g}}_{s'i} \tilde{\mathbf{g}}_{s'i}^H \tilde{\mathbf{g}}_{s'k}\right\} \\ &= P_n \sum_{s \in \tilde{\mathcal{S}}_{\text{MRT}}} \sum_{s' \in \tilde{\mathcal{S}}_{\text{MRT}}} \sqrt{\tilde{\eta}_{si}^{\text{MRT}} \tilde{\eta}_{s'i}^{\text{MRT}}} \tilde{\mathbf{g}}_{s'i}^H \mathbb{E}\{\tilde{\mathbf{g}}_{s'k} \tilde{\mathbf{g}}_{sk}^H\} \tilde{\mathbf{g}}_{si}, \quad (86) \end{aligned}$$

where

$$\begin{aligned} \mathbb{E}\{\tilde{\mathbf{g}}_{s'k} \tilde{\mathbf{g}}_{sk}^H\} &= \mathbb{E}\{(\mathbf{h}_k^H \mathbf{\Theta} \mathbf{H}_{2,s'})^H \mathbf{h}_k^H \mathbf{\Theta} \mathbf{H}_{2,s}\} \\ &= \mathbb{E}\{(\alpha_2 \tilde{\mathbf{H}}_{2,s'} + \beta_2 \tilde{\mathbf{H}}_{2,s'}) \mathbf{M}_{kk} (\alpha_2 \tilde{\mathbf{H}}_{2,s} + \beta_2 \tilde{\mathbf{H}}_{2,s})\} \\ &= \alpha_2^2 \tilde{\mathbf{H}}_{2,s'}^H \mathbf{M}_{kk} \tilde{\mathbf{H}}_{2,s} + \beta_2^2 \mathbb{E}\{\tilde{\mathbf{H}}_{2,s'}^H \mathbf{M}_{kk} \tilde{\mathbf{H}}_{2,s}\}, \quad (87) \end{aligned}$$

when $s = s'$, we have $\mathbb{E}\{\tilde{\mathbf{H}}_{2,s'}^H \mathbf{M}_{kk} \tilde{\mathbf{H}}_{2,s}\} = \text{tr}(\mathbf{M}_{kk}) \mathbf{I}_{M^*} = \zeta_k N \mathbf{I}_{M^*}$. Otherwise, $\mathbb{E}\{\tilde{\mathbf{H}}_{2,s'}^H \mathbf{M}_{kk} \tilde{\mathbf{H}}_{2,s}\} = \mathbf{0}_{M^* \times M^*}$. Therefore, we can obtain the following result

$$\mathbb{E}\{|\tilde{\mathbf{U}}_i|^2\} = P_n \sum_{s \in \tilde{\mathcal{S}}_{\text{MRT}}} \sum_{s' \in \tilde{\mathcal{S}}_{\text{MRT}}} \sqrt{\tilde{\eta}_{si}^{\text{MRT}} \tilde{\eta}_{s'i}^{\text{MRT}}} \tilde{\Xi}_{ki}^{ss'}. \quad (88)$$

Substituting (78), (82), (85), (88) into (18), we obtain the final result (21).

REFERENCES

- [1] Z. Wang *et al.*, "A tutorial on extremely large-scale MIMO for 6G: Fundamentals, signal processing, and applications," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 3, pp. 1560–1605, Thirdquarter 2024.
- [2] M. Matthaiou, O. Yurduseven, H. Q. Ngo, D. Morales-Jimenez, S. L. Cotton, and V. F. Fusco, "The road to 6G: Ten physical layer challenges for communications engineers," *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 64–69, Jan. 2021.
- [3] R. Deng *et al.*, "Reconfigurable holographic surfaces for ultra-massive MIMO in 6G: Practical design, optimization and implementation," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2367–2379, Aug. 2023.
- [4] X. Li, Z. Dong, Y. Zeng, S. Jin, and R. Zhang, "Multi-user modular XL-MIMO communications: Near-field beam focusing pattern and user grouping," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 13 766–13 781, Oct. 2024.
- [5] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.
- [6] Z. Zhang *et al.*, "Active RIS vs. passive RIS: Which will prevail in 6G?" *IEEE Trans. Commun.*, vol. 71, no. 3, pp. 1707–1725, Mar. 2023.
- [7] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3083–3098, May 2022.
- [8] M. Nerini, S. Shen, H. Li, and B. Clerckx, "Beyond diagonal reconfigurable intelligent surfaces utilizing graph theory: Modeling, architecture design, and optimization," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 9972–9985, Aug. 2024.
- [9] G. Torcolacci *et al.*, "Holographic imaging with XL-MIMO and RIS: Illumination and reflection design," *IEEE J. Sel. Topics Signal Process.*, vol. 18, no. 4, pp. 587–602, May 2024.
- [10] J. Lee, H. Chung, Y. Cho, S. Kim, and S. Hong, "Near-field channel estimation for XL-RIS assisted multi-user XL-MIMO systems: Hybrid beamforming architectures," *IEEE Trans. Commun.*, vol. 73, no. 3, pp. 1560–1574, Mar. 2025.
- [11] X. Zhang, H. Shao, W. Zhang, Z. Xie, X. Yang, and W. Jing, "RIS-aided XL-MIMO channel estimation based on expectation-maximization," *IEEE Commun. Lett.*, vol. 28, no. 12, pp. 2869–2873, Dec. 2024.
- [12] M. Cui, Z. Wu, Y. Lu, X. Wei, and L. Dai, "Near-field MIMO communications for 6G: Fundamentals, challenges, potentials, and future directions," *IEEE Commun. Mag.*, vol. 61, no. 1, pp. 40–46, Jan. 2023.
- [13] H. Lu and Y. Zeng, "Communicating with extremely large-scale array/surface: Unified modeling and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4039–4053, June 2022.
- [14] E. Björnson and L. Sanguinetti, "Power scaling laws and near-field behaviors of massive MIMO and intelligent reflecting surfaces," *IEEE Open J. Commun. Society*, vol. 1, pp. 1306–1324, Sept. 2020.
- [15] D. Dardari, "Communicating with large intelligent surfaces: Fundamental limits and models," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2526–2537, Nov. 2020.
- [16] E. D. Carvalho, A. Ali, A. Amiri, M. Angelichinoski, and R. W. Heath Jr., "Non-stationarities in extra-large-scale massive MIMO," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 74–80, Aug. 2020.
- [17] K. Zhi *et al.*, "Performance analysis and low-complexity design for XL-MIMO with near-field spatial non-stationarities," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 6, pp. 1656–1672, June 2024.
- [18] Y. Chen and L. Dai, "Non-stationarity channel estimation for extremely large-scale MIMO," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 7683–7697, July 2024.
- [19] H. Chen *et al.*, "ELAA near-field localization and sensing with partial blockage detection," in *Proc. IEEE PIMRC*, Sept. 2024, pp. 1–6.
- [20] W. Xu, A. Liu, M.-J. Zhao, and G. Caire, "Joint visibility region detection and channel estimation for XL-MIMO systems via alternating MAP," *IEEE Trans. Signal Process.*, vol. 72, pp. 4827–4842, Oct. 2024.
- [21] X. Cao, M. Mohammadi, H. Q. Ngo, and M. Matthaiou, "RIS-assisted XL-MIMO for coexistence of near-field and far-field communications," in *Proc. IEEE WCNC*, Apr. 2024, pp. 1–6.
- [22] A. Amiri, M. Angelichinoski, E. de Carvalho, and R. W. Heath Jr., "Extremely large aperture massive MIMO: Low complexity receiver architectures," in *Proc. IEEE GLOBECOM*, Dec. 2018, pp. 1–6.
- [23] J. C. Marinello, T. Abrão, A. Amiri, E. de Carvalho, and P. Popovski, "Antenna selection for improving energy efficiency in XL-MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13 305–13 318, Nov. 2020.
- [24] L. N. Ribeiro, S. Schwarz, and M. Haardt, "Low-complexity zero-forcing precoding for XL-MIMO transmissions," in *Proc. IEEE EUSIPCO*, Aug. 2021, pp. 1621–1625.
- [25] X. Yang, F. Cao, M. Matthaiou, and S. Jin, "On the uplink transmission of extra-large scale massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15 229–15 243, Dec. 2020.
- [26] Z. Lu, Y. Han, S. Jin, and M. Matthaiou, "Near-field localization and channel reconstruction for ELAA systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 7, pp. 6938–6953, July 2024.
- [27] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.
- [28] M. Mohammadi, Z. Mobini, H. Q. Ngo, and M. Matthaiou, "Ten years of research advances in full-duplex massive MIMO," *IEEE Trans. Commun.*, vol. 73, no. 3, pp. 1756–1786, Mar. 2025.
- [29] J. H. I. de Souza, A. Amiri, T. Abrão, E. de Carvalho, and P. Popovski, "Quasi-distributed antenna selection for spectral efficiency maximization in subarray switching XL-MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 7, pp. 6713–6725, July 2021.
- [30] Z. Liu, J. Zhang, Z. Liu, H. Xiao, and B. Ai, "Double-layer power control for mobile cell-free XL-MIMO with multi-agent reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4658–4674, May 2024.
- [31] Y. Liu, Z. Wang, J. Xu, C. Ouyang, X. Mu, and R. Schober, "Near-field communications: A tutorial review," *IEEE Open J. Commun. Society*, vol. 4, pp. 1999–2049, Sept. 2023.
- [32] K. Zhi, C. Pan, G. Zhou, H. Ren, M. El Kashlan, and R. Schober, "Is RIS-aided massive MIMO promising with ZF detectors and imperfect CSI?" *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 3010–3026, Oct. 2022.
- [33] M. Mohammadi, H. Q. Ngo, and M. Matthaiou, "Phase-shift and transmit power optimization for RIS-aided massive MIMO SWIPT IoT networks," *IEEE Trans. Commun.*, vol. 73, no. 1, pp. 631–647, Jan. 2025.

- [34] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends. Signal Process.*, vol. 11, no. 3-4, pp. 154–655, 2017.
- [35] M. Mohammadi, Z. Mobini, H. Q. Ngo, and M. Matthaiou, "Next-generation multiple access with cell-free massive MIMO," *Proc. IEEE*, vol. 112, no. 9, pp. 1372–1420, Sept. 2024.
- [36] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.
- [37] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 2022–2035, Mar. 2020.
- [38] T. Jiang and Y. Shi, "Over-the-air computation via intelligent reflecting surfaces," in *Proc. IEEE GLOBECOM*, Dec. 2019, pp. 1–6.
- [39] L. T. A. Beck, A. Ben-Tal, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J Glob Optim*, vol. 47, no. 1, p. 29–51, June. 2010.
- [40] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," <https://cvxr.com/cvx>, Mar. 2014.
- [41] M. Mohammadi, T. T. Vu, H. Q. Ngo, and M. Matthaiou, "Network-assisted full-duplex cell-free massive MIMO: Spectral and energy efficiencies," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 9, pp. 2833–2851, Sept. 2023.
- [42] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.
- [43] H. H. M. Tam, H. D. Tuan, D. T. Ngo, T. Q. Duong, and H. V. Poor, "Joint load balancing and interference management for small-cell heterogeneous networks with limited backhaul capacity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 872–884, Feb. 2017.
- [44] K. Zhi *et al.*, "Two-timescale design for reconfigurable intelligent surface-aided massive MIMO systems with imperfect CSI," *IEEE Trans. Inf. Theory*, vol. 69, no. 5, pp. 3001–3033, May 2023.