

Performance of Large Intelligent Surfaces in Multiuser Millimeter Wave MIMO-NOMA Systems

NOURAN ARAFAT¹, ENGY ALY MAHER¹, AHMED EL-MAHDY¹
AND FALKO DRESSLER², (Fellow, IEEE).

¹Faculty of Information Engineering and Technology, German University in Cairo, Cairo 16482, Egypt.

²School of Electrical Engineering and Computer Science, TU Berlin, 10623 Berlin, Germany.

Corresponding author: Nouran Arafat (e-mail: nouran.zaghlool@guc.edu.eg).

This work is supported by the DAAD and in part by the Federal Ministry of Education and Research (BMBF, Germany) within the 6G Research and Innovation Cluster 6G-RIC under Grant 16KISK020K, in co-operation between TU Berlin and the German University in Cairo.

ABSTRACT Large Intelligent Surfaces (LIS) are considered a candidate technology for 6G networks due to its simple deployment and relatively low cost. Recently, Millimeter Wave (mmWave) communication has attracted much attention due to its huge spectrum resources (30-300 GHz), which supports gigabits per second data rates. However, it suffers from severe path loss compared to path attenuation at low-frequency bands. Massive MIMO compensates this path loss by adding more antennas in a small area. Non-orthogonal multiple access (NOMA) is one of the multiple access techniques that increases the data rate of the system. In this paper, we integrate LIS with multi-user MIMO-NOMA in the mmWave band to enhance the overall system performance. Hybrid precoding is performed by having a zero-forcing digital precoder at the base station and an analog precoding at the LIS. Users are grouped into clusters to minimize multi-user interference. Joint optimization of power allocation of NOMA users, LIS phases, and gains is proposed to maximize the energy efficiency of the system. Closed form expressions for the optimum phases and gains of LIS are obtained using the Quadratic Constraint Quadratic Problem (QCQP). Due to the non-convexity of the formulated problem, alternating optimization is developed to solve the problem. Subsequently, the optimization parameters are decoupled through applying fractional programming. The problem is divided into sub problems solved alternatively. The performance of the system is evaluated in terms of energy efficiency and power consumption. Numerical results show the superior performance of LIS at the expense of higher power consumption.

INDEX TERMS Large Intelligent Surface (LIS), Millimeter Wave (mmWave) communication, NOMA, MIMO, Passive RIS, Active RIS, Spectral Efficiency, Energy Efficiency, Joint Optimization.

Glossary

ADC	Analog Digital Converter.	NOMA	Non Orthogonal Multiple Access.
AF	Amplify and Forward.	OMA	Orthogonal Multiple Access.
BS	Base Station.	QCQP	Quadratic Constraint Quadratic Problem.
DAC	Digital Analog Converter.	RE	Reflecting element.
DoF	Degree of Freedom.	RF	Radio Frequency.
FDD	Frequency Division Duplexing.	RIS	Reconfigurable Intelligent Surface.
FDMA	Frequency Division Multiple Access.	SC	Superposition Coding.
FP	Fractional Programming.	SIC	Successive Interference Cancellation.
LIS	Large Intelligent Surface.	TDMA	Time Division Multiple Access.

I. INTRODUCTION

Nowadays, Millimeter Wave (mmWave) communication has much attention as it supports data rates of gigabits per second due to its huge spectrum resources (30-300 GHz). However, it experiences severe path loss compared to lower frequency bands [1], [2]. Integrating massive MIMO with mmWave compensates the path loss by adding more antennas in a small area [3]. The key idea of massive MIMO is to deploy a large number of antennas N to serve M users simultaneously.

Non-Orthogonal Multiple Access (NOMA) is one of the most promising technologies, due to its high spectral and energy efficiencies [4]. In the same resource block, multiple users are served using power-domain NOMA. By using superposition coding (SC) at the transmitter and successive interference cancellation (SIC) at the receiver, NOMA improves the spectral efficiency of the wireless system.

Reconfigurable Intelligent Surface (RIS) has raised significant interest in the current research, because it is relatively cheap and easy to mount on building facades. A RIS is mainly a two dimensional meta-surface with the ability to control the wireless signal and reflect it in a desired direction. In other words, RIS consists of passive reflecting elements (REs). Each RE has the capability to reflect the incident electromagnetic waves using the reflecting coefficient (amplitude and phase) [5]. By adjusting the reflecting coefficients, the reflected signal is added constructively in the desired direction (receiver) and destructively in other directions (suppressing interference). Unlike other related technologies, such as relays and MIMO beamforming, RIS does not require its own power source [6]. Thus, RIS can enhance the spectral and energy efficiency of the wireless system. By deploying RIS in mmWave MIMO system, the performance of the channel is improved.

Based on power consumption, there are mainly two types of RIS: active and passive. Active RIS is sometimes referred to in the literature as Large Intelligent Surface (LIS), where each RE is connected to a power amplifier and phase shifter. On the other hand, passive RIS, or reflecting RIS, consists of passive elements with almost zero power consumption. Each element has a different phase shifter value to direct the signal in a certain direction [6]. A comparison between RIS and relay is addressed in [7], [8]. Furthermore, LIS (active RIS) differs from amplify and forward relay (AF) in terms of hardware complexity and power consumption. The AF relay amplifies and retransmits the signal, thus requiring two time slots. Additionally, relays require RF chains, analog to digital converters (ADC), digital to analog converters (DAC) and amplifiers [8].

The motivation behind integrating NOMA with LIS is to combine the benefits of both technologies into one system, enhancing spectral efficiency, user coverage and connectivity at a low cost. Since the wireless channel suffers from high fluctuations due to path loss and blockage, the combination of these technologies can control the wireless channel and guarantee specific users' order required by NOMA.

In this paper, we propose a joint iterative optimization

algorithm for LIS phase shifts and gains, as well as power allocation for users, to maximize energy efficiency. The main contributions of this paper can be summarized as follows:

- 1) An algorithm for clustering the users is proposed to minimize the multi-user interference such that users in the same cluster have a high channel correlation, while users in different clusters have a low correlation between their channels.
- 2) Hybrid precoding is performed by zero-forcing digital precoding at the BS and analog precoding at LIS. LIS elements are divided into sub-surfaces and each sub-surface is assigned to one antenna at the BS to realize an equivalent performance as the energy-efficient sub-connected architecture.
- 3) A joint optimization of power allocation among users, LIS power amplifier gain, and LIS phase shifts are proposed for the mmWave MIMO-NOMA multi-user system.

The rest of this paper is organized as follows. Section II introduces the related work. Section III presents the system and channel model. Then, Section IV shows the signal model and signal to interference noise ratio for the mmWave MIMO-NOMA aided LIS system. Afterwards, the energy efficiency maximization problem is given and transformed into simpler form in Section V. Section VI derives the optimum values for LIS parameters (phase and gain) and users power. Then, the numerical results are shown in Section VII. Finally, the conclusions are presented in Section VIII.

Notation Upper and lower bold case represent matrix and vector respectively. The notation of $(\cdot)^T, (\cdot)^H, (\cdot)^{-1}$ are the transpose, conjugate transpose (Hermitian) and the inverse of the matrix respectively. The $\|\cdot\|_p$ represents the l_p norm. If S is a set, then $|S|$ denotes the number of elements inside it. The difference between two sets is denoted by (\setminus) . The Kronecker product is denoted by \otimes . The dot product between two vectors is denoted by \cdot . The notation of $\mathcal{CN}(\mu; \Sigma)$ denotes the complex Gaussian distribution with mean μ and covariance Σ . The expectation of random variable is denoted by $E\{f(\cdot)\}$.

II. RELATED WORK

The concept of RIS has been, for example, summarized in [9]. The continuous and discrete RIS phases are explained. Integrating RIS into a two-user Non-Orthogonal Multiple access (NOMA) system is discussed. Moreover, the working principles and challenges of RIS-aided systems are presented in [10].

In [11], Intelligent Reflecting Surface (IRS) is proposed to support a multi-user millimeter wave (mmWave) system, in which multiple IRS are deployed to enhance the coverage of the system when the direct links between the base station and users are blocked. The authors present an optimization for the weighted sum rate maximization problem for joint active and passive beamforming. To address this issue, an alternating iterative algorithm is proposed to decouple the active and passive beamforming matrices. In [12], multiple

IRS are deployed in multi-user system to enhance the overall system performance. Double-manifold alternating optimization is proposed to maximize the weighted sum rate and IRS is addressed in [22]–[27]. In [22], the role of IRS in MIMO-NOMA is discussed in different cases. Different reflection coefficients of each IRS element. Results show that the proposed approach outperforms traditional methods, FDMA, TDMA, and NOMA [23]. Results reveal that in terms of spectral efficiency. In [13], RIS assisted aerial-ground communication scenario, exploring UAV trajectory efficiency. In [24], the spectral efficiency optimization is design and RIS' phase shift optimization aiming to maximize only addressed in case of passive RIS. DC programming method is used to solve the unit modulus constraint of the reflecting coefficient of RIS.

RIS aided frequency division duplexing (FDD) system for active RIS. Also, mmWave communications did not consider uplink and downlink transmission is discussed in [14]. The same phase shift matrix for both transmission directions is proposed. RIS is used as interference nulling in a multi-user environment [15]. In [16], joint optimization of digital precoder at BS and the phase shift at RIS is considered in order to enhance the received signal power.

There are many reports demonstrating the performance of ZF precoder with lower complexity. While in of passive RIS; however, active RIS is only addressed in a few works [17], [18]. A comparison between active and passive RIS is investigated in [18], [19]. In [17], two architectures for active RIS are compared: fully connected and sub-connected architectures. In fully connected architecture each RIS element has one power amplifier. However, in sub-connected architecture, a group of RIS elements share the same power amplifier. As the number of RIS elements increases, the fully connected architecture consumes huge amount of power consumption. Therefore, the sub-connected architecture reduces the power consumption at the expense of lower degree of freedoms (DoF) for beamforming design. The spectral and energy efficiencies are evaluated, and the results show that the latter architecture enjoys higher energy efficiency at the cost of lower sum rate than fully connected architecture. In [18], active and passive RIS are considered in case of a strong/weak direct link. Results show that passive RIS has only 3% gain compared to the case without RIS in case of a strong direct link, while active RIS achieves noticeable spectral efficiency gain of 67% in case of a strong direct link. Thus, by implementing active RIS, fading can be overcome.

Active and passive RIS aided systems are compared in [19]. For fair comparison, the same overall power budget is assumed at the base station (BS) and RIS. The system is evaluated based on the achievable sum rate. The results show that active RIS has superior performance when the power budget is not small and the number of RIS elements is small. This is because as the number of RIS elements increases, a huge amount of power is consumed to supply the power amplifiers, and the thermal noise also increases in active RIS. In [20], conventional hybrid precoder is replaced by an energy-efficient RIS to realize analog precoder. Also, a deep learning-based multiple discrete classification is proposed to solve the sum rate maximization optimization problem.

Non-Orthogonal Multiple Access (NOMA) is widely de-

handed due to the increasing number of devices and enormous connectivity [21]. The combination of NOMA with IRS is addressed in [22]–[27]. In [22], the role of IRS in MIMO-NOMA is discussed in different cases. Different Multiple Access (MA) techniques are investigated with RIS, e.g., FDMA, TDMA, and NOMA [23]. Results reveal that in terms of spectral efficiency. In [13], RIS assisted aerial-ground communication scenario, exploring UAV trajectory efficiency. In [24], the spectral efficiency optimization is only addressed in case of passive RIS. DC programming and successive convex approximation (SCA) are utilized to solve EE maximization problem in NOMA-RIS system [25]. However, authors of [23], [24] and [25] did not consider uplink and downlink transmission is discussed in [14]. In [23] and [25]. In [28], authors proposed two low complexity ZF approximation schemes for precoding and user grouping. Those schemes are based on dividing the antennas into sub-groups. In [16], joint optimization of digital precoder at BS and the phase shift at RIS is considered in order to enhance the received signal power. In which a precoding scheme is designed to achieve the maximum ratio transmission/combining (MRT/MRC), and minimum mean squared error (MMSE) are discussed for multi-user MIMO system. Results show that the highest EE is achieved using ZF precoding due to interference suppression at affordable complexity. A RIS partitioning technique is proposed in [26], in which each sub-surface serves one user and only a single antenna is considered at the BS. In [27], only two users NOMA aided IRS system are considered. A comparison between IRS-NOMA and IRS-NOMA is presented. Results show superior energy efficiency performance in the case of NOMA. Both [26], [27] do not consider mmWave band and consider passive NOMA with LIS in mmWave band, which is considered in this paper.

Based on the discussed related work, we believe that integrating MIMO-NOMA with LIS in mmWave band will enhance the overall system performance. In this paper, the three technologies are integrated for increasing the spectral efficiency, extending the coverage and minimizing the cost. Combining MIMO-NOMA with LIS in mmWave band is still missing in the literature. In addition, users grouping algorithm and hybrid precoding are proposed. The hybrid precoding is implemented by considering digital precoding at the BS, and LIS acts as an analog precoder.

In this section, the system model and the channel model are presented. We consider a downlink mmWave MIMO-

NOMA system assisted by LIS, where the base station is equipped with N antenna elements, and LIS has L elements to serve M single antenna users, as shown in Fig. 1. The base station transmits data to users through the direct link and the reflected link from LIS. Users are grouped into clusters to minimize the interference among them. The channel between the BS and the k^{th} user in the c^{th} cluster, the channel between BS and LIS, and the channel between LIS and the k^{th} user in the c^{th} cluster are represented by $\mathbf{g}_{c;k} \in \mathbb{C}^{N \times 1}$; $\mathbf{H} \in \mathbb{C}^{N \times L}$; $\mathbf{f}_{c;k} \in \mathbb{C}^{L \times 1}$.

A. CHANNEL MODEL

The mmWave channel model for the k^{th} user in the c^{th} cluster is given as [3]

$$\mathbf{g}_{c;k} = \sum_{p=1}^P \frac{1}{P_{c;k}} \mathbf{r}_{c;k}^{(p)} \mathbf{a}_{c;k}^{(p)}(\theta_{c;k}^{(p)}, \phi_{c;k}^{(p)}); \quad (1)$$

where $P_{c;k}$ is the number of paths for the k^{th} user in the c^{th} cluster group. The complex gain of the p^{th} path is given by $\mathbf{r}_{c;k}^{(p)}$. The azimuth angle of departure (AOD) and the elevation angle of arrival (AOA) are denoted by $\theta_{c;k}^{(p)}$, $\phi_{c;k}^{(p)}$ respectively. The steering vector is given by $\mathbf{a}_{c;k}^{(p)}(\theta_{c;k}^{(p)}, \phi_{c;k}^{(p)})$ with size $N \times 1$. A uniform planar array (UPA) is considered with N_1 antenna elements in horizontal direction and N_2 antennas in vertical direction where $N = N_1 N_2$, so $\mathbf{a}_{c;k}^{(p)}(\theta_{c;k}^{(p)}, \phi_{c;k}^{(p)}) = \mathbf{a}_{\text{azmi}}(\theta_{c;k}^{(p)}) \mathbf{a}_{\text{elev}}(\phi_{c;k}^{(p)})$, where $\mathbf{a}_{\text{azmi}}(\theta_{c;k}^{(p)}) = \frac{1}{\sqrt{N_1}} [1; e^{j2(\frac{d_1}{\lambda})\sin(\theta_{c;k}^{(p)})}; \dots; e^{j2(N_1-1)(\frac{d_1}{\lambda})\sin(\theta_{c;k}^{(p)})}]$, and $\mathbf{a}_{\text{elev}}(\phi_{c;k}^{(p)}) = \frac{1}{\sqrt{N_2}} [1; e^{j2(\frac{d_2}{\lambda})\sin(\phi_{c;k}^{(p)})}; \dots; e^{j2(N_2-1)(\frac{d_2}{\lambda})\sin(\phi_{c;k}^{(p)})}]$, λ is the signal wavelength and d_1 and d_2 are the antenna spacing in horizontal and vertical directions respectively, with $d_1 = d_2 = \lambda/2$.

B. USER CLUSTERING AND HYBRID PRECODING

Users are grouped into clusters, where one cluster head user is chosen for each cluster. Cluster head users are selected based on the maximum Euclidean distance between each other. First, the distance between all users are calculated where $\mathbf{h}_{n,l}$ is the channel between the n^{th} antenna at BS and the l^{th} LIS element.

$$\mathbf{H} = \begin{bmatrix} \underbrace{\{1; 1; \dots; 1\}}_{1:L_s} & 0 & 0 & \underbrace{\{1; 1; \dots; 1\}}_{1:L_s} \\ 0 & \underbrace{\{2; L_s+1; \dots; 2; 2L_s\}}_{2:L_s} & 0 & \underbrace{\{1; 1; \dots; 1\}}_{1:L_s} \\ 0 & 0 & 0 & \underbrace{\{N; L_s+1; \dots; N; L\}}_{N:L} \end{bmatrix}$$

The precoding matrix of LIS is denoted by $\mathbf{W} = \text{diag}(\mathbf{a}) \in \mathbb{C}^{L \times L}$. The phase vector of LIS elements is given by $\mathbf{a} = [a_1; a_2; \dots; a_L] \in \mathbb{C}^{L \times 1}$, where $a_l = e^{j\phi_l}$. The gain vector for LIS elements is denoted by $\mathbf{g} = [g_1; g_2; \dots; g_L] \in \mathbb{R}^{L \times 1}$. In passive RIS, the gain vector is assumed to be all ones ($\mathbf{g} = \mathbf{1}$).

Clustering algorithm is applied in order to group users. The total number of clusters is denoted by C . The set \mathcal{U}_i contains the index of users in i^{th} cluster. So $|\mathcal{U}_i|$ represents the total number of users in i^{th} cluster. Every cluster has at least one user. The received signal at LIS is modeled as

$$\mathbf{r} = \underbrace{\mathbf{H}^H \sum_{i=1}^C \sum_{j=1}^{|\mathcal{U}_i|} \mathbf{w}_i p_{i,j} s_{i,j}}_{\text{desired signal}} + \underbrace{\mathbf{z}}_{\text{dynamic noise}} + \underbrace{\mathbf{n}_a}_{\text{static noise}} \quad (2)$$

where $s_{i,j}$ is the transmitted signal with $\sum_{j=1}^{|\mathcal{U}_i|} |s_{i,j}|^2 = 1$ and $p_{i,j}$ is the transmitted power for j^{th} user in i^{th} cluster. The beamforming vector at BS for i^{th} cluster is denoted by \mathbf{w}_i , where the number of beams from BS is equal to the number of clusters C . The noise introduced due to active RIS can be divided into two parts: dynamic and static noise. The dynamic noise is denoted by \mathbf{z} , while the static noise is denoted by \mathbf{n}_a . The dynamic noise is introduced due to the active hardware components, i.e., reflection type amplifier, which amplifies the incident signal as well as the noise. This noise depends on the input power applied to the power amplifier circuit. It also depends on the range in which the power amplifier is operating. The dynamic noise has a significant value and cannot be neglected in the analysis. Where \mathbf{z} is as result of the input noise and the inherent device noise of active elements [30]. On the other hand, the static noise is because of the phase shifter circuit and it does not depend on \mathbf{r} . In consequence, static noise is usually neglected when compared to dynamic noise. The received signal $y_{c,k}$ for k^{th} user in c^{th} cluster is given as

$$\begin{aligned} y_{c,k} &= \mathbf{h}_{c,k}^H \sum_{i=1}^C \sum_{j=1}^{|\mathcal{U}_i|} \mathbf{w}_i p_{i,j} s_{i,j} + \mathbf{f}_{c,k}^H \mathbf{z} + n_{c,k} \\ &= \underbrace{\mathbf{h}_{c,k}^H \mathbf{w}_c p_{c,k} s_{c,k}}_{\text{desired signal}} \\ &\quad + \underbrace{\mathbf{h}_{c,k}^H \mathbf{w}_c \sum_{j=1}^{|\mathcal{U}_c|} p_{c,j} s_{c,j}}_{\text{intra-cluster interferences}} + \underbrace{\sum_{j=1}^{|\mathcal{U}_c|} p_{c,j} s_{c,j}}_{\text{intra-cluster interferences}} \\ &\quad + \underbrace{\mathbf{h}_{c,k}^H \sum_{i \in \mathcal{C}} \sum_{j=1}^{|\mathcal{U}_i|} \mathbf{w}_i p_{i,j} s_{i,j}}_{\text{inter-cluster interferences}} \\ &\quad + \underbrace{\mathbf{f}_{c,k}^H \mathbf{z}}_{\text{noise due to active RIS}} + \underbrace{n_{c,k}}_{\text{noise}} \end{aligned} \quad (3)$$

where $\mathbf{h}_{c,k} = \mathbf{g}_{c,k} + \mathbf{H} \mathbf{f}_{c,k}$ is the equivalent channel from BS to k^{th} user in the c^{th} cluster, the first term represents the direct link from BS to user, however the second path $\mathbf{H} \mathbf{f}_{c,k}$ represents the reflected link from LIS. The received signal $y_{c,k}$ at k^{th} user in c^{th} cluster in (3) is divided into: the desired signal, the interference from users in same cluster, interference due to users in other cluster, noise due to integrating LIS in the system and the noise at the receiver, respectively.

Digital Zero-Forcing (ZF) precoding vector is used at the BS and it is denoted by $\mathbf{w}_i \in \mathbb{C}^{N \times 1}$ for the i^{th} cluster. The precoding matrix $\mathbf{W} \in \mathbb{C}^{N \times C}$ is given by

$$\mathbf{W} = [\mathbf{w}_1; \mathbf{w}_2; \dots; \mathbf{w}_C] = \hat{\mathbf{H}} (\hat{\mathbf{H}}^H \hat{\mathbf{H}})^{-1} \quad (4)$$

where $\hat{\mathbf{H}} = [\mathbf{h}_{1,k}; \mathbf{h}_{2,k}; \dots; \mathbf{h}_{c,k}]$ and $\mathbf{h}_{c,k}$ represents the k^{th} user with the highest channel gain in the cluster.

The dynamic noise follows additive white Gaussian distribution, i.e., $\mathbf{z} \sim \mathcal{CN}(0, \sigma_z^2 \mathbf{I}_L)$ and the static noise of LIS (n_a) is neglected. While $n_{c,k}$ represents the noise at the user with $\mathcal{CN}(0, \sigma_{n_{c,k}}^2)$ distribution. Assuming users are ordered in each cluster based on channel gain in descending order. In (3), the second summation in the second term (intra-cluster interference) with summation greater than k is cancelled due to applying NOMA using successive interference cancellation (SIC). Then, the signal to noise and interference ratio (SINR) for k^{th} user in c^{th} cluster is given by

$$\begin{aligned} \gamma_{c,k} &= (\mathbf{h}_{c,k}^H \mathbf{w}_c)^2 p_{c,k} : (\mathbf{h}_{c,k}^H \mathbf{w}_c)^2 \sum_{j=1}^{|\mathcal{U}_c|} p_{c,j} \\ &\quad + \sum_{i \in \mathcal{C}} \sum_{j=1}^{|\mathcal{U}_i|} |\mathbf{h}_{c,k}^H \mathbf{w}_i|^2 p_{i,j} + \sum_{j=1}^{|\mathcal{U}_c|} |\mathbf{h}_{c,k}^H \mathbf{w}_c|^2 p_{c,j} + \sigma_z^2 + \sigma_{n_{c,k}}^2 \end{aligned} \quad (5)$$

The spectral efficiency of the k^{th} user in the c^{th} cluster is

$$R_{c,k} = \log_2(1 + \gamma_{c,k}) \quad (6)$$

The total spectral efficiency is written as

$$R_{\text{sum}} = \sum_{c=1}^C \sum_{k=1}^{|\mathcal{U}_c|} R_{c,k} \quad (7)$$

The energy efficiency is defined as the ratio of the total spectral efficiency to the total power consumption

$$EE = \frac{R_{\text{sum}}}{P_c^{\text{tot}}} \quad (\text{bps/Hz/W}); \quad (8)$$

Where P_c^{tot} is the total power consumption. The power consumption of active RIS is due to the switches, phase shift control circuit and the power amplification circuit. There are many factors that leads to power dissipation in active RIS, i.e., the hardware implementation, DC power, output power as well as load characteristics. While for the passive RIS, the power dissipation is only due to the switches and phase shift control circuit ($P_{\text{passive}} = L P_{\text{PS}}$).

Then, the total power consumption is given as the powerfunction, thus natural logarithmic is considered through the dissipated in the BS, active RIS and users as

$$P_c^{\text{tot}} = \underbrace{\frac{P_{\text{tx}} + P_B}{Z}}_{\text{power dissipated due to BS}} + \underbrace{\left(\sum_{i=1}^C \sum_{j=1}^J \eta_{i,j} \mathbf{h}_{i,j}^H \mathbf{w}_i \mathbf{j}^2 p_{i,j} + \sum_{j=1}^J \eta_j^2 \frac{2}{z} \right)}_{\text{power dissipated due to active RIS}} + \underbrace{\frac{(P_{\text{PS}} + P_{\text{PA}})}{Z}}_{\text{power dissipated by phase shift and power amplifier}} + \underbrace{\sum_{i=1}^C \frac{P_{i,j}}{Z}}_{\text{power dissipated by users}}; \quad (9)$$

where $\eta_{i,j}$ and η_j are the inverse energy conversion coefficients at the BS and LIS, respectively. In addition, P_B is the baseband power consumption. While, P_{tx} is the total transmitted power at BS, i.e. $P_{\text{tx}} = \sum_{i=1}^C \sum_{j=1}^J \eta_{i,j} \mathbf{w}_i \mathbf{j}^2 p_{i,j}$. The power dissipated per user is denoted by $P_{i,j}$. However the second term in (9) represents the power dissipated due to RIS, and P_{PA} are the static power of LIS hardware for the phase shift circuit and the power amplifier, respectively.

The optimization problem is presented in this section. The objective function is to maximize the energy efficiency with respect to the user power \mathbf{p} (and LIS phase ϕ) and gain \mathbf{a} . The maximization problem is written as

$$\max_{\mathbf{p}; \mathbf{a}} \text{EE} \quad (10a)$$

$$\text{st: } C_1 : p_{i,j} > 0; \quad \forall i, j \quad (10b)$$

$$C_2 : \sum_{i=1}^C \sum_{j=1}^J \eta_{i,j} \mathbf{w}_i \mathbf{j}^2 p_{i,j} \leq P_{\text{max}}^{\text{BS}}; \quad (10c)$$

$$C_3 : \left(\sum_{i=1}^C \sum_{j=1}^J \eta_{i,j} \mathbf{h}_{i,j}^H \mathbf{w}_i \mathbf{j}^2 p_{i,j} + \sum_{j=1}^J \eta_j^2 \frac{2}{z} \right) \leq P_{\text{max}}^{\text{RIS}}; \quad (10d)$$

$$C_4 : |j_{i,j}| = 1; \quad \forall i, j \quad (10e)$$

$$C_5 : a_i > 0; \quad \forall i \quad (10f)$$

where C_1 is to guarantee that the power assigned to each user must be positive. C_2 indicates that the total power consumption by the BS is less than the maximum power assigned to the BS ($P_{\text{max}}^{\text{BS}}$). However C_3 guarantees that the power consumed due to LIS does not exceed the maximum power dedicated for LIS ($P_{\text{max}}^{\text{RIS}}$). The last two constraints for the unit modulus phase shift constraint, and power amplifier can be any positive number.

Let us define $\mathbf{P}_{\text{max}}^{\text{BS}} = P_{\text{max}}^{\text{BS}}$, $\mathbf{P}_B = P_B$, $\mathbf{P}_{\text{max}}^{\text{RIS}} = P_{\text{max}}^{\text{RIS}}$, $\mathbf{L}(P_{\text{PS}} + P_{\text{PA}})$ and $\mathbf{p} = [p_{1,1}; p_{1,2}; p_{1,j}; \dots; p_{C,j}; p_{C,j}]^T$ is a vector of the power allocated for all users. The optimal solution does not affect by the base of the logarithmic

rest of the paper.

The problem in (10) is non-convex because the hessian matrix is negative since it is a non-linear function consisting of fractional and logarithmic functions. The non-convexity implies that the problem may have multiple local optima or discontinuities. To solve this non-convex problem, Dinkelbach's algorithm is applied to make the problem in a form of a non-fractional so the Hessian matrix becomes non-negative. Dinkelbach's algorithm is an efficient method to solve fractional problem by decoupling the numerator and denominator of a single ratio [31]. Thus, Dinkelbach's algorithm transformed (10a) into parameterized problem as

$$f_1(\mathbf{p}; \mathbf{a}) = R_{\text{sum}} - \gamma P_c^{\text{tot}} \quad (11)$$

where $\gamma = \frac{R_{\text{sum}}}{P_c^{\text{tot}}}$ is an auxiliary variable. The optimal energy efficiency γ^{opt} satisfies this condition $f_1(\mathbf{p}; \mathbf{a}) = 0$. The auxiliary variable γ is updated iteratively until the condition is satisfied. Now, the maximization problem is given as

$$\begin{aligned} \max_{\mathbf{p}; \mathbf{a}} \quad & f_1(\mathbf{p}; \mathbf{a}) \\ \text{st:} \quad & C_1; C_2; C_3; C_4; C_5 \end{aligned} \quad (12)$$

In order to solve (12), which is still non-convex problem. We apply fractional programming. A closed form for fractional programming (FP) was proposed to deal with sum of logarithms of ratio [32]. The SINR in (5), can be transformed to

$$c_{c,k} = \frac{j A_{c,k} j^2}{B_{c,k} - j A_{c,k} j^2} \quad (13)$$

where

$$A_{c,k} = \mathbf{h}_{c,k}^H \mathbf{w}_c \mathbf{p}_{c,k} \quad (14a)$$

$$\begin{aligned} B_{c,k} = & \sum_{j=1}^K \mathbf{h}_{c,k}^H \mathbf{w}_c \mathbf{p}_{c,j} j^2 \\ & + \sum_{i \in \mathcal{C}, j=1}^J \eta_{i,j} \mathbf{h}_{c,k}^H \mathbf{w}_i \mathbf{p}_{i,j} j^2 \\ & + \sum_{i \in \mathcal{C}} \eta_i^2 \left(\frac{2}{z} + \eta_i^2 \right) \end{aligned} \quad (14b)$$

By introducing an auxiliary variable $c_{c,k}$, the logarithmic function is tackled based on

$$\ln(1 + c_{c,k}) = \max_{c_{c,k}} \ln(1 + c_{c,k}) - c_{c,k} + (1 + c_{c,k}) \left[\frac{c_{c,k}}{1 + c_{c,k}} \right] \quad (15)$$

From (13), we can write:

$$\frac{c_{c,k}}{1 + c_{c,k}} = \frac{j A_{c,k} j^2}{B_{c,k}}$$

After substituting the previous expression in (15), the problem is written as

$$\begin{aligned} \max_{\mathbf{p}; \mathbf{a}; \mathbf{c}; \mathbf{g}} & \sum_{c=1}^P \sum_{k=1}^{U_c} \ln(1 + \frac{A_{c;k} j^2}{B_{c;k}}) \\ \text{st:} & C_1; C_2; C_3; C_4; C_5; \mathbf{c}; \mathbf{g} \geq 0 \end{aligned} \quad (16)$$

where $\mathbf{c} = [c_{1,1}; c_{1,2}; \dots; c_{1,U_1}; \dots; c_{j,U_c}]^T \in \mathbb{R}^{2 \times C^M - 1}$.

For a given \mathbf{c} , one can focus on the sum of ratio problem as

$$\max_{\mathbf{p}; \mathbf{a}; \mathbf{g}} \sum_{c=1}^P \sum_{k=1}^{U_c} \frac{A_{c;k} j^2}{B_{c;k}} \quad (17)$$

By introducing new auxiliary variable $\mathbf{g} = [g_{1,1}; \dots; g_{j,U_c}]^T \in \mathbb{R}^{2 \times C^M - 1}$, using quadratic transform, problem (17) is equivalent to

$$\max_{\mathbf{p}; \mathbf{a}; \mathbf{g}} \sum_{c=1}^P \sum_{k=1}^{U_c} 2 \text{Re} \sqrt{A_{c;k} g_{j,c;k} j^2 B_{c;k}} \quad (18)$$

After substituting by (18), the original optimization problem (12) is given by

$$\begin{aligned} \max_{\mathbf{p}; \mathbf{a}; \mathbf{g}; \mathbf{c}} & f_2(\mathbf{p}; \mathbf{a}; \mathbf{g}; \mathbf{c}) \\ \text{st:} & C_1; C_2; C_3; C_4; C_5 \end{aligned} \quad (19)$$

$$\begin{aligned} \text{where } f_2(\mathbf{p}; \mathbf{a}; \mathbf{g}; \mathbf{c}) = & \sum_{c=1}^P \sum_{k=1}^{U_c} [\ln(1 + \frac{A_{c;k} j^2}{B_{c;k}}) \\ & + \frac{P}{c=1} \sum_{k=1}^{U_c} \frac{P}{1 + \frac{A_{c;k} j^2}{B_{c;k}}} \text{Re} \sqrt{A_{c;k} g_{j,c;k}} \\ & + \frac{P}{c=1} \sum_{k=1}^{U_c} \frac{P}{c;k j^2 B_{c;k}} \mathbf{P}_c^{\text{tot}} \end{aligned}$$

The optimal solution of this problem is obtained using alternating joint optimization [33]. The main idea is to optimize each variable independently while assuming the other variables are fixed.

In this section, problem (19) is solved using alternating optimization. The objective is to obtain the optimum value for the user power allocation (\mathbf{p}^{opt}), the auxiliary variables (\mathbf{g}^{opt} & \mathbf{a}^{opt}) and LIS parameters (\mathbf{c}^{opt} & \mathbf{a}^{opt}).

A. OPTIMIZATION OF POWER ALLOCATION

In this section, the optimum user power is derived. Assume the other parameters (\mathbf{a} , \mathbf{g} , \mathbf{c}) are fixed. The optimization problem (19) can be rewritten as

$$\max_{\mathbf{p}} f_3(\mathbf{p}) \quad (20a)$$

$$\text{st: } C_1 : p_{i,j} \geq 0; \quad 8i \in 2; j \in 2; U_i \quad (20b)$$

$$C_2 : \sum_{i=1}^P \sum_{j=1}^{U_i} j j w_i j j^2 p_{i,j} \leq P_{\max}^{\text{BS}}; \quad (20c)$$

$$\text{where } f_3(\mathbf{p}) = \sum_{c=1}^P \sum_{k=1}^{U_c} A_{c;k} \text{Re} \sqrt{B_{c;k} \frac{P}{p_{c;k}}} g$$

$$\begin{aligned} & \sum_{c=1}^P \sum_{k=1}^{U_c} \frac{P}{c;k j^2} \sum_{j=1}^{U_c} j D_{c;k;c} \frac{P}{p_{c;j}} \\ & + \sum_{i \in C} \sum_{j=1}^{U_i} j D_{c;k;i} \frac{P}{p_{i,j}} + E_{c;k} \\ & \sum_{i=1}^P \sum_{j=1}^{U_i} j j w_i j j^2 p_{i,j} \leq Q \\ & \sum_{i=1}^P \sum_{j=1}^{U_i} Z_i p_{i,j} + Y + X \end{aligned}$$

where

$$\begin{aligned} A_{c;k} &= 2 \frac{P}{1 + \frac{A_{c;k}}{B_{c;k}}}; \quad D_{c;k;i} = h_{c;k}^H w_i; \quad B_{c;k} = \frac{P}{c;k} D_{c;k;c} \\ E_{c;k} &= j j f_{c;k}^H j j^2 \frac{2}{z} + \frac{2}{z}; \quad P_{\max}^{\text{BS}} = \frac{1}{P_{\max}^{\text{BS}}}; \end{aligned}$$

$$Q = P_B + M P_u + L (P_{\text{PS}} + P_{\text{PA}}); \quad Z_i = j j \sum_{j=1}^{U_i} h_{i,j}^H w_i j j^2;$$

$$X = \sum_{i=1}^P \sum_{j=1}^{U_i} \ln(1 + \frac{A_{c;k}}{B_{c;k}}) \frac{P}{c;k}; \quad Y = j j \sum_{j=1}^{U_i} \frac{2}{z}$$

In order to simplify the objective function in (20), $f_4(\mathbf{p})$ can be rewritten as

$$\begin{aligned} f_4(\mathbf{p}) = & \sum_{c=1}^P \sum_{k=1}^{U_c} A_{c;k} \text{Re} \sqrt{B_{c;k} \frac{P}{p_{c;k}}} \\ & \sum_{c=1}^P \sum_{k=1}^{U_c} \frac{P}{c;k j^2} \sum_{j=1}^{U_c} j D_{c;k;c} j^2 p_{c;j} \\ & + \sum_{i \in C} \sum_{j=1}^{U_i} j D_{c;k;i} j^2 p_{i,j} + E_{c;k} \\ & \sum_{i=1}^P \sum_{j=1}^{U_i} j j w_i j j^2 p_{i,j} \leq Q \\ & \sum_{i=1}^P \sum_{j=1}^{U_i} Z_i p_{i,j} + Y + X \end{aligned} \quad (21)$$

To solve this optimization problem, we use Lagrange multiplier. The Lagrangian function will be as follows

$$L(\mathbf{p}; \lambda) = f_4(\mathbf{p}) - \lambda \left(\sum_{i=1}^P \sum_{j=1}^{U_i} j j w_i j j^2 p_{i,j} - P_{\max}^{\text{BS}} \right)$$

where λ is the Lagrange multiplier for C_2 constraint, $\lambda \geq 0$. The dual objective function is given by

$$L_D(\lambda) = \max_{\mathbf{p}} L(\mathbf{p}; \lambda) \quad \text{s.t. } C_2$$

Then, the dual optimization problem can be written as

$$\lambda = \min_{\lambda \geq 0} L_D(\lambda)$$

User optimal power allocation is obtained by differentiating the Lagrange function $L(\mathbf{p}; \lambda)$ with respect to $p_{c;k}$ and λ as

$$\frac{\partial L(\mathbf{p}; \lambda)}{\partial p_{c;k}} = 0; \quad \frac{\partial L(\mathbf{p}; \lambda)}{\partial \lambda} = 0$$

The necessary and sufficient KKT conditions are given as where

$$\begin{aligned}
 1) \quad & \frac{\partial \mathcal{L}(\mathbf{p}; \lambda)}{\partial p_{c;k}} = 0 \\
 2) \quad & \mathbf{1}^T \left(\sum_{i=1}^C \mathbf{P}_c \sum_{j=1}^L \mathbf{w}_{ij} \mathbf{w}_{ij}^H \mathbf{p}_{ij} \right) - P_{\max}^{\text{BS}} = 0 \\
 3) \quad & \sum_{i=1}^C \mathbf{P}_c \sum_{j=1}^L \mathbf{w}_{ij} \mathbf{w}_{ij}^H \mathbf{p}_{ij} - P_{\max}^{\text{BS}} \\
 4) \quad & \lambda_1 \geq 0
 \end{aligned}$$

The solution for the optimal user power must satisfy those conditions.

B. OPTIMUM AUXILIARY VARIABLES

Consider fixed user power and LIS phase and gain, the optimal auxiliary variables (\mathbf{p}^{opt} ; λ^{opt}) are obtained by differentiating $\mathcal{L}(\mathbf{p}; \lambda; \mathbf{a}; \mathbf{g})$ given in (19) with respect to $p_{c;k}$ and $\lambda_{c;k}$ and equate to zero. So, the optimal auxiliary variables are given as follows

$$\begin{aligned}
 p_{c;k}^{\text{opt}} &= \frac{D_{c;k}}{p^2} (D_{c;k} + \sqrt{D_{c;k}^2 + 4}) \quad (22a) \\
 \lambda_{c;k}^{\text{opt}} &= \frac{p^2}{B_{c;k}} \quad (22b)
 \end{aligned}$$

where $D_{c;k} = \text{Re} \{ \mathbf{f}_{c;k}^H \mathbf{A}_{c;k} \mathbf{g} \}$, $A_{c;k}$ and $B_{c;k}$ are given in (14).

C. OPTIMIZATION OF LIS PHASE AND GAIN

For notation simplicity, let $\mathbf{D} = \text{diag}(\mathbf{d})$, \mathbf{a} is a vector of size $L \times 1$ and defined as $\mathbf{a} = [a_1; a_2; \dots; a_L]^T = [a_1; a_2; \dots; a_L]^T$.

In order to get the optimum values for phase shift and gain of LIS elements, problem (19) is formulated as

$$\begin{aligned}
 \max_{\mathbf{a}} \quad & \text{Re} \{ \mathbf{H} \mathbf{x} \mathbf{g} \} - \lambda^T \mathbf{S} \\
 \text{s.t.} \quad & \mathbf{C}_3: \mathbf{H} \mathbf{T} \leq P_{\max}^{\text{RIS}} \\
 & \mathbf{C}_4; \mathbf{C}_5
 \end{aligned} \quad (23)$$

$$\begin{aligned}
 \mathbf{x} &= \sum_{c=1}^C \sum_{k=1}^K \frac{p_{c;k}}{2(1 + \lambda_{c;k})} \text{diag}(\mathbf{f}_{c;k}^H) \mathbf{c} \\
 &+ \sum_{c=1}^C \sum_{k=1}^K \frac{\lambda_{c;k}}{2} \text{diag}(\mathbf{f}_{c;k}^H) \mathbf{c} p_{c;k} \\
 &+ \sum_{i \in \mathcal{C}} \sum_{j=1}^L \frac{\lambda_{ij}}{2} \text{diag}(\mathbf{f}_{c;k}^H) \mathbf{i} p_{ij} \\
 \mathbf{T} &= \sum_{i=1}^C \sum_{j=1}^L p_{ij} \text{diag}(\mathbf{d}_i) (\text{diag}(\mathbf{d}_i))^H + \frac{1}{2} \mathbf{I}_L \\
 \mathbf{S} &= \sum_{c=1}^C \sum_{k=1}^K \frac{p_{c;k}^2}{2} \text{diag}(\mathbf{f}_{c;k}^H) \text{diag}(\mathbf{f}_{c;k}) \\
 &+ \sum_{c=1}^C \sum_{k=1}^K \frac{\lambda_{c;k}}{2} p_{c;k} \text{diag}(\mathbf{d}_c) \mathbf{f}_{c;k} \mathbf{f}_{c;k}^H \text{diag}(\mathbf{d}_c) \\
 &+ \sum_{i \in \mathcal{C}} \sum_{j=1}^L p_{ij} \text{diag}(\mathbf{d}_i) \mathbf{f}_{c;k} \mathbf{f}_{c;k}^H \text{diag}(\mathbf{d}_i) + \mathbf{T}
 \end{aligned}$$

Since, $\mathbf{h}_{c;k} = \mathbf{g}_{c;k} + \mathbf{H} \mathbf{f}_{c;k}$, then, $\mathbf{h}_{c;k}^H \mathbf{w}_i = \mathbf{g}_{c;k}^H \mathbf{w}_i + \mathbf{f}_{c;k}^H \text{diag}(\mathbf{d}_i) \mathbf{w}_i$, where $\mathbf{g}_{c;k}^H \mathbf{w}_i = \mathbf{g}_{c;k}^H \mathbf{w}_i$; $\mathbf{f}_{c;k}^H \text{diag}(\mathbf{d}_i) \mathbf{w}_i = \mathbf{H}^H \mathbf{w}_i$; $\mathbf{d}_i = \mathbf{C}^L \mathbf{1}$.

Let

$$P_{\max}^{\text{RIS}} = P_{\max}^{\text{RIS}}$$

Note that problem (23) is a standard quadratic constraint quadratic programming (QCQP) problem, so it can be solved using Lagrange multiplier method [34], the optimal solution can be obtained by

$$\mathbf{p}^{\text{opt}} = (\mathbf{S} + \lambda_2 \mathbf{T})^{-1} \mathbf{x} \quad (24)$$

where λ_2 is Lagrange multiplier, which should be chosen such that complementary slackness conditions of power constraint \mathbf{C}_3 of the problem (23) is satisfied. The optimal Lagrange multiplier λ_2^{opt} can be obtained by binary search [34]. An initial range for λ_2 is chosen, i.e., $[\lambda_2^{\min}, \lambda_2^{\max}]$. Then, power constraint \mathbf{C}_3 of the problem (23) is calculated for λ_2^{\min} and λ_2^{\max} . If the calculated minimum power is less than the maximum RIS power, then the optimum value is reached. If both the calculated minimum power and maximum power are higher than the maximum RIS power (the constraint is not satisfied), the range is updated and the power is recalculated for the new maximum and minimum values. From [35], since λ_2 is non decreasing after each iteration, thus convergence of alternating algorithm to local optimal is guaranteed. So, if the updated values of \mathbf{p} ; \mathbf{a} ; and λ_2 are all optimal, a local optimal solution can be obtained by alternating optimizing these parameters until converges of problem (11).

Algorithm 1: Joint Iterative Optimization Algorithm

Input: System parameters and Channels.

Output: Optimized variables \mathbf{p}^{opt} ; \mathbf{a}^{opt} and \mathbf{d}^{opt} .

Initialization: \mathbf{p} ; \mathbf{a} ; \mathbf{d} , and $t = 1$

- 1: while no convergence of f_1 OR $t \geq T_{\max}$ do
 - 2: update \mathbf{t} and \mathbf{t} as in (22)
 - 3: update \mathbf{p}^t
 - 4: update \mathbf{a}_2 using binary search
 - 5: update \mathbf{d}^t as in (24)
 - 6: $t = t + 1$
 - 7: end while
 - 8: return \mathbf{p}^{opt} based on \mathbf{p} and \mathbf{t} .
-

(20a), (22a), (22b) and (24) respectively. The complexity of updating \mathbf{p} is $O(M^2N)$ and the complexity of updating \mathbf{a} is $O(MN)$. The complexity of updating \mathbf{d} is $O(M^2N + ML)$. For a given tolerance ϵ , the complexity of solving QCQP is $O(\log_2(\frac{1}{\epsilon})(\sqrt{L+1}(1+2L)L^3))$. Therefore, the overall complexity of the algorithm is $O(\log_2(\frac{1}{\epsilon})T(L^{4.5} + M^2N))$, where T is the number of iterations required for convergence.

In comparison with other computational complexity model, the complexity of LIS phase and gain and the auxiliary variables \mathbf{a} and \mathbf{d} are similar to the complexity computation of Algorithm 1 in [18]. However, the power complexity in our work can be compared with the precoding complexity which is $O(\log_2(\frac{1}{\epsilon})T(\sqrt{NM+2}(1+NM)M^3N^3))$. In addition, the overall complexity of the other model is $O(\log_2(\frac{1}{\epsilon})T(M^{4.5}N^{4.5} + L^{4.5}))$. It shows that our algorithm is much lower in complexity than the other model. Assuming $M=3$, $N=4$ and $L=16$, then the overall complexity of our algorithm is 262180 while the other model is 333975.

In this section, the simulation results are presented. The BS is located at (0,-20,0), RIS is located at (100,5,0) and users are uniform distributed in circle centered at (0,0) with radius 5m. The users positions are fixed over the simulations at $A = 200\text{m}$, where A represents the horizontal distance from the origin (0,0,0). Users are grouped in clusters, where cluster heads are first chosen then the remaining users are assigned to the nearest cluster head user as described in Section III. From [36], far-field assumption is valid, when the distance between the scattering source and the receiver is larger than Fraunhofer distance d_f . In [34], it is stated that when the number of elements is less than 361, the far field distance is about 5m. In this work, the distance between BS and RIS is 100m, the distance between RIS center of users distribution circle is 100m and the distance between BS and center of user distribution circle is 200m. All distances are larger than the far field distance, thus far field is considered.

We consider the BS employs Uniform Linear Array (ULA) with N antennas. Passive RIS is included for comparison purposes. RIS (active or passive) employs Uniform Planar Array (UPA) with $L = L_h L_v$ where L_h is the number of elements along horizontal axis and L_v is the number of elements along vertical axis. The mmWave channel model that is considered in the simulation, has three paths ($B_{c;k} = 3$), one line of sight $\mathbf{B}_{c;k}^{(1)}$ and two non-line of sight $\mathbf{B}_{c;k}^{(p)}$; $p \in \{2, 3\}$. The paths $\mathbf{B}_{c;k}^{(1)}$ and $\mathbf{B}_{c;k}^{(p)}$ follow $CN(0; 10^{-2})$ and $CN(0; 10^{-3})$ respectively. While $\mathbf{B}_{c;k}^{(p)}$ and $\mathbf{B}_{c;k}^{(p)}$ follow uniform distribution from $[\theta; \phi]$. The values of the simulation parameters are shown in Table 1.

In this paper, four cases are explored in the simulation:

- Case 1: Optimization of phase, gain of LIS elements and power allocation of users (proposed algorithm).
- Case 2: Optimization of phase, gain of LIS elements only (power is equally divided among users).

Energy Efficiency vs. the number iterations (T_{\max}) for LIS and passive RIS.

The joint iterative optimization algorithm is explained in Algorithm 1. The optimization parameters are first initialized. Then the auxiliary variables are obtained using (22). Afterwards, using the obtained auxiliary variables, the user power allocation is obtained when the necessary and sufficient conditions are satisfied. Then, the phase and gain of LIS elements are derived using (24). The algorithm is repeated from step 2 to 5 until f_1 in (11) is converged or the iteration index reached the maximum number of iterations T_{\max} . Finally, the optimal energy efficiency is obtained after substituting with the optimal parameters.

D. CONVERGENCE AND COMPLEXITY

To illustrate the convergence of Algorithm 1, we show in Fig. 3 the energy efficiency versus the number of iterations for LIS and passive RIS. The figure shows that, for passive RIS, the algorithm converges after about 2 iterations while, for the LIS, it converges after about 5 iterations. The reason behind that, is for the passive RIS, only phase and power are optimized while in LIS, three parameters are optimized, i.e., gain, phase, and power, which take more iterations.

The computational complexity of joint iteration algorithm depends on updating the four variables \mathbf{p} , \mathbf{a} , and \mathbf{d} using

Description	Values
BS location, $(x_1; y_1; z_1)$	(0m, -20m, 0m)
LIS location, $(x_2; y_2; z_2)$	(100m, 5m, 0m)
Number of antennas N	8
Number of users M	6
Number of LIS elements L	256
Number of clusters C	4
Inverse energy conversion at BS & LIS η	1.1
power dissipated per user P_u	10mW
static power consumption of phase shifter P_{PS}	10mW
static power consumption of power amplifier P_{PA}	10mW
noise variance at user & LIS σ^2	10^{-8} mW
Baseband power at BS P_B	6dB
maximum power at BS P_{max}^{BS}	[9dB, 18dB]
maximum power at RIS P_{max}^{RIS}	9dB

Simulation Parameters

Power Consumption vs. maximum power at BS (P_{max}^{BS}) for different cases of LIS.

Energy Efficiency vs. maximum power at BS (P_{max}^{BS}) for LIS and passive RIS.

Sum Rate vs. maximum power at BS (P_{max}^{BS}) for LIS and passive RIS.

Case 3: Optimization of users powers only (random phase and gain).

Case 4: Without any optimization (random phase, gain and equally divided power among users).

Energy efficiency comparison of passive RIS and LIS versus P_{max}^{BS} is shown in Fig. 4. The results show that as the maximum power dedicated for BS increases, the EE decreases. This is because of the increase of the total power consumption due to the increase of the BS power. Moreover, it can be observed that the EE obtained by jointly optimizing user powers, phase and gain of LIS elements has superior performance. This is also shown in the figure: case 4 has the highest power consumption.

The reason behind that is that by jointly optimizing phase, gain with power leads to minimize the power consumption shown in Fig. 6. As previously mentioned, as the maximum power dedicated for BS increases, the sum rate increases. For passive RIS, phase and user power optimization (case 1) outperforms passive RIS due to optimization of the gain. Similar to LIS, phase of passive RIS elements and user power optimization (case 1) has superior EE performance relative to case 3. For optimization of the user powers, when compared to the other optimization cases of passive RIS gain and phase (case 1) outperforms the optimization of RIS. There is a slight performance enhancement when user powers only (case 3). The figure shows that the sum

Sum Rate vs. number of RIS elements (L) for LIS and passive RIS.

Energy Efficiency vs. RIS Horizontal location for LIS and passive RIS.

Power Consumption vs. number of RIS elements (L) for LIS and passive RIS.

Sum Rate vs. RIS Horizontal location for LIS and passive RIS.

rate for LIS is higher than the sum rate for passive RIS for both cases. The figure also shows that, there is an increase in the LIS (case 1) relative to passive RIS (case 1) by 89% at $P_{max}^{BS} = 9\text{dB}$.

Sum rate versus the number of RIS elements for active and passive RIS is illustrated in Fig. 7. As the number of RIS elements increases, the sum rate increases. This is because as L increases, the reflected links to the users are increased and then the received signal is enhanced. In addition, sum rate of LIS outperforms passive RIS. This is due to the effect of the gain of the power amplifier of the LIS. Moreover as mentioned before, the optimization of RIS elements and user power give better sum rate when compared to the optimization of user power only for active/passive RIS.

The tradeoff between active and passive RIS in terms of power consumption is shown in Fig. 8. The figure shows that LIS suffers from high power consumption relative to passive RIS. This results from power dissipated in power amplifier in LIS. It is clear that as the number of RIS elements increases, the total power consumption increases. Since in passive RIS

the phase norm equals to one, thus the passive RIS power consumption for phase and user power optimization (case 1) is the same as the power consumption for user power optimization (case 3) indicated by two overlapping curves. However in LIS, a random gain for user power optimization (case 3) is considered, which leads to higher power consumption compared to case 1.

Both Fig. 9 and Fig. 10 illustrate the energy efficiency and sum rate versus the RIS horizontal location for LIS/passive RIS respectively. It is shown that when RIS is located 100m in the x-axis, the curves achieve high EE and sum rate relative to the other locations, which indicates that in this scenario, the optimum location for the LIS is at 100m (mid way between the BS and the center of the users distribution circle).

In this paper, we propose the joint optimization of LIS elements (phase, gain) and NOMA user powers. The performance of LIS and passive RIS is investigated in mmWave multi-user MIMO-NOMA system. Results show that LIS outperforms passive RIS in spectral efficiency. In contrast, the total power consumption in LIS is high compared to passive RIS. However, LIS still enjoys higher energy efficiency than passive RIS. Thus, the joint optimization of LIS elements have a significant effect on the overall system performance. For future work, a hybrid LIS may be considered, in which some elements are passive and others are active. This will decrease the total power consumption of the system. In future work, we will also consider user fairness by updating the minimum rate constraint for each user.

- [1] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, "Millimeter-Wave Communications: Physical Channel Models, Design Considerations, Antenna Constructions, and Link-Budgets," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 870–913, 2017.
- [2] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C.-L. I, and A. Ghosh, "Millimeter Wave Communications for Future Mobile Networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.
- [3] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.
- [4] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A Survey of Non-Orthogonal Multiple Access for 5G," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.
- [5] R. Long, Y.-C. Liang, Y. Pei, and E. G. Larsson, "Active Reconfigurable Intelligent Surface-Aided Wireless Communications," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 4962–4975, Aug. 2021.
- [6] T. Sharma, A. Chehri, and P. Fortier, "Reconfigurable Intelligent Surfaces for 5G and beyond Wireless Communications: A Comprehensive Survey," *Energies*, vol. 14, no. 24, p. 8219, Dec. 2021.
- [7] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable Intelligent Surfaces for Energy Efficiency in Wireless Communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.
- [8] Q. Gu, D. Wu, X. Su, J. Jin, Y. Yuan, and J. Wang, "Performance comparison between reconfigurable intelligent surface and relays: Theoretical methods and a perspective from operator," *arXiv preprint arXiv:2101.12091*, 2021.
- [9] M. Jian, G. C. Alexandropoulos, E. Basar, C. Huang, R. Liu, Y. Liu, and C. Yuen, "Reconfigurable intelligent surfaces for wireless communications: Overview of hardware designs, channel models, and estimation techniques," *Intelligent and Converged Networks*, vol. 3, no. 1, pp. 1–32, Mar. 2022.
- [10] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable Intelligent Surfaces for Wireless Communications: Principles, Challenges, and Opportunities," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 3, pp. 990–1002, Sep. 2020.
- [11] Y. Cao, T. Lv, and W. Ni, "Intelligent Reflecting Surface Aided Multi-User mmWave Communications for Coverage Enhancement," *31st IEEE Annual International Symposium on Personal, Indoor and Mobile Radio Communications - IEEE, 2020*, pp. 1–6.
- [12] L. Zhang, Q. Wang, and H. Wang, "Multiple Intelligent Reflecting Surface aided Multi-user Weighted Sum-Rate Maximization using Manifold Optimization," in *2021 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE, 2021, pp. 364–369.
- [13] J. Li and J. Liu, "Sum Rate Maximization via Reconfigurable Intelligent Surface in UAV Communication: Phase Shift and Trajectory Optimization," in *IEEE/CIC International Conference on Communications in China (ICCC 2020)*. IEEE, 2020, pp. 124–129.
- [14] B. Guo, C. Sun, and M. Tao, "Two-way Passive Beamforming Design for RIS-aided FDD Communication Systems," *IEEE Wireless Communications and Networking Conference (WCNC 2021)*. IEEE, 2021, pp. 1–6.
- [15] T. Jiang and W. Yu, "Interference Nulling Using Reconfigurable Intelligent Surface," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 5, pp. 1392–1406, May 2022.
- [16] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent Reflecting Surface-Assisted Millimeter Wave Communications: Joint Active and Passive Precoding Design," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 14960–14973, 2020.
- [17] K. Liu, Z. Zhang, L. Dai, S. Xu, and F. Yang, "Active Reconfigurable Intelligent Surface: Fully-Connected or Sub-Connected?" *IEEE Communications Letters*, vol. 26, no. 1, pp. 167–171, Jan. 2022.
- [18] Z. Zhang, L. Dai, X. Chen, C. Liu, F. Yang, R. Schober, and H. V. Poor, "Active RIS vs. Passive RIS: Which will prevail in 6G?" *arXiv preprint arXiv:2103.15154*, 2021.
- [19] K. Zhi, C. Pan, H. Ren, K. K. Chai, and M. Elkashlan, "Active RIS Versus Passive RIS: Which is Superior With the Same Power Budget?" *IEEE Communications Letters*, vol. 26, no. 5, pp. 1150–1154, May 2022.
- [20] Y. Lu, M. Hao, and R. Mackenzie, "Reconfigurable intelligent surface based hybrid precoding for THz communications," *Intelligent and Converged Networks*, vol. 3, no. 1, pp. 103–118, Mar. 2022.
- [21] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, Sep. 2015.
- [22] A. S. de Sena, D. Carrillo, F. Fang, P. H. J. Nardelli, D. B. da Costa, U. S. Dias, Z. Ding, C. B. Papadias, and W. Saad, "What Role Do Intelligent Reflecting Surfaces Play in Multi-Antenna Non-Orthogonal Multiple Access?" *IEEE Wireless Communications*, vol. 27, no. 5, pp. 24–31, Oct. 2020.
- [23] Y. Liu, X. Mu, X. Liu, M. Di Renzo, Z. Ding, and R. Schober, "Reconfigurable Intelligent Surface-Aided Multi-User Networks: Interplay Between NOMA and RIS," *IEEE Wireless Communications*, vol. 29, no. 2, pp. 169–176, Apr. 2022.
- [24] P. Liu, Y. Li, W. Cheng, X. Gao, and X. Huang, "Intelligent Reflecting Surface Aided NOMA for Millimeter-Wave Massive MIMO With Lens Antenna Array," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 4419–4434, May 2021.
- [25] A. Ihsan, W. Chen, M. Asif, W. U. Khan, and J. Li, "Energy-efficient IRS-aided NOMA beamforming for 6G wireless communications," *arXiv preprint arXiv:2203.16099*, 2022.
- [26] A. Khaleel and E. Basar, "A Novel NOMA Solution With RIS Partitioning," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 1, pp. 70–81, Jan. 2022.
- [27] F. Fang, Y. Xu, Q.-V. Pham, and Z. Ding, "Energy-Efficient Design of IRS-NOMA Networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 14088–14092, Nov. 2020.
- [28] L. N. Ribeiro, S. Schwarz, and M. Haardt, "Low-complexity zero-forcing precoding for x1-mimo transmissions," in *2021 29th European Signal Processing Conference (EUSIPCO)*. IEEE, 2021, pp. 1621–1625.
- [29] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Transactions on wireless communications*, vol. 14, no. 6, pp. 3059–3075, 2015.
- [30] J.-F. Bousquet, S. Magierowski, and G. G. Messier, "A 4-GHz Active Scatterer in 130-nm CMOS for Phase Sweep Amplify-and-Forward," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 3, pp. 529–540, Mar. 2012.
- [31] K. Shen and W. Yu, "Fractional programming for communication systems—part I: Power control and beamforming," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, 2018.
- [32] —, "Fractional programming for communication systems—Part II: Uplink scheduling via matching," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2631–2644, 2018.
- [33] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM Journal on imaging sciences*, vol. 6, no. 3, pp. 1758–1789, 2013.
- [34] S. Boyd, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*. Foundations and Trends in Machine Learning. Now Publishers Inc., 2010, vol. 3:1.

