Rate-splitting Multiple Access and Dynamic User Clustering for Sum-Rate Maximization in Multiple RISs-aided Uplink mmWave System

Mayur Katwe 1, Keshav Singh 2, Bruno Clerckx 2, and Chih-Peng Li 2

1 National Sun Yat-sen University
2 Affiliation not available

October 30, 2023

Abstract

In this paper, a reconfigurable intelligent surfaces (RISs)-aided millimeter wave (mmWave) uplink (UL) rate-splitting multiple access (RSMA) system is investigated which targets to achieve better rate performance and enhanced coverage capability for multiple users. The considered UL RSMA model splits the rate for each user by dividing their message into multiple parts and hence exploits all the necessary degrees of freedom to achieve maximum capacity region and high user fairness. In particular, we focus on the sum-rate maximization for considered UL RSMA system subject to joint optimization of power allocation to the UL users and beamforming design, i.e., active receive beamforming at the base-station (BS) and passive beamforming at multiple RISs. To efficiently mitigate high inter-node interference in multi-user scenario, we first provided a low-complex user pairing scheme based on k-means clustering and then develop an effective low-cost alternating optimization framework to solve the joint optimization problem sub-optimally by decoupling the problem into different sub-problems of power allocation and beamforming design. Specifically, the sub-problems of power allocation and beamforming design are solved using successive convex approximation, Riemannian manifold and fractional programming techniques. Later, the unified solution based on block coordinate descent (BCD) algorithm is proposed. Extensive numerical simulations validate that the user-clustering effectively significantly improves the performance gain and the considered RSMA system outperforms the conventional multiple schemes in terms rate and user-fairness. Also, the exploitation of spatial correlation among each RIS elements i.e., non-diagonal phase-matrices at each RIS achieve better performance that conventional diagonal phase-matrices setting.
Rate-splitting Multiple Access and Dynamic User Clustering for Sum-Rate Maximization in Multiple RISs-aided Uplink mmWave System

Mayur Katwe, Member, IEEE, Keshav Singh, Member, IEEE, Bruno Clerckx, Fellow, IEEE, and Chih-Peng Li, Fellow, IEEE

Abstract

In this paper, a reconfigurable intelligent surfaces (RISs)-aided millimeter wave (mmWave) uplink (UL) rate-splitting multiple access (RSMA) system is investigated which targets to achieve better rate performance and enhanced coverage capability for multiple users. The considered UL RSMA model splits the rate for each user by dividing their message into multiple parts and hence exploits all the necessary degrees of freedom to achieve maximum capacity region and high user fairness. In particular, we focus on the sum-rate maximization for considered UL RSMA system subject to joint optimization of power allocation to the UL users and beamforming design, i.e., active receive beamforming at the base-station (BS) and passive beamforming at multiple RISs. To efficiently mitigate high inter-node interference in multi-user scenario, we first provided a low-complex user pairing scheme based on k-means clustering and then develop an effective low-cost alternating optimization framework to solve the joint optimization problem sub-optimally by decoupling the problem into different sub-problems of power allocation and beamforming design. Specifically, the sub-problems of power allocation and beamforming design are solved using successive convex approximation, Riemannian manifold and fractional programming techniques. Later, the unified solution based on block coordinate descent (BCD) algorithm is proposed. Extensive numerical simulations validate that the user-clustering effectively significantly improves the performance gain and the considered RSMA system outperforms the conventional multiple schemes in terms rate and user-fairness. Also, the exploitation of spatial correlation among each RIS elements i.e., non-diagonal phase-matrices at each RIS achieve better performance that conventional diagonal phase-matrices setting.

Index Terms

Rate-splitting multiple access (RSMA), reconfigurable intelligent surfaces (RISs), uplink transmission, active receive and passive beamforming, millimeter wave (mmWave) communications, sum-rate maximization.

I. INTRODUCTION

RECENTLY, reconfigurable intelligent surfaces (RISs), also known as intelligent reflecting surfaces (IRS), have been considered as a promising solution to improve the network coverage and to resolve blockages issue in millimeter wave (mmWave) communications [1]–[3]. Despite the potential of providing
gigabits-per-second communication, the severe path loss and high directivity make mmWave communication vulnerable to severe attenuation and blockages, which can be frequent in dense urban environments [4]. An RIS is a planar array consisting of multiple low-cost reflecting elements which steers the incident signal by adjusting the phase shift and amplitude [5]. Owing to this ability, RISs can reconfigure the wireless channel to facilitate information transmission and the performance of mmWave communications can be significantly enhanced by deploying RISs on the exterior walls of buildings [6].

Besides, the ever-rising demand of high data rates and dense wireless networks in recent years have led to the inevitable search for very high spectrum and power efficient technologies [7], [8]. As a possible candidate technique to radio access of the future wireless networks, rate-splitting multiple access (RSMA) has recently been recognized as a promising technology due to its superior spectral efficiency and effective resource sharing [8]–[14]. Particularly, RSMA allows multiple users to access the same orthogonal resource block and splits the rate of each user into various sub-messages [13]–[15]. The rate-split of the users enable efficient resource allocation and hence, secure enhanced spectral efficiency as compared to existing multiple access schemes such as orthogonal multiple access (OMA), non-orthogonal multiple access (NOMA) and space-division multiple access (SDMA) [11], [14]–[16]. Specifically, the rate-splitting in RSMA exploits all the necessary degrees of freedom in power control to achieve the entire capacity region and the optimal user fairness in the multiple access channel even with sub-optimal user pairing [17].

A. Interplay of RIS and RSMA

Primarily, the RIS-aided communication network are capable of resolving severe blockage issues in mm-wave communications by enhancing the combined channel strengths and coverage capability for each users [18], [19]. While, the key feature of RSMA technique is to split the user messages into multiple sub-messages so as to achieve more flexible interference management, and thus provides better rate-throughput while ensuring delay-limited transmission, optimal user-fairness, massive connectivity, spectral efficient transmission as compared to conventional OMA and NOMA [17], [20]. The interplay between RIS and RSMA techniques can bring out a new paradigm for wireless communication architectures which effectively utilizes their individual merits of enhanced spectral efficiency and coverage capability and render a competent solution for better system performance in next-generation wireless networks. The underlying merits of the RIS aided UL RSMA system are detailed as follows:

1) Enhanced spectral efficiency and Coverage capability: The integration of RIS and RSMA captivate the individual merits in terms of high spectral efficiency and enhanced coverage for the users on both transmission and reflection sides. Mainly, the optimal phase-shift design overcomes the randomness effects of channel and provide high spectral efficiency especially for RIS-aided network and this further enhances the system performance for power-domain multiple schemes i.e., RSMA and NOMA [21].
2) Smart RSMA design: The passive array gain that comes with high number of STAR-RIS elements relaxes the need of massive antenna systems for RSMA system, and it also maximize the overall performance significantly even in limited spectral resources when compared to NOMA system [11]. Moreover, the channel diversity and flexible interference management facilitated by RIS and RSMA, respectively prohibits the high power (energy) consumption for dead zone users to maintain high quality of service (QoS) via direct link (without RIS) [22].

3) Robustness to imperfect SIC: RSMA is robust to imperfect SIC due to which it becomes an competent solution for RIS aided networks for better system performance. Besides, the performance of multi-layer RSMA system is prone to imperfect SIC decoding. However, the passive array gain provided by RIS compensates the deterioration caused by imperfect SIC decoding up to possible extent [23].

4) Low hardware overhead: The channel diversity provided by RIS enforces the single-layer/ two-layer RSMA systems to attain performance equivalent to higher-layer level RSMA systems (without RIS system). Moreover, the performance gain achieved by STAR-RIS aided networks can significantly relax the sophisticated SIC receiver design for RSMA system.

B. Related Works

Many efforts have been dedicated to improve the system performance with respect to spectral efficiency, energy efficiency, outage probability, network coverage capability in RIS-assisted mmWave communication systems [18], [24]–[28]. For instance, the authors in [24] studied joint optimization of power allocation and phase-shifters in RIS-aided mmWave systems. In [26], an RIS-aided massive multiple-input multiple-output (MIMO) architecture for mmWave communications was investigated in which two efficient precoder were designed by exploiting the sparsity of mmWave channels. Generally, the utilization of multi-RIS can resolve the blockages issue for mmWave communication by providing smart-radio propagation environment and flexibility when compared to conventional RIS system. Even though the user is present in single RIS’s reflection half-space, the single RIS deployment may not guarantee an effective blockage-free transmission path between user and base-station (BS) especially in dynamic and dense urban environments [29]. Due to the practical limitation on the number of RIS elements on RIS module [30], the single IRS may render limited passive beamforming gain and thus can refrain a high gain performance. The authors in [19] demonstrated that received signal power increases quadratically with the number of reflecting elements for both the single IRS and multi-IRS cases. Moreover, they also demonstrate that the joint optimization of active beamforming at base station (BS) and passive beamforming at each RIS can significantly improve the rate performance. Notably, few works [31]–[33] in literature showed that spatial correlation among channels or among the phase-shifters of each RIS render positive impact on the system performance of RIS aided communication networks.
Further, the resource allocation problem for multi-channel RIS-aided NOMA and RSMA systems have attracted significant attention in past years. The joint power allocation and phase shifts optimization problems for sum-rate maximization and power minimization were studied for the downlink RIS-NOMA multi-user systems in [34]–[41], respectively, which validated the superior performance of NOMA over conventional OMA scheme. The authors in [34], [42]–[44] proposed algorithms to improve the spectrum and energy-efficiency fairness for a multiuser RIS aided network. In [45], [46], the authors concluded that the high SINR and high number of RIS elements improve the outage probability, ergodic capacity and rate throughput. Overall, as compared to conventional OMA scheme, the NOMA schemes for RIS aided network provide better spectral efficiency and coverage capability especially when direct link between communicating nodes is absent [47]–[49]. Nevertheless, the recent studies reveal that the system performance can be further improved using RSMA which splits the rate of user into various sub-messages [11]–[14], [16], [43], [50]–[53]. The closed-form expressions for the outage probability of cell-edge users and near users were derived in [51] for RIS assisted downlink communication which target to gain optimal resource allocation. It was shown that RSMA performs better than NOMA under various system parameters such as the number of RIS reflecting elements and the node density. The authors in [52] considered the maximization of the minimum transmission rate among multiple single-antenna users for RIS-aided multi-user MISO RSMA downlink system by optimizing the transmit beamformers at the transmitter and the phase sifters at the RIS. In [53], the problem of energy efficiency maximization was investigated by controlling the active and passive beamforming at the BS and at the RIS, respectively. The works in [51]–[53] amalgamate RIS and RSMA technologies to achieve optimal network performance in terms of energy efficiency and rate maximization and outage analysis.

C. Motivations and Contributions

The multi-RIS deployment provide effective LOS and thus render robust performance against blockages in mmWave communication [19]. Alternatively, the power control at the low-cost users and efficient spectrum sharing in RSMA design will become critical aspects in resource-limited networks. Overall, the integration of multi-RIS deployment and the RSMA can be beneficial in terms of better spectral efficiency, improved flexibility and and enhanced coverage especially in mmWave communication. However, the deep insights and solid theoretical analysis for the integration of multi-RIS and RSMA is majorly missing in the previous works. The theoretical and practical investigation of the system performance of multi-RIS-aided RSMA system is a hot research topic and thus, is the prime motivation of this work. Nevertheless, the selection of optimal power-splitting, decoding order and beamforming design schemes becomes very pivotal for multi-RIS aided RSMA system which bring new optimization challenges for the effective resource allocation design.

Despite many potential of both multi-RIS deployment and RSMA schemes, the research on integrated
multi-RIS aided UL RSMA system is still in infancy. As far as the authors are aware, the UL-RSMA has not been investigated with multi-RIS in the literature yet from neither a theoretical nor a resource allocation perspective. Motivated by this background, we investigated multi-RIS-RISs-aided UL-RSMA framework where the UL users operate in RSMA fashion and communicate with the BS via multi-RISs. The important contributions of this work are as follows:

1) We theoretically analyze the system performance of multi-RIS aided RSMA system in terms of spectral efficiency and coverage capability of uplink users under mmWave communication. In the considered RSMA system, a message of each user is split into several sub-messages and each part contributes to the rate of that user and the BS decodes them using an appropriate decoding order. To the extent of our knowledge, this is the first paper looking at spectral efficient resource allocation design for multi-RIS aided UL RSMA system by exploiting the mmWave channel characteristics.

2) To effectively mitigate the INI in the considered multi-RISs aided multi-user UL RSMA system, we first perform an unconventional dynamic user clustering based on low-cost k-means clustering, Later, a resource allocation design problem which aim to maximize the overall rate-throughput of the system under the joint optimization of power allocation, decoding order, and beamforming (including active receive beamforming at the BS and passive beamforming at each RIS) under the consideration of discrete phase-shift models and spatial-correlated phase-shift matrices at each RIS.

3) To solve the formulated resource allocation design problem for the considered STAR-RIS aided RSMA design, we decouple the transformed problem into various sub-problems of power allocation, decoding order and beamforming design separately. The sub-problems of power allocation and beamforming design are solved independently using general convex approximation and fractional programming approaches and later, a unified solution based on alternating optimization algorithm is presented. Further, we select a SIC decoding order strategy for considered UL RSMA which ensure optimal user-fairness index.

4) Deep theoretical insights and extensive numerical simulations are provided based on stochastic geometry of the considered model under correlated and uncorrelated mmWave channels. The simulation results demonstrate that a) the proposed solution attains fast convergence and outperforms the considered conventional schemes in terms of overall rate-throughput and user-fairness b) the proposed usering clustering significantly improves the rate performance c) the considered RSMA system significantly outperform conventional multiple access schemes i.e., NOMA and OMA and d) the spatial correlation among each RIS elements i.e., non-diagonal phase-matrices at each RIS achieve better performance that conventional diagonal phase-matrices.

Notations: Throughout the paper, the scalar, vectors, matrices and sets are represented by regular, bold lowercase, bold uppercase and scripts, respectively. \(||I|||\) indicates the \(\ell_2\) norm of the vector \(I\) and \(|I|\)
indicates magnitude of $l$. $p^{(a)}$ denotes the value of parameter $p$ at $a^{th}$ iteration whereas $\{p_i\}$ indicates the accumulation of all variables $p_i, \forall i$. $p_a$ denotes the $a^{th}$ element of vector $p$. $n \sim \mathcal{CN}(\mu, \sigma^2)$ denotes that $n$ is circularly symmetric Gaussian random variable with $\mu$ mean and variance $\sigma^2$. $\text{diag}(I)$ represents diagonal matrix with diagonal element set as $I$. $\Re(z)$ denotes real part of complex variable.

II. SYSTEM MODEL: MULTI-RIS AIDED UL RSMA SYSTEM

Let us consider a multiple RIS-aided mmWave UL RSMA system as shown in Fig. 1, where $M$ RISs are deployed to simultaneously assist the data transmission from $K$ single-antenna UL users to a multi-antenna BS. A set of multiple RISs are deployed on the surface of the buildings to assist the severely blocked users in communication. Without loss of generality, we ignore the multi-reflection signals which are reflected by the RIS two or more times due to the high path loss of mmWave transmissions [19]. The BS is equipped with $N_r$ receiving antennas and each RIS has $N$ reflecting elements. In the considered UL RSMA system, each user splits its own message into $J$ parts (sub-messages) and transmits them to BS simultaneously within same time and frequency slot [12]. $J = \{1, \ldots, J\}$ denotes the set of sub-messages for each user.

A. Channel and Signal Model

Let $x_k$ represents the transmitted signal from the $k^{th}$ user such that

$$x_k = \sum_{j \in J} \sqrt{p_{kj}} s_{kj}, \forall k \in \mathcal{K} \doteq \{1, \ldots, K\},$$

where $s_{kj}$ is the $j^{th}$ sub-message of the $k^{th}$ user with $\mathbb{E} \left| s_{kj} \right|^2 = 1$ and $p_{kj}$ is the corresponding transmit power allocated to the sub-message $s_{kj}$. The total transmission signal power $p_k = \sum_j p_{kj}$ of each user is limited upto $P^\text{max}$, i.e., $p_k = \sum_{j \in J} p_{kj} \leq P^\text{max}, \forall k \in \mathcal{K}$.

1Note that the scope of the considered RIS aided UL RSMA is not restricted for only mm-wave system. The proposed theoretical analysis in this work can be also extended to other sub-6Ghz band communication applications as well.

2The multi-reflection links are generally encountered in multi-RIS configuration where the signal transmitted by user is reached to BS via direct reflection as well as inter-RIS-reflection from multiple RIS. We consider that RIS are sparsely deployed at the network peripherals such that the inter-RIS links are generally weak. However, the diverse reflection paths including inter-RIS reflections available in the multi-IRS-aided system boosts extra DoFs and flexibility in system performance condition that multi-RIS are optimally placed [28]. Nevertheless, the consideration of joint resource allocation and optimal placement of RIS to maximize the overall system performance of the considered mmWave communication is out of the scope of this paper and will be considered as a future work.
The BS receives transmitted signal from each user through different communication channels i.e., RIS-aided communication (user-RIS) and direct communication channels (user-BS) as shown in Fig. 1. Specifically, each reflecting element on the RIS behaves like a single physical point which combines all the received signals and then re-scatters the combined signal with a certain phase shift and unity amplitude gain. At the BS, the multi-antenna array receives the signal from the user and performs receive beamforming. It is assumed that the channel state information (CSI) of all channels involved is perfectly known at the BS and the IRS under the consideration of quasi-static flat-fading channel model [56] as [37], [47], [48], [57]–[59].

Denote $h_{bk} \in \mathbb{C}^{N_r \times 1}$, $h_{km} \in \mathbb{C}^{N \times 1}$ and $H_{bm} \in \mathbb{C}^{N \times N_r}$ as the mm-wave channel between $k^{th}$ user-BS, $k^{th}$ user -$m^{th}$ RIS and $m^{th}$ RIS-BS, respectively. Assuming that mmWave channels are sparsely scattered into multiple paths. The BS employs a uniform linear array (ULA) with $N_r$ antennas, and each RIS consists of a uniform rectangular array (URA). The channel gain between $m^{th}$ RIS-BS can be given using Saleh-Valenzuela (SV) channel model [60] as

$$H_{bm} = \sqrt{1/L_{bm}} \left( \zeta_{bm,0} a_{lm,0} (\varphi_{lm,0}) a_{r,m,0}^T (\varphi_{r,m,0}) + \sum_{l=1}^{L_{bm}} \zeta_{bm,l} a_{lm,l} (\varphi_{lm,l}) a_{r,m,l}^T (\varphi_{r,m,l}) \right), \quad (2)$$

where $L_{bm}$ are the number of scattering paths for the $m^{th}$ RIS-BS link, $a_{lm,l} (\varphi_{lm,l})$ and $a_{r,m,l} (\varphi_{r,m,l})$ are the normalized transmit and receive array response vector and $\varphi_{lm,l}^t, \varphi_{km,l}^r$ and $\varphi_{mb,l}^r$ are the angle of arrival and departure for the $l^{th}$ path, respectively. $\zeta_{bm,l} \sim \mathcal{CN} (0, 10^{-0.1k_0})$ and $g_{mb,l} \sim \mathcal{CN} (0, 10^{-0.1k_l})$ is the complex channel gain corresponding to $l^{th}$ path with $k_l = a_l + 10\gamma_l \log_{10} \frac{d}{\epsilon}$ and $\epsilon \sim \mathcal{N} (0, \sigma_e^2)$. Note that $l = 0$ corresponds to LOS path.

**Proposition 1:** Each RIS-BS mmWave channel can be approximated using as a rank-one matrix approximation when LOS path dominates [19] such that,

$$H_{bm} = \zeta_{bm} a_{lm} a_{r,m}^H, \forall m \in \mathcal{M} \quad (3)$$

where $\zeta_{bm}$ is the path-loss-dependent complex gain, $a_{lm}$ and $b_{lm}$ are the normalized array response vector associated with the LOS path of the $m^{th}$ RIS and BS, respectively.

**Proof:** Besides the sparse scattering characteristics, few works [61], [62] in the literature have shown that the received power for the LOS path in mmWave communications is nearly 13dB higher than the sum of power of NLOS paths when their exist direct LOS link between BS and each RIS. Hence, the consideration of only LOS link is straightforward.

Simulations results validate that the rank-one approximation with only LOS path achieve performance nearly equivalent to the case while considering all paths into account.

---

3The results in this paper serve as theoretical performance upper bounds for the considered UL mmWave system which can provide a benchmark for the system design under imperfect CSI. Some related works on CSI estimation can be found in [53], [55]. Nevertheless, for the multi-RIS aided system involving single and multi-reflection coupled channel coefficients, the accurate and efficient estimate of complete CSI is out of the scope and can be considered as a open research problem for future works.

4Without loss of generality, it is assumed that the RISs maintain direct LOS path between BS and each RIS which can be performed using prior information on the BS location [28].
Let, $\Theta_m = \text{diag}(\theta_m)$, $\theta_m = [\theta_{m1}, \ldots, \theta_{mN}]^T$, $\Theta_m \in \mathbb{C}^{N \times N}$ and $z = [z_1, \ldots, z_{N_r}]^T \in \mathbb{C}^{N_r \times 1}$ denote the phase-shift matrix of the $m^{th}$ RIS and the receive beamforming vector of the BS, respectively such that $\theta_m = [\exp(j\phi_{r_{m1}}), \ldots, \exp(j\phi_{r_{mN}})]^T$. Note that $\phi_{r_{mn}}$ is the phase shift associated with the $m^{th}$ passive element of the $m^{th}$ RIS and the $\tilde{n}^{th}$ antenna of the BS, $m \in \mathcal{M} \triangleq \{1, \ldots, M\}$, $n \in \mathcal{N} \triangleq \{1, \ldots, N\}$, $\tilde{n} \in \mathcal{N}_r \triangleq \{1, \ldots, N_r\}$.

The received signal at the BS $y$ can be given as

$$y = \sum_{k=1}^{K} \left\{ \left( h_{bk}^H + \sum_{m=1}^{M} h_{rk_m} H \Theta_m H_{bm} \right) z x_k \right\} + n = \sum_{k=1}^{K} \sum_{j=1}^{J} \left\{ h_k(\Theta, z) \sqrt{p_{kj}s_{kj}} \right\} + n, \quad (4)$$

where $h_k(\Theta, z) \triangleq \left( h_{bk}^H + \sum_{m=1}^{M} h_{rk_m} H \Theta_m H_{bm} \right) z$ is the effective channel gain between $k^{th}$ user and the BS and $\Theta = [\ldots, \Theta_m, \ldots]$. $n \sim \mathcal{CN}(0, \sigma^2)$ denotes the additive white Gaussian noise at the BS where $\sigma^2$ represents the noise power.

**B. SIC Decoding and Rate-Throughput**

The considered RSMA system model serves all the users in the same time and frequency slots and the signal for the users are decoded successively based on the ascending order of their decoding order such that the users with low decoding order (termed as strong users) are decoded first and the users with high decoding order (termed as weak users) are decoded later. After SIC, the rate throughput of strong users is not subject to IUI, instead, it depends on its power allocation and channel gain. However, the rate throughput for weak users depend more upon the IUI from other strong users and therefore it becomes very critical for low order users especially when there exists large number of users.

The BS utilizes SIC to decode all the sub-messages of the users from the received signal $y$. There exist total $S = K \times J$ sub-messages from $K$ users. Assuming that decoding order of sub-messages at the BS is denoted as the set $\pi = \{\pi_{kj} : k \in \mathcal{K}, j \in \mathcal{J}\}$ in which first element is decoded first, second element is decoded second and so on. $\pi_{kj} \in S \triangleq \{1, \ldots, S\}$ denotes the decoding order of the sub-message $s_{kj}$.

The permutation $\pi$ belong to the set $\Pi$ which is set of all the possible decoding order of sub-messages. Particularly, for the sub-message $s_{kj}$, the BS successfully decodes and eliminates all the sub-messages which have low decoding order than $s_{kj}$ and treats remaining sub-messages as the interference (other than $s_{kj}$). So, the signal to noise-ratio (SINR) in RSMA system for the sub-message $s_{kj}$ can be given as

$$\gamma_{kj} = |h_k(\Theta, z)|^2 p_{kj} \left( \sum_{(u,v) \in Q_{kj}} |h_u(\Theta, z)|^2 p_{uv} + \sigma^2 \right), \quad (5)$$

where $\forall k \in \mathcal{K}$, $\forall j \in \mathcal{J}$ and $Q_{kj}$ is the set of all the sub-messages which have greater decoding order than $s_{kj}$ i.e., $Q_{kj} = \{(u, v) : \pi_{uv} > \pi_{kj}\}$.

**Remark 1:** Overall, the interference in the received signal become highly dominant than desired signal which results into very poor SINR for weak users due to which the optimal system performance is not guaranteed. Moreover, it also becomes impractical to implement the single centralized RSMA model for
large scale network due to high complexity of SIC design [16]. Therefore, in this paper, we investigate a
dynamic user clustering algorithm of RSMA system which effectively mitigate the IUI among the users.

We adopt a hybrid multiplexing scheme such that the inter-cluster users operate in OMA fashion and
intra-cluster users are served using RSMA which significantly mitigate the IUI among the users. Note that
all the RISs are allowed to operate whole bandwidth and they assist every user for the UL transmission
irrespective of the cluster. Importantly, we allocate dynamic time/ frequency resource to the users based on
their cluster density such that any \( t \)th cluster, \( \varphi_t \), is assigned with \( \tau_t \approx |\varphi_t|/K \) portion of the total bandwidth
allocated to the system. Hence, the cluster with high user density is assigned with higher bandwidth and
low density cluster with lower bandwidth. This ensure fair bandwidth distribution among all the users.

Assuming that all the UL users into \( T \) clusters \((T \leq K)\) such that the size of each cluster \( S_t \) is less than
maximum allowable cluster size \( T_s \) and moreover, the cluster set belong to the set of all the possible clusters
\( S \), i.e., \( S_t \in S, t \in T = \{1, \ldots, T\} \). After user clustering, the achievable rate for the \( k \)th user belonging to
the \( t \)th cluster can be given using (5) as

\[
 r_{kj}^t = \tau_t \log_2 \left( 1 + \gamma_{kj,t} \right) = \tau_t \left( \log_2 \left( \sum_{(u,v) \in \tilde{Q}_{kj}^t} |h_u(\theta,z)|^2 p_{uv} + \tau_t \sigma^2 \right) - \log_2 \left( \sum_{(u,v) \in \tilde{Q}_{kj}^t} |h_u(\theta,z)|^2 p_{uv} + \tau_t \sigma^2 \right) \right),
\]

where \( \tilde{Q}_{kj}^t = \{ (u, v) : \pi_{uv}^t \geq \pi_{kj}^t, u, k \in \varphi_t \} \) is the set of sub-messages which contribute IUI for the \( j \)th
sub-message of the \( k \)th user belonging to the \( t \)th cluster and \( \pi_{kj}^t \) represents its decoding order and \( \tilde{Q}_{kj}^{t+} = \{ (u, v) : \pi_{uv}^t \geq \pi_{kj}^t, u, k \in \varphi_t \} \).

III. PROBLEM FORMULATION: RESOURCE ALLOCATION FOR SUM-RATE MAXIMIZATION

The prime motivation of the considered multi-layer RSMA system is to maximize the spectral efficiency
of all the UL users using multi-RIS aided RIS aided network. In particular, we formulate the problem the
sum-rate maximization under the joint consideration of user clustering power allocation, decoding order
and beamforming design (including receive beamforming at the BS and passive beamforming at each RIS).

Now, the original optimization problem for resource allocation design can be formulated as

\[
 (P0) : \max_{C, \pi, \Theta, S_t, p, z} \sum_{k \in K} \sum_{t \in T} c_{kt} r_k = \sum_{k \in K} \sum_{j \in J} \sum_{t \in T} c_{kt} \tau_t \log_2 \left( 1 + \gamma_{kj,t} \right)
\]

s.t. (C1): \( \sum_{j \in J} p_{kj} \leq P_{\text{max}}, \forall k \in \mathcal{K}, \)

(C2): \( \sum_{k \in K} r_k \geq r_{\text{min}}, \forall k \in \mathcal{K}, \)

(C3): \( \pi \in \Pi, \)

(C4): \( |\Theta_{m[n,n]}| = 1, \forall m \in \mathcal{M}, \forall n \in \mathcal{N}, \)

(C5): \( |z_{ii}| = 1/\sqrt{N_r}, \forall i \in \mathcal{N}_r, \)

(C6): \( S_t \in \mathcal{S}, |S_t| = T_s, c_{kt} \in \{0, 1\}, \sum_{t \in T} c_{kt} = 1, \forall k \in \mathcal{K}, \forall t \in T, \)
where $p = \{p_{kj}\}$ and $C = \{c_{kt}\}$ are the collections of power allocation coefficients and user clustering indicators. The constraints (C1), (C2), (C3), (C4) and (C7) are the power allocation, minimum QoS requirement, decoding order, passive beamforming, receive beamforming and user-clustering constraints, respectively. Specifically, the constraint (C1) provide the maximum UL power allocation to the users, (C2) ensure that the resource allocation should maintain minimum rate requirement for each user, (C3) and (C4) are the constant modulus beamforming constraints of for RISs’ and BS antenna elements, respectively. The constraint (C5) and (C6) represent the selection of decoding order $\pi$ and user clustering from the set of all possible decoding orders, $\Pi$ and clustering sets, $S$, respectively.

It can be observed that the formulated resource allocation design (P0) in (7) is a NP-hard mixed-integer non-linear programming (MINLP) problem due to the non-convex nature of its constraints and strong coupling of variables. Primarily, the decoding order and clustering setting in the constraint (C5) and (C6) is integer programming.

**Proposition 2:** The problem (P0) in (7) is non-deterministic polynomial- (NP-) hard even when the decoding order and user clustering is optimally or sub-optimally designed.

**Proof:** Firstly, the constraint (C2) is neither convex nor concave with respect to the optimization variables and other constraints for beamforming design (C4) and (C5) are unity modulus, highly coupled or binary expressions. Consequently, the formulated resource allocation problem in (7) even after decoding order and user clustering design is nearly intractable.

**Remark 2:** Although, the exhaustive search may solve it, however the implementation of generic exhaustive search algorithms are not practically feasible as its computational complexity grows exponentially over the total number of variables.

**Remark 3:** Therefore, it is necessary to transform (P0) into some tractable sub-problems that can be solved separately and alternatively over multiple iterations. Many research works [34], [52], [53], [57] have witnessed that the effective utilization of general convex approximations, fractional programming and alternating optimization can solve complex resource allocation design problem with ultra-low complexity. The aggregation of general convex relaxation to solve such complex resource allocation design for Multi-RIS aided UL RSMA can be considered as a meaningful and reasonable contribution.

Overall, the problem (P0) in (7) is nearly intractable and moreover, there exist no standard methods to solve such complex problems. Due to the intractability of (P0), it is necessary to transform the problem it into some tractable sub-problems that can be solved separately and alternatively over multiple iterations.

**IV. PROPOSED UNIFIED SOLUTION**

In this section, we provide an efficient and low complex sub-optimal solution framework to solve the MINLP problem (P0) in (7). We first perform user clustering to effectively mitigate INI and then decouple the transformed problem into sub-problems of decoding order selection, power allocation and
Algorithm 1 Dynamic user clustering for RSMA system

Input: $x_k^u, \forall k \in \mathcal{K}$, Initialize: $\text{MSD}^{(0)} = 0, t = 1$

1: while $\text{MSD}^{(t)} - \text{MSD}^{(t-1)} < \epsilon_t$ do
2: Randomly select $t$ cluster centers $c_i$ and set empty cluster sets $\varphi_i^{(0)} = \emptyset$, $i \in \{1, \ldots, t\}$
3: Calculate distance of the $k^{th}$ user from each cluster center, $d_{ki} = \|x_k^u - c_i\|$
4: Assign users to the nearest cluster centre, $i = \min_i \{d_{ki}\}$ and $\varphi_i = \varphi_i \cup \{k\}$
5: Recalculate the cluster center by finding centroid of the users assigned to that cluster
6: Repeat step 3-6 until convergence
7: $\text{MSD}^{(t)} = \frac{1}{t} \sum_{i=1}^{t} \frac{1}{|\varphi_i|} \sum_{k \in \varphi_i} ||c_i - x_k^u||^2$
8: Check cluster size such that $|\varphi_i| \leq T_s$, otherwise repeat 2-8.
9: $t = t + 1$
10: end while

Output: $\varphi_i, \forall \in \{1, \ldots, t\} = \mathcal{T}$

beamforming design which are solved using general convex approximation techniques. Specifically, for a given decoding order and power allocation, we design active and passive beamforming matrices using alternating optimization by Riemannian manifold algorithm and sequential fractional programming. Then, for fixed decoding order and beamforming matrices, we solve power allocation problem using the difference of two convex functions method and iterative successive convex approximation. The convergence of the proposed unified solution is clarified.

A. User Clustering

Under given decoding order, power allocation and beamforming design, the problem of user clustering can be formulated as

$$\begin{align*}
\text{(P1)} : \max_{C, S_t} \sum_{t \in \mathcal{T}} \sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{J}} C_{kt} \tau_t \log_2 (1 + \gamma_{kj}) \\
\text{s.t. (C6)}. 
\end{align*}$$

(8)

Even after decoupling the original problem into sub-problems, the user-clustering problem in (P1) is still non-convex due to integer programming constraint (C6) and thus hard to solve.

To resolve this problem, we adopt low-complex k-means clustering algorithm to divide the multiple users dynamically into the clusters. The proposed user clustering scheme is illustrated in Algorithm 1. Particularly, we exhibit k-means algorithm for varying (increasing) value of $T$. For each value of $T$, the mean square distance (MSD) of UNs from their respective sub-clusters centroid is calculated and the least value of $T$ is selected. Defining $\epsilon_t$ as the convergence parameter of MSD. Finally, $T$ sets of clusters i.e., $\varphi_t, t \in \mathcal{T} = \{1, \ldots, T\}$ are obtained. The proposed clustering provide cluster sets which have users associated with it. Algorithm 1 describes the proposed user clustering in detail. Although, the proposed
k-means algorithm does not ensure channel disparity (near-far strategy) among users, however, it primarily focuses on maintaining strong effective channel gain for each user with a low-complex user pairing scheme [63]. Moreover, it also relaxes the complicated user-pairing scheme design. Further, the convergence characteristics of the proposed user-clustering scheme in Algorithm 1 is complemented by the discussion of the following proposition.

**Proposition 3:** The proposed user clustering based on k-means in Algorithm 1 always converges in a finite number of iterations but not necessarily to a global optimal solution.

**Proof:** An equivalent convex optimization problem can be constructed for the considered k-means clustering method [64] whose sub-optimal solution can be obtained using Karush-Kuhn-Tucker (KKT) conditions. Algorithm 1 converges to a sub-optimal solution of problem (P1) in a finite number of iterations [64, see Theorem 5].

**B. Active and Passive Beamforming Design**

Under given clustering, power allocation, and decoding order, the joint problem of active and passive beamforming can be formulated as

\[
(P2) : \max_{\theta, z} f_{\theta, z} = \sum_{t \in T} \sum_{k \in \varphi_t} \sum_{j \in J} \tau_t \log_2 \left( 1 + \gamma_{kj}(\theta, z) \right)
\]

\[
\text{s.t. (C4) (C5)}.
\]

The problem (P2) is non-convex in nature due to the non-concave objective function and constant modulus constraints (C4) and (C5) under a continuous phase-shift. In order to solve this problem, we propose two alternate optimization methods based on manifold optimization and sequential fractional programming which are discussed as below:

1) **Alternate Riemannian Manifold Optimization:** The constraints (C4) and (C5) can be viewed as two different Riemannian manifolds [4], \( \mathcal{M}_\theta \) and \( \mathcal{M}_z \) such that

\[
\mathcal{M}_\theta = \left\{ \theta \in \mathbb{C}^{NM \times 1} : |\theta_{mn}| = 1, m \in \mathcal{M}, n \in \mathcal{N} \right\},
\]

\[
\mathcal{M}_z = \left\{ z \in \mathbb{C}^{N_r \times 1} : |z_{\vec{n}}| = 1/\sqrt{N_r}, \vec{n} \in \mathcal{N}_r \right\},
\]

where \( h_k(\theta, z) = \left( h^H_{b_k} + \theta^H \sum_{m=1}^{M} W_m \text{diag} \left( h^H_{r_m} \right) H_{p_m} \right) z, \theta = [\theta_{11}, \ldots, \theta_{M1}, \ldots, \theta_{MN}]^T, W_i = e_i \otimes I_N \in \mathbb{R}^{N \times NM}, i \in \mathcal{M}, NM \) is the arithmetic product of \( N \) and \( M \). and \( e_i \in \mathbb{R}^M \) is a vector with all elements zero but only \( i^{th} \) element as 1.

Denote \( T \mathcal{M}_\theta \) and \( T \mathcal{M}_z \) as the tangent space, i.e., the sets of all the tangent vectors which are tangent to the manifolds \( \mathcal{M}_\theta \) and \( \mathcal{M}_z \) at the point \( \theta \) and \( z \), respectively. Among all tangent vectors, the one yields the fastest increase of the objective function which is defined as the Riemannian gradient. The Riemannian
gradient $g_\theta$ of the function $f_{z,\theta}$ at point $\theta$ is calculated from orthogonal projection of the Euclidean gradient of $f_{z,\theta}$ w.r.t. $\theta$, i.e., $\Delta_\theta f_{z,\theta}$ onto tangent space $\mathcal{T}M_\theta$ which is defined as

$$g_\theta = P_\theta (\Delta_\theta f_{z,\theta}) = \Delta_\theta f_{z,\theta} - \text{Re} \left[ \Delta_\theta f_{z,\theta} \odot \theta \right], \quad (12)$$

where $P$ is the projection function defined as

$$P_\theta (X) = X - \text{Re} (X \odot x) \odot x \quad (13)$$

for any $X$ and $x$. Similarly, for $z$, Riemannian gradient $g_z$ can be determined as $g_z = P_z (\Delta_z f_{z,\theta}) = \Delta_z f_{z,\theta} - \text{Re} \left[ \Delta_z f_{z,\theta} \odot \theta \right] \odot z$.

Let $\bar{h}_{bk} = h_{bk}^H z$, $\bar{h}_{rk} = \sum_m W_m^T \text{diag} (h_{rkm}) H_{bm} z$, $\bar{h}_k = \left( h_{bk}^H + \theta^H \sum_{m=1}^M W_m^T \text{diag} (h_{rkm}) H_{bm} \right)$. The Euclidean gradient $\Delta_\theta f_{z,\theta} = \frac{\partial f_{z,\theta}}{\partial \theta^*}$ is calculated from orthogonal projection of the Euclidean gradient $\nabla \theta f_{z,\theta} = \sum_{j=1}^J p_{ikj}^T \text{diag} (\bar{h}_{k}) \bar{h}_{k}$. The update conjugate direction is used to search a maximum of the function w.r.t. $\theta$ and $z$.

After obtaining the Riemannian gradient $g_\theta$ and $g_z$, we exploit the alternate optimization approach based on conjugate gradient to tackle the manifold optimization problem. Particularly, we employ alternate conjugate gradient algorithm which alternatively optimize $\theta$ and $z$ to obtain local maximizer for the objective function $f_{z,\theta}$. The conjugate gradient method finds the local optimal point for $\theta$ and $z$ on the respective manifold iteratively. In order to restrict the search on the manifold, we apply the concept of the retraction obtained by projecting the Euclidean step onto the manifold.

$$\theta^{(a+1)} = \text{Ret}_\theta \left( \alpha_\theta d_\theta^{(a)} \right) \quad \text{and} \quad z^{(a+1)} = \text{Ret}_z \left( \alpha_z d_z^{(a)} \right) \quad (14)$$

where $\alpha_\theta$ and $\alpha_z$ are the search step sizes and $d_\theta^{(a)} = [d_{\theta_1}^{(a)}, \ldots, d_{\theta_{\theta_{MN}}}^{(a)}]^T$ and $d_z^{(a)} = [d_{z_1}^{(a)}, \ldots, d_{z_{z_{Nz}}}^{(a)}]^T$ are the search directions for $\mathcal{M}_\theta$ and $\mathcal{M}_z$ at $(a)^{th}$ iteration, respectively. The step sizes $\alpha_\theta$ and $\alpha_z$ can be chosen by a line-search algorithm using the Armijo backtracking line search as in [65 Definition 4.2.2].

The update conjugate direction is used to search a maximum of the function w.r.t. $\theta$ and $z$ and the update rule of the search direction for the $(a+1)^{th}$ iteration can be given by

$$d_\theta^{(a+1)} = g_\theta^{(a+1)} + \lambda_\theta^{(a+1)} P_{\theta^{(a+1)}} (d_\theta^{(a)}) \quad (16)$$
Algorithm 2: Alternate Riemannian Conjugate Gradient (RCG) based Manifold Optimization for Active and Passive Beamforming

**Input:** Set $\theta^{(0)}$ and $z^{(0)}$ randomly satisfying (C2) and (C3)

1: **Initialize:** $d^{(0)}_{\theta} = P_{\theta^{(0)}}(\Delta_{\theta} f_{z,\theta})|_{z=z^{(0)}}$, $f_{z,\theta}(z^{(0)}, \theta^{(-1)}) = 0$ and $a = 0$
2: **while** $(f_{z,\theta}(z^{(a)}, \theta) - f_{z,\theta}(z^{(a-1)}, \theta)) / f_{z,\theta}(z^{(a-1)}, \theta) \leq \epsilon_{\theta}$ **do**
3: Choose Armijo backtracking line search step size $\alpha^{(a)}_{\theta}$ and update $\theta^{(a+1)}$ using (14)
4: Compute Riemannian gradient $g^{(a+1)}_{\theta} = P_{\theta^{(a)}}(\Delta_{\theta} f_{z,\theta})$ using (12) ; Calculate Polak-Ribiere parameter $\lambda^{(a+1)}_{\theta}$ as (18) ; Determine conjugate direction $d^{(a+1)}_{\theta}$ as (16)
5: $a=a+1$
6: **end while**
7: **Output:** $\hat{\theta} = \theta^{(a)}$

1: **Initialize:** $d^{(0)}_{z} = \bar{P}_{z^{(0)}}(\Delta_{z} f_{z,\theta})|_{\theta=\hat{\theta}}, z^{(a-1)} = z^{(1)}$
2: **while** $(f_{z,\theta}(z^{(a)}, \hat{\theta}) - f_{z,\theta}(z^{(a-1)}, \hat{\theta})) / f_{z,\theta}(z^{(a-1)}, \hat{\theta}) \leq \epsilon_{\theta}$ **do**
3: Choose Armijo backtracking line search step size $\alpha^{(a)}_{z}$ and update $z^{(a+1)}$ using (15)
4: Compute Riemannian gradient $g^{(a+1)}_{z} = P_{z^{(a)}}(\Delta_{z} f_{z,\theta})$ using (13) ; Calculate Polak-Ribiere parameter $\lambda^{(a+1)}_{z}$ as (19) ; Determine conjugate direction $d^{(a+1)}_{z}$ as (17)
5: $a=a+1$
6: **end while**
7: **Output:** $\bar{z} = z^{(a)}$

$$d^{(a+1)}_{z} = g^{(a+1)}_{z} + \lambda^{(a+1)}_{z} P_{z^{(a+1)}}(d^{(a)}_{z}),$$

(17)

where $\lambda^{(a+1)}_{\theta}$ and $\lambda^{(a+1)}_{z}$ are the Polak-Ribiere parameters which are calculated as

$$\lambda^{(a+1)}_{\theta} = g^{(a+1)H}_{\theta} g^{(a+1)}_{\theta} - P_{\theta^{(a+1)}}(g^{(a)}_{\theta}) / \left(\|P_{\theta^{(a+1)}}(g^{(a)}_{\theta})\|_{F}^{2}\right),$$

(18)

$$\lambda^{(a+1)}_{z} = g^{(a+1)H}_{z} g^{(a+1)}_{z} - P_{z^{(a+1)}}(g^{(a)}_{z}) / \left(\|P_{z^{(a+1)}}(g^{(a)}_{z})\|_{F}^{2}\right),$$

(19)

and $g^{(a+1)}_{\theta}$ and $g^{(a+1)}_{z}$ denote the Riemannian gradients for $\theta$ and $z$, respectively.

The update points for the $\theta$ and $z$ are given by using retraction operations at each iteration. The alternating maximization algorithm for two different manifold optimization is summarized in Algorithm 2

**Proposition 4:** The proposed beamforming design based on two different Riemannian manifold optimization discussed in Algorithm 2 converges to a stationary point

**Proof:** See Theorem 4.3.1 in [65]

2) **Sequential Fractional Programming (SFP):** Besides RCG method, the problem (P1) can be solved using fractional programming (FP) approach [66]. Conceptually, the FP approach employs two key steps to
solve the optimization problem, i.e., Lagrangian dual transform and quadratic transform. Particularly, the
Lagrangian dual transform tackles the logarithmic function by introducing auxiliary variable as
\[
\max_x \log(1 + x) = \log(1 + \zeta) - \zeta + \frac{(1 + \zeta)x}{1 + x},
\] 
(20)
While, the quadratic transform converts the sum of ratios problem \(\sum_i \frac{|A_i(x)|^2}{B_i(x)}\) as
\[
\max_{x, \beta} \sum_i \left(2 \Re\{\beta_i A_i(x)\} - |\beta_i|^2 B_i(x)\right).
\] 
(21)
Using FP approach, the objective function in the problem (P2), i.e., \(f_{z, \theta}\) can be approximately written as
\[
\tilde{f}_{z, \theta}(\zeta, \beta, z, \theta) = \sum_{t=1}^{T} \sum_{k \in \varphi_t} \sum_{j=1}^{J} \tau_t (\log_2 (1 + \zeta_{kj}) - \zeta_{kj}) +
\sum_{t=1}^{T} \sum_{k \in \varphi_t} \sum_{j=1}^{J} 2\tau_t (1 + \zeta_{kj}) p_{kj} \Re\{\beta_{kj} h_k(\theta, z)\} - |\beta_{kj}|^2 \tau_t \sum_{(u,v) \in \tilde{Q}_{kj}^+} |h_u(\theta, z)|^2 p_{uv} + \sigma^2),
\] 
(22)
where \(\zeta = [\ldots, \zeta_{kj} \ldots]^T\) and \(\beta = [\ldots, \beta_{kj} \ldots]^T\).

Now, the problem (P2) can be written as
\[
\tilde{\text{(P2)}}: \max_{\theta, z} \tilde{f}_{z, \theta}(\zeta, \beta, z, \theta)|_{\pi = \tilde{\pi}, p = \tilde{p}}
\text{ s.t. (C2) (C3)}.
\] 
(23)
To solve (23), we employ iterative approach which solves and updates \(\zeta, \beta, z\) and \(\theta\) sequentially. Firstly, we obtain \(\zeta\) and \(\beta\) at the \((a + 1)^{th}\) iteration as [66]
\[
\zeta_{kj}^{(a+1)} = \frac{\psi_{kj}^{(a)} + \psi_{kj}^{(a)} \sqrt{2 \psi_{kj}^{(a)}} + 4}{2}, \beta_{kj}^{(a+1)} = \sqrt{\tau_t (1 + \zeta_{kj}^{(a)}) p_{kj} h_k\left(\theta^{(a)}, z^{(a)}\right)} \sum_{(u,v) \in \tilde{Q}_{kj}^+} |h_u(\theta^{(a)}, z^{(a)})|^2 p_{uv} + \sigma^2, \]
(24)
where \(\psi_{kj}^{(a)} = \frac{1}{\sqrt{\tau_t}} \Re\{\beta_{kj}^{(a)} h_k\left(\theta^{(a)}, z^{(a)}\right)\} p_{kj}\).

**Update \(\theta\):** After dropping irrelevant constant terms, the approximated maximization problem \(\tilde{\text{(P1)}}\) under the optimal \(\theta\) at the \((a + 1)^{th}\) iteration of FP approach can be simplified as
\[
\phi^{(a+1)} = \min_{\phi} f_\theta(\phi) \triangleq \exp(j \phi)^H U_1^{(a)} \exp(j \phi) - 2 \Re\{\exp(j \phi)^H v_1^{(a)}\},
\] 
(25)
where \(\phi = [\phi_1, \ldots, \phi_{MN}]^T\) and \(U_1^{(a)}\) and \(v_1^{(a)}\) are defined as
\[
U_1^{(a)} = \sum_{t=1}^{T} \sum_{k \in \varphi_t} \frac{1}{\sqrt{\tau_t}} \beta_{kj}^{(a)} \sum_{(u,v) \in \tilde{Q}_{kj}^+} p_{uv} h_{ra}^H h_{ra},
\] 
(26)
\[
v_1^{(a)} = \sum_{t=1}^{T} \sum_{k \in \varphi_t} \sum_{j=1}^{J} \sqrt{\tau_t (1 + \zeta_{kj}^{(a)}) \beta_{kj}^{(a)} h_{ra}} - |\beta_{kj}^{(a)}|^2 \sum_{(u,v) \in \tilde{Q}_{kj}^+} p_{uv} h_{ra} h_{ra}^H.
\] 
(27)
However, the problem in (25) is intractable as the objective function $f_\theta (\phi)$ is non-convex. Importantly, $f_\theta (\phi)$ can be equivalently written into its convex surrogate function at the $(a + 1)^{th}$ iteration as

$$f_\theta (\phi^{(a+1)}) \approx f_\theta (\phi^{(a)}) + \Delta_f f_\theta (\phi^{(a)})^T (\phi^{(a+1)} - \phi^{(a)}) + 0.5 (\kappa_\phi \|\phi^{(a+1)} - \phi^{(a)}\|^2),$$

(28)

where $\Delta_f f_\theta (\phi^{(a)}) = 2 \Re \{-j \theta^{(a)} \odot (U_1^{(a)} \theta^{(a)} - v_1^{(a)})\}$ and $\kappa_\phi$ is step size which follows

$$f_\theta (\phi^{(a)}) - f_\theta (\phi^{(a+1)}) \geq \nu \kappa_\phi \|\Delta_f f_\theta (\phi^{(a)})\|^2.$$  

(29)

Note that $\kappa_\phi$ can be determined by using Armijo rule with $0 < \nu < 0.5$. So, the problem in (25) can be solved using SCA technique such that

$$\phi^{(a+1)} = \phi^{(a)} - \Delta_f f_\theta (\phi^{(a)}) \bigg|_{\kappa_\phi},$$

(30)

which ultimately converges to a stationary solution.

**Update $z$:** Similar to $\theta$, $z = \frac{1}{\sqrt{N_r}} \exp (j \omega)$ can be updated as

$$\omega^{(a+1)} = \omega^{(a)} - \Delta_\omega f_z (\omega^{(a)}) \bigg|_{\kappa_\omega},$$

(31)

where $\omega = [\exp(j \phi_{b_1}), \ldots, \exp(j \phi_{b_{N_r}})]^T$ is the phase shift vector at the BS.

$$f_z (\omega) \triangleq \exp (j \omega)^H U_2^{(a)} \exp (j \omega) - 2 \Re \{\exp (j \omega)^H v_2^{(a)}\}, \Delta_\omega f_z (\omega^{(a)}) = 2 \Re \{-j z^{(a)} \odot (U_2^{(a)} z^{(a)} - v_2^{(a)})\},

U_2^{(a+1)} = \sum_{t=1}^{T} \sum_{k \in \mathcal{K}} \sum_{j=1}^{J} \tau_t (\beta_{k,j}^{(a+1)})^2 \sum_{(u,v) \in \mathcal{Q}_k^t} \tilde{p}_{uv}^z \tilde{h}_k \tilde{h}_k^H, v_2^{(a+1)} = \sum_{t=1}^{T} \sum_{k \in \mathcal{K}} \sum_{j=1}^{J} \sqrt{\tau_t (1 + \epsilon_{k,j}^{(a+1)})} \beta_{k,j}^{(a+1)} \tilde{h}_k.

Note that $\kappa_\omega$ is set using Armijo rule satisfying the following condition:

$$f_z (\omega^{(a)}) - f_z (\omega^{(a+1)}) \geq \nu \kappa_\omega \|\Delta_\omega f_z (\omega^{(a-1)})\|^2.$$  

(32)

**Proposition 5:** The proposed beamforming design based on SFP optimization discussed in Algorithm 3 converges to a stationary point [66].

**Proof:** Following the derivation in [67], see Appendix A], it can be shown that the sum rate objective function is monotonically non-decreasing after each iteration in Algorithm 3. Since the optimization problem in (P2) is bounded above, the algorithm must converge. At the convergence, the algorithm arrives at a local optimum of the reformulated problem of (P2).

**Remark 4:** Discrete Phase-Shift Model at RIS: In previous discussion, we consider continuous phase-shift setting (infinite phase-shift resolution) at each RIS. However, the phase-shift may not account a continuous phase-shift model in practical scenario. To satisfy the practical aspect of setting phase-shifters at RIS, we impose finite resolution constraint for each RIS elements such that the phase-shifter can take value from the set $P = \left\{ \frac{2 \pi a}{2b}, a \in \{0, 1, \ldots, 2^b - 1\} \right\}$ where $b$ denotes the resolution of the phase-shifter. The phase-shifters solution obtained from the proposed beamforming design are then mapped to the closest discrete value from $P$.  


The problem in \((C7)\) can be written as difference of convex functions (or approximate convex functions) as:

\[ f(z^0, \theta) - f(z^{a-1}, \theta^{a-1}) \leq \epsilon_\theta \]

sum-rate maximization in (7) can be recast as

\[ \max \sum_{t \in T} \sum_{k \in \varphi_t} \sum_{j \in J} \tau_t \log_2 (1 + \gamma_{kj}(p)) \]

s.t. (C1).

The problem in (P3) is non-convex in nature and thus, we utilize an epigraph method by introducing an auxiliary variable \(r \triangleq \left[ r_{k,j}^1, \ldots, r_{k,j}^T \right]^T \) such that:

\[ \max_{p, r} \sum_{t=1}^T \sum_{k \in \varphi_t} \sum_{j=1}^J r_{k,j}^t \]

s.t. (C1)

\[ (C7): \tau_t \log_2 (1 + \gamma_{kj,t}) \geq r_{kj,t}', \forall t \in T, \forall k \in \varphi_t, \forall j \in J, \]

Note that the problem \((P3)\) in (35) is still non-convex due to the constraint (C7). Particularly, the left side of (C7) can be written as difference of convex function (or approximate convex functions) as:

\[ \log_2 \left( \sum_{(u,v) \in \mathcal{Q}_{kj}^t} |h_{u} (\theta, z)|^2 p_{uv} + \sigma^2 \right) - \xi_{kj,t} \]

where \(\xi_{kj,t} = \log_2 \left( \sum_{(u,v) \in \mathcal{Q}_{kj}^t} |h_{u} (\theta, z)|^2 p_{uv} + \sigma^2 \right) \). Now, we adopt SCA method in order to convert the non-convexity of (C7) into a convex one by iteratively approximates \(\xi_{kj,t}\) into its equivalent convex form using given power allocation from the previous iteration.

Now, \(\xi_{kj,t}\) can be approximated at the \((b + 1)th\) iteration using first-order Taylor’s series expansion as:

\[ \xi_{kj,t} \approx \xi_{kj,t}^{(b+1)} = \xi_{kj,t}^{(b)} + |h_k (\theta, z)|^2 (p_{kj,t} - p_{kj,t}) \left( \log_2 (2^{\xi_{kj,t}^{(b)}}) \right) \]

and (C7) can be approximately written as:

\[ (C7)^{\prime}: \tau_t \log_2 \left( \sum_{(u,v) \in \mathcal{Q}_{kj}^t} |h_{u} (\theta, z)|^2 p_{uv} + \sigma^2 \right) - \tau_t \xi_{kj,t}^{(b+1)} \geq r_{kj,t}'. \]
Algorithm 4 Successive Convex Approximation for Power allocation

1: Initialize: $\theta = \bar{\theta}$, $z = \bar{z}$, $\pi = \bar{\pi}$, $p_{k,j}^{(0)} = \frac{P_{\text{max}}}{J}$, $\forall j \in J$, $k \in K$, $R^{(-1)} = 0$ and $b = 0$
2: while $\frac{R^{(b)} - R^{(b-1)}}{R^{(b-1)}} \leq \epsilon_p$ or $b \leq B_{\text{max}}$ do
3: Solve (38) and calculate the sum rate $R^{(b)} = \sum_k r^{(b)}_{k,j}$
4: $b = b + 1$
5: end while
6: Output: $\hat{R} = R^{(b)}$, $\hat{p} = p^{(b)}$

Ultimately, the problem in (35) can be written into an approximate convex form as

$$p^{(b+1)} = \max_{p,r} \sum_{t=1}^{T} \sum_{k \in \Phi_t} \sum_{j=1}^{J} r^{(b)}_{k,j}$$

s.t. (C1), (C7).

With given initial power allocation values $\{p_{k,j}^{(0)}\}$, the problem (38) is iteratively solved for maximum $B_{\text{max}}$ iterations or until convergence. Let $\epsilon_p$ denotes the convergence factor. The proposed SCA method for achieving the optimal power allocation under given $\theta = \bar{\theta}$, $z = \bar{z}$ and $\pi = \bar{\pi}$ is summarized in Algorithm 4.

**Proposition 6:** Given an initial feasible solution, the proposed SCA based power allocation method in Algorithm 4 converges to a stationary point satisfying the KKT conditions of original power allocation problem (P3).

**Proof:** Let $p^{(b)}_{k,j,t}$ be the power allocation of the $j^{th}$ sub-message of $k^{th}$ user which is obtained after the execution of $b^{th}$ SCA iteration which are utilized in next i.e., $(b + 1)^{th}$ iteration. It can be shown that $\hat{r}^{(b)}_{k,j,t} \leq \hat{r}^{(b+1)}_{k,j,t}$ and $\hat{r}^{(b+1)}_{k,j,t} \geq \hat{r}^{(b)}_{k,j,t}$. This implies that the total sum-rate function $R_k$ monotonically increases as $a$ increases. Due to the power constraints, the $R_k$ is bounded proposed and hence the Algorithm 4 converges.

D. Decoding order and Unified Solution

Algorithm 5 summarizes the proposed unified solution. In particular, the sub-problems of power allocation and beamforming design are solved alternatively in an iterative manner for all possible decoding order. With given decoding order and initial feasible power allocation, we first design beamforming matrices either using Algorithm 2 or Algorithm 3. Later, the obtained beamforming matrices is utilized for the power allocation in Algorithm 4. In next iteration, the obtained power value in the previous iteration is utilize for beamforming design and this process continues until a convergence ($\epsilon_c$) is achieved. The output of each optimization process in current iteration serves an input to the next iteration. Importantly, the optimal decoding order in UL RSMA ensures high user fairness in its capacity region [17]. Particularly, we consider Jain’s fairness index as performance indicator for user fairness for all possible decoding order. The user fairness i.e., Jain’s fairness index $J_i^{th}$ for $\pi_i$ decoding order is given as

$$J_i = \left( \sum_{k \in K} \hat{p}^{(i)}_{k} \right)^2 \left( K \sum_{k \in K} \hat{p}^{(i)}_{k} \right)^2 , i \in S = \{1, 2, \ldots, (KJ)\}$$

(39)
Algorithm 5 Block coordinate descent (BCD) for Unified Solution

1: **Input**: $p = \bar{p}, \theta = \bar{\theta}, z = \bar{z}$ and $C_{\text{max}}$

2: **Initialize**: $p^{(0)} = \bar{p}, \theta^{(0)} = \bar{\theta}, z^{(0)} = \bar{z}, c = 0, J = 0, \hat{R}_f = 0$ and $R^{(-1)} = 0$

3: Perform user clustering using Algorithm 1

4: for $\pi = \pi_1 : \pi_{S_f}$ do

5: \hspace{1em} while $\left(R^{(c)} - R^{(c-1)} \right) \leq \epsilon_c$ or $c \leq C_{\text{max}}$ do

6: \hspace{2em} Using $p^{(c)}$, design beamforming matrices using Algorithm 2 or Algorithm 3

7: \hspace{2em} Using $\theta^{(c)}$ and $z^{(c)}$, allocate power using Algorithm 4 and calculate overall rate-throughput $\hat{R}_i^{(c)}$ and rate-throughput for each user $\hat{r}_k^{(c)}$

8: \hspace{1em} Update $c = c + 1$

9: \hspace{1em} end while

10: Calculate Jain fairness index $J_i$ using $\hat{r}_k^{(c)}$ in (39)

11: if $J_i > J$ then $J = J_i$ and $\hat{R}_f = \hat{R}_i^c$

12: end if

13: end for

where $\hat{r}_k^{(i)}$ is the rate of $k^{th}$ for $i^{th}$ decoding order. Note that the rate $\hat{r}_k^{(i)}$ is rate obtained after executing unified solution of power allocation and beamforming design. The decoding order which provide high user fairness $J$ is selected.

Proposition 7: The proposed unified solution in Algorithm 5 for resource allocation design involving user clustering, power allocation and receive and passive beamforming design at BS and RIS, respectively is guaranteed to converge to local optimal stationary point.

Proof: Since the feasible solution set of (P0) is compact and its objective value is non-decreasing over iterations via solving the sub-problems in (P1), (P2) and (P3), iteratively, the solution of the proposed unified solution in Algorithm 5 is guaranteed to converge [69]. Since we adopted k-means, SCA and SFP techniques to solve the unified resource allocation design using alternating optimization algorithm, the obtained solution may be sub-optimal [69]–[71] for the original problem (P0).

The optimality and convergence of the proposed unified method is further complemented in simulation section through its performance comparison with optimal brute-force search algorithm.

E. Computational Complexity

Now, we describe the overall computational complexity analysis of the proposed unified solution in detail. The SCA based power allocation problem in (38) has $2L$ variables and $4L + K$ constraints where $S = KJ$. With consideration of worst case scenario, let the SCA algorithm converges at maximum iterations $B_{\text{max}}$. So, the worst case computational complexity of SCA based power allocation algorithm can be given as $O \left( B_{\text{max}} \cdot \left( (KJ + K)^2 (4KJ + K) \right) \right)$. Further, the RCG and SFP algorithms for beamforming design in
Algorithm 2 and Algorithm 3 entail worst-case complexity of \( O \left( A_{\theta_1}^{\max} \left( K^2 J^2 N^2 M^2 \right) + A_{\tau_1}^{\max} \left( K^2 J^2 N^2 \right) \right) \) and \( O \left( A_2^{\max} \left( 2KJNMN_r + K^2 J^2 N^2 M^2 + K^2 J^2 N^2 \right) \right) \), where \( A_{\theta_1}^{\max} \) (for \( \theta \)), \( A_{\tau_1}^{\max} \) (for \( \tau \)) and \( A_2^{\max} \) are their maximum iterations for convergence, respectively [66]. Note that SFP based beamforming design has slightly higher computational complexity than RCG, however, it is shown in the next section that the proposed solution with SFP based beamforming design provides better system performance as compared to the RCG algorithm.

Let, the unified solution converges in maximum \( C^{\max} \) iterations then the overall complexity of the proposed solution with RCG based and SFP based beamforming can be given as

\[
O \left( S_{\pi} C^{\max} \left\{ B^{\max} (KJ + K)^2 (KJ + K) + A_{\theta_1}^{\max} \left( K^2 J^2 N^2 M^2 \right) + A_{\tau_1}^{\max} \left( K^2 J^2 N^2 \right) \right\} \right),
\]

\[
O \left( S_{\pi} C^{\max} \left\{ B^{\max} (KJ + K)^2 (KJ + K) + A_2^{\max} \left( 2KJNMN_r + K^2 J^2 N^2 M^2 + K^2 J^2 N^2 \right) \right\} \right),
\]

respectively. We observe that the overall complexity mostly depends on the number of users, the sub-messages, number of RISs and number of RIS elements. The another important parameter affecting complexity of proposed solution is the convergence rate of the SCA, RCG (or SFP) and the proposed BCD algorithm. The fast convergence rate of the proposed algorithm is validated shortly in the simulation section.

V. EXTENSION TO NON-DIAGONAL PHASE-MATRICES AT RIS

Here, we provide a tractable asymptotic performance analysis for the considered Multi-RIS aided UL RSMA system under mmWave communication based on stochastic geometry of channels while accounting for the spatial correlation among each RIS element i.e., for non-diagonal phase-matrices at RIS. For a rectangular phase-shift array with \( N = N_vN_h \) elements at each RIS, the spatial correlation at each RIS under isotropic Rayleigh fading can be given as

\[
\mathbf{R}_c = d_h d_r \mathbf{R}, \quad \text{where } \mathbf{R}_{\{n,n'\}} = r_{n,n'} = \text{sinc} \left( 2u_{n,n'}/\lambda \right),
\]

\( u_{n,n'} \) is the distance between \( n^{th} \) and \( (n')^{th} \) elements of the RIS, \( d_h \times d_r \) is the dimension of each RISs’ element and \( N_v \) and \( N_h \) are the number of elements per column and per row, respectively [33].

Let \( h_{bk} \sim \mathcal{CN} \left( 0, \xi_b^2 \mathbf{I}_{N_r} \right) \) and \( h_{rk,m} \sim \mathcal{CN} \left( 0, \xi_{r_k,m}^2 \mathbf{I}_N \right) \) are the complex channel gain associated with \( k^{th} \) user-BS and \( k^{th} \) user-\( m^{th} \) RIS link, respectively where \( \xi_b \) and \( \xi_{r_k,m} \) are the corresponding distance dependent path loss model. We utilize the rank-1 approximation of RIS-BS mmWave where the normalized array response vector associated with the LOS path are considered as the complex Gaussian, \( \mathbf{a}_{r_m} \sim \mathcal{CN} \left( 0, \mathbf{I}_N \right) \), \( \mathbf{a}_{r_m} \sim \mathcal{CN} \left( 0, \mathbf{I}_{N_r} \right) \).

Before discussing further on spatial correlation, we derive the closed-form expression for the overall rate-throughput for the multi-RIS aided UL RSMA as

**Lemma 1:** The closed-form expression for overall rate-throughput of the UL RSMA system irrespective of any decoding order scheme under can be given as

\[
\sum_k \sum_j r_{kj} = \sum_{r \in T} \sum_{k \in \varphi_r} \sum_{j \in \mathcal{J}} \tau_r \log_2 \left( 1 + \gamma_{kj} \right) = \sum_{r \in T} \tau_r \log_2 \left( 1 + \frac{\sum_{k \in \varphi_r} \sum_{j \in \mathcal{J}} |h_k(\Theta, \mathbf{z})|^2 p_{kj}}{\tau_r \sigma^2} \right)
\]
Proof: See Appendix A

Defining $\gamma_t = \frac{\sum_{k \in \mathcal{F}_t} \sum_{j \in J} |h_k(\Theta, z)|^2 p_{kj}}{\tau_t \sigma^2}$ as the effective SINR of the users corresponding to the $i^{th}$ cluster in the given network. It can be shown that the effective SINR for each cluster depend upon the phase-shift setting at each RIS.

Theorem 1: The average effective SINR at the BS for the considered multi-RIS aided UL RSMA can be upper bounded as

$$\gamma_t \leq \frac{\sum_{k \in \mathcal{F}_t} \sum_{j \in J} |h_k(\Theta, z)|^2 p_{kj}}{\tau_t \sigma^2} \leq \sum_{k \in \mathcal{F}_t} \sum_{j \in J} \frac{h_k^2 p_{kj}}{\tau_t \sigma^2}, \text{ where}$$

$$h_k^* = \zeta_{b_k}^2 + \frac{\pi \sqrt{N_t} \zeta_{h_k} \xi_{r_k}}{8} \sum_{m=1}^{M} \rho_m \mathbb{E}[\xi_{br_m}] + \sum_{m=1}^{M} \left( \frac{\rho_m \pi}{4} + \frac{\rho_m (4 - \pi)}{4} \right) \frac{\xi_{r_m}^2}{2},$$

and $\bar{\rho}_m = \sum_{m=1}^{M} \sum_{n=1}^{N} |\Theta_{m,n,n'}|^2$, $\rho_m = \sum_{m=1}^{M} \sum_{n=1}^{N} \Theta_{m,n,n'}$ are the factors associated with spatial correlation among passive elements at the $m^{th}$ RIS.

Proof: See Appendix B

From Theorem 1, we observe that the factors $\bar{\rho}_m$ and $\rho_m$ associated with the phase-matrices at each RIS considerably affects the asymptotic performance of the considered system model. In particular, these factors can be considered as the correlation factors at each RIS. For uncorrelated phase-shift matrices i.e., diagonal phase-matrices, we obtain that $\rho_m = N$ and $\bar{\rho}_m = N^2$.

Remark 5: Primarily, the appropriate selection of phase-shifters for non-diagonal phase-matrices $\Theta_{m,n,n'} = \theta_{m,n,n'}, \forall m, \forall n, \forall n'$ can provide that $\rho_m \geq N$ and $\bar{\rho}_m \geq N^2$. Besides, the inappropriate selection or any random selection of non-diagonal phase-matrices may lead to performance even worse than diagonal phase-shifters. Hence, the consideration of spatial correlation at each RIS can provide better performance when compared with uncorrelated RIS phase-matrices as validated in [31]. The correlated phase-matrices at RISs corresponds to non-diagonal phase-matrices at RISs.

Therefore, we focus on the optimal (or sub-optimal) design of the non-diagonal phase-matrices to ameliorate the rate performance for the considered Multi-layer UL RSMA system. Using (43), we formulate the passive beamforming design problem for the case of non-diagonal phase-matrices at the RIS under given active receive beamforming and power allocation design as

$$(P4): \max_{\{\Theta_m\}} f_{\Theta} = \frac{\sum_{k \in \mathcal{F}_t} \sum_{j \in J} |h_k(\Theta, z)|^2 p_{kj}}{\tau_t \sigma^2} \quad \text{s.t.} \quad (C8): |\Theta_{m[n,n]}| \leq 1.$$  (46)

Proposition 8: The optimization problem (P4) in (46) can be equivalent transformed as

$$\tilde{(P4)}: \max_{\{\Theta_m\}} \tilde{f}_{\Theta} = \frac{\sum_{t \in T} \sum_{k \in \mathcal{F}_t} \sum_{j \in J} \left| \left( h_{b_k}^H + \sum_{m=1}^{M} h_{rm}^H \Theta_{m} H_{bm} \right) z \right|^2 p_{kj}}{(\tau_t \sigma^2)} \quad \text{s.t.} \quad (C8): |\Theta_{m[n,n]}| \leq 1.$$  (47)
Proof: The objective function in the problem (46) are the logarithmic functions of optimization variables which monotonically increases with \( \{ \Theta_m \} \), the corresponding objective function of the form \( \max \{ \Theta_m \} f_\Theta \) an be equivalently written as \( \max \{ \Theta_m \} \tilde{f}_\Theta \).

The passive beamforming problem (P4) is still non-convex due to its non-concave objective function. However, we can construct a surrogate function for the objective function \( \tilde{f}_\Theta \) and solve the optimization problem in (47). At \( a^{th} \) iteration, the objective function \( \tilde{f}_\Theta \) can be upper bounded using its surrogate function \( \tilde{f}_\Theta \) as \( \tilde{f}_\Theta \leq f_\Theta \).

\[
\tilde{f}_\Theta^{(a)} = \sum_{i \in T} \sum_{k \in \Phi_i} \sum_{j \in J} 2 \Re \left( (h_{bk}^H + \sum_{m=1}^M h_{rm}^H \Theta_{m}^{(a-1)} H_{bm}) z \zeta^H \left( h_{bk}^H + \sum_{m=1}^M h_{rm}^H \Theta_{m}^{(a-1)} H_{bm} \right)^H \right) - \left| \left( h_{bk}^H + \sum_{m=1}^M h_{rm}^H \Theta_{m}^{(a-1)} H_{bm} \right) z \| \sum_{k \neq j} p_{kj} \right|^2 (48)
\]

where \( \overline{p}_{kj} = p_{kj} / (\tau_1 \sigma^2) \). Now, the optimization problem in (47) can be solved iteratively using SCA as

\[
\{ \Theta_m^{(a+1)} \} = \max \{ \Theta_m^{(a)} \} \tilde{f}_\Theta
\]

s.t. (C8) : \( |\Theta_{m_{[n,a]}}| \leq 1 \).

(49)

Since the objective function is now affine in and the constraint (C8) is convex, the problem in (49) is convex in nature and can be solved until convergence.

Proposition 9: Given an initial feasible point \( \Theta \), the proposed SCA based passive beamforming design for non-diagonal phase-matrices converges to a stationary point.

Proof: It can be shown that \( \tilde{f}_\Theta^{(a)} \leq \tilde{f}_\Theta^{(a+1)} \). This implies that the total objective function monotonically increases as \( a \) increases. Due to the beamforming constraints, the objective function \( \tilde{f}_\Theta \) is bounded proposed and hence the SCA based passive beamforming design for non-diagonal phase-matrices converges.

VI. NUMERICAL SIMULATIONS AND DISCUSSION

In this section, the performance of the proposed RISs-aided UL RSMA system under mm-Wave channel communication is examined through extensive computer simulations. The simulation results are averaged by 1000 Monte-Carlo simulations. For simulations, \( M = 4 \) RIS and \( K = 20 \) UL users are uniformly deployed on ground plane of \( 300 \times 300 \) m\(^2\). The position of the BS is fixed on ground plane at \([0,0,0]^T\) m, while the RISs are pre-deployed on the buildings at height 30 m. The noise power \( \sigma^2 \) is set as \(-100\) dB. The number of phase-shifters at each RIS and BS are set to \( N = 60 \) and \( N_r = 15 \), respectively. The signal from each user is split into two sub-messages, i.e., \( J = 2 \). The convergence factors \( \epsilon_\Theta, \epsilon_p \) and \( \epsilon_c \) are set as \( 10^{-4} \) and the maximum number of allowed iterations for SCA and BCD algorithms are set to 10. The UL power budget is set as \( p^{\text{max}} = 100 \) mW. The beamforming matrices are initially selected as random value and the initial transmit power is set as \( p_{kj} = p^{\text{max}} / J, \forall j, \forall k \). The mmWave channel parameters are set as \( L_{bm} = 5, a_0 = 8.4, \gamma_0 = 2.2, a_1 = 40, \gamma_1 = 3.1, \sigma^2_e = 3.6 \).
We refer our proposed unified solution in the Algorithm 5 with random beamforming matrices, RCG and SFP based beamforming design as RIS-RSMA-RND, RIS-RSMA-RCG and RIS-RSMA-SFP, respectively.

For performance benchmark, we solve the equivalent sum-rate maximization problem with RCG and SPF based beamforming for counterpart NOMA and OMA schemes scheme which are referred as as RIS-NOMA and RIS-OMA, respectively. Moreover, we examine the performance of proposed algorithms without user clustering which are referred as RIS-RSMA-RCG-WC and RIS-RSMA-SFP-WC. Further, we also analyze the performance behaviour of the condensed RIS-RSMA system under spatial correlation and discrete phase-shift design at each RIS.

A. Convergence and optimality analysis

Firstly, we examine the convergence and optimality of the proposed unified solution, power allocation and beamforming matrices design algorithms w.r.t. the number of iterations. Fig. 2a validates the Proposition 4 and Proposition 5 of the proposed beamforming design algorithm for RCG based Algorithm 2 and SFP based Algorithm 3. We consider objective function $f_{z, \theta}$ in (P2) to evaluate convergence characteristics. Fig. 2a shows that the RCG algorithm converges quickly as compared to the SFP algorithm, while, the SFP algorithm provide better performance than RCG. In SFP algorithm, the active receive and passive beamforming matrix are updated sequentially in the current iteration, while Algorithm 2 first updates passive beamforming matrix under fixed receive beamforming matrix and later, active receive beamforming is updated under fixed receive beamforming matrix. Importantly, Algorithm 3 solves the beamforming design in block manner and Algorithm 5 execute the beamforming design in sequentially manner. Apparently, Algorithm 3 converges slowly but attains better performance than Algorithm 2. However, the computational complexity of the RCG algorithm is lower than the SFP algorithm as discussed in the Section IV-E.

Next, we examine the convergence analysis of the SCA based power allocation algorithm and the unified solution i.e., Algorithm 4 and Algorithm 5 w.r.t. varying number of iterations as shown in Fig. 2b and Fig. 2c respectively. The proposed power allocation algorithm converges within 5 to 6 iterations which conclude the convergence of the Algorithm 4 as mentioned in Proposition 6 and Proposition 7. Moreover, the unified solution with SFP and RCG based beamforming design, discussed in Algorithm 5, attains fast
Fig. 3: Impact of number of RIS elements

Fig. 4: Impact of number of receiving antennas

Fig. 5: Impact of maximum transmit power

convergence as it converges within 3 to 4 iterations. Simulation results also reveals that the proposed unified solution gain sub-optimal performance when compared to the computationally complex exhaustive search method. Nevertheless, the exhaustive search operations are not practically feasible as its computational complexity grows exponentially over the total number of variables. Hence, the proposed low-complex unified solution based on alternating optimization is desirable to solve the complex resource allocation design for the considered system effectively.

B. Impact of Number of RIS and BS antenna elements

Fig. 3 and Fig. 4 show the performance behaviour of the proposed RSMA system for different number of RIS elements and number of receive antennas, respectively. For both the scenarios, we fix alternate parameter. It is quite obvious that as the number of phase shifter increases, the effective channel gain between the users and the BS is improved and hence the achievable sum-rate increases. In other words, the extra phase shifters can reflect more power of the signal received from the UL users to the BS which leads to increased power gain. Moreover, it provides higher flexibility in resource allocation and stronger beamforming gain which improve the sum-rate throughput. Overall, the optimal phase determination achieves better sum-rate performance as phase-shifters increases when compared to the random phase shift.

C. Impact of Maximum Transmit Power and number of sub-messages

Fig. 5 shows the performance behaviour of the proposed RSMA solution with SFP based beamforming design in contrast with the equivalent NOMA and OMA schemes. It is quite obvious that increasing in maximum transmission power for each user increases the rate-throughput. However, RSMA system achieves better performance as compared to equivalent NOMA and OMA schemes due to the better resource allocation and power distribution to various sub-messages due to two main aspects. the multiple sub-messages in UL RSMA from each user under effective transmit power utilization among sub-messages by maintaining flexible inter-message interference with the optimal SIC and decoding order. However, in UL NOMA system [48], the transmit power is allocated to the single-message of the user such that the INI in the

For exhaustive search, we consider discrete phase-shift design at each RIS with resolution $b = 5$ and power allocation sample step size of 0.001 W for given user clustering and decoding order.
desired UL signal communication is minimum (or optimal) with minimum power utilization. Besides, the proposed user clustering scheme effectively mitigate the IUI by grouping and hence the proposed solution (RIS-RSMA-SFP) gain better performance when compared to without clustering (RIS-RSMA-SFP-WC). Moreover, the proposed RIS aided RSMA system outperforms the system without RIS.

We also analyze the performance behaviour of the proposed solution under consideration of varying decoding order such as ascending/descending order of effective channel gain, i.e., RIS-RSMA-SFP (ASC) and RIS-RSMA-SFP (DSC) and exhaustive search RIS-RSMA-SFP (EXH) with only 4 users and single cluster as shown in Fig. 6. RSMA effectively allocate the transmit power to various sub-messages which achieves maximal achievable rate region i.e., capacity region and it increases with increase in sub-messages as shown in Fig. 6. However, the increase in sub-messages increases the challenges for SIC design at receiver. Besides, the decoding order with descending order of effective channel gain provides better performance (equivalent to exhaustive search) as compared to ascending order as illustrated in Fig. 6. The ascending decoding order of users based on channel gain provides the optimal performance when the user channels are sufficiently aligned and there exit a large disparity in their channel strengths. While, the descending order of users based on channel gain achieves better overall system performance especially when there does not exits disparity among the channel gains. Generally, in RIS aided system, the channels are strong and there does not exist disparity among channel gain of users due to multiple RIS reflectors.

### D. Impact of Number of Users and Cluster Size

We illustrate the effect of the proposed user clustering on the performance behaviour of the proposed RSMA solution in comparison with equivalent NOMA and OMA schemes. Fig. 7 shows the effect of variation on the number of UL users in the network. As it can be seen that with increase in the number of users, the overall sum-rate throughput increases.

Nevertheless, the NOMA with time sharing gain rate capacity region equivalent to RSMA for multiple access Gaussian channel at the cost of increased resource utilization or increased computational complexity. Particularly, the NOMA with time-sharing performs all possible user-pairing schemes and perform NOMA communication in different time slots which gain higher user-fairness than conventional NOMA schemes. In order to provide fair comparison in terms of resource utilization, we provide comparative performance analysis of the proposed RSMA system with respect to NOMA without time-sharing only throughout this paper, at the cost of increased time resource utilization high implementation complexity and time-synchronization issue.
of UL users, the overall sum-rate throughput of the system increases. The power allocation of UL users are constrained individually, moreover, with increase in user-density each user operates considerably with high power to mitigate the IUI. Therefore, the overall rate for UL transmission increases with increase in UL users. Besides, we also compare the performance of the proposed solution without user clustering (RIS-NOMA-SFP-WC). The user clustering plays vital role in mitigating IUI when user density is considerably high as shown in Fig. 7. Importantly, the proposed algorithms with user clustering based on k-means algorithm outperform without clustering case. Note that user clustering based on k-means algorithm is easily scalable to large scale networks.

Fig. 8 depicts the rate performance of the proposed k-means user clustering with respect to the maximum allowed cluster-density $T_s$ i.e., number of users per clusters. We fix the resource bandwidth for given total number of users and vary the maximum number of users allowed in each clusters. Fig. 8 shows that the rate performance first increases with increases in $T_s$ and then decreases beyond particular value ($T_s = 3$). It is due to fact that low value of $\tau_t$ and high value of IUI respectively dominates for low value and high value of $T_s$. The increase in cluster-density increases both IUI as well as the bandwidth resource allocation factor $\tau_t$. Primarily, there always exits a trade-off between IUI and resource allocation factor in hybrid orthogonal and non-orthogonal schemes and the heuristic selection of the cluster-density can captivate better system performance gain [63], [72]. For our considered hybrid RSMA-OMA system, $T_s = 3$ gain better performance when compared to other values. Besides, the proposed k-means based clustering significantly outperforms the near-far user clustering scheme (RIS-RSMA-SFP-NF). It is due to fact that the effective rate-splitting in RSMA with optimal beamforming design at RIS captivate high performance gain with smart interference management even with higher number users and low channel disparity. Moreover, for without clustering, the performance of UL RSMA suffers from a significant loss in the overall rate-throughput. Overall, the proposed user clustering relaxes the sophisticated selection of user-pair for the required high channel-disparity in near-far user-clustering schemes.

E. Impact of QoS requirement and Noise Power

Fig. 9 illustrate the impact of noise power on the performance of both RSMA and NOMA depends. Intuitively, the system performance of both RSMA and NOMA degrades with an increase in noise power.

---

**Fig. 9: Impact of imperfect SIC decoding**

**Fig. 10: Impact of minimum rate requirement**

**Fig. 11: Impact of Spatially Correlated RIS**
However, the performance of RSMA degrades much faster than NOMA with an increase in imperfection. It is due to the fact that the high noise power reduces the SINR for both the sub-messages. This further increases the channel noise power over inter-node interference in RSMA and NOMA schemes. In a high noise environment, the performance of RSMA is approximately closer to the NOMA as the channel noise power dominates the performance as illustrated in Fig. 9.

We also analyze the system performance of the considered RSMA system w.r.t. varying minimum QoS requirement as shown in Fig. 10. For UL RSMA (or NOMA) system, the decoding of any sub-message \( s_{kj} \) (or message \( s_k \) ) depends on the successful SIC of its previously decoded sub-messages (messages) such that the sub-messages which does not meet QoS constraint then that sub-message \( s_{kj} \) and other subsequent sub-messages after \( s_{kj} \) cannot be decoded. In other words, if \( \gamma_{kj} < \gamma_{min} \) then \( \gamma_{uv} = \{0 : \pi_{uv} \geq \pi_{kj}\} \). Particularly, the increases in rate requirement decreases the probability of successful decoding of all the sub-messages which results in rate-performance degradation. However, the performance of RSMA still outperform NOMA due to rate-splitting and effective interference management. Fig. 9 and Fig. 10 validate that the performance of multi-layer RSMA is closer to that of NOMA. In other words, the performance of RSMA is trivial when compared to NOMA at high noise power and for low QoS requirement \( (r_k \leq 0.1 \text{ bps/Hz}) \). Moreover, NOMA inherit low computational cost and implementation cost and hence, it may be preferred over RSMA in such special cases.

F. Impact of Discrete Phase-shift and Spatial Correlated Phase-Matrices at RISs

Finally, we discuss the effect of the spatial correlation at each RIS on the performance behaviour of the considered multi-RIS aided RSMA system. Here, we also consider the impact of discrete phase-shift (Disc. PS) at RIS and Rank-1 approximation in Proposition 1 for mmWave-channel on the system performance and perform comparative analysis with continuous phase-shift (Cont. PS) and conventional mmWave-channel (without Rank-1 approximation ). Fig. 11 shows that the exploitation of spatial-correlation characteristics at RISs i.e. non-diagonal phase-matrices can significantly improves the system performance as compared to diagonal phase-matrices only when the phase-matrices are designed optimally (or sub-optimally). This validates Theorem 1. The extra-phase-shifters corresponding to non-diagonal elements provide extra channel gain-diversity \[31\], and thus boost the overall system performance of the considered RSMA system in terms of rate-metric. Moreover, the rank-one approximation which neglect NLOS paths and consider only LOS path has very little impact. Fig. 11 also validate that the discrete phase-shift model with high-resolution \( (b = 6) \) achieve performance near to the ideal continuous phase-shift case as it considerably reduces the performance loss due to discretization. Importantly, the low-resolution setting of discrete phase-shifters at RIS can lead to worst-performance especially when sub-optimal optimization techniques are utilized. Although the high resolution setting incurs extra hardware overhead and implementation bottleneck in practical systems, the
high resolution for discrete phase-shift are more advantageous over low-resolution model in terms of system performance.

VII. CONCLUSION

In this paper, the problem of sum-rate maximization for an RISs-aided mmWave UL RSMA system was investigated. To effectively reduce IUI among users, we performed dynamic user clustering and later solved the joint optimization problem by decoupling it into different sub-problems under given decoding order. Particularly, the sub-problem of power allocation was solved using successive convex approximation and beamforming design was solved using RCG and SFP algorithms. Later, the unified solution based on BCD algorithm was proposed. Simulation results validated that RSMA system achieved sum-rate throughput gain approximately up to 40% and 80% higher than conventional NOMA and OMA schemes, respectively. The descending order of users based on channel conditions when compared to ascending order attained better rate-throughput due to less channel gain disparity among the users. Moreover, the proposed k-means user clustering approach effectively mitigated the IUI and improved the sum-rate throughput by 50%. Besides, the spatial correlation among each RIS elements i.e., sub-optimal setting of non-diagonal phase-matrices at each RIS achieved better performance as compared to conventional diagonal phase-matrices.

APPENDIX A: PROOF OF LEMMA 1

For $t^{th}$ cluster, the sum-rate throughput can be re-expressed as

$$ r_t = \sum_k \sum_j \tau_t \left( \log_2 \left( \sum_{(u,v)\in \tilde{Q}_t} |h_u(\theta, z)|^2 p_{uv} + \tau_t \sigma^2 \right) - \log_2 \left( \sum_{(u,v)\in \tilde{Q}_t} |h_u(\theta, z)|^2 p_{uv} + \tau_t \sigma^2 \right) \right) $$

$$ = \tau_t \sum_{s=1}^S \left( \log_2 \left( \sum_{s' \geq s} |h_s(\theta, z)|^2 p_s + \tau_t \sigma^2 \right) - \log_2 \left( \sum_{s' > s} |h_s(\theta, z)|^2 p_s + \tau_t \sigma^2 \right) \right) $$

(A.1)

where $S = KJ$, $g_s$ and $p_s$ corresponds to channel gain and the power allocation of $j^{th}$ sub-message of $k^{th}$ user. The expression in (A.1) can be rewritten using telescoping product as

$$ r_t = \tau_t \log_2 \left( 1 + \sum_{s=1}^S |h_s(\theta, z)|^2 p_s / (\tau_t \sigma^2) \right) $$

(A.2)

Putting $(k,j)$ back instead of $s$ proves (43).

APPENDIX B: PROOF OF THEOREM 1

The effective channel gain between BS and $k^{th}$ user can be upper bounded as

$$ |h_k|^2 = \left| h_{bk}^H z + \sum_{m=1}^M h_{rk,m}^H \Theta_m H_{bm} z \right|^2 \leq \overline{h}_k, $$

$$ \overline{h}_k = \left| h_{bk}^H z \right|^2 + 2 \left| h_{bk}^H z \right| \sum_{m=1}^M \left| h_{rk,m}^H \Theta_m H_{bm} z \right| + \left( \sum_{m=1}^M \left| h_{rk,m}^H \Theta_m H_{bm} z \right| \right)^2 $$

(B.1)

Using the rank approximation for $H_m$ in Proposition 1, it can be obtained that

$$ h_{rk,m}^H \Theta_m H_{bm} z = \zeta_{bm} \sum_{n=1}^N \sum_{n'=1}^N \sum_{\pi=1}^{N_r} h_{r_{k,m}[n]} \Theta_{m[n,n']} a_{r_{m[\pi]}} a_{r_{m[\pi]}} $$

(B.2)
Using (B.4), the average maximum achievable effective SINR can be given as

\[
\gamma_i^\text{max} = \mathbb{E} \left[ \sum_{k \in K} \sum_{j \in J} \frac{h_{k}^{\text{max}} p_{kj}}{\sigma^2} \right] = \sum_{k \in K} \sum_{j \in J} p_{kj} \mathbb{E} \left[ h_{k}^{\text{max}} \right] / (\tau_i \sigma^2)
\]  

Using [19] see Appendix A, we now derived following closed form expression of (B.5)

\[
\mathbb{E} \left[ \sum_{m=1}^{M} \zeta_{b_m} \sum_{n=1}^{N} \sum_{n'=1}^{N} |h_{r_k,m|n'}| |\Theta_{m|n,n'}|^2 \right] = \frac{N_r}{4} \sum_{m=1}^{M} \rho_{m} \varsigma_{brm}^2 \mathbb{E} \left[ \varsigma_{brm}^2 \right] + \left( 2 - \frac{\pi}{2} \right) \rho_{m} \varsigma_{brm}^2 \mathbb{E} \left[ \varsigma_{brm}^2 \right] = \sum_{m=1}^{M} \left( \frac{\rho_{m} \pi}{4} + \frac{\rho_{m} (4 - \pi)}{4} \right) \zeta_{brm}^2 \mathbb{E} \left[ \varsigma_{brm}^2 \right]
\]

\[
\mathbb{E} \left[ 2 \left( h_{k}^{\text{max}} \right)^2 \right] = \frac{\pi \sqrt{\pi} N_r^2 \zeta_{k} \zeta_{brk}^2}{8} \sum_{m=1}^{M} \rho_{m} \mathbb{E} \left[ \zeta_{brm} \right]
\]

and \( \mathbb{E} \left[ \frac{1}{N_r} \sum_{n=1}^{N} |h_{k}^{\text{max}}|^2 \right] = \zeta_{k}^2 \), where \( \tilde{\rho}_{m} = \sum_{n=1}^{M} \sum_{n'=1}^{N} |\Theta_{m|n,n'}|^2, \rho_{m} = \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{n'=1}^{N} |\Theta_{m|n,n'}|^2 \). Note that we omit complete proof due to brevity of space. Combining (B.6) and (B.7) gives (44). Hence proved.

[1] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, “Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer,” IEEE J. Sel. Areas Commun., vol. 38, no. 8, pp. 1719–1734, 2020.

[2] Q. Wu and R. Zhang, “Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network,” IEEE Commun. Mag., vol. 58, no. 1, pp. 106–112, 2020.

[3] X. Yang, C.-K. Wen, and S. Jin, “MIMO detection for reconfigurable intelligent surface-assisted millimeter wave systems,” IEEE J. Sel. Areas Commun., vol. 38, no. 8, pp. 1777–1792, 2020.

[4] Y. Xiu, Y. Zhao, Y. Liu, J. Zhao, O. Yagan, and N. Wei, “IRS-assisted millimeter wave communications: Joint power allocation and beamforming design,” in Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW), 2021, pp. 1–6.

[5] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, “Wireless communications through reconfigurable intelligent surfaces,” IEEE Access, vol. 7, pp. 116753–116773, 2019.

[6] G. Zhou, C. Pan, H. Ren, K. Wang, M. Elkashlan, and M. Di Renzo, “Stochastic learning-based robust beamforming design for RIS-aided millimeter-wave systems in the presence of random blockages,” IEEE Trans. Veh. Technol., vol. 70, no. 1, pp. 1057–1061, 2021.

[7] W. Saad, M. Bennis, and M. Chen, “A vision of 6G wireless systems: Applications, trends, technologies, and open research problems,” IEEE Netw., vol. 34, no. 3, pp. 134–142, 2020.
[8] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, “Rate splitting for MIMO wireless networks: a promising PHY-layer strategy for LTE evolution,” *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 98–105, 2016.

[9] H. Joudeh and B. Clerckx, “Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach,” *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, 2016.

[10] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, “A rate splitting strategy for massive MIMO with imperfect CSIT,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4611–4624, 2016.

[11] B. Clerckx, Y. Mao, R. Schober, E. A. Jorswieck, D. J. Love, J. Yuan, L. Hanzo, G. Y. Li, E. G. Larsson, and G. Caire, “Is NOMA efficient in multi-antenna networks? a critical look at next generation multiple access techniques,” *IEEE Open J. Commun. Society*, vol. 2, pp. 1310–1343, 2021.

[12] B. Rimoldi and R. Urbanke, “A rate-splitting approach to the gaussian multiple-access channel,” *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 364–375, 1996.

[13] Y. Mao, B. Clerckx, and V. O. K. Li, “Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: Spectral and energy efficiency analysis,” *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8754–8770, 2019.

[14] Y. Mao, B. Clerckx, and V. O. Li, “Rate-splitting multiple access for downlink communication systems: Bridging, generalizing, and outperforming SDMA and NOMA,” *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, pp. 1–54, 2018.

[15] B. Clerckx, Y. Mao, R. Schober, and H. V. Poor, “Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: A simple two-user rate analysis,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 349–353, 2020.

[16] Z. Yang, M. Chen, W. Saad, W. Xu, and M. Shikh-Bahaei, “Sum-rate maximization of uplink rate splitting multiple access (RSMA) communication,” *IEEE Trans. Mobile Comput.*, pp. 1–1, 2020.

[17] H. Liu, T. A. Tsiftsis, K. J. Kim, K. S. Kwak, and H. V. Poor, “Rate splitting for uplink NOMA with enhanced fairness and outage performance,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4657–4670, 2020.

[18] Y. Cao, T. Lv, Z. Lin, and W. Ni, “Delay-constrained joint power control, user detection and passive beamforming in intelligent reflecting surface-assisted uplink mmwave system,” *IEEE Trans. Cogn. Commun. and Netw.*, vol. 7, no. 2, pp. 482–495, 2021.

[19] 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 Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14 960–14 973, 2020.

[20] J. Zeng, T. Lv, W. Ni, R. P. Liu, N. C. Beaulieu, and Y. J. Guo, “Ensuring max–min fairness of UL SIMO-NOMA: A rate splitting approach,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11 080–11 093, 2019.

[21] W. Mei, B. Zheng, C. You, and R. Zhang, “Intelligent reflecting surface-aided wireless networks: From single-reflection to multirefection design and optimization,” *Proceedings of the IEEE*, vol. Early Access, pp. 1–21, 2022.

[22] L. Yin, Y. Mao, O. Dizdar, and B. Clerckx, “Rate-splitting multiple access for 6G–Part II: Interplay with integrated sensing and communications,” *arXiv preprint arXiv:2205.02462*, 2022.

[23] H. Li, Y. Mao, O. Dizdar, and B. Clerckx, “Rate-Splitting Multiple Access for 6G–Part III: Interplay with Reconfigurable Intelligent Surfaces,” *arXiv preprint arXiv:2205.02036*, 2022.

[24] Y. Cao and T. Lv, “Intelligent reflecting surface enhanced resilient design for MEC offloading over millimeter wave links,” *arXiv preprint arXiv:1912.06361*, 2019.

[25] Y. Cao, T. Lv, and W. Ni, “Intelligent reflecting surface aided multi-user mmWave communications for coverage enhancement,” in *Proc. IEEE 31st Ann. Int. Symp. Pers., Indoor and Mobile Radio Commun.*, 2020, pp. 1–6.

[26] V. Jamali, A. M. Tulino, G. Fischer, R. Müller, and R. Schober, “Intelligent reflecting and transmitting surface aided millimeter wave massive mimo,” *arXiv preprint arXiv:1902.07670*, 2019.

[27] B. Ji, Y. Han, S. Liu, F. Tao, G. Zhang, Z. Fu, and C. Li, “Several key technologies for 6G: Challenges and opportunities,” *IEEE Commun. Std. Mag.*, vol. 5, no. 2, pp. 44–51, 2021.

[28] P.-Q. Huang, Y. Zhou, K. Wang, and B.-C. Wang, “Placement optimization for multi-irs-aided wireless communications: An adaptive differential evolution algorithm,” *IEEE Wireless Communications Letters*, vol. 11, no. 5, pp. 942–946, 2022.

[29] J. Xu, Y. Liu, X. Mu, J. T. Zhou, L. Song, H. V. Poor, and L. Hanzo, “Simultaneously transmitting and reflecting (STAR) intelligent omni-surfaces, their modeling and implementation,” *arXiv preprint arXiv:2108.06233*, 2021.
B. Zheng, C. You, W. Mei, and R. Zhang, “A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications,” *IEEE Commun. Surv. Tuts.*, vol. 24, no. 2, pp. 1035–1071, 2022.

Q. Li, M. El-Hajjar, I. A. Hemadeh, A. Shojaeifard, A. Mourad, B. Clerckx, and L. Hanzo, “Reconfigurable intelligent surfaces relying on non-diagonal phase shift matrices,” *IEEE Trans. Veh. Technol.*, 2022.

T. Van Chien, A. K. Papazafeiropoulos, L. T. Tu, R. Chopra, S. Chatzinotas, and B. Ottersten, “Outage probability analysis of IRS-assisted systems under spatially correlated channels,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1815–1819, 2021.

E. Bjornson and L. Sanguinetti, “Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 4, pp. 830–834, 2021.

Y. Liu, X. Mu, X. Liu, M. Di Renzo, Z. Ding, and R. Schober, “Reconfigurable intelligent surface (RIS) aided multi-user networks: Interplay between NOMA and RIS,” arXiv preprint arXiv:2011.13336, 2020.

Y. Xiu, J. Zhao, W. Sun, M. Di Renzo, G. Gui, Z. Zhang, N. Wei, and N. Wei, “Reconfigurable intelligent surfaces aided mmWave NOMA: Joint power allocation, phase shifts, and hybrid beamforming optimization,” *IEEE Trans. Wireless Commun.*, pp. 1–1, 2021.

B. Zheng, Q. Wu, and R. Zhang, “Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA?” *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 753–757, 2020.

G. Yang, X. Xu, and Y.-C. Liang, “Intelligent reflecting surface assisted non-orthogonal multiple access,” in *Proc. IEEE Wireless Commun., and Netw. Conf. (WCNC)*, 2020, pp. 1–6.

X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, “Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6884–6898, 2020.

J. Zuo, Y. Liu, Z. Qin, and N. Al-Dhahir, “Resource allocation in intelligent reflecting surface assisted NOMA systems,” *IEEE Trans. Commun.*, vol. 68, no. 11, pp. 7170–7183, 2020.

D. Zhang, Q. Wu, M. Cui, G. Zhang, and D. Niyato, “Throughput maximization for IRS-assisted wireless powered hybrid NOMA and TDMA,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 9, pp. 1944–1948, 2021.

X. Xie, F. Fang, and Z. Ding, “Joint optimization of beamforming, phase-shifting and power allocation in a multi-cluster IRS-NOMA network,” *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 7705–7717, 2021.

K. Liu and Z. Zhang, “On the energy-efficiency fairness of reconfigurable intelligent surface-aided cell-free network,” in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC2021-Spring)*, 2021, pp. 1–6.

G. Yang, Y. Liao, Y.-C. Liang, O. Tirkkonen, G. Wang, and X. Zhu, “Reconfigurable intelligent surface empowered device-to-device communication underlaying cellular networks,” *IEEE Trans. Commun.*, pp. 1–1, 2021.

Z. Yang, M. Chen, W. Saad, W. Xu, M. Shikh-Bahaei, H. V. Poor, and S. Cui, “Energy-efficient wireless communications with distributed reconfigurable intelligent surfaces,” *IEEE Trans. Wireless Commun.*, pp. 1–1, 2021.

C.-B. Le, D.-T. Do, X. Li, Y.-F. Huang, H.-C. Chen, and M. Voznak, “Enabling NOMA in backscatter reconfigurable intelligent surfaces-aided systems,” *IEEE Access*, vol. 9, pp. 33782–33795, 2021.

T. Hou, Y. Liu, Z. Song, X. Sun, Y. Chen, and L. Hanzo, “Reconfigurable intelligent surface aided NOMA networks,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2575–2588, 2020.

F. Fang, Y. Xu, Q.-V. Pham, and Z. Ding, “Energy-efficient design of IRS-NOMA networks,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 14088–14092, 2020.

M. Zeng, X. Li, G. Li, W. Hao, and O. A. Dobre, “Sum rate maximization for IRS-assisted uplink NOMA,” *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 234–238, 2021.

Y. Xu, G. Gui, H. Gacanin, and F. Adachi, “A survey on resource allocation for 5g heterogeneous networks: Current research, future trends, and challenges,” *IEEE Commun. Surv. Tuts.*, vol. 23, no. 2, pp. 668–695, 2021.

A. Rahmati, Y. Yapici, N. Rupasinghe, I. Guvenc, H. Dai, and A. Bhuyan, “Energy efficiency of RSMA and NOMA in cellular-connected mmWave UAV networks,” in *Proc. IEEE Int. Conf. Commun. (ICC Workshops)*, 2019, pp. 1–6.

A. Bansal, K. Singh, B. Clerckx, C.-P. Li, and M.-S. Alouini, “Rate-splitting multiple access for intelligent reflecting surface aided multi-user communications,” *IEEE Trans. Veh. Technol.*, pp. 1–1, 2021.
