Joint Optimization of Beamforming, Phase-Shifting and Power Allocation in a Multi-cluster IRS-NOMA Network

Document Version
Accepted author manuscript

Citation for published version (APA):

Published in:
IEEE Transactions on Vehicular Technology

Citing this paper

General rights

Takedown policy

Joint Optimization of Beamforming, Phase-Shifting and Power Allocation in a Multi-cluster IRS-NOMA Network

Ximing Xie, Fang Fang, *Member, IEEE*, and Zhiguo Ding, *Fellow, IEEE*

Abstract—The combination of non-orthogonal multiple access (NOMA) and intelligent reflecting surface (IRS) is an efficient solution to significantly enhance the energy efficiency of the wireless communication system. In this paper, a downlink multi-cluster NOMA network is considered, where each cluster is supported by one IRS. This paper aims to minimize the transmit power by jointly optimizing the beamforming, the power allocation and the phase shift of each IRS. The formulated problem is non-convex and challenging to be solved due to the coupled variables, i.e., the beamforming vector, the power allocation coefficient and the phase shift matrix. To address this non-convex problem, an alternating optimization based algorithm is proposed. Specifically, the primal problem is divided into two subproblems for beamforming optimization and phase shifting feasiblity, where the two subproblems are solved iteratively. Moreover, to guarantee the feasibility of the beamforming optimization problem, an iterative algorithm is proposed to search the feasible initial points. To reduce the complexity, a simplified algorithm based on partial exhaustive search for this system model is also proposed. Simulation results demonstrate that the proposed alternating algorithm can yield a better performance gain than the partial exhaustive search algorithm, NOMA with random IRS phase shift scheme and OMA-IRS scheme.

Index Terms—Intelligent reflective surface (IRS), non-orthogonal multiple access (NOMA), transmit power optimization.

I. INTRODUCTION

The 5G communication system has been commercialized world-widely, and the beyond 5G (B5G) system starts attracting more and more researchers' attention due to its low energy consumption, high spectrum efficiency and massive multi-device interconnections [1]–[3]. In order to satisfy the increasing demand caused by the fast-growing number of users, various techniques, including millimetre wave [4], massive multi-inputs and multi-outputs (MIMO) system [5], and small cell [6], have been investigated and extensively used in practice. As a potential technique of B5G, non-orthogonal multiple access (NOMA) has received widespread attention due to its high spectral efficiency [7], [8]. Different from conventional orthogonal multiply access (OMA), such as frequency division multiple access (FDMA), time division multiple access (TDMA), code division multiple

access (CDMA), and orthogonal frequency-division multiple access (OFDMA), NOMA allows multiple users to share the same time slot, frequency block and channel code, which dramatically increases the spectral efficiency. In particular, the users in a NOMA network usually adopt successive inference cancellation (SIC) to remove the inference from other NOMA users, which can efficiently improve the signal to interference and noise ratio (SINR) and reception reliability [9]. Recently, intelligent reflective surface (IRS) has also been proposed as a potential solution to further improve the performance of wireless networks, including enlarging the communication coverage, and improving transmission robustness. Specifically, the IRS can reflect the electromagnetic wave to extend the cover rage of the base station (BS). It also has the ability to tune the channel by adjusting the phase shift of each element, which will greatly improve the quality of users' received signal [10]. The typical architecture of IRS consists of a reflecting panel and a smart controller. The reflecting panel is composed of many reflecting elements and a control circuit. The control circuit is responsible for tuning the phase shift of each reflecting element. Moreover, the smart controller determines the reflection adaptation and also performs as a gateway to communicate with the BS. The smart controller can receive the control signal from the BS and then adjust the phase shift of each reflecting element [11].

A. Related Works

In literature, extensive research has been carried out for the NOMA technique, which has been combined with various state-of-the-art techniques such as MIMO and orthogonal time-frequency space modulation (OTFS) [12]–[16]. Recently, IRS has emerged as a kind of powerful equipment for wireless communication networks [17]–[19]. Among these works, IRS was proved as a perfect solution for a wireless communication network, where the channel will be intelligently reconfigured by the IRS [15], [20]–[22].

Motivated by the benefits from NOMA and IRS, the combination of NOMA and IRS has been recently proposed as a promising solution to improve the communication systems. There have been some ongoing works studying the combination of NOMA and IRS. Some recent research works such as [23], [24] considered a simple scenario where a single IRS serves two users in a downlink NOMA network. In [23], the authors minimized the transmit power at the BS by optimizing beamforming and IRS phase shift and also considered an improved quasi-degradation condition to guarantee that NOMA

can achieve the capacity region with high possibility. In [24], the authors analysed two kinds of phase shift designs, namely random phase shifting and coherent phase shifting.

Moreover, there are many works considering an IRS-assisted NOMA network where a signal IRS serves multiple users [25]–[29]. The problems which have been researched can be divided into two categories, one is about the transmit power minimization [25], [26] and the other is about the the sum-rate maximization [27]–[29]. For the transmit power minimization problem, the authors in [27] minimized the total transmit power by optimizing beamforming vectors of each user and the phase shift design of the IRS in an IRS-empowered downlink NOMA network. [26] considered a single IRS assisted downlink NOMA network and adopted reinforcement learning to design the beamforming vectors which minimized the transmit power at the BS. Regarding to the sum rate maximization problem, the authors of [27] optimized the beamforming design to maximize the sum rate in a downlink MISO IRS aided NOMA system. [28] discussed a multi-channel downlink communications IRS-NOMA framework, where the sum rate of multiple NOMA users served by one IRS was maximized by optimizing resource allocation to each user and jointly considering channel assignment and decoding order. [29] considered an IRS-assisted uplink NOMA system where multiple NOMA users can only transmit data through an IRS to the BS.

There are also some works considering a multi-cluster system mode, i.e., users are divided into different clusters [30], [31]. In [30], the authors discussed a downlink IRS-assisted NOMA network where two types of users named the central user and the cell edge user were assigned to different clusters. Each cluster had one central user, one cell edge user and one IRS serving all users. The authors minimized the transmit power at the BS by jointly optimizing the beamforming vectors of each user and the phase shift design of the IRS. In [31], the authors considered a multi-cluster and multi-BS IRS-aided NOMA network, where each cluster is served by its associated BS and one IRS serves all clusters. The sum rate was minimized by jointly optimizing power allocation and phase shift.

B. Motivation and Challenges

All the above works only consider one IRS. However the channel state of each user is related to its particular surrounding environment. Therefore, one single IRS might not be enough to reconfigure all users' channels simultaneously. Thus, multiple IRSs are deployed to assist the users whose channel conditions are bad. One IRS can adjust its phase shift dedicatedly for its associated user to generate a better channel condition. This paper considers a multi-cluster NOMA network, where each cluster has one IRS and the BS generates an unique beam for each cluster to serve all users located in this cluster.

With the considered scenario, there are a few challenges which need to be overcome. We consider a multi-user and multi-IRS scenario which increases the number of optimization variables and hence make the optimization more complicated than the case with a single IRS in the network. The joint optimization problem contains three coupled variables, which is a non-convex problem and highly intractable. Therefore, the primal problem is divided into subproblems. Those sub-problems are approximately transformed to the convex form but the feasibility of these subproblems cannot be guaranteed during the transformation. Moreover, due to the high quality of variables, the computing time of algorithms will be extensive.

C. Contributions

Different from the above mentioned works [30], [31], a new system model assisted with multiple IRSs is adopted in this paper. Then, a power minimization problem is formulated, which is non-convex and highly intractable. A novel alternating algorithm is proposed to solve this non-convex problem efficiently. Finally, a low-complexity algorithm, which achieves a reasonable performance, is also provided. The contributions are summarized as follows:

- A multi-cluster IRS-NOMA system is considered, where each cluster contains two users served by one IRS. The transmit power minimization problem with respect to the beamforming vector, the phase shifting matrix of IRSs and the power allocation coefficient of each cluster is formulated. Each IRS can accomplish channel reconfiguration according to the channel condition between the BS and the cell edge user it serves, which intuitively yields a better performance than the scenario with the single IRS serving the whole system.

- The formulated problem is non-convex because three variables are highly coupled together. To solve the proposed optimization problem, a novel alternating algorithm is proposed, where the primal problem is divided into the beamforming optimization problem and the phase shift feasibility problem. However, the beamforming optimization problem still has two variables coupled together, which causes the intractability. To address this challenge, the arithmetic and geometric means inequality is utilized to approximately transform the non-convex set to its convex upper bound. Then, the equivalence between Schur complement larger than zero and the positive semidefinite matrix and successive convex approximation (SCA) are applied to transfer another non-convex constraint to a convex form. Finally, the proposed alternating algorithm is ultilized to iteratively solve those two subproblems.

- Some fixed points are introduced during the approximation. It is essential to obtain the initial choice of the fixed points to guarantee the feasibility of the beamforming optimization problem. Therefore, a feasible initial points search algorithm is proposed, where an auxiliary variable forces all constraints to be feasible. Minimizing this auxiliary variable until it equals to zero will find the feasible fixed points. The values of the fixed points when this auxiliary variable equals to zero can be the initial choice of the fixed points for the proposed alternating algorithm.

- The complexity of the proposed alternating algorithm is high. To reduce the complexity, a simplified system model, where each cluster shares the same power allocation coefficient, is considered. With this assumption,

the previous problem will be degraded into a simpler one with two coupled variables. A partial exhaustive search algorithm is proposed to solve this new problem. Compared with the alternating algorithm, the complexity is reduced but the performance is still reasonable.

D. Organization

The rest of paper is organized as follows. In Section II, a multi-cluster IRS-assisted NOMA downlink network is introduced and a power minimization problem is formulated. In Section III, the solution to solve the formulated problem is introduced. In Section IV, the simplified optimization problem and the partial exhaustive search based algorithm are introduced. In Section V, the convergence analysis of the algorithms and the simulation results are provided. Finally, a conclusion is summarised in Section VI.

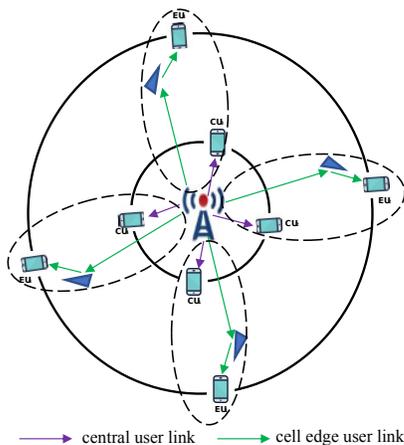## II. SYSTEM MODEL AND PROBLEM FORMULATION

A. System Model

Fig. 1: An IRS NOMA sytem model.

As shown in Fig.1, the multi-user downlink network contains two types of users, namely the central user and the cell edge user. They are served by the BS simultaneously. Generally, the central users are much closer to the base station than the cell edge users. It is assumed that there are K clusters and each cluster contains a central user, a cell edge user and an IRS. We use CU_k, EU_k and IRS_k to represent the central user, the cell edge user and the IRS in the k-th cluster, respectively. Each IRS is equipped with N passive reflecting elements and assists the cell edge user receiving signal from the BS. The BS is equipped with M ($K \leq M \leq 2K$) antennas and generates K beams to serve K clusters. It is assumed that the direct links between the BS and all the cell edge users are not available due to blockage, and the IRSs are implemented to reflect the signals sent by the BS to the cell edge users. Each cluster is far from others so the interference caused by the IRSs serving the other clusters can be reasonably ignored. In practice, the surface area of the IRS hardware is very limited, which can only reflect partial electromagnetic waves sent by the BS. The energy of the reflected signal will be greatly attenuated if there is severe path loss or fading attenuation [11]. In each cluster,

an IRS can be deployed carefully to ensure that it has strong connection to the cell edge user which does not have line-of-sight with the base station . As such, it is very likely that this IRS also has weak connections to those central users due to potential blockages [32]. The study for the case with direct links to those central users will be beyond the scope of this paper, but it will be an important direction for future research. The locations of each IRS and each user will also affect the total transmit power. For simplicity, the system model will be presented by assuming that the distances between the IRSs and the BS are the same. In Section V, simulation results will be presented to demonstrate the impact on the performance of the proposed algorithm with different BS-IRS distances. We note that the locations of the IRSs provide another dimension of system optimization, which is beyond the scope of this paper and will be treated as an important direction for future research.

To improve the spectrum efficiency, each cluster will use the same frequency-time resource block but with different beams, which is similar to the principle of spatial division multiple access (SDMA). NOMA is adopted within each mean to further improve the spectrum efficiency. The BS assigns different power levels to the signals being sent to the users in each cluster. The base station broadcasts the superposition signal $\sum_{i=1}^{K} \mathbf{w}_i (\sqrt{\alpha_i} s_{i,c} + \sqrt{1-\alpha_i} s_{i,e})$, where $\mathbf{w}_i \in \mathbb{C}^M$ denotes the beamforming vector in the i-th cluster and $i \in 1, 2, ..., K$. $s_{i,c}$ and $s_{i,e}$ denote the signals to be sent to the central user and the cell edge user in the i-th cluster, respectively, and α_i is the power allocation coefficient of CU_i, thus $1 - \alpha_i$ is the power coefficient of EU_i.

Without loss of generality, it is assumed that the channel between the IRS to the cell edge user follows Rayleigh fading in each cluster. Rician fading channel is adopted for the channel between the BS and the IRS and the channel between the BS and the central user in each cluster, which can be modelled as follows:

$$\mathbf{f} = \sqrt{\frac{\kappa}{1+\kappa}} \mathbf{f}^{\text{LoS}} + \sqrt{\frac{1}{1+\kappa}} \mathbf{f}^{\text{nLoS}}, \tag{1}$$

where κ is the Rician factor, $\mathbf{f}^{\text{LoS}}$ is the line-of-sight (LoS) component and $\mathbf{f}^{\text{nLoS}}$ is the non-Los (nLoS) component following the Rayleigh distribution.

Channel estimation is crucial for an IRS-assisted network to realize the beamforming design and phase shift design. If perfect channel information (CSI) is available, IRS is able to properly adjust the phase shift under the aid of a smart controller and the BS can generate the proper beams. In an IRS-assisted network, there are two types of channels, namely the BS-IRS channel and the IRS-user channel. These two channels are always coupled together at the receiver end. Individually estimating these two channels is the main challenge for IRS channel estimation. Typically, an IRS is implemented two operational modes, which are the estimation mode and the reflecting mode [25]. The IRS can be switched between these two modes. On the estimation mode, the IRS will adjust each reflecting element to a particular phase and then IRS channel estimation algorithms, e.g. passive channel estimation based on machine learning [33], are applied to

4

acquire CSI. On the reflecting mode, the IRS performs like a mirror to reflect the signal sent by the BS. This paper mainly focuses on the beamforming, power allocation and phase shift design, and it is assumed that the perfect CSI is available at every node. Therefore, the signal received at CU_k is given by

$$
y_{k,c} = \underbrace{\mathbf{h}_{k,c}^H \mathbf{w}_k \sqrt{\alpha_k} s_{k,c}}_{\text{signal}} + \underbrace{\mathbf{h}_{k,c}^H \mathbf{w}_k \sqrt{1-\alpha_k} s_{k,e}}_{\text{intra-cluster interference}}
$$
$$
+ \underbrace{\mathbf{h}_{k,c}^H \sum_{\substack{i=1\\i\neq k}}^{K} \mathbf{w}_i (\sqrt{\alpha_i} s_{i,c} + \sqrt{1-\alpha_i} s_{i,e})}_{\text{inter-cluster interference}} + w_{k,c}, \quad (2)
$$

where $\mathbf{h}_{k,c} \in \mathbb{C}^{M\times1}$ denotes the channel vector between the base station and CU_k, and $w_{k,c} \sim \mathcal{CN}(0,\sigma^2)$ is the additive white Gaussian noise (AWGN). Meanwhile, the signal received at EU_k is given by

$$
y_{k,e} = \underbrace{(\mathbf{h}_{k,e}^H \boldsymbol{\Theta}_k \mathbf{G}_k) \mathbf{w}_k \sqrt{1-\alpha_k} s_{k,e}}_{\text{signal}} + \underbrace{(\mathbf{h}_{k,e}^H \boldsymbol{\Theta}_k \mathbf{G}_k) \mathbf{w}_k \sqrt{\alpha_k} s_{k,c}}_{\text{intra-cluster interference}}
$$
$$
+ \underbrace{(\mathbf{h}_{k,e}^H \boldsymbol{\Theta}_k \mathbf{G}_k) \sum_{\substack{i=1\\i\neq k}}^{K} \mathbf{w}_i (\sqrt{\alpha_i} s_{i,c} + \sqrt{1-\alpha_i} s_{i,e})}_{\text{inter-cluster interference}} + w_{k,e},
$$
$$
(3)
$$

where $\mathbf{G}_k \in \mathbb{C}^{N\times M}$ denotes the channel matrix between the BS and IRS_k, $w_{k,e} \sim \mathcal{CN}(0,\sigma^2)$ denotes AWGN, $\mathbf{h}_{k,e} \in \mathbb{C}^{N\times1}$ denotes the channel vector between IRS_k and EU_k and $\boldsymbol{\Theta}_k = \text{diag}(\beta e^{j\theta_1^k}, ..., \beta e^{j\theta_n^k})$ is the phase shift matrix of IRS_k, where $\theta_n^k \in [0,2\pi), n \in \{1,...,N\}$ and $\beta \in [0,1]$ denote the phase shift of each reflecting element n and amplitude coefficient on the signal, respectively. It is assumed that $\beta = 1$ given the fact that each reflecting element can only change the phase but not the amplitude of the reflected signal [26]. It is worth to point out that the IRS with $\beta = 1$ may not be the optimal choice in some scenarios which are related to physical layer security. Therefore, the design of amplitudes is still crucial. More detailed discussions about the choices of the reflecting amplitude and the phase shift can be found in [34]. Due to path loss, we consider that the signal can be only efficiently reflected by the IRS once. Moreover, the long distance that geographically separates each cluster justifies the assumption that the IRS in one cluster will not infect other clusters. SIC strategies will directly affect the power allocation level. Since the cell edge user is far from the BS whose signal will suffer severe large scale path loss, the strategy that SIC is only performed at the central user to eliminate the interference from its intra-cluster edge user and the cell edge user decodes its data directly is adopted. In this case, the cell edge user will be allocated more power. It is necessary to point out that other SIC strategies will result in different power allocation,

which will be studied in the future research. Hence, the SINR of EU_k is given by

$$
\text{SINR}_{k,e} = \frac{|\mathbf{h}_{k,e}^H \boldsymbol{\Theta}_k \mathbf{G}_k \mathbf{w}_k|^2 (1-\alpha_k)}{|\mathbf{h}_{k,e}^H \boldsymbol{\Theta}_k \mathbf{G}_k \mathbf{w}_k|^2 \alpha_k + \sum_{\substack{i=1\\i\neq k}}^{K} |\mathbf{h}_{k,e}^H \boldsymbol{\Theta}_k \mathbf{G}_k \mathbf{w}_i|^2 + \sigma^2},
$$
$$
(4)
$$

where $|\mathbf{h}_{k,e}^H \boldsymbol{\Theta}_k \mathbf{G}_k \mathbf{w}_k|^2 \alpha_k$ is intra-cluster interference and $\sum_{\substack{i=1\\i\neq k}}^{K} |\mathbf{h}_{k,e}^H \boldsymbol{\Theta}_k \mathbf{G}_k \mathbf{w}_i|^2$ is inter-cluster interference. For the central users, they need to apply SIC to decode $s_{k,e}$ first and then remove it. Thus, the SINR of signal $s_{k,e}$ observed at CU_k can be expressed as follows:

$$
\text{SINR}_{k,c\to e} = \frac{|\mathbf{h}_{k,c}^H \mathbf{w}_k|^2 (1-\alpha_k)}{|\mathbf{h}_{k,c}^H \mathbf{w}_k|^2 \alpha_k + \sum_{\substack{i=1\\i\neq k}}^{K} |\mathbf{h}_{k,c}^H \mathbf{w}_i|^2 + \sigma^2}. \quad (5)
$$

The SINR of CU_k to decode its own signal is given by

$$
\text{SINR}_{k,c} = \frac{|\mathbf{h}_{k,c}^H \mathbf{w}_k|^2 \alpha_k}{\sum_{\substack{i=1\\i\neq k}}^{K} |\mathbf{h}_{k,c}^H \mathbf{w}_i|^2 + \sigma^2}. \quad (6)
$$

The design of beamforming vector is critical. Some existing works adopted block-diagonalization-based beamforming, e.g. using vectors from a FFT matrix, to cancel the inter-cluster interference [30]. However, in this paper, beamforming vectors are deigned by applying convex optimization directly, thus, the inter-cluster interference exists as noise which cannot be ignored.

B. Problem Formulation

In this section, a transmit power minimization problem is formulated by jointly optimizing the beamforming vector $(\mathbf{w}_k, k \in \{1,...K\})$, power allocation coefficients $(\alpha_k, k \in \{1,...K\})$ and phase shifting matrix $(\boldsymbol{\Theta}_k, k \in \{1,...K\})$, while considering the quality of service (QoS) requirement and the constraints of reflecting elements. The considered transmit power minimization problem can be formulated as follows:

$$
\text{P0}: \min_{\boldsymbol{\alpha},\mathbf{w},\boldsymbol{\Theta}} \sum_{k=1}^{K} ||\mathbf{w}_k||^2 \tag{7a}
$$
$$
\text{s.t. } \log_2(1+\text{SINR}_{k,c}) \geq R_{k,c}, \quad \forall k \tag{7b}
$$
$$
\log_2(1+\min(\text{SINR}_{k,e},\text{SINR}_{k,c\to e})) \geq R_{k,e}, \forall k \tag{7c}
$$
$$
0 \leq \theta_{k,n} \leq 2\pi, \quad \forall\, k,n \tag{7d}
$$
$$
|\boldsymbol{\Theta}_{k,n,n}| \leq 1, \quad \forall\, k,n \tag{7e}
$$

where $||\mathbf{w}_k||^2$ is the transmit power allocated to the cluster k, $R_{k,c}$ and $R_{k,e}$ denote the required minimum data rate of CU_k and EU_k, respectively. The constraints (7b) and (7c) indicate the QoS requirements of the central users and the cell edge users, (7d) defines the phase shift range of the reflecting elements and (7e) ensures that the IRS is a passive component.

Note that for many important applications for IRS, such as the next-generaiton Internet of Things (IoT), users, such as IoT sensors, can be severely energy constrained, which motivates the formulated power minimization problem. In particular, this formulated optimization problem can reduce the energy consumption at the IoT sensors, while guaranteeing their communication requirements [25], [30]. Alternatively, the energy efficiency optimization problem can also be formulated for the addressed IRS-NOMA network, which is beyond the scope of this paper but is an important direction for future research.

However, problem P0 is highly intractable due to the non-convex constraints (7b) and (7c). The non-convexity is caused by three highly coupled variables (i.e. $\mathbf{w}$, α and $\boldsymbol{\Theta}$). To efficiently solve this problem, SCA, SDR and the inequality approximation are adopted to develop an alternating algorithm to iteratively solve it.

III. OPTIMIZATION SOLUTION

As discussed in the previous section, it is difficult to find the optimal solution of P0 due to its non-convexity. In this section, an alternating optimization algorithm is proposed to solve P0 efficiently. The main idea of this algorithm is to divide the primal problem into two subproblems and solve them alternatively. In particular, P0 is divided to a beamforming optimization subproblem and a feasible phase shifting matrix search subproblem. As shown later, each of the two sub-problems is non-convex, and we will approximately transform them into the convex form, which can be solved efficiently by convex solvers, e.g., CVX in Matlab.

A. Beamforming Optimization

For a given phase shifting matrix $\boldsymbol{\Theta}$, the concatenated channel respond $\mathbf{h}_{k,e}^H \boldsymbol{\Theta}_k \mathbf{G}_k \in \mathbb{C}^{1 \times M}$ is fixed. Thus, the beamforming optimization problem can be formulated as

$$P1 : \min_{\alpha, \mathbf{w}} \quad \sum_{k=1}^{K} ||\mathbf{w}_k||^2 \tag{8a}$$

$$\text{s.t.} \quad \log_2(1 + \text{SINR}_{k,c}) \geq R_{k,c}, \quad \forall k \tag{8b}$$

$$\log_2(1 + \text{SINR}_{k,e}) \geq R_{k,e}, \quad \forall k \tag{8c}$$

$$\log_2(1 + \text{SINR}_{k,c \to e}) \geq R_{k,e}, \quad \forall k \tag{8d}$$

$$0 \leq \alpha_k \leq 1, \quad \forall k. \tag{8e}$$

P1 is non-convex because the beamforming vector and the power allocation coefficient are still coupled together in all constraints except (8e), which is challenging to be solved. It is noted that the rank-constrained semidefinite programming (SDP) problem can be approximated to a convex form. There-fore, after converting P1 into a SDP form, SDR can be applied to solve this problem.

First, the constraint (8c) needs to be approximately trans-formed into a convex form. According to (4), the constraint (8c) can be rewritten as follows:

$$\frac{|e_k^H \mathbf{D}_{k,e} \mathbf{G}_k \mathbf{w}_k|^2 (1 - \alpha_k)}{|e_k^H \mathbf{D}_{k,e} \mathbf{G}_k \mathbf{w}_k|^2 \alpha + \sum_{\substack{i=1 \\ i \neq k}}^{K} |e_k^H \mathbf{D}_{k,e} \mathbf{G}_k \mathbf{w}_i|^2 + \sigma^2} \geq r_{k,e}, \tag{9}$$

where $r_{k,e} = 2^{R_{k,e}} - 1$, e_k is an $N \times 1$ vector containing all the diagonal elements of $\boldsymbol{\Theta}_k^H$, and $\mathbf{D}_{k,e}$ is a diagonal matrix, whose main diagonal elements are from the channel vector $\mathbf{h}_{k,e}^H$. After some algebraic transformations, (9) can be equivalently expressed as follows:

$$|e_k^H \mathbf{D}_{k,e} \mathbf{G}_k \mathbf{w}_k|^2 (1 + r_{k,e}) \alpha_k \leq$$
$$|e_k^H \mathbf{D}_{k,e} \mathbf{G}_k \mathbf{w}_k|^2 - \sum_{\substack{i=1 \\ i \neq k}}^{K} |e_k^H \mathbf{D}_{k,e} \mathbf{G}_k \mathbf{w}_i|^2 r_{k,e} - \sigma^2 r_{k,e}. \tag{10}$$

Since the CSI is perfectly known by the BS, the channel $e_k^H \mathbf{D}_{k,e} \mathbf{G}_k$ is fixed with a given phase shifting matrix. For simply notation, we replace $e_k^H \mathbf{D}_{k,e} \mathbf{G}_k$ with $\mathbf{z}_{k,e}^H$ and rewrite (10) as follows:

$$\alpha_k |\mathbf{z}_{k,e}^H \mathbf{w}_k|^2 \leq \frac{|\mathbf{z}_{k,e}^H \mathbf{w}_k|^2}{1 + r_{k,e}} - \left(\sum_{\substack{i=1 \\ i \neq k}}^{K} |\mathbf{z}_{k,e}^H \mathbf{w}_i|^2 + \sigma^2 \right) \frac{r_{k,e}}{1 + r_{k.e}}, \tag{11}$$

where $\mathbf{z}_{k,e}^H = e_k^H \mathbf{D}_{k,e} \mathbf{G}_k$. The quadratic form $|\mathbf{z}_{k,e}^H \mathbf{w}_k|^2$ in (11) can be rewritten as $\mathbf{w}_k^H \mathbf{Z}_{k,e} \mathbf{w}_k$, where $\mathbf{Z}_{k,e} = \mathbf{z}_{k,e} \mathbf{z}_{k,e}^H$. A slack matrix $\mathbf{W}_k = \mathbf{w}_k \mathbf{w}_k^H$ is introduced, which is a rank-one positive semidefinite (PSD) matrix. It is known that $\mathbf{w}_k^H \mathbf{Z}_{k,e} \mathbf{w}_k = \text{Tr}(\mathbf{Z}_{k,e} \mathbf{W}_k)$ from SDR. Moreover, the rank of $\mathbf{W}_k$ has to be 1 and $\mathbf{W}_k$ is a PSD because of $\mathbf{W}_k = \mathbf{w}_k \mathbf{w}_k^H$. Then the constraint (11) can be equivalently rewritten as follows:

$$\alpha_k \text{Tr}(\mathbf{Z}_{k,e} \mathbf{W}_k) \leq$$
$$\frac{\text{Tr}(\mathbf{Z}_{k,e} \mathbf{W}_k)}{1 + r_{k,e}} - \left(\sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{Z}_{k,e} \mathbf{W}_i) + \sigma^2 \right) \frac{r_{k,e}}{1 + r_{k.e}} \tag{12}$$

$$\mathbf{W}_k \succeq 0 \tag{13}$$

$$\text{Rank}(\mathbf{W}_k) = 1. \tag{14}$$

From (12), it is noted that the right hand side of (12) is a liner combination of two convex terms with respect to $\mathbf{W}_k$, which is convex. The only obstacle is the left hand side, which is a bilinear term constructed by α_k and $\text{Tr}(\mathbf{Z}_{k,e} \mathbf{W}_k)$. To make this constraint a convex set, we need to approximately transform the non-convexity function $\alpha_k \text{Tr}(\mathbf{Z}_{k,e} \mathbf{W}_k)$ to a convex form. Inspired by the inequality of arithmetic and geometric means $2ab \leq a^2 + b^2$, where a and b are both non-negative numbers, we have the inequality that

$$2\alpha_k \text{Tr}(\mathbf{Z}_{k,e} \mathbf{W}_k) \leq (\alpha_k c_k)^2 + \left(\frac{\text{Tr}(\mathbf{Z}_{k,e} \mathbf{W}_k)}{c_k} \right)^2, \tag{15}$$

where c_k is a fixed point. From (15), it is noted that $(\alpha_k c_k)^2 + \left(\frac{\text{Tr}(\mathbf{Z}_{k,e} \mathbf{W}_k)}{c_k} \right)^2$ is an upper bound of $2\alpha_k \text{Tr}(\mathbf{Z}_{k,e} \mathbf{W}_k)$ and is a liner combination of two convex terms. Therefore, the non-convex feasible set of the left hand side term can be upper bounded by a convex set $\frac{1}{2}(\alpha_k^2 + \text{Tr}(\mathbf{Z}_{k,e} \mathbf{W}_k)^2)$. To tighten this upper bound in each iteration of the proposed iterative algorithm, updating rule of c_k in each iteration needs to be defined.

Lemma 1 *The fixed point c_k at the m-th iteration can be updated by:*

$$c_k^{(m)} = \sqrt{\frac{\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k^{(m-1)})}{\alpha_k^{(m-1)}}} \qquad (16)$$

Proof: A difference function of the original function $2\alpha_k\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k)$ and its approximated upper bound is defined as follows:

$$\mathcal{F}(c_k) = 2\alpha_k\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k) - (\alpha_k c_k)^2 - \left(\frac{(\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k))}{c_k}\right)^2. \qquad (17)$$

When the function (17) equals to 0, the equality of (15) holds, which tightens the upper bound. From (15), it is noted that the maximum value of function $\mathcal{F}(c_k)$ is 0. Since

$$\frac{\partial^2 \mathcal{F}(c_k)}{\partial c_k^2} = -2\alpha_k - \frac{6\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k)}{c_k^4} \le 0, \qquad (18)$$

when $\alpha_k \ge 0$ and $\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k) \ge 0$, the function $\mathcal{F}(c_k)$ is a concave function with respect to c_k. According to the Karush–Kuhn–Tucker (KKT) conditions, the maximum value of a concave function is obtained by letting the first order derivative equal to 0. Thus, the optimal value of c_k, defined as c_k^*, can be obtained by $\frac{\partial \mathcal{F}(c_k)}{\partial c_k} = 0$, then c_k^* can be given by

$$c_k^* = \sqrt{\frac{\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k)}{\alpha_k}}. \qquad (19)$$

Hence, the constraint (12) can be approximated as follows:

$$(\alpha_k c_k)^2 + \left(\frac{(\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k))}{c_k}\right)^2 \le$$
$$2\frac{\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k)}{1+r_{k,e}} - 2(\sum_{\substack{i=1 \\ i\neq k}}^{K}\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_i) + \sigma^2)\frac{r_{k,e}}{1+r_{k.e}}. \qquad (20)$$

It is noted that the left hand side of (20) is convex and the right hand side of (20) is an affine function, which means that the constraint (20) is a convex set. Eventually, (8c) can be approximated by (13), (14) and (20).

For handling with the next non-convex constraint (8d), after some algebraic manipulations, (8d) can be rewritten as follows:

$$\alpha_k|\mathbf{h}_{k,c}^H\mathbf{w}_k|^2 \le \frac{|\mathbf{h}_{k,c}^H\mathbf{w}_k|^2}{1+r_{k,e}} - (\sum_{\substack{i=1 \\ i\neq k}}^{K}|\mathbf{h}_{k,c}^H\mathbf{w}_i|^2 + \sigma^2)\frac{r_{k,e}}{1+r_{k.e}}. \qquad (21)$$

It is worth to point out that (21) has the same form as (11). Similarly, the method allied to (11) can be efficiently

applied to (21) to yield a convex form. Therefore, (21) can be approximately transformed to

$$(\alpha_k d_k)^2 + \left(\frac{(\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k))}{d_k}\right)^2 \le$$
$$2\frac{\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k)}{1+r_{k,e}} - 2(\sum_{\substack{i=1 \\ i\neq k}}^{K}\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i) + \sigma^2)\frac{r_{k,e}}{1+r_{k.e}}, \qquad (22)$$

(13),   (14),

where $\mathbf{H}_{k,c} = \mathbf{h}_{k,c}\mathbf{h}_{k,c}^H$, and d_k is a fixed point. At the m-th iteration, d_k can be updated as follows:

$$d_k^{(m)} = \sqrt{\frac{\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k^{(m-1)})}{\alpha_k^{(m-1)}}}. \qquad (23)$$

Eventually, (8d) can be approximated by (13), (14) and (22).

Now, we focus on the last non-convex constraint (8b). First, it can be rewritten as follows:

$$\alpha_k\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k) \ge \sum_{\substack{i=1 \\ i\neq k}}^{K}\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i)r_{k,c} + \sigma^2 r_{k,c} \qquad (24)$$

where $r_{k,c} = 2^{R_{k,c}} - 1$. Though (24) also has a bilinear term $\alpha_k\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k)$, the method which has been successfully applied to the constraint (8c) and (8d) cannot be straight-forwardly applied. Replacing $\alpha_k\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k)$ with the upper bound through the inequality of arithmetic and geometric means does not work because it is located at the left hand side of $\ge$ sign, which causes this inequality to be concave not convex. Hence, another method is proposed to deal with this constraint. First, we introduce a slack variable t_k and (8b) can be transformed to

$$\alpha_k\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k) \ge t_k^2 \qquad (25)$$

$$t_k^2 \ge \sum_{\substack{i=1 \\ i\neq k}}^{K}\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i)r_{k,c} + \sigma^2 r_{k,c}. \qquad (26)$$

It can be straightforwardly shown that neither of (25) and (26) is convex. According to the Schur complement theory [35], it is known that the sufficient and necessary condition for a matrix to be PSD is that its Schur complement is greater than zero. Moreover, a PSD matrix is a convex constraint. After a simple transformation, (25) can be rewritten as follows:

$$\alpha_k - \frac{t_k^2}{\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k)} \ge 0, \qquad (27)$$

which is equivalent to

$$\begin{bmatrix} \alpha_k & t_k \\ t_k & \text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i) \end{bmatrix} \succcurlyeq 0. \qquad (28)$$

Constraints (27) and (28) are mutually sufficient, and constraint (28) is convex. Now, we deal with the constraint (26). It is noted that t_k^2 is on the left hand side of the greater sign, which makes the whole constraint a non-convex set. SCA is utilized to deal with this, where the first order Taylor series

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2021.3090255

7

Algorithm 1 Initial Point Search Algorithm

1: **Initialize:** $c_k^{(0)}, d_k^{(0)}, t_{k,0}^{(0)} \quad \forall k, \epsilon = 0.00001, i = 0, q^{(0)} = 100$.
2: **while** $q^{(i)} > \epsilon$ **do**
3:     $i = i + 1$.
4:     Update $\mathbf{W}_k^{(i)}$, $\alpha_k^{(i)}$ and $q^{(i)}$ with fixed $c_k^{(i-1)}$, $d_k^{(i-1)}$, $t_{k,0}^{(i-1)}$ by solving P3.
5:     Update $c_k^{(i)}$, $d_k^{(i)}$, and $t_{k,0}^{(i)}$ based on (16), (23) and (30) respectively.
6: **end while**
7: **Output** $c_k^{(i)}$, $d_k^{(i)}$, and $t_{k,0}^{(i)}$.

---

approximation is adopted to approximate the quadratic form (26) as

$$t_{k,0}^2 + 2t_{k,0}(t_k - t_{k,0}) \geq \sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i)r_{k,c} + \sigma^2 r_{k,c} \tag{29}$$

where $t_{k,0}$ is a fixed point introduced by SCA. The updating rule of $t_{k,0}$ at the m-th iteration is given by

$$t_{k,0}^{(m)} = t_k^{(m-1)}. \tag{30}$$

The final obstacle to deal with this problem arises from the rank-one constraint (14). By applying SDR, the rank-one constraint is omitted to make the whole problem tractable. Thus, P1 is eventually transformed to

$$\text{P2}: \min_{\alpha, \mathbf{W}, t} \quad \sum_{k=1}^{K} \text{Tr}(\mathbf{W}_k) \tag{31a}$$

$$\text{s.t.} \quad (\alpha_k c_k)^2 + \left(\frac{(\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k))}{c_k}\right)^2 \leq$$
$$2\frac{\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k)}{1 + r_{k,e}} - 2(\sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_i) + \sigma^2)\frac{r_{k,e}}{1 + r_{k.e}}, \forall k \tag{31b}$$

$$(\alpha_k d_k)^2 + \left(\frac{(\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k))}{d_k}\right)^2 \leq$$
$$2\frac{\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k)}{1 + r_{k,e}} - 2(\sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i) + \sigma^2)\frac{r_{k,e}}{1 + r_{k.e}}, \forall k \tag{31c}$$

$$\begin{bmatrix} \alpha_k & t_k \\ t_k & \text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i) \end{bmatrix} \succcurlyeq 0, \forall k \tag{31d}$$

$$t_{k,0}^2 + 2t_{k,0}(t_k - t_{k,0}) \geq \sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i)r_{k,c} + \sigma^2 r_{k,c}, \forall k \tag{31e}$$

$$0 \leq \alpha_k \leq 1, \forall k. \tag{31f}$$

Since the restriction of rank one is removed, P2 is a convex problem and can be efficiently solved by convex optimization

toolboxes, for instance, CVX. It is noted that P1 and P2 have different optimization valuables. It is crucial to extract the optimal solution of P1 from the optimal solution of P2. We define the optimal solution of P2 as $\mathbf{W}_k^*, \forall k$, and each $\mathbf{W}_k^*$ is a positive semidefinite matrix. However, the optimal solution of P1 will not be obtained from the optimal solution of P2 unless the rank of $\mathbf{W}_k^*, \forall k$ is 1. If the rank of $\mathbf{W}_k^*$ is not 1, Gaussian randomization [36] is applied to alternatively obtain a suboptimal solution of P1. Specifically, several random vectors $\xi_k \sim \mathcal{N}(0, \mathbf{W}_k^*)$ will be generated and stored in a vector set. The one from this set which can satisfy all the constraints in P1 and also yield the best objective of P1 will be the suboptimal solution of P1.

Before solving P2, three fixed points, c_k, d_k and $t_{k,0}, \forall k$ need to be initialized. It is noted that initializing them randomly will make the formulated problem infeasible. Hence, a feasible initial points search algorithm is proposed to find the feasible fixed points to make P2 solvable. From P2, it is noted that the fixed points c_k. d_k and $t_{k,0}$ must satisfy the constraints (31b), (31c) and (31d). An auxiliary variable q, which intentionally relaxes the constraints to enlarge the feasible set, is introduced to address this problem. The initial point search problem can be formulated as follows:

$$\text{P3}: \min_{\alpha, \mathbf{W}, t, q} \quad q \tag{32a}$$

$$\text{s.t.} \quad (\alpha_k c_k)^2 + \left(\frac{(\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k))}{c_k}\right)^2 - q \leq$$
$$2\frac{\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k)}{1 + r_{k,e}} - 2(\sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_i) + \sigma^2)\frac{r_{k,e}}{1 + r_{k.e}}, \forall k \tag{32b}$$

$$(\alpha_k d_k)^2 + \left(\frac{(\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k))}{d_k}\right)^2 - q \leq$$
$$2\frac{\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k)}{1 + r_{k,e}} - 2(\sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i) + \sigma^2)\frac{r_{k,e}}{1 + r_{k.e}}, \forall k \tag{32c}$$

$$\begin{bmatrix} \alpha_k & t_k \\ t_k & \text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i) \end{bmatrix} \succcurlyeq 0, \forall k \tag{32d}$$

$$t_{k,0}^2 + 2t_{k,0}(t_k - t_{k,0}) + q \geq \sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i)r_{k,c} + \sigma^2 r_{k,c}, \forall k \tag{32e}$$

$$0 \leq \alpha_k \leq 1, \forall k \tag{32f}$$

$$q \geq 0. \tag{32g}$$

Specifically, when q equals to 0, all constraints in P3 are exactly the same as the constraints in P2 and the obtained values of c_k, d_k and $t_{k,0}$ can be the initial points of P2, which will guarantee the feasibility. It is noted that the objective function is an affine function and all constraints are convex so it can be solved easily by CVX. To solve P3 efficiently, an iterative algorithm shown as Algorithm 1 is proposed. It is worth to point out that, unlike P2, the initial points $c_k^{(0)}$,

---

Algorithm 2 The Beamforming Optimization Algorithm

1: **Initialize:** fixed feasible points $\{c_k^{*(0)}, d_k^{*(0)}, t_{k,0}^{*(0)},\}\ \forall k$, $\epsilon = 0.001$, $m = 0$.

2: **while** $\sum_{k=1}^{K} \text{Tr}(\mathbf{W}_k^{(m-1)}) - \sum_{k=1}^{K} \text{Tr}(\mathbf{W}_k^{(m)}) \geq \epsilon$ **do**

3:   Update beamforming matrix $\{\mathbf{W}_k^{(m)}, \alpha_k^{(m)}\}$, $\forall k$ by solving P2 with the fixed feasible point$\{c_k^{*(m)}, d_k^{*(m)}, t_{k,0}^{*(m)}\}$, $\forall k$.

4:   Update $\{c_k^{*(m)}, d_k^{*(m)}, t_{k,0}^{*(m)}\}$, $\forall k$ based on (16), (23) and (30) respectively.

5:   $m = m + 1$.

6: **end while**

7: Update $\alpha_k^* = \alpha_k^{(m)}, \forall k$

8: Update beamforming vector $\mathbf{w}_k^*, \forall k$ by decomposing $\mathbf{W}_k^{(m)}, \forall k$ based on Gaussian Randomization method.

9: **Output** $\{\mathbf{w}_k^*, \alpha_k^*\}$, $\forall k$

Algorithm 3 The Proposed Alternating Algorithm

1: **Initialize:** $\epsilon = 0.001, j = 0$.

2: **while** $\sum_{k=1}^{K} \text{Tr}(\mathbf{W}_k^{*(j-1)}) - \sum_{k=1}^{K} \text{Tr}(\mathbf{W}_k^{*(j)}) \geq \epsilon$ **do**

3:   Searching initial fixed feasible point $\{c_k^{*(j)}, d_k^{*(j)}, t_{k,0}^{*(j)}\}$, $\forall k$ based on Algorithm 1.

4:   Update $\{\mathbf{W}_k^{*(j)}, \mathbf{w}_k^{*(j)}, \alpha_k^{*(j)}\}$, $\forall k$ based on Algorithm 2.

5:   Update $\mathbf{V}_k^{*(j)}, \forall k$ by solving P5 with $\{\mathbf{w}_k^{*(j)}, \alpha_k^{*(j)}\}$, $\forall k$

6:   Update phase shift vector $\mathbf{e}_k^{*(j)}, \forall k$ by decomposing $\mathbf{V}_k^{*(j)}, \forall k$ based on Gaussian Randomization method.

7:   $j = j + 1$

8: **end while**

9: **Output** $\{\mathbf{w}_k^{*(j)}, \alpha_k^{*(j)}, \mathbf{e}_k^{*(j)}\}$, $\forall k$.

$d_k^{(0)}$ and $t_{k,0}^{(0)}$ in P3 can be generated randomly because the feasibility of P3 can be always guaranteed by q.

After deciding the fixed points, the last challenge for solving the beamforming optimization problem has been removed. To solve this problem efficiently, an iterative algorithm is designed to solve P2 iteratively. The details of the algorithm are shown in Algorithm 2. Specifically, the fixed initial points $\{c_k^{*(0)}, d_k^{*(0)}, t_{k,0}^{*(0)}\}\ \forall k$ are obtained from Algorithm 1.

B. Phase Shifting Optimization

In this section, we focus on the phase shifting optimization. The phase shifting optimization can be transformed to a feasibility problem since the objective function in the primal problem does not contain the phase shifting parameter $\mathbf{\Theta}_k, \forall k$. Only the constraints (7c), (7d) and (7e) in the primal problem contain the phase shifting parameter and (7c) can be equivalently divided into (8c) and (8d), where only (8c) contains the phase shifting parameter. Therefore, given the beamforming vectors, the phase shift feasibility problem can be written as follows:

$$\text{P4}: \text{find} \quad \mathbf{\Theta} \tag{33a}$$
$$\text{s.t.} \quad \log_2(1 + \text{SINR}_{k,e}) \geq R_{k,e}, \forall k \tag{33b}$$
$$0 \leq \theta_{k,n} \leq 2\pi, \forall k, n \tag{33c}$$
$$|\mathbf{\Theta}_{k,n,n}| = 1, \forall k, n. \tag{33d}$$

It is straightforward to find out that the non-convexity arises from the constraint (33b). The first step is to transform this non-convex constraint to be a convex constraint. Thus, (33b) can be rewritten as follows:

$$|\mathbf{h}_{k,e}^H \Gamma_{\mathbf{p}_k} \mathbf{e}_k|^2 (1 + r_{k,e}) \alpha_k \leq$$
$$|\mathbf{h}_{k,e}^H \Gamma_{\mathbf{p}_k} \mathbf{e}_k|^2 - \sum_{\substack{i=1 \\ i \neq k}}^{K} |\mathbf{h}_{k,e}^H \Gamma_{\mathbf{p}_i} \mathbf{e}_k|^2 - \sigma^2 r_{k,e}, \tag{34}$$

where $\Gamma_{\mathbf{p}_i}$ is a diagonal matrix whose main diagonal elements are from $\mathbf{p}_i = \mathbf{G}_k \mathbf{w}_i$ and $\mathbf{e}_k$ is the phase shifting vector. However, with $\mathbf{W}_k, \alpha_k, \forall k$ already obtained from the beamforming

optimization problem, the constraint (34) is a quartic form with respect to $\mathbf{e}_k$. For simplicity, we substitute $\mathbf{h}_{k,e}^H \Gamma_{\mathbf{p}_i}$ with $\mathbf{r}_{k,e}^{iH}$. From [36], it is known that a quartic form can be equivalently transformed to a linear form with a rank-one constraint. Thus, (34) can be expressed as follows:

$$\text{Tr}(\mathbf{R}_{k,e}^k \mathbf{V}_k)(1 + r_{k,e})\alpha_k \leq$$
$$\text{Tr}(\mathbf{R}_{k,e}^k \mathbf{V}_k) - \sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{R}_{k,e}^i \mathbf{V}_i) - \sigma^2 r_{k,e} \tag{35}$$

$$\mathbf{V}_k \succcurlyeq 0 \tag{36}$$
$$\text{Rank}(\mathbf{V}_k) = 1, \tag{37}$$

where $\mathbf{R}_{k,e}^i = \mathbf{r}_{k,e}^i \mathbf{r}_{k,e}^{iH}$ and $\mathbf{V}_i = \mathbf{e}_i \mathbf{e}_i^H$. Given $\mathbf{w}_k, \alpha_k, \forall k$, (35) is an affine constraint. The rank-one constraint will make the whole problem intractable, thus SDR is adopted again to remove this rank-one constraint. Then, P4 can be transformed as follows:

$$\text{P5}: \text{find} \quad \mathbf{V}_k, \quad \forall k \tag{38a}$$
$$\text{s.t.} \quad (35), \quad \forall k \tag{38b}$$
$$\mathbf{V}_k \succcurlyeq 0, \quad \forall k \tag{38c}$$
$$\mathbf{V}_{k,n,n} = 1, \quad \forall k, n. \tag{38d}$$

P5 is a convex problem, which can be solved by CVX efficiently. Since the rank-one constraint is removed, the optimal solution of P5 may not be the optimal solution of P4. Therefore, Gaussian randomization will be applied to achieve a suboptimal solution for P4.

C. Algorithm Design

The detail of the proposed alternating algorithm are illustrated in Algorithm 3, where P2 and P5 are alternately solved until the convergence metric is satisfied. At the i-th iteration of Algorithm 3, first, the initial points are obtained by Algorithm 1. Then, the algorithm begins to solve the beamforming optimization problem by solving P2 through Algorithm 2. Then, the algorithm starts to solve phase shifting feasibility problem by solving P5 (step 5 and step 6) to obtain a

feasible phase shift vector $\mathbf{e}_k^{*(i)}, \forall k$. The feasible phase shifting vector of this current iteration will be used as a given phase shift for the beamforming optimization in the next iteration. It is worth to point out that after each iteration, the channel state will change with the new obtained phase shifting vector $\mathbf{e}_k, \forall k$, so it is critical to search new feasible fixed points (step 3) before solving P2, which necessarily guarantees that P2 is always feasible.

It is worth to point out that there are three optimization variables coupled together in constraints (7b) and (7c), which are non-convex as well. Therefore, P0 is a NP-hard problem , i.e., it is difficult to solve it in polynomial time. It is difficult to find the global optimal solution of P0 by applying convex optimization. In Algorithm 3, the alternating algorithm and a few approximations are adopted to transform P0 to a solvable convex problem. Therefore, Algorithm 3 only provides a suboptimal solution for P0.

D. Complexity analysis

The worst complexity of solving a SDR problem through CVX provided by [36] is

$$\mathcal{O}(\max\{m,n\}^4 n^{1/2} \log(1/\epsilon_c)), \tag{39}$$

where n is the problem size, and m is the number of constraints and ϵ_c is the accuracy of the algorithm that CVX adopts. It is assumed that the problem size is greater than the number of constraints, then the complexity of CVX to solve a SDR problem can be expressed as

$$\mathcal{O}(n^{4.5} \log(1/\epsilon_c)). \tag{40}$$

Algorithm 1 is essentially to solve a SDR problem multiple times until the accuracy is satisfied. Thus, the complexity of Algorithm 1 is

$$\mathcal{O}\left(n_1^{4.5} \log\left(\frac{1}{\epsilon_c}\right) \log\left(\frac{1}{\epsilon_1}\right)\right), \tag{41}$$

where n_1 is the problem size of P3 and ϵ_1 is the accuracy of Algorithm 1. Algorithm 2 is similar to Algorithm 1, which is also to solve a SDR problem multiple times and hence P2 has the same size as P3. Thus, the complexity of Algorithm 2 can be expressed as follows:

$$\mathcal{O}\left(n_1^{4.5} \log\left(\frac{1}{\epsilon_c}\right) \log\left(\frac{1}{\epsilon_2}\right)\right), \tag{42}$$

where ϵ_2 is the accuracy of Algorithm 2. Now, we have the complexities of step 3 and step 4 in Algorithm 3. The last one is the complexity of step 5. It is easy to find out that a single SDR problem is solved in the step 5, so the complexity is

$$\mathcal{O}(n_2^{4.5} \log(1/\epsilon_c), \tag{43}$$

where n_2 is the problem size of P5. Finally, the complexity of the proposed algorithm is given by

$$\mathcal{O}(\mathcal{O}_1 \log(1/\epsilon_3)), \tag{44}$$

where

$$\mathcal{O}_1 = n_1^{4.5} \left(\log\left(\frac{1}{\epsilon_c}\right) \log\left(\frac{1}{\epsilon_1}\right) + \log\left(\frac{1}{\epsilon_c}\right) \log\left(\frac{1}{\epsilon_2}\right)\right) + n_2^{4.5} \log(1/\epsilon_c).$$

IV. Partial Exhaustive Search Algorithm

In this section, a simple algorithm based on partial exhaustive search is proposed, which can significantly reduce computation complexity. The main idea of this partial exhaustive search algorithm is to assume that all the clusters share the same power allocation coefficient, of which the optimal value can be obtained by an exhaustive search within the range [0, 1]. The primal problem can also be divided into the beamforming optimization problem and the phase shifting feasibility problem.

Since each cluster shares the same power coefficient, the power coefficient is first fixed in each searching progress so only the beamforming vector and the phase shifting vector need to be optimized in these two subproblems. It is noted that these two subproblems can be reduced to the QCQP problem, which is a classic form in convex optimization theory. SDR is widely used as one of the most common methods to efficiently solve the QCQP problem. Two subprobelms are formulated as P6 and P7. P6 and P7 can be obtained through the basic SDR theory and some simple algebraic transformations, where the derivation is omitted in this paper due to space limitations.

$$\text{P6}: \quad \min_{\mathbf{w}} \quad \sum_{k=1}^{K} \text{Tr}(\mathbf{W}_k) \tag{45a}$$

$$\text{s.t. } \alpha\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k) \geq \sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i)r_{k,c} + \sigma^2 r_{k,c}, \forall k \tag{45b}$$

$$\alpha\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k) \leq$$
$$\frac{\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k)}{1+r_{k,e}} - (\sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_i) + \sigma^2)\frac{r_{k,e}}{1+r_{k.e}}, \forall k \tag{45c}$$

$$\alpha\text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_k) \leq$$
$$\frac{\text{Tr}(\mathbf{Z}_{k,e}\mathbf{W}_k)}{1+r_{k,e}} - (\sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{H}_{k,c}\mathbf{W}_i) + \sigma^2)\frac{r_{k,e}}{1+r_{k.e}}, \forall k \tag{45d}$$

$$\mathbf{W}_k \succeq 0, \forall k \tag{45e}$$

where $\mathbf{Z}_{k,e}, \mathbf{W}_k, \forall k$ in P6 are the same as those in P2.

$$\text{P7}: \text{find} \quad \mathbf{V}_k, \quad \forall k \tag{46a}$$
$$\text{s.t.} \quad \text{Tr}(\mathbf{R}_{k,e}^k \mathbf{V}_k)(1+r_{k,e})\alpha \leq$$
$$\text{Tr}(\mathbf{R}_{k,e}^k \mathbf{V}_k) - \sum_{\substack{i=1 \\ i \neq k}}^{K} \text{Tr}(\mathbf{R}_{k,e}^i \mathbf{V}_i) - \sigma^2 r_{k,e}, \forall k \tag{46b}$$

$$\mathbf{V}_k \succeq 0, \quad \forall k \tag{46c}$$
$$\mathbf{V}_{k,n,n} = 1, \quad \forall k, n \tag{46d}$$

where $\mathbf{R}_{k,e}^i, \forall i, k$ and $\mathbf{V}_k, \forall k$ are the same as those in P5. The detail of the partial exhaustive search algorithm is illustrated in Algorithm 4.

In each search progress, the algorithm will solve two SDR problems with different sizes n_1 and n_2, which are the problem sizes of P6 and P7. Therefore, the complexity of Algorithm 4 can be expressed as follows:

$$\mathcal{O}\left(I\left(n_1^{4.5} \log(1/\epsilon_c) + n_2^{4.5} \log(1/\epsilon_c)\right)\right). \tag{47}$$

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This article is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2021.3090255

10

Algorithm 4 The Partial Exhaustive Search Algorithm

1: **Initialization** $P_{opt} = 10000, \alpha_{opt} = 0, \mathbf{w}_k^*, \mathbf{e}_k^*, \forall k$
2: **for** $\alpha = 0.1 : 0.1 : 0.9$ **do**
3:     **Initialization** $\epsilon = 0.001, i = 0, \mathbf{e}_k^{(0)}$
4:     **while** $\sum\limits_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k^{(i-1)}) - \sum\limits_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k^{(i)}) > \epsilon$ **do**
5:        Update $\mathbf{W}_k^{(i)}, \forall k$ by solving P6.
6:        Update $\mathbf{w}_k^{(i)}, \forall k$ by decomposing $\mathbf{W}_k^{(i)}, \forall k$ based on Gaussian Randomization method.
7:        Update $\mathbf{V}_k^{(i)}, \forall k$ by solving P7 based on given $\mathbf{w}_k^{(i)}, \forall k$.
8:        Update $\mathbf{e}_k^{(i)}, \forall k$ by decomposing $\mathbf{V}_k^{(i)}, \forall k$ based on Gaussian Randomization method.
9:        $i = i + 1$.
10:     **end while**
11:     **if** $P_{opt} > \sum\limits_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k^{(i)})$ **then**
12:        $P_{opt} = \sum\limits_{k=1}^{K} \mathrm{Tr}(\mathbf{W}_k^{(i)})$.
13:        $\alpha_{opt} = \alpha, \mathbf{w}_k^* = \mathbf{w}_k^{(i)}, \mathbf{e}_k^* = \mathbf{e}_k^{(i)}, \forall k$.
14:     **end if**
15: **end for**
16: **Output** $\alpha_{opt}, \mathbf{w}_k^*, \mathbf{e}_k^*, \forall k$.

I is the number of searches, which depends on the search step α. Obviously, the partial exhaustive search algorithm has a lower complexity than the complexity of the proposed alternating algorithm. It is worth to point out that the performance of this partial exhaustive search algorithm is related to the step size $\Delta\alpha$. A smaller step size will yield a better performance. However, according to (47), when the step size decreases, the complexity of the algorithm will increase. It is important to find a balance between performance and complexity. In numerical results, the performance of the partial exhaustive search algorithm with different step sizes is provided.

V. NUMERICAL RESULTS

In this section, we evaluate all simulation results of the proposed algorithms. In simulations, channel gains are generated by

$$\mathbf{h}_{k,e} = \frac{\mathbf{h}_{k,e}^*}{\sqrt{d_0^{\alpha_0}}} \quad \mathbf{G}_k = \frac{\mathbf{G}_k^*}{\sqrt{d_1^{\alpha_1}}} \quad \mathbf{h}_{k,c} = \frac{\mathbf{h}_{k,c}^*}{\sqrt{d_2^{\alpha_2}}} \quad (48)$$

where $k = 1, 2, ...K$, $\mathbf{h}_{k,e}^*$ and $\mathbf{h}_{k,c}^*$ are complex Reyleigh channel coefficients and $\mathbf{G}_k^*$ is complex Rician channel coefficient based on (1). $d_0 = 10$ m, $d_1 = 50$ m, $d_2 = 10$ m, respectively denote the distances between the IRS and the cell edge user, the distance between the BS and the IRS, and the distance between the BS and the cell center user. $\alpha_0, \alpha_1, \alpha_2$ are the path loss exponents of the corresponding links. It is assumed that all the cell central users are at the same distance from the BS, all the cell edge users are at the same distance from the related IRS and all the IRSs are at the same distance from the BS. We set $\alpha_0 = \alpha_2 = 1.8$ and $\alpha_1 = 2$. The noise power is $\sigma^2 = BN_0$, where the bandwidth $B = 100$ MHz
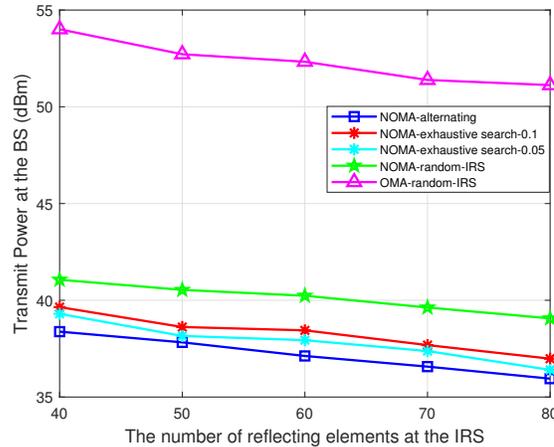

Fig. 2: The transmit power versus the number of reflecting elements at the IRS.

and the noise power spectral density is $N_0 = -80$ dBm. The number of clusters is $K = 4$, which means there are 8 users in the system. For the OMA benchmark scheme, OFDM is adopted where each user will occupy a specific spectrum and will not cause interference to other users. It is assumed that the channel bandwidth is normalized, thus each user in the OFDM scheme can occupy $1/2K$ bandwidth.

Fig. 2 shows the transmit power at the BS versus the number of each IRS's reflecting elements. We provide the performance of the proposed schemes compared with the random phase scheme in NOMA and OFDM. In Fig. 2, the number of antennas at the BS is $M = 6$, and the date rate requirement of all the users is 1 bps/Hz. Obviously, the transmit power at the BS of all schemes decreases with the increasing of the number of IRS's reflecting elements. From Fig. 2, we can see that both proposed algorithms requires a less transmit power than the benchmarks. Comparing the two proposed algorithms, the performance gap is very small and this gap will get smaller if the step size of the partial exhaustive search algorithm decreases. The result in Fig. 2 demonstrates that the alternating algorithm can yield the best performance among all the schemes but the partial exhaustive search can also yield competitive performance.

Fig. 3 shows the transmit power at the BS versus the minimum data rate of the central users. In this figure, it is assumed that each central user has the same date rate requirement, and all the cell edge users' date rate requirement is 1bps/Hz. In this figure, we set the number of antennas at the BS as $M = 6$ and the number of reflecting elements at each IRS as $N = 32$, respectively. According to the Shannon's capacity formula, it is well known that a higher date rate requires a higher transmit power at the BS. All schemes in Fig. 3 have the same trend, where the transmit power at the BS increases with the increasing of the central users' minimum date rate requirement. From Fig. 3, it is noted that the proposed alternating algorithm needs less power consumption under the same date rate requirement. Although, the partial exhaustive search algorithm cannot achieve the same performance as the proposed alternating algorithm, it has low complexity and still
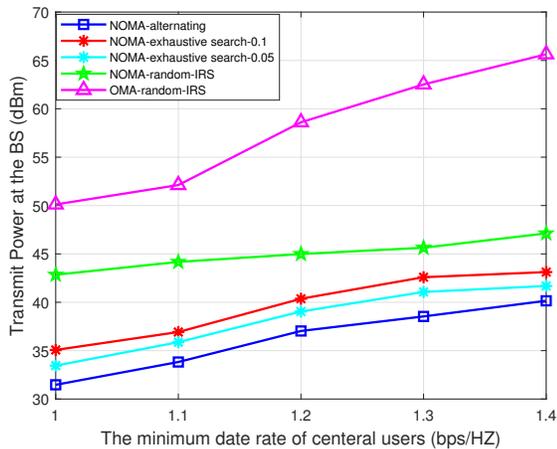
Fig. 3: The transmit power versus the minimum date rate of the central users.
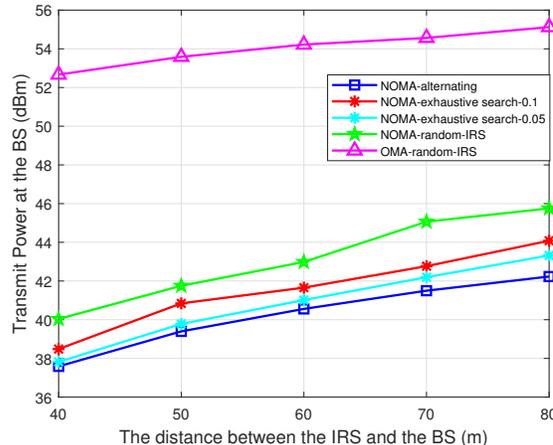

Fig. 4: The transmit power versus the number of antennas at the BS.

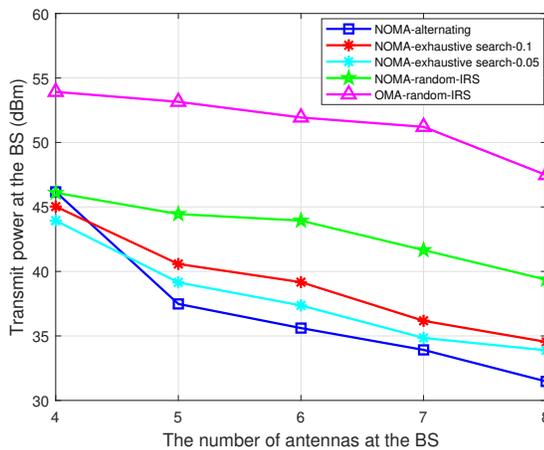Fig. 5: The transmit power versus the distance between the IRS and the BS.

Fig. 6: The value of q versus the iterative number.

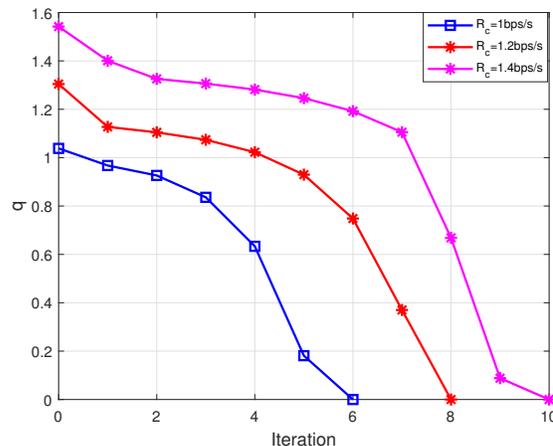Fig. 7: The transmit power at the BS versus the iterative number.

yields a better performance than NOMA with random IRS scheme and OFDM scheme.

Fig. 4 shows the transmit power versus the number of antennas at the BS. In this simulation, the relationship between the algorithm performance and the number of antennas at the BS is illustrated. Fig. 4 shows the performances of two proposed algorithms with NOMA and OFDM with random phase IRS. In Fig. 4, we set the number of reflecting elements at each IRS as $N = 32$ and the date rate requirement of all the users as 1 bps/Hz. From Fig. 4, it is noted that the alternating algorithm achieves better performance gain than other algorithms with the number of antennas at the BS increasing.

Fig. 5 shows the transmit power versus the distance between the IRS and the BS in each cluster. In Fig. 5, we set the number of antennas at the BS as $M = 6$, the number of each IRS's reflecting elements as $N = 32$. Each user's date rate requirement is 1bps/Hz. It has assumed that the IRS will not cause interference to the central user in each cluster. Since the central user in each cluster is close to the BS, thus the distance between the IRS and the BS cannot be short. Therefore, the starting point of the simulation is set as 40m. As expected,
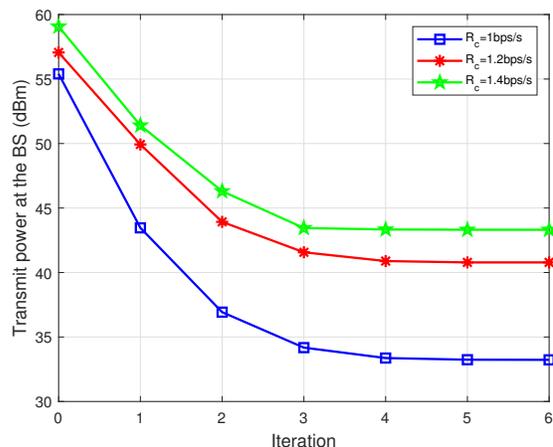
the transmit power of all schemes increases when the distance between the IRS in each cluster and the BS gets large. Similar to Fig. 4, the proposed alternating algorithm consumes less energy compared with all other schemes.

Fig. 6 shows the the value of q in the initial point search algorithm versus the iterative number. As previous discussion, the q represents the distance between the current problem and a feasible problem and q can enforce the current problem to be a feasible one. R_c denotes the data rate requirement of all the central users. Fig. 6 also shows that the value of q in the $R_c = 1.4$ bps/Hz scheme is larger than that in the $R_c = 1$ bps/Hz and $R_c = 1.2$ bps/Hz schemes at each iteration. Moreover, the scheme with $R_c = 1.4$ bps/Hz needs more iterations to converge, which indicates that a higher date rate requirement makes all constraints more difficult to be fulfilled.

Fig. 7 shows the transmit power at the BS versus the iterative number in Algorithm 2. R_c denotes the data rate requirement of all the central users. We evaluate the transmit power in different scenarios with the different data rate requirements of the central user. The data rate of all the cell edge users is 1 bps/Hz, the number of antennas at the BS is $M = 6$ and the number of each IRS is $N = 32$. From Fig. 7, we notice that the transmit power at the BS decreases with the number of iterations increasing, which also means this algorithm can converge with the algorithm proceeding.

VI. CONCLUSION

The joint optimization of beamforiming, power allocation and IRS phase shift in a NOMA-IRS assisted multi-cluster network is investigated in this paper. By introducing inequality approximation, SCA and SDR, an alternating algorithm is proposed to minimize the transmit power by iteratively solving beamforming optimization and phase shifting feasibility until the algorithm converges. Furthermore, an initial point search algorithm is proposed to guarantee the feasibility of the beamforming optimization subproblem. Moreover, a low-complexity solution is also provided for this scenario based on the partial exhaustive search. The simulation results demonstrated the alternating algorithm outperforms the partial exhaustive search algorithm but has a higher complexity. In the future research, the scenario that the IRS reconfigures the imperfect channel will be studied and the inter-cluster interference caused by IRS will also be considered.

REFERENCES

[1] Z. Zhang, Y. Xiao, Z. Ma, M. Xiao, Z. Ding, X. Lei, G. K. Karagiannidis, and P. Fan, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, 2019.

[2] 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, 2019.

[3] F. Tariq, M. Khandaker, K.-K. Wong, M. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," *arXiv preprint arXiv:1902.06700*, 2019.

[4] V. Jamali, A. Tulino, G. Fischer, R. Müller, and R. Schober, "Intelligent reflecting and transmitting surface aided millimeter wave massive MIMO," *arXiv e-prints*, 2019.

[5] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun Mag*, vol. 52, no. 2, pp. 186–195, 2014.

[6] Y. Guo, "Method and apparatus of small cell enhancement in a wireless communication system," Mar. 22 2016, uS Patent 9,295,077.

[7] Z. Ding, F. Adachi, and H. V. Poor, "The application of MIMO to non-orthogonal multiple access," *EEE Trans. Wirel. Commun.*, vol. 15, no. 1, pp. 537–552, 2015.

[8] M. Vaezi, Z. Ding, and H. V. Poor, *Multiple access techniques for 5G wireless networks and beyond*. Springer, 2019.

[9] K. Saito, A. Benjebbour, Y. Kishiyama, Y. Okumura, and T. Nakamura, "Performance and design of SIC receiver for downlink NOMA with open-loop SU-MIMO," in *2015 IEEE International Conference on Communication Workshop (ICCW)*. IEEE, 2015, pp. 1161–1165.

[10] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.

[11] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *arXiv preprint arXiv:2007.02759*, 2020.

[12] X. Sun, N. Yang, S. Yan, Z. Ding, D. W. K. Ng, C. Shen, and Z. Zhong, "Joint beamforming and power allocation in downlink NOMA multiuser MIMO networks," *EEE Trans. Wirel. Commun.*, vol. 17, no. 8, pp. 5367–5381, 2018.

[13] F. Fang, H. Zhang, J. Cheng, S. Roy, and V. C. Leung, "Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2874–2885, 2017.

[14] G. Surabhi, R. M. Augustine, and A. Chockalingam, "On the diversity of uncoded OTFS modulation in doubly-dispersive channels," *EEE Trans. Wirel. Commun.*, vol. 18, no. 6, pp. 3049–3063, 2019.

[15] W. Tang, M. Z. Chen, J. Y. Dai, Y. Zeng, X. Zhao, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with programmable metasurface: New paradigms, opportunities, and challenges on transceiver design," *arXiv preprint arXiv:1907.01956*, 2019.

[16] F. Fang, H. Zhang, J. Cheng, and V. C. Leung, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3722–3732, 2016.

[17] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun Mag*, 2019.

[18] M. Di Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: an idea whose time has come," *EURASIP J Wirel. Commun. Netw.*, vol. 2019, no. 1, pp. 1–20, 2019.

[19] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.

[20] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *arXiv preprint arXiv:1911.05326*, 2019.

[21] A. S. de Sena, D. Carrillo, F. Fang, P. H. J. Nardelli, D. B. da Costa, U. S. Dias, Z. Ding, C. B. Papadias, and W. Saad, "What role do intelligent reflecting surfaces play in non-orthogonal multiple access?" *IEEE Wireless Commun. Mag.*, July 2020 (Accept).

[22] F. Fang, Y. Xu, Q.-V. Pham, and Z.Ding, "Energy-efficient design of IRS-NOMA networks," *IEEE Trans. Veh. Technol*, Sept. 2020 (Accept).

[23] J. Zhu, Y. Huang, J. Wang, K. Navaie, and Z. Ding, "Power efficient IRS-assisted NOMA," *arXiv preprint arXiv:1912.11768*, 2019.

[24] Z. Ding, R. Schober, and H. V. Poor, "On the impact of phase shifting designs on IRS-NOMA," *IEEE Wireless Communications Letters*, vol. 9, no. 10, pp. 1596–1600, 2020.

[25] M. Fu, Y. Zhou, and Y. Shi, "Reconfigurable intelligent surface empowered downlink non-orthogonal multiple access," *arXiv preprint arXiv:1910.07361*, 2019.

[26] X. Liu, Y. Liu, Y. Chen, and H. V. Poor, "RIS enhanced massive non-orthogonal multiple access networks: Deployment and passive beamforming design," *arXiv preprint arXiv:2001.10363*, 2020.

[27] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in multi-antenna aided NOMA systems," *arXiv preprint arXiv:1910.13636*, 2019.

[28] J. Zuo, Y. Liu, Z. Qin, and N. Al-Dhahir, "Resource allocation in intelligent reflecting surface assisted NOMA systems," *arXiv preprint arXiv:2002.01765*, 2020.

[29] M. Zeng, X. Li, G. Li, W. Hao, and O. Dobre, "Sum rate maximization for IRS-assisted uplink NOMA," *arXiv preprint arXiv:2004.10791*, 2020.

This article has been accepted for publication in IEEE Transactions on Vehicular Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVT.2021.3090255

13

[30] Y. Li, M. Jiang, Q. Zhang, and J. Qin, "Joint beamforming design in multi-cluster MISO NOMA intelligent reflecting surface-aided downlink communication networks," *arXiv preprint arXiv:1909.06972*, 2019.

[31] W. Ni, X. Liu, Y. Liu, H. Tian, and Y. Chen, "Resource allocation for multi-cell irs-aided NOMA networks," *arXiv preprint arXiv:2006.11811*, 2020.

[32] Z. Ding and H. V. Poor, "A simple design of IRS-NOMA transmission," *IEEE Communications Letters*, vol. 24, no. 5, pp. 1119–1123, 2020.

[33] S. Yan, X. Zhou, D. W. K. Ng, J. Yuan, and N. Al-Dhahir, "Intelligent reflecting surface for wireless communication security and privacy," *arXiv preprint arXiv:2103.16696*, 2021.

[34] X. Zhou, S. Yan, Q. Wu, F. Shu, and D. W. K. Ng, "Intelligent reflecting surface (IRS)-aided covert wireless communication with delay constraint," *arXiv preprint arXiv:2011.03726*, 2020.

[35] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[36] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process Mag*, vol. 27, no. 3, pp. 20–34, 2010.
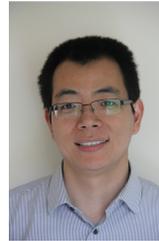
Zhiguo Ding (S'03-M'05-F'20) received his B.Eng in Electrical Engineering from the Beijing University of Posts and Telecommunications in 2000, and the Ph.D degree in Electrical Engineering from Imperial College London in 2005. From Jul. 2005 to Apr. 2018, he was working in Queen's University Belfast, Imperial College, Newcastle University and Lancaster University. Since Apr. 2018, he has been with the University of Manchester as a Professor in Communications. From Oct. 2012 to Sept. 2021, he has also been an academic visitor in Princeton University. Dr Ding' research interests are 5G networks, game theory, cooperative and energy harvesting networks and statistical signal processing. He is serving as an Area Editor for the *IEEE Open Journal of the Communications Society* an Editor for *IEEE Transactions on Vehicular Technology*, and *Journal of Wireless Communications and Mobile Computing*, and was an Editor for *IEEE Wireless Communication Letters*, *IEEE Transactions on Communications*, *IEEE Communication Letters* from 2013 to 2016. He recently received the EU Marie Curie Fellowship 2012-2014, the Top IEEE TVT Editor 2017, IEEE Heinrich Hertz Award 2018, IEEE Jack Neubauer Memorial Award 2018, IEEE Best Signal Processing Letter Award 2018 , and Friedrich Wilhelm Bessel Research Award 2020. He is a Fellow of the IEEE, a Distinguished Lecturer of IEEE ComSoc, and a Web of Science Highly Cited Researcher in two categories 2020.

Ximing Xie received the B.E. degree in communication engineering Harbin Institute of Technology, China, in 2017, and M.S degree in telecommunications from University Colloege London (UCL), UK, in 2018, respectively. He is currently pursuing the Ph.D. degree in electrical and electronic at the School of Electrical and Electronic Engineering engineering, The University of Manchester, UK. His current research interests include 5G and beyond wireless networks, machine learning, NOMA, IRS and backscattering techniques.

Fang Fang received her Ph.D. degree in electrical engineering from the University of British Columbia (UBC), Canada, in 2018. From 2018 to 2020, she was a Research Associate in the Department of Electrical and Electronic Engineering at the University of Manchester, UK. She is currently an Assistant Professor in the Department of Engineering at Durham University, Durham (UK). Her current research interests include intelligent beyond 5G/6G wireless communications, machine learning, non-orthogonal multiple access (NOMA), intelligent reflecting surface (IRS) and multi-access edge computing (MEC). Dr. Fang served as a technical program committee (TPC) member for IEEE flagship conferences, e.g., IEEE GLOBECOM, IEEE ICC, etc. She received the Exemplary Reviewer Certificate of the IEEE Transactions on Communications in 2017. Currently, she is an Associate Editor of IEEE Open Journal of The Communications Society.