Joint Hybrid Beamforming and Reconfigurable Intelligent Surface Design for MIMO Systems

Yuanli Ma[1], Bin Yan[1], Zheng Wang[1], and Zhen Gao[2]

[1]School of Information Science and Engineering, Southeast University, Nanjing, China
[2]School of Information and Electronics Beijing Institute of Technology, Beijing, China
E-mail: myl@seu.edu.cn; bin_yan@seu.edu.cn; wznuaa@gmail.com; gaozhen16@bit.edu.cn

***Abstract*—For reconfigurable intelligent surface (RIS) aided communication, hybrid beamforming structure has been applied due to its reduced hardware cost. In this paper, to maximize sum rate for multi-user MIMO (MU-MIMO) systems, we design the hybrid beamforming and RIS configuration by phase elimination in an alternate way. Specifically, the performance bounds of hybrid beamforming for RIS-aided MU-MIMO systems are investigated. Based on it, the problem of RIS design is approximated by a novel one to approach the performance bound. Then the hybrid beamforming matrices are jointly designed by phased-ZF, which leads to the proposed alternating optimization algorithm based on phase elimination (AO-PE). Finally, simulation results validate the advantage of proposed AO-PE over other algorithms for hybrid beamforming in RIS-aided MU-MIMO systems.**

***Index Terms*—Reconfigurable intelligent surface (RIS), hybrid beamforming, phase elimination.**

I. Introduction

As one of promising technologies for upcoming sixth-generation (6G) networks [1], the reconfigurable intelligent surface (RIS) has been carried out due to its ability to configure the wireless environment [2]. Specifically, RIS is a planar surface composed of reflecting elements, each of which can impose an independent phase shift [3]. To minimize the transmit power, [4] formulates the channel power gain maximization problem, and solves it by semidefinite relaxation (SDR). [5] aims to achieve sum rate maximization with the Riemannian conjugate gradient (RCG) algorithm or block coordinate descent (BCD) method.

Recently, hybrid beamforming in RIS-aided communication has been studied due to its considerable hardware cost reduction. For point-to-point MIMO communication, [6] utilizes sparse scattering nature of millimeter wave (mmWave) channels to approximate the rate maximization problem. Besides, for multi-user MIMO (MU-MIMO) communication, partially-connected structure for hybrid beamforming is considered in [7], in which the penalty-based algorithm and RCG-based algorithm are proposed. Besides, the works in [8], [9] solve the mean square error (MSE) minimization problem with accelerated RCG method and gradient projection (GP) method.

Different from the previous works, in this paper we consider the sum rate maximization problem of hybrid beamforming in RIS-aided MU-MIMO systems. In particular, we derive the bounds of hybrid beamforming in RIS-aided MU-MIMO systems. Based on it, an approximated problem of RIS configuration is proposed, the solution of which can approach the upper bound of sum rate by phase elimination. Besides, the analog beamforming is jointly designed with phase elimination and the digital beamforming is designed with minimum mean square error (MMSE) beamforming. Based on the alternating optimization (AO) scheme, the hybrid beamforming and RIS configuration are alternately optimized, which leads to the proposed AO algorithm based on phase elimination (AO-PE). Finally, simulation results illustrate the advantage of the proposed AO-PE.

Corresponding author: Zheng Wang (email: wznuaa@gmail.com)

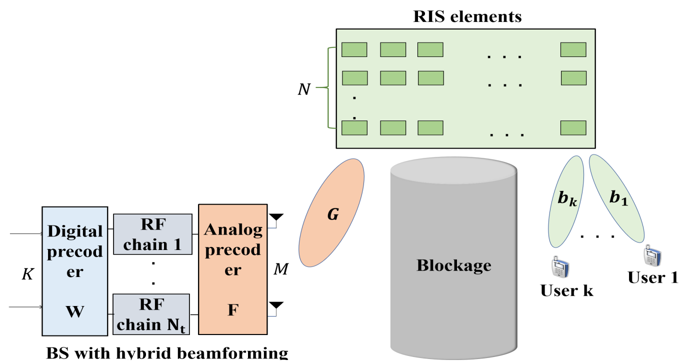

Fig. 1. System diagram of RIS-aided multi-user downlink.

II. System Model

We consider a RIS-assisted communication system, where the base station (BS) equipped with M transmit antennas and N_t RF chains serves K single-antenna users through N reflecting elements of RIS. The system architecture is shown in Fig. 1. The BS employs hybrid beamforming structure with analog beamforming matrix $\mathbf{F} = [\mathbf{f}_1, \cdots, \mathbf{f}_{N_t}] \in \mathbb{C}^{M\times N_t}$ and digital beamforming matrix $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_K] \in \mathbb{C}^{N_t\times K}$ with $M > N_t \geqslant K$. Each RF chain is connected to all M transmit antennas through phase shifters, which leads to the unit modulus constraint on $\mathbf{F}$ as $\mathbf{F} \in \mathcal{F} = \{\mathbf{F}||f_{m,t}| = 1, \forall m, t\}$. Furthermore, the digital beamforming matrix $\mathbf{W}$ should satisfy $||\mathbf{FW}||_F^2 = P$, where P is the transmit power constraint. Typically, the signal transmitted from BS antennas can be given by $\mathbf{x} = \mathbf{FWs} = \sum_k \mathbf{Fw}_k s_k$, where $\mathbf{s} = [s_1, \cdots, s_K]^H \in \mathbb{C}^{K\times 1}$ and s_k denotes the transmitted data symbol for user k with $\mathbb{E}[|s_k|^2] = 1$.

The RIS with N elements is deployed between the BS and users, characterized by a diagonal matrix $\mathbf{\Theta} = \mathrm{diag}(\mathbf{r}) \in \mathbb{C}^{N\times N}$, where $\mathbf{r} \in \mathcal{A} = \{\mathbf{r}||r_n| = 1, \forall n\}$. Assuming that the direct paths between the BS and users are blocked due to obstacle, the received signal for the k-th user is obtained as:

$$y_k = \mathbf{b}_k\mathbf{\Theta}\mathbf{G}\mathbf{F}\mathbf{w}_k s_k + \sum_{j\neq k}\mathbf{b}_k\mathbf{\Theta}\mathbf{G}\mathbf{F}\mathbf{w}_j s_j + n_k, \qquad (1)$$

where $n_k \sim \mathcal{CN}(0,\sigma^2)$ represents the additive white Gaussian noise. $\mathbf{b}_k \in \mathbb{C}^{1\times N}$, $\mathbf{G} \in \mathbb{C}^{N\times M}$ are Rayleigh fading channel vector from the RIS to user k and channel matrix from the BS to RIS, the elements of which are denoted as $b_{k,n} \sim \mathcal{CN}(0,b^2), g_{n,m} \sim \mathcal{CN}(0,g^2), \forall n,m$, respectively.

For simplicity, we denote $\mathbf{r}\mathbf{H}_k = \mathbf{b}_k\mathbf{\Theta}\mathbf{G}$, where $\mathbf{r} = \mathrm{diag}(\mathbf{\Theta}) \in \mathbb{C}^{1\times N}$ and $\mathbf{H}_k = \mathrm{diag}(\mathbf{b}_k)\mathbf{G} \in \mathbb{C}^{N\times M}$ is the combined RIS reflection channel for user k. According to (1), the achievable sum rate of user k can be formulated as:

$$R_k = \log_2\left(1 + \frac{|\mathbf{r}\mathbf{H}_k\mathbf{F}\mathbf{w}_k|^2}{\sum_{j\neq k}|\mathbf{r}\mathbf{H}_k\mathbf{F}\mathbf{w}_j|^2+\sigma^2}\right), \qquad (2)$$

such that the hybrid beamforming optimization problem for maximizing the sum rate $R = \sum_{k=1}^{K} R_k$ in RIS-aided communication systems can be given by:

$$\max_{\mathbf{r},\mathbf{F},\mathbf{w}_k} \quad \sum_{k=1}^{K} R_k \qquad (3a)$$

$$\text{s.t.} \quad |r_n| = 1, \quad \forall n, \qquad (3b)$$

$$|f_{m,t}| = 1, \quad \forall m,k, \qquad (3c)$$

$$\sum_{k=1}^{K}\|\mathbf{F}\mathbf{w}_k\|^2 = P, \qquad (3d)$$

where (3b) and (3c) represent the modulus constraints for the RIS elements and the analog beamforming matrix, (3d) is the total transmit power constraint.

III. Joint RIS Phase Shift and Hybrid Beamforming Design

A. Theory Analysis

To start with, regarding to the problem in (3), we first give the theoretical sum rate bounds of hybrid beamforming in RIS-aided MU-MIMO communication shown as *Theorem 1* and *Theorem 2*.

Theorem 1. *For hybrid beamforming in RIS-aided MU-MIMO systems, the sum rate R is upper bounded by*

$$\lim_{M\to\infty} R < K\log_2\left(1+\frac{\pi^2 PMN^2b^2g^2}{16\sigma^2 K}\right), \qquad (4)$$

if M tends to infinity.

Proof. First of all, we claim that when M tends to infinity, user channels are asymptotically orthogonal in MU-MIMO systems [10]. Based on it, the inter-user interference can converge to zero with zero forcing (ZF) digital beamforming. Then the sum rate shown in (3) becomes $\sum_{k=1}^{K}\log_2(1+\frac{1}{\sigma^2}|\mathbf{r}\mathbf{H}_k\mathbf{F}\mathbf{w}_k|^2)$. Furthermore, the asymptotically orthogonal property of user channels indicates that the ZF beamforming converges to conjugate beamforming [11] such that

$$|\mathbf{r}\mathbf{H}_k\mathbf{F}\mathbf{w}_k|^2 \approx ||\mathbf{r}\mathbf{H}_k\mathbf{F}||^2||\mathbf{w}_k||^2. \qquad (5)$$

In the spirit of [12], $\mathbf{F}^H\mathbf{F} \approx M\mathbf{I}$ is satisfied with $M \gg N_t$. With that, according to (3d), we have $\mathbb{E}[||\mathbf{w}_k||^2] = \frac{P}{KM}$.

Next, designing the analog beamforming matrix $\mathbf{F}$ by phase elimination, $|\mathbf{r}\mathbf{H}_k\mathbf{f}_t|^2, \forall t \neq k$ is negligible compared to $|\mathbf{r}\mathbf{H}_k\mathbf{f}_k|^2$ for user k at large M [11]. With that, $||\mathbf{r}\mathbf{H}_k\mathbf{F}||^2 = \sum_{t=1}^{N_t}|\mathbf{r}\mathbf{H}_k\mathbf{f}_t|^2$ can be written as $|\mathbf{r}\mathbf{H}_k\mathbf{f}_k|^2$. Denoting $f_{m,k}$, r_n and the elements of $\mathbf{H}_k$ as $e^{j\phi_{m,k}}$, $e^{j\theta_n}$ and $|h_{k,n,m}|e^{j\gamma_{k,n,m}}$, for user k, we can obtain

$$|\mathbf{r}\mathbf{H}_k\mathbf{f}_k| = \sum_{m=1}^{M}\sum_{n=1}^{N}|h_{k,n,m}|e^{j(\gamma_{k,n,m}+\theta_n+\phi_{m,k})}. \qquad (6)$$

Apparently, $|\mathbf{r}\mathbf{H}_k\mathbf{f}_k|$ can be maximized only if $\gamma_{k,n,m}+\theta_n+\phi_{m,k} = 0, \forall n,m,k$. However, when $N,M > 1$, the relation $KMN > KM+N$ holds, i.e., the number of $\gamma_{k,n,m}$ is larger than that of $\theta_n+\phi_{m,k}$. To this end, $\gamma_{k,n,m}+\theta_n+\phi_{m,k} = 0$ can not be satisfied for $\forall n,m,k$, and this case can be served as the upper bound:

$$|\mathbf{r}\mathbf{H}_k\mathbf{f}_k| < \sum_{m=1}^{M}\sum_{n=1}^{N}|h_{k,n,m}|. \qquad (7)$$

Since $b_{k,n} \sim \mathcal{CN}(0,b^2)$ and $g_{n,m} \sim \mathcal{CN}(0,g^2)$ are independent from each other, $|h_{k,n,m}| = |b_{k,n}||g_{n,m}|$ has mean value $\frac{\pi bg}{4}$. According to *central limit theorem*, $\sum_{m=1}^{M}\sum_{n=1}^{N}|h_{k,n,m}|$ follows Gaussian distribution with mean value $\frac{\pi MNbg}{4}$. Then the upper bound of channel power $|\mathbf{r}\mathbf{H}_k\mathbf{F}\mathbf{w}_k|^2$ is given by:

$$|\mathbf{r}\mathbf{H}_k\mathbf{F}\mathbf{w}_k|^2 < \frac{\pi^2 PMN^2b^2g^2}{16K}, \qquad (8)$$

which completes the proof. □

Similar to (8) in *Theorem 1*, we can easily derive the upper bound of channel power in fully digital (FD) RIS-aided MU-MIMO systems as $\frac{\pi PMN^2b^2g^2}{4K}$. It indicates that hybrid beamforming structure can achieve $\frac{\pi}{4}$ channel power gain of which in FD structure. Nevertheless, hybrid beamforming has advantage on hardware cost reduction. Consequently, the hybrid beamforming structure is promising for RIS-aided MU-MIMO systems.

In *Theorem 1*, the upper bound derived is based on the assumption that $\gamma_{k,n,m}+\theta_n+\phi_{m,k} = 0$ holds for $\forall n,m,k$, which can not be fulfilled in fact as discussed before. In order to investigate the impact of phase shifts on sum rate, we derive the lower bound in *Theorem 2*.

Theorem 2. *For hybrid beamforming in RIS-aided MU-MIMO systems, the sum rate is lower bounded by*

$$\lim_{M\to\infty} R > K\log_2\left(1+\frac{\pi P\left(\sqrt{MN-M-\frac{N}{K}}+\frac{\sqrt{\pi}}{2}(M+\frac{N}{K})\right)^2 b^2g^2}{4\sigma^2 MK}\right), \qquad (9)$$

if M tends to infinity.

Proof. As discussed before, the number of adjustable phases is $KM+N$. For each user, denoting $h_i = h_{k,n,m}, \forall i \in \{1,\cdots,MN\}$, we note that only the phases of $M+\frac{N}{K}$ terms of h_i can be completely eliminated. In the case of only considering phase elimination of $M+\frac{N}{K}$ terms, the other $MN-M-\frac{N}{K}$ terms are random phases. We claim that by properly optimizing $\{\theta_n, \phi_{m,k}, \forall n,m,k\}$, the value of $|\mathbf{r}\mathbf{H}_k\mathbf{f}_k|$ in (6) can be further improved compared to this case. In other words, it can be viewed as the lower bound shown as:

$$|\mathbf{r}\mathbf{H}_k\mathbf{f}_k| > \sum_{i=1}^{M+\frac{N}{K}} |h_i| + \left| \sum_{i=1}^{MN-M-\frac{N}{K}} h_i e^{j\psi_i} \right|, \quad (10)$$

where $\psi_i \in [0,2\pi)$ is random phase shift.

Since $b_{k,n} \sim \mathcal{CN}(0,b^2)$ and $g_{n,m} \sim \mathcal{CN}(0,g^2)$ are independent from each other, we have $h_i = h_{k,n,m} = b_{k,n}g_{n,m} \sim \mathcal{CN}(0,b^2g^2)$. As $h_i e^{j\psi_i}$ has the same distribution as h_i[3], $\left|\sum_{i=1}^{MN-M-\frac{N}{K}} h_i e^{j\psi_i}\right|$ has mean value $\frac{\sqrt{\pi\left(MN-M-\frac{N}{K}\right)}bg}{2}$. According to *central limit theorem*, $\sum_{i=1}^{M+\frac{N}{K}} |h_i|$ follows Gaussian distribution with mean value $\frac{\pi\left(M+\frac{N}{K}\right)bg}{4}$. Then the lower bound of channel power $|\mathbf{r}\mathbf{H}_k\mathbf{F}\mathbf{w}_k|^2$ is given by:

$$|\mathbf{r}\mathbf{H}_k\mathbf{F}\mathbf{w}_k|^2 > \frac{\pi P\left(\sqrt{MN-M-\frac{N}{K}} + \frac{\sqrt{\pi}}{2}(M+\frac{N}{K})\right)^2 b^2g^2}{4MK}, \quad (11)$$

which completes the proof. □

From (8) and (11), we can conclude that for hybrid beamforming in RIS-aided MU-MIMO systems, optimizing hybrid beamforming and RIS configuration can bring linear gain with transmit antenna number M and square gain with RIS element number N on channel power at large M and N. This conclusion expands the power scaling law in [4] to hybrid beamforming in RIS-aided MU-MIMO systems.

B. AO Scheme For Hybrid Beamforming and RIS Configuration

We claim that the sum rate upper bound derived in *Theorem 1* is based on the optimized choices of hybrid beamforming and RIS configuration. Put it in another way, approaching the upper bound of sum rate chiefly lies on the phase elimination by hybrid beamforming and RIS configuration. In this section, we alternately optimize $\{\mathbf{r},\mathbf{F},\mathbf{W}\}$ to approximately solve the problem in (3) in a better way.

With $\{\mathbf{F},\mathbf{W}\}$ fixed, we first ignore the inter-user interference in (3) and attempt to configure RIS by solving the problem shown as:

$$\begin{aligned} \max_{\mathbf{r}} \quad & \sum_{k=1}^{K} \log_2\left(1+\frac{1}{\sigma^2}|\mathbf{r}\mathbf{H}_k\mathbf{F}\mathbf{w}_k|^2\right) \\ \text{s.t.} \quad & |r_n| = 1, \quad \forall n. \end{aligned} \quad (12)$$

Denoting $\mathbf{m}_k = \mathbf{H}_k\mathbf{F}\mathbf{w}_k \in \mathbb{C}^{N\times 1}$, we have

$$\mathbf{r}\mathbf{H}_k\mathbf{F}\mathbf{w}_k = \mathbf{r}\mathbf{m}_k = \sum_{n=1}^{N} |m_{n,k}| e^{j(\omega_{n,k}+\theta_n)}, \quad (13)$$

where the n-th element of $\mathbf{m}_k$ is $|m_{n,k}|e^{j\omega_{n,k}}$. Clearly, $\mathbf{r}\mathbf{m}_k$ can be maximized only when $\omega_{n,k}+\theta_n = 0, \forall n$. To be more specific, $|\mathbf{r}\mathbf{m}_k|$ aims to be:

$$n_k = \sum_{n=1}^{N} |m_{n,k}|. \quad (14)$$

However, since θ_n can not be arranged to satisfy $\omega_{n,k}+\theta_n = 0$ for all $k \in \{1,\cdots,K\}$, only few items of them can be fulfilled at the same time. As shown in *Theorem 1* and *Theorem 2*, to approach the sum rate upper bound as possible, we can arrange the RIS phase shifts through proper phase elimination. Motivated by this, to minimize the difference between $|\mathbf{r}\mathbf{m}_k|$ and n_k for $\forall k$, we approximate the problem (12) by the following formula:

$$\begin{aligned} \min_{\mathbf{r},\xi_k} \quad & \sum_{k=1}^{K} \log_2\left(1+\frac{1}{\sigma^2}|n_k e^{j\xi_k} - \mathbf{r}\mathbf{m}_k|^2\right) \\ \text{s.t.} \quad & |r_n| = 1, \quad \forall n, \\ & \xi_k \in [0,2\pi), \forall k. \end{aligned} \quad (15)$$

Here, $\xi_k, \forall k$ is introduced auxiliary variable to increase degree of freedom of phase shift of $\mathbf{r}$ without influencing the modulus value of n_k, i.e., $|n_k e^{j\xi_k}| = n_k$.

To solve (15), we iteratively update $\{\mathbf{r}, \xi_k, \forall k\}$. On one hand, when updating $\mathbf{r}$, the computationally efficient GP method [13] can be applied as follows:

$$\begin{aligned} \nabla f(\mathbf{r}) &= -\frac{2}{\sigma^2 \ln 2} \sum_{k=1}^{K} \frac{n_k e^{j\xi_k} - \mathbf{r}\mathbf{m}_k}{1+\frac{1}{\sigma^2}|n_k e^{j\xi_k} - \mathbf{r}\mathbf{m}_k|^2} \mathbf{m}_k^H, \\ \mathbf{r} &= e^{j\angle(\mathbf{r}-\mu\nabla f(\mathbf{r}))}. \end{aligned} \quad (16)$$

On the other hand, when updating $\{\xi_k, \forall k\}$, the GP method is given by:

$$\begin{aligned} \nabla f(e^{j\xi_k}) &= -\frac{2n_k|n_k e^{j\xi_k} - \mathbf{r}\mathbf{m}_k|}{\sigma^2 \ln 2\left(1+\frac{1}{\sigma^2}|n_k e^{j\xi_k} - \mathbf{r}\mathbf{m}_k|^2\right)}, \\ \xi_k &= e^{j\angle\left(\xi_k - \lambda\nabla f(e^{j\xi_k})\right)}. \end{aligned} \quad (17)$$

Here, μ and λ are the step sizes given by *Armijo rule*. $\mathbf{r}$ and $\xi_k, \forall k$ are iteratively optimized until the objective function in (15) converges.

Next, with $\mathbf{r}$ fixed, we arrange the analog beamforming matrix $\mathbf{F}$ as:

$$f_{m,t} = e^{j\phi_{m,k}}, \quad (18)$$

where $\phi_{m,k}$ is the phase of the m-th element of $(\mathbf{r}\mathbf{H}_k)^H$.

Lastly, with $\{\mathbf{r},\mathbf{F}\}$ fixed, in order to eliminate inter-user interference, we apply MMSE beamforming to configure the digital beamforming matrix $\mathbf{W}$, given by:

$$\mathbf{W} = \left(\mathbf{H}_{eq}^H\mathbf{H}_{eq} + \frac{\sigma^2}{P}\mathbf{I}_{N_t}\right)^{-1}\mathbf{H}_{eq}^H, \quad (19)$$

where $\mathbf{H}_{eq} \in \mathbb{C}^{K\times N_t}$, the k-th row of which is $\mathbf{r}\mathbf{H}_k\mathbf{F}$.

To summarize, the proposed AO-PE algorithm for hybrid beamforming and RIS configuration is shown in *Algorithm 1*. Now, we analyze its computational complexity. On one hand, when optimizing $\mathbf{r}$, the complexity contains the computation of

Algorithm 1 Proposed AO-PE algorithm

```
Input: {G, b_k, ∀k}
Initialize {r, F, W, ξ_k, ∀k}.
repeat
    while convergence is not met do
        compute r using GDP shown in (16).
        compute {ξ_k, ∀k} using GDP shown in (17).
    end while
    arrange F with (18) when r is fixed.
    arrange W with (19) when {r, F} is fixed.
    normalize W so that ||FW||_F^2 = P.
until objective function in (15) converges
Output: {r, F, W}
```

$\{n_k, \mathbf{m}_k, \forall k\}$, and GP iteration in (16) and (17). Its computation complexity is given by $\mathcal{O}\left(K(3R_G N + M(N+N_t)+N)\right)$. On the other hand, when optimizing $\mathbf{F}$, (18) includes the computation of $\mathbf{r}\mathbf{H}_k, \forall k$, the complexity of which is given by $\mathcal{O}\left(KMN\right)$. When optimizing $\mathbf{W}$, MMSE beamforming in (19) has the computation of $\mathcal{O}\left(KMN_t + N_t^3\right)$. Due to the hybrid beamforming matrices and RIS configuration are alternately optimized, the total complexity of proposed algorithm is given by $\mathcal{O}\left(R_A(K(3R_G N + 2M(N+N_t)+N)+N_t^3)\right)$. *Table 1* lists the detail of computational complexities of proposed AO-PE and various algorithms, which reveals that the proposed AO-PE algorithm can converge rapidly and has reduced complexity.

IV. Simulation Results

In this section, we evaluate the sum rate achieved by the proposed AO-PE with different system settings. In order to compare the proposed AO-PE with other algorithms, we consider the following benchmark schemes: 1) FD structure at the BS with MMSE beamforming and designing RIS by solving (16) and (17). 2) AO scheme in [5] with hybrid beamforming designed by (18) and (19). 3) Alternating minimization (AM) scheme for hybrid beamforming and RIS in [8]. 4) Hybrid precoding and RIS phase shifts (HP-PSs) design in [9].

For Rayleigh fading channels, we set the variances $b^2 = g^2 = 1$. The transmitter power and receiver noise are set to be 20dBm and 10dBm. In Fig. 2 and Fig. 3, the transmit antenna and the RIS element numbers are varied to investigate the performance of proposed algorithm in Rayleigh fading channels. We can claim that regardless of the sizes of transmit antenna array and RIS, the proposed AO-PE always attain better performance compared to other hybrid beamforming algorithms. Note that though the FD architecture achieves slightly higher sum rate, it suffers expensive RF chain hardware cost with $N_t = M$.

Moreover, the upper bound and the lower bound are also illustrated in Fig. 2 and Fig. 3. It can be demonstrated that extending the sizes of transmit antenna and RIS brings linear gain and square gain on channel power, respectively. Then we can conclude that compared to adding transmit antenna, deploying the RIS not only can achieve lower hardware cost [8], but also can obtain significant gain for the system rate.

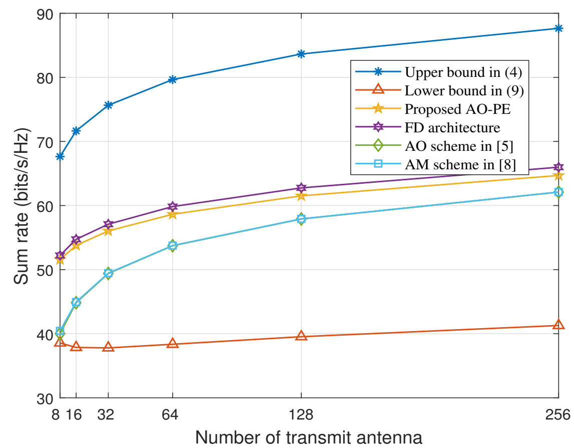

Fig. 2. Sum rate with the increasing number of transmit antenna for MU-MIMO Rayleigh fading systems with N = 100, $N_t = K$ = 4.

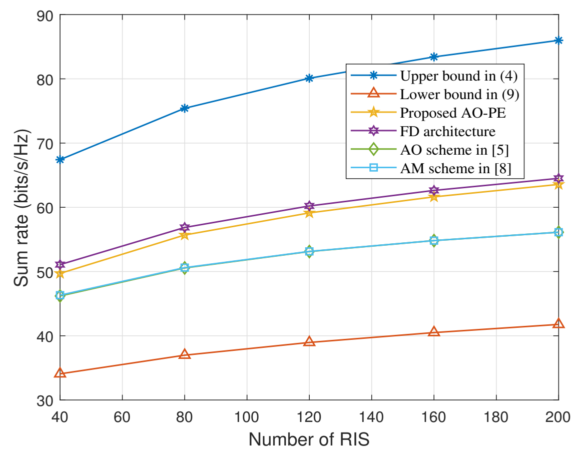

Fig. 3. Sum rate with the increasing number of RIS element for MU-MIMO Rayleigh fading systems with M = 48, $N_t = K$ = 4.

Besides, we explore the performance of proposed algorithm in mmWave channels, characterized by the Saleh-Valenzuela model [14] as:

$$\mathbf{G} = \sqrt{\frac{MN}{L_G}} \sum_{l=1}^{L_G} \alpha_l \boldsymbol{a}_{\mathrm{p}}\left(\phi_l^r, \theta_l^r\right) \boldsymbol{a}_l\left(\phi_l^b\right)^H,$$
$$\mathbf{b}_k = \sqrt{\frac{N}{L_B}} \sum_{l=1}^{L_B} \beta_{k,l} \boldsymbol{a}_{\mathrm{p}}\left(\phi_{k,l}^r, \theta_{k,l}^r\right), \forall k, \tag{20}$$

where $\alpha_l, \beta_{k,l} \sim \mathcal{CN}(0, \sigma_a^2)$. Here we set $L_G = L_B = 5, \sigma_a^2 = 1$. Moreover, we consider the path loss as $PL(d) = \beta_0 d^{-\alpha_p}$ [8] and $\beta_0 = -30dB, \alpha_p^{BR} = 2.1, \alpha_p^{RU} = 2.4$. The coordinate system with the BS and the RIS located at (0, 0)m and $(25\sqrt{2}, 25\sqrt{2})$m are introduced. Half of users are distributed randomly within a quarter-circle centered at the BS with 50m radius, and the other users are located evenly on a half circle centered at the RIS with 3m radius [8]. The transmitter power and receiver noise is set to be 20dBm and -100dBm.

TABLE I
THE COMPUTATIONAL COMPLEXITIES OF PROPOSED AO-PE AND OTHER ALGORITHMS($K = 4, M = 48, N = 100, N_t = 4$)

	Overall complexity	R_G	R_A	Complexity value
Proposed AO-PE	$\mathcal{O}\left(R_A(K(3R_GN + 2M(N + N_t) + N) + N_t^3)\right)$	3	5	220000
AO scheme in [5]	$\mathcal{O}\left(R_A(KMN + KMN_t + R_GK^2N^2 + N_t^3)\right)$	3	5	2500160
AM scheme in [8]	$\mathcal{O}\left(R_A(KN_t^3 + KM^2 + 2KMN + KMN_t)\right)$	–	5	243200

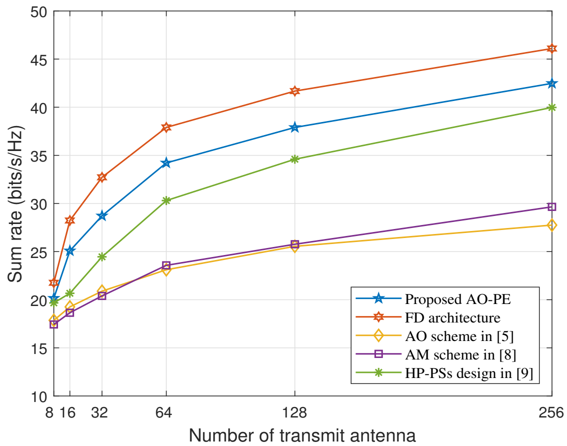

Fig. 4. Sum rate with the increasing number of transmit antenna for MU-MIMO mmWave systems with N = 100, $N_t = K$ = 4.

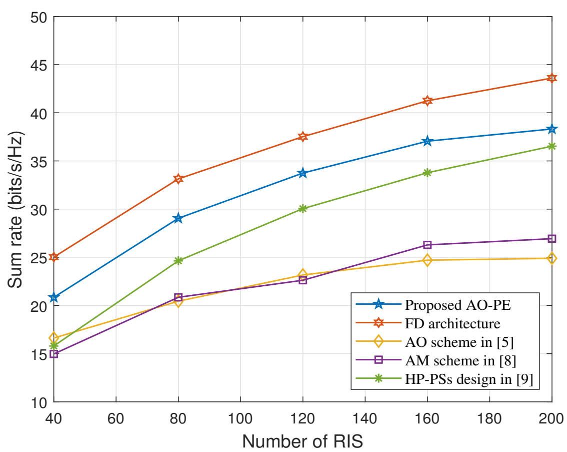

Fig. 5. Sum rate with the increasing number of RIS element for MU-MIMO mmWave systems with M = 48, $N_t = K$ = 4.

Fig. 4 and Fig. 5 show the behaviors of different algorithms in mmWave channels with various sizes of transmit antenna array and RIS. We confirm that the proposed AO-PE can always perform better than other algorithms due to the ability to obtain channel power gain.

V. CONCLUSION

In this paper we focus on hybrid beamforming and RIS design in an alternate way to maximize sum rate in MU-MIMO systems. Specifically, we derive the theoretical system sum rate bounds. According to it, we propose novel problem formulation of the RIS design to approach the upper bound, which can be solved with GP method. The analog beamforming is jointly designed by phase elimination to obtain the power gain while the digital beamforming is designed to eliminate interference.

ACKNOWLEDGMENT

This work was supported in part by National Science and Technology Major Projects of China under Grant 2025ZD1301800.

REFERENCES

[1] Y. Gao, Q. Chen, L. He, *et al.*, “Artificial Intelligence Enabled Joint Channel Estimation and Signal Detection for Massive MIMO Systems,” *Chin. J. Electron.*, vol. 35, no. 1, pp. 178-195, 2026.

[2] C. Huang et al., “Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends,” *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118-125, Oct. 2020.

[3] 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, Jan. 2020.

[4] Q. Wu and R. Zhang, “Intelligent Reflecting Surface Enhanced Wireless Network via Joint Active and Passive Beamforming,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394-5409, Nov. 2019.

[5] H. Guo, Y. -C. Liang, J. Chen, and E. G. Larsson, “Weighted Sum-Rate Maximization for Reconfigurable Intelligent Surface Aided Wireless Networks,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3064-3076, May 2020.

[6] P. Wang, J. Fang, L. Dai, and H. Li, “Joint Transceiver and Large Intelligent Surface Design for Massive MIMO mmWave Systems,” *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1052-1064, Feb. 2021.

[7] R. Li, B. Guo, M. Tao, Y. -F. Liu, and W. Yu, “Joint Design of Hybrid Beamforming and Reflection Coefficients in RIS-Aided mmWave MIMO Systems,” *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2404-2416, 2022.

[8] S. Gong, C. Xing, P. Yue, L. Zhao, and T. Q. S. Quek, “Hybrid Analog and Digital Beamforming for RIS-Assisted mmWave Communications,” *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1537-1554, Mar. 2023.

[9] C. Pradhan, A. Li, L. Song, B. Vucetic, and Y. Li, “Hybrid Precoding Design for Reconfigurable Intelligent Surface Aided mmWave Communication Systems,” *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 1041-1045, Jul. 2020.

[10] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[11] L. Liang, W. Xu, and X. Dong, “Low-complexity hybrid precoding in massive multiuser MIMO systems,” *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 653-656, Dec. 2014.

[12] F. Sohrabi, W. Yu, “Hybrid Digital and Analog Beamforming Design for Large-Scale Antenna Arrays,” *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501-513, Apr. 2016.

[13] Y. Ma, B. Yan, Z. Wang and Y. Huang, “Dynamic Metasurface Antennas-Based Beamforming for Holographic MIMO Systems,“ in *25th International Conference on Communication Technology (ICCT)*, 2025, pp. 1251-1255.

[14] Y. Ma, Q. Chen, B. Yan, Z. Wang and L. He, “Efficient Joint Hybrid Precoding And Analog Combining Scheme For Massive MIMO Systems,“ in *2024 International Wireless Communications and Mobile Computing (IWCMC)*, 2024, pp. 1455-1459.