Joint Power Allocation and User Association Optimization for IRS-Assisted mmWave Systems

Dan Zhao, Hancheng Lu, Yazheng Wang, Huan Sun, Yongqiang Gui

Abstract

Intelligent reflect surface (IRS) is a potential technology to build programmable wireless environment in future communication systems. In this paper, we consider an IRS-assisted multi-base station (multi-BS) multi-user millimeter wave (mmWave) downlink communication system, exploiting IRS to extend mmWave signal coverage to blind spots. Considering the impact of IRS on user association in multi-BS mmWave systems, we formulate a sum rate maximization problem by jointly optimizing passive beamforming at IRS, power allocation and user association. This leads to an intractable non-convex problem, for which to tackle we propose a computationally affordable iterative algorithm, capitalizing on alternating optimization, sequential fractional programming (SFP) and forward-reverse auction (FRA). In particular, passive beamforming at IRS is optimized by utilizing the SFP method, power allocation is solved through means of standard convex optimization method, and user association is handled by the network optimization based FRA algorithm. Simulation results demonstrate that the proposed algorithm can achieve significant performance gains, e.g., it can provide up to 175% higher sum rate compared with the benchmark and 140% higher energy efficiency compared with amplify-and-forward relay.

Index Terms

H. Lu, D. Zhao, Y. Wang, Y. Gui are with the Information Network Laboratory, Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China (e-mail: hclu@ustc.edu.cn; {zd2019, wang1997, yongqgui}@mail.ustc.edu.cn). Huan Sun is with the Wireless Technology Laboratory, Huawei Technologies Co.,Ltd., Shanghai, China (e-mail: sunhuan11@huawei.com).

Intelligent reflect surface, millimeter wave, user association, beamforming, power allocation, auction algorithm.

I. INTRODUCTION

The fifth generation mobile networks use key technologies such as millimeter wave (mmWave) technology, heterogeneous networks and massive multiple-input multiple-output (MIMO) to cope with increasing demand for higher data rates [1], [2]. As a promising technology, mmWave communication, with frequency ranging from $30\sim300$GHz, has the potential to provide gigabits-per-second communication data rates while supporting a wide range of applications [3]. However, due to the high carrier frequency, mmWave suffers more significant path loss compared with the path attenuation over the lower frequency bands. On the other hand, the short wavelength and high directivity cause that mmWave signal are difficult to pass through obstacles, which is highly susceptible to blocking. In order to overcome the blockage effect in mmWave systems, thereby improving the coverage of mmWave signal and maintaining robustness of the mmWave systems, some research work have made lots of attempts, such as employing relay [4]–[6], strategically placed reflectors [7], the reflections from walls and other surfaces [8] to overcome obstacles in mmWave systems.

Recently, intelligent reflect surface (IRS) becomes a new promising technology to improve the coverage enhancement when the line-of-sight (LOS) link is severely blocked by obstacles in mmWave systems [9]–[15]. IRS can intelligently reconfigure wireless propagation environment with a large number of passive reflect elements (e.g., low-cost printed dipoles) integrated on a planar surface. Specifically, each element of IRS receives a superimposed multipath signal from the transmitter and then independently reflects the incident signal by controlling the amplitude and/or phase to achieve passive beamforming (PBF). By smartly turning the phase shifts of the passive elements, the reflected signal can be added coherently at the desired receiver to create effective virtual LOS link while achieving directional signal enhancement, or destructively at non-intended receivers to suppress interference [16]. In addition, compared with relay, such as full-duplex amplify-and-forward (AF) realy, IRS is not equipped with a dedicated power source and an amplifier, which only uses PBF to forward signal, thus the power consumption and hardware costs of IRS are much lower than those of relay [17].

Many efforts have been dedicated to the research of IRS in mmWave systems [9]–[15]. A key problem for IRS-assited mmWave systems is to jointly optimize active beamforming at

base station (BS) and PBF at IRS to achieve different objectives and improve communication performance, such as optimization of channel capacity [9], received signal power [10], [11] and weighted sum rate (WSR) [12]. In [9], the authors firstly presented non-line-of-sight (NLOS) path, then proposed two optimization schemes utilizing IRS to maximize the channel capacity in an indoor MIMO mmWave system. More precisely, one exploits only the customizing capabilities of the IRS elements, while the other jointly optimizes PBF at IRS and transmits active beamforming, which can provide up to 200% channel capacity compared with basic MIMO system. A problem of maximizing the received signal power was investigated in [10], [11] for an multi-IRS assisted multiple input single output (MISO) system, and simulation results demonstrate that both the single IRS and multi-IRS can help to create effective virtual LOS paths and thus substantially improve robustness against blockages in mmWave systems. Considering the discrete phase shifts of IRS, the work in [12] proposed an alternate optimization algorithm to maximize the WSR of all users in an IRS-assisted multi-user MISO system. Another key issue is to design an efficient deployment strategy of IRS to establish robust mmWave connections for mobile mmWave systems [14], [15]. In [14], the authors developed a two-stage beam-searching algorithm and analyzed the reflect-array deployment to adequately utilize the smart reflectarray, thereby solving the problem of signal blockage in mmWave indoor communications. Based on Q-learning and neural networks, a reinforcement learning approach is proposed in [15] to maximize the downlink transmission capacity, where the location and reflection coefficient of unmanned aerial vehicle-carried intelligent reflector are jointly optimized.

However, existing work on IRS mostly focus on system performance optimization in a single BS scenario [9]–[12], [14]–[16], [18]–[23] and the multi-BS scenario has not been studied before. After introducing IRS in a multi-BS scenario, the channel reconfiguration brought by IRS can seriously affect the existing power allocation and user association (UA) algorithms [24]–[30]. Specifically, the channels involved by IRS are cascaded channels and related to the IRS reflection matrix, which leads to the coupling of the IRS reflection matrix with power allocation and UA. Thus, we have to redesign an effective PBF, power allocation and UA algorithm to enhance the performance of multi-BS mmWave systems. Furthermore, because of the erratic nature of mmWave propagation, the mmWave channel varies fastly and is susceptible to blockage effect, causing dramatic block of the communication on a short time scale. All the mentioned mmWave channel characteristics make the optimization of power allocation and UA much more complicated and challenging, and the use of existing power allocation and UA solvers [24]–[30]

cannot be applied to actual mmWave systems. Besides, the reflective elements of IRS usually shift the incident signal with limited discrete phase shifters for the ease of circuit implementation, and this further complicates the joint optimization problem due to the non-convex constraint of IRS. Nevertheless, lots of work related to IRS assume that each element of IRS is a continuous phase shifter [10], [18]–[20], which is difficult to implement in practice.

To tackle aforementioned issues, in this paper, we study the joint PBF at IRS (i.e., the design of IRS reflection matrix), power allocation and UA optimization problem in IRS-assisted multi-BS mmWave systems. To the best of our knowledge, this is the first attempt to investigate IRS-assisted multi-BS communication. It is worth noting that the preliminary design and results have been published in a conference version [13]. Different from the previous work, this paper considers power allocation to further enhance system performance, and provides more comprehensive analyses of the performance of IRS to expand the coverage of mmWave signal. By exploiting the benefits from IRS, power allocation and UA, our designed mmWave system can achieve high robust, low complex, and cost-effective communication. Our main contributions are summarized as follows:

1) We propose an IRS-assisted multi-BS multi-user mmWave system, where an IRS with discrete phase shifter is deployed to assist downlink MISO transmission. Based on the proposed system, we jointly take the IRS reflection matrix and transmit power into account so as to formulate the sum rate maximization problem more practically.

2) Given the non-convexity and complexity of the problem, we propose an iterative PBF, power allocation and UA (IPPU) algorithm based on alternating optimization technique to solve it. Particularly, the original problem is decomposed into three subproblems, including the non-convex subproblems of the IRS reflection matrix optimizationn and UA optimization, and the convex subproblem of optimizing power allocation.

3) It is worth noting that the non-convex subproblems of IRS optimization and UA optimization are very challenging, and we use two effective methods to solve them separately. For the optimization of IRS, we first deduce that the objective function has an upper bound and it satisfies the theoretical requirements of the sequential fractional programming (SFP) method. Then, we use the monotonic improvement properties of SFP to effectively find the approximate optimal solution of the IRS. For the optimization of UA, we propose a forward-reverse auction (FRA) algorithm to effectively solve this asymmetric multi-assignment problem. In particular, we prove that the sum rate of the obtained solution is strictly within 1 of being optimal.

4) We have further analyzed the performance of the proposed algorithm and compared it with different comparison algorithms. Our results show that the proposed algorithm can achieve significant gains in terms of sum rate and energy efficiency (EE). Specificilly, it can achieve 175% sum rate improvement compared with the benchmark and can also obtain 140% higher energy efficiency improvement compared with AF relay. Compared with the system assisted by AF relay and the system without IRS, the proposed algorithm has higher coverage and lower outage probability for communication.

The rest of this paper is organized as follows. In Section II, we start with the system and channel model, and then formulate a sum rate maximization problem. In Section III, an IPPU algorithm based on alternating optimization is designed to solve the problem. Simulation results of the proposed algorithm are discussed in Section IV. Finally, Section V concludes this paper and gives future research directions.

In this paper, we use italic letters to denote scalars, vectors and matrices are presented by bold-face lower-case and upper-case letters. For any general matrix $\mathbf{A}$, $a_{i,j}$ is the i-th row and j-th column element. $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$, $\mathbf{A}^+$ and $\|\mathbf{A}\|_F$ represent the transpose, Hermitian, inverse, pseudo-inverse and Frobenius norm of $\mathbf{A}$, respectively. $\text{Re}\{\cdot\}$, $|\cdot|$, $\arg(\cdot)$ and $\text{tr}(\cdot)$ denote the real part, modulus, the angle of a complex vector and the trace of a matrix, respectively. $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker products of $\mathbf{A}$ and $\mathbf{B}$, while $\text{vec}(\mathbf{A})$ is a vector stacking all the columns of $\mathbf{A}$. $\mathbf{I}_N$ represents the $N \times N$ identity matrix. For any general complex vector $\mathbf{x}$, x_i denotes the i-th element of $\mathbf{x}$, and $\text{diag}\{\mathbf{x}\}$ denotes a diagonal matrix with each diagonal element being the corresponding element in $\mathbf{x}$. $\|\mathbf{x}\|_0$ and $\|\mathbf{x}\|$ denote ℓ_0 and ℓ_2 norm of the vector $\mathbf{x}$, respectively. $\mathbb{C}^{x \times y}$ and $\mathbb{R}^{x \times y}$ represent the space of $x \times y$ complex and real number matrices. Notation $\mathbf{x} \sim \mathcal{CN}(0, \sigma^2)$ means that $\mathbf{x}$ is complex circularly symmetric Gaussian with zero mean and variance σ^2, and $E[\mathbf{x}]$ denotes the expected value of $\mathbf{x}$. $j \triangleq \sqrt{-1}$ presents the imaginary unit and the calligraphy upper-case letter $\mathcal{K}$ denotes a set.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this section, we describe an IRS-assisted multi-BS downlink multi-user MISO mmWave system model, where each user is served by a BS as illustrated in Fig. 1. It should be noted that an IRS is deployed on the surface of the building to assist a severely blocked BS in communication. Then we develop the mmWave channel model to accounts for the channel
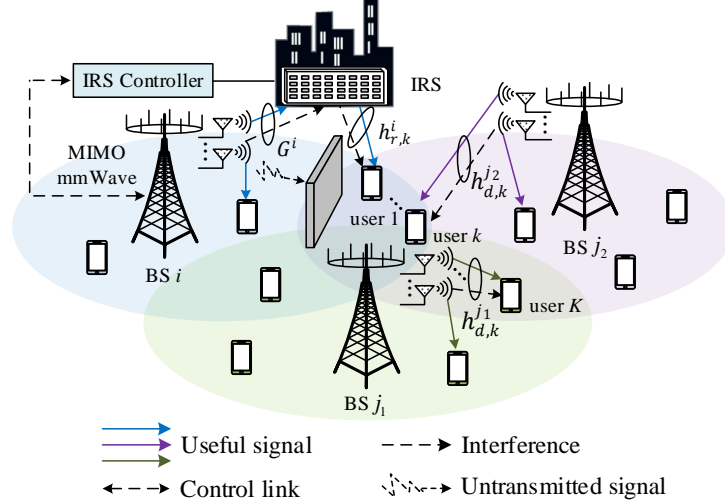
Fig. 1: An IRS-assisted multi-BS multi-user mmWave system.

characteristics of mmWave. This section also describes the problem formulation for the joint design of IRS reflection matrix, transmit power allocation at multi-BS and UA algorithm.

A. System Model

As shown in Fig. 1, K users are served by S BSs in the designed system, where the set of the user indexes and the BS indexes are denoted as $\mathcal{K} = \{1, 2, \cdots, K\}$ and $\mathcal{S} = \{1, 2, \cdots, S\}$, respectively. Each BS is equipped with $M > K$ antennas, and each user is equipped with only one antenna. Furthermore, we assume that an IRS[1] with N elements reflects the transmitted signal from the assisted BS $i \in \mathcal{S}$ for its associated users and other BSs $j \in \mathcal{S}, i \neq j$ are not equipped with IRS. Equipped with a smart controller, the IRS utilizes passive array to sense the state of the external environment and dynamically adjust the phase shift of each reflecting element. In downlink transmission, BS can transmit independent data steams to these K users simultaneously. We suppose the channel state information (CSI) is perfectly known at BSs and IRS[2], which is the same as [10]–[12], [18], [19].

---

[1]Existing research works [10]–[12] show that IRS-assisted BS communication has strong scalability and signals reflected by multi-IRSs can be superimposed at the receiver to further enhance the received signal power. Therefore, in order to simplify and focus on the impact of the introduction of IRS on the UA and system performance, we only consider one IRS, and multi-IRSs combined with power allocation and UA to enhance mmWave communication will be studied in future work.

[2]Though this is an idealistic assumption, it's still meaningful to study the performance gains brought by IRS and UA for the mmWave system. How to obtain CSI at IRS is out of the scope of this paper, and some related works can be found in [21], [22].

Since the mmWave links are highly susceptible to environmental blockages, we assume that the direct mmWave link between BS i and user $k, k \in \mathcal{K}$ is severely blocked by obstacles, such that BS i can only communicate with user k through the IRS reflection link. In contrast, the channel between BS j and user k is a direct channel. $\mathbf{G}^i \in \mathbb{C}^{N \times M}, \mathbf{h}_{r,k}^i \in \mathbb{C}^{1 \times N}, \mathbf{h}_{d,k}^j \in \mathbb{C}^{1 \times M}$ are equivalent channels from the BS i to the IRS, from the IRS to user k, and from the BS j to user k, respectively. Besides, let $\mathbf{h}_k^s \in \mathbb{C}^{1 \times M}$ denotes the channel between BS s and user k and $\mathbf{h}_k^s \in \{\mathbf{h}_{r,k}^i \boldsymbol{\Phi} \mathbf{G}^i, \mathbf{h}_{d,k}^j\}$, where $\mathbf{h}_{r,k}^i \boldsymbol{\Phi} \mathbf{G}^i$ represents the concatenation channel between BS i and user k. The reflection matrix of IRS is denoted as $\boldsymbol{\Phi} = \mathrm{diag}\{\beta_1 e^{j\varphi_1}, \cdots, \beta_N e^{j\varphi_N}\}$, where β_n and φ_n represent the amplitude reflection coefficient and the phase shift of the n-th element of the IRS, respectively. In practice, each element of the IRS is usually designed to maximize signal reflection [18], and thus we set $\beta_n = 1, \forall n$ in the sequel of this paper.

Furthermore, we define the UA matrix $\mathbf{A} \in \mathbb{C}^{S \times K}$ in (1), where $\tilde{\mathbf{a}}_s$ and $\|\tilde{\mathbf{a}}_s\|_0$ denote the UA vector at the BS s and the number of users served by s, respectively. And $\mathbf{a}_k = [a_{1,k}, a_{2,k}, ..., a_{S,k}]^T \in \mathbb{C}^{S \times 1}$ represents the UA vector of user k, where $a_{s,k}$ is a binary decision variable (i.e., $a_{s,k} \in \{0, 1\}, \forall s, \forall k$). In detail, $a_{s,k} = 1$ if user k is associated to BS s and $a_{s,k} = 0$ otherwise. Since each user is served by only one BS, $\|\mathbf{a}_k\|_0 = 1$ and we need to determine BS s that actually provides service for user k via UA vector $\mathbf{a}_k$. Meanwhile, we denote the set of users that BS s can serve as $\mathcal{A}(s)$ and the set of BSs that can serve user k as $\mathcal{B}(k)$. $\mathcal{D}$ is defined as a set of all possible BS-user assignment pairs (s, k) with $k \in \mathcal{A}(s)$.

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_K] = \begin{bmatrix} \underbrace{a_{11}, a_{12}, \cdots, a_{1K}}_{\tilde{\mathbf{a}}_1} \\ \vdots \\ \underbrace{a_{S1}, a_{S2}, \cdots, a_{SK}}_{\tilde{\mathbf{a}}_S} \end{bmatrix}, \tag{1}$$

The complex baseband transmitted signal at the BS can be expressed as $\mathbf{x}_s = \sum_{k \in \mathcal{A}(s)} \sqrt{p_k^s} \mathbf{w}_k^s t_k^s$, where p_k^s, $\mathbf{w}_k^s \in \mathbb{C}^{M \times 1}$ and t_k^s denote the transmit power, precoding vector and transmitted data symbol at BS s, of k-th user, respectively. It is assumed that $t_k^s, \forall s, \forall k$, are independent variables with zero mean and unit power. The signal received by user k from BS s can be expressed as

$$y_{s,k} = \mathbf{h}_k^s \sqrt{p_k^s} \mathbf{w}_k^s t_k^s + \mathbf{h}_k^s \sum_{j \neq k, j \in \mathcal{A}(s)} \sqrt{p_j^s} \mathbf{w}_j^s t_j^s + n_k, \tag{2}$$

where $n_k \sim \mathcal{CN}(0, \sigma^2)$ is complex additive white Gaussian noise with zero mean and variance σ^2. The power of the transmit signal from the BS s is restricted by the maximum transmit power

threshold P_{max}

$$E\left[|\mathbf{x}|^2\right] = \text{tr}(\mathbf{P}_s \mathbf{W}_s^H \mathbf{W}_s) \leq P_{max}, \tag{3}$$

where $\mathbf{W}_s = [\mathbf{w}_1^s, \cdots, \mathbf{w}_K^s] \in \mathbb{C}^{M \times K}$, $\mathbf{P}_s = \text{diag}\left[p_1^s, \cdots, p_K^s\right] \in \mathbb{R}^{K \times K}$.

In this paper, we suppose that neighboring BSs can be allocated orthogonal frequency band or employ enhanced inter-cell interference coordination techniques [31] to eliminate inter-cell interference. Accordingly, the SINR of user k is written as γ_k and it can be given by

$$\gamma_{s,k} = \frac{p_k^s \left|\mathbf{h}_k^s \mathbf{w}_k^s\right|^2}{\sum_{j \neq k, j \in \mathcal{A}(s)} \left|p_j^s \mathbf{h}_k^s \mathbf{w}_j^s\right|^2 + \sigma^2}, \forall s, \forall k. \tag{4}$$

Then, the achievable rate from BS s to user $k, k \in \mathcal{A}(s)$ is defined as $R_{s,k}$, which can be computed as

$$R_{s,k} = \log_2 \left(1 + \frac{p_k^s \left|\mathbf{h}_k^s \mathbf{w}_k^s\right|^2}{\sum_{j \neq k, j \in \mathcal{A}(s)} \left|p_j^s \mathbf{h}_k^s \mathbf{w}_j^s\right|^2 + \sigma^2}\right). \tag{5}$$

B. MmWave Channel Model

In mmWave systems, the BS-IRS channel $\mathbf{G}^i$ is modeled according to 3D SalehValenzuela channel model [12], [32] that has been widely used to characterize the mmWave channel:

$$\mathbf{G}^i = \sqrt{MN} \sum_{g=0}^{G_p} \alpha_g \xi_t \xi_r \mathbf{a}_N^H \left(\theta_{AoA}^{(g)}\right) \mathbf{a}_M \left(\theta_{AoD}^{(g)}\right), \tag{6}$$

where G_p denotes the number of NLOS paths, $g = 0$ denotes the LOS path[3], α_g indicates the complex gain associated with the g-th path, and ξ_r and ξ_t are the receive and transmit antenna gains, respectively. The parameters $\theta_{AoA}^{(g)}$ and $\theta_{AoD}^{(g)}$ represent the angle of arrival (AoA) and angle of departure (AoD) of the signal reflected by the IRS in the g-th path. We employ a uniform linear array (ULA) at the IRS [22], and thus the array response function of the IRS can be expressed as

$$\mathbf{a}_N(\theta_{AoA}) = \frac{1}{\sqrt{N}} \left[1, e^{j2\pi \frac{d}{\lambda} sin\theta_{AoA}}, \cdots, e^{j2\pi \frac{d}{\lambda}(N-1)sin\theta_{AoA}}\right], \tag{7}$$

where λ is the mmWave wavelength and d is the antenna spacing.

Assume that the IRS is coated on the buildings around users which provides a high probability of LOS propagation [12], [18]. Thus, the channel between the IRS and the k-th user can be described as

$$\mathbf{h}_{r,k}^i = \sqrt{N} \alpha_k \xi_t \xi_r \mathbf{a}_N \left(\theta_{AoD}\right), \tag{8}$$

---

[3]A series of studies have shown that mmWave channel normally consist of a limited number of main multipath components, while the scattering paths at sub-6 GHz is generally rich.

where α_k indicates the complex gain and θ_{AoD} is the AoD of the signal from the IRS to the k-th user. Similar to [10], we generate the direct BS-user channels $\mathbf{h}_{d,k}^j$ according to (8).

C. Problem Formulation

We further define the channel and beamforming matrix of BS s as $\mathbf{H}_s = [\mathbf{h}_1^{sT}, \cdots, \mathbf{h}_K^{sT}]^T \in \mathbb{C}^{K \times M}$. As there are more antennas than users, simple linear processing techniques such as maximum ratio transmission or zero forcing (ZF) are near optimal [26]. Therefore, we consider using ZF precoding at each BS by setting $\mathbf{W}_s^{opt} = |\mathbf{H}_s|^+$ to achieve perfect interference suppression. In this paper, our goal is to maximize the sum rate of all users by jointly optimizing the PBF at IRS, power allocation and UA in an IRS-assisted mmWave system under reasonable constraints. Accordingly, the sum rate maximization problem is formulated as

$$\max_{\mathbf{\Phi}, \{\mathbf{P}_s\}_{s=1}^S, \mathbf{A}} \quad \sum_{k \in \mathcal{K}} \sum_{s \in \mathcal{S}} a_{s,k} \log_2 \left(1 + p_k^s \sigma^{-2}\right) \tag{9a}$$

$$s.t. \quad \varphi_n \in \mathcal{F}, \forall n, \tag{9b}$$

$$R_k = \sum_{s \in \mathcal{S}} a_k^s \log_2 \left(1 + p_k^s \sigma^{-2}\right) \geq R_{min}, \forall s, \forall k, \tag{9c}$$

$$\mathrm{tr}((\mathbf{H}_s)^+ \mathbf{P}_s (\mathbf{H}_s)^{+H}) \leq P_{max}, \forall s, \tag{9d}$$

$$\sum_{s \in \mathcal{B}(k)} a_{s,k} = 1, \forall k, \tag{9e}$$

$$\sum_{k \in \mathcal{A}(s)} a_{s,k} \geq 1, \forall s, \tag{9f}$$

$$a_{s,k} \in \{0,1\}, \forall (s,k) \in \mathcal{D}, \tag{9g}$$

where $\mathcal{F} = \left\{0, \frac{2\pi}{2^b}, \frac{2\pi \times 2}{2^b}, \cdots, \frac{2\pi \times (2^b - 1)}{2^b}\right\}$ is the set of available phase shifts for the IRS, b is the resolution of the phase shifter at IRS, while R_k and R_{min} are the actual rate of user k and the minimum achievable rate of the all users, respectively. Also, constraint (9b) accounts for the fact that each IRS reflecting element only provides a discrete phase shift, (9c) represents the individual minimum rate constraint of user k (we use R_{min} to characterize users' QoS requirements in this paper), and (9d) ensures that the transmit power of BS s is kept below P_{max}. Then, constraint (9e) ensures that each user communicates with only one BS. Constraint (9f) indicates that each BS can serve multiple users simultaneously and at least one user. Given the non-convexity and complexity of the problem, we forward to solve problem (9) efficiently in the sequel.

III. PROPOSED SOLUTION

We cannot assert global optimality due to the non-convex objective function (9a) and the non-convex constraint (9b). Thus, a tractable algorithm employing the alternating optimization technique to separately and iteratively solve $\mathbf{\Phi}$, $\{\mathbf{P}_s\}_{s=1}^S$ and $\mathbf{A}$ is proposed. In particular, for given $\{\mathbf{P}_s\}_{s=1}^S$ and $\mathbf{A}$, we optimize the IRS reflection matrix $\mathbf{\Phi}$ based on the SFP method, and then find the optimum $\{\mathbf{P}_s\}_{s=1}^S$, when $\mathbf{\Phi}$ and $\mathbf{A}$ are fixed. Next, we solve $\mathbf{A}$ by utilizing the FRA alogrithm. Since (9a) is upper-bounded on the feasible set, iterating this solution process improves the sum rate value at each iteration and eventually converges in the approximate optimum value of the objective. Finally, we propose an IPPU algorithm and analyze the complexity of the proposed algorithm.

A. IRS Reflection Matrix Optimization

When the power allocation matrix $\{\mathbf{P}_s\}_{s=1}^S$ and UA matrix $\mathbf{A}$ are fixed, we only need to consider the channel between all users and IRS-assisted BS i in constraint (9d). Let $\mathbf{H}_r^i = [\mathbf{h}_{r,1}^{iT}, \cdots, \mathbf{h}_{r,K}^{iT}]^T \in \mathbb{C}^{K \times N}$ and then problem (9) can be represented as

$$\text{find} \quad \mathbf{\Phi} \tag{10a}$$

$$s.t. \quad \varphi_n \in \mathcal{F}, \forall n, \tag{10b}$$

$$\text{tr}((\mathbf{H}_r^i \mathbf{\Phi} \mathbf{G}^i)^+ \mathbf{P}_i (\mathbf{H}_r^i \mathbf{\Phi} \mathbf{G}^i)^{+H}) \leq P_{max}. \tag{10c}$$

The main challenge in solving problem (10) lies in the fact that the constraint (10b) is a non-convex constraint. We observe that problem (10) is feasible if and only if the optimal value of the following optimization problem is lower than P_{max}

$$\min_{\mathbf{\Phi}} \quad f_1(\mathbf{\Phi}) = \text{tr}((\mathbf{H}_r^i \mathbf{\Phi} \mathbf{G}^i)^+ \mathbf{P}_i (\mathbf{H}_r^i \mathbf{\Phi} \mathbf{G}^i)^{+H}) \tag{11a}$$

$$s.t. \quad \varphi_n \in \mathcal{F}, \forall n. \tag{11b}$$

***Theorem* 1:** The optimization problem (11) can be represented as

$$\min_{\mathbf{\Phi}} \quad \mathbf{y}^H \mathbf{B} \mathbf{y} \tag{12a}$$

$$s.t. \quad \varphi_n \in \mathcal{F}, \forall n, \tag{12b}$$

where

$$\mathbf{B} = ((\widetilde{\mathbf{H}}_r^i)^{+H} \otimes (\mathbf{G}^i)^+)^H ((\widetilde{\mathbf{H}}_r^i)^{+H} \otimes (\mathbf{G}^i)^+) \in \mathbb{C}^{N^2 \times N^2}, \tag{13a}$$

$$\mathbf{y} = \text{vec}(\boldsymbol{\Phi}^{-1}) \in \mathbb{C}^{N^2 \times 1}. \tag{13b}$$

Proof: The proof of Theorem 1 can be found in the Appendix A of this paper.

As we will show, the objective form in (12a) enables us to deal with the non-convex constraint (12b), if (12a) can be reformulated as a differentiable function. On this basis, we can use SFP method, also known as Majorization-Minimization method [19], [33] to solve problem (12a). The essence of the SFP method is an iterative approach that solves a non-convex problem by solving a sequence of approximate subproblems. If each approximation problem satisfies the following assumptions to the original problem in each iteration, the optimal sequence can decrease monotonically and converges.

1) In each iteration t, the feasible sequence $\{\mathbf{y}^{(t)}\}$ maximizes the upper bound of $\mathbf{y}^H \mathbf{B} \mathbf{y}$.

2) For any t-th iteration, the upper bound of the maximum value of the t-th iteration must be equal to the true objective $\mathbf{y}^H \mathbf{B} \mathbf{y}$, when evaluating at the maximum value calculated in $(t-1)$-th iteration.

The reason is that $\mathbf{y}^H \mathbf{B} \mathbf{y}$ is the lower-bounded over the feasible set of the problem (12). Nevertheless, the challenge of using the SFP method is to determine a suitable upper bound of $\mathbf{y}^H \mathbf{B} \mathbf{y}$, which is easier to minimize compared with the original objective (12a). Therefore, we provide a convenient upper bound of $\mathbf{y}^H \mathbf{B} \mathbf{y}$ in the following lemma which fulfills the theoretical requirements of the SFP method.

***Lemma* 1:** For any feasible $\mathbf{y}$ and given any feasible point $\mathbf{y}^{(t)}$, a suitable upper bound to employ the SFP method is:

$$\mathbf{y}^H \mathbf{B} \mathbf{y} \leq f_2(\mathbf{y}|\mathbf{y}^{(t)}) = \mathbf{y}^H \mathbf{C} \mathbf{y} + (\mathbf{y}^{(t)})^H (\mathbf{C} - \mathbf{B}) \mathbf{y}^{(t)} - 2 \operatorname{Re}\{\mathbf{y}^H (\mathbf{C} - \mathbf{B}) \mathbf{y}^{(t)}\}, \tag{14}$$

wherein, $\mathbf{C} = \lambda_{max} \mathbf{I}_{N^2}$, λ_{max} is the maximum eigenvalue of $\mathbf{B}$ and the matrix $\mathbf{C} - \mathbf{B}$ is positive semidefinite.

Proof: Consider the following inequality:

$$\|(\mathbf{C} - \mathbf{B})^{1/2} \mathbf{y} - (\mathbf{C} - \mathbf{B})^{1/2} \mathbf{y}^{(t)}\|^2 \geq 0, \tag{15}$$

then we perform the second-order Taylor expansion of (15)

$$\mathbf{y}^H (\mathbf{C} - \mathbf{B}) \mathbf{y} + (\mathbf{y}^{(t)})^H (\mathbf{C} - \mathbf{B}) \mathbf{y}^{(t)} - 2 \operatorname{Re}\{\mathbf{y}^H (\mathbf{C} - \mathbf{B}) \mathbf{y}^{(t)}\} \geq 0. \tag{16}$$

At last, the upper bound in (14) is obtained by separating and extracting $\mathbf{y}^H \mathbf{B} \mathbf{y}$, and the proof is complete.

Next, we use the variable $\mathbf{y}$ to rewrite the constraint (12b) in detail. Here, it should be noted that (12b) only constrains the unit modulus of the discrete phase of the diagonal elements of $\boldsymbol{\Phi}$, while $\mathbf{y}$ contains the elements of the vectorized $\boldsymbol{\Phi}$. Therefore, unit modulus of the discrete phase are only allocated to some elements in $\mathbf{y}$ and all other elements must be zero. Let y_i be the i-th element of $\mathbf{y}, i = 1, 2, ..., N, ..., N^2$. For y_i with unit modulus, $y_i = e^{j\theta_i}, \theta_i \in \mathcal{F}$. Then the element $y_i = (i-1)N + i$ must have a discrete phase unit modulus. From the above, each iteration of the SFP method needs to solve the optimal variable $\mathbf{y}$ through the following minimization problem

$$\min_{\mathbf{y}} \quad f_2(\mathbf{y}|\mathbf{y}^{(t)}) \tag{17a}$$

$$\text{s.t.} \quad \theta_i \in \mathcal{F}, \forall i = (n-1)N + n, n = 1, 2, ..., N, \tag{17b}$$

$$|y_i| = 1, \forall i = (n-1)N + n, n = 1, 2, ..., N, \tag{17c}$$

$$|y_i| = 0, \forall i \neq (n-1)N + n, n = 1, 2, ..., N, \tag{17d}$$

where

$$f_2(\mathbf{y}|\mathbf{y}^{(t)}) = \lambda_{max}\|\mathbf{y}\|^2 + (\mathbf{y}^{(t)})^H(\lambda_{max}\mathbf{I}_{N^2} - \mathbf{B})\mathbf{y}^{(t)} - 2\operatorname{Re}\{\mathbf{y}^H(\lambda_{max}\mathbf{I}_{N^2} - \mathbf{B})\mathbf{y}^{(t)}\}. \tag{18}$$

Observing (17c) and (17d), we can get $\|\mathbf{y}\|^2 = N$. By deleting items that are not related to $\mathbf{y}$, problem (17) can be equivalently transformed as

$$\max_{\mathbf{y}} \quad 2\operatorname{Re}\{\mathbf{y}^H(\lambda_{max}\mathbf{I}_{N^2} - \mathbf{B})\mathbf{y}^{(t)}\} \tag{19a}$$

$$\text{s.t.} \quad (17b) - (17d), \tag{19b}$$

Obviously, the only free variable is the phase θ_i of y_i. We define $\mathbf{d} = (\lambda_{max}\mathbf{I}_{N^2} - \mathbf{B})\mathbf{y}^{(t)}$, d_i is the i-th element of $\mathbf{d}$, and denote the argument of d_i by ϑ_i. Thus, the optimal phase solution $\theta_i, \forall i = (n-1)N + n, n = 1, ..., N$ of $\mathbf{y}$ is

$$\theta_i^{opt} = \arg\min_{\theta_i \in \mathcal{F}} |\vartheta_i - \theta_i|, \tag{20}$$

Hence, for any $\mathbf{y}^{(t)}$, the problem (17) is solved by

$$y_i^{opt} = \begin{cases} e^{j\theta_i^{opt}}, & \forall i = (n-1)N + n, n = 1, 2, ..., N. \\ 0, & \forall i \neq (n-1)N + n, n = 1, 2, ..., N. \end{cases} \tag{21}$$

B. Power Allocation Matrix Optimization

Since constraints (9b) and (9e)-(9g) are only determined by IRS reflection matrix $\mathbf{\Phi}$ and UA matrix $\mathbf{A}$, respectively, the problem (9) can be expressed as

$$\max_{\{\mathbf{P}_s\}_{s=1}^S} \quad \sum_{k \in \mathcal{K}} \log_2 \left(1 + p_k^s \sigma^{-2}\right) \tag{22a}$$

$$s.t. \quad p_k^s \geq \sigma^2 (2^{R_{min}} - 1), \forall s, \forall k, \tag{22b}$$

$$\mathrm{tr}((\mathbf{H}_s)^+ \mathbf{P}_s (\mathbf{H}_s)^{+H}) \leq P_{max}, \forall s. \tag{22c}$$

It can be seen that problem (22) is convex and can be handled with means of standard convex optimization [34]. Let $\mathcal{P} \triangleq \{\mathbf{P}_s = \mathrm{diag}\left[p_1^s, \cdots, p_K^s\right] : (22b) \ \& \ (22c), \forall s\}$, we have

$$\mathbf{P}_s^{opt} = \arg \max_{\mathbf{P}_s \in \mathcal{P}} \sum_{k \in \mathcal{K}} \log_2 \left(1 + p_k^s \sigma^{-2}\right), \forall s. \tag{23}$$

Furthermore, by analyzing the Karush–Kuhn–Tucker (KKT) optimality conditions of (23), the closed-form expression of the solution for $\mathbf{P}_s$ is obtained as

$$p_k^{opt} = \max\{0, \varpi \lambda_k - \sigma^2\} + \sigma^2 (2^{R_{min}} - 1) \lambda_k^{-1}, \forall k, \tag{24}$$

where water level ϖ is the Lagrange multiplier associated to (22c) and it follows

$$\varpi = \frac{1}{u} \left(P_{max} - \sigma^2 \left(2^{R_{min}} - 2\right) \sum_{k=1}^{K} \lambda_k^{-1} \right). \tag{25}$$

Here, λ_k is the k-th eigenvalue of $\mathbf{H}_s \mathbf{H}_s^H$ and u is the number of non-zero eigenvalues λ_k.

C. UA Matrix Optimization

In this subsection, we focus on problem (9) to optimize the UA matrix $\mathbf{A}$. To ensure the QoS constraints (9c) of all users, we assume that $R_{s,k} = \log_2(1 + p_k^s \sigma^{-2}) \geq R_{min}$ holds for all BSs and users. Considering fixed $\mathbf{\Phi}$ and $\{\mathbf{P}_s\}_{s=1}^S$, we pay attention to solving the asymmetric multi-assignment problem

$$\max_{\mathbf{A}} \quad \sum_{s \in \mathcal{S}} \sum_{k \in \mathcal{K}} \frac{R_{s,k}}{R_{min}} a_{s,k} \tag{26a}$$

$$s.t. \quad \sum_{s \in \mathcal{B}(k)} a_{s,k} = 1, \forall k, \tag{26b}$$

$$\sum_{k \in \mathcal{A}(s)} a_{s,k} \geq 1, \forall s, \tag{26c}$$

$$a_{s,k} \in \{0,1\}, \forall(s,k) \in \mathcal{D}. \tag{26d}$$

Note that $R_{s,k}/R_{min}$ is the benefit of assignment pair (s,k) and the objective function of (26a) is the total network benefit. We define $\mathcal{C}$ as a subset of $\mathcal{D}$, where each BS s is a part of at least one pair $(s,k) \in \mathcal{C}$ and each user k is a part of only one pair $(s,k) \in \mathcal{C}$. By setting $a_{s,k} = 1$ if $(s,k) \in \mathcal{C}$ and $a_{s,k} = 0$ otherwise, we can obtain a feasible solution of problem (26) as well defined as a feasible assignemnt $\mathcal{C}$. Since the structure of problem (26) is consistent with the typical minimum cost flow problem [35], [36], we convert problem (26) into a typical minimum cost flow problem by introducing a virtual node e connected to each BS as

$$\min_{\mathbf{A}} \quad \sum_{(s,k)\in\mathcal{C}} \frac{-R_{s,k}}{R_{min}} a_{s,k} \tag{27a}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{A}(s)} a_{s,k} - a_{e,s} = 1, \forall s, \tag{27b}$$

$$\sum_{s\in\mathcal{S}} a_{e,s} = K - S, a_{e,s} \geq 0, \forall s, \tag{27c}$$

$$\sum_{s\in\mathcal{B}(k)} a_{s,k} = 1, \forall k, a_{s,k} \geq 0, \forall(s,k) \in \mathcal{D}, \tag{27d}$$

where $a_{s,k}$ is extended to include the supernode e. By using the terminology of network optimization, we declare that $a_{s,k} = 1$ means there is one unit flow between s and k and the optimal solution to (27) is the same to the initial asymmetric multi-assignment problem (26). Constraint (27b) ensures that the flow supply of each BS s is one unit, and constraint (27c) declares that e is the source node with $K - S$ units of flows. The last constraint (27d) ensures that each user is served by only one BS.

To proceed further, we utilize the duality theory [36] for this minimum cost flow problem (27) and formulate the dual problem

$$\min_{\pi_s, q_k, \mu} \quad \sum_{s\in\mathcal{S}} \pi_s + \sum_{k\in\mathcal{K}} q_k + (K-S)\mu \tag{28a}$$

$$\text{s.t.} \quad \pi_s + q_k \geq \frac{R_{s,k}}{R_{min}}, \forall(s,k) \in \mathcal{D}, \tag{28b}$$

$$\mu \geq \pi_s, \forall s, \tag{28c}$$

where π_s, μ and q_k are all Lagrangian multipliers and they are associated with constraint (27b), (27c) and (27d), respectively. The parameter $-\pi_s$ represents the price of each BS s, μ stands for the price of source node e and q_k denotes the price of each user k. We can derive the optimal solution of problem (26) from the optimal solution of problem (28).

Next, we introduce $\epsilon - Complementary\ Slackness$ ($\epsilon - CS$) to solve problem (28). Let ϵ be a positive scalar. Then an assignment $\mathcal{C}$ and a pair (π, p) satisfy $\epsilon - CS$ if

$$\pi_s + q_k \geq \frac{R_{s,k}}{R_{min}} - \epsilon, \forall (s, k) \in \mathcal{D}, \tag{29a}$$

$$\pi_s + p_k = R_{s,k}, \forall (s, k) \in \mathcal{C}, \tag{29b}$$

$$\pi_s = \max_{l=1,2,\cdots,S} \pi_l, s \text{ has multi-pairs } (s, k) \in \mathcal{C}. \tag{29c}$$

Proposition **1:** Consider a dual variable pair (π, q) and let $\mathcal{C}$ be a feasible solution of problem (28). Assuming that $\epsilon < 1/S$ and R_{sk}/R_{min} is an integer $\forall s, k$, $\mathcal{C}$ is the optimal solution of problem (28), if $\epsilon - CS$ conditions (29) are satisfied by $\mathcal{C}$ and (π, q).

Proof: The proof naturally results from the Proposition 7.7 in [36] and Proposition 1 in [35], thus, it will not be elaborated in this paper for briefness.

Based on proposition 1, problem (28) can be solved by our proposed FRA based UA algorithm. As shown in Algorithm 1, it consists of processes of forward and reverse auction. The forward auction aims to associate each BS with one user, and the reverse one is to assign the remaining users to available BSs. In the forward auction process, to begin with, we select the unassociated BS s in $\mathcal{C}$ and find the best user k_s which can provides the maximum benefit $\max_{k \in \mathcal{A}(s)}\{\frac{R_{s,k}}{R_{min}} - q_k\}$ among all users of $\mathcal{A}(s)$ (line 4-7). Then, BS s gives a bid b_{s,k_s} for user k_s and the bidding process ends. Next, we assign the users who received bid to the corresponding BS s which provides the highest bid and update k's price (line 8-9). If user k is initially assigned to another BS s' at the beginning, remove (s', k) and add (s_k, k) to $\mathcal{C}$ (line 10). When all BSs are assigned to one user and satisfy $\epsilon - CS$ conditions, the forward auction is terminated (line 3-12). There are still some unassociated users due to $K > S$. In the reverse auction process, we assign the rest of the users to available BSs as shown in line 13-20. Specifically, select the unallocated user k, and we can find the best BS s_k that provides the maximum benefit $\max_{s \in \mathcal{B}(k)}\{R_{s,k}/R_{min} - \pi_s\}$ among all BS of $\mathcal{B}(k)$ (line 14-17). Then, user k decreases its price (i.e., increases the profit π_{s_k} for s_k) to attract BS s_k, and we associate k to s_k. After updating the profit of s_k and the price of k, respectively, the reverse auction process (line 13-20) is terminated and we can obtain an optimal assignment $\mathcal{C}$ by a finite number of iterations.

Note that $R_{s,k}/R_{min}$ is not the integer required by proposition 1 in geneural. Therefore, we

---

Algorithm 1: FRA based UA Algorithm

---

1 Input: Initial values of $\mathcal{C}, (\pi, q), \epsilon$ and μ.

2 Ensure: 1) $\frac{R_{s,k}}{R_{min}} - q_k \geq \max\limits_{l \in \mathcal{A}(s)} \left\{ \frac{R_{s,k}}{R_{min}} - q_l \right\} - \epsilon, \forall (s,k) \in \mathcal{C};$

2) (π, q) satisfy $\epsilon - CS$ conditions (29).

3 while *there are unassociated BSs* **do**

4   BS s is unassociated in $\mathcal{C}$, find the best user k_s that:

       $k_s = \arg\max_{k \in \mathcal{A}(s)} \left\{ \frac{R_{s,k}}{R_{min}} - q_k \right\}, \rho_s = \max_{k \in \mathcal{A}(s)} \left\{ \frac{R_{s,k}}{R_{min}} - q_k \right\},$

       $v_s = \max_{k \in \mathcal{A}(s), k \neq k_s} \left\{ \frac{R_{s,k}}{R_{min}} - q_k \right\};$

5   **if** k_s *is the only user in* $\mathcal{A}(s)$ **then**

6 $v_s \to -\infty;$

7   **end**

8 $b_{s,k_s} = q_{k_s} + \rho_s - v_s + \epsilon = R_{s,k_s}/R_{min} - v_s + \epsilon;$

9 $q_k = \max\limits_{s \in Q(k)} b_{s,k}$, where $Q(k)$ is the set of BSs that user k received a bid;

10   Remove any pair $(s', k) \in \mathcal{C}$ and add the pair (s_k, k) to $\mathcal{C}$ with $s_k = \arg\max\limits_{s \in Q(k)} b_{s,k};$

11   Update (π, q) and $\mu = \max_{s=1,\cdots,S} \pi_s.$

12 end

13 while *there are unassociated users* **do**

14   User k is unassociated in $\mathcal{C}$, find the best BS s_k that:

       $s_k = \arg\max_{s \in \mathcal{B}(k)} \left\{ \frac{R_{s,k}}{R_{min}} - \pi_s \right\}, \zeta_k = \max_{s \in \mathcal{B}(k)} \left\{ \frac{R_{s,k}}{R_{min}} - \pi_s \right\},$

       $v_k = \max_{s \in \mathcal{B}(k), s \neq s_k} \left\{ \frac{R_{s,k}}{R_{min}} - \pi_s \right\};$

15   **if** s_k *is the only BS in* $\mathcal{B}(k)$ **then**

16 $v_k \to -\infty;$

17   **end**

18 $\delta = \min \left\{ \mu - \pi_{s_k}, \zeta_k - v_k + \epsilon \right\}$, add (s_k, k) to $\mathcal{C};$

19   Update: $\pi_{s_k} = \pi_{s_k} + \delta, \ q_k = \zeta_k - \delta.$

20 end

    Output: A feasible optimal assignment $\mathcal{C}$.

---

first amplify $R_{s,k}/R_{min}$ by a certain factor, and then round to the closest integer value[4] before running Algorithm 1.

---

[4]In this way, the fractional part of $R_{s,k}/R_{min}$ is much smaller than the integer part of $R_{s,k}/R_{min}$, thus the influence of the optimal solution obtained after rounding on the true optimal value of problem (26) can be ignored.

***Proposition* 2:** When $\epsilon < 1/S$, Algorithm 1 terminates after a finite number of iterations with an optimal feasible assignment $\mathcal{C}$ of problem (28).

***Proof*:** First of all, using the theory of auction algorithm directly [36], if the $\epsilon - CS$ condition is satisfied before Algorithm 1 being executed, it is also satisfied at the end of the forward auction. This ensures that each BS is assigned to one user with the maximum benefit. Next, in order to prove the optimality and convergence of the improved reverse auction, we need to prove: 1) the reverse auction still satisfies $\epsilon - CS$ conditions after continuous iteration. 2) the reverse auction terminates after a limited number of iterations, and the optimal assignment $\mathcal{C}$ is obtained. We prove the above two points respectively as follows

1) Suppose the profit of BS s before and after iteration are π_s and π_s', respectively, BS s^* receives a bid from user k and is assigned to user k during the iteration, then $\pi_{s^*}' = R_{s^*,k}/R_{min} - \upsilon_k + \epsilon$ holds. Next, by substituting the expression of υ_k, we can deduce

$$\frac{R_{s^*,k}}{R_{min}} - \pi_{s^*}' = \upsilon_k - \epsilon = \max_{s \in \mathcal{B}(k), s \neq s_k} \left\{ \frac{R_{s,k}}{R_{min}} - \pi_s \right\} - \epsilon. \tag{30}$$

Since $\pi_s' \geq \pi_s$ holds for any s, (30) can be transformed into

$$\frac{R_{s^*,k}}{R_{min}} - \pi_{s^*}' \geq \max_{s \in \mathcal{B}(k)} \left\{ \frac{R_{s,k}}{R_{min}} - \pi_s' \right\} - \epsilon. \tag{31}$$

It can be seen that formula (31) shows after iterative assignment, all $(s^*, k) \in \mathcal{C}$ satisfy $\epsilon - CS$ conditions. Consider a pair (s^*, k) that belongs to both the assignment before and after iteration. If BS s^* does not receive a bid during the iteration, $\pi_{s^*}' = \pi_{s^*}$, and there is $\pi_s' \geq \pi_s$ holds for any s. Thus (s^*, k) (31) holds before and after iteration. In summary, all pairs $(s^*, k) \in \mathcal{C}$ after iteration satisfy $\epsilon - CS$ conditions.

2) Noting that user k is assigned to BS s_k without changing the original association of s_k in each iteration of the reverse auction, BS s will only receive a finite number of bids after meeting $\pi_s = \mu$. Since μ is the upper bound of the profit that all BSs can reach, at the end of each iteration, the profit π_s obtained by s is equal to μ or at least increased by ϵ. Therefore, if BS s receives infinite bids from users, the profit of s will grow to positive infinity, which contradicts $\mu = \max_{s=1,\cdots,S} \pi_s$. Thus, the reverse auction process will terminate after a finite number of iterations.

***Proposition* 3:** The final assignment $\mathcal{C}$ obtained by Algorithm 1 is within $S\epsilon$ of the optimal assignment benefit of problem (26).

Proof: Considering the forward auction of Algorithm 1, $\epsilon > 0$ is the increment of each bidding. For any assignment $\mathcal{U} = \{(s, k_s) | \ \forall s\}$, its total benefit satisfies

$$\sum_{s=1}^{S} \frac{R_{s,k_s}}{R_{min}} \leq \sum_{k=1}^{K} q_k + \sum_{s=1}^{S} \max_k \left\{ \frac{R_{s,k}}{R_{min}} - q_k \right\}. \tag{32}$$

Define X^* as the total benefit that is obtained by the optimal assignment of problem (26) and Y^* as the optimal minimum obtained by Algorithm 1 of dual problem (28), respectively:

$$X^* = \max_{k_s \in \mathcal{A}(s)} \sum_{s \in \mathcal{S}} \frac{R_{s,k_s}}{R_{min}}, \ \text{if } s \neq m, k_s \neq k_m, \tag{33a}$$

$$Y^* = \min_{q_k, k \in \mathcal{K}} \left\{ \sum_{k=1}^{K} q_k + \sum_{s=1}^{S} \max_{k_s \in \mathcal{A}(s)} \left\{ \frac{R_{s,k}}{R_{min}} - q_k \right\} \right\}. \tag{33b}$$

Noting that user k_s satisfies the $\epsilon - CS$ conditions, we can get $\frac{R_{s,k_s}}{R_{min}} - q_{k_s} \geq \max_{k \in \mathcal{A}(s)} \{ \frac{R_{s,k}}{R_{min}} - q_k \} - \epsilon$. Since the the price of unassigned user is 0, $\sum_{s=1}^{S} (q_{k_s} + \max_{k_s \in \mathcal{A}(s)} \{ \frac{R_{s,k}}{R_{min}} - q_k \}) = \sum_{k=1}^{K} q_k + \sum_{s=1}^{S} \max_{k_s \in \mathcal{A}(s)} \{ \frac{R_{s,k}}{R_{min}} - q_k \} \geq Y^*$ is satisfied. Accordingly, it can be deduced that

$$\begin{aligned} Y^* &\leq \sum_{s=1}^{S} \left(q_{k_s} + \max_{k_s \in \mathcal{A}(s)} \left\{ \frac{R_{s,k}}{R_{min}} - q_k \right\} \right) \\ &\leq \sum_{s=1}^{S} \left(\frac{R_{s,k_s}}{R_{min}} + \epsilon \right) \leq X^* + S\epsilon. \end{aligned} \tag{34}$$

It presents that the total benefit $\sum_{s=1}^{S} \frac{R_{s,k}}{R_{min}}$ is within $S\epsilon$ of the optimal value X^* of problem (26). Now consider $\epsilon < 1/S$, then the assignment $\mathcal{C}$ obtained is the optimal within 1 strictly. The reverse auction can be proved similarly and details are omitted.

D. Sum Rate Maximization

In this subsection, our proposed IPPU algorithm for the initial problem (9) is summarized in Algorithm 2. As stated before, we solve $\boldsymbol{\Phi}$, $\{\mathbf{P}_s\}_{s=1}^{S}$ and $\mathbf{A}$ by iterating continuously and alternately till reaching a stable optimal rate. Specifically, since the optimization of $\{\mathbf{P}_s\}_{s=1}^{S}$ and $\mathbf{A}$ will increase the sum rate value $R_{sum} = \sum_{k \in \mathcal{K}} R_k$ after each iteration and the objective is upper bound over the feasible set of (9), the convergence of Algorithm 2 is guaranteed.

The computational complexity of the proposed algorithm depends on the number of iterations T_{max} to the outermost layer alternation and the complexity required to solve each subproblem. For Algorithm 2, it can be seen that the complexity of IRS optimization depends on the number of iterations T_{sfp} of the SFP method, multiplied by the amount of operations performed by

---

Algorithm 2: IPPU Algorithm

---

1 **Input:** $\{\mathbf{H}_s\}_{s=1}^S$, K, M, N, b, σ^2, P_{max}, R_{min}, (π,q), ϵ, μ, the tolerance ξ, iteration number t,t' that are both set to 1 and the upper bound T_{max}, T_{sfp}; A feasible solution $\mathbf{\Phi}^{(0)}$, $(\{\mathbf{P}_s\}_{s=1}^S)^{(0)}$, $\mathbf{A}^{(0)}$ of (9).

2 **while** $|R_{sum}^{(t)} - R_{sum}^{(t-1)}|^2 > \xi$ *and* $t \leq T_{max}$ **do**

3   **repeat**

4   Obtain $\mathbf{B}$ and $\mathbf{y}$ via Eq.(13);

5   Calculate the optimal $\mathbf{y}$ as in (21);

6       $\mathbf{y}^{(t')}$=reshape($\mathbf{y}$); $t' = t' + 1$;

7   **until** $|\mathbf{\Phi}^{(t')} - \mathbf{\Phi}^{(t'-1)}|^2 < \xi$ *or* $t' > T_{sfp}$;;

8   Update $\mathbf{\Phi}^{(t)}$ via $\mathbf{y}^{(t')}$ with given $(\{\mathbf{P}_s\}_{s=1}^S)^{(t-1)}$ and $\mathbf{A}^{(t-1)}$;

9   **if** *(11a) evaluated at* $\mathbf{\Phi}^{(t)}$ *is lower than* P_{max} **then**

10   Update $(\{\mathbf{P}_s\}_{s=1}^S)^{(t)}$ with given $\mathbf{\Phi}^{(t)}$ and $\mathbf{A}^{(t-1)}$ by using (24);

11   **else**

12   Break and declare infeasibility;

13   **end**

14   **end**

15   Update $\mathbf{A}^{(t)}$ with fixed $\mathbf{\Phi}^{(t)}$ and $(\{\mathbf{P}_s\}_{s=1}^S)^{(t)}$ by using Algorithm 1;

16   Set $t = t + 1$.

17 **end**

   Output: The optimal $\mathbf{\Phi}^{(t)}$, $(\{\mathbf{P}_s\}_{s=1}^S)^{(t)}$ and $\mathbf{A}^{(t)}$.

---

each iteration. We notice that each iteration of IRS needs to solve a convex problem with N variables. Besides, the transmist power optimization only needs to solve a convex problem with K variables. Furthermore, we know that convex problems have polynomial complexity in the number of optimization variables, which is at most quartic complexity.

On the other hand, for UA optimization, the total calculation is mainly determined by the value of ϵ and the maximum absolute value $\Delta = \max\limits_{(s,k)\in\mathcal{D}} \frac{R_{s,k_s}}{R_{min}} - \min\limits_{(s,k)\in\mathcal{D}} \frac{R_{s,k_s}}{R_{min}}$ with $\forall(s,k) \in \mathcal{D}$ in the FRA process. In detail, for forward auction process, the total number of bid calculations for BS s is proportional to $\lceil\Delta/\epsilon\rceil|\mathcal{A}(s)|$. So the total running time of forward auction is $\lceil\Delta/\epsilon\rceil \sum_{s=1}^S |\mathcal{A}(s)| = \lceil\Delta/\epsilon\rceil|\mathcal{D}| = \mathcal{O}(\lceil\Delta/\epsilon\rceil SK)$, where $|\mathcal{D}|$ is the number of (s,k) in $\mathcal{D}$. Similarly, the total running time of the reverse auction is $\mathcal{O}(\lceil\Delta/\epsilon\rceil(K-S)S)$. Considering the fact $K > K - S$, the complexity of UA optimization mainly depends on forward auction.

Therefore, the asymptotic complexity of Algorithm 2 can be expressed as

$$\mathcal{O}\left(T_{max}\left(T_{sfp}N^z + K^z + \lceil \Delta/\epsilon \rceil SK\right)\right), \tag{35}$$

where $1 \leq z \leq 4$.

IV. PERFORMANCE EVALUATION

A. Simulation Scenario

We consider an IRS-assisted multi-BS multi-user MISO mmWave communication system which operates at 28 GHz with bandwidth $B = 500$ MHz. $K = 16$ single-antenna users are randomly distributed in a circular area at $(150\ m, 50\ m)$ with a radius of $30\ m$. There are three BSs in $(0\ m, 0\ m)$, $(200\ m, 200\ m)$ and $(300\ m, 0\ m)$, respectively, and each of them is equipped with $M = 32$ antennas. The IRS comprising $N = 32$ passive elements is deployed at a position $(50\ m, 100\ m)$ to assist BS i for communication. The complex gain α_k is generated according to a complex Gaussian distribution $\alpha_k \sim \mathcal{CN}(0, 10^{-0.1\kappa})$ and $\kappa = \kappa_a + 10\kappa_b \log_{10}(d) + \kappa_c$ with $\kappa_a = 61.4$, $\kappa_b = 2$, $\kappa_c \sim \mathcal{CN}(0, \sigma_c^2)$ and $\sigma_c = 5.8$ dB, and α_g is generated according to α_k. The LOS path gain α_0 is set the same as α_k, while the parameters of NLOS path gain are set as $\kappa_a' = 72$, $\kappa_b' = 2.92$ and $\sigma_c' = 8.7$ dB, and the values of these parameters follow from [37]. Other required parameters are set as follows: $b = 2$, $\sigma^2 = -65$ dBm, $P_{max} = 30$ dBm, $G_p = 5$, $\xi_t = 9.82$ dBi, $\xi_r = 0$ dBi, $\epsilon = 0.2$.

All simulation results are obtained by averaging 10^5 channels and users positions realizations, and we apply average performance metrics obtained in this way. Specifically, we first generate 10^3 scenes and randomly set the locations of all users in each scene. Then, we generate 100 independent channels for each scene to implement Algorithm 2 and optimize the sum rate.

In order to verify the effectiveness of our proposed IPPU algorithm, we consider different four comparison algorithms in vary scenarios:

- **AF relay:** We consider a relevant scheme to replace the IRS structure that includes a conventional N-antenna AF relay. Different from the IRS reflection matrix $\boldsymbol{\Phi}$, we assume $N = 8$ antennas are used at the AF relay and the diagonal elements of $N \times N$ complex AF matrix $\boldsymbol{\Gamma}$ are constrained to maximum relay power instead of unit modulus. More details of AF relay can be found in [17], [19] and omitted here for briefness. The power allocation and UA optimization are the same as our proposed algorithm.
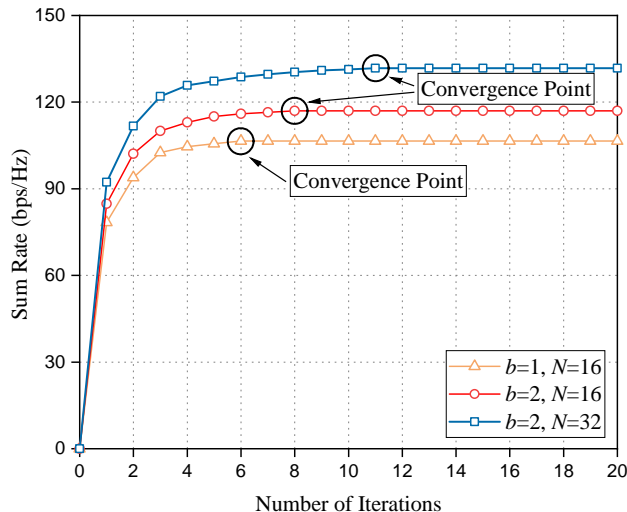
Fig. 2: Convergence performance of the proposed algorithm.

- **PBF + UAPC:** We utilize the UA and power control (UAPC) algorithm proposed in [26] which ensures that all users are treated fairly and the minimum QoS among users is maximized. Meanwhile, the proposed passive beamforming (PBF) is applied at IRS.

- **RPBF+NUBA:** The IRS reflection matrix is not optimized, in which the phases are randomly chosen from $\mathcal{F}$. Then, the nearest-based UA (NBUA) [27] algorithm under our proposed maximal transmit power (MP) method is applied at all BSs.

- **Without IRS:** Considering the sum rate of all users, each BS uses MP and the traditional received signal strength indicator (RSSI) based UA algorithm to serve users without IRS, which is utilized as a benchmark. Specifically, for the channels between BS i and all users, we generate the BS-user channels with only NLOS components and large path loss to replace BS-IRS and IRS-user channels.

B. Convergence Performance and Impact of IRS Parameter

Fig. 2 presents the convergence performance of the proposed algorithm under different number of reflect elements and phase resolutions at IRS. It can be seen that, the convergence of our proposed algorithm is confirmed in multiple simulated cases and all these three curves converge to stable solutions after no more than 11 iterations. When $b = 1$ and $N = 16$, our proposed algorithm converges after about 6 iterations. Meanwhile, we observe that the curve with the minimum b and N converges the fastest. With the increase of b and N, the convergence speed becomes slower but obtains significant improvement on sum rate. On the other hand, when b

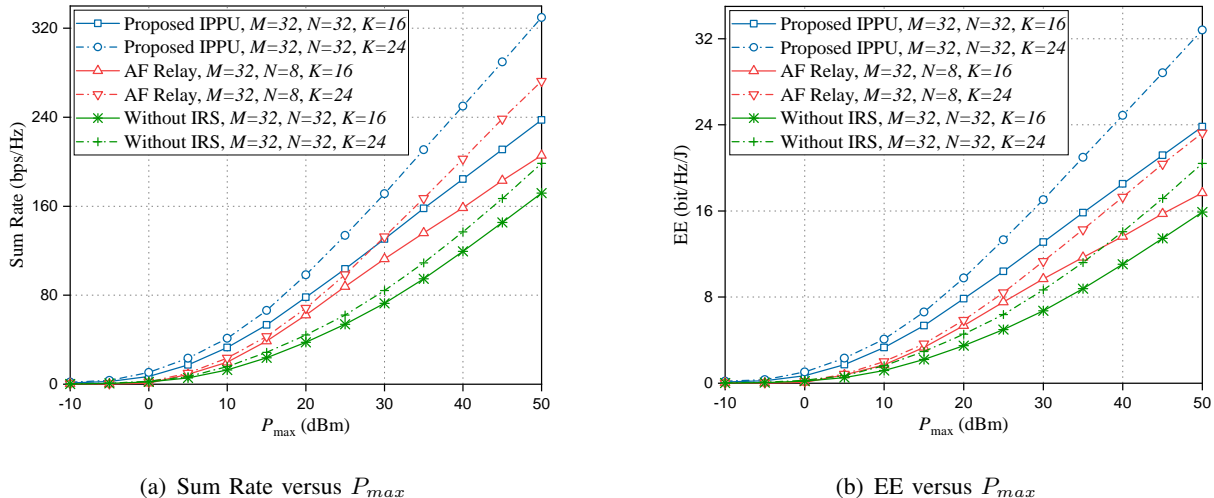(a) Sum Rate versus P_{max} (b) EE versus P_{max}

Fig. 3: The performance versus maximum transmit power P_{max} for $R_{min} = 0$ bps/Hz.

is equal, the curve with more reflect elements N can achieve higher average sum rate. This is because more signals can be reflected and more effective IRS reflection matrix can be obtained, when there are more reflect elements at IRS. Similarly, utilizing higher phase resolution, IRS can steer the incident signals with finer direction, and thus the curve with higher b can achieve more performance gains.

C. IRS versus AF Relay Performance Comparison

In this subsection, we compare the performance of achievable sum rate and EE with AF relay and benchmark schemes. The EE of the considered system enhanced by IRS can be describe as

$$EE = \frac{R_{sum}}{\eta \sum_{(s,k)\in\mathcal{C}} p_k^s + SP_{BS} + KP_u + NP_n},$$ (36)

where η is the circuit dissipated power coefficients, and P_{BS}, P_u and NP_n are the circuit dissipated power at BS, user and the IRS, respectively. We ignore NP_n for the benchmark scheme. In the contrast, for AF relay system, NP_n is replaced by $\eta_{AF}P_{AF}+NP_r$ which represents the sum of power consumed by AF realy. Meanwhile, we set $NP_n = 0$ for the system without IRS. Further, we adopt the typical values $\eta = 1.2$, $P_{BS} = 5$ dBW, $P_u = 10$ dBm, $P_n = 10$ dBm, while $\eta_{AF} = 1.2$, $P_{AF} = 2$ dBW and $P_r = 10$ dBm [19].

The sum rate and EE performances under different settings of P_{max} in dBm with $R_{min} =$

0 bps/Hz[5] are illustrated in Fig. 3. It can be seen from Fig. 3(a) that the sum rate performance of the IRS-assisted system outperforms other two schemes with vary settings. This is an obvious and predictable result mainly due to strong PBF gain brought by IRS. For AF relay, it is an active terminal with a dedicated transmitting amplifier circuit and is not restricted by the unit modulus that IRS has. Considering that the number of antennas of AF relay in practical applications is smaller than the number of antennas at BS, the reflection path gain brought by AF relay is still smaller than IRS. At the same time, we observe that the performance gap between IRS and AF relay becomes bigger as P_{max} increases. The main reason is that with the increase of P_{max}, the relay transmist power $P_{r,max}$ and sum rate become more and more irrelevant and both of them are actually affected by the BS transmist power. Specifically, when $P_{max} = 50$ dBm and $K = 24$, our proposed algorithm can reach up to 120% sum rate compared to AF relay and obtain 165% sum rate compared to the benchmark. It is also shown that the curve with $K = 24$ achieves higher sum rate than that of the curve with $K = 16$. This is because, as the number of users increases, the spectrum resource utilization rate and power efficiency of the system increase as well. Similar results can be observed in Fig. 3(b).

Fig. 3(b) presents the EE performance trend under the same settings as Fig. 3(a). In this simulation, our proposed algorithm can obtain the best EE, where the EE of our algorithm is 140% higher than that of relay when $P_{max} = 50$ dBm and $K = 16$. This is because the IRS-assisted system has strong PBF gain and lower power consumption, while the power consumption of relay is much greater. What's more, with increasing transmit power, the EE of all these algorithms increases, which is very similar to the trend of Fig. 3(a). Due to the channel characteristics of the mmWave, the power allocated by BS to each user is greatly affected by mmWave channel, which leads to a small change in the power allocated by each user within a large maximum power variation range. In other words, the performance of EE is mainly affected by the sum rate rather than the transmit power. However, we can foresee that the system performance cannot be infinitely improved by increasing transmit power in the actual mmWave system, since the excess BS transmit power is actually not used. In summary, our proposed algorithm can achieve significant EE gains within an acceptable sum rate range.

---

[5]$R_{min} = 0$ is equivalent to solving the initial problem (9) without considering the users' minimum rate constraints. Correspondingly, when solving the problem (26), R_{min} is set as a very small value to satisfy the use conditions of Algorithm 1.
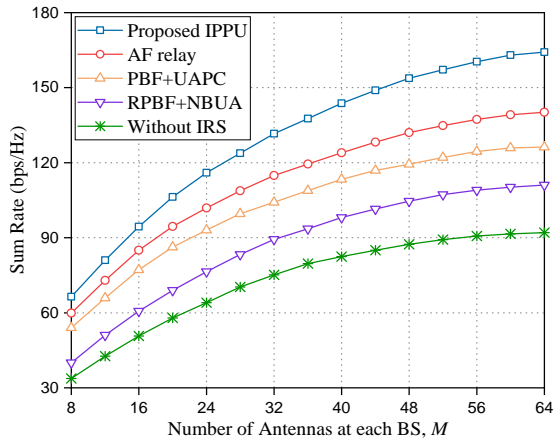
Fig. 4: Sum Rate versus M with $K = 16$, $N = 32$.
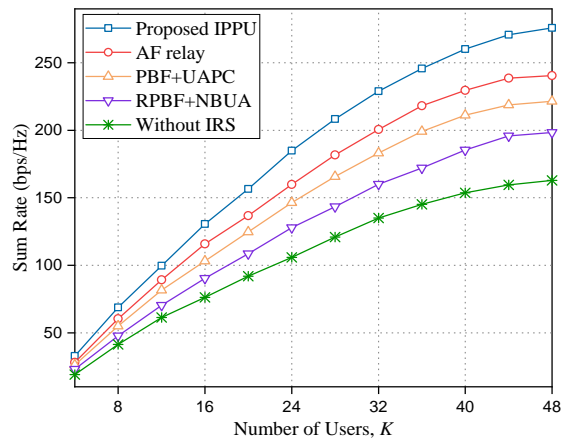
Fig. 5: Sum Rate versus K with $M = 32$, $N = 32$.

D. Impact of BS Parameter and User Parameter

In Fig. 4, we evaluate the achievable sum rate versus the number of antennas at each BS, where R_{min} is set to 2 bps/Hz, the following four figures (i.e., Fig. 4 - Fig. 8) are also the same setting. When the number of antennas increases, the sum rate increases and the proposed joint optimization IPPU algorithm gets the maximum sum rate compared with other three algorithms. The reason is that, as the M increases, the number of antennas available for beamforming increases as well, which implies active beamforming becomes more efficient and results in the achievable rate of users to be improved. Besides, the proposed algorithm achieves significant gain in sum rate due to better PBF, power allocation and UA gains, which validates the effectiveness of our joint optimization. In particular, when $M = 32$, the achievable sum rate of the IPPU algorithm is as high as 175% of the benchmark rate, and this conclusion can also correspond to the rate with $K = 16, P_{max} = 30$ dBm in Fig. 3(a).

Fig. 5 shows the impact of the number of users K on achievable sum rate of all users. We can observe that all four curves ascend as K increases within a certain range. This is because the more users are in the network, the higher utilization of network resources is obtained. Further, such phenomenon also results in greater total network benefit (i.e., sum rate of all users). Due to limited network resources, the sum rate cannot be infinitely improved by increasing K. Moreover, benefiting from the optimal gain brought by the PBF at IRS and the effective FRA-based UA algorithm, our proposed algorithm achieves optimal performance especially in high load conditions. The simulation result in Fig. 5 presents that when $K = 48$, our proposed algorithm can improve the performance compared with AF relay algorithm, UAPC algorithm,
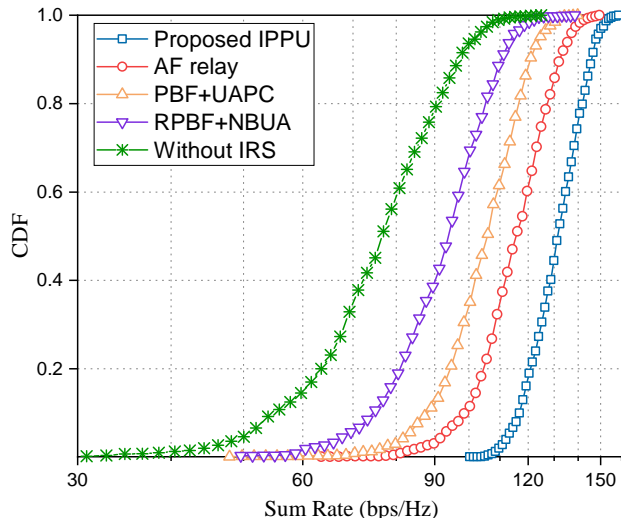
Fig. 6: CDFs of sum rate for four comparison algorithms with $K = 16$, $M = 32$ and $N = 32$.

NBUA algorithm and benchmark by 15%, 24%, 39% and 70%.

Next, we fix $K = 16$, $M = 32$, $N = 32$ and plot the cumulative distribution function (CDF) of sum rate as shown in Fig. 6. It can be seen that, the performance gains of the four algorithms are stable, and also keep consistent with their counterparts in Fig. 4 and Fig. 5. We can observe that our proposed algorithm has a probability of about 0.92 to obtain achievable rate $R_{sum} > 115.7$ bps/Hz (i.e., $R_k > 7.23$ bps/Hz). Therefore, we conclude that the performance of our proposed algorithm performs the best and can adapt to most of scenarios.

E. Coverage Performance and Average Rate Performance per User

Fig. 7 compares the average total number of users served by each BS of different algorithms, where we average the results of 10^3 scenarios and the same as below. Observing the number of users served by heavily blocked BS 1, it is obvious that our algorithm serves the most users compared with the other three algorithms except NBUA (UA is only related to physical distance) algorithm, and the benchmark without IRS serves the smallest number of users. This demonstrates that IRS can expand the coverage of mmWave signal and increase the number of users served by the blocked BS. In addition, the number of users served by BS 2 and BS 3 are relatively close under different algorithms, which is consistent with the actual situation.

Fig. 8 shows the average rate of users served by each BS. It can be seen that due to the optimization of the PBF at the IRS, both our algorithm and the UAPC algorithm can provide a higher communication rate for users served by BS 1 and AF relay also achieves a good average
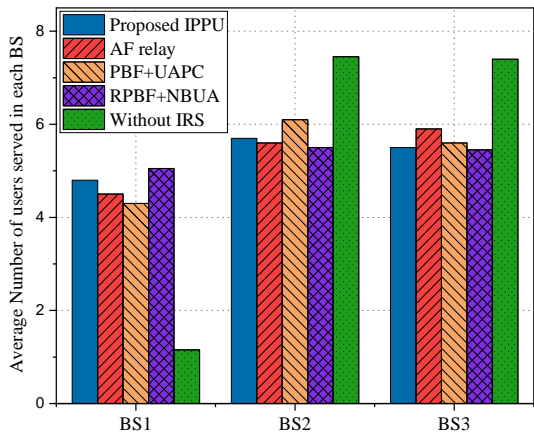
Fig. 7: Average number of users per BS served for four comparison algorithms with $K = 16$, $M = 32$ and $N = 32$.
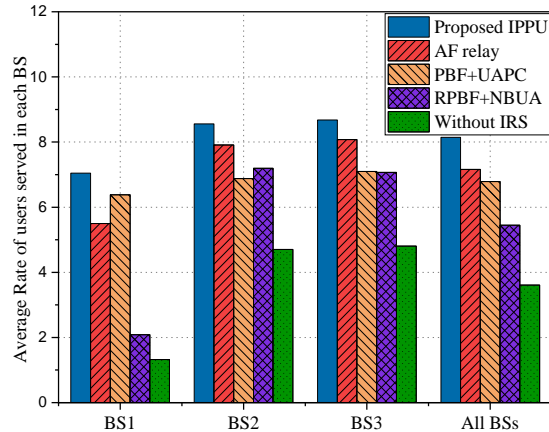
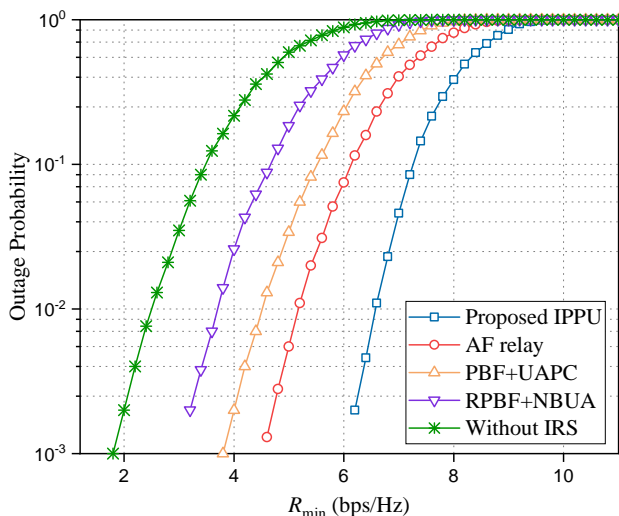Fig. 8: Average rate of users per BS served for four comparison algorithms with $K = 16$, $M = 32$ and $N = 32$.

Fig. 9: Outage probability versus R_{min} with $K = 16$, $M = 32$ and $N = 32$.

rate gain for similar reasons. In contrast, the remaining two algorithms result in lower average rates due to little or no PBF gain (i.e., the communication between users and BS 1 is greatly affected by obstacles). For the entire mmWave system, the relationship of average rate of users also corresponds to the results in previous subsections IV-C and IV-D.

F. Outage Probability Performance versus R_{min}

Finally, in order to investigate the impact of different minimum rate constraints on all algorithms and show the robustness of our IRS-assisted system against blockages, we calculate the

outage probability as follows

$$P_{out}(R_{min}) = \Pr\left(E[R_k] \leq R_{min}\right), \tag{37}$$

where $\Pr(\cdot)$ is the probability function. From Fig. 9, we observe that the outage probability can be substantially reduced by deploying IRS and exploiting our IPPU algorithm. Under the same minimum rate constraints, our proposed algorithm has the lowest outage probability of communication. Also, as R_{min} increases, the outage probability of user communication continues to rise. Especially, when $R_{min} = 8.4$ bps/Hz, the outage probability of these algorithms are close to 1 except our IPPU algorithm.

V. CONCLUSION

In this paper, we design an IRS-assisted multi-BS multi-user mmWave system to realize high robust, cost-effective and high coverage mmWave communication. With consideration of the impact of IRS on UA when multiple BSs are deployed, a maximization sum rate problem is formulated by jointly optimizing PBF at IRS with discrete phase constraint, power allocation with limited transmit power and UA with QoS constraints. Then, an efficient iterative algorithm based on alternate optimization is proposed to solve this non-convex problem. Extensive simulation results corroborate the feasibility and effectiveness of our proposed algorithm under various scenarios. Specifically, our proposed algorithm converges at an acceptable rate and obtains significant sum rate promotion. In particular, our proposed algorithm is able to provide up to 175% higher sum rate compared with the benchmark and 140% higher EE compared with AF relay.

In the future, it will be desirable to investigate multi-IRS assisted multi-BS mmWave communication with power allocation and UA. The collaborative work of multi-IRS can provide more optimization variables to tune, thereby further enhancing the performance of the system. Other meaningful research directions include the acquisition of CSI and the impact of channel estimation on the performance of IRS-assisted systems. Channel estimation at IRS is meaningful and important as we have mentioned, which will be more complicated in the IRS-assisted mmWave systems by jointly considering with power allocation and UA. And it is worthy to find a suitable channel estimation method for higher estimation accuracy with lower feedback overhead.

APPENDIX A

PROOF OF THEOREM 1

Rewrite the power allocation matrix $\mathbf{P}_i$ as $\mathbf{P}_i = \mathbf{Q}\mathbf{Q}^T$ with $\mathbf{Q} = \sqrt{\mathbf{P}_i}$, and let $\widetilde{\mathbf{H}}_r^i = \mathbf{Q}^{-1}\mathbf{H}_r^i$. Then the objective function $f_1(\mathbf{\Phi})$ in (11) can be rewritten as

$$
\begin{aligned}
f_1(\mathbf{\Phi}) &= \mathrm{tr}((\mathbf{Q}^{-1}\mathbf{H}_r^i\mathbf{\Phi}\mathbf{G}^i)^+(\mathbf{Q}^{-1}\mathbf{H}_r^i\mathbf{\Phi}\mathbf{G}^i)^{+H}) \\
&= \mathrm{tr}((\widetilde{\mathbf{H}}_r^i\mathbf{\Phi}\mathbf{G}^i)^+(\widetilde{\mathbf{H}}_r^i\mathbf{\Phi}\mathbf{G}^i)^{+H}) \\
&\overset{a}{=} \|(\mathbf{G}^i)^+\mathbf{\Phi}^{-1}(\widetilde{\mathbf{H}}_r^i)^+\|_F^2 \\
&\overset{b}{=} \|\,\mathrm{vec}((\mathbf{G}^i)^+\mathbf{\Phi}^{-1}(\widetilde{\mathbf{H}}_r^i)^+)\|^2 \\
&= \|(\widetilde{\mathbf{H}}_r^i)^{+H} \otimes (\mathbf{G}^i)^+\,\mathrm{vec}(\mathbf{\Phi}^{-1})\|^2 \\
&= \mathrm{vec}(\mathbf{\Phi}^{-1})^H((\widetilde{\mathbf{H}}_r^i)^{+H} \otimes (\mathbf{G}^i)^+)^H((\widetilde{\mathbf{H}}_r^i)^{+H} \otimes (\mathbf{G}^i)^+)\,\mathrm{vec}(\mathbf{\Phi}^{-1})
\end{aligned}
\tag{38}
$$

Here, we have applied the properties of the Frobenius matrix norm and the pseudo-inverse law of matrix product in step (a), whereas steps (b) follows from the vectorization operator. Thus, the optimization problem (12) is equivalent to (38).

REFERENCES

[1] S. Han, I. Chih-Lin, T. Xie, *et al.*, "Achieving high spectrum efficiency on high speed train for 5g new radio and beyond," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 62–69, 2019.

[2] L. Zhang, M. Xiao, G. Wu, M. Alam, Y.-C. Liang, and S. Li, "A survey of advanced techniques for spectrum sharing in 5g networks," *IEEE Wireless Communications*, vol. 24, no. 5, pp. 44–51, 2017.

[3] S. A. Busari, K. M. S. Huq, S. Mumtaz, *et al.*, "Millimeter-wave massive mimo communication for future wireless systems: A survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 2, pp. 836–869, 2017.

[4] Y. Niu, W. Ding, H. Wu, Y. Li, X. Chen, B. Ai, and Z. Zhong, "Relay-assisted and qos aware scheduling to overcome blockage in mmwave backhaul networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1733–1744, 2019.

[5] G. Yang, J. Du, and M. Xiao, "Maximum throughput path selection with random blockage for indoor 60 ghz relay networks," *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3511–3524, 2015.

[6] S. Zubair, S. Jangsher, Y. Mao, and V. O. Li, "Blockage-aware power allocation and relay selection in millimeter-wave small cell network," in *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pp. 1–5, IEEE, 2019.

[7] C. Yiu and S. Singh, "Empirical capacity of mmwave wlans," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1479–1487, 2009.

[8] Z. Genc, U. H. Rizvi, E. Onur, and I. Niemegeers, "Robust 60 ghz indoor connectivity: Is it possible with reflections?," in *2010 IEEE 71st vehicular technology conference*, pp. 1–5, IEEE, 2010.

[9]   N. S. Perović, M. Di Renzo, and M. F. Flanagan, "Channel capacity optimization using reconfigurable intelligent surfaces in indoor mmwave environments," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pp. 1–7, IEEE, 2020.

[10] P. Wang, J. Fang, and H. Li, "Joint beamforming for intelligent reflecting surface-assisted millimeter wave communications," *arXiv preprint arXiv:1910.08541*, 2019.

[11] P. Wang, J. Fang, X. Yuan, Z. Chen, H. Duan, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *arXiv preprint arXiv:1908.10734*, 2019.

[12] Y. Cao and T. Lv, "Intelligent reflecting surface aided multi-user millimeter-wave communications for coverage enhancement," *arXiv preprint arXiv:1910.02398*, 2019.

[13] D. Zhao, H. Lu, Y. Wang, and H. Sun, "Joint passive beamforming and user association optimization for irs-assisted mmwave systems," in *2020 IEEE Global Communications Conference (GLOBECOM), accepted, online*, 2020.

[14] X. Tan, Z. Sun, D. Koutsonikolas, and J. M. Jornet, "Enabling indoor mobile millimeter-wave networks based on smart reflect-arrays," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pp. 270–278, IEEE, 2018.

[15] Q. Zhang, W. Saad, and M. Bennis, "Reflections in the sky: Millimeter wave communication with uav-carried intelligent reflectors," in *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, IEEE, 2019.

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

[17] M. Di Renzo, K. Ntontin, J. Song, F. H. Danufane, X. Qian, F. Lazarakis, J. De Rosny, D.-T. Phan-Huy, O. Simeone, R. Zhang, *et al.*, "Reconfigurable intelligent surfaces vs. relaying: Differences, similarities, and performance comparison," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 798–807, 2020.

[18] 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.

[19] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, 2019.

[20] A. Kammoun, A. Chaaban, M. Debbah, M.-S. Alouini, *et al.*, "Asymptotic max-min sinr analysis of reconfigurable intelligent surface assisted miso systems," *IEEE Transactions on Wireless Communications*, 2020.

[21] Y. Cui and H. Yin, "An efficient csi acquisition method for intelligent reflecting surface-assisted mmwave networks," *arXiv preprint arXiv:1912.12076*, 2019.

[22] Y. Han, W. Tang, S. Jin, *et al.*, "Large intelligent surface-assisted wireless communication exploiting statistical csi," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 8238–8242, 2019.

[23] Y. Wang, H. Lu, D. Zhao, and H. Sun, "Energy efficiency optimization in irs-enhanced mmwave systems with lens antenna array," in *2020 IEEE Global Communications Conference (GLOBECOM), accepted, online*, 2020.

[24] T. Zhou, Z. Liu, J. Zhao, C. Li, and L. Yang, "Joint user association and power control for load balancing in downlink heterogeneous cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2582–2593, 2017.

[25] M. Feng, S. Mao, and T. Jiang, "Joint frame design, resource allocation and user association for massive mimo heterogeneous networks with wireless backhaul," *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1937–1950, 2017.

[26] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and user association optimization for massive mimo systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6384–6399, 2016.

[27] J. Zuo, J. Zhang, C. Yuen, *et al.*, "Energy efficient user association for cloud radio access networks," *IEEE Access*, vol. 4, pp. 2429–2438, 2016.

[28] F. Guo, H. Lu, D. Zhu, and Z. Gu, "Joint user association, grouping and power allocation in uplink noma systems with qos constraints," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2019.

[29] A. Alizadeh and M. Vu, "Load balancing user association in millimeter wave mimo networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 6, pp. 2932–2945, 2019.

[30] R. Liu, M. Lee, G. Yu, and G. Y. Li, "User association for millimeter-wave networks: A machine learning approach," *IEEE Transactions on Communications*, 2020.

[31] D. Astély, E. Dahlman, *et al.*, "Lte: the evolution of mobile broadband," *IEEE Communications magazine*, vol. 47, no. 4, pp. 44–51, 2009.

[32] Z. Lin, T. Lv, and P. T. Mathiopoulos, "3-d indoor positioning for millimeter-wave massive mimo systems," *IEEE Transactions on Communications*, vol. 66, no. 6, pp. 2472–2486, 2018.

[33] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2016.

[34] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna gaussian broadcast channels," *IEEE transactions on Information Theory*, vol. 51, no. 4, pp. 1570–1580, 2005.

[35] G. Athanasiou, P. C. Weeraddana, and C. Fischione, "Auction-based resource allocation in millimeterwave wireless access networks," *IEEE communications letters*, vol. 17, no. 11, pp. 2108–2111, 2013.

[36] D. P. Bertsekas, *Network optimization: continuous and discrete models*. Athena Scientific Belmont, MA, 1998.

[37] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave mimo systems," *IEEE transactions on wireless communications*, vol. 13, no. 3, pp. 1499–1513, 2014.