# Intelligent Reflecting Surface Enhanced Wireless Network via Joint Active and Passive Beamforming

Qingqing Wu, *Member, IEEE*, and Rui Zhang, *Fellow, IEEE*

***Abstract*— Intelligent reflecting surface (IRS) is a revolutionary and transformative technology for achieving spectrum and energy efficient wireless communication cost-effectively in the future. Specifically, an IRS consists of a large number of low-cost passive elements each being able to reflect the incident signal independently with an adjustable phase shift so as to collaboratively achieve three-dimensional (3D) passive beamforming without the need of any transmit radio-frequency (RF) chains. In this paper, we study an IRS-aided single-cell wireless system where one IRS is deployed to assist in the communications between a multi-antenna access point (AP) and multiple single-antenna users. We formulate and solve new problems to minimize the total transmit power at the AP by jointly optimizing the transmit beamforming by active antenna array at the AP and reflect beamforming by passive phase shifters at the IRS, subject to users' individual signal-to-interference-plus-noise ratio (SINR) constraints. Moreover, we analyze the asymptotic performance of IRS's passive beamforming with infinitely large number of reflecting elements and compare it to that of the traditional active beamforming/relaying. Simulation results demonstrate that an IRS-aided MIMO system can achieve the same rate performance as a benchmark massive MIMO system without using IRS, but with significantly reduced active antennas/RF chains. We also draw useful insights into optimally deploying IRS in future wireless systems.**

***Index Terms*— Intelligent reflecting surface, joint active and passive beamforming, phase shift optimization.**

## I. Introduction

TO ACHIEVE 1,000-fold network capacity increase and ubiquitous wireless connectivity for at least 100 billion devices in the forthcoming fifth-generation (5G) networks, a variety of wireless technologies have been proposed and thoroughly investigated in the last decade, including most prominently the ultra-dense network (UDN), massive multiple-input multiple-output (MIMO), and millimeter wave (mmWave) communication [2]. However, the network energy consumption and hardware cost still remain critical issues in practical systems [3]. For example, UDNs almost linearly scale up the circuit and cooling energy consumption with the number of deployed base stations (BSs), while costly radio frequency (RF) chains and complex signal processing are needed for achieving high-performance communication at mmWave frequencies, especially when massive MIMO is employed to exploit the small wavelengths. Moreover, adding an excessively large number of active components such as small-cell BSs/relays/remote radio heads (RRHs) in wireless networks also causes a more aggravated interference issue. As such, innovative research on finding both spectrum and energy efficient techniques with low hardware cost is still imperative for realizing a sustainable wireless network evolution with scalable cost in the future [4].

In this paper, intelligent reflecting surface (IRS) is proposed as a promising new solution to achieve the above goal. Specifically, IRS is a planar array consisting of a large number of reconfigurable passive elements (e.g., low-cost printed dipoles), where each of the elements is able to induce a certain phase shift (controlled by an attached smart controller) independently on the incident signal, thus collaboratively changing the reflected signal propagation [5]. Although passive reflecting surfaces have found a variety of applications in radar systems, remote sensing, and satellite/deep-space communications, they were rarely used in mobile wireless communication. This is mainly because traditional reflecting surfaces only have fixed phase shifters once fabricated, which are unable to cater to the dynamic wireless channels arising from user mobility. However, recent advances in RF micro electromechanical systems (MEMS) and metamaterial (e.g., metasurface) have made the reconfigurability of reflecting surfaces possible, even by controlling the phase shifters in real time [6]. By smartly adjusting the phase shifts of all passive elements at the IRS, the reflected signals can add coherently with the signals from other paths at the desired receiver to boost the received signal power or destructively at non-intended receivers to suppress interference as well as enhancing security/privacy [5].

It is worth noting that the proposed IRS differs significantly from other related existing technologies such as amplify-and-forward (AF) relay, backscatter communication, and active intelligent surface based massive MIMO. First, compared to the AF relay that assists in source-destination transmission by amplifying and regenerating signals, IRS does not use a transmitter module but only reflects the received signals as a passive array, which thus incurs no transmit power consumption.[1] Furthermore, active AF relay usually operates in half-duplex (HD) mode and thus is less spectrally efficient than the

Manuscript received May 15, 2019; revised August 6, 2019; accepted August 12, 2019. Date of publication August 23, 2019; date of current version November 11, 2019. This article was presented in part at the IEEE GLOBECOM 2018 [1]. The associate editor coordinating the review of this article and approving it for publication was K. Kansanen. *(Corresponding author: Qingqing Wu.)*

The authors are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583 (e-mail: elewuqq@nus.edu.sg; elezhang@nus.edu.sg).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2019.2936025

[1] Although using devices like MEMs as mentioned previously to adjust the phase shifts at the IRS requires some power consumption, it is practically negligible as compared to the much higher transmit power of active communication devices.

TABLE I
COMPARISON OF IRS WITH OTHER RELATED TECHNOLOGIES

| Technology | Operating mechanism | Duplex | No. of transmit RF chains needed | Hardware cost | Energy consumption | Role |
|---|---|---|---|---|---|---|
| IRS | Passive, reflect | Full duplex | 0 | Low | Low | Helper |
| Backscatter | Passive, reflect | Full duplex | 0 | Very low | Very low | Source |
| MIMO relay | Active, receive and transmit | Half/full duplex | $N$ | High | High | Helper |
| Massive MIMO | Active, transmit/receive | Half/full duplex | $N$ | Very high | Very high | Source/ Destination |

proposed IRS operating in full-duplex (FD) mode. Although AF relay can also work in FD, it inevitably suffers from the severe self-interference, which needs effective interference cancellation techniques. Second, different from the traditional backscatter communication of the radio frequency identification (RFID) tag that communicates with the receiver by reflecting the signal sent from a reader, IRS is utilized mainly to enhance the existing communication link performance instead of delivering its own information by reflection. As such, the direct-path signal (from reader to receiver) in backscatter communication is undesired interference and hence needs to be canceled/suppressed at the receiver. However, in IRS-enhanced communication, both the direct-path and reflect-path signals carry the same useful information and thus can be coherently added at the receiver to maximize the total received power. Third, IRS is also different from the active intelligent surface based massive MIMO [7] due to their different array architectures (passive versus active) and operating mechanisms (reflect versus transmit). A more detailed comparison between the above technologies and IRS is summarized in Table I, where $N$ denotes the number of active antennas in massive MIMO or MIMO relay.

On the other hand, from the implementation perspective, IRSs possess appealing advantages such as low profile, lightweight, and conformal geometry, which enable them to be easily attached/removed to/from the wall or ceiling, thus providing high flexibility for their practical deployment [8]. For example, by installing IRSs on the walls/ceilings which are in line-of-sight (LoS) with an access point (AP)/BS, the signal strength in the vicinity of each IRS can be significantly improved. In addition, integrating IRSs into the existing networks (such as cellular or WiFi) can be made transparent to the users without the need of any change in the hardware and software of their devices. All the above features make IRS a compelling new technology for future wireless networks, particularly in indoor applications with high density of users (such as stadium, shopping mall, exhibition center, airport, etc.). To validate the feasibility of IRS, an experimental testbed for a two-user setup was developed [9], where the spectral efficiency is shown to be greatly improved by using the IRS. However, the research on IRS design as well as the performance analysis and optimization for IRS-aided wireless communication systems is still in its infancy, which thus motivates this work.

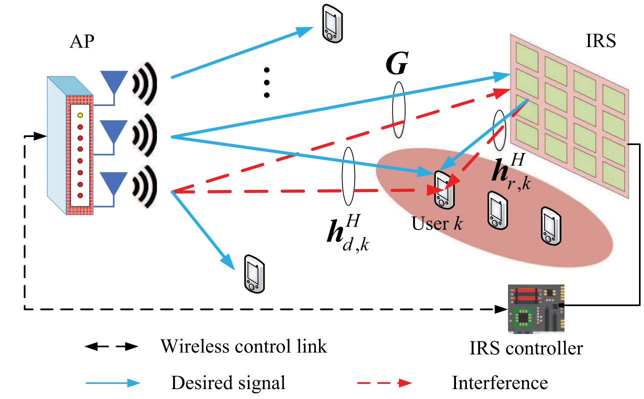


Fig. 1. An IRS-aided multiuser communication system.

In this paper, we consider an IRS-aided multiuser multiple-input single-output (MISO) communication system in a single cell as shown in Fig. 1, where a multi-antenna AP serves multiple single-antenna users with the help of an IRS. Since each user in general receives the superposed (desired as well as interference) signals from both the AP-user (direct) link and AP-IRS-user (reflected) link, we jointly optimize the (active) transmit beamforming at the AP and (passive) reflect beamforming by the phase shifters at the IRS to minimize the total transmit power at the AP, under a given set of signal-to-interference-plus-noise ratio (SINR) constraints at the user receivers. For the special case of single-user transmission without any interference, it is intuitive that the AP should beam toward the user directly if the channel of the AP-user link is much stronger than that of the AP-IRS link; while in the opposite case, especially when the AP-user link is severally blocked by obstacles (e.g., thick walls in indoor applications), the AP ought to adjust its beam toward the IRS to maximally leverage its reflected signal to serve the user (i.e., by creating a virtual LoS link with the user to bypass the obstacle). In this case, a large number of reflecting elements with adjustable phases at the IRS can focus the signal into a sharp beam toward the user, thus achieving a high beamforming gain similarly as by the conventional massive MIMO [10], but only via a passive array with significantly reduced energy consumption and hardware cost.

Moreover, under the general multiuser setup, an IRS-aided system will benefit from two main aspects: the beamforming of desired signal as in the single-user case as well as the spatial interference suppression among the users. Specifically, a user near the IRS is expected to be able to tolerate more interference from the AP as compared to the user farther away from the IRS, because the phase shifts of the IRS can be tuned such that the interference reflected by the IRS can add destructively with that from the AP-user link at the near user to suppress its overall received interference. This thus provides more flexibility for designing the transmit beamforming at the AP for serving the other users outside the IRS's covered region, so as to improve the SINR performance of all users in the system. Therefore, the transmit beamforming at the AP needs to be jointly designed with the phase shifts at the IRS based on all the AP-IRS, IRS-users, and AP-users channels in order to fully reap the network beamforming gain. However, this design problem is difficult to be solved optimally in general, due to the non-convex SINR constraints as well as the signal unit-modulus constraints imposed by passive phase shifters. Although beamforming optimization under unit-modulus constraints has been studied in the research on constant-envelope precoding [11], [12] as well as hybrid digital/analog processing [13], [14], such designs are mainly restricted to either the transmitter or the receiver side, which are not applicable to our considered joint active and passive beamforming optimization at both the AP and IRS.

To tackle this new problem, we first consider a single-user setup and apply the semidefinite relaxation (SDR) technique to obtain a high-quality approximate solution as well as a lower bound of the optimal value to evaluate the tightness of approximate solutions. To reduce the computational complexity, we further propose an efficient algorithm based on the alternating optimization of the phase shifts and transmit beamforming vector in an iterative manner, where their optimal solutions are derived in closed-form with the other being fixed. Then, we extend our designs for the single-user case to the general multiuser setting, and propose two algorithms to obtain suboptimal solutions that also offer different tradeoffs between performance and complexity. Numerical results demonstrate that the required transmit power at the AP to meet users' SINR targets can be considerably reduced by deploying the IRS as compared to the conventional setup without using IRS for both single-user and multiuser setups. In particular, for serving a single-user in the vicinity of the IRS, it is shown that the AP's transmit power decreases with the number of reflecting elements $N$ at the IRS in the order of $N^2$ when $N$ is sufficiently large, which is consistent with the performance scaling law derived analytically. Note that in [15], the authors also considered the use of passive intelligent mirror (analogous to IRS) to enhance the sum-rate in a multiuser system. This paper differs from [15] in the following two main aspects. First, to simplify the system model and algorithm design, [15] ignored the direct channels from the AP to users, while this paper considers the more general setting with the AP-user direct channels considered. Second, [15] adopted the suboptimal zero-forcing (ZF) based precoding at the AP to simplify the optimization of passive phase shifters, while in this paper we optimize AP transmit precoding jointly with IRS's phase shifts. As such, the algorithm proposed in [15] is not applicable to solving the formulated problems in this paper.

The rest of this paper is organized as follows. Section II introduces the system model and the problem formulation for designing the IRS-aided wireless network. In Sections III and IV, we propose efficient algorithms to solve the formulated problems in the single-user and multiuser cases, respectively. Section V presents numerical results to evaluate the performance of the proposed designs. Finally, we conclude the paper in Section VI.

*Notations:* Scalars are denoted by italic letters, vectors and matrices are denoted by bold-face lower-case and upper-case letters, respectively. $\mathbb{C}^{x\times y}$ denotes the space of $x \times y$ complex-valued matrices. For a complex-valued vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|$ denotes its Euclidean norm, $\arg(\boldsymbol{x})$ denotes a vector with each element being the phase of the corresponding element in $\boldsymbol{x}$, and $\text{diag}(\boldsymbol{x})$ denotes a diagonal matrix with each diagonal element being the corresponding element in $\boldsymbol{x}$. The distribution of a circularly symmetric complex Gaussian (CSCG) random vector with mean vector $\boldsymbol{x}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{CN}(\boldsymbol{x}, \boldsymbol{\Sigma})$; and $\sim$ stands for "distributed as". For a square matrix $\boldsymbol{S}$, $\text{tr}(\boldsymbol{S})$ and $\boldsymbol{S}^{-1}$ denote its trace and inverse, respectively, while $\boldsymbol{S} \succeq \boldsymbol{0}$ means that $\boldsymbol{S}$ is positive semi-definite. For any general matrix $\boldsymbol{M}$, $\boldsymbol{M}^H$, $\text{rank}(\boldsymbol{M})$, and $\boldsymbol{M}_{i,j}$ denote its conjugate transpose, rank, and $(i,j)$th element, respectively. $\boldsymbol{I}$ and $\boldsymbol{0}$ denote an identity matrix and an all-zero matrix, respectively, with appropriate dimensions. $\mathbb{E}(\cdot)$ denotes the statistical expectation. $\text{Re}\{\cdot\}$ denotes the real part of a complex number.

## II. System Model and Problem Formulation

### A. System Model

As shown in Fig. 1, we consider the IRS-aided downlink communications in a single-cell network where an IRS is deployed to assist in the communications from a multi-antenna AP to $K$ single-antenna users over a given frequency band. The set of the users is denoted by $\mathcal{K}$. The number of transmit antennas at the AP and that of reflecting units at the IRS are denoted by $M$ and $N$, respectively. The IRS is equipped with a controller that coordinates its switching between two working modes, i.e., receiving mode for channel estimation and reflecting mode for data transmission [8]. Due to the high path loss, it is assumed that the power of the signals that are reflected by the IRS two or more times is negligible and thus ignored. To characterize the theoretical performance gain brought by the IRS, we assume that the channel state information (CSI) of all channels involved is perfectly known at the AP. In addition, the quasi-static flat-fading model is adopted for all channels. Since the IRS is a passive reflecting device, we consider a time-division duplexing (TDD) protocol for uplink and downlink transmissions and assume channel reciprocity for the CSI acquisition in the downlink based on the uplink training.

The baseband equivalent channels from the AP to IRS, from the IRS to user $k$, and from the AP to user $k$ are denoted by $\boldsymbol{G} \in \mathbb{C}^{N\times M}$, $\boldsymbol{h}_{r,k}^H \in \mathbb{C}^{1\times N}$, and $\boldsymbol{h}_{d,k}^H \in \mathbb{C}^{1\times M}$, respectively, with $k = 1, \cdots, K$. It is worth noting that the reflected channel from the AP to each user via the IRS is usually referred to as a dyadic backscatter channel in RFID communications [16], which behaves different from the AP-user direct channel. Specifically, each element of the IRS receives the superposed multi-path signals from the transmitter, and then scatters the combined signal with adjustable amplitude and/or phase as if from a single point source. Let $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_N]$ and define a diagonal matrix $\boldsymbol{\Theta} = \text{diag}(\beta_1 e^{j\theta_1}, \cdots, \beta_N e^{j\theta_N})$ (with $j$ denoting the imaginary unit) as the reflection-coefficients matrix of the IRS, where $\theta_n \in [0, 2\pi)$ and $\beta_n \in [0, 1]$ denote the phase shift[2] and the amplitude reflection coefficient[3] of the $n$th element of the IRS, respectively. The composite AP-IRS-user channel is thus modeled as a concatenation of three components, namely, the AP-IRS link, IRS reflection with phase shifts, and IRS-user link.

In this paper, we consider linear transmit precoding at the AP where each user is assigned with one dedicated beamforming vector. Hence, the complex baseband transmitted signal at the AP can be expressed as $\boldsymbol{x} = \sum_{k=1}^{K} \boldsymbol{w}_k s_k$, where $s_k$ denotes the transmitted data for user $k$ and $\boldsymbol{w}_k \in \mathbb{C}^{M\times 1}$ is the corresponding beamforming vector. It is assumed that $s_k$, $k = 1, \cdots, K$, are independent random variables with zero mean and unit variance (normalized power). The signal received at user $k$ from both the AP-user and AP-IRS-user channels is then expressed as

$$y_k = (\boldsymbol{h}_{r,k}^H \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{h}_{d,k}^H) \sum_{j=1}^{K} \boldsymbol{w}_j s_j + n_k, \quad k = 1, \cdots, K, \tag{1}$$

where $n_k \sim \mathcal{CN}(0, \sigma_k^2)$ denotes the additive white Gaussian noise (AWGN) at the user $k$'s receiver. Accordingly, the SINR of user $k$ is given by

$$\text{SINR}_k = \frac{|(\boldsymbol{h}_{r,k}^H \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{h}_{d,k}^H) \boldsymbol{w}_k|^2}{\sum_{j\neq k}^{K} |(\boldsymbol{h}_{r,k}^H \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{h}_{d,k}^H) \boldsymbol{w}_j|^2 + \sigma_k^2}, \quad \forall k. \tag{2}$$

### *B. Problem Formulation*

Let $\boldsymbol{W} = [\boldsymbol{w}_1, \cdots, \boldsymbol{w}_K] \in \mathbb{C}^{M\times K}$, $\boldsymbol{H}_r = [\boldsymbol{h}_{r,1}, \cdots, \boldsymbol{h}_{r,K}] \in \mathbb{C}^{N\times K}$, and $\boldsymbol{H}_d = [\boldsymbol{h}_{d,1}, \cdots, \boldsymbol{h}_{d,K}] \in \mathbb{C}^{M\times K}$. In this paper, we aim to minimize the total transmit power at the AP by jointly optimizing the transmit beamforming at the AP and reflect beamforming at the IRS, subject to individual SINR constraints at all users. Accordingly, the problem is formulated as

$$\text{(P1)}: \min_{\boldsymbol{W}, \boldsymbol{\theta}} \sum_{k=1}^{K} \|\boldsymbol{w}_k\|^2 \tag{3}$$

$$\text{s.t.} \ \frac{|(\boldsymbol{h}_{r,k}^H \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{h}_{d,k}^H) \boldsymbol{w}_k|^2}{\sum_{j\neq k}^{K} |(\boldsymbol{h}_{r,k}^H \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{h}_{d,k}^H) \boldsymbol{w}_j|^2 + \sigma_k^2} \geq \gamma_k, \ \forall k, \tag{4}$$

$$0 \leq \theta_n \leq 2\pi, n = 1, \cdots, N, \tag{5}$$

where $\gamma_k > 0$ is the minimum SINR requirement of user $k$. Although the objective function of (P1) and constraints in (5) are convex, it is challenging to solve (P1) due to the non-convex constraints in (4) where the transmit beamforming and phase shifts are coupled. In general, there is no standard method for solving such non-convex optimization problems optimally. Nevertheless, in the next section, we apply the SDR and alternating optimization techniques, respectively, to solve (P1) approximately for the single-user case, which are then generalized to the multiuser case. Prior to solving problem (P1), we present a sufficient condition for its feasibility as follows. Let $\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_K] \in \mathbb{C}^{M\times K}$ where $\boldsymbol{h}_k^H = \boldsymbol{h}_{r,k}^H \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{h}_{d,k}^H$, $\forall k$.

*Proposition 1:* Problem (P1) is feasible for any finite user SINR targets $\gamma_k$'s if $\text{rank}(\boldsymbol{G}^H \boldsymbol{H}_r + \boldsymbol{H}_d) = K$.

*Proof:* If $\text{rank}(\boldsymbol{G}^H \boldsymbol{H}_r + \boldsymbol{H}_d) = K$, the (right) pseudo inverse of $\boldsymbol{H}^H = \boldsymbol{H}_r^H \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{H}_d^H$ exists with $\boldsymbol{\Theta} = \boldsymbol{I}$ and the precoding matrix $\boldsymbol{W}$ at the AP can be set as

$$\boldsymbol{W} = \boldsymbol{H}(\boldsymbol{H}^H \boldsymbol{H})^{-1} \text{diag}(\gamma_1 \sigma_1^2, \cdots, \gamma_k \sigma_k^2)^{\frac{1}{2}}. \tag{6}$$

It is easy to verify that the above solution allows all users to achieve their corresponding $\gamma_k$'s and thus (P1) is feasible. ■

Thanks to the additional AP-IRS-user link, the rank condition in Proposition 1 is practically easier to be satisfied in an IRS-aided system, as compared to that in the case without the IRS, i.e., $\text{rank}(\boldsymbol{H}_d) = K$. For instance, if the AP-user direct channels of two users lie in the same direction, then $\text{rank}(\boldsymbol{H}_d) = K$ does not hold. While the rank condition in an IRS-aided system may still hold since the combined AP-user channels (including both the AP-user direct and AP-IRS-user reflected links) of these two users are unlikely to be aligned too, due to the additional IRS reflected paths.

## III. Single-User System

In this section, we consider the single-user setup, i.e., $K = 1$, to draw important insights into the optimal joint beamforming design. In this case, no inter-user interference is present, and thus (P1) is simplified to (by dropping the user index)

$$\text{(P2)}: \min_{\boldsymbol{w}, \boldsymbol{\theta}} \|\boldsymbol{w}\|^2 \tag{7}$$

$$\text{s.t.} \ |(\boldsymbol{h}_r^H \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{h}_d^H) \boldsymbol{w}|^2 \geq \gamma \sigma^2, \tag{8}$$

$$0 \leq \theta_n \leq 2\pi, \quad n = 1, \cdots, N. \tag{9}$$

Although much simplified, problem (P2) is still a non-convex optimization problem since the left-hand-side (LHS) of (8) is not jointly concave with respect to $\boldsymbol{w}$ and $\boldsymbol{\theta}$. In the next

[2] To characterize the fundamental performance limits of IRS, we assume that the phase shifts can be continuously varied in $[0, 2\pi)$, while in practice they are usually selected from a finite number of discrete values from 0 to $2\pi$ for the ease of circuit implementation. The design of IRS with discrete phase shifts is addressed in our follow-up work [17].

[3] In practice, each element of the IRS is usually designed to maximize the signal reflection. Thus, we set $\beta_n = 1, \forall n$, in the sequel of this paper for simplicity. Note that this scenario is different from the traditional backscatter communication where the RFID tags usually need to harvest a certain amount of energy from the incident signals for powering their circuit operation and thus a much smaller amplitude reflection coefficient than unity is resulted in general.

two subsections, we solve (P2) by applying the SDR and alternating optimization techniques, respectively, which will be extended to the general multiuser system in the next section.

### *A. SDR*

We first apply SDR to solve problem (P2), which also helps obtain a lower bound of the optimal value of (P2) for evaluating the performance gaps from other suboptimal solutions. For any given phase shift $\boldsymbol{\theta}$, it is known that the maximum-ratio transmission (MRT) is the optimal transmit beamforming solution to problem (P2) [18], i.e., $\boldsymbol{w}^* = \sqrt{P}\frac{(\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_d^H)^H}{\|\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_d^H\|}$, where $P$ denotes the transmit power of the AP. Substituting $\boldsymbol{w}^*$ to problem (P2) yields the following problem

$$\min_{P,\boldsymbol{\theta}} \ P \tag{10}$$

$$\text{s.t. } P\|\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_d^H\|^2 \geq \gamma\sigma^2, \tag{11}$$

$$0 \leq \theta_n \leq 2\pi, \forall n. \tag{12}$$

It is not difficult to verify that the optimal transmit power satisfies $P^* = \frac{\gamma\sigma^2}{\|\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_d^H\|^2}$. As such, minimizing the transmit power is equivalent to maximizing the channel power gain of the combined channel, i.e.,

$$\max_{\boldsymbol{\theta}} \ \|\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_d^H\|^2 \tag{13}$$

$$\text{s.t. } 0 \leq \theta_n \leq 2\pi, \forall n. \tag{14}$$

Let $\boldsymbol{v} = [v_1, \cdots, v_N]^H$ where $v_n = e^{j\theta_n}$, $\forall n$. Then, constraints in (14) are equivalent to the unit-modulus constraints: $|v_n|^2 = 1, \forall n$. By applying the change of variables $\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G} = \boldsymbol{v}^H\boldsymbol{\Phi}$ where $\boldsymbol{\Phi} = \text{diag}(\boldsymbol{h}_r^H)\boldsymbol{G} \in \mathbb{C}^{N\times M}$, we have $\|\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_d^H\|^2 = \|\boldsymbol{v}^H\boldsymbol{\Phi}+\boldsymbol{h}_d^H\|^2$. Thus, problem (13) is equivalent to

$$\max_{\boldsymbol{v}} \ \boldsymbol{v}^H\boldsymbol{\Phi}\boldsymbol{\Phi}^H\boldsymbol{v} + \boldsymbol{v}^H\boldsymbol{\Phi}\boldsymbol{h}_d + \boldsymbol{h}_d^H\boldsymbol{\Phi}^H\boldsymbol{v} + \|\boldsymbol{h}_d^H\|^2 \tag{15}$$

$$\text{s.t. } |v_n|^2 = 1, \quad \forall n. \tag{16}$$

Note that problem (15) is a non-convex quadratically constrained quadratic program (QCQP), which can be reformulated as a homogeneous QCQP [19]. Specifically, by introducing an auxiliary variable $t$, problem (15) is equivalently written as

$$\max_{\bar{\boldsymbol{v}}} \ \bar{\boldsymbol{v}}^H\boldsymbol{R}\bar{\boldsymbol{v}} + \|\boldsymbol{h}_d^H\|^2 \tag{17}$$

$$\text{s.t. } |\bar{v}_n|^2 = 1, \quad n = 1, \cdots, N+1, \tag{18}$$

where

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{\Phi}\boldsymbol{\Phi}^H & \boldsymbol{\Phi}\boldsymbol{h}_d \\ \boldsymbol{h}_d^H\boldsymbol{\Phi}^H & 0 \end{bmatrix}, \quad \bar{\boldsymbol{v}} = \begin{bmatrix} \boldsymbol{v} \\ t \end{bmatrix}.$$

However, problem (17) is still non-convex in general [19]. Note that $\bar{\boldsymbol{v}}^H\boldsymbol{R}\bar{\boldsymbol{v}} = \text{tr}(\boldsymbol{R}\bar{\boldsymbol{v}}\bar{\boldsymbol{v}}^H)$. Define $\boldsymbol{V} = \bar{\boldsymbol{v}}\bar{\boldsymbol{v}}^H$, which needs to satisfy $\boldsymbol{V} \succeq \boldsymbol{0}$ and $\text{rank}(\boldsymbol{V}) = 1$. Since the rank-one constraint is non-convex, we apply SDR to relax this constraint. As a result, problem (17) is reduced to

$$\max_{\boldsymbol{V}} \ \text{tr}(\boldsymbol{R}\boldsymbol{V}) + \|\boldsymbol{h}_d^H\|^2 \tag{19}$$

$$\text{s.t. } \boldsymbol{V}_{n,n} = 1, \quad n = 1, \cdots, N+1, \tag{20}$$

$$\boldsymbol{V} \succeq 0. \tag{21}$$

As problem (19) is a convex semidefinite program (SDP), it can be optimally solved by existing convex optimization solvers such as CVX [20]. Generally, the relaxed problem (19) may not lead to a rank-one solution, i.e., $\text{rank}(\boldsymbol{V}) \neq 1$, which implies that the optimal objective value of problem (19) only serves an upper bound of problem (17). Thus, additional steps are needed to construct a rank-one solution from the obtained higher-rank solution to problem (19), while the details can be found in [1] and thus are omitted here. It has been shown that such an SDR approach followed by a sufficiently large number of randomizations guarantees at least a $\frac{\pi}{4}$-approximation of the optimal objective value of problem (19) [19].

### *B. Alternating Optimization*

To achieve lower complexity than the SDR-based solution presented in the preceding subsection, we propose an alternative suboptimal algorithm in this subsection based on alternating optimization. Specifically, the transmit beamforming direction and transmit power at the AP are optimized iteratively with the phase shifts at the IRS in an alternating manner, until the convergence is achieved.

Let $\boldsymbol{w} = \sqrt{P}\bar{\boldsymbol{w}}$ where $\bar{\boldsymbol{w}}$ denotes the transmit beamforming direction and $P$ is the transmit power. For fixed transmit beamforming direction $\bar{\boldsymbol{w}}$, (P2) is reduced to a joint transmit power and phase shifts optimization problem which can be formulated as (similar to (10) and (13)),

$$\max_{\boldsymbol{\theta}} \ |(\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_d^H)\bar{\boldsymbol{w}}|^2 \tag{22}$$

$$\text{s.t. } 0 \leq \theta_n \leq 2\pi, \quad n = 1, \cdots, N. \tag{23}$$

Although being non-convex, the above problem admits a closed-form solution by exploiting the special structure of its objective function. Specifically, we have the following inequality:

$$\begin{aligned} |(\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_d^H)\bar{\boldsymbol{w}}| &= |\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}\bar{\boldsymbol{w}}+\boldsymbol{h}_d^H\bar{\boldsymbol{w}}| \\ &\overset{(a)}{\leq} |\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}\bar{\boldsymbol{w}}| + |\boldsymbol{h}_d^H\bar{\boldsymbol{w}}|, \end{aligned} \tag{24}$$

where $(a)$ is due to the triangle inequality and the equality holds if and only if $\arg(\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}\bar{\boldsymbol{w}}) = \arg(\boldsymbol{h}_d^H\bar{\boldsymbol{w}}) \triangleq \varphi_0$. Next, we show that there always exists a solution $\boldsymbol{\theta}$ that satisfies $(a)$ with equality as well as the phase shift constraints in (23). Let $\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}\boldsymbol{w} = \boldsymbol{v}^H\boldsymbol{a}$ where $\boldsymbol{v} = [e^{j\theta_1}, \cdots, e^{j\theta_N}]^H$ and $\boldsymbol{a} = \text{diag}(\boldsymbol{h}_r^H)\boldsymbol{G}\bar{\boldsymbol{w}}$. With (24), problem (22) is equivalent to

$$\max_{\boldsymbol{v}} \ |\boldsymbol{v}^H\boldsymbol{a}|^2 \tag{25}$$

$$\text{s.t. } |v_n| = 1, \forall n = 1, \cdots, N, \tag{26}$$

$$\arg(\boldsymbol{v}^H\boldsymbol{a}) = \varphi_0. \tag{27}$$

It is not difficult to show that the optimal solution to the above problem is given by $\boldsymbol{v}^* = e^{j(\varphi_0 - \arg(\boldsymbol{a}))} = e^{j(\varphi_0 - \arg(\text{diag}(\boldsymbol{h}_r^H)\boldsymbol{G}\bar{\boldsymbol{w}}))}$. Thus, the $n$th phase shift at the IRS is given by

$$\begin{aligned} \theta_n^* &= \varphi_0 - \arg(h_{n,r}^H\boldsymbol{g}_n^H\bar{\boldsymbol{w}}) \\ &= \varphi_0 - \arg(h_{n,r}^H) - \arg(\boldsymbol{g}_n^H\bar{\boldsymbol{w}}), \end{aligned} \tag{28}$$

where $h_{n,r}^H$ is the $n$th element of $\boldsymbol{h}_r^H$ and $\boldsymbol{g}_n^H$ is the $n$th row vector of $\boldsymbol{G}$. Note that $\boldsymbol{g}_n^H\bar{\boldsymbol{w}}$ combines the transmit beamforming and the AP-IRS channel, which can be regarded as the effective channel perceived by the $n$th reflecting element at the IRS. Therefore, (28) suggests that the $n$th phase shift should be tuned such that the phase of the signal that passes through the AP-IRS and IRS-user links is aligned with that of the signal over the AP-user direct link to achieve coherent signal combining at the user. Furthermore, it is interesting to note that the obtained phase $\theta_n^*$ is independent of the amplitude of $h_{n,r}$. As a result, the optimal transmit power is given by $P^* = \frac{\gamma\sigma^2}{\|(\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_d^H)\bar{\boldsymbol{w}}\|^2}$ from (P2). Next, we optimize the transmit beamforming direction for given $\boldsymbol{\theta}$ in (28). As in Section III-A, the combined AP-user channel is given by $\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G} + \boldsymbol{h}_d^H$ and hence MRT is optimal, i.e., $\bar{\boldsymbol{w}}^* = \frac{(\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_d^H)^H}{\|\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_d^H\|}$. The above alternating optimization approach is practically appealing since both the transmit beamforming and phase shifts are obtained in closed-form expressions, without invoking the SDP solver. Its convergence is guaranteed by the following two facts. First, for each subproblem, the optimal solution is obtained which ensures that the objective value of (P2) is non-increasing over iterations. Second, the optimal value of (P2) is bounded from below due to the SNR constraint. Thus, the proposed algorithm is guaranteed to converge.

### C. Power Scaling Law With Infinitely Large Surface

Next, we characterize the scaling law of the average received power at the user with respect to the number of reflecting elements, $N$, in an IRS-aided system with $N \to \infty$. For simplicity, we assume $M = 1$ with $\boldsymbol{G} \equiv \boldsymbol{g}$ to obtain essential insight. By ignoring the AP-user direct channel, the user's received power is given by $P_u = P|h^H|^2 = P|\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{g}|^2$. We consider three different phase shift solutions, i.e., 1) unit phase shift where $\boldsymbol{\Theta} = \boldsymbol{I}$; 2) random phase shift where $\theta_n$'s in $\boldsymbol{\Theta}$ are uniformly and randomly distributed in $[0, 2\pi)$; and 3) optimal phase shift which is obtained by the above two proposed algorithms (both optimal for $M = 1$).

*Proposition 2:* Assume $\boldsymbol{h}_r^H \sim \mathcal{CN}(\boldsymbol{0}, \varrho_h^2\boldsymbol{I})$ and $\boldsymbol{g} \sim \mathcal{CN}(\boldsymbol{0}, \varrho_g^2\boldsymbol{I})$. As $N \to \infty$, it holds that

$$P_u \to \begin{cases} NP\varrho_h^2\varrho_g^2, & \text{for } \boldsymbol{\Theta} = \boldsymbol{I} \text{ or random } \boldsymbol{\Theta}, \\ N^2\dfrac{P\pi^2\varrho_h^2\varrho_g^2}{16}, & \text{for optimal } \boldsymbol{\Theta}. \end{cases} \tag{29}$$

*Proof:* The three cases are discussed as follows:

- When $\boldsymbol{\Theta} = \boldsymbol{I}$, we have $h^H = \boldsymbol{h}_r^H\boldsymbol{g}$. By invoking the Lindeberg-Lévy central limit theorem [21], we have $\boldsymbol{h}_r^H\boldsymbol{g} \sim \mathcal{CN}(0, N\varrho_h^2\varrho_g^2)$ as $N \to \infty$. As the equivalent channel $h^H$ is a random variable, the average user received power is given by $P_u = P\mathbb{E}(|h^H|^2) \to NP\varrho_h^2\varrho_g^2$. For the case of random phase shifts with $\theta_n \in [0, 2\pi)$, we have $h^H = \boldsymbol{h}_r^H\bar{\boldsymbol{g}}$ where $\bar{\boldsymbol{g}} = \boldsymbol{\Theta}\boldsymbol{g}$. As $\boldsymbol{\Theta}$ is a unitary matrix, it follows that $\bar{\boldsymbol{g}}$ has the same distribution as $\boldsymbol{g}$, i.e., $\bar{\boldsymbol{g}} \sim \mathcal{CN}(\boldsymbol{0}, \varrho_g^2\boldsymbol{I})$. Thus we attain the same result as $\boldsymbol{\Theta} = \boldsymbol{I}$.
- For optimal $\boldsymbol{\Theta}$ where the solution is given by (28), we have $|h^H| = |\boldsymbol{h}_r^H\boldsymbol{\Theta}\boldsymbol{g}| = \sum_{n=1}^{N}|h_{r,n}||g_n|$, where $h_{r,n}$ and $g_n$ are the $n$-th elements in $\boldsymbol{h}_r^H$ and $\boldsymbol{g}$, respectively. The received power is then given by
$$P_u = P\Big|\sum_{n=1}^{N}|h_{r,n}||g_n|\Big|^2. \tag{30}$$
Since $|h_{r,n}|$ and $|g_n|$ are statistically independent and follow Rayleigh distribution with mean values $\sqrt{\pi}\varrho_h/2$ and $\sqrt{\pi}\varrho_g/2$, respectively, we have $\mathbb{E}(|h_{r,n}||g_n|) = \pi\varrho_h\varrho_g/4$. By using the fact that $\sum_{n=1}^{N}|h_{r,n}||g_n|/N \to \pi\varrho_h\varrho_g/4$ as $N \to \infty$, it follows that
$$P_u \to N^2\frac{P\pi^2\varrho_h^2\varrho_g^2}{16}. \tag{31}$$

This thus completes the proof. ■

The power scaling law with the optimal IRS phase design in Proposition 2 is highly promising since it implies that by using a large number of reflecting units at the IRS, we can scale down the transmit power of the AP by a factor of $1/N^2$ without compromising the user received SNR. The fundamental reason behind such a "squared gain" is that the IRS not only achieves the transmit beamforming gain of order $N$ in the IRS-user link as in the conventional massive MIMO [10], but also captures an inherent aperture gain of order $N$ by collecting more signal power in the AP-IRS link, which, however, cannot be achieved by scaling up the number of transmit antennas in massive MIMO due to the fixed total transmit power. Moreover, for the two benchmark cases with unit and random phase shifts at the IRS, a received power gain of order $N$ is also achieved. This shows the practical usefulness of the IRS, even without requiring any channel knowledge for optimally setting the phase shifts. Note that the received noise power in the IRS-aided system remains constant as $N$ increases and thus the corresponding user receive SNR also has the same squared gain as the received signal power with increasing $N$.

Next, we show the performance scaling law of an FD AF relay aided system under the same setup as the above IRS-aided system. The relay is equipped with $N$ transmit and $N$ receive antennas and the direct channel from the AP to the user can be similarly ignored when $N$ is asymptotically large. We assume that the relay adopts linear receive and transmit beamforming vectors, denoted by $\boldsymbol{x}_r^H$ and $\boldsymbol{x}_t$, respectively. In addition, perfect self-interference cancellation (SIC) is assumed at the relay so that the obtained performance serves as an upper bound for the practical case with imperfect SIC. In this case, the user receive SNR can be expressed as

$$\gamma_{FD} = \frac{\frac{PP_r\|\boldsymbol{x}_r^H\boldsymbol{g}\|^2\|\boldsymbol{h}_r^H\boldsymbol{x}_t\|^2}{\|\boldsymbol{x}_t\|^2\|\boldsymbol{x}_r\|^2}}{\frac{P_r\sigma_r^2\|\boldsymbol{x}_r\|^2\|\boldsymbol{h}_r^H\boldsymbol{x}_t\|^2+P\sigma^2\|\boldsymbol{x}_t\|^2\|\boldsymbol{x}_r^H\boldsymbol{g}\|^2}{\|\boldsymbol{x}_t\|^2\|\boldsymbol{x}_r\|^2} + \sigma_r^2\sigma^2}, \tag{32}$$

where $P_r$ and $\sigma_r^2$ denote the transmit power and the noise power at the relay, respectively. It is not difficult to show that the optimal solution maximizing $\gamma_{FD}$ satisfies $\boldsymbol{x}_t^* = \frac{\boldsymbol{h}_r}{\|\boldsymbol{h}_r\|}$ and $\boldsymbol{x}_r^* = \frac{\boldsymbol{g}}{\|\boldsymbol{g}\|}$. Substituting $\boldsymbol{x}_t^*$ and $\boldsymbol{x}_r^*$ into (32), we have

$$\gamma_{FD} = \frac{PP_r\|\boldsymbol{g}\|^2\|\boldsymbol{h}_r\|^2}{P_r\sigma_r^2\|\boldsymbol{h}_r\|^2 + P\sigma^2\|\boldsymbol{g}\|^2 + \sigma_r^2\sigma^2}. \tag{33}$$

Then, we have the following proposition.

*Proposition 3:* Assume $\boldsymbol{h}_r^H \sim \mathcal{CN}(\mathbf{0}, \varrho_h^2 \boldsymbol{I})$ and $\boldsymbol{g} \sim \mathcal{CN}(\mathbf{0}, \varrho_g^2 \boldsymbol{I})$. As $N \to \infty$, it holds that

$$\gamma_{FD} \to \frac{P P_r \varrho_g^2 \varrho_h^2 N}{P_r \sigma_r^2 \varrho_h^2 + P \sigma^2 \varrho_g^2}. \tag{34}$$

*Proof:* Since $\frac{\|\boldsymbol{g}\|^2}{N} \to \varrho_g^2$ and $\frac{\|\boldsymbol{h}_r\|^2}{N} \to \varrho_h^2$ as $N \to \infty$ [10], it follows that

$$\begin{aligned}\gamma_{FD} &\to \frac{P P_r \varrho_g^2 \varrho_h^2 N^2}{P_r \sigma_r^2 \varrho_h^2 N + P \sigma^2 \varrho_g^2 N + \sigma_r^2 \sigma^2} \\ &\approx \frac{P P_r \varrho_g^2 \varrho_h^2 N}{P_r \sigma_r^2 \varrho_h^2 + P \sigma^2 \varrho_g^2}. \end{aligned} \tag{35}$$

This thus completes the proof. ■

Proposition 3 shows that even with perfect SIC, the receive SNR by using the FD AF relay increases only linearly with $N$ when $N$ is asymptotically large. This is fundamentally due to the noise effect at the AF relay. To be specific, although the signal power in the FD AF relay system scales in the order of $N^2$ same as that in the IRS-aided system, its effective noise power at the receiver also scales linearly with $N$ (see (35)) in contrast to the constant noise power $\sigma^2$ in the IRS-aided system, thus resulting in a lower SNR gain order with $N$. Last, it is worth mentioning that for the HD AF relay system, its receive SNR scaling order with $N$ can be shown to be identical to that of the FD AF relay system given in Proposition 3.

## IV. Multiuser System

In this section, we consider the general multiuser setup. Specifically, we propose two efficient algorithms to solve (P1) suboptimally by generalizing the two approaches in the single-user case.

### A. Alternating Optimization Algorithm

This algorithm leverages the alternating optimization similarly as in the single-user case, while the transmit beamforming at the AP is designed by applying the well-known minimum mean squared error (MMSE) criterion to cope with the multiuser interference instead of using MRT in the single-user case without interference. For given phase shift $\boldsymbol{\theta}$, the combined channel from the AP to user $k$ is given by $\boldsymbol{h}_k^H = \boldsymbol{h}_{r,k}^H \boldsymbol{\Theta} \boldsymbol{G} + \boldsymbol{h}_{d,k}^H$. Thus, problem (P1) is reduced to

$$\text{(P3)}: \min_{\boldsymbol{W}} \sum_{k=1}^{K} \|\boldsymbol{w}_k\|^2 \tag{36}$$

$$\text{s.t.} \ \frac{|\boldsymbol{h}_k^H \boldsymbol{w}_k|^2}{\sum_{j\neq k}^{K} |\boldsymbol{h}_k^H \boldsymbol{w}_j|^2 + \sigma_k^2} \geq \gamma_k, \quad \forall k. \tag{37}$$

Note that (P3) is the conventional power minimization problem in the multiuser MISO downlink broadcast channel, which can be efficiently solved by using second-order cone program (SOCP) [22], SDP [23], or a fixed-point iteration algorithm based on the uplink-downlink duality [24], [25]. In addition, it is easy to verify that at the optimal solution to problem (P3), all the SINR constraints in (37) are met with equalities.

On the other hand, for given transmit beamforming $\boldsymbol{W}$, problem (P1) is reduced to a feasibility-check problem. Let $\boldsymbol{h}_{d,k}^H \boldsymbol{w}_j = b_{k,j}$ and $v_n = e^{j\theta_n}, n = 1, \cdots, N$. By applying the change of variables $\boldsymbol{h}_{r,k}^H \boldsymbol{\Theta} \boldsymbol{G} \boldsymbol{w}_j = \boldsymbol{v}^H \boldsymbol{a}_{k,j}$ where $\boldsymbol{v} = [e^{j\theta_1}, \cdots, e^{j\theta_N}]^H$ and $\boldsymbol{a}_{k,j} = \text{diag}(\boldsymbol{h}_{r,k}^H)\boldsymbol{G}\boldsymbol{w}_j$, problem (P1) is reduced to

$$\text{Find } \boldsymbol{v} \tag{38}$$

$$\text{s.t.} \ \frac{|\boldsymbol{v}^H \boldsymbol{a}_{k,k} + b_{k,k}|^2}{\sum_{j\neq k}^{K} |\boldsymbol{v}^H \boldsymbol{a}_{k,j} + b_{k,j}|^2 + \sigma_k^2} \geq \gamma_k, \quad \forall k, \tag{39}$$

$$|v_n| = 1, \quad n = 1, \cdots, N. \tag{40}$$

While the above problem appears similar to the relay beamforming optimization problem for multi-antenna relay broadcast channel [26], it cannot be directly transformed into an SOCP optimization problem because the phase rotation of the common vector $\boldsymbol{v}$ may not render $\boldsymbol{v}^H \boldsymbol{a}_{k,k} + b_{k,k}$'s in (39) to be real numbers for all users. Moreover, it has non-convex unit-modulus constraints in (40). However, by observing that constraints (39) and (40) can be transformed into quadratic constraints, we apply the SDR technique to approximately solve problem (38) efficiently.

Specifically, by introducing an auxiliary variable $t$, (38) can be equivalently written as

$$\text{Find } \boldsymbol{v} \tag{41}$$

$$\begin{aligned}\text{s.t.} \ &\bar{\boldsymbol{v}}^H \boldsymbol{R}_{k,k} \bar{\boldsymbol{v}} + |b_{k,k}|^2 \geq \gamma_k \sum_{j\neq k}^{K} \bar{\boldsymbol{v}}^H \boldsymbol{R}_{k,j} \bar{\boldsymbol{v}} \\ &\quad + \gamma_k (\sum_{j\neq k}^{K} |b_{k,j}|^2 + \sigma_k^2), \quad \forall k, \end{aligned}\tag{42}$$

$$|v_n|^2 = 1, \quad n = 1, \cdots, N+1, \tag{43}$$

where

$$\boldsymbol{R}_{k,j} = \begin{bmatrix} \boldsymbol{a}_{k,j} \boldsymbol{a}_{k,j}^H & \boldsymbol{a}_{k,j} b_{k,j}^H \\ \boldsymbol{a}_{k,j}^H b_{k,j} & 0 \end{bmatrix}, \quad \bar{\boldsymbol{v}} = \begin{bmatrix} \boldsymbol{v} \\ t \end{bmatrix}.$$

Note that $\bar{\boldsymbol{v}}^H \boldsymbol{R}_{k,j} \bar{\boldsymbol{v}} = \text{tr}(\boldsymbol{R}_{k,j} \bar{\boldsymbol{v}} \bar{\boldsymbol{v}}^H)$. Define $\boldsymbol{V} = \bar{\boldsymbol{v}} \bar{\boldsymbol{v}}^H$, which needs to satisfy $\boldsymbol{V} \succeq \mathbf{0}$ and $\text{rank}(\boldsymbol{V}) = 1$. Since the rank-one constraint is non-convex, we relax this constraint and problem (41) is then transformed to

$$\text{(P4)}: \ \text{Find } \boldsymbol{V} \tag{44}$$

$$\begin{aligned}\text{s.t.} \ &\text{tr}(\boldsymbol{R}_{k,k} \boldsymbol{V}) + |b_{k,k}|^2 \geq \gamma_k \sum_{j\neq k}^{K} \text{tr}(\boldsymbol{R}_{k,j} \boldsymbol{V}) \\ &\quad + \gamma_k (\sum_{j\neq k}^{K} |b_{k,j}|^2 + \sigma_k^2), \quad \forall k, \end{aligned}\tag{45}$$

$$\boldsymbol{V}_{n,n} = 1, n = 1, \cdots, N+1, \tag{46}$$

$$\boldsymbol{V} \succeq 0. \tag{47}$$

It is not difficult to observe that problem (P4) is an SDP and hence it can be optimally solved by existing convex optimization solvers such as CVX [20]. While the SDR may not be tight for problem (41), the Gaussian randomization can be similarly used to obtain a feasible solution to problem (41) based on the higher-rank solution obtained by solving (P4). In addition, it is worth pointing out that the SINR constraints in (45) are not necessarily to be met with equality for a feasible solution of (P4), due to the common phase shifting matrix ($\boldsymbol{V}$) for all users.

In the proposed alternating optimization algorithm, we solve problem (P1) by solving problems (P3) and (P4) alternately in an iterative manner, where the solution obtained in each iteration is used as the initial point of the next iteration. The details of the proposed algorithm are summarized in Algorithm 1. In particular, the algorithm starts with solving problem (P3) for given $\boldsymbol{\theta}$ instead of solving (P4) for given $\boldsymbol{W}$. This is deliberately designed since (P3) is always feasible for any arbitrary $\boldsymbol{\theta}$, provided that $\mathrm{rank}(\boldsymbol{G}^H\boldsymbol{\Theta}\boldsymbol{H}_r+\boldsymbol{H}_d)=K$, while this may not be true for (P4) with arbitrary $\boldsymbol{W}$. On the other hand, as solving (P4) only attains a feasible solution, it remains unknown whether the objective value of (P3) will monotonically decrease or not over iterations in Algorithm 1. Intuitively, if the feasible solution obtained by solving (P4) achieves a strictly larger user SINR than the corresponding SINR target $\gamma_k$ for user $k$, then the transmit power of user $k$ and hence the total transmit power in problem (P3) can be properly reduced without violating all the SINR constraints. More rigorously, the convergence of Algorithm 1 is ensured by the following proposition.

*Proposition 4:* The objective value of (P3) is non-increasing over the iterations by applying Algorithm 1.

*Proof:* Denote the objective value of (P3) based on a feasible solution $(\boldsymbol{\theta},\boldsymbol{W})$ as $f(\boldsymbol{\theta},\boldsymbol{W})$. As shown in step 4 of Algorithm 1, if there exists a feasible solution to problem (P4), i.e., $(\boldsymbol{\theta}^{r+1},\boldsymbol{W}^r)$ exists, it is also feasible to problem (P3). As such, $(\boldsymbol{\theta}^{r},\boldsymbol{W}^r)$ and $(\boldsymbol{\theta}^{r+1},\boldsymbol{W}^{r+1})$ in step 3 are the feasible solutions to (P3) in the $r$th and $(r+1)$th iterations, respectively. It then follows that $f(\boldsymbol{\theta}^{r+1},\boldsymbol{W}^{r+1})\overset{(a)}{\geq}f(\boldsymbol{\theta}^{r+1},\boldsymbol{W}^{r})\overset{(b)}{=}f(\boldsymbol{\theta}^{r},\boldsymbol{W}^{r})$, where $(a)$ holds since for given $\boldsymbol{\theta}^{r+1}$ in step 3 of Algorithm 1, $\boldsymbol{W}^{r+1}$ is the optimal solution to problem (P3); and $(b)$ holds because the objective function of (P3) is regardless of $\boldsymbol{\theta}$ and only depends on $\boldsymbol{W}$. ■

To achieve better converged solution, we further transform problem (P4) into an optimization problem with an explicit objective to obtain a generally more efficient phase shift solution to reduce the transmit power. The rationale is that for the transmit beamforming optimization problem, i.e., (P3), all the SINR constraints are active at the optimal solution. As such, optimizing the phase shift to enforce the user achievable SINR to be larger than the SINR target in (P4) directly leads to the transmit power reduction in (P3) (e.g., by simply scaling down the power of transmit beamforming). To this end, problem (P4) is transformed into the following problem

$$\text{(P4')}:\ \max_{\boldsymbol{V},\{\alpha_k\}}\ \sum_{k=1}^{K}\alpha_k \tag{48}$$

$$\begin{aligned}\text{s.t. } \mathrm{tr}(\boldsymbol{R}_{k,k}\boldsymbol{V})+|b_{k,k}|^2 &\geq \gamma_k\sum_{j\neq k}^{K}\mathrm{tr}(\boldsymbol{R}_{k,j}\boldsymbol{V})\\ &\quad+\gamma_k\Big(\sum_{j\neq k}^{K}|b_{k,j}|^2+\sigma_k^2\Big)+\alpha_k,\quad\forall k,\end{aligned} \tag{49}$$

$$\boldsymbol{V}_{n,n}=1,\quad n=1,\cdots,N+1, \tag{50}$$

$$\boldsymbol{V}\succeq 0,\alpha_k\geq 0,\quad\forall k, \tag{51}$$

**Algorithm 1** Alternating Optimization Algorithm

```
1: Initialize the phase shifts $\boldsymbol{\theta}=\boldsymbol{\theta}^1$ and set the iteration number $r=1$.
2: repeat
3:   Solve problem (P3) for given $\boldsymbol{\theta}^r$, and denote the optimal solution as $\boldsymbol{W}^r$.
4:   Solve problem (P4) or (P4') for given $\boldsymbol{W}^r$, and denote the solution after performing Gaussian randomization as $\boldsymbol{\theta}^{r+1}$.
5:   Update $r=r+1$.
6: until The fractional decrease of the objective value is below a threshold $\epsilon>0$ or problem (P4)/(P4') becomes infeasible.
```

where the slack variable $\alpha_k$ can be interpreted as the "SINR residual" of user $k$ in phase shift optimization. Note that (P4) and (P4') have the same set of feasible $\boldsymbol{V}$, while (P4') is more efficient than (P4) in terms of the converged solution, as will be verified in Section V-B by simulation.

### *B. Two-Stage Algorithm*

Inspired by the combined channel gain maximization problem (13) in the single-user case, we next propose a two-stage algorithm with lower complexity compared to the alternating optimization algorithm by decoupling the joint beamforming design problem (P1) into two beamforming subproblems, for optimizing the phase shifts and transmit beamforming, respectively. Specifically, the phase shifts at the IRS are optimized in the first stage by solving a weighted effective channel gain maximization problem. This aims to align with the phases of different user channels so as to maximize the beamforming gain of the IRS, especially for the users near to the IRS. In the second stage, we solve problem (P3) to obtain the optimal MMSE-based transmit beamforming with given phase shifts $\boldsymbol{\theta}$.

Let $\boldsymbol{v}=[e^{j\theta_1},\cdots,e^{j\theta_N}]^H\in\mathbb{C}^{N\times 1}$ and $\boldsymbol{\Phi}_k=\mathrm{diag}(\boldsymbol{h}_{r,k}^H)\boldsymbol{G}\in\mathbb{C}^{N\times M}$, $\forall k$. The weighted sum of the combined channel gain of all users is expressed as

$$\begin{aligned}\sum_{k=1}^{K}t_k\|\boldsymbol{h}_{r,k}^H\boldsymbol{\Theta}\boldsymbol{G}+\boldsymbol{h}_{d,k}^H\|^2&=\sum_{k=1}^{K}t_k\|\boldsymbol{v}^H\mathrm{diag}(\boldsymbol{h}_{r,k}^H)\boldsymbol{G}+\boldsymbol{h}_{d,k}^H\|^2\\&=\sum_{k=1}^{K}t_k\|\boldsymbol{v}^H\boldsymbol{\Phi}_k+\boldsymbol{h}_{d,k}^H\|^2,\end{aligned} \tag{52}$$

where we set the weights to be $t_k=\frac{1}{\gamma_k\sigma_k^2}, k=1,...,K$, motivated by constraint (11). Based on (52), the phase shifts can be obtained by solving the following problem

$$\text{(P5)}:\ \max_{\boldsymbol{v}}\ \sum_{k=1}^{K}t_k\|\boldsymbol{v}^H\boldsymbol{\Phi}_k+\boldsymbol{h}_{d,k}^H\|^2 \tag{53}$$

$$\text{s.t. } |v_n|=1,\quad n=1,\cdots,N. \tag{54}$$

Note that for $K=1$, (P5) is equivalent to problem (13) for the single-user case in Section III-A. However, in the multiuser case, due to the same set of phase shifts applied for all users with different channels, the combined channel power gains of different users cannot be maximized at the same time in general, which thus need to be balanced for optimally solving (P5). Nevertheless, since problem (P5) is a non-convex QCQP, it can be similarly reformulated as a homogeneous QCQP as in Section III-A and then solved by applying the SDR and Gaussian randomization techniques. The details are omitted here for brevity. With the phase shifts obtained from (P5), the MMSE-based transmit bemaforming is then obtained by solving (P3). Compared to the alternating optimization based algorithm proposed in Section IV-A, the two-stage algorithm has lower computational complexity as (P5) and (P3) only need to be respectively solved for one time, but may suffer from certain performance loss, which will be evaluated in the next section.

## V. Simulation Results

In this section, numerical examples are provided to validate the effectiveness of the proposed algorithms. We consider a three-dimensional (3D) coordinate system where a uniform linear array (ULA) at the AP and a uniform rectangular array (URA) at the IRS are located in $x$-axis and $x$-$z$ plane, respectively. The antenna spacing is half wavelength and the reference (center) antennas at the AP and IRS are respectively located at $(0,0,0)$ and $(0,d_0,0)$, where $d_0>0$ is the distance between them. For the IRS, we set $N=N_xN_z$ where $N_x$ and $N_z$ denote the numbers of reflecting elements along the $x$-axis and $z$-axis, respectively. For the purpose of exposition, we fix $N_x=5$ and increase $N_z$ linearly with $N$. The distance-dependent path loss model is given by

$$L(d)=C_0\left(\frac{d}{D_0}\right)^{-\alpha}, \tag{55}$$

where $C_0$ is the path loss at the reference distance $D_0=1$ meter (m), $d$ denotes the individual link distance, and $\alpha$ denotes the path loss exponent. To account for small-scale fading, we assume the Rician fading channel model for all channels involved. Thus, the AP-IRS channel $\boldsymbol{G}$ is given by

$$\boldsymbol{G}=\sqrt{\frac{\beta_{\mathrm{AI}}}{1+\beta_{\mathrm{AI}}}}\boldsymbol{G}^{\mathrm{LoS}}+\sqrt{\frac{1}{1+\beta_{\mathrm{AI}}}}\boldsymbol{G}^{\mathrm{NLoS}}, \tag{56}$$

where $\beta_{\mathrm{AI}}$ is the Rician factor, and $\boldsymbol{G}^{\mathrm{LoS}}$ and $\boldsymbol{G}^{\mathrm{NLoS}}$ represent the deterministic LoS (specular) and Rayleigh fading components, respectively. In particular, the above model is reduced to the LoS channel when $\beta_{\mathrm{AI}}\to\infty$ or Rayleigh fading channel when $\beta_{\mathrm{AI}}=0$. The elements in $\boldsymbol{G}$ are then multiplied by the square root of the distance-dependent path loss in (55) with the path loss exponent denoted by $\alpha_{\mathrm{AI}}$. The AP-user and IRS-user channels are also generated by following the similar procedure. The path loss exponents of the AP-user and IRS-user links are denoted by $\alpha_{\mathrm{Au}}$ and $\alpha_{\mathrm{Iu}}$, respectively, and the Rician factors of the two links are denoted by $\beta_{\mathrm{Au}}$ and $\beta_{\mathrm{Iu}}$, respectively. Due to the relatively large distance and random scattering of the AP-user channel, we set $\alpha_{\mathrm{Au}}=3.5$ and $\beta_{\mathrm{Au}}=0$ while their counterparts for AP-IRS and IRS-user channels will be specified later to study their effects on the system performance. Without loss of generality, we assume that all users have the same SINR target, i.e., $\gamma_k=\gamma,\forall k$. The number of random vectors used for the Gaussian randomization is set to be 1000 and the stopping threshold for the alternating optimization algorithms is set as $\epsilon=10^{-4}$. Other system parameters are set as follows: $C_0=-30$ dB, $\sigma_k^2=-80$ dBm, $\forall k$, and $d_0=51$ m (if not specified otherwise).

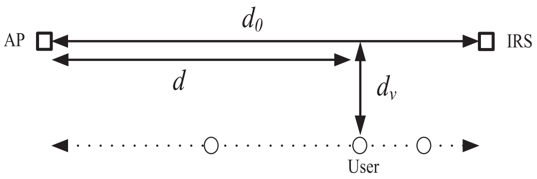


Fig. 2. Simulation setup of the single-user case (top view).

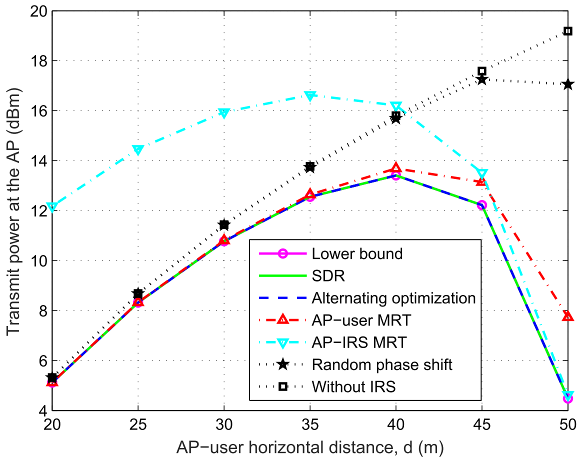


Fig. 3. AP transmit power versus AP-user horizontal distance.

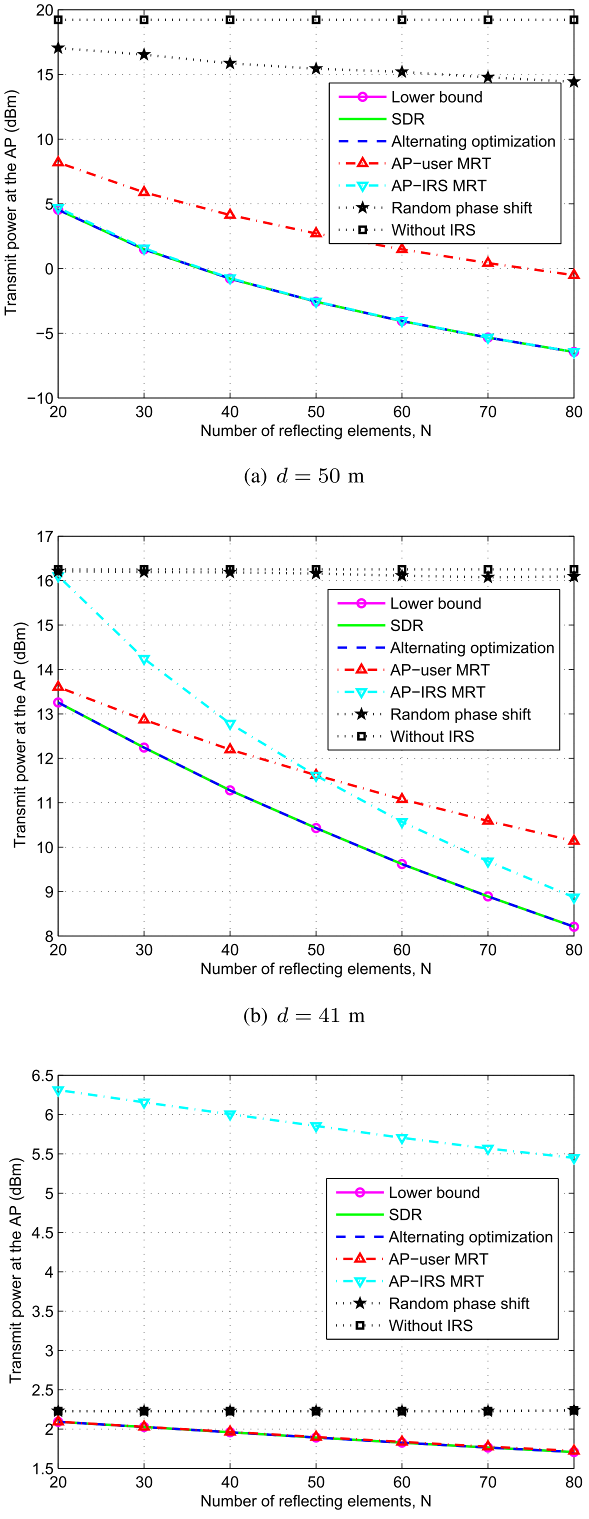
### A. Single-User System

First, we consider a single-user system with the user SNR target $\gamma=10$ dB and $M=4$. As shown in Fig. 2, the user lies on a horizontal line that is in parallel to the one that connects the reference antennas of AP and IRS, with the vertical distance between these two lines being $d_v=2$ m. Denote the horizontal distance between the AP and user by $d$ m. Accordingly, the AP-user and IRS-user link distances are given by $d_1=\sqrt{d^2+d_v^2}$ and $d_2=\sqrt{(d_0-d)^2+d_v^2}$, respectively. By varying the value of $d$, we can study the transmit power required for serving the user located between the AP and IRS, under the given SNR target. The path loss exponents and Rician factors are set as $\alpha_{\mathrm{AI}}=2$, $\alpha_{\mathrm{Iu}}=2.8$, $\beta_{\mathrm{Iu}}=0$, and $\beta_{\mathrm{AI}}=\infty$, respectively, where $\boldsymbol{G}$ is of rank one, i.e., an LoS channel between the AP and IRS. We compare the following schemes: 1) Lower bound: the minimum transmit power based on the optimal solution of the SDP problem (19); 2) SDR: the solution obtained by applying SDR and Gaussian randomization techniques in Section III-A; 3) Alternating optimization: the solution proposed in Section III-B; 4) AP-user MRT: we set $\bar{\boldsymbol{w}}=\boldsymbol{h}_d/\|\boldsymbol{h}_d\|$ to achieve MRT based on the AP-user direct channel; 5) AP-IRS MRT: we set $\bar{\boldsymbol{w}}=\boldsymbol{g}/\|\boldsymbol{g}\|$ to achieve MRT based on the AP-IRS rank-one channel, with $\boldsymbol{g}^H$ denoting any row in $\boldsymbol{G}$; 6) Random phase shift: we set the elements in $\boldsymbol{\theta}$ randomly in $[0,2\pi]$ and then perform MRT at the AP based on the combined channel; 7) Benchmark scheme without the IRS by setting $\boldsymbol{w} = \sqrt{\gamma\sigma^2}\boldsymbol{h}_d/\|\boldsymbol{h}_d\|^2$. Note that for scheme 3), the transmit beamforming is initialized by using the AP-user MRT, and for schemes 4) and 5) with given $\bar{\boldsymbol{w}}$, the transmit power and phase shifts are optimized by using the results in Section III-B.

*1) AP Transmit Power Versus AP-User Distance:* In Fig. 3, we compare the transmit power required by all schemes versus the horizontal distance between the AP and user, $d$. First, it is observed that the proposed two schemes both achieve near-optimal transmit power as compared to the transmit power lower bound, and also significantly outperform other benchmark schemes. Second, for the scheme without the IRS, one can observe that the user farther away from the AP requires higher transmit power at the AP due to the larger signal attenuation. However, this problem is alleviated by deploying an IRS, which implies that a larger AP-user distance does not necessarily lead to a higher transmit power in IRS-aided wireless networks. This is because the user farther away from the AP may be closer to the IRS and thus it is able to receive stronger reflected signal from the IRS. As a result, the user near either the AP (e.g., $d = 23$ m) or IRS (e.g., $d = 47$ m) requires lower transmit power than a user far away from both of them (e.g., $d = 40$ m). This phenomenon also suggests that the signal coverage can be effectively extended by deploying only a passive IRS rather than installing an additional AP or active relay. For example, for the same transmit power about 13 dBm, the coverage of the network without the IRS is about 33 m whereas this value is improved to be beyond 50 m by applying the proposed joint beamforming designs with an IRS.

On the other hand, it is observed from Fig. 3 that the AP-user MRT scheme performs close to optimal when the user is nearer to the AP, while it incurs considerably higher transmit power when the user is nearer to the IRS. This is expected since in the former case, the user received signal is dominated by the AP-user direct link whereas the IRS-user link is dominant in the latter case. Moreover, it can be observed that the AP-IRS MRT scheme behaves oppositely as the user moves away from the AP toward IRS. Finally, Fig. 3 also shows that if the transmit beamforming is not designed properly, the performance achieved by using the IRS may be even worse than that of the case without the IRS, e.g., with the AP-IRS MRT scheme for $d \leq 35$ m. This further demonstrates that the proposed joint beamforming designs can dynamically adjust the AP's beamforming to strike an optimal balance between the signal power transmitted directly to the user and that to the IRS, to achieve the maximum received power at the user.

*2) AP Transmit Power Versus Number of Reflecting Elements:* In Fig. 4, we compare the AP's transmit power of all the above schemes versus the number of reflecting elements at the IRS when $d = 50$, 41, and 15 m, respectively. From Fig. 4 (a), it is observed that for the case of $d = 50$ m (i.e., the user is very close to the IRS), the AP-IRS MRT scheme achieves near-optimal transmit power since the signal reflected by the IRS is much stronger than that directly from the AP at the user. Furthermore, it is interesting to note that the transmit power required by the proposed schemes scales down with the number of reflecting elements $N$ approximately in the order of $N^2$ in this case, which is in accordance with Proposition 2 (even when the AP-IRS channel is LoS rather than Rayleigh fading). For example, for the same user SNR, a



(a) $d = 50$ m

(b) $d = 41$ m

(c) $d = 15$ m

Fig. 4. AP transmit power versus the number of reflecting elements at the IRS, $N$.

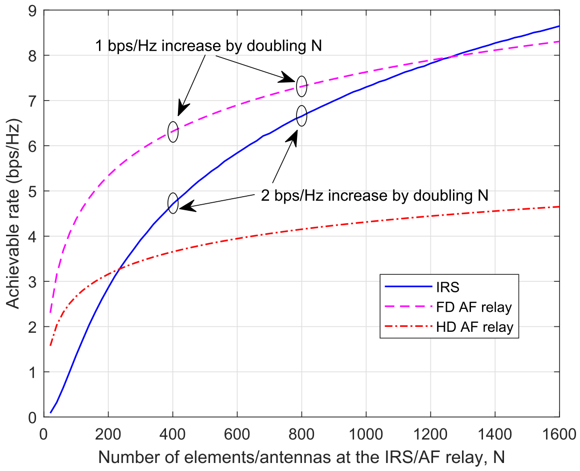


Fig. 5. Comparison between the IRS and the FD/HD AF relay.

transmit power of 2 dBm is required at the AP when $N = 30$ while this value is reduced to $-4$ dBm when $N = 60$, which suggests an around 6 dB gain by doubling the number of reflecting elements. In contrast, the transmit power required by using the random phase shift decreases with increasing $N$ in a much slower rate, because without reflect beamforming the average signal power of the reflected signal is comparable to that of the signal from the AP-user direct link in this case. Finally, it is observed that the above gains diminish as the user moves away from the IRS. For example, for the case of $d = 15$ m shown in Fig. 4 (c) where the AP-user direct link signal is much stronger than that of the IRS-user link, the required transmit power is insensitive to the number of reflecting elements. For the case of $d = 41$ m shown in Fig. 4 (b) when the user is neither close to the AP nor close to the IRS, it is observed that the transmit power gain of the proposed schemes is generally lower than $N^2$. This is because in this case the signal power received at the IRS is compromised as the AP transmit beamforming is steered to strike a balance between the AP-IRS link and the AP-user direct link. In practice, the number of reflecting elements can be properly selected depending on the IRS's location as well as the target user SNR/AP coverage range.

*3) Comparison With AF Relay:* Next, we compare the achievable rates of the IRS versus the FD AF relay based on the results derived in Section III. We consider the setup in Fig. 2 with $d_0 = d = 100$ m, $d_v = 1$ m, $\sigma_r^2 = -80$ dBm, $\alpha_{\rm AI} = 3.2$, $\alpha_{\rm Iu} = 2$, $\beta_{\rm AI} = 0$, $\beta_{\rm Iu} = \infty$, and $M = 1$. To focus on the comparison with large $N$, the direct link from the AP to the user is ignored, and perfect SIC is assumed for the FD AF relay. As such, the SNRs and the corresponding achievable rates for the IRS-aided and the FD AF relay-aided systems can be obtained based on (30) and (32), respectively. For a fair comparison, we assume that both systems have the same total transmit power budget $P = 5$ mW (for single link only). Since the IRS is passive, all the transmit power is used at the AP, whereas since the AF relay is active like the AP, an optimal power allocation between them is required which can be obtained by exhaustive search.

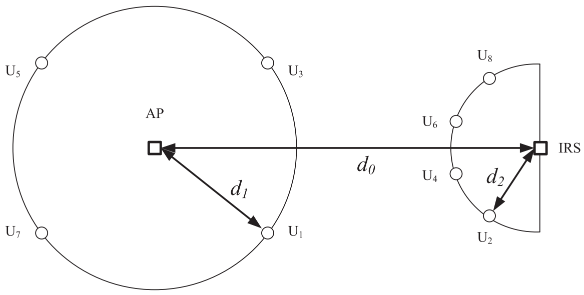


Fig. 6. Simulation setup of the multiuser case (top view).

Under the above setup, we plot the achievable rate in bits/second/Hertz (bps/Hz) versus $N$ in Fig. 5, where the HD AF relay is also considered as a benchmark. It is observed that when $N$ is small, the IRS-aided system is able to achieve the same rate as the FD/HD AF relay-aided system by using more reflecting elements. However, since the IRS's elements are passive, no transmit RF chains are needed for them and thus the cost is much lower as compared with that of active antennas for the AF relay requiring transmit RF chains. Furthermore, one can observe that by doubling $N$ from 400 to 800, the achievable rate of using IRS increases about 2 bps/Hz whereas that of using the FD AF relay only increases about 1 bps/Hz. This is due to their different SNR gains ($N^2$ versus $N$) with increasing $N$ as revealed in Propositions 2 and 3. As a result, it is expected that the IRS-aided system will eventually outperform the FD/HD AF relay-aided system when $N$ is sufficiently large, as shown in Fig. 5.

### *B. Multiuser System*

Next, we consider a multiuser system with eight users, denoted by $U_k$, $k = 1, \cdots, 8$, and their locations are shown in Fig. 6. Specifically, $U_k$'s, $k = 2, 4, 6, 8$, lie evenly on a half circle centered at the reference antenna of the IRS with radius $d_2 = 3$ m, which are usually considered as "cell-edge" users, as compared to $U_k$'s, $k = 1, 3, 5, 7$, which lie evenly on a circle centered at the reference antenna of the AP with radius $d_1 = 20$ m. Since the IRS can be practically deployed in LoS with the AP and "cell-edge" users, we set $\alpha_{\rm AI} = \alpha_{\rm Iu} = 2.8$, $\beta_{\rm AI} = \beta_{\rm Iu} = 3$ dB, respectively. We compare our proposed two algorithms (named as Alternating optimization w/ IRS and Two-stage algorithm w/ IRS, respectively) in Section IV with the two conventional designs in the case without the IRS, i.e., MMSE and zero-forcing (ZF) based beamforming [23], [27]. Specifically, the transmit power of the MMSE-based scheme without the IRS is obtained by solving (P3) with $\boldsymbol{\Theta} = \mathbf{0}$, while that of the ZF-based scheme without the IRS is given by $\mathrm{tr}(\boldsymbol{P}(\boldsymbol{H}_d^H \boldsymbol{H}_d)^{-1})$ where $\boldsymbol{P} = \mathrm{diag}(\sigma_1^2\gamma_1, \cdots, \sigma_K^2\gamma_K)$. The transmit power required by using the random phase shift at the IRS and MMSE beamforming at the AP is also plotted as a benchmark. Before comparing their performances, we first show the convergence behaviour of the proposed Algorithm 1 in Fig. 7 by setting $M = 4$ and considering that only $U_k$, $k = 1, 2, 3, 4$, are active (need to be served) with $\gamma = 20$ dB. The phase shifts are initialized using the two-stage algorithm. It is observed that the transmit

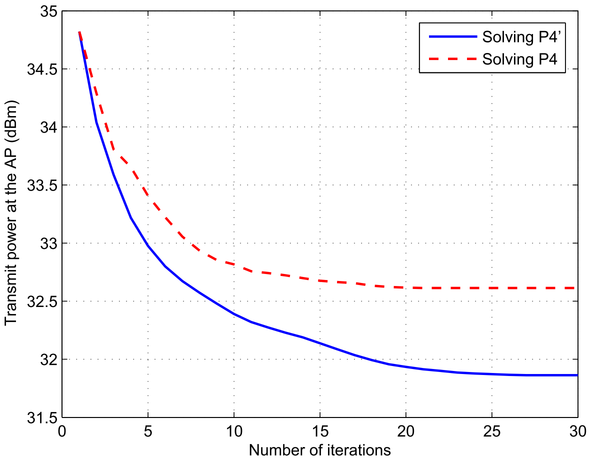


Fig. 7. Convergence behaviour of the proposed alternating optimization algorithm.

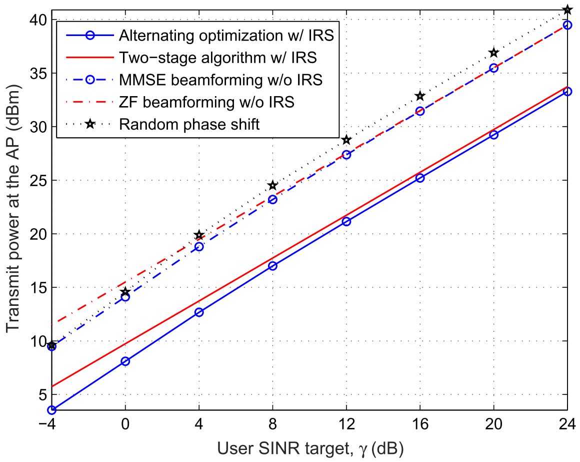


Fig. 8. AP's transmit power versus the SINR target for the two-user case.

power required by the proposed algorithm decreases quickly with the number of iterations and solving (P4') instead of (P4) in Algorithm 1 achieves lower converged power. Thus, in the following simulations, we use (P4') instead of (P4) in Algorithm 1.

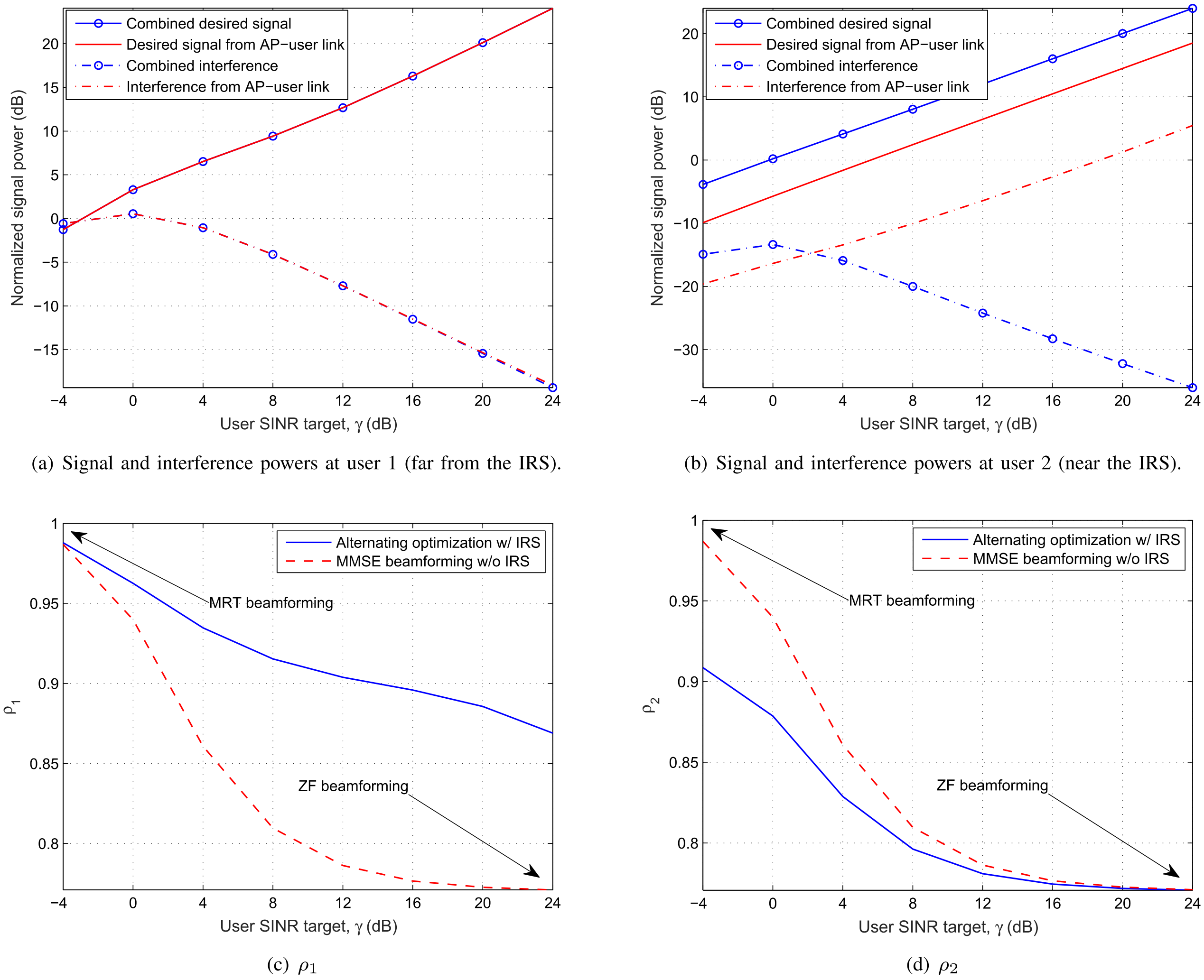
*1) AP Transmit Power Versus User SINR Target:* In Fig. 8, we show the transmit power at the AP versus the SINR target by considering that only two users, namely $U_k$'s, $k = 1, 2$, are active, which are far from and near the IRS, respectively. It is observed that by adding the IRS, the transmit power required by applying the proposed algorithms is significantly reduced, as compared to the case without the IRS. In addition, one can observe that the transmit power of the proposed two-stage algorithm asymptotically approaches that of the alternating optimization algorithm as $\gamma$ increases.

Next, we show how the AP and IRS serve the two users by collaboratively steering the transmit beamforming and adjusting the phase shifts based on the proposed Algorithm 1 (i.e., Alternating optimization w/ IRS). To visualize the transmit beamforming direction at the AP, we use the effective angle between the transmit beamforming (i.e., $\boldsymbol{w}_k$) and the AP-user direct channel (i.e., $\boldsymbol{h}^H_{d,k}$) for user $k$, which is defined as [18], [27]

$$\rho_k \triangleq \mathbb{E}\left(\frac{|\boldsymbol{h}^H_{d,k}\boldsymbol{w}_k|}{\|\boldsymbol{h}^H_{d,k}\|\|\boldsymbol{w}_k\|}\right), \quad \forall k. \tag{57}$$

In particular, $\rho_k = 1$ when the AP steers $\boldsymbol{w}_k$ to align with $\boldsymbol{h}^H_{d,k}$ perfectly, i.e., MRT transmission, whereas in general $\rho_k < 1$ when the AP sets $\boldsymbol{w}_k$ orthogonal to $\boldsymbol{h}^H_{d,j}$, $j \neq k$, i.e., ZF transmission. Generally, a higher $\rho_k$ implies that the transmit beamforming direction is closer to that of the AP-user direct channel. Figs. 9 (a) and (b) show both the desired signal power and interference power of the two users, respectively. For each user $k$, we plot the following four power terms: 1) Combined desired signal power, i.e., $|(\boldsymbol{h}^H_{r,k}\boldsymbol{\Theta}\boldsymbol{G} + \boldsymbol{h}^H_{d,k})\boldsymbol{w}_k|^2$; 2) Desired signal power from AP-user link, i.e., $|\boldsymbol{h}^H_{d,k}\boldsymbol{w}_k|^2$; 3) Combined interference power, i.e., $|(\boldsymbol{h}^H_{r,k}\boldsymbol{\Theta}\boldsymbol{G} + \boldsymbol{h}^H_{d,k})\boldsymbol{w}_j|^2$, $j \neq k$; 4) Interference power from AP-user link, i.e., $|\boldsymbol{h}^H_{d,k}\boldsymbol{w}_j|^2$, $j \neq k$. For the purpose of exposition, both the desired signal and interference powers are normalized by the noise power. It is observed from Fig. 9 (a) that for user 1 (far from IRS), the desired signal and interference powers received from the combined channel are almost the same as those from the AP-user direct link. This is expected since user 1 is far away from the IRS and hence the signal (both desired and interference) from the IRS is negligible. However, the case of user 2 (near IRS) is quite different from that of user 1. As shown in Fig. 9 (b), the desired signal power from the combined channel is remarkably higher than that from the AP-user direct link, due to the reflect beamforming gain by the IRS. Furthermore, the IRS also helps align the interference from the AP-IRS-user reflect link oppositely with that from the AP-user direct link to suppress the interference at user 2. This is shown by observing that the interference power from the AP-user direct link monotonically increases with the SINR target whereas that from the combined channel decreases as SINR target increases.

In Figs. 9 (c) and (d), we plot $\rho_1$ and $\rho_2$, respectively, for both the cases with and without the IRS. For user 1, it is observed in Fig. 9 (c) that for low SINR (noise-limited) regime, the AP in the case with the IRS steers its beam direction toward the AP-user direct channel as in the case without the IRS (i.e., MRT beamforming), while for high SINR (interference-limited) regime where the beam direction in the latter case is adjusted to null out the interference to user 2 (i.e, ZF beamforming), the former case still keeps a high correlation between user 1's beam direction and the corresponding AP-user direct channel. This is expected since the IRS is not able to enhance the desired signal for user 1 (see Fig. 9 (a)), thus keeping a high $\rho_1$ helps reduce the transmit power for serving user 1. However, user 2 inevitably suffers more interference from user 1 in the AP-user direct link (see Fig. 9 (b)), which, nevertheless, is significantly suppressed at user 2, thanks to the IRS-assisted interference cancellation. In contrast, from Fig. 9 (d) it is observed that the user 2's beam direction in the case with the IRS is steered toward the AP-user combined channel rather than the AP-user direct channel at low SINR regime, and then converges to the same ZF beamforming as in the case without the IRS at high SINR regime. This is expected since the IRS is not



(a) Signal and interference powers at user 1 (far from the IRS).

(b) Signal and interference powers at user 2 (near the IRS).

(c) $\rho_1$

(d) $\rho_2$

Fig. 9. Illustration for the collaborative working mechanism of the AP and IRS.

able to help cancel the AP-user direct interference at user 1, thus nulling out the interference caused by user 2 is the most effective way for meeting the high SINR target. From the above, it is concluded that the transmit beamforming directions for the users with the IRS are drastically different from those in the case without the IRS, depending on their different distances with the IRS. The above results further demonstrate the necessity of jointly optimizing the transmit beamforming and phase shifts in IRS-aided multiuser systems.

*2) AP Transmit Power Versus Rician Factor of* $\boldsymbol{G}$*:* In Fig. 10, we plot the AP's transmit power required by the proposed algorithms with IRS against the benchmark schemes without IRS versus the Rician factor of the AP-IRS channel by assuming all eight users are active and setting $M = 8$, $N = 40$, and $\gamma = 10$ dB. One can observe that when the average power of the LoS component is comparable to that of the fading component (e.g., $\beta_{\rm AI} \geq -10$ dB), the transmit power of both proposed algorithms increases with the increasing Rician factor, for meeting the same set of SINR targets. This is because for users near the IRS, the reflected signals generally dominate the signals from AP-user direct links. As such, a higher Rician factor of $\boldsymbol{G}$ results in higher correlation among the combined channels of these users, which is detrimental to the spatial multiplexing gain due to the more severe multiuser interference. The implication of this result is that (somehow surprisingly) it is practically favorable to

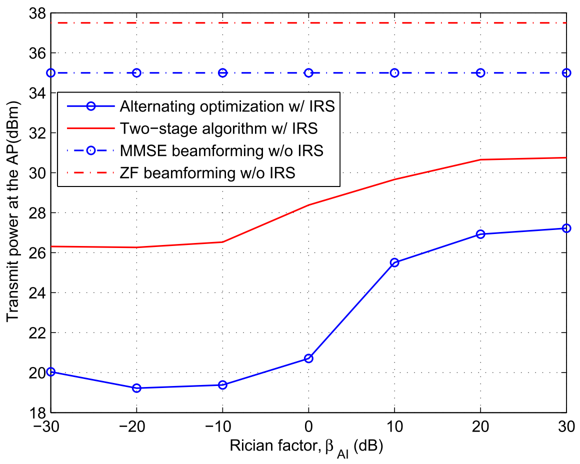


Fig. 10. AP's transmit power versus the Rician factor of the AP-IRS channel.

deploy the IRS in a relatively rich scattering environment to avoid strong LoS (low-rank $\boldsymbol{G}$) with the AP, so as to serve multiple users by the IRS which requires sufficient spatial degrees of freedom from the AP to the IRS.

*3) Comparison With Massive MIMO:* To compare with the existing TDD-based massive MIMO system (without the use of IRS), we consider there are sixteen users with eight users randomly distributed within 60 m from the AP and the other eight users randomly distributed within 6 m from the IRS. Other channel parameters are set to be the same as those for Fig. 8. To facilitate the channel estimation, we assume that the IRS is equipped with receive RF chains. The transmission protocol for the IRS-aided wireless system is described as follows: 1) all the users send orthogonal pilot signals concurrently as in the TDD-based massive MIMO system; 2) the AP and the IRS estimate the AP-user and IRS-user channels, respectively[4]; 3) the AP starts to transmit data to the users and in the meanwhile sends its estimated AP-user channels to the IRS controller via a separate control link (see Fig. 1), so that the IRS can jointly optimize the AP transmit beamforming vectors and its phase shifts by using the proposed algorithms in this paper; 4) the IRS controller sends optimized transmit beamforming vectors to the AP and sets its phase shifts accordingly; and 5) the AP and IRS start to transmit data to the users collaboratively. As such, different from the massive MIMO system, the IRS-aided system generally incurs additional delay in steps 3) and 4) due to information exchange and algorithm computation.

We denote the channel coherence time as $T_c$ and the total delay caused by 3) and 4) as $\tau$, where we assume that $\tau < T_c$. The delay ratio is thus given by $\rho = \frac{\tau}{T_c}$. Since steps 1) and 2) are required in both the IRS-aided system and massive MIMO system, the time overhead for channel training is omitted for both schemes for a fair comparison. Note that users are served by the AP only over $\tau$ with the achievable SINR of user $k$, $k \in \mathcal{K}$, denoted by $\mathrm{SINR}_k^1$, while when they are served by both the AP and IRS during the remaining time of $T_c - \tau$, the SINR of user $k$ is denoted by $\mathrm{SINR}_k^2$. Accordingly, the achievable rates of user $k$ in the above two phases can be expressed as $R_k^1 = \log_2(1+\mathrm{SINR}_k^1)$ and $R_k^2 = \log_2(1+\mathrm{SINR}_k^2)$ in bps/Hz, respectively. The average achievable rate of user $k$ over $T_c$ is thus given by $r_k = \frac{\tau}{T_c}R_k^1 + (1-\frac{\tau}{T_c})R_k^2 = \rho R_k^1 + (1-\rho)R_k^2$. Given the transmit power constraint at the AP, we aim to maximize the minimum achievable rate of $r_k$ among all the users, which is a dual problem of (P1) and thus can be solved by using the proposed alternating optimization algorithm together with an efficient bisection search over the AP transmit power [22]. For the case of massive MIMO, we adopt the optimal MMSE-based transmit beamforming at the AP.

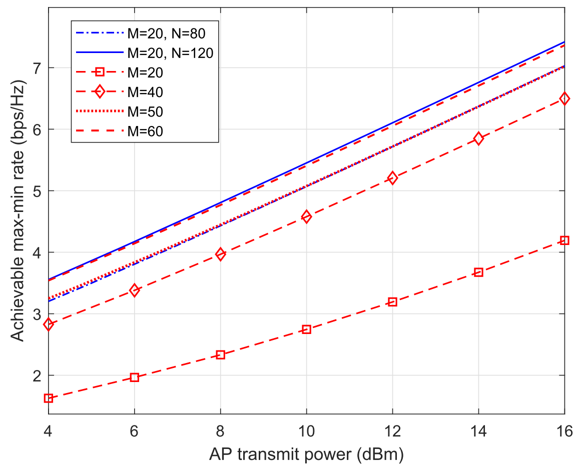


Fig. 11. Achievable max-min rate versus transmit power at AP.

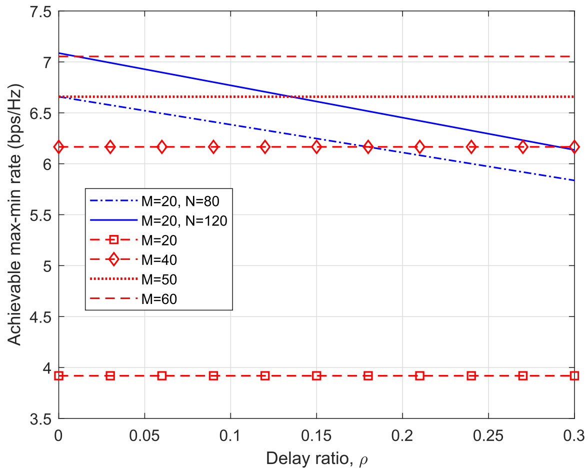


Fig. 12. Impact of IRS delay on achievable max-min rate.

In Fig. 11, we show the achievable max-min rate versus the transmit power at the AP for the ideal case with negligible delay (i.e., $\rho = 0$) in the IRS-aided system, which provides a throughput upper bound for practical implementation. From Fig. 11, it is observed that a hybrid deployment of an AP with $M = 20$ active antennas and an IRS with $N = 80$ passive reflecting elements achieves nearly the same performance as deploying only an AP with $M = 50$ but without using the IRS, which implies that the IRS is indeed effective in reducing the number of active antennas required in conventional massive MIMO. To take into account the practical delay due to IRS-AP coordination, we show in Fig. 12 the achievable max-min rate versus the delay ratio $\rho$ by fixing the transmit power at the AP as 15 dBm. It is observed that as the delay ratio increases, the achievable rate of the IRS-aided system decreases since users are served by a relatively small MIMO system (with $M = 20$) for more time before the IRS is activated for signal reflection. However, it is also observed that even with a large delay of $0.18T_c$, the IRS-aided system with $M = 20$ and $N = 80$ can still achieve better performance than the AP-only system with $M = 40$. This validates the practical throughput gain of IRS even by taking into account a moderate delay for its coordination with the AP.

[4] Since in practice the AP and IRS are deployed at fixed locations, we assume for simplicity that the channel $\boldsymbol{G}$ between them is quasi-static and changes much slower as compared to the AP-user and IRS-user channels, and thus is constant and known for the considered communication period of interest.

## VI. CONCLUSION

In this paper, we proposed a novel approach to enhance the spectrum and energy efficiency as well as reducing the implementation cost of future wireless communication systems by leveraging the passive IRS via smartly adjusting its signal reflection. Specifically, given the user SINR constraints, the active transmit beamforming at the AP and passive reflect beamforming at the IRS were jointly optimized to minimize the transmit power in an IRS-aided multiuser system. By applying the SDR and alternating optimization techniques, efficient algorithms were proposed to trade off between the system performance and computational complexity. It was shown for the single-user system that the receive SNR increases quadratically with the number of reflecting elements of the IRS, which is more cost efficient than the conventional massive MIMO or multi-antenna AF relay. While for the multiuser system, it was shown that IRS-enabled interference suppression can be jointly designed with the AP transmit beamforming to improve the performance of all users in the system, even for those that are far away from the IRS. Extensive simulation results under various practical setups demonstrated that by deploying the IRS and jointly optimizing its reflection with the AP transmission, the wireless network performance can be significantly improved in terms of energy consumption, coverage as well as achievable rate, as compared to the conventional systems without using the IRS. Useful insights on optimally deploying the IRS and its delay-performance trade-off were also drawn to provide useful guidance for practical design and implementation.

In practice, if IRS is equipped with receive RF chains, the commonly used pilot-assisted channel estimation methods can be similarly applied to the IRS as shown in Section V-B; otherwise, it is infeasible for the IRS to directly estimate the channels with its associated AP/users. For the latter (more challenging) case, a viable approach may be to design the IRS's passive beamforming based on the feedback from the AP/users that receive the signals reflected by the IRS, which is worth investigating in the future work. In addition, after the initial version of this paper was submitted on October 12, 2018, we became aware of another parallel work [28], which shows that the IRS-aided wireless system is more energy-efficient than the conventional multi-antenna AF relay system with HD operation. Although the spectrum efficiency can be further improved by using the FD AF relay, effective SIC is required, which incurs additional energy consumption. As such, it is worthy of further comparing the energy efficiency of IRS with the FD AF relaying in future work.

## REFERENCES

[1] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE GLOBECOM*, Dec. 2018, pp. 1–6.

[2] F. Boccardi, R. W. Heath, Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] S. Zhang, Q. Wu, S. Xu, and G. Y. Li, "Fundamental green tradeoffs: Progresses, challenges, and impacts on 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 33–56, 1st Quat., 2017.

[4] Q. Wu, G. Y. Li, W. Chen, D. W. K. Ng, and R. Schober, "An overview of sustainable green 5G networks," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 72–80, Aug. 2017.

[5] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," 2019, *arXiv:1905.00152*. [Online]. Available: https://arxiv.org/abs/1905.00152

[6] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng, "Coding metamaterials, digital metamaterials and programmable metamaterials," *Light, Sci. Appl.*, vol. 3, no. 10, p. e218, Oct. 2014.

[7] S. Hu, F. Rusek, and O. Edfors, "Beyond massive MIMO: The potential of data transmission with large intelligent surfaces," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2746–2758, May 2017.

[8] L. Subrt and P. Pechac, "Intelligent walls as autonomous parts of smart indoor environments," *IET Commun.*, vol. 6, no. 8, pp. 1004–1010, May 2012.

[9] X. Tan, Z. Sun, J. M. Jornet, and D. Pados, "Increasing indoor spectrum sharing capacity using smart reflect-array," in *Proc. IEEE ICC*, May 2016, pp. 1–6.

[10] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[11] S. K. Mohammed and E. G. Larsson, "Single-user beamforming in large-scale MISO systems with per-antenna constant-envelope constraints: The doughnut channel," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3992–4005, Nov. 2012.

[12] S. Zhang, R. Zhang, and T. J. Lim, "Constant envelope precoding for MIMO systems," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 149–162, Jan. 2018.

[13] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[14] F. Sohrabi and 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.

[15] C. Huang, A. Zappone, M. Debbah, and C. Yuen, "Achievable rate maximization by passive intelligent mirrors," in *Proc. IEEE ICASSP*, Apr. 2018, pp. 3714–3718.

[16] C. Boyer and S. Roy, "Backscatter communication and RFID: Coding, energy, and MIMO analysis," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 770–785, Mar. 2014.

[17] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," 2019, *arXiv:1906.03165*. [Online]. Available: https://arxiv.org/abs/1906.03165

[18] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[19] A. M.-C. So, J. Zhang, and Y. Ye, "On approximating complex quadratic optimization problems via semidefinite programming relaxations," *Math. Program.*, vol. 110, no. 1, pp. 93–110, 2007.

[20] M. Grant and S. Boyd. (2016). *CVX: MATLAB Software for Disciplined Convex Programming*. [Online]. Available: http://cvxr.com/cvx

[21] H. Cramér, *Random Variables and Probability Distributions*, vol. 36. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[22] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[23] M. Bengtsson and B. Ottersten, "Optimal and suboptimal transmit beamforming," in *Handbook of Antennas in Wireless Communications*. Boco Raton, FL, USA: CRC Press, 2001.

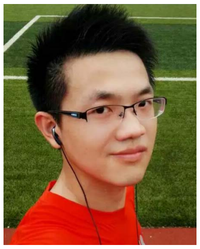
[24] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, Jan. 2004.

[25] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1426–1438, Aug. 2006.

[26] R. Zhang, C. C. Chai, and Y. C. Liang, "Joint beamforming and power control for multiantenna relay broadcast channel with QoS constraints," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 726–737, Feb. 2009.

[27] T. Yoo and A. Goldsmith, "On the optimality of multi-antenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

[28] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," 2018, *arXiv:1810.06934*. [Online]. Available: https://arxiv.org/abs/1810.06934



**Qingqing Wu** (S'13–M'16) received the B.Eng. degree in electronic engineering from the South China University of Technology in 2012 and the Ph.D. degree in electronic engineering from Shanghai Jiao Tong University (SJTU) in 2016. He is currently a Research Fellow with the Department of Electrical and Computer Engineering, National University of Singapore. His current research interests include intelligent reflecting surface (IRS), unmanned aerial vehicle (UAV) communications, and green communications. He was a recipient of the Outstanding Ph.D. Thesis Funding in SJTU in 2016, the Outstanding Ph.D. Thesis Award of the China Institute of Communications in 2017, and received the IEEE WCSP Best Paper Award in 2015. He was honored as the Exemplary Reviewer of the IEEE COMMUNICATIONS LETTERS in 2016 and 2017, the IEEE WIRELESS COMMUNICATIONS LETTERS in 2018, the IEEE TRANSACTIONS ON COMMUNICATIONS in 2017 and 2018, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS in 2017 and 2018. He serves as an Associate Editor for the IEEE COMMUNICATIONS LETTERS and the IEEE Open Journal of Communications Society and the Workshop Co-Chair for the IEEE ICC 2019 Workshop on Integrating UAVs into 5G.

**Rui Zhang** (S'00–M'07–SM'15–F'17) received the B.Eng. (Hons.) and M.Eng. degrees from the National University of Singapore, Singapore, and the Ph.D. degree from Stanford University, Stanford, CA, USA, all in electrical engineering.

From 2007 to 2010, he was a Research Scientist with the Institute for Infocomm Research, ASTAR, Singapore. Since 2010, he has been with the Department of Electrical and Computer Engineering, National University of Singapore, where he is currently a Dean's Chair Associate Professor with the Faculty of Engineering. He has authored over 300 papers. He has been listed as a Highly Cited Researcher (also known as the World's Most Influential Scientific Minds) by Thomson Reuters (Clarivate Analytics) since 2015. His research interests include UAV/satellite communication, wireless information and power transfer, multiuser MIMO, smart and reconfigurable environment, and optimization methods.

Dr. Zhang was an elected member of the IEEE Signal Processing Society SPCOM Technical Committee from 2012 to 2017 and the SAM Technical Committee from 2013 to 2015. He serves as a member of the Steering Committee of the IEEE WIRELESS COMMUNICATIONS LETTERS. He was a recipient of the 6th IEEE Communications Society Asia–Pacific Region Best Young Researcher Award in 2011 and the Young Researcher Award of the National University of Singapore in 2015. He was a co-recipient of the IEEE Marconi Prize Paper Award in Wireless Communications in 2015, the IEEE Communications Society Asia–Pacific Region Best Paper Award in 2016, the IEEE Signal Processing Society Best Paper Award in 2016, the IEEE Communications Society Heinrich Hertz Prize Paper Award in 2017, the IEEE Signal Processing Society Donald G. Fink Overview Paper Award in 2017, and the IEEE Technical Committee on Green Communications and Computing (TCGCC) Best Journal Paper Award in 2017. His coauthored paper received the IEEE Signal Processing Society Young Author Best Paper Award in 2017. He served for over 30 international conferences as the TPC co-chair or an organizing committee member and as the Guest Editor for three special issues in the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He served as the Vice Chair of the IEEE Communications Society Asia–Pacific Board Technical Affairs Committee from 2014 to 2015. He served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2012 to 2016, the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS: Green Communications and Networking Series from 2015 to 2016, and the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2013 to 2017. He is currently an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING. He is a Distinguished Lecturer of the IEEE Signal Processing Society and the IEEE Communications Society.