Delay-Constrained Joint Power Control, User Detection and Passive Beamforming in Intelligent Reflecting Surface Assisted Uplink mmWave System

Yashuai Cao, Student Member, IEEE, Tiejun Lv, Senior Member, IEEE, Zhipeng Lin, Wei Ni, Senior Member, IEEE, and Norman C. Beaulieu, Fellow, IEEE

Abstract

Millimeter-wave (mmWave) communications provide access to spectra with bandwidths and in abundance. However, the high susceptibility of mmWave to blockage imposes crucial challenges, especially to low-latency services. In this paper, a novel intelligent reflecting surface (IRS) assisted mmWave scheme is proposed to overcome the impact of blockage. New methods are developed to minimize the user power for a multi-user mmWave system by jointly optimizing individual device power, multi-user detection matrix and passive beamforming, subject to delay requirements. An alternating optimization framework is delivered so that the joint optimization problem can be decomposed into three subproblems iteratively optimized till convergence. In particular, closed-form expressions are devised for the update of the powers and multi-user detection vectors. The configuration of the IRS is formulated as a sum-of-inverse minimization (SIMin) fractional programming problem and solved by developing a new solution based on the alternating direction method of multipliers (ADMM). The configuration is also interpreted as a latency residual maximization problem, and solved efficiently by designing a new complex circle manifold optimization (CCMO) method. Numerical results corroborate the feasibility and effectiveness of our algorithms in terms of power saving, as compared with an existing semidefinite relaxation technique.

Y. Cao, T. Lv, Z. Lin and N. C. Beaulieu are with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications (BUPT), Beijing 100876, China (e-mail: {yashcao, lvtiejun, linlzp, nborm}@bupt.edu.cn).
W. Ni is with Data61, Commonwealth Scientific and Industrial Research, Sydney, NSW 2122, Australia (e-mail: wei.ni@data61.csiro.au).
I. INTRODUCTION

Millimeter-wave (mmWave) communications are perceived as a pivotal technology to provide gigabit data-rate in fifth generation of mobile network (5G) [1]–[5]. In 5G beyond, using spectrum at mmWave frequencies is expected to enable the diverse service-oriented applications in Big Data and Internet of Things (IoT), allowing wireless networks to support unprecedented low latency services in an efficient manner. Apart from quality of service (QoS) requirements in terms of connectivity, reliability, capacity and latency determined by vertical-specific services, 5G use cases also expect solutions for long-range coverage and low power consumption communication devices.

Evolving physical-layer communications technologies, such as massive multi-input multi-output (MIMO) and ultra-dense deployment, an integration mmWave system is promising to achieve significant network performance improvements. MmWave communications provide an appealing solution for meeting the stringent latency and power consumption requirements of tasks processed in IoT mobile devices. As one example of uplink mmWave system, a joint Europe/Japan H2020 Project is promoting the use of mmWave communications for mobile edge computing (MEC) [6]–[10]. The reason is that mmWave can offer high-capacity links, thereby leading to low-latency task offloading in the uplink. While mmWave can provide unrivaled data rates, the channel intermittency can adversely increase more uplink transmission latency and additional power consumption since mmWave links are highly susceptible to blockage. Barbarossa et al. [6] proposed two countermeasures to address the issue of blockage, by adopting the plethora of wireless access points and overbooking of communication resources. Pietro et al. [7] extended the two countermeasures to optimize uplink transmit power for mobile users under required latency constraints. However, these solutions require the a-priori knowledge of blocking probabilities for real-time prediction. Moreover, the over-provisioning of network resources and radio equipments has also presented the need for unpalatable costs, penalizing the green and sustainable design of 5G and beyond [11]. Intelligent reflecting surface (IRS) has been recently proposed as a promising solution for mitigating the impact of intermittent mmWave links on uplink transmission. Different from active massive MIMO or large intelligent surface [12], IRS, also termed as reconfigurable intelligent surface (RIS), is composed of nearly passive low-cost reflecting elements. IRS origins from reflectarrays and software-defined metamaterials [13]. It can adaptively change the signal propagation by adjusting phase shifters according to
dynamic wireless environments. To be specific, each IRS element is controlled by a smart controller and reflects incident signals towards desired direction [14], thus strengthening the signals at the desired receiver and suppressing the interference. These configurability and real-time control of phase shifters are attributed to the development and breakthrough of radio frequency micro-electromechanical systems (RF-MEMS) and metamaterial field [15]. In contrast, conventional reflectarrays widely applied in radar and satellite communications, and they are typically fabricated to have persistent phase shifts. Some important terminologies frequently used in the literature are active beamforming and passive beamforming: passive beamforming refers to reflect beamforming by the phase shifters of IRS; and active beamforming refers to precoding at the base station (BS).

The design of IRS-aided wireless transmission has been increasingly studied recently [15]–[20]. From the perspective of energy efficiency (EE), the authors of [16] jointly optimized the transmit power allocation of the BS and the passive beamforming of the IRS in an IRS aided downlink multi-user multi-input single-output (MISO) system, where an optimal trade-off between EE and number of reflecting elements was observed. Further, the BS transmit power control schemes were proposed with some recommendations on the deployment of IRS in [15]. The BS transmit power was also optimized in multi-IRS assisted simultaneous wireless information and power transfer (SWIPT) systems [17]. Later, IRS was applied to create virtual line-of-sight (LoS) paths to combat the blockages of downlink mmWave links [18], where the signal-to-interference-plus-noise ratio (SINR) was maximized to improve robustness of mmWave transmission. In [21], [22], IRS was applied to enhance physical-layer security. To achieve efficient focus signal transmissions, an online phase-shift configuration of IRS based on deep learning was introduced by [23]. The impact of phase shifts design of IRS on the network transmission performance were analyzed in [19], [20]. The above works have been focused on the BS power control and transmission design in downlink transmission systems. Very recently, Bai et al. [24] applied IRS in an MEC scenario for uplink transmissions. The radio and computation resource allocation were jointly optimized to minimize the offloading latency, under the assumption of equal uplink transmit power among users. Till now, the potential of IRS to improve uplink power allocation is yet to be well studied.

In this paper, we jointly optimize the uplink power control and passive beamforming in an IRS assisted multi-user mmWave communications system. Different from [6], [7], [10], supplementary links provided by IRS are introduced to address the blockage of mmWave channels
and meet the predefined upload latency requirements. We propose to minimize the total uplink user power in a single-input multiple-output (SIMO) system, where IRS configuration are utilized to assist intermittent mmWave channels for high-speed uplink scenarios. Different from existing studies which are based on an assumption of equal power allocation in the uplink [24], the problem to be solved in this paper, is a joint optimization of total uplink power, multi-user detectors at the access point (AP) and IRS phase shifts, where the uplink cochannel interference [25] is considered. Our key contributions are listed as follows:

- A new scenario is proposed, where IRS acts as a passive reflector to provide configurable reflecting paths between an AP and multiple users to overcome blockage in mmWave systems. With the aid of IRSs, we build up an uplink SIMO model where the user power is minimized under stringent uplink latency constraints and cochannel interference is suppressed by reconfigurations of IRS.

- An alternating optimization framework is put forth to decouple the originally intractable problem of interest, resulting in three separate tangible subproblems of uplink power control, multi-user detection and IRS reconfiguration. Closed-form expressions are derived for the user power and multi-user detection matrix. The overall convergence of our framework is analyzed.

- We also develop two efficient algorithms to reconfigure the passive beamforming of IRS. Specifically, we propose a fraction transform based alternating direction method of multipliers (ADMM) algorithm to optimize phase shifts in parallel. Moreover, by exploiting the geometric interpretation of constant modulus constraints pertaining to the passive beamforming, we also transform the configuration of IRS into a latency residual maximization problem. A new Complex Circle Manifold Optimization (CCMO) method is developed, where the original problem with unit modulus constraints is reformulated as an unconstrained optimization on a Riemannian manifold and solved efficiently using Riemannian gradient descent.

The remainder of this paper is organized as follows. Section II outlines the system model and formulates the problem of interest. Section III proposes the alternating optimization framework to solve the problem, and presents the closed-form solutions for uplink power control and multi-user detectors. Section IV develops the two efficient algorithms to configure the phase shifts of IRS. Section V provides the simulation results, demonstrating the gains of the proposed approach.
over a semidefinite relaxation (SDR) based alternative. The conclusion is provided in Section VI.

Notations: Lower-case and upper-case boldface letters denote vector and matrix, respectively; \{\cdot\}^*, \{\cdot\}^T, and \{\cdot\}^H stand for conjugate, the transpose, and the conjugate transpose; \text{tr}\{\cdot\} and \text{diag}\{\cdot\} stand for trace and diagonalization; \left[\cdot\right]_{i,j} denotes the \((i, j)\)-th entry of a matrix; \(j = \sqrt{-1}\); \text{Re}\{\cdot\}, \text{Im}\{\cdot\} and \text{arg}\{\cdot\} denote the real part, imaginary part and phase of a complex value; \otimes and \odot denote the Kronecker and Hadamard products, respectively; \mathbb{C} and \mathbb{R} denote the complex space and real space, respectively; \mathbb{E}\{\cdot\} stands for the expectation; \|\cdot\| denotes the \(\ell_2\) norm of vector \(\cdot\); and superscript \((\cdot)^{(t)}\) indicates the \(t\)-th iteration step.

II. SYSTEM MODEL

A. IRS Assisted Uplink MmWave SIMO system

Fig. 1 illustrates the proposed IRS-assisted uplink mmWave system. The system includes an SIMO wireless communication scenario where \(K\) single-antenna users communicate with the AP equipped with an \(M\)-element uniform linear array (ULA). The IRS consists of \(N_{az}\) horizontally arranged elements and \(N_{el}\) vertically arranged passive reflective elements. \(N = N_{az} \times N_{el}\) is the total number of passive elements. The AP configures the phase shifts of the IRS, and then the results are fed back from the AP to an IRS controller via dedicated control link. The phase shifters of IRS is configured in real-time by a smart controller. We assume that the AP has perfect channel state information (CSI) of all relevant mmWave links, as typically assumed in the literature \cite{15, 16, 21, 24}.

Denote the transmit signal from the \(k\)-th user by \(x_k = \sqrt{p_k} s_k\) with \(s_k\) and \(p_k\) representing the normalized power information symbol and transmit power. The received signal at the AP from the \(k\)-th user can be written as

\[
y_k = f_k^H \left( \left( h_{d,k} + G\Theta h_{r,k} \right) x_k + \sum_{j \neq k}^{K} \left( h_{d,j} + G\Theta h_{r,j} \right) x_j + u_k \right),
\]

where \(f_k \in \mathbb{C}^{M \times 1}\) is the multi-user detection vector of the \(k\)-th user and \(F = [f_1, f_2, \cdots, f_K]\); \(h_{d,k} \in \mathbb{C}^{M \times 1}\) is the channel between AP and the \(k\)-th user; \(G \in \mathbb{C}^{M \times N}\) is the mmWave channel matrix between the AP and IRS, and \(h_{r,k} \in \mathbb{C}^{N \times 1}\) is the channel between IRS and the \(k\)-th user; the phase shift matrix of IRS is denoted by \(\Theta = \sqrt{\mu}\text{diag}([\theta_1, \cdots, \theta_N]^T)\), where \(\mu\) indicates the
reflection coefficient\(^1\) and \(\theta_n = e^{j\varphi_n}\) with \(\varphi_n\) being the reflection phase shift; and \(u_k \in \mathbb{C}^{M \times 1}\) is the noise vector which follows the circularly symmetric complex Gaussian (CSCG) distribution of \(\mathcal{CN}(0, \sigma_u^2 I)\).

### B. Wireless Channel Model

Based on the widely used 3D Saleh-Valenzuela channel model [26], [27] for mmWave communications, the channel between AP and the \(k\)-th user can be characterized as

\[
h_{d,k} = \sqrt{\frac{M}{L + 1}} \left[ \xi_{k,0} g_B g_U a_M (\phi_{d,k,0}) + \sum_{\ell=1}^{L} \xi_{k,\ell} g_B g_U a_M (\phi_{d,k,\ell}) \right],
\]

where \(L\) denotes the number of non-line-of-sight (NLoS) paths\(^2\); \(\ell = 0\) represents the LoS path; \(\xi_{k,\ell}\) expresses the complex channel gain of the \(\ell\)-th path and \(\phi_{d,k,\ell} \in [-\pi/2, \pi/2]\) is the associated angle-of-arrivals (AoA); \(g_U\) and \(g_B\) indicate the transmit and receive antenna element gain, respectively; and \(a_M \in \mathbb{C}^{M \times 1}\) is the normalized array steering vector of ULA. For the mmWave channel between IRS and users, the IRS can be properly deployed in the hotspot area, thus causing a high LoS probability. Due to severe path loss, the transmit power of two or more

---

\(^1\)We set \(\mu = 1\) in the sequel for simplicity, since the incident signal energy is not absorbed to drive the circuit for IRS, which is different from the backscatter communications.

\(^2\)There have shown that mmWave channels normally consist of only a small number of dominant multipath components, and typically exhibit 3-5 paths in realistic environments [28], [29], while the scattering at sub-6 GHz is generally rich.
reflections can be ignored so that only LoS path is considered [15], [30]. As such, we make the simplifying assumption that the channel for each user is an LoS case, as given by

$$h_{r,k} = \sqrt{N} \xi_k \delta \Theta \delta \Upsilon \alpha_N \left( \phi_{r,k} \right),$$

where $\xi_k$ indicates the complex channel gain for $k$-th user and $\alpha_N \left( \phi_{r,k} \right)$ is also defined in the same manner as the channel parameters mentioned above.

Accordingly, it is a reasonable assumption that the channel between the AP and IRS can be modeled as a rank-one matrix, which can be mathematically expressed as

$$G = \sqrt{MN} \xi \delta B \delta \Upsilon \alpha_M \left( \phi \right) \alpha^H_N \left( \vartheta_{az}, \vartheta_{el} \right),$$

where $\xi$ denotes the channel gain; $\alpha_M \left( \phi \right) \in \mathbb{C}^{M \times 1}$ is the receiver array steering vector at the AP along the direction $\phi$, and $\alpha_N \left( \vartheta_{az}, \vartheta_{el} \right) \in \mathbb{C}^{N \times 1}$ is the transmitter antenna array steering vector for elevation angle $\vartheta_{el}$ and azimuth angle $\vartheta_{az}$ at the IRS. In (4),

$$\alpha_M \left( \phi \right) = \frac{1}{\sqrt{M}} \left[ e^{-j\frac{2\pi d}{\lambda} \phi} \right]_{i \in \mathcal{I}(M)},$$

$$\alpha_N \left( \vartheta_{az}, \vartheta_{el} \right) = \alpha_{Na} \left( \vartheta_{az} \right) \otimes \alpha_{Ne} \left( \vartheta_{el} \right),$$

where $\lambda$ is the mmWave wavelength, $d$ is the antenna spacing, and $\mathcal{I}(N_\delta) = \{ n - (N_\delta - 1)/2, n = 0, 1, \cdots, N_\delta - 1 \}^3$.

C. Problem Statement

In the uplink SIMO scenario, the upload data amount of the $k$-th user can be characterized by a parameter pair $\{ D_k, T_k \}$, where $D_k$ and $T_k$ denote the data size in number of nats and maximum latency, respectively. To reduce system complexity, we assume that $D_k = D \ (k = 1, 2, \cdots, K)$ and $T_1 \leq T_2 \leq \cdots \leq T_K$.

In this paper, the time-division duplexing (TDD) protocol is adopted, and all the CSIs can be acquired by leveraging the channel reciprocity. We assume the quasi-static flat-fading model for all mmWave channels. Next, let us concentrate on the total user power minimization problem in the multi-user mmWave SIMO system. To be specific, the sum transmit power $\sum_{k=1}^{K} p_k$, multi-
user detection matrix $F$ and passive beamforming matrix $\Theta$ should be designed to meet the predefined upload latency requirements for each user. Accordingly, the problem is formulated as

$$
(P1): \min_{p, F, \Theta} \sum_{k=1}^{K} p_k,
$$

subject to:

1. $p_k \geq 0, \quad \forall k$; (7a)
2. $\theta_n \in F, \quad \forall n$; (7b)
3. $\frac{D}{W \log (1 + \Gamma_k)} \leq T_k, \quad \forall k$. (7c)

where $p = [p_1, p_2, \ldots, p_K]^T$ denotes the vector of allocated mobile powers, and the continuous feasible set for $\theta_n$ is given by

$$
F = \{\theta_n = e^{j\varphi_n} | \varphi_n \in [0, 2\pi)\}. \quad (8)
$$

In the latency constraint (7c), the uplink transmission rate of the $k$-th user is $W \log (1 + \Gamma_k)$, where $W$ is the channel bandwidth, and the SINR of the $k$-th user is given by

$$
\Gamma_k(p, F, \Theta) = \frac{p_k |f_k^H(h_{d,k} + G\Theta h_{r,k})|^2}{\sum_{j=1,j\neq k}^{K} p_j |f_k^H(h_{d,j} + G\Theta h_{r,j})|^2 + \sigma_u^2 \| f_k \|^2}. \quad (9)
$$

### III. MULTI-USER MMWAVE SIMO SYSTEM

#### A. Power Control And Multi-User Detection

In the uplink wireless transmission system, the joint optimization problem (P1) is untractable due to its nonconvex constant-modulus constraint in (7b) and coupled optimization variables. Fundamentally, this problem cannot be solved globally. To tackle these challenges, we devise an alternating optimization algorithm as the main design approach. Actually, this approach is applicable to a set of problems whose subproblems can be solved by developing standard algorithms in each step. The key idea is to alternately optimize each of variables by fixing other variables, thus achieving the decomposition of complicated problems and multi-variable decoupling.

We start with the uplink transmit power control when given $F$ and $\Theta$. For notational brevity, we define $h_k = h_{d,k} + G\Theta h_{r,k}$. Thus, (9) becomes

$$
\Gamma_k(p) = \frac{p_k |f_k^H h_k|^2}{\sum_{j=1,j\neq k}^{K} p_j |f_k^H h_j|^2 + \sigma_u^2 \| f_k \|^2}. \quad (10)
$$
Plugging (10) into (7c), the latency constraint can be reformulated as

\[-p_k \|f_k^H h_k\|^2 + \bar{T}_k \left( \sum_{j=1,j\neq k}^{K} p_j \|f_k^H h_j\|^2 + \sigma_u^2 \|f_k\|^2 \right) \leq 0, \quad \forall k, \] (11)

where \(\Gamma_k(p) \geq \bar{T}_k\) with \(\bar{T}_k = e^{\frac{P}{W_k}} - 1\) being the minimum protection ratio for \(k\)-th user. This constraint can be presented in a matrix form as

\[(I - Q)p \succeq \tau,\] (12)

where

\[[Q]_{i,j} = \begin{cases} 0, & \text{if } j = i, \\ \frac{\bar{T}_i |f_i^H h_i|^2}{|f_i^H h_i|^2}, & \text{otherwise}, \end{cases} \] (13)

\[\tau = \begin{bmatrix} \sigma_u^2 \bar{T}_1 f_1^H f_1 & \sigma_u^2 \bar{T}_2 f_2^H f_2 & \cdots & \sigma_u^2 \bar{T}_K f_K^H f_K \\ \frac{\|f_1^H h_1\|^2}{\|f_1^H h_1\|^2} & \frac{\|f_2^H h_2\|^2}{\|f_2^H h_2\|^2} & \cdots & \frac{\|f_K^H h_K\|^2}{\|f_K^H h_K\|^2} \end{bmatrix}^T.\] (14)

As such, the corresponding optimization problem of (P1) is reformulated as

\[(P2) : \min_p \quad 1^T p \] s.t. \( (I - Q)p \succeq \tau, \) \hspace{1cm} (15a)

\[p \succeq 0. \] (15b)

It should be pointed out that \(\tau\) in the inequality constraint (15a) is the vector associated with minimum protection ratio for all users. This implies that we can minimize the objective function in (P2) by identifying the critical points of the inequality (15a). Further, as long as a feasible solution \(F\) is found to admit the spectral radius of \(Q\) less than unity\(^4\), the matrix \(I - Q\) is proven to be invertible [31]. Then, based on this assumption, we can update \(p\) by

\[p = (I - Q)^{-1} \tau. \] (16)

The assumption property of \(Q\) provokes a Neumann series expansion [31], i.e., \((I - Q)^{-1} = \sum_{w=1}^{\infty} Q^w\). Defining the \(w\)-th iteration of power as \(p^{(w)} = Qp^{(w-1)} + \tau\), the optimal \(p\) can be

\(^4\)In fact, this assumption property will be satisfied when specified \(F\) obtained in the following parts.
obtained with an arbitrary start vector $p^{(0)}$ in an iterative manner:

$$\lim_{w \to \infty} p^{(w)} = \lim_{w \to \infty} \left\{ Q^{w} p^{(0)} + \left[ Q^{w-1} + Q^{w-2} + \cdots + Q + I \right] \tau \right\}.$$  \hspace{1cm} (17)

Inspired by the results in (17), an equivalent solution can be obtained by taking the limit of the following iterative update process:

$$p_k^{(t+1)} = \sum_{j=1,j \neq k}^{K} \frac{\tilde{T}_k \left| f_k^H h_j \right|^2 p_j^{(t)} + \sigma_u^2 \tilde{T}_k \left| f_k \right|^2}{\left| f_k^H h_k \right|^2}, \quad k = 1, 2, \cdots, K. \hspace{1cm} (18)$$

This iteration scheme for the solution of transmit power implies that, the matrix inversion operation in (16) can be bypassed, thus resulting in an efficient computation of power minimization.

Next, we proceed with the optimal solution of $F$ for the minimization of (18). This subproblem is given by

$$f_k = \arg \min_{f_k} \left\{ \frac{\tilde{T}_k \left( \sum_{j=1,j \neq k}^{K} \left| f_k^H h_j \right|^2 p_j + \sigma_u^2 \left| f_k \right|^2 \right)}{\left| f_k^H h_k \right|^2} \right\},$$

$$= \arg \min_{f_k} \left\{ \frac{f_k^H \left( \sum_{j=1,j \neq k}^{K} p_j D_j + \sigma_u^2 I \right) f_k}{f_k^H D_k f_k} \right\}, \quad k = 1, 2, \cdots, K, \hspace{1cm} (19)$$

where $D_j = h_j h_j^H$ for $j = 1, 2 \cdots, K$ are complex Hermitian matrices, and hence the square matrix $\left( \sum_{j=1,j \neq k}^{K} p_j D_j + \sigma_u^2 I \right)$ is Hermitian. It is shown that, (19) is, in essence, in the form of Rayleigh quotient minimization [32]. Thus, the optimal problem of $f_k$ is equivalent to the following generalized eigenvalue problem:

$$(P3): \min_{f_k} \frac{f_k^H \left( \sum_{j=1,j \neq k}^{K} p_j h_j h_j^H + \sigma_u^2 I \right) f_k}{f_k^H h_k h_k^H f_k}$$

s.t. $f_k^H h_k = 1, \quad k = 1, 2, \cdots, K.$  \hspace{1cm} (20a)

Based on the minimum variance distortionless response (MVDR) [33] scheme, the optimal $f_k$ in (P3) can be given by

$$f_k^{(t+1)} = \frac{\left( \sum_{j=1,j \neq k}^{K} p_j^{(t+1)} h_j h_j^H + \sigma_u^2 I \right)^{-1} h_k}{h_k^H \left( \sum_{j=1,j \neq k}^{K} p_j^{(t+1)} h_j h_j^H + \sigma_u^2 I \right)^{-1} h_k}. \hspace{1cm} (21)$$

In the following, we present the convergence analysis of the above iteration method. Defining
the $k$-th element of the power mapping as

$$\mathcal{M}_k(p) = \min_{f_k} \left\{ \sum_{j \neq k} \frac{\widetilde{T}_k |f_k^H h_j|^2}{|f_k^H h_k|^2} p_j + \frac{\sigma_u^2 |\widetilde{T}_k|^2}{|f_k^H h_k|^2} \right\}, \quad \text{s.t.} \quad f_k^H h_k = 1, \quad (22)$$

we denote the iteration procedure of uplink power vector as $p^{(t+1)} = \mathcal{M}(p^{(t)})$. Note that the optimal multi-user detection vectors are given in a closed form, which can always keep the mapping $\mathcal{M}_k$ non-increasing by updating $f_k$. Furthermore, it also provides the guarantee for the conditions that spectral radius of $Q$ is less than one.

**Lemma 1.** The mapping $\mathcal{M}$ for iterative power update has a unique fixed point.

*Proof:* See Appendix A.

---

**B. Problem Transformation of Passive Beamforming**

After the completion of joint transmit power control and multi-user detection subproblems, we investigate the passive beamforming subproblem. Given $F$ and $p$, problem (P1) is reduced to optimize the reflecting coefficients. We denote the reflecting coefficients as $\theta = \text{tr}(\Theta) = [\theta_1, \cdots, \theta_N]^T$. Then, we have the following variable substitution:

$$G\Theta h_{r,j} = G \cdot \text{diag}(h_{r,j}) \cdot \theta = G_{h,j} \theta. \quad (23)$$

For notational simplicity, we define the variables independent of $\theta$:

$$b_{k,j} = f_k^H h_{d,j}, \quad (24)$$

$$g_{k,j}^H = f_k^H G_{h,j}. \quad (25)$$

Using the new variables as defined above, (9) can be rewritten as

$$\Gamma_k(\theta) = \frac{p_k |f_k^H (h_{d,k} + G_{h,k} \theta)|^2}{\sum_{j=1,j \neq k}^K p_j |f_k^H (h_{d,j} + G_{h,j} \theta)|^2 + \sigma_u^2 \|f_k\|^2},$$

$$= \frac{p_k |b_{k,k} + g_{k,k}^H \theta|^2}{\sum_{j=1,j \neq k}^K p_j |b_{k,j} + g_{k,j}^H \theta|^2 + \sigma_u^2 \|f_k\|^2}. \quad (26)$$
By substituting (26) into (18), the optimization of \( \theta \), given \( F \) and \( p \), can be equivalently transformed to

\[
\theta^{(t+1)} = \arg \min_{\theta \in F} \sum_{k=1}^{K} \left\{ \sum_{j=1, j \neq k}^{K} \tilde{T}_k \left| b_{k,j} + g_{k,j}^H \theta \right|^2 p_j + \sigma^2 \left| b_{k,k} + g_{k,k}^H \theta \right|^2 \right\}.
\]  

(27)

It is very arduous to solve this optimization problem due to the non-convexity of the feasible set associated with \( \theta \). We will provide two efficient algorithms to deal with this passive beamforming subproblem in the next section. Eventually, these three optimization processes corresponding to \( p \), \( F \) and \( \Theta \) repeat in an alternating manner until a stop condition is met. For ease of illustration our proposed alternating optimization framework, the uplink transmit power minimization method is described in Algorithm 1.

**Algorithm 1** The proposed alternating optimization framework for mobile power allocation.

**Initialize:** Set feasible values of \( \{p^{(0)}, F^{(0)}, \Phi^{(0)}\} \) and iteration index \( t = 0 \).

1: repeat

2: \( t \leftarrow t + 1 \);
3: With given \( \Theta^{(t-1)} \), obtain the superimposed channel \( h_k^{(t-1)} = h_{d,k} + G \Theta^{(t-1)} h_{r,k} \) for all \( k \);
4: Set iteration index \( l = 0 \), \( p^{(l)} = p^{(t-1)} \) and \( F^{(l)} = F^{(t-1)} \);
5: repeat

6: With given \( F^{(l)} \), \( p^{(l)} \) and \( h_k^{(t-1)} \), update \( p_k^{(l+1)} \) using (18);
7: With given \( p^{(l+1)} \) and \( h_k^{(t-1)} \), update \( f_k^{(l+1)} \) using (21);
8: \( l \leftarrow l + 1 \);
9: until \( \|1^T p^{(l)} - 1^T p^{(l-1)}\| \) converges.
10: Obtain \( p^{(l)} = p^{(l)} \) and \( F^{(l)} = F^{(l)} \).
11: With given \( p^{(l)} \) and \( F^{(l)} \), update \( \Theta^{(l)} \) by solving problem (27).
12: until \( \|1^T p^{(l)} - 1^T p^{(l-1)}\| \) converges.
13: Output optimal \( \{p, F, \Phi\} \) and calculate \( \sum_{k=1}^{K} p_k \).

**IV. REFLECTION BEAMFORMING**

A. *SIMin Fraction Transform Based ADMM*

Considering that IRS tends to consist of massive reflecting elements in realistic environment, there is a strong need for parallel computation algorithms that can be extended to large-scale IRS case. Therefore, in this paper, the original passive beamforming problem (27) is first transformed into a fractional programming problem. Then, we take advantage of the problem specifics to propose a novel algorithm that merges the fraction transform based alternating optimization and ADMM method.
Given \( p \) and \( F \), the optimization problem of \( \Theta \) is reframed as

\[
(P6): \quad \min_{\theta} \quad \sum_{k=1}^{K} \tilde{T}_k \cdot \sum_{j=1, j \neq k}^{K} p_j \frac{|b_{k,j} + g_{k,j}^H \theta|^2 + \sigma_u^2 \|f_k\|^2}{|b_{k,k} + g_{k,k}^H \theta|^2}
\]

s.t. (7b).

Problem (P6) is a weighted sum of inverse SINR minimization problem. Typically, this set of fractional programming problems are addressed by decoupling their numerators and denominators. However, conventional decoupling methods in fractional programming, such as Dinkelbach’s algorithms, cannot resolve the sum-of-ratios case directly [34]–[36]. Moreover, the involvement of complex variables \( \theta \) make the sum-of-inverse minimization (SIMin) problem more complicated. To this end, we present a novel fraction transform technique which is termed as SIMin fraction transform, as given in the following theorem.

**Theorem 1.** Given \( K \) pairs of positive functions \( A_k(\theta) \) and \( B_k(\theta) \), the sum-of-inverse fractional minimization problem is defined to be of the form:

\[
\min_{\theta \in \mathcal{F}} \quad \sum_{k=1}^{K} \frac{A_k(\theta)}{B_k(\theta)},
\]

which is equivalent to

\[
\min_{\theta \in \mathcal{F}} \quad \sum_{k=1}^{K} z_k A_k(\theta)^2 + \sum_{k=1}^{K} \frac{1}{4z_k} B_k(\theta)^2,
\]

where \( z = [z_1, z_2, \cdots, z_K]^T \) is a newly introduced auxiliary vector.

**Proof:** See Appendix B.

With the aid of SIMin fraction transform technique, (P4) can be equivalently cast as

\[
\min_{\theta} \quad J(\theta) = \sum_{k=1}^{K} J_{A,k}(\theta) + \sum_{k=1}^{K} J_{B,k}(\theta),
\]

where

\[
J_{A,k}(\theta) = \beta_k \tilde{T}_k^2 \left( \sum_{j=1, j \neq k}^{K} p_j \frac{|b_{k,j} + g_{k,j}^H \theta|^2 + \sigma_u^2 \|f_k\|^2}{\|b_{k,k} + g_{k,k}^H \theta\|^2} \right)^2,
\]

\[
J_{B,k}(\theta) = \frac{1}{4\beta_k} \frac{1}{\|b_{k,k} + g_{k,k}^H \theta\|^4},
\]
where $\beta = [\beta_1, \beta_2, \cdots, \beta_K]^T$ is the auxiliary vector introduced by the fraction transform technique. Then, according to Theorem 1 and Appendix B, the optimal $\beta_k$ is computed by

$$
\beta_k = \frac{1}{2 \tilde{T}_k |b_{k,k} + g_{k,k}^H \theta|^2} \left( \sum_{j=1,j \neq k}^{K} p_j |b_{k,j} + g_{k,j}^H \theta|^2 + \sigma_u^2 \|f_k\|^2 \right).
$$

(34)

Next, in order to split the optimization variables between terms of $J_{A,k}(\theta)$ and $J_{B,k}(\theta)$, we employ the ADMM method as stated in the following. The augmented Lagrangian of (31) can be expressed as

$$
L_\rho(\theta, q, r) = \sum_{k=1}^{K} J_{A,k}(\theta) + J_{B,k}(q) + \frac{\rho}{2} \|\theta - q + r\|^2 - \sum_{n=1}^{N} \mathbb{1}_F(\theta_n),
$$

(35)

where $\rho$ is the penalty parameter, and indicator function associated with the feasible set of $\theta$ is given by

$$
\mathbb{1}_F(\theta_n) = \begin{cases} 
0, & \theta_n \in F, \\
+\infty, & \text{otherwise.}
\end{cases}
$$

(36)

Thus, ADMM for solving (35) then is the following sequential iterative form:

$$
\theta^{(l+1)} := \arg \min_{\theta_n \in F} \sum_{i=1}^{K} J_{A,k}(\theta) + \frac{\rho}{2} \|\theta - q^{(l)} + r^{(l)}\|^2,
$$

(37)

$$
q^{(l+1)} := \arg \min_{q} \sum_{i=1}^{K} J_{B,k}(q) + \frac{\rho}{2} \|\theta^{(l+1)} - q + r^{(l)}\|^2,
$$

(38)

$$
r^{(l+1)} := r^{(l)} + \theta^{(l+1)} - q^{(l+1)}.
$$

(39)

Since the update rule in (39) is quite straightforward, there is no need for the discussion of the solutions for $r$. Accordingly, we just need to investigate the detailed methodologies of solving (37) and (38) in the following parts.

To solve (37), we first relax the constraint (7b) as

$$
\theta^H e_n e_n^H \theta \leq 1, \quad \forall n = 1, 2, \cdots, N.
$$

(40)

Replacing (7b) by (40), the optimization problem in (37) can be rewritten as

$$
\min_{\theta, \varepsilon} G_1(\theta, \varepsilon) = \sum_{k=1}^{K} J_{A,k}(\theta) + \frac{\rho}{2} \|\theta - q + r\|^2 + \sum_{n=1}^{N} \varepsilon_n (\theta^H e_n e_n^H \theta - 1),
$$

(41)
where $\epsilon = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N]^T$ is the associated dual variable vector for (40). The function $G_1(\theta, \varepsilon)$, obviously, is convex and can be solved by the CVX solver in each iteration of (37). Then, we can perform the projection operation to update $\theta$ and its update process is given by

$$
\theta^o = \arg \min_{\theta} G_1(\theta, \varepsilon) = [\theta^o_1, \theta^o_2, \ldots, \theta^o_N]^T,
$$

(42)

$$
\theta^{(l+1)} = [e^{-j\arg\{\theta^o_1\}}, e^{-j\arg\{\theta^o_2\}}, \ldots, e^{-j\arg\{\theta^o_N\}}]^T.
$$

(43)

It should be noted that the variable splitting effect of ADMM, forces the optimization of (38) to rely on $q$, not $\theta$. As a result, in the iteration of entire ADMM framework, the separate optimization of $q$ is modeled as an unconstrained minimization problem:

$$
\min_q G_2(q) = \sum_{i=1}^{K} J_{B,k}(q) + \frac{\rho}{2} \| \theta - q + r \|^2.
$$

(44)

According to Newton’s method, we can iteratively update $q$ by

$$
q^{(i+1)} = q^{(i)} - \iota \left[ \nabla^2 G_2(q^{(i)}) \right]^{-1} \nabla G_2(q^{(i)}),
$$

(45)

where $\iota$ is a step size factor of Newton’s method. In order to avoid Hessian matrix inversion, we adopt a Quasi-Newton method approximate and update the inverse Hessian matrix at each iteration to reduce the amount of computation. For ease of understanding, the proposed SIMin fraction transform based ADMM method is described in Algorithm 2.

**Algorithm 2** The proposed SIMin fraction transform based ADMM framework.

**Initialize:** Set feasible values of $\{\beta^{(0)}, \theta^{(0)}, q^{(0)}, r^{(0)}\}$ and iteration index $t = 0$.

1: repeat
2: \hspace{0.1cm} Set $t \leftarrow t + 1$;
3: \hspace{0.1cm} With given $\theta^{(t-1)}$, update $\beta^{(t)}$ using (34);
4: \hspace{0.1cm} Set iteration index $l = 0$;
5: \hspace{0.1cm} repeat
6: \hspace{0.1cm} \hspace{0.1cm} Set $l \leftarrow l + 1$;
7: \hspace{0.1cm} \hspace{0.1cm} Update $\theta^{(l)}$ according to (42) and (43);
8: \hspace{0.1cm} \hspace{0.1cm} Update $q^{(l)}$ according to the Quasi-Newton method;
9: \hspace{0.1cm} \hspace{0.1cm} Perform $r^{(l)} = r^{(l-1)} + \theta^{(l)} - q^{(l)}$;
10: \hspace{0.1cm} \hspace{0.1cm} until The value of $\| \theta^{(l)} - q^{(l)} \|$ converges.
11: \hspace{0.1cm} \hspace{0.1cm} Obtain $\theta^{(l)} = \theta^{(l)}$;
12: until The function in (P6) converges.
13: Output $\theta$ and set $\Theta = \text{diag}\{\theta\}$. 
B. Manifold Optimization Scheme

Although the SIMin fraction transform-based ADMM method performs efficiently and it is more suitable for the original weighted sum of inverse SINR minimization problem, dealing with the unit modulus constraints by constraint relaxation technique incurs some performance loss. Additionally, the fraction transform method requires solving the extra auxiliary variables. In the following, we revisit the passive beamforming problem from another point of view. A feasible perspective is to find proper IRS reflection coefficients which can reduce the transmit power while keeping the latency constraints active. To guarantee that the achievable latency of each user is adjusted by passive beamforming to be larger than the latency threshold, the reflection beamforming subproblem is initially transformed to a new optimization problem. Specifically, using (26), constraint (11) can be reformulated as

$$p_k \left| b_{k,k} + g_{k,k}^H \theta \right|^2 \geq \tilde{T}_k \left( \sum_{j=1,j \neq k}^{K} p_j \left| b_{k,j} + g_{k,j}^H \theta \right|^2 + \sigma_u^2 \| \mathbf{f}_k \|^2 \right), \quad \forall k. \quad (46)$$

Then, by unfolding the squared terms in (46), we define the “latency residual” of the $k$-th user terminal as

$$\alpha_k = p_k \left| b_{k,k} + g_{k,k}^H \theta \right|^2 - \tilde{T}_k \left( \sum_{j=1,j \neq k}^{K} p_j \left| b_{k,j} + g_{k,j}^H \theta \right|^2 + \sigma_u^2 \| \mathbf{f}_k \|^2 \right),$$

$$= \theta^H \left( p_k g_{k,k} g_{k,k}^H - \tilde{T}_k \sum_{j \neq k}^{K} p_j g_{k,j} g_{k,j}^H \right) \theta + 2 \text{Re} \left\{ \left( p_k b_{k,k}^* g_{k,k}^H - \tilde{T}_k \sum_{j \neq k}^{K} p_j b_{j,k}^* g_{k,j}^H \right) \theta \right\}$$

$$+ p_k b_{k,k}^2 - \tilde{T}_k \sum_{j \neq k}^{K} \left( p_j b_{k,j}^2 + \| \mathbf{f}_k \|^2 \right), \quad (47)$$

where the latency residual vector is denoted as $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_K]^T$. Therefore, the passive beamforming subproblem (27) becomes a latency residual maximization problem, i.e.,

$$\max_{\theta, \alpha} \sum_{k=1}^{K} \alpha_k.$$  

One straightforward idea is to convert constraints (46) and unit modulus constraints into quadratic constraints. Then, by ignoring the rank-one constraint involved in constructing symmetric matrices, problem (27) can be recast into an SDR form, which is typically solved by eigen-decomposition. However, the transformation into SDR problem incurs the main drawback that the number of optimization variables increases quadratically with the number of IRS elements. For this reason, we develop a CCMO method which is suitable for directly solving the reformulated latency residual maximization problem.
As stated before, the relaxation of constant modulus constraint still leads to mathematically intractable expressions. Further, the special geometry structure of constraint $|\theta_n| = 1$ inspires us to resort to Riemannian-Geometric optimization tools [37]. Indeed, the feasible region of the \textit{latency residual maximization} problem geometrically constitutes a complex circle manifold. Moreover, manifold representation can get a relatively concise form. In the following, we briefly introduce the rationale of manifold optimization while providing our solutions. Riemannian manifold optimization breaks the confinement of the Euclidean space to generalize the gradient descent algorithm on a manifold established from the geometric properties of constraints. By exploiting the manifold formed by constraints, the original constrained optimization problem can be transformed into an unconstrained optimization problem on the manifold, which can be minimized by gradient descent algorithms.

Referring to the latency residual representation (47), the subproblem (27) can be readily reframed as

\[
(P4) : \max_{\theta} \ f_0(\theta) = \theta^H U \theta + 2 \text{Re} (\theta^H v) + C
\]

s.t. $|\theta_n| = 1, \ n = 1, 2, \ldots, N,$ \hspace{1cm} (48a)

where

\[
U = \sum_{k=1}^{K} \left( p_k g_{k,k} g_{k,k}^H - \bar{T}_k \sum_{j \neq k}^{K} p_j g_{k,j} g_{k,j}^H \right), \hspace{1cm} (49)
\]

\[
v = \sum_{k=1}^{K} \left( p_k b_{k,k} g_{k,k} - \bar{T}_k \sum_{j \neq k}^{K} p_j b_{k,j} g_{k,j} \right), \hspace{1cm} (50)
\]

\[
C = \sum_{k=1}^{K} \left[ p_k |b_{k,k}|^2 - \bar{T}_k \left( \sum_{j \neq k}^{K} p_j |b_{k,j}|^2 + \sigma_u^2 \| f_k \|^2 \right) \right]. \hspace{1cm} (51)
\]

According to the notion of manifold optimization, problem (P4) can be reformulated as:

\[
(P5) : \min_{\theta \in S^N} \ f(\theta) = -\theta^H U \theta - 2 \text{Re} (\theta^H v), \hspace{1cm} (52)
\]

where $S^N$ indicates the manifold space defined in the constant modulus constraints of problem (P4). Here, $S^N$ can be expressed as

\[
S^N = \{ \theta \in \mathbb{C}^N : |\theta_1| = |\theta_2| = \cdots = |\theta_N| = 1 \}, \hspace{1cm} (53)
\]
where $\mathcal{S} = \{\theta_n \in \mathbb{C} : \theta_n \theta_n^* = \text{Re}\{\theta_n\}^2 + \text{Im}\{\theta_n\}^2 = 1\}$ is known as a complex circle and can be viewed as a sub-manifold of $\mathbb{C}$. The search space $S^N$ can be viewed as the product of $N$ complex circles. Likewise, this manifold is a sub-manifold of $\mathbb{C}^N$, and is termed as the complex circle manifold.

The main idea of CCMO algorithm is to perform gradient descent algorithm on a complex circle manifold space. The procedure of gradient descent on Riemannian manifold is similar to the counterpart in the Euclidean space. Both of their general frameworks mainly consists of two phases: the first phase, which determines the descent direction of the current solution by computing the negative gradient, and the second phase, which decreases the value of objective function via the line search method [37]. These two main phases are iteratively performed until its stopping criterion is met. Nevertheless, instead of the Euclidean gradient, manifold optimization requires us to calculate the Riemannian gradient as the search direction. The Riemannian gradient of $f(\theta)$ at the current iteration point $\theta^{(i)} \in S^N$ is defined as a projection of search direction in Euclidean space onto the tangent space $T_{\theta^{(i)}}S^N$, which can be expressed as

$$T_{\theta^{(i)}}S^N = \{\eta \in \mathbb{C}^{N+1} : \text{Re}\{\eta^* \odot \theta^{(i)}\} = 0\}.$$  \hspace{1cm} (54)

Then, the Euclidean gradient of $f(\theta^{(i)})$ at $\theta^{(i)}$ can be readily computed by

$$\nabla f(\theta^{(i)}) = -2U\theta^{(i)} - 2v.$$  \hspace{1cm} (55)
By performing the projection operator on the Euclidean gradient, the Riemannian gradient of \( f(\theta(i)) \) is obtained as

\[
\nabla_{SN} f(\theta(i)) = \text{Proj}_{T_{\theta(i)}SN} (\nabla f(\theta(i))) = \nabla f(\theta(i)) - \text{Re} \{ \nabla f(\theta(i))^* \circ \theta(i) \} \circ \theta(i). \tag{56}
\]

Hence, the current point \( \theta(i) \) in the tangent space \( T_{\theta(i)}SN \) is updated as

\[
\theta(i)^{\circ} = \theta(i) - \zeta \nabla_{SN} f(\theta(i)), \tag{57}
\]

where \( \zeta > 0 \) is a well-chosen constant step size\(^5\). It should be noted that \( \theta(i)^{\circ} \) is still in the tangent space \( T_{\theta(i)}SN+1 \) but it leaves manifold \( SN \). Therefore, a Retraction mapping operator is applied to move the point \( \theta(i)^{\circ} \) back to the manifold \( SN \). Finally, the point \( \theta(i+1) \) updated by using the Retraction mapping operator is given by

\[
\theta(i+1) = \text{Ret}_{\theta(i)} \left( -\zeta \nabla_{SN} f(\theta(i)) \right) = \frac{\theta(i) - \zeta \nabla_{SN} f(\theta(i))}{\|\theta(i) - \zeta \nabla_{SN} f(\theta(i))\|} = \theta(i)^{\circ} \odot \frac{1}{|\theta(i)^{\circ}|}. \tag{58}
\]

These operations above are illustrated in Fig. 2 and the details are summarised in Algorithm 3.

**Algorithm 3** The proposed CCMO algorithm for passive beamforming.

**Initialize:** Set feasible values of \( \{\theta(0)\} \) and iteration index \( i = 0 \).

1: repeat
2: \( \text{Set } i \leftarrow i + 1; \)
3: Calculate the Euclidean gradient \( \nabla f(\theta(i)) \) at \( \theta(i) \) using (55);
4: Construct the tangent space \( T_{\theta(i)}SN \) and calculate the current Riemannian gradient \( \nabla_{SN} f(\theta(i)) \) using (56);
5: Perform gradient descent algorithm over the current tangent space using (57);
6: Update \( \theta(i+1) \) using the Retraction mapping operator according to (58);
7: until The value of \( |f(\theta(i)) - f(\theta(i-1))| \) in (P5) converges.
8: Output \( \theta \) and obtain \( \Theta = \text{diag} \{\theta\} \).

\[\text{V. Numerical And Simulation Results}\]

**A. Simulation Setup**

In this section, numerical simulations are carried out to verify the effectiveness of above proposed uplink transmit power control method and potential benefits of deploying IRS in mmWave SIMO systems. We consider an IRS-aided mmWave system illustrated in Fig. 3. More specifically, the AP is equipped with an ULA consisting of \( M = 32 \) antennas, and it is located in

\(^5\)To ensure stability and convergence of the CCMO algorithm, the step size \( \zeta \) should be selected to satisfy \( \zeta \leq 1/\lambda_U \) where \( \lambda_U \) represents the largest eigenvalue of the matrix \( U \) in problem (P5). This optimization problem can be solved by leveraging the Manopt toolbox in MATLAB [38]–[40].
the origin of coordinate system. The IRS is implemented with a URA where the vertical length is set as $N_{az} = 5$ and horizontal length $N_{el}$ varies in different evaluations. The IRS central element is placed at $(80 \text{ m}, 0)$. Here, both single-user and multi-user scenarios are investigated. In the single-user scenario, we assume that only User 1 exists in this system and its coordinate position is taken as $(d_{x1}, d_{y1})$. In the multi-user scenario, both User 1 and User 2 are considered to concurrently upload their data where the coordinate position of User 2 is taken as $(d_{x2}, -d_{y2})$.

As suggested by real-world channel measurements [18], [26], the channel gain $\xi$ follows a complex Gaussian distribution:

$$\xi \sim \mathcal{CN}(0, 10^{-\text{PL}(R)})$$

where $\text{PL}(R)$ expresses the path-loss over a distance $R(\text{m})$, and it is expressed as:

$$\text{PL}(R) = \chi_a + 10\chi_b \log_{10}(R) + \kappa,$$

where $\kappa \sim \mathcal{N}(0, \sigma_{\kappa}^2)$ is the lognormal shadowing variance. Note that for mmWave communications at 28 GHz, the channel gain is generated in two cases. As regards characterization of LoS path, the parameter values of $\chi_a$, $\chi_b$ and $\sigma_{\kappa}$ are set to be $\chi_a = 61.4$, $\chi_b = 2$ and $\sigma_{\kappa} = 5.8$ dB, respectively. As regards NLoS path, $\chi_a$, $\chi_b$ and $\sigma_{\kappa}$ are taken as $\chi_a = 72$, $\chi_b = 2.92$ and $\sigma_{\kappa} = 8.7$ dB, respectively. To investigate the role of IRS in the uplink transmission mode,

the mmWave channel between AP and users considered in this paper can be categorized into following two groups:

- LoS scenario is defined as the case where only pure LoS signal is received.
- Obstructed-line-of-sight (OLoS) scenario is defined as the scenario where the optical LoS
component is blocked and only NLoS components exist, where $\rho_b$ indicates the blockage probability of LoS path.

According to [18], [41], in typical uplink mmWave communications, the antenna gain values for AP-user channel are set as $\varrho_U = 0$ dBi and $\varrho_B = 9.82$ dBi. We take $\varrho_U = 0$ dBi for IRS-user link, and $\varrho_B = 9.82$ dBi for IRS-AP link. Note that the cascaded channel of IRS-user and AP-IRS becomes severely weakened by double-fading effect [20], [42] and high path loss of mmWave. Fortunately, with the tremendous advancements in meta-materials, the reflection gain of IRS elements can compensate channel attenuation completely. Here, the relative reflection gain of IRS is defined as $\nu = \frac{\varrho_I}{\sqrt{\varrho_U \varrho_B}}$.

Unless otherwise stated in this paper, other parameters are set as follows: noise variance is $\sigma_u^2 = -85$ dBm; transmission bandwidth is $W = 500$ MHz; the data transmitted from each user is $D = 5000$ nats; the relative reflection gain is $\nu = 15$ dB; location parameters are taken as $d_{x1} = 40$ m, $d_{y1} = 40$ m, $d_{x2} = 50$ m, $d_{y2} = 20$ m; the number of NLoS paths is $L = 3$; and the upload latency requirement for each user follows the uniform distribution $T_k \sim U(400$ ms, 600 ms).

B. Single-User Scenario

In this subsection, we study the special case of single-user mmWave system\(^6\). The optimal mobile power required for several various parameter settings is verified through Monte Carlo simulations. For comparison of various schemes, two baseline schemes are considered as follows:

- **Without IRS**: We consider the uplink mobile power allocation based on the direct link of AP-user only.
- **SDR**: In this scheme, we adopt the SDR based iterative optimization approach to obtain reflection coefficients.

First, we investigate the impact of the number of IRS elements on uplink transmit power attained by all schemes. In Figs. 4(a) and (b), the mobile powers allocated are shown for the LoS and OLoS scenarios, respectively. As reasonably expectable, the total power required by the mmWave SIMO system without IRS is kept constant in both scenarios. In contrast, all the three schemes with IRS exhibit the superiority over the baseline scheme without IRS. It is observed that the schemes with IRS can substantially reduce the mobile power in OLoS scenario whereas

\(^6\)For the case of single-user transmission, any interference terms are eliminated from the above analysis.
Fig. 4. Transmit power versus number of IRS elements, $N$. (a) LoS scenario; (b) OLoS scenario.

Fig. 5. Transmit power versus AP-user horizontal distance, $d_{x1}$. (a) LoS scenario; (b) OLoS scenario.

only a slight improvement can be achieved in LoS scenario. This can be explained because the
direct link of strong LoS incurs considerably higher channel gain than the supplementary link
provided by IRS. Moreover, due to the dominance of LoS path, the three schemes with IRS
give about the same yield. However, Fig. 4(b) reveals that the supplementary link is dominant in
LOoS scenario thanks to the IRS-enhanced received power. It can be also seen that, as a tendency,
the optimal mobile power decreases with $N$ increasing since large-scale IRS can provide the
beneficial array gain as well as passive beamforming gain. Especially for OLoS scenarios, our
proposed CCMO and ADMM methods stably precede the SDR method.

Fig. 5 examines the uplink transmit power versus the the horizontal distance between AP and
User 1. In the current settings, it is noted that with $d_{x1}$ increasing from 10 to 70, the distance
between AP and User 1 gradually increases and the distance between IRS and User 1 gradually
decreases. Evidently, the power required by the scheme without IRS increases rapidly as user moves further from AP. Meanwhile, the prominent benefits brought by IRS are shown in both LoS and OLoS scenarios, which is different from Fig. 4. This conclusion can be inferred from the high propagation loss relying critically on transmission distance. Additionally, significant advantage achieved by our proposed ADMM and CCMO algorithms over two baselines is illustrated in Fig. 5. As seen in Fig. 5(a), the performance gap between the scheme without IRS and the other schemes of IRS-aided system begins to rise from the distance of 40 m. This is because the resulting high propagation loss between AP and user enables the reflection gain of IRS. As expected, in Fig. 5(b), this performance gap keeps getting wider for the same reason. More importantly, due to the weak AP-user link in OLoS scenario, the IRS-user link maintains the dominance. This implies an intriguing insight that the closer the user moves to the IRS, the easier both the antenna and passive beamforming gain of the IRS can be exploited. Put another way, IRS instead of the user device transmits the uploading data to AP intrinsically.

Then, we compare the optimal power performance achieved by all schemes versus the data size of task to be uploaded. In the above evaluations, both the amount of data and latency distribution are assumed to be fixed. Here, the amount of data varies from 4000 to 9000 nats. Fig. 6(a) depicts the LoS scenario where the powers achieved by all the schemes increase as the amount of data enlarges. As expected, the total power required by the IRS-aided system is lower than the system without IRS. But, this power gap becomes quite large for OLoS scenario shown in Fig. 6(b). We also observe that the proposed CCMO and ADMM methods significantly outperform the SDR method, which coincides well with the analysis of Fig. 4.
In Fig. 7, we assess the impact of relative reflection gain on transmit power for all schemes. As depicted, with relative reflection gain increasing from 10 to 20 dB, all the schemes with IRS exhibit the same downward trend. It is intuitive that the channel quality corresponding to the IRS-related links becomes stronger when increasing the reflection gain. Hence, the user is prone to exploit the better links to avoid allocating more power. More crucially, IRS can be deployed in an economical way to further save power consumption.

To further elaborate the appealing benefits of merging IRS with mmWave system, Fig. 8 plots the achievable minimum user power under different probabilities of LoS blockage. These results clearly show that, the larger blockage probability incurs considerably higher transmit powers achieved by all schemes. For the case of $\rho_b = 0$ indicating the LoS scenario, the performance gain induced by the IRS can be neglected, while for the case of $\rho_b = 1$ indicating the OLoS scenario, our proposed IRS-assisted uplink mmWave transmission design is capable of saving transmission power consumption while meeting predefined upload latency requirements. Fig. 8 also shows the performance superiority of the proposed ADMM and CCMO approaches over the SDR methods.

C. Multi-User Scenario

We now analyze a more interesting multi-user mmWave system, where User 1 and User 2 simultaneously upload their computing tasks via millimeter wave channels. In a real-world situation, the direct AP-user channel qualities are not all the same due to mobility and different geographic locations of each user. For ease of illustration, we presume the blocking events to
be statistically independent modeled as in [43]. Next, we present some curious observations of mobile power allocation in various multi-user scenarios.

In Fig. 9 and Fig. 10, we compare the proposed algorithms with two baselines for two users. Fig. 9 considers that both two users have identified available LoS links. In this way, the schemes with IRS have shown a little improvement compared with the system without IRS. As the intuition suggests, User 1 requires more power than User 2 due to the latency relationship $T_1 \leq T_2$, and for this reason IRS contributes mainly to the power optimization of User 1. Afterwards, Fig. 10 considers that both two users have identified the blocking events. Similar to the previous analysis, the supplementary links provided by IRS play a dominant role in joint optimization of mobile powers for two users.

In Fig. 11, we consider that User 1 has identified an available LoS link while User 2 has identified the blocking events. We can notice the substantial power performance gain attained
by the schemes with IRS for User 2, whereas the involvement of IRS has little effect on User 1. This again validates the benefits brought by IRS are prominent when direct links are weaker. Although the latency requirement of User 1 is higher than that of user 2 (i.e. User 1 has higher service priority), the available LoS channel is strong enough to cause the lower transmit power than User 2. For the case shown in Fig. 12 where User 1 has identified the blocking events while User 2 has identified the strong LoS link, we can come to exactly the opposite observations for Fig. 11. This can be easily explained by the analysis of Fig. 11. To conclude, IRS mainly acts as a dominant link when weak direct AP-user channel incurs in multi-user scenarios.

VI. CONCLUSION

In this paper, we have proposed a novel IRS-assisted uplink mmWave transmission scheme, where IRS can resolve link blockage problems by creating supplementary links to guarantee the real-time data upload under stringent latency requirements. Our objective is to minimize the uplink transmit power for all users while satisfying the predefined latency targets, where individual device power, multi-user detection matrix and passive beamforming coefficients should be jointly optimized. First, we have designed an alternating optimization framework to transform the joint optimization problem into three tractable subproblems. Then, the closed-form solutions of uplink power and multi-user detection matrix in each iteration have been derived. Considering the high computational complexity of SDR method, we have developed two algorithms well suited to the passive beamforming problems of large-scale IRS. Our simulation results have validated the benefits brought by IRS in the uplink mmWave SIMO system in the presence of channel intermittency, where some interesting insights on the role of IRS in uplink power allocation for different scenarios have been provided. Furthermore, our proposed algorithms were shown to be superior to the baseline schemes.

APPENDIX A

PROOF OF LEMMA 1

Assume two sets of positive power vectors \(\hat{\mathbf{p}}\) and \(\mathbf{p}^*\) are the fixed points of the mapping \(\mathcal{M}\). Without loss of generality, we suppose that there exists an index \(k\) satisfying \(\hat{p}_k > p_k^*\), and let \(\max_j \left(\hat{p}_j/p_j^*\right) = \gamma > 1\). Thus, we have the fact \(\gamma \mathbf{p}^* \geq \hat{\mathbf{p}}\). Here, we can find an index \(i\) such that
\( \gamma p^*_i = \hat{p}_i \). According to the fact that both \( \hat{p} \) and \( p^* \) are the fixed points of the mapping \( \mathcal{M} \), we have

\[
\hat{p}_i = \min_{f_i} \left\{ \sum_{j=1,j\neq i}^{K} \frac{\bar{T}_i \left| f_i^H h_j \right|^2}{\left| f_i^H h_i \right|^2} \hat{p}_j + \frac{\sigma_i^2 \bar{T}_i \left| f_i \right|^2}{\left| f_i^H h_i \right|^2} \right\},
\]

\[
\leq \min_{f_i} \left\{ \sum_{j=1,j\neq i}^{K} \frac{\bar{T}_i \left| f_i^H h_j \right|^2}{\left| f_i^H h_i \right|^2} \gamma p^*_j + \frac{\sigma_i^2 \bar{T}_i \left| f_i \right|^2}{\left| f_i^H h_i \right|^2} \right\},
\]

\[
< \gamma \left( \min_{f_i} \left\{ \sum_{j=1,j\neq i}^{K} \frac{\bar{T}_i \left| f_i^H h_j \right|^2}{\left| f_i^H h_i \right|^2} p^*_j + \frac{\sigma_i^2 \bar{T}_i \left| f_i \right|^2}{\left| f_i^H h_i \right|^2} \right\} \right),
\]

\[
= \gamma p^*_i. \tag{61}
\]

From the explicit contradiction derived above, we conclude that the fixed point of mapping \( \mathcal{M} \) is unique. \( \blacksquare \)

**APPENDIX B**

**PROOF OF THEOREM 1**

Let us introduce the following equivalence relationship:

\[
z_k A_k(\theta)^2 + \frac{1}{4z_k B_k(\theta)^2} = \left( \sqrt{z_k A_k(\theta)} - \frac{1}{2\sqrt{z_k B_k(\theta)}} \right)^2 + A_k(\theta) \frac{\left( \frac{1}{B_k(\theta)} \right)}{B_k(\theta)},
\]

(62)

It is facile to see that minimizing the right-hand side of (62) with respect to \( z \) and \( \theta \) is equivalent to the minimization of its left-hand side. Meanwhile, we observe that the optimal \( z \) in minimizing the right-hand side of (62) can be always found by

\[
z_k = \frac{1}{2A_k(\theta)B_k(\theta)},
\]

so that the squared term in (62) becomes zero. Therefore, the optimal \( \theta \) for the left-hand side of (62) is always part of the solutions of minimizing \( \frac{A_k(\theta)}{B_k(\theta)} \).

\( \blacksquare \)

**REFERENCES**

[1] T. K. Vu, C. Liu, M. Bennis, M. Debbah, M. Latva-aho, and C. S. Hong, “Ultra-reliable and low latency communication in mmWave-enabled massive MIMO networks,” IEEE Commun. Lett., vol. 21, no. 9, pp. 2041–2044, Sep. 2017.

[2] M. Alonzo, S. Buzzi, A. Zappone, and C. D’Elia, “Energy-efficient power control in cell-free and user-centric massive MIMO at millimeter wave,” IEEE Trans. Green Commun. and Networking, vol. 3, no. 3, pp. 651–663, Sep. 2019.

[3] S. Guo, H. Zhang, P. Zhang, P. Zhao, L. Wang, and M. Alouini, “Generalized beamspace modulation using multiplexing: A breakthrough in mmWave MIMO,” IEEE J. Select. Areas Commun., vol. 37, no. 9, pp. 2014–2028, Sep. 2019.
[4] N. S. Perović, M. Di Renzo, and M. F. Flanagan, “Channel capacity optimization using reconfigurable intelligent surfaces in indoor mmWave environments,” arXiv preprint arXiv:1910.14310, 2019.

[5] F. Boccardi, R. W. Heath, 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.

[6] S. Barbarossa, E. Ceci, and M. Merluzzi, “Overbooking radio and computation resources in mmW-mobile edge computing to reduce vulnerability to channel intermittency,” in Proc. Euro. Conf. Networks and Commun. (EuCNC), Jun. 2017, pp. 1–5.

[7] N. di Pietro, M. Merluzzi, E. C. Strinati, and S. Barbarossa, “Resilient design of 5G mobile-edge computing over intermittent mmWave links,” arXiv preprint arXiv:1901.01894, 2019.

[8] N. Nouri, J. Abouei, M. Jaseemuddin, and A. Anpalagan, “Joint access and resource allocation in ultra-dense mmWave NOMA networks with mobile edge computing,” IEEE Internet Thing J., pp. 1–1, Nov. 2019.

[9] C. Zhao, Y. Cai, M. Zhao, and Q. Shi, “Joint hybrid beamforming and offloading for mmWave mobile edge computing systems,” in Proc. IEEE Wireless Commun. Networking Conf. (WCNC), Apr. 2019, pp. 1–6.

[10] S. Barbarossa, E. Ceci, M. Merluzzi, and E. Calvanese-Strinati, “Enabling effective mobile edge computing using millimeterwave links,” in Proc. Int. Conf. Commun. Work. (ICC Workshops), May 2017, pp. 367–372.

[11] C. Huang, G. C. Alexandropoulos, A. Zappone, M. Debbah, and C. Yuen, “Energy efficient multi-user MISO communication using low resolution large intelligent surfaces,” in Proc. IEEE GLOBECOM Workshops (GC Workshops), Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[12] 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 2018.

[13] M. D. Renzo, M. Debbah, D. T. P. Huy, A. Zappone, M. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin, J. de Rosny, A. Bounceur, G. Lerosey, and M. Fink, “Smart radio environments empowered by reconfigurable AI meta-surfaces: an idea whose time has come,” EURASIP J. Wireless Comm. and Networking, vol. 2019, no. 129, May 2019.

[14] S. Hu, K. Chitti, F. Rusek, and O. Edfors, “User assignment with distributed large intelligent surface (LIS) systems,” in Proc. IEEE Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC), Bologna, Italy, Sep. 2018, pp. 1–6.

[15] Q. Wu and R. Zhang, “Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming,” IEEE Trans. Wireless Commun., pp. 1–1, Aug. 2019.

[16] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, “Reconfigurable intelligent surfaces for energy efficiency in wireless communication,” IEEE Trans. Wireless Commun., vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[17] Q. Wu and R. Zhang, “Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints,” arXiv preprint arXiv:1910.06220, 2019.

[18] P. Wang, J. Fang, X. Yuan, Z. Chen, H. Duan et al., “Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design,” arXiv preprint arXiv:1908.10734, 2019.

[19] Y. Han, W. Tang, S. Jin, C. Wen, and X. Ma, “Large intelligent surface-assisted wireless communication exploiting statistical CSI,” IEEE Trans. Veh. Technol., vol. 68, no. 8, pp. 8238–8242, Aug. 2019.

[20] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, “Weighted sum-rate optimization for intelligent reflecting surface enhanced wireless networks,” arXiv preprint arXiv:1905.07920, 2019.

[21] M. Cui, G. Zhang, and R. Zhang, “Secure wireless communication via intelligent reflecting surface,” IEEE Wireless Commun. Lett., pp. 1–1, May 2019.

[22] H. Shen, W. Xu, S. Gong, Z. He, and C. Zhao, “Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications,” IEEE Commun. Lett., vol. 23, no. 9, pp. 1488–1492, Sep. 2019.
[23] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, “Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces,” in Proc. IEEE Int. Work. on Signal Process. Advances for Wireless Commun. (SPAWC), Cannes, France, Jul. 2019, pp. 1–5.

[24] T. Bai, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, “Latency minimization for intelligent reflecting surface aided mobile edge computing,” arXiv preprint arXiv:1910.07990, 2019.

[25] A. V. Guglielmi, M. Levorato, and L. Badia, “A bayesian game theoretic approach to task offloading in edge and cloud computing,” in Proc. IEEE Int. Conf. Commun. Work. (ICC Wkshps), May 2018, pp. 1–6.

[26] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan et al., “Millimeter wave channel modeling and cellular capacity evaluation,” IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[27] Z. Lin, T. Lv, and P. T. Mathiopoulos, “3-D indoor positioning for millimeter-wave massive MIMO systems,” IEEE Trans. Commun., vol. 66, no. 6, pp. 2472–2486, Jun. 2018.

[28] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 831–846, Oct. 2014.

[29] Z. Gao, L. Dai, D. Mi, Z. Wang, M. A. Imran et al., “MmWave massive-MIMO-based wireless backhaul for the 5G ultra-dense network,” IEEE Wireless Commun., vol. 22, no. 5, pp. 13–21, Oct. 2015.

[30] Q. Wu and R. Zhang, “Beamforming optimization for intelligent reflecting surface with discrete phase shifts,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Brighton, United Kingdom, May 2019, pp. 7830–7833.

[31] F. Hiai and D. Petz, Introduction to matrix analysis and applications. Springer Science & Business Media, 2014.

[32] D. Hwang, D. I. Kim, and T. Lee, “Throughput maximization for multiuser MIMO wireless powered communication networks,” IEEE Trans. Veh. Technol., vol. 65, no. 7, pp. 5743–5748, Jul. 2016.

[33] F. Rashid-Farrokhi, L. Tassiulas, and K. J. R. Liu, “Joint optimal power control and beamforming in wireless networks using antenna arrays,” IEEE Trans. Commun., vol. 46, no. 10, pp. 1313–1324, Oct. 1998.

[34] K. Shen and W. Yu, “Fractional programming for communication systems-part I: Power control and beamforming,” IEEE Trans. Signal Process., vol. 66, no. 10, pp. 2616–2630, May 2018.

[35] L. Zheng and C. W. Tan, “Optimal algorithms in wireless utility maximization: Proportional fairness decomposition and nonlinear Perron-Frobenius theory framework,” IEEE Trans. Wireless Commun., vol. 13, no. 4, pp. 2086–2095, Apr. 2014.

[36] Z. Wang, L. Vandendorpe, M. Ashraf, Y. Mou, and N. Janatian, “Minimization of sum inverse energy efficiency for multiple base station systems,” arXiv preprint arXiv:1909.04355, 2019.

[37] P.-A. Absil, R. Mahony, and R. Sepulchre, Optimization Algorithms on Matrix Manifolds. Princeton, NJ: Princeton University Press, 2008.

[38] J. Chen, “Manifold optimization approach for data detection in massive multiuser MIMO systems,” IEEE Trans. Veh. Technol., vol. 67, no. 4, pp. 3652–3657, Apr. 2018.

[39] K. Alhujaili, V. Monga, and M. Rangaswamy, “Transmit MIMO radar beampattern design via optimization on the complex circle manifold,” IEEE Trans. Signal Process., vol. 67, no. 13, pp. 3561–3575, Jul. 2019.

[40] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, “Manopt, a matlab toolbox for optimization on manifolds,” J. Mach. Learn. Res., vol. 15, no. 1, pp. 1455–1459, Jan. 2014.

[41] I. A. Hemadeh, K. Satyanarayana, M. El-Hajjar, and L. Hanzo, “Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget,” IEEE Commun. Surveys Tuts., vol. 20, no. 2, pp. 870–913, Dec. 2018.

[42] J. D. Griffin and G. D. Durgin, “Complete link budgets for backscatter-radio and RFID systems,” IEEE Antennas Propagat. Mag., vol. 51, no. 2, pp. 11–25, Apr. 2009.
[43] S. Singh, R. Mudumbai, and U. Madhow, “Interference analysis for highly directional 60-GHz mesh networks: The case for rethinking medium access control,” *IEEE/ACM Trans. Netw.*, vol. 19, no. 5, pp. 1513–1527, Oct. 2011.