Intelligent Reflecting Surface Enhanced Resilient Design for MEC Offloading over Millimeter Wave Links

Yashuai Cao, Student Member, IEEE, and Tiejun Lv, Senior Member, IEEE

Abstract

The merge of mobile edge computing (MEC) and millimeter-wave (mmWave) communications will hopefully enable the fast access for computational resources, where these two technologies can benefit from each other’s potentials. However, the high susceptibility to blocking in mmWave networks imposes crucial challenges for further development of mmWave-MEC vision. In this paper, a novel intelligent reflecting surface (IRS) assisted mmWave-MEC scheme is proposed to overcome the disruptive effect caused by blockage events. In this context, we investigate new methods to minimize mobile power for a multi-user mmWave-MEC system, thus efficiently orchestrating the uplink mobile power resources for latency-constrained computation offloading. In particular, the mobile power is optimized by joint design of individual device power, multi-user detection matrix and passive beamforming. To tackle this issue, we develop an alternating optimization framework so that the joint optimization can be decomposed into tractable subproblems. First, we provide closed-form expressions for the update of powers and multi-user detection vectors in each iteration step. Then, we reformulate the passive beamforming problem. To cater for large-scale IRS scenario, we propose two efficient algorithms including complex circle manifold optimization (CCMO) method and sum-of-inverse minimization (SIMin) fraction transform based alternating direction method of multipliers (ADMM) method. Finally, numerical results corroborate the merits of our proposed IRS assisted mmWave-MEC scheme, and demonstrate the feasibility and effectiveness of our algorithms.

I. INTRODUCTION

The proliferation of various new services and applications in fifth generation of mobile systems (5G), e.g., Industry 4.0, the virtual reality (VR), Internet of Things (IoT) and automated driving,
has brought forward higher requirements on the data rate and system capacity. To enable such diverse service-oriented 5G use cases, the desired quality of service (QoS) requirements in terms of connectivity, reliability, capacity and latency determined by 5G vertical-specific services, are driving the design paradigm shift for mobile systems. Toward the realization of supporting different service types (i.e., distinct QoS requirements) within the identical network infrastructure, network slicing has emerged where tailored virtualized sub-networks can be created [1]. To achieve the stringent latency, reliability and power consumption requirements of processing tasks performed in IoT mobile devices, mobile edge computing (MEC) as a pivotal role in new computing paradigms, aims to empower the network edge with available computing resources. In particular, computation offloading can transfer computation-intensive tasks from resource-hungry devices to edge servers, thus spurring a great deal of interest among researchers.

Within this context, the collaborations of communications, caching and computing (3C), as suggested in [2], have opened up new possibilities for improving the overall mobile system performance. In such 3C framework, evolving physical-layer communications technologies, such as millimeter-wave (mmWave), massive multi-input multi-output (MIMO) and ultra-dense deployment, have provided a strong impulse for computation offloading. Recently, a joint Europe/Japan H2020 Project called 5G-MiEdge (MmWave Edge Cloud as an Enabler for 5G Ecosystem) has been actively conducting to promote 5G ecosystem [3], [4]. The motivation behind these efforts is that mmWave is capable of contributing to high-capacity links, thereby leading to low-latency services by task offloading. Based on this mmWave-MEC scheme, the benefits of merging MEC with mmWave communications are further investigated in [5]–[7]. While the unrivaled data rates of mmWave cannot be overemphasized, the channel intermittency inevitably cause more offloading latency and additional power consumption since the mmWave links are highly susceptible to blocking events. To pave the way for mmWave-MEC scheme, Barbarossa et al. [3] proposed the solutions of overbooking of computation and communication resources and adopting multiple radio access points. By the deployment of multiple radio access points, a further study was conducted by Pietro et al. in [5] to optimize uplink transmit power for mobile users under required latency constraints. However, these solutions call for an a priori knowledge of blocking probabilities and real-time prediction. Moreover, the over-provisioning of resources and radio equipments has also presented the need for unpalatable costs.

In 5G and beyond wireless networks, the demand for green and sustainable design turns out to be increasingly critical. Fortunately, intelligent reflecting surface (IRS) emerged to this end, is
a promising and potential candidate for reducing network energy consumption while effectively improving the spectral efficiency [8], [9]. Intrinsically differently from the active massive MIMO antenna or large intelligent surface (LIS), IRS also termed as reconfigurable intelligent surface (RIS), is composed of nearly passive reflecting elements, which has its origin in reflectarrays and software-defined metamaterials [10], [11]. Furthermore, conventional reflectarrays widely applied in radar and satellite communications, present fixed phase shifts once fabricated. On the contrary, IRS can adaptively change the signal propagation by judiciously adjusting phase shifters according to dynamic wireless environment. To be specific, each IRS element reflects the incident signals via a smart controller, thus achieving the beamforming effects and enhance the holistic communication link performance corresponding to the desired receiver. These features enable the configurability and real-time control of phase shifters, which should be attributed to the development and breakthrough of radio frequency micro-electromechanical systems (RF-MEMS) and metamaterial field [12]. Strictly speaking, resembling the half-duplex amplify-and-forward (AF) relay assisted communications, IRS can avoid the drawbacks of additional power consumption and noise introduced in AF relay. Besides, IRS enjoys the advantages of full-duplex mode without causing self-interference. In practice, IRS can be intensively deployed in a large-scale and low energy consumption manner, since the low-cost passive elements (e.g., printed dipoles) removes requirements for RF chains.

A series of initial efforts to enhance the performance of wireless networks with diverse optimization goals has been extensively studied [8], [12]–[19]. The terminologies frequently used in most literature are active beamforming and passive beamforming, where passive beamforming refers to the reflect beamforming by the phase shifters of IRS and passive beamforming refers to the precoding operation at the base station (BS). From the perspective of energy efficiency (EE), the authors of [8] investigated the joint BS transmit power allocation and passive beamforming for downlink multi-user multi-input single-output (MISO) system, where an optimal trade-off between the EE and number of reflecting elements was observed. Later, recent research works have explored the joint active and passive beamforming design in downlink MISO systems. In [12], the BS transmit power control schemes were proposed while providing some ideas of deploying IRS. Similar to [12], the scheme of multi-IRS was verified to effectively reduce the BS transmit power in simultaneous wireless information and power transfer (SWIPT) systems [13]. In [14], the signal-to-interference-plus-noise ratio (SINR) maximization problem was considered to improve robustness of mmWave systems. In [15], the authors discussed the impact
of phase shifts on the maximum ergodic capacity. By considering a practical IRS phase shift model, Abeywickrama et al. [20] have explored a balance between the amplitude and phase alignment. More interestingly, an online phase-shift configuration of IRS based on deep learning was introduced by [21]. In addition, a distributed beamforming algorithm for achievable rate maximization were proposed in [16]. Later, IRS beamforming design problems are extended to physical-layer security in [17], [18]. Very recently, Bai et al. [19] have studied the applications of IRS in MEC systems, and notably this project mainly focuses on computing resource allocation to minimize the task latency. Till now, the exploitation potential of IRS in improving MEC resource allocation is still in a nascent research stage.

Motivated by the 5G-MiEdge vision, in this paper, we investigate a joint problem of mobile power control and beamforming in mmWave-MEC scenarios. Be in stark contrast with [3]–[5], supplementary links between radio access point (RAP) and IRS are adopted to against the blockage events of mmWave channel while minimizing the sum mobile power. To our best knowledge, merging mmWave-MEC and reflective radio techniques of IRS has not been addressed yet. Through this paper, we propose a uplink mobile power minimization method in the single-input multiple-output (SIMO) system, wherein mmWave links are utilized to support the high-speed upload of computational tasks. This study differs from the downlink MISO system in previous literature since cochannel interference [22] is caused when multi-user transmissions are active, which expands our recent contribution to the beamforming design of IRS-assisted mmWave systems [23]. It is noteworthy that we mainly dedicate to the radio resource allocation in the context of MEC with given maximum predefined latency requirement, and the optimization of offloading decisions is not the focus of current manuscript. Specifically, the problem to be solved is the joint optimization of mobile power sum, multi-user detectors and IRS phase shifts, and the latency requirement constraints is involved in computation offloading, which is distinguished from the model of identical-power mobile devices in [19]. Our main contributions of this paper are listed as follows:

- First, by exploiting the properties of metasurface communications, we introduce supplementary links between radio access point (RAP) and IRS to overcome the blocking events in mmWave-MEC systems, instead of expensive solutions such as multi-RAP or over-provisioning of radio resources. With the aid of cheap IRSs, we establish a Uplink SIMO model for our proposed IRS-enhanced mmWave-MEC sceraio. Then, we investigate the mobile power minimization problem under stringent offloading latency constraints, where
cochannel interference caused by multi-user MEC case can be disrupted by reflection beamforming of IRS.

- Next, we propose an alternating optimization method to solve the formulated problem. In view of non-convex constraints, we decouple the optimization variables, thereby enabling the resulting subproblems tractable to be optimized. Based on this iterative optimization manner, we provide closed-form expressions for the mobile power and multi-user detection matrix, respectively. Besides, the algorithm for these two subproblems belongs to a distributed implementation approach. At MEC RAP, the multi-user detection vector can be updated in a decentralized solution for given mobile powers. For the same reason, the mobile powers are also in a decentralized solution. The convergence of the proposed iterative power update algorithm is also provided.

- Finally, we develop two efficient algorithms to solve the reflection beamforming problem. Considering semidefinite relaxation (SDR) technique generally requires an eigenvalue decomposition per iteration, the high complexity prevents its application in large-scale IRS case. Thus, a novel reflection beamforming algorithm is developed from Riemannian manifold optimization, the original problem with unit modulus constraints of all phase-shifts can be reframed as a constraint-free optimization on a manifold space. To further extend to the large-scale IRS system, we propose a fraction transform based alternating direction method of multipliers (ADMM) to optimize phase shifts in parallel.

The remainder of this paper is organized as follows. Section II outlines the system model and formulates the mobile power minimization problem. Section III provides an alternating optimization framework to solve the formulated problem for multi-user mmwave-MEC system. Section IV develops two efficient algorithms to optimize the phase shifts of IRS. Section V provides the simulation results, aiming to demonstrate the performance gains achieved by the proposed schemes of IRS-enhanced mmWave-MEC. The conclusion is summarized in Section VI.

Notations: The lower-case and upper-case boldface letters denote vectors and matrices, respectively; \{\cdot\}^*, \{\cdot\}^T, and \{\cdot\}^H represent the conjugate, the transpose, and the conjugate transpose; \text{tr}\{\cdot\} and \text{diag}\{\cdot\} return the trace and diagonalization; \lfloor \cdot \rfloor_{i,j} represents the \((i,j)\)-th entry of a matrix; \(j = \sqrt{-1}\), \text{Re}\{\cdot\}, \text{Im}\{\cdot\} and \arg\{\cdot\} denote the imaginary unit, the real part, the real imaginary and the phase of a complex number; \otimes and \odot denote the Kronecker and Hadamard products, respectively; \mathbb{C} and \mathbb{R} denote the space of complex and real numbers, respectively;
\(\mathbb{E}\{\cdot\}\) indicates the expectation operator.

### II. System Model

#### A. IRS Assisted mmWave-MEC Offloading

Let us start with an IRS-assisted uplink mmWave-MEC system, as shown in Fig. 1. The system includes a uplink SIMO wireless communication scenario where \(K\) single-antenna users communicates with the MEC RAP equipped with an \(M\)-element uniform linear array (ULA). In this mmWave-MEC system, the execution of computationally heavy applications can be transferred from a mobile user device to an edge server, which is termed as Mobile Edge Host (MEH). The IRS consists of \(N_{az}\) elements in horizon and \(N_{el}\) elements in vertical. Here, let \(N = N_{az} \times N_{el}\) denotes the total number of reflecting elements. Note that IRS is managed by a smart controller [12], [17], [24], which exchanges CSI and coordinates the reflecting modes.

Denoting 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 MEC AP for the \(k\)-th user can be expressed as

\[
y_k = f_k^H \left( \mathbf{h}_{d,k} + \mathbf{G} \Theta \mathbf{h}_{r,k} \right) x_k + \sum_{j \neq k}^{K} \left( \mathbf{h}_{d,j} + \mathbf{G} \Theta \mathbf{h}_{r,j} \right) x_j + \mathbf{u}_k,
\]  

(1)

where \(f_k \in \mathbb{C}^{M \times 1}\) refers to the multi-user detection vector of the \(k\)-th user and \(F = [f_1, f_2, \cdots, f_K]\); \(\mathbf{h}_{d,k} \in \mathbb{C}^{M \times 1}\) indicates the channel between AP and the \(k\)-th user; \(\mathbf{G} \in \mathbb{C}^{M \times N}\) is the mmWave channel matrix between AP and IRS, and \(\mathbf{h}_{r,k} \in \mathbb{C}^{N \times 1}\) represents 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 $\eta$ indicates the reflection coefficient\(^1\) and $\theta_n = e^{j\varphi_n}$ with $\varphi_n$ being the reflection phase shift; $u_k \in \mathbb{C}^{M \times 1}$ is the noise vector which follows the circularly symmetric complex Gaussian (CSCG) distribution of $\mathcal{CN}(0, \sigma_u^2 \mathbb{I})$.

### B. Wireless Channel Model

In this paper, according to the widely used 3D Saleh-Valenzuela channel model [25], [26] for mmWave communications, the channel between MEC AP and the $k$-th user can be characterized as

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

where $L$ denotes the number of non-line-of-sight (NLoS) paths\(^2\); $\ell = 0$ represents the line-of-sight (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); $\vartheta_U$ and $\vartheta_B$ 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 is densely distributed in the hotspot spaces, which gives rise to a high probability of LoS propagation. Due to severe path loss, the transmit power of 2 or more reflections can be ignored so that only LoS is considered [12], [29]. As such, we make the simplifying assumption that the channel for each user is a LoS case, as given by

$$h_{r,k} = \sqrt{N} \xi_k \vartheta_B \vartheta_U a_N(\phi_{r,k}),$$

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

In reality, the IRS can be properly deployed in the hotspot area to realize the strong LoS dominant channel condition, so that better coverage are provided. Accordingly, it is a reasonable assumption that the channel between the AP and IRS can be modeled as a rank-one matrix,

\(^1\)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.

\(^2\)A series of works 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 [27], [28], while the scattering at sub-6 GHz is generally rich.
which can be mathematically expressed as

\[
G = \sqrt{MN} \xi \vartheta \vartheta a_M(\phi) a_N(\vartheta_{az}, \vartheta_{el}),
\]

(4)

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

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

(5)

\[
a_N(\vartheta_{az}, \vartheta_{el}) = a_{N_{az}}(\vartheta_{az}) \otimes a_{N_{el}}(\vartheta_{el}),
\]

(6)

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

C. Problem Statement

As shown in Fig. 1, the AP is connected to an integrated MEC server (i.e., MEH) through high-capacity links, which provides users with low-latency computation services. In this mmWave-MEC scenario, we assume the limited computational capabilities for each mobile user device. All mobile users can offload their computationally intensive and latency-critical tasks to the MEH.

In our computation offloading case, the computation task of the \( k \)-th user can be characterized by a parameter pair \( \{D_k, T_k\} \), where \( D_k \) and \( T_k \) denote the size of the input data in number of nats and maximum latency, respectively. To reduce system complexity, we assume that \( D_k = D \) \((k = 1, 2, \ldots, K)\) and \( T_1 \leq T_2 \leq \cdots \leq T_K \). In addition, we presume that the total number of CPU cycles necessary to complete each user’s task is identical and denoted by \( w \). Therefore, the latency of the \( k \)-th device can be computed as

\[
\frac{D}{W \log (1 + \Gamma_k)} + \frac{w}{f_{MEH}} + \Delta_R,
\]

(7)

\footnote{The array element spacing of both ULA and IRS is assumed to be \( \lambda/2 \), and IRS is implemented with discrete antenna elements [9], [24], just like uniform rectangular array (URA).}

\footnote{The following conditions of computation offloading for each user device are always satisfied in mmWave-MEC scenario: i) the computational capability required to run specific applications exceeds what the user device alone is capable of employing; ii) when the latency constraints are considered, the energy required to run an application locally is greater than the energy required to offload it; iii) the local execution latency is much larger than the latency resulting from task offloading.}
where $\Gamma_k$ is the SINR of the $k$-th user, $f_{\text{MEH}}$ represents the computational rate of MEH measured in terms of CPU cycles per second and $\Delta R$ is the time necessary to send the result back to the mobile user. In this paper, the latency and energy cost for MEH to return the computation task results to users can be ignored since the size of task results is usually minuscule [30] and the battery capacity is not constrained on the MEH platform. Hence, we remove the latter two terms in (7) to reformulate the latency.

Let us concentrate on the total mobile power consumption minimization problem in the mmWave-MEC offloading system. To be specific, the transmitted power $p_k$, multi-user detection matrix $F$ and passive beamforming matrix $\Theta$ should be designed to meet the quality of service (QoS) requirements for each user. Accordingly, the problem is formulated as

$$\begin{align*}
\text{(P1)} : \quad & \text{minimize} \quad \sum_{k=1}^{K} p_k \\
\text{s.t.} \quad & p_k \geq 0, \quad \forall k, \quad (8a) \\
& \theta_n \in F, \quad \forall n, \quad (8b) \\
& \frac{D}{W \log (1 + \Gamma_k)} \leq T_k, \quad \forall k, \quad (8c)
\end{align*}$$

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

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

(9)

In SINR constraint (8c), the uplink transmission rate of the $k$-th user is computed by $W \log (1 + \Gamma_k)$, where $W$ is the channel bandwidth, and the SINR of the $k$-th user is computed by

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

(10)

We adopt the continuous phase shifts of IRS in this paper [31], while the set of discrete phase shifts [9] at each element is given by $\mathcal{F}_d = \left\{ \theta_n = e^{j \varphi_n} | \varphi_n \in \left\{ \frac{2\pi i}{2^B} \right\}_{i=0}^{2^B - 1} \right\}$, where $B$ denotes the phase resolution in number of bits. Note that we can obtain the discrete phase shifts by the quantized phase projection method [16].
III. MULTI-USER mmWAVE-MEC SYSTEM

A. Power Control And Multi-User Detection

In this paper, we consider a total mobile transmit power minimization problem by jointly optimizing the allocated mobile power vector $p$, multi-user detection matrix $F$ and phase shift matrix $\Theta$ while meeting the required latency targets. However, this optimization problem is untractable due to its non-convexity introduced by (8b) and coupling variables. To tackle these challenges, we propose an iterative optimization algorithm based on the above analysis. The key idea is to alternately optimize each of variables by fixing other variables, thus achieving variables decoupling.

First, we focus our attention on the mobile power control with fixed $F$ and $\Theta$. For notational brevity, by defining $h_k = h_{d,k} + G\Theta h_{r,k}$, we have

$$\Gamma_k(p) = \frac{p_k |f_k^H h_k|^2}{\sum_{j=1, j \neq k}^K p_j |f_k^H h_j|^2 + \sigma^2_u f_k^H f_k}.$$  \hspace{1cm} (11)

Plugging (11) into (8c), the latency constraint can be reformulated as

$$-p_k |f_k^H h_k|^2 + \tilde{T}_k \left( \sum_{j=1, j \neq k}^K p_j |f_k^H h_j|^2 + \sigma^2_u f_k^H f_k \right) \leq 0, \forall k,$$  \hspace{1cm} (12)

where $\tilde{T}_k = e^{\frac{\beta_p}{\gamma_k}} - 1$, $k = 1, 2, \ldots, K$. This constraint can be presented in a matrix form as

$$\begin{pmatrix} I - Q \end{pmatrix} \begin{pmatrix} p \end{pmatrix} \succeq \tau,$$  \hspace{1cm} (13)

where

$$[Q]_{i,j} = \begin{cases} 0, & \text{if } j = i, \\ \frac{\tilde{T}_i |f_i^H h_i|^2}{|f_i^H h_i|^2}, & \text{otherwise}, \end{cases}$$  \hspace{1cm} (14)

$$\tau = \begin{bmatrix} \sigma^2_u \tilde{T}_1 f_1^H f_1 & \sigma^2_u \tilde{T}_2 f_2^H f_2 & \cdots & \sigma^2_u \tilde{T}_K f_K^H f_K \\ \frac{|f_1^H h_1|^2}{f_1^H h_1^2} & \frac{|f_2^H h_2|^2}{f_2^H h_2^2} & \cdots & \frac{|f_K^H h_K|^2}{f_K^H h_K^2} \end{bmatrix}^T.$$  \hspace{1cm} (15)
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, \quad (16a)\]
\[p \succeq 0. \quad (16b)\]

It should be pointed out that \(\tau\) in inequality constraint (16a) implies the vector of \textit{minimum protection ratio} for all users. When the fixed \(F\) and \(\Theta\) is given, we can minimize the objective function in (P2) by identifying the critical points of the inequality (16a). One reasonable assumption is that we can easily find a feasible solution \(F\) to admit the spectral radius of \(Q\) less than 1, and hence the matrix \(I - Q\) is invertible [32]. Then we can update \(p\) by

\[p = (I - Q)^{-1} \tau, \quad (17)\]

which yields a specific iterative update process:

\[p_k^{(t+1)} = \sum_{j=1, j \neq k}^{K} \frac{T_k |f_k^H h_j|^2}{|f_k^H h_k|^2} p_j^{(t)} + \frac{\sigma_u^2 T_k f_k^H f_k}{|f_k^H h_k|^2}, \quad k = 1, 2, \ldots, K. \quad (18)\]

In the following, we present the convergence analysis for the minimization mapping of allocated mobile power. Denoting the above iteration procedure in (18) as \(p^{(t+1)} = \mathcal{M}(p^{(t)})\), we define the \(k\)-th element of the mapping \(\mathcal{M}\) as

\[\mathcal{M}_k(p) = \min f_k \left\{ \sum_{j=1, j \neq k}^{K} \frac{T_k |f_k^H h_j|^2}{|f_k^H h_k|^2} p_j + \frac{\sigma_u^2 T_k f_k^H f_k}{|f_k^H h_k|^2} \right\}, \quad k = 1, 2, \ldots, K. \quad (19)\]

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

**Proof:** See Appendix A.

Next, we concentrate on the optimal solution of \(F\) when we fix \(p\) and \(\Theta\) for power minimization problem. For clarity of problem formulation, rearranging (19) leads to

\[f_k^{(t+1)} = \arg \min f_k \left\{ \frac{T_k \left( \sum_{j=1, j \neq k}^{K} |f_k^H h_j|^2 p_j + \sigma_u^2 f_k^H f_k \right)}{|f_k^H h_k|^2} \right\}, \quad k = 1, 2, \ldots, K, \quad (20)\]

which is, in essence, in the form of Rayleigh quotient minimization [33]. Thus, the optimal
problem of $f_k$ can be reformulated as

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

s.t. $f_k^H h_k = 1, \quad k = 1, 2, \ldots, K.

(21a)

By converting to the generalized eigenvalue problem, we adopt the minimum variance distortionless response (MVDR) beamforming [34] scheme to solve (P3). Therefore, the optimal $f_k$ can be computed by

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

(22)

Notice that the optimal multi-user detection vectors are given a unique solution in a closed form, which can always keep the total power non-increasing by updating $f_k$.

**B. Problem Transformation of Passive Beamforming**

Lastly, for given $F$ and $p$, problem (P1) is reduced to a reflecting coefficient optimization problem. Denoting the reflecting coefficients by $\theta = [\theta_1, \ldots, \theta_N]^T$, we get the following change of variables:

$$
G \Theta_{h_{r,j}} = G \cdot \text{diag} (h_{r,j}) \cdot \theta = G_{h_{r,j}} \theta.
$$

(23)

For notational simplicity, let us proceed to define the variables independent of $\theta$:

$$
b_{k,j} = f_k^H h_{d,j},
$$

(24)

$$
g_{k,j}^H = f_k^H G_{h_{r,j}}.
$$

(25)

Using the new variables as defined above, the expression in (10) becomes

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

(26)

$$
= \frac{p_k \left| b_{k,k} + g_{k,k}^H \theta \right|^2}{\sum_{j=1, j \neq k}^{K} p_j \left| b_{k,j} + g_{k,j}^H \theta \right|^2 + \sigma^2_u \|f_k\|^2}.
$$
Therefore, from (18), by applying the re-expression of SINR in (26), the optimization of $\theta$ when given fixed $F$ and $p$ can be equivalently transformed to

$$
\theta^{(t+1)} = \arg \min_{\theta \in F} \sum_{k=1}^{K} \left\{ \sum_{j=1,j \neq k}^{K} |T_k b_{k,j} + g_{k,j}^H \theta|^2 p_j + \frac{\sigma_n^2 T_k}{|b_{k,k} + g_{k,k}^H \theta|^2} \right\}.
$$

(27)

Fundamentally, it is very arduous to solve this optimization problem due to the non-convex unit modulus constraints. We will provide two efficient methods to deal with this problem in the following sections. These three optimization processes corresponding to $p$, $F$ and $\Theta$ are repeated alternately until the stop condition is achieved. To facilitate the understanding of alternating optimization algorithms, we summarise the proposed mobile power minimization method 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: \textbf{repeat}
2: \hspace{1em} Set $t \leftarrow t + 1$;
3: \hspace{1em} 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: \hspace{1em} Set iteration index $l = 0$, $p^{(l)} = p^{(t-1)}$ and $F^{(l)} = F^{(t-1)}$;
5: \hspace{1em} \textbf{repeat}
6: \hspace{2em} With given $F^{(l)}$, $p^{(l)}$ and $h_k^{(t-1)}$, update $p_k^{(l+1)}$ using (18);
7: \hspace{2em} With given $p^{(l+1)}$ and $h_k^{(t-1)}$, update $f_k^{(l+1)}$ using (22);
8: \hspace{1em} Set $l \leftarrow l + 1$;
9: \hspace{1em} \textbf{until} $|1^T p^{(l)} - 1^T p^{(t-1)}|$ converges.
10: \hspace{1em} Obtain $p^{(l)} = p^{(l)}$ and $F^{(l)} = F^{(l)}$;
11: \hspace{1em} With given $p^{(l)}$ and $F^{(l)}$, update $\Theta^{(l)}$ by solving problem (27);
12: \hspace{1em} \textbf{until} $|1^T p^{(l)} - 1^T p^{(l-1)}|$ converges.
13: \hspace{1em} Output optimal $\{p, F, \Phi\}$ and calculate $\sum_{k=1}^{K} P_k$.

IV. REFLECTION BEAMFORMING

A. Manifold Optimization Scheme

To guarantee that all the offloading latency constraints are active at the optimal solution, in this subsection, the passive beamforming problem is initially transformed to a new optimization problem. This solution is expected to reduce the transmitted power while provoking significant gain of phase shifting scheme. To achieve a effective solution for this optimization problem, we provide a background on optimization over manifolds and subsequently develop a new technique to optimize directly over the complex circle manifold.
From (26), the constraint (12) can be reformulated as

\[ p_k |b_{k,k} + g_{k,k}^H \theta|^2 \geq \bar{T}_k \left( \sum_{j = 1, j \neq k}^K p_j |b_{k,j} + g_{k,j}^H \theta|^2 + \sigma_\theta^2 \|f_k\|^2 \right), \quad \forall k. \quad (28) \]

One straightforward idea is to convert constraints (28) and unit modulus constraints into quadratic constraints. Then, by ignoring the rank-one constraint involved in constructing \( \Phi \), problem (27) can be recast into a SDR form, which can be solved by eigen-decomposition.

However, SDR is notoriously difficult to extend to large-scale IRS beamforming problems since the number of variables to be optimized increases quadratically with the number of IRS elements. Besides the alternating SDR iteration method, we develop a gradient based method which can be applied to the constant modulus constraint of complex-valued case. By invoking principles of gradient descent over non-convex manifolds, we embody a drastically different approach, i.e., the Complex Circle Manifold Optimization (CCMO) method to solve (P4). As previously stated, the optimization over the non-convex constant modulus constraint \(|\theta_n| = 1\) remains elusive.

In the following, we briefly introduce the rationale of manifold optimization while providing our solutions. The rationale behind Riemannian manifold optimization breaks the confine 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 will be minimized by gradient descent algorithms.

Referring to the latency constraint (28), problem (P4) can be readily reframed as

\[
\text{(P5)} : \quad \max_{\theta} \quad f_0(\theta) = \theta^H U \theta + 2 \text{Re} (\theta^H v) + C \\
\text{s.t.} \quad |\theta_n| = 1, \quad n = 1, 2, \ldots, N. \quad (29a)
\]

where

\[
U = \sum_{k=1}^{K} \left( p_k g_{k,k} g_{k,k}^H - \bar{T}_k \sum_{j \neq k}^K p_j g_{k,j} g_{k,j}^H \right), \quad (30)
\]

\[
v = \sum_{k=1}^{K} \left( p_k b_{k,k} g_{k,k} - \bar{T}_k \sum_{j \neq k}^K p_j b_{k,j} g_{k,j} \right), \quad (31)
\]

\[
C = \sum_{k=1}^{K} \left[ p_k |b_{k,k}|^2 - \bar{T}_k \left( \sum_{j \neq k}^K p_j |b_{k,j}|^2 + \sigma_\theta^2 \|f_k\|^2 \right) \right]. \quad (32)
\]
According to the notion of manifold optimization, the problem (P5) can be reformulated as:

\[(P6) : \text{minimize}_{\theta \in \mathcal{S}^N} \ f(\theta) = -\theta^H U \theta - 2 \text{Re}\left(\theta^H v\right),\]  

(33)

where \(\mathcal{S}^N\) indicates the manifold space defined in the constant modulus constraints for the problem in (P5). Here, \(\mathcal{S}^N\) can be expressed as

\[\mathcal{S}^N = \{\theta \in \mathbb{C}^N : |\theta_1| = |\theta_2| = \cdots = |\theta_N| = 1\},\]  

(34)

in which \(\mathcal{S} = \{\theta_n \in \mathbb{C} : \theta_n \theta_n^* = \text{Re}\{\theta_n\}^2 + \text{Im}\{\theta_n\}^2 = 1\}\) is known as a complex circle and can be viewed as a sub-manifold of \(\mathbb{C}\). Naturally, the search space \(\mathcal{S}^N\) 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.

Since the main idea of CCMO algorithm is to derive a gradient descent algorithm based on the complex circle manifold space, we need to find the counterpart on Riemannian manifold by emulating the gradient based unconstrained optimization algorithms in Euclidean space. The general framework of gradient descent algorithm in Euclidean space mainly consists of two phases: the first phase, which determines the descent direction of the current solution by computing the Euclidean gradient, and the second phase, which decreases the value of objective function via the line search method [35]. These two main phases are iteratively performed until its stopping criterion is met. Similarly, Riemannian manifold optimization also requires the above iterative form. Nevertheless, instead of standard (Euclidean) gradient, Riemannian manifold
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

$$T_{\theta^{(i)}} S^N = \{ \eta \in \mathbb{C}^{N+1} : \text{Re}\{\eta^* \odot \theta^{(i)}\} = 0 \}. \quad (35)$$

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

$$\nabla f(\theta^{(i)}) = -2U\theta^{(i)} - 2v. \quad (36)$$

By performing the projection operator on the Euclidean gradient, the Riemannian gradient of $f(\theta^{(i)})$ is obtained as

$$\nabla_{S^N} f(\theta^{(i)}) = \text{Proj}_{T_{\theta^{(i)}} S^N} \left( \nabla f(\theta^{(i)}) \right) = \nabla f(\theta^{(i)}) - \text{Re}\{\nabla f(\theta^{(i)})^* \odot \theta^{(i)}\} \odot \theta^{(i)}. \quad (37)$$

Hence, the current point $\theta^{(i)}$ in the tangent space $T_{\theta^{(i)}} S^N$ is updated as

$$\theta^{(i)^o} = \theta^{(i)} - \zeta \nabla_{S^N} f(\theta^{(i)}), \quad (38)$$

where $\zeta > 0$ is a well-chosen constant step size$^6$. It should be noted that $\theta^{(i)^o}$ is still in the tangent space $T_{\theta^{(i)}} S^{N+1}$ but it leaves manifold $S^N$. Therefore, a Retraction mapping operator is applied to move the point $\theta^{(i)^o}$ back to the manifold $S^N$. Finally, the point $\theta^{(i+1)}$ updated by using the Retraction mapping operator is given by

$$\theta^{(i+1)} = \text{Ret}_{\theta^{(i)}} \left( -\zeta \nabla_{S^N} f(\theta^{(i)}) \right) = \frac{\theta^{(i)} - \zeta \nabla_{S^N} f(\theta^{(i)})}{|\theta^{(i)} - \zeta \nabla_{S^N} f(\theta^{(i)})|} = \theta^{(i)^o} \odot \frac{1}{|\theta^{(i)^o}|}. \quad (39)$$

Notably, the point updated in each iteration satisfies the unit constant modulus constraints. The details are summarised in Algorithm 2.

**B. SIMin Fraction Transform Based ADMM**

In realistic environment, IRS tends to consist of massive reflecting elements. Although the manifold optimization method performs more efficiently and it is suitable for constant modulus

$^6$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 (P6). This optimization problem can be computed by leveraging the Manopt toolbox in MATLAB [36], [37].
Algorithm 2 The proposed CCMO algorithm for passive beamforming.

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

1: repeat
2: \( i \leftarrow i + 1 \);
3: Calculate the Euclidean gradient \( \nabla f(\theta^{(i)}) \) at \( \theta^{(i)} \) using (36);
4: Construct the tangent space \( T_{\theta^{(i)}}S^N \) and calculate the current Riemannian gradient \( \nabla_{S^N} f(\theta^{(i)}) \) using (37);
5: Perform gradient descent algorithm over the current tangent space using (38);
6: Update \( \theta^{(i+1)} \) using the Retraction mapping operator according to (39);
7: until The criteria of \( |f(\theta^{(i)}) - f(\theta^{(i-1)})| < \epsilon_2 \) in (P6) converges.
8: Output \( \theta \) and obtain \( \Theta = \text{diag}\{\theta\} \).

constraint problem, there is a strong need for algorithms of parallel computation that can be extended to large-scale IRS case. Therefore, in this paper, the original passive beamforming problem is first transformed into a fractional minimization problem, and then we propose a projection-based ADMM algorithm to solve it.

For fixed \( p \) and \( F \), the optimization problem of \( \Theta \) is reframed as

\[
(P7) : \begin{aligned}
\text{minimize} & \quad \sum_{k=1}^{K} \frac{T_k \left( \sum_{j=1,j\neq k}^{K} P_j b_{k,j} + g_{k,k}^H \theta \right)^2 + \sigma_k^2 \| f_k \|^2}{\| b_{k,k} + g_{k,k}^H \theta \|^2} \\
\text{s.t.} & \quad (8b).
\end{aligned}
\]

which is a sum-of-inverse fractional minimization problem [38]. Typically, this fractional programming problem is addressed by decoupling its numerators and denominators, thus enables the separate optimization of the numerator part and denominator part. To facilitate this mathematically intractable issue, we present a novel fraction transform technique which is termed as sum-of-inverse minimization (SIMin) fraction transform, as given in the following theorem.

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

\[
\text{minimize}_{x \in \mathcal{X}} \quad \sum_{k=1}^{K} \frac{A_k(x)}{B_k(x)}, \tag{41}
\]

which is equivalent to

\[
\text{minimize}_{x \in \mathcal{X}} \quad \sum_{k=1}^{K} y_k A_k(x)^2 + \sum_{k=1}^{K} \frac{1}{4y_k} B_k(x)^2. \tag{42}
\]

**Proof:** See Appendix B.
By introducing a novel SIMin fraction transform technique, (P4) can be equivalently cast as

$$
\text{minimize } J(\theta) = \sum_{k=1}^{K} J_{A,k}(\theta) + \sum_{k=1}^{K} J_{B,k}(\theta),
$$

where

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

$$
J_{B,k}(\theta) = \frac{1}{\beta_k} \left( \frac{1}{b_{k,k} + g_{k,k}^H \theta} \right)^{1/2},
$$

where $\beta = [\beta_1, \beta_2, \cdots, \beta_K]^T$ is the auxiliary vector during the fraction transform process. Then, according to Theorem 1, the optimal $\beta_k$ is computed by

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

Next, in order to decouple the terms of $J_{A,k}(\theta)$ and the terms of $J_{B,k}(\theta)$ to optimize $J_{B,k}(\theta)$ in a distributed manner, we employ the ADMM method as stated in the following. The augmented Lagrangian of (43) can be expressed as

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

where

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

Thus, the scaled form of ADMM can be readily obtained as

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

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

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

It is evident that the update rule in (51) is quite straightforward. Accordingly, we just need to investigate the detailed methodologies of solving (49) and (50) in the following parts.
Note that the constraint (8b) is non-convex. Herein, we start with a convex case so that the optimization problem can be extended to the non-convex constraint subsequently. At first, we can relax the constraint (8b) as
\[
\theta^H e_n e_n^H \theta \leq 1, \quad \forall n = 1, 2, \cdots, N.
\] (52)
Replacing (8b) by (52), the Lagrangian form of (49) can be written as
\[
\mathcal{G}_1(\theta, \varepsilon) = \sum_{k=1}^{K} J_{A,k}(\theta) + \frac{\rho}{2} \| \theta - q + r \|^2 + \sum_{n=1}^{N} \varepsilon_n (\theta^H e_n e_n^H \theta - 1),
\] (53)
where \( \varepsilon = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_N]^T \) is the associated dual variable vector for (52). The function \( \mathcal{G}_1(\theta, \varepsilon) \), obviously, is a convex optimization problem. So, we can attain the optimal \( \theta \) by the CVX solver in each iteration. Then, we can perform the projection operation to update \( \theta \) and its update process is given by
\[
\theta^o = \arg\min_{\theta} \mathcal{G}_1(\theta, \varepsilon) = [\theta^o_1, \theta^o_2, \cdots, \theta^o_N]^T,
\] (54)
\[
\theta^{(l+1)} = [e^{-j\arg\{\theta^o_1\}}, e^{-j\arg\{\theta^o_2\}}, \cdots, e^{-j\arg\{\theta^o_N\}}]^T.
\] (55)
Particularly, thanks to the decouple effect of ADMM method, all constraints of the problem in (50) are on \( q \) instead of \( \theta \). In the concurrent update process, the optimization of \( q \) is modeled as an unconstrained minimization problem:
\[
\min_{q} \mathcal{G}_2(q) = \sum_{i=1}^{K} J_{B,k}(q) + \frac{\rho}{2} \| \theta - q + r \|^2.
\] (56)
According to Newton’s method, we can iteratively update \( q \) by
\[
q^{(i+1)} = q^{(i)} - \lambda \left[ \nabla^2 \mathcal{G}_2(q^{(i)}) \right]^{-1} \nabla \mathcal{G}_2(q^{(i)}),
\] (57)
where \( \lambda \) is a step size factor introduced in 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 3.
Algorithm 3 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: \(t \leftarrow t + 1\);
3: With given \(\theta(t-1)\), update \(\beta_k^{(t)}\) using (46);
4: Set iteration index \(l = 0\);
5: repeat
6: \(l \leftarrow l + 1\);
7: Update \(\theta(l)\) according to (54) and (55);
8: Update \(q(l)\) according to the Quasi-Newton method;
9: Perform \(r(l) = r(l-1) + \theta(l) - q(l)\);
10: until The value of \(\|\theta(l) - q(l)\|\) converges.
11: Obtain \(\theta(t) = \theta(l)\);
12: until The function in (P7) converges.
13: Output \(\theta\) and set \(\Theta = \text{diag}\{\theta\}\).

V. NUMERICAL AND SIMULATION RESULTS

A. Simulation Setup

In this section, numerical simulations are carried out to verify the effectiveness of above proposed transmit power control method and potential benefits of deploying IRS in MEC systems. We consider an IRS-aided mmWave-MEC system illustrated in Fig. 3. More specifically, the MEC AP is equipped with an ULA consisting of \(M = 32\) antennas, and it is located in 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. And 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 [14], [26], the channel gain \(\xi\) follows a complex Gaussian distribution:

\[
\xi \sim \mathcal{CN}\left(0, 10^{-\text{PL}(R)}\right),
\]

where \(\text{PL}(R)\) is 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 \(\kappa \sim \mathcal{N}(0, \sigma_k^2)\) is the lognormal shadowing variance. Note that for mmWave communica-
tions 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 computation offloading, the mmWave channel between MEC 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 [14], [39], [40], 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. Besides, the relative reflection gain of IRS is defined as $\nu = \frac{\varrho_b}{\sqrt{\varrho_U \varrho_B}}$. Fortunately, with the tremendous advancements in meta-materials, the reflection gain can compensate channel attenuation completely, though the cascaded channel of IRS-user and AP-IRS becomes severely weakened by double-fading effect [16], [41] and high path loss of mmWave.

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 offloading latency requirement for each user follows the uniform distribution.
Fig. 4. Transmit power versus number of IRS elements, \(N\). (a) LoS scenario; (b) OLoS scenario.

\[T_k \sim \mathcal{U}(400 \text{ ms}, 600 \text{ ms}).\]

### B. Single-User Scenario

In this subsection, we study the special case of single-user mmWave-MEC\(^7\). 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 and passive beamforming 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-MEC 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 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

\(^7\)For the case of single-user transmission, any interference terms are eliminated from the above analysis.
Fig. 5. Transmit power versus AP-user horizontal distance, $d_{x_1}$. (a) LoS scenario; (b) OLoS scenario.

give about the same yield. However, Fig. 4(b) reveals that the supplementary link is dominant in OLoS 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 $d_{x_1}$ 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 are 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 enable 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 offloading data to AP intrinsically.
Then, we compare the optimal power performance achieved by all schemes versus the task data size of to be offloaded. 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 increases 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 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-MEC 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 mmWave-MEC designs is capable of saving transmission power consumption while meeting predefined offloading 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-MEC system, where User 1 and User 2 simultaneously offload 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 [42]. 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 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 for weak direct AP-user channel in multi-user scenarios.

VI. CONCLUSION

In this paper, we have proposed a novel IRS-assisted mmWave-MEC scheme, where IRS can resolve link blockage problems in an economical manner to guarantee the real-time offloading of computing tasks. Our objective is to minimize the uplink mobile power for all users while satisfying the offloading latency constraints, 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 several tractable subproblems. Then, the closed-form solutions of 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 lending themselves to easy extension to the passive beamforming problem of large-scale IRS. In the end, our simulation results have validated the benefits brought by IRS in the mmWave-MEC system in the presence of channel intermittency, and demonstrated that our proposed algorithms are superior to the baseline schemes. Furthermore, we have provided some interesting insights on the role of IRS in mobile power allocation for different scenarios.

APPENDIX A

PROOF OF LEMMA 1

Assume two sets of positive power vectors \( \hat{p} \) and \( p^* \) 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 \( \max_j (\hat{p}_j / p^*_j) = \gamma > 1 \). Thus, we have the fact \( \gamma p^* \geq \hat{p} \). 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

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

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

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

\[
= \gamma p^*_i. \tag{60}
\]

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:

\[ \frac{y_k A_k(x)}{4y_k B_k(x)^2} + \frac{1}{2 \sqrt{y_k B_k(x)}} = \left( \frac{\sqrt{y_k A_k(x)} - \frac{1}{2 \sqrt{y_k B_k(x)}}}{\sqrt{y_k B_k(x)}} \right)^2 + \frac{A_k(x)}{B_k(x)}. \] (61)

It is facile to see that minimizing the right-hand side of (61) with respect to \( y \) and \( x \) must guarantee the minimization of its left-hand side. Meanwhile, we observe that the optimal \( y \) in minimizing the right-hand side of (61) can be always found by

\[ y_k = \frac{1}{2 A_k(x) B_k(x)}, \] (62)

so that the squared term in (61) becomes zero. Therefore, the optimal \( x \) for the left-hand side of (61) is always part of the solutions of minimizing \( \frac{A_k(x)}{B_k(x)} \).

\[ \blacksquare \]

REFERENCES

[1] V. Sciancalepore, M. Di Renzo, and X. Costa-Perez, “STORNS: Stochastic radio access network slicing,” in **ICC 2019 - 2019 IEEE International Conference on Communications (ICC)**, May 2019, pp. 1–7.

[2] H. Liu, Z. Chen, and L. Qian, “The three primary colors of mobile systems,” **IEEE Communications Magazine**, vol. 54, no. 9, pp. 15–21, Sep. 2016.

[3] S. Barbarossa, E. Ceci, and M. Merluzzi, “Overbooking radio and computation resources in mmW-mobile edge computing to reduce vulnerability to channel intermittency,” in **2017 European Conference on Networks and Communications (EuCNC)**, Jun. 2017, pp. 1–5.

[4] S. Barbarossa, E. Ceci, M. Merluzzi, and E. Calvaneze-Strinati, “Enabling effective mobile edge computing using millimeterwave links,” in **2017 IEEE International Conference on Communications Workshops (ICC Workshops)**, May 2017, pp. 367–372.

[5] 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.

[6] 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 of Things Journal**, pp. 1–1, 2019.

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

[8] 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.

[9] 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 Wkshps)**, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[10] 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.
[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

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

[18] 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.

[19] 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.

[20] S. Abeywickrama, R. Zhang, and C. Yuen, “Intelligent reflecting surface: Practical phase shift model and beamforming optimization,” arXiv preprint arXiv:1907.06002, vol. abs/1907.06002, 2019.

[21] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, “Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces,” in IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Cannes, France, Jul. 2019, pp. 1–5.

[22] A. V. Guglielmi, M. Levorato, and L. Badia, “A bayesian game theoretic approach to task offloading in edge and cloud computing,” in 2018 IEEE International Conference on Communications Workshops (ICC Workshops), May 2018, pp. 1–6.

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

[24] 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.

[25] 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.

[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] 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.

[28] 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.

[29] 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.
[30] Z. Ding, J. Xu, O. A. Dobre, and H. V. Poor, “Joint power and time allocation for NOMA-MEC offloading,” IEEE Trans. Veh. Technol., vol. 68, no. 6, pp. 6207–6211, Jun. 2019.

[31] F. Liu, O. Tsilipakos, A. Pitilakis, A. C. Tasolamprou, M. S. Mirmoosa et al., “Intelligent metasurfaces with continuously tunable local surface impedance for multiple reconfigurable functions,” Physical Review Applied, vol. 11, no. 4, p. 044024, 2019.

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

[33] D. Hwang, D. I. Kim, and T. Lee, “Throughput maximization for multiuser MIMO wireless powered communication networks,” IEEE Transactions on Vehicular Technology, vol. 65, no. 7, pp. 5743–5748, Jul. 2016.

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

[35] 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.

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

[37] 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.

[38] 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.

[39] 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 Communications Surveys Tutorials, vol. 20, no. 2, pp. 870–913, Secondquarter 2018.

[40] J. Kim, L. Xian, and A. S. Sadri, “Numerical simulation study for frequency sharing between micro-cellular systems and fixed service systems in millimeter-wave bands,” IEEE Access, vol. 4, pp. 9847–9859, 2016.

[41] J. D. Griffin and G. D. Durgin, “Complete link budgets for backscatter-radio and rfid systems,” IEEE Antennas and Propagation Magazine, vol. 51, no. 2, pp. 11–25, Apr. 2009.

[42] 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.