MISO-BASED MEC SYSTEM WITH IRS: JOINT COMPUTATION OFFLOADING AND BEAMFORMING OPTIMIZATION

Zhiguo Xu; Shucheng Wang; Zhongming Ji; Jun Xu; Dongxu Wang; Shengju Zhang; Lingkun Meng

China Mobile Communications Group Co., Ltd, China

ABSTRACT

Intelligent reflecting surface (IRS) has appeared as an innovative green and economical technique to markedly enhance the spectrum and energy efficiency in mobile edge computing (MEC) systems. However, limited by single antenna setting, the potential of IRS-aided MEC systems has not been fully exerted. To further improve the system computational performance, the paper introduces multiple-input single-output (MISO) technology, and proposes a MISO-based MEC system with IRS. In this system, IRS is utilized to support the communication between users and a multi-antenna access point integrated with edge servers. Also, we adopt non-orthogonal multiple access (NOMA) strategy for information transmission. The computation rate is maximized to optimally design receiver beamforming, CPU frequency, transmit power, and IRS phase shifts. This formulated problem is non-convex and challenging to solve due to multi-variable coupling. To tackle it, we exploit alternating optimization manner to address the four decoupled subproblems until convergence. Simulation results highlight the superior computation rate performance of our proposed IRS-aided MISO-based MEC system.

Keywords - Computation rate, IRS, MISO, MEC, NOMA

1. INTRODUCTION

With the popularity of mobile devices, the demands for high information transmission rate and QoS are increasing [1]. Also, the continuous emergence of computation-intensive and latency-sensitive applications (e.g., image/biometric recognition, and augmented/virtual reality) has brought new challenges to B5G/6G wireless systems [2]. Aiming at coping with these challenges, mobile edge computing (MEC), as an up-and-coming technique, has been proposed to significantly improve the computational capability for mobile devices with limited computing resources [3]. In MEC systems, small-size and power-limited devices offload their intensive tasks to nearby powerful edge servers, and these tasks can be executed remotely and downloaded, thus efficiently reducing latency and expanding computational capacity [4].

Additionally, intelligent reflecting surface (IRS), an innovative green and economical technology that has emerged lately, can markedly extend communication coverage and enhance energy-efficiency [5]. In general, IRS consists of massive electromagnetic units, which can independently change the amplitude and/or phase, thus achieving reconfigurable wireless propagation environment [6]. Compared to traditional active relays which regenerate and retransmit signals, IRS only passively reflect incident signals with low energy consumption. Besides, IRS operates in full duplex (FD) mode with suppressing interference [7]. Because of these advantages, IRS is capable of improving the communication rates, and mitigating latency of MEC systems. Thus, the combination of IRS and MEC can reinforce their respective advantages and further enhance the potential of MEC systems. Also, the new paradigm will be applied in diverse scenarios of B5G/6G wireless communication systems. [8] minimize the computational latency of an IRS assisted MEC system by jointly optimize computing and communications in different scenarios. [9] formulated an energy minimization problem of IRS-aided MEC systems to optimally design the local CPU frequency, receiver beamforming, offloading schedules and IRS phase shifts. [10] jointly optimized time assignment, transmission power, and IRS phase shifts to maximize total computational bits of IRS-aided MEC system, and presented the introduction of IRS can achieve significant advantages.

Meanwhile, the advent of non-orthogonal multiple access (NOMA) technique alleviate the problem of limited spectrum resources in IRS-assisted MEC systems, which can improve spectral efficiency, and support large scale connectivity [11]. Therefore, the involvement of NOMA can complement and synergize with the advantages of MEC, effectively ascending the system transmission rate and computational efficiency. [12] introduced IRS to assist task offloading for two users in MEC system, in which users transmit data in NOMA and time division multiple access (TDMA) strategies, and the IRS phase shifts and offloading scheduling was optimized to minimized user delay. [13] considered an IRS-assisted MEC system where users communicated with access point (AP) through the NOMA protocol to effectively utilize spectrum resources. In particular, [13] jointly optimized IRS phase shifts, decoding order, transmission data, power and time to minimize the sum energy consumption of all users, and demostrated the proposed scheme achieved marked performance gains compared to the IRS-aided MEC system with TDMA protocol. However, the abovementioned works all adopts single antenna technology, which still cannot fully explore the potential of IRS-aided MEC systems. Compared with single antenna technique, multiple-input single-output(MISO) technology can efficiently strengthen communication quality, and increase the capacity and spectral efficiency [14]. Besides, computational rate is also the

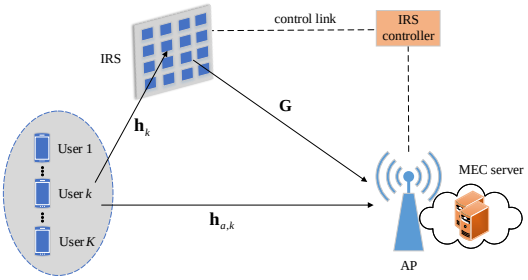

Figure 1 – MISO-based MEC system with IRS.

most direct performance metric to evaluate MEC systems [15, 16]. To our best knowledge, extremely few studies have considered the introduction of MISO technology and optimization analysis of computational rate performance in NOMA-based IRS-aided MEC systems.

In this paper, we consider a IRS-aided MEC system, where AP is equipped with multiple antennas and NOMA strategy is adopt. Also, the receiver beamforming matrix at the AP is optimized. A computation rate maximization problem is established, which optimizes receiver beamforming, CPU frequency, transmission power, as well as IRS phase shifts, subject to constraints on energy, transmit power, signal-to-noise ratio (SNR), IRS phase shifts, and receiver beamforming matrix. To handle this multi-variable coupled non-convex problem, we decouple the problem into four subproblems and adopt alternating optimization (AO) approach. In particular, closed-form optimal solutions are proposed to optimize the receiving beamforming matrix and CPU frequency, respectively. And then successive convex approximation (SCA)-based iterative algorithm and semi-definite relaxation (SDR) algorithm based on Gaussian randomization are exploited to optimize transmit power and IRS phase shifts, respectively. Numerical results demonstrate that our MISO-based MEC system with IRS can achieve superior performance gains.

2. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a MISO-based MEC system with IRS in Fig.1, in which an IRS comprised of N reflecting elements is installed near edge users to assist computation offloading from K users to AP equipped with M antennas. Let $\mathcal{K} = \{1, 2, \cdots, K\}$ and $\mathcal{N} = \{1, 2, \cdots, N\}$ denote the set of all users and reflecting elements of IRS, respectively. Each user partially or completely offloads tasks to AP for edge computing via direct link and reflected link, while the remain portion can be locally computed. Also, we adopt NOMA strategy for information transmission. We assume that channel state information of all involved channels is fully available at AP.

2.1 Communication Model

Let $\mathbf{h}_{a,k} \in \mathbb{C}^{M\times 1}$ denote the direct channel from user k to AP. The cascade channel of IRS divides into tripartite: the channel from user k to IRS, the reflection-coefficient matrix of IRS and from IRS to AP, are respectively denoted by $\mathbf{h}_k \in \mathbb{C}^{N\times 1}$, $\mathbf{\Theta} \in \mathbb{C}^{N\times N}$ and $\mathbf{G} \in \mathbb{C}^{M\times N}$. Wherein, the reflection-coefficient matrix is $\mathbf{\Theta} = \text{diag}\left(\beta_1 e^{j\theta_1}, \cdots, \beta_N e^{j\theta_N}\right)$, in which $\theta_n \in [0, 2\pi]$, $\forall n \in \mathcal{N}$ represent the phase shift, and the amplitude reflection coefficient is fixed as $\beta_n = 1, \forall n \in \mathcal{N}$. Thus, the effective channel between user k and AP can be represented as

$$\mathbf{h}_k(\eta) = \mathbf{G}\mathbf{\Theta}\mathbf{h}_k + \mathbf{h}_{a,k}, \tag{1}$$

in which η represents nonzero elements of matrix $\mathbf{\Theta}$. Accordingly, the received signal of AP is given by

$$\mathbf{y}_{AP} = \sum_{k=1}^{K} \mathbf{h}_k(\eta)\sqrt{p_k}s_k + \mathbf{n}, \tag{2}$$

in which $\mathbf{n} \sim \mathcal{CN}(0, N_0\mathbf{I}_M)$ denotes zero-mean additive white Gaussian noise at AP. p_k denotes transmission power of user k, subjecting to $p_k \le P_k^{max}$, in which P_k^{max} denotes maximum available transmission power. s_k represents the transmitted signal of user k, satisfying $\mathbb{E}\left[|s_k|^2\right] = 1, k \in \mathcal{K}$. Since the received antenna at the AP is multi-antenna M, multiplying $\mathbf{y}_{AP}$ by $\mathbf{w}_k$, we can obtain

$$\mathbf{w}_k^H\mathbf{y}_{AP} = \mathbf{w}_k^H\mathbf{h}_k(\eta)\sqrt{p_k}s_k + \mathbf{w}_k^H\sum_{i\neq k}^{K}\mathbf{h}_i(\eta)\sqrt{p_i}s_i + \mathbf{w}_k^H\mathbf{n}, \tag{3}$$

where $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$ is the normalized beamforming vector, satisfying $|\mathbf{w}_k|^2$=1. Note that uplink NOMA strategy is adopt. Similar to [17], we assume that the channel gain of K users can be sorted as

$$\|\mathbf{G}\mathbf{\Theta}\mathbf{h}_1 + \mathbf{h}_{a,1}\|_2 \ge \cdots \ge \|\mathbf{G}\mathbf{\Theta}\mathbf{h}_K + \mathbf{h}_{a,K}\|_2. \tag{4}$$

Thus, the SNR of user k is given by

$$\gamma_k = \frac{|\mathbf{w}_k^H\mathbf{h}_k(\eta)|^2 p_k}{\sum_{i=k+1}^{K}|\mathbf{w}_k^H\mathbf{h}_i(\eta)|^2 p_i + \sigma_k^2}, \tag{5}$$

where $\sigma_k^2 = |\mathbf{w}_k^H\mathbf{n}|^2$. Then, the achievable rate of user k is $R_k = B\log_2(1+\gamma_k)$, where B denotes the signal bandwidth.

2.2 Partial Computational Offloading Model

In this subsection, partial computational offloading model is considered, which indicate that a portion of computation tasks are offloaded to the AP for remote execution, while the remaining portion is computed locally [4]. Then, we denote T as the maximum tolerated computation latency. Next, we will introduce the four parts of completing each user's computing task, including local computing, task offloading, edge computing and result downloading.

2.2.1 Local Computing

In the stage of local computing, the time required for processing task is T. Then, we denote f_k as the CPU frequency of user k, and C as CPU cycles per bit required for local computing. Then, the computational bits locally computed at user k can be given by $L_k^{loc} = \frac{Tf_k}{C}$.

Correspondingly, the CPU energy consumed per second by user k is modeled as $\xi_k f_k^3$, where ξ_k is corresponding coefficient related to the chip architecture [18]. Then, the energy consumption for local computing at user k is expressed as $E_k^{loc} = T\xi_k f_k^3$.

2.2.2 Task Offloading

Besides, in the stage of task offloading, K users need to transmit a portion of their computation tasks to AP by applying NOMA strategy. Also, simliar to [19], we set the time consumption of task offloading as T. Then, we can obtain the computational bits of task offloading at user k as $L_k^{off} = TR_k$. Accordingly, the energy consumption for task offloading is expressed as $E_k^{off} = Tp_k$.

2.2.3 Edge Computing and Result Downloading

In the stage of edge computing and result downloading, the MEC server calculate computational task and subsequently return results to K users. Due to the abundant computation resources of MEC servers and the typically very small size of computing result, the energy and time consumption can be ignored [20].

Based on the above analysis, we can get that the total computational bits completed by user k as

$$L_k = \frac{Tf_k}{C} + TR_k. \tag{6}$$

Correspondingly, the total energy consumed by user k to complete its computational bits can be represented as $E_k = E_k^{loc} + E_k^{off} = T\xi_k f_k^3 + Tp_k$.

2.3 Problem Formulation

As introduced above, we aim to maximize the total computation bits of K users through appropriate receiver beamforming matrix, transmission power, CPU frequency, and IRS phase shifts control. To proceed, the considered problem is formulated as:

$$\max_{\mathbf{W},\{f_k\},\mathbf{p},\boldsymbol{\Theta}} \quad \sum_{k=1}^{K}\left(\frac{Tf_k}{C} + BTR_k\right) \tag{7}\text{a}$$

$$\text{s.t.} \quad \xi_k f_k^3 + p_k \le P_k^{max}, \forall k \in \mathcal{K} \tag{7}\text{b}$$

$$0 \le p_k \le P_k^{max}, \forall k \in \mathcal{K}, \tag{7}\text{c}$$

$$\gamma_k \ge \gamma_k^{min}, \forall k \in \mathcal{K}, \tag{7}\text{d}$$

$$|\mathbf{w}_k|^2 = 1, \forall k \in \mathcal{K} \tag{7}\text{e}$$

$$0 \le f_k \le f_k^{max}, \forall k \in \mathcal{K}, \tag{7}\text{f}$$

$$0 \le \theta_n \le 2\pi, \quad \forall n \in \mathcal{N}, \tag{7}\text{g}$$

where $\mathbf{W} = [\mathbf{w}_1, \cdots, \mathbf{w}_k]$ denotes the receiver beamforming matrix of AP, $\mathbf{p} = [p_1, \cdots, p_K]$ denotes the transmission power vector. The objective function represents the total computation bits of K users. The constraints *(7)*b and *(7)*c indicates that the power consumption and transmission power of user k should be lower than the maximum available power P_k^{max}. The constraint *(7)*d states the received SNR at AP to detect s_k successfully should be no less than the target SNR. The constraint *(7)*e shows the receive beamforming matrix restriction of user k. In addition, the range of CPU frequency and IRS phase shifts is characterized by constraint *(7)*f and *(7)*g, respectively.

3. ALTERNATING OPTIMIZATION SOLUTION

Because of coupling variables with regard to receiver beamforming matrix, phase matrix and transmission power in *(7)*a and *(7)*d, the optimization problem *(7)* is not jointly convex and challenging to solve. To make the problem tractable, an AO-based algorithm is adopted to efficiently tackle it. Specifically, we split the problem *(7)* into receiver beamforming optimization, CPU frequency optimization, transmission power optimization and IRS phase shift optimization subproblems, and then handle these four subproblems alternately. In particular, we first derive the closed form expression of receiver beamforming matrix and CPU frequency, then exploit SCA-based iterative algorithm to optimize transmission power, and finally adopt SDR-based iterative algorithm to optimize phase matrix.

3.1 Optimize receiver beamforming

For given $\mathbf{p}$, $\{f_k\}$, and $\boldsymbol{\Theta}$, it is easy to verify that maximizing the total computational bits of K users is equivalent to maximizing the sum rate of K users in the system. Furthermore, the maximum achievable sum rate of K users is given by [21, 22]:

$$\begin{aligned}&\sum_{k=1}^{K}\log_2\left(1+\frac{|\mathbf{w}_k^H\mathbf{h}_k(\eta)|^2 p_k}{\sum_{i=k+1}^{K}|\mathbf{w}_k^H\mathbf{h}_i(\eta)|^2 p_i+\sigma_k^2}\right)\\&=\log_2\det\left(\mathbf{I}_M+\sum_{k=1}^{K}\frac{p_k\mathbf{h}_k(\eta)\mathbf{h}_k(\eta)^H}{\sigma_k^2}\right).\end{aligned} \tag{8}$$

According to [23], the SIC receiver with minimum MSE can achieve the maximum sum rate, then we obtain

$$\bar{\mathbf{w}}_k = \left(N_0\mathbf{I}_M + \sum_{i=k+1}^{K} p_i\mathbf{h}_k(\eta)\mathbf{h}_k(\eta)^H\right)^{-1}\mathbf{h}_k(\eta). \tag{9}$$

By standardizing $\bar{\mathbf{w}}_k$, $\mathbf{w}_k$ can be given by $\mathbf{w}_k = \frac{\bar{\mathbf{w}}_k}{|\bar{\mathbf{w}}_k|}$. Correspondingly, we can obtain the rate of user k as

$$\begin{aligned}R_k =& \log_2\det\left(\mathbf{I}_M+\sum_{i=k}^{K}\frac{p_k\mathbf{h}_i(\eta)\mathbf{h}_i(\eta)^H}{\sigma_k^2}\right)\\&-\log_2\det\left(\mathbf{I}_M+\sum_{i=k+1}^{K}\frac{p_k\mathbf{h}_i(\eta)\mathbf{h}_i(\eta)^H}{\sigma_k^2}\right).\end{aligned} \tag{10}$$

3.2 Optimize CPU Frequency

Secondly, we optimize the CPU frequency $\{f_k\}$ for local computing. Given $\mathbf{W}$, $\mathbf{p}$, and $\boldsymbol{\Theta}$, the CPU frequency

optimization subproblem can be formulated as:

$$\max_{\{f_k\}} \quad \sum_{k=1}^{K} \frac{T f_k}{C} \tag{11}\text{a}$$

$$\text{s.t.} (7)b, (7)f \tag{11}\text{b}$$

We can observe that the optimization problem *(11)* satisfy linear programming. Therefore, we can directly derive CPU frequency in closed-form expression. Based on constraints *(11)*b, the range of f_k can be denoted by

$$0 \le f_k \le \min\left\{ f_k^{max}, \sqrt[3]{\frac{P_k^{max} - p_k}{\xi_k}} \right\}. \tag{12}$$

Since the objective function *(11)*a monotonically increases with f_k, the optimal f_k^* of local CPU frequency is the upper bound of f_k, which is given by

$$f_k^* = \min\left\{ f_k^{max}, \sqrt[3]{\frac{P_k^{max} - p_k}{\xi_k}} \right\}. \tag{13}$$

3.3 Optimize Transmission Power

Thirdly, we focus on transmission power optimization. Given $\mathbf{W}$, $\{f_k\}$, and $\mathbf{\Theta}$, the transmission power optimization subproblem can be rephrased as

$$\max_{\mathbf{p}} \quad \sum_{k=1}^{K} \log_2\left(1 + \frac{|\mathbf{w}_k^H \mathbf{h}_k(\eta)|^2 p_k}{\sum_{i=k+1}^{K} |\mathbf{w}_k^H \mathbf{h}_i(\eta)|^2 p_i + \sigma_k^2}\right) \tag{14}\text{a}$$

$$\text{s.t.} (7)b, (7)c, (7)d \tag{14}\text{b}$$

Because objective function *(14)*a is nonconvex, problem *(14)* is still difficult to handle. By defining $g_{k,i} = \frac{|\mathbf{w}_k \mathbf{h}_i(\eta)|^2}{\sigma_k^2}$, the objective function *(14)*a is reexpressed as

$$\sum_{k=1}^{K} \log_2\left(\sum_{i=k}^{K} p_i g_{k,i} + 1\right) - \sum_{k=1}^{K} \log_2\left(\sum_{i=k+1}^{K} p_i g_{k,i} + 1\right) \tag{15}$$

Subsequently, by substituting the expression of γ_k into *(7)*d, the constraint *(7)*d can be rephrased as

$$p_k g_{k,k} \ge \gamma_k^{min}\left(\sum_{i=k}^{K} p_i g_{k,i} + 1\right), \forall k \in \mathcal{K}. \tag{16}$$

Obviously, *(16)* is an affine constraint, that is, a convex constraint. Correspondingly, we define functions

$$\begin{aligned} z_1(\mathbf{p}) &= \sum_{k=1}^{K} \log_2\left(\sum_{i=k}^{K} p_i g_{k,i} + 1\right), \\ z_2(\mathbf{p}) &= \sum_{k=1}^{K} \log_2\left(\sum_{i=k+1}^{K} p_i g_{k,i} + 1\right). \end{aligned} \tag{17}$$

Hence, the transmission power optimization subproblem can be reformulated as

$$\max_{\mathbf{p}} \quad z_1(\mathbf{p}) - z_2(\mathbf{p}) \tag{18}\text{a}$$

$$\text{s.t.} (7)b, (7)c, (16) \tag{18}\text{b}$$

We can observe that *(18)*b are all convex, and *(18)*a is the difference between two concave functions, the optimization problem *(18)* is a typical DC programming problem. For $k \in \mathcal{K}$, we define function $u_k(i)$ that satisfies $u_k(k) = 0$ as

$$u_k(i) = \frac{g_{k,i}}{\ln 2}, i \ge k+1, \tag{19}$$

then the gradient of z_2 at $\mathbf{p}$ can be expressed as

$$\nabla z_2(\mathbf{p}) = \sum_{k=1}^{K} \frac{u_k}{\sum_{i=k+1}^{K} p_i g_{k,i} + 1}, \tag{20}$$

Then, by defining $(m-1)$-th iteration feasible solution $\{\mathbf{p}^{(m-1)}\}$, and initializing a feasible solution $\{\mathbf{p}^{(0)}\}$, the optimal solution $\{\mathbf{p}^{(m)}\}$ of m-th iteration can be obtained by converting problem *(18)* to

$$\max_{\mathbf{p}} z_1(\mathbf{p}) - z_2(\mathbf{p}^{(m-1)}) - \left\langle \nabla z_2(\mathbf{p}^{(m-1)}), \mathbf{p} - \mathbf{p}^{(m-1)} \right\rangle \tag{21}\text{a}$$

$$\text{s.t.} (18)b \tag{21}\text{b}$$

where $\langle ., . \rangle$ is the operation of inner product. Obviously, problem *(21)* is convex. Therefore, we can efficiently solved it via utilizing CVX toolbox [24].

3.4 Optimize Phase Shift

Finally, we optimize the phase matrix $\mathbf{\Theta}$. Given $\mathbf{W}$, $\{f_k\}$ and $\mathbf{p}$, the IRS phase shifts optimization subproblem can be reformulated as

$$\max_{\mathbf{\Theta}} \sum_{k=1}^{K} \log_2(1 + \gamma_k) \tag{22}\text{a}$$

$$\text{s.t.} (7)d, (7)g \tag{22}\text{b}$$

To transform the terms $|\mathbf{w}_k^H(\mathbf{G}\mathbf{\Theta}\mathbf{h}_i + \mathbf{h}_{a,i})|^2$ into tractable form, we define $\mathbf{v} = [v_1, \cdots, v_N]^H$, where $v_n = e^{j\theta_n}, \forall n \in \mathcal{N}$. Then, we define $\hat{\mathbf{h}}_{k,i} = (\mathbf{w}_k^H \mathbf{G}) \circ \mathbf{h}_i$, where $\circ$ denotes the operation of Hadamard product. Correspondingly, we can obtain

$$|\mathbf{w}_k^H \mathbf{G}\mathbf{\Theta}\mathbf{h}_i|^2 = |\mathbf{v}^H \hat{\mathbf{h}}_{k,i}|^2. \tag{23}$$

Besides, by defining $b_{k,i} = \mathbf{w}_k^H \mathbf{h}_{a,i}$, the problem *(22)* can be reformulated as

$$\max_{\mathbf{v}} \sum_{k=1}^{K} \log_2\left(1 + \frac{|\mathbf{v}^H \hat{\mathbf{h}}_{k,k} + b_{k,k}|^2 p_k}{\sum_{i=k+1}^{K} |\mathbf{v}^H \hat{\mathbf{h}}_{k,i} + b_{k,i}|^2 p_i + \sigma_k^2}\right) \tag{24}\text{a}$$

$$\begin{aligned} \text{s.t.} & |\mathbf{v}^H \hat{\mathbf{h}}_{k,k} + b_{k,k}|^2 p_k \\ & \ge \gamma_k^{min}\left(\sum_{i=k+1}^{K} |\mathbf{v}^H \hat{\mathbf{h}}_{k,i} + b_{k,i}|^2 p_i + \sigma_k^2\right), \forall k \in \mathcal{K} \end{aligned} \tag{24}\text{b}$$

$$|v_n| = 1, \forall n \in \mathcal{N}, \tag{24}\text{c}$$

Nevertheless, problem *(24)* is still non-convex. Then, by defining

$$\mathbf{\Phi}_{k,i} = p_i \begin{bmatrix} \hat{\mathbf{h}}_{k,i}^H \hat{\mathbf{h}}_{k,i} & \hat{\mathbf{h}}_{k,i}^H b_{k,i}^H \\ \hat{\mathbf{h}}_{k,i} b_{k,i} & b_{k,i}^H b_{k,i} \end{bmatrix}, \bar{\mathbf{v}} = \begin{bmatrix} \mathbf{v} \\ 1 \end{bmatrix}, \tag{25}$$

problem *(24)* can be represented as

$$\max_{\mathbf{v}} \sum_{k=1}^{K} \log_2 \left(1 + \frac{\bar{\mathbf{v}}^H \mathbf{\Phi}_{k,k} \bar{\mathbf{v}}}{\sum_{i=k+1}^{K} \bar{\mathbf{v}}^H \mathbf{\Phi}_{k,k} \bar{\mathbf{v}} + \sigma_k^2}\right) \quad (26)\text{a}$$

$$\text{s.t.} \bar{\mathbf{v}}^H \mathbf{\Phi}_{k,k} \bar{\mathbf{v}} \geq \gamma_k^{min} \left(\sum_{i=k+1}^{K} \bar{\mathbf{v}}^H \mathbf{\Phi}_{k,i} \bar{\mathbf{v}} + \sigma_k^2\right), \forall k \in \mathcal{K} \quad (26)\text{b}$$

$$|v_n| = 1, \forall n \in \mathcal{N}. \quad (26)\text{c}$$

Then, we define $\mathbf{V} = \bar{\mathbf{v}}\bar{\mathbf{v}}^H$. By utilizing the property of matrix traces $\bar{\mathbf{v}}^H \mathbf{\Phi}_{k,i} \bar{\mathbf{v}} = Tr\left(\mathbf{\Phi}_{k,i} \bar{\mathbf{v}}\bar{\mathbf{v}}^H\right)$ and SDR to forcibly omit rank-one constraint [25], problem *(26)* can be relaxed as

$$\max_{\mathbf{V}} \sum_{k=1}^{K} \log_2 \left(1 + \frac{Tr\left(\mathbf{\Phi}_{k,k}\mathbf{V}\right)}{\sum_{i=k+1}^{K} Tr\left(\mathbf{\Phi}_{k,i}\mathbf{V}\right) + \sigma_k^2}\right) \quad (27)\text{a}$$

$$\text{s.t.} Tr\left(\mathbf{\Phi}_{k,k}\mathbf{V}\right) \geq \sum_{i=k+1}^{K} Tr\left(\mathbf{\Phi}_{k,i}\mathbf{V}\right) + \sigma_k^2, \forall k \in \mathcal{K} \quad (27)\text{b}$$

$$\mathbf{V}_{n,n} = 1, \forall n \in \{1, \cdots, N+1\}. \quad (27)\text{c}$$

$$\mathbf{V} \succeq 0, \quad (27)\text{d}$$

where $\mathbf{V}_{n,n}$ denotes the diagonal elements of matrix $\mathbf{V}$. Although the constraints *(27)*b-*(27)*d are all convex, the problem remains non-convex because of *(27)*a. Thus, by using the properties of matrix traces [25], we can equivalently transform *(27)*a into

$$\begin{aligned} &\sum_{k=1}^{K} \log_2 \left(1 + \frac{Tr\left(\mathbf{\Phi}_{k,k}\mathbf{V}\right)}{\sum_{i=k+1}^{K} Tr\left(\mathbf{\Phi}_{k,i}\mathbf{V}\right) + \sigma_k^2}\right) \\ &= \sum_{k=1}^{K} \log_2 \left(Tr\left(\mathbf{V}\sum_{i=k}^{K} \mathbf{\Phi}_{k,i}\right) + \sigma_k^2\right) \\ &- \sum_{k=1}^{K} \log_2 \left(Tr\left(\mathbf{V}\sum_{i=k+1}^{K} \mathbf{\Phi}_{k,i}\right) + \sigma_k^2\right). \end{aligned} \quad (28)$$

And we define

$$\begin{aligned} z_3(\mathbf{V}) &= \sum_{k=1}^{K} \log_2 \left(Tr\left(\mathbf{V}\sum_{i=k}^{K} \mathbf{\Phi}_{k,i}\right) + \sigma_k^2\right), \\ z_4(\mathbf{V}) &= \sum_{k=1}^{K} \log_2 \left(Tr\left(\mathbf{V}\sum_{i=k+1}^{K} \mathbf{\Phi}_{k,i}\right) + \sigma_k^2\right). \end{aligned} \quad (29)$$

We can easily obtain that *(28)* is a DC function, so we can use DC programming to solve it. Nevertheless, we cannot take partial derivatives with respect to the complex variables as in *(20)* by reason of the complex matrix of $\mathbf{V}$. Moreover, since $z_4(\mathbf{V})$ is not an analytic function, its derivative with respect to $\mathbf{V}$ does not exist [26]. To address this issue, we derive $z_4(\mathbf{V})$ with regard to the real and imaginary parts of $\mathbf{V}$. Since $\mathbf{V}$ is a symmetric matrix, we only need to calculate the real and imaginary parts of the lower triangular elements of $\mathbf{V}$. Specifically, for $V_{n,s}, s \leq n, \forall n \in \mathcal{N}$, the partial derivatives concerning the real and imaginary parts are respectively given by

$$\frac{\partial z_4}{\partial \mathrm{Re}\left(\mathbf{V}_{n,s}\right)} = \frac{1}{\ln 2} \sum_{k=1}^{K} \frac{\sum_{i=k+1}^{K} \mathbf{\Phi}_{k,i_{(n,s)}} + \mathbf{\Phi}_{k,i_{(n,s)}}^H}{Tr\left(\mathbf{V}\sum_{i=k+1}^{K} \mathbf{\Phi}_{k,i}\right) + \sigma_k^2}, \quad (30)$$

Algorithm 1 The Gaussian Randomization Method For Solving (41).

1: **Initialize**: The optimal solution $\bar{\mathbf{v}}^*$ of the problem (41), the random number Ω and $t = 1, 2, \cdots, \Omega$.
2: **repeat**
3: Generate random vectors $\mathbf{q} \sim \mathcal{CN}(0, \mathbf{I}_{N+1})$.
4: Obtain random values for the solution of optimization problem (41) $\bar{\mathbf{v}}_t = \bar{\mathbf{v}}^* \cdot \mathbf{q}$.
5: Standardize $\bar{\mathbf{v}}_t$ to obtain $\bar{\mathbf{v}}_t^{norm}$.
6: Transform $\bar{\mathbf{v}}_t^{norm}$ into $N \times N$ dimensional matrix $\mathbf{\Theta}_t = \mathrm{diag}\left(\bar{\mathbf{v}}_{\mathrm{t}}^{\mathrm{norm}}\right)_{\mathrm{N\times N}}$ matrix.
7: Compute total computational bits L_t^{total} according to $\mathbf{\Theta}_t$.
8: $t \leftarrow t + 1$.
9: **until** $t = \Omega$.
10: **Output**: index t which maximizes L_t^{total} and its corresponding $\mathbf{\Theta}_{t^*}$.

$$\frac{\partial z_4}{\partial \mathrm{Im}\left(\mathbf{V}_{n,s}\right)} = \frac{-i}{\ln 2} \sum_{k=1}^{K} \frac{\sum_{i=k+1}^{K} \mathbf{\Phi}_{k,i_{(n,s)}} - \mathbf{\Phi}_{k,i_{(n,s)}}^H}{Tr\left(\mathbf{V}\sum_{i=k+1}^{K} \mathbf{\Phi}_{k,i}\right) + \sigma_k^2}, \quad (31)$$

where $\mathbf{\Phi}_{k,i_{(n,s)}}$ represents $(n,s) - th$ element of $\mathbf{\Phi}_{k,i}$. Based on the above derivations, we can reformulate problem *(27)* in m-th iteration as

$$\begin{aligned} \max_{\mathbf{V}} \quad & z_3(\mathbf{V}) - z_4(\mathbf{V}^{(m-1)}) \\ & - \sum_{n=1}^{N}\sum_{s=1}^{n-1} \mathrm{Re}\left(\mathbf{V}_{n,s} - \mathbf{V}_{n,s}^{(m-1)}\right) \times \frac{\partial z_4}{\partial \mathrm{Re}\left(\mathbf{V}_{n,s}^{(m-1)}\right)} \\ & - \sum_{n=1}^{N}\sum_{s=1}^{n-1} \mathrm{Im}\left(\mathbf{V}_{n,s} - \mathbf{V}_{n,s}^{(m-1)}\right) \times \frac{\partial z_4}{\partial \mathrm{Im}\left(\mathbf{V}_{n,s}^{(m-1)}\right)} \end{aligned} \quad (32)\text{a}$$

$$\text{s.t.} (27)b - (27)d. \quad (32)\text{b}$$

In this case, optimization problem *(32)* becomes a SDP problem. Thus, we can efficiently solved by applying convex optimization tools, e.g., interior point methods or the CVX package [24]. However, since the obtained optimal solution $\mathbf{V}^*$ is not necessarily rank-one, we need to further apply Gaussian randomization to obtain $\mathbf{\Theta}$. The detailed steps of Gaussian randomization are outlined in Algorithm 1.

In summary, this section proposes an AO-based algorithm to maximize computation rate. Firstly, we derive the closed-form solution for $\mathbf{W}$. Afterwards we achieve the closed-form optimal solution of f_k by solving the subproblem *(11)*. Then, we optimize $\mathbf{p}$ by using variable substitution and SCA-based iterative algorithm. Finally, $\mathbf{\Theta}$ is optimized by applying variable substitution and SDR-based iterative algorithm, where the four subproblems are optimized alternately.

3.5 Complexity Analysis

We define the computational complexity of each iteration as J^{iter}, and the maximum number of iterations as J. Then, the computational complexity upper bound can be denoted by $O(JJ^{iter})$. Note that the computational complexity comes

from part "repeat" in Algorithm 1. Firstly, in the process of optimizing $\mathbf{V}^{(iter)}$, its complexity is $O((KM)^4)$. Secondly, in the process of optimizing $\{f_k\}^{(iter)}$, because of constant operation, its computational complexity is K. Furthermore, in the process of optimizing $\mathbf{p}^{(iter)}$, we define the maximum number of iterations for the optimization problem (21) as J_1, resulting in a computational complexity of $O(J_1K^2)$. Finally, there are two parts involved in the process of optimizing $\mathbf{\Theta}^{(iter)}$. One part is the optimization of $\mathbf{V}$, where we define the maximum iterations of problem (32) as J_2, the complexity is $O(J_2(K+N^2)^{3.5})$ according to reference [24]. The other part is to recover $\mathbf{\Theta}^{(iter)}$ by using Gaussian randomization. There are Ω loops and N-dimensional vector multiplications, thus the complexity is $O(\Omega N)$. Therefore, the complexity of optimizing $\mathbf{\Theta}^{(iter)}$ is $O(J_2(K+N^2)^{3.5})+O(\Omega N)$. As a result, the total complexity of the proposed algorithm is $O(J((KM)^4+K+J_1K^2+J_2(K+N^2)^{3.5}+\Omega N))$.

4. NUMERICAL RESULTS

In this section, numerical results are presented to evaluate the superior performance of the proposed MISO-based MEC system with IRS in terms of computation rate. We consider using the Cartesian coordinate system to illustrate the relative locations of user k, IRS, and AP, which is shown in Fig.2. Specifically, the relative coordinates of AP, IRS, and user k are $(x_{AP},0)$, $(0,y_{IRS})$ and (x_k,y_k), respectively, and K users are uniformly distributed in a circular cell centered at $(x_D,0)$ with radius r_d. Besides, the distance between user k and IRS, between IRS and AP and between user k and AP are denoted as $d_1=\sqrt{(y_{IRS}-y_k)^2+x_k^2}$, $d_2=\sqrt{x_{AP}^2+x_k^2}$ and $d_3=\sqrt{(x_{AP}-x_k)^2+y_k^2}$, respectively. The distance-dependent path loss is modeled by $P_L=\beta d^{-\alpha}$. Here, Euclidean distance, the path loss and corresponding exponent are denoted by d, β, and α, respectively.

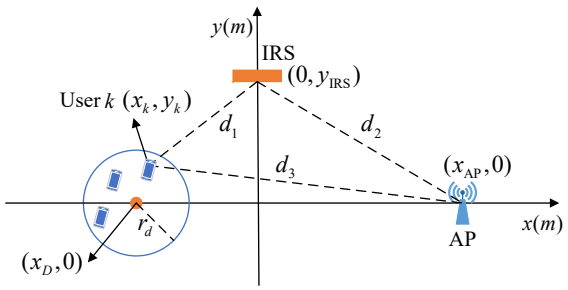

Figure 2 – The relative locations of user k, IRS, and AP.

In addition, we assume the small-scale fading between user k and IRS, and that between IRS and AP, following Rayleigh fading channel model, while that between user and AP experiences Rician fading channel model. Meanwhile, the Rician factor is set as 3. Without otherwise specified, we set $K=3$, and the path loss exponents for user-to-AP, user-to-IRS, and IRS-to-AP links are 3, 2.5 and 2.2, respectively. Besides, we set A=-30dB, x_{AP}=40m, y_{IRS}=12m, x_D=-15m, r_d=8m, N=40, B=2MHz, M=4, P_k^{max}=30dBm, and the noise power is -174dBm/Hz.

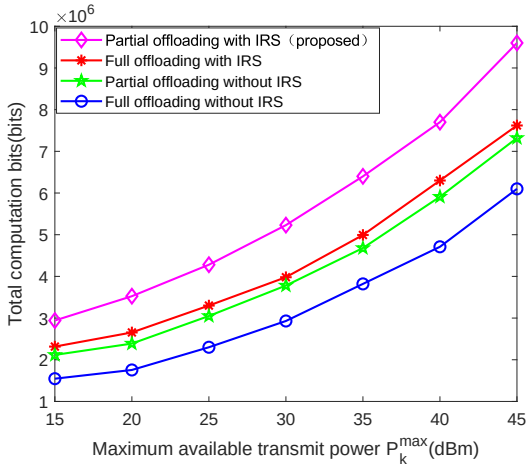

Figure 3 – The total computational bits versus the maximum available transmission power.

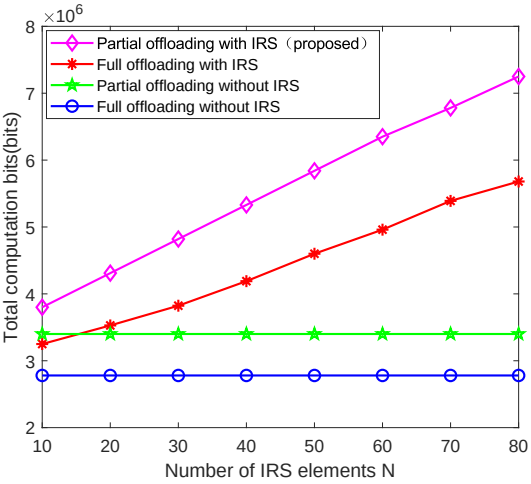

Figure 4 – The total computational bits versus the number of IRS elements N.

Moreover, the maximum CPU frequency is $f_k^{max}=10^9$cycles/s, the CPU chip coefficient of the CPU is $\xi_k=10^{-28}$, and $C=10^3$ cycles/bit.

Fig.3 shows the total computational bits versus the maximum available transmission power $P_k^{\max}$. We can observe that the total computational bits increases with the increase of $P_k^{\max}$ for all schemes, which clearly indicates that all schemes have consistent variation characteristics. Moreover, our proposed scheme obtain a substantial gain over the other three benchmark schemes, and the gap becomes larger as the maximum available transmission power increases. It implies that the proposed partial offloading scheme can intelligently allocate computation tasks, thus improve computing efficiency compared to the full offloading schemes. In addition, the introduction of IRS adds a new dimension for improving the diversity gain of the AP receiver, enhancing the signal reflection and reception efficiency. As a result, the proposed scheme can bring extra performance gain in comparison with the schemes without IRS.

Fig.4 further shows the total computational bits versus N. We can observe that the schemes without IRS do not change as N increases. Next, for the IRS-aided schemes, the

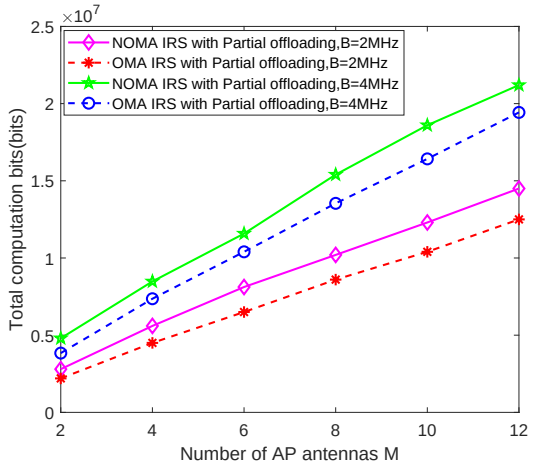

Figure 5 – The total computational bits versus the number of receiver antennas M.

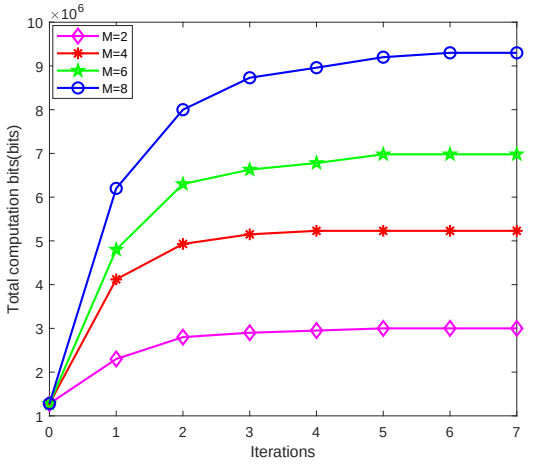

Figure 6 – The total computational bits versus the iterations.

total computational bits increases with N, which clearly demonstrates that the proposed scheme and the IRS-aided full offloading scheme with NOMA exhibit consistent trends. The reason is that the addition number of N can further improve the offloading efficiency, thereby effectively improving the computational offloading performance. Moreover, IRS plays a vital part in improving computational bits, as it can increase the transmission rate and thus enhancing the computational offloading performance.

Fig.5 plots the total computational bits versus M under different bandwidths B and transmission strategies. We can observe that the total computational bits increases with the bandwidth increases. This is because introducing MISO and optimizing the receive beamforming matrix can effectively improve the system capacity, enhancing the offloading efficiency. Meanwhile, increasing the number of antennas and bandwidth improves the offloading efficiency, thereby enhancing the overall system performance. In addition, we can see that the NOMA-based scheme significantly outperforms the partial offloading scheme with OMA in terms of the total computational bits. The reason is that under the same bandwidth, compared to OMA, NOMA transmission strategy can dramatically improve the spectral efficiency, thus increasing the offloading capability and the total computational bits.

Fig.6 presents the total computational bits versus the iterations under different number of receiving antennas at AP. We set $N = 40$ and $P_k^{\max} = 30$dBm. It can be observed that under different antenna numbers, the total computational bits can converge within 7 iterations, and the value after 4 iteration is already close to the final converged value. This validates that the proposed algorithm can converge quickly and effectively. Besides, we can see that the total computational bits improves as M increases.

5. CONCLUSION

In this paper, we considered a MISO-based MEC system with IRS. To assess the computational performance, we studied the joint receiver beamforming, CPU frequency, transmission power, and the IRS phase shifts problem to maximize the computation rate. An AO algorithm was developed to solve the challenging non-convex problem. Firstly, we split the orginal problem into four subproblems and then solve the subproblems alternatively. Then, closed form optimal solution were derived to optimize receiver beamforming matrix and CPU frequency. Finally, we exploited SCA-based iterative algorithm and SDR-based iterative algorithm to optimize transmission power and IRS phase shifts, respectively. It was proved that our proposed MISO-based MEC system with IRS can achieve superior computational performance compared to four benchmark schemes.

The substantial gains in computational rate achieved by the proposed system highlight its potential to support emerging applications with computation-intensive and delay-sensitive in B5G/6G networks. For example, in augmented reality services, the enhanced computational capacity can enable real-time rendering of high-quality graphics and seamless user interaction. In intelligent transportation systems, the reduced latency can facilitate real-time processing of vast sensor data for rapid decision-making in autonomous driving.

REFERENCES

[1] F. Fang, K. Wang, Z. Ding, and V. C. Leung, "Energy-efficient resource allocation for noma-mec networks with imperfect csi," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3436–3449, 2021.

[2] L. A. Haibeh, M. C. Yagoub, and A. Jarray, "A survey on mobile edge computing infrastructure: Design, resource management, and optimization approaches," *IEEE Access*, vol. 10, pp. 27 591–27 610, 2022.

[3] B. Ji, Y. Wang, K. Song, C. Li, H. Wen, V. G. Menon, and S. Mumtaz, "A survey of computational intelligence for 6g: Key technologies, applications and trends," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 7145–7154, 2021.

[4] F. Zhou and R. Q. Hu, "Computation efficiency maximization in wireless-powered mobile edge

computing networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3170–3184, 2020.

[5] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE communications magazine*, vol. 58, no. 1, pp. 106–112, 2019.

[6] S. Hong, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Artificial-noise-aided secure mimo wireless communications via intelligent reflecting surface," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7851–7866, 2020.

[7] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE transactions on communications*, vol. 69, no. 5, pp. 3313–3351, 2021.

[8] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Latency minimization for intelligent reflecting surface aided mobile edge computing," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2666–2682, 2020.

[9] C. Sun, W. Ni, Z. Bu, and X. Wang, "Energy minimization for intelligent reflecting surface-assisted mobile edge computing," *IEEE Transactions on Wireless Communications*, vol. 21, no. 8, pp. 6329–6344, 2022.

[10] Z. Chu, P. Xiao, M. Shojafar, D. Mi, J. Mao, and W. Hao, "Intelligent reflecting surface assisted mobile edge computing for internet of things," *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 619–623, 2020.

[11] X. Pei, Y. Chen, M. Wen, H. Yu, E. Panayirci, and H. V. Poor, "Next-generation multiple access based on noma with power level modulation," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 4, pp. 1072–1083, 2022.

[12] F. Zhou, C. You, and R. Zhang, "Delay-optimal scheduling for irs-aided mobile edge computing," *IEEE wireless communications letters*, vol. 10, no. 4, pp. 740–744, 2020.

[13] Z. Li, M. Chen, Z. Yang, J. Zhao, Y. Wang, J. Shi, and C. Huang, "Energy efficient reconfigurable intelligent surface enabled mobile edge computing networks with noma," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 2, pp. 427–440, 2021.

[14] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "Capacity comparison between mimo-noma and mimo-oma with multiple users in a cluster," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2413–2424, 2017.

[15] S. Mao, N. Zhang, L. Liu, J. Wu, M. Dong, K. Ota, T. Liu, and D. Wu, "Computation rate maximization for intelligent reflecting surface enhanced wireless powered mobile edge computing networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 10, pp. 10 820–10 831, 2021.

[16] S. Bi and Y. J. Zhang, "Computation rate maximization for wireless powered mobile-edge computing with binary computation offloading," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 4177–4190, 2018.

[17] M. Zeng, X. Li, G. Li, W. Hao, and O. A. Dobre, "Sum rate maximization for irs-assisted uplink noma," *IEEE Communications Letters*, vol. 25, no. 1, pp. 234–238, 2020.

[18] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, "Mobile-edge computing: Partial computation offloading using dynamic voltage scaling," *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4268–4282, 2016.

[19] X. Yu, F. Xu, J. Cai, X.-y. Dang, and K. Wang, "Computation efficiency optimization for millimeter-wave mobile edge computing networks with noma," *IEEE Transactions on Mobile Computing*, 2022.

[20] X. Hu, K.-K. Wong, and Y. Zhang, "Wireless-powered edge computing with cooperative uav: Task, time scheduling and trajectory design," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 8083–8098, 2020.

[21] B. Ning, Z. Chen, W. Chen, and J. Fang, "Beamforming optimization for intelligent reflecting surface assisted mimo: A sum-path-gain maximization approach," *IEEE Wireless Communications Letters*, vol. 9, no. 7, pp. 1105–1109, 2020.

[22] M. Zeng, E. Bedeer, O. A. Dobre, P. Fortier, Q.-V. Pham, and W. Hao, "Energy-efficient resource allocation for irs-assisted multi-antenna uplink systems," *IEEE Wireless Communications Letters*, vol. 10, no. 6, pp. 1261–1265, 2021.

[23] G. Li, M. Zeng, D. Mishra, L. Hao, Z. Ma, and O. A. Dobre, "Energy-efficient design for irs-empowered uplink mimo-noma systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 9490–9500, 2022.

[24] M. Grant and S. Boyd, "Cvx: Matlab software for disciplined convex programming, version 2.1," 2014.

[25] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.

[26] K. Kreutz-Delgado, "The complex gradient operator and the cr-calculus," *arXiv preprint arXiv:0906.4835*, 2009.