

IRS-Based MEC for Delay-Constrained QoS Over RF-Powered 6G Mobile Wireless Networks

Yuedong Li, Fei Wang, Xi Zhang, *Fellow, IEEE*, and Songtao Guo, *Senior Member, IEEE*

Abstract—Mobile-edge computing (MEC) and intelligent reflecting surface (IRS) have been recognized as two key technologies for 6G mobile networks. Consequently, we propose the IRS-based MEC schemes for the delay-constrained quality-of-service (QoS) provisioning over radio frequency (RF) powered 6G mobile networks with *non-linear* energy harvesting (EH) model. Using the IRS technique, multiple mobile users (MUs) first harvest energy from a multi-antenna base station (BS) equipped with an MEC server and then transmit their data to the BS for data-processing. We first formulate a delay minimization problem for MUs under their QoS requirements, by jointly optimizing the IRS's phase-shift matrices, the MEC server's finite computation resource allocation, the MIMO based BS's multi-MU detection (MUD) coefficients, and the energy/data transmission time and task allocation coefficient of each MU. We define the total delay of each MU as the sum of its EH time, data-transmission time, and data-processing time. Since our formulated joint-optimization problem is non-convex with multiple coupled variables, we apply the *block coordinate descending* (BCD) method to decompose it into several subproblems which then can be iteratively solved by a low complexity algorithm. Moreover, we also extend our proposed scheme to IRS-based MEC over Terahertz (THz) wideband mobile networks. Finally, we validate and evaluate our developed delay minimization schemes through numerical analyses, which show that the total delay of the RF-powered MUs can be significantly reduced by using our proposed schemes.

Index Terms—IRS-Based MEC, Delay-Bounded QoS, RF-Powered 6G Mobile Wireless Networks, MIMO, THz.

I. INTRODUCTION

NOWADAYS, a variety of intelligent services, e.g, smart home and telemedicine, have imposed great challenges for energy and computing-power constrained mobile users (MUs) [1]–[2]. Mobile-edge computing (MEC) together with radio frequency (RF) based wireless energy transfer have been recognized as two promising techniques to solve the above-mentioned issues [2]–[3]. However, due to the poor propagation environments between MUs and energy sources (or MEC servers), the data offloading rates and the amount of harvested

energy of MUs may still be restrained, and thus the advantage of RF-powered MEC cannot be fully utilized.

Recently, the intelligent reflecting surface (IRS) mechanism has been considered as an innovative 6G technique to reconfigure and enhance the wireless propagation environments [2]–[5] for supporting the delay constrained QoS provisioning [6]–[8]. An IRS consists of a large number of low-cost passive or active reflecting elements that can be controlled by an intelligent controller. The controller can automatically adjust the phase shifts of the reflecting elements and then reconfigure and enhance wireless propagation environments dynamically. The IRS is able not only to reflect and enhance signals from different paths, but also to reflect signals at a certain angle to avoid obstacles. Therefore, the strength of the received signals at the information receivers can be significantly enhanced.

There have been a large number of research works studying the potentials of IRSs in enhancing wireless information and energy transmissions. By using IRSs to help information transmission, the authors of [9]–[10] studied the MUs' sum rate maximization problem by optimizing the IRS's phase shifts and the precoding matrices of the base station (BS). In [11]–[12], the authors studied the BS's transmit power minimization problem by optimizing the BS's power allocation and the IRS's phase shifts. Besides, the authors of [13] addressed the system energy efficiency maximization problem by optimizing the BS's beamforming vectors and the IRS's phase shifts. Unlike [13], the authors of [14] maximized the minimum energy efficiency for an IRS-based heterogeneous network by taking into account channel uncertainties and transceivers' hardware impairments. The authors of [15] proposed a joint radio resource allocation and phase shifts optimization scheme for an IRS-based wireless-powered communication system to improve system energy efficiency, where the IRS is used to help multiple MUs harvest RF energy. Moreover, under the energy efficiency and spectral efficiency constraints, the authors of [16] minimized the number of reflecting elements required in an IRS-based wireless system with one source node and one destination node. However, how to utilize IRSs to improve both information and energy transmissions is not considered in [9]–[16]. Therefore, in [17]–[19], the authors leveraged the IRS to enhance both the energy harvesting (EH) efficiency of the EH MUs and the information transmission rates of the information decoding MUs. In [20]–[21], the IRS was first utilized to improve the amount of energy harvested by MUs in the wireless power transfer (WPT) phase, and then was used to improve the information transmission rates of MUs in wireless information transmission (WIT) phase, where the IRS's phase shifts in the WPT and WIT phases were jointly optimized. However, the authors of [17]–[21] did not consider MEC and the beneficial role of IRSs for MEC over RF-powered 6G mobile networks. Besides, since

The work of Fei Wang was supported in part by Zhejiang Lab under Grant 2021LC0AB01. The work of Xi Zhang was supported in part by the U.S. National Science Foundation under Grant CCF-2008975, Grant ECCS-1408601, and Grant CNS-1205726. The work of Songtao Guo was supported in part by the Natural Science Key Foundation of Chongqing under Grant cstc2020jcyj-zdxmX0026, in part by the Fundamental Research Funds for the Central Universities under Grant 2020CDCGJSJ071, Grant 2020CDCGJSJ038, and Grant 2019CDYGD004, and in part by Zhejiang Lab under Grant 2021LC0AB01. (*Corresponding Author: Fei Wang.*)

Yuedong Li and Fei Wang are with College of Electronic and Information Engineering, Southwest University, Chongqing, 400715 China (E-mail: lyd_learning@163.com; wsf0107@163.com).

Xi Zhang is with the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA (E-mail: xizhang@ece.tamu.edu).

Songtao Guo is with the Key Laboratory of Dependable Service Computing in Cyber Physical Society (Chongqing University), Ministry of Education, Chongqing 400054, China, College of computer science, Chongqing University, Chongqing 400044, China (E-mail: songtao_guo@163.com).

MUs' computation offloading and IRSs' phase shifts affect each other, the IRSs' phase shifts selection schemes and/or resource allocation schemes proposed in [17]–[21] may not be suitable for IRS-aided MEC.

In [5], the authors provided an overview of the IRS-aided MEC. The authors of [22]–[23] minimized the maximum learning error of all MUs by optimizing MUs' transmit powers, the BS's beamforming vector, and the IRS's phase shifts, where the IRS is employed to transmit MUs' machine learning tasks to the MEC server co-located with the BS. The authors of [24] studied the MUs' delay minimization problem by jointly optimizing system computation resources and the IRS's phase shifts. In [25], the authors maximized the total completed task-input bits of MUs by jointly optimizing the IRS's phase shifts, the receive beamforming vectors at the access point (AP), and MUs' energy partition for local computing and data offloading. In [26]–[28], the authors minimized the energy consumption of IRS-aided MEC by optimizing the MEC server's computation resource allocation and the IRS's phase shifts. However, in [5] and [22]–[28], IRSs are only used for enhancing data transmission between MUs and MEC servers. Therefore, utilizing IRSs for both data and energy transmissions, the authors of [29] maximized MUs' task computation rates by jointly optimizing the WPT-time allocation, the IRS's phase shifts, and MUs' offloading decision, where each MU can process task locally or offload task to the MEC server. While the authors of [2] studied the system energy minimization for IRS-aided and RF-powered MEC in the case of partial offloading, by jointly optimizing the IRS's phase shifts in both energy and data transmission phases, and the system communication and computation resources. Similarly, the authors of [30] considered the total computation bits maximization problem for IRS-aided and RF-powered MEC, where the IRS's phase shifts in energy transfer and data transmission phases are also jointly optimized. However, for simplicity, the authors of [2], [29]–[30], and the above-mentioned [17] and [20]–[21] all calculated MUs' harvested energy based on the ideal EH model, where each MU's energy conversion efficiency is a constant. In fact, the practical RF-based EH circuits exhibit the *non-linear* characteristics of wireless energy harvesting, where each MU's RF energy conversion efficiency changes with the received RF power level [31]. Besides, the authors of [2] and [29]–[30] ignored the time and energy required for data (the offloaded data) processing at the MEC servers, and then did not consider the MEC servers' resource allocation which is very important for MEC with finite computation resources [3]. In addition, the authors of [2] and [30] only considered the scenarios when the BS and/or AP is equipped with a single antenna, and then multiple MUs transmit data to the MEC server based on the orthogonal multiple access technique, e.g., the frequency division multiple access (FDMA), to avoid the mutual interference among MUs.

Consequently, for MEC over RF-powered mobile networks with *non-linear* EH model, we employ an IRS to help to improve both the EH efficiency and transmission rates of multiple MUs. For the considered IRS-aided MEC, all MUs offload data simultaneously to a multi-antenna BS equipped with an MEC server, and the BS applies the multi-MU detection

(MUD) technique to reduce the co-channel interference among MUs. Then, taking into account the quality-of-service (QoS) requirements of MUs, we develop delay minimization schemes for the IRS-aided and RF-powered MEC by jointly optimizing the IRS's phase-shift matrices, the MEC server's computation resource, the MIMO based BS's MUD coefficients, and the energy/data transmission time and task allocation coefficient of each MU. Here, we define the total delay of each MU as the sum of MU's EH time, data-transmission time, and data-processing time. Also, we extend our work to IRS-aided MEC over Terahertz (THz) wideband mobile networks. Our main contributions can be summarized as follows:

- Taking into account the QoS requirements of MUs, we minimize the MUs' delay for IRS-aided MEC over RF-powered 6G mobile networks with *non-linear* EH model, where an IRS is used to improve both the EH efficiency and data-transmission rates between multiple MUs and a multi-antenna BS equipped with an MEC server. In our considered multi-antenna communication scenario, all MUs offload data to the multi-antenna BS simultaneously over the same spectrum, where the MUD technique is used to reduce the co-channel interference among MUs.
- Since our formulated joint-optimization problem is non-convex with multiple coupled variables, we apply the block coordinate descending (BCD) method to decompose it into several subproblems which can be iteratively solved by a low complexity algorithm. Then, we propose a delay minimization scheme to jointly optimize the IRS's phase-shift matrices, the MEC server's finite computation resource allocation, the MIMO-based BS's MUD coefficients, and the energy/data transmission time and task allocation coefficient of each MU.
- To further reduce MUs' total delay, under the *non-linear* EH model, we also extend our work to IRS-aided MEC over THz wideband mobile networks, where using FDMA, we divide the total bandwidth of the used THz band into several orthogonal sub-bands to combat the frequency-selective fading. Using the BCD method, we also propose a delay minimization scheme.
- Finally, we validate and evaluate our proposed delay minimization schemes through numerical analyses, which show that using our proposed schemes, the EH efficiency and data-transmission rates of all MUs can be significantly improved, and then MUs' total delay can be significantly reduced. Besides, numerical results also show that MUs' total delay can be reduced generally by employing the multi-antenna BS and the *linear* EH model.

The rest of the paper is organized as follows: Section II builds up the system model. Section III develops the delay minimization schemes using MUD. Section IV investigates the delay minimization schemes under THz communications architectures. Section V conducts the numerical analyses. The paper concludes with Section VI. Moreover, for ease of reading, we list the abbreviations in TABLE I.

Notation: $\mathcal{C}^{N \times M}$ is the space of $N \times M$ complex-valued matrices. $\mathcal{R}_+^{N \times M}$ is the space of $N \times M$ positive real-valued matrices. The bold-face lower-case and bold-face upper-case letters, e.g., \mathbf{x} and \mathbf{X} , denote the vector and matrix, respectively.

TABLE I
SUMMARY OF ABBREVIATIONS

MEC	Mobile-edge computing
IRS	Intelligent reflecting surface
MUs	Mobile users
QoS	Quality-of-service
EH	Energy harvesting
BS	Base station
MUD	Multi-MU detection
BCD	Block coordinate descending
Thz	Terahertz
RF	Radio frequency
WPT	Wireless power transfer
WIT	Wireless information transmission
AP	Access point
FDMA	Frequency division multiple access
MIMO	Multiple-input multiple-output
CSI	Channel state information
AWGN	Additive white Gaussian noise
i.i.d.	Independent and identically distributed
CPU	Central processing unit
KKT	Karush-Kuhn-Tucker
BSM	Bisection search method
MSE	Mean-square error
SCA	Successive convex approximation
SDP	Semi-definite programming
UHF	Ultra High Frequency
TDMA	Time division multiple access
NOMA	Non-orthogonal multiple access
RP	Relaxation and projection
SIC	Successive interference cancellation

\mathbf{X}^H is the conjugate transpose matrix of \mathbf{X} . $|\mathbf{x}|$ denotes the vector element-wise absolute value of \mathbf{x} . $\|\mathbf{x}\|$ denotes the Euclidean norm of vector \mathbf{x} and $\text{diag}(\mathbf{x})$ denotes a diagonal matrix with each diagonal entry being the corresponding entry in \mathbf{x} . For a matrix \mathbf{X} , $\text{tr}(\mathbf{X})$ stands for its trace, whereas $[\mathbf{X}]_{pq}$ represents its element in the p -th row and q -th column. Also, $\arg(\mathbf{x})$ denotes the phase extraction operation, and $\mathcal{CN}(\mu, \sigma^2)$ denotes the complex Gaussian distribution with mean μ and variance σ^2 .

II. SYSTEM ARCHITECTURE MODELS

A. Network Architecture

We consider an IRS aided MEC over an RF powered mobile wireless network as shown in Fig. 1, which is composed of one BS equipped with M antennas using the MIMO techniques, K single-antenna-equipped MUs, and an IRS consisting of N reflecting elements [19]. We define the set of the BS's antennas as $\mathcal{M} \triangleq \{1, 2, \dots, m, \dots, M\}$ and the set of MUs as $\mathcal{K} \triangleq \{1, 2, \dots, k, \dots, K\}$. Because of limited computing capability, each MU may offload a fraction of its computing tasks to the MEC server for data processing via the BS. We assume that the distance between the BS and the MEC server is very short, so that the transmission time between them can be ignored. Moreover, we use the terms ‘‘the BS’’ and ‘‘the MEC server’’ interchangeably for simplicity. Due to the substantial path loss, we only consider the first-time signal reflection by the IRS and ignore signals that are reflected thereafter in a way similar to the one used in [21].

The system is time-slotted, and each slot is divided into two phases. In the first phase with duration t_1 , the BS sends RF energy signals to all MUs in \mathcal{K} and each MU harvests energy. In the meantime, the IRS reflects the energy signals to all MUs. In

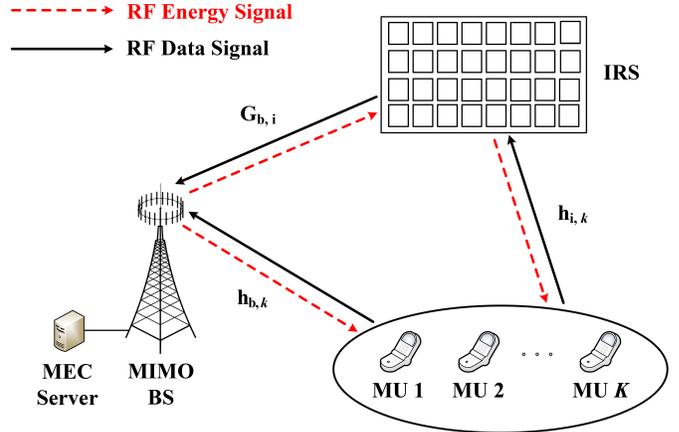


Fig. 1. The system architecture model for our proposed IRS-aided MEC over RF-powered 6G mobile networks, where an IRS assists the computation offloading and EH between K single-antenna MUs and a multi-antenna BS.

the second phase, the MUs use the harvested energy to offload information to the BS, and the IRS reflects MUs' signals to the BS. Then, the MEC server helps to process data for the MUs. Let $t_{2,k}$ and $t_{3,k}$ denote the data-transmission time of MU k and the data-processing time at the MEC server for MU k , respectively. Furthermore, we assume that all channels are quasi-static within each time slot duration [21].

Let $\mathbf{G}_{b,i} \in \mathcal{C}^{M \times N}$, $\mathbf{h}_{b,k} \in \mathcal{C}^{M \times 1}$, and $\mathbf{h}_{i,k} \in \mathcal{C}^{N \times 1}$ denote the channel matrices or vectors between the BS and the IRS, between the BS and MU k , and between the IRS and MU k , respectively. We assume that system channel state information (CSI), i.e., $\mathbf{G}_{b,i}$, $\mathbf{h}_{b,k}$, and $\mathbf{h}_{i,k}$, $\forall k \in \mathcal{K}$, can be perfectly known at the BS using the recently proposed channel estimation techniques in [32] for IRS-aided systems. Hence, the results obtained in this paper actually provide theoretical performance upper bounds for the considered system. In the performance evaluation section, we will evaluate the effect of CSI estimation errors on system performance. We define the phase-shift matrices of the IRS in the first and second phases respectively, as follows:

$$\begin{cases} \Phi_1 \triangleq \text{diag}(\beta_{1,1}e^{j\theta_{1,1}}, \beta_{1,2}e^{j\theta_{1,2}}, \dots, \beta_{1,N}e^{j\theta_{1,N}}), & (1) \\ \Phi_2 \triangleq \text{diag}(\beta_{2,1}e^{j\theta_{2,1}}, \beta_{2,2}e^{j\theta_{2,2}}, \dots, \beta_{2,N}e^{j\theta_{2,N}}), & (2) \end{cases}$$

where $\theta_{1,n}, \theta_{2,n} \in [0, 2\pi], \forall n \in \mathcal{N} \triangleq \{1, 2, \dots, N\}$, are the phase shifts of the n -th reflecting element of the IRS in the first and second phases, respectively. The reflection amplitudes of each reflecting element in the first and second phases are denoted as $\beta_{1,n}$ and $\beta_{2,n}$, respectively, which can be adjusted for different purposes. In this paper, each reflecting element n of the IRS is typically designed to maximize the reflection strength. Thus, similar to [20], we assume that $\beta_{1,n} = \beta_{2,n} = 1$, $\forall n \in \mathcal{N}$. Moreover, we assume that the channel reciprocity holds between the forward and reverse links, and consider both the small scale fading and the large scale path loss for the transmission channel [24]. The main symbols used in this paper are listed in TABLE II.

B. Communication Model

During the first phase with duration t_1 , the BS transmits energy signals with a fixed power level P_b . We denote the

TABLE II
SYSTEM VARIABLES

Symbol	Description
$\mathbf{G}_{b,i}$	Channel coefficients between the BS and the IRS
$\mathbf{h}_{b,k}$	Channel coefficients between the BS and MU k
$\mathbf{h}_{i,k}$	Channel coefficients between the IRS and MU k
N	Number of the IRS's reflecting elements
K	Number of MUs
M	Number of BS's antennas
Φ_1	IRS's phase-shift matrix in the first phase
Φ_2	IRS's phase-shift matrix in the second phase
P_b	BS's transmit power
P_k^t	MU k 's transmit power
t_1	Energy harvesting time
$t_{2,k}$	MU k 's signal transmission time
$t_{3,k}$	Information processing time at the MEC server for MU k
\mathbf{w}_k	MU k 's MUD coefficients
s_k	MU k 's data/information signal
E_k^h	Harvested energy at MU k
D_k^e	Edge computing time for MU k
D_k^l	MU k 's local computing time
L_k	MU k 's total information (in bits)
c_k	Proportion of MU k 's data offloaded to the MEC server
a_k	MU k 's CPU cycles to process one bit of data
L_k^e	MU k 's information volume offloaded to the MEC server (in bits)
L_k^l	MU k 's local computation volume (in bits)
f_k^e	Computation resource allocated to MU k
f_k^l	MU k 's local computing capability
χ_k	Computation energy efficiency of MU k 's processor chip
R_k	Achievable information rate from MU k to the BS
\mathcal{T}	Set of MUs' information transmission time
\mathcal{C}	Set of data allocation proportions of MUs
\mathcal{W}	Set of MUD coefficients for MUs
\mathcal{F}	Set of computation resources allocated to MUs
E_k^{nl}	Amount of harvested energy at MU k
P_k^{max}	MU k 's maximum transmit power
P_k^{min}	MU k 's minimum transmit power
ε_k	Ratio of channel bandwidth allocated to MU k

energy signal as x_0 ($\mathbb{E}[|x_0|^2] = 1$), where $\mathbb{E}[\cdot]$ is the expectation operation. Then, we can express the received energy signal at MU k , denoted by y_k , as:

$$y_k = (\mathbf{G}_{b,i}\Phi_1\mathbf{h}_{i,k} + \mathbf{h}_{b,k})\sqrt{P_b}x_0 + n_k, \quad (3)$$

where $n_k \sim \mathcal{CN}(0, \sigma_b^2)$ denotes the independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) at MU k with zero mean and variance σ_b^2 [19]. Then, we can write the RF power received at MU k , denoted by P_k^h , as:

$$\begin{aligned} P_k^h &= \mathbb{E}[|y_k|^2] \\ &= |\mathbf{G}_{b,i}\Phi_1\mathbf{h}_{i,k} + \mathbf{h}_{b,k}|^2 P_b. \end{aligned} \quad (4)$$

Unlike the traditional *linear* EH model, similar to [31], we apply a more general and more practical *non-linear* EH model, which is given by:

$$E_k^{\text{nl}} = \frac{\Psi_k^{\text{nl}} - \Lambda_k \Omega_k}{1 - \Omega_k}, \quad (5)$$

where

$$\begin{cases} \Psi_k^{\text{nl}} = \frac{\Lambda_k}{1 + e^{-g_k(P_k^h - v_k)}}, \\ \Omega_k = \frac{1}{1 + e^{g_k v_k}}. \end{cases} \quad (6)$$

Moreover, E_k^{nl} denotes the amount of harvested energy at MU k , Ψ_k^{nl} is the traditional logistic function, Ω_k is a constant to guarantee a zero input/output response, and Λ_k is the maximum

harvested power of MU k when its *non-linear* EH circuit saturates. By adjusting the parameters Λ_k , g_k , and v_k , the *non-linear* EH model given by Eq. (5) is able to characterize the joint effects of various *non-linear* phenomena caused by hardware limitations [33]. Therefore, we can express the harvested energy at MU k in the first phase, denoted by E_k^h , as:

$$E_k^h = t_1 E_k^{\text{nl}}. \quad (8)$$

In the second phase with duration $t_{2,k}$, MU k uses a part of harvested energy to transmit information. We denote the amount of energy used by MU k for local computing as $E_{2,k}$. By exhausting the harvested energy, the transmit power of MU k , denoted by P_k^t , can be given by:

$$P_k^t = \frac{E_k^h - E_{2,k}}{t_{2,k}} = \frac{t_1 E_k^{\text{nl}} - E_{2,k}}{t_{2,k}}. \quad (9)$$

Let s_k , where $\mathbb{E}[|s_k|^2] = 1$, denotes the data/information signal of MU k . Moreover, we use the conventional continuous linear MUD technique [24] to reduce the interference among different MUs, and denote $\mathcal{W} \triangleq \{\mathbf{w}_k, \mathbf{w}_k \in \mathcal{C}^{M \times 1}, \forall k \in \mathcal{K}\}$ as the set of MUD vectors for MUs, where \mathbf{w}_k is the MUD vectors for MU k . Then, the signal received at the BS from MU k , denoted by $y_{k,b}$, can be written as:

$$\begin{aligned} y_{k,b} &= \mathbf{w}_k^H \left[(\mathbf{G}_{b,i}\Phi_2\mathbf{h}_{i,k} + \mathbf{h}_{b,k})\sqrt{P_k^t}s_k \right. \\ &\quad \left. + \sum_{j \in \mathcal{K}, j \neq k} \sqrt{P_j^t}s_j (\mathbf{G}_{b,i}\Phi_2\mathbf{h}_{i,k} + \mathbf{h}_{b,j}) + n_b \right], \end{aligned} \quad (10)$$

where $n_b \sim \mathcal{CN}(0, \sigma_b^2)$ denotes the i.i.d. AWGN at the BS with zero mean and variance σ_b^2 . Therefore, we can write the received SINR at the BS for MU k as:

$$\gamma_k = \frac{P_k^t |\mathbf{w}_k^H (\mathbf{G}_{b,i}\Phi_2\mathbf{h}_{i,k} + \mathbf{h}_{b,k})|^2}{\sum_{j \in \mathcal{K}, j \neq k} P_j^t |\mathbf{w}_k^H (\mathbf{G}_{b,i}\Phi_2\mathbf{h}_{i,k} + \mathbf{h}_{b,j})|^2 + \sigma_b^2 |\mathbf{w}_k^H|^2}. \quad (11)$$

Let the bandwidth of the system be B , and then the achievable information rate from MU k to the BS is

$$R_k = B \log_2(1 + \gamma_k). \quad (12)$$

C. Computation Model

In the following, we introduce the detailed computation models for the local computing and edge computing, respectively.

- **Edge Computing:** For MU k , we denote L_k , a_k , and L_k^e as the total information (in bits) to be processed, the number of *central processing unit* (CPU) cycles required to process one bit of data, and the information volume offloaded to the MEC server, respectively. For the MEC server, we denote its maximum computing capability in terms of CPU cycles per second as f_{total}^e , and use f_k^e ($\sum_{j \in \mathcal{K}} f_j^e \leq f_{\text{total}}^e$) to denote the computation resource allocated to MU k . Moreover, we denote c_k as the proportion of MU k 's data offloaded to the MEC server, and denote $\mathcal{C} \triangleq \{c_k, c_k \in [0, 1], \forall k \in \mathcal{K}\}$ as the set of data allocation proportions of MUs. Then, for MU k , the

information volume offloaded to the MEC server, i.e., L_k^e , can be given by $L_k^e = c_k L_k$. Besides, we can write the data-transmission time of MU k as $t_{2,k} = c_k L_k / R_k$, and the data-processing time of the MEC server for MU k as $t_{3,k} = c_k L_k a_k / f_k^e$. Given that the computation result is typically of a small size [34], we neglect the feedback delay. Defining the total delay of MU k as the sum of its EH time, data-transmission time, and data-processing time. Then, the total delay of MU k in edge computing, denoted by D_k^e , is given by $D_k^e = t_1 + t_{2,k} + t_{3,k}$.

- **Local Computing:** For MU k , we use L_k^l to denote the local computation volume of MU k , where $L_k^l = (1 - c_k) L_k$. Besides, we denote the computing capability at MU k in terms of the number of CPU cycles per second as f_k^l . Since we aim to minimize the total delay of MUs and the timeliness of information processing is highly required, we take f_k^l as MU k 's maximum computing capability. Then, the data-processing time and the energy consumption of MU k in local computing are $t_k^l = L_k^l a_k / f_k^l$ and $E_{2,k} = \chi_k (f_k^l)^3 t_k^l$, respectively, where χ_k is the computation energy efficiency of the processor chip. Letting D_k^l denote the total delay of MU k in edge computing, we then have $D_k^l = t_1 + t_k^l$.

Based on the above discussion, we can write the total delay of MU k , denoted by D_k , as follows:

$$\begin{aligned} D_k &= \max \{ D_k^e, D_k^l \} \\ &= \max \{ t_1 + t_{2,k} + t_{3,k}, t_1 + t_k^l \} \\ &= \max \left\{ t_1 + \frac{c_k L_k}{R_k} + \frac{c_k L_k a_k}{f_k^e}, t_1 + \frac{(1 - c_k) L_k a_k}{f_k^l} \right\}. \end{aligned} \quad (13)$$

III. DELAY MINIMIZATION SCHEME FOR MUS USING MUD

A. Optimization Problem Formulation

In this paper, we aim to minimize the total delay of all MUs under their QoS requirements, by jointly optimizing the EH time t_1 , the data-transmission time in $\mathcal{T} \triangleq \{t_{2,k}, t_{3,k} \geq 0, \forall k \in \mathcal{K}\}$, the task allocation proportions in \mathcal{C} , the MUD coefficients in \mathcal{W} , the IRS's phase-shift matrices Φ_1 and Φ_2 , and the MEC server's computation resource allocation coefficients in $\mathcal{F} \triangleq \{f_k^e, f_k^e \in [0, f_{\max}^e], \forall k \in \mathcal{K}\}$, where f_{\max}^e is the maximum computation resource that the MEC server can allocate to MU k . Then, we can formulate the considered optimization problem as follows:

$$\begin{aligned} & \min_{t_1, \mathcal{T}, \mathcal{C}, \mathcal{F}, \mathcal{W}, \Phi_1, \Phi_2} \left\{ \sum_{k \in \mathcal{K}} \tau_k D_k \right\} \\ \text{s.t.} \quad & C1: t_1, t_{2,k}, f_k^e \geq 0, \forall k \in \mathcal{K}, \\ & C2: \sum_{k \in \mathcal{K}} f_k^e \leq f_{\text{total}}^e, \\ & C3: 0 \leq \theta_{z,n} \leq 2\pi, |e^{j\theta_{z,n}}| = 1, z = 1, 2, \forall n \in \mathcal{N}, \\ & C4: 0 \leq P_k^h \leq P_h^{\max}, \forall k \in \mathcal{K}, \\ & C5: P_k^{\min} |\mathbf{w}_k^H (\mathbf{G}_{b,i} \Phi_2 \mathbf{h}_{i,k} + \mathbf{h}_{b,k})|^2 \geq X_1, \forall k \in \mathcal{K}, \\ & C6: \sum_{j \in \mathcal{K}, j \neq k} P_k^{\max} |\mathbf{w}_k^H (\mathbf{G}_{b,i} \Phi_2 \mathbf{h}_{i,k} + \mathbf{h}_{b,j})|^2 \leq X_2, \\ & \quad \forall k \in \mathcal{K}, \end{aligned} \quad (14)$$

where τ_k is a constant, which represents the weight of MU k , P_k^{\max} and P_k^{\min} are the upper and lower bounds of MU k 's transmit power, respectively, and P_h^{\max} denotes the upper bound of the RF power received at MU k , whose specific value is related to MU k 's *non-linear* circuit, etc. Moreover, $C1$ indicates that t_1 , $t_{2,k}$, and f_k^e are non-negative. $C2$ ensures that the sum of computation resources allocated to all MUs should not be larger than the maximum computing capability f_{total}^e of the MEC server. $C3$ specifies the range of the phase shifts of the IRS. $C4$ indicates the range of P_k^h . $C5$ and $C6$ are QoS constraints of MU k . $C4$ ensures that the received signal power at the BS from MU k must be not less than the threshold X_1 , while $C6$ guarantees that the interference to MU k 's data transmission from other MUs must not be larger than the threshold X_2 .

To tackle the formulated optimization problem specified by Eq. (14), we first fix $t_1 = \bar{t}_1$, and apply the classic BCD method to decouple the optimization problem given by Eq. (14) into several subproblems, by alternatively optimizing all variables. Specifically, we firstly optimize the computation resource allocation variables in \mathcal{F} and the task allocation proportions in \mathcal{C} , for given MUD coefficients in \mathcal{W} , the IRS's phase-shift matrices Φ_1 and Φ_2 , and the transmission time in \mathcal{T} . Then, we optimize $\{\mathcal{W}, \Phi_1, \Phi_2, \mathcal{T}\}$ for given \mathcal{F} and \mathcal{C} .

B. Joint Optimization of Computation Resource Allocation \mathcal{F} and Task Allocation \mathcal{C}

We first optimize the computation resource allocation variables in \mathcal{F} and the task allocation variables in \mathcal{C} for given MUD coefficients in \mathcal{W} , IRS's phase-shift matrices Φ_1 and Φ_2 , and transmission time in \mathcal{T} . Then, we can simplify the optimization problem given by Eq. (14) as:

$$\begin{aligned} & \min_{\mathcal{F}, \mathcal{C}} \left\{ \sum_{k \in \mathcal{K}} \tau_k D_k \right\} \\ \text{s.t.} \quad & C1, C2. \end{aligned} \quad (15)$$

Relying on the aforementioned BCD technique, we can optimize \mathcal{F} and \mathcal{C} as follows:

1) *Optimization of \mathcal{C} :* We can optimize \mathcal{C} based on the following proposition similar to [24].

Proposition 1: For given \mathcal{W} , Φ_1 , Φ_2 , \mathcal{T} , and \mathcal{F} , the optimal allocation proportion c_k^* for MU k is given by:

$$c_k^* = \arg \min_{c_k} \{ D_k \}, \quad (16)$$

where $(\cdot)^*$ denotes the optimal solution or suboptimal solution of one function. Moreover, c_k^* ensures that the value of D_k^l is equivalent to that of D_k^e , i.e., $D_k^l = D_k^e$. Using Eq. (13), we can obtain that

$$c_k^* = \frac{a_k R_k f_k^e}{a_k R_k (f_k^e + f_k^l) + f_k^e f_k^l}. \quad (17)$$

Proof: Please see Appendix A.

2) *Optimization of \mathcal{F} :* We fix \mathcal{W} , Φ_1 , Φ_2 , and \mathcal{T} to optimize the computation resource allocation variables in \mathcal{F} . We plug c_k^* into the optimization problem given by Eq. (17). Since the EH time t_1 is fixed, we can reformulate the optimization problem

given by Eq. (15) as follows:

$$\begin{aligned} \min_{\mathcal{F}} \left\{ \sum_{k \in \mathcal{K}} \frac{\tau_k (a_k L_k f_k^e + L_k a_k^2 R_k)}{a_k R_k (f_k^e + f_k^L) + f_k^e f_k^L} \right\} \\ \text{s.t. } C1 - C2. \end{aligned} \quad (18)$$

Proposition 2: The optimization problem given by Eq. (18) is a convex optimization problem.

Proof: Please see Appendix B.

Since the optimization problem given by Eq. (18) is convex, it satisfies the Slater's condition [35]. Then, we can obtain the optimal solution to the optimization problem given by Eq. (18) by using the Karush-Kuhn-Tucker (KKT). Specifically, we can write the Lagrangian function of the optimization problem given by Eq. (18) as follows:

$$\begin{aligned} \mathcal{L}(f_k^e, \mu) = \sum_{k \in \mathcal{K}} \frac{\tau_k (a_k L_k f_k^e + L_k a_k^2 R_k)}{a_k R_k (f_k^e + f_k^L) + f_k^e f_k^L} \\ + \mu \left(\sum_{k \in \mathcal{K}} f_k^e - f_{\text{total}}^e \right), \end{aligned} \quad (19)$$

where the variable $\mu \geq 0$ is the Lagrange multiplier associated with $C2$. For given \mathcal{W} , Φ_1 , Φ_2 , and \mathcal{T} , the optimal computation resource $(f_k^e)^*$ allocated to MU k by the MEC server and the optimal Lagrange multiplier μ^* satisfy the following KKT conditions:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial f_k^e} = \frac{-\tau_k L_k a_k^3 R_k^2}{a_k R_k (f_k^e + f_k^L) + f_k^e f_k^L} + \mu^* = 0, \forall k \in \mathcal{K}, & (20) \\ \mu^* \left(\sum_{k \in \mathcal{K}} (f_k^e)^* - f_{\text{total}}^e \right) = 0, \forall k \in \mathcal{K}, & (21) \\ (f_k^e)^* \geq 0, \forall k \in \mathcal{K}. & (22) \end{cases}$$

The optimal Lagrange multiplier μ^* is not equal to 0, i.e., $\mu^* \neq 0$. Otherwise using Eq. (20), we can obtain $R_k = 0$, $\forall k \in \mathcal{K}$, which is contrary to Proposition 1 from which we can obtain $D_k^L = D_k^e$ and $R_k \neq 0$, $\forall k \in \mathcal{K}$. For given μ , the value of f_k^e can be directly derived from Eq. (20) and is given by:

$$f_k^e = \frac{\sqrt{\frac{\tau_k a_k^3 L_k R_k^2}{\mu} - a_k R_k f_k^L}}{f_k^L + a_k R_k}, \quad \forall k \in \mathcal{K}. \quad (23)$$

To ensure $f_k^e \geq 0$ in Eq. (23), we have $\sqrt{\frac{\tau_k a_k^3 L_k R_k^2}{\mu} - a_k R_k f_k^L} \geq 0$, and then we can obtain $\mu \leq \frac{\tau_k a_k L_k}{(f_k^L)^2}$. Since $\mu \neq 0$ in Eq. (23), the optimal μ^* belongs to the interval $(\mu_{\min}, \mu_{\max}] = \left(0, \min_k \left\{ \frac{\tau_k a_k L_k}{(f_k^L)^2} \right\} \right]$. Since f_k^e is monotonically decreasing with respect to (w.r.t.) μ , we can use the well-known bisection search method (BSM) to derive μ^* , and then derive $(f_k^e)^*$ and c_k^* , $\forall k \in \mathcal{K}$, for given \mathcal{W} , Φ_1 , Φ_2 , and \mathcal{T} . We summarize the algorithm for solving the optimization problem specified by Eq. (15) in **BSM-CRA/TA**, where the BSM is used to determine the computation resource allocation variables in \mathcal{F} and the task allocation variables in \mathcal{C} . Moreover, in **BSM-CRA/TA**, ρ denotes the iteration index and $\text{obj}(f_k^e)$ denotes the objective function of the optimization problem given by Eq. (18).

BSM-CRA/TA: Joint Optimization of \mathcal{F} and \mathcal{C} for given Φ_1 , Φ_2 , \mathcal{W} , and \mathcal{T}

- 1 **Input:** The EH time $t_1 = \bar{t}_1$, the phase-shift matrices Φ_1 and Φ_2 , MUD coefficients in \mathcal{W} , the transmission time in \mathcal{T} , and the termination coefficient ϵ .
 - 2 **Output:** \mathcal{F}^* and \mathcal{C}^* for given Φ_1 , Φ_2 , \mathcal{W} , and \mathcal{T} ;
 - 3 **Initialization**
 - 4 Initialize: $\rho \leftarrow 0$, (f_k^e) , μ_{\min} , and μ_{\max} ;
 - 5 **Repeat**
 - 6 Calculate $c_k^{(\rho)}$ by using Eq. (17) for given $(f_k^e)^{(\rho)}$;
 - 7 $\mu^{(\rho)} = \frac{\mu_{\min} + \mu_{\max}}{2}$;
 - 8 Calculate $(f_k^e)^{(\rho+1)}$ by using Eq. (23) for given $\mu^{(\rho)}$;
 - 9 **if** $\text{obj}((f_k^e)^{(\rho+1)}) - \text{obj}((f_k^e)^{(\rho)}) \leq 0$;
 - 10 $\mu_{\max} = \mu^{(\rho)}$;
 - 11 **else**
 - 12 $\mu_{\min} = \mu^{(\rho)}$;
 - 13 **end**
 - 14 $\rho = \rho + 1$;
 - 15 **Until** The optimal objective value of the optimization problem given by Eq. (15) converges;
 - 16 **Return** \mathcal{F}^* and \mathcal{C}^* for given Φ_1 , Φ_2 , \mathcal{W} , and \mathcal{T} .
-

C. Joint Optimization of MUD Coefficients in \mathcal{W} , IRS Phase-Shift Matrices Φ_1 and Φ_2 , and Transmission Time in \mathcal{T}

Under the independence assumption of s_k and n_2 [20], [36], we can write the mean-square error (MSE) of MU k , denoted by e_k , $k \in \mathcal{K}$, as follows:

$$\begin{aligned} e_k &= \mathbb{E}\{|y_{k,b} - s_k|^2\} \\ &= \left| \mathbf{w}_k^H (\mathbf{G}_{b,i} \Phi_2 \mathbf{h}_{i,k} + \mathbf{h}_{b,k}) \sqrt{P_k^t} - 1 \right|^2 \\ &\quad + \sum_{j \in \mathcal{K}, j \neq k} \left| \mathbf{w}_k^H (\mathbf{G}_{b,i} \Phi_2 \mathbf{h}_{i,j} + \mathbf{h}_{b,j}) \right|^2 P_j^t \\ &\quad + \sigma_b^2 \left| \mathbf{w}_k^H \right|^2. \end{aligned} \quad (24)$$

Moreover, based on Eq. (13), to minimize D_k for given t_1 , \mathcal{C} , and \mathcal{F} , we only need to maximize R_k . Then, we can convert the optimization problem given by Eq. (14) into the following problem:

$$\begin{aligned} \max_{\Phi_1, \Phi_2, \mathcal{W}, \mathcal{T}} \left\{ \sum_{k \in \mathcal{K}} R_k \right\} \\ \text{s.t. } C1 - C6. \end{aligned} \quad (25)$$

Based on the above introduced e_k and the problem given by Eq. (25), we first give the following Proposition 3.

Proposition 3: By using the MSE method, the maximization optimization problem given by Eq. (25) is equivalent to the following optimization problem:

$$\begin{aligned} \min_{\Phi_1, \Phi_2, \mathcal{W}, \mathcal{T}, \mathcal{Q}} \left\{ \sum_{k \in \mathcal{K}} [q_k e_k - \log q_k - 1] \right\} \\ \text{s.t. } C1, C3 - C6, \end{aligned} \quad (26)$$

where q_k is a positive auxiliary variable and $\mathcal{Q} \triangleq \{q_k, q_k \geq$

$0, \forall k \in \mathcal{K}$).

Proof: Please see [20] and references therein.

With the above transformation, we can use the BCD method to solve the optimization problem given by Eq. (26) to obtain \mathcal{Q} , \mathcal{T} , \mathcal{W} , Φ_1 , and Φ_2 .

1) *Optimization of the transmission time \mathcal{T} for given $\{\mathcal{W}, \Phi_1, \Phi_2, \mathcal{C}, \mathcal{F}\}$:* we first optimize \mathcal{T} and the auxiliary variables in \mathcal{Q} for given \mathcal{W} , Φ_1 , Φ_2 , \mathcal{C} , and \mathcal{F} . For given \mathcal{W} , Φ_1 , Φ_2 , \mathcal{C} , and \mathcal{F} , we can re-write the optimization problem given by Eq. (26) as the following equivalent optimization problem:

$$\begin{aligned} \min_{\mathcal{T}, \mathcal{Q}} \left\{ \sum_{k \in \mathcal{K}} [q_k e_k - \log q_k - 1] \right\} \\ \text{s.t. } C1, \end{aligned} \quad (27)$$

where we re-write e_k defined by Eq. (24) as follows:

$$\begin{aligned} e_k &= \left| \mathbf{w}_k^H \mathbf{b}_k \sqrt{P_k^t} - 1 \right|^2 + \sum_{j \in \mathcal{K}, j \neq k} |\mathbf{w}_k^H \mathbf{b}_j|^2 P_j^t + \sigma_b^2 |\mathbf{w}_k^H|^2 \\ &= \sum_{j \in \mathcal{K}} |\mathbf{w}_k^H \mathbf{b}_j|^2 P_j^t - 2 \mathbf{w}_k^H \mathbf{b}_k \sqrt{P_k^t} + 1 + \sigma_b^2 |\mathbf{w}_k^H|^2, \end{aligned} \quad (28)$$

where $\mathbf{b}_k \triangleq \mathbf{G}_{b,i} \Phi_2 \mathbf{h}_{i,k} + \mathbf{h}_{b,k}$, $\forall k \in \mathcal{K}$. Then, the objective function of the optimization problem given by Eq. (27) is convex w.r.t. each of the optimization variables in \mathcal{T} and \mathcal{Q} . For simplicity of illustration, we denote the objective function of the problem given by Eq. (27) as follows:

$$\mathcal{H}(\mathcal{T}, \mathcal{Q}) \triangleq \sum_{k \in \mathcal{K}} [q_k e_k - \log q_k - 1]. \quad (29)$$

The function $\mathcal{H}(\mathcal{T}, \mathcal{Q})$ given by Eq. (29) is continuously differentiable. Then, by checking the first-order optimality conditions for minimizing the objective function w.r.t. q_k and $t_{2,k}$, respectively, we can obtain that

$$\begin{cases} \frac{\partial \mathcal{H}}{\partial q_k} = e_k - \frac{1}{q_k}, \\ \frac{\partial \mathcal{H}}{\partial t_{2,k}} = q_k \left(\mathbf{w}_k^H \mathbf{b}_k t_{2,k}^{-\frac{3}{2}} \sqrt{E_{3,k}} - \sum_{j \in \mathcal{K}} |\mathbf{w}_k^H \mathbf{b}_j|^2 \frac{E_{3,k}}{t_{2,k}^2} \right), \end{cases}$$

where $E_{3,k} \triangleq E_k^h - E_{2,k}$, $\forall k \in \mathcal{K}$. Letting $\frac{\partial \mathcal{H}}{\partial q_k} = 0$ and $\frac{\partial \mathcal{H}}{\partial t_{2,k}} = 0$, we can then obtain the optimal q_k^* and $t_{2,k}^*$, $k \in \mathcal{K}$, for given \mathcal{W} , Φ_1 , Φ_2 , \mathcal{C} , \mathcal{F} , which are given by:

$$\begin{cases} q_k^* = \frac{1}{e_k}, \quad \forall k \in \mathcal{K}, \\ t_{2,k}^* = \left(\frac{\sum_{j \in \mathcal{K}} \sqrt{E_{3,k}} |\mathbf{w}_k^H \mathbf{b}_j|^2}{\mathbf{w}_k^H \mathbf{b}_k} \right)^2, \quad \forall k \in \mathcal{K}. \end{cases} \quad (30)$$

2) *Optimization of the MUD coefficients in \mathcal{W} for given $\{\mathcal{T}, \Phi_1, \Phi_2, \mathcal{C}, \mathcal{F}\}$:* Here, we optimize \mathcal{W} for given \mathcal{Q} , \mathcal{T} , Φ_1 , Φ_2 , \mathcal{C} , and \mathcal{F} . For given \mathcal{Q} , \mathcal{T} , Φ_1 , Φ_2 , \mathcal{C} , and \mathcal{F} , we can

re-write the optimization problem given by Eq. (26) as follows:

$$\begin{aligned} \min_{\mathcal{W}} \left\{ \sum_{k \in \mathcal{K}} [q_k e_k - \log q_k - 1] \right\} \\ \text{s.t. } C7: P_k^{\min} |\mathbf{w}_k^H \mathbf{b}_k|^2 \geq X_1, \quad \forall k \in \mathcal{K}, \\ C8: \sum_{j \in \mathcal{K}, j \neq k} P_k^{\max} |\mathbf{w}_k^H \mathbf{b}_j|^2 \leq X_2, \quad \forall k \in \mathcal{K}, \end{aligned} \quad (32)$$

where $C7$ - $C8$ are derived from $C5$ - $C6$, respectively, based on the definition of $\mathbf{b}_k \triangleq \mathbf{G}_{b,i} \Phi_2 \mathbf{h}_{i,k} + \mathbf{h}_{b,k}$. Since $C7$ is a non-convex constraint, the optimization problem given by Eq. (32) is non-convex, which is difficult to solve. Therefore, we use the first-order Taylor expansion to convert $C7$ into a linear constraint, and use the successive convex approximation (SCA) method to solve the problem given by Eq. (32). Converting $C7$ into a linear constraint around $\mathcal{W}^{(l-1)} \triangleq \{\mathbf{w}_k^{(l-1)}, \mathbf{w}_k^{(l-1)} \in \mathcal{C}^{M \times 1}, \forall k \in \mathcal{K}\}$, we can re-write the optimization problem given by Eq. (32) as follows:

$$\begin{aligned} \min_{\mathcal{W}} \left\{ \sum_{k \in \mathcal{K}} [q_k e_k - \log q_k - 1] \right\} \\ \text{s.t. } C8, \\ C9: P_k^{\min} \left| \left(\mathbf{w}_k^{(l-1)} \right)^H \mathbf{b}_k \right|^2 - \nu_k \left(\mathbf{w}_k - \mathbf{w}_k^{(l-1)} \right) \geq X_1, \\ \quad \forall k \in \mathcal{K}, \end{aligned} \quad (33)$$

where $\mathcal{W}^{(l-1)}$ and $\mathbf{w}_k^{(l-1)}$ denote the values of \mathcal{W} and \mathbf{w}_k , respectively, in the $(l-1)$ -th iteration of the SCA method, and $\nu_k \in \mathcal{C}^{1 \times M}$ is given by:

$$\nu_k = 2 \left| \left(\mathbf{w}_k^{(l-1)} \right)^H \mathbf{b}_k \right| \mathbf{b}_k^H.$$

Moreover, $C9$ is derived from $C7$ by using the first-order Taylor expansion to the left-hand side of $C7$ around $\mathcal{W}^{(l-1)}$. Then, the optimization problem given by Eq. (33) is convex w.r.t. \mathcal{W} around $\mathcal{W}^{(l-1)}$, and thus it can be efficiently solved by using CVX.

3) *Optimization of the IRS's phase-shift matrices $\{\Phi_1, \Phi_2\}$ for given $\{\mathcal{C}, \mathcal{W}, \mathcal{T}, \mathcal{F}\}$:* Now, we optimize the IRS phase-shift matrices Φ_1 and Φ_2 by solving the optimization problem given by Eq. (26) for given \mathcal{Q} , \mathcal{C} , \mathcal{W} , \mathcal{T} , and \mathcal{F} . However, since P_k^t is a non-convex function w.r.t. Φ_1 , analyzing Eq. (11), we can know that the $N \times N$ complex-valued matrices Φ_1 and Φ_2 are tightly coupled. Moreover, $C3$ - $C5$ are non-convex constraints w.r.t. Φ_1 and Φ_2 . Hence, it is challenging to directly obtain Φ_1 and Φ_2 . For ease of calculation, we define $v_{1,n} \triangleq e^{j\theta_{1,n}}$ and $v_{2,n} \triangleq e^{j\theta_{2,n}}$, $\forall n \in \mathcal{N}$, which satisfy $|v_{1,n}| = 1$ and $|v_{2,n}| = 1$. Then, the IRS's phase-shift matrices Φ_1 and Φ_2 can be re-written as $\Phi_1 = \text{diag}(v_{1,1}, \dots, v_{1,N})$ and $\Phi_2 = \text{diag}(v_{2,1}, \dots, v_{2,N})$, respectively. Besides, we denote $\mathbf{v}_z \triangleq [v_{z,1}, \dots, v_{z,n}, \dots, v_{z,N}]^T$, $z = 1, 2$, and $\mathbf{H}_k \triangleq \mathbf{G}_{b,i} \text{diag}(\mathbf{h}_{i,k}^H) \in \mathcal{C}^{M \times N}$, $\forall k \in \mathcal{K}$. Then, we can re-write $\mathbf{G}_{b,i} \Phi_1 \mathbf{h}_{i,k} + \mathbf{h}_{b,k}$ in Eq. (4) and $\mathbf{G}_{b,i} \Phi_2 \mathbf{h}_{i,k} + \mathbf{h}_{b,k}$ in Eq. (11),

based on which we can obtain the suboptimal Φ_1^* and Φ_2^* for given \mathcal{C} , \mathcal{W} , \mathcal{T} , and \mathcal{F} .

We summarize the algorithm to solve the optimization problem given by Eq. (14) in **IRS-DT/EH**, where the IRS is employed to help both data transmission and energy harvesting. In **IRS-DT/EH**, we use $\tilde{\rho}$ to denote the iteration index. We decompose the optimization problem given by Eq. (14) into two optimization sub-problems given by Eq. (15) and Eq. (26), respectively. The optimization problem given by Eq. (26) is further divided into three optimization sub-problems given by Eq. (27), Eq. (33), and Eq. (41), respectively. Specifically, given $t_1 = \bar{t}_1$, the algorithm starts with certain feasible values of $\Phi_1^{(0)}$, $\mathcal{W}^{(0)}$, $\mathcal{T}^{(0)}$, and $\Phi_2^{(0)}$. Given a fixed solution $\{\Phi_1^{(\tilde{\rho})}, \Phi_2^{(\tilde{\rho})}, \mathcal{W}^{(\tilde{\rho})}, \mathcal{T}^{(\tilde{\rho})}\}$ in the $\tilde{\rho}$ -th iteration, we can obtain the optimal values $\mathcal{C}^{(\tilde{\rho}+1)}$ and $\mathcal{F}^{(\tilde{\rho}+1)}$ by using **BSM-CRA/TA**. Based on the obtained $\mathcal{C}^{(\tilde{\rho}+1)}$ and $\mathcal{F}^{(\tilde{\rho}+1)}$, we obtain $\mathcal{T}^{(\tilde{\rho}+1)}$ for given $\{\mathcal{W}^{(\tilde{\rho})}, \Phi_1^{(\tilde{\rho})}, \Phi_2^{(\tilde{\rho})}, \mathcal{C}^{(\tilde{\rho}+1)}, \mathcal{F}^{(\tilde{\rho}+1)}\}$, and then we can obtain $\mathcal{W}^{(\tilde{\rho}+1)}$ for given $\{\Phi_1^{(\tilde{\rho})}, \Phi_2^{(\tilde{\rho})}, \mathcal{C}^{(\tilde{\rho}+1)}, \mathcal{F}^{(\tilde{\rho}+1)}, \mathcal{T}^{(\tilde{\rho}+1)}\}$. In particular, we update the phase-shift matrices $\Phi_1^{(\tilde{\rho}+1)}$ and $\Phi_2^{(\tilde{\rho}+1)}$ by solving the problem given by Eq. (41) using the SDP approach in the $(\tilde{\rho} + 1)$ -th iteration. Finally, we obtain the optimal EH time t_1 that minimizes the objective function of the problem given by Eq. (14) via a simple one-dimensional search method over the interval $(0, T)$, e.g., the golden-section search or the data-driven-based search [38], where T is an upper-bound of t_1 .

IRS-DT/EH: Joint Optimization of Φ_1 , Φ_2 , \mathcal{F} , \mathcal{W} , \mathcal{C} , and \mathcal{T} for given $t_1 = \bar{t}_1$

- 1 **Input:** The EH time $t_1 = \bar{t}_1$, the weight value τ_k , the channel coefficients $\mathbf{G}_{b,i}$, $\mathbf{h}_{i,k}$, and $\mathbf{h}_{b,k}$.
 - 2 **Output:** The solution $\{\Phi_1, \Phi_2, \mathcal{F}, \mathcal{T}, \mathcal{W}, \mathcal{C}\}$ to the problem given by Eq. (14) for given $t_1 = \bar{t}_1$;
 - 3 **Initialization**
 - 4 Initialize: $\tilde{\rho} \leftarrow 0$, $\Phi_1^{(0)}$, $\Phi_2^{(0)}$, $\mathbf{w}_k^{(0)}$, $t_{2,k}^{(0)}$, and $(f_k^e)^{(0)}$;
 - 5 **Repeat**
 - 6 Calculate $c_k^{(\tilde{\rho}+1)}$ and $(f_k^e)^{(\tilde{\rho}+1)}$ using **BSM-CRA/TA**;
 - 7 Calculate $t_{2,k}^{(\tilde{\rho}+1)}$ and $q_k^{(\tilde{\rho}+1)}$ for given $c_k^{(\tilde{\rho}+1)}$, $(f_k^e)^{(\tilde{\rho}+1)}$, $\mathbf{w}_k^{(\tilde{\rho})}$, $\Phi_1^{(\tilde{\rho})}$, and $\Phi_2^{(\tilde{\rho})}$;
 - 8 Calculate $\mathbf{w}_k^{(\tilde{\rho}+1)}$ for given $c_k^{(\tilde{\rho}+1)}$, $(f_k^e)^{(\tilde{\rho}+1)}$, $t_{2,k}^{(\tilde{\rho}+1)}$, $\Phi_1^{(\tilde{\rho})}$, and $\Phi_2^{(\tilde{\rho})}$;
 - 9 Calculate $\mathbf{V}_1^{(\tilde{\rho}+1)}$ and $\mathbf{V}_2^{(\tilde{\rho}+1)}$ by solving the optimization problem given by Eq. (41);
 - 10 Calculate $\Phi_1^{(\tilde{\rho}+1)}$ and $\Phi_2^{(\tilde{\rho}+1)}$ using $\mathbf{V}_1^{(\tilde{\rho}+1)}$ and $\mathbf{V}_2^{(\tilde{\rho}+1)}$;
 - 11 $\tilde{\rho} = \tilde{\rho} + 1$;
 - 12 **Until** The objective value of the optimization problem given by Eq. (14) for given $t_1 = \bar{t}_1$ converges;
 - 13 **Return** The obtained solution $\{\Phi_1, \Phi_2, \mathcal{F}, \mathcal{T}, \mathcal{W}, \mathcal{C}\}$ to the problem given by Eq. (14) for given $t_1 = \bar{t}_1$.
-

D. Convergence and Complexity Analyses

In **IRS-DT/EH**, we first optimize $\{\mathcal{C}, \mathcal{F}\}$ for given $\{\mathcal{T}, \mathcal{W}, \Phi_1, \Phi_2\}$ by solving the problem given by Eq. (15)

using **BSM-CRA/TA**, and then optimize $\{\mathcal{T}, \mathcal{W}, \Phi_1, \Phi_2\}$ for given $\{\mathcal{C}, \mathcal{F}\}$ by solving the problem given by Eq. (26). In **BSM-CRA/TA**, we need to solve the problem given by Eq. (18) to obtain \mathcal{F} for given $\{\mathcal{T}, \mathcal{W}, \Phi_1, \Phi_2\}$. According to [24], since the problem given by Eq. (18) is a minimization problem which is convex and the obtained value of its objective function is bounded, the obtained $\{\mathcal{F}^{(\tilde{\rho})}\}$ is a monotone decreasing sequence and can converge to the optimal \mathcal{F}^* for given $\{\mathcal{T}, \mathcal{W}, \Phi_1, \Phi_2\}$. Then, the optimal \mathcal{C}^* for given $\{\mathcal{T}, \mathcal{W}, \Phi_1, \Phi_2\}$ can be obtained by using Eq. (17). Therefore, **BSM-CRA/TA** can converge. Similarly, since the optimization problem given by Eq. (26) is a minimization problem, and the obtained value of the objective function decreases as the iteration number of **IRS-DT/EH**, i.e., $\tilde{\rho}$, increases following Theorem 3 in [36]. Moreover, due to C1 and C3 – C6, the objective function of the problem given by Eq. (26) is bounded. Therefore, we can conclude that the proposed **IRS-DT/EH** converges.

We assume that the iteration numbers of **BSM-CRA/TA** and **IRS-DT/EH** are I_a and I_b , respectively. The computational complexity of **BSM-CRA/TA** mainly lies in calculating f_k^e and μ by using Eq. (23) and the bisection search method, respectively. Then, the total computational complexity of **BSM-CRA/TA** is $\mathcal{O}(I_a \log_2(\frac{\mu_{\max} - \mu_{\min}}{\epsilon}) K)$ [24], where ϵ is a termination coefficient, and μ_{\max} and μ_{\min} can be obtained through **BSM-CRA/TA**. We assume that the iteration numbers of the SCA for solving the optimization sub-problems given by Eqs. (33) and (41) are I_c and I_d , respectively. Since the computational complexities of the optimization sub-problems given by Eqs. (27) and (33) are proportional to the number of variables, their computational complexities are $\mathcal{O}(2K)$ and $\mathcal{O}(I_c MK)$, respectively [39]. The SDP problem given by Eq. (41) can be solved with a worst-case computational complexity of $\mathcal{O}(I_d(N+1)^{4.5})$ [20]. Since the number of BS's antennas is not less than 2, i.e., $M \geq 2$, the total computational complexity of **IRS-DT/EH** is $\mathcal{O}(I_b((I_a \log_2(\frac{\mu_{\max} - \mu_{\min}}{\epsilon}) K) + 2K + I_c MK + I_d(N+1)^{4.5}))$.

IV. DELAY MINIMIZATION SCHEME FOR MUS UNDER THZ COMMUNICATION

Because of the large bandwidth, Terahertz (THz) communication can be used to support ultra-high data transmission [40]. However, due to the features of Ultra High Frequency (UHF), THz waves experience severe signal attenuations in territorial communications and will lose a lot of energy when meeting obstructions, which makes it challenging to efficiently use THz communication. Fortunately, the above issue can be solved by using IRSs, which can significantly enhance the strength of THz signals by improving wireless propagation environments and changing THz signals' transmission direction to bypass obstacles [41]–[42].

Consequently, we extend our work to IRS-aided MEC over THz wideband mobile networks with *non-linear* EH, and study the MUS' delay minimization problem. Since THz communication operates over broadband channels, unlike the problem given by Eq. (14), we divide system channel into K orthogonal channels to combat the frequency-selective fading similar to [43]. Each of the K channels can only be utilized by one of

the K MUs, and then there is no interference among all the K MUs. Hence, unlike in the problem specified by Eq. (14), the BS need not employ the MUD technique to reduce the interference among MUs. Thus, the signal received at the BS from MU k is given by:

$$y_{k,b} = (\mathbf{G}_{b,i}\Phi_2\mathbf{h}_{i,k} + \mathbf{h}_{b,k})\sqrt{P_k^t s_k} + n_b. \quad (42)$$

Then, the received SINR at the BS is

$$\gamma_k = \frac{P_k^t |(\mathbf{G}_{b,i}\Phi_2\mathbf{h}_{i,k} + \mathbf{h}_{b,k})|^2}{\sigma_b^2}. \quad (43)$$

The bandwidth of the system is B . We denote the ratio of channel bandwidth allocated to MU k as ε_k , and define $\varepsilon \triangleq \{\varepsilon_k, \varepsilon_k \in [0, 1], \forall k \in \mathcal{K}\}$, where $\sum_{k \in \mathcal{K}} \varepsilon_k = 1$. Then, the channel bandwidth of MU k is $\varepsilon_k B$. Therefore, the achievable information rate from MU k to the BS is

$$R_k = \varepsilon_k B \log_2(1 + \gamma_k). \quad (44)$$

In this section, we consider MUs' delay minimization for IRS-aided MEC over THz wideband mobile networks with *non-linear* EH model, where an IRS is used to improve both the EH efficiency and data transmission rates between multiple MUs and a multi-antenna BS equipped with an MEC server. Like the optimization problem given by Eq. (14), we can formulate the optimization problem for THz communication as follows:

$$\begin{aligned} & \min_{\Phi_1, \Phi_2, t_1, \mathcal{T}, \mathcal{F}, \mathcal{C}, \varepsilon} \left\{ \sum_{k \in \mathcal{K}} \tau_k D_k \right\} \\ \text{s.t. } & C1 - C4, \\ & C16 : \sum_{k \in \mathcal{K}} \varepsilon_k = 1, \forall k \in \mathcal{K}, \\ & C17 : P_k^{\min} |(\mathbf{G}_{b,i}\Phi_2\mathbf{h}_{i,k} + \mathbf{h}_{b,k})|^2 \geq X_1, \forall k \in \mathcal{K}. \end{aligned} \quad (45)$$

As mentioned above, unlike the optimization problem given by Eq. (14) in Section III, we apply the FDMA technique to divide the considered THz channel into K orthogonal channels, so that we do not have to consider the interference among MUs. Therefore, unlike the problem specified by Eq. (14), we need not consider $C6$ in the problem specified by Eq. (45). Also, we need not consider the optimization of the MUD coefficients at the BS, and instead we optimize the channel allocation coefficient $\varepsilon_k, \forall k \in \mathcal{K}$, for each MU. Since the variables are coupled together, similar to the problem given by Eq. (14), we still use the BCD method to solve the problem given by Eq. (45) by dividing all variables into the two sets $\{\mathcal{F}, \mathcal{C}\}$ and $\{\Phi_1, \Phi_2, \mathcal{T}, \varepsilon\}$ and optimizing them iteratively. Because $C16$ related to ε_k is linear and there is no interference among MUs, the algorithm complexity does not increase.

We summarize the algorithm for solving the optimization problem given by Eq. (45) in **IRS-DT/EH-THz**, where the IRS is utilized to help both data transmission and energy harvesting for THz communication. Moreover, we use $\hat{\rho}$ to denote the iteration index of **IRS-DT/EH-THz**.

V. PERFORMANCE EVALUATIONS

In this section, we evaluate the performance of our proposed **IRS-DT/EH** and **IRS-DT/EH-THz**. As for the transmission

IRS-DT/EH-THz: Joint Optimization of $\Phi_1, \Phi_2, \mathcal{F}, \varepsilon, \mathcal{C}$, and \mathcal{T} for given $t_1 = \bar{t}_1$

- 1 **Input:** The EH time $t_1 = \bar{t}_1$, the weight value τ_k , the channel coefficients $\mathbf{G}_{b,i}$, $\mathbf{h}_{i,k}$, and $\mathbf{h}_{b,k}$.
 - 2 **Output:** The solution $\{\Phi_1, \Phi_2, \mathcal{F}, \mathcal{T}, \varepsilon, \mathcal{C}\}$ to the problem given by Eq. (45) for given $t_1 = \bar{t}_1$;
 - 3 **Initialization**
 - 4 Initialize: $\hat{\rho} \leftarrow 0, \Phi_1^{(0)}, \Phi_2^{(0)}, \varepsilon_k^{(0)}, t_{2,k}^{(0)}$, and $(f_k^e)^{(0)}$;
 - 5 **Repeat**
 - 6 Calculate $c_k^{(\hat{\rho}+1)}$ and $(f_k^e)^{(\hat{\rho}+1)}$;
 - 7 Calculate $t_{2,k}^{(\hat{\rho}+1)}$ for given $\varepsilon_k^{(\hat{\rho})}, \Phi_1^{(\hat{\rho})}$, and $\Phi_2^{(\hat{\rho})}$;
 - 8 Calculate $q_k^{(\hat{\rho})}$ for given $\varepsilon_k^{(\hat{\rho})}, \Phi_1^{(\hat{\rho})}$, and $\Phi_2^{(\hat{\rho})}$;
 - 9 Calculate $\varepsilon_k^{(\hat{\rho}+1)}$ for given $\Phi_1^{(\hat{\rho})}, \Phi_2^{(\hat{\rho})}$, and $t_{2,k}^{(\hat{\rho}+1)}$;
 - 10 Calculate $q_k^{(\hat{\rho}+1)}$ for given $\Phi_1^{(\hat{\rho})}, \Phi_2^{(\hat{\rho})}, \varepsilon_k^{(\hat{\rho}+1)}$, and $t_{2,k}^{(\hat{\rho}+1)}$;
 - 11 Calculate $\mathbf{V}_1^{(\hat{\rho}+1)}$ and $\mathbf{V}_2^{(\hat{\rho}+1)}$ by solving SDP for given $\varepsilon_k^{(\hat{\rho}+1)}$ and $t_{2,k}^{(\hat{\rho}+1)}$;
 - 12 Calculate $\Phi_1^{(\hat{\rho}+1)}$ and $\Phi_2^{(\hat{\rho}+1)}$ using $\mathbf{V}_1^{(\hat{\rho}+1)}$ and $\mathbf{V}_2^{(\hat{\rho}+1)}$;
 - 13 $\hat{\rho} = \hat{\rho} + 1$;
 - 14 **Until** The optimal objective value of the optimization problem given by Eq. (45) converges;
 - 15 **Return** The obtained solution $\{\Phi_1, \Phi_2, \mathcal{F}, \mathcal{T}, \varepsilon, \mathcal{C}\}$ to the problem given by Eq. (45) for given $t_1 = \bar{t}_1$.
-

channel, we consider both the small scale fading and the large scale path loss. Specifically, like [17], the small scale fading is i.i.d. and follows the complex Gaussian distribution with zero mean and unit variance, and the path loss in dB is given by:

$$\text{PL} = \text{PL}_0 - 10\alpha \log_{10} \left(\frac{d}{d_0} \right)$$

where PL_0 is the path loss at the reference distance d_0 ,

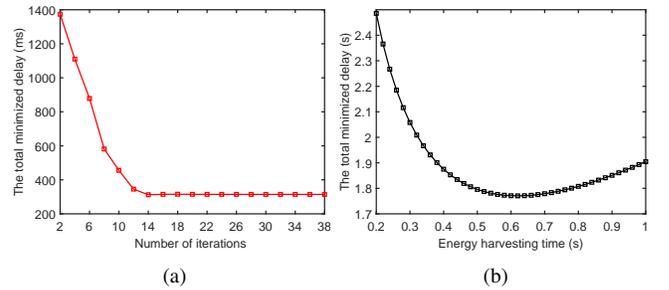


Fig. 2. Performance of our proposed schemes with *non-linear* EH model. (a) Convergence performance of our proposed **IRS-DT/EH**; (b) The total minimized delay versus the EH time of **IRS-DT/EH**.

and d and α represent the length of the communication link and the path loss exponent, respectively. We set $\text{PL}_0 = 30$ dB, $d_0 = 1$ m, and use $\alpha_{b,k} = 3.5, \alpha_{b,i} = 2.2$, and $\alpha_{i,k} = 2.2$ to denote the path loss exponents of the link between the BS and MU k , the link between the BS and the IRS, and the link between the IRS and MU k , respectively. Besides, we use $d_{b,k}, d_{b,i}$, and $d_{i,k}$ to denote the average distances from

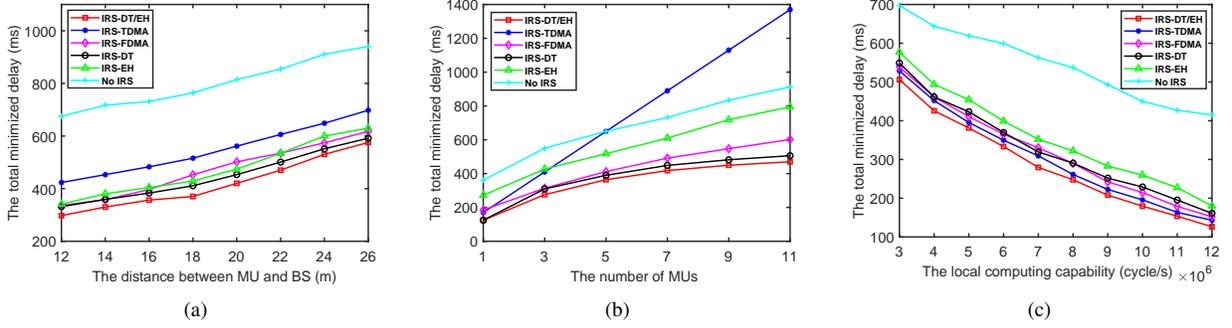


Fig. 3. The total minimized delay versus the average distance between the BS and MUs, the number of MUs, and the local computing capability of each MU. (a) The total minimized delay versus the average distance between the BS and MUs; (b) The total minimized delay versus the number of MUs; (c) The total minimized delay versus the local computing capability of each MU.

the BS to MUs, from the BS to the IRS, and from the IRS to MUs, respectively. Without loss of generality, we assume $d = d_{b,k} = d_{b,i} = d_{i,k} = 10$ m. Moreover, we take the frequency of the local computing $f_k = 5 \times 10^8$ Hz, the variance of noise $\sigma_b^2 = 3.98 \times 10^{-14}$, the total computing capability of the MEC server $f_{\text{total}}^e = 5 \times 10^9$ cycle/s, the QoS thresholds $X_1 = 1$ and $X_2 = 4$, and the number of reflecting elements $N = 100$.

Furthermore, for performance comparisons, we also consider the following 10 baseline schemes (the scheme names in the bold-face fonts are referring to our developed schemes):

- **IRS-TDMA**: Similar to [30], all MUs transmit information based on time division multiple access (TDMA), and the IRS is employed to help both EH and data transmission of MUs.
- **IRS-FDMA**: This scheme is similar to IRS-TDMA except that all MUs transmit data based on FDMA as in [2].
- **IRS-DT**: All MUs transmit data to the BS simultaneously, while the IRS is only used to help data transmission of MUs as in [24].
- **IRS-EH**: This scheme is similar to IRS-DT except that the IRS is only used to help EH of MUs as in [18].
- **No IRS**: This scheme is also similar to IRS-DT, while no IRS is used to help EH and data transmission of MUs.
- **IRS-DT-THz**: All MUs communicate with the BS over THz channel based on FDMA similar to our proposed **IRS-DT/EH-THz**, while the IRS is only used to help data transmission of MUs similar to [42].
- **IRS-EH-THz**: This scheme is similar to IRS-DT-THz except that the IRS is only used to help EH of MUs as in [18].
- **No IRS-THz**: This scheme is also similar to IRS-DT-THz, while no IRS is utilized to help EH and data transmission of MUs.
- **IRS-NOMA**: This scheme is similar to our proposed **IRS-DT/EH**, except that all MUs transmit data to the BS based on non-orthogonal multiple access (NOMA).
- **IRS-RP**: This scheme is also similar to our proposed **IRS-DT/EH**, except that we use the relaxation and projection (RP) method [44] to optimize the IRS's phase shifts.

Fig. 2(a) shows the convergence performance of our proposed **IRS-DT/EH**, where we set the number of IRS's reflecting

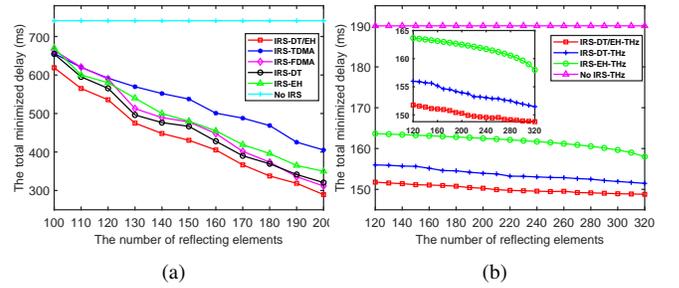


Fig. 4. The total minimized delay versus the number of the IRS's reflecting elements, i.e., N . (a) The total minimized delay caused by **IRS-DT/EH** and baseline schemes versus N ; (b) The total minimized delay caused by **IRS-DT/EH-THz** and baseline schemes versus N for THz communication.

elements as $N = 100$. It can be observed that our proposed **IRS-DT/EH** has a good convergence performance, which can converge within about 14 iterations. Moreover, Fig. 2(b) shows the total minimized delay of all MUs versus the EH time t_1 for our proposed **IRS-DT/EH**. Analyzing Fig. 2(b), we can observe that the total delay caused by **IRS-DT/EH** decreases when the EH time increases from $t_1 = 0.2$ s to $t_1 = 0.6$ s, and then increases when $t_1 > 0.6$ s. This is because when the EH time t_1 becomes too large, for MUs, the amount of energy harvested is larger than the amount of energy consumed, which shows that it is necessary to optimize t_1 so as to minimize the total delay of MUs.

Figs. 3(a)-3(c) show the total minimized delay of all MUs versus the average distance, i.e., $d_{b,k}$, between MUs and the BS, the number of MUs, i.e., K , and the local computing capability of each MU, i.e., f_k^L , respectively. First, it can be seen that the total delay posed by our proposed **IRS-DT/EH** is much smaller than that caused by IRS-TDMA and IRS-FDMA, which are similar to our **IRS-DT/EH** while MUs transmit information to the BS based on TDMA and FDMA, respectively. The reason for this is that for our **IRS-DT/EH**, the interference among MUs can be efficiently reduced by optimizing the MUD coefficients even all MUs transmit information simultaneously over the same channel. Then, system spectrum efficiency can be significantly improved as compared with IRS-TDMA and IRS-FDMA especially when K takes a large value. Moreover, we can observe that the total delay posed by our **IRS-DT/EH** is also smaller or significantly smaller than that caused by

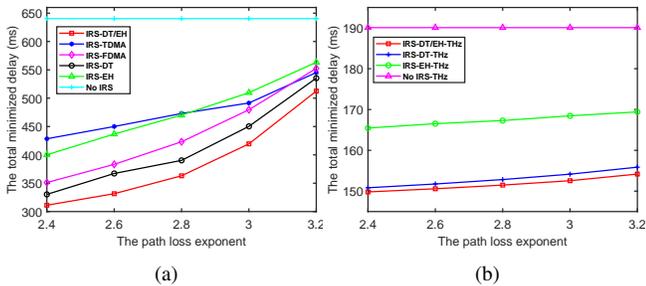


Fig. 5. The total minimized delay versus the path loss exponent $\alpha_{i,k}$. (a) The total minimized delay posed by **IRS-DT/EH** and baseline schemes versus $\alpha_{i,k}$; (b) The total minimized delay posed by **IRS-DT/EH-THz** and baseline schemes versus $\alpha_{i,k}$ for THz communication.

the schemes IRS-DT, IRS-EH, and No IRS, where the IRS is only used to help data/energy transmission or no IRS is utilized. This indicates that IRS can efficiently improve both data transmission and EH efficiencies, by enhancing wireless propagation environments. In addition, analyzing Fig 3(c), we can observe that the minimized delay by each scheme decreases with the increase of f_k^L , since more and more data can be computed locally at MUs instead of being transmitted to the MEC server with a long period of transmission time.

Figs. 4(a)-4(b) show the minimized delay versus the number of the IRS's reflecting elements, i.e., N . Analyzing Fig. 4(a), we can see that the minimized delay caused by each scheme decreases with the increase of N , and the performance gains of our proposed **IRS-DT/EH** over the other schemes increase as N increases. This suggests that the sophisticated design of the IRS's phase shifts can achieve a higher reflection-based beamforming gain, and then significantly improve both data and energy transmission efficiencies to reduce MUs' total delay especially when N takes a large value. Similarly, Fig. 4(b) shows that compared with the scheme No IRS-THz, the total delay of MUs posed by our **IRS-DT/EH-THz** can be greatly reduced by employing the IRS to help data and energy transmission for THz communications. Moreover, analyzing Fig. 4(b), we can observe that compared with the delay shown in Fig. 4(a), the delay obtained for THz communication is much smaller. This is because the bandwidth of THz communication is very large, and then much more data can be transmitted to the MEC server simultaneously. In addition, from Fig. 4(b), we can also see that the minimized delay by each scheme in Fig. 4(b) has only a little change with the increase of N . The reason for this is that due to the large bandwidth of THz band, a limited number of passive reflecting elements can only achieve limited reflection-based beamforming gains for THz communication.

Fig. 5(a) shows the total minimized delay caused by **IRS-DT/EH** versus the path loss exponent, i.e., $\alpha_{i,k}$, of the link between the IRS and each MU k . First, it can be seen that the total resulting delay by each scheme decreases with the increase of $\alpha_{i,k}$. The reason for this is that the channel gain between the IRS and MU k decreases as $\alpha_{i,k}$ increases. Thus, the signal transmission time, i.e., $t_{2,k}$, and then the total delay of MUs increases as $\alpha_{i,k}$ increases. Moreover, we can observe that our proposed **IRS-DT/EH** always outperforms the baseline schemes in Fig. 5(a) no matter what values $\alpha_{i,k}$ takes, and

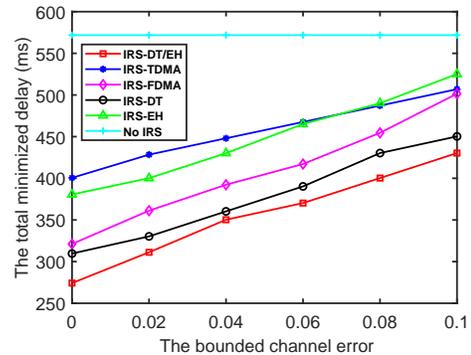


Fig. 6. The total minimized delay versus the bounded channel error Γ_k .

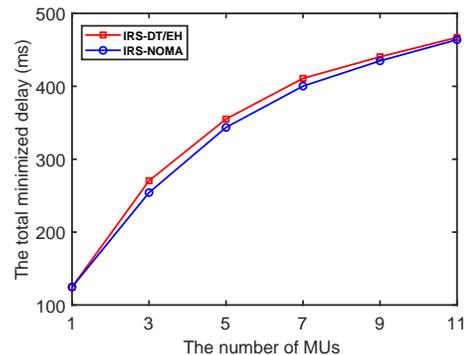


Fig. 7. Comparison of the total minimized delay caused by **IRS-DT/EH** and **IRS-NOMA**.

the performance gains over these baseline schemes decrease as $\alpha_{i,k}$ increases. These are also because the IRS can improve the efficiencies of data transmission and EH by reconfiguring wireless transmission environments. However, when the channel is in deep fading, deploying one IRS may not be able to significantly improve channel quality just as in the case when $\alpha_{i,k}$ takes a large value, e.g., $\alpha_{i,k} = 3.2$. Furthermore, Fig. 5(b) also shows the minimized delay versus $\alpha_{i,k}$, for THz communication. Due to the same reasons as above, we can observe that our proposed **IRS-DT/EH-THz** outperforms the baseline schemes in Fig. 5(b) regardless of whatever values $\alpha_{i,k}$ takes and the performance gains decrease as $\alpha_{i,k}$ increases.

In practice, it is difficult to perfectly estimate the CSI of the cascaded link between the BS and MU k via the IRS, due to channel estimation errors, the limited signal processing capabilities of the passive IRS, etc. Therefore, similar to [45]-[46], we model the channel of cascaded link of MU k , i.e., $\mathbf{H}_k \triangleq \mathbf{G}_{b,i} \text{diag}(\mathbf{h}_{1,k}^H) \in \mathcal{C}^{M \times N}$, $\forall k \in \mathcal{K}$, as $\mathbf{H}_k = \hat{\mathbf{H}}_k + \Delta \mathbf{H}_k$, where $\hat{\mathbf{H}}_k$ is the estimation of \mathbf{H}_k and $\Delta \mathbf{H}_k$ denotes the corresponding channel error vector. We assume that $\Delta \mathbf{H}_k$ satisfies the constraint $\|\Delta \mathbf{H}_k\| \leq \Gamma_k$, where Γ_k is the radius of the uncertainty region known by the BS. In Fig. 6, we show the minimized delay versus the bounded CSI error Γ_k . Analyzing Fig. 6, we can observe that the delay posed by all schemes (except the scheme No IRS) increases with the increase of Γ_k . The reason for this is that a large Γ_k generally increases the channel fading of the cascaded links, which then reduces the gains of the IRS for energy acquisition and information

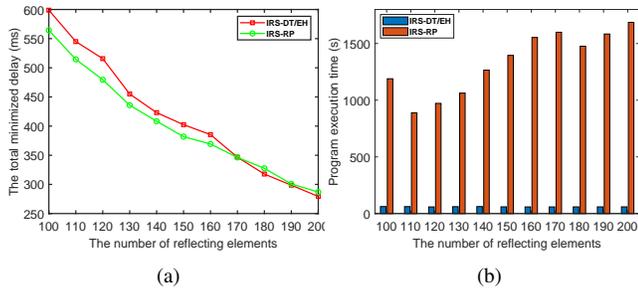


Fig. 8. Performance comparison between **IRS-DT/EH** and **IRS-RP**. (a) The total minimized delay; (b) Program execution time.

transmission. Moreover, we can see that no matter what Γ_k is, our proposed **IRS-DT/EH** still outperforms the other schemes and the performance gains decrease as Γ_k increases.

In Fig. 7, we also show the total resulting delay under our proposed **IRS-DT/EH** and a new scheme, i.e., **IRS-NOMA**, which is similar to **IRS-DT/EH** except that all MUs transmit data to the BS based on NOMA. Analyzing Fig. 7, we can see that the minimized delay caused by **IRS-DT/EH** is slightly larger than that caused by **IRS-NOMA**, and the difference becomes smaller and smaller as the number of MUs, i.e., K , increases. This is because when all MUs use NOMA to access the network, a part of interference can be eliminated for each MU by using the successive interference cancellation (SIC) technique. Therefore, the data-transmission rate of each MU caused by **IRS-NOMA** can be increased so that the data-transmission time can be decreased. However, as K increases, the data-transmission rate of each MU caused by **IRS-NOMA** becomes more and more close to that posed by **IRS-DT/EH**, and then the performance gains of **IRS-NOMA** over **IRS-DT/EH** decrease as K increases. Since it is difficult to determine the optimal decoding order for MUs at the BS and the decoding complexity is very high when employing NOMA [47], we still assume that all MUs transmit data simultaneously over the same spectrum and the BS uses the MUD technique to reduce interference among MUs in this paper.

In Fig. 8, we also show the performance comparison between our proposed **IRS-DT/EH** and another scheme, i.e., **IRS-RP**. **IRS-RP** is similar to **IRS-DT/EH** except that the RP method, which is very suitable to address optimization problems with complex constraints, e.g., QoS constraints, is utilized to optimize the IRS's phase shifts. Analyzing Fig. 8(a), we can observe that the minimized delay caused by **IRS-DT/EH** is larger than that caused by **IRS-RP**, when the number of reflecting elements, i.e., N , takes a relatively small value. As N increases, the total minimized delay posed by **IRS-DT/EH** becomes more and more close to that posed by **IRS-RP**. However, analyzing Fig. 8(b), we can observe that the execution time of **IRS-DT/EH** is about 60 s, while the execution time of **IRS-RP** is more than 800 s. Hence, the computation complexity of our proposed **IRS-DT/EH** is significantly lower than that of **IRS-RP**. Therefore, we choose to optimize the IRS's phase shifts based on SDP.

Fig. 9 shows the minimized delay caused by our proposed **IRS-DT/EH** under the *non-linear* and *linear* EH models,

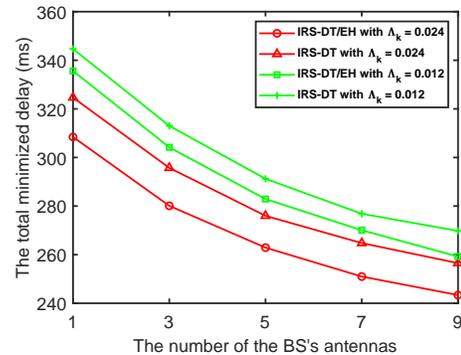


Fig. 9. The total minimized delay caused by **IRS-DT/EH** under the *non-linear* and *linear* EH models.

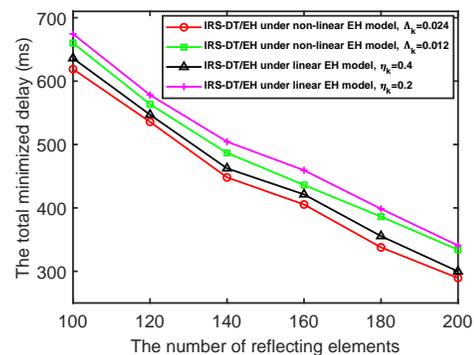


Fig. 10. The total minimized delay versus the number of BS's antennas.

where Λ_k characterizes the maximum harvested power of MU k when its *nonlinear* EH circuit saturates and η_k is the fixed energy conversion efficiency of MU k under the *linear* EH model. From Fig. 9, we can see that the minimized delay obtained under the *non-linear* EH model increases with the decrease of Λ_k , since MU k cannot harvest enough power when Λ_k takes a small value. Moreover, we can observe that the minimized delay obtained under the *non-linear* model is always lower than that obtained under the *linear* model when $\Lambda_k = 0.024$. However, when Λ_k takes a relatively small value, e.g., $\Lambda_k = 0.012$, the minimized delay obtained under the *linear* EH model may be lower than that obtained under the *non-linear* EH model. This is also because MU k cannot harvest enough power when $\Lambda_k = 0.012$. Therefore, the parameters Λ_k and η_k of EH circuits can significantly affect system performance. Consequently, it is very necessary to reasonably model the energy harvesting of RF-based MUs and select suitable circuit parameters.

Fig. 10 shows the minimized delay versus the number of BS's antennas, i.e., M , posed by our proposed **IRS-DT/EH** and the baseline **IRS-DT**, which can yield lower delay than other baseline schemes as shown in Figs. 4(a)-4(b). From Fig. 10, we can observe that the minimized delay of MUs decreases as M increases, and the minimized delay posed by our proposed **IRS-DT/EH** can be significantly reduced by using a multi-antenna BS. Besides, the minimized delay posed by our proposed **IRS-DT/EH** is always lower than that caused by **IRS-DT** whatever values M takes.

VI. CONCLUSIONS

We proposed delay minimization schemes for IRS-aided MEC over RF-powered 6G mobile networks with *non-linear* EH model, where using an IRS, multiple MUs first harvest energy from a multi-antenna BS and then transmit data to the BS for data-processing. We first formulated a delay minimization problem for MUs under their QoS requirements, by jointly optimizing the IRS's phase-shift matrices, the MEC server's computation resource allocation, the BS's MUD coefficients, and the energy/data transmission time and task allocation coefficient of each MU. Since our formulated joint-optimization problem is non-convex with multiple coupled variables, we applied the BCD method to decompose it into several subproblems which then can be iteratively solved with low complexity. Furthermore, we also extended our work to IRS-aided MEC over THz wideband mobile networks. Finally, we validated and evaluated the proposed delay minimization schemes through numerical analyses, which show that the total delay of the RF-powered MUs can be significantly reduced by using our proposed schemes.

APPENDIX A

THE PROOF OF PROPOSITION 1

Proof: For given $c_k \in [0, 1]$, $\forall k \in \mathcal{K}$, if $0 \leq c_k \leq (a_k f_k^e R_k) / (a_k R_k (f_k^e + f_k^L) + f_k^L f_k^e)$, we can re-write the delay D_k defined in Eq. (13) as follows:

$$D_k(c_k) = t_1 + \frac{(1 - c_k)L_k a_k}{f_k^L}, \quad (46)$$

otherwise if $(a_k f_k^e R_k) / (a_k R_k (f_k^e + f_k^L) + f_k^L f_k^e) < c_k \leq 1$,

$$D_k(c_k) = t_1 + \frac{c_k L_k}{R_k} + \frac{c_k L_k a_k}{f_k^e}. \quad (47)$$

Using Eqs. (46)-(47), we can obtain that D_k decreases with the increase of c_k when $c_k \in \left[0, \frac{a_k f_k^e R_k}{a_k R_k (f_k^e + f_k^L) + f_k^L f_k^e}\right]$, while D_k increases with the increase of c_k when $c_k \in \left[\frac{a_k f_k^e R_k}{a_k R_k (f_k^e + f_k^L) + f_k^L f_k^e}, 1\right]$. Therefore, D_k achieves its minimum value, when $D_k^L = D_k^e$. Using $D_k^L = D_k^e$, we can obtain the optimal $c_k^* = \frac{a_k f_k^e R_k}{a_k R_k (f_k^e + f_k^L) + f_k^L f_k^e}$, i.e., $c_k^* = \arg \min_{c_k} \{D_k(c_k)\}$. This completes the proof. ■

APPENDIX B

THE PROOF OF PROPOSITION 2

Proof: We denote the second derivative of the objective function of the optimization problem given by Eq. (18) w.r.t. f_k^e by Q , which is expressed as:

$$Q = \frac{2\tau_k L_k a_k^3 R_k^2 (f_k^L + a_k R_k)}{[a_k R_k (f_k^e + f_k^L) + f_k^L f_k^e]^3}. \quad (48)$$

Since the values of R_k and f_k^L are positive, and $L_k, f_k^e \geq 0$, Q must be non-negative. Hence the objective function is a convex function w.r.t. f_k^e . In addition, since $C1$ and $C2$ are linear, the optimization problem given by Eq. (18) is a convex optimization problem. ■

REFERENCES

- [1] D. Han and T. Shi, "Secrecy Capacity Maximization for a UAV-Assisted MEC System," *China Communications*, vol. 17, no. 10, pp. 64-81, October 2020.
- [2] T. Bai, C. Pan, H. Ren, Y. Deng, M. ElKashlan, and A. Nallanathan, "Resource Allocation for Intelligent Reflecting Surface Aided Wireless Powered Mobile Edge Computing in OFDM Systems," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 5389-5407, August 2021.
- [3] Y. Xu, G. Gui, H. Gacanin, and F. Adachi, "A Survey on Resource Allocation for 5G Heterogeneous Networks: Current Research, Future Trends, and Challenges," *IEEE Communications Surveys and Tutorials*, vol. 23, no. 2, pp. 668-695, Secondquarter 2021.
- [4] 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, January 2020.
- [5] T. Bai, C. Pan, C. Han, and L. Hanzo, "Reconfigurable Intelligent Surface Aided Mobile Edge Computing," *IEEE Wireless Communications*, vol. 28, no. 6, pp. 80-86, December 2021.
- [6] Xi Zhang, Jia Tang, Hsiao-Hwa Chen, Song Ci, and M. Guizani, "Cross-Layer-Based Modeling for Quality of Service Guarantees in Mobile Wireless Networks," *IEEE Communications Magazine*, vol. 44, no. 1, pp. 100-106, January 2006.
- [7] J. Tang and X. Zhang, "Quality-of-Service Driven Power and Rate Adaptation over Wireless Links," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 3058-3068, August 2007.
- [8] H. Su and X. Zhang, "Cross-Layer Based Opportunistic MAC Protocols for QoS Provisionings Over Cognitive Radio Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 118-129, January 2008.
- [9] C. Pan, H. Ren, K. Wang, W. Xu, M. ElKashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO Communications Relying on Intelligent Reflecting Surfaces," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5218-5233, August 2020.
- [10] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent Reflecting Surface Meets OFDM: Protocol Design and Rate Maximization," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4522-4535, July 2020.
- [11] H. Wang, C. Liu, Z. Shi, Y. Fu, and R. Song, "On Power Minimization for IRS-Aided Downlink NOMA Systems," *IEEE Wireless Communications Letters*, vol. 9, no. 11, pp. 1808-1811, November 2020.
- [12] B. Zheng, Q. Wu, and R. Zhang, "Intelligent Reflecting Surface-Assisted Multiple Access With User Pairing: NOMA or OMA?," *IEEE Communications Letters*, vol. 24, no. 4, pp. 753-757, April 2020.
- [13] F. Fang, Y. Xu, Q. -V. Pham, and Z. Ding, "Energy-Efficient Design of IRS-NOMA Networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 14088-14092, November 2020.
- [14] Y. Xu, H. Xie, Q. Wu, C. Huang, and C. Yuen, "Robust Max-Min Energy Efficiency for RIS-Aided HetNets With Distortion Noises," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1457-1471, February 2022.
- [15] Y. Xu, Z. Gao, Z. Wang, C. Huang, Z. Yang, and C. Yuen, "RIS-Enhanced WPCNs: Joint Radio Resource Allocation and Passive Beamforming Optimization," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7980-7991, August 2021.
- [16] D. Li, "How Many Reflecting Elements Are Needed for Energy- and Spectral-Efficient Intelligent Reflecting Surface-Assisted Communication," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1320-1331, February 2022.
- [17] C. Pan, H. Ren, K. Wang, M. ElKashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent Reflecting Surface Aided MIMO Broadcasting for Simultaneous Wireless Information and Power Transfer," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1719-1734, August 2020.
- [18] Q. Wu and R. Zhang, "Weighted Sum Power Maximization for Intelligent Reflecting Surface Aided SWIPT," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 586-590, May 2020.
- [19] J. Liu, K. Xiong, Y. Lu, D. W. K. Ng, Z. Zhong, and Z. Han, "Energy Efficiency in Secure IRS-Aided SWIPT," *IEEE Wireless Communications Letters*, vol. 9, no. 11, pp. 1884-1888, November 2020.
- [20] Y. Zheng, S. Bi, Y. -J. A. Zhang, X. Lin, and H. Wang, "Joint Beamforming and Power Control for Throughput Maximization in IRS-Assisted MISO WPCNs," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8399-8410, May 2021.
- [21] Y. Zheng, S. Bi, Y. J. Zhang, Z. Quan, and H. Wang, "Intelligent Reflecting Surface Enhanced User Cooperation in Wireless Powered Communication Networks," *IEEE Wireless Communications Letters*, vol.

- 9, no. 6, pp. 901-905, June 2020.
- [22] S. Huang, S. Wang, R. Wang, M. Wen, and K. Huang, "Reconfigurable Intelligent Surface Assisted Mobile Edge Computing With Heterogeneous Learning Tasks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 2, pp. 369-382, June 2021.
- [23] S. Huang, S. Wang, R. Wang, M. Wen, and K. Huang, "Reconfigurable Intelligent Surface Assisted Edge Machine Learning," *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1-6.
- [24] 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, November 2020.
- [25] X. Hu, C. Masouros, and K. K. Wong, "Reconfigurable Intelligent Surface Aided Mobile Edge Computing: From Optimization-Based to Location-Only Learning-Based Solutions," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 3709-3725, June 2021.
- [26] 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, June 2021.
- [27] H. Mei, K. Yang, J. Shen, and Q. Liu, "Joint Trajectory-Task-Cache Optimization With Phase-Shift Design of RIS-Assisted UAV for MEC," *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1586-1590, July 2021.
- [28] P. D. Lorenzo, M. Merluzzi, and E. C. Strinati, "Dynamic Mobile Edge Computing empowered by Reconfigurable Intelligent Surfaces," *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2021, pp. 526-530.
- [29] A. Li, Y. Liu, M. Li, Q. Wu, and J. Zhao, "Joint Scheduling Design in Wireless Powered MEC IoT Networks Aided by Reconfigurable Intelligent Surface," *2021 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*, 2021, pp. 159-164.
- [30] 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. 10820-10831, October 2021.
- [31] S. Zargari, A. Khalili, Q. Wu, M. Robot Mili, and D. W. K. Ng, "Max-Min Fair Energy-Efficient Beamforming Design for Intelligent Reflecting Surface-Aided SWIPT Systems With Non-Linear Energy Harvesting Model," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 5848-5864, June 2021.
- [32] Q. Wu, X. Guan, and R. Zhang, "Intelligent Reflecting Surface-Aided Wireless Energy and Information Transmission: An Overview," *Proceedings of the IEEE*, vol. 110, no. 1, pp. 150-170, January 2022.
- [33] P. Zeng, Q. Wu, and D. Qiao, "Energy Minimization for IRS-Aided WPCNs With Non-Linear Energy Harvesting Model," *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2592-2596, November 2021.
- [34] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A Survey on Mobile Edge Computing: The Communication Perspective," *IEEE Communications Surveys and Tutorials*, vol. 19, no. 4, pp. 2322-2358, Fourthquarter 2017.
- [35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [36] Q. Shi, M. Razaviyayn, Z. Q. Luo, and C. He, "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331-4340, September 2011.
- [37] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite Relaxation of Quadratic Optimization Problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20-34, May 2010.
- [38] Z. Hou, R. Chi, and H. Gao, "An Overview of Dynamic-Linearization-Based Data-Driven Control and Applications," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 4076-4090, May 2017.
- [39] I. M. Johnstone and A. Y. Lu, "Sparse Principal Components Analysis," 2009. [Online]. Available: <https://arxiv.org/abs/0901.4392>.
- [40] Y. Pan, K. Wang, C. Pan, H. Zhu, and J. Wang, "UAV-Assisted and Intelligent Reflecting Surfaces-Supported Terahertz Communications," *IEEE Wireless Communications Letters*, vol. 10, no. 6, pp. 1256-1260, June 2021.
- [41] Y. Pan, K. Wang, C. Pan, H. Zhu, and J. Wang, "Sum-Rate Maximization for Intelligent Reflecting Surface Assisted Terahertz Communications," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 3, pp. 3320-3325, March 2022.
- [42] W. Hao, G. Sun, M. Zeng, Z. Chu, Z. Zhu, O. A. Dobre, and P. Xiao, "Robust Design for Intelligent Reflecting Surface-Assisted MIMO-OFDMA Terahertz IoT Networks," *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 13052-13064, August 2021.
- [43] H. Yuan, N. Yang, K. Yang, C. Han, and J. An, "Hybrid Beamforming for Terahertz Multi-Carrier Systems Over Frequency Selective Fading," *IEEE Transactions on Communications*, vol. 68, no. 10, pp. 6186-6199, October 2020.
- [44] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. D. Renzo, A. L. Swindlehurst, R. Zhang, and A. Y. Zhang, "An Overview of Signal Processing Techniques for RIS/IRS-aided Wireless Systems," 2021. [Online]. Available: <https://arxiv.org/abs/2112.05989v2>.
- [45] G. Zhou, C. Pan, H. Ren, K. Wang, M. D. Renzo, and A. Nallanathan, "Robust Beamforming Design for Intelligent Reflecting Surface Aided MISO Communication Systems," *IEEE Wireless Communications Letters*, vol. 9, no. 10, pp. 1658-1662, October 2020.
- [46] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "A Framework of Robust Transmission Design for IRS-Aided MISO Communications With Imperfect Cascaded Channels," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5092-5106, 2020.
- [47] G. Li, Z. Ming, 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*, doi: 10.1109/TVT.2022.3177132.



Yuedong Li received the B.E. degree in Communication Engineering from Henan Polytechnic University, Jiaozuo, China, in 2020. He is currently pursuing the M.E. degree with Southwest University, Chongqing, China. His research interests include mobile edge computing and intelligent reflecting surface in wireless networks.



Fei Wang received the B.S. degree in Mathematics from Liaocheng University, Liaocheng, China, in 2005, the M.S. degree from Wuhan University, Wuhan, China, in 2007, and the Ph.D. degree in Computer Science from Chongqing University, China, in 2012. She is currently an Associate Professor with Southwest University, Chongqing, China. She was a visiting Ph.D. student under the supervision of Professor Xi Zhang in the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas, USA, from March 2017 to March 2018. Her research interests include resource allocation in wireless networks, network congestion control, and distributed algorithm design in wireless networks.



Xi Zhang (Fellow, IEEE) received the B.S. and M.S. degrees from Xidian University, Xi'an, China, the M.S. degree from Lehigh University, Bethlehem, PA, USA, all in electrical engineering and computer science, and the Ph.D. degree in electrical engineering and computer science (Electrical Engineering-Systems) from The University of Michigan, Ann Arbor, MI, USA.

He is currently a Full Professor and the Founding Director of the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas, USA. He is a Fellow of the IEEE for contributions to quality of service (QoS) theory in mobile wireless networks. He was with the Networks and Distributed Systems Research Department, AT&T Bell Laboratories, Murray Hill, NJ, USA, and AT&T Laboratories Research, Florham Park, NJ, USA, in 1997. He was a Research Fellow with the School of Electrical Engineering, University of Technology, Sydney, Australia, and the Department of Electrical and Computer Engineering, James Cook University, Australia. He has published more than 400 research articles on wireless networks and communications systems, network protocol design and modeling, statistical communications, random signal processing, information theory, and control theory and systems. He received the U.S. National Science Foundation CAREER Award in 2004 for his research in the areas of mobile wireless and multicast networking and systems. He received six best paper awards at IEEE GLOBECOM 2020, IEEE ICC 2018, IEEE GLOBECOM 2014, IEEE GLOBECOM 2009, IEEE GLOBECOM 2007, and IEEE WCNC 2010, respectively. One of his IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS papers has been listed as the IEEE Best Readings Paper (receiving the highest citation rate among all IEEE TRANSACTIONS/JOURNAL articles in the area) on wireless cognitive radio networks and statistical QoS provisioning over mobile wireless networking. He is an IEEE Distinguished Lecturer of the IEEE Communications Society and the IEEE Vehicular Technology Society. He received the TEES Select Young Faculty Award for Excellence in Research Performance from the College of Engineering at Texas A&M University, College Station, in 2006, and the Outstanding Faculty Award from Texas A&M University, College Station, in 2020.

Professor Zhang is serving or has served as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, and IEEE TRANSACTIONS ON NETWORK AND SCIENCE AND ENGINEERING. He served twice as a Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS for Special Issue on "Broadband Wireless Communications for High Speed Vehicles" and Special Issue on "Wireless Video Transmissions." He was an Associate Editor for IEEE COMMUNICATIONS LETTERS. He served twice as the Lead Guest Editor for *IEEE Communications Magazine* for Special Issue on "Advances in Cooperative Wireless Networking" and Special Issue on "Underwater Wireless Communications and Networks: Theory and Applications." He served as a Guest Editor for *IEEE Wireless Communications Magazine* for Special Issue on "Next Generation CDMA vs. OFDMA for 4G Wireless Applications." He served as an Editor for *Wireless Communications and Mobile Computing* (Wiley), *Journal of Computer Systems, Networks, and Communications*, and *Security and Communication Networks* (Wiley). He served as an Area Editor for *Computer Communications* (Elsevier), and among many others. He is serving or has served as the TPC Chair for IEEE GLOBECOM 2011, the TPC Vice-Chair for IEEE INFOCOM 2010, the TPC Area Chair for IEEE INFOCOM 2012, the Panel/Demo/Poster Chair for ACM MobiCom 2011, the General Chair for IEEE WCNC 2013, and the TPC Chair for IEEE INFOCOM 2017-2019 Workshops on "Integrating Edge Computing, Caching, and Offloading in Next Generation Networks."



Songtao Guo received his B.S., M.S., and Ph.D. degrees in Computer Software and Theory from Chongqing University, Chongqing, China, in 1999, 2003, and 2008, respectively. He was a professor from 2011 to 2012 at Chongqing University and a professor from 2013 to 2018 at Southwest University. At present, he is a full professor at Chongqing University, China. He was a senior research associate at the City University of Hong Kong from 2010 to 2011, and a visiting scholar at Stony Brook University, New York, USA, from May 2011 to May 2012. His research

interests include mobile edge computing, federated learning, wireless sensor networks, and wireless ad hoc networks. He has published more than 250 scientific papers in leading refereed journals and conferences. He has received many research grants as a Principal Investigator from the National Science Foundation of China and Chongqing and the Postdoctoral Science Foundation of China. He is a senior member of the IEEE and ACM.