IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 July 2024, accepted 25 July 2024, date of publication 29 July 2024, date of current version 14 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3435483

RESEARCH ARTICLE

Computation Offloading and Resource Allocation Optimization for Mobile Edge Computing-Aided UAV-RIS Communications

PHUC Q. TRUONG[1], TAN DO-DUY[1], (Member, IEEE), ANTONINO MASARACCHIA[2,3], (Senior Member, IEEE), NGUYEN-SON VO[4,5], (Senior Member, IEEE), VAN-CA PHAN[1], DAC-BINH HA[4,5], AND TRUNG Q. DUONG[2,6], (Fellow, IEEE)

[1]Department of Computer and Communication Engineering, Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology and Education, Ho Chi Minh City 700000, Vietnam
[2]School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, BT7 1NN Belfast, U.K.
[3]School of Electronic Engineering and Computer Science, Queen Mary University of London, E1 4NS London, U.K.
[4]Institute of Fundamental and Applied Sciences, Duy Tan University, Ho Chi Minh City 700000, Vietnam
[5]Faculty of Electrical-Electronic Engineering, Duy Tan University, Da Nang 550000, Vietnam
[6]Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada

Corresponding author: Dac-Binh Ha (hadacbinh@duytan.edu.vn)

This work was supported by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant 102.04-2021.11.

ABSTRACT The concept of Mobile Edge Computing (MEC) has been recently highlighted as a key enabling technology for the deployment of sixth-generation (6G) wireless network services. On the other hand, the possibility of combining Unmanned Aerial Vehicles (UAV) with Reconfigurable Intelligent Surfaces (RIS) has also been recognized as a powerful communication paradigm able to provide improved propagation characteristics of wireless communication channels, as well as increased capacity and extended coverage. Then, the possibility of merging the characteristics of such a communication paradigm with the one provided through MEC represents a valid solution to fulfill the main requirements of 6G networks. In this paper, we consider the combination of computation offloading and resource allocation in an MEC-based system where the MEC server is hosted by a massive MIMO base station, which serves multiple macro-cells assisted by a UAV-equipped RIS. In this context, we focus on minimising the latency for executing tasks of all user equipment (UE) within the considered scenario. To tackle this problem, we formulate an optimisation problem that jointly optimises computation offloading from user equipment (UE) towards the MEC server, and communication resources in the underlying UAV-assisted and RIS-aided network. The extensive simulation results demonstrate how the proposed method outperforms in terms of providing reduced latency for the considered system when compared with other conventional schemes.

INDEX TERMS Computation offloading, mobile edge computing, reconfigurable intelligent surfaces, resource allocation, unmanned aerial vehicle.

I. INTRODUCTION

Wthin the last two decades, wireless communication technologies have undergone a rapid advancement process, leading to the development of smaller, more portable, and intelligent mobile devices. This process has marked the dawn of the Internet of Things (IoT) era. On the one hand, the widespread use of such devices has paved and still continues to pave the way for the deployment of innovative services, which constantly simplify and enhance our daily lives. However, according to the International Telecommunication Union (ITU), such exponential diffusion of portable devices is expected to lead to a significant rise in global mobile subscribers, projected to reach approximately 17 billion by 2030, with a corresponding generation of data traffic envisaged to soar to around 5 zettabytes per month [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Yeon-Ho Chung.

This represents clear evidence that shortly we may assist to a collapse of the fifth-generation (5G) network technology if adequate actions are not taken. Indeed, such collapse will be mainly caused by the deployment of new capacity-hungry communication use cases and services such as multi-sensory extended reality, autonomous vehicles, industrial automation, healthcare systems, and video streaming, which are envisioned to be delivered within the upcoming years. [2]. However, the evolution of wireless networks must extend beyond the possibility to create more capacity to accommodate this surge. Indeed, next-generation networks must also deliver real-time communication with near-zero latency (communication lags less than 1 ms) and ultra-reliable transmission, i.e., less than 10^{-5} of communication error probability. These features are expected to become a reality through the full roll-out of 6G mobile communication systems [3], [4].

Furthermore, 6G technology is also expected to provide improved communication efficiency and intelligent data processing features for smart connected devices [5]. Within this regard, one approach that has gained particular attention within the last few years is the so-called MEC paradigm. With such an approach, computationally intensive tasks can be either partly or entirely offloaded and executed at the MEC servers, which are usually put at the edge of networks [6], [7], [8]. In this way, IoT devices with limited computational capabilities can offload the execution of the task to the edge server, reducing then the latency of the application, as well as increasing their operational lifetime since they are also energy-constrained. Then, one can easily notice why MEC has been highlighted as a valid candidate to provide improvements to next-generation wireless networks in terms of reduced latency and improved energy efficiency [8].

Nevertheless, the full deployment of 6G networks is also strongly dependent on providing innovative technologies that can improve the propagation characteristics of the wireless channel. Indeed, 6G communication scenarios will be highly complex and subject to strong high penetration losses of communication signals since THz communications are expected to be supported. This problem has been partly addressed with the introduction of massive multiple-input multiple-output (mMIMO) and hybrid analog and digital beamforming technologies [9], [10]. However, designing highly efficient multi-antenna transceivers for beamforming on THz bandwidth is challenging. To this end two main solutions have been identified as valid candidates to provide improvements at the physical layer: *i)* UAV-based communications, and *ii)* RIS-assited communication environments. Indeed, the main distinctive feature of UAV-based communications, when compared to conventional static base station (BS) communication, is the possibility of establishing line-of-sight (LoS) communication between UAV, acting as flying BS, and ground users. In this way, it will be possible to offer increased signal strength, which in turn enables the possibility to increase network performances [11]. On the other hand, RIS are entirely programmable metasurfaces, typically placed on a building facade, that through the usage of appropriate external signals allow to reflect the wireless signal in the desired direction. In this way, RIS can be employed to provide additional sources of links with the main aim of compensating for path loss and channel sparsity, enhancing then the effective connections between the base station and users [12]. Interestingly, the possibility of merging the benefits of these two physical layer solutions by realizing RIS-assisted UAV communications is also receiving a lot of attention [13], [14], [15].

Then, from the above discussions, one can easily observe how the possibility of implementing MEC-based solutions over underlying RIS-assisted UAV communications holds great potential for the deployment of the next generation of wireless networks. Some of the most relevant work presented in the literature on these novel research areas are discussed in the next section.

A. RELATED WORKS

As previously mentioned, the idea of integrating the advantages of LoS transmissions, through the adoption of UAVs, with the potential of implementing RIS to create a smart and controllable propagation environment is gaining attention as a compelling future research direction contributing to the deployment of next-generation wireless networks.

For example, authors in [13] considered a communication scenario where multiple UAVs equipped with an onboard RIS are used to support transmission subject to Ultra-reliable and low-latency communication (URLLC) constraints. In this case, each UAV acts as a repeater aimed at reflecting the signal from a macro BS to all users in the networks located in different areas far away from the BS. For such a communication scenario, authors formulated an optimisation problem for jointly optimising UAVs' deployment, power allocation at BS, phase-shift of RIS, and blocklength of URLLC transmission blocks. Such complex and non-convex optimisation problem, aimed at maximizing communication reliability and fairness among users, has been solved by adopting a deep neural network (DNN) based solution. Through numerical it has been highlighted the great potentialities of aerial RIS in supporting stringent URLLC demands.

Another work focused on showing the potentialities of aerial RIS in further extending the coverage range of massive multiple-input multiple-output (mMIMO) networks has been presented in [16]. In this case, the authors considered an optimisation problem to maximize the total network throughput by finding the optimal power control coefficients at the BS and the phase shift coefficients of the multiple RISs used in the system. By solving this optimisation problem through an iterative algorithm, authors illustrated how aerial RISs can achieve higher levels of network throughput as well as improvement for the users with worst-case throughput and less computational complexity when compared with other benchmark schemes.

On the other hand, the investigation on how UAV-enabled communications can contribute to further improving the performance of MEC systems represents another important research direction. Under this perspective, the optimal computation and communication resource allocation problem for UAV-assisted MEC systems under a non-orthogonal multiple access (NOMA) scheme has been considered in [17]. More specifically, they considered a communication scenario where a UAV serves as a MEC server-equipped flying base station (UAV-MEC). Under these assumptions, an iterative algorithm for jointly optimising user association, transmit power, and computing capacity allocation in order to minimise the total latency of UEs was proposed. Through numerical simulations, it has been highlighted how the proposed scheme outperforms other benchmark schemes in terms of offering overall reduced communication latency for the underlying UEs. Noteworthy, the usage of underlying NOMA communications resulted to provide better performances when compared with conventional orthogonal frequency-division multiple access (OFDMA) systems.

The possibility of reducing task offloading latency in MEC-based systems through the adoption of RIS has been studied in [18]. In this case, they considered a RIS-aided wireless MEC system for heterogeneous networks (HetNet). For such a setting, they formulated the optimisation problem for minimizing the overall system delay by jointly optimising caching, task offloading, and computing resources for the MEC system, as well as resource allocation for the RIS and BS sides. To deal with the resulting NP-hard mixed integer nonlinear programming problem, they proposed a two-stage optimisation algorithm. Through numerical results, it has been shown how the adoption of RIS represents a very effective solution to greatly reduce the task computing delay in MEC HetNet systems.

Recently, possibilities of fully integrating RIS and UAV technologies into MEC-based systems have been also investigated. A novel RIS-enhanced and UAV-assisted MEC framework with underlying NOMA communication has been investigated in [19]. More specifically, in this case, authors supposed that a single-antenna UAV is employed to offload the computation tasks to single antenna ground access points (APs) with the assistance of a RIS. To maximise the UAV's computation capacity, they proposed a two steps optimisation algorithm that jointly optimise the reflecting phase shift of the RIS, communication, and computation (2C) resource allocation, decoding order, and UAV's deployment which was supposed to be static. The numerical results provided within this study demonstrated that the computation capacity is greatly improved with such an approach when compared with other solutions proposed in the literature. A similar work has been presented in [20]. In this case, it has been supposed that the UAV acts as a relay node for supporting multiple offloading computation tasks to remote access points from ground users, through the assistance of the RIS. Also in this case the usage of NOMA as a communication paradigm was considered. However, in this case, authors considered the possibility for the UAV to dynamically moving within an optimal trajectory. Under these assumptions, a method for jointly optimising computation and offloading bits, RIS phase shift design, bandwidth allocation, and the trajectory of the UAV was proposed. Through this study, authors illustrated how the considered system can provide enhanced computation capacity, as well as how the inclusion of RIS and NOMA impact finding the optimal trajectory for the UAV.

B. MOTIVATION AND CONTRIBUTIONS

Based on the previous discussion, is evident how the possibility of implementing MEC systems assisted by the usage of both UAV and RIS technologies is gaining a lot of attention by the research community. More specifically, it has been highlighted how the complete integration of both technologies [19], [20] can provide higher benefits when compared with the exclusive inclusion of a single technology, i.e., UAV-based MEC systems [13], [16] or RIS-assisted MEC systems [18]. However, to the best of the author's knowledge, the majority of works focused on building such unified network, i.e., including both UAV and RIS, have considered RIS installed on the facade of buildings.

Under these perspectives, in this paper, we considered the optimisation problem of computation offloading and resource allocation for MEC systems assisted by a RIS-equipped UAV. A similar work in terms of communication environment has been presented in [21], in which authors aimed at maximising the energy efficiency of a single-antenna communication system. In contrast, this paper provides the following contributions to the current state of the art:

- We propose a novel optimisation framework for a MEC system, hosted within a massive MIMO Base Station (MBS), assisted by the usage of a RIS-equipped UAV able to fly within the coverage area. For such a system, we formulated an optimisation problem aimed at minimising the system latency by jointly optimising the power allocation of each user, user association, phase shift configuration of RIS reflecting elements, and computing resource allocation at the MBS subject to the MBS's computing resource constraints and QoS requirements.
- We design an iterative algorithm to efficiently solve the proposed optimisation problem by applying some approximation and inequalities, path following, and block coordinate descent (BCD) methods. An algorithm for determining the UAV trajectory based on the density of ground users is also provided.
- By means of extensive simulation results, we show that our proposed method outperforms the benchmark strategies indicating the effectiveness of our proposed method.

The remainder of this paper is organized as follows. The considered system model is presented in Section II. Section III provides the formulation of the optimisation problem for latency reduction, as well as the proposed methodology.

The effectiveness of the proposed algorithm in minimising the total system latency is illustrated in Section IV. Finally, conclusions and future research directions are provided in Section V.

II. MEC SYSTEM MODEL

A. SYSTEM MODEL

As illustrated in Figure 1, we consider a mMIMO communication system, which provides coverage extension between the MBS and distributed users (UEs) (e.g., mobile users, vehicles, internet-of-thing (IoT) sensors) through the assistance of assisted a RIS-equipped UAV. In this case, the support of RIS, combined with the flexible deployment of UAV, allows to provide enhanced network performances, in terms of reliable wireless network operation with high quality-of-service (QoS) to UEs in areas that are seriously impacted by propagation blockage (directly to/from the MBS) such as shadowing and blockage geometry [22], [23], [24]. Within the considered scenario, we suppose that a large antenna array consisting of L-elements is used by the MBS to provide service to K single-antenna users. We also assume that the UEs are grouped into M clusters represented by the set $\mathbf{K_U} = \{\mathcal{K}_1, \ldots, \mathcal{K}_M\}$ with $\mathcal{K}_m = \{1, \ldots, K_m\}, m = 1, \ldots, M$, with cluster having different numbers of users. To support the M clusters of UEs, we use one RIS-equipped UAV working as a small-cell base station where the RIS panel is comprised of N discrete elements to reflect the signal from each group of UEs to the MBS. Hence, in general, we denote with (m, k)-th UE as the k-th UE of the m-th group with $m = 1, \ldots, M$ and $k \in \mathcal{K}_m$. We use $u_{m,k}$ as the user association indicator to denote whether the kth UE of the m-th cluster offloads its computing task to the MEC server as follow:

$$u_{m,k} = \begin{cases} 1 & \text{if computing task offloading is needed,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We denote $u_m = [u_{m,k}]_{k=1}^{K_m}$ the association vector within the cluster, while $\mathbf{u} = [u_m]_{m=0}^{M}$ the entire user-association matrix. As the RIS-equipped UAV reaches the m-th cluster, users that needs to offload a task to the MBS are served according to time division multiple access (TDMA) technique. In this case, the main functionality of the RIS is to steer the beam from the ground user to the MBS. In this way we guarantee that RIS will only reflect the desired signal towards the MBS and no signal form other users [25], [26].

B. CHANNEL MODEL

By considering a 3D Cartesian coordinate system, we indicate the MBS's position, the UAV-RIS position, and users' positions as, (x_0, y_0, H_0), (x_U, y_U, H_U) and $(x_k, y_k, 0)$, $k = 1, 2, \ldots, K$, respectively. Within this notation, H_0 represents the MBS's antenna height while H_U is the height at which the RIS-equipped UAV is flying. These are supposed to be retrieved by the usage of the Global Positioning System (GPS) and locally stored at the MBS. Without loss of generality, we also assume the existence of a line-of-sight (LoS) communication between the MBS and the UAV-RIS, meaning that the path loss between the MBS and the UAV-RIS can be modelled by using free-space path loss model [27] as follow:

$$\beta_{m,0} = \frac{\beta_0}{d_{m,0}^2 + (H_0 - H_m)^2}, \; m = 1, \ldots, M, \tag{2}$$

where $d_{m,0} = \sqrt{(x_0 - x_m)^2 + (y_0 - y_m)^2}$, d_0 is the reference distance, and β_0 represents the power gain of wireless channel. In this case, H_m represents the height of the UAV when it flies on top of the m-th cluster as explained later in section II-E.

As regards the communication channel between the UAV-RIS and the (m, k)-th UE as the air-to-ground (ATG) channels, they are more complicated because of the propagation blockage effects. To this end, we have the path-loss formulation including the air-to-air (ATA) link and the air-to-ground (ATG) link are denoted as follows [28]:

$$\beta_{m,k} = PL_{m,k} + \eta_{LoS} P_{m,k}^{LoS} + \eta_{NLoS} P_{m,k}^{NLoS}, \tag{3}$$

Let us denote the average additional losses for the LoS and NLoS paths as η_{LoS} and η_{NLoS}, respectively. Since we can derive the distance path loss as

$$PL_{m,k} = 10 \log \left(\frac{4\pi f_c D_{m,k}}{c} \right)^{\alpha}, \tag{4}$$

Here, we have α as the path loss exponent with the value $\alpha \geq 2$, likely, c and f_c are the speed of light in m/s unit and the carrier frequency in Hz unit, respectively. Let us consider the Euclidean distance from the UAV-RIS to the (m, k)-th UE and the Euclidean distance from the MBS to the $(0, k)$th UE as $d_{m,k}$, then, we have $D_{m,k} = \sqrt{d_{m,k}^2 + H_m^2}$. To this end, we have the probability of LoS given as follows [29]

$$P_{m,k}^{LoS} = \frac{1}{1 + a \exp\left[-b\left(\arctan\left(\frac{H_m}{d_{m,k}}\right) - a\right)\right]} \tag{5}$$

where the values of both constants a and b depend on the environment. Then, we can express $P_{m,k}^{NLoS} = 1 - P_{m,k}^{LoS}$.

For the UEs that need the help of the RISs to reach the MBS, the small-scale fading coefficients for the channels from the (m, k)-th UE to the UAV-RIS and the UAV-RIS to the MBS ($m \in \mathcal{M}, k \in \mathcal{K}_m$), denoted by $\mathbf{h}_{m,k} \in \mathbb{C}^{N \times 1}$ and $\mathbf{h}_{m,0}^H \in \mathbb{C}^{N \times L}$, respectively. It worth to be noted that the coefficients are assumed as independent and identically distributed (i.i.d.) random variables with zero mean and unit variance, while the superscript H denotes the conjugate transpose operation. Furthermore, we denote $\mathbf{H}_{m,k} \in \mathbb{C}^{N \times 1}$ and $\mathbf{H}_{m,0}^H \in \mathbb{C}^{L \times N}$ as the channel matrix from the (m, k)-th UE to the UAV-RIS and the UAV-RIS to the MBS, respectively, where $\mathbf{H}_{m,k} = \sqrt{\beta_{m,k}} \mathbf{h}_{m,k}$ and $\mathbf{H}_{m,0}^H = \sqrt{\beta_{m,0}} \mathbf{h}_{m,0}^H$. Hence, the cascaded channel matrix of

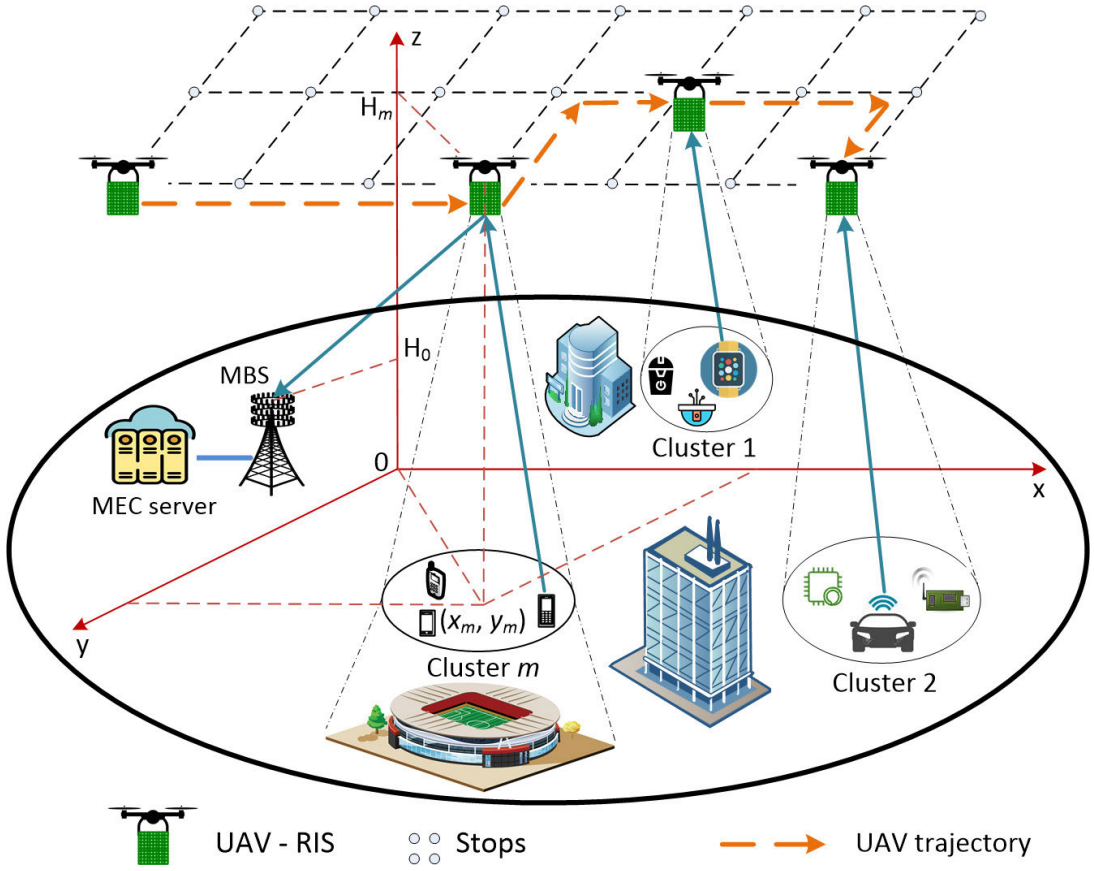


FIGURE 1. Illustration of a MEC system with UAV-RIS.

the link from the (m, k)-th UE to the MBS via the UAV-RIS, $\mathbf{G}_{m,k} \in \mathbb{C}^{L\times 1}$, can be shown as [23]

$$\mathbf{G}_{m,k} = \mathbf{H}_{m,0}^H \Phi_m \mathbf{H}_{m,k}, \ m \in \mathcal{M}, \tag{6}$$

where $\Phi_m = diag[\phi_{1m}, \phi_{2m}, \ldots, \phi_{Nm}]$ is the phase shift matrix at the UAV-RIS; $\phi_{nm} = \alpha_{nm} e^{j\theta_{nm}}$ with $\alpha_{nm} \in [0, 1]$ and $\theta_{nm} \in [0, 2\pi]$ $(\forall n = 1, 2, \ldots, N, \ m \in \mathcal{M})$ denotes the reflection amplitude and phase shift of the n-th reflecting element, respectively. Assuming that only the phase of reflected signals is changed by the RIS reflecting elements, then we can set $\alpha_{nm} = 1$ [30].

C. TRANSMISSION SCHEME

Since the (m, k)-th UE in the m-th group does not have a direct link with the MBS due to propagation blockage such as large buildings, it offloads its computing task to the MBS, it transmits the signal to the MBS via the RIS-equipped UAV. Hence, the signal received at the MBS from the (m, k)-th UE can be expressed as:

$$y_{m,k} = \sqrt{P_{m,k}} \mathbf{G}_{m,k}^H \boldsymbol{f}_{m,k} s_{m,k} + \underbrace{+ \sum_{l=1, l\neq k}^{K_m} \sqrt{P_{l,m}} \mathbf{G}_{l,m}^H \boldsymbol{f}_{l,m} s_{l,m}}_{\text{intra-cell interference}} + n_0, \tag{7}$$

where $P_{m,k}$ is the transmit power of the (m, k)-th UE; $\boldsymbol{f}_{m,k} \in \mathbb{C}^{L\times 1}$ is the beamforming vector of the MBS with respect to the (m, k)-th UE; $s_{m,k}$ is offloading information transmitted by the (m, k)-th UE with $||s_{m,k}||^2 \leq 1$; $n_0 \sim \mathcal{CN}(0, \sigma_0^2)$ is the AWGN at the MBS. The maximum transmit power of the (m, k)-th UE is denoted as $P_{m,k}^{max}$. The first term in (7) denotes the signal transmitted from the (m, k)-th UE via the RIS panel. On the other hand, since the users within the same cluster are supposed to transmit to the MBS at the same time, the second term in (7) represents the intra-cell interference inflicted by the other UEs in the m-th group. To overcome the interference in (7), we apply zero-forcing (ZF) technique [31]. More specifically, we define $\mathbf{G}_m = \left[\mathbf{G}_{m,1}, \ldots, \mathbf{G}_{m,K_m}\right] \in \mathbb{C}^{L\times K_m}$ $(m = 0, 1, \ldots, M)$. As illustrated in [32] and [33], eigenvalue distribution of the square matrix $\mathbf{G}_m^H \mathbf{G}_m \in \mathbb{C}^{K_m \times K_m}$ becomes more deterministic as L increases. Based on this favorable propagation of the mMIMO system, we develop a beamforming vector $\boldsymbol{f}_{m,k}$ by applying ZF as follows. Let

$$\bar{\boldsymbol{f}}_m = \mathbf{G}_m (\mathbf{G}_m^H \mathbf{G}_m)^{-1}, \tag{8}$$

where $\bar{\boldsymbol{f}}_m = \left[\bar{\boldsymbol{f}}_{m,1}, \cdots, \bar{\boldsymbol{f}}_{m,K_m}\right] \in \mathbb{C}^{L\times K_m}$, $\bar{\boldsymbol{f}}_{m,k} \in \mathbb{C}^{L\times 1}$, $m = 0, 1, \ldots, M$, $k \in \mathcal{K}_m$. We then normalize $\tilde{\boldsymbol{f}}_{m,k} = \bar{\boldsymbol{f}}_{m,k} / \|\bar{\boldsymbol{f}}_{m,k}\|$ and calculate $\boldsymbol{f}_{m,k}$ as

$$\boldsymbol{f}_{m,k} = \sqrt{p_{m,k}} \tilde{\boldsymbol{f}}_{m,k}, \ m = 0, 1, \ldots, M, \ k \in \mathcal{K}_m, \tag{9}$$

where $p_{m,k}$ is power control coefficient of the (m, k)-th UE.

Hence, the equation (7) becomes

$$y_{m,k} = \sqrt{P_{m,k}}\sqrt{p_{m,k}}\mathbf{G}_{m,k}^H \tilde{\boldsymbol{f}}_{m,k} s_{m,k} + n_0, \tag{10}$$

where the intra-cell interference in (7) has been cancelled.

Let $p_m = [p_{m,k}]_{k=1}^{K_m}$ and $\boldsymbol{p} = [p_m]_{m=0}^{M}$ denote the power control coefficients, and $\Phi = [\Phi_m]_{m=1}^{M}$ denote the phase shifts of RIS panels, the achievable throughput (in bits per second) at the MBS with respect to the transmission of the (m, k)-th UE can be given by

$$R_{m,k}\left(p_{m,k}, \Phi_m\right) = W \log_2\left(1 + \frac{P_{m,k} p_{m,k} \left|\mathbf{G}_{m,k}^H \tilde{\boldsymbol{f}}_{m,k}\right|^2}{\sigma_0^2}\right), \tag{11}$$

where W is the bandwidth allocated to the (m, k)-th UE.

D. OFFLOADING MODEL

In terms of computation modelling, we suppose that a particular task of size $\mathcal{I}_{m,k}$ can be executed either locally by the (m, k)-th UE or remotely through the assistance of the MEC server located within the MBS. To this end, we define two models of the computation latency as detailed below.

1) LOCAL COMPUTING

Indicating with $\mathcal{F}_{m,k}$ represents the number of CPU cycles required to compute each bit of the task by the (m, k)-th UE, the required time to execute the task locally is obtained as [34]:

$$T_{m,k}^l = \frac{\mathcal{I}_{m,k}\mathcal{F}_{m,k}}{c_{m,k}}, \ m = 0, 1, \ldots, M, \ k \in \mathcal{K}_m, \tag{12}$$

where $c_{m,k}$ denotes the maximum computing resource of the (m, k)-th UE.

2) OFFLOADING TO MBS

On the other hand, if the task is offloaded from the (m, k)-th UE to the MBS, we need first to take into account the offloading transmission time expressed as [34]:

$$T_{m,k}^{tx}(p_{m,k}, \Phi_m) = \frac{\mathcal{I}_{m,k}}{R_{m,k}\left(p_{m,k}, \Phi_m\right)}, \tag{13}$$

where $R_{m,k}\left(p_{m,k}, \Phi_m\right)$ is the communication rate expressed in (11). Once the task reaches the MBS, the computing time for the offloaded task at the MBS can be given as

$$T_{m,k}^{com}(\zeta_{m,k}^{bs}) = \frac{\mathcal{I}_{m,k}\mathcal{F}_{m,k}}{\zeta_{m,k}^{bs}}, \ m = 0, 1, \ldots, M, \ k \in \mathcal{K}_m, \tag{14}$$

where $\zeta_{m,k}^{bs}$ denotes the computing capacity of the MBS allocated to process the task of the (m, k)-th UE. For convenience, let $\zeta_m = [\zeta_{m,k}^{bs}]_{k=1}^{K_m}$ and $\zeta = [\zeta_m]_{m=0}^{M}$ denote the MBS computing capacity allocation. From (12)-(14), hence, the total latency for executing the task of the (m, k)-th UE can be written as

$$T_{m,k}^{tot}(p_{m,k}, u_{m,k}, \Phi_m, \zeta_{m,k}^{bs}) = (1 - u_{m,k})T_{m,k}^l + \\ + u_{m,k}\left(T_{m,k}^{tx}(p_{m,k}, \Phi_m) + T_{m,k}^{com}(\zeta_{m,k}^{bs})\right). \tag{15}$$

Algorithm 1 Shortest Trajectory

Require: $(0, 0, H_m)$, (x_m, y_m, H_m), $d_{\max}$
Ensure: $\mathcal{T}$

1: Generate all the permutations of the original stops in the sequence of $\{1, 2, \ldots, M\}$ to obtain the matrix $\mathcal{M}$ of $M!$ rows and M columns. Each permutation is a row representing a potential solution, i.e., flight sequence or a trajectory.
2: **for** each row $\mathcal{R}$ in $\mathcal{M}$ **do**
3: Compute the distance d from the UAV to the final stop
4: **if** $d < d_{\max}$ **then**
5: $\mathcal{T} \leftarrow \mathcal{R}$
6: $d_{\max} = d$
7: **end if**
8: **end for**

Note that we can ignore the time required for transmitting the computation results from the MBS back to the UEs since such latency is much less than the total latency for executing the task [34], [35].

E. CLUSTERS CREATION AND UAV TRAJECTORY

In terms of cluster creation and relative flying path optimization for the UAV we used an approach similar to the one provided in our previous work presented in [36] . More specifically, the creation of the clusters is based on the UE channel gains. Once the clusters are created, in order to save energy, the UAV will fly from its original position $(0, 0, H_0)$ to the closest cluster center within a straight line. Once it reaches the cluster center, it adjusts its flying height H_m within the range (H^{min}, H^{max}) in order to satisfy the QoS requirements as illustrated in our previous work (see Eq (22) of [36].

III. PROBLEM FORMULATION AND PROPOSED APPROACH

For the UAV-RIS assisted MEC model illustrated in Section II, we formulated an optimisation problem aimed at minimizing the total latency for executing the tasks of all the users in the considered area. This will be achieved through the joint optimisation of all the most relevant system variables, i.e., the power allocated by each user ($\boldsymbol{p}$), user association ($\mathbf{u}$), phase shift matrix of the RIS (Φ), and computing resource allocation at the MBS (ζ) subject to the MBS computing resource constraints and QoS requirements. More in detail, such optimisation problem is formulated as follows:

$$\min_{\boldsymbol{p}, \mathbf{u}, \Phi, \zeta} \sum_{m=0}^{M}\sum_{k=1}^{K_m} T_{m,k}^{tot}\left(p_{m,k}, u_{m,k}, \Phi_m, \zeta_{m,k}^{bs}\right) \tag{16a}$$

$$\text{s.t.} \quad 0 \le p_{m,k} \le 1, \tag{16b}$$

$$R_{m,k}\left(p_{m,k}, \Phi_m\right) \ge \bar{r}_0, \ m = 0, 1, \ldots, M, \ k \in \mathcal{K}_m, \tag{16c}$$

$$0 \le \theta_{nm} \le 2\pi, \forall n = 1, 2, \ldots, N, \ m \in \mathcal{M}, \tag{16d}$$

$$\sum_{k=1}^{K_m} u_{m,k}\zeta_{m,k}^{bs} \leq \zeta_{max}, \tag{16e}$$

where $u_{m,k}$ represents the user association coefficient as defined in (1). As regards the other constraints, (16b) represents the range of the power that can be used by each user, while (16c) the QoS requirements of each user in terms of minimum achievable uplink rate $\bar{r}_0$. On the other hand, the value range of each RIS phase-shift coefficient is expressed through constraint (16d). Finally, (16e) reflects the limit of computing resources at the MEC server that can be allocated at each UE.

Due to the non-convexity of such sum-latency minimization problem, we proposed a three-step optimisation framework that firstly finds the optimal value of power coefficient for a fixed value for RIS coefficients, ζ and $\mathbf{u}$. Subsequently, we formulate another optimisation problem aimed at finding the optimal value of the phase-shift coefficient at RIS. Finally, the optimal value for ζ is obtained. All the optimisation steps, described within the subsequent subsections, are repeated at each iteration until a stop condition is met (see Algorithm 4).

A. OPTIMAL POWER ALLOCATION

For any given $\mathbf{u}, \Phi, \zeta$, the original problem (16) can be reformulated as follow:

$$\min_{\boldsymbol{p}} \sum_{m=0}^{M}\sum_{k=1}^{K_m} T_{m,k}^{tot}\left(p_{m,k}\right)$$
$$\text{s.t. (16b), (16c).} \tag{17a}$$

To solve this new problem with make use of the following inequality [37], [38]:

$$f(x) = \log_2(1+\frac{1}{x}) \geq \hat{f}(x), \tag{18}$$

where $\hat{f}(x)$ is defined as follow:

$$\begin{aligned}\hat{f}(x) =& \log_2(1+\frac{1}{\bar{x}}) + \left(\frac{\partial f(\bar{x})}{\partial x}\right)(x-\bar{x}) \\ =& \log_2\left(1+\frac{1}{\bar{x}}\right) + \frac{1}{1+\bar{x}} - \frac{x}{(1+\bar{x})\bar{x}}.\end{aligned} \tag{19}$$

To this end, it is worth mentioning that (18) holds $\forall x > 0$ and $\forall \bar{x} > 0$. In our case, at the i-th iteration of Algorithm 2, both x and $\bar{x}$ are represented by the following quantities: WW

$$x = \frac{\sigma_0^2}{P_{m,k}p_{m,k}\left|\mathbf{G}_{m,k}^H \tilde{\boldsymbol{f}}_{m,k}\right|^2},$$
$$\bar{x} = x^{(i)} = \frac{\sigma_0^2}{P_{m,k}p_{m,k}^{(i)}\left|\mathbf{G}_{m,k}^H \tilde{\boldsymbol{f}}_{m,k}\right|^2},$$

These are used for approximating the throughput of each (m,k)-th UE as follows:

$$R_{m,k}\left(p_{m,k}\right) \geq \hat{R}_{m,k}^{(i)}\left(p_{m,k}\right), \ \forall m \in \mathcal{M}, \ \forall k \in \mathcal{K}_m, \tag{20}$$

with

$$\hat{R}_{m,k}^{(i)}\left(p_{m,k}\right) = W\left(\log_2\left(1+\frac{1}{\bar{x}}\right) + \frac{1}{1+\bar{x}} - \frac{x}{(1+\bar{x})\bar{x}}\right). \tag{21}$$

Subsequently, by introducing a new variables $\mathbf{r} \triangleq \{r_{m,k}\}$ $(\forall m \in \mathcal{M}, \forall k \in \mathcal{K})$, that satisfies the condition $\frac{1}{R_{m,k}(p_{m,k})} \leq r_{m,k}$, we can provide the following upper-bound for the objective function $T_{m,k}^{tot}\left(r_{m,k}\right)$:

$$\begin{aligned}T_{m,k}^{tot}\left(r_{m,k}\right) &\leq \hat{T}_{m,k}^{tot}\left(r_{m,k}\right) \\ &= (1-u_{m,k})T_{m,k}^{l} + u_{m,k}\left(r_{m,k}\mathcal{I}_{m,k} + T_{m,k}^{com}\right).\end{aligned} \tag{22}$$

As a result, we can rewrite the problem (17) as

$$\min_{\boldsymbol{p},\mathbf{r}} \sum_{m=0}^{M}\sum_{k=1}^{K_m} \hat{T}_{m,k}^{tot}\left(r_{m,k}\right) \tag{23a}$$
$$\text{s.t.} \quad p_{m,k} \leq 1, \ P_{m,k} \leq P^{\max}, \tag{23b}$$
$$\hat{R}_{m,k}^{(i)}\left(p_{m,k}\right) \geq \bar{r}_0, \tag{23c}$$
$$\frac{1}{\hat{R}_{m,k}^{(i)}} \leq r_{m,k}, \ m = 0, 1, \ldots, M, \ k \in \mathcal{K}_m, \tag{23d}$$

Consequently, problem (23) is now in the form of a standard convex optimisation problem that can be efficiently solved by using convex optimisation solvers like CVX [39]. The proposed power allocation procedure for solving problem (23) is summarised in Algorithm 2.

Algorithm 2 Optimal Power Allocation Procedure for Solving Problem (23)

Input:
Set $\mathbf{u}$, Φ, ζ and initial point $\boldsymbol{p}^{(0)}$;
Set the tolerance $\varepsilon = 10^{-3}$, the maximum iterations $I_{max} = 20$ to stop the algorithm;
for $m = 1$ to M do
Set $i = 0$
repeat
Solve problem (23) for the feasible solution $(\boldsymbol{p}^{(i+1)})$;
Set $i = i + 1$;
until Convergence or $i > I_{max}$;
end for
Output: Optimal power control coefficients $(\boldsymbol{p}^*)$

B. PHASE SHIFT OPTIMISATION

At this stage, we use the values of $\boldsymbol{p}$ obtained through Algorithm 2, while $\mathbf{u}$, and ζ have the same original value set during the power resource optimisation. Then, for these fixed values problem (16) can be rewritten as:

$$\min_{\Phi} \sum_{m=0}^{M}\sum_{k=1}^{K_m} T_{m,k}^{tot}\left(\Phi_m\right) \tag{24a}$$
$$\text{s.t.} \quad \text{(16c), (16d).} \tag{24b}$$

In this case, for a given beamforming vector ($\boldsymbol{f}_{m,k},\ \forall n = 1, 2, \ldots, N,\ m \in \mathcal{M}$) at the MBS, the main objective will be to search the optimal set of phase-shift coefficients for the RIS. Indicating the vector of phase-shift coefficients as $v_m = [v_m^1, \ldots, v_m^N]^H$ with $v_m^n = e^{j\theta_{nm}}$ ($\forall n = 1, 2, \ldots, N$), one can easily notice that the constraint (16d) is equivalent to the unit-modulus constraint i.e., $|v_m^n|^2 = 1$ [40]. At this stage, by defining the new variables $\chi_{m,k} = diag(\mathbf{H}_{m,0}^H)\mathbf{H}_{m,k}\tilde{\boldsymbol{f}}_{m,k}$, that leads to $\mathbf{H}_{m,0}^H \Phi_m \mathbf{H}_{m,k}\tilde{\boldsymbol{f}}_{m,k} = v_m^H \chi_{m,k}$, and by also applying the approximation as in (20)-(21), we can define the following inequality:

$$R_{m,k}\left(\Phi_m\right) \geq \tilde{R}_{m,k}^{(i)}\left(\Phi_m\right),\ \forall m \in \mathcal{M},\ \forall k \in \mathcal{K}_m, \tag{25}$$

where

$$\tilde{R}_{m,k}^{(i)}\left(\Phi_m\right) = W\left(\log_2\left(1+\frac{1}{\bar{y}}\right) + \frac{1}{1+\bar{y}} - \frac{y}{(1+\bar{y})\bar{y}}\right),$$

$$y = \frac{\sigma_0^2}{P_{m,k}p_{m,k}\left|v_m^H \chi_{m,k}\right|^2},$$

$$\bar{y} = y^{(i)} = \frac{\sigma_0^2}{P_{m,k}p_{m,k}\left|v_m^{H(i)} \chi_{m,k}\right|^2}.$$

In addition, by the mean of introducing a new variable $\tilde{\mathbf{r}} \triangleq \{\tilde{r}_{m,k}\}$ ($\forall m \in \mathcal{M}, \forall k \in \mathcal{K}$), that satisfies $\frac{1}{R_{m,k}(\Phi_m)} \leq \tilde{r}_{m,k}$, the objective function $T_{m,k}^{tot}\left(\tilde{r}_{m,k}\right)$ can be upper-bounded as:

$$T_{m,k}^{tot}\left(\tilde{r}_{m,k}\right) \leq \tilde{T}_{m,k}^{tot}\left(\tilde{r}_{m,k}\right) = \\ = (1-u_{m,k})T_{m,k}^{l} + u_{m,k}\left(\tilde{r}_{m,k}\mathcal{I}_{m,k} + T_{m,k}^{com}\right). \tag{26}$$

As result, problem (24) can be reformulated as follow:

$$\min_{v_m,\ \tilde{\mathbf{r}},\ m\in\mathcal{M}} \sum_{m=0}^{M}\sum_{k=1}^{K_m} \tilde{T}_{m,k}^{tot}\left(\tilde{r}_{m,k}\right) \tag{27a}$$

$$\text{s.t.}\quad v_m^H \chi_{m,k}\chi_{m,k}^H v_m \geq \left(2^{\bar{r}_0}-1\right)/a_k,\ m \in \mathcal{M},\ k \in \mathcal{K}_m, \tag{27b}$$

$$|v_m^n|^2 = 1, \forall n = 1, 2, \ldots, N,\ m \in \mathcal{M}, \tag{27c}$$

$$\frac{1}{\tilde{R}_{m,k}^{(i)}} \leq \tilde{r}_{m,k},\ m = 0, 1, \ldots, M,\ k \in \mathcal{K}_m \tag{27d}$$

where the constraint (27b) is equivalent to (16c), and $a_k = P_{m,k}p_{m,k}/\sigma_0^2$. However, the new optimisation problem (27) is still a non-convex quadratically constrained quadratic programming (QCQP) problem. To obtain a more easy formulation, we introduce additional transformations. More specifically, we define $\mathbf{X}_{m,k} = \chi_{m,k}\chi_{m,k}^H$ and $v_m^H \mathbf{X}_{m,k} v_m = tr\left(\mathbf{X}_{m,k} v_m v_m^H\right) = tr\left(\mathbf{X}_{m,k} V_m\right)$ where $V_m = v_m v_m^H$ must satisfy $V_m \succeq 0$ and rank(V_m)=1 [40], [41]. These allow us to obtain the following equivalent transformation for problem (27):

$$\min_{v_m,\ \tilde{\mathbf{r}},\ m\in\mathcal{M}} \sum_{m=0}^{M}\sum_{k=1}^{K_m} \tilde{T}_{m,k}^{tot}\left(\tilde{r}_{m,k}\right) \tag{28a}$$

$$\text{s.t.}\quad tr\left(\mathbf{X}_{m,k} V_m\right) \geq \left(2^{\bar{r}_0}-1\right)/a_k,\ m \in \mathcal{M},\ k \in \mathcal{K}_m, \tag{28b}$$

$$V_{m(n,n)} = 1, \forall n = 1, 2, \ldots, N,\ m \in \mathcal{M}, \tag{28c}$$

$$V_m \succeq 0, \tag{28d}$$

$$\frac{1}{\tilde{R}_{m,k}^{(i)}} \leq \tilde{r}_{m,k},\ m = 0, 1, \ldots, M,\ k \in \mathcal{K}_m, \tag{28e}$$

with $y = \frac{\sigma_0^2}{P_{m,k}p_{m,k}tr\left(\mathbf{X}_{m,k}V_m\right)}$. As one can easily notice, problem (28) is a convex semi-definite program (SDP) [40], [42], and then easy to solve by using CVX. The entire Block Coordinate Descent (BCD)-based procedure for phase shift searching is summarized in Algorithm 3.

Algorithm 3 Phase Shift Searching Procedure for Solving Problem (28)

Input:
Set $\mathbf{u}$, ζ, $\boldsymbol{p}$, and initial $\boldsymbol{f}_{m,k}^{(0)}$.
Set the tolerance $\varepsilon = 10^{-3}$, the maximum iterations $I_{max} = 20$ to stop the algorithm.
for $m = 1$ to M do
  Set $i = 0$
  repeat
    Solve problem (28) for the feasible solution ($\Phi_m^{(i+1)}$).
    Update $\boldsymbol{f}_{m,k}^{(i+1)}$.
    Set $i = i + 1$.
  until Convergence or $i > I_{max}$.
end for
Output: Optimal phase shift (Φ_M^*)

C. COMPUTATION OFFLOADING OPTIMISATION

Finally, using the optimal values of $\boldsymbol{p}$, $\mathbf{u}$, Φ obtained through the optimisation steps described in the previous subsection, we obtain the following optimisation problem with respect to ζ:

$$\min_{\zeta} \sum_{m=0}^{M}\sum_{k=1}^{K_m} T_{m,k}^{tot}\left(\zeta_{m,k}^{bs}\right) \tag{29a}$$

$$\text{s.t. (16e)},$$

which can also be easily through CVX since both the objective function and constraint (16e) are both convex with respect to ζ.

D. ITERATIVE OPTIMISATION ALGORITHM

Finally, we propose an iterative optimisation problem to jointly identify the optimal power allocation, phase shift searching, and computation offloading. The entire optimisation process is summarised in Algorithm 4 showing all the optimisation flow where the solution in each iteration is the initial point in the next iteration.

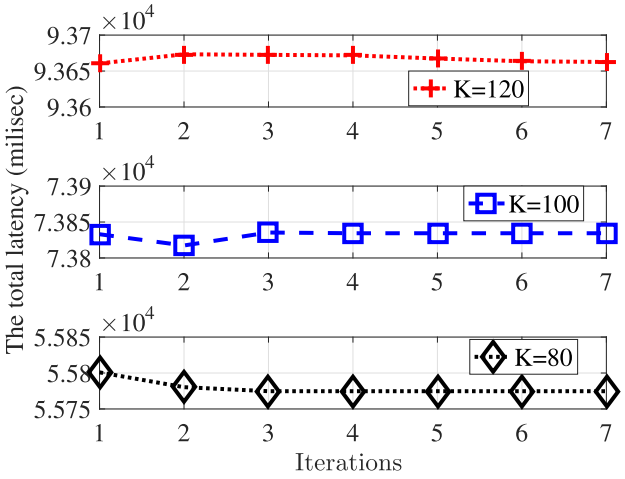


FIGURE 2. Convergence of the proposed optimization framework when changing the number of user K in the communication scenario.

Algorithm 4 Iterative Optimisation Algorithm for Jointly Solving Problem (16)

Input:
Set $\mathbf{u}$, Φ, ζ and initial point $\boldsymbol{p}^{(0)}, \boldsymbol{f}_{m,k}^{(0)}, \zeta^{(0)}$;
Set the tolerance $\varepsilon = 10^{-3}$, the maximum iterations $I_{max} = 20$ to stop the algorithm;
Set $j = 0$
repeat
Solve problem (23) for the feasible solution $(\boldsymbol{p}^{(j+1)})$;
Solve problem (28) for the feasible solution $(\boldsymbol{f}_{m,k}^{(j+1)})$;
Solve problem (29) for the feasible solution $(\zeta^{(j+1)})$;
Set $j = j + 1$;
until Convergence or $j > I_{max}$;
Output: $(\boldsymbol{p}^*, \boldsymbol{f}_{m,k}^*, \zeta^*)$

IV. SIMULATION RESULTS

This section provides the performance evaluations in terms of convergence of the proposed optimisation framework, as well as in term of the latency minimization when compared with other conventional scheme where only optimal resource allocation at the MEC server is performed, i.e., neither optimal power allocation nor phase-shift coefficients optimisation for the RIS is performed. To carry out these performance evaluations we considered the following parameters. The MBS is supposed to be placed at the center of a circular area and provide service to users located within 500 m distance through a direct connection. In addition, it is also assumed that there is a need to serve users distant up to 2000 m from the MBS. Within such a scenario, the 3D Cartesian coordinates of the MBS are (0, 0, 30), while all the UEs are randomly distributed within the whole area. As regards the RIS-equipped UAV, it is supposed to fly within this area and an altitude range (H^{min}, H^{max}) set to (50, 150) m. In terms of physical transmission parameters, in addition to the ones adopted in [37], it is assumed that each UE can use up to $P_{m,k}^{max} = 30$ dBm for the uplink transmissions over a communication channel with central frequency $f_c = 2.4$ GHz, bandwidth $W = 1$ MHz, and subject to white noise spectral density $\sigma_0^2 = -130$ dBm/Hz. In terms of QoS requirements, the minimum achievable rate for each UE in uplink is set to $\bar{r}_0 = 1$ Mbps. In terms of task computing, we assume that each UE needs to perform task computing, either locally or at the MBS, with a size of $D_m = 100$ kB and inner computation complexity of $\mathcal{F}_{m,k} = 600$ cycles/bit. The maximum computing resource (CPU cycle frequency) of the MEC server and the UEs are $\zeta_{max} = 30$ Giga cycles/s and $c_{m,k} = 0.5$ Giga cycles/s, respectively [43].

Under these assumptions, we evaluate the performance of our proposed optimisation framework, when compared with the conventional scheme detailed above, in terms of:

- Capability of reducing the total network latency, which is defined as $\sum_{m=0}^{M} \sum_{k=1}^{K_m} T_{m,k}^{tot}$.
- Ability of reducing the worst-case total latency defined as $\sum_{m=0}^{M} \max_{k \in \mathcal{K}_m} \{T_{m,k}^{tot}\}$.

A. CONVERGENCE ANALYSIS

The convergence characteristics of the proposed optimization framework have been evaluated by changing the number of users considered within the communication scenario. This is because, an increase in the number of users, corresponds to an increase in the variables to optimize, i.e., the number of clusters, power coefficients at the MBS, optimal sets of phase-shift coefficients, and computation optimal computation policy at the MEC server.

As illustrated in Figure 2 the proposed optimization framework requires only a few iterations to solve the optimization problem in line with Algorithm 4. More specifically, when $K = 80$ or $K = 100$ the algorithm converges after 3 iterations, while it requires more iterations when $K = 120$. This is because, as stated before, an increase in users corresponds to an increase in the optimization variables, which in turn requires more time for the algorithm to converge.

On the other hand, one can easily notice how an increase in the number of users also corresponds to an increase in the total latency for the worst-case user. This can be explained by the fact that, for a fixed amount of communication resources at the MBS and computation resources at the MEC, the higher the number of users that offload their tasks to the MEC server, the lower the resources that will be allocated to each user, requiring then more time at the MEC server receive the task, complete it and transmit back to the user. These aspects will be more evident in the next subsections.

B. THE WORST-CASE LATENCY VS COMPUTATION COMPLEXITY $\mathcal{F}_{M,K}$

In Figure 3, we demonstrate the total network latency of all UEs for different values of $\mathcal{F}_{m,k}$, with $K = 80$ and $\zeta_{max} = 30$ Giga cycles/s. The total latency is evaluated with difference values of CPU cycles, $\mathcal{F}_{m,k}$, ranging from 600 to

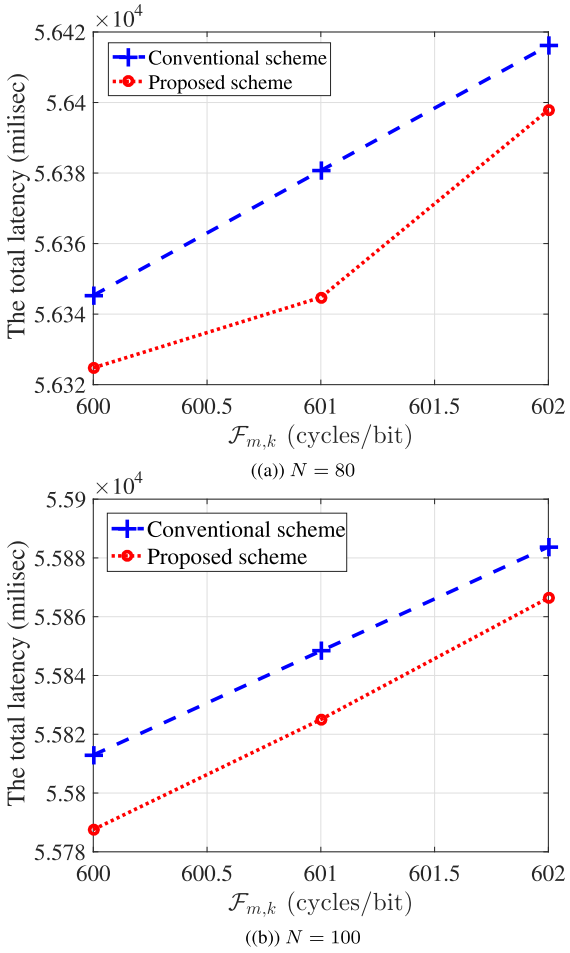


FIGURE 3. The total network latency according to different resource allocation schemes versus a range of $\mathcal{F}_{m,k}$, with $K = 80$ and $\zeta_{max} = 30$ Giga cycles/s.

602 cycles/bit, while the number of reflecting elements of RIS is $N = 80$ and $N = 100$, respectively.

Particularly, Figures 3(a) and 3(b) show the outperformance of the proposed method compared with the conventional scheme in terms of the total network latency. For instance, with $N = 80$ and $\mathcal{F}_{m,k} = 600$, the total latency with the proposed scheme and the conventional scheme is 5.632×10^4 ms and 5.635×10^4 ms, respectively. Moreover, the higher the number of reflecting elements, the lower the total network latency. To this end, it is worth mentioning that these results are valid for the range of considered reflective elements and do not represent a general rule [44] Furthermore, in Figure 4, we evaluate the total worst-case latency with different values of CPU cycles, $\mathcal{F}_{m,k}$. From the figure, we can observe that the total worst-case latency increases with the number of CPU cycles required to compute each bit of the task, i.e., the task complexity. Hence, the UEs should offload their local computing tasks to the MEC server to reduce the latency. By making jointly optimal power allocation, phase shift, and computation offloading, the proposed scheme can provide a better performance than the conventional scheme in terms of the total worst-case latency. On the other hand, when $\mathcal{F}_{m,k}$

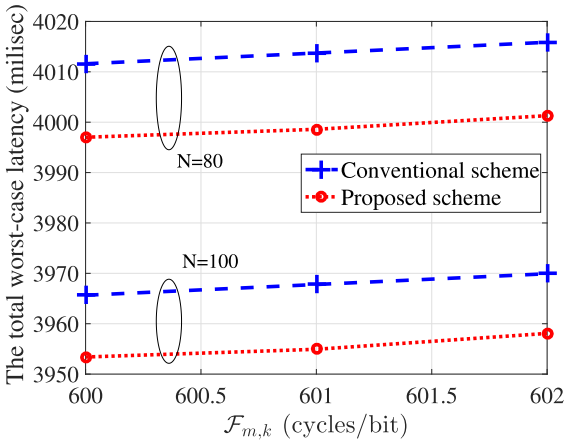


FIGURE 4. The total worst-case latency according to different resource allocation schemes versus a range of $\mathcal{F}_{m,k}$, with $K = 80$ and $\zeta_{max} = 30$ Giga cycles/s.

increases from 600 to 602 Giga cycles/s, with $N = 80$, the total worst-case latency with the proposed scheme goes up from approximately 3997 ms to 4001 ms while with $N = 100$, the total worst-case latency rises from some 3953 ms to 3958 ms, respectively. This again confirms that, increasing the number of RIS elements within the considered range, provides improvements in the system performance in terms of reduced latency.

C. THE WORST-CASE LATENCY VS COMPUTATION CAPABILITIES ζ_{MAX}

In Figure 5, we evaluate the total network latency with the proposed scheme in comparison with the conventional scheme for a range of computing capacity of MBS (Mega cycles/s), ζ_{max}. The number of UEs and CPU cycles is set at $K = 80$ and $\mathcal{F}_{m,k} = 600$ cycles/bit, respectively. Specifically, Figure 5 shows how an increase in the number of RIS elements and the computing capacity of MBS improves the total network latency. As observed from Figure 5, the total latency goes down remarkably with the computing capacity of MBS, ζ_{max}. For example, with $N = 80$, when the computing capacity of MBS increases from 4.5×10^4 to 4.53×10^4, the total latency reduces from 1.826×10^4 to 1.816×10^4. On the other hand, in Figure 6, we prove the benefits of the proposed method for UAV-RIS aided MEC system in terms of the total worst-case latency. It can be seen that, in general, the total worst-case latency declines significantly with the increase of the computing capacity. Most importantly, the proposed scheme always achieves a better latency performance than that of the benchmark. This again proves the efficiency of the proposed optimisation scheme i.e., jointly optimal power allocation, phase shift, and computation offloading for the UAV-RIS aided MEC system.

D. EXPERIMENTAL SUMMARY

In light of illustrations and relative discussions provided in subsections IV-B and IV-C, one can easily notice how the proposed optimization scheme is able to reach higher

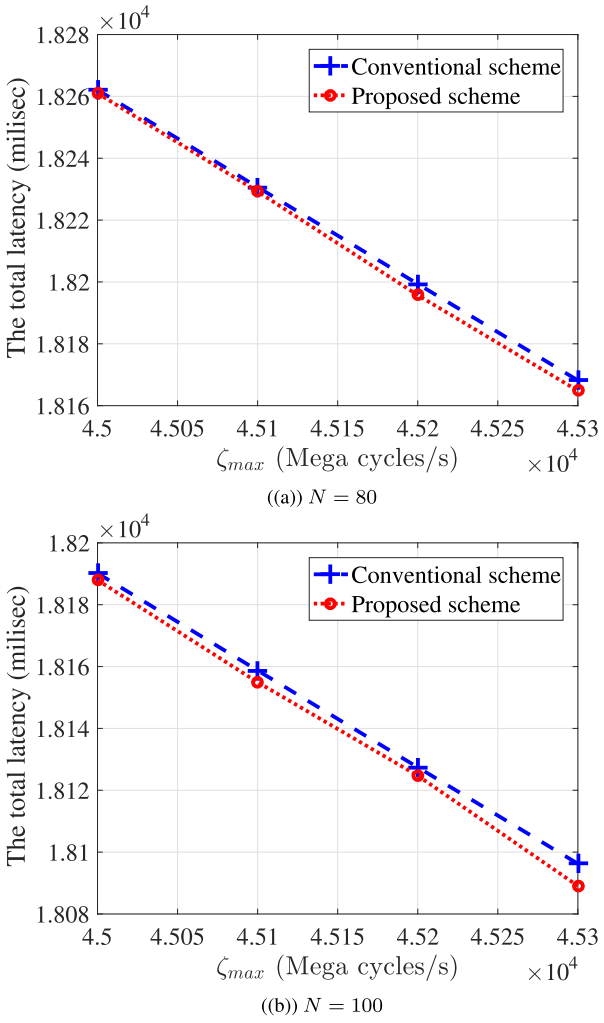


FIGURE 5. The total network latency according to different resource allocation schemes versus various values of ζ_{max}, with $K = 80$ and $\mathcal{F}_{m,k} = 600$ cycles/bit.

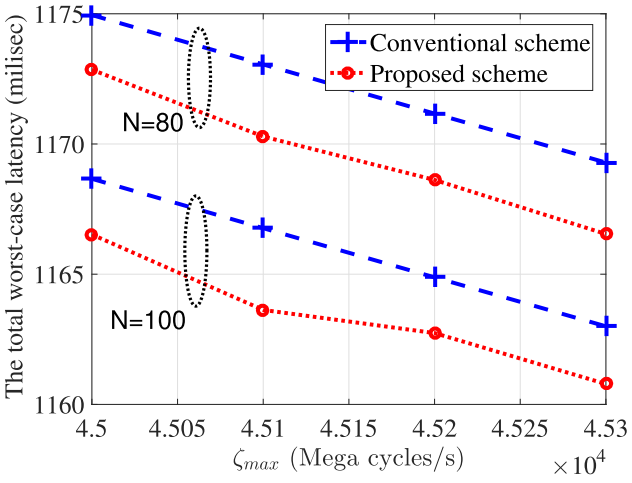


FIGURE 6. The total worst-case latency according to different resource allocation schemes versus various values of ζ_{max}, with $K = 80$ and $\mathcal{F}_{m,k} = 600$ cycles/bit.

performances in terms of reduced network latency when compared with a conventional transmission scheme where neither optimal power allocation at the MBS nor phase-shift coefficients optimisation for the RIS is performed. Indeed, the proposed optimization scheme has been tested within different configurations obtained by varying the main influential parameters of the considered communication scenarios such as the number of reflective elements N at the RIS panel and the computation complexity of the task $\mathcal{F}_{m,k}$ required by each user, and the CPU frequency at the MEC server ζ_{max}. Based on the obtained results we can summarize the main key point of this study as follows:

- In all the considered scenarios, the proposed optimization framework can find the optimal configurations of power allocation at the MBS and phase-shift coefficients for the RIS, as well as optimal path for the UAV, that permits to obtain lower levels of communication latency, while guaranteeing the QoS requirements for each user in the network.
- There is clear evidence that the communication latency increases when the task complexity $\mathcal{F}_{m,k}$ increases, while it decreases with an increase of either the number N of RIS elements or CPU frequency at the MEC server ζ_{max} increases. In any case cases, the proposed framework can provide the optimal resource allocation which minimizes the overall communication latency.
- For the considered scenario, the proposed optimization framework can provide lower levels of communication latency also for the worst-case user. This highlights the importance of how the usage of UAV equipped with RIS panels and joint optimization of UAV path, power allocation at the MBS, and phase-shift coefficients for the reflective elements represents a powerful solution to foster the deployment of 6G-oriented services with low-latency requirements.

The obtained results represent then a clear contribution to the current state of the art of 6G-based communication and services.

V. CONCLUSION

In this paper, we have proposed an MEC system hosted within a massive MIMO base station, serving M groups of users with the assistance of RIS-equipped UAV to enhance the coverage of the whole communication system. For such a communication scenario, we have considered the optimization problem of minimising the total latency for executing tasks of all UEs in the proposed system. More specifically, we have formulated the min-sum latency of all UEs by jointly optimising the user power allocation, user association, phase shift of reflecting elements of RIS, and computing allocation at the MBS subject to the QoS, and the MBS computing capacity. Additionally, we have designed the trajectory for UAV to save total fly time throughout M stops associated with M clusters of UEs for task offloading. The effectiveness of the proposed scheme has been demonstrated through numerical simulations, which highlighted how the

proposed scheme outperforms in terms of reducing the total network latency and the total worst-case latency of all UEs when compared with the conventional scheme. In the future, we plan to investigate the jittering effect of UAVs and extend to multiple UAVs in Internet of Things scenarios. Last but not least, we are also planning to design an optimization algorithm aimed at finding the optimal number of clusters and UAV-related trajectory to minimize the communication latency and save energy on the UAV side.

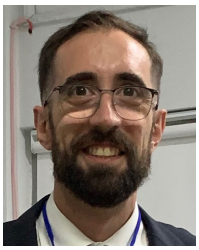
REFERENCES

[1] *IMT Traffic Estimates for the Years 2020 to 2030*, document ITU-R M.2370-0, Jul. 2015.
[2] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 334–366, 2021.
[3] H. Viswanathan and P. E. Mogensen, "Communications in the 6G era," *IEEE Access*, vol. 8, pp. 57063–57074, 2020.
[4] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May 2020.
[5] F. Guo, F. R. Yu, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, "Enabling massive IoT toward 6G: A comprehensive survey," *IEEE Internet Things J.*, vol. 8, no. 15, pp. 11891–11915, Aug. 2021.
[6] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 450–465, Feb. 2018.
[7] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A survey on mobile edge networks: Convergence of computing, caching and communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.
[8] D.-B. Ha, V.-T. Truong, and Y. Lee, "Performance analysis for RF energy harvesting mobile edge computing networks with SIMO/MISO-NOMA schemes," *EAI Endorsed Trans. Ind. Netw. Intell. Syst.*, vol. 8, no. 27, Jun. 2021, Art. no. 169425.
[9] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.
[10] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3060–3097, 4th Quart., 2018.
[11] A. Masaracchia, Y. Li, K. K. Nguyen, C. Yin, S. R. Khosravirad, D. B. D. Costa, and T. Q. Duong, "UAV-enabled ultra-reliable low-latency communications for 6G: A comprehensive survey," *IEEE Access*, vol. 9, pp. 137338–137352, 2021.
[12] S. Basharat, S. A. Hassan, H. Pervaiz, A. Mahmood, Z. Ding, and M. Gidlund, "Reconfigurable intelligent surfaces: Potentials, applications, and challenges for 6G wireless networks," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 184–191, Dec. 2021.
[13] Y. Li, C. Yin, T. Do-Duy, A. Masaracchia, and T. Q. Duong, "Aerial reconfigurable intelligent surface-enabled URLLC UAV systems," *IEEE Access*, vol. 9, pp. 140248–140257, 2021.
[14] K. K. Nguyen, A. Masaracchia, V. Sharma, H. V. Poor, and T. Q. Duong, "RIS-assisted UAV communications for IoT with wireless power transfer using deep reinforcement learning," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 1086–1096, Aug. 2022.
[15] K. K. Nguyen, A. Masaracchia, and C. Yin, "Deep reinforcement learning for intelligent reflecting surface-assisted D2D communications," *EAI Endorsed Trans. Ind. Netw. Intell. Syst.*, vol. 10, no. 1, p. e1, Jan. 2023.
[16] M. T. Nguyen, E. Garcia-Palacios, T. Do-Duy, O. A. Dobre, and T. Q. Duong, "UAV-aided aerial reconfigurable intelligent surface communications with massive MIMO system," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 4, pp. 1828–1838, Dec. 2022.
[17] T. Do-Duy, D. V. Huynh, E. Garcia-Palacios, T.-V. Cao, V. Sharma, and T. Q. Duong, "Joint computation and communication resource allocation for unmanned aerial vehicle NOMA systems," in *Proc. IEEE 28th Int. Workshop Comput. Aided Modeling Design Commun. Links Netw. (CAMAD)*, Edinburgh, U.K., Nov. 2023, pp. 1–6.
[18] Y. Wang, J. Niu, G. Chen, X. Zhou, Y. Li, and S. Liu, "RIS-aided latency-efficient MEC HetNet with wireless backhaul," *IEEE Trans. Veh. Technol.*, vol. 73, no. 6, pp. 8705–8719, Jun. 2024.
[19] Y. Xu, T. Zhang, Y. Zou, and Y. Liu, "Reconfigurable intelligence surface aided UAV-MEC systems with NOMA," *IEEE Commun. Lett.*, vol. 26, no. 9, pp. 2121–2125, Sep. 2022.
[20] H. Hu, Z. Sheng, A. A. Nasir, H. Yu, and Y. Fang, "Computation capacity maximization for UAV and RIS cooperative MEC system with NOMA," *IEEE Commun. Lett.*, vol. 28, no. 3, pp. 592–596, Mar. 2024.
[21] Z. Zhai, X. Dai, B. Duo, X. Wang, and X. Yuan, "Energy-efficient UAV-mounted RIS assisted mobile edge computing," *IEEE Wireless Commun. Lett.*, vol. 11, no. 12, pp. 2507–2511, Dec. 2022.
[22] Z. Chu, P. Xiao, M. Shojafar, D. Mi, J. Mao, and W. Hao, "Intelligent reflecting surface assisted mobile edge computing for Internet of Things," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 619–623, Mar. 2021.
[23] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020.
[24] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14960–14973, Dec. 2020.
[25] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng, "Coding metamaterials, digital metamaterials and programmable metamaterials," *Light: Sci. Appl.*, vol. 3, no. 10, p. e218, Oct. 2014.
[26] H. Taghvaee, A. Cabellos-Aparicio, J. Georgiou, and S. Abadal, "Error analysis of programmable metasurfaces for beam steering," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 10, no. 1, pp. 62–74, Mar. 2020.
[27] M.-N. Nguyen, L. D. Nguyen, T. Q. Duong, and H. D. Tuan, "Real-time optimal resource allocation for embedded UAV communication systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 225–228, Feb. 2019.
[28] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage," *IEEE Commun. Lett.*, vol. 20, no. 8, pp. 1647–1650, Aug. 2016.
[29] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.
[30] X. Xie, F. Fang, and Z. Ding, "Joint optimization of beamforming, phase-shifting and power allocation in a multi-cluster IRS-NOMA network," *IEEE Trans. Veh. Technol.*, vol. 70, no. 8, pp. 7705–7717, Aug. 2021.
[31] H. Q. Ngo, M. Matthaiou, T. Q. Duong, and E. G. Larsson, "Uplink performance analysis of multicell MU-SIMO systems with ZF receivers," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4471–4483, Nov. 2013.
[32] A. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. Delft, The Netherlands: Now Publishers Inc., 2004.
[33] L. D. Nguyen, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Beamforming and power allocation for energy-efficient massive MIMO," in *Proc. 22nd Int. Conf. Digit. Signal Process. (DSP)*, Aug. 2017, pp. 1–5.
[34] Y. Zhou, P. L. Yeoh, C. Pan, K. Wang, M. Elkashlan, Z. Wang, B. Vucetic, and Y. Li, "Offloading optimization for low-latency secure mobile edge computing systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 480–484, Apr. 2020.
[35] J. Xu and J. Yao, "Exploiting physical-layer security for multiuser multicarrier computation offloading," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 9–12, Feb. 2019.
[36] T. Do-Duy, L. D. Nguyen, T. Q. Duong, S. R. Khosravirad, and H. Claussen, "Joint optimisation of real-time deployment and resource allocation for UAV-aided disaster emergency communications," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 11, pp. 3411–3424, Nov. 2021.
[37] L. D. Nguyen, H. D. Tuan, T. Q. Duong, O. A. Dobre, and H. V. Poor, "Downlink beamforming for energy-efficient heterogeneous networks with massive MIMO and small cells," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3386–3400, May 2018.
[38] L. D. Nguyen, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Multi-user regularized zero-forcing beamforming," *IEEE Trans. Signal Process.*, vol. 67, no. 11, pp. 2839–2853, Jun. 2019.
[39] M. Grant and S. Boyd. (Mar. 2014). *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.1*. [Online]. Available: http://cvxr.com/cvx
[40] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.
[41] H. Yu, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "Joint design of reconfigurable intelligent surfaces and transmit beamforming under proper and improper Gaussian signaling," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2589–2603, Nov. 2020.

[42] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.
[43] T. Liu, L. Tang, W. Wang, Q. Chen, and X. Zeng, "Digital-twin-assisted task offloading based on edge collaboration in the digital twin edge network," *IEEE Internet Things J.*, vol. 9, no. 2, pp. 1427–1444, Jan. 2022.
[44] D. Tyrovolas, P.-V. Mekikis, S. A. Tegos, P. D. Diamantoulakis, C. K. Liaskos, and G. K. Karagiannidis, "Energy-aware design of UAV-mounted RIS networks for IoT data collection," *IEEE Trans. Commun.*, vol. 71, no. 2, pp. 1168–1178, Feb. 2023.

PHUC Q. TRUONG was born in Can Tho City, Vietnam. He received the B.Eng. degree in electronics and telecommunication engineering and the M.Eng. degree in electronics engineering from Ho Chi Minh City University of Technology and Education, Vietnam, in 2011 and 2014, respectively, where he is currently pursuing the Ph.D. degree with the Faculty of Electrical and Electronics Engineering. He is also a Senior Lecturer with the Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology and Education. His research interests include convex optimization techniques, heterogeneous networks, the Internet of Things, and intelligent reflecting surfaces (IRS).

TAN DO-DUY (Member, IEEE) received the B.S. degree from the HCMC University of Technology, Ho Chi Minh City, Vietnam, in 2010, the M.S. degree from the Kumoh National Institute of Technology, Gumi, South Korea, in 2013, and the Ph.D. degree from the Autonomous University of Barcelona, Barcelona, Spain, in 2019. He is currently a Lecturer with the Department of Computer and Communication Engineering, Ho Chi Minh City University of Technology and Education, Vietnam. His main research interests include wireless cooperative communications and network coding applications for wireless networking.



ANTONINO MASARACCHIA (Senior Member, IEEE) was a Research Fellow with Queens University Belfast, U.K. He is currently a Lecturer with the Queen Mary University of London. His research interests include 6G networks, digital twin, generative AI and applied machine learning techniques to wireless communications, reconfigurable intelligent surfaces (RIS), UAV-enabled networks, and ultra-reliable low-latency communications (URLLC).

NGUYEN-SON VO (Senior Member, IEEE) received the Ph.D. degree in communication and information systems from the Huazhong University of Science and Technology, China, in 2012. He is currently with the Institute of Fundamental and Applied Sciences, Duy Tan University, Ho Chi Minh City, Vietnam. His research interests include self-powered multimedia wireless communications, quality of experience provision in wireless networks for smart cities, and the IoT for disaster and environment management. He received the Best Paper Award from the IEEE Global Communications Conference, in 2016; and the Prestigious Newton Prize, in 2017. He has been serving as an Associate Editor for IEEE COMMUNICATIONS LETTERS, since 2019; and a Guest Editor for *Physical Communication* (Elsevier), Special Issue on Mission Critical Communications and Networking for Disaster Management, in 2019; *IET Communications*, Special Issue on Recent Advances on 5G Communications, in 2018; *Mobile Networks and Applications* (Springer), Special Issues on The Key Trends in B5G Technologies, Services and Applications, in 2021; *Wireless Communications and Networks for 5G and Beyond*, in 2018; and *Wireless Communications and Networks for Smart Cities*, in 2017.

VAN-CA PHAN received the Ph.D. degree in electronics and radio engineering from Kyung Hee University, Republic of Korea, in 2010. He is currently an Associate Professor with the Department of Computer and Communications Engineering, Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology and Education, Vietnam. His diverse research interests include wireless and mobile communication networks, heterogeneous networks, game theory, machine learning, reinforcement learning, dynamic programming techniques, the Internet of Things, cyber-physical systems, and embedded systems design, with a particular focus on optimization techniques for wireless and mobile communications.

DAC-BINH HA received the B.S. degree in radio techniques and the M.S. and Ph.D. degrees in communications and information systems from the Huazhong University of Science and Technology (HUST), China, in 1997, 2006, and 2009, respectively. He is currently an Associate Professor with the School of Engineering and Technology, Duy Tan University, Da Nang, Vietnam. His research interests include secure physical layer communications, cooperative communications, cognitive radio, RF energy harvesting networks, B5G/6G networks, mobile edge computing, and quantum computing and communications.

TRUNG Q. DUONG (Fellow, IEEE) is currently a Canada Excellence Research Chair (CERC) and a Full Professor with the Memorial University of Newfoundland, Canada. He is also an Adjunct Chair Professor in telecommunications with Queen's University Belfast, U.K., and the Research Chair of the Royal Academy of Engineering, U.K. His current research interests include quantum communications, wireless communications, signal processing, machine learning, and real-time optimization.

Dr. Duong received the Best Paper Award at the IEEE VTC-Spring 2013, IEEE ICC 2014, IEEE GLOBECOM 2016, 2019, 2022, IEEE DSP 2017, IWCMC 2019, 2023, 2024, and IEEE CAMAD 2023. He was a recipient of the prestigious Newton Prize 2017. He has also received the two prestigious awards from the Royal Academy of Engineering (RAEng): RAEng Research Chair (2021–2025) and the RAEng Research Fellow (2015–2020). He has served as an Editor/a Guest Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE COMMUNICATIONS LETTERS, IEEE WIRELESS COMMUNICATIONS LETTERS, IEEE WIRELESS COMMUNICATIONS, *IEEE Communications Magazines*, and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.