Multi-Agent UAV-aided URLLC Mobile Edge Computing Systems: A Joint Communication and Computation Optimisation Approach

Yijiu Li, *Graduate Student Member, IEEE*, Dang Van Huynh, *Member, IEEE*, Van-Linh Nguyen, *Member, IEEE*, Dac-Binh Ha, *Member, IEEE*, Hans-Jürgen Zepernick, *Senior Member, IEEE* and Trung Q. Duong, *Fellow, IEEE*

***Abstract*—In this paper, we consider a multi-agent unmanned aerial vehicle (UAV)-aided system employing mobile edge computing (MEC) servers to satisfy the requirement of ultra-reliable low latency communications (URLLC) in intelligent autonomous transport applications. Our MEC architecture aims to guarantee quality-of-service (QoS) by investigating task offloading and caching implemented in the nearby UAVs. To enhance system performance, we propose to minimise the network energy consumption by jointly optimising communication and computation parameters. This includes decisions on task offloading, edge caching policies, uplink transmission power, and processing rates of users. Given the non-convex nature and high computational complexity of this optimisation problem, an alternating optimisation algorithm is proposed, where the three sub-problems of caching, offloading, and power allocation are solved in an alternating manner. Our simulation results demonstrate the efficacy of the proposed method, showcasing significant reductions in user energy consumption and optimal resource allocation. This work serves as an initial exploration of the transformative potential of cutting-edge technologies, such as UAVs, URLLC, and MEC, in shaping the future landscape of intelligent autonomous transport systems.**

***Index Terms*—Mobile edge computing, task caching, multi-agent unmanned aerial vehicles, intelligent autonomous transport systems**

I. Introduction

Unmanned aerial vehicle (UAV) communication technology is an effective solution for intelligent systems, offering unparalleled flexibility and availability in extending wireless connectivity. Consequently, UAV-assisted communications has garnered considerable attention within the research community [1]–[6]. A holistic approach encompassing UAV deployment and resource allocation has been a focal point of investigation, particularly within the context of UAV-aided disaster emergency communications. For instance, in [1], a joint optimisation framework was proposed, leveraging a K-means-based user clustering model to enhance resource allocation efficiency in real time, thereby maximising energy efficiency. Similarly, in [2], a resource allocation scheme for UAV communications was devised which optimally managed energy-harvesting time and power control for device-to-device (D2D) communications in real time through the application of a path-following algorithm. Moreover, practical path planning optimisation solutions were explored in [4], [5], offering UAVs optimal trajectories for collecting Internet-of-Things (IoT) data that satisfy the requirements of completion time and energy consumption. UAVs and intelligent reflective surfaces were investigated to support terahertz communications in [3], where the authors presented an iterative algorithm relying on successive convex approximation to improve the system performance. Collectively, these studies highlight the multifaceted nature of UAV-assisted communications and their implications across various domains, from disaster response to IoT data collection and beyond. Continued research in this domain promises to unlock new possibilities and efficiencies in wireless communication systems.

With the rapid growth of network technology and advanced computing architecture, the new generation of wireless communication networks change people's everyday life drastically. Recently, the demand to run compute-intensive applications on user equipment under the stringent constraints of performance metrics can be met by mobile edge computing (MEC), a technology that brings computation and storage resources closer to the users [7]–[15]. Many research studies have been investigated from this perspective [16], [17]. In particular, the authors in [16] considered joint communication and computation to tackle the double near-far effect in a basic wireless-powered MEC network aided by non-orthogonal multiple access (NOMA). Accordingly, to minimise the access point (AP)'s transmit power consumption, the near user assists the far user in computing and forwarding the latter's partial computing task to the AP. In [18], the authors examined a three-node MEC system and jointly optimised the communication and computation resources in partial/binary offloading cases. The study in [19] investigated an MEC-enabled D2D task offloading system with the active users being able to offload their computation tasks to the idle user devices nearby. By jointly optimising communication and computing resource allocation, the system gets the best performance. In [20], the joint optimisation of transmit power

Y. Li and T. Q. Duong are with Queen's University Belfast, UK (e-mail: {yli84, trung.q.duong}@qub.ac.uk).

V.-L. Nguyen is with National Chung Cheng University (CCU), Taiwan (e-mail: nvlinh@cs.ccu.edu.tw).

D.-B. Ha is with Duy Tan University, Da Nang, Vietnam, (e-mail: hadacbinh@duytan.edu.vn).

H.-J. Zepernick is with the Blekinge Institute of Technology, Karlskrona, Sweden (e-mail: hans-jurgen.zepernick@bth.se).

D. V. Huynh and T. Q. Duong are with Memorial University, Canada (e-mail: {vdhuynh, tduong}@mun.ca).

This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.04-2021.11.

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

and computing resources was proposed for a multi-user full-duplex communication system with the employment of MEC and simultaneous wireless information and power transfer, aiming to achieve high computation speed and long-lasting self-sustainability. More recently, a joint communication and computation resource allocation scheme for a multi-user and multi-AP radio access network has been introduced in [21], which was obtained by jointly optimising user pairing and AP assignment. In [22], considering an MEC-based wireless network with the presence of energy harvesting devices and backscatter devices, the authors minimised the total energy consumption of the wireless devices by proposing a jointly optimal computation offloading and resource allocation. Furthermore, in [23], the authors considered a multi-user massive multiple-input multiple-output (MIMO) system in which each radio frequency chain was associated with a restricted number of antennas. The network energy efficiency was minimised by jointly optimising the computation and communication power. In [24], the total energy consumption in a two-user two-way relay MEC network was minimised, subject to computational delay. The two users cooperated in computing and shared their computational results. In [25], by using a layered algorithm, the study in jointly optimised the communication and computing resource allocation in an IoT system supported by multi-access MEC with NOMA. For an Internet-of-Vehicles edge computing network, the authors in [26] proposed a multi-objective reinforcement learning method to find the optimal solution of communication and computation resource allocation, thus effectively reducing the total system cost. In [17], artificial intelligence (AI) and cloud/edge computing was integrated to form an AI-enabled smart edge architecture suitable for IoT networks. Then, an algorithm was proposed to minimise the delay and optimise offloading decision and strategies in the network.

Ultra-reliable low-latency communications (URLLC), as a component of the 5G network architecture, is under the spotlight thanks to its ability to assure more effective data transfer scheduling. Characterised by its exceptionally low latency, ranging from 1 millisecond to just a few milliseconds, coupled with a remarkable reliability exceeding 99.999%, URLLC stands as a transformative technology. It serves as a cornerstone for mission-critical communications across an array of domains, including industrial automation, disaster response and communications, environmental monitoring, and beyond. The stringent standards of URLLC necessitate the utilisation of short-packet transmissions as a means to ensure and uphold the desired quality-of-service (QoS) [27]–[34]. This paradigm shift towards URLLC signifies a profound transformation in the way communication networks operate, promising unprecedented levels of reliability and responsiveness for a broad spectrum of applications, and it continues to be a focal point of research and development in the telecommunications arena [35]–[37]. However, these works often overlook the combined impact of MEC and UAV technologies on URLLC performance. For instance, [37] and [38] have explored joint communication and computation offloading but do not fully address the energy constraints and latency requirements in a UAV-assisted setting. Our study fills this gap by providing a comprehensive solution that optimizes power control, offloading decisions, and caching policies under strict URLLC constraints.

Inspired by the advances in URLLC, a joint communication and computation offloading scheme of URLLC has been investigated in an edge-cloud system [38]. To ensure the QoS, the authors employed alternating optimisation (AO) and inner approximation (IA) to solve the problem of minimising the end-to-end (e2e) latency. The digital twin (DT) concept has often been mentioned with URLLC in recent years to support time-sensitive applications. DT can create virtual twins of physical objects, and the DT models will support more accurate estimation and optimisation for the system. There is a growing body of research that investigates these areas [39]–[41]. In particular, an MEC-based URLLC digital twin framework has been presented in [39] to deal with the problem of reducing latency. Because of the high complexity of the non-convex objective function, an AO-based solution is applied to solve the optimisation problem. A URLLC multi-tier computing framework is proposed in [40] to support DT networks for metaverse applications. Another computation offloading problem under the DT paradigm via URLLC link has been addressed in [41]. To achieve better performance of the system, a latency minimisation problem comprised of three subproblems was solved by AO-IA.

In addition, the convergence of UAVs and MEC is attracting research attention due to their flexible configuration, mobile characteristic, and powerful computing ability. UAVs can fly up into the air to extend wireless coverage, improve transmission efficiency, and serve as flying edge servers. Thus, extensive studies [42], [43] have been conducted in this research area. More specifically, the authors in [42] jointly optimised the energy and latency cost for a UAV-assisted network where MEC servers were mounted on the UAVs to serve terminal devices. In [43], an on-board computation and communication resource allocation problem was jointly optimised, which was based on the Dinkelbach-like algorithm. The work in [44] discussed architectures and design considerations of UAV-enabled aerial computing and demonstrated its benefits over conventional MEC. In [45], an iterative method depending on successive convex approximation was used to jointly optimise the allocation of communication and computation resource in a UAV-assisted MEC vehicle platoon, where the vehicles supplied power to the UAV by employing wireless power transmission. In the UAV-assisted MEC system considered in [46], the mobile users have the opportunity to compute the tasks locally or offload a portion of tasks to the UAV, and the UAV can offload tasks on behalf of the devices and act as a computational relay to the terrestrial AP. In [47], taking into account the UAV's mobility, air-to-ground channels, and computation dynamics, the energy efficiency of the UAV in a UAV-oriented computation offloading system was maximised by using a gradient projection-based iterative algorithm.

The aforementioned studies have made their own contributions to the research on joint communication and computation resource allocation. However, to the best of your knowledge, there is a lack of research that jointly considers MEC, UAV, and URLLC. The novelty of this study lies in the integration

of MEC, UAV, and URLLC technologies to optimise energy consumption in UAV-assisted IoT systems. Unlike previous works that have primarily focused on individual aspects of these technologies, our approach jointly considers offloading decisions, edge caching policies, transmission power, and processing rates. This comprehensive approach allows for a holistic optimisation that minimises total energy consumption while meeting stringent URLLC requirements. The iterative algorithm based on inner approximation, which decomposes the original complex problem into manageable subproblems, represents a novel methodological advancement in this domain.

The main contributions of this paper are as follows:

- We formulate an energy consumption minimisation problem with joint communication, computation, and task caching in UAV-enabled URLLC for intelligent autonomous transport systems. The formulated problem not only takes into consideration wireless factors such as transmission power, but also jointly optimises the computing rate, and caching policies of UAV-MECs to minimise user equipment (UEs)' energy consumption.
- Under the constraints of various related parameters, i.e., latency constraints, energy budget, caching budget, we propose an alternating algorithm based on inner approximation to deal with this challenging problem. The original problem is divided into three subproblems and solved in an alternating fashion.
- Extensive simulation results confirm the effectiveness of the proposed solution in terms of minimising the energy consumption of UEs as well as optimising power control, offloading portions, computing rate, and caching policies under the stringent requirement of URLLC-based transmissions.

The remainder of this paper is organised as follows. Section II fully describes the system model and problem formulation, including the URLLC-based edge network model, task offloading model, latency, and energy model of the considered system, and problem formulation. Section III proposes the optimisation solution by decomposing the original problem into three subproblems and providing the development of the iterative algorithm. Section IV provides the numerical results and discussion. Finally, Section V highlights the main contributions of this paper and presents promising future research directions.

Notations: Throughout the paper, scalar variables and parameters are denoted in lowercase, (e.g., x, p); matrices are represented as bold uppercase and vectors are bold lowercase letters (e.g., $\mathbf{H}$ and $\mathbf{x}, \mathbf{p}$). $\|\cdot\|$ presents the vector's Euclidean norm, while $x \sim \mathcal{CN}(.,.)$ represents that x is a complex circularly symmetric Gaussian distribution, and $\mathbb{C}$ represents the collection of complex numbers. Finally, x_{mk} denotes a variable or parameter of x in connection with the m-th UE within the k-th UAV-MEC's network.

II. System Model and Problem Formulation

This section introduces the system model of using UAV to support URLLC ground users, the transmission model and the computing policy at the UAV. In this regard, the details of latency and energy models for UAV-aided URLLC MEC are also provided. We then introduce the minimisation problem for UEs' energy consumption.

A. System and Transmission Models

A URLLC edge network supported by multiple UAVs is shown in Fig. 1. In this system, the UEs are connected to the UAV's network via URLLC links to guarantee the reliability and low-latency requirements.

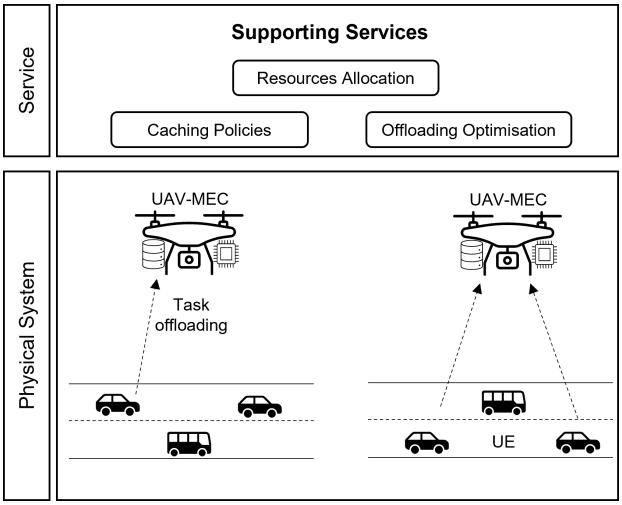

Fig. 1: A UAV-enabled URLLC network for intelligent autonomous transport systems.

There are M UEs defined as a collection of $\mathcal{M} = \{1, 2, ..., M\}$. There are totally K UAVs in the system, defined as a set of $\mathcal{K} = \{1, 2, ..., K\}$. The UAVs are formed into K clusters, where the k-th UAV covers the connections of M_k UEs in their group. The UAVs serve as flying APs and can also perform as edge servers (ES). In order to meet the requirement on the latency of computational tasks offloaded from UEs, the UAVs can support the mobile computing and mobile caching at the edge.

Each UAV is equipped with multiple (L) antennas, whereas the UEs are equipped with single-antenna.

1) Channel Model: The links between the UAVs and UEs can be considered as air-to-ground (ATG) communication, which is supported by the strong signal from line-of-sight (LoS) channels. However, due to the UAV's mobility, these ATG communication links may experience propagation attenuation under the severe blockage geometry and high shadowing effect, which leads to the fact that the channel modelling is complex [3]. As a result, the path loss of the transmissions from the k-th UAV to the (m, k)-th UE can be modelled as the combination between average additional losses for LoS (η^{LoS}) and non-LoS (NLoS) (η^{NLoS}) as

$$g_{mk} = PL_{mk} + \eta^{LoS} P_{m,k}^{LoS} + \eta^{NLoS} P_{mk}^{NLoS}. \quad (1)$$

We further elaborate the individual components in the path loss formula. Firstly, the conventional path loss (PL_{mk}) is

described as a function of the distance between the transmitters and receivers as

$$PL_{mk} = 10\log\left(\frac{4\pi f_c r_{mk}}{c}\right)^{\beta}, \tag{2}$$

in which f_c is the carrier frequency (Hz) of the transmitted signal, c is the speed of light (m/s), $\beta \geq 2$ is the exponent parameter of the path loss. In addition, we have

$$r_{mk} = \sqrt{d_{mk}^2 + Z_k^2}, \tag{3}$$

where d_{mk} is the distance from the m-th UE to the k-th UAV whose antenna height is Z_k. We next examine the probabilities of the path loss components as

$$P_{mk}^{LoS} = \frac{1}{1 + a\exp\left[-b\left(\arctan\left(\frac{Z_k}{d_{mk}}\right) - a\right)\right]}, \tag{4}$$

$$P_{mk}^{NLoS} = 1 - P_{mk}^{LoS}, \tag{5}$$

in which a and b are constants and determined by the specific arrangement of the environment.

The channel vector between the k-th UAV and the m-th UE is given by

$$\mathbf{h}_{mk} = \sqrt{g_{mk}}\bar{\mathbf{h}}_{mk}, \mathbf{h}_{mk} \ \in \ \mathcal{C}^{L\times 1}, \tag{6}$$

in which $\bar{\mathbf{h}}_{mk}$ is the small-scale fading for the channel from the k-th UAV to the m-th UE, and g_{mk} is the large-scale channel coefficient defined in (1) [1]. Given that the transmitted message, s_{mk}, from the (m,k)-th UE is a Gaussian signal with zero mean and unit variance, the received signal at the k-th UAV is expressed as

$$\mathbf{y}_k = \sum_{m=1}^{M_k} \mathbf{h}_{mk}\sqrt{p_{mk}}s_{mk} + \mathbf{n}_k, \tag{7}$$

where $\mathbf{h}_k$ is the channel matrix of the links between the M_k UEs and the k-th UAV; p_{mk} is the power of the (m,k)-th UE; and $\mathbf{n}_k \ \sim \ \mathcal{CN}(\mathbf{0}, N_0\mathbf{I}_L)$ is the additive white Gaussian noise (AWGN) with variance N_0.

Due to the complex nature of interference in the considered network, to maintain fairness amongst all the UEs and enhance the overall network performance, we apply the successive interference cancellation technique at each UAV to decode the signals from the UEs. Consequently, the signal-to-interference-plus-noise ratio (SINR) for the transmission from the (m,k)-th UE to the k-th UAV becomes:

$$\gamma_{mk}(\mathbf{p}) = \frac{p_{mk}||\mathbf{h}_{mk}||^2}{\mathcal{I}_{mk}(\mathbf{p}) + N_0}, \tag{8}$$

where $\mathcal{I}_{mk}(\mathbf{p}) = \sum_{n>m}^{M} p_{nk}\frac{|\mathbf{h}_{mk}^H\mathbf{h}_{nk}|^2}{||\mathbf{h}_{mk}||^2}$ is the interference imposed by the UEs $n > m$.

2) Transmission rate for URLLC link: The exact transmission rate of URLLC link with respect to the finite blocklength design is complicated. Let $N = \delta B$ denote the blocklength with B being the bandwidth and δ being the transmission time interval, and ϵ denote package error probability. The achievable transmission rate (bits/s) can be approximated as [32], [48]

$$R_{mk}(\mathbf{p}) \approx B\log_2\left[1 + \gamma_{mk}(\mathbf{p})\right] - B\sqrt{\frac{V_{mk}(\mathbf{p})}{N}}\frac{Q^{-1}(\epsilon)}{\ln 2}, \tag{9}$$

where $Q^{-1}(.)$ is the inverse function of $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty}\exp^{\left(\frac{-t^2}{2}\right)}dt$, while V is the channel dispersion given by

$$V_{mk}(\mathbf{p}) = 1 - \left[1 + \gamma_{mk}(\mathbf{p})\right]^{-2}. \tag{10}$$

It can be observed that the approximated transmission rate of URLLC in (9) for the finite blocklength N can be considered as a special case of maximum transmission rate. Specifically, as the blocklength approaches to infinity, (9) converges to Shannon's formula $B\log_2(1 + \gamma_{mk}(\mathbf{p}))$, i.e., the channel capacity.

B. Computational Task Offloading Model

To characterise a computational task I_{mk} from the (m,k)-th UE, we use two parameters ψ and T_{mk}. In this regard, the task complexity is given as $\psi = \frac{C_{mk}}{D_{mk}}$ (cycles/bit), where D_{mk} denotes task size (bits), C_{mk} is the required computing resource (CPU cycles), and T_{mk} (s) is the requirement of latency to complete the task.

Let $\mathbf{x} = \{x_{mk}\}_{\forall m,k}$, which satisfies $0 \leq x_{mk} \leq 1$, be the portion of local processing tasks. Then, there is a portion (1-x_{mk}) of the task that is offloaded to the UAV-MEC.

1) Local computing: The (m,k)-th UE processes x_{mk} probability of task J_{mk} with the computing rate f_{mk}^{ue}. Thus, the local computing latency is calculated as

$$T_{mk}^{\mathsf{ue}}(x_{mk}, f_{mk}^{\mathsf{ue}}) = \frac{x_{mk}C_{mk}}{f_{mk}^{\mathsf{ue}}}. \tag{11}$$

2) Edge computing: Let f_{mk}^{uav} be the computing rate of the k-th UAV-MEC for handling the offloaded task of the (m,k)-th UE. The latency associated with the execution of task J_m by the k-th UAV-MEC can be obtained as

$$T_{mk}^{\mathsf{uav}}(x_{mk}, f_k^{\mathsf{uav}}) = \frac{(1 - x_{mk})C_{mk}}{f_{mk}^{\mathsf{uav}}}. \tag{12}$$

C. Latency and Energy for Mobile Edge Computing

In this paper, we follow the traditional caching policy in MEC. If the task is popular, it will be cached at the UAV. Therefore, when the UE requests this task, it can be directly transmitted from the UAV to save the communications time and the edge processing latency is computed. In contrast, if the task is not cached at the UAV, the UE will conduct the task offloading process as normal. In addition, the feedback latency between the UAV and UEs through the control channel can be neglected compared with the trasmission latency of the uplink. As a result, the total latency for the caching policy of

the UE can be shown as

$$T_{mk}^{\mathsf{tot}} = \frac{\pi_{mk} C_{mk}}{f_{mk}^{\mathsf{uav}}} + (1-\pi_{mk})(T_{mk}^{\mathsf{ue}} + T_{mk}^{\mathsf{co}} + T_{mk}^{\mathsf{uav}}) = \frac{\pi_{mk} C_{mk}}{f_{mk}^{\mathsf{uav}}} + (1-\pi_{mk})\Big[\frac{x_{mk} C_{mk}}{f_{mk}^{\mathsf{ue}}} + \frac{(1-x_{mk}) D_{mk}}{R_{mk}(\mathbf{p})} + \frac{(1-x_{mk}) C_{mk}}{f_{mk}^{\mathsf{uav}}}\Big]. \quad (13)$$

We next investigate the total energy consumed by the m-th UE as

$$E_{mk}^{\mathsf{ue}}(\pi_{mk}, x_{mk}, f_{mk}^{\mathsf{ue}}, \mathbf{p}) = (1-\pi_{mk})(E_{mk}^{\mathsf{cp}} + E_{mk}^{\mathsf{co}}), \quad (14)$$

where $\pi \triangleq \pi_{mk}$, which satisfies $\pi \in \{0,1\}$ for $\forall m,k$. Here, π specifies whether the task J_{mk} is cached at the UAV ($\pi_{mk}=1$) or not ($\pi_{mk}=0$). It is observed that this energy consists of the energy for computation (E_{mk}^{cp}) and communication (E_{mk}^{co}), respectively,

$$E_{mk}^{\mathsf{cp}} = \frac{\theta_m}{2} x_{mk} C_{mk} \big(f_{mk}^{\mathsf{ue}}\big)^2, \quad (15)$$

$$E_{mk}^{\mathsf{co}} = \frac{(1-x_{mk}) p_{mk} D_{mk}}{R_{mk}(\mathbf{p})} \quad (16)$$

where the constant $\theta_m/2$ is the power parameter for the UEs' computing energy consumption.

Similarly, the energy that is consumed by the k-th UAV to process the offloaded tasks is modelled as

$$E_{mk}^{\mathsf{uav}}(\pi_{mk}, x_{mk}) = \pi_{mk}[\frac{\theta_k}{2} C_{mk} \big(f_{mk}^{\mathsf{uav}}\big)^2] + (1-\pi_{mk})(1-x_{mk})[\frac{\theta_k}{2} C_{mk} \big(f_{mk}^{\mathsf{uav}}\big)^2], \quad (17)$$

where $\theta_k/2$ is the power parameter for the UAVs' computing energy consumption.

D. Problem Formulation

$$\min_{\mathbf{x},\pi,\mathbf{p},\mathbf{f}} \sum_{m=1}^{M}\sum_{k=1}^{K} E_{mk}^{\mathsf{ue}}(x_{mk}, \pi_{mk}, f_{mk}^{\mathsf{ue}}, \mathbf{p}), \quad (18a)$$

$$\text{s.t. } E_{mk}^{\mathsf{ue}}(x_{mk}, \pi_{mk}, f_{mk}^{\mathsf{ue}}, \mathbf{p}) \le E_{\max}^{\mathsf{ue}}, \forall m,k, \quad (18b)$$

$$\sum_{\forall m} E_{mk}^{\mathsf{uav}}(x_{mk}, \pi_{mk}) \le E_{\max}^{\mathsf{uav}}, \forall k, \quad (18c)$$

$$T_{mk}^{\mathsf{tot}}(x_{mk}, \pi_{mk}, f_{mk}^{\mathsf{ue}}, \mathbf{p}) \le T_{mk}^{\max}, \forall m,k, \quad (18d)$$

$$R_{mk}(\mathbf{p}) \ge R_{\min}, \forall m,k, \quad (18e)$$

$$\sum_{\forall m,k} \Big[\pi_{mk} f_{mk}^{\mathsf{uav}} + (1-\pi_{mk})(1-x_{mk}) f_{mk}^{\mathsf{uav}}\Big] \le F_{\max}^{\mathsf{uav}}, \quad (18f)$$

$$\sum_{\forall m} \pi_{mk} D_{mk} \le S_{\max}^{\mathsf{uav}}, \forall k, \quad (18g)$$

$$\mathbf{x} \in \mathscr{D}, \mathbf{p} \in \mathscr{P}, \mathbf{f} \in \mathscr{F}. \quad (18h)$$

In this paper, we aim to minimise the total energy consumption of the UEs by jointly considering offloading policies, caching strategies, transmit power, and computing rates of IoT and UAV-MEC under the constraints of URLLC QoS. Firstly, we define the following notations: $\mathscr{D} = \{x_{mk}\}_{,\forall m,k} | 0 \le x_{mk} \le 1, \forall m,k$, $\mathscr{P} = \{p_{mk}\}_{\forall m,k} | 0 \le p_{mk} \le P^{\max} mk, \forall m,$ and $\mathscr{F} = \{fmk^{\mathsf{ue}}\}_{\forall m,k} | 0 \le f_{mk}^{\mathsf{ue}} \le F_{\max}^{\mathsf{ue}}, \forall m$, which represent the feasible domains of the optimisation variables. With these considerations, the problem of minimising energy consumption is formulated as (18).

In (18), the constraints (18b) and (18c) indicate the maximum energy consumption requirements for the UE and UAV, respectively. Constraint (18d) ensures the maximum latency for each incoming task, while constraint (18e) guarantees the minimum transmission rate for uplink transmission. Additionally, constraints (18f) and (18g) define the upper limits for the computing and caching capacity of the ES.

III. Proposed Solution

Our main purpose is to minimise the UEs' energy consumption. However, it is challenging to solve problem (18) directly due to the non-convexity of the objective function and constraints. To solve problem (18), we propose an alternating optimisation algorithm (AOA)-based solution in which a subset of variables is solved while keeping other variables fixed. Hence, we have broken down our optimisation problem into three distinct subproblems, namely, the optimisation of caching policies, the optimisation of offloading policies, and the optimisation of communication and computation resources in a coordinated manner. The subsequent sections will provide a comprehensive presentation of how our proposed solution evolves.

A. Caching Policy Optimisation

In this subsection, our purpose is to find the optimal solution π^* for π with given current solutions of $(\mathbf{x}, \mathbf{f}, \mathbf{p})$. Our goal is to solve the following problem:

$$\min_{\pi_{mk} | \mathbf{x}^{(i)}, \mathbf{f}^{(i)}, \mathbf{p}^{(i)}} \sum_{\forall m}\sum_{\forall k} E_{mk}^{\mathsf{ue}}(\pi_{mk}), \quad (19a)$$

$$\text{s.t. (18b), (18c), (18d), (18f), (18g).} \quad (19b)$$

As defined before, $\pi \triangleq \{\pi_{mk}\}$ is an integer variable. Thus, (19) is obviously non-convex. To address this particular issue, we introduce the notation $e_{mk}^{\pi} = E_{mk}^{\mathsf{ue}}\Big(x_{mk}^{(i)}, f_{mk}^{ue(i)}, p_{mk}^{(i)}\Big)$ for all m and k. Then, to determine the optimal values of π at the i-th iteration, we arrange e_{mk}^{π} in descending order among the M UEs and select the task with a higher e^{π} (i.e., indicating greater energy consumption) for caching until the constraint (18g) is breached relative to the other constraints. It is worth noting that this method only requires a limited number of constraint checks (less than M).

B. Offloading Policy Optimisation

Within this subsection, our aim is to determine the subsequent iterative point of $\mathbf{x}$ with given $(\pi, \mathbf{f}, \mathbf{p})$. The considered optimisation can be represented as follows:

$$\min_{\mathbf{x}_{mk} | \pi^{(i+1)}, \mathbf{f}^{(i)}, \mathbf{p}^{(i)}} \sum_{\forall m}\sum_{\forall k} E_{mk}^{\mathsf{ue}}(\pi_{mk}), \quad (20a)$$

$$\text{s.t. (18b), (18c), (18d), (18f), (18h).} \quad (20b)$$

As all the constraints mentioned earlier are linear, (20) is evidently a convex program. It can be effectively solved

by using well-known optimisation packages, e.g., CVX [49], CVXPY [50].

C. Joint Communication and Computation Resources Optimisation

In this subsection, our objective is to determine the subsequent iterative point of the subset $(\mathbf{f}, \mathbf{p})$ with fixed subset variables of $(\pi, \mathbf{x})$. The considered optimisation problem in this subsection is expressed as

$$\min_{\mathbf{f},\mathbf{p}|\pi^{(i+1)},\mathbf{x}^{(i+1)}} \sum_{\forall m}\sum_{\forall k} E^{\mathsf{ue}}_{mk}\left(\pi_{mk}\right), \tag{21a}$$

$$\text{s.t. (18b), (18c), (18d), (18e), (18f), (18h).} \tag{21b}$$

Given that the constraints (18b), (18c), (18d), and (18e) are non-convex, we develop the solution by convexifying for these constraints with inner approximations as follows.

We first handle the constraint (18e) since it connects to transmission rate, latency, and energy equations. To do this, we rewrite (8) as $\gamma mk\left(\mathbf{p}\right) = \frac{p_{mk}}{q_{mk}(\mathbf{p})}$, where $q_{mk}\left(\mathbf{p}\right)$ is given by $q_{mk}\big(\mathbf{p}\big) = \big(\mathcal{I}_{mk}(\mathbf{p}) + N_0\big)\big/||\mathbf{h}_{mk}||^2$

Then, we develop the following approximation for the transmission rate:

$$R_{mk}\big(\mathbf{p}\big) \geq R^{(i)}_{mk}\big(\mathbf{p}\big) = \frac{B}{\ln 2}\left[W^{(i)}_{mk}\left(\mathbf{p}\right) - \omega Z^{(i)}_{mk}\left(\mathbf{p}\right)\right] \tag{22}$$

under the trusted regions $\forall m, i, k$:

$$\left(q_m(\mathbf{p}) + p_{mk}\right)\left[q_m(\mathbf{p}^{(i)}) + p^{(i)}_{mk}\right]^{-1} \leq 2\frac{q_m(\mathbf{p})}{q_m(\mathbf{p}^{(i)})}, \tag{23}$$

$$\left(q_m(\mathbf{p}) + p_{mk}\right)\left[q_m(\mathbf{p}^{(i)}) + p^{(i)}_{mk}\right]^{-1} \leq 2, \tag{24}$$

where $W^{(i)}_{mk}\left(\mathbf{p}\right)$, and $Z^{(i)}_{mk}\left(\mathbf{p}\right)$ are defined as in the Appendix, $\omega = \frac{\mathcal{Q}^{-1}(\epsilon)}{\sqrt{N}}$. Consequently, we innerly approximate constraint (18e) as

$$R^{(i)}_{mk}\big(\mathbf{p}\big) \geq R_{\min}, \forall m, i, k. \tag{25}$$

To handle the non-convex constraint (18b), (18c), and (18d), we introduce variable $\zeta \triangleq \{\zeta_{mk}\}_{\forall m,k}$, satisfying equation

$$\frac{1}{R^{(i)}_{mk}} \leq \zeta_{mk}. \tag{26}$$

Thus, (18d) can be approximately expressed as

$$T^{\mathsf{tot}}_{mk} \leq \frac{\pi^{(i+1)}_{mk} C_{mk}}{f^{\mathsf{uav}}_{mk}} + (1-\pi^{(i+1)}_{mk})\Big[\frac{x^{(i+1)}_{mk} C_{mk}}{f^{\mathsf{ue}}_{mk}} + (1-x^{(i+1)}_{mk}) D_{mk}\zeta_{mk} + \frac{(1-x^{(i+1)}_{mk})C_{mk}}{f^{\mathsf{uav}}_{mk}}\Big] \triangleq \mathcal{T}^{(i)}_{mk}. \tag{27}$$

By using ζ, we can rewrite constraint (18b) as

$$E^{\mathsf{ue}}_{mk} \leq (1-\pi^{(i+1)}_{mk})\big[\frac{\theta_m}{2} x^{(i+1)}_{mk} C_{mk}\big(f^{\mathsf{ue}}_{mk}\big)^2 + (1-x^{(i+1)}_{mk}) p_{mk} D_{mk}\zeta_{mk}\big]. \tag{28}$$

However, as (28) is still non-convex, we continue by following the inequality (29) given $x = p_{mk}$, $\bar{x} = p^{(i)}_{mk}$, $y = \zeta_{mk}$, $\bar{y} = \zeta^{(i)}_{mk}$:

$$xy \leq \frac{1}{2}\left(\frac{\bar{y}}{\bar{x}}x^2 + \frac{\bar{x}}{\bar{y}}y^2\right). \tag{29}$$

By applying (29) to non-convex equation (28), the non-convex objective function of total energy consumption thus can be approximately expressed as

$$\begin{aligned} E^{\mathsf{ue}}_{mk} &\leq (1-\pi^{(i+1)}_{mk})\big[\frac{\theta_m}{2} x^{(i+1)}_{mk} C_{mk}\big(f^{\mathsf{ue}}_{mk}\big)^2 \\ &+ \frac{(1-x^{(i+1)}_{mk})D_{mk}}{2}\Big(\frac{\zeta^{(i)}_{mk}}{p^{(i)}_{mk}} p^2_{mk} + \frac{p^{(i)}_{mk}}{\zeta^{(i)}_{mk}}\zeta^2_{mk}\Big)\big] \\ &\triangleq \mathcal{E}^{(i)}_{mk}. \end{aligned} \tag{30}$$

As a result, the resources allocation problem (21) is equivalent to the following convex program:

$$\min_{\mathbf{f},\mathbf{p}|\pi^{(i+1)},\mathbf{x}^{(i+1)}} \sum_{\forall m}\sum_{\forall k} \mathcal{E}^{(i)}_{mk}, \tag{31a}$$

$$\text{s.t.} \mathcal{E}^{(i)}_{mk}\left(\mathbf{x}^{(i+1)}_{mk}, \pi^{(i+1)}_{mk}, \mathbf{f}, \mathbf{p}\right) \leq E^{\mathsf{ue}}_{\max}, \tag{31b}$$

$$\text{(18f), (18h), (25), (26), (27).} \tag{31c}$$

D. Proposed AOA-based Algorithm

Based on above development, we present Algorithm 1 as a proposed solution for solving problem (18). To do this, for the i-th iteration, we denote $\mathcal{S}_1(\pi^{(i)})$, $\mathcal{S}_2(\mathbf{x}^{(i)})$, and $\mathcal{S}_3(\mathbf{f}^{(i)}, \mathbf{p}^{(i)})$ as updating subsets of subproblems (19), (20), and (31), respectively. We note that the choice of the convergence tolerance ε and the maximum iteration index $I_{\max}$ is to ensure that the algorithm efficiently obtains an optimal solution. When the difference between the energy consumption of current iteration and that in the previous iteration is sufficiently small (e.g., less than ε times), the algorithm is considered having reached convergence. The convergence behaviour of the algorithm will be clearly discussed in simulation results.

Algorithm 1 : Proposed AOA-based algorithm to solve (18).

1: **Initialisation**: Initialise $i = 0$ and generate initial points of optimisation variables $\mathcal{S}^{(0)}_1$, $\mathcal{S}^{(0)}_2$ and $\mathcal{S}^{(0)}_3$ satisfying the constraints in (19), (20) and (31); set the convergence tolerance as $\varepsilon = 10^{-3}$ and the maximum iteration index as $I_{\max} = 20$.
2: **while** (not convergence or $i < I_{\max}$) **do**
3: Solve problem (19) with the given $\mathcal{S}^{(i)}_2$, $\mathcal{S}^{(i)}_3$ to acquire the next points of (π^*) and update $\mathcal{S}^{(i+1)}_1$;
4: Solve problem (20) with the given $\mathcal{S}^{(i+1)}_1$, $\mathcal{S}^{(i)}_3$ to acquire the next points of $(\mathbf{x}^*)$ and update $\mathcal{S}^{(i+1)}_2$;
5: Solve problem (31) with the given $\mathcal{S}^{(i+1)}_1$, $\mathcal{S}^{(i+1)}_2$ to acquire the next points of $(\mathbf{p}^*, \mathbf{f}^*)$ and update $\mathcal{S}^{(i+1)}_3$;
6: $i := i + 1$;
7: **end while**
8: **Solution**: $\{\pi^*, \mathbf{x}^*, \mathbf{p}^*, \mathbf{f}^*\}$ and $\min\{\sum E^{\mathsf{ue}}_{mk}\}_{\forall m,k}$.

IV. Numerical Results

A. Parameter Settings

Our simulations involve a system model comprising two UAV-MECs and eight UEs. In terms of computation, the

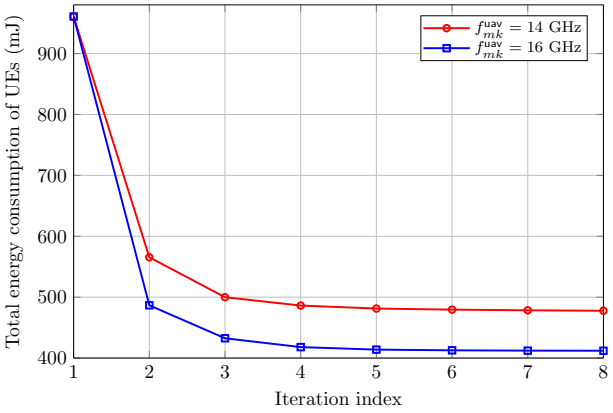

Fig. 2: Convergence behaviour of the proposed algorithm.

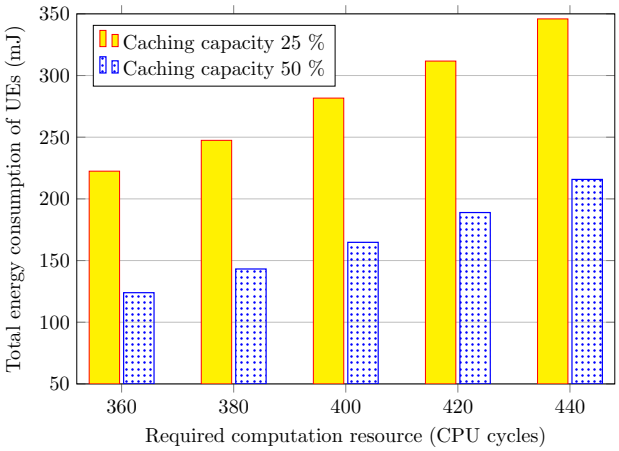

Fig. 3: Impact of caching capability.

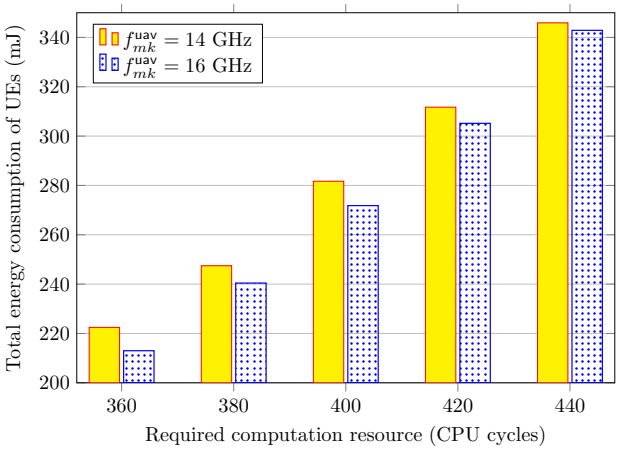

Fig. 4: Impact of UAVs' processing rate.

required computation resource of a task (CPU cycles) is varied in a range of $[360, 440]$ cycles. The data size of the task is 1354 bytes [27]. The processing rate of UAV-MEC is set to 14 GHz and 16 GHz. The maximum processing rate of UEs limits at 500 MHz. The coefficient parameter of computation energy is set to 10^{-26} Watt.s^3/cycles^3 [39]. The energy budget of the UEs and UAV-MECs are set to 800 mJ and 20 J, respectively. For communications, we assume that the UAV-UE networks are established in advance. The maximum transmission power of the UEs is 23 dBm and the system bandwidth is 5 MHz. The latency requirement for task processing is set to 2.5 ms. The URLLC decoding error probability is 10^{-5} and the blocklength is 100 [27], [31].

B. Numerical Results

In this subsection, we present simulation results to validate the effectiveness of the proposed algorithm. In particular, the convergence behaviour of Algorithm 1 is displayed in Fig. 2 while the impact of system parameters on the energy consumption are demonstrated in Fig. 4-8.

1) Convergence behaviour: To show the convergence behaviour of the proposed algorithm, we monitored the total energy consumption of UEs over iterations. Fig. 2 provides a clear visual representation of Algorithm 1 successfully reaching optimal energy consumption. Specfically, with the system model of 8 UEs and 2 UAV-MECs, our algorithm converged after just 8 iterations. Fig. 2 also reveals a noticeable reduction in UEs' energy consumption after the first iteration, followed by gradual decreases until convergence. This behaviour confirms that the proposed algorithm works appropriately and effectively to minimise the energy consumption of the UEs. Notably, the convergence is well established after only 8 iterations, making it practical for real-world implementation. Furthermore, Fig. 2 highlights the impact of UAV-MEC processing capacity on UEs' energy consumption. With the same setting at the initialisation (i.e., same task size), the higher processing rate of UAV-MEC is, the lower energy is consumped by UEs for task execution. This result underlines the efficiency of our proposed UAV-MEC architecture in facilitating task offloading and reducing UEs' energy usage.

2) Impact of caching capability: In order to show the impact of UAV-MEC caching capability on the optimal energy consumption of UEs, we present simulation results in Fig. 3 across various levels of required computational resources (CPU cycles), considering two distinct settings of UAV-MEC caching capacity. Evidently, when the UAV-MEC possesses a higher caching capability, UEs are enabled to offload a larger portion of their computational tasks to the UAV-MEC. This subsequently leads to a reduction in the energy consumption of UEs, reflecting the significance of the constraint (18g) and the UEs energy model defined in (14). Furthermore, Fig. 3 also illustrates the impact of required computational resources (C_{mk}) on UEs' energy consumption for task processing. As depicted in the graph, an increase in the value of C_{mk} corresponds to UEs consuming progressively more energy for both local processing and task offloading. For instance, when C_{mk} increases from 360 cycles to 440 cycles, the optimal total energy consumption experiences an approximate increase of 100 mJ.

3) Impact of UAVs' computing capacity: In the MEC-based architecture, the processing rate of edge servers plays an important role in reducing completion time as well as saving UEs energy consumption. To show that, Fig. 4 illustrates the

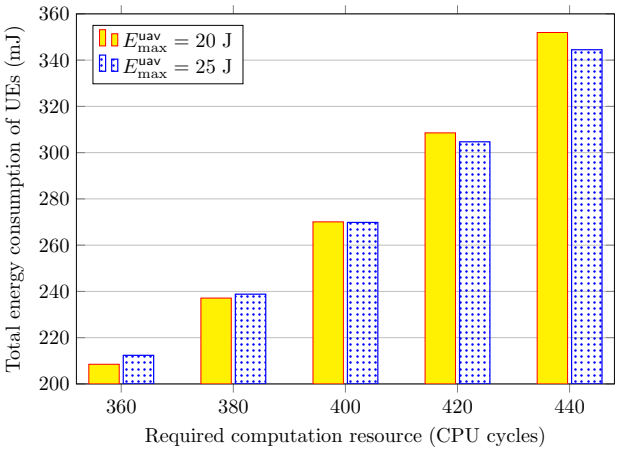

Fig. 5: Impact of UAVs' energy budget.

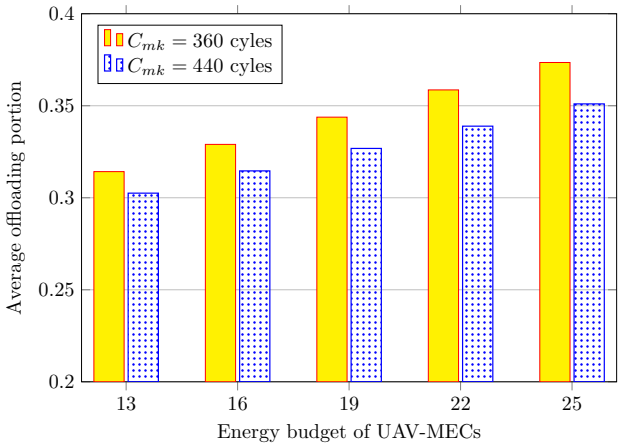

Fig. 6: Optimal task offloading versus UAVs' energy budget.

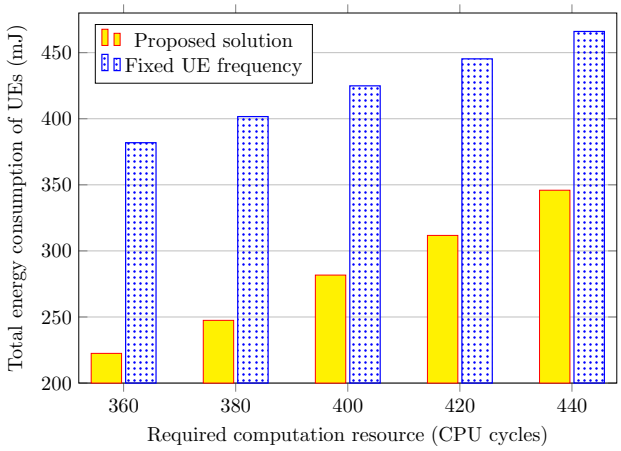

Fig. 7: Effectiveness of optimal UEs' processing rate.

obtained total energy consumption of UEs among various values of required computation resource (C_{mk}) with different settings of UAV-MEC's processing rate. In this regard, when the UAV-MECs are equipped with more powerful processors, the UEs consume less energy locally. This behaviour is because UEs can offload a larger portion of the tasks to UAV-MECs for edge processing. However, due to the limitation of the UAV's energy and computation resource budget (i.e., (18c), (18f)), it is challenging for the UAV-MECs to deal with huge computational tasks. This is why the gain of UEs' energy consumption between two f^{uav}_{mk} settings gradually reduces when the computational tasks become more complicated.

4) Impact of UAVs' energy budget: To show how UAVs' energy budget affects UEs' energy consumption in task processing, we conduct simulations with varied settings of $E^{\mathsf{uav}}_{\max}$. Fig. 5 compares the total energy consumption of UEs with different settings of $E^{\mathsf{uav}}_{\max}$ a range of C_{mk} values. As shown in the graph, the energy budget of UAV-MEC considerably impacts on the energy consumption of UEs for task execution, especially with highly complicated tasks. When a task requires more computing resource, it is likely to be offloaded to UAV-MEC to meet latency requirement, i.e., (18d) and UE's energy budget, i.e., (18b). This leads to an increase of UAV's energy consumption. Therefore, the higher UAV's energy budget is, the higher performance gain can be attained.

5) Optimal task offloading versus UAVs' energy budget: To illustrate the offloading pattern of UEs in our design, Fig 6 displays the average offloading portions of UEs among different level of the UAV-MECs' energy budget. As modelled in problem (18) and the energy consumption model of UAV-MECs (17), when the task is not cached in UAV-MECs, UEs can offload a portion of the task to UAV-MECs to meet latency requirement as well as to save energy for computation. Normally, when the energy budget of UAV-MEC increases, UEs can offload a higher portion of the task to the UAV-MECs. This is clearly demonstrated in Fig. 6. As we can see from the figure, when the energy budget of UAV-MECs increases from 13 J to 25 J, the average offloading portion gradually increases. Interestingly, under the same setting, the task which requires more CPU cycles, offloads less portion to UAV-MECs due to the UAV-MEC's constraints on processing capability and energy budget.

6) Effectiveness of optimal UEs' processing rate: To validate the performance gain of the proposed solution, we compare the total energy consumption of UEs between the proposed scheme and the conventional scheme (i.e., fixed UE processing rate). As formulated in problem (18), the processing rate of UEs (frequency) contributes to the energy for local processing of UEs, which is presented in (14). When increasing the processing rate of UEs, UEs can process the task more quickly to meet the latency requirement; however, the computation energy also increases exponentially. Therefore, it is important to solve for optimal control of UEs' processing rate, which can guarantee the latency constraint but save energy for UEs. Fig. 7 clearly shows that the proposed solution significantly outperforms the benchmark scheme for all settings of task complexity. More specifically, the proposed solution saves approximately 70% energy consumed by UEs to process all the tasks among varied settings of C_{mk}.

7) Effectiveness of optimal UEs power allocation: Similar to the UEs' processing rate, the optimal power allocation also vitally contributes to the minimised energy consumption of UEs. As modelled in (14), the transmission power of the UEs appears in the communication energy component. The

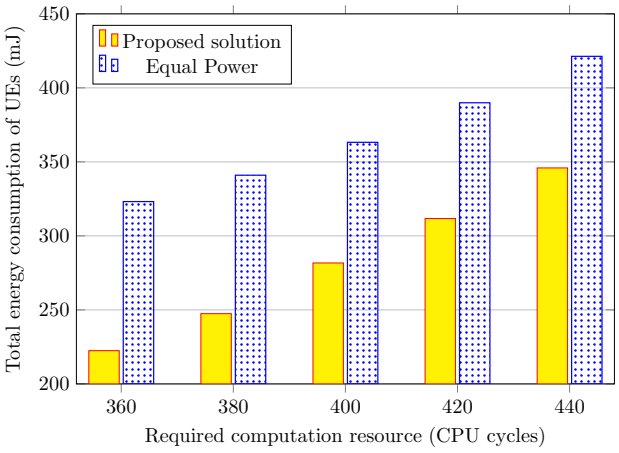

Fig. 8: Effectiveness of optimal UEs power allocation.

transmission power should be optimised to meet the QoS requirement of transmission rate (i.e., (18e)) as well as the latency requirement (18d). However, the transmission power also needs to be adjusted appropriately to save energy. Fig. 8 illustrates the energy gain of the proposed solution compared to a benchmark scheme with equal power allocation (i.e., $p_{mk} = 20$ dBm). As we can see from the graph, the optimal power scheme reduces appropriately 40% total UEs' energy consumption for most simulations. Fig. 7 and Fig. 8 can be reflected in the numerical results as shown in Table I.

TABLE I: Effectiveness of the proposed solution compared with the benchmark schemes.

CPU cycles	Alg. 1	Equal Fre.	Equal Power
360	222.466548	381.868948	323.178365
380	247.455066	401.629659	340.993792
400	281.704842	424.88549	363.205196
420	311.721097	445.309921	389.924318
440	345.906494	466.028229	421.295347

V. Conclusion

In conclusion, this paper has investigated the application of UAVs communications in enabling task offloading for URLLC-enabled intelligent autonomous transport systems. The formulated optimisation problem has taken into consideration various practical variables such as computing rate, task caching policies, and transmission power of UEs. To deal with the problem, we have decomposed the original problem into three subproblems and hae solved the problems iteratively based on successive convex approximations. The extensive simulation results clearly demonstrate the effectiveness of the proposed solution in terms of convergence speed, minimisation of UE energy consumption, and optimisation of resource allocation. For future work, it is promising to additionally consider UAVs deployment and path planning problems to fully realise the potentials of UAV-assisted intelligent autonomous transport systems. Development of efficient, real-time optimisation solutions is also an important direction to enhance the adaptability and flexibility in real-world deployment.

Acknowledgement

This research is funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.04-2021.11.

Appendix

We first rewrite the SINR (8) as $\gamma_{mk}(\mathbf{p}) = \frac{p_{mk}}{q_{mk}(\mathbf{p})}$. Then, we apply the following inequality [31, Eq. (74)]:

$$\ln\left(1+\frac{u}{v}\right) \geq a - \frac{b}{u} - cv, \tag{32}$$

in which

$$a = \ln\left(1+\bar{u}/\bar{v}\right) + 2\bar{u}/(\bar{u}+\bar{v}) \tag{33}$$

$$0 < b = \frac{\bar{u}^2}{\bar{u}+\bar{v}}, 0 < c = \frac{\bar{u}}{(\bar{u}+\bar{v})\bar{v}}. \tag{34}$$

By updating $u = p_{mk}$, $v = q_{mk}(\mathbf{p})$, $\bar{u} = p_{mk}^{(i)}$, and $\bar{v} = q_{mk}(\mathbf{p}^{(i)})$ to (33), we obtain the convex approximation of $W_{mk}(\mathbf{p})$.

To find a convex approximation of $Z_{mk}(\mathbf{p})$, we apply the inequality [31, Eq. (81)], we have

$$Z_{mk}(\mathbf{p}, \pi_{mk}^{(i)}) \leq= \phi_{mk}^{(i)} - \varphi_{mk}^{(i)} \frac{q_{mk}^2(\mathbf{p})}{\left(q_{mk}(\mathbf{p}) + p_{mk}\right)^2} \tag{35}$$

where

$$\phi_{mk}^{(i)} = \frac{\sqrt{1-\rho}}{2} + \frac{1}{2\sqrt{1-\rho}}, \tag{36}$$

$$\varphi_{mk}^{(i)} = \frac{1}{2\sqrt{1-\rho}}. \tag{37}$$

where $\rho = \left[1+\gamma_{mk}(\mathbf{p}^{(i)})\right]^{-2}$. We can see that $\frac{q_{mk}^2(\mathbf{p})}{\left(q_{mk}(\mathbf{p})+p_{mk}\right)^2}$ in (35) is still not convex [31] with respect to $(\mathbf{p})$. Therefore, we further approximate it by using inequalities [31, Eq. (76)] and [31, Eq. (77)] as

$$\begin{aligned}\frac{x^2}{y^2} = \frac{x^2}{y}\frac{1}{y} &\geq \frac{x^2}{y}\left(\frac{2}{\bar{y}} - \frac{y}{\bar{y}^2}\right) = \frac{2}{\bar{y}}\frac{x^2}{y} - \frac{x^2}{\bar{y}^2} \\ &\geq \frac{2}{\bar{y}}\left(\frac{2\bar{x}x}{\bar{y}} - \frac{\bar{x}^2 y}{\bar{y}^2}\right) - \frac{x^2}{\bar{y}^2}\end{aligned} \tag{38}$$

by replacing $x = q_{mk}^2(\mathbf{p})$, $\bar{x} = q_{mk}^2(\mathbf{p}^{(i)})$, $y = q_{mk}(\mathbf{p}) + p_{mk}$, and $\bar{y} = q_{mk}(\mathbf{p}^{(i)}) + p_{mk}^{(i)}$ regarding the regions given by (24) and (23), we finally obtain the convex approximation of $Z_{mk}(\mathbf{p})$.

References

[1] 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.

[2] 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. 2018.

[3] Y. Pan, K. Wang, C. Pan, H. Zhu, and J. Wang, "UAV-assisted and intelligent reflecting surfaces-supported terahertz communications," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1256–1260, Jun. 2021.

[4] T. Q. Duong, L. D. Nguyen, and L. K. Nguyen, "Practical optimisation of path planning and completion time of data collection for UAV-enabled disaster communications," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Tangier, Morocco, Jun. 2019, pp. 372–377.

[5] D. V. Huynh, T. Do-Duy, L. D. Nguyen, M.-T. Le, N.-S. Vo, and T. Q. Duong, "Real-time optimized path planning and energy consumption for data collection in unmanned ariel vehicles-aided intelligent wireless sensing," *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2753–2761, Apr. 2022.

[6] A. A. Nasir, "Latency optimization of UAV-enabled MEC system for virtual reality applications under Rician fading channels," *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1633–1637, Aug. 2021.

[7] D. V. Huynh, V.-D. Nguyen, S. R. Khosravirad, and T. Q. Duong, "Minimising latency for edge-cloud systems with ultra-reliable and low-latency communications," in *Proc. IEEE Int. Conf. Commun. (ICC'22)*, Seoul, Korea, May16–20 2022.

[8] F. Wang, J. Xu, and Z. Ding, "Multi-antenna NOMA for computation offloading in multiuser mobile edge computing systems," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2450–2463, Mar. 2019.

[9] C. You, K. Huang, H. Chae, and B.-H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1397–1411, Mar. 2017.

[10] 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 Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5389–5407, Oct. 2021.

[11] T. T. Vu, D. N. Nguyen, D. T. Hoang, E. Dutkiewicz, and T. V. Nguyen, "Optimal energy efficiency with delay constraints for multi-layer cooperative fog computing networks," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3911–3929, Jun. 2021.

[12] T. Q. Dinh, J. Tang, Q. D. La, and T. Q. S. Quek, "Offloading in mobile edge computing: Task allocation and computational frequency scaling," *IEEE Trans. Commun.*, vol. 65, no. 8, pp. 3571–3584, Aug. 2017.

[13] Y. Xiao and M. Krunz, "Distributed optimization for energy-efficient fog computing in the tactile Internet," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2390–2400, Nov. 2018.

[14] D. V. Huynh, V.-D. Nguyen, S. Chatzinotas, S. R. Khosravirad, H. V. Poor, and T. Q. Duong, "Joint communication and computation offloading for ultra-reliable and low-latency with multi-tier computing," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 2, pp. 521–537, Feb. 2023.

[15] 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 Transactions on Industrial Networks and Intelligent Systems*, vol. 8, no. 27, Apr. 2021.

[16] S. Zeng, X. Huang, and D. Li, "Joint communication and computation cooperation in wireless powered mobile edge computing networks with NOMA," *IEEE Internet Things J.*, vol. 10, no. 11, pp. 9849–9862, 2023.

[17] Y. Hao, Y. Miao, L. Hu, M. S. Hossain, G. Muhammad, and S. U. Amin, "Smart-Edge-CoCaCo: AI-enabled smart edge with joint computation, caching, and communication in heterogeneous IoT," *IEEE Netw.*, vol. 33, no. 2, pp. 58–64, 2019.

[18] X. Cao, F. Wang, J. Xu, R. Zhang, and S. Cui, "Joint computation and communication cooperation for energy-efficient mobile edge computing," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4188–4200, 2018.

[19] D. Wu, F. Wang, X. Cao, and J. Xu, "Joint communication and computation optimization for wireless powered mobile edge computing with D2D offloading," *J. Commun. Inf. Netw.*, vol. 4, no. 4, pp. 72–86, 2019.

[20] F. Chen, J. Fu, Z. Wang, Y. Zhou, and W. Qiu, "Joint communication and computation resource optimization in FD-MEC cellular networks," *IEEE Access*, vol. 7, pp. 168 444–168 454, 2019.

[21] S. Mostafa, C. W. Sung, and Y. Guo, "Joint computation and communication resource allocation with NOMA and OMA offloading for multi-server systems in F-RAN," *IEEE Access*, vol. 10, pp. 24 456–24 466, 2022.

[22] Y. Xu, B. Gu, R. Q. Hu, D. Li, and H. Zhang, "Joint computation offloading and radio resource allocation in MEC-based wireless-powered backscatter communication networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 6, pp. 6200–6205, 2021.

[23] X. Ge, Y. Sun, H. Gharavi, and J. Thompson, "Joint optimization of computation and communication power in multi-user massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4051–4063, 2018.

[24] B. Xie, Q. Zhang, and J. Qin, "Joint optimization of cooperative communication and computation in two-way relay MEC systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4596–4600, 2020.

[25] L. P. Qian, B. Shi, Y. Wu, B. Sun, and D. H. Tsang, "NOMA-enabled mobile edge computing for Internet of Things via joint communication and computation resource allocations," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 718–733, 2019.

[26] Y. Cui, L. Du, H. Wang, D. Wu, and R. Wang, "Reinforcement learning for joint optimization of communication and computation in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 13 062–13 072, 2021.

[27] 3GPP, "Study on scenarios and requirements for next generation access technologies," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.913, 2018, Version 15.0.0.

[28] 3GPP, "Release 16 description," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 21.916, 2020, Version 1.0.0.

[29] K. S. Kim, D. K. Kim, C. B. Chae, S. Choi, Y. C. Ko, J. Kim, Y. G. Lim, M. Yang, S. Kim, B. Lim, K. Lee, and K. L. Ryu, "Ultrareliable and low-latency communication techniques for tactile internet services," *Proc. IEEE*, vol. 10, no. 2, pp. 376–393, 2019.

[30] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[31] A. A. Nasir, H. D. Tuan, H. Nguyen, M. Debbah, and H. V. Poor, "Resource allocation and beamforming design in the short blocklength regime for URLLC," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1321–1335, Feb. 2021.

[32] H. Ren, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint pilot and payload power allocation for massive-MIMO-enabled URLLC IIoT networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 5, pp. 816–830, May 2020.

[33] H. Ren, C. Pan, K. Wang, Y. Deng, M. Elkashlan, and A. Nallanathan, "Achievable data rate for URLLC-enabled UAV systems with 3-D channel model," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1587–1590, 2019.

[34] K. M. S. Huq, S. Mumtaz, J. Rodriguez, P. Marques, B. Okyere, and V. Frascolla, "Enhanced C-RAN using D2D network," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 100–107, 2017.

[35] D. V. Huynh, S. R. Khosravirad, L. D. Nguyen, and T. Q. Duong, "Multiple relay robots-assisted URLLC for industrial automation with deep neural networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM'21)*, Madrid, Spain, Dec. 7–11 2021.

[36] A. A. Nasir, H. D. Tuan, H. Q. Ngo, T. Q. Duong, and H. V. Poor, "Cell-free massive MIMO in the short blocklength regime for URLLC," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5861–5871, Sep. 2021.

[37] H. Ren, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, "Joint power and blocklength optimization for URLLC in a factory automation scenario," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1786–1801, Mar. 2020.

[38] D. V. Huynh, V.-D. Nguyen, S. Chatzinotas, S. R. Khosravirad, H. V. Poor, and T. Q. Duong, "Joint communication and computation offloading for ultra-reliable and low-latency with multi-tier computing," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 2, pp. 521–537, Feb. 2022.

[39] D. V. Huynh, S. R. Khosravirad, A. Masaracchia, O. A. Dobre, and T. Q. Duong, "Edge intelligence-based ultra-reliable and low-latency communications for digital twin-enabled metaverse," *IEEE Wireless Commun. Lett.*, vol. 11, no. 8, pp. 1733–1737, Aug. 2022.

[40] T. Q. Duong, D. V. Huynh, S. R. Khosravirad, V. Sharma, O. A. Dobre, and H. Shin, "From digital twin to metaverse: The role of 6G ultra-reliable and low-latency communications with multi-tier computing," *IEEE Wireless Commun. Mag.*, vol. 30, no. 3, pp. 140–146, Jun. 2023.

[41] D. V. Huynh, V.-D. Nguyen, S. R. Khosravirad, V. Sharma, O. A. Dobre, H. Shin, and T. Q. Duong, "URLLC edge networks with joint optimal user association, task offloading and resource allocation: A digital twin approach," *IEEE Trans. Commun.*, vol. 70, no. 11, pp. 7669–7682, Nov. 2022.

[42] F. Pervez, A. Sultana, C. Yang, and L. Zhao, "Energy and latency efficient joint communication and computation optimization in a multi-UAV assisted MEC network," *IEEE Trans. Wireless Commun.*, vol. 23, no. 3, pp. 1728–1741, Mar. 2024.

[43] N. Ye, L. Chen, Q. Ouyang, and J. An, "Time-efficient data download for emergency UAV: Joint optimization of on-board computation and communication under energy constraint," *IEEE Trans. Veh. Technol.*, vol. 72, no. 10, pp. 13 718–13 722, Oct. 2023.

[44] J. Ji, K. Zhu, and D. Niyato, "Joint communication and computation design for UAV-enabled aerial computing," *IEEE Commun. Mag.*, vol. 59, no. 11, pp. 73–79, Nov. 2021.
[45] Y. Liu, J. Zhou, D. Tian, Z. Sheng, X. Duan, G. Qu, and V. C. Leung, "Joint communication and computation resource scheduling of a UAV-assisted mobile edge computing system for platooning vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8435–8450, Jul. 2021.
[46] T. Zhang, Y. Xu, J. Loo, D. Yang, and L. Xiao, "Joint computation and communication design for UAV-assisted mobile edge computing in IoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5505–5516, Feb. 2019.
[47] J. Zhou, D. Tian, Z. Sheng, X. Duan, and X. Shen, "Joint mobility, communication and computation optimization for UAVs in air-ground cooperative networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2493–2507, Mar. 2021.
[48] C. She, C. Yang, and T. Q. S. Quek, "Radio resource management for ultra-reliable and low-latency communications," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 72–78, Jun. 2017.
[49] M. Grant and S. Boyd, "CVX: MATLAB software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.
[50] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *J. Mach. Learning Res.*, vol. 17, no. 83, pp. 1–5, 2016.

Yijiu Li (Graduate Student Member, IEEE) is currently a PhD student at Queen's University Belfast, UK. Her research interests include ultra-reliable and low latency communications, unmanned aerial vehicle (UAV) communications, and mobile edge computing.

Dang Van Huynh (Member, IEEE) is currently a Postdoctoral Fellow at Memorial University, Canada. His research interests include resource allocation in wireless communications, edge/cloud computing, convex optimization, and applied artificial intelligence.

Van-Linh Nguyen (Member, IEEE) is with the Department of Computer Science and Information Engineering, National Chung Cheng University (CCU), Taiwan. His current research interests include cybersecurity, network/edge intelligence, autonomous driving, and vehicular networks.

Dac-Binh Ha (Member, IEEE) is 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.

Hans-Jurgen Zepernick (Senior Member, IEEE) is with the Blekinge Institute of Technology, Karlskrona, Sweden. His research interests include cooperative communications, cognitive radio networks, immersive mobile multimedia, and perceptual multimedia quality assessment.

Trung Q. Duong (Fellow, IEEE) is a Canada Excellence Research Chair (CERC) and a Full Professor at Memorial University, Canada. He is also the adjunct Chair Professor in Telecommunications at Queen's University Belfast, UK and a Research Chair of Royal Academy of Engineering, UK. He was a Distinguished Advisory Professor at Inje University, South Korea (2017-2019), an Adjunct Professor and the Director of Institute for AI and Big Data at Duy Tan University, Vietnam (2012-present), and a Visiting Professor (under Eminent Scholar program) at Kyung Hee University, South Korea (2023-2025). His current research interests include quantum communications, wireless communications, quantum machine learning, and optimisation.