

Energy-Efficient UAV-Enabled MEC Systems: NOMA, FDMA, or TDMA Offloading?

Qingjie Wu, Miao Cui, Guangchi Zhang, Beixiong Zheng, *Senior Member, IEEE*,
Xiaoli Chu, *Senior Member, IEEE*, and Qingqing Wu, *Senior Member, IEEE*

Abstract—Unmanned aerial vehicle (UAV)-enabled mobile edge computing (MEC) systems can use different multiple access schemes to coordinate multi-user task offloading. However, it is still unknown which scheme is the most energy-efficient, especially when the offloading blocklength is finite. To answer this question, this paper minimizes and compares the MEC-related energy consumption of non-orthogonal multiple access (NOMA), frequency division multiple access (FDMA), and time division multiple access (TDMA)-based offloading schemes within UAV-enabled MEC systems, considering both infinite and finite blocklength scenarios. Through theoretical analysis of the minimum energy consumption required by these three schemes, two novel findings are presented. First, TDMA consistently achieves lower energy consumption than FDMA in both infinite and finite blocklength cases, due to the degrees of freedom afforded by sequential task offloading. Second, NOMA does not necessarily achieve lower energy consumption than FDMA when the offloading blocklength is finite, especially when the channel conditions and the offloaded task data sizes of two user equipments (UEs) are relatively symmetric. Furthermore, an alternating optimization algorithm that jointly optimizes the portions of task offloaded, the offloading times of all UEs, and the UAV location is proposed to solve the formulated energy consumption minimization problems. Simulation results verify the correctness of our analytical findings and demonstrate that the proposed algorithm effectively reduces MEC-related energy consumption compared to benchmark schemes that do not optimize task offloading portions and/or offloading times.

The work of Miao Cui and Guangchi Zhang was supported in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515011980. The work of Beixiong Zheng was supported in part by the National Natural Science Foundation of China under 62571193, Grant 62201214, Grant 62331022, and Grant 62301171, the Natural Science Foundation of Guangdong Province under Grant 2023A1515011753, Grant 2024A1515010013, and Grant 2022A1515110484, the Guangdong program under Grant 2023QN10X446 and Grant 2023ZT10X148, the GJYC program of Guangzhou under Grant 2024D01J0079 and Grant 2024D03J0006, and the Fundamental Research Funds for the Central Universities under Grant 2024ZYGXZR087. The work of Qingqing Wu was supported by NSFC 62371289 and Shanghai JXW Fund JJ-GFWPT-01-24-0030. (*Corresponding authors: Miao Cui; Guangchi Zhang.*)

Qingjie Wu was with the School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China, and is currently with the School of Microelectronics, South China University of Technology, Guangzhou 511442, China (e-mail: qingjiwu55@163.com).

Miao Cui and Guangchi Zhang are with the School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China (e-mail: cuimiao@gdut.edu.cn; gc Zhang@gdut.edu.cn).

Beixiong Zheng is with the School of Microelectronics, South China University of Technology, Guangzhou 511442, China (e-mail: bxzheng@scut.edu.cn).

Xiaoli Chu is with the School of Electrical and Electronic Engineering, The University of Sheffield, S1 3JD Sheffield, U.K. (e-mail: x.chu@sheffield.ac.uk).

Qingqing Wu is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: qingqingwu@sjtu.edu.cn).

Corresponding authors: Miao Cui and Guangchi Zhang.

Index Terms—Mobile edge computing (MEC), multiple access, unmanned aerial vehicle (UAV), infinite/finite blocklength, energy consumption minimization.

I. INTRODUCTION

A. Background and Motivation

Mobile edge computing (MEC) is regarded as a promising solution to resolving the conflict between resource-constrained user equipment (UE) and resource-hungry mobile applications by offloading computation tasks of UE to nearby computing servers at the network edge [1]. In a conventional terrestrial MEC system, the task offloading performance is mainly limited by the poor wireless channels that typically suffer from various impairments such as shadowing and fading in addition to path loss. Moreover, in scenarios of limited or no communication infrastructure, such as disaster relief and emergency response [2], a fixed terrestrial MEC server may not be able to sustain computation offloading for all UEs in a large area. Due to the advantages of unmanned aerial vehicles (UAVs) in terms of high mobility, easy deployment, line-of-sight (LoS) connections, and wide coverage [3], [4], UAV-enabled MEC has been proposed as a potential technique to tackle the above challenges. Compared with its terrestrial counterpart, a UAV-carried MEC server can further enhance the flexibility and resilience of MEC and improve the UEs' experience [5].

The energy efficiency is a key design aspect of a UAV-enabled MEC system, since both the UAV and the UEs are powered by batteries of a limited capacity [6]. The MEC-related energy consumption includes the offloading energy consumption (i.e., communication energy consumption) of the UEs and the computation energy consumption of the UEs and the UAV [7]. In a multi-UE UAV-enabled MEC system, both the communication and computation energy consumption will be affected by the multiple-access offloading scheduling and resource allocation [8]. However, the energy efficiency of different multiple access offloading schemes has not been sufficiently studied for a UAV-enabled MEC system.

In addition, the fifth generation (5G) and beyond 5G (B5G) networks are expected to support enhanced mobile broadband (eMBB) and ultra-reliable low latency communication (URLLC) services [9]. For eMBB services, the tasks offloaded by the UEs are packetized into long packets, e.g., thousands of bytes per packet, where the transmission blocklength can be assumed to be infinite and the decoding error can be assumed to be zero [10]. In this case, the Shannon capacity theorem

TABLE I
COMPARISON OF THIS PAPER WITH RELATED LITERATURE.

Reference	UAV	Multiple access scheme			Blocklength		Key findings
		NOMA	FDMA	TDMA	Infinite	Finite	
[24]		✓	✓		✓		NOMA achieves lower energy consumption than FDMA for tasks characterized by large data sizes and stringent delay requirements.
[25]		✓		✓	✓		Hybrid NOMA is more energy-efficient than TDMA for time-demanding tasks, but this performance advantage diminishes for delay-tolerant tasks.
[26]		✓		✓	✓		NOMA achieves substantial energy consumption performance gains over TDMA-based offloading in multi-antenna MEC systems.
[27]		✓	✓	✓	✓		When minimizing a linear combination of completion time and energy consumption, NOMA outperforms both TDMA and FDMA.
[28]	✓	✓		✓	✓		In terms of total mobile energy consumption, NOMA offloading is preferable to TDMA, unless the computation delay requirements are stringent.
[29]	✓	✓	✓		✓		NOMA requires lower energy consumption than FDMA, and this advantage becomes more pronounced as the number of users increases.
[33]		✓		✓	✓		Due to the flexible IRS beamforming, TDMA can achieve a computation rate comparable to or even exceeding that of NOMA in an IRS-aided wireless-powered MEC system.
[30]		✓				✓	IRS-aided NOMA transmissions have been designed considering the practical implications of the finite-blocklength regime in MEC networks.
This paper	✓	✓	✓	✓	✓	✓	TDMA achieves lower energy consumption than FDMA. Furthermore, NOMA does not necessarily outperform FDMA when the offloading blocklength is finite, particularly when the channel conditions and offloaded task data sizes of the UEs are relatively symmetric.

is applicable to characterize the offloading rate. In contrast, for URLLC services, the size of a data packet is quite small, e.g., 20 or 32 bytes per packet, and the decoding errors in offloading transmission cannot be ignored due to the finite blocklength [11]. Thus, the Shannon capacity cannot well characterize the offloading rate in the finite blocklength case [12]. Since the finite-blocklength transmission suffers from a lower data rate and a higher decoding error probability than the infinite-blocklength transmission, the multiple access and resource allocation schemes should be designed for these two transmission modes separately in a UAV-enabled MEC system, which remains a gap in the current literature. In particular, for non-orthogonal multiple access (NOMA), different from the infinite-blocklength transmission, the successive interference cancellation (SIC) failure probability should be considered in the finite-blocklength transmission [13], [14], where the decoding performance of one UE's signal will be affected by the decoding error of the other UEs' signals that have been decoded earlier during SIC.

Motivated by the above, and to determine which multiple access offloading scheme achieves the highest energy efficiency in UAV-enabled MEC, this work aims to minimize and compare the MEC-related energy consumption. This is achieved by investigating various multiple access offloading schemes for both infinite and finite blocklength transmissions in a UAV-enabled MEC system.

B. Related Works

Conventional orthogonal multiple access (OMA) schemes, such as frequency division multiple access (FDMA) and time division multiple access (TDMA), are commonly used in MEC systems to support multi-UE task offloading [2], [15]–[20]. The average weighted energy consumption of the UEs and the UAV in a UAV-enabled MEC system was minimized in [2], where all UEs were allocated the same bandwidth in FDMA-based task offloading. The computation efficiency of

a multi-UAV-enabled MEC system was maximized in [15], where each UAV serves its associated UEs with FDMA. In [16], for a UAV-enabled MEC system with finite-blocklength FDMA offloading, the computation latency was minimized by jointly optimizing the computing times and central processing unit (CPU) frequencies of the UEs and the UAV, the offloading bandwidths of the UEs, and the three-dimensional location of the UAV. In the wireless-powered multi-UE MEC system proposed in [17], the UEs offloaded their tasks based on TDMA. The resource allocation strategies for both TDMA and orthogonal frequency-division multiple access (OFDMA) multi-UE MEC systems were studied in [18] to maximize the computation capacity while reducing energy consumption and latency. In [19], the authors proposed a new solution that UAVs offload computation-intensive tasks to multiple edge servers with an effective multi-leader multi-follower Stackelberg game strategy. In [20], the UAV-aided MEC system allows the UAV to connect to only one UE at a time. In these OMA-based MEC systems, only a single UE is permitted to perform task offloading within each designated time/frequency resource block.

Due to the superposition coding at the transmitter and the SIC at the receiver, NOMA can achieve high spectral efficiency and massive connectivity for MEC [21]. The resource allocation for latency minimization in NOMA-based MEC networks was investigated in [22] and [23]. In [22], the hybrid NOMA-MEC policy allows two UEs to offload their tasks in a shared time slot, wherein only one UE completes its task offloading while the other UE will offload its remaining task in the next time slot solely occupied by itself. While [23] considered a pure NOMA offloading scheme, where multiple UEs can offload their tasks to the MEC server simultaneously in the same frequency band. The NOMA-based offloading was applied in [24] and [25] to minimize the energy consumption. Specifically, both the task uploading and result downloading in [24] benefited from the advantages of NOMA, achiev-

ing energy-efficient MEC. In [25], the energy consumption of three MEC offloading strategies, including TDMA, pure NOMA, and hybrid NOMA, was investigated for a general and arbitrary network status. The multi-antenna NOMA technique was integrated into multi-UE computation offloading in [26], which jointly optimized the resource allocation and the decoding order for SIC to minimize the weighted sum energy consumption of all UEs. In [27], a linear combination of the computation time and total energy consumption was minimized for a user-group NOMA-MEC system. In [28], the total energy consumption of the UEs in a frequency-division-duplex UAV-enabled MEC system was minimized for both the TDMA and NOMA offloading schemes. To extend the MEC coverage and improve the system flexibility, NOMA was also considered in the multi-UAV-enabled MEC system [29]. The aforementioned studies demonstrated the performance advantages of NOMA over OMA in MEC systems. However, these investigations typically assumed an infinite offloading blocklength and perfect SIC for NOMA. Although [30] studied the energy efficiency of NOMA-based MEC in the finite blocklength regime, SIC was still assumed to be error free.

However, we note that NOMA does not always achieve performance advantages over OMA. For example, the throughput regions of OMA and NOMA are identical in the symmetric channel case [31]. In the intelligent reflecting surface (IRS)-assisted downlink communication network [32], TDMA is more energy-efficient than NOMA when two nearby users are paired and their rates are symmetric, since the IRS phase shift vector can be set differently for the two users in different time slots of TDMA. Further, an IRS-aided wireless-powered MEC system was studied in [33], and the authors unveiled that the TDMA offloading scheme can achieve a better computation rate than that of NOMA when the IRS beamforming is flexibly adapted for uplink offloading. Table I provides a comparative summary of system models and contributions, contrasting our work with existing state-of-the-art multiple-access MEC literature. From this comparison, it is evident that comprehensive studies on UAV-enabled MEC systems employing finite-blocklength NOMA offloading are still missing. On the other hand, it remains unknown yet whether NOMA can outperform OMA or not (e.g., in terms of energy consumption) in a UAV-enabled MEC system when the transmission blocklength is finite.

C. Contributions

In this paper, we investigate the MEC-related energy consumption minimization problem for a two-UE UAV-enabled MEC system while considering three different multiple access offloading schemes, i.e., NOMA, FDMA, and TDMA, for both the infinite and finite blocklength cases. Specifically, the MEC-related energy consumption is defined as the sum of the energy consumption caused by the local computing at the UEs, remote computing at the UAV, and the UEs' transmission for task offloading, and it is minimized by jointly optimizing the portions of task offloaded, the offloading times of all UEs, and the UAV location for each multiple access scheme. Our main contributions are summarized as follows:

- For each considered offloading scheme, the expression of the minimum required transmit power of each UE that can satisfy the information-causality constraint for computation offloading is derived, and the energy minimization problem that subjects to the computation latency constraint, maximum CPU frequency constraint, and maximum transmit power constraint is formulated. Notably, in the finite-blocklength NOMA offloading scheme, since the perfect SIC cannot be guaranteed, both SIC success and failure cases are considered, and a closed-form expression for the average minimum transmit power of these two cases is obtained.
- The relationships of minimum MEC-related energy consumption required by the considered multiple access offloading schemes are analyzed, and the analytical results show that TDMA can always achieve a lower energy consumption than FDMA due to the freedom of remote-computing time brought by the sequential task offloading of the UEs. In addition, NOMA can achieve a lower energy consumption than FDMA in the infinite blocklength case, but the same conclusion cannot be reached in the finite blocklength case since the SIC failure should be considered, especially when the channel conditions and the offloaded task data sizes of the two UEs are relatively symmetric.
- Since the formulated MEC-related energy consumption minimization problem is non-convex, it is solved by devising an alternating optimization algorithm, where the original problem is decomposed into three subproblems that optimize the portion of task offloaded by each UE, the offloading time of each UE, and the UAV location, respectively, based on the block coordinate descent (BCD) method. Furthermore, in the subproblems, the expressions of the minimum transmit power are transformed into a form containing the difference of two convex functions, which are handled by the successive convex approximation (SCA) technique.
- Simulation results verify the analytical results and show that the proposed algorithm can effectively reduce the MEC-related energy consumption, as compared with the benchmark schemes without optimizing the portions of task offloaded or the offloading times. The results show that the advantage of joint optimization is more remarkable when the computational burden is heavier or the computation latency requirement is more stringent for each UE. The results also show that the decoding error probability has a significant impact on NOMA offloading with finite-blocklength, due to nonnegligible SIC failures in this case.

The rest of this paper is organized as follows. Section II presents the system models and problem formulations of all the considered multiple access offloading schemes in both infinite blocklength and finite blocklength cases. Section III provides the theoretical performance comparison of difference offloading schemes. Section IV proposes the alternating optimization algorithm. Section V shows the simulation results to evaluate the effectiveness of the proposed algorithm. Section

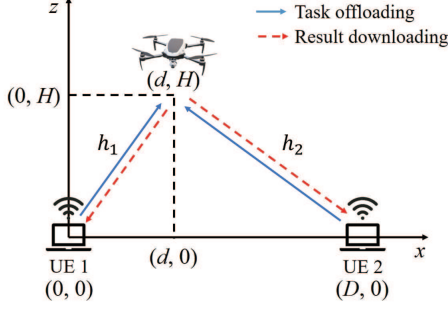


Fig. 1. A two-UE UAV-enabled MEC system.

VI concludes the paper.

II. SYSTEM MODEL AND PROBLEM FORMULATION

As shown in Fig. 1, we consider a two-UE UAV-enabled MEC system, where the two UEs on the ground can offload parts of their computation tasks to a rotary-wing UAV that carries an MEC server for remote computing, and the UAV and UEs are each equipped with a single antenna. UE $k \in \{1, 2\}$ is required to compute an L_k -bit task in a given time length $T_{\max}$. To minimize the sum transmission distance of the two offloading links, the UAV hovers above the horizontal line between the two UEs. Accordingly, the locations of the UAV and UEs are contained within a vertical two-dimensional Cartesian coordinate system. Without loss of generality, we consider that UE 1 is located at the origin $(0, 0)$, and the locations of the UAV and UE 2 are (d, H) and $(D, 0)$, respectively, where d , D , and H denote the horizontal distance between the UAV and UE 1, the distance between the two UEs, and the fixed altitude of the UAV, respectively.

It is worth pointing out that the considered two-UE system can be extended to scenarios of more than two UEs, where the UEs can be grouped into pairs and our proposed system can be applied for each pair. And if different UE pairs are allocated with different orthogonal time/frequency resources, they can offload their task to the UAV without multi-pair interference. Furthermore, by designing a two-timescale framework, we can also extend the considered system to a mobility scenario where the UAV provides computation offloading services for a wide region through cruising flight. Specifically, the UAV's trajectory is optimized in the large timescale, while the computation offloading and resource allocation are optimized in the small timescale.

Since UAV typically has a high probability of establishing an LoS air-to-ground communication link [3], the free space channel model can be adopted. Thus, the channel power gains from UEs 1 and 2 to the UAV can be represented as

$$h_1 = \frac{\beta_0}{H^2 + d^2}, \quad h_2 = \frac{\beta_0}{H^2 + (D - d)^2}, \quad (1)$$

respectively, where β_0 is the channel power gain at a reference distance of $d_0 = 1$ m. To design the communication and computation coordination between the UAV and UEs, a central controller can be deployed in the MEC network to collect

the global channel state information (CSI) and computation-related information. Assuming a quasi-static channel environment for the duration of low-latency task offloading, perfect CSI is assumed to be known for the remainder of this paper to facilitate tractable analysis and derive fundamental performance benchmarks.¹

The partial computation offloading mode is considered in the MEC. Denoting the portion of task offloaded by UE k by $0 \leq \rho_k \leq 1$, the data sizes of local-computing and remote-computing can be expressed as $(1 - \rho_k)L_k$ and $\rho_k L_k$, respectively. We consider scenarios such as factory automation and wireless federated learning, where the computation results are usually some control commands or model parameters whose data size is typically negligible compared to that of the offloaded task [14], [34]. For example, event-triggered control based on binary neural networks for image recognition outputs a binary parameter to indicate the presence of a target object [35], and the sparse gradient triggering mechanism in federated learning uploads a 1-bit signal to indicate the need for global model update [36]. Thus, we do not consider the latency and energy consumption caused by the result downloading process from the UAV to UEs. Thus, if UE k 's task offloading time is t_k , the available execution times of its local and remote computing are $T_{\max}$ and $T_{\max} - t_k$, respectively.

By applying the dynamic voltage and frequency scaling (DVFS) technique, the UEs and UAV can control the energy consumption for local and remote computing by adjusting their CPU frequencies for each cycle, respectively [37]. To minimize the computation-related energy consumption, based on [17], the CPU frequencies for UE k 's local and remote computing during their available execution times should be adjusted as

$$f_k^{\text{loc}} = \frac{(1 - \rho_k)c_k L_k}{T_{\max}}, \quad f_k^{\text{rem}} = \frac{\rho_k c_k L_k}{T_{\max} - t_k}, \quad (2)$$

respectively, where c_k denotes the number of CPU cycles required to compute 1 bit of UE k 's task. Then, the energy consumption of UE k 's local computing and remote computing at the UAV can be characterized as

$$E_k^{\text{loc}} = \kappa_k T_{\max} (f_k^{\text{loc}})^3, \quad (3a)$$

$$E_k^{\text{rem}} = \kappa_U (T_{\max} - t_k) (f_k^{\text{rem}})^3, \quad (3b)$$

respectively, where κ_k and κ_U denote the effective capacitance coefficients of UE k and the UAV, respectively.

The energy consumption of each UE includes local-computing energy consumption and task-offloading energy consumption, and that of UAV mainly includes remote-computing energy consumption and its propulsion energy consumption. According to [38], the propulsion energy consumption of the UAV is determined by physical parameters like aircraft weight, air density, rotor disc area, and so on and cannot be optimized by communication and computation

¹To improve the robustness and account for CSI error, a bounded error model can be considered to minimize the worst-case MEC-related energy consumption.

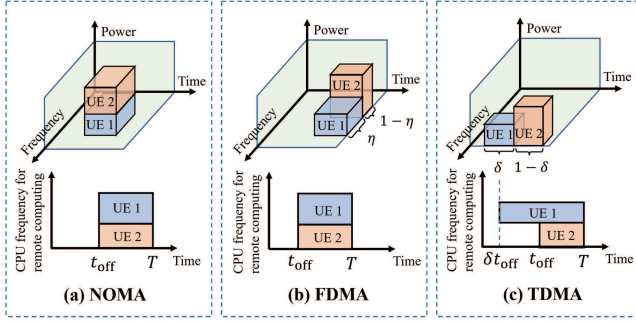


Fig. 2. Illustrations of the time-frequency transmit power allocations and the remote computing CPU frequency allocations in different offloading schemes.

resource allocation, so we only consider the communication-related and computation-related energy consumption and define the MEC-related energy consumption of UE k as

$$E_k^{\text{MEC}} = E_k^{\text{loc}} + E_k^{\text{rem}} + E_k^{\text{off}}, \quad (4)$$

where $E_k^{\text{off}} = P_k t_k$ is the energy consumption of offloading, and P_k is UE k 's transmit power.

It can be observed in (4) that the MEC-related energy consumption of the UEs is determined by the portions of task offloaded, transmit powers, and offloading times of the UEs, which are related to the multiple access schemes in the task offloading. We consider three main multiple access schemes, namely NOMA, FDMA, and TDMA. These three multiple access offloading schemes will be sequentially modeled and analyzed below in both the infinite blocklength and finite blocklength cases, which correspond to the computation-intensive tasks in eMBB services and the mission-critical tasks in URLLC services, respectively. Additionally, the UAV location directly impacts the quality of the offloading channels, making it a crucial factor in designing a computation offloading strategy that balances communication-related energy consumption and computation-related energy consumption.

A. NOMA Offloading

By using NOMA, the UEs can offload tasks to the UAV simultaneously over the same time-frequency resource block as shown in Fig. 2(a).² As such, the offloading bandwidths and times of two UEs can be expressed as $B_1 = B_2 = B$ and $t_1 = t_2 = t$, respectively, where B is the total bandwidth and t is the offloading time. By letting $x_k \in \mathbb{C}$ denote the UE k 's task-bearing signal with unit power for uplink offloading, i.e., $\mathbb{E}[|x_k|^2] = 1$, the received signal $y \in \mathbb{C}$ at the UAV is then expressed as

$$y_{\text{NOMA}} = \sqrt{P_1} h_1 x_1 + \sqrt{P_2} h_2 x_2 + z_{\text{NOMA}}, \quad (5)$$

where $z_{\text{NOMA}} \sim \mathcal{CN}(0, \sigma_{\text{NOMA}}^2)$ denotes the additive white Gaussian noise (AWGN) at the UAV with $\sigma_{\text{NOMA}}^2 = BN_0$ and N_0 being the noise power and the noise power spectrum density, respectively. The UAV performs SIC to decode two

UEs' signals in descending order of their channel power gains. Note that since UEs 1 and 2's channel power gains h_1 and h_2 depend on the UAV location, we may have $2! = 2$ permutations for the decoding order, i.e., (i) the UAV first decodes UE 1's signal and then decodes UE 2's signal, and (ii) the UAV first decodes UE 2's signal and then decodes UE 1's signal. In the following, we choose order (i) to analyze, and the other order can be analyzed similarly.

1) *Infinite Blocklength Case:* In this case, the transmission blocklength of two UEs is assumed to be infinite, i.e., $N \rightarrow \infty$. Based on Shannon's channel coding theorem, the decoding error probability can be made arbitrarily small if the channel codes are infinitely long. Thus, the decoding error probability is negligible during the SIC process in the infinite-blocklength NOMA offloading scheme [39], [40]. Specifically, the UAV first decodes the signal from the stronger UE (i.e., UE 1), treating the signal from the weaker UE (i.e., UE 2) as interference. Once UE 1's signal is successfully decoded, it is reconstructed and canceled from the composite signal. The UAV then proceeds to decode UE 2's signal without suffering from any interference. Thus, the signal-to-interference-plus-noise ratio (SINR) of UE 1's signal and the signal-to-noise ratio (SNR) of UE 2's signal at UAV are expressed as

$$\gamma_{1,1}^{\text{NOMA}} = \frac{P_1 h_1}{P_2 h_2 + BN_0}, \quad \gamma_{1,2}^{\text{NOMA}} = \frac{P_2 h_2}{BN_0}, \quad (6)$$

respectively.³

For remote computing at the UAV, the information-causality constraint should be satisfied, i.e., the offloaded task data can only be computed at the UAV if it has already been previously received from the UEs. The information-causality constraint for UE k 's computation offloading can be expressed as

$$Bt \log_2 (1 + \gamma_{1,k}^{\text{NOMA}}) \geq \rho_k L_k, \quad \forall k, \quad (7)$$

where the left-hand side (LHS) of (7) represents the effective offloading throughput from UE k to the UAV, in which the offloading data rate is measured by the Shannon capacity formula, while the right-hand side (RHS) denotes the data size of UE k 's remote-computing task.

To ensure that UEs' offloaded tasks can be efficiently received by the UAV for remote computing in the quantity required, according to (7), the minimum transmit powers of UEs 1 and 2 are respectively obtained as

$$P_{1,1}^{\text{NOMA}} = \frac{\Upsilon_1(\rho_1, t)(\Upsilon_2(\rho_2, t) + 1)}{\bar{h}_1}, \quad (8a)$$

$$P_{1,2}^{\text{NOMA}} = \frac{\Upsilon_2(\rho_2, t)}{\bar{h}_2}, \quad (8b)$$

where $\Upsilon_k^1(\rho_k, t) \triangleq \exp\left(\frac{\ln 2 \rho_k L_k}{Bt}\right) - 1$ is the minimum SINR/SNR, and $\bar{h}_k = \frac{h_k}{BN_0}$ denotes the normalized channel power gain from UE k to the UAV. Note that we do not constrain any power allocation order between the two UEs, since the power allocation in NOMA may not necessarily follow the reverse order of the UEs' channel gains [43].

²Power-domain NOMA is considered in this paper, where the UAV exploits differences in UEs' received power levels to distinguish between different UEs based on SIC.

³Accounting for the imperfection in SIC resulting from practical factors such as channel estimation errors, $\gamma_{1,2}^{\text{NOMA}}$ should be remodeled to include a residual interference parameter similar to [41] and [42].

As shown in (8), the transmit power of the UE depends on various factors including the UAV location, offloading time, and offloading task data sizes.

Thus, according to (2)–(4) and (8), the MEC-related energy consumption of NOMA offloading in the infinite blocklength case is expressed as

$$E_I^{\text{NOMA}}(\rho, t, d) = \sum_{k=1}^2 E_k^{\text{MEC}} = \sum_{k=1}^2 \left(\kappa_U \frac{(\rho_k c_k L_k)^3}{(T_{\max} - t)^2} + \kappa_k \frac{((1 - \rho_k) c_k L_k)^3}{T_{\max}^2} + P_{I,k}^{\text{NOMA}} t \right), \quad (9)$$

and the problem that jointly optimizes the UAV location d , the portion of task offloaded by each UE $\rho \triangleq \{\rho_1, \rho_2\}$, and the offloading time t to minimize $E_I^{\text{NOMA}}(\rho, t, d)$ can be formulated as⁴

$$\text{(NOMA-I)} \quad \min_{\rho, t, d} E_I^{\text{NOMA}}(\rho, t, d) \quad (10a)$$

$$\text{s.t.} \quad f_1^{\text{rem}} + f_2^{\text{rem}} \leq f_{U, \max}, \quad (10b)$$

$$f_k^{\text{loc}} \leq f_{k, \max}, \quad \forall k, \quad (10c)$$

$$P_{I,k}^{\text{NOMA}} \leq P_{k, \max}, \quad \forall k, \quad (10d)$$

$$0 \leq \rho_k \leq 1, \quad \forall k, \quad (10e)$$

$$0 \leq t \leq T_{\max}, \quad (10f)$$

$$d \leq \frac{1}{2}D, \quad (10g)$$

where $f_{U, \max}$ and $f_{k, \max}$ are the maximum available CPU frequencies of the UAV and UE k , respectively, and $P_{k, \max}$ is the maximum transmit power of UE k . (10b) ensures that the sum of remote-computing CPU frequencies allocated to the UEs does not exceed the maximum available CPU frequencies of the UAV. (10c) and (10d) mean that UE k 's CPU frequency f_k^{loc} and transmit power $P_{I,k}^{\text{NOMA}}$ are upper bounded by $f_{k, \max}$ and $P_{k, \max}$, respectively. (10e) and (10f) are the value ranges of $\{\rho_k\}$ and t , respectively, ensure the tasks required to be computed will be completed within the given time. (10g) exists to satisfy the assumptions that $h_1 \geq h_2$ and the UAV first decodes UE 1's signal during SIC as we analyzed above.

2) *Finite Blocklength Case*: In this case, the offloading transmission blocklength of two UEs is finite and can be given by $N_1 = N_2 = N = Bt$ [11]. Since the blocklength is short and the decoding error probability cannot be ignored, the offloading data rate should be characterized by the data rate expression derived in [12]. For a given decoding error probability ϵ_k , the information-causality constraint for computation offloading can be expressed as

$$(1 - \epsilon_k)N \underbrace{\left(\log_2(1 + \gamma_{F,k}^{\text{NOMA}}) - \sqrt{\frac{V_k}{N}} \frac{Q^{-1}(\epsilon_k)}{\ln 2} \right)}_{\text{finite-blocklength offloading data rate}} \geq \rho_k L_k, \quad \forall k, \quad (11)$$

where the LHS of (11) represents the effective offloading throughput from UE k to the UAV, V_k is the channel dispersion

⁴Note that since the altitude of the UAV H is assumed to be fixed, the location of the UAV can be expressed by its horizontal distance d only.

that can be approximated as 1 for convenience [9], and $Q^{-1}(\cdot)$ is the inverse function of $Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$.

Note that the perfect SIC cannot be guaranteed if the blocklength is finite [13], so both SIC success and failure should be considered. The UAV first decodes UE 1's signal while the interference caused by UE 2 is treated as noise, so the SINR of UE 1's signal at the UAV is expressed as

$$\gamma_{F,1}^{\text{NOMA}} = \frac{P_1 h_1}{P_2 h_2 + B N_0}. \quad (12)$$

If SIC succeeds, UE 1's signal is correctly decoded and completely canceled, the probability of which is $1 - \epsilon_1$, and the SNR of UE 2's signal is expressed as

$$\hat{\gamma}_{F,2}^{\text{NOMA}} = \frac{P_2 h_2}{B N_0}. \quad (13)$$

Otherwise, if SIC fails, UE 1's signal is not decoded correctly, the probability of which is ϵ_1 , it cannot be successfully canceled, and thus the SINR of UE 2's signal is expressed as

$$\tilde{\gamma}_{F,2}^{\text{NOMA}} = \frac{P_2 h_2}{P_1 h_1 + B N_0}. \quad (14)$$

By letting $\gamma_{F,2}^{\text{NOMA}} = \hat{\gamma}_{F,2}^{\text{NOMA}}$ and combining (11), (12), and (13), the minimum transmit powers of UEs 1 and 2 if the SIC succeeds can be obtained as

$$\hat{P}_{F,1}^{\text{NOMA}} = \frac{\Upsilon_1^F(\rho_1, N)(\Upsilon_2^F(\rho_2, N) + 1)}{\bar{h}_1}, \quad (15a)$$

$$\hat{P}_{F,2}^{\text{NOMA}} = \frac{\Upsilon_2^F(\rho_2, N)}{\bar{h}_2}, \quad (15b)$$

respectively, where $\Upsilon_k^F(\rho_k, N) \triangleq \exp\left(\frac{\ln 2 \rho_k L_k}{N(1 - \epsilon_k)} + \frac{Q^{-1}(\epsilon_k)}{\sqrt{N}}\right) - 1$ is the minimum SINR/SNR. If the SIC fails, by letting $\gamma_{F,2}^{\text{NOMA}} = \tilde{\gamma}_{F,2}^{\text{NOMA}}$ and combining (11), (12), and (14), the minimum transmit powers of UEs 1 and 2 can be respectively derived as

$$\tilde{P}_{F,1}^{\text{NOMA}} = \frac{1}{\bar{h}_1} \frac{\Upsilon_1^F(\rho_1, N)(\Upsilon_2^F(\rho_2, N) + 1)}{1 - \Upsilon_1^F(\rho_1, N)\Upsilon_2^F(\rho_2, N)}, \quad (16a)$$

$$\tilde{P}_{F,2}^{\text{NOMA}} = \frac{1}{\bar{h}_2} \frac{\Upsilon_2^F(\rho_2, N)(\Upsilon_1^F(\rho_1, N) + 1)}{1 - \Upsilon_1^F(\rho_1, N)\Upsilon_2^F(\rho_2, N)}. \quad (16b)$$

By combining the two cases, the average minimum transmit power of UE k can be characterized as

$$\bar{P}_{F,k}^{\text{NOMA}} = (1 - \epsilon_1)\hat{P}_{F,k}^{\text{NOMA}} + \epsilon_1\tilde{P}_{F,k}^{\text{NOMA}}. \quad (17)$$

Thus, according to (2)–(4) and (17), the MEC-related energy consumption of NOMA offloading in the finite blocklength case is expressed as

$$E_F^{\text{NOMA}}(\rho, t, d) = \sum_{k=1}^2 \left(\kappa_U \frac{(\rho_k c_k L_k)^3}{(T_{\max} - t)^2} + \kappa_k \frac{((1 - \rho_k) c_k L_k)^3}{T_{\max}^2} + \bar{P}_{F,k}^{\text{NOMA}} t \right), \quad (18)$$

and the joint optimization problem that minimizing $E_F^{\text{NOMA}}(\rho, t, d)$ can be formulated as

$$\text{(NOMA-F)} \quad \min_{\rho, t, d} E_F^{\text{NOMA}}(\rho, t, d) \quad (19a)$$

$$\text{s.t.} \quad \bar{P}_{F,k}^{\text{NOMA}} \leq P_{k, \max}, \quad \forall k, \quad (19b)$$

$$\Upsilon_1^F(\rho_1, N)\Upsilon_2^F(\rho_2, N) < 1, \quad (19c)$$

$$(10b), (10c), (10e) - (10g), \quad (19d)$$

where (19c) ensures that $\check{P}_{F,k}^{\text{NOMA}} \geq 0, \forall k$.

B. FDMA Offloading

By using FDMA, two UEs offload their tasks to the UAV simultaneously over two different frequency-domain resource blocks as shown in Fig. 2(b). Thus, they have the same offloading time, i.e., $t_1 = t_2 = t$, and the bandwidths allocated to UEs 1 and 2 can be written as $B_1 = \eta B$ and $B_2 = (1 - \eta)B$, respectively, where $0 \leq \eta \leq 1$ denotes the bandwidth allocation coefficient. The signal transmitted from UE k and received by the UAV is given by

$$y_{k,\text{FDMA}} = \sqrt{P_k}h_k x_k + z_{k,\text{FDMA}}, \quad \forall k, \quad (20)$$

where $z_{k,\text{FDMA}} \sim \mathcal{CN}(0, \sigma_{k,\text{FDMA}}^2)$ with $\sigma_{1,\text{FDMA}}^2 = \eta B N_0$ and $\sigma_{2,\text{FDMA}}^2 = (1 - \eta) B N_0$. Accordingly, the SNRs of UEs 1 and 2's signals received by the UAV are expressed as

$$\gamma_1^{\text{FDMA}} = \frac{P_1 h_1}{\eta B N_0}, \quad \gamma_2^{\text{FDMA}} = \frac{P_2 h_2}{(1 - \eta) B N_0}, \quad (21)$$

respectively.

1) *Infinite Blocklength Case*: In this case, the information-causality constraint for computation offloading can be expressed as

$$B_k t \log_2(1 + \gamma_{I,k}^{\text{FDMA}}) \geq \rho_k L_k, \quad \forall k, \quad (22)$$

where $\gamma_{I,k}^{\text{FDMA}} = \gamma_k^{\text{FDMA}}$. According to (22), the minimum transmit powers of UEs 1 and 2 are respectively obtained as

$$P_{I,1}^{\text{FDMA}} = \frac{\eta}{h_1} \Upsilon_1^I(\rho_1, \eta t), \quad (23a)$$

$$P_{I,2}^{\text{FDMA}} = \frac{1 - \eta}{h_2} \Upsilon_2^I(\rho_2, (1 - \eta)t). \quad (23b)$$

Thus, according to (2)–(4) and (23), the MEC-related energy consumption of FDMA offloading in the infinite blocklength case is expressed as

$$E_I^{\text{FDMA}}(\rho, t, d) = \sum_{k=1}^2 \left(\kappa_U \frac{(\rho_k c_k L_k)^3}{(T_{\max} - t)^2} + \kappa_k \frac{((1 - \rho_k) c_k L_k)^3}{T_{\max}^2} + P_{I,k}^{\text{FDMA}} t \right), \quad (24)$$

and the joint optimization problem that minimizing $E_I^{\text{FDMA}}(\rho, t, d)$ can be formulated as

$$(\text{FDMA-I}) \quad \min_{\rho, t, d} \quad E_I^{\text{FDMA}}(\rho, t, d) \quad (25a)$$

$$\text{s.t.} \quad P_{I,k}^{\text{FDMA}} \leq P_{k,\max}, \quad \forall k, \quad (25b)$$

$$(10b), (10c), (10e), (10f). \quad (25c)$$

2) *Finite Blocklength Case*: In this case, the blocklength of UEs 1 and 2 can be written as $N_1 = B_1 t = \eta N$ and $N_2 = B_2 t = (1 - \eta)N$, respectively. And for a given decoding error probability ϵ_k , the information-causality constraint for computation offloading can be expressed as

$$(1 - \epsilon_k) N_k \left(\log_2(1 + \gamma_{F,k}^{\text{FDMA}}) - \sqrt{\frac{V_k}{N_k}} \frac{Q^{-1}(\epsilon_k)}{\ln 2} \right) \geq$$

$$\rho_k L_k, \quad \forall k, \quad (26)$$

where $\gamma_{F,k}^{\text{FDMA}} = \gamma_k^{\text{FDMA}}$. According to (26), the minimum transmit powers of UEs 1 and 2 are respectively obtained as

$$P_{F,1}^{\text{FDMA}} = \frac{\eta}{h_1} \Upsilon_1^F(\rho_1, \eta N), \quad (27a)$$

$$P_{F,2}^{\text{FDMA}} = \frac{1 - \eta}{h_2} \Upsilon_2^F(\rho_2, (1 - \eta)N). \quad (27b)$$

Thus, according to (2)–(4) and (27), the MEC-related energy consumption of FDMA offloading in the finite blocklength case is expressed as

$$E_F^{\text{FDMA}}(\rho, t, d) = \sum_{k=1}^2 \left(\kappa_U \frac{(\rho_k c_k L_k)^3}{(T_{\max} - t)^2} + \kappa_k \frac{((1 - \rho_k) c_k L_k)^3}{T_{\max}^2} + P_{F,k}^{\text{FDMA}} t \right), \quad (28)$$

and the joint optimization problem that minimizing $E_F^{\text{FDMA}}(\rho, t, d)$ can be formulated as

$$(\text{FDMA-F}) \quad \min_{\rho, t, d} \quad E_F^{\text{FDMA}}(\rho, t, d) \quad (29a)$$

$$\text{s.t.} \quad P_{F,k}^{\text{FDMA}} \leq P_{k,\max}, \quad \forall k, \quad (29b)$$

$$(10b), (10c), (10e), (10f). \quad (29c)$$

C. TDMA Offloading

By using TDMA, as shown in Fig. 2(c), two UEs offload their tasks to the UAV over two successive time-domain resource blocks, sharing the same bandwidth. Thus, the bandwidths of UEs 1 and 2 can be written as $B_1 = B_2 = B$, and their offloading times can be written as $t_1 = \delta t$ and $t_2 = (1 - \delta)t$, respectively, where $0 \leq \delta \leq 1$ denotes the time allocation coefficient. The signal transmitted from UE k and received by the UAV is given by

$$y_{k,\text{TDMA}} = \sqrt{P_k}h_k x_k + z_{\text{TDMA}}, \quad \forall k, \quad (30)$$

where $z_{\text{TDMA}} \sim \mathcal{CN}(0, \sigma_{\text{TDMA}}^2)$ with $\sigma_{\text{TDMA}}^2 = B N_0$. Accordingly, the SNR of UE k 's signal received by the UAV is

$$\gamma_k^{\text{TDMA}} = \frac{P_k h_k}{B N_0}, \quad \forall k. \quad (31)$$

Note that it may have $2! = 2$ permutations for the offloading order, i.e., (i) UE 1 offloads its task before UE 2 and (ii) UE 2 offloads its task before UE 1. In the following, we choose order (i) to analyze, and the other order can be analyzed similarly. Different from the NOMA and FDMA offloadings, where the UAV computes two UEs' offloaded tasks simultaneously, the UAV can compute UE 1's offloaded task during UE 2 offloads its task in TDMA offloading. Thus, UEs 1 and 2's remote-computing times can be written as $T_{\max} - \delta t$ and $T_{\max} - t$, respectively, and the CPU frequencies allocated for their remote computing by the UAV can be expressed as

$$f_{\text{TDMA},1}^{\text{rem}} = \frac{\rho_1 c_1 L_1}{T_{\max} - \delta t}, \quad f_{\text{TDMA},2}^{\text{rem}} = \frac{\rho_2 c_2 L_2}{T_{\max} - t}. \quad (32)$$

1) *Infinite Blocklength Case*: In this case, the information-causality constraint for computation offloading can be expressed as

$$B_k t \log_2(1 + \gamma_{I,k}^{\text{TDMA}}) \geq \rho_k L_k \quad \forall k, \quad (33)$$

where $\gamma_{I,k}^{\text{TDMA}} = \gamma_k^{\text{TDMA}}$. According to (33), the minimum transmit powers of UEs 1 and 2 are respectively obtained as

$$P_{I,1}^{\text{TDMA}} = \frac{1}{h_1} \Upsilon_1^I(\rho_1, \delta t), \quad (34a)$$

$$P_{I,2}^{\text{TDMA}} = \frac{1}{h_2} \Upsilon_2^I(\rho_2, (1-\delta)t). \quad (34b)$$

Thus, according to (3), (4), (32), and (34), the MEC-related energy consumption of TDMA offloading in the infinite blocklength case is expressed as

$$\begin{aligned} E_I^{\text{TDMA}}(\rho, t, d) &= \kappa_U \frac{(\rho_1 c_1 L_1)^3}{(T_{\max} - \delta t)^2} + \kappa_U \frac{(\rho_2 c_2 L_2)^3}{(T_{\max} - t)^2} \\ &\quad + \delta P_{I,1}^{\text{TDMA}} t + (1-\delta) P_{I,2}^{\text{TDMA}} t \\ &\quad + \sum_{k=1}^2 \kappa_k \frac{((1-\rho_k) c_k L_k)^3}{T_{\max}^2}, \end{aligned} \quad (35)$$

and the joint optimization problem that minimizing $E_I^{\text{TDMA}}(\rho, t, d)$ can be formulated as

$$\text{(TDMA-I)} \quad \min_{\rho, t, d} E_I^{\text{TDMA}}(\rho, t, d) \quad (36a)$$

$$\text{s.t.} \quad f_{\text{TDMA},1}^{\text{rem}} + f_{\text{TDMA},2}^{\text{rem}} \leq f_{U,\max}, \quad (36b)$$

$$P_{I,k}^{\text{TDMA}} \leq P_{k,\max}, \quad \forall k, \quad (36c)$$

$$(10c), (10e), (10f). \quad (36d)$$

2) *Finite Blocklength Case:* In this case, the blocklength of UEs 1 and 2 can be written as $N_1 = Bt_1 = \delta N$ and $N_2 = Bt_2 = (1-\delta)N$, respectively. And for a given decoding error probability ϵ_k , the information-causality constraint for computation offloading can be expressed as

$$(1 - \epsilon_k) N_k \left(\log_2(1 + \gamma_{F,k}^{\text{TDMA}}) - \sqrt{\frac{V_k}{N_k}} \frac{Q^{-1}(\epsilon_k)}{\ln 2} \right) \geq \rho_k L_k, \quad \forall k, \quad (37)$$

where $\gamma_{F,k}^{\text{TDMA}} = \gamma_k^{\text{TDMA}}$. According to (37), the minimum transmit powers of UEs 1 and 2 are respectively obtained as

$$P_{F,1}^{\text{TDMA}} = \frac{1}{h_1} \Upsilon_1^F(\rho_1, \delta N), \quad (38a)$$

$$P_{F,2}^{\text{TDMA}} = \frac{1}{h_2} \Upsilon_2^F(\rho_2, (1-\delta)N). \quad (38b)$$

Thus, according to (3), (4), (32), and (38), the MEC-related energy consumption of TDMA offloading in the finite blocklength case is expressed as

$$\begin{aligned} E_F^{\text{TDMA}}(\rho, t, d) &= \kappa_U \frac{(\rho_1 c_1 L_1)^3}{(T_{\max} - \delta t)^2} + \kappa_U \frac{(\rho_2 c_2 L_2)^3}{(T_{\max} - t)^2} \\ &\quad + \delta P_{F,1}^{\text{TDMA}} t + (1-\delta) P_{F,2}^{\text{TDMA}} t \\ &\quad + \sum_{k=1}^2 \kappa_k \frac{((1-\rho_k) c_k L_k)^3}{T_{\max}^2}, \end{aligned} \quad (39)$$

and the joint optimization problem that minimizing $E_F^{\text{TDMA}}(\rho, t, d)$ can be formulated as

$$\text{(TDMA-F)} \quad \min_{\rho, t, d} E_F^{\text{TDMA}}(\rho, t, d) \quad (40a)$$

$$\text{s.t.} \quad P_{F,k}^{\text{TDMA}} \leq P_{k,\max}, \quad \forall k, \quad (40b)$$

$$(10c), (10e), (10f), (36b). \quad (40c)$$

III. PERFORMANCE COMPARISON

To explore which multiple access offloading schemes is more capable of reducing the MEC-related energy consumption of the considered UAV-enabled MEC system and whether the differences in performance benefit among different schemes are the same in both infinite blocklength and finite blocklength cases, we compare the minimum MEC-related energy consumption obtained by NOMA, FDMA, and TDMA offloadings, respectively, in both infinite blocklength and finite blocklength cases. The analysis results of this section remain applicable in a multi-UE system, where UEs are grouped into pairs, and each UE pair is allocated with orthogonal time/frequency communication resources and independent CPU frequency for remote computing.

A. Comparison between FDMA and TDMA Offloadings

First, we compare the MEC-related energy consumption of FDMA and TDMA offloadings with finite blocklength. For the special case that $\eta = \delta$, let $\{\rho_F^*, t_F^*, d_F^*\}$ denote the optimal solution to problem (FDMA-F) in (29), the energy gap between FDMA and TDMA, i.e., the difference between (28) and (39), is given by

$$\begin{aligned} &E_F^{\text{FDMA}}(\rho, t, d) - E_F^{\text{TDMA}}(\rho, t, d) \\ &\stackrel{(a)}{\geq} E_F^{\text{FDMA}}(\rho_F^*, t_F^*, d_F^*) - E_F^{\text{TDMA}}(\rho_F^*, t_F^*, d_F^*) \\ &= \kappa_U \left(\frac{(\rho_{F,1}^* c_1 L_1)^3}{(T_{\max} - t_F^*)^2} - \frac{(\rho_{F,1}^* c_1 L_1)^3}{(T_{\max} - \delta t_F^*)^2} \right) \stackrel{(b)}{\geq} 0, \end{aligned} \quad (41)$$

where the equality in (a) holds if $\{\rho_F^*, t_F^*, d_F^*\}$ is also the optimal solution to problem (TDMA-F), and the equality in (b) holds if $\delta = 1$. Similarly, a non-negative energy gap between FDMA and TDMA offloadings with infinite blocklength can be also obtained. For the equality relation in (41), we find that FDMA and TDMA offloadings have the same task-offloading energy consumption $\{E_k^{\text{off}}\}$ with any given $\{\rho, t, d\}$ and $\eta = \delta$. Furthermore, for the inequality relationship in (41), we find that TDMA can achieve a lower MEC-related energy consumption than FDMA due to the freedom of remote-computing time brought by the sequential task offloading of the UEs.

B. Comparison between NOMA and FDMA Offloadings with Infinite Blocklength

Next, we compare the MEC-related energy consumption of NOMA and FDMA offloadings with infinite blocklength. For a special case that $h_1 \geq h_2$ (the UAV first decodes UE 1's signal) and $\eta = \frac{1}{2}$, let $\{\tilde{\rho}_F^*, \tilde{t}_F^*, \tilde{d}_F^*\}$ denote the optimal solution to the problem (FDMA-I) in (25), and $\{\tilde{h}_{F,k}^*\}$ denotes the corresponding normalized channel power gains, the energy gap between NOMA and FDMA, i.e., the difference between (9) and (24), can be expressed as

$$\begin{aligned} &E_I^{\text{NOMA}}(\rho, t, d) - E_I^{\text{FDMA}}(\rho, t, d) \\ &\stackrel{(a)}{\leq} E_I^{\text{NOMA}}(\tilde{\rho}_F^*, \tilde{t}_F^*, \tilde{d}_F^*) - E_I^{\text{FDMA}}(\tilde{\rho}_F^*, \tilde{t}_F^*, \tilde{d}_F^*) \\ &= \left[\frac{1}{\tilde{h}_{F,1}^*} \Upsilon_1^I(\tilde{\rho}_{F,1}^*, \tilde{t}_F^*) \Upsilon_2^I(\tilde{\rho}_{F,2}^*, \tilde{t}_F^*) + \left(\frac{1}{\tilde{h}_{F,2}^*} - \frac{1}{\tilde{h}_{F,1}^*} \right) \right] \end{aligned}$$

$$\begin{aligned}
& \times \left(\bar{\Upsilon}_2^I(\tilde{\rho}_{F,2}^*, \tilde{t}_F^*) - \frac{1}{2} \bar{\Upsilon}_2^I(\tilde{\rho}_{F,2}^*, \frac{\tilde{t}_F^*}{2}) \right) \\
& - \frac{1}{2\tilde{h}_{F,1}^*} \sum_{k=1}^2 \bar{\Upsilon}_k^I(\tilde{\rho}_{F,k}^*, \frac{\tilde{t}_F^*}{2}) + \left(\frac{1}{2\tilde{h}_{F,1}^*} - \frac{1}{2\tilde{h}_{F,2}^*} \right) \tilde{t}_F^* \\
& \stackrel{(b)}{\leq} \left[\bar{\Upsilon}_1^I(\tilde{\rho}_{F,1}^*, \tilde{t}_F^*) \bar{\Upsilon}_2^I(\tilde{\rho}_{F,2}^*, \tilde{t}_F^*) - \sum_{k=1}^2 \frac{\bar{\Upsilon}_k^I(\tilde{\rho}_{F,k}^*, \frac{\tilde{t}_F^*}{2})}{2} \right] \frac{\tilde{t}_F^*}{\tilde{h}_{F,1}^*} \\
& = \left[\sqrt{\bar{\Upsilon}_1^I(\tilde{\rho}_{F,1}^*, \frac{\tilde{t}_F^*}{2}) \bar{\Upsilon}_2^I(\tilde{\rho}_{F,2}^*, \frac{\tilde{t}_F^*}{2})} - \sum_{k=1}^2 \frac{\bar{\Upsilon}_k^I(\tilde{\rho}_{F,k}^*, \frac{\tilde{t}_F^*}{2})}{2} \right] \\
& \times \frac{\tilde{t}_F^*}{\tilde{h}_{F,1}^*} \stackrel{(c)}{\leq} 0, \tag{42}
\end{aligned}$$

where $\bar{\Upsilon}_k^I(\rho_k, t) \triangleq \Upsilon_k^I(\rho_k, t) + 1 = \exp(\frac{\ln 2 \rho_k L_k}{Bt})$. In (42), the inequality (b) holds due to $\tilde{h}_{F,1}^* \geq \tilde{h}_{F,2}^*$ and $\tilde{t}_F^* \bar{\Upsilon}_2^I(\tilde{\rho}_{F,2}^*, \tilde{t}_F^*) < \frac{\tilde{t}_F^*}{2} \bar{\Upsilon}_2^I(\tilde{\rho}_{F,2}^*, \frac{\tilde{t}_F^*}{2})$, where the latter can be proved by the fact that $f(x) = xe^{-x}$ is a decreasing function when $x \geq 1$, and then $B \left(\tilde{t}_F^* \bar{\Upsilon}_2^I(\tilde{\rho}_{F,2}^*, \tilde{t}_F^*) - \frac{\tilde{t}_F^*}{2} \bar{\Upsilon}_2^I(\tilde{\rho}_{F,2}^*, \frac{\tilde{t}_F^*}{2}) \right) = e^{\ln 2 \tilde{\rho}_{F,2}^* L_2} \left(B \tilde{t}_F^* e^{-B \tilde{t}_F^*} - \frac{B \tilde{t}_F^*}{2} e^{-\frac{B \tilde{t}_F^*}{2}} \right) < 0$ since $\frac{B \tilde{t}_F^*}{2} \geq 1$ can be satisfied. The inequality (c) holds due to the inequality $\sqrt{xy} \leq \frac{x+y}{2}$. In addition, the equality in (a) holds if $\{\tilde{\rho}_F^*, \tilde{t}_F^*, \tilde{d}_F^*\}$ is also the optimal solution to the problem (NOMA-F); the equality in (b) holds if $\tilde{h}_{F,1}^* = \tilde{h}_{F,2}^*$; the equality in (c) holds if $\bar{\Upsilon}_1^I(\tilde{\rho}_{F,1}^*, \frac{\tilde{t}_F^*}{2}) = \bar{\Upsilon}_2^I(\tilde{\rho}_{F,2}^*, \frac{\tilde{t}_F^*}{2})$. Note that the similar result can be also proved if $h_1 < h_2$. We can find that NOMA and FDMA offloadings have the same computation-related energy consumption, i.e., $\{E_k^{\text{loc}}\}$ and $\{E_k^{\text{rem}}\}$, with any given $\{\rho, t\}$, so the inequality relationship in (42) implies that NOMA can achieve a lower MEC-related energy consumption than FDMA with infinite blocklength due to the spectrum resource sharing and the perfect SIC.

C. Comparison between NOMA and FDMA Offloadings with Finite Blocklength

Then, we compare the MEC-related energy consumption of NOMA and FDMA offloadings with finite blocklength. For a special case that $\eta = \frac{1}{2}$, let $\{\rho_N^*, t_N^*, d_N^*\}$ denote the optimal solution to problem (NOMA-F), and $\{\bar{h}_{N,k}^*\}$ and N_N^* denote the corresponding normalized channel power gains and blocklength, respectively, the energy gap between NOMA and FDMA, i.e., the difference between (18) and (28), can be expressed as

$$\begin{aligned}
& E_F^{\text{NOMA}}(\rho, t, d) - E_F^{\text{FDMA}}(\rho, t, d) \\
& \stackrel{(a)}{\geq} E_F^{\text{NOMA}}(\rho_N^*, t_N^*, d_N^*) - E_F^{\text{FDMA}}(\rho_N^*, t_N^*, d_N^*) \\
& \stackrel{(b)}{>} \left[\frac{\Upsilon_1^F(\rho_{N,1}^*, N_N^*)(\Upsilon_2^F(\rho_{N,2}^*, N_N^*) + 1)}{\bar{h}_{N,1}^*} + \frac{\Upsilon_2^F(\rho_{N,2}^*, N_N^*)}{\bar{h}_{N,2}^*} \right] \frac{N_N^*}{B} \\
& \quad - \left[\frac{\Upsilon_1^F(\rho_{N,1}^*, \frac{N_N^*}{2})}{2\bar{h}_{N,1}^*} + \frac{\Upsilon_2^F(\rho_{N,2}^*, \frac{N_N^*}{2})}{2\bar{h}_{N,2}^*} \right] \frac{N_N^*}{B} \\
& = \left[\frac{\Upsilon_1^F(\rho_{N,1}^*, N_N^*) \bar{\Upsilon}_2^F(\rho_{N,2}^*, N_N^*)}{\bar{h}_{N,1}^*} - \sum_{k=1}^2 \frac{\Upsilon_k^F(\rho_{N,k}^*, \frac{N_N^*}{2})}{2\bar{h}_{N,k}^*} \right]
\end{aligned}$$

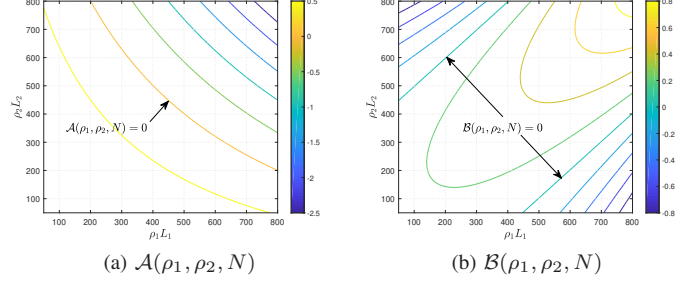


Fig. 3. Isolines of $\mathcal{A}(\rho_1, \rho_2, N)$ and $\mathcal{B}(\rho_1, \rho_2, N)$ over mesh $(\rho_1 L_1, \rho_2 L_2)$ for $N = 300$.

$$+ \left(\frac{1}{\bar{h}_{N,2}^*} - \frac{1}{\bar{h}_{N,1}^*} \right) \left(\bar{\Upsilon}_2^F(\rho_{N,2}^*, N_N^*) + \frac{1}{2} \right) \frac{N_N^*}{B} \triangleq \Delta_F^N, \tag{43}$$

where $\bar{\Upsilon}_k^F(\rho_k, N) \triangleq \Upsilon_k^F(\rho_k, N) + 1 = \exp(\frac{\ln 2 \rho_k L_k}{N(1-\epsilon_k)} + \frac{Q^{-1}(\epsilon_k)}{\sqrt{N}})$. In (43), the equality in (a) holds if $\{\rho_N^*, t_N^*, d_N^*\}$ is also the optimal solution to problem (FDMA-F), and the inequality (b) holds because $\bar{P}_{F,k}^{\text{NOMA}} > \hat{P}_{F,k}^{\text{NOMA}}, \forall k$. Since $\bar{\Upsilon}_k^F(\rho_k, N)$ is complex and the value of (43) depends on the relative magnitude between $\rho_{N,1}^*$ and $\rho_{N,2}^*$ and that between $\bar{h}_{N,1}^*$ and $\bar{h}_{N,2}^*$, it is hard to determine whether Δ_F^N is positive or negative directly.

In order to facilitate the analysis of the value of Δ_F^N within its feasible domain for a given N , we first define $\mathcal{A}(\rho_1, \rho_2, N) \triangleq 1 - \Upsilon_1^F(\rho_1, N) \Upsilon_2^F(\rho_2, N)$ and $\mathcal{B}(\rho_1, \rho_2, N) \triangleq \bar{\Upsilon}_1^F(\rho_1, N) \bar{\Upsilon}_2^F(\rho_2, N) - \frac{1}{2} \sum_{k=1}^2 \bar{\Upsilon}_k^F(\rho_k, \frac{N}{2})$, and show their isolines over mesh $(\rho_1 L_1, \rho_2 L_2)$ for $N = 300$ in Fig. 3. In Fig. 3(a), we can observe that $\mathcal{A}(\rho_1, \rho_2, N)$ decreases with the increase of $\rho_1 L_1 + \rho_2 L_2$ and $\mathcal{A}(\rho_1, \rho_2, N) > 0$ should be satisfied to ensure $\bar{P}_{F,k}^{\text{NOMA}} \geq 0, \forall k$, since for the given SNR $\gamma_{F,k}^{\text{NOMA}}$ and blocklength N , the decoding error probability ϵ_k increases as the data rate $\frac{\rho_k L_k}{N}$ increases [13], the sum data size of two UEs' offloaded tasks, i.e., $\rho_1 L_1 + \rho_2 L_2$, cannot exceed a certain value that satisfies $\mathcal{A}(\rho_1, \rho_2, N) = 0$ to ensure the decoding error probabilities of two UEs' offloaded tasks do not exceed ϵ_k . In Fig. 3(b), we can find that $\mathcal{B}(\rho_1, \rho_2, N)$ decreases with the increase of $|\rho_1 L_1 - \rho_2 L_2|$. Since $\Delta_F^N = \frac{N_N^*}{\bar{h}_{N,1}^* B} \mathcal{B}(\rho_{N,1}^*, \rho_{N,2}^*, N_N^*)$ when $\bar{h}_{N,1}^* = \bar{h}_{N,2}^*$, and $\bar{P}_{F,k}^{\text{NOMA}}$ in (17) increases as ρ_k increases, NOMA offloading should keep an offloaded task gap between two UEs, i.e., $|\rho_1 L_1 - \rho_2 L_2|$ to leverage the advantages of spectrum efficiency due to the power domain multiplexing.

Thus, if $(\rho_{N,1}^* L_1, \rho_{N,2}^* L_2)$ is at the overlap area of $\mathcal{A}(\rho_1, \rho_2, N) > 0$ in Fig. 3(a) and $\mathcal{B}(\rho_1, \rho_2, N) \geq 0$ in Fig. 3(b), $\Delta_F^N \geq 0$ when $\bar{h}_{N,1}^* = \bar{h}_{N,2}^*$. For the most special case that $\bar{h}_{N,1}^* = \bar{h}_{N,2}^*, \rho_{N,1}^* L_1 = \rho_{N,2}^* L_2$, and $\epsilon_1 = \epsilon_2$, we have

$$\begin{aligned}
\Delta_F^N &= \frac{N_N^*}{\bar{h}_{N,k}^* B} \left(\bar{\Upsilon}_k^F(\rho_{N,k}^*, N_N^*)^2 - \bar{\Upsilon}_k^F(\rho_{N,k}^*, \frac{N_N^*}{2}) \right) \\
&> 0, \quad k = 1 \text{ or } 2. \tag{44}
\end{aligned}$$

Note that because (43) is a strict inequality, and the inequality (b) in (43) makes Δ_F^N only measures the difference between

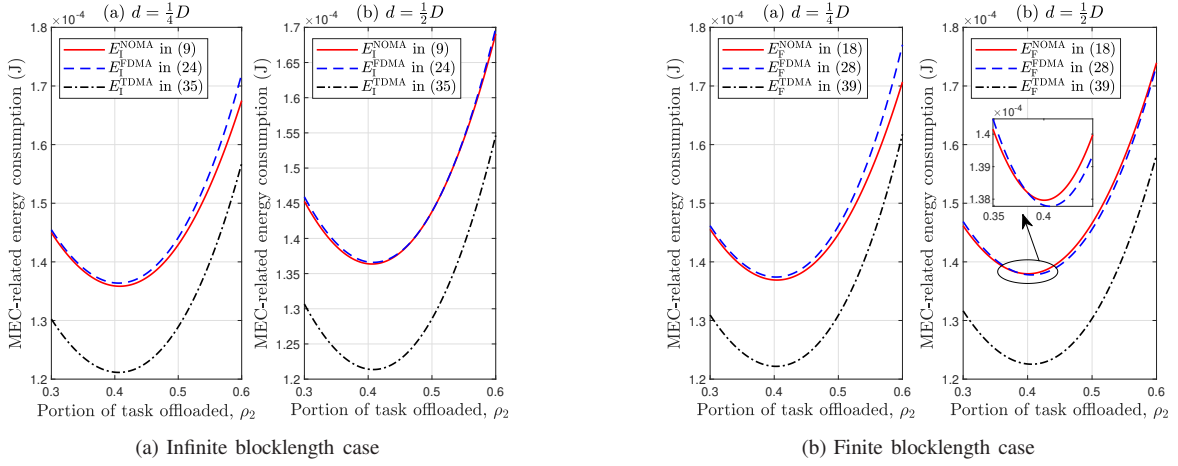


Fig. 4. MEC-related energy consumption required by different multiple access offloading schemes versus the portion of task offloaded by UE 2 ρ_2 for different values of d when $\eta = \delta = 0.5$, $D = 100$ m, $\rho_1 = 0.5$, $B = 0.5$ MHz, $t = 0.3$ ms, $L_1 = L_2 = 1.2$ kbit, and $\epsilon_1 = \epsilon_2 = 10^{-5}$.

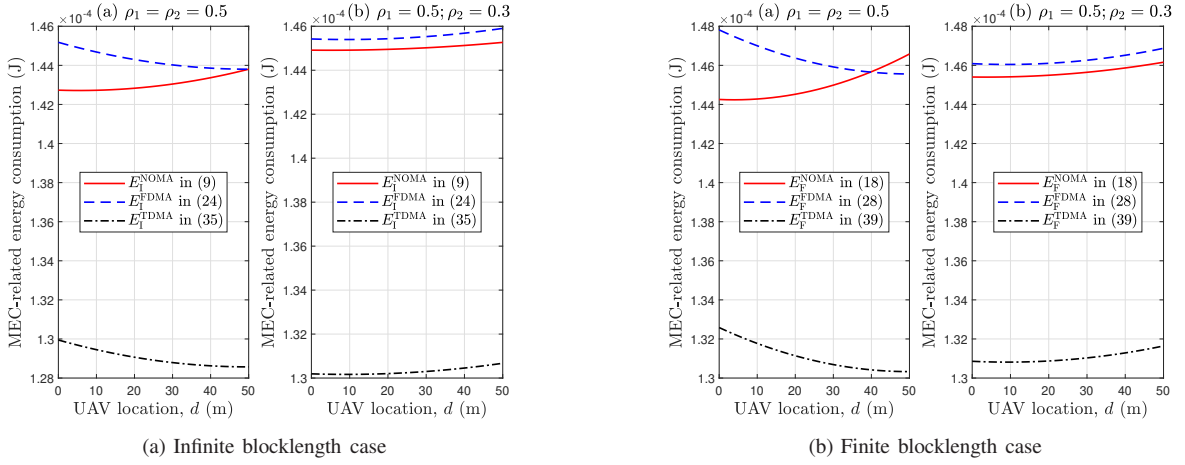


Fig. 5. MEC-related energy consumption required by different multiple access offloading schemes versus the UAV location d for different cases of ρ when $\eta = \delta = 0.5$, $D = 100$ m, $B = 0.5$ MHz, $t = 0.3$ ms, $L_1 = L_2 = 1.2$ kbit, and $\epsilon_1 = \epsilon_2 = 10^{-5}$.

the MEC-related energy consumption required by SIC success case of NOMA offloading and that required by FDMA offloading, so the above result is conservative.

The correctness of the above theoretical conclusion can be verified through the numerical result shown in Fig. 4, which presents the calculation results of the MEC-related energy consumption required by different multiple access offloading schemes in both the infinite and finite blocklength cases versus the portion of task offloaded by UE 2 ρ_2 for $\rho_1 = 0.5$. As the comparison result in (41), TDMA achieves a lower energy consumption than FDMA in all cases in Fig. 4. As the comparison result in (42), if the blocklength is assumed to be infinite, NOMA requires less energy consumption than FDMA, and the performance advantage of NOMA when $d = \frac{1}{4}D$ is more obvious than that when $d = \frac{1}{2}D$ due to the quality difference of the offloading links caused by the UAV location. For the finite-blocklength offloading, NOMA can still achieve a lower energy consumption than FDMA when $d = \frac{1}{4}D$ as shown in Fig. 4(a), while the performance advantage of NOMA over FDMA cannot be obtained when $0.4 \leq \rho_2 \leq 0.58$ and $d = \frac{1}{2}D$, i.e., $\bar{h}_1 = \bar{h}_2$, in Fig. 4(b), which is consistent with the analysis results in Section III-C.

To investigate the impact of UAV location on the performance gap between NOMA and FDMA, we present the MEC-related energy consumption versus UAV location d in Fig. 5. For the symmetric task offloading case (i.e., $\rho_1 = \rho_2 = 0.5$, where the computation-related energy consumption of two UEs is identical), as d decreases from $\frac{D}{2} = 50$ m to 0 m, the energy consumption gap between NOMA and FDMA increases, as shown in both Fig. 5(a) and Fig. 5(b). This is because the performance benefits of NOMA over FDMA become more pronounced when the channel conditions of the users differ significantly [31]. When the offloading channels of two UEs tend to be symmetric (i.e., $d = \frac{D}{2} = 50$ m), the performance gap becomes zero if the blocklength is infinite. However, in the finite blocklength case, NOMA requires more energy consumption than FDMA due to the consideration of decoding error probabilities and SIC failure. In contrast, for the asymmetric task offloading case (i.e., $\rho_1 = 0.5$ and $\rho_2 = 0.3$), NOMA consistently achieves lower energy consumption compared to FDMA with both infinite and finite blocklength, thanks to the efficiency provided by the power domain multiplexing.

To sum up, TDMA can achieve a lower MEC-related energy

consumption than FDMA in both infinite blocklength and finite blocklength cases due to the sequential task offloading of the UEs, and NOMA consumes less energy compared to FDMA in infinite blocklength case. However, the same conclusion cannot be obtained if the blocklength is finite, especially when the channel power gains and the offloaded task data sizes of two UEs are relatively symmetrical, the energy consumption of NOMA will be larger than that of FDMA.

IV. PROPOSED ALGORITHM FOR PROBLEM (NOMA-F)

Note that the six MEC-related energy consumption minimization problems of different offloading schemes formulated in Section II are all non-convex due to the non-convexity of their objective functions and constraints, as well as the coupling of the optimization variables, especially in the expressions of transmit power, which make these problems difficult to be solved directly. We can find that the mathematical forms of these six problems are similar except for the specific expressions of the transmit power. To avoid redundancy, we only present the solution to problem (NOMA-F) in the following, which is the most complex and challenging, and the other five problems can be solved similarly. To overcome the above difficulties, we decouple variables ρ , t , and d in problem (NOMA-F) by the BCD technique, which decomposes problem (NOMA-F) into three subproblems: 1) Subproblem 1 optimizes the portions of task offloaded ρ for given t and d ; 2) Subproblem 2 optimizes the offloading time t for given ρ and d ; 3) Subproblem 3 optimizes the UAV location d for given ρ and d . These three subproblems are solved in an alternating manner until the objective value $E_F^{\text{NOMA}}(\rho, t, d)$ converges. The solutions to the three subproblems are presented as follows.

A. Optimizing the Portions of Task Offloaded

For given d and t , problem (NOMA-F) reduces to

$$\min_{\rho} \sum_{k=1}^2 \zeta_k \rho_k^3 + \xi_k (1 - \rho_k)^3 + \bar{P}_{F,k}^{\text{NOMA}} t \quad (45a)$$

$$\text{s.t.} \quad (10b), (10c), (10e), (19b), (19c), \quad (45b)$$

where $\zeta_k = \frac{\kappa_U (c_k L_k)^3}{(T_{\max} - t)^2}$ and $\xi_k = \frac{\kappa_k (c_k L_k)^3}{T_{\max}^2}$. The expressions of transmit power in (15) and (16) can be simplified as

$$\hat{P}_{F,1}^{\text{NOMA}}(\rho) = \frac{1}{h_1} (e^{\omega_1 \rho_1 + \omega_2 \rho_2 + v_1 + v_2} - e^{\omega_2 \rho_2 + v_2}), \quad (46a)$$

$$\hat{P}_{F,2}^{\text{NOMA}}(\rho) = \frac{1}{h_2} (e^{\omega_2 \rho_2 + v_2} - 1), \quad (46b)$$

$$\check{P}_{F,1}^{\text{NOMA}}(\rho) = \frac{1}{h_1} \left(\frac{1}{1 - e^{\omega_2 \rho_2} (e^{v_2} - e^{-\omega_1 \rho_1})} - 1 \right), \quad (46c)$$

$$\check{P}_{F,2}^{\text{NOMA}}(\rho) = \frac{1}{h_2} \left(\frac{1}{1 - e^{\omega_1 \rho_1} (e^{v_1} - e^{-\omega_2 \rho_2})} - 1 \right), \quad (46d)$$

where $\omega_k \triangleq \frac{\ln 2 L_k}{N(1 - \epsilon_k)}$ and $v_k \triangleq \frac{Q^{-1}(\epsilon_k)}{\sqrt{N}}$. To handle the complex forms of $\{\hat{P}_{F,k}^{\text{NOMA}}(\rho)\}$, we introduce slack variables $\tilde{\lambda} \triangleq \{\tilde{\lambda}_1, \tilde{\lambda}_2\}$, and $\tilde{\mu} \triangleq \{\tilde{\mu}_1, \tilde{\mu}_2\}$, and problem (45) can be equivalently transformed into

$$\min_{\rho, \tilde{\lambda}, \tilde{\mu}} \sum_{k=1}^2 \zeta_k \rho_k^3 + \xi_k (1 - \rho_k)^3 + \mathcal{P}_k(\rho, \tilde{\mu}_k) t \quad (47a)$$

$$\text{s.t.} \quad \tilde{\mu}_k \leq 1 - e^{-\tilde{\lambda}_k}, \quad \forall k, \quad (47b)$$

$$0 \geq \tilde{\lambda}_k \geq \lambda_k(\rho), \quad \forall k, \quad (47c)$$

$$\mathcal{P}_k(\rho, \tilde{\mu}_k) \leq P_{k,\max}, \quad \forall k, \quad (47d)$$

$$(10b), (10c), (10e), \quad (47e)$$

where $\mathcal{P}_k(\rho, \tilde{\mu}_k) \triangleq (1 - \epsilon_1) \hat{P}_{F,k}^{\text{NOMA}}(\rho) + \frac{\epsilon_1}{h_k} \left(\frac{1}{\tilde{\mu}_k} - 1 \right)$, $\lambda_1(\rho) \triangleq \ln(e^{\omega_2 \rho_2} (e^{v_2} - e^{-\omega_1 \rho_1}))$, and $\lambda_2(\rho) \triangleq \ln(e^{\omega_1 \rho_1} (e^{v_1} - e^{-\omega_2 \rho_2}))$. Note that $\lambda_k \leq 0$ in (47c) is equivalent to (19c) to guarantee that $\hat{P}_{F,k}^{\text{NOMA}}(\rho) \geq 0, \forall k$. Problem (47) is still non-convex due to the non-convexity of $\hat{P}_{F,1}^{\text{NOMA}}(\rho)$ in (47a) and $\{\lambda_k(\rho)\}$ in (47c), which all contain a difference of two convex functions. This form can be effectively handled by the SCA technique [44] that replaces the non-convex objective function (47a) and constraint (47c) with suitable convex approximations, e.g., approximating $e^{\omega_2 \rho_2 + v_2}$ in $\hat{P}_{F,1}^{\text{NOMA}}(\rho)$ and $e^{-\omega_k \rho_k}$ in $\lambda_k(\rho)$ in the i th iteration with their first-order Taylor expansion at the given local point $\rho_k^{(i-1)}$ that denotes the solution of ρ_k obtained in the $(i-1)$ th iteration, and solves problem (47) iteratively until its objective value converges.

B. Optimizing the Offloading Time

For given ρ and d , problem (NOMA-F) reduces to

$$\min_t \sum_{k=1}^2 \kappa_U \frac{(\rho_k c_k L_k)^3}{(T_{\max} - t)^2} + \bar{P}_{F,k}^{\text{NOMA}} t \quad (48a)$$

$$\text{s.t.} \quad (10b), (10f), (19b), (19c). \quad (48b)$$

The expressions of transmit power in (15) and (16) can be simplified as

$$\hat{P}_{F,1}^{\text{NOMA}}(t) = \frac{1}{h_1} \left(e^{\frac{z_1 + z_2}{t} + \frac{s_1 + s_2}{\sqrt{t}}} - e^{\frac{z_2}{t} + \frac{s_2}{\sqrt{t}}} \right), \quad (49a)$$

$$\hat{P}_{F,2}^{\text{NOMA}}(t) = \frac{1}{h_2} \left(e^{\frac{z_2}{t} + \frac{s_2}{\sqrt{t}}} - 1 \right), \quad (49b)$$

$$\check{P}_{F,1}^{\text{NOMA}}(t) = \frac{1}{h_1} \left(\frac{1}{1 - e^{\frac{z_2}{t} + \frac{s_2}{\sqrt{t}}} + e^{\frac{z_2 - z_1}{t}}} - 1 \right), \quad (49c)$$

$$\check{P}_{F,2}^{\text{NOMA}}(t) = \frac{1}{h_2} \left(\frac{1}{1 - e^{\frac{z_1}{t} + \frac{s_1}{\sqrt{t}}} + e^{\frac{z_1 - z_2}{t}}} - 1 \right), \quad (49d)$$

where $z_k = \frac{\ln 2 \rho_k L_k}{B(1 - \epsilon_k)}$ and $s_k = \frac{Q^{-1}(\epsilon_k)}{\sqrt{B}}$. Similarly, to handle the complex forms of $\{\check{P}_{F,k}^{\text{NOMA}}(t)\}$, by introducing slack variables $\check{\mathbf{p}} \triangleq \{\check{p}_1, \check{p}_2\}$, $\check{\zeta} \triangleq \{\check{\zeta}_1, \check{\zeta}_2\}$, and $\check{\mathbf{e}} \triangleq \{\check{e}_1, \check{e}_2\}$, problem (48) can be equivalently transformed into

$$\min_{t, \check{\mathbf{p}}, \check{\zeta}, \check{\mathbf{e}}} \sum_{k=1}^2 \kappa_U \frac{(\rho_k c_k L_k)^3}{(T_{\max} - t)^2} + \check{e}_k \quad (50a)$$

$$\text{s.t.} \quad \check{p}_k \geq \mathcal{G}_k(t, \check{\zeta}_k), \quad \forall k, \quad (50b)$$

$$\check{\zeta}_k \leq \zeta_k(t), \quad \forall k, \quad (50c)$$

$$0 \leq \check{\zeta}_k \leq 1, \quad \forall k, \quad (50d)$$

$$\check{e}_k \geq \check{p}_k t, \quad \forall k, \quad (50e)$$

$$\check{p}_k \leq P_{k,\max}, \quad \forall k, \quad (50f)$$

$$(10b), (10f), \quad (50g)$$

where $\mathcal{G}_k(t, \zeta_k) \triangleq (1 - \epsilon_1) \hat{P}_{F,k}^{\text{NOMA}}(t) + \frac{\epsilon_1}{h_k} \left(\frac{1}{\zeta_k} - 1 \right)$, $\varsigma_1(t) \triangleq 1 - e^{-\frac{z_2}{t} + \frac{s_2}{\sqrt{t}}} + e^{-\frac{z_2 - z_1}{t}}$, and $\varsigma_2(t) \triangleq 1 - e^{-\frac{z_1}{t} + \frac{s_1}{\sqrt{t}}} + e^{-\frac{z_1 - z_2}{t}}$, and note that (50d) is equivalent to (19c). Problem (50) is non-convex due to the non-convexities of constraints (50b) and (50c) as well as the bilinear constraints in (50e). To overcome this difficulty, we will approximate problem (50) as a convex one and solve it in an iterative manner.

In the i th iteration, $\check{p}_k t$ in (50e) can be first rewritten as a difference of two convex functions, i.e., $\check{p}_k t = \left(\frac{\check{p}_k + t}{2} \right)^2 - \left(\frac{\check{p}_k - t}{2} \right)^2$. Then, by applying the first-order Taylor expansion on $\left(\frac{\check{p}_k - t}{2} \right)^2$ with the given local point $\{\check{p}_k^{(i-1)}, t^{(i-1)}\}$, where $\check{p}_k^{(i-1)}$ and $t^{(i-1)}$ respectively denote the solutions of $\check{p}_k$ and t obtained in the $(i-1)$ th iteration, constraint (50e) can be approximated by

$$\check{e}_k \geq \frac{1}{4}(\check{p}_k + t)^2 + \frac{1}{4}(\check{p}_k^{(i-1)} - t^{(i-1)})^2 - \frac{1}{2}(\check{p}_k^{(i-1)} - t^{(i-1)})(\check{p}_k - t), \forall k. \quad (51)$$

Finally, the non-convexities of $\hat{P}_{F,1}^{\text{NOMA}}(t)$ in (50b) and $\{\varsigma_k(t)\}$ in (50c) that caused by their forms of the difference of two convex functions can be effectively handled by SCA technique, i.e., approximating problem (50) by replacing $e^{-\frac{z_2}{t} + \frac{s_2}{\sqrt{t}}}$ in $\hat{P}_{F,1}^{\text{NOMA}}(t)$, $e^{-\frac{z_2 - z_1}{t}}$ in $\varsigma_1(t)$, and $e^{-\frac{z_1 - z_2}{t}}$ in $\varsigma_2(t)$ with their first-order Taylor expansion at the given local points $t^{(i-1)}$.

C. Optimizing the UAV Location

For given ρ and t , problem (NOMA-F) reduces to

$$\min_d \quad \bar{P}_{F,1}^{\text{NOMA}} + \bar{P}_{F,2}^{\text{NOMA}} \quad (52a)$$

$$\text{s.t.} \quad (10g), (19b). \quad (52b)$$

The objective function in (52a) is convex with respect to d , and the solution of problem (52) can be derived as

$$d^* = \left[\frac{\Psi(\rho_1, \rho_2, N)}{\Phi(\rho_1, \rho_2, N) + \Psi(\rho_1, \rho_2, N)} D \right]_{d_{\min}}^{d_{\max}}, \quad (53)$$

where $[x]_b^a \triangleq \min\{a, \max\{b, x\}\}$, $\Phi(\rho_1, \rho_2, N) \triangleq (1 - \epsilon_1) \times (\Upsilon_1^F(\rho_1, N)(\Upsilon_2^F(\rho_2, N) + 1)) + \epsilon_1 \frac{\Upsilon_1^F(\rho_1, N)(\Upsilon_2^F(\rho_2, N) + 1)}{1 - \Upsilon_1^F(\rho_1, N)\Upsilon_2^F(\rho_2, N)}$, and $\Psi(\rho_1, \rho_2, N) \triangleq (1 - \epsilon_1)\Upsilon_2^F(\rho_2, N) + \epsilon_1 \frac{\Upsilon_2^F(\rho_2, N)(\Upsilon_1^F(\rho_1, N) + 1)}{1 - \Upsilon_1^F(\rho_1, N)\Upsilon_2^F(\rho_2, N)}$.

In addition, $d_{\min} = D - \sqrt{\frac{P_{2,\max}\beta_0}{BN_0\Psi(\rho_1, \rho_2, N)}} - H^2$ and $d_{\max} = \min\{\frac{1}{2}D, \sqrt{\frac{P_{1,\max}\beta_0}{BN_0\Phi(\rho_1, \rho_2, N)}} - H^2\}$ are obtained from (10g) and (19b).

D. Overall Algorithm

We summarize the overall algorithm for solving problem (NOMA-F) in Algorithm 1, where three subproblems will be solved alternatively and iteratively. Since the SCA-based algorithm is guaranteed to converge based on its global convergence and stability [44], in the i th iteration, we have $E_F^{\text{NOMA}}(\rho^{(i)}, t^{(i-1)}, d^{(i-1)}) \leq E_F^{\text{NOMA}}(\rho^{(i-1)}, t^{(i-1)}, d^{(i-1)})$ after solving problem (47), and $E_F^{\text{NOMA}}(\rho^{(i)}, t^{(i)}, d^{(i-1)}) \leq E_F^{\text{NOMA}}(\rho^{(i)}, t^{(i-1)}, d^{(i-1)})$

Algorithm 1 Proposed Algorithm for Problem (NOMA-F).

- 1: **Initialization:** Set initial values for $\rho^{(0)}$, $t^{(0)}$, $d^{(0)}$, accuracy σ , the iteration number $i = 0$, and calculate $E_F^{\text{NOMA}}(\rho^{(0)}, t^{(0)}, d^{(0)})$ according to (18).
 - 2: **repeat**
 - 3: Set $i = i + 1$.
 - 4: Given $\rho^{(i-1)}$, $t^{(i-1)}$, and $d^{(i-1)}$, obtain $\rho^{(i)}$ by solving problem (47) with the SCA-based algorithm.
 - 5: Given $\rho^{(i)}$, $t^{(i-1)}$, and $d^{(i-1)}$ and calculate $\check{p}_k^{(i-1)}$ according to (17), obtain $t^{(i)}$ by solving problem (50) with the SCA-based algorithm.
 - 6: Given $\rho^{(i)}$ and $t^{(i)}$, obtain $d^{(i)}$ according to (53).
 - 7: Calculate $E_F^{\text{NOMA}}(\rho^{(i)}, t^{(i)}, d^{(i)})$ according to (18).
 - 8: **until** $\left| \frac{E_F^{\text{NOMA}}(\rho^{(i)}, t^{(i)}, d^{(i)}) - E_F^{\text{NOMA}}(\rho^{(i-1)}, t^{(i-1)}, d^{(i-1)})}{E_F^{\text{NOMA}}(\rho^{(i-1)}, t^{(i-1)}, d^{(i-1)})} \right| \leq \sigma$.
 - 9: **Output:** $\rho^* = \rho^{(i)}$, $t^* = t^{(i)}$, $d^* = d^{(i)}$, and $E_F^{\text{NOMA}*} = E_F^{\text{NOMA}}(\rho^{(i)}, t^{(i)}, d^{(i)})$.
-

after solving problem (50). And since d^* in (53) is the optimal solution to problem (52), we furtherly have $E_F^{\text{NOMA}}(\rho^{(i)}, t^{(i)}, d^{(i)}) \leq E_F^{\text{NOMA}}(\rho^{(i-1)}, t^{(i-1)}, d^{(i-1)})$. Thus, Algorithm 1 is guaranteed to converge since the optimal value is non-increasing and must have a lower bound during each iteration.

The computational complexity of the proposed algorithm mainly depends on the application of CVX solver for solving subproblems that optimize the portions of task offloaded and the offloading time, i.e., problems (47) and (50), in an iterative manner. These two subproblems can be solved by a primal-dual interior point method with complexity $\mathcal{O}(n^{3.5}\log(1/\sigma))$, where n and σ denote the number of optimization variables and the given convergence accuracy, respectively [45], [46]. Since there are $n_1 = 6$ and $n_2 = 5$ optimization variables in problems (47) and (50), respectively, and the computational complexity for calculating the UAV location according to (53) can be ignored here, the total computational complexity of Algorithm 1 can be calculated as $\mathcal{O}(I(n_1^{3.5} + n_2^{3.5})\log(1/\sigma))$, where I is the required number of iterations for convergence.⁵

V. SIMULATION RESULTS

In this section, we provide some simulation results to compare the performances of different multiple access offloading schemes and demonstrate the effectiveness of the proposed algorithm. The system parameters are set as follows: $N_0 = -169$ dBm/Hz, $\beta_0 = -60$ dB, $D = 100$ m, $H = 50$ m, $f_{U,\max} = 9$ GHz, $\kappa_U = 10^{-28}$, $\eta = \delta = 0.5$, and for both $k \in \{1, 2\}$, $f_{k,\max} = 1$ GHz, $P_{k,\max} = 0.1$ W, $c_k = 1000$ cycles/b, $\kappa_k = 10^{-28}$, and $\epsilon_k = \epsilon$, where ϵ denotes a specific decoding error probability that may take different values in the simulations. The proposed algorithm will be compared with the following three benchmark algorithms for problem (NOMA-F):

⁵To mitigate the potential conflict between the computational burden of the optimization algorithm and the strict task computation latency constraints, the optimal computation offloading strategies for various predicted UE locations and task data size ratios can be pre-calculated offline and stored in a database on the network controller.

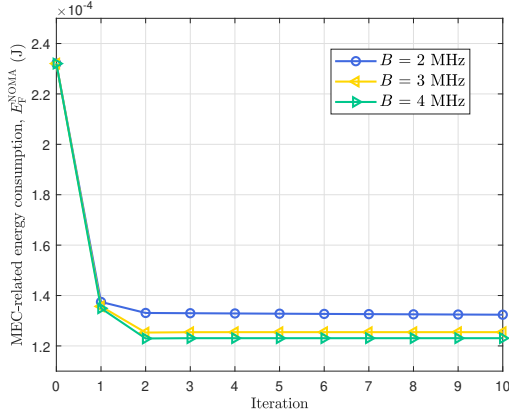


Fig. 6. Convergence behavior of the proposed algorithm for problem (NOMA-F) for $B = 2, 3, 4$ MHz.

- *Without Optimizing the Portions of Task Offloaded:* It fixes the portions of task offloaded as 0.5, i.e., $\rho_k = 0.5, \forall k$, and optimizes the offloading time and the UAV location by solving problems (50) and (52), respectively. The MEC-related energy consumption obtained by this scheme is denoted by $E_{F, NP}^{NOMA}$.
- *Without Optimizing the Offloading Time:* It fixes the offloading time as $0.5T_{\max}$, and optimizes the portions of task offloaded and the UAV location by solving problems (47) and (52), respectively. The MEC-related energy consumption obtained by this scheme is denoted by $E_{F, NT}^{NOMA}$.
- *Exhaustive Search Method:* It minimizes $E_F^{NOMA}(\rho, t, d)$ by searching for the portions of task offloaded and the offloading time in the feasible domain of problem (NOMA-F) and calculating the UAV location according to (53). The MEC-related energy consumption obtained by this scheme is denoted by $E_{F, ES}^{NOMA}$, which serves as the globally optimal value of problem (NOMA-F).

First, in Fig. 6, for $L_1 = L_2 = 1.2$ kbit, $T_{\max} = 1$ ms, and $\epsilon = 10^{-5}$, we illustrate the convergence behavior of the proposed algorithm (Algorithm 1). It is observed that E_F^{NOMA} with different B decreases over iterations and converges after three iterations. This indicates that the proposed AO algorithm can converge quickly and demonstrates its effectiveness. Additionally, E_F^{NOMA} decreases with the increase of B since a larger blocklength $N = Bt$ leads to a higher offloading data rate for each UE. The convergence behaviors of the proposed algorithms for other five offloading schemes are similar, and we do not present them here due to space limitations.

In Fig. 7, by letting $L_1 = L_2 = L$, we show the MEC-related energy consumption of different multiple access offloading schemes and three benchmark schemes versus data size of UEs' computation tasks L . As expected, the MEC-related energy consumption of all considered schemes increases as the data size of UEs' computation tasks increases. E_F^{NOMA} is substantially lower compared with $E_{F, NP}^{NOMA}$ and $E_{F, NT}^{NOMA}$, and always approaches $E_{F, ES}^{NOMA}$, which demonstrates the effectiveness of the proposed joint optimization algorithm. The performance improvement achieved by the proposed algorithm over the first two benchmark schemes increases as L

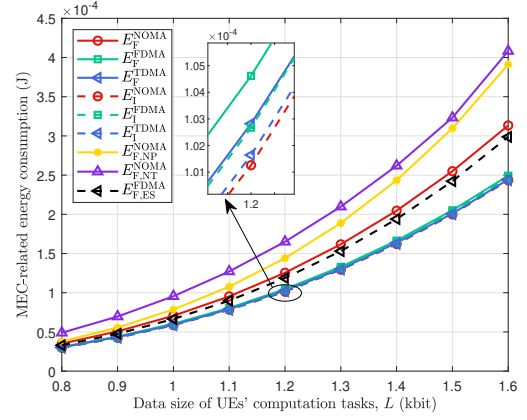


Fig. 7. MEC-related energy consumption of different schemes versus data size of UEs' computation tasks L for $B = 3$ MHz, $T_{\max} = 1$ ms, and $\epsilon = 10^{-5}$.

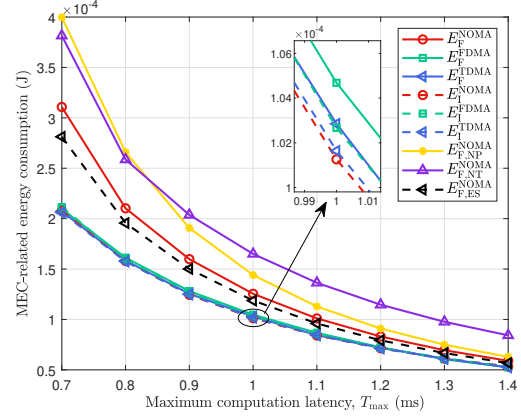


Fig. 8. MEC-related energy consumption of different schemes versus maximum computation latency $T_{\max}$ for $L_1 = L_2 = 1.2$ kbit, $B = 3$ MHz, and $\epsilon = 10^{-5}$.

increases, which indicates that the joint optimization of the portions of task offloaded and the offloading time is more important when the computational burden of UEs is heavier.

Fig. 8 shows the MEC-related energy consumption of different schemes versus maximum computation latency $T_{\max}$. It is observed that the MEC-related energy consumption of all schemes decreases with the increase of $T_{\max}$. This is because as the available time for both local computing and remote computing increases, the UEs and UAV can adjust their CPU frequencies to a smaller state according to the principle of the DVFS technique so that the computation-related energy consumption can be effectively reduced. We can also observe that E_F^{NOMA} is lower than $E_{F, NP}^{NOMA}$ and $E_{F, NT}^{NOMA}$ regardless of $T_{\max}$. And the gap between $E_{F, NP}^{NOMA}$ and $E_{F, NT}^{NOMA}$ becomes smaller with the increase of $T_{\max}$, while $E_{F, NT}^{NOMA}$ always maintains a larger performance gap with E_F^{NOMA} . This indicates that when the latency requirement is relatively severe, i.e., $T_{\max}$ is small, the joint optimization of the portion of task offloaded and the offloading time is extremely important to reduce energy consumption. Whereas when the latency requirement is relatively loose, i.e., $T_{\max}$

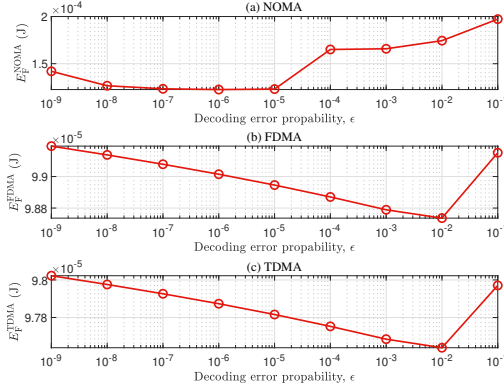


Fig. 9. MEC-related energy consumption of different schemes with finite blocklength versus decoding error probability ϵ for $L_1 = L_2 = 1.2$ kbit, $B = 4$ MHz, and $T_{\max} = 1$ ms.

is large, the performance gap caused by the lack of the optimization of the portions of task offloaded can be mitigated by the optimization of offloading time and the moderation effect of the DVFS technique. And the optimization of the offloading time is always important since it should cautiously balance the energy consumption caused by task offloading and remote computing.

In addition, in both Fig. 7 and Fig. 8, E_F^{NOMA} is larger than E_F^{FDMA} and E_F^{TDMA} regardless of L and $T_{\max}$. It indicates that the task offloading capability of NOMA is weaker than that of FDMA and TDMA when the blocklength is finite since the SIC failure should be considered additionally for NOMA in this case. Whether in infinite blocklength case or finite blocklength case, TDMA always achieves a lower MEC-related energy consumption compared to FDMA, i.e., $E_I^{\text{TDMA}} < E_I^{\text{FDMA}}$ and $E_F^{\text{TDMA}} < E_F^{\text{FDMA}}$, thanks to the freedom of remote-computing time brought by sequential task offloading of the UEs. Moreover, E_I^{NOMA} is always lower than E_I^{FDMA} due to the spectrum efficiency brought by time-frequency resource sharing and the perfect SIC. The above simulation results verify the correctness of the theoretical analysis conclusions obtained in Section III. We can also observe that the MEC-related energy consumption required by the finite blocklength case is always larger than that required by the infinite blocklength case for each considered multiple access offloading scheme because the Shannon capacity services as an upper bound for the finite-blocklength transmission data rate derived in [12].

Fig. 9 shows the MEC-related energy consumption required by different multiple access offloading schemes with finite blocklength versus decoding error probability ϵ . For any value of ϵ , it can also be observed that E_F^{NOMA} is larger than E_F^{FDMA} , and E_F^{FDMA} is larger than E_F^{TDMA} . In Fig. 9(a), it is observed that E_F^{NOMA} decreases with the increase of ϵ when $\epsilon \leq 10^{-5}$, and increases with the increase of ϵ when $\epsilon > 10^{-5}$. This is because ϵ will affect both the achievable data rate and the SIC failure probability in NOMA offloading as we consider in Section II-A-2. When the reliability requirement is strict, i.e., ϵ takes an extremely small value, the SIC failure

probability is low and the average transmit power derived in (17) is dominated by the transmit power required for SIC success case, increasing ϵ enables the achievable data rate of the offloading link to reach a larger value, which provides a larger computation offloading capacity for UEs to obtain a more flexible task computing strategy and promotes the reduction of the MEC-related energy consumption. However, when the reliability requirement is slack, i.e., ϵ takes a large value, although increasing ϵ can still result in a larger achievable data rate, the SIC failure probability is already so large in this case that the transmit power required for SIC failure already occupies a significant portion of the average transmit power, and increasing ϵ will instead result in a lower effective offloading throughput and a larger communication-related energy consumption due to $\hat{P}_{F,k}^{\text{NOMA}} > \hat{P}_{F,k}^{\text{NOMA}}, \forall k$. In Fig. 9(b) and Fig. 9(c), as ϵ increases, both E_F^{FDMA} and E_F^{TDMA} decrease slightly when $\epsilon \leq 10^{-2}$ and increase when $\epsilon > 10^{-2}$. Since the UEs offload their tasks to the UAV over orthogonal communication resources in FDMA and TDMA offloadings, the UAV does not need to perform SIC to decode the received UEs' signal, so ϵ mainly affects the achievable data rate during task offloading transmission, and increasing ϵ will decrease the energy consumption. However, when ϵ is too large, i.e., when $\epsilon > 10^{-2}$, the impact of ϵ on the effective offloading throughput defined in the LHS of (26) and (37) is significant, increasing ϵ instead will increase the energy consumption. Based on the above discussion, we can find that the decoding error probability has a more significant impact on NOMA than FDMA and TDMA since both SIC success and failure need to be considered simultaneously when the blocklength is finite.

In Fig. 10, we show the optimal portions of task offloaded, offloading times, UAV locations, and transmit power of different schemes versus the data size of UE 2's computation task L_2 for $L_1 = 1.2$ kbit. In Fig. 10(a) and Fig. 10(b), we can observe that the finite-blocklength NOMA offloading scheme obtains the lowest portion of task offloaded and the largest offloading time, which decreases and increases as L_2 increases, respectively. This indicates that the finite-blocklength NOMA offloading needs to restrict the offloading throughput and allocate sufficient offloading time to ensure the reliability of offloading transmission can be satisfied even in the SIC failure case as we analyzed in Section III-C. Different from NOMA and FDMA in which ρ_1 and ρ_2 are very close, ρ_1 is higher than ρ_2 in both infinite-blocklength and finite-blocklength TDMA offloading schemes since UE 1 offloads its task before UE 2 and then UE 1 can have more time to perform remote computing at the UAV. In Fig. 10(c), the UAV locations obtained by all schemes become closer to UE 2 as L_2 increases to improve the quality of UE 2's offloading link to meet UE 2's task offloading demand. The UAV locations in FDMA and TDMA schemes vary significantly and the ratios of distance from the UAV to UEs 1 and 2 are proportional to the ratios of task data size between UEs 1 and 2. However, in NOMA offloading with both infinite blocklength and finite blocklength, the UAV only moves slightly and it is always closer to UE 1 to ensure $h_1 \geq h_2$ under the assumption that

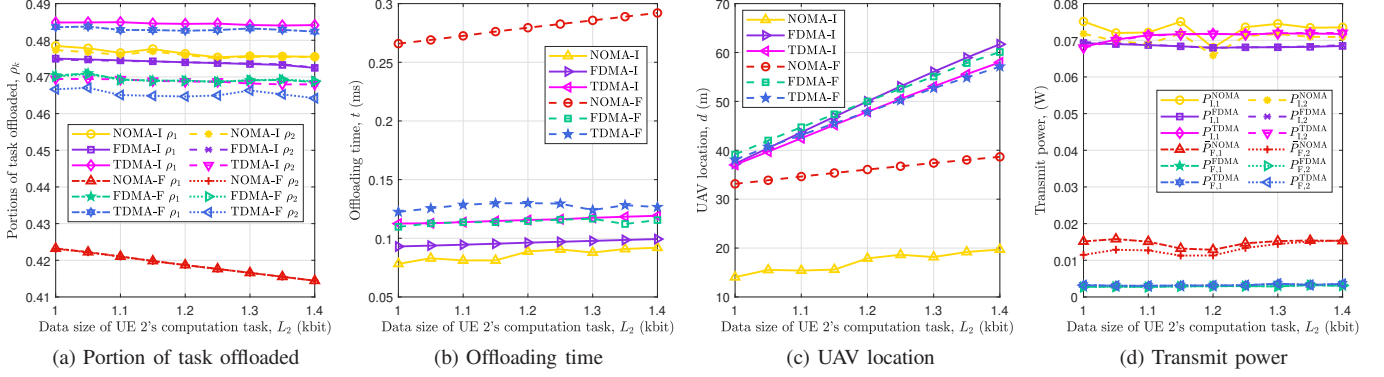


Fig. 10. Optimized portions of task offloaded, offloading times, UAV locations, and transmit power of different schemes versus data size of UE 2's computation task L_2 for $L_1 = 1.2$ kbit, $B = 3$ MHz, $T_{\max} = 1$ ms, and $\epsilon = 10^{-5}$.

the UAV first decodes UE 1's signal. In Fig. 10(d), in order to distinguish the signals of two UEs in the power domain, the transmit power of UEs 1 and 2 is different in NOMA schemes, which is different from FDMA and TDMA offloadings. Based on the optimization results presented in Fig. 10, we can find that whether in infinite blocklength case or finite blocklength case, NOMA mainly balances the difference in offloading demand between two UEs due to the difference between L_1 and L_2 by adjusting the portions of task offloaded and the offloading time, while FDMA and TDMA mainly balances that by adjusting the UAV locations.

VI. CONCLUSION

In this paper, we have considered the minimum MEC-related energy consumption for different multiple access offloading schemes, including NOMA, FDMA, and TDMA in both infinite blocklength and finite blocklength cases. Through theoretical analysis, we have proved that TDMA consistently achieves lower energy consumption than FDMA, and NOMA achieves lower energy consumption than FDMA in the infinite blocklength case. In contrast, since both SIC success and failure cases were considered in finite-blocklength NOMA offloading, NOMA could not be guaranteed to achieve lower energy consumption than FDMA in this case, especially when the channel conditions and the offloaded task data sizes of two UEs are relatively symmetric. Furthermore, we have proposed an alternating optimization algorithm to minimize the MEC-related energy consumption of different schemes by jointly optimizing the portions of task offloaded, the offloading times of all UEs, and the UAV location. Simulation results have validated that the proposed algorithm can effectively reduce MEC-related energy consumption compared to other benchmark schemes and verified the correctness of the theoretical analysis results. In addition, our obtained results have shown that the decoding error probability has a more significant impact on the energy consumption of NOMA than both FDMA and TDMA in the finite blocklength case.

REFERENCES

- [1] D. Sabella, A. Vaillant, P. Kuure, *et al.*, "Mobile-edge computing architecture: The role of MEC in the Internet of Things," *IEEE Consum. Electron. Mag.*, vol. 5, no. 4, pp. 84–91, Oct. 2016.
- [2] J. Zhang *et al.*, "Stochastic computation offloading and trajectory scheduling for UAV-assisted mobile edge computing," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3688–3699, Apr. 2019.
- [3] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.
- [4] G. Zhang, Q. Wu, M. Cui, and R. Zhang, "Securing UAV communications via joint trajectory and power control," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1376–1389, Feb. 2019.
- [5] M. Abrar, U. Ajmal, Z. M. Almohaimeed, X. Gui, R. Akram, and R. Masroor, "Energy efficient UAV-enabled mobile edge computing for IoT devices: A review," *IEEE Access*, vol. 9, pp. 127779–127798, 2021.
- [6] C. Xu, C. Zhan, H. Yang, and L. Xiao, "Pareto-optimal aerial-ground energy minimization for aerial 3D mobile edge computing networks," *IEEE Trans. Veh. Technol.*, vol. 73, no. 5, pp. 7218–7233, May 2024.
- [7] C. Zhan, H. Hu, X. Sui, Z. Liu, and D. Niyato, "Completion time and energy optimization in the UAV-enabled mobile-edge computing system," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7808–7822, Aug. 2020.
- [8] M. Salmani and T. N. Davidson, "Energy-optimal multiple access computation offloading: Signalling structure and efficient communication resource allocation," *IEEE Trans. Signal Process.*, vol. 68, pp. 1646–1661, 2020.
- [9] Q. Wu *et al.*, "A comprehensive overview on 5G-and-beyond networks with UAVs: From communications to sensing and intelligence," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 10, pp. 2912–2945, Oct. 2021.
- [10] S. K. Singh, K. Agrawal, K. Singh, Y. -M. Chen, and C. -P. Li, "Performance analysis and optimization of RSMA enabled UAV-aided IBL and FBL communication with imperfect SIC and CSI," *IEEE Trans. Wireless Commun.*, vol. 22, no. 6, pp. 3714–3732, Jun. 2023.
- [11] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.
- [12] 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.
- [13] X. Sun, S. Yan, N. Yang, Z. Ding, C. Shen, and Z. Zhong, "Short-packet downlink transmission with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4550–4564, Jul. 2018.
- [14] 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.
- [15] J. Zhang *et al.*, "Computation-efficient offloading and trajectory scheduling for multi-UAV assisted mobile edge computing," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 2114–2125, Feb. 2020.
- [16] Q. Wu, M. Cui, G. Zhang, F. Wang, Q. Wu, and X. Chu, "Latency minimization for UAV-enabled URLLC-based mobile edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 3298–3311, Apr. 2024.
- [17] F. Wang, J. Xu, X. Wang, and S. Cui, "Joint offloading and computing optimization in wireless powered mobile-edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1784–1797, Mar. 2018.
- [18] 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.
- [19] J. Kang *et al.*, “Tiny multiagent DRL for twins migration in UAV metaverses: A multileader multifollower stackelberg game approach,” *IEEE Internet Things J.*, vol. 11, no. 12, pp. 21021–21036, Jun. 2024.
- [20] M. Hua, Y. Wang, C. Li, Y. Huang, and L. Yang, “UAV-aided mobile edge computing systems with one by one access scheme,” *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 3, pp. 664–678, Sep. 2019.
- [21] A. Kiani and N. Ansari, “Edge computing aware NOMA for 5G networks,” *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1299–1306, Apr. 2018.
- [22] Z. Ding, D. W. K. Ng, R. Schober, and H. V. Poor, “Delay minimization for NOMA-MEC offloading,” *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1875–1879, Dec. 2018.
- [23] F. Fang, Y. Xu, Z. Ding, C. Shen, M. Peng, and G. K. Karagiannidis, “Optimal resource allocation for delay minimization in NOMA-MEC networks,” *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7867–7881, Dec. 2020.
- [24] Y. Pan, M. Chen, Z. Yang, N. Huang, and M. Shikh-Bahaei, “Energy-efficient NOMA-based mobile edge computing offloading,” *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 310–313, Feb. 2019.
- [25] W. Huang and Z. Ding, “New insight for multi-user hybrid NOMA offloading strategies in MEC networks,” *IEEE Trans. Veh. Technol.*, vol. 73, no. 2, pp. 2918–2923, Feb. 2024.
- [26] 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.
- [27] Z. Yang, C. Pan, J. Hou, and M. Shikh-Bahaei, “Efficient resource allocation for mobile-edge computing networks with NOMA: Completion time and energy minimization,” *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7771–7784, Nov. 2019.
- [28] S. Jeong, O. Simeone, and J. Kang, “Mobile edge computing via a UAV-mounted cloudlet: Optimization of bit allocation and path planning,” *IEEE Trans. Veh. Technol.*, vol. 67, no. 3, pp. 2049–2063, Mar. 2018.
- [29] X. Zhang, J. Zhang, J. Xiong, L. Zhou, and J. Wei, “Energy-efficient multi-UAV-enabled multiaccess edge computing incorporating NOMA,” *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5613–5627, Jun. 2020.
- [30] Y. Yang, Y. Hu, and M. C. Gursoy, “Energy efficiency of RIS-assisted NOMA-based MEC networks in the finite blocklength regime,” *IEEE Trans. Commun.*, vol. 72, no. 4, pp. 2275–2291, Apr. 2024.
- [31] K. Higuchi and A. Benjebbour, “Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access,” *IEICE Trans. Commun.*, vol. 98, no. 3, pp. 403–414, Mar. 2015.
- [32] B. Zheng, Q. Wu, and R. Zhang, “Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA?,” *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 753–757, Apr. 2020.
- [33] G. Chen, Q. Wu, W. Chen, D. W. K. Ng, and L. Hanzo, “IRS-aided wireless powered MEC systems: TDMA or NOMA for computation offloading?,” *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1201–1218, Feb. 2023.
- [34] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. Shikh-Bahaei, “Energy efficient federated learning over wireless communication networks,” *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1935–1949, Mar. 2021.
- [35] H. Jiang, Y. Zhang, M. Liu, and X. Wang, “Binary neural networks for edge-side object detection,” *ACM Trans. Embed. Comput. Syst.*, vol. 21, no. 5, pp. 1–22, Nov. 2022.
- [36] S. U. Stich, J. B. Cordonnier, and M. Jaggi, “Sparsified SGD with memory,” in *Proc. Adv. Neural Inf. Process. Syst.*, Montreal, QC, Canada, Dec. 2018, pp. 4448–4459.
- [37] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, “Energy-optimal mobile cloud computing under stochastic wireless channel,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4569–4581, Sep. 2013.
- [38] Y. Zeng, J. Xu, and R. Zhang, “Energy minimization for wireless communication with rotary-wing UAV,” *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2329–2345, Apr. 2019.
- [39] Z. Ding *et al.*, “Application of non-orthogonal multiple access in LTE and 5G networks,” *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.
- [40] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, “A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems,” *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7244–7257, Nov. 2016.
- [41] M. Gupta and S. Prakriya, “Performance of multiuser uplink underlay NOMA networks with channel knowledge,” *IEEE Trans. Veh. Technol.*, vol. 73, no. 4, pp. 6000–6005, Apr. 2024.
- [42] N. S. Mouni, A. Kumar, and P. K. Upadhyay, “Adaptive user pairing for NOMA systems with imperfect SIC,” *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1547–1551, Jul. 2021.
- [43] M. Vaezi, R. Schober, Z. Ding, and H. V. Poor, “Non-orthogonal multiple access: Common myths and critical questions,” *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 174–180, Oct. 2019.
- [44] Q. Dinh *et al.*, “Local convergence of sequential convex programming for nonconvex optimization,” in *Recent Advances in Optimization and its Applications in Engineering*. Berlin, Germany: Springer, 2010, pp. 93–102.
- [45] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [46] S. Bubeck, “Convex optimization: Algorithms and complexity,” *Found. Trends Mach. Learn.*, vol. 8, nos. 3–4, pp. 231–357, 2015.