JOURNAL OF MECHANICS OF CONTINUA AND
MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online): 2454 -7190 Vol.-15, No.-9, September (2025) pp 38-51 ISSN (Print) 0973-8975

HYBRID NOMA-BASED RESOURCE ALLOCATION FOR MULTI-ACCESS EDGE COMPUTING IN HETNETS

Hind S. Ghazi [1], Ali Al-Shuwaili [2], Adham R. Azeez [3], Maciej Krasicki[4]

[1,2,3] Media Technology and Communication Engineering Department,
University of Information Technology and Communications, Iraq.

[4] Computing and Telecommunications, Institute of Radiocommunications,
Poznan University of Technology, Poland.

Email: hind.salim@uoitc.edu.iq[1], ali.najdi@uoitc.edu.iq[2]
adham.azeez@uoitc.edu.iq[3], maciej.krasicki@put.poznan.pl[4]

Corresponding Author: **Hind S. Ghazi**

https://doi.org/10.26782/jmcms.2025.09.00003

(Received: June 27, 2025; Revised: August 14, 2025; Accepted: September 02, 2025)

Abstract

This paper investigates the optimal offloading policy in heterogeneous networks where radio resources are provided, in both uplink and downlink, via two distinct types of BSs, namely Macro-cell Base Stations (MBSs) and Small-cell Base Stations (SBSs), in a multi-tier Multi-Access Edge Computing (MEC) -assisted scenario. Since the feasibility of the offloading problem is a function of radio connectivity in uplink and downlink, we propose to assign radio links using the flexible hybrid NOMA scheme that leverages both the limited interference of OMA as well as the faster data rates of NOMA. To this end, we formulate an optimization problem aiming to optimize the allocation of both radio and computation resources while minimizing the offloading energy across all users. The formulated problem is then tackled by means of decomposition and relation. The numerical results show that the Hybrid NOMA scheme balances subchannel allocation and power control to maintain high spectral efficiency and low offloading latency without the decoding overhead of Full NOMA (high interference) or the inefficiency of No NOMA (low subchannel reuse).

Keywords: Multi-Access Edge Computing (MEC), Non-Orthogonal Multiple Access (NOMA), fifth-generation (5G), Heterogeneous Networks (HetNets), Hybrid NOMA.

I. Introduction

In the era of fifth-generation (5G) and beyond, smart mobile devices have become essential tools enabling innovative mobile applications, including virtual reality (VR), augmented reality (AR), autonomous vehicles, and the Internet of Things (IoT), which, in turn, have recently driven the advancement of wireless communication networks. [I]. However, most of them usually require intensive computation and real-time responses, which pose unprecedented challenges to mobile devices constrained by limited computing capabilities, storage, and battery life [II]. In the context of these challenges, various advanced solutions have been considered. Among them, MEC is presented as a transformative approach to improve the computational power of mobile devices and address the needs of latency-sensitive tasks [III]. The primary idea of MEC is to provide cloud computing and IT capabilities within the radio access network (RAN) (e.g., base stations, BS) in proximity to mobile users (MUs); It authorizes them to transform their delay-sensitive and computation-intensive tasks to MEC servers that are positioned at the edge of radio access networks (RANs) [IV, V]. Taking advantage of this mechanism can significantly reduce the total network transmission latency and energy consumption of smart mobile devices [VL].

To achieve further enhancement in spectrum efficiency and system throughput, the concept of Heterogeneous Networks (HetNets) and Non-Orthogonal Multiple Access (NOMA) has been introduced [VIL]. HetNets can achieve more spectrum-efficient communications by positioning small cells, such as picocells and femtocells, within the existing macrocells. Both co-tier and cross-tier interference have resulted due to spectrum sharing across multi-tier cells, which can severely deteriorate communication quality in these networks [VIIL]. As a result, exploiting the NOMA technique in HetNets has attracted significant research interest and emerged as an optimal solution to overcome the resource allocation and interference management challenges for HetNets. In the NOMA framework, improved spectrum efficiency and enhanced system performance can be achieved by concurrently sharing the same frequency-time resource between multiple users using either the power domain (PD-NOMA) or the code domain (CD-NOMA) [IX, X].

Numerous studies have inspected the integration of computation offloading and resource allocation strategies in NOMA-enabled MEC within HetNets, aspiring to minimize the weighted sum cost of delay and energy consumption in multicell MEC networks. In [XL, XIL], a joint optimization problem for radio resource allocation and task offloading has been introduced to leverage the advantages of NOMA, considering key factors affecting its efficiency, such as intra-cell and inter-cell interference, power control, and subchannel allocation; where an iterative algorithm was proposed to reduce latency and energy consumption for participating users. In like manner, the authors of [VIL, XIIL] suggested an efficient approach to jointly optimize task offloading decisions, local CPU frequency scheduling, power control, and computational resource allocation, focusing on mitigating user energy consumption and enhancing overall system performance while maintaining manageable complexity. Moreover, the NOMA-enabled dynamic task offloading problem in heterogeneous MEC networks has been studied in [XIV]. The authors proposed a dynamic task offloading (NDTO) algorithm leveraging NOMA and stochastic optimization techniques to diminish system energy consumption while sustaining the stability of the task queue. In [XV], a NOMA-based vehicular edge computing (VEC) network model has been investigated, and a cost minimization problem has been proposed. By jointly optimizing offloading decisions, VUE clustering, subchannel assignment, computational resource allocation, and transmission power control, the system cost was reduced while ensuring the delay tolerance requirements of all VUEs. Two heuristic algorithms have been suggested to solve the task-offloading problem and the MEC resource assignment problem.

To maximize energy efficiency while simultaneously minimizing the latency of cell edge devices that lie at longer distances from BS and experience degraded channels due to multipath, shadowing, and fading effects; the authors of [XVL] have proposed a cooperative offloading multiaccess edge computing (COMEC) scheme which integrates NOMA aided joint processing coordinated multipoint (JP-CoMP) with HetNet for distributed offloading of tasks by cell edge users.

A different scenario has been applied in this article. It presents a framework that integrates the principle of hybrid NOMA for Multi-Access Edge Computing in HetNets to minimize the total energy expenditure of mobile devices and offloading latency, as well as optimize the allocation of radio and computation resources. In the proposed system, both computation and radio resources are considered. For computation resources, MEC servers in each tier are used in cooperation to offer mobile tasks offloading, whereas the radio resources for both uplink and downlink are provided by employing two distinct types of BSs, specifically MBS and SBS in the heterogeneous multi-tier scenario. A hybrid NOMA scheme has also been suggested for improving system connectivity and overcoming interference challenges, as well as achieving higher data rates.

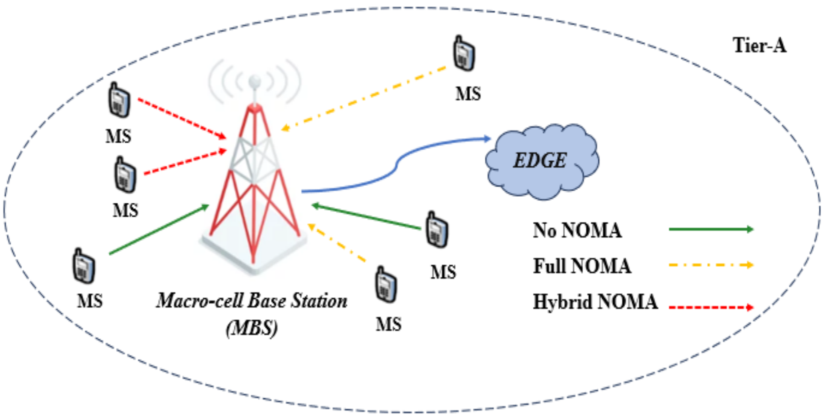

Fig. 1. Tier-A of a heterogeneous cloudlet-aided mobile computing network that consists of one MEC server and one Macro-Cell Base Station (MBS). Note that tier-B shows a similar arrangement while having SBSs instead of MBSs.

II. System Model

The network under consideration is a heterogeneous cloud-aided mobile computing system comprising a central cloud and a two-tier architecture featuring $\boldsymbol{K}$ Macro-cell Base Stations (MBSs) and Small-cell Base Stations (SBSs). MBSs and SBSs are directly linked to local computing servers, often referred to as cloudlets [XVIL], which may execute certain applications on behalf of Mobile Users (MUs) (see Fig. 1). Both tiers use frequency division duplex (FDD) and operate on the same frequency band [XVIIL]. All Base Stations (BSs) within the same tier are supposed to transmit at identical power levels, and are denoted by $\boldsymbol{P^{k,dl}}$ with $\boldsymbol{k} \in \{\boldsymbol{M}, \boldsymbol{S}\}$. We denote by $\boldsymbol{\mathcal{K}} =$

$\{\mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{K}\}$ the set of BSs in both tiers. The set of single-antenna mobile devices of the MUs is denoted as $\boldsymbol{\mathcal{I}} = \{\mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{I}\}$. All MDs may offload mobile applications to a cloudlet and/or cloud servers using the allocated subchannel $\boldsymbol{j} \in \boldsymbol{\mathcal{J}}$, where $\boldsymbol{\mathcal{J}} = \{\mathbf{1}, \mathbf{2}, \ldots, \mathbf{J}\}$ is the set of available subchannels in the uplink direction. Note that the index $\boldsymbol{i_j^k}$ refers to the $\boldsymbol{i} - \mathbf{th}$ MU connected to a BS in tier $\boldsymbol{k}$ and scheduled on the $\boldsymbol{j}\mathbf{th}$ subchannel. We assume the number of MUs is larger than the number of available subchannels. Accordingly, each MU is either allocated one dedicated subchannel in an OMA-like fashion or a pair of MUs is scheduled on one subchannel as a non-orthogonal pair in a NOMA-like fashion, i.e., hybrid NOMA. To this end, the subset of orthogonal MUs $\boldsymbol{I_O} \in \boldsymbol{\mathcal{I}}$ is scheduled in the uplink using the subchannels indexed by $\boldsymbol{\mathcal{J}_O} = \{\mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{J_O}\}$ while the remaining $\mathbf{2}\boldsymbol{I_N} = \mathbf{I} - \boldsymbol{I_O}$ MUs are allocated in pairs from the subset $\boldsymbol{\mathcal{J}_\mathcal{N}} = \{\boldsymbol{J_{O+1}}, \ldots, \boldsymbol{J_N}\}$. In each tier, a local computing server, also referred to as a "cloudlet," is directly connected to a single-antenna base station (BS) (see Fig. 1). It is generally assumed that the cloudlets associated with MBSs possess higher computational capacity than those connected to SBSs [XIX, XX]. Let $\boldsymbol{F^k}$ presents the computation capacity of the cloudlets, measured in CPU cycles per second, for the BSs in tier k, we then have $\boldsymbol{F^M} \geq \boldsymbol{F^S}$. Each mobile user $\boldsymbol{i}$ aims to execute a computational task within a specified allowable time $\boldsymbol{T_i^{max}}$. The task to be offloaded is described by the number $\boldsymbol{V_i^k}$ of the required number of CPU cycles for completion, by the number $\boldsymbol{B_i^{k,I}}$ of input bits, and by the number $\boldsymbol{B_i^{k,O}}$ of output bits representing the outcome of the remote execution. The MU can offload its computations to the $\boldsymbol{k}\mathbf{BS}$ with $\boldsymbol{k} \in \{\boldsymbol{M}, \boldsymbol{S}\}$ in the same cell. Then, each $\mathbf{kBS}$ can either execute the computation task on behalf of the MU or offload it to the cloud as long as the latency constraint is satisfied. The offloading latency is comprised of three components: $\boldsymbol{T_{i_j}^{k,ul}}$, which is the time required by the MU to upload the input bits to its base station; $\boldsymbol{\Delta_{i_j}^k}$, the time required for the edge servers to perform the instructions, and finally $\boldsymbol{T_{i_j}^{k,dl}}$ the necessary time that is taken to transmit the outcome bits back to the user device in the downlink direction. Accordingly, the total offloading latency $\boldsymbol{L_{i_j}^k}$ experienced by each MU $\boldsymbol{i}$ reads

$$L_{i_j}^k = T_{i_j}^{k,ul} + \Delta_{ij}^k + T_{i_j}^{k,dl} \tag{1}$$

Next, we derive the energy and latency associated with the offloading decision of all MUs. The energy $E_{i_j}^{k,ul}$ for each MU i depends only on the power utilized for uplink transmission. These latency and energy terms are computed as a function of the radio and computational resources in the following.

1) ***Uplink time*:** The achievable rate, in bits/s, for sending the input bits of user i connected to a BS in tier k via orthogonal subchannel j in the uplink is given by:

$$R_{i_j}^{k,ul}\left(P_{i_j}^{k,ul}\right) = W^{ul}\, log_2\left(1 + \frac{P_{i_j}^{k,ul} h_{i_j}^k}{\sigma^2}\right), \tag{2}$$

where $P_{i_j}^{k,ul}$ is the transmit power of the mobile device of user i connected to tier k in the uplink ; W^{ul} is the uplink channel bandwidth; $h_{i_j}^{k}$ is the uplink channel power gains of user i over subchannel j; and σ^2 is the receiver noise power. Note that we assume the interference cancellation technique is deployed by the receiver such that the formula in (2) holds for both OMA users and also users in NOMA with $h_{i_j} \geq \bar{h}_{i_j}$ where $\bar{h}_{i_j}$ denotes the channel gain of the interferer in NOMA pair. However, if $h_{i_j} < \bar{h}_{i_j}$, then the uplink rate formula reads

$$R_{i_j}^{k,ul}\left(P_{i_j}^{k,ul}\right) = W^{dl}\log_2\left(1+\frac{P_{i_j}^{k,ul}h_{i_j}^{k}}{\bar{P}_{i_j}^{k,ul}\bar{h}_{i_j}^{k}+|\sigma^2|}\right) \tag{3}$$

The time needed for user i to transmit $B_i^{k,I}$bits is $T_{i_j}^{k,ul}\left(P_{i_j}^{k,ul}\right) = B_i^{k,I}/R_{i_j}^{k,ul}\left(P_{i_j}^{k,ul}\right)$. The corresponding mobile energy consumption due to uplink transmission is

$$E_{i_j}^{k,ul}\left(P_{i_j}^{k,ul}\right) = B_i^{k,I}.\frac{P_{i_j}^{k,ul}}{R_{i_j}^{k,ul}\left(P_{i_j}^{k,ul}\right)} \tag{4}$$

Note that in (2) and all subsequent equations, the parameter between the parentheses denotes the variable under optimization.

2) ***Processing time*:** Let the capacity of the cloudlet server attached to each BS in both tiers be denoted as F^k. Also, let $f_{i_j}^{k} \geq 0$ be the fractions, to be optimized, of the processing power F^k, assigned to user i in tier k via subchannel j, so that $\sum_{i=1}^{I} f_{i_j}^{k} \leq 1$.The cloudlet execution time for V_i^k CPU cycles are

$$\Delta_{i_j}^{k} = \frac{V_i^k}{f_{i_j}^{k}F^k} \tag{5}$$

3) ***Downlink time***: Similar to uplink, the B_i^Ooutput bits intended for NOMA user i are sent in the downlink with rate

$$R_{i_j}^{k,dl}\left(P_{i_j}^{k,dl}\right) = W^{dl}\log_2\left(1+\frac{P_{i_j}^{k,dl}g_{i_j}^{k}}{\bar{P}_{ij}^{k,ul}\bar{g}_{ij}^{k}+}\right) \tag{6}$$

Where $g_{i_j}^{k}$ represents the downlink channel gain, and for $g_{i_j}^{k} \geq \bar{g}_{ij}^{k}$ with $P_{i_j}^{k,dl}$ being the BS transmit power allocated to serve user i.The downlink transmission time to transmit $B_i^{k,O}$ bits can hence be computed as

$$T_{i_j}^{k,dl}\left(P_{i_j}^{k,dl}\right) = \frac{B_i^{k,O}}{R_{i_j}^{k,dl}\left(P_{i_j}^{k,dl}\right)} \tag{7}$$

Please note that for the conventional OMA pair, the downlink rate is readily available using the standard Shannon-Hartley formula in (2).

III. Problem Formulation

The optimal offloading problem aims to minimize the total energy consumed by all mobile devices to offload their applications to the MEC server, involving individual latency and power constraints. This problem can be expressed mathematically as:

$$\underset{\mathbf{P}^{ul},\mathbf{P}^{dl},\mathbf{f},\boldsymbol{a}}{\mathbf{minimize}} \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} a_{i_j}^{k} E_{i_j}^{k,ul}\left(P_{i_j}^{k,ul}\right)$$

subject to

C.1 $a_{i_j}^{k}\left(T_{i_j}^{k,ul}+\Delta_{i_j}^{k}+T_{i_j}^{k,dl}\right)\leq L^{max}, \forall\, i\in\,\mathcal{I},$

C.2 $f_{i_j}^{k}\geq 0, \sum_{i\in\mathcal{I}} f_{i_j}^{k}\leq 1, \forall\, k\in\,\mathcal{K},$

C.3 $P_{i_j}^{k,ul}\leq\, P_{max}^{ul}\ , \forall\, i\in\,\mathcal{I},$

C.4 $\sum_{i\in\mathcal{I}} P_{i_j}^{k,dl}\ \leq\, P_{max}^{dl},\ k\in\,\mathcal{K},$

C. 5 $\sum_{i\in\mathcal{I}} a_{ij}^{k}=2,\ \forall\, j\in\,\mathcal{J}_{\mathcal{N}}, \sum_{i\in\mathcal{I}} a_{ij}^{k}=\quad 1,\ \forall\, j\in\,\mathcal{J}_{\mathcal{O}},$

C. 6 $\sum_{i\in\mathcal{I}} a_{ij}^{k}=1,\ \forall\, i\in\,\mathcal{I}, a_{ij}^{k}\ \in\{0,1\},$

Where a_{ij}^{k} is a binary variable governing the subchannel allocation among MUs, that is $a_{ij}^{k}=1$ indicates that the $j-$th subchannel is dedicated to $i-$th MU for uplink radio operation. In that regard, three distinct cases were considered:

1) **H-NOMA**: Each subchannel supports exactly two users, modelled as:

$$\sum_{i=1}^{I}\sum_{k=1}^{K} a_{i_j}^{k}=2,\quad \forall j \tag{8}$$

2) **NOMA:** No restriction on the number of users per subchannel, allowing more flexible allocation.
3) **OMA**: Each subchannel is exclusively allocated to a single user:

$$\sum_{\mathrm{i}=1}^{\mathrm{I}}\sum_{\mathrm{k}=1}^{\mathrm{K}} \mathrm{a}_{\mathrm{i_j}}^{\mathrm{k}}=1,\quad \forall \mathrm{j} \tag{9}$$

Constraint C.1 imposes that the offloading time for each MU i be less than or equal to the application deadline of L_{max} seconds (otherwise the offloading is infeasible); C.2 imposes the practical limit on the MEC computational resources in each tier k; Constraints C.3 and C.4 guarantee that the power budget constraint on the uplink and downlink radio resources are satisfied, respectively; C.5 and C.6 are the natural bounds on the binary subchannel allocation variable. Problem (P.1) is non-convex because of the non-convex nature of both the objective function and the constraint (C.1). Consequently, in the following section, we investigate an efficient algorithm that aims at obtaining an effective suboptimal solution.

IV. Solution via Relaxation and Decomposition

This section presents the algorithmic solution for the optimal offloading problem in MEC-enabled heterogeneous networks (HetNets) using a hybrid NOMA approach. Recall that the Problem (P.1) is NP-hard due to the non-convexity of the objective function and the constraint (C.1). Also, the existence of the binary allocation variables a_{ij}^k makes the problem not mathematically tractable. Therefore, the optimization problem is approached with a relaxation and decomposition method. We first relax the binary allocation variable a_{ij}^k to continuous values and then apply a heuristic rounding technique to approximate an integer solution.

The procedures outlined in Algorithm 1 start with Step 1, which calculates the power allocation to maximize uplink and downlink rates for each MU, ensuring power limits are not exceeded. In step 2, the processing latency is computed based on the MEC server capacity and CPU cycles required for each task. Next, subchannel assignment is done in Step 3, where we assign subchannels to MUs via the hybrid NOMA scheme, ensuring binary assignment using a heuristic. Finally, in Step 4, the offloading latency is checked to ensure that the total latency for each MU remains within the allowable maximum latency L_{max}.Table I summarizes the key simulation parameters used in the study.

Table 1: Simulation Parameters

Parameter	Value
Number of BS tiers (K)	2 (Macro and Small cells)
Number of subchannels (J)	5
Uplink bandwidth (W_{ul})	1 MHz
Downlink bandwidth (W_{dl})	1 MHz
Noise power (σ^2)	1×10^{-9} W
Maximum uplink power $(P_{ul,max})$	0.2 W
Maximum downlink power ($P_{dl,max}$)	0.5 W
Maximum allowable latency (L_{max})	0.05 s
Processing capacities of Macro and Small BSs (F)	5 GHz, 1 GHz

Algorithm 1 Efficient Offloading Policy in Hybrid NOMA MEC-enabled HetNets

---

Input: Channel gains $h_{i_j}^k$, $g_{i_j}^k$ and system parameters from Table 1.

Output: Optimal power allocations $\mathbf{P}^{ul}, \mathbf{P}^{dl}$, processing fraction $\mathbf{f}$, and subchannel assignments $\boldsymbol{a}$

Initialization: Relax $a_{ij}^k \in \{0,1\}$ to be in $\in [0,1]$. Initialize power and processing fractions to feasible values.

Repeat

  Step 1: Update Power Allocation

  for each MU i and BS tier k **do**

    Solve for $P_{i_j}^{k,ul}$ and $P_{i_j}^{k,dl}$ that minimize transmission latency using (2) subject to **C.3**

  end for

  Step 2: Compute Processing Latency

  for each BS tier k **do**

    Calculate processing time $\Delta_{i_j}^k$ using (5) subject to **C.2**

  end for

  Step 3: Update Subchannel Assignment

  for each subchannel j **do**

    Allocate subchannels based on hybrid NOMA

    $a^k = \{1,$ if MU i assigned to subchannel j

  end for

  Step 4: Check Latency Constraints

  for each MU i **do**

    Calculate total offloading latency $L_{i_j}^k$ using (1) and check **C.1**

  end for

until convergence criteria are met

V. Numerical Results

This section presents the simulation results obtained from evaluating the energy consumption and latency under different NOMA scenarios: Hybrid NOMA, Full NOMA, and No NOMA. The Hybrid NOMA scenario limits subchannel assignment to two users per subchannel, while the Full NOMA scenario removes this restriction, allowing more than two users to share a subchannel. Conversely, the No NOMA scenario assigns each user a unique subchannel, thus preventing any resource sharing among users. The performance metrics examined include energy consumption and average latency, with variations analyzed across different user counts.

To simulate realistic wireless environments, the evaluation adopts a multipath Rayleigh fading channel model, which accurately reflects signal fluctuations in mobile and urban deployments with non-line-of-sight conditions. This model is extensively used for performance evaluation in both MEC systems and NOMA-based wireless networks [XXL, XXIL]. Furthermore, the simulation parameters such as transmission power, subchannel bandwidth, channel gain, noise power, and MEC server computation capacities, and task input size are based on 3GPP TR 36.814 [XXIIL] and validated with experimental setups from recent studies on 5G heterogeneous networks [XXIV, XXV]. These configurations ensure the practical relevance of the results by accounting for non-deterministic user mobility, varying wireless conditions, and stochastic task arrivals.

Figures 2 and 3 demonstrate the comparative performance in terms of energy consumption and average latency across varying numbers of users in the three NOMA scenarios. These results highlight the practical advantages of the Hybrid NOMA scheme in balancing spectral efficiency and system cost under realistic propagation conditions.

- **Energy Consumption**: As shown in Figure 2, the energy consumption increases with the number of users for all scenarios. The OMA scenario exhibits the highest energy consumption due to the increased interference among users sharing the same subchannels, while the Hybrid NOMA scenario achieves the lowest energy consumption by eliminating subchannel interference. The NOMA scenario provides a balanced trade-off. Numerically, the proposed Hybrid NOMA schemes exhibit the lowest energy consumption with a 25% and 50% decrease compared to Full NOMA and No NOMA schemes, respectively.
- Figure 3 indicates that average latency increases with the number of users, especially in the Full NOMA scenario, due to the high level of resource sharing. No NOMA experiences the lowest latency by isolating each user on a unique subchannel. This is seen in the reduction of approximately 29% and 46% when compared to Hybrid NOMA and No NOMA, respectively. It is also evident that Hybrid NOMA achieves a compromise between resource sharing and latency. Spectral efficiency (SE) is a measure of how efficiently the available bandwidth is utilized for communication. It is defined as:

$$SE = \frac{\text{Total Throughput (bps)}}{\text{Total Bandwidth (Hz)}} \tag{10}$$

Where the total throughput includes both uplink and downlink rates, and the total bandwidth is the sum of the uplink and downlink bandwidths.

As seen in Figure 4, Hybrid NOMA improves spectral efficiency by dynamically combining power domain multiplexing (as in Full NOMA) and interference avoidance (as in No NOMA). Specifically, compared to No NOMA, Hybrid NOMA enables multiple users to share subchannels, leading to a higher spectral efficiency through better spectrum utilization. Conversely, compared to Full NOMA, Hybrid NOMA minimizes excessive interference by limiting multiplexing to users with significant channel disparities. This ensures that the achievable rates for each user are not severely degraded by interference. Hybrid NOMA is particularly beneficial in scenarios with diverse channel conditions or moderate user density, as it balances interference mitigation with bandwidth efficiency.

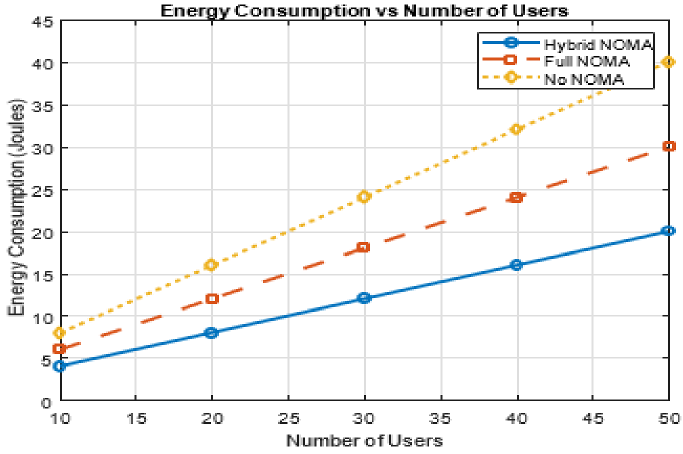

Fig. 2. Energy Consumption vs. Number of Users for Standard NOMA, Full NOMA, and No NOMA

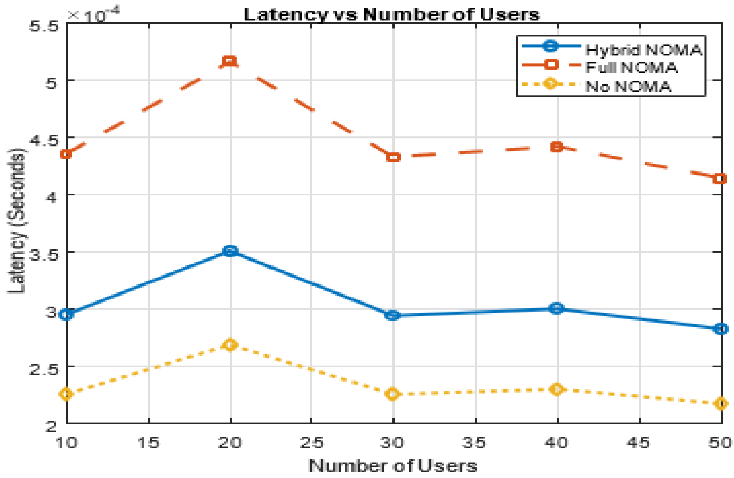

Fig. 3. Latency vs. Number of Users for Standard NOMA, Full NOMA, and No NOMA

Hybrid NOMA exhibits lower latency than Full NOMA while achieving comparable latency to No NOMA, as depicted in Figure 5. This is achieved through its adaptive resource allocation: By reducing interference, Hybrid NOMA decreases the decoding delay commonly associated with Full NOMA since Hybrid NOMA reduces interference levels, leading to faster decoding times. Its dynamic subchannel allocation avoids the strict and often inefficient resource division seen in No

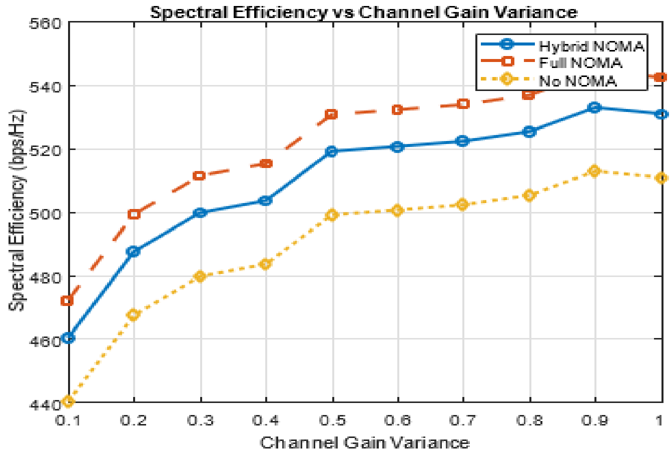

Fig. 4. Spectral Efficiency vs. Channel Gain Variance for H-NOMA, NOMA, and OMA.

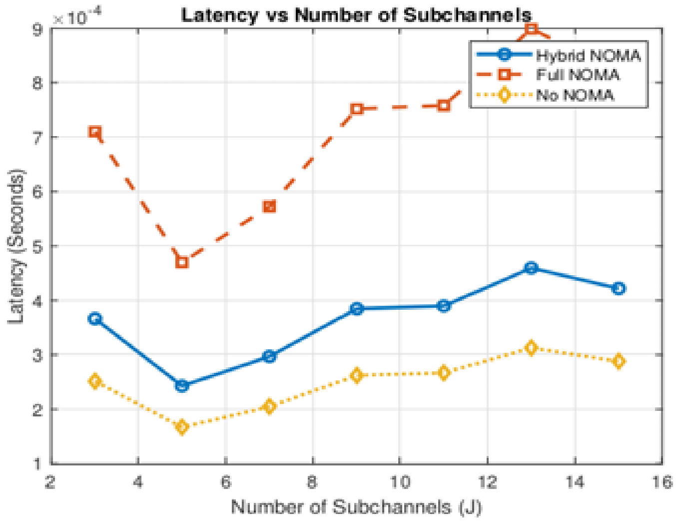

Fig. 5. Latency vs. Number of Users for Hybrid NOMA, Full NOMA, and No NOMA.

Table 2. Performance Comparison of Hybrid NOMA

Metric	Hybrid NOMA vs. Full NOMA	Hybrid NOMA vs. No NOMA
Spectral Efficiency	Comparable or slightly lower	Higher

Latency	Lower	Comparable or slightly higher
Energy Consumption	Comparable or slightly lower	Lower

NOMA, thereby maintaining low latency. The flexible use of shared and dedicated subchannels ensures that Hybrid NOMA strikes a balance between resource utilization and latency minimization.

Hybrid NOMA avoids the excessive interference seen in Full NOMA and thus achieves lower energy consumption by reducing interference and minimizing the need for high transmission power and retransmissions. Unlike NO NOMA, Hybrid NOMA avoids the inefficiencies of rigid subchannel allocation, ensuring better power utilization.

Table 2 summarizes the performance comparison of Hybrid NOMA versus Full NOMA and NO NOMA in the context of spectral efficiency, latency, and energy consumption.

VI. Conclusions

Spectral efficiency, latency, as well as energy consumption can optimally be balanced by Hybrid NOMA, which offers higher spectral efficiency compared to No NOMA while achieving lower latency and energy consumption than Full NOMA. This makes Hybrid NOMA a promising candidate for next-generation wireless networks; its advantages become particularly significant in scenarios where its inherent strengths, such as its flexibility in resource allocation and ability to reduce interference, become evident, as in heterogeneous channel conditions and diverse QoS requirements.

Although the proposed solution provides promising performance in terms of bandwidth usage, computational delay, and power efficiency, there remain important areas for future exploration. One critical direction involves the integration of machine learning (ML) and reinforcement learning techniques to enable adaptive, intelligent resource management under dynamic network conditions.

Additionally, future work should explore the impact of user mobility, channel variability, and traffic fluctuation on system performance. It is also essential to consider security and privacy-preserving mechanisms in offloading processes, particularly when dealing with sensitive user data in distributed edge environments.

Finally, the suggested framework can be extended to accommodate heterogeneous edge architectures and support emerging technologies such as 6G, UAV-assisted MEC, and intelligent reflecting surfaces (IRS).

Conflict of Interest:

There was no relevant conflict of interest regarding this article.

References

I. Ding, J., Han, L., Li, J., & Zhang, D. (2023). Resource allocation strategy for blockchain-enabled NOMA-based MEC networks. Journal of Cloud Computing, 12(1), 142.

II. Song, Z., Liu, Y., & Sun, X. (2020). Joint task offloading and resource allocation for NOMA-enabled multi-access mobile edge computing. IEEE Transactions on Communications, 69(3), 1548-1564.

III. Pham, H. G. T., Pham, Q. V., Pham, A. T., & Nguyen, C. T. (2020). Joint task offloading and resource management in NOMA-based MEC systems: A swarm intelligence approach. IEEE Access, 8, 190463-190474.

IV. Lin, L., Liu, J., Zhang, D., & Xie, Y. (2019). Joint offloading decision and resource allocation for multiuser NOMA-MEC systems. IEEE Access, 7, 181100-181116.

V. Wu, Y., Ni, K., Zhang, C., Qian, L. P., & Tsang, D. H. (2018). NOMA-assisted multi-access mobile edge computing: A joint optimization of computation offloading and time allocation. IEEE Transactions on Vehicular Technology, 67(12), 12244-12258.

VI. Liu, Y., Peng, M., Shou, G., Chen, Y., & Chen, S. (2020). Toward edge intelligence: Multiaccess edge computing for 5G and Internet of Things. IEEE Internet of Things Journal, 7(8), 6722-6747.

VII. Zheng, G., Xu, C., Long, H., & Zhao, X. (2021, March). MEC in NOMA-HetNets: A joint task offloading and resource allocation approach. In 2021, IEEE Wireless Communications and Networking Conference (WCNC) (pp. 1-6). IEEE.

VIII. Zhao, J., Liu, Y., Chai, K. K., Nallanathan, A., Chen, Y., & Han, Z. (2017). Spectrum allocation and power control for non-orthogonal multiple access in HetNets. IEEE Transactions on Wireless Communications, 16(9), 5825-5837.

IX. Fang, F., Cheng, J., & Ding, Z. (2018). Joint energy efficient subchannel and power optimization for a downlink NOMA heterogeneous network. IEEE Transactions on Vehicular Technology, 68(2), 1351-1364.

X. Fang, F., Wang, K., Ding, Z., & Leung, V. (2020). Energy-efficient resource allocation for NOMA-enabled MEC networks with imperfect CSI. arXiv preprint arXiv:2009.06234.

XI. Song, Z., Liu, Y., & Sun, X. (2018). Joint radio and computational resource allocation for NOMA-based mobile edge computing in heterogeneous networks. IEEE Communications Letters, 22(12), 2559-2562.

XII. Wu, L., Sun, P., Chen, H., Zuo, Y., Zhou, Y., & Yang, Y. (2023). NOMA-Enabled Multiuser Offloading in Multicell Edge Computing Networks: A Coalition Game-Based Approach. IEEE Transactions on Network Science and Engineering.

XIII. Xu, C., Zheng, G., & Zhao, X. (2020). Energy-minimization task offloading and resource allocation for mobile edge computing in NOMA heterogeneous networks. IEEE Transactions on Vehicular Technology, 69(12), 16001-16016.

XIV. Li, K., Xu, J., Xing, H., Chen, Y., & Huang, J. (2022). Dynamic Task Offloading for NOMA-Enabled Mobile Edge Computing with Heterogeneous Networks. Security and Communication Networks, 2022(1), 7258236.

XV. Du, J., Sun, Y., Zhang, N., Xiong, Z., Sun, A., & Ding, Z. (2022). Cost-effective task offloading in NOMA-enabled vehicular mobile edge computing. IEEE Systems Journal, 17(1), 928-939.

XVI. Saleem, M., Jangsher, S., & Khaliq, H. (2025), "Cooperative Offloading Multi-Access Edge Computing (Comec) for Cell-Edge users in Heterogeneous Dense Networks ", Available at SSRN 4968395.

XVII. M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies (2009, Oct.), "The case for VM-based cloudlets in mobile computing," IEEE Pervasive Comput., vol. 8, no. 4, pp. 14–23.

XVIII. A. Al-Shuwaili, M. Dh Hassib, T. M. Jamel (2022, Aug.), "Ball-Based Adaptive Channel Estimation Scheme Using RLS Family-Types Algorithms," in *International Journal of Computing and Digital Systems*, vol. 12, no. 1, pp. 469-477.

XIX. Agarwal, Bharat, et al. (2022) "A comprehensive survey on radio resource management in 5G HetNets: Current solutions, future trends and open issues." *IEEE Communications Surveys & Tutorials* 24.4: 2495-2534.

XX. T. M Jamel, A. Al-Shuwaili, and B. M Mansoor (2020), "Novelty Study of the Window Length Effects on the Adaptive Beam-forming Based-FEDS Approach," *International Journal of Computing and Digital Systems*, vol. 9, no. 6, pp. 1221–1227.

XXI. Van Truong, Truong, and Anand Nayyar (2023), "System performance and optimization in NOMA mobile edge computing surveillance network using GA and PSO." Computer Networks 223: 109575

XXII. Van Truong, Truong, and Anand Nayyar (2024), "Enhancing security offloading performance in NOMA heterogeneous MEC networks using access point selection and meta-heuristic algorithm." Computer Networks 254: 110802.

XXIII. 3GPP TR 36.814 V9.0.0 (2010, Mar), "Further advancements for E-UTRA physical layer aspects (Release 9).

XXIV. Mao, Yuyi, et al. "A survey on mobile edge computing (2017), " The communication perspective." IEEE communications surveys & tutorials 19.4:2322-2358.

XXV. Sinop, Hayriye, Mehmet Berke Beyaz, and Asuman Savaşçıhabeş (2024), "Performance Analysis of NOMA Systems over Rayleigh Fading Channels in 5G Communication." Orclever Proceedings of Research and Development 5.1: 238-252.