

Intelligent Resource Optimization in Integrated Aerial Terrestrial Networks



Muhammad Awais

School of Computing and Communications

Lancaster University

A thesis submitted for the degree of

Doctor of Philosophy

February, 2025

Copywrite © 2024

by Muhammad Awais

This thesis is dedicated to my loving parents.

Declaration

I declare that the work presented in this thesis is, to the best of my knowledge and belief, original and my work. The material has not been submitted, either in whole or in part, for a degree at this or any other university. This thesis does not exceed the maximum permitted word length of 80,000 words, including appendices and footnotes, but excluding the bibliography.

Muhammad Awais

Intelligent Resource Optimization in Integrated Aerial Terrestrial Networks

Muhammad Awais

Submitted for the degree of Doctor of Philosophy

February, 2025.

Abstract

Developing sixth generation (6G) wireless communication necessitates low-power consumption, high-reliability, and massive-connectivity. One of the most promising solutions to address these requirements is aerial base station (ABS) based communication systems that employ both in-the-air (aerial) and on-the-ground (terrestrial) components. ABSs enhance line of sight (LoS) connections, fulfilling escalating quality of service (QoS) demands. Nevertheless, integrating aerial and terrestrial networks into future three dimensional (3D) networks introduces emerging requirements for resource allocation and new functional challenges, such as latency, reliability, energy consumption, and QoS. Motivated by the above observations, this thesis investigates the challenges of intelligent resource optimization in integrated aerial terrestrial networks.

An integrated aerial and terrestrial network is initially examined to design a bisection-based low-complexity adaptation (BLCA) algorithm for optimal resource allocation. A joint optimization problem that involves sub-carrier (SC) assignment, blocklength, and power allocation (PA) subject to delay, reliability, and QoS constraints is investigated to enhance system performance in a finite blocklength (FBL) regime. The proposed solution includes sub-carrier allocation based on matching theory, optimal blocklength allocation using the bisection algorithm, and a two-step projected gradient descent power distribution by optimizing the power budget on each sub-carrier. Case studies on flexible blocklength and PA are also examined under the next generation of multiple access techniques.

The second part integrates digital twin (DT) technology with mobile edge computing (MEC) to facilitate mobile offloading in an integrated aerial-terrestrial network. An advanced bisection sampling-based stochastic solution enhancement (BSSE) algorithm is specifically tailored to jointly optimize transmit power, central processing unit (CPU) frequency, and the task offloading policy to minimize the system's energy-time cost against benchmarks. The proposed solution includes a one-climb policy to narrow the search space, a closed-form solution for calculating the optimal CPU frequency and transmit power for given offloading decisions, and an inequality condition formulated to manage dependent tasks efficiently. The scalability of the proposed scheme is also analyzed.

In the final part, machine learning techniques are adopted to improve the system performance in an integrated aerial-terrestrial wireless network. The proposed solution employs unsupervised learning techniques for the grouping of internet of things smart devices (ISDs), Q-learning (a type of reinforcement learning) for the intelligent ABS placement, and deep learning for power allocation. A closed-form expression is also derived for PA among multiplexed devices based on their QoS requirements. Numerical results indicate that the proposed scheme significantly outperforms existing benchmark schemes.

This thesis presents valuable insights into innovative, sustainable, and energy-efficient resource optimization in integrated aerial-terrestrial future-generation networks, setting the stage for further advancements in resource allocation to enhance reliability, QoS, and latency.

Acknowledgement

First and foremost, I am deeply grateful to my supervisors, Dr. Haris Pervaiz, Dr. Wenjuan Yu, and Prof. Qiang Ni, for their unwavering support, guidance, and inspiration throughout my PhD journey. Their expertise and encouragement have been invaluable to my research and personal development. I also extend my heartfelt thanks to my colleagues and friends at Lancaster University for their camaraderie, insights, and collaboration. Your support has made this experience enriching and pleasant. A special note of gratitude goes to my colleague, Naif Alshammari, for consistently motivating me to strive for excellence and being a steadfast source of support through good and challenging times.

Beyond my academic circle, I would like to express my deepest thanks to my parents, Muhammad Ashraf and Mukhtran Bibi, and my brothers, Ali Raza and Zain Ali, whose unconditional love and encouragement have been a constant source of strength. Their unwavering belief in me has been a driving force throughout my life. I owe my partner, Aqsa Shabir, immense gratitude for her patience, understanding, and unwavering support. Your love and encouragement have been my anchor during the most challenging times of this journey.

Finally, I express my gratitude to all those who have supported me, directly or indirectly, in completing this thesis. Your contributions have made this achievement possible and I will be forever grateful.

List of Publications

- **M. Awais**, H. Pervaiz, W. Yu, Q. Ni, “Intelligent Resource Allocation for URLLC in Green Aerial Terrestrial Wireless Networks”, in *IEEE Transactions on Green Communications and Networking*, 2025 (submitted).
- **M. Awais**, H. Pervaiz, Q. Ni, W. Yu, “Task Dependency Aware Optimal Resource Allocation for URLLC Edge Network: A Digital Twin Approach Using Finite Blocklength”, in *IEEE Transactions on Green Communications and Networking*, pp. 1–1, 2024.
- **M. Awais**, H. Pervaiz, M. A. Jamshed, W. Yu, Q. Ni, “Energy-Aware Resource Optimization for Improved URLLC in Multi-hop Integrated Aerial Terrestrial Networks”, pp. 548–553, in *IEEE Transactions on Green Communications and Networking*, vol. 8, no. 1, pp. 252–264, 2023.
- **M. Awais**, H. Pervaiz, M. A. Jamshed, W. Yu, Q. Ni, “Enhancing URLLC in Integrated Aerial Terrestrial Networks: Design Insights and Performance Trade-offs”, in *23rd IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pp. 548–553, 2022.
- N. Sharma, **M. Awais**, H. Pervaiz, H. Malik, Q. Ni, “Network-assisted Unmanned Aerial Vehicle Communication for Smart Monitoring of Lock-down”, in *Autonomous Airborne Wireless Networks*, pp. 195–216, Wiley 2021.

Contents

Copyright	ii
Dedication	iii
Declaration	iv
Acknowledgements	vii
Publications	viii
List of Tables	xiii
List of Figures	xiv
List of Algorithms	xvii
List of Acronyms	xviii
1 Introduction	1
1.1 Motivations	2
1.2 Contributions and Structure of the Thesis	7
1.3 Structure of the Thesis	12
2 Theoretical Background and Literature Review	14
2.1 Summary	26

3	Energy Aware Optimal Resource Allocation Using Greedy Approaches	28
3.1	Introduction	29
3.2	System Model	30
3.2.1	Architecture of the Aerial Terrestrial Network	30
3.2.2	Decoding Error Probability for the NOMA Phase	36
3.3	Problem Formulation	38
3.4	Proposed Solution	41
3.4.1	Sub-carrier Allocation and Selection of Cooperative Aerial Base Stations within the Coverage of Macro Base Station	41
3.4.1.1	Sub-carrier Allocation using Stable Matching	41
3.4.1.2	Selection of the Best Cooperative Aerial Base Stations	43
3.4.2	Sub-carrier Allocation and Selection of IoT Devices within the Coverage of Serving Aerial Base Station	43
3.4.3	Joint Blocklength and Power Optimization	44
3.4.4	Bisection-based Low-Complexity Adaption Algorithm	48
3.5	Results and Discussion	50
3.5.1	Simulations Setup	50
3.5.2	Performance Comparison	52
3.6	Summary	59
4	Task Dependency Aware Optimal Resource Allocation Using One-climb Policy	61
4.1	Introduction	62
4.2	System Model	63
4.2.1	Architecture of Digital Twin Edge Network	63
4.2.2	Architecture of Mobile Edge Computing based on Ultra-reliable low-latency communication	64
4.2.3	Task Offloading Model in Edge Network	65
4.2.4	Digital Twin Model	66

4.2.5	Communication Model	67
4.2.6	Computational Model	68
4.2.6.1	Local Computing	68
4.2.6.2	Edge Computing	68
4.2.7	Latency Model	69
4.2.8	Energy Consumption Model	69
4.2.9	Task Dependency Use Cases	70
4.3	Problem Formulation	70
4.4	Proposed Solution	74
4.4.1	Optimal Transmit Power and Local CPU Frequency for given φ	75
4.4.2	Bisection Method for given φ to Obtain Optimal Power \mathbf{p} and Frequency \mathbf{f}	79
4.4.3	Optimize Offloading Decision φ using given \mathbf{p} and \mathbf{f}	82
4.4.4	Complexity Analysis	82
4.5	Multi-Device Scenario	83
4.6	Simulation and Results	84
4.6.1	Simulations Parameters	84
4.6.2	Performance Comparison	86
4.7	Summary	93
5	Connectivity Aware Optimal Resource Allocation Using Unsuper-	
	vised Learning	94
5.1	Introduction	95
5.2	System Model	96
5.2.1	Multi-ABS Aerial Terrestrial Network Architecture	97
5.2.2	Channel Model	97
5.2.2.1	Ground-to-Ground Channel Model	98
5.2.2.2	Ground-to-Air Channel Model	98
5.2.2.3	Air-to-Ground Channel Model	99
5.2.3	NOMA-based Transmission Model	99

5.2.3.1	NOMA-based Transmission	100
5.2.4	URLLC-based Downlink Transmission Rate	101
5.3	Problem Formulation	102
5.4	Proposed Solution	104
5.4.1	Unsupervised Learning-based Internet of Things Smart Device Grouping	104
5.4.2	Reinforcement Learning-based Aerial Base Station Placement	107
5.4.2.1	Agent	109
5.4.2.2	States ($s \in \mathcal{S}$)	109
5.4.2.3	Actions ($\psi \in \Theta$)	110
5.4.2.4	Observations	110
5.4.2.5	Rewards (r)	110
5.4.2.6	Environment	110
5.4.3	Power Allocation	112
5.4.3.1	Power Allocation among Sub-carriers	112
5.4.3.2	Power Allocation between ISDs	113
5.4.4	Complexity Analysis	117
5.5	Simulation and Results	117
5.5.1	Simulations Settings	117
5.5.2	Performance Comparison	118
5.6	Summary	126
6	Conclusions and Future Work	127
6.1	Conclusion	128
6.2	Limitations of the Study	129
6.3	Future Improvements	130

List of Tables

3.1	Simulation Parameters.	51
4.1	Simulation Parameters.	85
5.1	Simulation Parameters.	118

List of Figures

1.1	Estimation of Global Traffic From 2020 to 2030 (source: [3], Fig. 9). . .	2
1.2	Past, Present, and Future of Mobile Networks (source: [5], Fig. 1). . .	3
1.3	Two Devices-based OMA Transmission in the Downlink.	4
1.4	Two Devices-based NOMA Transmission in the Downlink.	4
1.5	ABS-aided Future-Generation Mobile Networks.	5
3.1	Considered Multi-hop Scenario: $\mathcal{U}_h = \{x, y\}$, where $(x, y) \in \mathcal{U}_h$, $\mathcal{U}_u = \{p, q\}$, where $(p, q) \in \mathcal{U}_u$, $ \mathcal{I} $ IoT Devices are Grouped into \mathcal{L} Communities Located within the Coverage of Serving ABSs, and $\mathcal{M} = \{a, b\}$, where $(a, b) \in \mathcal{M}$ Located within the Coverage of MBS. .	31
3.2	Systematic Diagram depicting Blocklength and DEPs for the Consid- ered Scenario Setting.	37
3.3	Sum-rate versus M_{\max} with Optimal Blocklength and Power Alloca- tion, where $b_1 = k \times b_2$, $b_2 = 3$, $M_{\max} = k \times 10$, and $k = \{3, 4, \dots, 9\}$. .	52
3.4	Sum-rate versus M_{\max} using Matching-based Sub-carrier Allocation and Optimal Blocklength Allocation with different Power Allocation Strategies, where $b_1 = k \times b_2$, $b_2 = 3$, $M_{\max} = k \times 10$, and $k =$ $\{1, 2, \dots, 10\}$	53
3.5	Sum-rate versus different Number of IoT Devices with Matching- based Sub-carrier Allocation against LDDP [19], where $M_{\max} = 100$. .	54

3.6	Complexity of the BLCA Scheme and LDDP [19] with Stable Matching-based Sub-carrier Allocation, Considering the basic Number of Operations, where $M_{\max} = 100$	55
3.7	Sum-rate versus the ratio of Blocklengths with Stable Matching based Sub-carrier Allocation and Optimal Power Allocation, where $b_1 = k \times b_2$, $b_2 = 3$, and $M_{\max} = 100$	56
3.8	Sum-rate versus Altitudes using Stable Matching based Sub-carrier Allocation and Optimal Power Allocation, where $b_1 = 13$, $b_2 = 13$, $z_{u_u} = 50$, $z_{u_h} = \{50, 100, \dots, 250\}$ and $M_{\max} = 100$	57
3.9	Spectral Efficiency versus Circuit Power using Stable Matching based Sub-carrier Allocation and Optimal Power Allocation, where $b_1 = 13$, $b_2 = 13$, and $M_{\max} = 100$	58
4.1	DT-enabled Integrated Aerial-Terrestrial Edge Network denoting a Two-device Framework with an Intermediate Task Node m_t	63
4.2	Task Dependent Computational Model Incorporating Outputs from Multiple Devices.	83
4.3	The Trade-off between the Total Execution Time and Energy for each IoT Device when $\partial_{k_2}^\lambda$ varies, where $d_{k_1} = d_{k_2} = 15$ Meters.	87
4.4	Impact of Intermediate Task t on the Computational Delay, when $(L, M, t) = (3, 5, t)$	88
4.5	Energy-time Cost of Benchmark and Proposed Algorithms, when $\mathcal{A} = 2$, $\mathcal{K} = 2$ and $d_{k_1} = d_{k_2} = 15$ Meters.	89
4.6	Overall Energy-time Cost versus Value of d_1 , when $d_2 = 10$ Meters.	90
4.7	Energy-time Cost for Increasing value of (L , M , t) in Three Algorithms.	91
4.8	Proposed Task Dependency Model for a Multi-device Case Study with $k = 4$	92
5.1	Scenario of NOMA-based Multi-UAV Aerial Terrestrial Network.	96

5.2	Growth Rate.	98
5.3	A Framework using Dual Criteria for Selecting the Optimal Number of Clusters.	105
5.4	Illustration of 2-ISD NOMA Grouping for 12 Active ISDs in a Region.	107
5.5	Deep Neural Network Infrastructure For Proposed Work.	113
5.6	Convergence of Normalized Reward for UAVs using Q-learning based Intelligent Positioning over Iterations.	119
5.7	Sum-rate of NOMA-DeepFusion-PA [Optimal UAV position] Scheme under different Power Levels and Power Allocation Coefficients ζ_1 for both UAVs with Perfect SIC.	120
5.8	Performance Comparison of NOMA-DeepFusion-PA [Optimal UAV position] Scheme under Perfect and Imperfect SIC Conditions with Residual of Coefficient ν for both UAVs.	121
5.9	System Sum-rate under different Power Allocation and UAV Posi- tioning Strategies with Perfect SIC for OMA and NOMA Systems. . .	122
5.10	Evaluation of Blocklength Impact on Sum-rate for NOMA-DeepFusion- PA with Optimal UAV Positioning Compared to Benchmark Schemes on both UAVs.	123
5.11	Sum-rate of NOMA-DeepFusion-PA, NOMA-FTPA, and OMA Schemes with Optimal UAV Positioning across different Horizontal Distances.	124

List of Algorithms

3.1	Computing Optimal Power on $[0, \bar{p}^k]$	44
3.2	Power Distribution on given IoT Devices	48
3.3	Proposed BLCA Algorithm	49
4.1	Bisection Method to Compute Optimal Power and Local Frequency	79
4.2	Proposed BSSE Algorithm	81
5.1	Machine learning-based ISD Grouping/Sub-carrier Allocation	108
5.2	Mapping Function $\mathcal{F}(y)$	109
5.3	Q-Learning based ABS Placement in a Virtual Space	111
5.4	Power Allocation and Sum-rate Computation	115
5.5	Proposed Algorithmic Solution for Intelligent Resource Allocation	116

List of Abbreviations

ABS	Aerial base station.
AWGN	Additive white Gaussian noise.
B	Bandwidth.
BLCA	Bisection-based low-complexity adaption.
BSSE	Bisection sampling-based stochastic solution enhancement.
CU	Cellular users.
CPU	Central processing unit.
CP	Circuit power.
CSI	Channel state information.
DEP	Decoding error probability.
DT	Digital twins.
DL	Deep learning.
DITEN	Digital twin edge network.
DNN	Deep neural network.
ES	Edge server.
EC	Energy consumption.
eMBB	Enhanced mobile broadband.
5G	Fifth-generation.
FBL	Finite blocklength.
FPA	Fixed power allocation.
FTPA	Fractional transmit power allocation.
FPTAS	Fully polynomial time approximation scheme.

GPR	Gaussian process regression.
GRAD-JASPA	Gradient descent-dependent heuristic.
IoT	Internet of things.
ISD	Internet of things smart device.
IoV	Internet of vehicles.
IRS	Intelligent reflecting surfaces.
JSPA	Joint sub-carrier and power allocation.
KKT	Karush kuhn tucker.
LDDP	Lagrangian duality and dynamic programming.
LoS	Line of sight.
L	Lagrangian.
ML	Machine learning.
MBS	Macro base station.
MINLP	Mixed integer nonlinear programming.
MEC	Mobile edge computing.
MC-NOMA	Multicarrier non-orthogonal multiple access.
mMTC	Massive machine- type communication.
MIMO	Massive multiple input multiple outputs.
MF-SIC	Match filtering successive interference cancellation.
ME-MEC	Multi-access mobile edge computing.
NDT	NOMA-based on-demand transmission.
NOMA	Non-orthogonal multiple access.
N-LoS	Non-line of sight.
OFDMA	Orthogonal frequency division multiple access.
OMA	Orthogonal multiple access.
PA	Power allocation.
PD-NOMA	Power-domain NOMA.
PEC	Pervasive edge computing.
OPT-JSPA	Pseudo-polynomial time optimal algorithm.

QoS	Quality of service.
Q-learning	Quality-learning.
RL	Reinforcement learning.
6G	Sixth-generation.
SIC	Successive interference cancellation.
SC	Sub-carrier.
SCPC	Single carrier power control.
SCUS	Single carrier user selection.
SGD	Stochastic gradient descent.
SNR	Signal to noise ratio.
SINR	Signal to interference plus noise ratio.
3D	Three-dimensional.
UAV	Unmanned aerial vehicles.
URLLC	Ultra-reliable low-latency communication.
WSR	Weighted sum-rate.

Chapter 1

Introduction

1.1 Motivations

The emergence of advanced wireless infrastructure has transformed the way information is generated, disseminated, received, and perceived [1]. The capacity is expected to increase by up to 1000 times to support the growing number of wireless users and internet of things (IoTs) devices [2] (see Fig. 1.1).

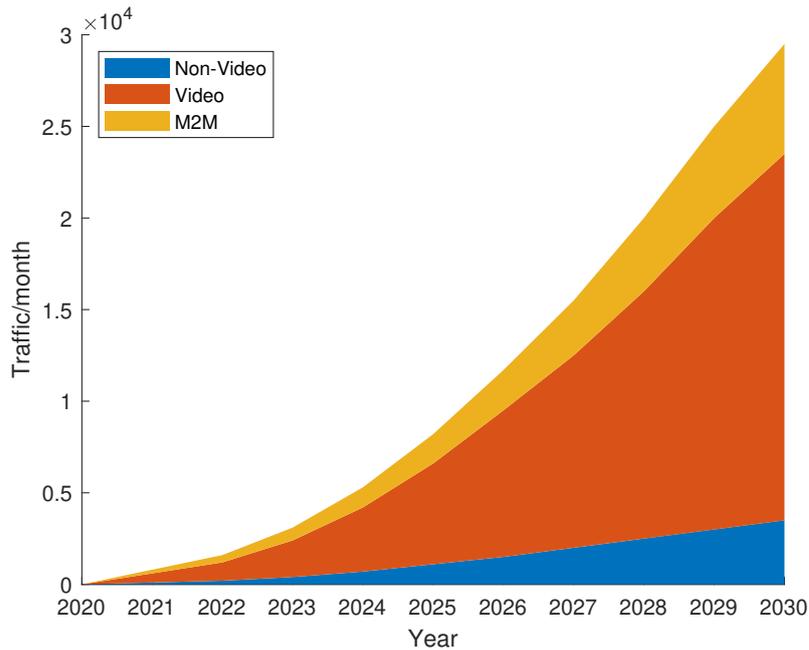


Figure 1.1: Estimation of Global Traffic From 2020 to 2030 (source: [3], Fig. 9).

Therefore, a few novel communication paradigms are needed to address three key connectivity types that align with the new technical requirements, i.e., enhanced mobile broadband (eMBB) to provide high throughput to demanding clients, massive machine-type communication (mMTC) to support low-cost, low-power IoT devices, and ultra-reliable low-latency communication (URLLC) to support mission-critical IoT devices, such as the tactile internet and autonomous vehicles, which require stringent quality of service (QoS) requirements to achieve a delay of less than one millisecond and reliability more significant than 99.9999%.

The rapid evolution of IoTs has also facilitated the cost-effective connection of billions of wireless users, introducing a new era of connectivity. However, the

limited battery life and computational power of IoT devices have emerged as a significant barrier, particularly in supporting computationally intensive applications within future generation networks [4]. These limitations are primarily attributed to concerns about production costs and stringent size restrictions. Therefore, advances in various technologies are required to meet the increasing demand for wireless users and the continuously evolving requirements for system capacity, energy consumption, and massive connectivity (see Fig. 1.2).

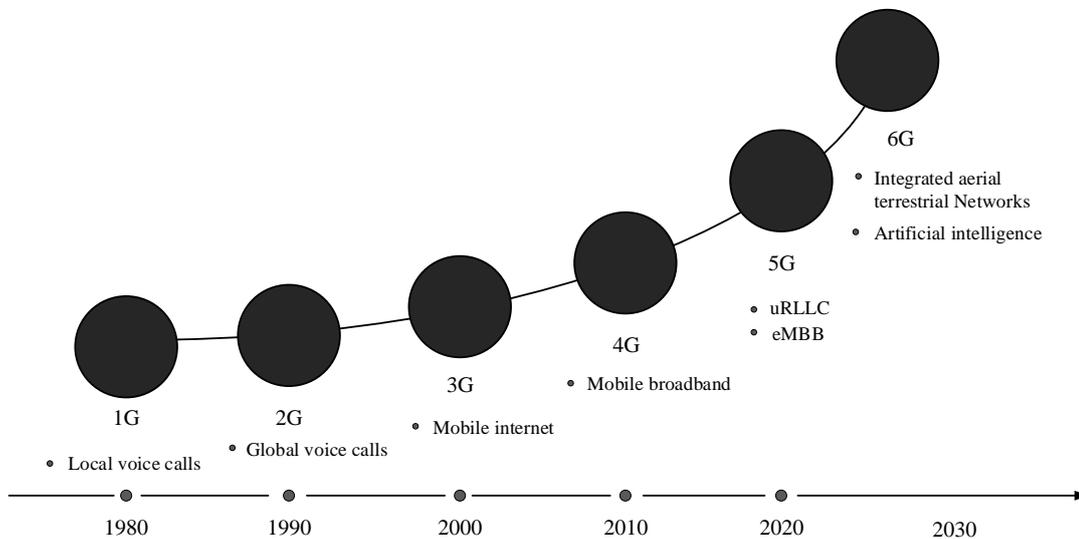


Figure 1.2: Past, Present, and Future of Mobile Networks (source: [5], Fig. 1).

Given the scarcity of radio spectrum, it is crucial to explore how it can be utilized most efficiently [6]. It is well-known that the radio channel is broadcast in nature, which can lead to inevitable interference between different users' transmissions. Therefore, there is a strong need for multiple access schemes to coordinate these transmissions. The existing literature demonstrates that multiple access schemes play a significant role in the design of next-generation cellular networks. Non-orthogonal multiple access (NOMA) is a promising technology in cellular networks and beyond to address the problem of scarcity of shared spectrum resources [7].

In contrast to orthogonal multiple access (OMA), NOMA with finite blocklength (FBL) follows the exact mechanism as typical NOMA, particularly superposition

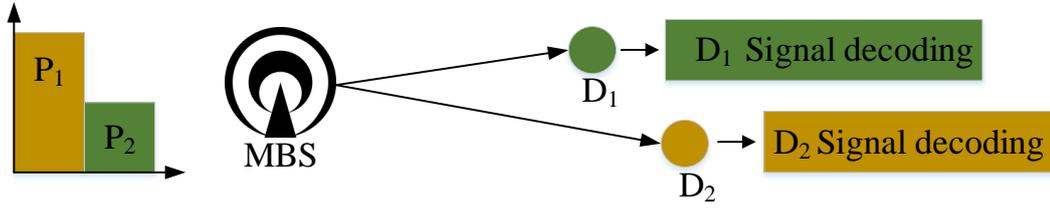


Figure 1.3: Two Devices-based OMA Transmission in the Downlink.

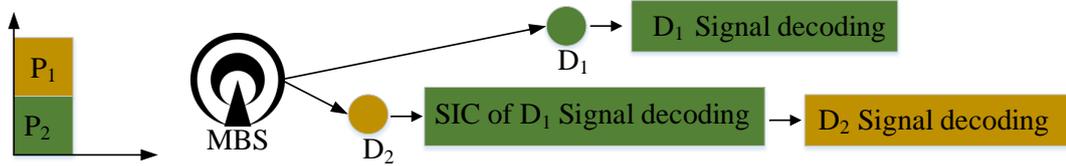


Figure 1.4: Two Devices-based NOMA Transmission in the Downlink.

coding at the transmitter and successive interference cancellation (SIC) at the receiver [8]. It permits devices to share the same time-frequency resources, enhancing the network's energy and spectral efficiency. However, the traditional Shannon formula does not approximate the maximum achievable rate when considering short-packet communication. This thesis exploits the intrinsic attributes of power domain NOMA to accommodate maximum IoT devices without compromising the sum rate, referred to as NOMA for simplicity (see Figs. 1.3 and 1.4).

The mode of communication is as critical as the associated technologies. For example, terrestrial infrastructure is commonly at risk of being destroyed by natural disasters. During 2019, at least one type of disaster affected approximately 27% of the world's road and railway systems [9]. Therefore, timely warnings and relief operations can actively control these damages, which require effective management. One promising solution is to use emerging future-generation communication and control systems. Future generation networks envision seamless connectivity of both ground and aerial users. However, practical implementation of future-generation

networks faces significant challenges, including high data rates, massive connectivity, and high energy demands.

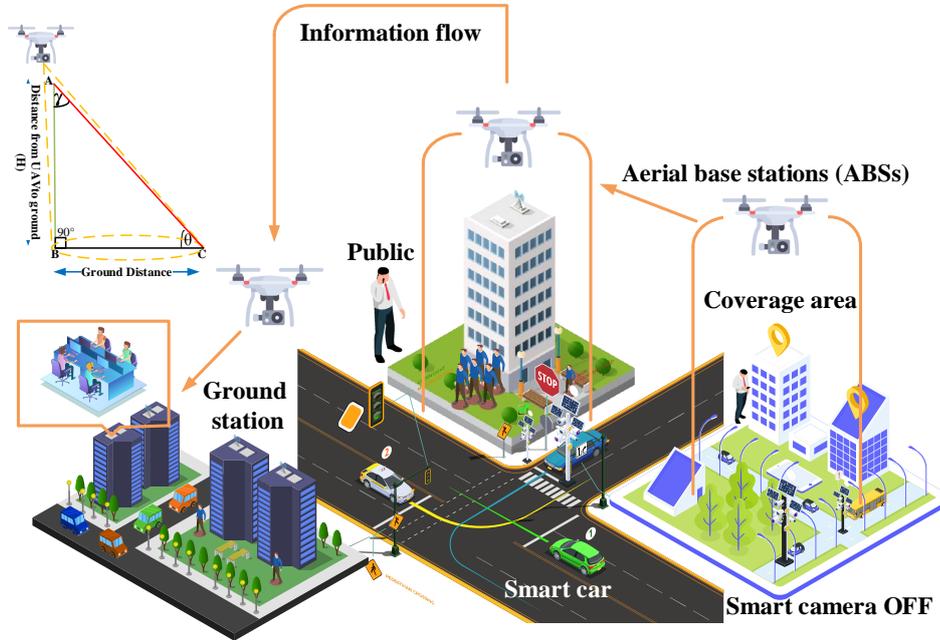


Figure 1.5: ABS-aided Future-Generation Mobile Networks.

Aerial base stations (ABSs) have recently emerged as highly adaptable airborne wireless technology due to ease of mobility and to ensure massive connectivity with minimal human intervention (see Fig. 1.5). They provide high data rates and a wide range of services, including monitoring the Internet of agricultural things, surveillance during natural disasters, and data offloading in different hotspots [10]. Furthermore, ABSs can also act as mobile base stations and extend their services to support growing traffic without geographical limitations. These features make ABS an excellent possible solution in hot-spot areas.

To meet the growing demand of an ever-increasing number of mobile users, efficient use of radio spectrum and provision of massive connectivity with high-reliability, low-latency, and other QoS requirements are essential. However, the subsequent challenge lies in the computing capability of IoT devices, which have small on-chip computing units with low-performance [11]. These drawbacks

fundamentally hinder the support of computation-based applications in future-generation networks. To this end, mobile edge computing (MEC) came as an emerging and cost-effective paradigm that leverages computational and storage capabilities to support wireless devices with limited resources.

Task offloading is the most significant feature of MEC, which enables resource-constrained IoT devices to offload their computation-intensive tasks to high-performance edge servers, either binary or partially. Each task is processed locally or offloaded to the edge server in binary offloading. In contrast, each task is partitioned and executed locally and on the edge server [12]. This thesis focuses primarily on binary offloading within the context of FBL, which is frequently used in IoT systems to process tasks that cannot be partitioned. This approach is instrumental in satisfying the increasing quality of service demands in edge networks, particularly in the context of resource allocation [13]. Furthermore, ABS-assisted communication can be effectively combined with mobile edge computing (ABS-MEC) to enable promising solutions to overcome challenges, such as offering support for time-sensitive applications and computation-intensive services, encompassing a broad spectrum of IoT applications, i.e., from security to actuation and monitoring IoT systems.

Using the new capabilities of future-generation networks under the strict requirements of URLLC links, another innovative technology that has recently gained substantial prominence is a digital twin (DT) [14]. This technology creates a virtual replica of physical systems, enabling the simulation of optimal solutions prior to real-time implementation. Its application significantly improves system performance while minimizing downtime [15]. The integration of URLLC with other emerging technologies, including MEC and DT, is also crucial in accommodating a wide range of mobile applications, covering Industry 4.0, future smart cities, and intelligent communication systems [16]. However, its practical implementation raises intriguing and challenging concerns due to the complex relationship between reliability and end-to-end delay in next-generation wireless networks [17].

The major challenges in future-generation networks powered by intelligent edge include resource allocation, adaptive learning, intelligent decision-making, privacy, low latency in terms of communication and computation, and limited computing power. Machine learning is a key enabler to perform better parameter estimation and efficient decision-making. Therefore, there is a strong need to design sub-optimal algorithms with performance guarantees (i.e., low-complexity solutions) by utilizing virtual platforms to simulate high-computing, machine learning-based training and to facilitate active interaction between the physical and virtual worlds. These techniques significantly impact algorithmic solutions, providing optimal radio resource management in URLLC edge networks to improve the performance and reliability of communication systems [18]. This thesis advocates a novel set of guidelines for future-generation networks.

This thesis aims to design optimal resource allocation policies for integrated aerial-terrestrial edge networks, investigating the correlation and interdependency of tasks executed by different IoT devices. In addition, various unsupervised techniques based on machine learning are employed to enhance system performance. The research addresses new and technical challenges in future-generation wireless networks, including energy efficiency, reliability, latency, and other QoS requirements. In particular, this thesis uses users, IoT devices, and ISDs interchangeably. In addition, unmanned aerial vehicles (UAVs) and ABSs are used interchangeably.

1.2 Contributions and Structure of the Thesis

This thesis provides optimal resource allocation solutions through innovative, sustainable, and energy-efficient resource scheduling in integrated non-terrestrial network systems for 6G. Our approach aims to enhance QoS and latency aspects, offering significant improvements over existing state-of-the-art methods, particularly: 1) by using a traditional heuristic approach, a low-complexity, connectivity-aware optimal resource allocation policy is developed for integrated aerial-terrestrial

networks; 2) task dependencies among IoT devices in the context of DT-aided edge computing for URLLC are examined to optimize task offloading and resource allocation; and 3) machine learning-based strategies are adopted for optimal resource allocation and intelligent UAV positioning to support scattered communities and enhance network invulnerability cooperatively.

In Chapter 2, the underlying literature related to system design for FBL regimes is examined. Initially, a background knowledge of UAV-assisted communication, including OMA and NOMA, is introduced. Subsequently, the concept of DTs in conjunction with mobile edge computing is explained. The chapter then discusses existing studies on intelligent UAV positioning and optimal resource allocation, providing readers with a comprehensive understanding of the research background.

In Chapter 3, a low-complexity, connectivity-aware optimal resource allocation policy is developed using traditional heuristic approaches to enhance network performance. The strong coupling between optimization variables poses a challenge, mainly when grouping IoT devices in multi-carrier transmission. Consequently, we decided to investigate the development of a resource allocation strategy specifically for multi-carrier communication. It involves a fast network formation where integrated aerial terrestrial networks¹ can be used to ensure the connectivity of the IoT device in a signal dead zone. The primary contributions of this part of the thesis are summarized below.

- A problem in an integrated aerial terrestrial network that involves mixed-integer non-linear programming is investigated. To address this issue, the problem of maximizing the sum-rate is reframed by utilizing its decomposition property and jointly optimizing the channel allocation, power allocation (PA), and blocklength allocation for both OMA and NOMA systems.

¹It refers to the integration and convergence of communication systems that employ both in the air (aerial) on the ground (terrestrial). This concept helps to establish a unified network infrastructure that inherits the capabilities of aerial platforms such as ABSs and airborne systems with conventional terrestrial communication systems.

- An alternating optimization method is later utilized to present an iterative bisection-based low-complexity adaption (BLCA) algorithm to optimize the resource allocation problem subject to delay, reliability, and QoS constraints. The problem formulated is solved in three steps. Firstly, the matching theory is employed to allocate sub-carriers and select the best cooperative ABSs. Secondly, the optimal blocklength is computed using the bisection algorithm. Finally, dynamic and geometric programming is applied to perform power distribution by optimizing the power budget on each sub-carrier with a two-tier projected gradient descent-based algorithm.
- In the end, the optimality of the BLCA algorithm is evaluated against a high complexity benchmark scheme, namely Lagrangian duality and dynamic programming (LDDP), which employs Lagrangian dual to relax the individual power constraint [19]. Monte Carlo simulations are conducted to compare the performance of the proposed algorithmic solution against the LDDP scheme. Moreover, two benchmark algorithms are also implemented for comparative analysis, where the worst sub-carrier is prevented from being assigned to a transmitting node [20, 21]. The proposed algorithmic solution is also analyzed against random and fixed blocklength approaches using legacy OMA and NOMA with different PA schemes. The results show that the proposed algorithmic solution significantly achieves a near-optimal solution and outperforms the LDDP scheme.

In Chapter 4, the interdependence of tasks executed by different IoT devices is analyzed [22]- [23]. This interdependence significantly influences decisions related to the offloading and allocation of resources in an integrated aerial-terrestrial edge network [24–26]. For example, a device might need to offload its tasks to an ABS-assisted edge server that another device urgently needs. Due to this strong coupling and the combinatorial nature of the problem, identifying an optimal solution is challenging. This chapter presents a pioneering examination of the dependency of tasks between devices within the context of DT-aided edge computing for URLLC

to optimize task offloading and resource allocation. The key contributions of this part are summarized as follows:

- A mixed-integer non-linear programming (MINLP) problem is formalized within a DT-enabled integrated aerial-terrestrial network, novelly considering task interdependencies. Due to its inherently combinatorial link with task-offloading decisions and strong correlation with resource allocation, this problem poses significant computational challenges. Therefore, an enhanced bisection sampling-based stochastic solution enhancement (BSSE) algorithm is introduced to minimize the system's energy-time cost iteratively, offering a solution that closely matches the performance of the most effective existing scheme.
- To efficiently narrow the search space, a 'one-climb policy' is opted, where a device offloads its data to the edge server at the optimum time only once. The proposed algorithm is designed to jointly optimize transmit power, central processing unit (CPU) frequency, and task offload policy, thus minimizing the weighted sum of energy consumption and execution time of the devices. Later, a closed-form solution is derived to calculate the optimal CPU frequency and transmit power for given offloading decisions. Then, an inequality condition is formulated to manage dependent tasks efficiently. The proposed algorithm commences with a random task offloading configuration and iteratively updates it to reduce the system's energy-time cost.
- The scalability of the proposed model is analyzed by varying the number of IoT devices with the sequential number of tasks, i.e., from a simplified two-device framework to multiple devices, incorporating different intermediate tasks. The trade-off between the system's energy-time cost is also analyzed to validate the effectiveness of our approach.
- Although the proposed algorithm can manage a diverse range of tasks, the computational complexity of our proposed approach is significantly lower than

the benchmark schemes. The proposed BSSE approach is also compared with three sophisticated benchmark schemes: the bisection algorithm [22], the Gibbs sampling algorithm based on a one-climb policy [22], and the exhaustive search algorithm. A comparative analysis with the bisection algorithm reveals that our approach reduces the total energy-time cost by 15.35% to 33.12% when the weighting parameter $\partial_{k_2}^\lambda$ increases from 0.1 to 0.3, respectively.

In Chapter 5, it is observed that the existing literature on UAV communication either focuses on stationary UAVs hovering above hotspot centers or optimizes UAV trajectories without considering reliability, delay, and QoS constraints to maximize system throughput [27]. Neglecting kinematic constraints, intelligent UAV positioning, and static propulsion are critical factors in maintaining user connectivity during malicious attacks [28]. This chapter introduces a novel NOMA-DeepFusion-PA approach for optimal resource allocation and UAV deployment to maximize the achievable rate. Consequently, a collaborative system assisted by drones is designed to support scattered communities and cooperatively enhance network invulnerability. The key features include machine learning-based intelligent UAV placement and optimal resource allocation to reduce co-channel interference with improved system performance. The major contributions of our work are:

- A connectivity-aware aerial-terrestrial edge network is considered, where a UAV hovers at the optimal position to provide on-demand services. Later, an optimization problem is formulated by employing unsupervised learning-based ISD grouping, reinforcement learning-based intelligent UAV placement, and DeepFusion-based PA that integrates the deep-learning based PA in conjunction with fractional transmit PA (FTPA) to provide an additional layer of optimization to ensure power efficiency and fairness among sub-carriers to improve overall system performance.
- Due to the inherent non-convexity, the problem is decoupled into sub-problems and solved iteratively to obtain optimal solutions. In contrast to [29]- [30],

a data-driven approach is proposed that integrates the elbow method with the F-test method to determine the optimal number of clusters for a given channel condition, thus avoiding arbitrary selections. Afterward, the k-means clustering algorithm is utilized to group devices into clusters. Unlike in [29], the proposed data-driven approach offers superior control over the number of devices per sub-carrier and more efficient bandwidth utilization. A mathematical definition is also derived to calculate the value of the F-test for a given number of clusters.

- To address the ample search space, a Q-learning based strategy is utilized to maximize the utility function by adjusting the UAV position relative to the location of ISDs. The goal is to extend the coverage area for each UAV while maintaining fairness among the ISDs. Subsequently, a successive convex approximation-based iterative algorithm is proposed to address non-convexity in the rate function within the ensuing sub-problem.
- The effectiveness of the proposed scheme is validated through numerical results. Simulation results demonstrate a significant improvement in the system's achievable rate for different transmit powers, that is, the sum rate for NOMA-DeepFusion-PA [Optimal UAV position] can be increased by 28.5762% than NOMA with fixed PA method, namely: NOMA-FPA [Optimal UAV position] [31], and 38.3119% higher than legacy OMA [32].

1.3 Structure of the Thesis

The remainder of the thesis is organized as follows. Chapter 1 outlines the motivations for the study. Chapter 2 presents the background knowledge, including a brief review of the literature. Chapter 3 provides a detailed discussion of energy-aware resource allocation using greedy approaches. Chapter 4 focuses on optimal resource allocation based on task dependency through the one-climb policy. Subsequently, Chapter 5 explores optimal resource allocation based on connectivity

employing unsupervised learning techniques. Finally, Chapter 6 concludes the thesis, addressing its limitations and proposing directions for future research.

Chapter 2

Theoretical Background and Literature Review

NOMA is an emerging paradigm in cellular networks and beyond applications to address the problem of scarcity in shared spectrum resources. It helps maintain high link quality ubiquitously and increases spectral efficiency by exploiting available resources more efficiently. The details are given below.

The performance of multi-carrier NOMA (MC-NOMA) systems mainly depends on two interacted factors, i.e., power control and sub-carrier allocation. The authors in [33] demonstrate that in real-time LTE cellular systems, MC-NOMA outperforms OFDMA in network-level performance, for example, in both wide-band and sub-band scheduling, including high and low mobility scenarios. Moreover, the authors proposed a predefined user grouping and fixed PA to minimize complexity. Different optimization methods are commonly used to achieve optimal PA. However, these methods restrict the algorithm's performance because they can assist at most two users on each sub-carrier.

Considering the low-latency requirements for future generation networks, a joint sub-carrier and PA algorithm is proposed to maximize the sum rate for the downlink MC-NOMA problem [34], consisting of a three-step resource allocation framework. The proposed algorithm operates in three phases: 1) the problem is relaxed by allowing each user to access all sub-carriers, 2) greedy sub-carrier allocation is performed based on the power vector from the previous step, and 3) to improve system performance, the power control scheme from the first step is re-implemented with the updated sub-carrier allocation from the second step. One centralized and two distributed PA methods are examined. The proposed algorithm performs almost as LDDP and outperforms standard power-controlled OMA schemes.

A distinctive technique to improve the weighted sum rate (WSR) is proposed in a power-constrained environment [35]. To this end, two basic building blocks are introduced: single carrier power control (SCPC) and single carrier user selection (SCUS). Additionally, based on these building blocks and a joint sub-carrier and PA (JSPA) scheme, a method is proposed that achieves near-optimal WSR and user fairness through a proper resource allocation policy. This work further exploits enhanced

versions of SCPC and SCUS, featuring low and practical computational complexity [36]. These pre-computations improve the performance of JSPA and are named gradient descent-dependent heuristics (GRAD-JASPA). Subsequently, a pseudo-polynomial-time approximation scheme (OPT-JSPA) and a fully polynomial-time approximation scheme (FPTAS) are proposed. The proposed schemes provide a theoretical performance guarantee with manageable computational complexity, although their performance lies within a specific threshold value.

The link layer rate is investigated for a two-user NOMA in an FBL regime, focusing on short packet communication [37]. This research uses the effective capacity framework to analyze the performance of the two-user NOMA network considering the overall reliability, including the probability of transmission error and the probability of violation of queue delay. The study is then extended to multi-user NOMA networks, noting that NOMA users with different channel state information (CSI) achieve better effective capacity. The impact of SNR, delay exponent, and transmission error probability on effective capacity is also analyzed. Simulations validate that a high delay exponent results in a latency violation probability that cannot be improved beyond a certain point due to the significant impact of transmission error probability.

Research in [38] focuses on downlinking NOMA systems using the FBL regime to enhance resource allocation. An efficient PA algorithm is proposed, and its performance is compared against a fixed PA algorithm. The results validate the efficacy of the proposed algorithm in terms of resource block usage and energy efficiency. The authors in [39] focus on factory automation scenarios to optimize blocklength and PA by minimizing the decoding error probability (DEP) while maintaining reliability, delay, and total energy constraints. Four downlink transmission schemes are considered: OMA, NOMA, relay-assisted, and cooperative NOMA. The study concludes that relay-assisted transmission significantly outperforms OMA, while NOMA is effective with limited blocklength. Moreover, relay-assisted transmission outperforms cooperative NOMA due to its

larger feasible region.

In industrial IoT networks, low latency and high reliability are crucial factors. There is a strong need to address the heterogeneous requirements of different applications operating on the same channel, such as transmission capacity, reliability, and energy efficiency. To address these issues within limited spectrum resources, a NOMA-based on-demand transmission (NDT) technique is proposed [40]. The proposed method achieves a higher sum data rate and transmission reliability by exploiting the relationship between the SIC decoding order and the application priority, thereby minimizing the complexity of the solution. It is done by isolating the compact coupling of channel allocation, power control, and NOMA clustering. The results validate the effectiveness of the proposed technique compared to existing methods.

Multi-access mobile edge computing (ME-MEC) is visualized as a critical approach to enable the delay-sensitive and computation-intensive tasks for the future generation of industrial IoT. Therefore, NOMA-assisted ME-MEC is exploited in static and dynamic channel scenarios for offloading computational tasks; e.g., an IoT device can unload its workload (specific part) to edge-computing servers via NOMA. At this end, a distributed procedure is proposed for optimal offloading solutions by jointly optimizing the problems, including computation resource allocation, multi-task computational offloading, and NOMA-dependent transmission by utilizing the minimum energy of the IoT device. In addition, a deep reinforcement learning-based online algorithm is proposed to meet the time-varying channel power gains between IoT devices and edge computing servers. Numerical results show the efficacy of the proposed work compared to conventional orthogonal multi-access-based offloading schemes.

In future-generation industrial IoT networks, intelligent transportation is a critical focus in the internet of vehicles (IoV). System delay and communication security are prominent concerns. The work in [41] proposes a two-layer NOMA-dependent pervasive edge computing (PEC) resource allocation framework to

address these issues and enhance the system's performance. The problem is formulated as a PA problem, considering imperfect CSI, queuing models, and vehicle speed. In addition, a Frank-Wolfe procedure is suggested to achieve optimal transmission power. Comparative analysis demonstrates the superiority of the proposed framework over conventional OFDMA counterparts.

In the leading study of [42], the potential of intelligent reflecting surfaces (IRS)-based UAV is investigated. It is tested in real-world settings, including constrained QoS requirements. However, the proposed method is unsuitable for need-based networks in disaster response and mission-critical applications (where the existing infrastructure may be severely damaged or unavailable). Hence, a small malfunction or failure in reflective elements can significantly impact overall network performance and reliability.

The author in [43] proposes a low-complexity algorithm to position the UAV and plan an efficient route for data collection, resulting in improved performance subject to delay. The researchers observed that the placement of ABSs is complex and requires attention. Therefore, the authors in [44] developed a framework that uses Markov chain and Gibbs sampling. The study is extended to use clustering to deploy ABSs and user association using NOMA. However, the number of covered users served by a single ABS is limited [45].

Considering these factors, it is better to rely on more traditional and reliable communication technologies to better align with the requirements and objectives of a need-based network. Motivated by the benefits of ABS and NOMA, their integration is analyzed and investigated. In [46], opportunistic channel gain disparities against each IoT device are identified, and the positions of ABSs are optimized, which becomes more challenging when power limitations constrain the problem. To maximize the minimum rate, [47] investigates aerial jamming and PA to improve security and reliability in ABS-assisted NOMA communication. Furthermore, in the latter work, a relay selection strategy is explored to optimize the PA of ABSs and to maximize energy efficiency under the NOMA scheme [48].

The emergence of short packets with FBL is a crucial enabler to support emerging technologies such as intelligent transportation systems and virtual reality [49]. Advanced wireless networks require reliable and efficient transmission, so studying communication systems in the FBL regime becomes crucial. However, the maximum achievable rate cannot be approximated using the Shannon capacity formula, which requires an alternative solution [50]. Previous work highlights inherent problems such as channel estimation errors and SIC and introduces additional challenges such as the spatial distribution of ABS, which poses security concerns, and flexible mobility, which complicates the channel [51]. Therefore, there is a strong need to explore the potential applications of integrated aerial-terrestrial communication in the FBL regime utilizing next-generation multiple access techniques ¹.

To meet the increasing requirements of URLLC, NOMA is being investigated in the FBL regime with reliability constraints [52]. Due to the benefits of FBL, efficient bandwidth allocation schemes that consider delay constraints have also been developed. However, using multiple hops adds complexity to resource allocation and decision-making processes. In [53], the authors optimize the amount of information transmitted from the control station of an ABS-aided system by simultaneously optimizing blocklength and transmit power. This work is extended to optimize ABS placement and transmission power to reduce DEP jointly [54].

The research is then expanded to an optimal resource allocation technique for heterogeneous communication links that use both OMA and NOMA [55]. However, it should be noted that the proposed approach is limited to throughput maximization and does not provide a closed-form expression. Hence, developing a low-complexity and connectivity-aware optimal resource allocation policy is crucial for enhancing network performance. However, the strong coupling between optimization variables poses a challenge, especially when grouping IoT devices in multi-carrier transmission.

¹It refers to innovative strategies that surpass conventional approaches. In the context of 5G and beyond, these techniques explore novel methodologies, i.e., NOMA to optimize resource allocation, resulting in improved URLLC characterized by higher data rates, reduced latency, and enhanced connectivity.

To address the high data traffic load, researchers perform a joint optimization of content placement, co-channel link, maintenance of QoS requirements, and spectrum allocation to enhance energy efficiency in aerial-terrestrial vehicular networks [56]. Innovative user clustering, joint power optimization, and placement of UAVs are applied to maximize energy efficiency in a full duplex NOMA system [57]. Joint optimization of UAV trajectory, beamforming, power, and time allocation to maximize throughput subject to minimum average user rate [58].

The authors in [59] strive to improve the system's minimum achievable rate by jointly optimizing the user association, transmit power, and UAV trajectory. Reference [60] classified the devices according to channel quality and strategic deployment of multiple UAVs with minimal energy usage. The study in [61] aims to optimize transmit power by scheduling ISDs, bandwidth allocation, power control, and UAV deployment while maintaining URLLC constraints. In [62], a UAV-assisted heterogeneous fisher-sidecar composite fading channel model is proposed to improve the capacity and energy efficiency of the system. The objective in [63] is to minimize congestion at the macro base station (MBS) by maintaining a stable transmission rate with minimum delay. In a study [64], a three-layer iterative algorithm is proposed to optimize the scheduling of UAV tasks, transmit power, and 3D flight parameters to improve performance. An innovative method to improve the efficacy of the URLLC system is proposed in [65], with the aim of maximizing the end-to-end achievable rate by jointly optimizing the UAV trajectory and blocklength allocation in a multi-user UAV relay system.

Traditional wireless networks are not suitable for providing sustainable communication links in critical situations such as disasters, wildfires, and emergencies. Therefore, the author in [9] proposes a connectivity-aware network aided by DTs to ensure reliable and real-time system communication. The author focuses on two objectives: maximizing the data rate for users within scattered communities and minimizing the transmitted power of the UAV. To address the limited resources of UAVs, the author utilizes the DT virtual space to train a reinforcement learning

model. The proposed model is evaluated on the basis of the physical and virtual trajectory and normalized reward. The results validate that the proposed solution provides a nearly optimal approach for the deployment of autonomous UAVs, reducing the energy-time cost of the network.

The concept of DTs is further explored in the context of offloading to UAV-aided edge servers. In [66], the authors address the problem of minimizing end-to-end delay using DT-aided offloading to MEC, supported by UAVs within URLLC links. To achieve this, the inner approximation is combined with alternative optimization, simultaneously optimizing the edge servers and IoT devices' offloading policies, transmit power, and processing rate. The performance of the proposed model is evaluated on the basis of worst-case end-to-end latency and total computational resources.

The authors in [67] extended the study of DTs to the metaverse and proposed a powerful computing platform employing MEC-based URLLC. They exploit the concept of task caching and task offloading to a nearby edge server, ensuring high-reliability and low-latency for future-generation networks within the metaverse. The performance of the proposed model is evaluated based on system reliability and latency by jointly optimizing the edge policies, the offloading factors, and the computing and communication environment. The results validate that the implemented scheme effectively enhances QoS using edge caching policies and bandwidth allocation mechanisms.

Machine learning and the fifth-generation (5G) paradigm open possibilities for extending industrial IoT. However, limited resources and immense data present significant hurdles, increasing transmission overhead and compromising privacy. To improve the quality of industrial IoT, a federated learning-based DT edge network (DITEN) is proposed in [68]. The proposed architecture integrates DTs with edge computing to optimize industrial IoT. Additionally, the scheme reduces transmission overhead and protects data privacy. It updates the model asynchronously, thereby reducing communication and computation costs. A deep neural network (DNN)

selects the best strategies for assigning optimal resources. The results demonstrate the efficacy of the implemented technique in terms of communication efficiency and energy cost.

Edge systems in industrial IoT devices face high latency issues, affecting URLLC systems' productivity [69]. Therefore, Dang et al. propose a mechanism to minimize system latency by jointly optimizing edge policies, offloading factors, and computing and communication variables. Therefore, an alternating optimization scheme with an inner convex approximation is utilized to develop an iterative algorithm that reduces system delay. Numerical results validate the efficacy of the proposed solution.

Considering the DT as a powerful computing platform, the problem of minimizing delay during computational offloading for industrial IoT is addressed in [24]. The proposed work optimizes resource allocation for the entire system by using the DT to model the computing capabilities of the edge server. The authors consider local processing time, transmission time, and edge processing time, subject to computing resources, to jointly optimize transmit power, processing rates of users and the edge server, offloading decisions, and user association. The current model also employs an alternating optimization approach with an inner approximation scheme to reduce latency compared to benchmarks.

Research in [70] states that UAV-based aerial communication significantly provides seamless connectivity to the IoVs. However, the IoVs system is vulnerable to various resource allocation challenges due to its dynamic nature. Therefore, the authors propose a two-stage, incentives-based mechanism for on-demand resource allocation using dynamic DTs, ultimately ensuring efficient resource allocation. The fundamental benefit of this scheme is its parallel processing on multiple roadside units, which minimizes computational burdens on the UAVs. The proposed model simultaneously increases fairness among the vehicles and improves the system's energy efficiency.

In [71], the authors highlight the significant role of edge collaboration in enhancing the performance of edge computing systems. Specifically, they integrate emerging DT technology with mobile edge servers to efficiently perform intelligent task offloading. Initially, the best MEC is selected by leveraging the capabilities of a DT-aided decentralized blockchain mechanism and CSI. Subsequently, the Markov decision process is adopted for task offloading to the selected edge server. The results validate that DT technology enables mobile users to efficiently minimize power-time costs instead of directly executing high-complexity machine learning algorithms.

Authors in reference [72] state that the sixth-generation (6G) system envisions wireless communication systems with high connectivity and computational efficiency. However, it also highlights the limitations of MEC, as the environment of edge systems is unpredictable and complicated due to user mobility. Later work presents a new version of DT technology functioning in two folds: the DT of the edge server estimates the edge server state. At the same time, the system's DT handles training data to make offloading decisions. Specifically, the migration cost of mobile users is minimized, resulting in reduced offloading latency. The efficacy of the proposed model is validated against performance metrics, including offloading latency, failure rate, and migration rate.

The authors of [73] extended the idea of users' mobility and the unpredictable MEC environment, proposing an intelligent task offloading scheme in UAV-enabled MEC with the support of DTs. They perform a joint optimization of UAV trajectory, user association, and resource allocation to efficiently reduce the energy consumption of the entire MEC system. Results validate the effectiveness of the proposed scheme compared to benchmarks.

It is commonly observed that many distributed machine learning models enhance model training by using parallel architectures. In these models, the gradient computation is distributed across sub-samples. However, the fundamental issues with this approach are data privacy and locality. Consequently, there is a strong need to preserve the features' data locality and confidentiality.

The authors in paper [74] have proposed the concept of ephemeral edge computing. The proposed technique leverages multiple neighboring nodes for task computation instead of relying on a single edge node. It performs device-to-device communication for computational task offloading rather than offloading tasks to the base station connected to the server. Additionally, the scheme makes on-the-fly decisions to elect one of the neighboring nodes to which the task should be allocated. The significant advantages of the scheme are the minimization of communication latency and the maximization of the number of computed tasks. Simulations and results demonstrate that the technique is nearly optimal compared to the offline solution.

Later research noticed that the training data is both large in quantity and privacy-sensitive in edge learning. Therefore, uploading data from various devices to an edge server for centralized model training raises data privacy concerns and incurs exorbitant communication and computation costs. These concerns prompted the development of an innovative edge-based learning framework called federated learning [75]. The authors use a distributed learning mechanism in this technique and update the model accordingly. Experiments show that the proposed approach reduces communication costs by 10-100 times compared to synchronized stochastic gradient descent. However, the above rate-driven classical techniques fail to accommodate actual learning tasks.

An asynchronous stochastic gradient descent (SGD) is proposed to learn from distributed attributes mutually [76]. A parameter server system primarily inspires this method. The main advantage of the proposed algorithm is that it avoids sharing the primary attributes. Comparative analysis shows that the proposed algorithm has enhanced application recommendations in the Tencent Application. However, the efficacy of the proposed SGD depends on user items and features from other apps. Managing the order sequence for each party's sample data remains challenging.

It is commonly observed that the performance of an intelligent transportation system mainly relies on the connection between vehicles with minimal delays and

errors. Therefore, it is essential to design an approach that can achieve high reliability and minimal delay. To this end, a federated learning-based scheme is proposed to jointly optimize power and resource allocation to meet these QoS requirements [77]. Results show that the proposed methodology outperforms the centralized solution, with a 70% reduction in data exchange and a 60% reduction in vehicular users with longer queue lengths.

An active learning-based approach is essential for designing information-aware, low-latency, and reliable networks to manage the dynamically changing nature of vehicular networks, such as interference and wireless channels [78]. Therefore, there is a need to actively learn the network dynamics to allocate power and resources efficiently. The proposed scheme achieves the desired results through an online decentralized strategy. The solution is mainly composed of the following steps: observe the recent information age, update the dataset, actively learn the scalability of the network using Gaussian process regression (GPR), and allocate resources by minimizing the objective. Numerical results validate the efficacy of the proposed technique compared to counterparts.

A federated learning-based scheme is proposed in [79]. This model operates in three stages: first, it locally trains the federated learning model; then, it applies the model globally; and finally, it transfers the model back to the users for updates. The main focus of this strategy is "efficient user selection" to minimize the federated learning loss function. It includes optimizing each user's transmit power and resource allocation, considering the impact of transmission errors and constrained bandwidth. Results show that the proposed algorithm successfully minimizes the federated learning loss function by 10-16% compared to benchmarks.

A multi-agent deep Q-learning network with a grouping mechanism is developed in [80] to support massive URLLC by minimizing delay in an FBL regime, formulating the problem subject to the adopted blocklength, transmission time interval, and bandwidth allocation. In reference [81], an unsupervised trajectory optimization is proposed by implementing a sequential deep reinforcement learning

model to minimize the system's energy consumption.

A lower-bound methodology for maximizing data rate using massive multiple input multiple outputs (MIMO) is proposed for uplink URLLC, as detailed in [82]. This optimization is achieved by jointly fine-tuning the payload and pilot transmission strategies for zero-fading and maximum ratio combining designs. Reference [83] addresses the challenge of resource optimization in URLLC, employing a game theoretic approach to enhance the offloading factor within a multi-agent edge network [84]. A distributed solution to optimize the average response time for computational offloading is presented in [85]. The study is extended to a distributed framework that solves the NP-hard energy efficiency problem, leveraging parallel processing [86].

Channel characteristics are also crucial in the next generation of wireless networks. It is not always recommended to completely offload computational tasks to the edge server because it may lead the network towards a low offloading data rate due to the possibility of deep fading [87]. An illustrative example is an adaptive search algorithm, which minimizes EC through joint optimization of offloading factors, resource allocation, and user association [88]. Similarly, the authors in [89] investigate resource allocation and task offloading from an economic perspective to enhance computing efficiency. The research in [90] focuses on developing optimal binary offloading policies for single-user tasks, later extended to multi-user scenarios in subsequent studies [91]. This concept is further applied to multiple independent tasks in the system where a single user offloads its task to different edge servers [92], and multiple users offload their tasks to a single edge server [93].

2.1 Summary

The literature highlights key findings and gaps in aerial terrestrial wireless communication systems. For example, unlike orthogonal schemes, NOMA is an emerging technology with significant potential to improve spectral efficiency and enable

massive connectivity in next-generation wireless networks. A key challenge in radio resource management for NOMA systems is the joint optimization of power control and subcarrier allocation. It is essential to carefully optimize the transmit power levels to manage intra-carrier interference caused by superposed signals and to maximize achievable data rates. In addition, the allocation of subcarriers for each transmission must also be optimized. Similarly, it is observed that the interdependence of tasks executed by different IoT devices greatly influences decisions related to offloading and resource allocation. Therefore, identifying an optimal solution is challenging, and researchers have yet to address these task dependencies between devices for optimal task offloading and resource allocation. Most existing studies on UAV communication focus primarily on stationary UAVs hovering above hotspot centers or optimizing UAV trajectories, often overlooking critical reliability, delay, and QoS constraints. It should be noted that most of the solutions in the literature are heuristic-based, lacking theoretical performance guarantees. In this context, sub-optimal algorithms with performance guarantees could significantly impact the design of radio resource management schemes. Hence, identifying such algorithms remains an open challenge.

This thesis systematically addresses these challenges by providing optimal solutions to each of the problems identified above. The following sections present detailed analyses and methodologies.

Chapter 3

^{1 2} Energy Aware Optimal Resource Allocation Using Greedy Approaches

¹The part of this chapter was presented in part at the IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Belfast, U.K., 2022, pp. 548–553 [DOI: 10.1109/WoWMoM54355.2022.00085].

²This complete chapter was also presented in IEEE Transactions on Green Communications and Networking 2023, Print ISSN: 2473-2400, Online ISSN: 2473-2400, Digital Object Identifier: 10.1109/TGCN.2023.3330018.

3.1 Introduction

Unlike conventional multimedia streaming services, URLLC³ services transmit control commands and sensing information in brief packets from source to destination. Delivering these packets quickly and reliably imposes heterogeneous requirements on both latency and reliability. In highly dynamic UAV-based communication, the swift information flow required by demand-based networks imposes strict requirements on these constraints above. Subsequent research suggests that a single URLLC link cannot address these challenges. Therefore, integrated aerial-terrestrial⁴ URLLC networks are needed to enable the trade-off flexibility required to improve reliability and shorter delay.

UAV-based communication is widely acknowledged for its role in supporting integrated aerial-terrestrial URLLC. UAVs offer superior LoS connection compared to conventional terrestrial communication systems, where even a minor failure can significantly impact overall network performance. Hence, relying on reliable communication technologies is beneficial for better alignment with the requirements and goals of demand-based networks. Alternatively, UAVs can operate as physical platforms for user devices, enabling them to access networks that require URLLC services. Traditional infrastructure does not adequately account for air-to-ground, ground-to-air, and air-to-air communications. These limitations result in greater delays, which presents major challenges to UAV communication.

The emergence of short packets with FBL is a key enabler to support emerging technologies such as intelligent transportation systems. Future-generation wireless networks require reliable and efficient transmission, so studying communication

³It refers to support applications that require extremely high reliability and low latency, where delays and failures can have severe consequences. It aims to provide reliability levels of 99.999% or higher and latency as low as 1 millisecond or less.

⁴It refers to the integration and convergence of communication systems that employ both in the air (aerial) on the ground (terrestrial). This concept helps to establish a unified network infrastructure that inherits the capabilities of aerial platforms such as ABSs and airborne systems with conventional terrestrial communication systems.

systems in the FBL regime becomes important. For URLLC, minimizing latency is crucial, so using smaller blocklengths can be beneficial to ensure minimum delay and high reliability. In this chapter, we decided to investigate the development of an optimal resource allocation strategy for URLLC systems, which involves fast network formation where integrated aerial terrestrial networks can be used to ensure IoT device connectivity in a signal dead zone. A low-complexity and connectivity-aware optimal resource allocation policy is designed to solve the mixed-integer non-linear programming problem. It jointly optimizes channel allocation, PA, and blocklength allocation for both OMA and NOMA systems. The proposed algorithm optimizes resource allocation that is subject to delay, reliability, and QoS constraints.

The rest of this chapter is organized as follows. Section 3.2 presents the case of interest and a mathematical framework for the proposed work. Section 3.3 formulates the problem, and Section 3.4 presents the proposed solution. Section 3.5 provides the simulation results with in-depth analysis. Finally, this chapter is concluded in Section 3.6.

3.2 System Model

In this section, the conceptual architecture of an integrated aerial terrestrial multi-hop downlink network is introduced first. Later, the DEP for the NOMA phase is presented.

3.2.1 Architecture of the Aerial Terrestrial Network

Fig. 3.1 illustrates the conceptual architecture of an integrated aerial terrestrial multi-hop downlink network. The architecture consists of a single MBS that utilizes next-generation multiple access schemes to establish communication with multiple IoT devices. The system comprises a set of ABSs denoted by $u \in \mathcal{U} = \{1, 2, \dots, |\mathcal{U}|\}$. This set \mathcal{U} is further categorized into two subsets based on different QoS: the cooperative ABS set and the serving ABS set denoted by \mathcal{U}_c and \mathcal{U}_s , respectively.

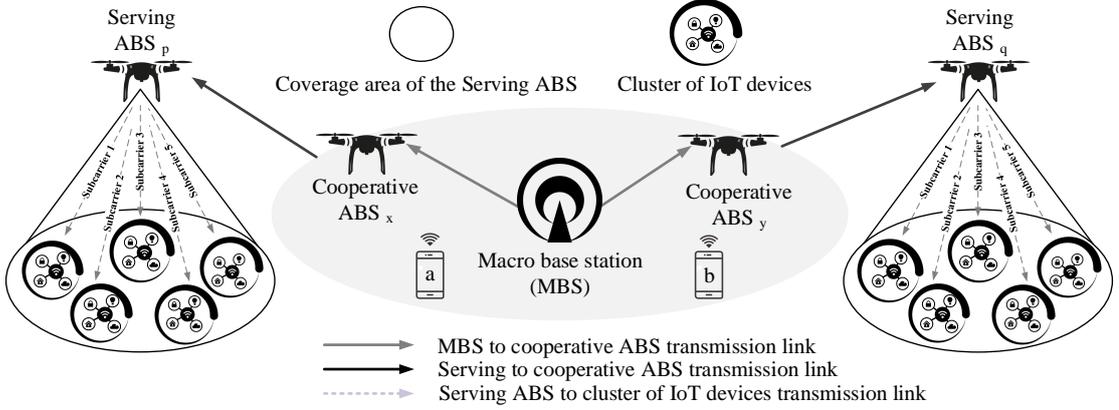


Figure 3.1: Considered Multi-hop Scenario: $\mathcal{U}_h = \{x, y\}$, where $(x, y) \in \mathcal{U}_h$, $\mathcal{U}_u = \{p, q\}$, where $(p, q) \in \mathcal{U}_u$, $|\mathcal{I}|$ IoT Devices are Grouped into \mathcal{L} Communities Located within the Coverage of Serving ABSs, and $\mathcal{M} = \{a, b\}$, where $(a, b) \in \mathcal{M}$ Located within the Coverage of MBS.

The cooperative ABSs⁵ act as relay nodes to facilitate the successful transmission of messages to the i^{th} IoT device. The set of \mathcal{U}_h is represented by $u_h \in \mathcal{U}_h = \{x, y, \dots, |\mathcal{U}_h|\}$ and the set of \mathcal{U}_u is expressed as $u_u \in \mathcal{U}_u = \{p, q, \dots, |\mathcal{U}_u|\}$. The IoT devices within the MBS coverage are denoted by the set $m \in \mathcal{M} = \{a, b, \dots, |\mathcal{M}|\}$ and IoT devices within the coverage of each serving ABS are represented by the set $i \in \mathcal{I} = \{u_k^{(1)}, u_k^{(2)}, \dots, |\mathcal{I}|\}$. It is important to note that all IoT devices are positioned on the ground and can be served directly from MBS or through ABS using multi-hop communication.

In a demand-based network, it is reasonable to consider that it is resource-constrained and has limited bandwidth. We divide the total bandwidth (W) into $|\mathcal{C}|$ orthogonal sub-carriers, denoted by $c \in \mathcal{C} = \{1, 2, \dots, |\mathcal{C}|\}$, i.e., $\sum_{c \in \mathcal{C}} w_c = W$. In addition, all devices (IoT devices and ABSs) are grouped into \mathcal{L} communities. The set \mathcal{L} is expressed as $l \in \mathcal{L} = \{1, 2, \dots, |\mathcal{L}_c|\}$, where $c \in \mathcal{C}$ and L_c denote the maximum number of devices that can be served on the given sub-carrier c . If

⁵We assume all cooperative ABSs are situated within the coverage of MBS, and the most optimal cooperative ABSs among them will be elected as relay nodes.

$L_c = 1$, it means incorporating the novel concepts of the orthogonal scheme, while $1 < L_c \leq |S|$ means the incorporation of the NOMA scheme. Hence, we define two sets, $n \in \mathcal{N} = \{c|c \in \mathcal{C}, L_c = 1\}$ and $k \in \mathcal{K} = \{c|c \in \mathcal{C}, 1 < L_c \leq |S|\}$ containing the indexes of the OMA sub-carriers and NOMA sub-carriers, respectively. The notation $|S|$ presents the threshold value for the maximum number of devices on a sub-carrier.

Assume the set \mathcal{U}_k and \mathcal{U}_n containing the indexes of the devices assigned to NOMA sub-carrier ($k \in \mathcal{K}$) and OMA sub-carrier ($n \in \mathcal{N}$). We define a set $U_{k,[u_k]} = \{\cup i, \|h_{u_u,[i]}\|^2 \geq \|h_{u_u,[u_k]}\|^2, i, u_k \in \mathcal{U}_k\}$ that contains the indexes of the IoT devices that impose interference on the IoT device u_k allocated to the sub-carrier k within the same community, where $h_{u_u,[i]}$ denotes the channel of the i^{th} IoT device served by the ABS u_u . The priority to provide fairness between each IoT device is given by $\sum_{u_k \in \mathcal{U}_k} \omega_{k,[u_k]} = 1$. The SIC decoding order is also important for the power domain NOMA. This work considers the optimal decoding error, where we decode the signals of the IoT device from the highest to the lowest normalized noise power [94]. If the decoding order on the given sub-carrier is $\pi_k : \{U_{k,[u_k]}, u_k \in \mathcal{U}_k\}$. For $i \in \{1, 2, \dots, |U_k|\}$, $\pi_k(i)$ gives the location of the i^{th} decoded device on the k^{th} sub-carrier while $\pi_k^{-1}(i)$ gives its decoding order. The IoT device $\pi_k(i)$ first decodes the signals from the IoT devices $\pi_k(1)$ to $\pi_k(i-1)$ and before decoding the needed signal, subtracts them from the overlaid signal. The intervention of the IoT devices $\pi_k(\bar{i})$ for $\bar{i} > i$ is considered noise.

It is assumed that the subcarriers belonging to the set \mathcal{N} can serve at-most single cooperative ABS or IoT devices (within the MBS coverage). Still, one sub-carrier can serve multiple links simultaneously, i.e., the link between MBS and the IoT device and between cooperative to serving ABS. However, each transmission link can be assigned to only one sub-carrier between MBS and IoT devices (within its coverage) and the link between MBS and cooperative ABSs. The sub-carriers belonging to the set \mathcal{K} can be used by a maximum of one serving ABS, and that sub-carrier can only be allocated among $|S|$ IoT devices within the community, where

$s \in \mathcal{S} = \{2, 3, \dots, |\mathcal{S}|\}$. A matrix of size $(|U_h|+|M|) \times N$ is defined to describe the sub-carrier allocation indicator for MBS to cooperative ABSs and MBS to IoT devices within its vicinity. It is denoted by $\boldsymbol{\psi} = [\psi_{\text{mbs},j}^n]$. For $j \leq |U_h|$, $\psi_{\text{mbs},j}^n = 1$ means that a sub-carrier n is assigned to u_h^{th} cooperative ABS, otherwise $\psi_{\text{mbs},j}^n = 0$. Whereas, for $j > |U_h|$, $\psi_{\text{mbs},j}^n = 1$ means that a sub-carrier n is assigned to an IoT device m , otherwise $\psi_{\text{mbs},j}^n = 0$. We define a matrix of size $|U_u| \times K$ to describe the sub-carrier allocation indicator to serve ABS to IoT devices within its vicinity, shown by $\boldsymbol{\varphi} = [\varphi_{u_u,[u_k]}^k]$. Therefore, $\varphi_{u_u,[u_k]}^k = 1$ means that a sub-carrier k is assigned to the IoT device u_k , otherwise $\varphi_{u_u,[u_k]}^k = 0$.

The physical locations of MBS, m^{th} IoT device, u_h^{th} cooperative ABS, u_u^{th} serving ABS, and u_k^{th} IoT device are indicated as $(x = 0, y = 0, z = 0)$, (x_m, y_m, z_m) , $(x_{u_h}, y_{u_h}, z_{u_h})$, $(x_{u_u}, y_{u_u}, z_{u_u})$ and $(x_{u_k}, y_{u_k}, z_{u_k})$, respectively⁶. We have used different channel models due to the different channel characteristics in the LoS and non-LoS (N-LoS) probabilities for air-to-ground, ground-to-ground, and ground-to-air propagation models [95]. The distance between MBS and the cooperative ABS u_h^{th} and the distance between MBS and the m^{th} IoT device are calculated as $d_{\text{mbs},u_h} = \sqrt{(x - x_{u_h})^2 + (y - y_{u_h})^2 + (z - z_{u_h})^2}$ and $d_{\text{mbs},m} = \sqrt{(x - x_m)^2 + (y - y_m)^2 + (z - z_m)^2}$, respectively. The pathloss for the given link is given by $l_{\text{mbs},m} = 55.9 + 38 \log(d_{\text{mbs},m}) + (24.5 + 1.5f/925) \log(f)$, where f represents the carrier frequency. The distance between u_h^{th} cooperative ABS and u_u^{th} serving ABS is given as $d_{u_h,u_u} = \sqrt{(x_{u_h} - x_{u_u})^2 + (y_{u_h} - y_{u_u})^2 + (z_{u_h} - z_{u_u})^2}$, where $d_{u_h,u_u}^{-\alpha_{\text{mbs}}}$ represents pathloss, where α_{mbs} represents the pathloss exponent.

The probability of LoS between the IoT device u_k and the ABS u_u is given as $p_{u_u,[u_k]}^{\text{LoS}} = \frac{1}{1+a \exp\left(-b \left[\left(\frac{180}{\pi}\right) \left(\sin^{-1}\left(\frac{z_{u_u}}{d_{u_u,u_k}}\right)\right) - a\right]\right)}$, where d_{u_u,u_k} is the distance between the given IoT device and serving ABS, calculated as $d_{u_u,u_k} = \sqrt{(x_{u_u} - x_{u_k})^2 + (y_{u_u} - y_{u_k})^2 + (z_{u_u} - z_{u_k})^2}$, and a and b are the constant values

⁶ABSs have diverse applications beyond communication. However, the current work focuses primarily on aspects that do not facilitate device-to-device communication. ABSs predetermine the locations of IoT devices, whereas pilot signals are used to determine the CSI.

depending on environmental factors. The probability of establishing a non-LoS link is $p_{u_u, [u_k]}^{\text{N-LoS}} = 1 - p_{u_u, [u_k]}^{\text{LoS}}$. The pathloss between the IoT device u_k and the ABS u_u for the connection of LoS and N-LoS is given by $l_{u_u, [u_k]}^{\text{LoS}} = l_{\text{fsp}} + 20 \log_{10}(d_{u_u, u_k}) + \eta_{\text{LoS}}$, and $l_{u_u, [u_k]}^{\text{N-LoS}} = l_{\text{fsp}} + 20 \log_{10}(d_{u_u, u_k}) + \eta_{\text{N-LoS}}$, respectively. The free space pathloss is given as $l_{\text{fsp}} = 20 \log_{10}(f) + 20 \log_{10}\left(\frac{4\pi}{c}\right)$, where c represents the speed of light, η_{LoS} and $\eta_{\text{N-LoS}}$ presents the attenuation due to LoS and N-LoS connection, respectively. Thus, the average pathloss is given by $pl_{u_u, [u_k]}^{\text{avg}} = p_{u_u, [u_k]}^{\text{LoS}} l_{u_u, [u_k]}^{\text{LoS}} + p_{u_u, [u_k]}^{\text{N-LoS}} l_{u_u, [u_k]}^{\text{N-LoS}}$ [95].

Let h_{mbs, u_h}^n be the channel between MBS and cooperative ABS u_h . It is computed as $h_{\text{mbs}, u_h}^n = \frac{g_{\text{mbs}, u_h}^n}{[(x-x_{u_h})^2 + (y-y_{u_h})^2 + (z-z_{u_h})^2]}$, where g_{mbs, u_h}^n is the channel power gain on the given sub-carrier. The signal-to-noise ratio (SNR) at the u_h^{th} cooperative ABS on the given sub-carrier is computed as $\varrho_{\text{mbs}, u_h}^n = \frac{\psi_{\text{mbs}, u_h}^n p_{\text{mbs}, u_h}^n \|h_{\text{mbs}, u_h}^n\|^2}{\delta^2}$, where δ^2 is the noise spectral density, and p_{mbs, u_h}^n shows the power allocated to the given cooperative ABS. The achievable rate for the given link is computed by normalizing over the sub-carriers bandwidth w_n [69].

$$r_{\text{mbs}, u_h}^n = \log_2 \left(1 + \varrho_{\text{mbs}, u_h}^n \right) - \sqrt{\frac{V_{\text{mbs}, u_h}^n Q^{-1}(\epsilon_{\text{mbs}, u_h})}{b_{\text{mbs}, u_h} \ln 2}}, \quad (3.1)$$

where, b_{mbs, u_h} is the adopted blocklength, and Q is the Gaussian Q-function, i.e., $Q(x) = \frac{1}{2\pi} \int_x^\infty \exp(-\frac{t^2}{2}) dt$ [10]. The DEP for the link between MBS and cooperative ABS u_h is approximately $\epsilon_{\text{mbs}, u_h} \approx Q(f(\varrho_{\text{mbs}, u_h}^{\min}, r_{\min}^{u_h}, b_{\text{mbs}, u_h}))$, $\forall u_h \in \mathcal{U}_h$, where $\varrho_{\text{mbs}, u_h}^{\min}$ is the minimum SNR received in all sub-carriers allocated for the link between MBS and cooperative ABS u_h and $r_{\min}^{u_h}$ is the minimum achievable rate. The overall DEP for this hop is given by $\epsilon_{\text{mbs}, u_h}, \forall u_h \in \mathcal{U}_h$. The channel dispersion for the given link is computed by $V_{\text{mbs}, u_h}^n = 1 - (1 + \varrho_{\text{mbs}, u_h}^n)^{-2}$. The sum-rate for the given cooperative ABS is computed as $r_{\text{mbs}, u_h} = \sum_{n \in \mathcal{N}} \psi_{\text{mbs}, u_h}^n r_{\text{mbs}, u_h}^n, \forall u_h \in \mathcal{U}_h$.

Let $h_{\text{mbs}, m}^n$ be the channel between the MBS and the m^{th} IoT device, which is defined as $h_{\text{mbs}, m}^n = \frac{g_{\text{mbs}, m}^n}{l_{\text{mbs}, m}}$, where $g_{\text{mbs}, m}^n$ is the channel power gain. The received signal to interference plus noise ratio (SINR) at the m^{th} device is calculated as $\varrho_{\text{mbs}, m}^n = \frac{\psi_{\text{mbs}, m}^n p_{\text{mbs}, m}^n \|h_{\text{mbs}, m}^n\|^2}{\delta^2 + I_{u_h, m}^n}$. where $p_{\text{mbs}, m}^n$ denotes the transmitted power for the m^{th} device, and $I_{u_h, m}^n$ is the interference power caused by the re-used link between

cooperative to serving ABSs. It is defined as $I_{u_h, m}^n = p_{u_h, u_u}^n h_{u_h, m}^n$, where p_{u_h, u_u}^n is the allocated power for the link between cooperative to serving ABS and $h_{u_h, m}^n$ is the channel between u_h^{th} cooperative ABS and the m^{th} IoT device. It is given by $h_{u_h, m}^n = g_{u_h, m}^n l_{u_h, m}^{-\alpha_{\text{mbs}}}$, where $g_{u_h, m}^n$ is the channel gain and $l_{u_h, m}^{-\alpha_{\text{mbs}}}$ is the pathloss for the following channel with pathloss exponent α_{mbs} . The achievable rate for the given link is calculated by normalizing over the sub-carrier's bandwidth w_n .

$$r_{\text{mbs}, m}^n = \log_2 \left(1 + \varrho_{\text{mbs}, m}^n \right) - \sqrt{\frac{V_{\text{mbs}, m}^n Q^{-1}(\epsilon_{\text{mbs}, m})}{b_{\text{mbs}, m} \ln 2}}, \quad (3.2)$$

where, $b_{\text{mbs}, m}$ is the adopted blocklength. The DEP for the link between MBS and the IoT device m is approximated as $\epsilon_{\text{mbs}, m} \approx Q(f(\varrho_{\text{mbs}, m}^{\min}, r_{\text{min}}^m, b_{\text{mbs}, m}))$, $\forall m \in \mathcal{M}$, where $\varrho_{\text{mbs}, m}^{\min}$ is the minimum received SNR for all subcarriers assigned for the link between MBS and the IoT device m and r_{min}^m is the minimum achievable rate. The overall DEP for this hop is given by $\epsilon_{\text{mbs}, m}, \forall m \in \mathcal{M}$. The channel dispersion for the given link is computed by $V_{\text{mbs}, m}^n = 1 - (1 + \varrho_{\text{mbs}, m}^n)^{-2}$. The sum-rate for the given IoT device is computed as $r_{\text{mbs}, m} = \sum_{n \in \mathcal{N}} \psi_{\text{mbs}, m}^n r_{\text{mbs}, m}^n, \forall m \in \mathcal{M}$.

Let h_{u_h, u_u}^n be the channel between cooperative ABS u_h and serving ABS u_u , which is given by $h_{u_h, u_u}^n = g_{u_h, u_u}^n d_{u_h, u_u}^{-\alpha_{\text{mbs}}}$, where g_{u_h, u_u}^n represents the channel gain. The SINR is calculated as $\varrho_{u_h, u_u}^n = \frac{\psi_{u_h, u_u}^n p_{u_h, u_u}^n \|h_{u_h, u_u}^n\|^2}{\delta^2 + I_{u_h, u_u}^n}$ at the u_u^{th} serving ABS. The interference power caused by the IoT devices in the set \mathcal{M} is defined as $I_{u_h, u_u}^n = \sum_{m \in \mathcal{M}} p_{\text{mbs}, m}^n h_{\text{mbs}, u_u}^n$, where h_{mbs, u_u}^n represents the channel between MBS and ABS u_u . It is calculated as $h_{\text{mbs}, u_u}^n = g_{\text{mbs}, u_u}^n \times l_{\text{mbs}, u_u}^{-\alpha_{\text{mbs}}}$, where the terms g_{mbs, u_u}^n and $l_{\text{mbs}, u_u}^{-\alpha_{\text{mbs}}}$ represent the gain and the pathloss between MBS and the given ABS, respectively. The achievable rate for the given link is calculated as [69]

$$r_{u_h, u_u}^n = \log_2 \left(1 + \varrho_{u_h, u_u}^n \right) - \sqrt{\frac{V_{u_h, u_u}^n Q^{-1}(\epsilon_{u_h, u_u})}{b_{u_h, u_u} \ln 2}}, \quad (3.3)$$

where, b_{u_h, u_u} is the allocated blocklength. $\epsilon_{u_h, u_u} \approx Q(f(\varrho_{u_h, u_u}^{\min}, r_{\text{min}}^{u_u}, b_{u_h, u_u}))$, $\forall u_u \in \mathcal{U}_u$ is the DEP for the link between cooperative ABS u_h and serving ABS u_u , where $\varrho_{u_h, u_u}^{\min}$ is the minimum received SNR across all allocated sub-carriers for the link between cooperative ABS u_h and serving ABS u_u and $r_{\text{min}}^{u_u}$ is the minimum

achievable rate. The total DEP for the link between MBS and the serving ABS u_u is given by $\epsilon_2 \approx \epsilon_{\text{mbs},u_h} + (1 - \epsilon_{\text{mbs},u_h}) \cdot \epsilon_{u_h,u_u}, \forall u_u \in \mathcal{U}_u$. The channel dispersion at the given link is calculated as $V_{u_h,u_u}^n = (1 - (1 + \varrho_{u_h,u_u}^n)^{-2})$. The sum-rate for the link between cooperative ABS u_h and serving ABS u_u is calculated as $r_{u_h,u_u} = \sum_{n \in \mathcal{N}} \psi_{u_h,u_u}^n r_{u_h,u_u}^n, \forall u_u \in \mathcal{U}_u$.

Let $h_{u_u,[u_k]}$ is the channel between serving ABS u_u and the IoT device u_k . It is defined as $h_{u_u,[u_k]} = \frac{g_{u_u,[u_k]}^k}{pl_{u_u,[u_k]}^{\text{avg}}}$, where $g_{u_u,[u_k]}^k$ is the channel gain for the given sub-carrier, and $pl_{u_u,[u_k]}^{\text{avg}}$ is the average pathloss. The SINR computed on the u_k^{th} IoT device is expressed as $\varrho_{u_u,[u_k]}^k = \frac{\varphi_{u_u,[u_k]}^k p_{u_u,[u_k]}^k \|h_{u_u,[u_k]}\|^2}{\delta^2 + I_{u_u,[u_k]}^k}$, where $p_{u_u,[u_k]}^k$ is the transmitted power. The interference power caused by other IoT devices is given by $I_{u_u,[u_k]}^k = \sum_{u_k \in \mathcal{U}_k, u_k \neq u_k} g_{u_u,[u_k]}^k p_{u_u,[u_k]}^k$. The rate for the IoT device u_k is computed by normalizing over the given sub-carrier's bandwidth w^k .

$$r_{u_u,u_k}^k = \omega_{[u_k]} \log_2(1 + \varrho_{u_u,[u_k]}^k) - \sqrt{\frac{V_{u_u,[u_k]}^k Q^{-1}(\epsilon_{u_u,[u_k]})}{b_{l_k,[u_k]}} \frac{1}{\ln 2}}, \quad (3.4)$$

where, $\omega_{[u_k]}$ represents the priority of the given IoT device, and $b_{l_k,[u_k]}$ denotes the blocklength allocated to l^{th} community served by k^{th} sub-carrier. The channel dispersion for the given link is computed as $V_{u_u,[u_k]}^k = (1 - (1 + \varrho_{u_u,[u_k]}^k)^{-2})$. The sum-rate for the given IoT device is computed as $r_{u_u,[u_k]} = \sum_{k \in \mathcal{K}} \varphi_{u_u,[u_k]}^k r_{u_u,u_k}^k, \forall u_k \in \mathcal{U}_k$. The energy efficiency of the system is defined as the ratio of the achievable rate of the system to the total power consumed by the system, including flexible transmit power and circuit power (CP) [96].

3.2.2 Decoding Error Probability for the NOMA Phase

Considering NOMA $|\mathcal{S}|=2$ in Fig. 3.2, where two IoT devices namely: $u_k^{(1)}$ and $u_k^{(2)}$ are allocated to the sub-carrier k within the same community within the coverage of serving ABS u_u . The device $u_k^{(1)}$ is considered as a stronger user and device $u_k^{(2)}$ is considered weaker user. IoT device $u_k^{(1)}$ as a stronger device can perform SIC and first decodes the message of IoT device $u_k^{(2)}$ while treating its message as interference. If this is successful, then it decodes its message. Therefore, the total

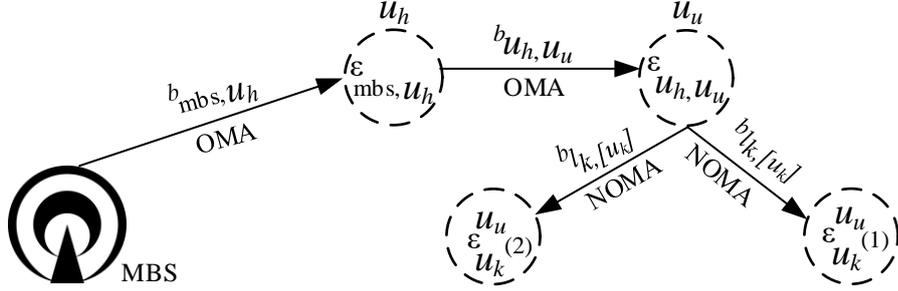


Figure 3.2: Systematic Diagram depicting Blocklength and DEPs for the Considered Scenario Setting.

DEP of $u_k^{(1)}$ depends on the DEP of previous transmission links and successful SIC at $u_k^{(1)}$. Whereas the IoT device $u_k^{(2)}$ directly decodes its signal while treating the message of IoT device $u_k^{(1)}$ as noise. Therefore, the total DEP of $u_k^{(2)}$ only depends on the DEP of previous transmission phases. The DEP for detecting the data of IoT device $u_k^{(2)}$ at IoT device $u_k^{(1)}$ is approximated as

$$\epsilon_{[u_k^{(1)}, u_k^{(2)}]}^{u_u} \approx Q \left(f \left(\varrho_{[u_k^{(1)}, u_k^{(2)}]}^{\min}, r_{\min}^{[u_k]}, b_{l_k, [u_k]} \right) \right), \quad (3.5)$$

where $r_{\min}^{[u_k]}$ is the minimum achievable rate of the IoT device. The minimum received SINR across all the allocated sub-carriers for the IoT device $u_k^{(1)}$ related to detecting data of IoT device $u_k^{(2)}$ is computed as $\varrho_{[u_k^{(1)}, u_k^{(2)}]}^{\min} = \min \left(\varrho_{[u_k^{(1)}, u_k^{(2)}]}^1, \varrho_{[u_k^{(1)}, u_k^{(2)}]}^2, \dots, \varrho_{[u_k^{(1)}, u_k^{(2)}]}^{|U_k|} \right)$, where SIC is applied at the receiver end

and $\varrho_{[u_k^{(1)}, u_k^{(2)}]}^k = \frac{p_{u_u, [u_k^{(2)}]}^k \left\| h_{u_u, [u_k^{(1)}]} \right\|^2}{\left(p_{u_u, [u_k^{(1)}]}^k \times \left\| h_{u_u, [u_k^{(1)}]} \right\|^2 \right) + \delta^2}$, $\forall k \in \mathcal{K}$. The DEP for detecting the data

of IoT device $u_k^{(1)}$ at the IoT device $u_k^{(1)}$ is given by

$$\epsilon_{[u_k^{(1)}, u_k^{(1)}]}^{u_u} \approx Q \left(f \left(\varrho_{[u_k^{(1)}, u_k^{(1)}]}^{\min}, r_{\min}^{[u_k]}, b_{l_k, [u_k]} \right) \right), \quad (3.6)$$

where, $\varrho_{[u_k^{(1)}, u_k^{(1)}]}^{\min}$ is the minimum received SNR across all allocated sub-carrier for the IoT device $u_k^{(1)}$ related to detecting the data of IoT device $u_k^{(1)}$, and $\varrho_{[u_k^{(1)}, u_k^{(1)}]}^k =$

$\frac{p_{u_u, [u_k^{(1)}]}^k \left\| h_{u_u, [u_k^{(1)}]} \right\|^2}{\delta^2}$. Similarly, the DEP for detecting the data of IoT device $u_k^{(2)}$ at IoT device $u_k^{(2)}$ is given by

$$\epsilon_{[u_k^{(2)}, u_k^{(2)}]}^{u_u} \approx Q \left(f \left(\varrho_{[u_k^{(2)}, u_k^{(2)}]}^{\min}, r_{\min}^{[u_k]}, b_{l_{k, [u_k]}} \right) \right), \quad (3.7)$$

where, $\varrho_{[u_k^{(2)}, u_k^{(2)}]}^{\min}$ is the minimum received SNR across all allocated sub-carriers for the IoT device $u_k^{(2)}$ related to detecting the data of IoT device $u_k^{(2)}$, and $\varrho_{[u_k^{(2)}, u_k^{(2)}]}^k =$

$\frac{p_{u_u, [u_k^{(2)}]}^k \left\| h_{u_u, [u_k^{(2)}]} \right\|^2}{\left(p_{u_u, [u_k^{(1)}]}^k \times \left\| h_{u_u, [u_k^{(2)}]} \right\|^2 \right) + \delta^2}$ after employing successful SIC. The overall DEPs for both IoT devices from MBS are given by

$$\epsilon_{[u_k^{(1)}]}^{u_u} = \epsilon_2 + (1 - \epsilon_2) \cdot \left(\epsilon_{[u_k^{(1)}, u_k^{(2)}]}^{u_u} + \left(1 - \epsilon_{[u_k^{(1)}, u_k^{(2)}]}^{u_u} \right) \cdot \epsilon_{[u_k^{(1)}, u_k^{(1)}]}^{u_u} \right), \quad (3.8)$$

$$\epsilon_{[u_k^{(2)}]}^{u_u} = \epsilon_2 + (1 - \epsilon_2) \cdot \epsilon_{[u_k^{(2)}, u_k^{(2)}]}^{u_u}. \quad (3.9)$$

3.3 Problem Formulation

This work aims to optimize the sub-carrier allocation, blocklength allocation, and PA to maximize the minimum feasible rates while ensuring that the delay, reliability, and QoS constraints are met. The proposed optimization problem does not consider the transmission link between the MBS and IoT device $m \in \mathcal{M}$, as it only focuses on maximizing the minimum rate across each hop involved in transmitting information from the MBS to the IoT device $u_k \in \mathcal{U}_k$ ⁷.

The notation $\mathbf{b} = \{b_{\text{mbs}, u_h}, b_{\text{mbs}, m}, b_{u_h, u_u}, b_{l_{k, [u_k]}}, \forall u_h \in \mathcal{U}_h, \forall m \in \mathcal{M}, \forall u_u \in$

⁷Thereby, $\sum_{m \in \mathcal{M}} \overbrace{b_{\text{mbs}, m}}^{b_0}, \sum_{u_h \in \mathcal{U}_h} \overbrace{b_{\text{mbs}, u_h}}^{b_1}, \sum_{u_u \in \mathcal{U}_u} \overbrace{b_{u_h, u_u}}^{b_2}, \sum_{l \in \mathcal{L}} \overbrace{b_{l_{k, [u_k]}}}^{b_3}$, where the notations $b_0, b_1, b_2,$ and b_3 denote the sum of the blocklengths of IoT devices belonging to set \mathcal{M} , cooperative ABSs in set \mathcal{U}_h , serving ABSs in set \mathcal{U}_u , and IoT devices in community l assigned to sub-carrier set \mathcal{K} , respectively. It is important to note that all devices within the same community share the same blocklength.

$\mathcal{U}_u, \forall l \in \mathcal{L}$ represent the variable of blocklengths, $\phi = \{\psi_{\text{mbs},j}^n, \varphi_{u_u,[u_k]}^k, \forall j \in \{(|\mathcal{U}_h|+|\mathcal{M}|) \times \mathcal{N}\} \forall k \in \{(|\mathcal{U}_u| \times \mathcal{K})\}\}$ represent the variable of sub-carrier associations, and the variable for allocated transmit powers is notated as $\mathbf{p} = \{P_{\text{mbs},m}^n, p_{\text{mbs},u_h}^n, p_{\text{mbs},m}^n, p_{u_h,u_u}^n, p_{u_u,[u_k]}^k, \forall m \in \mathcal{M} \forall u_k \in \mathcal{U}_k \forall u_h \in \mathcal{U}_h \forall u_u \in \mathcal{U}_u\}$. The term D_{max} denotes the maximum tolerable delay, T_{block} represents the duration of the time required to convey one unit of blocklength, α_{u_u,u_k}^k is an association-based binary variable, where $\alpha_{u_u,u_k}^k = 1$ means that the IoT device u_k is served by the ABS u_u on the given sub-carrier k , otherwise 0. The optimization problem can be formulated as follows.

$$\max_{\mathbf{p}, \mathbf{b}, \phi} \min (r_{\text{mbs},u_h}^n, r_{u_h,u_u}^n, r_{u_u,u_k}^k)$$

subject to

$$C_1 : \sum_{j \in \mathcal{M}} \psi_{\text{mbs},j}^n \leq 1, \sum_{j \in \mathcal{U}_h} \psi_{\text{mbs},j}^n \leq 1, \sum_{k \in \mathcal{K}} \varphi_{u_u,[u_k]}^k \leq |\mathcal{S}|,$$

$$C_2 : \max \left(b_{\text{mbs},m}, \left(b_{\text{mbs},u_h} + b_{u_h,u_u} + b_{l_k,[u_k]} \right) \right) \leq M_{\text{max}},$$

$$\text{where, } M_{\text{max}} = (D_{\text{max}}/T_{\text{block}}),$$

$$\text{and } \left(b_{\text{mbs},m}, b_{\text{mbs},u_h}, b_{u_h,u_u}, b_{l_k,[u_k]} \right) \in \mathbb{Z}^+,$$

$$C_3 : (x_{u_h}^2 + y_{u_h}^2) \leq r_{\text{max}}^2,$$

$$C_4 : 0 < p_{\text{mbs},m} \leq p_{\text{mbs},m}^{\min}, 0 < p_{\text{mbs},u_h} \leq p_{\text{mbs},u_h}^{\min}, \quad (3.10)$$

$$0 < p_{u_h,u_u} \leq p_{u_h,u_u}^{\min}, \forall m \in \mathcal{M}, \forall u_h \in \mathcal{U}_h, \forall u_u \in \mathcal{U}_u,$$

$$C_5 : 0 \leq \sum_{u_k \in \mathcal{U}_k} \alpha_{u_u,u_k}^k p_{u_u,[u_k]}^k \leq \bar{p}^k, \forall k \in \mathcal{K},$$

$$C_6 : (\epsilon_{\text{mbs},m}, \epsilon_{\text{mbs},u_h}, \epsilon_{u_h,u_u}, \epsilon_{u_k}^{u_u}) \leq \epsilon_{\text{threshold}},$$

$$C_7 : r_{\text{mbs},m} \geq r_{\text{min}}^m, r_{\text{mbs},u_h} \geq r_{\text{min}}^{u_h}, r_{u_h,u_u} \geq r_{\text{min}}^{u_u},$$

$$r_{u_u,[u_k]} \geq r_{\text{min}}^{[u_k]}, \forall m \in \mathcal{M}, \forall u_h \in \mathcal{U}_h \forall u_u \in \mathcal{U}_u \forall u_k \in \mathcal{U}_k,$$

$$C_8 : \left(\frac{g_{u_u,[u_k]}^k p_{u_u,[u_k]}^k}{I_{u_u,[u_k]}^k} \right) \geq \bar{h}, \forall u_k \in \mathcal{U}_k \forall k \in \mathcal{K},$$

$$C_9 : \|h_{u_u,[i]}\|^2 \geq \|h_{u_u,[u_k]}\|^2, i, u_k \in \mathcal{U}_k.$$

The constraints in (3.10) are defined as follows.

- Constraint C_1 assures the maximum number of devices multiplexed on each sub-channel.
- Constraint C_2 restricts the blocklength and satisfies the end-to-end transmission delay for a single communication link.
- Constraint C_3 ensures that given cooperative ABSs lie within the radius of the MBS, denoted as r_{\max} .
- The constraint C_4 encompasses the minimum power requirements for various entities. Specifically, it represents the minimum power required by the IoT device m to meet the minimum rate requirement r_{\min}^m , the minimum power required by cooperative ABS u_h to meet the minimum rate requirement $r_{\min}^{u_h}$, and the minimum power required for the link between cooperative ABS u_h and serving ABS u_u to meet the minimum rate requirement $r_{\min}^{u_u}$.
- Constraint C_5 ensures that the power allocated to all the IoT devices within the same community should not be more than the total power \bar{p}^k allocated to the given sub-carrier for that community.
- The restriction C_6 guarantees that the DEP of each user (i.e., IoT devices and ABSs) should not violate their threshold $\epsilon_{\text{threshold}}$.
- Constraint C_7 ensures that each device's achievable rate should be more than or equal to its minimum rate requirement.
- Constraint $C_8 - C_9$ ensures that the SIC decoding is done successfully.

The objective function is a mixed-integer non-linear programming problem; therefore, it cannot be solved in polynomial time [94]. It results from the non-convexity of the non-convex normal approximation and the combinatorial constraint C_2 . The problem (3.10) can be solved by combining a penalty technique with monotonic optimization at a high computational cost [97]. Alternatively, it can be resolved by leveraging the problem's decomposition property. Therefore, a

common relaxation strategy is used to divide the maximization problem into two sub-problems [95]. The proposed solution is clearly explained in the following section.

3.4 Proposed Solution

3.4.1 Sub-carrier Allocation and Selection of Cooperative Aerial Base Stations within the Coverage of Macro Base Station

To obtain the subsequent iterative solution of sub-carrier allocation, i.e., $(\phi)^{i+1}$, the problem (3.10) is solved with fixed values of $(\mathbf{b}^{(i)}, \mathbf{p}^{(i)})$.

$$\begin{aligned} & \max_{\phi} \min (r_{\text{mbs}, u_h}^n, r_{u_h, u_u}^n, r_{u_u, u_k}^k) \\ & \text{subject to } C_1 - C_3 \text{ and } C_6 - C_7. \end{aligned} \quad (3.11)$$

3.4.1.1 Sub-carrier Allocation using Stable Matching

A traditional way to compute the best solution to the above sub-problem is to exhaustively search for every potential combination of sub-carriers and IoT devices. However, it is time-consuming and computationally expensive. Alternatively, it can be reformulated using matching theory with a low-complexity algorithm. The basic concepts are given below.

Definition 1 (*Two Way Matching*): The problem (3.11) is a two-way matching problem because a maximum of one IoT device should be allocated to sub-carrier from its priority order based on their rate⁸ values. For better understanding, preference order introduced for given IoT device m with any two sub-carriers (in its preference order) $j, j' \in (|U_h| + |M|) \times N$, $j, j' > |U_h|$, the two matchings τ and τ'

⁸This rate is calculated based on initial PA, which is to be optimized later to achieve better rates.

are defined as

$$(j, \tau) \succ_m (j', \tau') \Leftrightarrow r_{\text{mbs},m}^n(\tau) > r_{\text{mbs},m}^n(\tau'), \quad (3.12)$$

which implies that if m^{th} IoT device achieves a higher rate than sub-carrier j' , then device m prioritizes sub-carrier j in τ against sub-carrier j' in τ' . The terms swap matching and swap blocking pair are introduced and defined below to demonstrate the impact of externalities (peer effects).

Definition 2 (Swap Matching): Considering two IoT devices (m, m') and two sub-carriers (j, j') , the current matching state is denoted as $\tau(m) = j$ and $\tau(m') = j'$. A swap matching will be performed between (m, m') and (j, j') if m prefers sub-carrier j' over its current match and sub-carrier j' prefers m over its current match. The swap matching operation is defined as follows:

$$\tau_m^{m'} = \left\{ \tau \setminus \{(m, j), (m', j')\} \cup \{(m, j'), (m', j)\} \right\}, \quad (3.13)$$

where $\tau_m^{m'}$ represents the updated matching state, indicating that IoT device m is now matched with sub-carrier j' , and vice versa. Therefore, the swap-blocking pair is defined as follows.

Definition 3 (Swap Blocking Pair): From the given matching state $\tau(m) = j$, $\tau(m') = j'$, an IoT pair of devices (m, m') is a swap pair if there exists

1. $\forall q \in \{m, m', j, j'\}, \tau_m^{m'}(q) \geq_q \tau(q)$,
2. $\exists q \in \{m, m', j, j'\}, \tau_m^{m'}(q) \succ_q \tau(q)$,

where q shows the involved player (either sub-carrier or IoT device). It means that swap matching $\tau_m^{m'}$ is approved, and both IoT devices (m, m') can switch their sub-carriers in τ by following these two conditions: 1) rate should not reduce after swapping and 2) the rate of the at least one IoT device increases. The process continues until the swap-blocking pair does not exist, resulting in a globally converged solution. However, if the optimal matching is $\{(m, j), (m', j')\}$ and the current matching is $\{(m, j'), (m', j)\}$, then the solution may not converge and stick to a local optimum. The same procedure is adopted for the sub-carrier allocation to each cooperative ABS.

3.4.1.2 Selection of the Best Cooperative Aerial Base Stations

The selection of the best cooperative ABS from the set \mathcal{U}_h (to relay the information to neighbor serving ABS) is based on the maximum achievable rate, which is given by

$$u_h^{\text{opt}} = \arg \max_{u_h \in \mathcal{U}_h} r_{\text{mbs}, u_h}. \quad (3.14)$$

3.4.2 Sub-carrier Allocation and Selection of IoT Devices within the Coverage of Serving Aerial Base Station

Dynamic programming is utilized to compute the sub-carrier allocation and selection of IoT devices under multiplexing constraint C_1 , power constraint C_5 , and SIC constraints $C_8 - C_9$. The idea is to recursively compute three auxiliary vectors to keep the record of the current value of power, optimal solution, and backtracking, i.e., V, Q and T , respectively. Assuming \bar{p}^k as fixed power budget for the sub-carrier k , if $s \in \mathcal{S} = \{1, 2, \dots, |S|\}$, $u_k \in \mathcal{U}_k = \{1, 2, \dots, |U_k|\}$ and $f \geq u_k$, the term $V[s, u_k, f]$ is computed as an optimal power value after satisfying the constraints as mentioned earlier. The recurrence relation is given by $V[s, u_k, f] = \max(v_a, v_b, v_c)$, where (v_a, v_b, v_c) represents PAs. These are defined as follows.

$$\begin{cases} v_a = V[s, u_k, f] \\ v_b = \begin{cases} V[s-1, u_k-1, u_k-1] + \bar{A} - \bar{B}, \\ \text{if } 0 < p^{\text{opt}} < Q[s-1, u_k-1, u_k-1] \\ 0, \text{ otherwise} \end{cases} \\ v_c = V[s, u_k-1, f]. \end{cases} \quad (3.15)$$

The variables $\bar{A} = \sum_{f \in u_k} r_{u_u, f}^k(p^{\text{opt}})$ and $\bar{B} = \sum_{f \in u_k} r_{u_u, f}^k(0)$. The pseudocode for computing the optimal power p^{opt} within the range of $[0, \bar{p}^k]$ is provided in Algorithm (3.1). The algorithm first assigns the variables s and t with the values of $\pi_k(u_k)$ and $\pi_k(u_k - 1)$, respectively. Then it checks whether $u_k = 1$ or if the value of $\omega[s]$ is greater than or equal to $\omega[t]$. If either condition is true, it returns the value of

Algorithm 3.1 Computing Optimal Power on $[0, \bar{p}^k]$

```

1: Input:  $u_k, \pi_k, (g_{u_u, [u_k]}^k)_{u_k \in \mathcal{U}_k}, \bar{p}^k$ 
2: Output:  $p^{\text{opt}}$ 
3:  $s \leftarrow \pi_k(u_k), t \leftarrow \pi_k(u_k - 1)$ 
4: if  $u_k = 1$  or  $\omega_{[s]} \geq \omega_{[t]}$  then
5:   return  $\bar{p}^k$ 
6: else
7:   return  $\max\left(0, \min\left(\frac{\omega_{[t]} g_{u_u, [s]}^k - \omega_{[s]} g_{u_u, [t]}^k}{\omega_{[s]} - \omega_{[t]}}, \bar{p}^k\right)\right)$ 
8: end if

```

\bar{p}^k as the optimal power (line 4). Otherwise, the optimal power is computed using the formula specified in line 7. Algorithm (3.1) performs a fixed number of basic operations; therefore, its complexity is $\mathcal{O}(1)$.

3.4.3 Joint Blocklength and Power Optimization

For clarity, the sub-problem (3.11) can be explicitly articulated by sub-problems (3.16) and (3.17), which implies that the PA and blocklength allocation are done solely. To obtain the next best value of $\mathbf{b}^{(i+1)}$, we first solve the sub-problem (3.16) with fixed values of $(\phi^{(i+1)}, \mathbf{p}^{(i)})$.

$$\begin{aligned} & \max_{\mathbf{b}} \min (r_{\text{mbs}, u_h}^n, r_{u_h, u_u}^n, r_{u_u, u_k}^k) \\ & \text{subject to } C_1 - C_9. \end{aligned} \quad (3.16)$$

Concerning blocklength constraint C_2 , the bisection-based optimal value of blocklength is computed to minimize the complexity of the proposed solution. We assume $b_{\text{lb}}^k = 1, b_{\text{ub}}^k = M_{\text{max}} - B$, where B is a fixed value calculated as $B = b_1 + b_2$. Subsequently, the optimal value of blocklength $b_{l_k, [u_k]}$ is computed, defined as $b_{\text{opt}} = \arg \max_{\{ \lfloor b_{\text{mid}}^k \rfloor, \lceil b_{\text{mid}}^k \rceil \}} (r_{u_u, u_k}^k)$, which is upper bounded by a threshold value $\bar{\xi}$. We set the initial value of $b_{\text{mid}}^k = \frac{(b_{\text{ub}}^k + b_{\text{lb}}^k)}{2}$ and then update the value of $b_{\text{ub}}^k = b_{\text{mid}}^k$ if $r_{u_u, u_k}^k(b_{\text{opt}})|_{b_{\text{opt}}=b_{\text{mid}}^k} > \bar{\xi}$. Otherwise, it is considered as $b_{\text{lb}}^k = b_{\text{mid}}^k$. This

process continues until $b_{\text{ub}}^k - b_{\text{lb}}^k > \bar{\sigma}$ is achieved. The complexity of these steps is $\mathcal{O}(\log_2(M_{\text{max}}/\bar{\sigma}))$, where $\bar{\sigma} = 0.01$. Subsequent analysis reveals that the worst-case computing complexity of the exhaustive search method is $\mathcal{O}(\mathcal{K}^3)$, which is significantly higher compared to our proposed steps.

Afterwards, the problem (3.17) is solved with the fixed values of $(\boldsymbol{\phi}^{(i+1)}, \mathbf{b}^{(i+1)})$ to determine the next best value of $\mathbf{p}^{(i+1)}$.

$$\begin{aligned} & \max_{\mathbf{p}} \min (r_{\text{mbs},u_h}^n, r_{u_h,u_u}^n, r_{u_u,u_k}^k) \\ & \text{subject to } C_1 - C_9. \end{aligned} \quad (3.17)$$

To solve the above-mentioned sub-problem, the minimum power required by the given device ϑ on the given sub-carrier n is computed to achieve its minimum rate requirement under constraints C_4 and C_7 , where $\vartheta \in \{m, u_h\}$ like that in [98]. We define the overall minimum power required to the given device as $p_{\text{mbs},\vartheta}^{\text{min}} = \sum_{n \in \mathcal{N}} p_{\text{mbs},\vartheta}^{\text{min},n}$. We compute the minimum power on sub-carrier n as below.

$$p_{\text{mbs},\vartheta}^{\text{min},n} = \left(\mu_{\vartheta} - \frac{1}{g_{\text{mbs},\vartheta}^n} \right)^+, \forall n \in \mathcal{N}, \vartheta \in \{m, u_h\}, \quad (3.18)$$

$$\sum_{n \in \{n \in \mathcal{N} | p_{\text{mbs},\vartheta}^{\text{min},n} > 0\}} w_n \log_2(\mu_{\vartheta} g_{\text{mbs},\vartheta}^n) = r_{\text{min}}^{\vartheta}, \vartheta \in \{m, u_h\}, \quad (3.19)$$

where $(x)^+$ represents $\max(x, 0)$, μ and μ_{ϑ} are the intermediate variables. The optimal powers for the device, such as IoT device m and cooperative AB u_h on a given sub-carrier, can be computed using the water-filling algorithm [98].

$$p_{\text{mbs},\vartheta}^n = p_{\text{mbs},\vartheta}^{\text{min},n} + \left(\mu - \frac{1}{g_{\text{mbs},\vartheta}^n} - p_{\text{mbs},\vartheta}^{\text{min},n} \right)^+, \vartheta \in \{m, u_h\}, \quad (3.20)$$

$$\sum_{\vartheta \in \downarrow} \sum_{n \in \{n \in \mathcal{N} | p_{\text{mbs},\vartheta}^n > p_{\text{mbs},\vartheta}^{\text{min},n}\}} \left(\mu - \frac{1}{g_{\text{mbs},\vartheta}^n} - p_{\text{mbs},\vartheta}^{\text{min},n} \right) = P_{\text{mbs}} - \sum_{\vartheta \in \downarrow} \sum_{n \in \mathcal{N}} p_{\text{mbs},\vartheta}^{\text{min},n}, \quad (3.21)$$

where, P_{mbs} is the sum of all the powers allocated to the given sub-carriers defined as $\sum_{n \in \mathcal{N}} \psi_{\text{mbs},\vartheta}^n p_{\text{mbs},\vartheta}^n = P_{\text{mbs}}$, where $\psi_{\text{mbs},\vartheta}^n$ is a binary indicator for sub-carrier allocation. So $\psi_{\text{mbs},\vartheta}^n = 1$ if the given sub-carrier is allocated to the device ϑ ; otherwise $\psi_{\text{mbs},\vartheta}^n = 0$. This work guarantees adherence to the minimum QoS criteria, ensuring that

every communication link satisfies its specific minimum rate requirement. Therefore, the received SINR of the IoT device m from the MBS should be greater than or equivalent to its minimum SINR threshold $\varrho_{\text{mbs},m}^{\text{min},n}$ for the following link. It is given by

$$\left(\frac{\psi_{\text{mbs},m}^n p_{\text{mbs},m}^n h_{\text{mbs},m}^n}{\delta^2 + p_{u_h, u_u}^n h_{u_h, m}^n} \right) \geq \varrho_{\text{mbs},m}^{\text{min},n}. \quad (3.22)$$

Hence, the total achievable rate of the IoT device m computed across all the allocated sub-carriers should be greater than or equal to r_{min}^m . The maximum power allocated to the cooperative to the serving ABS communication link must also be restricted to achieve the minimum QoS criteria for the IoT devices within the coverage of MBS. Hence, the power allocated to the link between the cooperative to the serving ABS should be subject to the following constraints.

$$p_{u_h, u_u}^{\text{min},n} \leq p_{u_h, u_u}^n \leq \left(\frac{p_{\text{mbs},m}^n h_{\text{mbs},m}^n}{\varrho_{\text{mbs},m}^{\text{min},n} h_{u_h, m}^n} - \frac{\delta^2}{h_{u_h, m}^n} \right) \leq p_{u_u}^{\text{max}}, \quad (3.23)$$

The allocated power for the link between the given cooperative and serving ABS should also meet its minimum QoS requirement, as given below.

$$\left(\frac{\psi_{\text{mbs},m}^n p_{u_h, u_u}^{\text{min},n} g_{u_h, u_u}^n d_{u_h, u_u}^{-\alpha_{\text{mbs}}}}{\delta^2 + \sum_{m \in \mathcal{M}} \psi_{\text{mbs},m}^n p_{\text{mbs},m}^n h_{u_u, m}^n} \right) \geq \varrho_{u_h, u_u}^{\text{min},n}, \quad (3.24)$$

$$p_{u_h, u_u}^{\text{min},n} \leq \frac{\left(\delta^2 + \sum_{m \in \mathcal{M}} \psi_{\text{mbs},m}^n p_{\text{mbs},m}^n h_{u_u, m}^n \right) \varrho_{u_h, u_u}^{\text{min},n}}{g_{u_h, u_u}^n d_{u_h, u_u}^{-\alpha_{\text{mbs}}}}, \quad (3.25)$$

where $\varrho_{u_h, u_u}^{\text{min},n}$ can be computed by setting $p_{u_h, u_u}^n = p_{u_h, u_u}^{\text{opt}}$.

$$p_{u_h, u_u}^{\text{opt}} = \begin{cases} 0, & \text{if } p_{u_h, u_u}^{\text{min},n} > p_{u_h, u_u}^n \\ p_{u_h, u_u}^n, & \text{if } p_{u_h, u_u}^{\text{min},n} < p_{u_h, u_u}^n \\ p_{u_u}^{\text{max}}, & \text{if } p_{u_u}^{\text{max}} \in [p_{u_h, u_u}^{\text{min},n}, p_{u_h, u_u}^n] \\ \min(p_{u_h, u_u}^n, \max(p_{u_u}^{\text{max}}, p_{u_h, u_u}^{\text{min},n})), & \text{Otherwise.} \end{cases} \quad (3.26)$$

Thus, the total achievable rate for the link between cooperative ABS and serving ABS computed across all the allocated sub-carriers should be greater than or equal

to $r_{\min}^{u_u}$. Basically, the idea is to divide the minimum rate requirement for each device across all the allocated sub-carriers to ensure that the total power allocated across all the allocated sub-carriers to the given device should result in a rate better than the minimum rate requirement for that device⁹. The overall minimum power required for the following link is given below.

$$p_{u_h, u_u}^{\min} = \sum_{n \in \mathcal{N}} p_{u_h, u_u}^{\min, n}. \quad (3.27)$$

Relevant to the constraint C_5 , we distribute the power to the given number of IoT devices within the serving ABS allocated to sub-carrier k within the same community. It is worth mentioning that the sum of the powers allocated to each IoT device within a community assigned to a sub-carrier k must be less than or equal to \bar{p}^k . It is given by

$$\bar{p}^k \geq \sum_{u_k \in \mathcal{U}_k} \alpha_{u_u, u_k}^k p_{u_u, [u_k]}^k, \forall k \in \mathcal{K} \text{ and } \forall u_u \in \mathcal{U}_u. \quad (3.28)$$

The feasible set containing the feasible powers for these devices is given as

$$\mathcal{R} = \left\{ \bar{\mathbf{P}} : \sum_{k \in \mathcal{K}} \bar{p}^k \leq p_{u_u}^{\max} \text{ and } 0 \leq \bar{p}^k \leq p^k, \forall k \in \mathcal{K} \right\}. \quad (3.29)$$

The set \mathcal{R} can also be expressed as Cartesian's product of all the user's feasible sets, and p^k represents the power limit to the given sub-carrier. We determine the optimal value on line 4 employing a for loop for each IoT device. If the constraint C_5 is satisfied, then $p_{u_u, u_k [u_k]}^k = p^{\text{opt}}$. Otherwise, the algorithm backtracks and finds the highest index i such that $p_{u_u, u_k [i]}^k \geq p^{\text{opt}}$.

In this way, the optimal vector containing the power values for each IoT device can be retrieved, i.e., $p_{u_u, [i+1]}^k, \dots, p_{u_u, [u_k]}^k \leftarrow p^{\text{opt}}$ in line 10. Consequently, the complexity of the algorithm is $\mathcal{O}(S^2)$. However, if the optimal power is calculated for D different power budgets, the complexity will be $\mathcal{O}(S^2 + DS)$.

⁹We consider the sub-carrier allocated to the device with the lowest SINR to calculate the minimum power requirement per sub-carrier for that device.

Algorithm 3.2 Power Distribution on given IoT Devices

```

1: Input:  $(\mathcal{U}_k)_{u_k \in \mathcal{U}_k}, (\mathcal{U}_k)_{u_k \in \mathcal{U}_k}, (g_{u_u, [u_k]}^k)_{u_k \in \mathcal{U}_k}, \bar{p}^k$ 
2: Output:  $p_{u_u, [1]}^k, p_{u_u, [2]}^k, \dots, p_{u_u, [U_k]}^k$ 
3: for  $u_k \in |U_k|$  do
4:    $p^{\text{opt}} \leftarrow \text{OptimalPower}(u_k, u_k, w_c, g_{u_u, [u_k]}^k, \bar{p}^k)$ 
5:    $i \leftarrow (u_k - 1)$ 
6:   while  $i \geq 1$   $p_{u_u, [i]}^k < p^{\text{opt}}$  do
7:      $p^{\text{opt}} \leftarrow \text{OptimalPower}(i, u_k, w_c, g_{u_u, [u_k]}^k, \bar{p}^k)$ 
8:      $i \leftarrow (i - 1)$ 
9:   end while
10:   $p_{u_u, [i+1]}^k, \dots, p_{u_u, [u_k]}^k \leftarrow p^{\text{opt}}$ 
11: end for

```

3.4.4 Bisection-based Low-Complexity Adaption Algorithm

Algorithm (3.3) is designed to perform sub-carrier allocation utilizing matching theory with fixed values of power and blocklength in line 5. The best cooperative ABSs are selected based on the derived results in line 6. A bisection search is conducted within the specified range to determine the optimal blocklengths for the subsequent iteration, as indicated in line 7. Subsequently, the available power is distributed using water-filling while adhering to the power constraints C_4 and C_5 to meet the minimum QoS requirements, as stated in line 9.

In line 10, the power is allocated to the links between cooperative ABSs and serving ABSs using equations (3.22) to (3.26). The power distribution is achieved by optimizing the power budget on each sub-carrier through dynamic and geometric programming. This process involves a two-tier projected gradient descent-based algorithm that distributes the power among the devices. The algorithm iterates for each sub-carrier to optimize the power budget until the condition $\|\bar{\mathbf{P}}' - \bar{\mathbf{P}}\|_2^2 \leq \lambda$ is satisfied, as described in lines 11-19. The term λ corresponds to the error tolerance for the termination condition, whose value is chosen from reference paper [36].

Algorithm 3.3 Proposed BLCA Algorithm

-
- 1: **Input:** $\sigma, (\mathcal{U}_k)_{u_k \in \mathcal{U}_k}, (\mathcal{N})_{n \in \mathcal{N}}, (\mathcal{K})_{k \in \mathcal{K}}, \lambda, |\mathcal{S}|, p_{u_u}^{\max}$, recursive index $i = 1$, highest amount of iterations possible t_{\max} , and randomly choose feasible values $p_c, \phi^{(0)}, \mathbf{b}^{(0)}$ and $\mathbf{p}^{(0)}$.
 - 2: **Output:** $\mathbf{p}^*, \mathbf{b}^*, \phi^*$
 - 3: Suppose the starting point $\bar{\mathbf{P}} = 0$
 - 4: **while** Convergence or $i > t_{\max}$ **do**
 - 5: Solve (3.11) for fixed $(\mathbf{b}^{(i)}, \mathbf{p}^{(i)})$ to find $(\phi)^{i+1}$
 - 6: Selection of best cooperative ABSs (3.14)
 - 7: Solve (3.16) using bisection-based algorithmic steps
 - 8: Solve (3.17) for fixed $(\phi^{(i+1)}, \mathbf{b}^{(i+1)})$ to find $\mathbf{p}^{(i+1)}$
 - 9: Power distribution using water-filling (3.18 – 3.21)
 - 10: PA for cooperative to the serving ABS communication link (3.22 – 3.26)
 - 11: Power distribution to IoT devices on the sub-carrier k within the same community (3.28 – 3.29)
 - 12: **while** $\|\bar{\mathbf{P}}' - \bar{\mathbf{P}}\|_2^2 \leq \lambda$ **do**
 - 13: $\bar{\mathbf{P}}' \leftarrow \bar{\mathbf{P}}$ saving previous power vector
 - 14: $\Delta = \Delta \sum_{u_k \in \mathcal{U}_k} r_{u_u, u_k}^k(\bar{p}^k)$ and update step size σ
 - 15: $\bar{\mathbf{P}} = \text{Projection of } \bar{\mathbf{P}} + (\sigma \Delta \text{ on } \mathcal{R})$
 - 16: **end while**
 - 17: **for** $k \in \mathcal{K}$ **do**
 - 18: Allocate power to the IoT device u_k by algorithm (3.2)
 - 19: **end for**
 - 20: Compute rates using (3.1), (3.3) and (3.4)
 - 21: Set $i : i + 1$
 - 22: **end while**
-

The search direction in line 14 is calculated using the exact gradient method, and the step size is determined by backtracking using the exact line search method. In line 16, the projection of $\bar{\mathbf{P}}$ onto the feasible set \mathcal{R} is calculated, as presented in [99]. The power distribution among the devices within a sub-carrier is performed from line 17 to line 19, while the rates are computed in line 20. The algorithm stops when the difference between iterations is small enough, indicating convergence or the algorithm repeats until the maximum number of iterations $i > t_{\max}$ are reached. The proposed algorithm converges within $\mathcal{O}(\log_2(1/\lambda))$ iterations.

3.5 Results and Discussion

A comparative analysis is conducted to evaluate the BLCA (blocklength-constrained algorithm). Three PA use cases are also analyzed: minimum PA, where each IoT device satisfies its minimum rate, and dynamic PA, where low-priority IoT devices first fulfill their minimum rate requirements compared to high-priority IoT devices. The remaining power is then optimally distributed among the high-priority IoT devices.

3.5.1 Simulations Setup

We configured the MBS to transmit at a power of 40 watts with a coverage radius of 500 meters [100]. Within this setup, we deployed a total of $\mathcal{M} = 5$ IoT devices at a minimum distance of 30 meters, $\mathcal{U}_h = 5$ cooperative ABSs at a minimum distance of 350 meters, and $\mathcal{U}_s = 2$ serving ABSs at a minimum distance of 80 meters [101]. The serving ABSs are filled with $\mathcal{I}_q = 9$ and $\mathcal{I}_q = 11$ IoT devices, respectively. The circular coverage area of the serving ABSs has a radius of 50 meters, and their maximum transmit power is limited to 1 watt [102]. The cooperative ABS is approximately 100 meters from the serving ABS. We consider a maximum of $\mathcal{C} = 20$ sub-carriers. Unless specifically stated otherwise, we assume the following parameter values: the threshold for DEP is $\epsilon_{\text{threshold}} = 10^{-5}$, the path loss exponent

Parameters	Values
Altitude of the serving ABS and cooperative ABS (z_{u_u}, z_{u_h}) [10]	(50, 50) Meters
Coefficients for LoS and N-LoS (η_{LoS}, η_{N-LoS})	(1dB, 20dB)
Density and height of building (a, b) [103]	(12, 0.135)
Error tolerance λ	10^{-4}
Minimum rate requirement for each device r_{\min} [102]	2 bits/s/Hz
Number of cooperative and serving ABSs ($\mathcal{U}_h, \mathcal{U}_u$) [101]	(5, 2)
Number of IoT devices within the coverage of cooperative ABS p and q ($\mathcal{I}_p, \mathcal{I}_q$)	(9,11)
Number of sub-carriers and IoT devices within the coverage of MBS (\mathcal{C}, \mathcal{M}) [94]	(20, 5)
Noise power density [10]	-174 dBm/Hz
Power of the serving ABS $p_{u_u}^{\max}$ [102]	1 Watt
Power of the MBS P_{mbs} [100]	40 Watts
Pathloss exponent (α_{mbs}) [95]	2
Radius of the MBS r_{\max} [10]	500 Meters
Time needed to convey one unit of blocklength T_{block} [10]	0.01 Millisecond
Transmission delay D_{\max} [10]	1 Millisecond

Table 3.1: Simulation Parameters.

is $\alpha_{\text{mbs}} = 2$, the speed of light is $c = 3 \times 10^8$ meters per second, the circuit power is 10 watts, and the noise power density is $\delta^2 = -174$ dBm/Hz [10]. The altitude of both serving and cooperative ABSs is set to $z_{u_u} = 50$ meters and $z_{u_h} = 50$ meters, respectively [95]. The attenuation for the LoS and N-LoS connection is assumed to be 1 dB and 20 dB, respectively. The channel parameters, including the density and height of the building, are $a = 12$ and $b = 0.135$, respectively [103]. The minimum rate requirement for each link is set as $r_{\min} = r_{\min}^m = r_{\min}^{u_h} = r_{\min}^{u_u} = r_{\min}^{[u_k]} = 2$

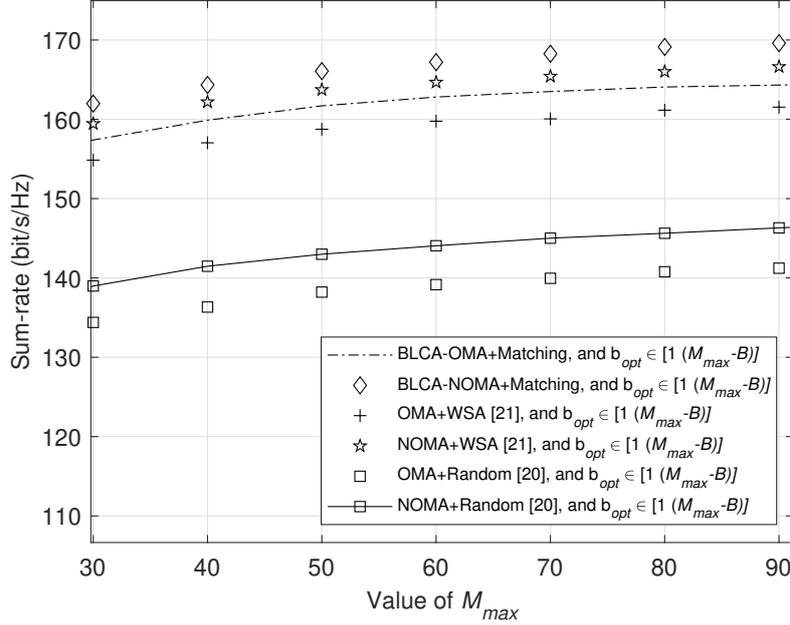


Figure 3.3: Sum-rate versus M_{max} with Optimal Blocklength and Power Allocation, where $b_1 = k \times b_2$, $b_2 = 3$, $M_{max} = k \times 10$, and $k = \{3, 4, \dots, 9\}$.

bits/s/Hz [102]. We use the radio propagation channel model provided in [101]. For simplicity, we assume that the sum of the blocklengths of each communication link within each hop is equal to the blocklength of individual links. We compare our proposed scheme, BLCA (Blocklength constrained algorithm), under two multiple access techniques, i.e., OMA and NOMA, which are named BLCA-OMA and BLCA-NOMA, respectively. Additionally, this work investigates them under two distinct scenarios of FBL, i.e., fixed-blocklength and random-blocklength approaches. In the fixed blocklength approach, we select a fixed value of $b_{opt} \in [1, 2, \dots, (M_{max} - B)]$. In the random blocklength approach, we randomly select a value of $b_{opt} \in [1, 2, \dots, (M_{max} - B)]$. The simulation parameters are summarized in Table 3.1.

3.5.2 Performance Comparison

In Fig. 3.3, the impact of heterogeneous delay on the time required to transmit a unit blocklength on the system sum-rate is analyzed. This work compares the

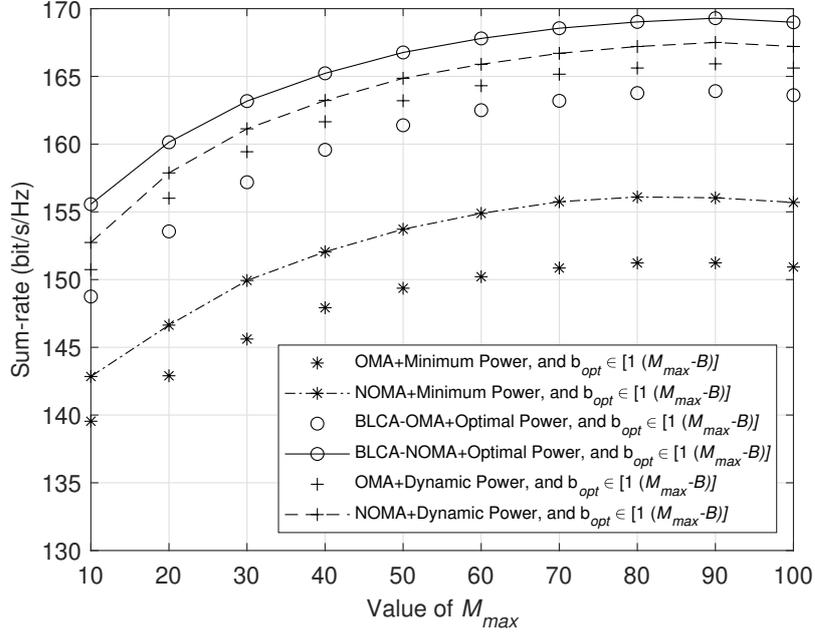


Figure 3.4: Sum-rate versus M_{max} using Matching-based Sub-carrier Allocation and Optimal Blocklength Allocation with different Power Allocation Strategies, where $b_1 = k \times b_2$, $b_2 = 3$, $M_{max} = k \times 10$, and $k = \{1, 2, \dots, 10\}$.

proposed BLCA algorithm with two baseline resource allocation schemes, namely random matching [20] and WSA matching [21]. The achievable rate is observed to increase with an increase in M_{max} because it depends on the maximum transmission delay. In addition, the proposed scheme demonstrates superior performance over the benchmark schemes, and the performance gap between the proposed scheme and the WSA matching and random matching schemes widens as the value of M_{max} increases. The enhanced throughput in the proposed scheme can be attributed to the significant improvement in both channel qualities and achievable SNR per sub-carrier achieved through stable matching. In contrast, the random matching approach [20] involves devices randomly selecting sub-carriers, which can result in sub-carrier assignments with inferior channel qualities. Similarly, the WSA scheme [21] may assign sub-carriers to devices with lower channel qualities. The effectiveness of the proposed scheme is evaluated using two different multiplexing techniques: OMA

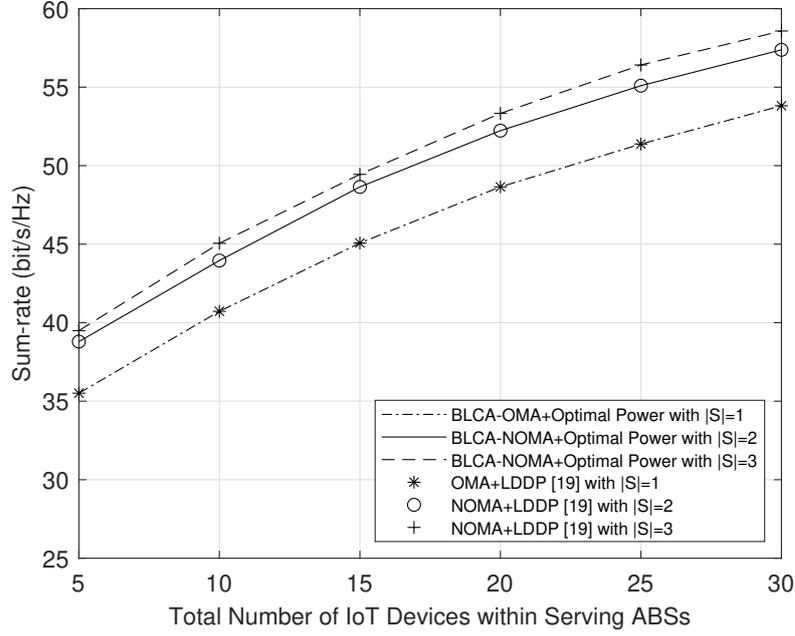


Figure 3.5: Sum-rate versus different Number of IoT Devices with Matching-based Sub-carrier Allocation against LDDP [19], where $M_{\max} = 100$.

with ($|S| = 1$) and NOMA with ($|S| = 2$). The superior performance of the NOMA scheme can be attributed to its fundamental principles, such as superposition coding at the transmitter (multiplexing two IoT devices per sub-carrier) and SIC at the receiver (demultiplexing based on power levels). The simulation results demonstrate that the system sum-rate is relatively low when M_{\max} is set to 30. However, it gradually increases to 169.59 bits/s/Hz, representing a 3.21% improvement (for BLCA-NOMA), after which it remains relatively constant.

Fig. 3.4 illustrates the trade-off between heterogeneous delays over blocklength and different PA techniques for OMA and NOMA systems. The following observations can be made: 1) the system throughput increases with an increase in the value of M_{\max} for all PA approaches, and 2) the proposed scheme outperforms both other PA techniques (OMA and NOMA with minimum power and dynamic PA) for both OMA and NOMA systems. The effectiveness of the NOMA system is significantly higher than that of the OMA system. This increase in effectiveness

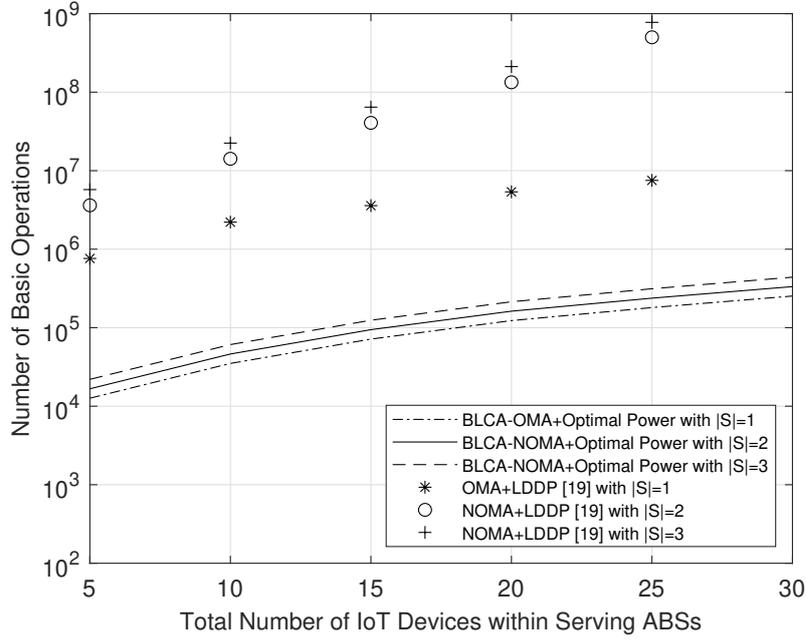


Figure 3.6: Complexity of the BLCA Scheme and LDDP [19] with Stable Matching-based Sub-carrier Allocation, Considering the basic Number of Operations, where $M_{\max} = 100$.

can be attributed to superposition coding at the transmitting node and SIC at the receiving node in conventional NOMA. The results indicate that the sum-rate for NOMA with optimal power is 4.58% higher than that for legacy OMA with optimal power and 2.68% higher than that for NOMA with dynamic PA.

The impact of an increasing number of IoT devices on the sum-rate is analyzed in Fig. 3.5. The proposed BLCA algorithm is compared with the near-optimal high-complexity benchmark scheme, LDDP [19]. Both schemes are simulated by varying the number of IoT devices in the vicinity of each serving ABS from 5 to 30 due to high computational run-time. As expected, the following observations are made: throughput increases with increased IoT devices, and greater participation of devices further elevates system throughput. The throughput gain of NOMA (with two and three IoT devices multiplexed per sub-carrier) is greater than that of OMA (with one IoT device per sub-carrier). There is a constant gap between NOMA

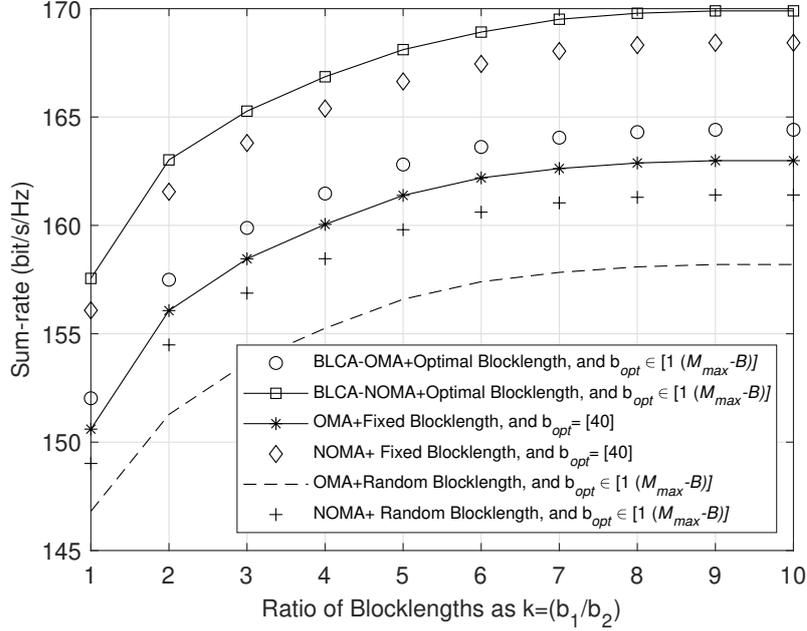


Figure 3.7: Sum-rate versus the ratio of Blocklengths with Stable Matching based Sub-carrier Allocation and Optimal Power Allocation, where $b_1 = k \times b_2$, $b_2 = 3$, and $M_{\max} = 100$.

with $|S|=2$ and $|S|=3$. Furthermore, the performance gain of BLCA and LDDP is almost the same for any number of IoT devices, indicating that the proposed BLCA algorithm is near-optimal. It should be noted that the proposed BLCA algorithm runs in seconds on a computer with specifications such as a core i5, 8th generation for $\mathcal{I} \leq 30$. In contrast, LDDP [19] requires 1600 times more operations for $\mathcal{I}=20$ and $|S|=2$ (as shown in Fig. 3.6), validating its low computational cost¹⁰ toward an optimal solution.

Fig. 3.7 shows the impact of the ratio of blocklengths, $k = (b_1/b_2)$, on the achievable system sum-rate. The results demonstrate that an increased blocklength ratio corresponds to a higher system throughput. This effect is because the degree of freedom to transmit data packets depends mainly on the blocklength. Consequently,

¹⁰We compute the computational complexity of each algorithm by assessing the number of basic operations (including additions, multiplications, and comparisons) performed.

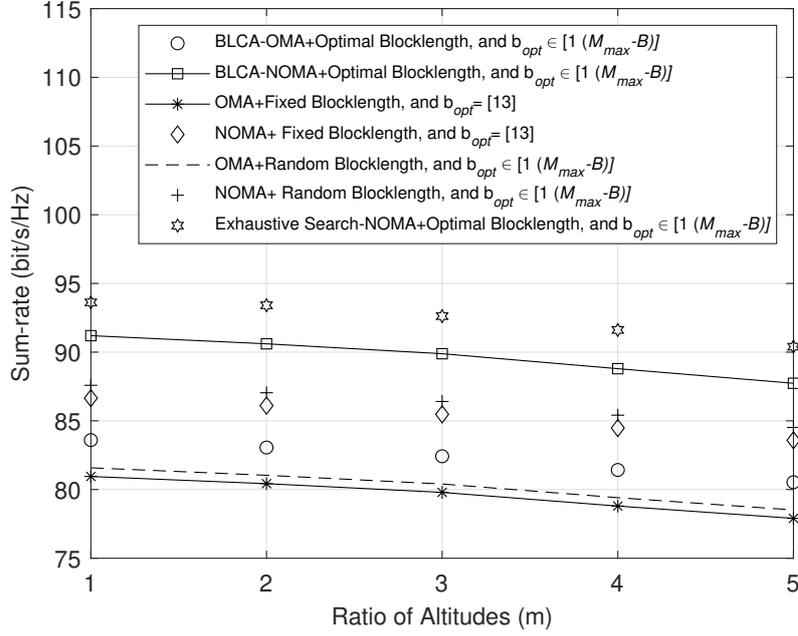


Figure 3.8: Sum-rate versus Altitudes using Stable Matching based Sub-carrier Allocation and Optimal Power Allocation, where $b_1 = 13$, $b_2 = 13$, $z_{u_u} = 50$, $z_{u_h} = \{50, 100, \dots, 250\}$ and $M_{\max} = 100$.

greater blocklength values lead to enhanced system sum rates. The proposed BLCA algorithm employing an optimal blocklength consistently yields better results than scenarios involving fixed or random blocklengths, emphasizing the importance of efficient blocklength allocation to maximize performance. When combining the advantages of the NOMA scheme with optimal blocklength, it emerges as the optimal choice, surpassing NOMA with fixed or random blocklengths. Hence, the NOMA scheme outperforms OMA due to its efficient utilization of spectrum resources, thereby accommodating multiple devices within resource constraints. It is important to note that NOMA with optimal blocklength surpasses OMA with an optimal blocklength. Similarly, NOMA with fixed or random blocklengths outperforms their respective OMA counterparts in their corresponding scenarios. The results validate that the throughput of BLCA-NOMA with optimal blocklength is 3.63% higher than that of BLCA-OMA with optimal blocklength, in all scenarios.

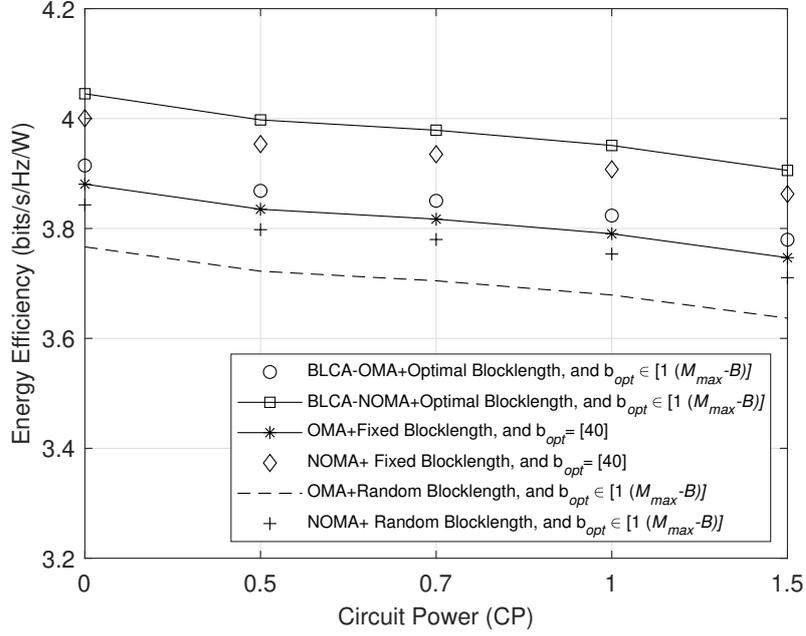


Figure 3.9: Spectral Efficiency versus Circuit Power using Stable Matching based Sub-carrier Allocation and Optimal Power Allocation, where $b_1 = 13$, $b_2 = 13$, and $M_{\max} = 100$.

Fig. 3.8 investigates the impact of ratio of altitudes, i.e., $H = (z_{u_h}/z_{u_u})$ on the achievable system sum-rate. Our assumption posits that the cooperative ABSs are strategically positioned at higher altitudes than the serving ABSs. In particular, the achievable rate of the proposed scheme decreases as the altitude increases due to higher channel fading and increased LoS interference. Regardless of the considered blocklength scenario, whether optimal, random, or fixed, the NOMA scheme consistently outperforms OMA. In all cases, the NOMA curve maintains a higher position than the OMA curve. In addition, the scheme employing an optimal blocklength consistently yields better results when compared to scenarios involving fixed or random blocklengths. NOMA with fixed or random blocklengths outperforms their respective OMA counterparts within their respective scenarios. The simulation results solidify that the sum-rate for BLCA-NOMA, employing optimal blocklength allocation, exceeds that of BLCA-OMA with optimal

blocklength allocation by a margin of 9.09%. Notably, the curve for NOMA with a fixed blocklength allocation is lower than that for NOMA with a random blocklength allocation. This difference arises from choosing a higher random blocklength value than the fixed blocklength. Monte-Carlo simulations are also conducted to compute the best possible solution. To validate our plotted curves, this work includes an upper-bound solution curve (computed using the exhaustive search method, also known as the brute force method). This curve demonstrates the proximity of our proposed solution to the optimal one.

Fig. 3.9 evaluates the current energy efficiency values of the proposed solution by analyzing the total energy efficiency of the system against CP. In our evaluation, this work emphasizes the distinction between the BLCA scheme under two distinct multiple-access techniques: OMA and NOMA. This work evaluates the efficacy of our proposed algorithm in varying blocklength scenarios. The simulation and results illustrate that increasing the value of CP results in a decrease in the total energy efficiency of the system. It is important to note that NOMA with an optimal blocklength surpasses OMA with an optimal blocklength within their respective scenarios. Similarly, NOMA with fixed or random blocklengths outperforms their respective OMA counterparts in their corresponding scenarios. Comparative analysis shows that the proposed BLCA-NOMA achieves a 5.25% (resp. BLCA-OMA 3.39%) improvement in energy efficiency for NOMA with random blocklength and a 1.12% improvement for NOMA with fixed blocklength (resp. OMA 0.87%). The fundamental reason behind this minimal increase is selecting a fixed blocklength value closer to its optimal value.

3.6 Summary

This study explores a mixed-integer non-linear programming problem for optimizing joint resource allocation in an integrated aerial-terrestrial wireless network to maximize the system sum-rate. A novel BLCA (blocklength constrained) algorithm

is proposed, which utilizes alternating optimization and a two-step projected gradient descent-based strategy to optimize the resource allocation policy while considering delay, reliability, and QoS constraints through dynamic and geometric programming. This work compares the proposed algorithm with benchmark algorithms that avoid allocating the worst sub-carrier to transmitting devices using various techniques. The study concludes that NOMA with optimal blocklength surpasses OMA with optimal blocklength, and NOMA with fixed or random blocklengths outperforms their respective OMA counterparts in their corresponding scenarios. Simulation results demonstrate the efficacy of the proposed algorithm, which requires 1600 times less computational cost than baseline approaches. The next chapter investigates the correlation between IoT devices using the concept of digital twins. For simplicity, only the OMA system is analyzed.

Chapter 4

¹Task Dependency Aware Optimal Resource Allocation Using One-climb Policy

¹This complete chapter was also presented in IEEE Transactions on Green Communications and Networking 2024, Print ISSN: 2473-2400, Online ISSN: 2473-2400, Digital Object Identifier: 10.1109/TGCN.2024.3425442.

4.1 Introduction

The rapid evolution of IoTs has facilitated the cost-effective connection of billions of wireless devices, introducing a new era of connectivity [104]. However, the limited battery life and computational power of IoT devices have emerged as a significant barrier, particularly in supporting computationally intensive applications within future-generation wireless networks. These limitations are primarily attributed to concerns about production costs and stringent size restrictions [105]. MEC has attracted significant attention to address these challenges and support wireless devices with low computing power [106].

Task offloading is MEC's most significant feature, enabling resource-constrained IoT devices to offload their computation-intensive tasks to high-performance edge servers, either binary or partially [107]. Each task is processed locally or offloaded to the edge server in binary offloading. In contrast, each task is partitioned and executed locally and on the edge server. Our research focuses primarily on binary offloading within the context of FBL, which is frequently used in IoT systems to process tasks that cannot be partitioned. This approach is instrumental in satisfying the increasing quality of service demands in edge networks, particularly in the context of resource allocation.

Tasks executed by different IoT devices are usually correlated. This correlation significantly influences decisions related to offloading and resource allocation. Identifying an optimal solution is challenging, as mentioned in the literature review. This chapter examines task dependency among devices for the first time within the context of DT-aided edge computing. We formalize a mixed-integer non-linear programming problem in a DT-enabled aerial-terrestrial network [66]. The proposed algorithm jointly optimizes transmit power, CPU frequency, and the task offloading policy, subject to delay and QoS constraints, thus minimizing the energy-time cost of the devices. The model is also extended from a simplified two-device framework to multiple devices, incorporating intermediate tasks.

The remainder of this chapter is organized as follows. Section 4.2 describes the

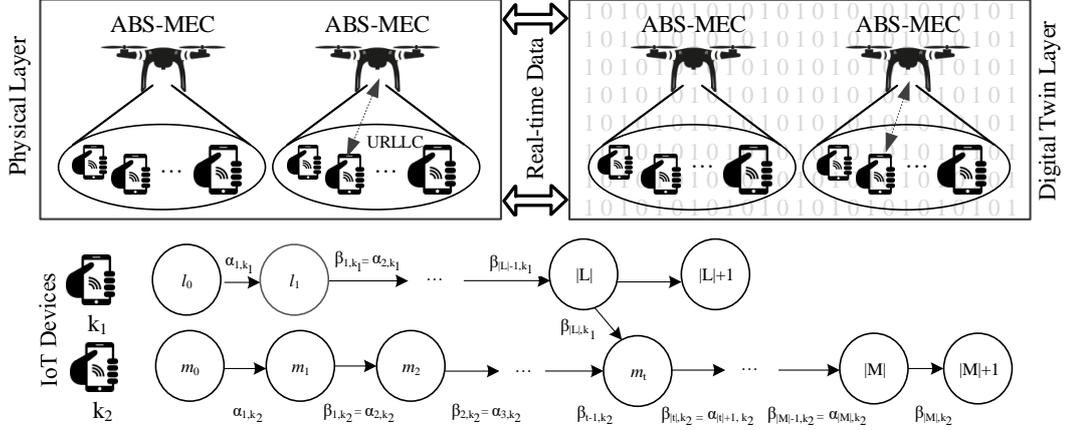


Figure 4.1: DT-enabled Integrated Aerial-Terrestrial Edge Network denoting a Two-device Framework with an Intermediate Task Node m_t

proposed system model. Section 4.3 provides the problem formulation. Section 4.4 explains the proposed solution, later extended to a multi-device scenario in Section 4.5. The simulation results are given in Section 4.6. Finally, this chapter is concluded in Section 4.7.

4.2 System Model

First, the architecture of DT in an edge network is presented. Next, the architecture of MEC based on URLLC is explained. Afterwards, the task offloading model is introduced. Finally, the DT, communication, computational, latency, and energy consumption models are briefly explained.

4.2.1 Architecture of Digital Twin Edge Network

Fig. 4.1 illustrates a DT-enabled edge computing network including two layers. The physical layer includes IoT devices and ABSs acting as access points (capable of working as edge servers, i.e., executing back-haul processing to mitigate the constraints of limited storage and computing resources). The devices are connected

via URLLC links to ensure stringent high-reliability and low-latency communication requirements in mission-critical applications. Each access point is assumed to be integrated with a multi-processor architecture with a fixed service rate. This configuration allows every access point to process a predetermined number of tasks simultaneously. The DT layer is a virtual replica of the physical layer, facilitating real-time monitoring of the physical system's operations.

4.2.2 Architecture of Mobile Edge Computing based on Ultra-reliable low-latency communication

This architecture considers a set of ABSs denoted by $a \in \mathcal{A} = \{a_1, a_2, \dots, |A|\}$. Each ABS serves a distinct non-overlapping region within its coverage area. It is assumed that each ABS has completed its deployment and networking planning beforehand. Within these coverage areas, This work considers different numbers of IoT devices represented by the set $k \in \mathcal{K} = \{k_1, k_2, \dots, k_j, \dots, |K|\}$. A binary variable is introduced to represent the connection between the device and ABS, i.e.,

$$\pi_{k,a} = \begin{cases} 1, & \text{if there is an association between the } (a)\text{-th} \\ & \text{ABS and the } (k)\text{-th IoT device} \\ 0, & \text{otherwise} \end{cases} \quad (4.1)$$

For easy understanding, the number of IoT devices within each serving ABS is limited to two, i.e., IoT devices k_1 and k_2 , exhibiting limited mobility. Later, we extend the proposed model to accommodate multi-user scenarios in Section (4.5). The sequential order of tasks to be performed by each respective device is denoted by the set $l \in \mathcal{L} = \{l_0, l_1, l_2, \dots, |L|, |L|+1\}$ and the set $m \in \mathcal{M} = \{m_0, m_1, m_2, \dots, |M|, |M|+1\}$. Two auxiliary nodes, l_0 and m_0 , are allocated as the starting points for the tasks assigned to each IoT device. Meanwhile, additional two nodes $|L|+1$ and $|M|+1$ are designated as termination points for the tasks assigned to each IoT device k_1 and k_2 , respectively. This chapter assumes a framework of task dependency among IoT devices, where the intermediary task $t \in \{t_1, t_2, \dots, |M|\}$

of IoT device k_2 requires the output from the final task of device k_1 . All tasks must be started and completed on the same device, e.g., the IoT device must execute two additional tasks locally. However, the remaining $|L + M|$ tasks may be processed locally or remotely.

It is reasonable to consider that the network under consideration is resource-constrained with limited bandwidth (W). The bandwidth is equally distributed among orthogonal sub-carriers, represented by the set $s \in \mathcal{S} = \{s_1, s_2, \dots, |\mathcal{S}|\}$, where $W = \sum_{s \in \mathcal{S}} w_s$. A binary variable for sub-carrier allocation is defined, where $\phi_{s,k} = 1$ indicates successful sub-carrier allocation; otherwise, $\phi_{s,k} = 0$. In numerous power-efficient IoT structures, the rate required to offload the task is often minimal and generally necessitates only a narrow bandwidth; e.g., a narrow-band IoT system utilizing a 10 MHz bandwidth can accommodate over fifty IoT devices through orthogonal transmission. Consequently, every device is assigned an orthogonal sub-carrier, having equal bandwidth. Our assumption to offload the task to an edge server only once minimizes the probability of offloading all the data simultaneously.

4.2.3 Task Offloading Model in Edge Network

This work represent the task i originating from the k^{th} device as a tuple $T_{i,k} = (\delta_{i,k}, \alpha_{i,k}, \beta_{i,k})$, where $i \in (l, m)$ and $k \in (k_1, k_2)$. In this context, $\delta_{i,k}$ specifies the computing resource requirement (cycles) necessary to execute the task, $\alpha_{i,k}$ indicates the input size of the task, and $\beta_{i,k}$ represents the output size of the task (bits). The computing resource requirement for auxiliary nodes corresponding to each IoT device is zero. Additionally for the given device k_1 , the input of task l is equal to the output of the preceding task $l - 1$, i.e., $\alpha_{l,k_1} = \beta_{l-1,k_1}$. For IoT device k_2 , it is assumed that

$$\alpha_{m,k_2} = \begin{cases} \beta_{m-1,k_2} + \beta_{|L|,k_1} & \text{if } m = t, \\ \beta_{m-1,k_2} & \text{Otherwise.} \end{cases} \quad (4.2)$$

It is important to mention that the initial node is characterized by an input value of $\alpha_{i,k} = 0$, and the final node by an output value of $\beta_{i,k} = 0, \forall k \in \mathcal{K}$. To determine

the feasibility of task offloading, This chapter introduces a computational offloading decision variable, where $\varphi_{i,k} = 1$ means task i is processed at the edge; otherwise, $\varphi_{i,k} = 0$. Our analysis operates under non-causal channel information, implying that the access point has full access to the CSI while downloading or offloading tasks. These assumptions are critical in optimal offloading decisions.

4.2.4 Digital Twin Model

The DT of the ABS-assisted MEC within the URLLC edge network is articulated as follows [69].

$$DT = \{(K, \tilde{K}), (A, \tilde{A})\}. \quad (4.3)$$

The replica of the physical system is defined as \tilde{K} and \tilde{A} . The following equation represents the DT for the k^{th} device.

$$DT_{i,k} = \left(f_{i,k}^{\text{lo}}, \hat{f}_{i,k}^{\text{lo}} \right). \quad (4.4)$$

The actual processing rate of the DT layer replicates the behavior of the physical IoT device is denoted by $f_{i,k}^{\text{lo}}$. Any deviation from the performance of the corresponding physical device is given by $\hat{f}_{i,k}^{\text{lo}}$ [67]. The DT model for the a^{th} ABS is articulated as follows.

$$DT_a = \left(f_a^{\text{es}}, \hat{f}_a^{\text{es}} \right). \quad (4.5)$$

The actual processing rate at which the physical ABS distributes the computing power of the edge server is denoted by f_a^{es} , and any deviation from the performance of the corresponding physical device is indicated by \hat{f}_a^{es} . It helps to minimize the processing latency gap at the DT layer by facilitating the adjustment of computing resource allocation. After gathering real-time data from the physical system, the digital services within the DT layer perform visualization and analysis to streamline and optimize decision-making, thereby enhancing system performance.

4.2.5 Communication Model

When device k offloads its task i to the access point a , the SINR is determined by the expression $\varrho_{i,k}^{\text{es},s} = \left(\frac{\pi_{k,a} \phi_{s,k} \varphi_{i,k} p_{i,k}^s \|h_{i,k}^s\|^2}{\bar{\sigma}^2 + \sigma_{i,k}} \right)$, where $p_{i,k}^s$ denotes the transmission power required for offloading task i by device k at sub-carrier s , $\bar{\sigma}^2$ represents the spectral noise density, and $\sigma_{i,k}$ indicates the interference power caused by neighboring IoT devices. The wireless channel between the k^{th} device and a^{th} ABS on the given sub-carrier is represented by $h_{i,k}^s$ and can be modelled as $h_{i,k}^s = \sqrt{g_{i,k}^s} \tilde{h}_{i,k}^s$, where $\tilde{h}_{i,k}^s$ describes the small-scale fading with zero mean and uniform variance and $g_{i,k}^s = PL_{a,k} + \eta^{\text{LoS}} \rho_{a,k}^{\text{LoS}} + \eta^{\text{N-LoS}} \rho_{a,k}^{\text{N-LoS}}$ is the large-scale channel co-efficient, where ρ^{LoS} and $\rho^{\text{N-LoS}}$ are the additional loss for the LoS and N-LoS respectively. The pathloss between the given ABS and the IoT device is calculated by $PL_{a,k} = 10 \log \left(\frac{4\pi f_c d_{a,k}}{c} \right)^2$, where $d_{a,k}$ is the Euclidean distance between the ABS a and (a,k) -th IoT device, f_c is the carrier frequency, and c is the speed of light. The probability of LoS is computed by $\rho_{a,k}^{\text{LoS}} = \frac{1}{1+a \exp \left[-b \left(\arctan \left(\frac{h_a}{d_{a,k}} \right) - v \right) \right]}$ at height h_a , and fixed environmental factors b and v . The probability of N-LoS is given by $\rho_{a,k}^{\text{N-LoS}} = 1 - \rho_{a,k}^{\text{LoS}}$ [24]. If the device k offloads the task i to the edge server, then the uplink data rate is given by

$$r_{i,k}^{\text{es},s} = w_s \log_2 \left(1 + \varrho_{i,k}^{\text{es},s} \right) - w_s \sqrt{\frac{v_{i,k}^s}{b_{i,k}}} \frac{Q^{-1}(\epsilon_{i,k})}{\ln 2}, \quad (4.6)$$

where, $i \in \{l, m\}, k \in \{k_1, k_2\}$.

The bandwidth allocated to the given sub-carrier s is denoted by w_s , the channel dispersion is calculated using $v_{i,k}^s = 1 - \left(1 + \varrho_{i,k}^{\text{es},s} \right)^{-2}$, and the blocklength is denoted by the variable $b_{i,k}$. The Gaussian Q-function is defined as $Q(x) = \frac{1}{2\pi} \int_x^\infty \exp(-\frac{t^2}{2}) dt$ [10]. The transmission time required to offload the task i from the device k to the access point is expressed as $\eta_{i,k}^{\text{es},s} = \frac{\beta_{i-1,k}}{r_{i,k}^{\text{es},s}}$, where $i \in \{l, m\}, k \in \{k_1, k_2\}$. Assuming that task i for device k is downloaded from the access point, the signal-to-noise ratio (SNR) is given by $\varrho_{i,k}^{\text{dl},s} = \left(\frac{\pi_{k,a} \phi_{s,k} \varphi_{i,k} p_a \|h_{i,k}^s\|^2}{\bar{\sigma}^2 + \sigma_{i,k}} \right)$, where p_a denotes the transmission power of the access point. The downlink data rate of task i for device k from the

edge server is given by

$$r_{i,k}^{dl,s} = w_s \log_2 \left(1 + \varrho_{i,k}^{dl,s} \right) - w_s \sqrt{\frac{v_{i,k}^s Q^{-1}(\epsilon_{i,k})}{b_{i,k} \ln 2}}, \quad (4.7)$$

where, $i \in \{l, m\}, k \in \{k_1, k_2\}$.

The channel dispersion for this link is calculated by $v_{i,k}^s = 1 - (1 + \varrho_{i,k}^{dl,s})^{-2}$, and the time required for downlink transmission is calculated as $\eta_{i,k}^{dl,s} = \frac{\beta_{i-1,k}}{r_{i,k}^{dl,s}}$, where $i \in \{l, m\}, k \in \{k_1, k_2\}$.

4.2.6 Computational Model

4.2.6.1 Local Computing

The time required for the local computation of the task i on device k using the actual processing rate is expressed as [69].

$$\eta_{i,k}^{lo,s} = \frac{\delta_{i,k}}{f_{i,k}^{lo}}, i \in \{l, m\}, k \in \{k_1, k_2\}. \quad (4.8)$$

Assuming the deviation between the physical value and its DT representation is predetermined [69], the latency gap for task i on device k is given by

$$\Delta\eta_{i,k}^{lo,s} = \frac{\delta_{i,k} \hat{f}_k^{lo}}{f_k^{lo} (f_{i,k}^{lo} - \hat{f}_{i,k}^{lo})}, i \in \{l, m\}, k \in \{k_1, k_2\}. \quad (4.9)$$

Consequently, the total local computation time for task i on device k is computed as

$$\eta_{i,k}^{lc,s} = \eta_{i,k}^{lo,s} + \Delta\eta_{i,k}^{lo,s}, i \in \{l, m\}, k \in \{k_1, k_2\}. \quad (4.10)$$

4.2.6.2 Edge Computing

The anticipated execution time for task i on device k when processed at the edge server is given by

$$\tilde{\eta}_{i,k}^{es,s} = \frac{\delta_{i,k}}{f_{i,k}^{es}}, i \in \{l, m\}, k \in \{k_1, k_2\}, \quad (4.11)$$

where $f_{i,k}^{\text{es}}$ denotes the fixed frequency of the CPU at the edge server. The latency gap between the real value and its DT is given by

$$\Delta\eta_{i,k}^{\text{es},s} = \frac{\delta_{i,k} \hat{f}_{i,k}^{\text{es}}}{f_{i,k}^{\text{es}} \left(f_{i,k}^{\text{es}} - \hat{f}_{i,k}^{\text{es}} \right)}, i \in \{l, m\}, k \in \{k_1, k_2\}. \quad (4.12)$$

4.2.7 Latency Model

The end-to-end latency within the network is given by

$$\eta_{i,k}^{\text{tot},s} = \eta_{i,k}^{\text{lc},s} + \eta_{i,k}^{\text{es},s} + \overbrace{\tilde{\eta}_{i,k}^{\text{es},s}}^{\eta_{i,k}^{\text{ec},s}} + \Delta\eta_{i,k}^{\text{es},s}. \quad (4.13)$$

4.2.8 Energy Consumption Model

The local EC required for the computing task i is calculated as follows [22].

$$\xi_{i,k}^{\text{lo},s} = \mu \frac{(\delta_{i,k})^3}{\left(\eta_{i,k}^{\text{lo},s} \right)^2} \approx \mu \delta_{i,k} \left(f_{i,k}^{\text{lo}} - \hat{f}_{i,k}^{\text{lo}} \right)^2, \quad (4.14)$$

where, $i \in \{l, m\}, k \in \{k_1, k_2\}$,

where μ represents the switched capacitance coefficient of the IoT device [69]. Let me define $f(x) = \sigma^2 \left(2^{\left(\frac{x}{w_s} \right)} - 1 \right)$, we get using (4.6) as

$$p_{i,k}^s = \frac{1}{\|h_{i,k}^s\|^2} f \left(\frac{\beta_{i-1,k}}{\eta_{i,k}^{\text{es},s}} \right), i \in \{l, m\}, k \in \{k_1, k_2\}. \quad (4.15)$$

Equation (15) shows the transmit power depends on the channel conditions and the data requirements. For example, an increase in the distance leads to a higher pathloss; therefore, it requires more transmission power to offload tasks effectively or to maintain a constant level of received signal power. The transmission energy required to offload task i at the access point is computed as

$$\xi_{i,k}^{\text{es},s} = p_{i,k}^s \eta_{i,k}^{\text{es},s} = \frac{\eta_{i,k}^{\text{es},s}}{\|h_{i,k}^s\|^2 f \left(\frac{\beta_{i-1,k}}{\eta_{i,k}^{\text{es},s}} \right)} \quad (4.16)$$

where, $i \in \{l, m\}, k \in \{k_1, k_2\}$.

The total EC is computed as

$$\zeta_{i,k}^{\text{tot},s} = \xi_{i,k}^{\text{lo},s} + \xi_{i,k}^{\text{es},s}. \quad (4.17)$$

4.2.9 Task Dependency Use Cases

Task dependency can be one of the four cases.

- Case I: When both IoT devices k_1 and k_2 offload tasks $|L|$ and t on edge, i.e., $\varphi_{|L|,k_1} = 1$ and $\varphi_{t,k_2} = 1$, then it is not necessary to offload the task or download the task.
- Case II: When device k_1 offloads task $|L|$ on the edge and device k_2 computes task t locally, i.e., $\varphi_{|L|,k_1} = 1$ and $\varphi_{t,k_2} = 0$, the resulting data from task $|L|$ are transmitted to the IoT device k_2 after completion of its computational processing at the edge node.
- Case III: When devices k_2 offload task t at the edge and device k_1 computes task $|L|$ locally, i.e., $\varphi_{t,k_2} = 1$ and $\varphi_{|L|,k_1} = 0$, the IoT device k_1 needs to offload the results before calculating task t at the edge node.
- Case IV: When both IoT devices k_1 and k_2 execute tasks $|L|$ and t locally, that is, $\varphi_{|L|,k_1} = 0$ and $\varphi_{t,k_2} = 0$, device k_1 first offloads its output to the edge, then the edge forwards it information to IoT device k_2 . Therefore, the offloading transmission time is calculated as $\eta_{|L|+1,k_1}^{\text{es},s} = \frac{\beta_{|L|,k_1}}{r_{|L|+1,k_1}^{\text{es},s}}$, where $r_{|L|+1,k_1}^{\text{es},s}$ denotes the corresponding offloading data rate. The offloading EC is given by $\xi_{|L|+1,k_1}^{\text{es},s} = \left(p_{|L|+1,k_1}^s \times \eta_{|L|+1,k_1}^{\text{es},s} \right)$, where $p_{|L|+1,k_1}^s$ represents the offloading transmission power. The downlink transmission time is computed as $\eta_{t',k_2}^{\text{dl},s} = \frac{\beta_{|L|,k_1}}{r_{t,k_2}^{\text{dl},s}}$, where $r_{t,k_2}^{\text{dl},s}$ is the corresponding downlink rate.

4.3 Problem Formulation

The computational time required by IoT device k_1 involves local processing time at the edge server, which is given by

$$\lambda_{k_1}^{\text{tcom}} = \sum_{l=1}^{|L|} \left[\overbrace{(1 - \varphi_{l,k_1}) \eta_{l,k_1}^{\text{lo},s}}^{\text{local}} + \overbrace{\varphi_{l,k_1} \tilde{\eta}_{l,k_1}^{\text{es},s}}^{\text{edge}} \right]. \quad (4.18)$$

The total computational delay for IoT device k_1 on offloading or downloading the task to/from the edge server is given by

$$\lambda_{k_1}^{trans} = \sum_{l=1}^{|\mathcal{L}|+1} \left[\overbrace{\varphi_{l,k_1}(1 - \varphi_{l-1,k_1})\eta_{l,k_1}^{es,s}}^{\text{offloading}} + \overbrace{(1 - \varphi_{l,k_1})\varphi_{l-1,k_1}\eta_{l,k_1}^{dl,s}}^{\text{downloading}} \right]. \quad (4.19)$$

The transmission delay is zero when both tasks are processed on the same IoT device, i.e., $\varphi_{l-1,k_1} = \varphi_{l,k_1}$. In contrast, a delay occurs during the offloading process if $\varphi_{l-1,k_1} = 0$ and $\varphi_{l,k_1} = 1$. Similarly, there is a delay in downloading when $\varphi_{l-1,k_1} = 1$ and $\varphi_{l,k_1} = 0$. Therefore, the total execution time for the device k_1 is calculated as $\lambda_{k_1}^{tot} = \lambda_{k_1}^{tcom} + \lambda_{k_1}^{trans}$. The total EC of the IoT device k_1 is the cumulative sum of the EC of the tasks $|\mathcal{L}|$ and the EC required to offload the output of the final result if executed locally. The total EC is expressed as follows.

$$\xi_{k_1} = \overbrace{\sum_{l=1}^{|\mathcal{L}|} \left[(1 - \varphi_{l,k_1})\xi_{l,k_1}^{lo,s} + \varphi_{l,k_1}(1 - \varphi_{l-1,k_1})\xi_{l,k_1}^{es,s} \right]}^{\text{EC of the } |\mathcal{L}| \text{ tasks}} + \overbrace{(1 - \varphi_{|\mathcal{L}|,k_1})\xi_{|\mathcal{L}|+1,k_1}^{es,s}}^{\text{EC while offloading final result, if } \varphi_{|\mathcal{L}|,k_1}=0}. \quad (4.20)$$

It should be mentioned that the energy required to perform a given task, indicated by $\xi_{l,k_1}^{es,s}$ occurs only when $\varphi_{l,k_1} = 1$ and $\varphi_{l-1,k_1} = 0$. Similarly, the total EC for the IoT device k_2 is calculated as follows.

$$\xi_{k_2} = \sum_{m=1}^{|\mathcal{M}|} \left[(1 - \varphi_{m,k_2})\xi_{m,k_2}^{lo,s} + \varphi_{m,k_2}(1 - \varphi_{m-1,k_2})\xi_{m,k_2}^{es,s} \right]. \quad (4.21)$$

For the execution time of the IoT device k_2 , the waiting time is to reach the final output of the IoT device k_1 to the IoT device k_2 . It consists of the total execution time to compute $|\mathcal{L}|$ computational tasks from the IoT device k_1 and the transmission time to offload the final output of the device k_1 (refer to Fig. 4.1) is given in (4.22).

$$\begin{aligned}
 \xi_{k_1}^{hold} = & \overbrace{\sum_{l=1}^{|\mathcal{L}|} \left[(1 - \varphi_{l,k_1}) \eta_{l,k_1}^{lo,s} + \varphi_{l,k_1} (\tilde{\eta}_{l,k_1}^{es,s} + \eta_{l,k_1}^{es,s}) + \varphi_{l-1,k_1} \eta_{l,k_1}^{dl,s} - \varphi_{l-1,k_1} \varphi_{l,k_1} (\eta_{l,k_1}^{dl,s} + \eta_{l,k_1}^{es,s}) \right]}^{\text{Total amount of time for } |\mathcal{L}| \text{ tasks}} \\
 & + \overbrace{\left((1 - \varphi_{|\mathcal{L}|,k_1}) \eta_{|\mathcal{L}+1,k_1}^{es,s} + (1 - \varphi_{t,k_2}) \eta_{t',k_2}^{dl,s} \right)}^{\text{Transmission time of the output of } |\mathcal{L}| \text{ tasks by IoT device } k_1}.
 \end{aligned} \tag{4.22}$$

The time required for IoT device k_2 , until the output of the task $|t|-1$ is ready is expressed in (4.23).

$$\begin{aligned}
 \xi_{k_2}^{hold} = & \overbrace{\sum_{m=1}^{|t|-1} \left[(1 - \varphi_{m,k_2}) \eta_{m,k_2}^{lo,s} + \varphi_{m,k_2} (\tilde{\eta}_{m,k_2}^{es,s} + \eta_{m,k_2}^{es,s}) \right.}^{\text{Total execution time for first } |t|-1 \text{ tasks}} \\
 & \left. + \varphi_{m-1,k_2} \eta_{m,k_2}^{dl,s} - \varphi_{m-1,k_2} \varphi_{m,k_2} (\eta_{m,k_2}^{dl,s} + \eta_{m,k_2}^{es,s}) \right]} \\
 & + \overbrace{\left(\varphi_{t,k_2} \eta_{t,k_2}^{es,s} + \varphi_{t-1,k_2} \eta_{t,k_2}^{dl,s} - \varphi_{t-1,k_2} \varphi_{t,k_2} (\eta_{t,k_2}^{dl,s} + \eta_{t,k_2}^{es,s}) \right)}^{\substack{\text{Trans. time to offload task } |t|, \text{ i.e., } \varphi_{t-1,k_2}=0, \varphi_{t,k_2}=1 \\ \text{or to download output for } |t|-1 \text{ task to device } k_2, \text{ i.e., } \varphi_{t-1,k_2}=1, \varphi_{t,k_2}=0}}
 \end{aligned} \tag{4.23}$$

Using (4.22) and (4.23), the total time required before the t^{th} task of device k_2 is ready to execute is given by $\xi^{hold} = \max\{\xi_{k_1}^{hold}, \xi_{k_2}^{hold}\}$. The total execution time of device k_2 is formulated by adding ξ^{hold} and the time required to complete the tasks from t to $|\mathcal{M}|$ is given in (4.24).

$$\begin{aligned}
 \xi_{k_2} = & \xi^{hold} + \sum_{m=t}^{|\mathcal{M}|} \left[(1 - \varphi_{m,k_2}) \eta_{m,k_2}^{lo,s} + \varphi_{m,k_2} (\tilde{\eta}_{m,k_2}^{es,s}) \right] \\
 & + \sum_{m=t+1}^{|\mathcal{M}|+1} \left[\varphi_{m,k_2} \eta_{m,k_2}^{es,s} + \varphi_{m-1,k_2} \eta_{m,k_2}^{dl,s} - \varphi_{m-1,k_2} \varphi_{m,k_2} (\eta_{m,k_2}^{dl,s} + \eta_{m,k_2}^{es,s}) \right].
 \end{aligned} \tag{4.24}$$

This study aims to reduce the energy-time cost for URLLC edge networks. Such a reduction is achieved by optimizing the offloading policy and allocating the resources, e.g., transmission power, to offload the task and CPU frequency. Let $0 \leq \partial_{k_1}^{\xi} \leq 1$

and $0 \leq \partial_{k_1}^\lambda \leq 1$ be the weighting factors for the EC and the execution time of the given device k_1 . In this context, if $\partial_{k_1}^\xi + \partial_{k_1}^\lambda = 1$ then $\partial_{k_1}^\xi = 1 - \partial_{k_1}^\lambda$. Following real-time requirements, each IoT device can select weights (higher or lower) to fulfill user-oriented needs. So, the energy-time cost for the device k_1 is given by $\zeta_{k_1} = \partial_{k_1}^\xi \xi_{k_1} + \partial_{k_1}^\lambda \lambda_{k_1}^{\text{tot}}$. Let $0 \leq \partial_{k_2}^\xi \leq 1$ and $0 \leq \partial_{k_2}^\lambda \leq 1$ be the weighting factors for the EC and the execution time of the given device k_2 . In this context, if $\partial_{k_2}^\xi + \partial_{k_2}^\lambda = 1$ then $\partial_{k_2}^\xi = 1 - \partial_{k_2}^\lambda$. The energy-time cost for the device k_2 is given by $\zeta_{k_2} = \partial_{k_2}^\xi \xi_{k_2} + \partial_{k_2}^\lambda \lambda_{k_2}^{\text{tot}}$. We assume $\boldsymbol{\varphi} \triangleq \{\varphi_{i,k}\}_{\forall i,k}$, $\mathbf{p} \triangleq \{p_{i,k}\}_{\forall i,k}$, and $\mathbf{f} \triangleq \{f_{i,k}^{\text{lo}}\}_{\forall i,k}$ as the set constraints of the computational offloading decision, transmit power, and CPU frequency, respectively. Therefore, the problem is formulated below.

$$\begin{aligned}
 \text{(P1)} \quad & \min_{\boldsymbol{\varphi}, \mathbf{p}, \mathbf{f}} \sum_{k \in \mathcal{K}} \zeta_k \\
 \text{s.t.} \quad & C_1 : p_{i,k} \leq p_{\max}, \forall i, k, \\
 & C_2 : f_{i,k}^{\text{lo}} \leq f_{\max}^{\text{lo}}, \forall i, k, \\
 & C_3 : f_{i,k}^{\text{es}} \leq f_{\max}^{\text{es}}, \forall i, k, \\
 & C_4 : \sum_{k \in \mathcal{K}} \pi_{k,a} \leq \chi_{\max}, \forall i, k, \\
 & C_5 : \lambda_k^{\text{tot}} \leq \lambda_k^{\max}, \forall k, \\
 & C_6 : \xi_k \leq \xi_k^{\max}, \forall k, \\
 & C_7 : r_{i,k}^{\text{dl},s} \geq r_{\min}, \forall i, k, \\
 & C_8 : \varphi_{i,k} \in (0, 1), \forall i, k, \\
 & C_9 : 0 \leq i \leq t_k, 1 \leq k \leq 2.
 \end{aligned} \tag{4.25}$$

The constraints are described as follows: constraint C_1 denotes the upper limit of transmission power, denoted by p_{\max} . The restriction C_2 limits the maximum CPU frequency of the specified device to f_{\max}^{lo} . The restriction C_3 restricts the computational resources to f_{\max}^{es} . The constraint C_4 restricts the number of IoT devices that each edge server can serve at χ_{\max} . The constraint C_5 sets the maximum latency threshold at λ_k^{\max} . The constraint C_6 specifies the maximum energy threshold as ξ_k^{\max} . The constraint C_7 establishes the minimum rate requirement, represented as r_{\min} . The constraint C_8 characterizes the computational offloading decision. The constraint C_9 specifies the number of tasks executed on the k^{th} device.

4.4 Proposed Solution

The problem outlined is a MINLP combinatory optimization problem, which is computationally intractable. It is due to the combinatorial nature of binary variable φ and strong coupling with other continuous optimization variables, e.g., transmit power and CPU frequency. This problem is challenging to solve directly using an exhaustive search, especially in a dense network. It is important to note that there is a direct relationship between CPU frequency and local computational time, as indicated in (4.8). Similarly, there is a direct relationship between power and transmission time while offloading, as indicated in (4.15). Therefore, optimizing the problem concerning the power and CPU frequency is equivalent to optimizing it over the time allocation variables, i.e., $\{\eta_{i,k}^{\text{lo},s}\}$, $\{\eta_{i,k}^{\text{es},s}\}$, respectively. This inequality implies that if it is understandable how the time allocation variables can be optimized, we can infer the optimal power and CPU frequency, simplifying the original problem. By introducing a temporary variable $\xi_t = \max\{\zeta_{k_1}^{\text{hold}} + \zeta_{k_2}^{\text{hold}}\}$, we equivalently transform (4.25) to

$$\begin{aligned}
 \text{(P2)} \quad & \min_{\varphi, \{\eta_{i,k}^{\text{lo},s}\}, \{\eta_{i,k}^{\text{es},s}\}, \xi_t} \sum_{k \in \mathcal{K}} \zeta_k \\
 & \text{s.t.} \quad C_1 - C_9, \\
 & C_{10} : \xi_t \geq \zeta_{k_1}^{\text{hold}}, \xi_t \geq \zeta_{k_2}^{\text{hold}}, \\
 & C_{11} : \eta_{i,k}^{\text{lo},s} \geq \frac{\delta_{i,k}}{f_{\max}^{\text{lo}}}, \\
 & C_{12} : \eta_{i,k}^{\text{es},s} \geq \frac{\beta_{i-1,k}}{r_{i,k}^{\text{es},s}}, \\
 & \text{where } p_{i,k}^s = p_{\max}.
 \end{aligned} \tag{4.26}$$

Constraint C_{10} establishes an upper bound for ξ_t . This equation implies that if the optimal solution $\{\varphi^*, \{\eta_{i,k}^{\text{lo},s}\}^*, \{\eta_{i,k}^{\text{es},s}\}^*\}$ of (4.26), then we can easily compute power and CPU frequency in (4.25). The problem (4.26) is non-convex due to the combinatorial nature of binary variable φ . For given φ , the remaining optimization over $(\{\eta_{i,k}^{\text{lo},s}\}, \{\eta_{i,k}^{\text{es},s}\}, \xi_t)$ becomes a convex problem.

4.4.1 Optimal Transmit Power and Local CPU Frequency for given φ

We solve the above problem for the given offloading decisions to compute a closed-form solution of optimal transmit power and local CPU frequency. We approximate the objective function as an unbounded optimization problem by utilizing the concept of partial Lagrangian subject to constraint C_{10} .

$$L_p \left(\{\eta_{i,k}^{\text{lo},s}\}, \{\eta_{i,k}^{\text{es},s}\}, \xi_t, \mu_{k_1}, \mu_{k_2} \right) = \zeta_{k_1} + \zeta_{k_2} + \mu_{k_1} (\xi_{k_1}^{\text{hold}} - \xi_t) + \mu_{k_2} (\xi_{k_2}^{\text{hold}} - \xi_t), \quad (4.27)$$

$\mu_{k_1} \geq 0$ and $\mu_{k_2} \geq 0$ are the Lagrange multipliers computed to satisfy all remaining constraints. If $\mu_{k_1}^{\text{opt}}$ and $\mu_{k_2}^{\text{opt}}$ represent the optimal values for Lagrangian's multiplier, we derive the analytical expressions to compute the optimal power and CPU frequencies for both IoT devices as in [22]. To analyze device k_1 , the derivative of (4.27) with respect to $\eta_{i,k_1}^{\text{lo},s}$ is computed as follows.

$$L'_p = \frac{\partial L_p}{\partial \eta_{i,k_1}^{\text{lo},s}}, \forall i, k, \quad (4.28)$$

$$\frac{\partial L_p}{\partial \eta_{i,k_1}^{\text{lo},s}} = \partial_{k_1}^\lambda - \frac{2\mu \partial_{k_1}^\xi (\delta_{i,k_1})^3}{(\eta_{i,k_1}^{\text{lo},s})^3} + \mu_{k_1}, \forall i, k, \quad (4.29)$$

where the above derivation indicates that it is a non-decreasing function for $\eta_{i,k_1}^{\text{lo},s} \in \left[\frac{\delta_{i,k_1}}{f_{\text{max}}^{\text{lo}}}, +\infty \right)$. Hence, if the first derivative is positive, then we have $f_{i,k_1}^{\text{lo},\text{opt}} = f_{\text{max}}^{\text{lo}}$. Otherwise,

$$\mu_{k_1} + \partial_{k_1}^\lambda = \frac{2\mu \partial_{k_1}^\xi (\delta_{i,k_1})^3}{(\eta_{i,k_1}^{\text{lo},s})^3}, \forall i, \quad (4.30)$$

$$(\eta_{i,k_1}^{\text{lo},s})^3 = \frac{2\mu \partial_{k_1}^\xi (\delta_{i,k_1})^3}{\mu_{k_1} + \partial_{k_1}^\lambda}, \forall i, \quad (4.31)$$

$$\eta_{i,k_1}^{\text{lo},s} = \delta_{i,k_1} \left(\frac{2\mu \partial_{k_1}^\xi}{\mu_{k_1} + \partial_{k_1}^\lambda} \right)^{\frac{1}{3}}, \forall i. \quad (4.32)$$

$$f_{i,k_1}^{\text{lo},\text{opt}} \Rightarrow \frac{\delta_{i,k_1}}{\eta_{i,k_1}^{\text{lo},s}} = \left(\frac{\mu_{k_1}^{\text{opt}} + \partial_{k_1}^\lambda}{2\mu \partial_{k_1}^\xi} \right)^{\frac{1}{3}}, \forall i. \quad (4.33)$$

The optimal CPU frequencies are equivalent to the frequencies computed as follows.

If $(\varphi_{i,k})_{\forall i,k} = 0$ and $i \in l$ then

$$f_{i,k_1}^{\text{lo,opt}} = \min \left(\sqrt[3]{\frac{\mu_{k_1}^{\text{opt}} + \partial_{k_1}^\lambda}{2\mu\partial_{k_1}^\xi}}, f_{\text{max}}^{\text{lo}} \right), \quad (4.34)$$

for IoT device k_1 . If $i \in \{1, 2, \dots, |t|-1\}$ then

$$f_{i,k_2}^{\text{lo,opt}} = \min \left(\sqrt[3]{\frac{\mu_{k_2}^{\text{opt}}}{2\mu\partial_{k_2}^\xi}}, f_{\text{max}}^{\text{lo}} \right), \quad (4.35)$$

for IoT device k_2 . Otherwise, for $i \in \{|t|, \dots, |M|\}$, we have

$$f_{i,k_2}^{\text{lo,opt}} = \min \left(\sqrt[3]{\frac{\partial_{k_2}^\lambda}{2\mu\partial_{k_2}^\xi}}, f_{\text{max}}^{\text{lo}} \right). \quad (4.36)$$

Hence, we have the following observations:

- The optimal CPU frequencies for all similar tasks are identical. Specifically for device k_1 , $i \in \{1, 2, \dots, |L|\}$. For device k_2 , the relevant tasks are either $i \in \{1, 2, \dots, |t|-1\}$ or $i \in \{|t|, \dots, |M|\}$, regardless of channel conditions.
- When the value of $\partial_{k_1}^\lambda$ or $\mu_{k_1}^{\text{opt}}$ is higher, the optimal strategy shifts towards accelerating local computations. Conversely, if $\partial_{k_1}^\xi$ is higher, the optimal strategy shifts involve conserving energy with a lower optimal CPU frequency.
- The value of $\mu_{k_2}^{\text{opt}}$ does not affect the optimal CPU frequency for tasks where $i \in \{|t|, \dots, |M|\}$. Otherwise, the optimal CPU frequency increases proportionally with the value of the Lagrange multiplier.

The optimal power for IoT device k_1 is calculated by computing the derivative of the problem (4.27) with respect to $\eta_{i,k_1}^{\text{es,s}}$, when $i \leq |M|$, which is given by

$$\frac{\partial L'_p}{\partial \eta_{i,k_1}^{\text{es,s}}} = \partial_{k_1}^\lambda + \partial_{k_1}^\xi \left[\frac{1}{\|h_{i,k_1}^s\|^2} f \left(\frac{\beta_{i-1,k_1}}{\eta_{i,k_1}^{\text{es,s}}} \right) + \frac{\eta_{i,k_1}^{\text{es,s}}}{\|h_{i,k_1}^s\|^2} f' \left(\frac{\beta_{i-1,k_1}}{\eta_{i,k_1}^{\text{es,s}}} \right) \right] + \mu_{k_1}, \quad (4.37)$$

$$\frac{\partial L'_p}{\partial \eta_{i,k_1}^{\text{es,s}}} = \partial_{k_1}^\lambda + \partial_{k_1}^\xi \left[\frac{\bar{\sigma}^2}{\|h_{i,k_1}^s\|^2} 2 \left(\frac{\beta_{i-1,k_1}}{w_s \eta_{i,k_1}^{\text{es,s}}} \right) \left(1 - \frac{\beta_{i-1,k_1}}{w_s \eta_{i,k_1}^{\text{es,s}}} \ln 2 \right) - \frac{\bar{\sigma}^2}{\|h_{i,k_1}^s\|^2} \right] + \mu_{k_1}. \quad (4.38)$$

It is worth mentioning that the above function is an increasing function. To check the curvature, the second order derivative for the problem (4.27) concerning $\eta_{i,k_1}^{\text{es},s}$ is given by

$$\frac{\partial L_p''}{\partial (\eta_{i,k_1}^{\text{es},s})^2} = \partial_{k_1}^\xi \left[\frac{\bar{\sigma}^2}{\|h_{i,k_1}^s\|^2} \frac{(\beta_{i-1,k_1})^2}{(w_s)^2 (\eta_{i,k_1}^{\text{es},s})^3} 2^{\left(\frac{\beta_{i-1,k_1}}{w_s \eta_{i,k_1}^{\text{es},s}}\right)} (\ln 2)^2 \right] + \mu_{k_1} > 0, \quad (4.39)$$

which shows that the function is concave-up and non-decreasing for $\eta_{i,k_1}^{\text{es},s} \in \left[\frac{\beta_{i-1,k_1}}{r_{i,k_1}^{\text{es},s}}, +\infty \right)$, where where $p_{i,k_1}^s = p_{\max}$. Therefore, we have $\frac{\partial L_p'}{\partial \eta_{i,k_1}^{\text{es},s}} \in \left[\frac{\beta_{i-1,k_1}}{r_{i,k_1}^{\text{es},s}}, \partial_{k_1}^\lambda + \mu_{k_1} \right]$ if $\frac{\partial L_p'}{\partial \eta_{i,k_1}^{\text{es},s}} > 0$, when $p_{i,k_1}^s = p_{\max}$. The optimal transmit powers are equivalent to the powers computed below. If $(\varphi_{i,k})_{\forall i,k} = 1$, the optimal power for device k_1 is given in (4.40).

$$\left\{ \begin{array}{l} \text{if } i \in \{1, \dots, |L|\}, p_{i,k_1}^{s,\text{opt}} = \begin{cases} p_{\max}, & \|h_{i,k_1}^s\|^2 < v_{i,k_1}^{\text{threshold}}, \\ \frac{\bar{\sigma}^2}{\|h_{i,k_1}^s\|^2} \left[\frac{\psi_1}{\omega(\psi_1 e^{-1})} - 1 \right], \\ \psi_1 = \left[\frac{\|h_{i,k_1}^s\|^2 (\partial_{k_1}^\lambda + \mu_{k_1}^{\text{opt}})}{\partial_{k_1}^\xi \bar{\sigma}^2} \right] - 1, & \text{otherwise.} \end{cases} \\ \text{if } i = |L|+1, p_{i,k_1}^{s,\text{opt}} = \begin{cases} p_{\max}, & \|h_{i,k_1}^s\|^2 < v_{i,k_1}^{\text{threshold}}, \\ \frac{\bar{\sigma}^2}{\|h_{i,k_1}^s\|^2} \left[\frac{\psi_2}{\omega(\psi_2 e^{-1})} - 1 \right], \\ \psi_2 = \left[\frac{\|h_{i,k_1}^s\|^2 \mu_{k_1}^{\text{opt}}}{\partial_{k_1}^\xi \bar{\sigma}^2} \right] - 1, & \text{otherwise.} \end{cases} \end{array} \right. \quad (4.40)$$

where $\omega(\psi_1)$ and $\omega(\psi_2)$ denote special inverse functions. These functions are defined as the inverse of the given function $y = f(x) = x e^x$, where x is expressed as $x = \omega(y)$. The computation of threshold values is performed as follows. For $i \in \{1, \dots, |L|\}$:

$$v_{i,k_1}^{\text{threshold}} = \frac{\bar{\sigma}^2}{p_{\max}} \left[\frac{\nu_1}{\omega(-\nu_1 e^{-\nu_1})} - 1 \right], \text{ where } \nu_1 = 1 + \left(\frac{\partial_{k_1}^\lambda + \mu_{k_1}^{\text{opt}}}{\partial_{k_1}^\xi p_{\max}} \right). \quad (4.41)$$

For $i = |L|+1$:

$$v_{i,k_1}^{\text{threshold}} = \frac{\bar{\sigma}^2}{p_{\max}} \left[\frac{\nu_2}{\omega(-\nu_2 e^{-\nu_2})} - 1 \right], \text{ where } \nu_2 = 1 + \left(\frac{\mu_{k_1}^{\text{opt}}}{\partial_{k_1}^\xi p_{\max}} \right). \quad (4.42)$$

We compute the optimal power for the IoT device k_2 in (4.43).

$$\left\{ \begin{array}{l} \text{if } i \in \{1, \dots, |t|\}, p_{i,k_2}^{s,\text{opt}} = \begin{cases} p_{\max}, & \|h_{i,k_2}^s\|^2 < v_{i,k_2}^{\text{threshold}}, \\ p_{i,k_2}^{s,\text{opt}} = \frac{\bar{\sigma}^2}{\|h_{i,k_2}^s\|^2} \left[\frac{\psi_3}{\omega(\psi_3 e^{-1})} - 1 \right], \\ \psi_3 = \left[\frac{\|h_{i,k_2}^s\|^2 \mu_{k_2}^{\text{opt}}}{\partial_{k_2}^\xi \bar{\sigma}^2} \right] - 1, & \text{otherwise.} \end{cases} \\ \text{if } i = \{|t|+1, \dots, |M|\}, p_{i,k_2}^{s,\text{opt}} = \begin{cases} p_{\max}, & \|h_{i,k_2}^s\|^2 < v_{i,k_2}^{\text{threshold}} \\ p_{i,k_2}^{s,\text{opt}} = \frac{\bar{\sigma}^2}{\|h_{i,k_2}^s\|^2} \left[\frac{\psi_4}{\omega(\psi_4 e^{-1})} - 1 \right], \\ \psi_4 = \left[\frac{\|h_{i,k_2}^s\|^2 \partial_{k_2}^\lambda}{\partial_{k_2}^\xi \bar{\sigma}^2} \right] - 1, & \text{otherwise,} \end{cases} \end{array} \right. \quad (4.43)$$

where $\omega(\psi_3)$ and $\omega(\psi_4)$ represent special functions, as defined above. We compute the threshold values as follows. For $i \in \{1, \dots, |t|\}$:

$$v_{i,k_2}^{\text{threshold}} = \frac{\bar{\sigma}^2}{p_{\max}} \left[\frac{\nu_3}{\omega(-\nu_3 e^{-\nu_3})} - 1 \right], \text{ where } \nu_3 = 1 + \left(\frac{\mu_{k_2}^{\text{opt}}}{\partial_{k_2}^\xi p_{\max}} \right). \quad (4.44)$$

For $i = \{|t|+1, \dots, |M|\}$:

$$v_{i,k_2}^{\text{threshold}} = \frac{\bar{\sigma}^2}{p_{\max}} \left[\frac{\nu_4}{\omega(-\nu_4 e^{-\nu_4})} - 1 \right], \text{ where } \nu_4 = 1 + \left(\frac{\partial_{k_2}^\lambda}{\partial_{k_2}^\xi p_{\max}} \right). \quad (4.45)$$

Hence, we have the following observations:

- The channel gain $\|h_{i,k_1}^s\|^2$ is inversely proportional to the optimal transmit power when $\|h_{i,k_1}^s\|^2 > v_{i,k_1}^{\text{threshold}}$.
- The transmission power has been set to the maximum power, i.e., $p_{i,k_1}^{s,\text{opt}} = p_{\max}$ when $\|h_{i,k_1}^s\|^2 < v_{i,k_1}^{\text{threshold}}$.
- When the maximum transmit power p_{\max} increases, it leads to a decrease in the squared magnitude of the channel gain h_{i,k_1}^s , indicating that the device is adapting to meet the conditions of a weaker channel.

Algorithm 4.1 Bisection Method to Compute Optimal Power and Local Frequency

-
- 1: **Input:** φ , precision factor $\epsilon = 0.001$.
 - 2: **Output:** Optimal \mathbf{f} and \mathbf{p}
 - 3: Initialize $\varphi'_{ub} \leftarrow \partial_{k_2}^\lambda$, $\varphi'_{lb} \leftarrow 0$
 - 4: **if** $(\xi_{k_1}^{hold} - \xi_{k_2}^{hold})|_{\varphi'=\varphi'_{lb}} < 0$ **then**
 - 5: Set $\varphi' = \varphi'_{lb}$, $\mu_{k_1} = \varphi'$, $\mu_{k_2} = \partial_{k_2}^\lambda - \varphi'$
 - 6: Compute \mathbf{f} using (4.33) and (4.36), and \mathbf{p} using (4.40) and (4.43)
 - 7: **else**
 - 8: **while** $(\xi_{k_1}^{hold} - \xi_{k_2}^{hold}) < \epsilon$ **do**
 - 9: $\varphi' = (\varphi'_{ub} + \varphi'_{lb}) / 2$
 - 10: Set $\mu_{k_1} = \varphi'$, $\mu_{k_2} = \partial_{k_2}^\lambda - \varphi'$
 - 11: Compute \mathbf{f} using (4.33) and (4.36), and \mathbf{p} using (4.40) and (4.43)
 - 12: **if** $(\xi_{k_1}^{hold} - \xi_{k_2}^{hold}) < 0$ **then**
 - 13: Set $\varphi'_{ub} \leftarrow \varphi'$
 - 14: **else**
 - 15: Set $\varphi'_{lb} \leftarrow \varphi'$
 - 16: **end if**
 - 17: **end while**
 - 18: **end if**
-

4.4.2 Bisection Method for given φ to Obtain Optimal Power \mathbf{p} and Frequency \mathbf{f}

We can state that $\xi_{k_1}^{hold} \leq \xi_{k_2}^{hold}$ and $\mu_{k_2}^{\text{opt}}$ hold at the optimum of (4.26). It can be proved through contradiction. Assume that $\{\eta_{i,k}^{\text{lo},s}, \tilde{\eta}_{i,k}^{\text{es},s}\}$ are the optimal solutions with $\xi_{k_1}^{hold} > \xi_{k_2}^{hold}$. According to Karush Kuhn Tucker (KKT) conditions $\mu_{k_1}^{\text{opt}}(\xi_{k_1}^{hold} - \xi_t) = 0$, and $\mu_{k_2}^{\text{opt}}(\xi_{k_2}^{hold} - \xi_t) = 0$. Given that $\mu_{k_1}^{\text{opt}} > 0$ and $\mu_{k_2}^{\text{opt}} = 0$, it follows from Eq. (4.33) and (4.40) that the optimal CPU frequency $f_{i,k_1}^{\text{lo},\text{opt}}$ and $p_{i,k_1}^{s,\text{opt}}$ are finite. This implies that $\{(\eta_{i,k_1}^{\text{lo},s})^*, (\tilde{\eta}_{i,k_1}^{\text{es},s})^*\}$ are also finite for each task, resulting in finite $\xi_{k_1}^{hold}$. However, when $\mu_{k_2}^{\text{opt}} = 0$, we will have the optimal value

$(\eta_{i,k_2}^{\text{lo},s})^* \Rightarrow \infty$, i.e., for $i \in \{1, 2, 3, \dots, |t|-1\}$ from $f_{i,k_2}^{\text{lo,opt}}$. Similarly, it follows from Eq. (4.43) the optimal value $(\tilde{\eta}_{i,k_2}^{\text{es},s})^* \Rightarrow \infty$, i.e., for $i \in \{1, 2, 3, \dots, |t|\}$, resulting in infinite $\xi_{k_1}^{\text{hold}}$. This contradicts our initial assumption of $\xi_{k_1}^{\text{hold}} > \xi_{k_2}^{\text{hold}}$. Hence, we have $\xi_{k_1}^{\text{hold}} \leq \xi_{k_2}^{\text{hold}}$, and $\max(\xi_{k_1}^{\text{hold}}, \xi_{k_2}^{\text{hold}}) = \xi_{k_2}^{\text{hold}}$. Therefore, the optimization problem (4.28) is simplified as follows.

$$\begin{aligned}
 \text{(P3)} \quad & \min_{\boldsymbol{\varphi}, \mathbf{p}, \mathbf{f}} \sum_{k \in \mathcal{K}} \zeta_k \\
 \text{s.t.} \quad & C_1 - C_9 \\
 & C_{13} : \xi_{k_1}^{\text{hold}} \leq \xi_{k_2}^{\text{hold}}.
 \end{aligned} \tag{4.46}$$

Similarly, the Lagrangian for the problem defined in Eq. (4.46) subject to constraint C_{13} is given by

$$L_2^p(\mathbf{p}, \mathbf{f}, \wp') = \zeta_{k_1} + \zeta_{k_2} + \wp'(\xi_{k_1}^{\text{hold}} - \xi_{k_2}^{\text{hold}}), \tag{4.47}$$

whereas $\wp' \geq 0$ represents the Lagrange multiplier that satisfies all remaining constraints. Various iterative solutions can be applied to solve Eq. (4.47), but we apply the KKT conditions [22]. The details are omitted here. Algorithm (4.1) presents the bisection search method to compute optimal power and frequency. This algorithm accepts the matrix $\boldsymbol{\varphi}$ and a precision factor ϵ as input and produces the optimal values of \mathbf{f} and \mathbf{p} as outputs. In line 3, the upper and lower bounds are initialized. Then, the constrained inequality C_{13} is checked, and the values for Lagrange multipliers are updated at line 5. The initial values for frequency and power are computed in line 6. The loop then iterates until the condition $(\xi_{k_1}^{\text{hold}} - \xi_{k_2}^{\text{hold}}) < \epsilon$ is satisfied, as outlined in lines 8-17. The value \wp' is updated on line 9 during this iteration. Line 11 involves computing the optimal frequency and power. Importantly, from lines 12 to 16, if the difference $\xi_{k_1}^{\text{hold}}$ and $\xi_{k_2}^{\text{hold}} < 0$ results in a negative value, the algorithm updates the upper bound of the variable, designated as \wp'_{ub} . In contrast, the algorithm converges towards the lower bound variable, \wp'_{lb} . The overall complexity of the Algorithm (4.1) is $\mathcal{O}(L + M)$.

Algorithm 4.2 Proposed BSSE Algorithm

```

1: Input: Maximum iterations ( $i_{\max}$ ), Number of task of each device  $(t_k)_{\forall k \in \{k_1, k_2\}}$ .
2: Output: Matrix  $\varphi$ 
3: Set  $t \leftarrow 0$ ,  $\zeta_{k_1} \leftarrow \infty$ , and  $\zeta_{k_2} \leftarrow \infty$ 
4: We initialize  $\varphi$  randomly either with the value 0 or 1
5: while Convergence or  $t < i_{\max}$  do
6:   Set  $i \leftarrow 0$ 
7:   while  $i < t_{k_1}$  do
8:     Set flag  $\leftarrow$  True
9:      $j \leftarrow 0$ 
10:    while  $j < t_{k_2}$  do
11:      if flag == False then
12:        break
13:      end if
14:       $\varphi' \leftarrow \varphi$ 
15:      Update  $(\varphi, i, j) \leftarrow$  Swap  $\varphi[i, k_1]$  and  $\varphi[j, k_2]$ 
16:      Compute  $\mathbf{f}$  and  $\mathbf{p}$  using Algorithm (4.1)
17:      if  $(\zeta_{k_1}^{\text{new}} + \zeta_{k_2}^{\text{new}}) \leq (\zeta_{k_1} + \zeta_{k_2})$  then
18:        Set  $\zeta_{k_1} \leftarrow \zeta_{k_1}^{\text{new}}$ 
19:        Set  $\zeta_{k_2} \leftarrow \zeta_{k_2}^{\text{new}}$ 
20:        flag  $\leftarrow$  False
21:      else
22:         $\varphi \leftarrow \varphi'$ 
23:      end if
24:      Set  $j \leftarrow j + 1$ 
25:    end while
26:    Set  $i \leftarrow i + 1$ 
27:  end while
28:  Set  $t \leftarrow t + 1$ 
29: end while

```

4.4.3 Optimize Offloading Decision φ using given p and f

The optimal offloading scheme adheres to the one climb policy [22], implying that there will be at most one instance where $\varphi_{i,k} = 1$ and $\varphi_{i-1,k} = 0$. The proposed algorithmic solution is given in Algorithm (4.2). Initially, we set the maximum iteration count to zero and the energy-time cost for both IoT devices to infinity, denoted as $\zeta_{k_1} = \zeta_{k_2} = \infty$ in line 3. Afterward, the proposed algorithm initiates with a random offloading scheme and progressively refines it toward minimizing the overall energy-time cost, as detailed in lines 5 to 29. The process includes swapping $\varphi[i, k_1]$ and $\varphi[j, k_2]$ within the matrix φ and updating the tuple (φ, i, j) at lines 14 and 15. Line 16 invokes the bisection algorithm, which computes optimal frequency and power using equations (4.33), (4.36), (4.40), and (4.43), respectively. The results of the bisection search are then evaluated and compared with previously optimized energy-time cost values, as shown in lines 17-23. If $(\zeta_{k_1}^{\text{new}} + \zeta_{k_2}^{\text{new}}) \leq (\zeta_{k_1} + \zeta_{k_2})$ is true, then $\zeta_{k_1} \leftarrow \zeta_{k_1}^{\text{new}}$ and $\zeta_{k_2} \leftarrow \zeta_{k_2}^{\text{new}}$; otherwise, the matrix φ reverts to φ' . This process iterates until the solution converges at optimized values.

4.4.4 Complexity Analysis

The first benchmark is the naive search algorithm corresponding to the complexity of $\mathcal{O}(16^{L+M})$, since it considers each possible case for both devices with L and M tasks, i.e., 2^4 binary decisions. This algorithm becomes increasingly inefficient with a more significant number of tasks, especially when $|L| \approx 10$ and $|M| \approx 10$. To mitigate this, we effectively narrow the search space for both devices, that is, $\left(\frac{(L+1)L}{2} + 1\right) \left(\frac{(M+1)M}{2} + 1\right)$. This includes the scenario in which the device cannot offload data during execution. This approach yields an optimal offloading scheme with a complexity of $\mathcal{O}(L^2M^2)$. However, its efficiency diminishes for large values, precisely when $|L| > 100$ and $|M| > 100$. Upon examining the pseudocode provided, it becomes evident that the overall complexity of the BSSE algorithm is $\mathcal{O}(LM)$, significantly lower than the benchmarks.

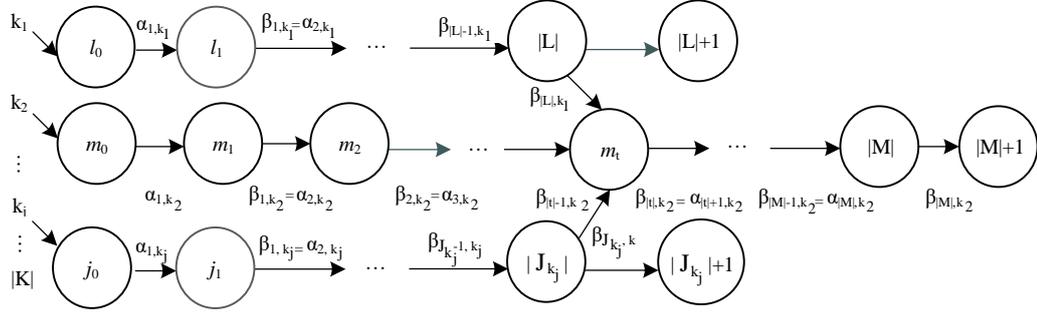


Figure 4.2: Task Dependent Computational Model Incorporating Outputs from Multiple Devices.

4.5 Multi-Device Scenario

In Fig. 4.2, we extend the proposed model to support multiple devices. In this scenario, the inputs for an intermediate task t on IoT device k_2 require the final task outputs from all other $|K|-1$ devices. For example, for the IoT device k_2 , we define $\alpha_{t,k_2} = \beta_{|t|-1,k_2} + \sum_{k \in \mathcal{K}, k_j \neq k_2} \beta_{J_{k_j},k_j}$, where J_{k_j} denotes the sequential number of tasks that must be executed on the device k_j . The first terms on the right side of the equation represent the output of the previous task $|t|-1$. Meanwhile, the summation term adds the output of the final tasks of all other devices (excluding k_2), which are required for the intermediate task t . The waiting time for the output of the J_{k_j} -th task to reach the IoT device k_2 is calculated in (4.48).

$$\begin{aligned}
 \xi_{|k_j|}^{hold} &= \overbrace{\sum_{j=1}^{J_{k_j}} \left[(1 - \varphi_{j,k_j}) \eta_{j,k_j}^{lo,s} + \varphi_{j,k_j} (\tilde{\eta}_{j,k_j}^{es,s} + \eta_{j,k_j}^{es,s}) + \varphi_{j-1,k_j} \eta_{j,k_j}^{dl,s} - \varphi_{j-1,k_j} \varphi_{j,k_j} (\eta_{j,k_j}^{dl,s} + \eta_{j,k_j}^{es,s}) \right]}^{\text{Total amount of time for } J_{k_j} \text{ task}} \\
 &\quad + \overbrace{\left(1 - \varphi_{J_{k_j},k_j} \right) \eta_{J_{k_j}+1,k_j}^{es,s} + (1 - \varphi_{t,k_2}) \frac{\beta_{J_{k_j},k_j}}{r_{t,k_2}^{dl,s}}}_{\text{Transmission time of the output of } |J| \text{ tasks by IoT device } k_j}. \quad (4.48)
 \end{aligned}$$

The waiting time needed before the joint task can be executed is determined by the total waiting time $\xi^{hold} = \max\{\xi_{k_1}^{hold}, \xi_{k_2}^{hold}, \dots, \xi_{|k_j|}^{hold}, \dots, \xi_{|K|}^{hold}\}$. Therefore, the

problem (4.26) is reformulated as follows.

$$\begin{aligned}
 \text{(P4)} \quad & \min_{\boldsymbol{\varphi}, \{\eta_{i,k_j}^{\text{lo},s}\}, \{\eta_{i,k_j}^{\text{es},s}\}, \boldsymbol{\xi}_t} \sum_{k \in \mathcal{K}} \zeta_k \\
 & \text{subject to} \\
 & C_1 - C_9 \\
 & C_{10} : \xi_t \geq \xi_{k_1}^{\text{hold}}, \xi_t \geq \xi_{k_2}^{\text{hold}}, \dots, \xi_t \geq \xi_{k_j}^{\text{hold}}, \dots, \xi_t \geq \xi_{|\mathcal{K}|}^{\text{hold}}, \\
 & C_{11} : \eta_{i,k_j}^{\text{lo},s} \geq \frac{\delta_{i,k_j}}{f_{i,k_j}^{\text{lo}} - \hat{f}_{i,k_j}^{\text{lo}}}, \\
 & \text{where } f_{i,k_j}^{\text{lo}} = f_{i,k_j}^{\text{lo},\text{max}}, \\
 & C_{12} : \eta_{i,k_j}^{\text{es},s} \geq \frac{\beta_{i-1,k_j}}{r_{i,k_j}^{\text{es},s}}, \\
 & \text{where } p_{i,k_j}^s = p_{\text{max}}.
 \end{aligned} \tag{4.49}$$

We assume that the edge server is equipped with c_r processor cores, where each core is exclusively assigned to the execution of an individual task operating at a constant frequency of $f_{\text{max}}^{\text{es}}$. Therefore, if J_{max} represents the upper limit on the number of tasks that can be processed simultaneously on the edge server, the total count of cores must be adequate to meet this requirement, for example, $J_{\text{max}} \leq c_r$.

4.6 Simulation and Results

4.6.1 Simulations Parameters

Regarding communication considerations, the IoT devices are uniformly distributed within the ABS coverage area, which includes a radius ranging from 10 to 30 meters. The computing requirements for the corresponding IoT devices are represented by the values $\{\delta_{i,k_1}\} = [65.5, 40.3, 96.6]$ (Mcycles) and $\{\delta_{i,k_2}\} = [70.8, 95.3, 86.4, 18.6, 158.6]$ (Mcycles). The total power budget for the MEC-enabled ABS under consideration is set at $(p_a)_{a \in \mathcal{A}} = 1$ W. The maximum transmit power of each IoT device is capped at $p_{\text{max}} = 200$ mW. It is assumed that the input of the fourth task at IoT device k_2 depends on the final output of the task from the other

Parameters	Values
Blocklength value $(b_{i,k})_{k \in \{k_1, K_2\}}$ [10]	100
Computing workload for device k_1 $\{\delta_{i,k_1}\}$ [22]	[65.5,40.3,96.6] (Mcycles)
Computing workload for device k_2 $\{\delta_{i,k_2}\}$ [22]	[70.8,95.3,86.4, 18.6,158.6] (Mcycles)
Distances of the respective devices (d_{k_1}, d_{k_2})	(15, 15) Meters
Deviation between real value of edge server speed $\hat{f}_{i,k}^{\text{es}}$ [22]	2 %
Edge server speed $f_{\text{max}}^{\text{es}}$ [22]	10^{10} Cycles/s
Error probability $\epsilon_{i,k}$ [10]	10^{-5}
Intermediate task t [22]	4
Maximum delay requirement λ_k^{max} [69]	2 Seconds
Maximum EC requirement ξ_k^{max} [69]	1 Joule
Minimum data rate requirement r_{min} [69]	2 Bits/s/Hz
Maximum IoT devices served by each edge server χ_{max} [69]	6
Noise power (for simplicity) σ^2 [22]	10^{-10} W
Output data size for device k_1 $\{\alpha_{i,k_1}\}$ [22]	[1500,1000,1600, 1000, 0] Kbytes
Output data size for device k_2 $\{\alpha_{i,k_2}\}$ [22]	[2000,1500,1000, [1400,1000],1500, 1000, 0] Kbytes
Peak computational frequency of device $f_{\text{max}}^{\text{lo}}$ [69]	10^8 Cycles/s
Peak transmit power of each IoT device p_{max}	200 mW
Tasks for device (k_1 and k_2) (L, M) [22]	(3, 5)

Table 4.1: Simulation Parameters.

IoT devices. Assuming $f_{i,k}^{\text{es}} > f_{\text{max}}^{\text{lo}}$, the maximum computational frequency for each IoT device and the processing speed of each edge server are established at $f_{\text{max}}^{\text{lo}} = 10^8$ Cycles/s and $f_{\text{max}}^{\text{es}} = 10^{10}$ Cycles/s, respectively. The FBL is set at 100, and the computing efficiency parameter for each device is defined as $\mu = 10^{-26}$ [22]. The noise spectral density is specified as $\sigma^2 = 10^{-10}$ Watts to compare the proposed scheme with a benchmark. For further analysis, it can be assumed that -174 dBm/Hz. The proportion of the anticipated processing rate is a pre-configured value, and its selection is based on established computing models and standard practices in wireless networking [66]. The URLLC decoding error probability is defined as $\epsilon_{i,k} = 10^{-5}$. The maximum EC is limited to $\xi_k^{\text{max}} = 0.5$ Joule. The experimental scenario is simulated using a single ABS $\mathcal{A} = 1$ along with multiple IoT devices $\mathcal{K} = \{1, 2\}$ unless explicitly stated otherwise. The analytical assessment considers a maximum latency of 1 millisecond. The minimum time required to transmit a unit blocklength is 0.01 milliseconds, as indicated in [10]. The altitude of the serving ABs is fixed at 50 meters. The pathloss models used are detailed in [66]. The details of the simulation parameters are provided in Table 4.1.

4.6.2 Performance Comparison

In Fig. 4.3, we study the correlation between devices. Specifically, we analyze the trade-off between total execution time and energy for each IoT device by varying the $\partial_{k_2}^\lambda$. For given $\partial_{k_1}^\lambda$, an increase in $\partial_{k_2}^\lambda$ results in increased EC while simultaneously resulting reducing the total execution time. This trade-off is also observable for IoT device k_1 . The trade-off curve for IoT device k_1 converges to a critical point at $\partial_{k_1}^\lambda = 0.4$. This convergence indicates that beyond this value, the optimal execution time and energy of IoT device k_1 remains constant despite increases in $\partial_{k_2}^\lambda$. The underlying rationale behind this phenomenon can be attributed to the fact that with an increase in $\partial_{k_1}^\lambda$, the device k_1 operates in a dual capacity, i.e., assisting IoT device k_2 while concurrently minimizing its processing delay.

In Fig. 4.4, we illustrate the correlation between the devices and analyze the

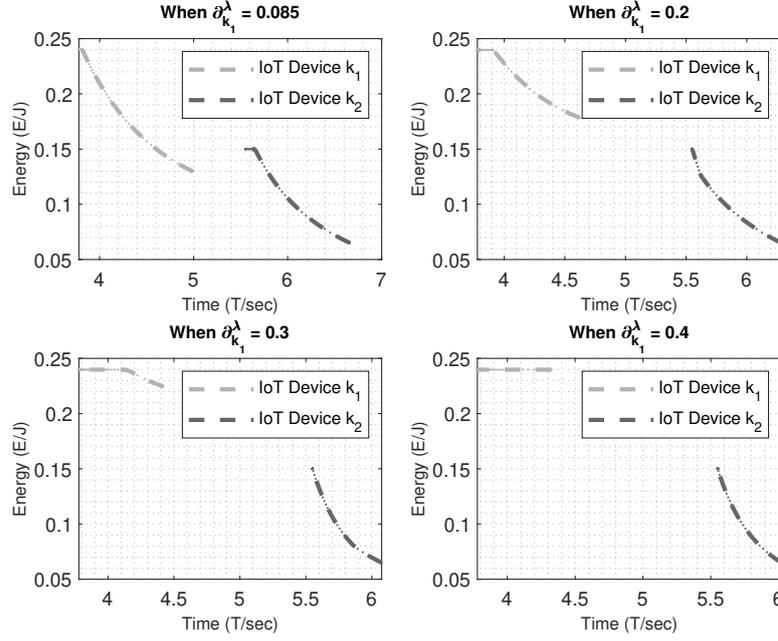


Figure 4.3: The Trade-off between the Total Execution Time and Energy for each IoT Device when $\partial_{k_2}^\lambda$ varies, where $d_{k_1} = d_{k_2} = 15$ Meters.

proposed scenario in depth. Specifically, the effect of the intermediate task t on the computational delay is explored. We observe that the waiting time for device k_1 increases with an increase in t when $\partial_{k_1}^\lambda$ is kept small, i.e., $\partial_{k_1}^\lambda = 0.05$ or 0.3 . In contrast, the waiting time for device k_2 decreases. This is because device k_1 needs to complete all three tasks to meet the finish time for the first t tasks of device k_2 , especially for smaller values of t , i.e., when $t = 1$ or 2 . Meanwhile, device k_2 only needs to slow down its computations to get the final output from device k_1 . Generally, this results in larger $\lambda_{k_1}^{\text{tot}}$ and smaller $\lambda_{k_2}^{\text{tot}}$ with increasing t . When $\partial_{k_1}^\lambda$ is further increased, i.e., $\partial_{k_1}^\lambda = 0.5$, computations for both devices become independent and are optimized separately, instead of minimizing its computational delay to meet the stringent requirements of t task at device k_2 . Therefore, there is no change in the computational delay for both devices when t increases from 1 to 5.

Recall that in IoT device k , the weight are related by the formula $\partial_k^\xi = 1 - \partial_k^\lambda$. The system's energy-time cost objective value, varying with $\partial_{k_2}^\lambda$ is analyzed in Fig.

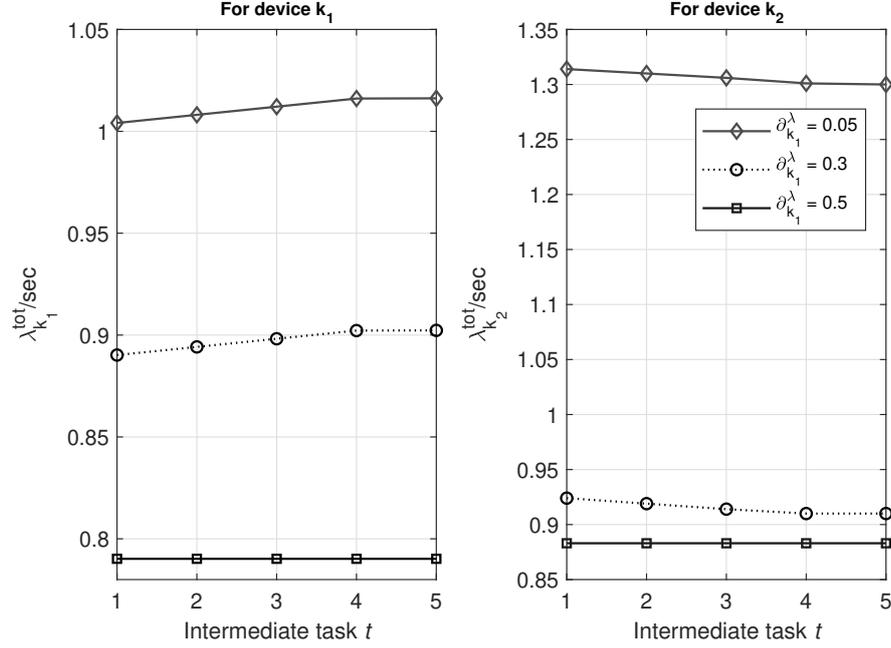


Figure 4.4: Impact of Intermediate Task t on the Computational Delay, when $(L, M, t) = (3, 5, t)$.

4.5. We employ two benchmarks against our proposed scheme to ensure a fair comparison. Our proposed BSSE algorithm is a modified version of the bisection algorithm, as referenced in [22]. It follows the same steps as defined in [22], except for line 3 in Algorithm (4.2), which is used to set the offloading factor to a random offloading scheme, and then update it to minimize the total energy-time cost objective. As anticipated, an increase in $\partial_{k_2}^\lambda$ leads to a higher energy-time cost of the system. Compared to the baseline algorithm in [22], our proposed BSSE algorithm is more efficient in terms of the energy-time cost of the system. The solutions obtained from the benchmark algorithm and those generated by the proposed BSSE algorithm have a significant difference. Furthermore, both the benchmark and the proposed algorithms satisfy the one-climb policy, i.e., (0 followed by 1) can appear only once. This finding underscores the potential of the proposed BSSE algorithm as a highly advantageous alternative to the benchmark algorithms, particularly in identifying the optimal offloading scheme in the counterparts.

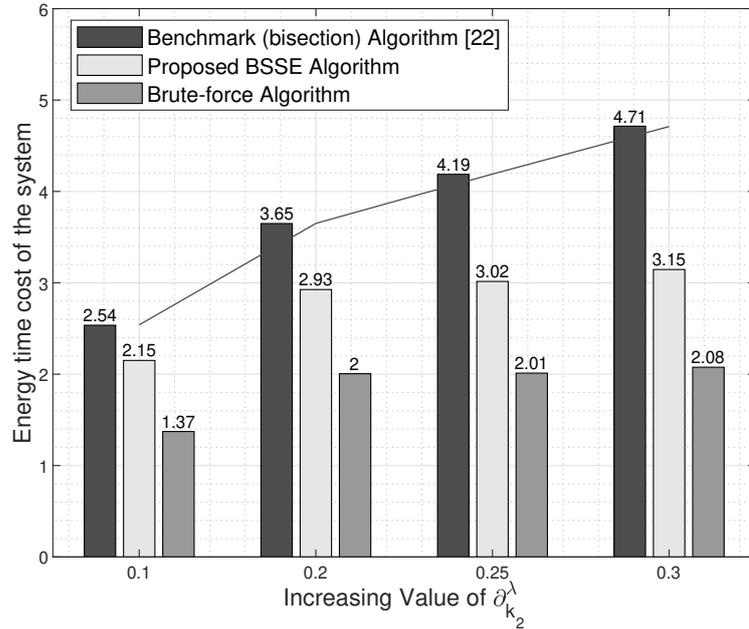


Figure 4.5: Energy-time Cost of Benchmark and Proposed Algorithms, when $\mathcal{A} = 2$, $\mathcal{K} = 2$ and $d_{k_1} = d_{k_2} = 15$ Meters.

In Fig. 4.6, we illustrate the impact of the increasing value of d_{k_1} on the overall energy-time cost of the system, where the distance of IoT device k_2 is fixed at $d_{k_2} = 10$ meters. The values of $\{\delta_{i,k_1}\}$ and $\{\delta_{i,k_2}\}$ are uniformly distributed between 10 and 200 Mcycles. The average performance of twenty independent iterations has been computed to plot this figure. It can be seen that the overall energy-time cost value increases with the distance of IoT device k_2 for all four schemes. As mentioned in Section 4.2.9, this increase is attributed to the system's reliance on local computing. In this case, the IoT device k_1 needs to offload the output of the final task to the edge server, forwarding this information to IoT device k_2 , resulting in a higher energy-time cost. Numerically, the energy-time cost of the proposed BSSE is slightly higher than the brute-force algorithm, though the difference is not substantial. Moreover, the effectiveness of the proposed BSSE algorithm surpasses that of both bisection algorithm [22] and Gibbs sampling algorithm [22], i.e., around 89.39% to 138.96% lower and around 33.61% to 45.28% lower, respectively. This

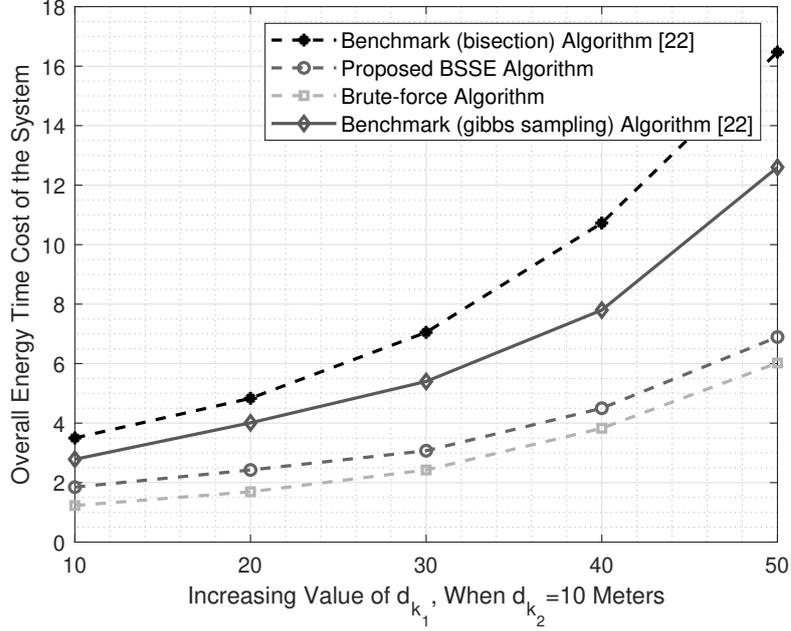


Figure 4.6: Overall Energy-time Cost versus Value of d_1 , when $d_2 = 10$ Meters.

highlights the benefits of employing a strategy that optimizes resource allocation and offloading decisions for both IoT devices.

In Fig. 4.7, we further study the impact of different case studies (task dependency models) on system performance, considering various topological call graphs ($|L|, |M|, |t|$). Each IoT device is limited to a distance of 10 meters, where $p_{k_1} = p_{k_2} = 0.2$ W, $\partial_{k_1}^\lambda = 0$, and $\partial_{k_2}^\lambda = 0.5$. We initially examined scenarios that involve three and four tasks on each device, with task four on device k_2 requiring the final output from device k_1 . Subsequently, we varied the number of tasks and the position of the intermediate node to observe how this inter-device dependency influences the overall energy-time cost of the network at the optimal configuration. It is observed that the proposed BSSE algorithm achieves a lower computational cost than the bisection algorithm and performs comparably to the brute-force algorithm. The effectiveness of the proposed algorithm can also be seen from its stable increase in the energy-time cost value. Specifically, when we extend the call graph, the brute-force algorithm shows an exponential increase in the

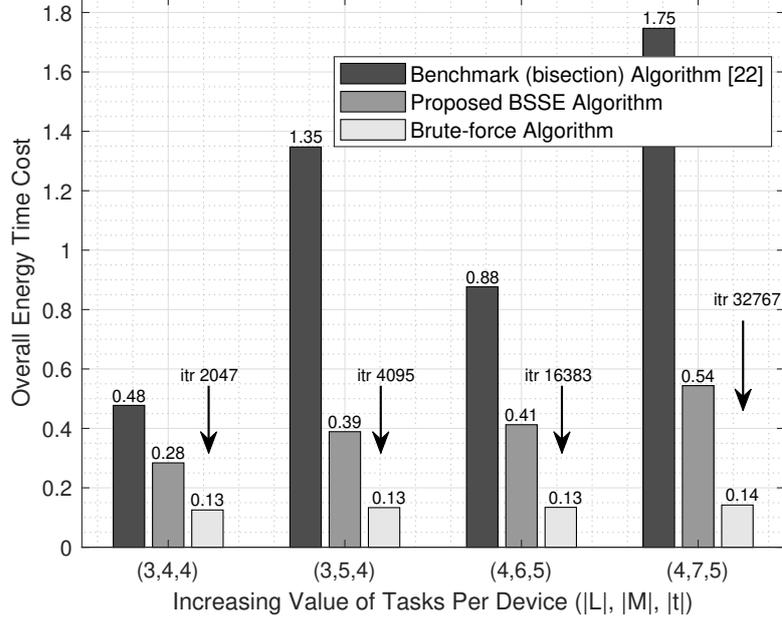


Figure 4.7: Energy-time Cost for Increasing value of $(|L|, |M|, |t|)$ in Three Algorithms.

complexity growth. In contrast, the bisection algorithm solves it in polynomial time but with a higher energy-time cost of the system. However, the BSSE algorithm maintains lower fluctuations and stable increases. For example, with the call graph $(|L|, |M|, |t|) = (3, 5, 4)$, the optimal solution computed by the bisection algorithm involves 4095 calls, resulting in $\varphi_{l,k_1} = \{0 \ 1 \ 1\}$ and $\varphi_{m,k_2} = \{1 \ 1 \ 0 \ 1 \ 1\}$ with an energy-time cost of $\zeta_{k_1} + \zeta_{k_2} = 0.13$. In contrast, the proposed BSSE algorithm requires only five calls to the bisection algorithm to find a near-optimal solution, with $\varphi_{l,k_1} = \{1 \ 0 \ 1\}$ and $\varphi_{m,k_2} = \{0 \ 0 \ 1 \ 1 \ 1\}$, resulting in an energy-time cost of $\zeta_{k_1} + \zeta_{k_2} = 0.28$. Hence, the proposed algorithm is less computationally intensive and achieves a 41.67% lower energy-time cost objective value than the benchmark bisection algorithm [22].

In Fig. 4.8, we extend the analysis of the overall energy-time cost of the proposed task dependency model to a multi-user case for all three schemes. The IoT devices are uniformly distributed at distances ranging from 10 to 30 meters.

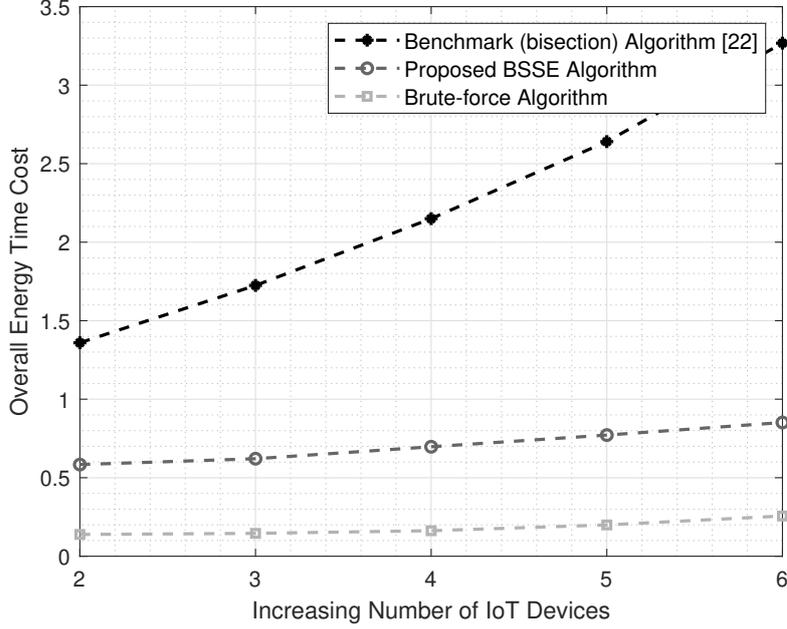


Figure 4.8: Proposed Task Dependency Model for a Multi-device Case Study with $k = 4$.

We assume the following computational requirements $\{\delta_{i,k_3}\} = [50.5, 45.3, 86.6]$ (Mcycles) and $\{\alpha_{i,k_3}\} = [1400, 1200, 1500, 1300]$ (Mcycles) for IoT device 3, $\{\delta_{i,k_4}\} = [65.5, 50.3, 75.6]$ (Mcycles) and $\{\alpha_{i,k_4}\} = [1500, 1400, 1000, 1500]$ (Mcycles) for IoT device 4, $\{\delta_{i,k_5}\} = [55.5, 42.3, 90.6]$ (Mcycles) and $\{\alpha_{i,k_5}\} = [1600, 1500, 1300, 1700]$ (Mcycles) for IoT device 5, and $\{\delta_{i,k_6}\} = [58.5, 47.3, 82.6]$ (Mcycles) and $\{\alpha_{i,k_6}\} = [1200, 1300, 1600, 1600]$ (Mcycles) for IoT device 6. The input required for the fourth task on the IoT device k_2 requires the final output of the task of all other devices $|K|-1$. It should be noted that the BSSE algorithm exhibits superior performance compared to the baseline algorithm, as referenced in [22]. More precisely, the total energy-time cost metric of the system shows a curve rise, increasing from 0.7758 to 2.4160 when we increase the number of IoT devices from 2 to 6. This observed pattern depicts the intensified interdependence of tasks between devices, which yields more substantial system performance improvements.

4.7 Summary

This study examines how task dependencies between IoT devices affect task offloading and resource allocation decisions. It does so by digitizing real-time edge networks and integrating aerial terrestrial networks. The presented problem is characterized as a mixed-integer non-linear programming problem. It becomes computationally intractable because of its inherently combinatorial link with task-offloading decisions and strong correlation with resource allocation. We propose a joint optimization approach that optimizes transmit power, CPU frequency, and task offloading policy to minimize energy-time cost. Our proposed scheme can accommodate various tasks, sometimes exceeding one hundred. The energy-time cost of the system closely approximates that of the brute-force algorithm that delivers the optimal solution. A notable discovery is that our proposed BSSE algorithm demonstrates approximately equal energy-time cost for both devices, that is, $\zeta_{k_1} \sim \zeta_{k_2}$, then brute force algorithm where $\zeta_{k_1} > \zeta_{k_2}$. This similarity is advantageous, as it ensures equitable energy-time costs for IoT devices, regardless of their task loads, leading to significant cost savings for devices handling more tasks. Furthermore, the BSSE algorithm achieves convergence in just five iterations using the bisection method, a marked improvement over the requirement of the brute force algorithm of 4096 iterations. Our simulation results corroborate the effectiveness of the proposed algorithm compared to benchmark approaches. The next chapter investigates the potential of machine learning techniques for optimal resource allocation in both the OMA and NOMA systems.

Chapter 5

¹Connectivity Aware Optimal Resource Allocation Using Unsupervised Learning

¹This chapter is submitted to IEEE Transactions on Green Communications and Networking, a leading journal in the wireless communication domain.

5.1 Introduction

ISD network plays a vital role in daily life, including intelligent transportation, agricultural monitoring, and disaster management [108]- [109]. ISDs primarily rely on terrestrial base stations to manage small-scale communication with inadequate mobility. Consequently, the efficacy of the network is compromised when the target coverage area is affected by natural disasters [110]. Under these circumstances, UAVs have become highly adaptable wireless airborne technology, offering a promising solution to this problem with minimal mobility, human intervention, and infrastructure costs in different hot spots [111].

A clear LoS channel makes UAVs an ideal candidate for UAV-assisted ISD networks, leading to improved QoS for every device and offering high data rates and a broad spectrum of services during natural disasters. It should be noted that existing literature on UAV communication either focuses on stationary UAVs hovering above the hotspot center or optimizes the UAV trajectory without considering the reliability, delay, and QoS constraints to maximize the system's throughput. Ignoring kinematic constraints, intelligent UAV positioning, and static propulsion pose significant challenges to UAV communication in maintaining user connectivity during malicious attacks.

This chapter introduces a novel approach to optimal resource allocation and UAV deployment to maximize the achievable rate for URLLC (requiring extremely high reliability and low latency). A collaborative system designed with UAVs is examined to support scattered communities and enhance network invulnerability cooperatively. The key features involve machine learning-based intelligent UAV placement and optimal resource allocation to reduce co-channel interference and improve system performance. An optimization solution employing unsupervised learning-based ISD grouping, reinforcement learning-based intelligent UAV placement, and DeepFusion-based PA (integrating deep learning-based PA with FTPA to provide an additional layer of optimization) is proposed. This approach ensures power efficiency and fairness among sub-carriers, improving overall system performance.

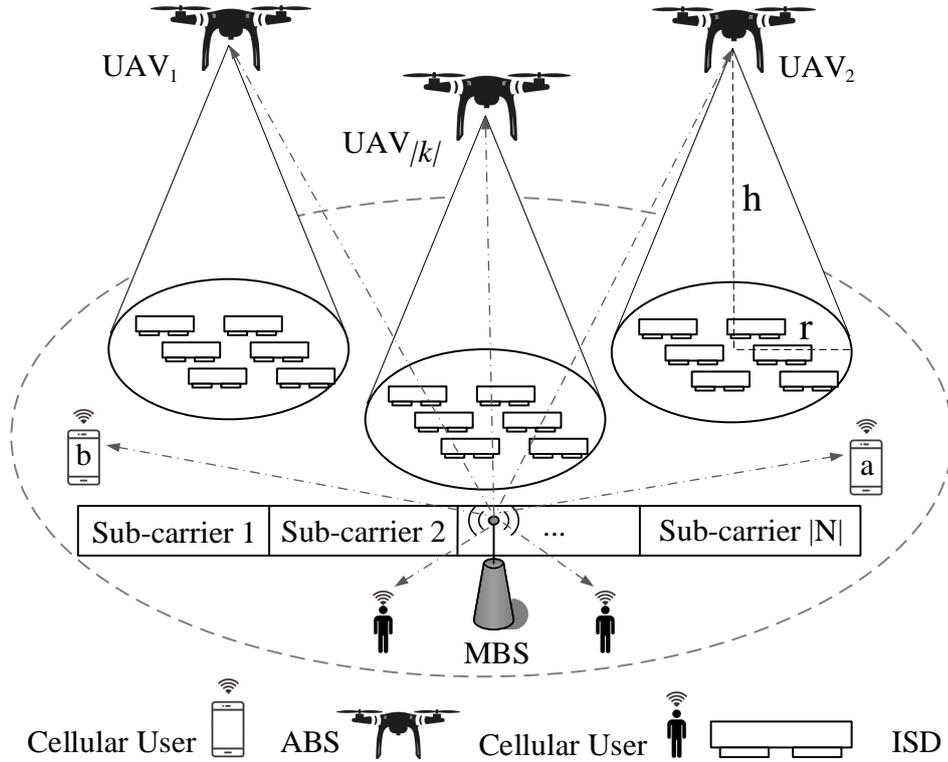


Figure 5.1: Scenario of NOMA-based Multi-UAV Aerial Terrestrial Network.

The remainder of this chapter is organized as follows. Section 5.2 presents the system model of the proposed work. Section 5.3 formulates the resource optimization problem. In Section 5.4, the initial problem is decoupled into sub-problems to provide optimal solutions. Section 5.5 presents the simulation results. Finally, this chapter is concluded in Section 5.6.

5.2 System Model

Fig. 5.1 shows a single-cell aerial terrestrial downlink cellular network comprising two layers. The ground layer contains an MBS, stationary ISDs, and cellular users (CUs). The second layer contains multiple ABSs with limited computing resources and battery capacity.

5.2.1 Multi-ABS Aerial Terrestrial Network Architecture

Let ISDs be randomly deployed within certain regions of the ground in the time slot t . CUs are also randomly distributed within the MBS coverage. We define the set of ISDs as $i \in \mathcal{I} = \{1, 2, \dots, |I|\}$ and the set of CUs as $j \in \mathcal{J} = \{1, 2, \dots, |J|\}$, respectively. It is further assumed that the ABSs have the CSI of all the ISDs in advance by the MBS using pilot signals. Since ISDs cannot communicate directly with the MBS, ABSs are deployed to hover at fixed strategic locations to transmit the data from MBS to ISDs. The set of ABSs is designated as $k \in \mathcal{K} = \{1, 2, \dots, |K|\}$, where each ABS serves a group of non-overlapping ISDs. It is important to note that the ISDs under the same ABS k are denoted by \mathcal{I}_k . Therefore,

$$\bigcup_{k \in \mathcal{K}} \mathcal{I}_k = \mathcal{I}, \quad \mathcal{I}_{k_1} \cap \mathcal{I}_{k_2} = \emptyset, \quad \forall k \in (k_1, k_2), k_1 \neq k_2. \quad (5.1)$$

The transmission bandwidth of the network is divided into N orthogonal sub-carriers, designated by the set $n \in \mathcal{N} = \{1, 2, \dots, |N|\}$. We reserve the sub-carrier percentage depending on the traffic demand to accommodate the CUs and ABSs. It is worth mentioning that each ABS subsystem uses $|N|$ orthogonal sub-carriers. Therefore, there will be no inter-cell interference. Meanwhile, the ISD using the same sub-carrier within a single ABS subsystem results in intra-cell interference. We assume that each sub-carrier can be allocated to a maximum γ_n number of ISDs. However, each ABS subsystem can assign one ISD to no more than χ_{\max} sub-carriers. A big community requires more sub-carriers; however, the growth rate $(\log_2(e)/(1 + \mathcal{I}_k))$ declines with an increase in community size, as given in Fig. 5.2. Therefore, each ABS subsystem serves a limited number of IoT devices.

5.2.2 Channel Model

We assume that the channels among the devices are independent and suffer Rayleigh fading. We define the following three models.

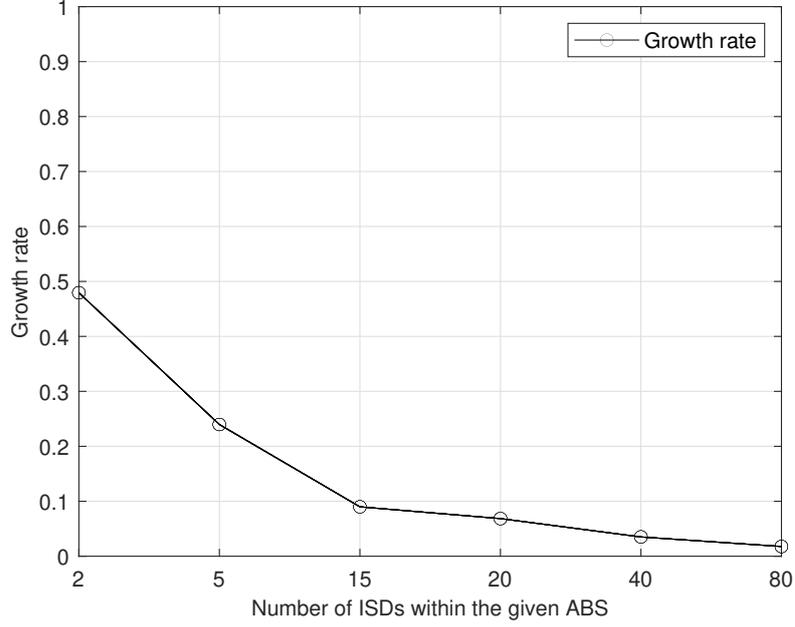


Figure 5.2: Growth Rate.

5.2.2.1 Ground-to-Ground Channel Model

In the ground-to-ground communication model, the MBS communicates with the CUs within its coverage area. The channel gain of the CU j over the sub-carrier n is given by $g_{\text{mbs},j}^n = \left(\frac{\beta_0 |h_{\text{mbs},j}|}{d_{\text{mbs},j}^2} \right)$, where β_0 is the channel power gain at the reference distance of one meter and $h_{\text{mbs},j}$ is the coefficient of fading. The squared distance between MBS and CU j is given by $d_{\text{mbs},j}^2 = \left[(X_{\text{mbs}} - X_j)^2 + (Y_{\text{mbs}} - Y_j)^2 + (Z_{\text{mbs}} - Z_j)^2 \right]$, where X, Y and Z are the coordinates x, y, and z of MBS and CU, respectively. For CU j , the pathloss is defined by

$$PL_{\text{mbs},j} = -55.9 + 38 \log(d_{\text{mbs},j}) + (24.5 + 1.5f/925) \log(f). \quad (5.2)$$

5.2.2.2 Ground-to-Air Channel Model

The MBS communicates with ABS within its coverage area in the ground-to-air communication model. The channel gain of the link between MBS and ABS k over the sub-carrier n is given by $g_{\text{mbs},k}^n = \left(\frac{\beta_0 |h_{\text{mbs},k}|}{d_{\text{mbs},k}^2} \right)$, where $h_{\text{mbs},k}$ is the

fading coefficient. The squared distance between MBS and ABS k is given by $d_{\text{mbs},k}^2 = \left[(X_{\text{mbs}} - X_k)^2 + (Y_{\text{mbs}} - Y_k)^2 + (Z_{\text{mbs}} - Z_k)^2 \right]$, where X , Y , and Z are the coordinates x , y , and z of MBS and ABS, respectively. The pathloss for the following model is the same as for the ABS-to-Ground communication model given below.

5.2.2.3 Air-to-Ground Channel Model

In the air-to-ground communication model, the ABS communicates with the ISD within its coverage area. The channel gain of ISD i over the sub-carrier n is given by $g_{k,i}^n = \left(\frac{\beta_0 |h_{k,i}|}{d_{k,i}^2} \right)$, where $h_{k,i}$ is the fading coefficient. We computed the squared distance between ABS and ISD i using the following equation $d_{k,i}^2 = \left[(X_k - X_i)^2 + (Y_k - Y_i)^2 + (Z_k - Z_i)^2 \right]$, where X , Y , and Z are the coordinates x , y , and z of MBS and ISD, respectively. According to [108], the pathloss for LoS and N-LoS for the given communication link is given by

$$PL_{\text{LoS},k} = L_{FS,k} + 20 \log(d_{k,i}) + \eta_{\text{LoS}}, \quad (5.3)$$

$$PL_{\text{N-LoS},k} = L_{FS,k} + 20 \log(d_{k,i}) + \eta_{\text{N-LoS}}, \quad (5.4)$$

where $L_{FS,k} = 20 \log(f) + 20 \log\left(\frac{4\pi}{c}\right)$ is the free space pathloss, c is the speed of light, and f is the carrier frequency. The additional attenuation factors due to LoS and N-LoS are η_{LoS} and $\eta_{\text{N-LoS}}$, respectively. The probability of LoS is given by

$$P_{\text{LoS},k} = \frac{1}{1 + a \exp(-b((\theta_i) - a))}, \quad (5.5)$$

where, $\theta_i = \sin^{-1}\left(\frac{Z_k}{d_{k,i}}\right)$ is the angle of elevation, a and b are parameters that reflect the density of buildings in the urban environment. The average pathloss is given by

$$PL_{\text{avg},k} = P_{\text{LoS},k} PL_{\text{LoS},k} + (1 - P_{\text{LoS},k}) PL_{\text{N-LoS},k}, \quad (5.6)$$

where, $P_{\text{N-LoS},k} = 1 - P_{\text{LoS},k}$ is the probability of N-LoS.

5.2.3 NOMA-based Transmission Model

We consider a downlink multi-carrier NOMA system in which multiple ISDS can be assigned to a unique orthogonal sub-carrier to improve spectral efficiency [112]-

[113]. Unlike NOMA, OMA allocates a unique orthogonal sub-carrier to each ISD (introduced as a benchmark scheme).

5.2.3.1 NOMA-based Transmission

If γ_n is the maximum number of ISDs allocated to the sub-carrier n , then the corresponding received signal at the given ISD i is given by

$$y_{k,i}^n = \sum_{i \in \gamma_n} \sqrt{p_{k,i}^n g_{k,i}^n} x_{k,i}^n + N_0. \quad (5.7)$$

Where, $x_{k,i}^n$ is the information signal transmitted to ISD i with transmitted power $p_{k,i}^n$ and $N_0 \approx \mathcal{CN}(0, \sigma^2)$ is the Additive White Gaussian Noise (AWGN), having zero mean and a variance of σ^2 . The transmit power of either ABS or ISD is practically non-negative and limited. A feasible set of PAs to various ISDs served by ABSs, i.e., $\mathcal{P} = \left[p_{k,i}^n \right]_{K \times I \times N}$ should be subject to

$$p \in \mathcal{P} \left\{ \left[p_{k,i}^n \right]_{K \times I \times N} \left| 0 \leq p_{k,i}^n \leq p_{k,i}^{\max}, \forall i \in \mathcal{I}, \forall k \in \mathcal{K} \right. \right\}; \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{I}_{k,n}} p_{k,i}^n \leq P_{\max}^k, \forall k \in \mathcal{K}. \quad (5.8)$$

Since NOMA allows multiple ISDs to share the same sub-carrier simultaneously, interference occurs. The total interference experienced by ISD i of sub-carrier n is given by

$$\tilde{\tau}_{k,i}^n = \sum_{l \in \gamma_n, l \neq i} p_{k,l}^n g_{k,l}^n. \quad (5.9)$$

In power domain NOMA (PD-NOMA), superposition coding is performed at the transmitter and match filtering successive interference cancellation (MF-SIC) at the receiver to detect overlapping signals [114]. It is worth mentioning that the decoding order significantly impacts the performance of the MF-SIC process. Conventionally, the decoding order of SIC is defined as the permutation over the active ISDs. Our decoding order consists of decoding the signals from the best to the worst channel conditions. If $q_n \in \mathcal{Q}_n = \{1, 2, \dots, |Q|\}$ is the set of decoding sequences over the given

sub-carrier, then $q_n(i)$ returns the decoding index² of the device i . In contrast, the decoding order of the device i is given by $q_n^{-1}(i)$.

Multiplexing in PD-NOMA depends on the ability of SIC. It can be achieved for any given device if its incoming signal's strength over residual interference is above or equal to δ , i.e., we have:

$$\left(\frac{p_{k,i}^n g_{k,i}^n}{\tau_{k,i}^n} \right) \geq \delta, \quad (5.10)$$

$\delta \geq 1$, where:

$$\tau_{k,i}^n = \sum_{l=q_n^{-1}(i)+1}^{\mathcal{I}_{k,i}^n} p_{k,l}^n g_{k,l}^n, \quad (5.11)$$

where $\mathcal{I}_{k,i}^n = \{\cup d \mid d \in \mathcal{I}_{k,n}, g_{k,i}^n \geq g_{k,d}^n\}$ is a set containing ISDs that have inferior channel conditions against ISD i and $\mathcal{I}_{k,n} = \{\cup i \mid \text{device } i \text{ is allocated sub-carrier } n, \forall i \in \mathcal{I}, \forall n \in \mathcal{N}\}$ is the set of ISDs that occupy the same subcarrier n within the serving ABS k . The interference is completely removed if the SIC constraint in (5.10) is successful.

5.2.4 URLLC-based Downlink Transmission Rate

The signal to noise (SNR) at device f on the given sub-carrier n is computed by

$$\gamma_{\text{mbs},f}^n = \left(\frac{\varphi_{\text{mbs},f}^n p_{\text{mbs},f}^n g_{\text{mbs},f}^n}{N_0} \right), \forall f \in (j, k), \quad (5.12)$$

where, $p_{\text{mbs},f}^n$ is the transmitted power of the given device f , $\varphi_{\text{mbs},f}^n$ is the binary indicator, i.e., if device f is allocated to sub-carrier n within the coverage of MBS then $\varphi_{\text{mbs},f}^n = 1$; otherwise 0. For the given sub-carrier, the achievable rate is determined by

$$r_{\text{mbs},f}^n = w_n \log_2 (1 + \gamma_{\text{mbs},f}^n) - w_n \sqrt{\frac{\partial_{\text{mbs},f}^n Q^{-1}(\epsilon)}{m_1 \ln 2}}, \quad (5.13)$$

where, w_n is the bandwidth of the given sub-carrier, ϵ is the DEP, m_1 is the blocklength [66], $\partial_{\text{mbs},f}^n$ is the channel dispersion on the given sub-carrier, and Q^{-1}

²In the decoding set \mathcal{Q}_n , if $q_n(1) = 5$ (showing that the device 1 is the fifth to be decoded), then $q_n^{-1}(5) = 1$

is the inverse of the Gaussian Q-function [108]. The SINR for decoding the signal of the device i on the given sub-carrier n transmitted by ABS $k \in \mathcal{K}$ is computed by

$$\gamma_{k,i}^n = \left(\frac{\varphi_{k,i}^n P_{k,i}^n g_{k,i}^n}{\tau_{k,i}^n + N_0} \right), \quad (5.14)$$

where, $\varphi_{k,i}^n$ is the binary indicator, i.e., if the device i is assigned to sub-carrier n within the serving ABS k then $\varphi_{k,i}^n = 1$; otherwise 0. A feasible set of sub-carrier allocations to various ISDs served by ABSs, i.e., $\varphi = [\varphi_{k,i}^n]_{K \times I \times N}$ should satisfy

$$\begin{aligned} \varphi \in \varphi \left\{ [\varphi_{k,i}^n]_{K \times I \times N} \middle| \sum_{n \in \mathcal{N}} \varphi_{k,i}^n \leq \chi_{\max}, \forall k \in \mathcal{K}, \right. \\ \left. \forall i \in \mathcal{I}; \varphi_{k,i}^n \in \{0, 1\}, \forall i \in \mathcal{I}, \forall k \in \mathcal{K}, \forall n \in \mathcal{N}. \right. \end{aligned} \quad (5.15)$$

The achievable rate on the given sub-carrier in URLLC is computed as [66]:

$$r_{k,i}^n = w_n \log_2 (1 + \gamma_{k,i}^n) - w_n \sqrt{\frac{\partial_{k,i}^n}{m_2} \frac{Q^{-1}(\epsilon)}{\ln 2}}, \quad (5.16)$$

where, m_2 is the blocklength for this hop, $\partial_{k,i}^n$ is the channel dispersion on the given sub-carrier. The total achievable rate for the given device within the serving ABS k is denoted by $r_{k,i} = \sum_{n \in \mathcal{N}} \varphi_{k,i}^n r_{k,i}^n$. The total data rate on the given sub-carrier is computed by $\gamma_k^n = \sum_{i \in \mathcal{I}_{k,n}} \gamma_{k,i}^n$. The total achievable rate of a single subsystem is computed by $\gamma_k = \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{I}_{k,n}} \gamma_{k,i}^n$. The sum-rate of the network is given by

$$\chi_{\text{Rate}} = \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{I}_{k,n}} \varphi_{k,i}^n r_{k,i}^n. \quad (5.17)$$

5.3 Problem Formulation

This work aims to maximize the utility function by optimizing the ABS location, sub-carrier allocation, and transmit power while ensuring reliability and QoS constraints. The proposed optimization problem overlooks the connection between the MBS and the CUs, focusing solely on each link that transmits data from the MBS to ISDs within the coverage of serving UAVs. The notation $(x_k)_{\forall k \in \mathcal{K}} = (X_k, Y_k, Z_k)$ is the

location of the serving ABS k , and $(x_i)_{\forall i \in \mathcal{I}_k} = (X_i, Y_i, Z_i)$ the location of the ISDs i within the coverage of serving ABS. We denote the set of locations for all ISDs within the given serving ABS by $\mathcal{X} = \{\cup x_i, \forall i \in \mathcal{I}_k\}$. The formulation of the optimization problem is as follows.

$$\begin{aligned}
 & \max_{\mathbf{p}, \boldsymbol{\varphi}, \boldsymbol{\chi}} \chi_{\text{Rate}} \\
 \text{s.t. } & C_1 : \sum_{n \in \mathcal{N}} \gamma_{k,i}^n = \gamma_{k,i}, \forall i \in \mathcal{I}, \forall k \in \mathcal{K}, \\
 & C'_2 : \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{I}_{k,n}} p_{k,i}^n \leq P_k^{\max}, \forall k \in \mathcal{K}, \\
 & C''_2 : \sum_{i \in \mathcal{I}_{k,n}} p_{k,i}^n \leq P_n, \forall n \in \mathcal{N}, \\
 & C_3 : \sum_{n \in \mathcal{N}} \varphi_{k,i}^n \leq \chi_{\max}, \forall k \in \mathcal{K}, \forall i \in \mathcal{I}, \\
 & C_4 : \sum_{i \in \mathcal{I}} \varphi_{k,i}^n \leq \gamma_n, \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \\
 & C_5 : X_k^2 + Y_k^2 + Z_k^2 \leq r_{\max}^2, \forall k \in \mathcal{K}, \\
 & C_6 : \left(\frac{p_{k,i}^n g_{k,i}^n}{\tau_{k,i}^n} \right) \geq \delta, \forall k \in \mathcal{K}, \forall i \in \mathcal{I}, \forall n \in \mathcal{N}, \\
 & C_7 : \sum_{n \in \mathcal{N}} \varphi_{k,i}^n r_{k,i}^n = \gamma_{\min}, \forall i \in \mathcal{I}, \forall k \in \mathcal{K}.
 \end{aligned} \tag{5.18}$$

The constraints are described as follows: constraint C_1 maintains the QoS of each device. The constraint C'_2 represents the maximum transmit power P_k^{\max} budget in the ABS, and C''_2 denotes the power limit of P_n^{\max} for each sub-carrier. The constraint C_3 shows that each ISD occupies maximum χ_{\max} sub-carriers. The restriction C_4 ensures that each sub-carrier can be allocated to a maximum γ_n number of devices. The restriction C_5 ensures that the position of the ABS is within the radius of the MBS r_{\max} . The constraint C_6 ensures that the SIC decoding is performed successfully at the receiver end. The restriction C_7 limits the minimum rate requirement to γ_{\min} .

We define our objective function as a mixed-integer non-convex programming problem. Consequently, finding a solution in polynomial time is impractical due to

the non-affine nature of the constraint C_2 and the binary nature of $\varphi_{k,i}^n$. The problem can be addressed by combining the penalty approach with a monotonic optimization technique, incurring a substantial computational cost. Alternatively, it can be addressed by a standard relaxation mechanism. Therefore, we decouple the initial problem into sub-problems to get the sub-optimal solution: 1) machine-learning-based ISD grouping and reinforcement learning-based intelligent UAV placement, 2) joint optimization of sub-carrier allocation and PA.

5.4 Proposed Solution

5.4.1 Unsupervised Learning-based Internet of Things Smart Device Grouping

The maximum achievable rate governs Sub-carrier allocation in the cellular coverage area. For the sub-carrier allocation/ISD grouping within the UAV coverage, the following sub-problem is derived from the original problem (5.18) by treating power as constant.

$$\begin{aligned}
 & \max_{\varphi, \mathbf{x}_\Gamma} \chi_{\text{Rate}} \\
 \text{s.t. } & C_1 : \sum_{n \in \mathcal{N}} \gamma_{k,i}^n = \gamma_{k,i}, \forall i \in \mathcal{I}, \forall k \in \mathcal{K}, \\
 & C_2 : \sum_{n \in \mathcal{N}} \varphi_{k,i}^n \leq \chi_{\max}, \forall k \in \mathcal{K}, \forall i \in \mathcal{I}, \\
 & C_3 : \sum_{i \in \mathcal{I}} \varphi_{k,i}^n \leq \gamma_n, \forall k \in \mathcal{K}, \forall n \in \mathcal{N}, \\
 & C_4 : X_k^2 + Y_k^2 + Z_k^2 \leq r_{\max}^2, \forall k \in \mathcal{K}, \\
 & C_5 : \left(\frac{p_{k,i}^n g_{k,i}^n}{\tau_{k,i}^n} \right) \geq \delta, \forall k \in \mathcal{K}, \forall i \in \mathcal{I}, \forall n \in \mathcal{N}, \\
 & C_6 : \sum_{n \in \mathcal{N}} \varphi_{k,i}^n r_{k,i}^n = \gamma_{\min}, \forall i \in \mathcal{I}, \forall k \in \mathcal{K}.
 \end{aligned} \tag{5.19}$$

PD-NOMA improves the spectral efficiency of communication systems, which demultiplexes multiple devices that share the same sub-carrier, differentiated by

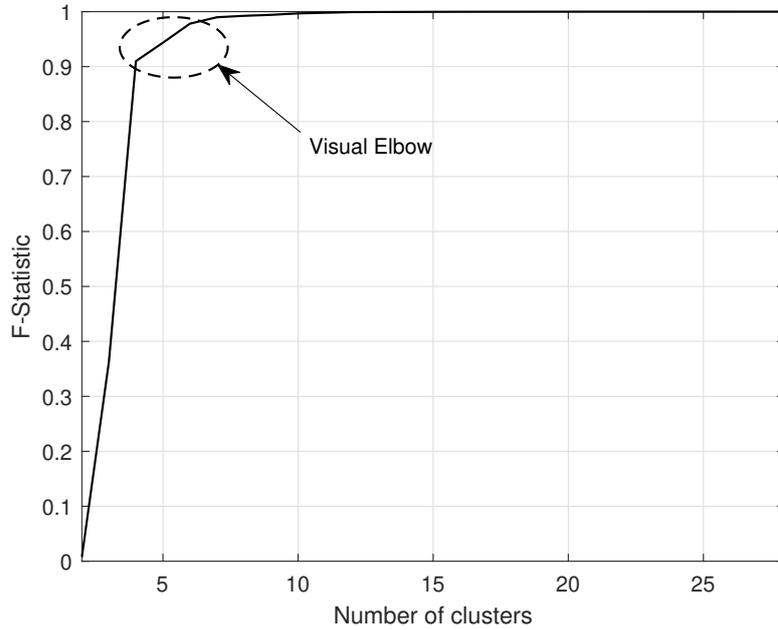


Figure 5.3: A Framework using Dual Criteria for Selecting the Optimal Number of Clusters.

their respective received power levels [114]. Its effectiveness mainly depends on the receiver's ability to distinguish these allocated power levels, which also depends on how the devices are grouped for multiplexing. While grouping the devices, machine-learning algorithms substantially decrease the computational complexity of the subsystem and enhance the probability of achieving an optimal solution compared to non-machine learning algorithms.

The Integration of K-means clustering has been proven to be a highly effective unsupervised technique for grouping users. It partitions data into multiple clusters, where it is necessary to specify the preferred/optimal number of clusters. Therefore, identifying the optimal number of clusters is a crucial aspect of this process [115]. Traditionally, the number should be relatively small compared to the data size intended for clustering. However, the authors in [116] used a basic version of the elbow method [117] to identify the optimal number of clusters, using K-means to determine the average internal cluster distance for the specified clusters. Our

proposed algorithm uses k-means alongside a dual criterion selection framework that integrates the elbow method with the F-test method [118] to determine the optimal number of clusters against each sub-carrier. The F-test method determines the optimal number of clusters using (5.20). It calculates the proportion of diversity between clusters in relation to the total diversity within and between clusters.

$$F_{k,n}(C_k) = \frac{\sum_{c=1}^{C_k} \mathcal{I}_{k,c} (\bar{g}_{k,c,n} - \bar{g}_{k,n})^2 / (C_k - 1)}{\sum_{c=1}^{C_k} \sum_{d=1}^{\mathcal{I}_{k,c}} (g_{k,c(d),n} - \bar{g}_{k,c,n})^2 / (\mathcal{I}_{k,c} - C_k) + \sum_{c=1}^{C_k} \mathcal{I}_{k,c} (\bar{g}_{k,c,n} - \bar{g}_{k,n})^2 / (C_k - 1)}, \quad (5.20)$$

where, $\bar{g}_{k,n} = \sum_i^{\mathcal{I}_k} g_{k,i}^n / \mathcal{I}_k$ indicates the mean channel gain $g_{k,i}^n, \forall n \in \mathcal{N}$, across all ISDs within ABS k , $\bar{g}_{k,c,n} = \sum_{d=1}^{\mathcal{I}_{k,c}} g_{k,c(d),n} / \mathcal{I}_{k,c}$ represents the mean of channel gain $g_{k,i}^n, \forall n \in \mathcal{N}$, across all ISDs belonging to the cluster c within the ABS k , denoted by the set $\mathcal{I}_{k,c}$.

Fig. 5.3 presents the graphical representation of the F-test method while employing K-means with a C_k number of clusters as input. It shows that the F-statistic rises with an increase in the number of clusters, i.e., the similarity within the channel gains of each cluster and the difference between channel gains of distinct clusters both increase. The curve exhibits a visually evident elbow shape between clusters 4 and 7, indicating that the optimal value likely falls within this range. Therefore, selecting $C_k = 6$ as an optimal value of clusters with an F-value of 97% appears to be a rational choice for C_k . The proposed framework is adaptable to numerous other requirements and constraints, such as adding more devices during an emergency. After calculating the optimal value of C_k , the sub-carrier allocation for each ISD is performed based on the normalized channel gain to remain fair with all ISDs. Therefore, the number of ISDs assigned to the given sub-carrier is $\mathcal{I}_{k,n} = C_k, \forall n \in \mathcal{N}$. The two-step ISD grouping is shown in Fig. 5.4, where $C_k = 2$. To improve the throughput of ISDs with lower channel gain, these are paired with ISDs having higher channel gains within the same NOMA cluster. Meanwhile, the

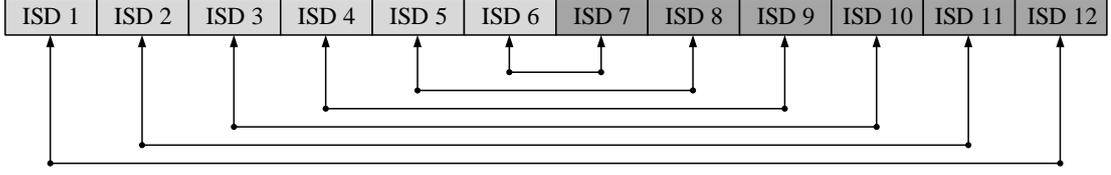


Figure 5.4: Illustration of 2-ISD NOMA Grouping for 12 Active ISDs in a Region.

device with the second highest channel gain is paired with the second lowest channel gain device in a separate NOMA cluster, and this pairing sequence is continued. The pseudo-code for ISD grouping and sub-carrier allocation is given in Algorithm (5.1).

5.4.2 Reinforcement Learning-based Aerial Base Station Placement

Adjusting the geographical positions of ABSs inherently forms an optimization problem with the aim of maximizing the objective function. This optimization can be approached from the following two perspectives: achievable data rate or transmit power of the ABS. In the current study, our focus is on achievable data rates. Consequently, the optimization problem is formulated as follows.

$$\gamma_k(\mathcal{X}, x_k) = \eta \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{I}_{k,n}} r_{k,i}(\mathcal{X}, x_k) + (1 - \eta) \min_{i \in \mathcal{I}_{k,n}} \{r_{k,i}(\mathcal{X}, x_k)\} \quad (5.21)$$

$$\begin{aligned} & x_k^{\text{opt}} \arg \max_{x_k} \gamma_k(\mathcal{X}, x_k), \\ & \text{subject to } X_k^2 + Y_k^2 + Z_k^2 \leq r_{\max}^2, \forall k \in \mathcal{K}, \end{aligned} \quad (5.22)$$

where η is the weighting factor. If $\eta \approx 1$, the achievable rate is prioritized; otherwise, the minimum rate is prioritized. It is important to mention that the traditional way to explore the large search space makes it difficult to optimize the position of the ABS. Because of limited resources at the ABS, it is assumed that the MBS can enable computation-intensive training for Q-learning beforehand and find an optimal policy

Algorithm 5.1 Machine learning-based ISD Grouping/Sub-carrier Allocation

- 1: **Input:** $(\mathcal{I})_{i \in \mathcal{I}}, (\mathcal{N})_{n \in \mathcal{N}}, (g_{k,i}^n)_{i \in \mathcal{I}, n \in \mathcal{N}}$
 - 2: **Output:** $\boldsymbol{\mu}, \boldsymbol{\phi}, \mathcal{I}_{k,n}$
 - 3: Initialize the matrix $\boldsymbol{\mu} = 0$ to keep track of allocated devices per sub-carrier
 - 4: Set $G_{k,i}^n = g_{k,i}^n / (\bar{g}_i = \sum_{n \in \mathcal{N}} g_{k,i}^n / |\mathcal{I}_k|), \forall k \in \mathcal{K}, \forall i \in \mathcal{I}_k, \forall n \in \mathcal{N}$
 - 5: **Step 1:** ISD grouping and computing the optimal number of clusters
 - 6: **for** $k = 1 : K$ **do**
 - 7: **for** $n = 1 : N$ **do**
 - 8: **for** $i = 2 : \mathcal{I}_k - 1$ **do**
 - 9: Employ k-means for clustering $\mathbf{g}_{k,i}^n = [g_{k,1}^n, g_{k,2}^n, \dots, g_{k,\mathcal{I}_k}^n]$ into i clusters
 - 10: Compute F-Statistics for current number of clusters using (5.20) and save its values for each i
 - 11: **end for**
 - 12: Select optimal number of clusters C_k using F-Statistics and Elbow method
 - 13: Set ISDs per sub-carrier equivalent to an optimal cluster's value, i.e., $\mathcal{I}_{k,n} = C_k$
 - 14: **end for**
 - 15: **end for**
 - 16: **Step 2:** Allocating sub-carriers per ISD
 - 17: Set $\phi_i = \lfloor \sum_{n \in \mathcal{N}} \mathcal{I}_{k,n} / \mathcal{I}_k \rfloor, \forall i \in \mathcal{I}_k$
 - 18: **Step 3:** Sub-carrier allocation
 - 19: **for** $k = 1 : K$ **do**
 - 20: **for** $n = 1 : N$ **do**
 - 21: Employ k-means for making $\mathcal{I}_{k,n}$ clusters based on $\mathbf{G}_{k,i}^n = [G_{k,1}^n, G_{k,2}^n, \dots, G_{k,\mathcal{I}_k}^n]$
 - 22: Associate the device with lower channel gain with a higher and configure the corresponding $(\phi_{k,i}^n)_{\forall i \in \mathcal{I}_{k,n}}$, if allocated sub-carriers are below ϕ_i
 - 23: **end for**
 - 24: **end for**
-

to maximize the cumulative reward. Afterward, this information is shared with the ABS using pilot signals to minimize frequent movements and data exchanges in the training process [9]. Basic concepts are given below.

Algorithm 5.2 Mapping Function $\mathcal{F}(y)$

```

1: Input: Achievable data rate ( $\gamma_k$ ), Number of states ( $N_s$ )
2: Output: Mapping function  $\mathcal{F}(y)$ 
3: Initialize the number of states
4:  $\vartheta = \text{Zeros}(1, N_s - 1)$ 
5: if  $y < \vartheta(1)$  then
6:    $ns = 0$ 
7: else if  $y \geq \vartheta(N_s - 1)$  then
8:    $ns = N_s$ 
9: else
10:  for  $i \in \text{len}(\vartheta) - 1$  do
11:    if  $y \geq \vartheta(i) \ \&\& \ y < \vartheta(i + 1)$  then
12:       $ns = i$ 
13:    end if
14:  end for
15: end if

```

5.4.2.1 Agent

The ABS k is designated as an agent in the current scenario. The agent determines its action based on the observation when modifying the Q-value in (5.24).

5.4.2.2 States ($s \in \mathcal{S}$)

We define the number of states by using a value mapping function (refer to Algorithm (5.2) using predefined thresholds, i.e., $\{\vartheta_i = i \in 1, 2, \dots, |N_s - 1|\}$, where $\vartheta_i < \vartheta_{i'}$ if $i < i'$.

5.4.2.3 Actions ($\psi \in \Theta$)

We consider an action ψ as a horizontal move within an action space Θ , comprising eight moves of uniform step size, where an angle of 45° differentiates each direction. Considering the Q-function in (5.24), one of the actions is performed in each iteration depending on the defined action rule.

$$\psi = \begin{cases} \arg \max_{\psi \in \Theta} Q(s, \psi), & \text{with probability } 1 - \epsilon, \\ \text{random step in } \Theta, & \text{with probability } \epsilon, \text{ where } \epsilon \in (0, 1). \end{cases} \quad (5.23)$$

5.4.2.4 Observations

The observations include positions (\mathcal{X}, x_k) and the received signal strength indicator [9].

5.4.2.5 Rewards (r)

We define the reward function in (5.18) to encourage ABS to move toward the optimal position during the hit-and-trial process. The agent learns with time to get a higher reward by selecting an action ψ in each iteration.

5.4.2.6 Environment

The environment serves as a platform in which the agent learns and makes decisions. It processes the agent's actions as inputs, generating observations and rewards as output. The agent's objective is to navigate the spectrum of potential states, execute actions, secure rewards, and iteratively refine the value of the Q function using the Bellman equation (5.24) to incorporate newly acquired information. Through comprehensive exploration and continuous learning, the Q-value gradually aligns with the optimal value from the perspective of long-term rewards. Initially, the methodology sets $Q(s^0, \psi^0)$ random. Subsequently, an ϵ -greedy strategy balances exploration and exploitation during the learning phase. It is important to mention

Algorithm 5.3 Q-Learning based ABS Placement in a Virtual Space

-
- 1: **Input:** $\mathcal{X}, x_k, \vartheta_i, N_0, \alpha, \beta, \epsilon_0, M_{\text{itr}}, \eta$, decay rate $\tau = 0.9999$
 - 2: **Output:** x_k^{opt}
 - 3: Initialize $Q(s, \psi), \forall s \in \mathcal{S}, \forall \psi \in \Theta$ using initial or previously learned values
 - 4: Initialize the state s ($s \in \mathcal{S}$) and action a ($\psi \in \Theta$) based on initial or previous values
 - 5: Find current location of ABS and ISDs \mathcal{X}, x_k
 - 6: Calculate the distance $(d_{k,i})_{\forall i \in \mathcal{I}_k}$ between the given ABS and devices using \mathcal{X}, x_k
 - 7: Calculate achievable rate for all ISDs using (5.16) based on current $(d_{k,i})_{\forall i \in \mathcal{I}_k}$
 - 8: $s' =$ Mapping Function $\mathcal{F}(y)$ depending on $(d_{k,i})_{\forall i \in \mathcal{I}_k}$
 - 9: $r = \eta \sum_{n \in \mathcal{N}} \sum_{i \in \mathcal{I}_{k,n}} r_{k,i}(\mathcal{X}, x_k) + (1 - \eta) \min_{i \in \mathcal{I}_{k,n}} \{r_{k,i}(\mathcal{X}, x_k)\}$
 - 10: $Q(s, \psi) \leftarrow (1 - \alpha)Q(s, \psi) + \alpha [r + \beta \max_{\psi'} Q(s', \psi')]$
 - 11: $\epsilon = \epsilon_0$
 - 12: **for** $t = 1 : M_{\text{itr}}$ **do**
 - 13: Select action ψ' using epsilon-greedy policy (5.23)
 - 14: Compute new location x_k^{opt} based on recent action ψ'
 - 15: Compute distance $(d'_{k,i})_{\forall i \in \mathcal{I}_k}$ between the UAV and ISDs using $\mathcal{X}, x_k^{\text{opt}}$
 - 16: Calculate achievable rate $r_{k,i}$ for all ISDs using (5.16) based on current $(d'_{k,i})_{\forall i \in \mathcal{I}_k}$
 - 17: $s' =$ Mapping Function $\mathcal{F}(y)$ depending on $(d'_{k,i})_{\forall i \in \mathcal{I}_k}$
 - 18: Compute r with updated distance values using step 9
 - 19: Compute updated Q-values using step 10
 - 20: $d_{k,i} = (d'_{k,i})_{\forall i \in \mathcal{I}_k}$
 - 21: $\psi = \psi', s = s'$
 - 22: $\epsilon \leftarrow \tau \epsilon$
 - 23: **end for**
-

that ϵ should decline with time, i.e., $\epsilon \leftarrow \tau\epsilon$, where $\tau \in (0, 1)$.

$$Q(s^t, \psi^t) \leftarrow (1 - \alpha)Q(s^t, \psi^t) + \alpha \left[r^t + \beta \max_{\psi} Q(s^{t+1}, \psi) \right], \quad (5.24)$$

where, β is the discount factor, α is the learning rate, and t is the iteration count. The pseudo-code for ABS placement is given in Algorithm (5.3).

5.4.3 Power Allocation

The water-filling algorithm calculates the optimal power levels for each device within the MBS coverage [32]. A deep learning algorithm is implemented to allocate power among the sub-carriers within the UAV coverage area. Afterward, power is distributed among multiplexed users while QoS constraints are ensured. The details are as follows.

5.4.3.1 Power Allocation among Sub-carriers

$$\begin{aligned} \max \quad & \chi_{\text{Rate}} \\ \text{s.t.} \quad & C_1 : \sum_{n \in \mathcal{N}} P_n \leq P_k^{\max}, \end{aligned} \quad (5.25)$$

where P_n denotes the power allocated to the sub-carrier. Constraint C_1 shows that the cumulative sum of all the sub-carrier's power is the total power of the given ABS. Today, many researchers are using deep learning algorithms to allocate resources. The deep neural network is the most promising solution for allocating power between sub-carriers. Assuming the ISD grouping/sub-carrier allocation, the maximum channel gain among each ISD on the given sub-carrier is selected for a more balanced and fair PA.

We design a deep neural network with one input/output and two hidden layers (as shown in Fig. 5.5). Compared with traditional deep neural networks, the channel gain matrix is adopted as an input to train the model. The size of the input layer is set to be equivalent to the number of sub-carriers. Assuming a neural network with fifteen neurons in the first hidden layer and ten neurons in the second hidden

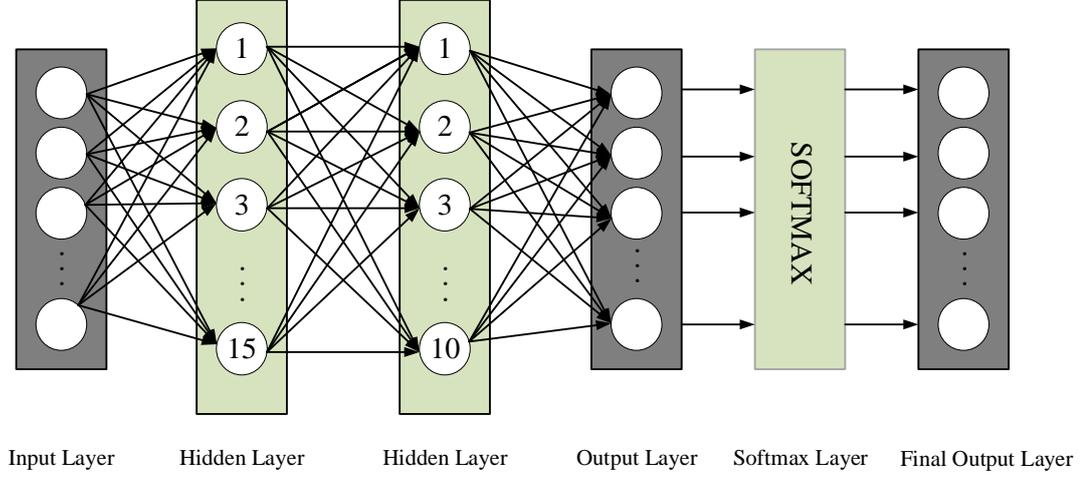


Figure 5.5: Deep Neural Network Infrastructure For Proposed Work.

layer, the network is trained to minimize the loss function, resulting in PA for the sub-carriers as the output, i.e., $\mathbf{P}_n = \tau_n P_k^{\max}$, where, \mathbf{P}_n denotes a power vector of P_n for each \mathcal{N} . Afterward, the fractional transmit (FT) PA provides an additional layer of optimization that ensures power efficiency and fairness among sub-carriers. The sigmoid function is utilized as an activation function, i.e., $y = 1/(1 + e^{-x})$ with input x and output y . A softmax layer is connected to normalize the powers in the final production, defined as $\text{softmax}(y)_n = \tau_n = \frac{e^{y_n}}{\sum_{t \in \mathcal{N}} e^{y_t}}$, where y_n is the power of the given sub-carrier and $\sum_{n \in \mathcal{N}} \tau_n = 1$. Using AdaGrad optimizer, weights, and biases are updated, while backpropagation minimizes the loss function. Given the primary objective to maximize the sum-rate, the loss function is formulated as the reciprocal of the sum-rate.

5.4.3.2 Power Allocation between ISDs

Once the power is allocated to each sub-carrier, it is distributed among the NOMA pairs. We assume that the ISD set on the given sub-carrier is $\mathcal{I}_{k,n} = \{i_1, i_2\}$, and the channel condition of ISD i_1 is better than the channel condition of ISD i_2 , i.e., $g_{k,i_1}^n \geq g_{k,i_2}^n$. According to the SIC detection order, ISD i_2 will directly demodulate its signal by treating the signal of ISD i_1 as interference noise. Therefore, the data

rate for ISD i_2 is given by

$$r_{k,i_1}^n = w_n \log_2 \left(1 + \frac{\varphi_{k,i_1}^n p_{k,i_1}^n g_{k,i_1}^n}{N_0} \right) - w_n \sqrt{\frac{\partial_{k,i_1}^n}{B}} \frac{Q^{-1}(\epsilon)}{\ln 2}, \quad (5.26)$$

where ∂_{k,i_1}^n is the channel dispersion for the ISD i_1 . The ISD i_1 must first perform SIC and decode its signal. The corresponding achievable data rate is given by

$$r_{k,i_2}^n = w_n \log_2 \left(1 + \frac{\varphi_{k,i_2}^n p_{k,i_2}^n g_{k,i_2}^n}{p_{k,i_1}^n g_{k,i_2}^n + N_0} \right) - w_n \sqrt{\frac{\partial_{k,i_2}^n}{B}} \frac{Q^{-1}(\epsilon)}{\ln 2}, \quad (5.27)$$

where ∂_{k,i_2}^n is the channel dispersion of ISD i_2 . Assuming that the PA coefficient of the device 1 is ζ_1 and the PA coefficient of device 2 is $\zeta_2 = (1 - \zeta_1)$, then $p_{k,i_1}^n = \zeta_1 P_n$ and $p_{k,i_2}^n = \zeta_2 P_n$, where $\zeta_2 > \zeta_1$. Then the total transmission rate for the given sub-carrier is $\gamma_k^n(\zeta_1, \zeta_2) = \sum_{i \in \mathcal{I}_{k,n}} \gamma_{k,i}^n, \forall i \in \{i_1, i_2\}$. Therefore,

$$\begin{aligned} \gamma_k^n(\zeta_1, \zeta_2) &= \left[w_n \log_2 \left(1 + \frac{\varphi_{k,i_1}^n \zeta_1 P_n g_{k,i_1}^n}{N_0} \right) - w_n \sqrt{\frac{\partial_{k,i_1}^n}{B}} \frac{Q^{-1}(\epsilon)}{\ln 2} \right] + \\ &\quad \left[w_n \log_2 \left(1 + \frac{\varphi_{k,i_2}^n \zeta_2 P_n g_{k,i_2}^n}{\zeta_1 P_n g_{k,i_2}^n + N_0} \right) - w_n \sqrt{\frac{\partial_{k,i_2}^n}{B}} \frac{Q^{-1}(\epsilon)}{\ln 2} \right], \\ &= \left[w_n \log_2 \left(\frac{\zeta_1 P_n g_{k,i_1}^n + N_0}{\zeta_1 P_n g_{k,i_2}^n + N_0} \right) - w_n \sqrt{\frac{\partial_{k,i_1}^n}{B}} \frac{Q^{-1}(\epsilon)}{\ln 2} \right] + \\ &\quad \left[w_n \log_2 \left(\frac{P_n g_{k,i_2}^n + N_0}{N_0} \right) - w_n \sqrt{\frac{\partial_{k,i_2}^n}{B}} \frac{Q^{-1}(\epsilon)}{\ln 2} \right]. \end{aligned} \quad (5.28)$$

Considering latter part of the (5.28), We set $f(\zeta_1, \zeta_2) = \left(\frac{\zeta_1 P_n g_{k,i_1}^n + N_0}{\zeta_1 P_n g_{k,i_2}^n + N_0} \right)$. By taking the first derivative, we observe that $f'(\zeta_1, \zeta_2) > 0$, where

$$f'(\zeta_1, \zeta_2) = \frac{(g_{k,i_1}^n - g_{k,i_2}^n) P_n N_0}{(\zeta_1 P_n g_{k,i_2}^n + N_0)^2}, \quad (5.29)$$

which shows that it is a monotonously increasing function. Hence, $\gamma_k^n(\zeta_1, \zeta_2)$ is also monotonously increasing function. We assume that the minimum rate requirement of the device i_2 is equivalent to its target rate, i.e., $r_{k,i_2}^n = \gamma_{\min}$, and the target SINR is given by $\partial_{\text{target}} = 2^{\gamma_{\min}} - 1$. We set the limit for ζ_1 and compute it by placing the target SINR value in (5.27).

$$\zeta_1 = \min \left(1, \frac{\partial_{\text{target}} (P_n g_{k,i_2}^n + N_0)}{g_{k,i_2}^n (1 + \partial_{\text{target}})} \right), \quad (5.30)$$

Algorithm 5.4 Power Allocation and Sum-rate Computation

-
- 1: **Input:** $\mathcal{I}_{k,n}, (g_{k,i}^n)_{i \in \mathcal{I}, n \in \mathcal{N}}, (\mathcal{N})_{n \in \mathcal{N}}, B, N_0, P_k^{\max}$, learning rate = 0.01.
 - 2: **Output:** $\mathcal{P}, P_n = P_n(1, 2, \dots, |N|)$
 - 3: **Step 1:** Power allocation among sub-carriers:
 - 4: Initialize the input layer, two hidden layers, output layer
 - 5: Initialize the random weights (w_1, w_2, w_3)
 - 6: Initialize the random biases (b_1, b_2, b_3)
 - 7: **Step 2:** Define the activation functions
 - 8: $\text{sigmoid}(x) \leftarrow \frac{1}{1+e^{-x}}$
 - 9: $\text{softmax}(y)_n \leftarrow \tau_n = \frac{e^{y_n}}{\sum_{t \in \mathcal{N}} e^{y_t}}$
 - 10: **Step 3:** Training loop
 - 11: **for** itr = 1 : 1000 **do**
 - 12: $a_1 = \text{sigmoid}((w_1 \times h) + b_1), a_2 = \text{sigmoid}((w_2 \times a_1) + b_2), y = \text{softmax}((w_3 \times a_2) + b_3)$
 - 13: Calculate PA, i.e., $P_n = y P_k^{\max}$
 - 14: Compute the loss function
 - 15: Perform back propagation
 - 16: Update weights and biases
 - 17: **end for**
 - 18: $\tau_n = \text{softmax}(y)_n$
 - 19: $P_n = \tau_n P_k^{\max}$
 - 20: $P_n = \text{FTPA}(P_k^{\max}, P_n)$ [32]
 - 21: **Step 4:** Power allocation between devices and sum-rate computing $\forall k \in \mathcal{K}$
 - 22: **for** $n = 1 : N$ **do**
 - 23: **for** $i = 1 : \mathcal{I}_{k,n}$ **do**
 - 24: Compute ζ_1 using (5.30), where $\zeta_2 = (1 - \zeta_1)$
 - 25: Allocate power to each device, assuming two devices per sub-carrier
 - 26: Compute sum-rate for all device on given sub-carrier
 - 27: **end for**
 - 28: **end for**
-

Algorithm 5.5 Proposed Algorithmic Solution for Intelligent Resource Allocation

- 1: **Input:** $(\mathcal{I})_{i \in \mathcal{I}}, (\mathcal{N})_{n \in \mathcal{N}}, (g_{k,i}^n)_{i \in \mathcal{I}, n \in \mathcal{N}}, \mathcal{X}, x_k, \vartheta_i, N_0, \alpha, \beta, \epsilon_0, M_{\text{itr}}, \eta$, decay rate $\tau = 0.9999, B, N_0, P_k^{\max}$, learning rate = 0.01
 - 2: **Output:** $\phi, \mathcal{I}_{k,n}, \mathbf{x}_k^{\text{opt}}, \mathcal{P}, \mathbf{P}_n, \mathbf{r}_{k,i}^n$
 - 3: Initialize the matrix $\mu = 0$ to keep the track of allocated devices per sub-carrier
 - 4: Set $G_{k,i}^n = g_{k,i}^n / (\bar{g}_i = \sum_{n \in \mathcal{N}} g_{k,i}^n / |\mathcal{I}_k|), \forall k \in \mathcal{K}, \forall i \in \mathcal{I}_k, \forall n \in \mathcal{N}$
 - 5: **Step 1:** ISD Grouping (ϕ)
 - 6: Find optimal clusters C_k via F-Statistics and Elbow method
 - 7: using step 1 in Algorithm (5.1)
 - 8: Set $\mathcal{I}_{k,n} = C_k$
 - 9: Allocate the sub-carriers per ISD using step 2 in Algorithm (5.1)
 - 10: Set $\phi_i = \lfloor \sum_{n \in \mathcal{N}} \mathcal{I}_{k,n} / \mathcal{I}_k \rfloor, \forall i \in \mathcal{I}_k$
 - 11: **Step 2:** Sub-carrier Allocation ($\mathcal{I}_{k,n}$)
 - 12: Perform the sub-carrier allocation using step 3 in Algorithm (5.1)
 - 13: **Step 3:** Reinforcement learning based optimal ABS placement ($\mathbf{x}_k^{\text{opt}}$)
 - 14: Compute optimal ABS position using Algorithm (5.2) and Algorithm (5.3)
 - 15: **Step 4:** Power Allocation and Sum-rate Computation
 - 16: Compute the optimal powers (\mathbf{P}_n) for each sub-carrier using steps 1-3 in
 - 17: Algorithm (5.4)
 - 18: $\tau_n = \text{softmax}(y)_n$
 - 19: $P_n = \tau_n P_k^{\max}$
 - 20: $P_n = \text{FTPA}(P_k^{\max}, P_n)$ [32]
 - 21: **Step 5:** Power allocation between devices and sum-rate computing $\forall k \in \mathcal{K}$
 - 22: Allocate the power and compute sum-rate for each ISD using steps 4 in
 - 23: Algorithm (5.4)
-

The pseudocode for PA is given in Algorithm (5.4). The comprehensive algorithmic solution for intelligent resource allocation integrates all the algorithms discussed in this chapter and is presented in Algorithm (5.5).

5.4.4 Complexity Analysis

The proposed solution is slightly more complex than other benchmarks. However, the primary goal of this research is to maximize the sum-rate of the edge network, so the solution's complexity is ignored. In the next section, the simulation and results of the proposed model are discussed in detail.

5.5 Simulation and Results

5.5.1 Simulations Settings

We have considered one MBS in a single cell, where the MBS, with radius $r_{\max} = 500$ meters, is located in the center of the cell. We consider \mathcal{J} cellular users randomly deployed within the cell, and \mathcal{K} auxiliary ABSs are deployed at optimal locations to extend their coverage area. We have considered the \mathcal{I} number of ISDs that need to be covered by the serving ABSs using \mathcal{N} sub-carriers. We assume that all ISDs adhere to the Federal Communications Commission guidelines. The minimum rate requirement for each ISD is assumed to be $\gamma_{\min} = 2$ bits/s/Hz [108].

The maximum power of the ABS is 1 watt, with a height of 50 – 80 meters. The network bandwidth is 10 MHz for comparison purposes with a carrier frequency of 5 MHz. Other parameters includes the learning rate $\alpha = 0.7$, blocklength for first hop $m_1 = 100$, blocklength for second hop $m_2 = 100$, discount factor $\beta = 0.5$, maximum number of episodes in Q-learning step $M_{\text{itr}} = 1000$, error probability $\epsilon = 0.001$, decay rate $\tau = 0.9999$, $a = 9.61$, $b = 0.16$, noise spectral density $N_0 = -174$ dBm/Hz, number of states $N_s = 2$, $\eta = 0.05$, with a single sub-carrier allocated to a maximum of $\chi_{\max} = 2$ ISDs, and each sub-carrier supporting a maximum of $\gamma_n = (1 \text{ (OMA)}, 2 \text{ (NOMA)})$ ISDs, unless otherwise stated. The pathloss model is defined in [108]. Detailed simulation parameters are also listed in Table 5.1.

Parameters	Values
Additive White Gaussian Noise N_0 [108]	-174 dBm/Hz
Discount factor β [9]	0.5
Decay rate τ [9]	0.9999
Error probability ϵ [9]	0.001
Height of the given ABS Z_k [9]	50-80 Meter
Leaning rate α [9]	0.7
Maximum number of ISDs allocated to single sub-carriers γ_n [108]	1 (OMA), 2 (NOMA)
Minimum rate requirement for each ISD γ_{\min} [108]	2 Bits/s/Hz
Maximum episodes M_{itr} [9]	1000
Maximum power of the UAV P_k^{\max} [108]	1 Watt
Number of states N_s [9]	2
Number of Cellular ISDs $ J $	5
Number of ISDs within serving UAV's coverage $ I_k $	15
Number of sub-carriers $ N $	30
Number of serving ABSs $ K $	2
Radius of the MBS r_{\max} [108]	500 Meters
Maximum numbers of sub-carriers allocated to single ISD χ_{\max} [108]	2
Step length of the given ABS Z_k [9]	5 Meter
Tuning weight η [9]	0.05

Table 5.1: Simulation Parameters.

5.5.2 Performance Comparison

Fig. 5.6 shows the convergence of the normalized reward for UAV 1 and UAV 2 over 1000 iterations using a Q-learning based intelligent positioning scheme. Furthermore, the following figure demonstrates how reward values increase and

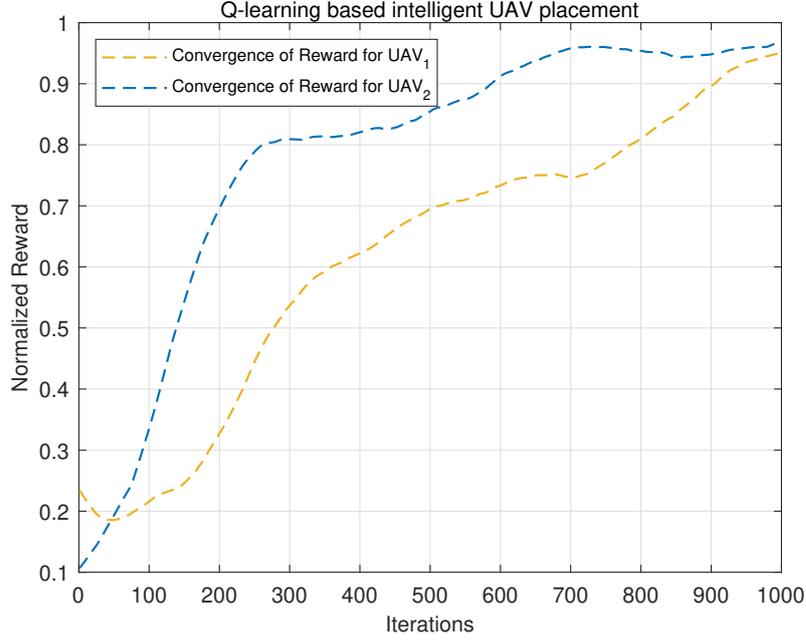


Figure 5.6: Convergence of Normalized Reward for UAVs using Q-learning based Intelligent Positioning over Iterations.

stabilize over time, indicating the effectiveness of the Q-learning algorithm in optimizing UAV positioning. Our proposed algorithmic solution can converge to the optimum in a few iterations, indicating a low computational solution. This implies that UAVs only need a few moves to reach their optimal positions.

Fig. 5.7 illustrates the sum-rate performance of the NOMA-DeepFusion-PA [Optimal UAV position] scheme for both UAVs with different power levels and PA coefficient for the weak ISD ζ_1 , which has an inferior channel condition compared to other devices allocated to the given sub-carrier. The sum-rate increases as the PA coefficient ζ_1 increases from 0 to 0.8 for both UAVs, indicating a positive correlation between ζ_1 and the sum-rate. For both UAVs, the sum-rate is consistently higher when the power level is $P_k^{\max} = 1$ or 3 watts. The rate of increase in the sum-rate is more significant for the higher power level, suggesting that higher PA enhances the sum-rate more effectively. Therefore, since a weak ISD is the farthest from the UAV, this device should be assigned the highest power to ensure fairness. Based on the

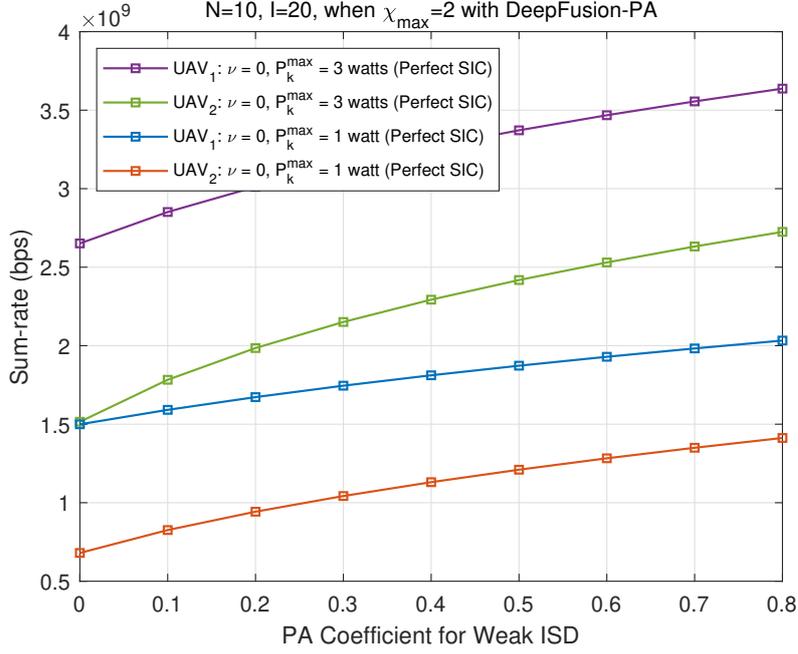


Figure 5.7: Sum-rate of NOMA-DeepFusion-PA [Optimal UAV position] Scheme under different Power Levels and Power Allocation Coefficients ζ_1 for both UAVs with Perfect SIC.

results shown in this figure, the PA coefficient for the weak ISD, i.e., $\zeta_1 = 0.6$, is used for further analysis. In general, increasing the PA coefficient and maximum power level positively impacts the sum-rate performance of the NOMA-DeepFusion-PA [Optimal UAV position] scheme.

Fig. 5.8 evaluates the performance of the proposed NOMA-DeepFusion-PA [Optimal UAV position] scheme against the varying power constraint P_n for each NOMA sub-carrier under different SIC conditions, i.e., perfect SIC and imperfect SIC for both UAVs. It shows the highest sum-rate across all power levels for both UAVs, indicating that perfect SIC significantly enhances the sum-rate performance. This positive correlation indicates that higher transmit power enhances the sum-rate performance in NOMA schemes. When power increases, that is, from 0.4 to 0.8, the performance gap between perfect SIC and imperfect SIC becomes notably greater, especially for $\nu = 10^0$, and $\nu = 10^{-1}$. In other words, the sum-rate decreases as

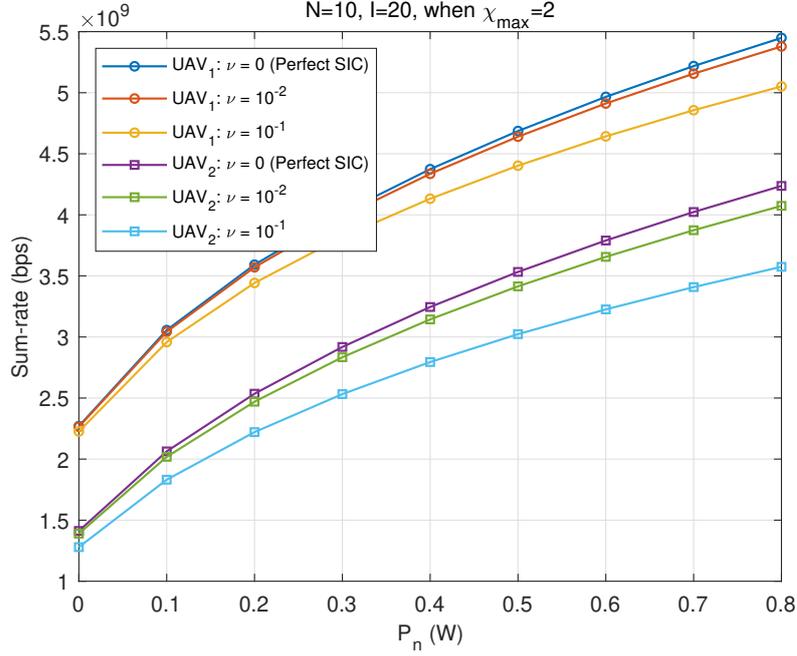


Figure 5.8: Performance Comparison of NOMA-DeepFusion-PA [Optimal UAV position] Scheme under Perfect and Imperfect SIC Conditions with Residual of Coefficient ν for both UAVs.

the residual coefficient ν increases, highlighting the impact of imperfect SIC on the system's performance. Specifically, the sum-rate is higher for $\nu = 10^{-2}$ compared to $\nu = 10^{-1}$, demonstrating that a lower imperfection in SIC leads to better performance. In conclusion, UAV 1 consistently achieves a higher sum-rate than UAV 2 in all scenarios and power levels. Although UAV 2's sum-rate is lower than UAV 1's, the trend is similar. The numerical results show that the sum-rate for UAV 1 with $\nu = 0$ (perfect SIC) is 7.86442% higher than with $\nu = 10^{-1}$ (imperfect SIC) and 18.5332% higher with $\nu = 0$ (perfect SIC) than with $\nu = 10^{-1}$ (imperfect SIC) for UAV 2, when $P_n = 0.8$ (W).

Fig. 5.9 compares the sum-rate of the proposed NOMA-DeepFusion-PA [Optimal UAV position] against the benchmark OMA scheme under various transmit power levels ($1 \leq P_k^{\max} \leq 5$ (W)). The analysis is carried out under perfect SIC conditions, with the system parameters $\mathcal{N} = 10$, $\mathcal{I}_k = 20$, and $\chi_{\max} = 2$. The results are derived

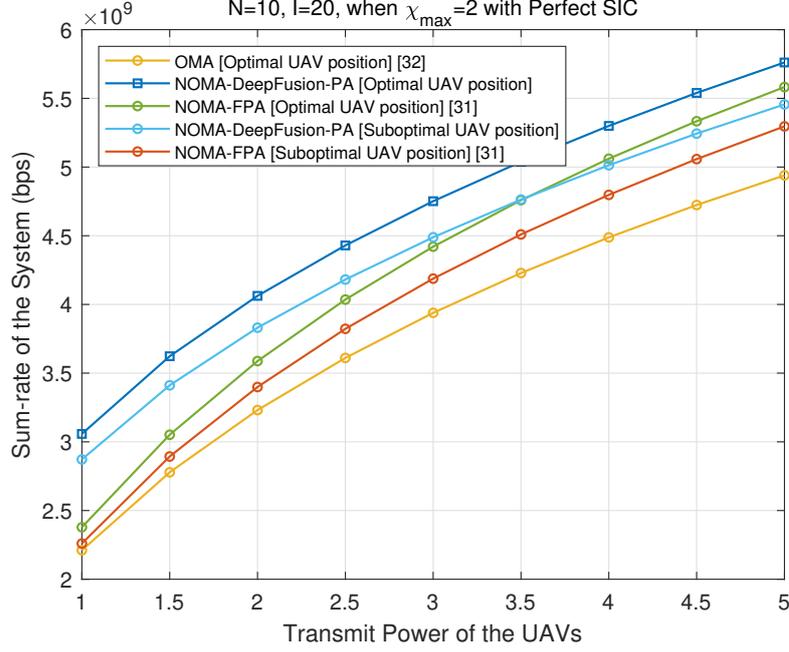


Figure 5.9: System Sum-rate under different Power Allocation and UAV Positioning Strategies with Perfect SIC for OMA and NOMA Systems.

by averaging the Monte-Carlo simulations with 10^5 realizations at each iteration. It can be observed that the proposed NOMA schemes are superior to conventional OMA regardless of the position of the UAV and PA schemes employed by the system. This performance is attributed to the benefits of NOMA, including superposition coding at the transmitter end and SIC at the receiver end, which allow efficient spectrum utilization and higher sum-rates. According to the results, the proposed NOMA-DeepFusion-PA [Optimal UAV position] achieves a sum rate that is 38.3119% higher compared to the legacy OMA scheme.

The proposed NOMA-DeepFusion-PA [Optimal UAV position] consistently performs better than NOMA-FPA [Optimal UAV position]. This demonstrates the advantage of the DeepFusion PA, which optimally distributes the power among sub-carriers, enhancing the overall system's performance. The figure shows that the rate is higher when the UAVs are in optimal positions compared to suboptimal positions. This is mainly due to the intelligent placement based on reinforcement learning,

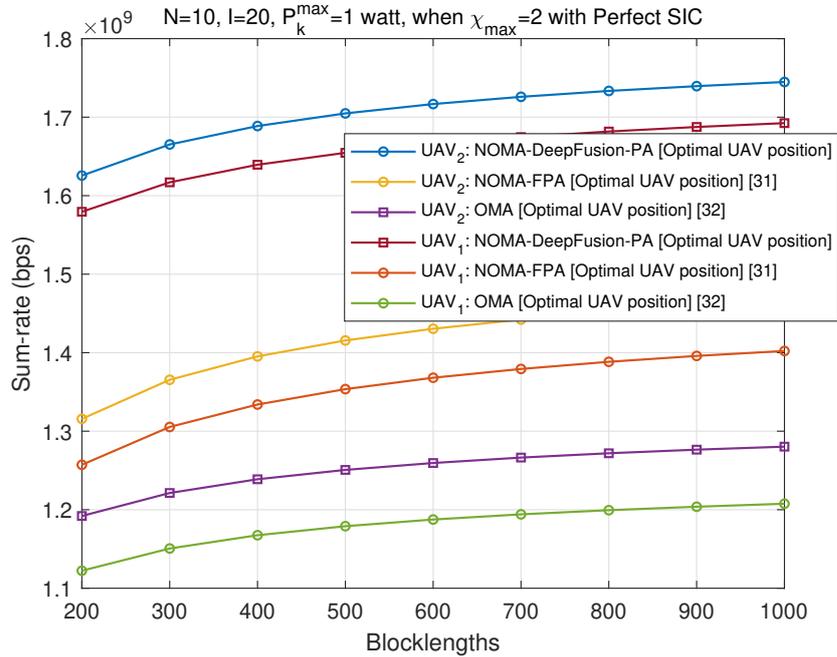


Figure 5.10: Evaluation of Blocklength Impact on Sum-rate for NOMA-DeepFusion-PA with Optimal UAV Positioning Compared to Benchmark Schemes on both UAVs.

which finds optimal positions relative to NOMA devices to maximize the utility function of maximizing the rate. Specifically, the NOMA-DeepFusion-PA [Optimal UAV position] outperforms all other schemes, indicating the combined benefit of optimal PA and UAV positioning. Overall, the sum-rate of the systems improves for all schemes by increasing the transmit power of UAVs. The rate increase is more pronounced in the NOMA-DeepFusion-PA [Optimal UAV position], further highlighting its efficiency in leveraging higher power levels to enhance performance. Numerical results reveal that the system's sum-rate for NOMA-DeepFusion-PA [Optimal UAV position] is 28.5762% higher than NOMA-FTPA [Optimal UAV position], and 6.46565% higher than NOMA-DeepFusion-PA [Suboptimal UAV position].

Fig. 5.10 evaluates the sum-rate performance of the proposed NOMA-DeepFusion-PA scheme with optimal UAV positioning in comparison to the NOMA-FTPA and the benchmark OMA scheme under perfect SIC and over a range of blocklengths

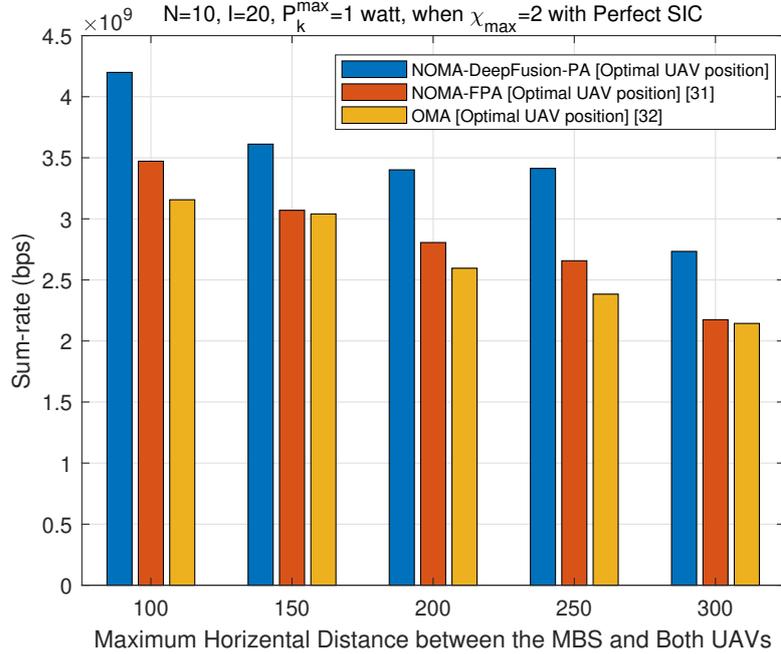


Figure 5.11: Sum-rate of NOMA-DeepFusion-PA, NOMA-FTPA, and OMA Schemes with Optimal UAV Positioning across different Horizontal Distances.

ranging from 200 to 1000. The key observations are as follows: 1) NOMA-DeepFusion-PA [Optimal UAV position] consistently outperforms the other schemes in all blocklengths, indicating the effectiveness of DeepFusion PA and optimal UAV positioning in improving system performance. 2) NOMA-FPA [Optimal UAV position] performs better than the legacy OMA scheme but is outperformed by NOMA-DeepFusion-PA, demonstrating the advantage of the DeepFusion approach over fixed PA.

It is also noted that an increase in blocklength results in an improved system rate across all schemes. This positive trend suggests that longer blocklengths enhance the system's ability to process and transmit data, thereby increasing the sum-rate. This rate increase is more evident for NOMA-DeepFusion-PA, indicating that it leverages the advantages of a longer blocklength more effectively than the other approaches. For both UAVs 1 and 2, the NOMA-DeepFusion-PA with optimal positioning achieves the highest sum-rate, followed by NOMA-FTPA and then

the OMA scheme. Importantly, this performance gap widens as the blocklength increases, further emphasizing the superior scalability and efficiency of the proposed approach. Numerical results validate that the sum-rate for NOMA-DeepFusion-PA [Optimal UAV position] is 9.0259% higher than NOMA-FPA [Optimal UAV position] and 36.2806% higher than OMA for UAV 2 at a blocklength of 1000. Similarly, the system's sum-rate for NOMA-DeepFusion-PA [Optimal UAV position] is 20.696% higher than NOMA-FPA [Optimal UAV position] and 40.1457% higher than OMA for UAV 1 at a blocklength of 1000.

Fig. 5.11 compares the sum-rate performance of three schemes: NOMA-DeepFusion-PA, NOMA-FTPA, and OMA, across different horizontal distances between MBS and UAVs. The key observations are: 1) NOMA-DeepFusion-PA scheme achieves the highest sum-rate across all distances, demonstrating the superior efficacy of Deepfusion PA and optimal UAV positioning in enhancing the system's performance. The sum-rate generally decreases as the horizontal distance increases from 100 to 300 meters. This trend suggests that the greater distance reduces the system's ability to maintain higher data rates because of increased path loss, including shadowing, fading, and signal degradation. 2) NOMA-FTPA schemes outperform legacy OMA schemes; however, short of the performance achieved by NOMA-DeepFusion-PA. This indicates the advantage of NOMA over OMA, particularly when DeepFusion PA is used.

The performance gap between NOMA-DeepFusion-PA against benchmarks becomes more evident as the distance increases. The increasing gap demonstrates the scalability and robustness of the DeepFusion-PA approach in sustaining higher data rates over extended distances. Moreover, the figure illustrates that optimal positioning is essential in maximizing the sum-rate, as evident from the consistently outperforming NOMA over OMA. According to numerical results, the system's sum-rate for NOMA-DeepFusion-PA [Optimal UAV position] is 20.95179% higher than NOMA-FPA [Optimal UAV position] and 33.041% higher than OMA, when the distance is around 100 meters. Likewise, the system's sum-rate for NOMA-

DeepFusion-PA [Optimal UAV position] is 25.7152% greater than NOMA-FPA [Optimal UAV position] and 27.5098% greater than legacy OMA when the distance is 300 meters.

5.6 Summary

This article introduces a novel framework for the optimal deployment of UAVs and resource allocation using FBL coding in URLLC networks. This framework optimizes ISD grouping/sub-carrier allocation, UAV placement, and PA, subject to delay, reliability, and other QoS requirements to maximize the system's performance. The inherited non-convex optimization problem is divided into solvable sub-problems for optimal solutions. The proposed solution involves the following steps: 1) integrating the elbow method with the F-test method to control the maximum number of devices on each sub-carrier, 2) implementing a Q-learning (type of reinforcement learning) technique for optimal UAV placement, and 3) DeepFusion based PA by integrating the deep-learning PA method with the fractional transmit PA method to provide an additional layer of optimization to ensure power efficiency and fairness among sub-carriers. The comparative analysis demonstrates that our proposed algorithmic solution achieves a performance improvement of 28.5762% than NOMA-FPA [Optimal UAV position], and 38.3119% higher than traditional OMA [Optimal UAV position].

Chapter 6

Conclusions and Future Work

6.1 Conclusion

This concluding chapter of the thesis evaluates how the research findings align with the intended objectives and outlines potential directions for future research. This research primarily addresses optimal resource optimization solutions by utilizing innovative, sustainable, and energy-efficient resource scheduling in integrated aerial-terrestrial networks to enhance QoS and latency aspects, offering significant improvements over existing state-of-the-art methods. The key contributions of this thesis are summarized as follows.

Chapter 3 investigates a mixed-integer non-linear programming problem to optimize resource allocation in an integrated aerial-terrestrial wireless network for maximizing the system sum-rate. A novel low-complexity algorithm is proposed, which applies alternating optimization and a two-step projected gradient descent-based strategy to optimize the resource allocation policy while incorporating delay, reliability, and QoS constraints. Simulation results demonstrate that the proposed algorithm requires 1600 times less computational cost than baseline approaches. This chapter concludes that NOMA with optimal blocklength outperforms OMA with optimal blocklength. NOMA with fixed or random blocklengths surpasses their respective OMA counterparts in their corresponding scenarios.

Chapter 4 emphasizes the significance of task dependencies between devices that influence task offloading and resource allocation decisions. A closed-form solution is derived to optimize transmit power, CPU frequency, and task offloading policy, thereby minimizing the energy-time cost. This chapter concludes that the proposed solution can handle multiple tasks, sometimes exceeding 100. Additionally, the system's energy-time cost closely approximates the brute-force algorithm's, ensuring equitable energy-time costs for IoT devices regardless of task loads. Our algorithmic solution converges in five iterations, a stark improvement over the brute-force algorithm requiring 4096 iterations.

Chapter 5 investigate intelligent approaches to provide optimal resource allocation and UAV deployment aimed at maximizing the achievable rate. This chapter

leverages unsupervised learning to optimize ISD grouping/sub-carrier allocation, UAV placement, and deep-learning based power allocation, subject to delay, reliability, and other QoS constraints. Additionally, it ensures power efficiency and fairness among sub-carriers, thereby improving overall system performance. A mathematical definition is also derived to calculate the value of the F-test for a given number of clusters. Simulation results validate that the proposed solutions outperform well-recognized benchmark schemes.

This research underscores the potential for optimization techniques in integrating aerial-terrestrial networks and their application to future-generation networks, setting the stage for further advancements in resource allocation, system efficiency, and QoS enhancement.

6.2 Limitations of the Study

- Many URLLC algorithms have been developed in UAV-based communication systems under the assumption of perfect hardware. However, various hardware imperfections present in practical UAV systems must be considered. The computer simulations for each part of this Ph.D. thesis represent the most feasible options.
- Current research integrates NOMA with UAV-based communication. However, when NOMA is used, the application of SIC poses a significant security risk, as one user can decode the signal intended for another. Therefore, it is crucial to study this area from a security perspective to fully leverage the advantages of NOMA-based wireless networks without compromising transmission integrity.
- Current research assumes perfect SIC. However, a more realistic understanding of channel conditions or an accurate channel prediction is challenging. It requires a pilot-based training process, which results in increased overhead and complexity, ultimately leading to delayed feedback. Consequently, various

robust resource allocation strategies should be developed to effectively address the challenges of adopting the NOMA concept in practical scenarios.

6.3 Future Improvements

It is anticipated from this thesis that some promising areas can be further developed in future research. These are summarized as follows.

- **Innovative synergy:** It is clear that UAVs will perform for aerial and ground users, with computational resources distributed on the ground and in the air. However, matching the time and spatially varying AI demands with distributed data supplies remains uncharted territory.
- **Secure wireless communication:** It is obvious that UAVs will be extensively deployed and seamlessly integrated into terrestrial communication systems. Relying solely on onboard embedded sensors for sensing will be insufficient. Therefore, integrated sensing, which combines UAV-embedded and infrastructure-based sensing, is required. Moreover, future networks must develop advanced security measures to protect against sophisticated cyber threats and unauthorized access.
- **Empowering collaborative intelligence:** we study the integration of machine learning in future wireless networks. However, there is a strong need for interplay between AI and edge networks, i.e., intelligent user access, adaptive decision-making, and more realistic knowledge of channel conditions or accurate channel prediction for AI-aided wireless transmission. Moreover, machine learning's predictive capability allows better resource allocation, capacity planning, and optimization of network parameters to meet changing demands.

References

- [1] M. Awais, H. Pervaiz, M. Jamshed, W. Yu, and Q. Ni, “Enhancing URLLC in Integrated Aerial Terrestrial Networks: Design Insights and Performance Trade-offs,” in *IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Belfast, Uk, Jun 14-17, pp. 548-553, 2022.
- [2] P. Rahimi, C. Chrysostomou, H. Pervaiz, V. Vassiliou, and Q. Ni, “Joint Radio Resource Allocation and Beam Forming Optimization for Industrial Internet of Things in Software-Defined Networking-Based Virtual Fog-Radio Access Network 5G-and-Beyond Wireless Environments,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 4198-4209, 2022.
- [3] M Series, “IMT Traffic Estimates for the Years 2020 to 2030,” *Report ITU-RM.2370-0*, 2024. Available at: <https://www.itu.int/pub/R-REP-M.2370-2015> (Accessed: 03 September 2024).
- [4] L. Xiaocui, Z. Zhou, Q. He, Z. Shi, W. Gaaloul, and S. Yangui. “Re-Scheduling IoT Services in Edge Networks,” *IEEE Transactions on Networks and Service Management*, vol. 20, no. 3, pp 1-1, 2023.
- [5] McLellan, P. EDPS: Transitioning from 5G to 6G, Semiconductor Engineering. Available at: <https://semiengineering.com/edps-transitioning-from-5g-to-6g/> (Accessed: 03 February 2025).
- [6] A. Bhowmick, S. Dhar Roy, and S. Kundu, “Throughput Maximization of a UAV Assisted CR Network With NOMA-Based Communication and Energy-

- Harvesting,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 1, pp. 362-374, 2022.
- [7] W. New, C.Y. Leow, K. Navaie, and Z. Ding, “Aerial-Terrestrial Network NOMA for Cellular-Connected UAVs,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 6, pp. 6559-6573, 2022.
- [8] L. Qian, Y. Wu, F. Jiang, N. Yu, W. Lu, and B. Lin, “NOMA Assisted Multi-Task Multi-Access Mobile Edge Computing via Deep Reinforcement Learning for Industrial Internet of Things,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5688-5698, 2021.
- [9] T. N. Guo, “Connectivity-Aware Fast Network Forming-aided by Digital Twin for Emergency Use,” *IEEE conference on Computing Communication Workshops*, pp. 1-6, 2022.
- [10] Q. Wu, J. Xu, Y. Zeng, D.W.K. Ng, N. Al-Dhahir, R. Schober, and A.L. Swindlehurst, “A Comprehensive Overview on 5G-and-Beyond Networks With UAVs: From Communications to Sensing and Intelligence,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 10, pp. 2912-2945, 2021.
- [11] L. Baozhu, F. Hou, G. Yang, H. Zhao, and S. Chen, “Data Analysis-Oriented Stochastic Scheduling for Cost-Efficient Resource Allocation in NFV Based MEC Network,” *IEEE Transactions on Vehicular Technology*, vol. 72, no. 5, pp 1-14, 2023.
- [12] C. Yao, Y. Zhang, H. Luo, T. Chen, G. Chen, H. Chen, Y. Wang, H. Wei, and C. Chou, “Management and Orchestration of Edge Computing for IoT: A Comprehensive Survey,” *IEEE Internet of Things Journal*, vol. 10, no. 16, pp 14307-14331, 2023.
- [13] A. A. Nasir, H. D. Tuan, H. H. Nguyen, M. Debbah, and H. V. Poor, “Resource Allocation and Beamforming Design in the Short Blocklength Regime

-
- for URLLC,” *IEEE Transactions Wireless Communication*, vol. 20, no. 2, pp. 1321–1335, Feb. 2021.
- [14] L. Qi, X. Xu, X. Wu, Q. Ni, Y. Yuan, and X. Zhang. “Digital-Twin-Enabled 6G Mobile Network Video Streaming Using Mobile Crowdsourcing,” *IEEE Journal on Selected Area in Communication (JSAC)*, Vol. 41, no. 10, pp. 3161 - 3174, Oct 2023.
- [15] Z Yining, H. Zhang, X. Liu, C. Zhao, and F. Xu, “Digital Twin-enabled Deep Reinforcement Learning for Joint Scheduling of Ultra-reliable Low-latency Communication and Enhanced Mobile Broadband: A Reliability-Guaranteed Approach,” *Transactions on Emerging TeleCommunication Technology* Vol. 34, no. 3, pp. e4705, 2023.
- [16] S. M. Muhammad, M. A. Tariq, J. Seo, and D. Kim, “An Overview of 3GPP Release 17 & 18 Advancements in the Context of V2X Technology,” *International Conference on Artificial Intelligence in Information and Communication*, pp. 057-062, 2023.
- [17] Z. Hosein, N. Gholipour, M. R. Mili, M. Rasti, H. Tabassum, and E. Hossain, “Resource Management for Multiplexing eMBB and URLLC Services Over RIS-Aided THz Communication,” *IEEE Transactions on Communication*, Vol. 71, no. 2, pp. 1207–1225, 2023.
- [18] W. Yuntao, Z. Su, S. Guo, M. Dai, T. H. Luan, and Y. Liu., “A Survey on Digital Twins: Architecture, Enabling Technologies, Security and Privacy, and Future Prospects,” *IEEE Internet of Things Journal*, Vol. 10, no. 17, pp 1-1, 2023.
- [19] L. Lei, D. Yuan, C. Ho, and S. Sun, “Power and Channel Allocation for Non-Orthogonal Multiple Access in 5G Systems: Tractability and Computation,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8580-8594, 2016.

- [20] B. Su, Z. Qin, and Q. Ni, “Energy Efficient Uplink Transmissions in LoRa Networks,” *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 4960-4972, 2020.
- [21] J. Shi, and L. Yang, “Novel Subcarrier-Allocation Schemes for Downlink MC DS-CDMA Systems,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5716-5728, 2014.
- [22] J. Yan, S. Bi, Y. J. Zhang, M. Tao, “Optimal Task Offloading and Resource Allocation in Mobile-edge computing with Inter-user Task Dependency,” *IEEE Transactions on Wireless Communication*, vol. 19, no. 1, pp. 235–250.
- [23] J. Yan, S. Bi, and Y. A. Zhang. “Optimal Offloading and Resource Allocation in Mobile Edge Computing with Inter-User Task Dependency,” *IEEE Global Communication Conference (GLOBECOM)*, pp. 1-8. IEEE, 2018.
- [24] D. V. Huynh, V. Nguyen, S. R. Khosravirad, V. Sharma, O. A. Dobre, H. Shin, and T. Q. Duong, “URLLC Edge Networks With Joint Optimal User Association, Task Offloading, and Resource Allocation: A Digital Twin Approach,” *IEEE Transactions on Communication*, vol. 70, no. 11, pp. 7669–7682, 2022.
- [25] Y. Chen, J. Zhao, J. Hu, S. Wan, and J. Huang, “Distributed Task Offloading and Resource Purchasing in NOMA-enabled Mobile Edge Computing: Hierarchical Game Theoretical Approaches,” *ACM Transactions in Embedded Computing Systems*, vol. 23, no. 1, pp. 1–28, 2023.
- [26] Y. Li, D. V. Huynh, T. Do-Duy, E. Garcia-Palacios, and T. Q. Duong, “Unmanned Aerial Vehicle-aided Edge Networks with Ultra-reliable Low-latency Communications: A Digital Twin Approach,” *IET Signal Processing*, vol. 16, no. 8, pp: 897-908, 2022.
- [27] Y. Li, H. Zhang, K. Long, C. Jiang, and M. Guizani, “Joint Resource Allocation and Trajectory Optimization with QoS in UAV-based NOMA Wireless Networks,”

-
- IEEE Transactions on Wireless Communication*, vol. 20, no. 10, pp. 6343-6355, 2021.
- [28] X. Liu, Z. Liu, and M. Zhou, “Fair Energy-Efficient Resource Optimization for Green Multi-NOMA-UAV assisted Internet of Things,” *IEEE Transactions on Green Communication and Networking*, vol.7, no. 2, pp. 904-915, 2021.
- [29] Q. Huang, W. Wang, W. Lu, N. Zhao, A. Nallanathan, and X. Wang, “Resource Allocation for Multi-Cluster NOMA-UAV Networks,” *IEEE Transactions on Communications*, vol. 70, no. 12, pp. 8448–8459, 2022.
- [30] T. M. C. Chu, H. Zepernick, and T. Q. Duong. “NOMA-Based Full-Duplex UAV Network with K-Means Clustering for Disaster Scenarios,” *In 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, pp. 1-7. IEEE, 2022.
- [31] W. Bai, T. Yao, H. Zhang, V. C. M. Leung, “Research on Channel Power Allocation of Fog Wireless Access Network based on NOMA,” *IEEE Access*, vol. 7, pp. 32867–32873, 2019.
- [32] M. Awais, H. Pervaiz, M. A. Jamshed, W. Yu, and Q. Ni, “Energy-Aware Resource Optimization for Improved URLLC in Multi-hop Integrated Aerial Terrestrial Networks,” *IEEE Transactions on green Communication and Networking*, vol. 8, no. 11, pp. 252-264, 2023.
- [33] A. Benjebbour, Anxin Li, Y. Saito, Y. Kishiyama, A. Harada, and T. Nakamura, “System-Level Performance of Downlink NOMA for Future LTE Enhancements,” *IEEE Globecom Workshops (GC Wkshps)*, pp. 706-710, 2014.
- [34] Y. Fu, L. Salaun, C. W. Sung, and C. S. Chen, “Subcarrier and Power Allocation for the Downlink of Multicarrier NOMA Systems,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 12, pp. 11833–11847, 2018.

- [35] L. Salaun, M. Coupechoux, and C. S. Chen, “Weighted Sum-rate Maximization in Multi-carrier NOMA with Cellular Power Constraint,” *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pp. 451-459, 2019.
- [36] L. Salaun, M. Coupechoux, and C. S. Chen, “Joint Subcarrier and Power Allocation in NOMA: Optimal and Approximate Algorithms,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 2215–2230, 2020.
- [37] M. Amjad, L. Musavian, and S. Aissa, “Link-Layer Rate of NOMA with Finite Blocklength for Low-Latency Communications,” *IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, UK*, pp. 1-6, 2020.
- [38] W. J. Ryu and S. Y. Shin, “Power Allocation for URLLC Using Finite Blocklength Regime in Downlink NOMA Systems,” *In 2019 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 770-773, 2019.
- [39] H. Ren, C. Pan, Y. Deng, M. ElKashlan, and A. Nallanathan, “Joint Power and Blocklength Optimization for Ultra-reliable Low-latency Communication in a Factory Automation Scenario,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 1786–1801, 2020.
- [40] L. Lyu, C. Chen, N. Cheng, S. Zhu, X. Guan, and X. Shen, “NOMA-assisted On-demand Transmissions for Monitoring Applications in Industrial IoT Networks,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp.12264-12276, 2020.
- [41] X. Pei, H. Yu, X. Wang, Y. Chen, M. Wen, and Y.-C. Wu, “NOMA-Based Pervasive Edge Computing: Secure Power Allocation for IoV,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 5021–5030, 2020.

- [42] L. Haiquan, Y. Zeng, S. Jin, and R. Zhang. “Enabling Panoramic Full-Angle Reflection via Aerial Intelligent Reflecting Surface,” *IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1-6, 2020.
- [43] D. Van Huynh, T. Do-Duy, L. D. Nguyen, M. T. Le, N. S. Vo, and T. Q. Duong, “Real-time Optimized Path Planning and Energy Consumption for Data Collection in Unmanned Aerial Vehicles-aided Intelligent Wireless Sensing,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 4, pp. 2753–2761, 2022.
- [44] Z. Kang, C. You, and R. Zhang, “Learning for Multi-UAV Relaying: A Gibbs Sampling Approach,” *IEEE International Conference on Communications Workshops (ICC Workshops)*, 2020.
- [45] J. Wang, M. Liu, J. Sun, G. Gui, H. Sari, and F. Adachi, “Multiple Unmanned-Aerial-Vehicles Deployments and User Pairing for Non-orthogonal Multiple Access Schemes,” *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1883–1895, 2021.
- [46] S. Arzykulov, A. Celik, G. Nauryzbayev, and A. Eltawil, “UAV-Assisted Cooperative & Cognitive NOMA: Clustering, and Resource Allocation,” *IEEE Transactions on Cognitive Communication and Networks*, vol. 8, no. 1, pp. 263-281, 2022.
- [47] D. Diao, B. Wang, K. Cao, R. Dong, and T. Cheng, “Enhancing Reliability and Security of UAV-enabled NOMA Communications with Power Allocation and Aerial Jamming,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 8, pp. 8662-8674, 2022.
- [48] S. Mirbolouk, M. Valizadeh, M. Amirani, and S. Ali, “Relay Selection and Power Allocation for Energy Efficiency Maximization in Hybrid Satellite-UAV Networks With CoMP-NOMA Transmission,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 5087-5100, 2022.

- [49] F. Salehi, N. Neda, M.H. Majidi, and H. Ahmadi, “Cooperative NOMA-Based User Pairing for URLLC: A Max–Min Fairness Approach,” *IEEE Systems Journal*, vol. 16, no. 3, pp. 3833-3843, 2021.
- [50] M. Amjad, L. Musavian, and S. Aissa, “NOMA Versus OMA in Finite Blocklength Regime: Link-Layer Rate Performance,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 16253-16257, 2020.
- [51] H. Lei, C. Zhu, K. Park, W. Lei, I. Ansari, and T. Tsiftsis, “On Secure NOMA-Based Terrestrial and Aerial IoT Systems,” *IEEE IoTs Journal*, vol. 9, no. 7, pp. 5329-5343, 2022.
- [52] T. Vu, T. Nguyen, T. Nguyen, B. Vo-Nguyen, and S. Kim, “Short-Packet Communications in NOMA-CDRT IoT Networks With Cochannel Interference and Imperfect SIC,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 5552-5557, 2022.
- [53] L. Yuan, N. Yang, F. Fang, and Z. Ding, “Performance Analysis of UAV-Assisted Short-Packet Cooperative Communications,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 4, pp. 4471–4476, 2022.
- [54] B. Yu, X. Guan, and Y. Cai, “Joint Blocklength and Power Optimization for Half Duplex Unmanned Aerial Vehicle Relay System with Short Packet Communications,” *International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 981-986, 2020.
- [55] Y. Hu, G. Sun, G. Zhang, M. C. Gursoy, and A. Schmeink, “Optimal Resource Allocation in-Ground Wireless Networks Supporting Unmanned Aerial Vehicle Transmissions,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8972–8984, 2020.
- [56] W. Qi, Q. Song, and A. Jamalipour, “Energy-Efficient Resource Allocation for UAV-Assisted Vehicular Networks With Spectrum Sharing,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7691–7702, 2022.

- [57] J. Lei, T. Zhang, X. Mu, and Y. Liu, "NOMA for STAR-RIS Assisted UAV Networks," *IEEE Transactions on Communications*, vol. 72, no. 3, pp. 1732-1745, 2024.
- [58] M. Katwe, K. Singh, P. K. Sharma, and C. P. Li, "Energy Efficiency Maximization for UAV-Assisted Full-Duplex NOMA System: User Clustering and Resource Allocation," *IEEE Transactions on Green Communication and Networking*, vol. 6, no. 2, pp. 992-1008, 2022.
- [59] J. Wang, Z. Na, and X. Liu, "Collaborative Design of Multi-UAV Trajectory and Resource Scheduling for 6G-enabled Internet of Things," *IEEE Internet of Things Journal*, vol. 8, no. 20, pp. 15096-15106, 2020.
- [60] W. Ding, Z. Yang, M. Chen, J. Hou, and M. Shikh-Bahaei, "Resource Allocation for UAV Assisted Wireless Networks with QoS Constraints," *IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1-7, 2020.
- [61] K. Chen, Y. Wang, J. Zhao, X. Wang, and Z. Fei, "URLLC Oriented Joint Power Control and Resource Allocation in UAV-assisted Networks," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 10103-10116, 2021.
- [62] Z. Yao, W. Cheng, W. Zhang, and H. Zhang, "Resource Allocation for 5G-UAV-Based Emergency Wireless Communication," *IEEE Journal on Selected Areas in Communication*, vol. 39, no. 11, pp. 3395-3410, 2021.
- [63] G. Yang and Y. Yao, "Resource Allocation Control of UAV-Assisted IoT Communication Device," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 11, pp. 13341-13349, 2023.
- [64] Z. Liu, X. Liu, Y. Liu, Victor, and T. S. Durrani, "UAV Assisted Integrated Sensing and Communication for Internet of Things: 3D Trajectory Optimization And Resource Allocation," *IEEE Transactions on Wireless Communications*, pp. 1-1, 2024.

- [65] H. Di, X. Zhu, Z. Liu, and X. Tu, “Joint Blocklength and Trajectory Optimizations for URLLC-enabled UAV Relay System,” *IEEE Communication Letters*, vol. 28, no. 1, pp. 118-122, 2024.
- [66] T. Q. Duong, D. V. Huynh, Y. Li, E. G. Palacios, K. Sun, “Digital Twin-Enabled 6G Aerial Edge Computing with Ultra-reliable and Low-latency Communications : (invited paper),” *International conference on 6G Networks (6GNet)*, pp. 1-5, 2022.
- [67] V. H., Dang, R. K. Saeed, M. Antonino, A. D. Octavia, and Q. D. Trung, “Edge Intelligence-based Ultra-reliable and Low-latency Communications for Digital Twin-enabled Metaverse,” *IEEE Wireless Communication Letters*, vol. 11, no. 8, pp. 1733–1737, 2022.
- [68] Y. Lu, X. Huang, K. Zhang, S. Maharjan, and Y. Zhang, “Communication-Efficient Federated Learning for Digital Twin Edge Networks in Industrial IoT,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5709-5718, 2020.
- [69] V. H. Dang, V. Nguyen, V. Sharma, O. A. Dobre, and T. Q. Duong. “Digital Twin Empowered Ultra-reliable and Low-latency Communications-based Edge Networks in Industrial IoT Environment,” *IEEE International Conference on Communications*, pp. 5651-5656, 2022.
- [70] W. Sun, P. Wang, N. Xu, G. Wang, and Y. Zhang, “Dynamic Digital Twin and Distributed Incentives for Resource Allocation in Aerial-assisted Internet of Vehicles,” *IEEE Internet of Things Journal*, vol. 9, no. 8, pp. 5839-5852, 2021.
- [71] T. Liu, L. Tang, W. Wang, Q. Chen, and X. Zeng, “Digital-Twin-Assisted Task Offloading Based on Edge Collaboration in the Digital Twin Edge Network,” *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1427–1444, 2022.
- [72] W. Sun, H. Zhang, R. Wang, and Y. Zhang, “Reducing Offloading Latency for Digital Twin Edge Networks in 6G,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 12240–12251, 2020.

-
- [73] B. Li, Y. Liu, L. Tan, H. Pan, and Y. Zhang, “Digital Twin Assisted Task Offloading for Aerial Edge Computing and Networks,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 10, pp. 10863–10877, 2022.
- [74] G. Lee, W. Saad, M. Bennis, C. Kim, and M. Jung, “An Online Framework for Ephemeral Edge Computing in the Internet of Things,” *IEEE transactions on wireless communications*, vol. 22, no. 3, pp. 1992–2007, 2023.
- [75] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas. “Communication-Efficient Learning of Deep Networks from Decentralized Data,” *In Artificial Intelligence and Statistics*, pp. 1273-1282, 2017.
- [76] Y. Hu, D. Niu, Y. Jianming, and Z. Shengping, “FDML: A Collaborative Machine Learning Framework for Distributed Features,” *In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2232-2240, 2019.
- [77] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, “Distributed Federated Learning for Ultra-Reliable Low-Latency Vehicular Communications,” *IEEE Transactions on Communications*, vol. 68, no. 2, pp. 1146-1159, 2019.
- [78] [M. K. Abdel-Aziz, S. Samarakoon, M. Bennis, and W. Saad, “Ultra-Reliable and Low-Latency Vehicular Communication: An Active Learning Approach,” *IEEE Communications Letters*, vol. 24, no. 2, pp. 367–370, 2020.
- [79] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, “A Joint Learning and Communications Framework for Federated Learning over Wireless Networks,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 269-283, 2020.
- [80] Y. Zhang, W. Cheng, and W. Zhang, “Multiple Access Integrated Adaptive Finite Blocklength for Ultra-low Delay in 6G Wireless Networks,” *IEEE Transactions on Wireless Communication*, vol. 23, no. 3, pp. 1670-1683, 2023.

- [81] B. Zhu, E. Bedeer, H. H. Nguyen, R. Barton, and J. Henry, “Joint Cluster Head Selection and Trajectory Planning in UAV-Aided IoT Networks by Reinforcement Learning With Sequential Model,” *IEEE Internet of Things Journal*, vol. 9, no. 14, pp. 12071–12084, 2022
- [82] H. Ren, C. Pan, Y. Deng, M. ElKashlan, and A. Nallanathan, “Joint Pilot and Payload Power Allocation for Massive-MIMO-enabled URLLC IIoT Networks,” *IEEE Journal on Selected Areas in Communication*, vol. 38, no. 5, pp. 816–830, 2020.
- [83] Y. Wang, and Z. Jun, ‘A Survey of Mobile Edge Computing for the Metaverse: Architectures, Applications, and Challenges,’ *International Conference on Collaboration and Internet Computing (CIC)*, pp. 1-9, 2022.
- [84] J. Zhou, D. Tian, Z. Sheng, X. Duan, and X. Shen, “Distributed Task Offloading Optimization with Queueing Dynamics in Multiagent Mobile Edge Computing Networks,” *IEEE Internet Things Journal*, vol. 8, no. 15, pp. 12311–12328, 2021.
- [85] R. Lin et al., “Distributed Optimization for Computation Offloading in Edge Computing,” *IEEE Transactions Wireless Communication*, vol. 19, no. 12, pp. 8179–8194, 2020.
- [86] T. T. Vu, D. N. Nguyen, D. T. Hoang, E. Dutkiewicz, and T. V. Nguyen, “Optimal Energy Efficiency with Delay Constraints for Multi-layer Cooperative Fog Computing Networks,” *IEEE Transactions Communication*, vol. 69, no. 6, pp. 3911–3929, 2021.
- [87] K. U. Latif, Z. Han, W. Saad, E. Hossain, M. Guizani, and C. S. Hong. “Digital Twin of Wireless Systems: Overview, taxonomy, Challenges, and Opportunities,” *IEEE Communication Surveys & Tutorials*, vol. 24, no. 4, pp 2230-2254, 2022.
- [88] T. Zhou, Y. Yue, D. Qin, X. Nie, X. Li, and C. Li, “Joint Device Association, Resource Allocation, and Computation Offloading in Ultra-dense Multi-Device

- and Multi-Task IoT Networks,” *IEEE Internet Things Journal*, vol.9, no. 199, pp 18695-18709, 2022.
- [89] D. T. Nguyen, L. B. Le, and V. Bhargava, “Price-based Resource Allocation for Edge Computing: A Market Equilibrium Approach,” *IEEE Transactions Cloud Computation*, vol. 9, no. 1, pp. 302–317, 2021.
- [90] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. O. Wu, “Energy-optimal Mobile Cloud Computing Under the Stochastic Wireless Channel,” *IEEE Transactions Wireless Communication*, vol. 12, no. 9, pp. 4569–4581, 2013.
- [91] S. Bi and Y. Zhang, “Computation Rate Maximization for Wireless Powered Mobile-Edge Computing with Binary Computation Offloading,” *IEEE Transactions Wireless Communication*, vol. 17, no. 6, pp. 4177–4190, 2018.
- [92] T. Q. Dinh, J. Tang, Q. D. La, and T. Q. S. Quek, “Offloading in MEC: Task Allocation and Computational Frequency Scaling,” *IEEE Transactions Communication*, vol. 65, no. 8, pp. 3571–3584, 2017.
- [93] M. H. Chen, B. Liang, and M. Dong, “Joint Offloading Decision and Resource Allocation for Multi-user Multi-task Mobile Cloud,” *IEEE International Conference on Communication (ICC)*, pp. 1–6, 2016.
- [94] S. Lou, C. S. Chen, and M. Coupechoux. “Optimal Joint Subcarrier and Power Allocation in NOMA is Strongly NP-Hard,” *In 2018 IEEE International Conference on Communications (ICC)*, pp. 1-7, 2018.
- [95] S. Zhang, H. Zhang, B. Di, and L. Song, “Cellular UAV-to-X communications: Design and Optimization for Multi-UAV Networks,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 1346–1359, 2019.
- [96] K. Aryan, J. Thompson, E. Vlachos, C. Tsinos, and S. Chatzinotas, “Dynamic RF Chain Selection for Energy Efficient and Low Complexity Hybrid

- Beamforming in Millimeter Wave MIMO Systems,” *IEEE Transactions on Green Communication and Networking*, vol. 3, no. 4, pp. 886-900, 2019.
- [97] W. Ghanem, V. Jamali, Y. Sun, and R. Schober, “Resource Allocation for Multi-User Downlink MISO OFDMA-URLLC Systems,” *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 7184-7200, 2020.
- [98] J. Zhang, Y. Jiang, and X. Li, “Energy-Efficient Resource Allocation in Multiuser Relay-based OFDMA Networks,” *Concurrency and Computation: Practice and Experience*, vol. 25, no. 9, pp. 1113-1125, 2012.
- [99] S. Boyd, and L. Vandenberghe, “Convex Optimization,” *Cambridge university press*, 2004.
- [100] Y. Chen, L. Xun, and S. Zhang, “The Energy Efficiency of Heterogeneous Cellular Networks based on the Poisson Hole Process,” *Future Internet*, vol. 15, no. 2, pp. 56, 2023.
- [101] A. Khalili, E. M. Monfared, S. Zargari, M. R. Javan, M. N. Yamchi, E. A. Jorswieck, “Resource Management for Transmit Power Minimization in UAV-assisted RIS HETNETS supported by dual connectivity,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 3, pp. 1806–1822, 2021.
- [102] J. Chakareski, S. Naqvi, N. Mastrorarde, J. Xu, F. Afghah, A. Razi, “An Energy Efficient Framework for UAV-Assisted Millimeter Wave 5G Heterogeneous Cellular Networks,” *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 1, pp. 37–44, 2019.
- [103] K. Minsu, and J. Lee. “Outage probability of UAV Communications,” *IEEE GLOBECOM*, pp. 1-6, 2018.
- [104] X. Chi, H. Yu, P. Zeng, and Y. Li, “Towards Critical Industrial Wireless Control: Prototype Implementation and Experimental Evaluation on URLLC,” *IEEE Communication Magazine*, vol. 61, no. 9, pp. 193-199, 2023.

-
- [105] S. Sundaresan, S. Maruthu, L. S. Kumar, and D. N. K. Jayakody, "Outage Analysis of Sparse Vector Coding based Downlink Multicarrier NOMA for URLLC," *IEEE Internet of Things Journal*, vol. 10, no. 14, pp. 12393-12400, 2023.
- [106] X. Yuncong, P. Ren, and D. Xu, "Security-Oriented Pilot and Data Transmission for URLLC in Mission-Critical IoT Scenarios," *IEEE Internet of Things Journal*, vol. 10, no. 15, pp. 13697-13715, 2023.
- [107] T. Liu, L. Tang, W. Wang, Q. Chen, X. Zeng, "Digital Twin Assisted Task Offloading based on Edge Collaboration in the Digital Twin Edge Network, " *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1427-1444, 2021.
- [108] M. A. Jamshed, F. Heliot, and T. W. C. Brown, "Unsupervised Learning Based Emission-Aware Uplink Resource Allocation Scheme for Non-Orthogonal Multiple Access Systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 8, pp. 7681-7691, 2021.
- [109] Y. Chen, W. Cheng, and W. Zhang, "Reconfigurable Intelligent Surface Equipped UAV in Emergency Wireless Communication: A New Fading-Shadowing Model and Performance Analysis," *IEEE Transactions on Communication*, vol. 72, no. 3, pp. 1821-1834, 2024.
- [110] Z. Yao, W. Cheng, W. Zhang, T. Zhang, and H. Zhang, "The Rise of UAV Fleet Technologies for Emergency Wireless Communication in Harsh Environments," *IEEE Network*, vol. 36, no. 4, pp. 28-37, 2022.
- [111] Q. Wu, Y. Zeng, and R. Zhang, "Joint Trajectory and Communication Design for Multi-UAV Enabled Wireless Networks," *IEEE Transactions on Wireless Communication*, vol. 17, no. 3, pp. 2109-2121, 2018.
- [112] Z. Liu, X. Liu, Victor, and T. S. Durrani, "Energy-Efficient Resource Allocation for Dual-NOMA-UAV Assisted Internet of Things," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 3, pp. 3532-3543, 2023.

- [113] M. Katwe, K. K. Singh, P. K. Sharma, C. P. Li, and Z. Ding, “Dynamic User Clustering and Optimal Power Allocation in UAV-Assisted Full-Duplex Hybrid NOMA System,” *IEEE Transactions on Wireless Communication*, vol. 21, no. 4, pp. 2573–2590, 2022.
- [114] S. Rezvani, E. A. Jorswieck, R. Joda, and H. Yanikomeroglu, “Optimal Power Allocation in Downlink Multicarrier NOMA Systems: Theory and Fast Algorithms,” *IEEE Journal on Selected Areas in Communication*, vol. 40, no. 4, pp. 1162–1189, 2022.
- [115] P. Luong, F. Gagnon, L. Tran, F. Labeay, “Deep Reinforcement Learning-Based Resource Allocation in Cooperative UAV-Assisted Wireless Networks,” *IEEE Transactions on Wireless Communication*, vol. 20, no. 11, pp. 7610–7625, 2021.
- [116] J. Cui, Z. Ding, P. Fan, and N. Al-Dhahir, “Unsupervised Machine Learning-Based User Clustering in Millimeter-Wave-NOMA Systems,” *IEEE Transactions on Wireless Communication*, vol. 17, no. 11, pp. 7425–7440, 2018.
- [117] R. Tibshirani, G. Walther, T. Hastie, “Estimating the Number of Clusters in a Data Set via the Gap Statistic,” *Journal of the Royal Statistic Society: Series B (Statistic Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.
- [118] C. Dougherty, Introduction to Econometric, *Oxford; New York*, Oxford University Press, 2011.