Unmanned Aerial Vehicles and Edge Computing in Wireless Networks

Bodong Shang

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Electrical Engineering

Lingjia Liu, Chair
Harpreet S. Dhillon
Yaling Yang
Yang (Cindy) Yi
Zhenyu (James) Kong

December 10, 2021
Blacksburg, Virginia

Keywords: Wireless Communications, Unmanned Aerial Vehicle, Edge Computing, Reconfigurable Intelligent Surface.

Unmanned Aerial Vehicles and Edge Computing in Wireless Networks

Bodong Shang

ABSTRACT

Unmanned aerial vehicles (UAVs) attract increasing attention for various wireless network applications by using UAVs' reliable line-of-sight (LoS) paths in air-ground connections and their flexible placement and movement. As such, the wireless network architecture is becoming three-dimensional (3D), incorporating terrestrial and aerial network nodes, which is more dynamic than the traditional terrestrial communications network. Despite the UAVs' advantages of high LoS path probability and flexible mobility, the challenges of UAV communications need to be considered in the design of integrated air-ground networks, such as spectrum sharing, air-ground interference management, energy-efficient and cost-effective UAV-assisted communications. On the other hand, in wireless networks, users request not only reliable communication services but also execute computation-intensive and latency-sensitive tasks. As one of the enabling technologies in wireless networks, edge computing is proposed to offload users' computation tasks to edge servers to reduce users' latency and energy consumption. However, this requires efficient utilization of both communication resources and computation resources. Furthermore, integrating UAVs into edge computing networks brings many benefits, such as enhancing offloading ability and extending offloading coverage region. This dissertation makes a series of fundamental contributions to UAVs and edge computing in wireless networks that include: 1) Reliable UAV communications, 2) Efficient edge computing schemes, and 3) Integration of UAV and edge computing.

In the first contribution, we investigate UAV spectrum access and UAV swarm-enabled aerial

reconfigurable intelligent surface (SARIS) for achieving reliable UAV communications. On the one hand, we study a 3D spectrum sharing between device-to-device (D2D) and UAVs communications. Specifically, UAVs perform spatial spectrum sensing to opportunistically access the licensed channels occupied by the D2D communications of ground users. The results show that UAVs' optimal spatial spectrum sensing radius can be obtained given specific network parameters. On the other hand, we study the beamforming and placement design for SARIS networks in downlink transmissions. We consider that the direct links between the ground base station (BS) and mobile users are blocked due to obstacles in the urban environment. SARIS assists the BS in reflecting the signals to randomly distributed mobile users. The results show that the proposed SARIS network significantly improves the weighted sum-rate for ground users, and the placement design plays an essential role in the overall system performance.

In the second contribution, we develop a joint communication and computation resource allocation scheme for vehicular edge computing (VEC) systems. The full channel state information (CSI) in VEC systems is not always available at roadside units (RSUs). The channel varies fast due to vehicles' mobility, and it is pretty challenging to estimate CSI and feed back the RSUs for processing the VEC algorithms. To address the above problem, we introduce a large-scale CSI-based partial computation offloading scheme for VEC systems. Using deep learning and optimization tools, we minimize the users' energy consumption while guaranteeing their offloading latency and outage constraints. The results demonstrate that the introduced resource allocation scheme can significantly reduce the total energy consumption of users compared with other computation offloading schemes.

In the third contribution, we present novel frameworks for integrating UAVs to edge computing networks to achieve improved computing performance. We study mobile edge computing (MEC) in air-ground integrated wireless networks, including ground computational access

points (GCAPs), UAVs, and user equipment (UE), where UAVs and GCAPs cooperatively provide computation resources for UEs. The resource allocation algorithm is developed based on the block coordinate descent method by optimizing the subproblems of users' association, power control, bandwidth allocation, computation capacity allocation, and UAV placement. The results show the advantages of the introduced iterative algorithm regarding the reduced total energy consumption of UEs.

Finally, we highlight directions for future works to advance the research presented in this dissertation and discuss its broader impact across the wireless networks industry and standard-making.

Unmanned Aerial Vehicles and Edge Computing in Wireless Networks

Bodong Shang

GENERAL AUDIENCE ABSTRACT

The fifth-generation (5G) cellular network aims to achieve a high data rate by having greater bandwidth, deploying denser networks, and multiplying the antenna links' capacity. However, the current wireless cellular networks are fixed on the ground and thus pose many disadvantages. Moreover, the improved system performance comes at the cost of increased capital expenditures and operating expenses in wireless networks due to the enormous energy consumption at base stations (BS) and user equipment (UE). More spectrum and energy-efficient yet cost-effective technologies need to be developed in next-generation wireless networks, i.e., beyond-5G or sixth-generation (6G) networks.

Recently, unmanned aerial vehicle (UAV) has attracted significant attention in wireless communications. Due to UAVs' agility and mobility, UAVs can be quickly deployed to support reliable communications, resorting to its line-of-sight-dominated connections in the air-ground channels. However, the sufficient available spectrum for extensive UAV communications is scarce, and the co-channel interference in air-air and air-ground connections need to be considered in the design of UAV networks. In addition to users' communication requests, users also need to execute intensive computation tasks with specific latency requirements. As such, edge computing has been proposed to integrate wireless communications and computing by offloading users' computation tasks to edge servers in proximity, reducing users' computation energy consumption and latency. Besides, integrating UAVs into edge computing networks makes efficient offloading schemes by leveraging the advantages of UAV communications. This dissertation makes several contributions that enhance UAV communications and edge computing systems performance, respectively, and present novel frameworks for UAV-assisted three-dimensional (3D) edge computing systems.

This dissertation addresses the fundamental challenges in UAV communications, including spectrum sharing, interference management, UAV 3D placement, and beamforming, allowing broadband, wide-scale, cost-effective, and reliable wireless connectivity. Furthermore, this dissertation focuses on the energy-efficient vehicular edge computing systems and mobile edge computing systems, where the UAVs are applied to achieve 3D edge computing systems. To this end, various mathematical frameworks and efficient joint communication and computation resource allocation algorithms are proposed to design, analyze, optimize, and deploy UAV and edge computing systems. The results show that the proposed air-ground integrated networks can deliver spectrum-and-energy-efficient yet cost-effective wireless services, thus providing ubiquitous wireless connectivity and green computation offloading in the future beyond-5G or 6G wireless networks.

To my parents.

Acknowledgments

Writing this dissertation has come at the end of my substantial and fulfilling Ph.D. journey. I would like to acknowledge some extraordinary people who I cherish during my Ph.D. program.

I owe my gratitude to Dr. Lingjia Liu for his consistent support during my Ph.D. program. I would like to sincerely thank him for his time and efforts that made my Ph.D. journey extremely valuable and joyful. Without his support and advice for me in pursuing research problems of my interest, I could not achieve the research results and finish this dissertation.

I am also thankful to my Ph.D. dissertation committee members Dr. Lingjia Liu, Dr. Harpreet S. Dhillon, Dr. Yaling Yang, Dr. Yang (Cindy) Yi, and Dr. Zhenyu (James) Kong, for their valuable comments and suggestion that help me improve the quality of this dissertation.

I am indebted to the excellent professors at Virginia Tech. I sincerely thank Dr. R. Michael Buehrer for the fantastic course on digital communications. The diverse teaching modes have enabled me to have a deeper understanding of digital communications. I am grateful to Dr. Harpreet S. Dhillon for teaching the course of stochastic signals and systems most interestingly. These courses have been beneficial in building my professional knowledge background and helping my research.

As I am pursuing a Ph.D. degree, the world is facing a global pandemic COVID 19. I would like to thank all the health care workers on the front lines, the medical researchers, and everyone who makes efforts to fight the pandemic and keep the essential services up and running.

Last but not least, it is time to thank all my friends and my parents. Thank you for your unwavering support throughout the long road until this day and the days to come. Without their constant support, I could have never been able to finish this dissertation.

Contents

List of Figures

List of Tables

List of Abbreviations

5G	Fifth-Generation
6G	Sixth-Generation
A2A	Air-to-Air
A2G	Air-to-Ground
ARIS	Aerial Reconfigurable Intelligent Surface
ASE	Area Spectral Efficiency
BS	Base Station
CPU	Central Processing Unit
CSI	Channel State Information
D2D	Device-to-Device
G2A	Ground-to-Air
GCAP	Ground Computational Access Point
IoT	Internet-of-Things
LoS	Line-of-Sight
MEC	Mobile Edge Computing
MIMO	Multiple-Input Multiple-Output

mmWave	millimeter-wave
NLoS	Non-Line-of-Sight
QoS	Quality-of-Service
RF	Radio Frequency
RIS	Reconfigurable Intelligent Surface
RSU	Roadside Unit
SARIS	UAV Swarm-Enabled ARIS
TRIS	Terrestrial Reconfigurable Intelligent Surface
UAV	Unmanned Aerial Vehicle
UDN	Ultra-Dense Network
UE	User Equipment
V2X	Vehicle-to-Everything
VEC	Vehicular Edge Computing

Chapter 1

Introduction

1.1 Background

The fifth-generation (5G) cellular network aims to achieve high data rate, low latency, massive connectivity by having greater bandwidth, deploying denser networks, and multiplying the antenna links' capacity [1]. The construction of 5G terrestrial network contains heterogeneous infrastructures, e.g., cellular network, mobile ad hoc network, and wireless local area network. In particular, the cellular network has been widely developed in recent decades to support various services. With the help of the 5G cellular network's key technologies, such as ultra-dense network (UDN), massive multiple-input-and-multiple-output (MIMO), and millimeter-wave (mmWave) communication, the terrestrial network is capable of providing high data rate (10 Gbits/s) with relatively low latency (1 ms). However, the terrestrial network has obvious restrictions on limited global coverage. Moreover, the improved system performance comes at the cost of increased capital expenditures and operating expenses due to the enormous energy consumption generated by active hardware components. In future wireless networks, i.e., beyond-5G or the sixth-generation (6G) network, more spectrum and energy-efficient yet cost-effective technologies need to be developed.

In next-generation wireless networks, it is expected for the network to provide global coverage, utilize all spectrum, support full applications, ensure strong security, etc. Future networks have to exploit all possible resources by interconnecting space, air, and ground network segments to provide seamless wireless coverage and support various and stringent service requirements. Unmanned aerial vehicle (UAV) and satellite communication networks are important for fast response in harsh and difficult environments. In addition to the demand for communications, users also increasingly need adequate computation resources to efficiently process their computation tasks to save on-board energy and reduce latency, especially in vehicular networks. Due to the limited computing capability of the central processing unit (CPU) at some local devices, their latency requirements for processing computation tasks may not be guaranteed. Moreover, due to the severe energy consumption of locally computing, merely processing all tasks at one device reduces its battery lifetime. Mobile edge computing (MEC) is proposed to be a promising solution for tackling the above peculiarities, where edge servers offload users' computation tasks through wireless communications. Recently, a new paradigm of reconfigurable intelligent surface (RIS)-assisted wireless networks has drawn extensive attention due to its low-cost characteristic and high spectral and energy efficiency. RIS is a planar surface comprising a large number of low-cost passive reflecting elements. By adjusting the amplitudes and phase shifts of the reflecting elements, RIS can achieve fine-grained reflection-beamforming. Since RIS does not require radio frequency (RF) chains to transmit or receive the signals, it enjoys low energy consumption but achieves significantly improved spectral efficiency.

1.2 Unmanned Aerial Vehicle

In future wireless networks, as UAVs become more available, mobile users will not be restricted to terrestrial mobile stations. There are many applications for UAVs in wireless network such as UAV swarm networks in disasters, UAV-assisted vehicle-to-everything (V2X) communications [2], UAV enabled smart city, traffic offloading in hotspots, and surveillance

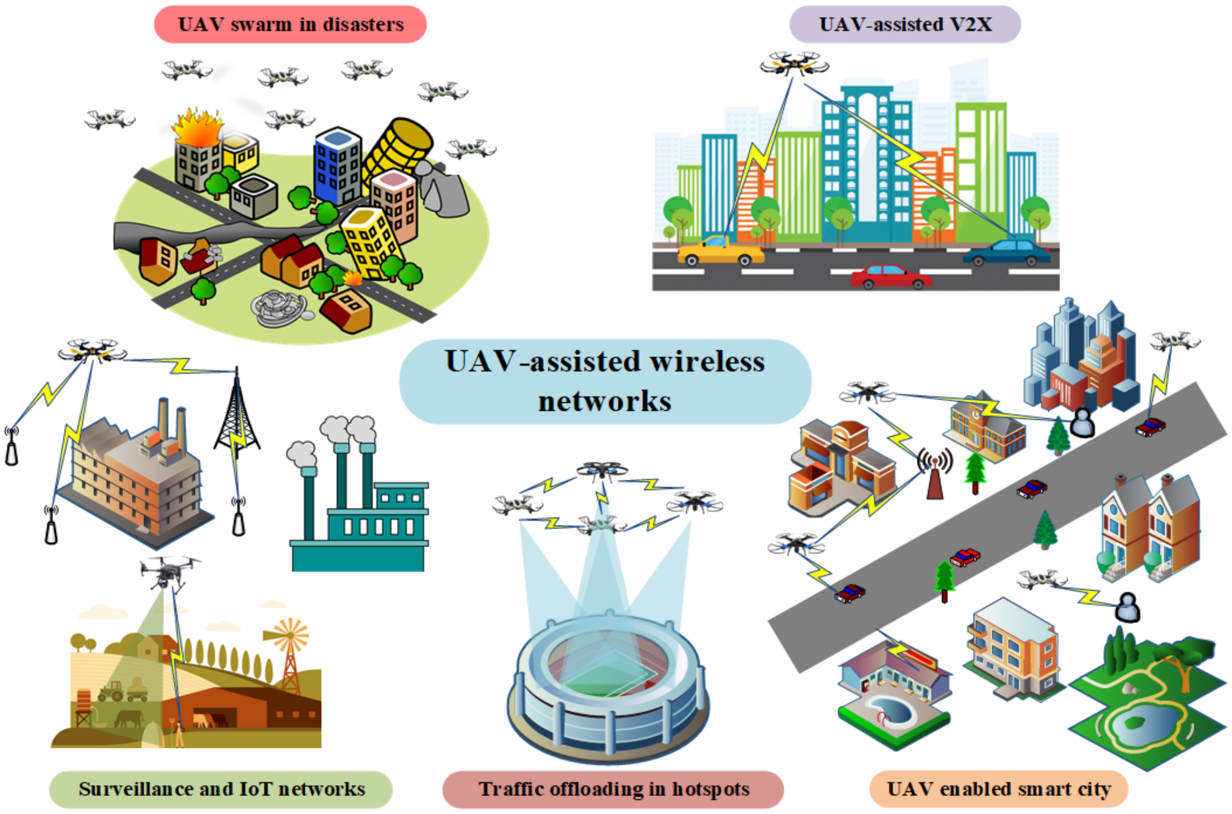


Figure 1.1: Applications of UAV in wireless networks.

and Internet-of-Things (IoT) networks, as shown in Fig. 1.1.

The wireless network architecture will become a three-dimensional (3D) structure, incorporating terrestrial and aerial network nodes that are more dynamic than the fixed terrestrial communications network that we have today. In the development of aerial platforms, the spectrum access for UAV communications is of significance in the design and management of the holistic communications network. Different from the non-line-of-sight (NLoS) transmissions in most ground communications, aerial communications including air-to-air (A2A), air-to-ground (A2G) and ground-to-air (G2A) enjoy reliable wireless transmissions resorting to the lower signal attenuation due to fewer obstacles. In A2A communications, the signal experiences approximately free-space propagation. In A2G and G2A communications, the occurrence probability of line-of-sight (LoS) connection or NLoS connection is a function of the elevation angle between the UAV and ground node and the environment such as rural, urban, dense urban or others.

UAV communications typically happen in unlicensed spectrum, including the 2.4 GHz and 5.8 GHz ISM bands [3]. For small and medium range UAV applications, multi-hop 802.11 or Zigbee technologies are considered according to their throughput and range demands [4]. For high throughput applications, additional wireless technologies and spectrum need to be considered. When operating in unlicensed spectrum, UAVs may suffer from security threats and attacks which impact the transmission of confidential information. Moreover, with the drastic increase in the number of wireless devices (such as tablets, smartphones, and sensors) which also operate in unlicensed spectrum, the unlicensed spectrum is becoming overcrowded and UAVs will be facing spectrum scarcity in near future.

The use of licensed spectrum, on the other hand, would enable wide-scale and high-quality connectivity for UAVs with enough capacity to support various services and increasing usage levels. Sharing the licensed spectrum that is used for cellular communications, for instance, with UAVs can significantly improve the communication performance of UAVs. However, the interference generated by the UAVs needs to be well managed to limit its effect on primary users. Therefore, network designers need to take the negative impact of implementing UAV communications into account and come up with efficient spectrum sharing strategies for the coexistence of UAVs and terrestrial communications devices. Enabling spectrum sensing for UAVs will allow UAVs to opportunistically exploit licensed spectrum holes and improve the area spectral efficiency (ASE) of the overall wireless ecosystem.

1.3 Edge Computing

In the IoT and the 5G wireless networks, enormous computation tasks are generated by the users [5]. These computation tasks are usually computation-intensive, which require a large amount of computation capacity, and also latency-sensitive with stringent latency requirements. However, users' computation capacity is limited, and thus the latency requirements would not be guaranteed when users locally execute tasks. In addition, entirely computing the tasks at users consumes much energy. Due to the limited battery capacity of device, locally processing the computation tasks will reduce the device's lifetime. Therefore, it is imperative to explore innovative, energy-efficient, yet Quality-of-Service (QoS) guaranteed solutions for computing in future IoT and beyond-5G wireless networks.

MEC has emerged as an effective solution to help users deal with computation-intensive and latency-sensitive tasks [6]. MEC enables users to offload their computational tasks to edge servers, thus reducing users' computation burden. As such, MEC helps users to save their energy and ensure their latency requirements. Unlike traditional cloud computing, offloading users' computational tasks to edge servers reduces data transmission delay. By carefully designing MEC algorithms that jointly optimize communication and computation resources allocation, one can improve network performance.

However, there are some drawbacks to terrestrial MEC networks. First, the edge servers and base stations (BSs) are usually deployed at fixed locations or far away from users located in remote areas, which poses disadvantages for dynamically and randomly distributed users. Second, the data transmissions in terrestrial MEC networks usually experience significant signal attenuation due to the large-scale path loss and shadow fading resulting from obstacles like buildings and trees. Thus, the uplink data rate for computation offloading can be unsatisfactory, allowing fewer data to be offloaded while guaranteeing the latency require-

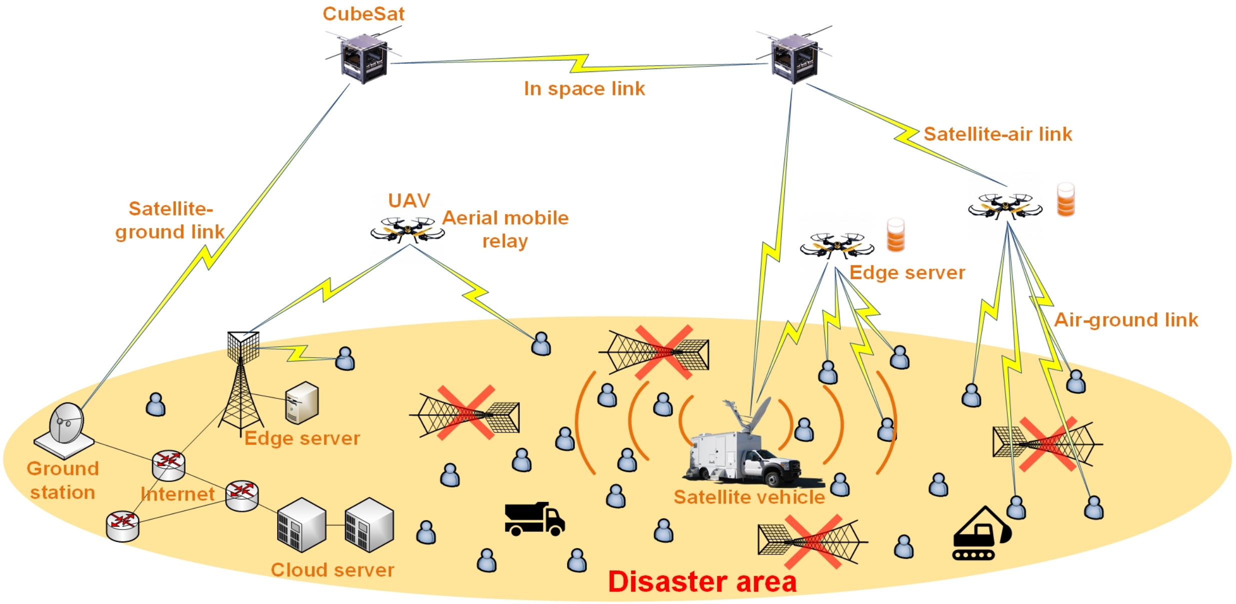


Figure 1.2: UAVs in MEC networks where some BSs are disabled.

ment. Third, terrestrial MEC networks can be disabled during disasters or in remote areas. Moreover, the edge servers usually incur the site-rent cost.

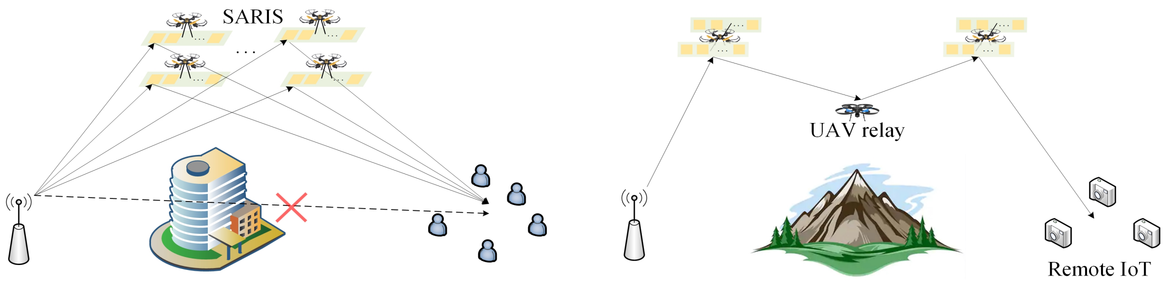
Based on the drawbacks of terrestrial MEC networks, UAV-enabled MEC has been proposed, where the edge server is mounted on UAV to provide computation offloading for users [7, 8]. Compared with the traditional terrestrial MEC networks, the UAV-enabled MEC network supports reliable LoS paths to users and can be deployed with flexible mobility [9]. Moreover, in disasters, the terrestrial networks are destroyed due to damage. The terrestrial MEC network may be disabled. In this case, we can deploy UAVs to assist terrestrial computing and/or enable aerial computing for ground users, as shown in Fig. 1.2.

However, there are still some critical issues of UAV-enabled MEC. First, mounting the edge server on a UAV increases the UAV's load and propulsion-related energy consumption. Since the UAV's battery capacity is scarce but limited, this reduces the UAV's service time. Second, offloading users' computational tasks to the UAV increases the UAV's computation-related energy consumption. In general, computation-related energy consumption is greater than communication-related energy consumption, emphasizing the importance of considering an energy-efficient UAV-enabled MEC design when aiming to prolong the UAV's service time. Third, deploying an additional edge server on a UAV increases the UAV's production cost. Therefore, research on finding new cost-effective yet energy-efficient technologies for MEC is still of considerable interest.

1.4 Aerial Reconfigurable Intelligent Surface

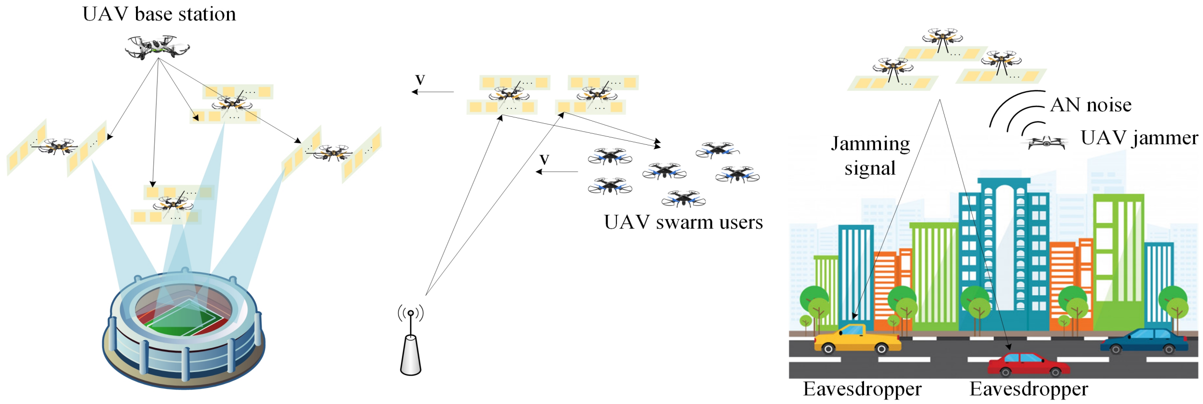
RIS has been introduced as a new technology to improve wireless networks' spectrum and energy efficiency [10, 11]. An RIS is a planar surface comprising a large number of low-cost passive reflecting elements. By adjusting the amplitudes and phase shifts of the reflecting elements, an RIS can achieve fine-grained reflection-beamforming. Moreover, with the full-duplex mode in operation and no noise-addition characteristics, an RIS is more spectrum-efficient than the conventional relay technology. In addition, an RIS does not require RF chains to transmit or receive signals, and thus it enjoys low energy consumption but achieves significantly improved spectral efficiency. Therefore, RIS can be deployed in wireless networks to achieve significantly improved overall system performance.

Instead of being limited to only terrestrial deployment, wireless networks are gradually evolving to air-ground integrated networks to achieve ubiquitous wireless connectivity and upgraded network capacity. Recently, UAVs have attracted significant attention in wireless communication. Due to the agility and mobility, UAVs can be quickly deployed in hotspots or disaster regions to support reliable communication, resorting to its LoS-dominated connections in the air-to-ground channels [2]. As an aid to this, an RIS can be mounted on UAVs to enable aerial RIS (ARIS) to achieve 3D signal reflection. Such an ARIS is not re-



Cooperative panoramic full-angle reflection

Multiple reflections for remote IoT



Air-to-air reflection for hotspots/disasters

Enhanced signals for UAV swarm

Trackable physical layer security

Figure 1.3: Applications of UAV swarm-enabled aerial RIS in wireless network.

stricted to the 180° half-space reflection, but instead, it provides a 360° panoramic full-angle reflection [12]. With the UAV's flight ability in 3D space, an ARIS is more flexible in deployment than the conventional terrestrial RIS (TRIS), which is usually deployed on facades of a building or at a dedicated site. Attaining an appropriate place for the TRIS installation would not be easy in practice due to excessive site-rent and urban landscape impact. Moreover, the ARIS's cascaded reflection channel is more desirable than the TRIS's, which shows the potential to improve system performance further. From the above discussion, it is practically appealing to investigate the combination between UAVs and RISs. In Fig. 1.3, we illustrate the applications of UAV swarm-enabled aerial RIS in wireless networks.

One of the challenges in designing ARIS-assisted networks is the trade-off between NLoS

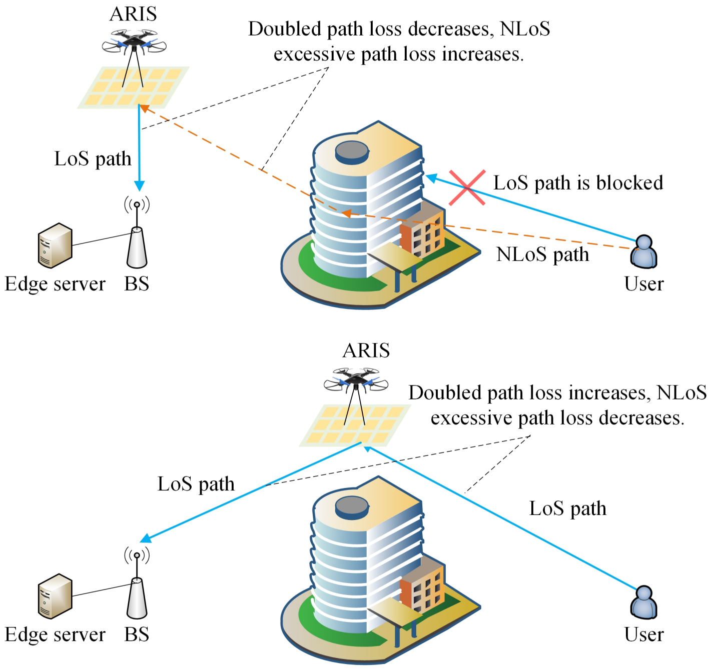


Figure 1.4: An illustration of the trade-off between NLoS excessive path loss and doubled path loss.

excessive path loss and the doubled path loss due to signal reflection. In the ARIS-assisted network, two links are cascaded together, i.e., one is between the user and ARIS, the other is between the ARIS and BS, shown in Fig. 1.4. In the cascaded channel, the doubled path loss severely deteriorates the network performance [13]. In terrestrial RIS-assisted wireless networks, it is better to deploy an RIS near the user or BS to minimize the doubled path loss in the cascaded channel. However, this deployment strategy may not be directly applied to an ARIS in air-ground channels. This is because an ARIS's 3D position determines the large-scale path loss of the cascaded channel, and it also changes the LoS and NLoS probabilities in air-ground channels.

Specifically, the probability of the NLoS path decreases with the ARIS's elevation angle. As such, deploying the ARIS near the BS reduces the elevation angle between the user and the ARIS. Thus, this increases the probability of NLoS path between user and ARIS, as shown in Fig. 1.4. It is worth noting that the NLoS excessive path loss deteriorates the user's signal strength. On the other hand, deploying the ARIS near the user decreases the elevation angle between the ARIS and BS. This increases the probability of NLoS path between the ARIS and the BS. Therefore, there is a trade-off between the NLoS excessive path loss and the doubled path loss. Nevertheless, the above trade-off analysis is under the single-user case. If there are multiple users in ARIS-assisted networks, the complexity of the analysis will increase. Despite the trade-off between NLoS excessive path loss and doubled path loss, the trade-off between LoS probability and NLoS probability also needs to be considered in the network design, where the sum of these two probabilities equals one. It is worth noting that in the LoS path, the aperture gain of ARIS can be achieved based on the array response. As a result, the 3D deployment of ARIS is significant to the entire network performance.

1.5 Contributions

This dissertation's main contribution is to develop analytical foundations for design, performance analysis, and optimization of UAV and edge computing wireless networks. In particular, we introduce various and novel network frameworks for efficient spectrum sharing, UAV placement and movement, performance analysis, and resource allocation of UAV communications and edge computing systems. Based on the proposed frameworks, we improve the network performance in terms of spectral efficiency, system capacity, latency, and energy consumption. This dissertation weaves together advanced mathematical tools such as stochastic geometry, optimization theory, machine learning, deep learning, and reinforcement learning. As such, this dissertation develops in-depth performance analysis and efficient algorithms to design, analyze, optimize, deploy, and operate UAV and edge computing wireless networks. In summary, so far, our contributions are shown as follows:

- First, in Chapter 2, we analyze a 3D spectrum sharing between device-to-device (D2D) and UAVs communications. We consider that UAVs perform spatial spectrum sensing to opportunistically access the licensed channels that are occupied by the D2D communications of ground users. The objective of the considered 3D spectrum sharing networks is to maximize the ASE of UAV networks while guaranteeing the required minimum ASE of D2D networks. Using the tools from machine learning, we obtain the probability of spatial false alarm and the probability of spatial missed detection at the UAV, which helps us to characterize the density of active UAVs. Then, based on the Neyman-Pearson criterion, we further derive the coverage probability of D2D and UAV communications by leveraging the tools from stochastic geometry. In addition, the ASE of the D2D and UAV networks are also obtained. Simulation results show that the optimal spatial spectrum sensing radius of UAVs can be obtained given specific network parameters.

- Second, in Chapter 3, we study MEC in air-ground integrated wireless networks including ground computational access points (GCAPs), UAVs, and user equipment (UEs), where UAVs and GCAPs cooperatively provide computing resources for UEs. Our goal is to minimize the total energy consumption of UEs by jointly optimizing users' association, uplink power control, channel allocation, computation capacity allocation, and UAV placement, subject to the constraints on deterministic binary offloading, UEs' latency requirements, computation capacity, UAV power consumption, and available bandwidth. Due to the non-convexity of the primary problem and the coupling of variables, we introduce a coordinate descent algorithm that decomposes the UEs' energy consumption minimization problem into several subproblems which can be efficiently solved. Simulation results demonstrate the advantages of the proposed algorithm in terms of the reduced total energy consumption of UEs.

- Third, in Chapter 4, we propose an energy-efficient computation offloading scheme for vehicular edge computing (VEC) systems, where multiple roadside units assist vehicular users to offload computation tasks to edge servers. Our goal is to minimize the users' energy consumption by optimizing users' association, data partition, transmit power, and computation resources, subject to the constraints of partial tasks offloading, users' latency, maximum transmit power, outage performance, and computation capacity of edge servers. We utilize deep learning for obtaining users' association and develop an efficient optimization algorithm to optimize other variables. Moreover, the complexity and convergence of the algorithm are analyzed. Simulation results demonstrate that the introduced resource allocation algorithm can significantly reduce the total energy consumption of users.

- Finally, in Chapter 5, we study a UAV swarm-enabled aerial RIS (SARIS)-assisted downlink communication system, where RISs can be mounted on UAVs to enable 3D signal reflections, reliable air-ground connections, and higher configuration flexibility. The objective of the considered SARIS system is to maximize the weighted sum-rate of ground users by designing the transmit beamforming at the base station (BS), the phase shifts of SARIS reflecting elements, and SARIS 3D placement. For joint BS and SARIS beamforming design, we introduce two beamforming schemes with low computational complexity. For SARIS placement design, the optimal SARIS 3D position is obtained by leveraging the tools from stochastic geometry and considering the distributions of ground users. Simulation results confirm the validity of the analytical derivations. In particular, the SARIS placement plays a vital role in the system performance when the distances between users and the BS increase.

1.6 List of Publications

As a byproduct of the above contributions, this dissertation has led to the following first-author publications:

1.6.1 Journal Publications

[J12] **B. Shang**, E. Bentley, and L. Liu, ”UAV Swarm-Enabled Aerial Reconfigurable Intelligent Surface: Modeling, Analysis, and Optimization”, submitted to IEEE for possible publication.

[J11] **B. Shang**, R. Shafin, and L. Liu, ”UAV Swarm Enabled Aerial Reconfigurable Intelligent Surface (SARIS)”, in *IEEE Wireless Communications*, 2021, doi: 10.1109/MWC.010.2000526.

[J10] **B. Shang**, H. V. Poor, and L. Liu, “Aerial Reconfigurable Intelligent Surfaces Meet Mobile Edge Computing”, submitted to IEEE for possible publication.

[J9] **B. Shang**, L. Liu, H. Song, B. Xu, S. Pudlewski, and E. Bentley, ”Trade-offs in Reliability, Latency, and Energy for Random Network Coding-Enabled Networks”, in *IEEE Communications Letters*, vol. 25, no. 8, pp. 2768-2772, Aug. 2021, doi: 10.1109/LCOMM.2021.3081686.

[J8] **B. Shang**, L. Liu, and Z. Tian, ”Deep Learning-Assisted Energy-Efficient Task Offloading in Vehicular Edge Computing Systems”, in *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9619-9624, Sept. 2021, doi: 10.1109/TVT.2021.3090179.

[J7] **B. Shang**, Y. Yi, and L. Liu, ”Computing over Space-Air-Ground Integrated Networks: Challenges and Opportunities”, in *IEEE Network*, vol. 35, no. 4, pp. 302-309, July/August 2021, doi: 10.1109/MNET.011.2000567.

[J6] **B. Shang**, L. Liu, H. Chen, J. C. Zhang, S. Pudlewski, E. S. Bentley and J. D. Ashdown, “Spatial Spectrum Sensing in Uplink Two-Tier User-Centric Deployed HetNets”, in *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 7957-7972, Dec. 2020, doi: 10.1109/TWC.2020.3018408.

[J5] **B. Shang**, L, Liu, R. M. Rao, V. Marojevic and J. H. Reed, ”3D Spectrum Sharing for Hybrid D2D and UAV Networks”, in *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5375-5389, Sept. 2020, doi: 10.1109/TCOMM.2020.2997957.

[J4] **B. Shang**, L. Liu, ”Mobile Edge Computing in the Sky: Energy Optimization for Air-Ground Integrated Networks”, in *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7443-7456, Aug. 2020, doi: 10.1109/JIOT.2020.2987600.

[J3] **B. Shang**, V. Marojevic, Y. Yi, A. Abdalla, and L. Liu, ”Spectrum Sharing for UAV Communications: Spatial Spectrum Sensing and Open Issues”, in *IEEE Vehicular Technology Magazine*, vol. 15, no. 2, pp. 104-112, June 2020, doi: 10.1109/MVT.2020.2980020.

[J2] **B. Shang**, L. Liu, ”Machine Learning Meets Point Process: Spatial Spectrum Sensing in User-Centric Networks,” in *IEEE Wireless Communications Letters*, vol. 9, no. 1, pp. 34-37, Jan 2020, doi: 10.1109/LWC.2019.2940442.

[J1] **B. Shang**, L. Liu, J. Ma and P. Fan, ”Unmanned Aerial Vehicle Meets Vehicle-to-Everything in Secure Communications”, in *IEEE Communications Magazine*, vol. 57, no. 10, pp. 98-103, Oct 2019, doi: 10.1109/MCOM.001.1900170.

1.6.2 Conference Publications

[C2] **B. Shang**, S. Liu, S. Lu, Y. Yi, W. Shi, and L. Liu, ”A Cross-Layer Optimization Framework for Distributed Computing in IoT Networks”, in Workshop on Edge Computing and Communications (EdgeComm) of 2020 *IEEE/ACM Symposium on Edge Computing*, San Jose, CA, USA, 2020, pp. 440-444, doi: 10.1109/SEC50012.2020.00067.

[C1] **B. Shang**, L. Liu, H. Chen, J. C. Zhang, S. Pudlewski, E. S. Bentley and J. D. Ashdown, "Spatial Spectrum Sensing-Based D2D Communications in User-Centric Deployed HetNets", 2019 *IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, 2019, pp. 1-6, doi: 10.1109/GLOBECOM38437.2019.9013184.

Chapter 2

3D Spectrum Sharing for Hybrid D2D and UAV Networks

2.1 Introduction

Unmanned aerial vehicles (UAVs) have attracted great attention as they enable various applications and services [14]. Cellular operators consider UAVs as users or network support nodes in cellular networks and vehicle-to-everything communications [2]. In addition to operator deployed UAVs, UAVs may belong to third-party organizations or individuals who want to enjoy broadband data transmissions for video streaming, content delivery, surveillance report, etc. In the meantime, wireless communication systems for UAV data transmission need to be carefully designed. Due to the congested unlicensed spectrum, it is desirable for UAVs to transmit in licensed or shared spectrum [15]. Licensing spectrum for massive/broadband UAV communications is not feasible; rather, we consider spectrum sharing between UAVs and ground licensed users to be a viable option.

In the licensed band, device-to-device (D2D) communications enable mobile users that are close to each other to communicate directly [16, 17]. The D2D operation in the licensed band includes the underlay mode (using the same spectrum as cellular communications links) and the overlay mode (using orthogonal spectrum to cellular communications) [18, 19]. D2D communications in the unlicensed band, on the other hand, need to coexist with WiFi networks and typically suffer severe interference [20]. Since D2D transceivers are in proximity of one another, the received interference power at D2D receivers (D2D-Rxs) from UAVs is lower, compared to that of the long range regime which occurs when the receiver lies outside the protection zone of D2D transmitters (D2D-Txs). Therefore, we consider the three dimensional (3D) spatial spectrum sharing between UAV and D2D communications where UAVs play the role of secondary users and D2D as primary users. To fully utilize the licensed spectrum in such hybrid D2D and UAV networks, we aim to maximize the area spectral efficiency (ASE) of UAV networks while guaranteeing the minimum required ASE for D2D users. From operators' perspective, sharing the licensed spectrum with UAVs can help them increase profit margins by charging fees from UAV users. From the UAVs' perspective, communications can achieve the desired performance in the licensed band, as opposed to the congested and insecure unlicensed band. Considering that mobile users need to pay monthly fees to the operator, D2D transmitters (D2D-Txs) are considered the primary and UAVs as the secondary users in the hybrid D2D-UAV spectrum sharing network. Since it is intractable to numerically calculate the conditional interference distributions (conditioned on the presence of D2D-Rx in the intersection of UAV sensing sphere and ground) at a UAV under the ground-to-air (G2A) channel models, we leverage machine learning tools, i.e., Gaussian kernel nonlinear regression, to approximately obtain the conditional interference distributions, which are used in the derivation of spatial false alarm and missed detection probabilities in UAV spatial spectrum sensing. Leveraging tools from stochastic geometry, we analytically derive the coverage probability of D2D communications and UAV communications. Based on our proposed model, we can maximize the area spectral efficiency (ASE) of UAV networks by optimizing the UAV spatial spectrum sensing radius under the constraint of a minimum ASE of D2D networks.

2.1.1 Related Works

Spectrum sharing in terrestrial networks has been investigated in [21, 22, 23, 24, 25, 26, 27]. In [21], the authors studied spectrum sharing for D2D communications in cellular networks. A paradigm for spectrum sharing between cellular communications and radio astronomy systems was introduced in [22]. In [23], the approach of guard zones (protection regions) around cellular BSs was introduced. The spatial spectrum sensing-based D2D communications have been studied in [24, 26] and have been extended to the D2D spectrum access in user-centric deployed heterogeneous networks [27].

On the other hand, 3D spectrum sharing for UAV networks has largely been unexplored. In [28], the authors derived the optimal density of spectrum sharing drone networks to maximize the throughput of the small cell UAV network. However, the considered channel model is rather simplistic to facilitate closed-form derivations. In [29], a spectrum sharing planning problem for a full-duplex UAV and underlaid D2D communications was studied, where a mobile UAV assists the communications between separated nodes without a direct link. In [30], the performance of a static UAV and a mobile UAV coexisting with D2D users in a finite area was studied, where the UAVs and D2D communications have the same spectrum access priority. [31, 32] also studied the coexistence of D2D and UAV communications. However, these works considered only one UAV in the sky in the absence of mutual interference between multiple UAVs with various flight heights. To the best of our knowledge, this is the first work that studies overlay spectrum sharing between UAVs and D2D communications from a system-level perspective.

2.1.2 Contributions

The main contributions of this chapter are summarized as follows:

- *3D UAV Spatial Spectrum Sensing Model:* A 3D UAV spatial spectrum sensing model for coexisting D2D and UAV networks is designed. The spatial spectrum sensing is conditioned on the sensing of D2D-Txs lying in the UAV's spatial spectrum sensing sphere. The conditional distributions of the received signal strength at the UAV have been approximated by log-normal distributions. Given network paramters, a machine learning-assisted approach is introduced to obtain the approximated distribution parameters (mean and standard deviation) to interpret the G2A channels. The spatial false alarm probability and the spatial missed detection probability of a typical UAV can be obtained using the approximated conditional distributions.

- *Coverage Probability and ASE Analysis:* We model and analyze the 3D hybrid spectrum sharing network from a system-level perspective. Our model is flexible enough to capture any density distributions of UAVs in 3D space, rather than fixing the heights of UAVs. The coverage probability of D2D and UAV communications are derived by considering co-tier and cross-tier interference. Based on these analyses, we obtain the ASE of the D2D and UAV communications, respectively.

- *Network Design Insights:* The analysis and simulation results provide important network design insights: The optimal spatial spectrum sensing radius of UAVs is obtained to maximize the ASE of UAV networks under the constraint of a minimum ASE of D2D networks. It is observed that a decrease in the spatial spectrum sensing radius of UAVs has a contrasting effect on the coverage probability of UAV communications and the ASE of UAV networks. The optimal transmit power of UAVs can be also obtained which maximizes the ASE of UAV networks under the constraint of a minimum ASE of D2D networks.

The chapter is organized as follows. Chapter 2.2 presents the system model. Chapter

2.3 gives the probabilities of spatial detection of a typical UAV. Chapter 2.4 shows the coverage probabilities of D2D and UAV communications, and presents the ASE of UAV and D2D networks, respectively. Simulation and numerical results are discussed in Chapter 2.5. Chapter 2.6 concludes the chapter.

2.2 System Model

2.2.1 Network Layout

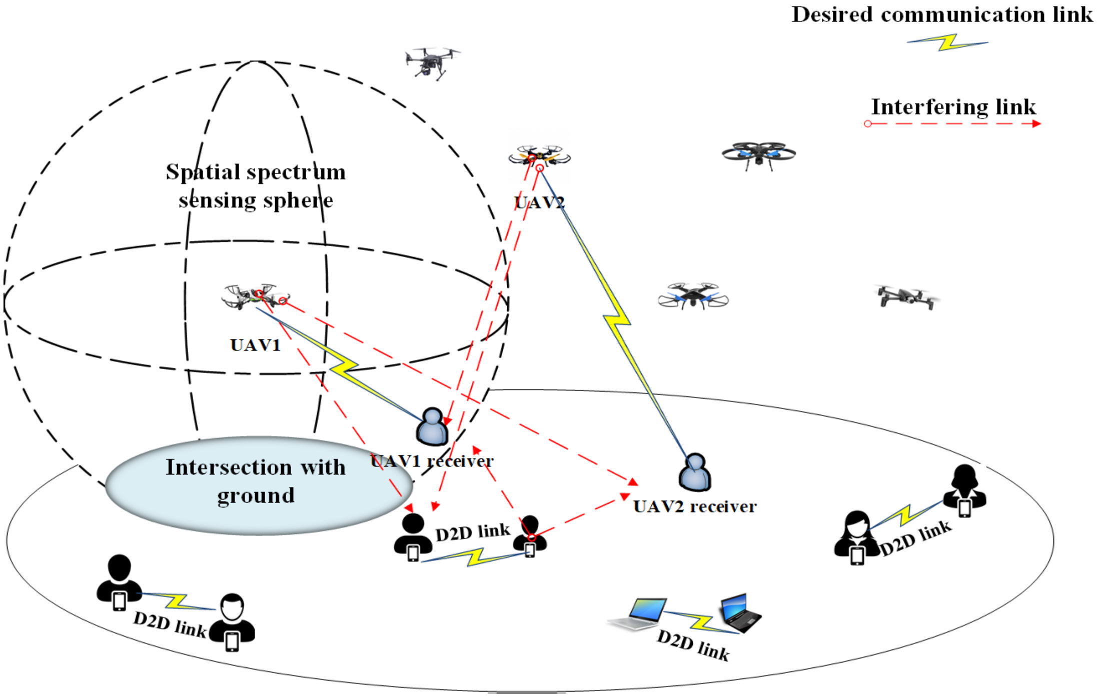


Figure 2.1: Spectrum sharing for D2D and UAV coexisting networks.

The network architecture is shown in Fig. 2.1. The locations of D2D transmitters (D2D-

Txs) are modeled as a homogeneous Poisson point process (PPP) on the two dimensional (2D) ground with the density of λ_D and the set of D2D-Txs are denoted as Φ_D. The signal-to-interference-plus-noise ratio (SINR) threshold of a D2D receiver (D2D-Rx) is γ_D^{th}. The transmit power of a D2D-Tx is P_D. Without loss of generality, we assume that UAVs are distributed in 3D space within the height range of $[H_{\min}, H_{\max}]$, where the densities of UAVs on different horizontal planes follow a certain distribution $f_H(h)$. Due to their flexible mobility, UAVs can fly at different heights and change their heights dynamically. Our model realistically captures the various heights of UAVs, unlike the state-of-the-art that assumes UAVs at a fixed height. We denote the 3D allowable flight space for UAVs as $\mathbb{V}_A = \{(x, y, z)|\, H_{\min} \leqslant z \leqslant H_{\max}\}$. The density of UAVs in $\mathbb{V}_A$ is λ_V which is the density of the ground projection points of UAVs. The set of all UAVs is represented by Φ_V. In this chapter, we consider the case where UAVs intend to transmit data to ground users in a same channel that is used by the overlaid D2D communications network. The transmit power of a UAV is denoted by P_V and the SINR threshold of the associated ground receiver is γ_V^{th}. The transmit power of UAV can be adapted according to the channel conditions and the user's quality of service requirement to further enhance the communication performance. Our model is applicable for many power control mechanisms such as the semi-static power control mechanism [33]. For ease of analysis, in this chapter we assume fixed transmit power at UAVs.

2.2.2 Radio Propagation Model

The SINR of a typical D2D-Rx u_k^d associated with the corresponding D2D-Tx d_k is given by

$$SINR\left(u_k^d\right) = \frac{P_D h_{d_k u_k^d} \left\| d_k - u_k^d \right\|^{-\alpha_{GG}}}{I_{u_k^d}^V + I_{u_k^d}^D + {\sigma_n}^2}$$

where

$$\begin{aligned} I_{u_k^d}^V &= \sum_{v_j \in \Phi_V^{act}} \frac{P_{LOS}\left(v_j, u_k^d\right) P_V g_{v_j u_k^d}}{\left\| v_j - u_k^d \right\|^{\alpha_{AG}}} + \sum_{v_j \in \Phi_V^{act}} \frac{P_{NLOS}\left(v_j, u_k^d\right) \eta P_V g_{v_j u_k^d}}{\left\| v_j - u_k^d \right\|^{\alpha_{AG}}} \\ I_{u_k^d}^D &= \sum_{d_j \in \Phi_D, d_j \neq d_k} P_D h_{d_j u_k^d} \left\| d_j - u_k^d \right\|^{-\alpha_{GG}}, \end{aligned} \tag{2.1}$$

where h_{xy} denotes the channel power gain between ground nodes x and y which follows the Rayleigh distribution, $\|a - b\|$ the distance between a and b, and α_{GG} the path loss exponent from a ground node to another ground node. $I_{u_k^d}^V$ is the aggregate interference power at u_k^d from active UAVs, $I_{u_k^d}^D$ is the aggregate interference power at u_k^d from D2D-Txs, ${\sigma_n}^2$ is the noise power, and Φ_V^{act} denotes the set of active UAVs which successfully access the channel through spatial spectrum sensing. For a ground D2D-Rx, the received interfering signals from active UAVs include LOS signals, Non-LOS (NLOS) signals, and multiple reflected components which cause multipath fading [34, 35]. $P_{LOS}\left(v_j, u_k^d\right)$ and $P_{NLOS}\left(v_j, u_k^d\right)$ denote the occurrence probabilities of the LOS and NLOS links between the j^{th} UAV v_j and the typical D2D-Rx u_k^d, where the summation of these two occurrence probabilities equals to one. Parameter α_{AG} is the path loss exponent of the A2G link. The path loss is higher in a NLOS than in a LOS connection because of shadowing and indirect signal paths. Parameter $\eta < 1$ is the excessive attenuation factor for NLOS. The Nakagami distribution can be used to describe the small scale fading in A2G and G2A channels [36, 37]. We use g_{xy} to denote the channel power gain between x and y in A2G and G2A connections which follows a normalized gamma distribution with parameter M.

Specifically, according to [35], we have

$$\begin{aligned}
&P_{LOS}\left(v_j, u_k^d\right) = \frac{1}{1 + C\exp\left[-B\left(\theta - C\right)\right]},\\
&\text{where}\\
&\theta = \frac{180}{\pi}\arctan\left(\frac{h_{v_j}}{r_{v_j u_k^d}}\right),
\end{aligned} \tag{2.2}$$

where C and B are constant values depending on the communications environment, e.g. rural, urban, or dense urban, h_{v_j} is the height of the UAV v_j, and $r_{v_j u_k^d}$ denotes the horizontal distance between UAV v_j and D2D-Rx u_k^d. In addition, we have

$$P_{NLOS}\left(v_j, u_k^d\right) = 1 - P_{LOS}\left(v_j, u_k^d\right). \tag{2.3}$$

The SINR of a typical UAV receiver (UAV-Rx) u_i^v associated with UAV v_i is given by

$$\begin{aligned}
&SINR\left(u_i^v\right) = \frac{P_V L\left(v_i, u_i^v\right) g_{v_i u_i^v} \|v_i - u_i^v\|^{-\alpha_{AG}}}{I_{u_i^v}^V + I_{u_i^v}^D + {\sigma_n}^2}\\
&\text{where}\\
&L\left(v_i, u_i^v\right) = P_{LOS}\left(v_i, u_i^v\right) + \eta P_{NLOS}\left(v_i, u_i^v\right),\\
&I_{u_i^v}^D = \sum_{d_j \in \Phi_D} P_D h_{d_j u_i^v} \|d_j - u_i^v\|^{-\alpha_{GG}}\\
&I_{u_i^v}^V = \sum_{v_j \in \Phi_V^{act}, v_j \neq v_i} \frac{P_{LOS}\left(v_j, u_i^v\right) P_V g_{v_j u_i^v}}{\|v_j - u_i^v\|^{\alpha_{AG}}} + \sum_{v_j \in \Phi_V^{act}, v_j \neq v_i} \frac{P_{NLOS}\left(v_j, u_i^v\right) \eta P_V g_{v_j u_i^v}}{\|v_j - u_i^v\|^{\alpha_{AG}}},
\end{aligned} \tag{2.4}$$

where $P_{LOS}\left(v_i, u_i^v\right)$ and $P_{NLOS}\left(v_i, u_i^v\right)$ are the occurrence probabilities of LOS and NLOS connections between the typical UAV v_i and its associated receiver u_i^v. The term $I_{u_i^v}^V$ in (2.4) represents the aggregate interference power at the typical ground UAV-Rx u_i^v caused by the active UAVs . $I_{u_i^v}^D$ is the aggregate interference power at u_i^v caused by the D2D-Txs.

2.2.3 Spatial Spectrum Sensing

We define the spatial spectrum sensing sphere of a typical UAV v_i at the height of h_{v_i} as

$$\begin{aligned}&\mathcal{S}_{v_i} = \left\{ \upsilon\left(x_\upsilon, y_\upsilon, z_\upsilon\right) \in \mathbb{V}^3 \middle| \left\|v_i - \upsilon\right\| \leqslant \mathcal{R}_s \right\} \\ &\text{and } \left\|v_i - \upsilon\right\| = \sqrt{\left(x_{v_i} - x_\upsilon\right)^2 + \left(y_{v_i} - y_\upsilon\right)^2 + \left(h_{v_i} - z_\upsilon\right)^2},\end{aligned} \tag{2.5}$$

where x_υ, y_υ and z_υ represent the coordinates of the point υ, $\mathbb{V}^3$ denotes the 3D space, $\left\|v_i - \upsilon\right\|$ is the spatial distance between UAV v_i and $\upsilon\left(x_\upsilon, y_\upsilon, z_\upsilon\right)$, and $\mathcal{R}_s$ is the spatial spectrum sensing radius of UAV. The intersection between $\mathcal{S}_{v_i}$ and the ground is denoted by $\mathcal{A}_{v_i} = \mathcal{S}_{v_i} \cap \mathbb{R}^2$, where $\mathbb{R}^2$ is the horizontal ground plane. More specifically, we have

$$\begin{aligned}&\mathcal{A}_{v_i} = \left\{ m\left(x_m, y_m\right) \in \mathbb{R}^2 \middle| \left\|v_i - m\right\| \leqslant \sqrt{{\mathcal{R}_s}^2 - {h_{v_i}}^2} \right\} \\ &\text{and } \left\|v_i - m\right\| = \sqrt{\left(x_{v_i} - x_m\right)^2 + \left(y_{v_i} - y_m\right)^2},\end{aligned} \tag{2.6}$$

where $m\left(x_m, y_m\right)$ denotes a point located in $\mathcal{A}_{v_i}$, and $\left(x_m, y_m\right)$ are its coordinates. Expression $\left\|v_i - m\right\|$ is the horizontal distance between UAV v_i and ground point m. Fig. 2.2 illustrates the geometrical setup and parameters.

Let $\mathcal{H}^0$ be the event that there is no D2D-Tx in $\mathcal{A}_{v_i}$, and $\mathcal{H}^1$ be the event that there is at least one D2D-Tx in $\mathcal{A}_{v_i}$. We assume that in each time slot of duration T, all UAVs first perform spatial spectrum sensing of duration τ, and the UAVs which access the channel transmit data in the remaining time duration $T - \tau$. At UAV v_i, the received signals $y\left[n\right]$ during spatial spectrum sensing for the events $\mathcal{H}^0$ and $\mathcal{H}^1$ are given in (2.7) and (2.8), where n is the sample index, $s_{d_k}\left[n\right]$ is the n^{th} sample from D2D-Tx d_k, and the noise $n_0\left[n\right]$ is i.i.d. circularly symmetric complex Gaussian with zero mean and variance ${\sigma_n}^2$:

$$\begin{aligned}\mathcal{H}^0 : y\left[n\right] = &\sum_{d_k \in \Phi_D, d_k \notin \mathcal{A}_{v_i}} P_{LOS}\left(d_k, v_i\right) \sqrt{\frac{P_D g_{d_k v_i}}{\left\|d_k - v_i\right\|^{\alpha_{GA}}}} s_{d_k}\left[n\right] \\ &+ \sum_{d_k \in \Phi_D, d_k \notin \mathcal{A}_{v_i}} P_{NLOS}\left(d_k, v_i\right) \sqrt{\frac{P_D \eta g_{d_k v_i}}{\left\|d_k - v_i\right\|^{\alpha_{GA}}}} s_{d_k}\left[n\right] + n_0\left[n\right],\end{aligned} \tag{2.7}$$

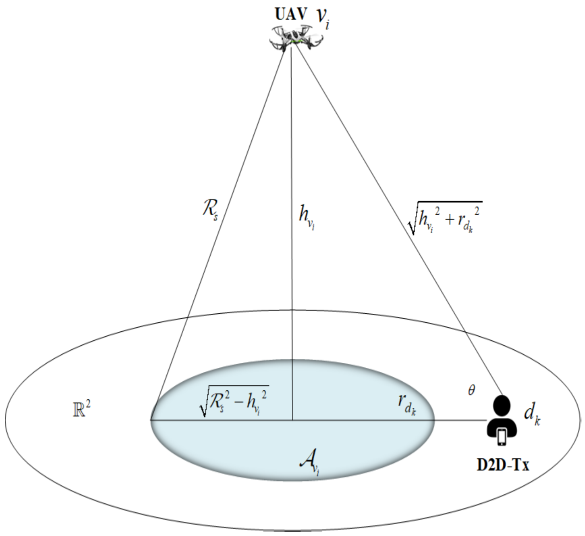


Figure 2.2: Ground intersection of spectrum sensing region of a typical UAV.

$$\begin{aligned}\mathcal{H}^1 : y\left[n\right] = &\sum_{d_k \in \Phi_D, \Phi_D \cap \mathcal{A}_{v_i} \neq \emptyset} P_{LOS}\left(d_k, v_i\right) \sqrt{\frac{P_D g_{d_k v_i}}{\left\| d_k - v_i \right\|^{\alpha_{GA}}}} s_{d_k}\left[n\right] \\ &+ \sum_{d_k \in \Phi_D, \Phi_D \cap \mathcal{A}_{v_i} \neq \emptyset} P_{NLOS}\left(d_k, v_i\right) \sqrt{\frac{P_D \eta g_{d_k v_i}}{\left\| d_k - v_i \right\|^{\alpha_{GA}}}} s_{d_k}\left[n\right] + n_0\left[n\right].\end{aligned} \tag{2.8}$$

The test statistics of the received signals at a typical UAV are given by

$$\Gamma |\, I^{\varpi} = \frac{1}{N_s} \sum\nolimits_{n=0}^{N_s - 1} \left| y\left[n\right] \right|^2, \quad \varpi = \left\{0, 1\right\}, \tag{2.9}$$

where N_s denotes the number of samples. When N_s is large, the distribution approaches a conditional Gaussian distribution because of the central limit theorem, i.e.,

$$\Gamma |\, I^{\varpi} \sim \mathcal{N}\left(I^{\varpi} + {\sigma_n}^2, \frac{\left(I^{\varpi} + {\sigma_n}^2\right)^2}{N_s} \right). \tag{2.10}$$

The mean and the variance of the conditional Gaussian distribution depend on the hypothesis [26, 38]. Note that I^0 and I^1 are random variables that depend on the network topology, D2D-Tx density and transmit powers, channel conditions, height of the sensing-based UAV and its spatial spectrum sensing radius. I^0 and I^1 are expressed as

$$I^0 = \sum_{d_k \in \Phi_D, d_k \notin \mathcal{A}_{v_i}} P_{LOS}\left(d_k, v_i\right) \frac{P_D g_{d_k v_i}}{\|d_k - v_i\|^{\alpha_{GA}}} + \sum_{d_k \in \Phi_D, d_k \notin \mathcal{A}_{v_i}} P_{NLOS}\left(d_k, v_i\right) \frac{P_D \eta g_{d_k v_i}}{\|d_k - v_i\|^{\alpha_{GA}}}, \tag{2.11}$$

$$I^1 = \sum_{d_k \in \Phi_D, \Phi_D \cap \mathcal{A}_{v_i} \neq \emptyset} P_{LOS}\left(d_k, v_i\right) \frac{P_D g_{d_k v_i}}{\|d_k - v_i\|^{\alpha_{GA}}} + \sum_{d_k \in \Phi_D, \Phi_D \cap \mathcal{A}_{v_i} \neq \emptyset} P_{NLOS}\left(d_k, v_i\right) \frac{P_D \eta g_{d_k v_i}}{\|d_k - v_i\|^{\alpha_{GA}}}. \tag{2.12}$$

The spatial false alarm probability and the spatial missed detection probability of a UAV are given by

$$P_{fa} = \mathbb{E}_{I^0}\left\{\mathbb{P}\left(\Gamma > \varepsilon \middle| \mathcal{H}^0\right)\right\}, \tag{2.13}$$

$$P_{md} = \mathbb{E}_{I^1}\left\{\mathbb{P}\left(\Gamma < \varepsilon \middle| \mathcal{H}^1\right)\right\}, \tag{2.14}$$

where ε is the energy detection threshold. The essential difference of spectrum sharing between 3D and 2D networks is the unknown conditional distribution of the aggregated received power at the UAV which determines the false alarm and missed detection probabilities.

If the test statistics received power Γ at a UAV is greater than ε, the UAV will transmit with probability β_1, otherwise, it will transmit with probability β_0, where $\beta_0 > \beta_1$. Therefore, for event $\mathcal{H}^0$, a UAV will access the licensed channel with probability

$$P^0 = P_{fa}\beta_1 + \left(1 - P_{fa}\right)\beta_0. \tag{2.15}$$

For event $\mathcal{H}^1$, a UAV will access the licensed channel with probability

$$P^1 = \left(1 - P_{md}\right)\beta_1 + P_{md}\beta_0. \tag{2.16}$$

Note that P_{fa} and P_{md} are key performance metrics for the UAVs spatial spectrum sensing, affecting the density of active UAVs and the co-channel interference to UAV and D2D communications. This will provide useful insights to design efficient spectrum sharing strategies that balance the aggressive spectrum reuse and the resulting co-channel interference.

Compared with the modeling and analysis in conventional heterogeneous networks [39], the technical challenges of analyzing spectral sharing opportunities between UAV and D2D networks include modeling the height dependent spectrum access, determining the distributions of the aggregated received power during spectrum sensing, and characterizing network interference in 3D.

2.3 False Alarm Probability and Missed Detection Probability Analysis

In this section, we provide the intermediate technical results for the system-level performance analysis, where we characterize the probability of spatial false alarm and the probability of spatial missed detection of a typical UAV. These probabilities will be used to determine the density of active UAVs and the interference from active UAVs.

2.3.1 Probability of Spatial False Alarm

For a typical UAV v_i at the height of h_{v_i}, since I^0 in (2.11) is a random variable relying on the network parameters, the probability of spatial false alarm P_{fa} in (2.13) is given by

$$P_{fa} = \int_0^\infty \mathbb{P}\left(\Gamma > \varepsilon \middle| \mathcal{H}^0, I^0 = x\right) f_{I^0}(x) dx, \tag{2.17}$$

where $f_{I^0}(x)$ is the probability density function (PDF) of I^0.

Then we derive the Laplace transform of I^0 when $\mathcal{R}_s > h_{v_i}$ to determine its distribution as

$$\begin{aligned}
&\mathcal{L}_{I^0|\mathcal{R}_s>h_{v_i}}(s) \\
&= \mathbb{E}\left\{\exp\left(-s\, I^0\middle|\, \mathcal{R}_s > h_{v_i}\right)\right\} \\
&= \mathbb{E}_{\Phi_\mathrm{D}}\left\{\prod_{d_k\in\Phi_D, d_k\notin\mathcal{A}_{v_i}} \mathbb{E}_g\left\{\exp\left(-s\frac{P_{LOS}\left(d_k, v_i\right)P_D g_{d_k v_i}}{\left\|d_k - v_i\right\|^{\alpha_{GA}}} - s\frac{P_{NLOS}\left(d_k, v_i\right)P_D \eta g_{d_k v_i}}{\left\|d_k - v_i\right\|^{\alpha_{GA}}}\right)\right\}\right\} \\
&\overset{(a)}{=} \exp\left\{-2\pi\lambda_D \int_{\sqrt{{\mathcal{R}_s}^2 - {h_{v_i}}^2}}^{\infty} \left[1 - \left(1 + \xi\left(x, P_D\right) + \Theta\left(x, P_D\right)\right)^{-M}\right] x dx\right\} \\
&\text{where} \\
&\xi\left(x, P_D\right) = \frac{sP_D\left(x^2 + {h_{v_i}}^2\right)^{-\frac{\alpha_{GA}}{2}} \Big/ M}{1 + C\exp\left[-B\left(E\arctan\left(h_{v_i}/x\right) - C\right)\right]} \\
&\Theta\left(x, P_D\right) = \frac{sP_D\eta\left(x^2 + {h_{v_i}}^2\right)^{-\frac{\alpha_{GA}}{2}}}{M} - \eta\xi\left(x, P_D\right),
\end{aligned} \tag{2.18}$$

where (a) is obtained from the expectation of the normalized gamma distribution. In addition, when $\mathcal{R}_s \leqslant h_{v_i}$, we have

$$\mathcal{L}_{I^0|\mathcal{R}_s\leqslant h_{v_i}}(s) = \exp\left\{-2\pi\lambda_D \int_0^{\infty} \left[1 - \left(1 + \xi\left(x, P_D\right) + \Theta\left(x, P_D\right)\right)^{-M}\right] x dx\right\}. \tag{2.19}$$

Using the Probability Density Function (PDF) of I^0, i.e., $f_{I^0|\cdot}(t)$, the Laplace transform of I^0 is expressed as

$$\mathcal{L}_{I^0|\cdot}(s) = \mathbb{E}\left\{e^{-sI_0}\right\} = \int_0^{\infty} e^{-st} f_{I^0|\cdot}(t) dt, \tag{2.20}$$

where $I^0|\cdot$ denotes the event I^0 under $\mathcal{R}_s > h_{v_i}$ or $\mathcal{R}_s \leqslant h_{v_i}$. The PDF can then be derived by taking the inverse Laplace transform:

$$f_{I^0|\cdot}(t) = \mathcal{L}^{-1}\left\{\mathcal{L}_{I^0|\cdot}(s)\right\} = \frac{1}{2\pi i}\lim_{T\to\infty}\int_{\gamma - iT}^{\gamma + iT} e^{st}\mathcal{L}_{I^0|\cdot}(s) ds. \tag{2.21}$$

Note that there are several other methods for deriving the distributions of I_0 and I_1 such as the inverse transform of characteristic function. However, these analytical methods involve multiple integrals and integrals in a complex domain which impede the tractable algorithms design.

The probability of spatial false alarm can be calculated by combining (2.21) into (2.17), as follows:

$$P_{fa} = \begin{cases} \int_0^\infty \zeta(x)\, f_{I^0|\mathcal{R}_s > h_{v_i}}(x)dx, \mathcal{R}_s > h_{v_i} \\ \int_0^\infty \zeta(x)\, f_{I^0|\mathcal{R}_s \leqslant h_{v_i}}(x)dx, \mathcal{R}_s \leqslant h_{v_i} \end{cases}, \tag{2.22}$$

where $\zeta(x) = Q\left(\frac{\varepsilon - x - {\sigma_n}^2}{x + {\sigma_n}^2}\sqrt{N}\right)$ and $Q(\cdot)$ is the Q-function.

2.3.2 Probability of Spatial Missed Detection

Similar to the derivation of P_{fa}, the Laplace transform of I^1 when $\mathcal{R}_s > h_{v_i}$ is obtained as

$$\begin{aligned}
&\mathcal{L}_{I^1|\mathcal{R}_s > h_{v_i}}(s) \\
&= \mathbb{E}\left\{\exp\left(-s\, I^1 \middle| \mathcal{R}_s > h_{v_i}\right)\right\} \\
&= \mathbb{E}_{\Phi_D}\left\{\prod_{d_k \in \Phi_D \cap \mathcal{S}_{v_i}, \Phi_D \cap \mathcal{A}_{v_i} \neq \emptyset} \mathbb{E}_g\left\{\exp\left(-s\frac{P_{LOS}(d_k, v_i)\, P_D g_{d_k v_i} + P_{NLOS}(d_k, v_i)\, P_D \eta g_{d_k v_i}}{\left\|d_k - v_i\right\|^{\alpha_{GA}}}\right)\right\}\right\} \\
&\quad \cdot \mathbb{E}_{\Phi_D}\left\{\prod_{d_k \in \Phi_D \cap \mathcal{A}_{v_i}^c} \mathbb{E}_g\left\{\exp\left(-s\frac{P_{LOS}(d_k, v_i)\, P_D g_{d_k v_i} + P_{NLOS}(d_k, v_i)\, P_D \eta g_{d_k v_i}}{\left\|d_k - v_i\right\|^{\alpha_{GA}}}\right)\right\}\right\} \\
&= \frac{e^{-2\pi\lambda_D \int_0^\infty \left[1-\left(1+\xi(x,P_D)+\Theta(x,P_D)\right)^{-M}\right]xdx}}{1-\exp\left[-\lambda_D \pi\left({\mathcal{R}_s}^2 - {h_{v_i}}^2\right)\right]} - \frac{\mathcal{L}_{I^0|\mathcal{R}_s > h_{v_i}}(s)\exp\left(-\lambda_D \pi\left({\mathcal{R}_s}^2 - {h_{v_i}}^2\right)\right)}{1-\exp\left[-\lambda_D \pi\left({\mathcal{R}_s}^2 - {h_{v_i}}^2\right)\right]},
\end{aligned} \tag{2.23}$$

where $\xi(x)$ and $\Theta(x)$ are given in (2.18).

Therefore, the probability of spatial missed detection can be expressed as

$$P_{md} = \begin{cases} \int_0^\infty \left(1 - \zeta(x)\right) f_{I^1|\mathcal{R}_s > h_{v_i}}(x)dx, \mathcal{R}_s > h_{v_i} \\ 0, \mathcal{R}_s \leqslant h_{v_i} \end{cases}. \tag{2.24}$$

Remark 2.1. In terrestrial communications with 2D Poisson distributed interfering nodes, the closed form PDF of I^0 can be approximated by the inverse Gaussian distribution [40] and the log-normal distribution [41] by taking the first, second and/or third order cumulants. However, these approximations do not match well with the exact values of I^0 and I^1 in 3D networks. Therefore, it is crucial to obtain the simple and tractable approximated distributions of the aggregate interference for 3D UAV spectrum sensing to provide useful system design guidelines.

2.3.3 Machine Learning-Assisted Approach

It is difficult to compute P_{fa} and P_{md} numerically in (2.22) and (2.24) due to multiple integrals in the calculation of the inverse Laplace transform. The nonlinear regression technique can be used to obtain the approximated PDFs of I^0 and I^1 to facilitate system-level performance analysis [42]. Based on the distance-dependent path-loss, the PDFs of the aggregated received signal strength I^0 and I^1 have more values within a specific region and have a long tail. By comparing some well-known continuous distributions, we select the Log-normal distribution with appropriate mean and standard deviation values to approximate the distributions of I^0 and I^1.

Proposition 2.2. *The PDFs of the received signal strength I^0 and I^1 for the events $\mathcal{H}^0$ and $\mathcal{H}^1$ can be well approximated by the log-normal distribution with appropriate mean and standard deviation values. More specifically, we have*

$$f_{I^{\varpi}}\left(x\right) \approx \frac{1}{x\sigma_{I^{\varpi}}\sqrt{2\pi}} \exp\left[-\frac{\left(\ln x - \mu_{I^{\varpi}}\right)^2}{2{\sigma_{I^{\varpi}}}^2}\right], \tag{2.25}$$

where $\mu_{I^{\varpi}}$ and $\sigma_{I^{\varpi}}$ are the mean and the standard deviation variables.

Note that other families of distributions such as the inverse gamma or scaled inverse chi-squared can be used to approximate the distributions and show similar performance. We do not focus on the optimal approximation and distribution selection, but we are more interested in the system-level performance evaluation.

In terrestrial communications with 2D Poisson distributed interfering nodes, the values of μ_{I^0} and σ_{I^0} can be obtained by calculating the first and second cumulants of I^0 [40]. However, the results obtained by the first and second cumulants of I^0 are no longer accurate under 3D UAV channels due to the LOS and NLOS nature of G2A connections. The Gaussian kernel nonlinear regression is used to explore the relationship between input (network parameters) and output (mean and the standard deviation of log-normal distribution). In Fig. 2.3, the input network parameters are the density λ_D of D2D-Txs, UAV's flight height h_{v_i}, UAV's spatial spectrum sensing radius $\mathcal{R}_s$, and the channel power gain M of G2A transmission. The output is the mean or the standard deviation of the approximated log-normal distribution of $I^{\varpi}, \varpi = \{0, 1\}$.

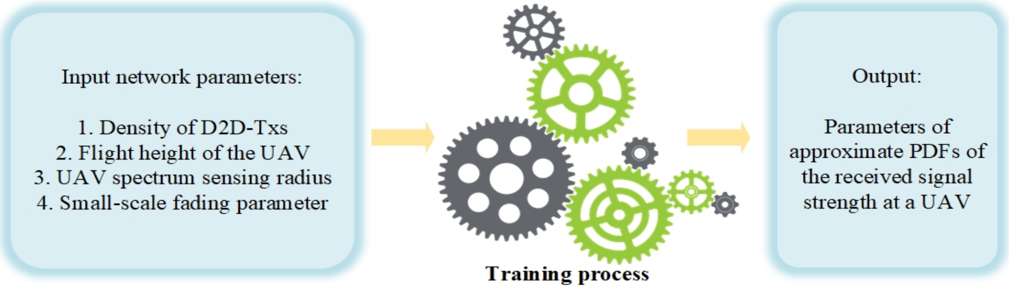


Figure 2.3: Input and output of the machine learning-based approach.

Remark 2.3. In the training process, the transmit power of D2D-Txs is not set as an input. We use the scaling property of the log-normal distribution to update the mean value of the approximated log-normal distribution, i.e, if $I^{\varpi} \sim \text{Lognormal}\left(\mu_{I^{\varpi}}, \sigma_{I^{\varpi}}\right)$, then we have $aI^{\varpi} \sim \text{Lognormal}\left(\mu_{I^{\varpi}} + \ln a, \sigma_{I^{\varpi}}\right)$.

We use Monte Carlo simulations with diverse input values to generate the data set for training. For this data set generation, the mean and the standard deviation values of the log-normal distribution can be obtained as follows

$$\mu_{I^{\varpi}} = \ln\left(\frac{\mathbb{E}\{I^{\varpi}\}^2}{\sqrt{\mathbb{E}\left\{(I^{\varpi})^2\right\}}}\right), \tag{2.26}$$

$$\sigma_{I^{\varpi}} = \left[\ln\left(\frac{\mathbb{E}\left\{(I^{\varpi})^2\right\}}{\mathbb{E}\{I^{\varpi}\}^2}\right)\right]^{\frac{1}{2}}. \tag{2.27}$$

We denote the input variables as $\mathbf{x}^{(i)} = \left[x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, x_4^{(i)}\right]$ (corresponding to the network parameters λ_D, h_{v_i}, $\mathcal{R}_s$, M). The output variable is denoted by $y^{(i)}$ (corresponding to the $\mu_{I^{\varpi}}$ or $\sigma_{I^{\varpi}}, \varpi = \{0, 1\}$), where (i) indicates the data set index. To balance the weights of the different inputs, we normalize the input variables between 0 and 1 as follows

$$\widetilde{x}_j^{(i)} = \frac{x_j^{(i)} - \min(\mathbf{x}_j)}{\max(\mathbf{x}_j) - \min(\mathbf{x}_j)}, j = 1, \cdots, 4, \tag{2.28}$$

where $\mathbf{x}_j = \left[x_j^{(1)}, \cdots, x_j^{(L)}\right]$ and L is the number of input-output pairs in the generated data set. We then obtain the normalized input variables $\widetilde{\mathbf{x}}^{(i)} = \left[\widetilde{x}_1^{(i)}, \widetilde{x}_2^{(i)}, \widetilde{x}_3^{(i)}, \widetilde{x}_4^{(i)}\right]$.

To estimate the output value given the input network parameters, we use the Gaussian Kernel function to calculate the weighted average output values as follows

$$f(x) = \frac{1}{\sqrt{2\pi{\sigma_G}^2}} \exp\left[-\frac{(x - \mu_G)^2}{2{\sigma_G}^2}\right], \tag{2.29}$$

where μ_G and σ_G are the mean and the standard deviation of the Gaussian kernel function.

The query point is denoted by $\mathbf{q} = [q_1, q_2, q_3, q_4]$, which collects the local information of the data set. The distance between the query point $\mathbf{q}$ and the normalized data point $\widetilde{\mathbf{x}}^{(i)}$, i.e., $D\left(\mathbf{q}, \widetilde{\mathbf{x}}^{(i)}\right)$, is given by

$$D\left(\mathbf{q}, \widetilde{\mathbf{x}}^{(i)}\right) = \sqrt{\left(\mathbf{q} - \widetilde{\mathbf{x}}^{(i)}\right)\left(\mathbf{q} - \widetilde{\mathbf{x}}^{(i)}\right)^{\mathrm{T}}}. \tag{2.30}$$

Note that the term $\frac{1}{\sqrt{2\pi{\sigma_G}^2}}$ in (2.29) will not impact the weighted average value at the query point $\mathbf{q}$. Therefore, the kernel function in the training process can be expressed as

$$K\left(\mathbf{q},\widetilde{\mathbf{x}}^{(i)}\right)=\exp\left(-\frac{D\left(\mathbf{q},\widetilde{\mathbf{x}}^{(i)}\right)^{2}}{2{\sigma_G}^{2}}\right), \tag{2.31}$$

where $D\left(\mathbf{q},\widetilde{\mathbf{x}}^{(i)}\right)$ is given in (2.30). Then, the estimate output value at the query point $\mathbf{q}$ is

$$\widehat{y}\left(\mathbf{q}\right)=\frac{\sum\limits_{i=1}^{L}\left(K\left(\mathbf{q},\widetilde{\mathbf{x}}^{(i)}\right)y^{(i)}\right)}{\sum\limits_{i=1}^{L}K\left(\mathbf{q},\widetilde{\mathbf{x}}^{(i)}\right)}, \tag{2.32}$$

where $K\left(\mathbf{q},\widetilde{\mathbf{x}}^{(i)}\right)$ represents the weight value of the input-output pair $\left(\widetilde{\mathbf{x}}^{(i)},y^{(i)}\right)$. After evaluating (2.32) over a range of query points, given an input combination of the network parameters, we can search the nearest query point to the inputs and obtain the corresponding output value. After that, we use the approximated PDFs of I^0 and I^1 in (2.22) and (2.24) to obtain the approximated spatial false alarm probability and spatial missed detection probability.

In detection theory, the Neyman-Pearson criterion says that one can minimize the spatial missed detection probability P_{md} while not allowing the spatial false alarm probability P_{fa} to exceed a predefined value ${P_{fa}}^{*}$, i.e., $P_{fa}\leqslant{P_{fa}}^{*}$, or minimize P_{fa} subject to a constraint on P_{md}. In this chapter, we assume a constant $P_{fa}={P_{fa}}^{*}$. Therefore, the spatial spectrum sensing radius and the energy detection threshold of a UAV are coupled through the UAV's spatial spectrum sensing. In practical engineering design, based on the policy $P_{fa}={P_{fa}}^{*}$, we can adjust the energy detection threshold to change the spatial spectrum sensing radius. It is worth noting that the proposed machine learning-assisted approach is applicable to other user distributions such as Poisson cluster process and Poisson hole process. However, the exact closed-form analytical expression of energy detection threshold cannot be derived for these, which provides few information in performance optimization.

2.4 System-Level Performance Analysis

In this section, we derive the coverage probability of a typical D2D communication network and the coverage probability of a typical UAV communication network conditioned on the UAV's flight height h_{v_i} and the distance r_i^v between UAV-Rx u_i^v and the projection of its associated UAV v_i on the ground. These results will be used to determine the ASE of the D2D and UAV networks.

Note that a UAV can access the licensed channel with probability $P^0 = P_{fa}\beta_1 + (1 - P_{fa})\,\beta_0$ for event $\mathcal{H}^0$ and with probability $P^1 = (1 - P_{md})\,\beta_1 + P_{md}\beta_0$ for event $\mathcal{H}^1$, where P_{fa} and P_{md} are given in (2.22) and (2.24), respectively, and the PDFs of I^0 and I^1 are given in Proposition 2.2 with its parameters trained using a Gaussian nonlinear kernel-based machine learning method.

Lemma 2.4. *In the spatial spectrum sensing process, the channel access probability P^0 under event $\mathcal{H}^0$ is greater than the channel access probability P^1 under event $\mathcal{H}^1$.*

Proof. Please refer to Appendix A.1 for the proof. □

It's worth noting that UAVs at different heights experience different aggregated received powers from D2D transmissions because of the varying elevation angles and intersection region between the spatial spectrum sensing sphere and ground. The specific aggregated received power at each UAV leads to different secondary channel access probabilities. In other words, UAVs at different flight heights will generate different levels of interference to the primary D2D network on the ground. For analytical tractability, we uniformly divide the height $\mathbb{V}_A$ into N sub-regions with a common thickness of ΔH, i.e.,

$$H_{\max} - H_{\min} = N \cdot \Delta H. \tag{2.33}$$

H_n approximates the height of UAVs in the n^{th} sub-region $\mathbb{V}(x, y, H_n)$. In $\mathbb{V}(x, y, H_n)$, we assume that UAVs are uniformly distributed on the horizontal plane at height H_n. The set of UAVs in $\mathbb{V}(x, y, H_n)$ is denoted as Φ_{V,H_n}. Considering the randomness of UAV locations, we assume that Φ_{V,H_n} follows a homogeneous PPP with density

$$\lambda_{V,H_n} = \lambda_V \int_{H_n}^{H_n + \Delta H} f_H(h) dh. \tag{2.34}$$

If $\mathcal{R}_s > H_n$, the active probability of UAVs in Φ_{V,H_n} is statistically equivalent to the case where all UAVs in Φ_{V,H_n} transmit with probability P^1 and additionally the UAVs in Φ_{V,H_n}, which have no D2D-Txs within their spatial spectrum sensing spheres, transmit with probability $P^0 - P^1$. Thus, the average channel access probability of UAVs at height H_n is

$$P_{UAV}^{act,n} = P^1 + \left(P^0 - P^1\right) e^{-\pi\lambda_D\left({\mathcal{R}_s}^2 - {H_n}^2\right)}. \tag{2.35}$$

For analytical tractability, we approximate the locations of active UAVs in $\mathbb{V}(x, y, H_n)$ as randomly distributed with the density of $P_{UAV}^{act,n}\lambda_{V,H_n}$. Φ_{V,H_n}^L denotes the set of active UAVs in $\mathbb{V}(x, y, H_n)$.

If $\mathcal{R}_s \leqslant H_n$, all UAVs in $\mathbb{V}(x, y, H_n)$ will transmit with probability P^0. Thus, the locations of active UAVs are modeled by a PPP Φ_{V,H_n}^S with density $P^0\lambda_{V,H_n}$.

We are now in the position of computing the coverage probability of a typical D2D communication network and a typical UAV communication network.

2.4.1 Coverage Probability of D2D Communications

We obtain the coverage probability of a typical D2D-Rx in the following theorem.

Theorem 2.5. *The coverage probability of a typical D2D-Rx* u_k^d *conditioned on the D2D* *serving distance* l_k^d *is given in equation (2.36).*

$$
\begin{aligned}
&\mathbb{P}^c_D |\, l_k^d \\
&= \exp\left(-\frac{2\pi^2 \lambda_D {\gamma_D^{th}}^{\frac{2}{\alpha_{GG}}} \left(l_k^d\right)^2}{\alpha_{GG} \sin\left(\frac{2\pi}{\alpha_{GG}}\right)} - 2\pi \sum_{n=1}^{N_h} P_{UAV}^{act,n} \lambda_{V,H_n} \int_{\sqrt{{\mathcal{R}_s}^2 - {H_n}^2}}^{\infty} x\Omega\left(x, H_n\right) dx \right) \\
&\quad \cdot \exp\left(-2\pi \sum_{n=N_h+1}^{N} P^0 \lambda_{V,H_n} \int_0^{\infty} x\Omega\left(x, H_n\right) dx - \frac{\gamma_D^{th} {\sigma_n}^2}{P_D \left(l_k^d\right)^{-\alpha_{GG}}} \right)
\end{aligned}
$$

where

$$
\begin{aligned}
&P_{UAV}^{act,n} = P^1 + \left(P^0 - P^1\right) e^{-\pi\lambda_D\left({\mathcal{R}_s}^2 - {H_n}^2\right)} \\
&\Omega\left(x, H_n\right) = 1 - \left(1 + \Delta\left(x, H_n\right) \frac{\eta_M \gamma_D^{th} P_V \left(x^2 + {H_n}^2\right)^{-\frac{\alpha_{AG}}{2}}}{M P_D \left(l_k^d\right)^{-\alpha_{GG}}} \right. \\
&\qquad\qquad \left. + \left(1 - \Delta\left(x, H_n\right)\right) \frac{\eta_M \gamma_D^{th} P_V \eta \left(x^2 + {H_n}^2\right)^{-\frac{\alpha_{AG}}{2}}}{M P_D \left(l_k^d\right)^{-\alpha_{GG}}} \right)^{-M} \\
&\Delta\left(x, H_n\right) = \frac{1}{1 + C \exp\left[-B\left(E \arctan\left(H_n/x\right) - C\right)\right]} \\
&\eta_M = M (M!)^{-\frac{1}{M}}, .
\end{aligned}
\tag{2.36}
$$

Proof. Please refer to Appendix A.2 for the proof. ☐

2.4.2 Coverage Probability of UAV Communications

A typical ground UAV-Rx receives interference from D2D and UAV communications. Note that the desired serving link distance of a typical UAV v_i is dependent on UAV's flight height and distance r_i^v which is the distance between UAV-Rx and the projection of UAV on the ground.

Theorem 2.6. *The coverage probability of the attached UAV-Rx* u_i^v *conditioned on the* *distance* r_i^v *and the UAV's flight height* h_{v_i} *is given in (2.37).*

$$\begin{aligned}
&\mathbb{P}_V^c | \, r_i^v, h_{v_i} \\
&\approx \sum_{m=1}^{M} \left\{ \left(\begin{array}{c} M \\ m \end{array} \right) (-1)^{m+1} \exp \left(- \frac{2\pi^2 \lambda_D \left(\frac{m \varsigma_V \gamma_V^{th} P_D}{P_V L\left(v_i, u_i^v\right)} \right)^{\frac{2}{\alpha_{GG}}}}{\alpha_{GG} \sin \left(\frac{2\pi}{\alpha_{GG}} \right)} \left(\left(r_i^v\right)^2 + \left(h_{v_i}\right)^2 \right)^{\frac{\alpha_{AG}}{\alpha_{GG}}} \right) \right. \\
&\quad \cdot \exp \left(-2\pi \sum_{n=1}^{N_h} P_{UAV}^{act,n} \lambda_{V,H_n} \int_0^\infty x \Xi \left(m, x, H_n, r_i^v, h_{v_i} \right) dx \right) \\
&\quad \cdot \exp \left(-2\pi \sum_{n=N_h+1}^{N} P^0 \lambda_{V,H_n} \int_0^\infty x \Xi \left(m, x, H_n, r_i^v, h_{v_i} \right) dx \right) \\
&\quad \left. \cdot \exp \left(- \frac{m \eta_M \gamma_V^{th} {\sigma_n}^2}{P_V L \left(v_i, u_i^v \right) \left(\left(r_i^v\right)^2 + \left(h_{v_i}\right)^2 \right)^{-\frac{\alpha_{AG}}{2}}} \right) \right\} \qquad (2.37)
\end{aligned}$$

where

$$\begin{aligned}
L\left(r_i^v, h_{v_i}\right) &= \frac{1}{1 + C \exp \left[-B \left(E \arctan \left(\frac{h_{v_i}}{r_i^v} \right) - C \right) \right]} \\
&\quad + \eta \left(1 - \frac{1}{1 + C \exp \left[-B \left(E \arctan \left(\frac{h_{v_i}}{r_i^v} \right) - C \right) \right]} \right) \\
\Xi \left(m, x, H_n, r_i^v, h_{v_i} \right) &= 1 - \left[1 + \Delta(x, H_n) \frac{m \eta_M \gamma_V^{th} \left(x^2 + {H_n}^2 \right)^{-\frac{\alpha_{AG}}{2}}}{M L \left(v_i, u_i^v \right) \left(\left(r_i^v\right)^2 + \left(h_{v_i}\right)^2 \right)^{-\frac{\alpha_{AG}}{2}}} \right. \\
&\qquad \left. + \left(1 - \Delta(x, H_n) \right) \frac{m \eta_M \gamma_V^{th} \eta \left(x^2 + {H_n}^2 \right)^{-\frac{\alpha_{AG}}{2}}}{M L \left(v_i, u_i^v \right) \left(\left(r_i^v\right)^2 + \left(h_{v_i}\right)^2 \right)^{-\frac{\alpha_{AG}}{2}}} \right]^{-M},
\end{aligned}$$

Proof. Please refer to Appendix A.3 for the proof. □

2.4.3 Area Spectral Efficiency

The ASE of UAV and D2D networks can be obtained from the previous results. We denote the PDF of the distance between UAV-Rx and the ground projection of its associated UAV as $f_{r_i^v}(r)$. For example, if we consider that r_i^v is uniformly distributed in a circular region centered at the projection of UAV v_i with radius $R_{\max}$, then we have $f_{r_i^v}(r) = \frac{2r}{{R_{\max}}^2}, r \geqslant 0$.

The ASE of UAV networks is given by

$$\begin{aligned} & ASE_V \\ & = \frac{T-\tau}{T}\left(\sum_{n=1}^{N_h} P_{UAV}^{act,n}\lambda_{V,H_n}\, \overline{\mathbb{P}_V^c}\big|\, H_n + \sum_{n=N_h+1}^{N} P^0\lambda_{V,H_n}\, \overline{\mathbb{P}_V^c}\big|\, H_n\right)\log_2\left(1+\gamma_V^{th}\right), \end{aligned} \tag{2.38}$$

where N_h satisfies that $H_{N_h} \approx \mathcal{R}_s$, $\overline{\mathbb{P}_V^c}\big|\, H_n = \mathbb{E}_r\left\{\mathbb{P}_V^c\big|\, r, H_n\right\}$, $\mathbb{P}_V^c\big|\, r, H_n$ is given in Theorem 2, and $P_{UAV}^{act,n}$ is obtained in (2.35).

The ASE of D2D networks is given by

$$ASE_D = \lambda_D \overline{\mathbb{P}_D^c}\log_2\left(1+\gamma_D^{th}\right), \tag{2.39}$$

where $\overline{\mathbb{P}_D^c} = \mathbb{E}_l\left\{\mathbb{P}_D^c\big|\, l\right\}$ and $\mathbb{P}_D^c\big|\, l$ is given in (2.36).

Based on our analytical framework, we can maximize the ASE of UAV networks while guaranteeing that the ASE of D2D networks is not below a certain value ϑ, as follows

$$\begin{aligned} \max_{\mathcal{R}_s} \;\; & ASE_V \\ \text{s.t.} \;\; & ASE_D \geqslant \vartheta, \end{aligned} \tag{2.40}$$

where ASE_V and ASE_D are given in (2.38) and (2.39), respectively, ϑ denotes the ASE threshold of D2D networks. It is worth noting that decreasing $\mathcal{R}_s$ leads to a more aggressive spectrum reuse for UAV communications; however, it also generates more severe co-channel interference, which reduces the D2D communications performance. Therefore, there is a clear trade-off between aggressive spectrum reuse and co-channel interference. Alternatively, based on our model, one can evaluate the entire network performance using the weighted Tchebycheff method [43] and obtain the Pareto-optimality criterion for spectrum sharing between UAV and D2D communications. To solve the problem (2.40), we leverage the one-dimensional numerical search method using the closed-form expressions of ASEs for both the UAV and D2D networks. It is worth noting that the ASE of both the UAV and D2D networks can be approximated by the machine learning approach. However, it involves many input network parameters which makes the data set generation difficult. In addition, when the operation of networks changes, the scalability of the approach that directly approximates the ASE by machine learning is insufficient.

2.5 Simulation Results and Discussion

In this section, we verify our analysis by simulations and evaluate the performance of the D2D and UAV spectrum sharing networks. The simulated network uses the following parameter settings, unless otherwise stated: $\lambda_D = 1 \times 10^{-5}/\text{m}^2$, $\lambda_V = 1 \times 10^{-5}/\text{m}^2$, $P_D = 20\text{mW}$, $P_V = 20\text{mW}$, $M = 1$, $N = 10$, $H_{min} = 10\text{m}$, $H_{max} = 100\text{m}$, $\alpha_{GG} = 4$, $\alpha_{GA} = \alpha_{AG} = 2.1$, ${P_{fa}}^* = 0.1$, ${\sigma_n}^2 = -110\text{dBm}$, $\beta_0 = 0.8$, $\beta_1 = 0.16$ [26], $\eta = 0.1$, $B = 0.136$ and $C = 11.95$ [35].

In Fig. 2.4, we show a 3D network for the coexistence of D2D and UAV communications. D2D-Txs are uniformly distributed on the ground with the density $\lambda_D = 1 \times 10^{-5}/\text{m}^2$. D2D-Rxs are located at positions with random directions and distances to its associated D2D-Txs between 20m and 80m. The UAVs are uniformly distributed in the aerial 3D space at a height of $[10\text{m}, 100\text{m}]$. The density of the projections of UAVs on the ground is $\lambda_V = 1 \times 10^{-5}/\text{m}^2$.

Fig. 2.5 compares the PDF of the received signal strength at a UAV for the event $\mathcal{H}^0$, where the UAV is at a height of 60m and has the spatial spectrum sensing radius of $\mathcal{R}_s = 150\text{m}$. The results are obtained from Monte-Carlo simulations, whereas the red dashed line is obtained by the proposed machine learning-based approximation in Section 2.3.3. We observe that the proposed approximation method accurately depicts the PDF of the received aggregated power generated from ground D2D communications for the event $\mathcal{H}^0$ during spatial spectrum sensing. The complexity of the Monte-Carlo simulations are more time-consuming than the

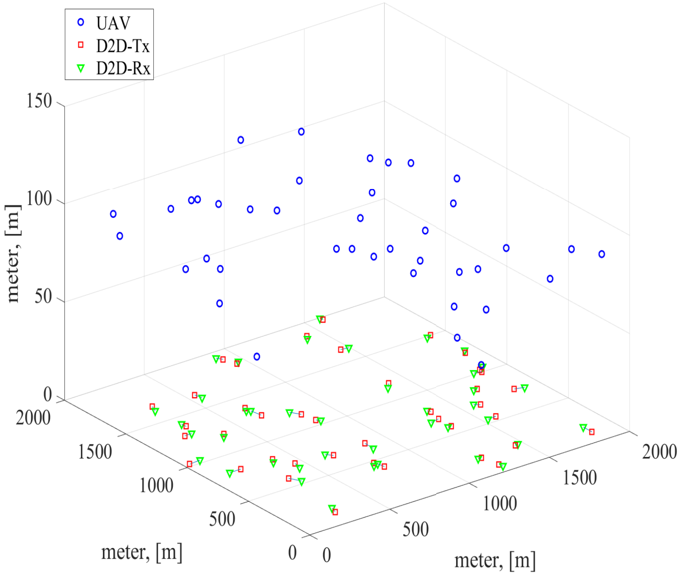


Figure 2.4: Simulation scenario where D2D-Txs are uniformly distributed on the ground and UAVs are uniformly distributed in the allowable flight space within the height of [10m, 100m].

proposed ML-based approach. This is because the ML-based approach needs to collect the data set in the early stage before the training process, whereas the trained distributions can be used permanently. However, for the Monte-Carlo simulations, we need to average multiple independent trials to obtain a stable performance, and, in each trial, the calculation of the network interference at each node is time-consuming.

Fig. 2.6 shows the cumulative distribution functions (CDFs) of the received signal strength at a UAV for the event $\mathcal{H}^0$. We observe that the machine learning-based approximation method well approximates the distribution of the received aggregate power from D2D communications for the event $\mathcal{H}^0$. In addition, we observe that when the UAV flight height increases, the received signal strength improves. This is because an increase in UAV height results in a decrease of the radius of $\mathcal{A}_{v_i}$ defined in (2.6) and improves the probability of LOS connections between D2D-Txs and UAV due to the increased elevation angle. Besides,

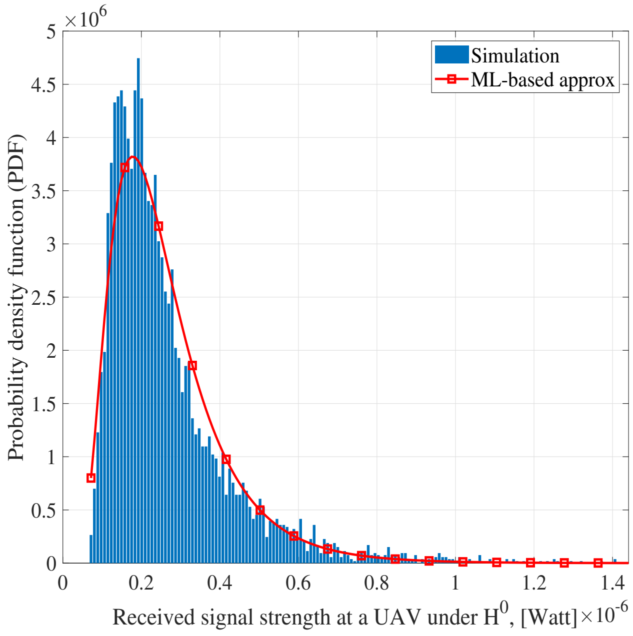


Figure 2.5: PDF of a the received signal strength at the UAV during spatial spectrum sensing for the false alarm case and the flight height of $h_v = 60$m, $\mathcal{R}_s = 150$m.

a higher value of small-scale fading parameter M improves the received signal strength.

In Fig. 2.7, the CDFs of received signal strength at a UAV for the event $\mathcal{H}^1$ are shown. When $\mathcal{R}_s$ increases, the received signal strength at a UAV under $\mathcal{H}^1$ decreases, because of the longer distances between D2D-Txs and the UAV and the higher probability of their NLOS connections in these scenarios. In addition, when a UAV's flight height increases, the received signal strength increases. The reason behind this is similar to that of Fig. 2.6.

In Fig. 2.8, we compare the UAV energy detection threshold ε with respect to UAV spatial spectrum sensing radius $\mathcal{R}_s$. Using the Neyman-Pearson criterion and assuming $P_{fa} = P^*_{fa}$, there exists a mapping between $\mathcal{R}_s$ and ε. It can be observed that ε decreases with $\mathcal{R}_s$. This is because the received signal strength I^0 under $\mathcal{H}^0$ decreases as $\mathcal{R}_s$ increases. To achieve the target P^*_{fa}, the energy detection threshold needs to be accordingly reduced. In addition, when the UAV's flight height h_v increases, its energy detection threshold increases, because

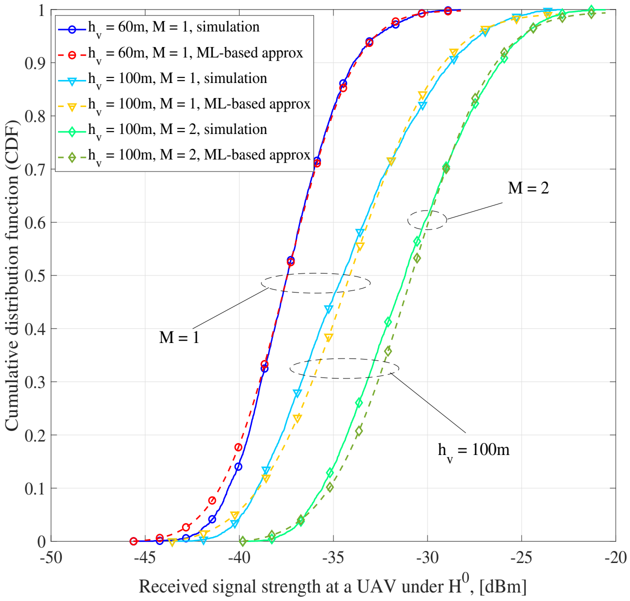


Figure 2.6: CDF of a UAV's received signal strength during spatial spectrum sensing for the false alarm case, where $\mathcal{R}_s = 150$m.

UAVs at higher heights receive more power from D2D communications under the event $\mathcal{H}^0$ as shown in Fig. 2.6. In Fig. 2.8, we observe that the proposed machine learning-based approximation cannot exactly describe the simulated ε. The reason for this is that the distribution of I^0 is not coincidentally the assumed log-normal distribution. However, we train the log-normal distribution parameters to approach the real I^0 distribution.

Fig. 2.9 plots the UAV channel access probability which is given in (2.35) over the UAV spatial spectrum sensing radius $\mathcal{R}_s$. It can be observed that the UAV channel access probability decreases with $\mathcal{R}_s$. This can be explained as follows: under the assumption of $\beta_0 > \beta_1$, increasing $\mathcal{R}_s$ results in a decrease of the UAV energy detection threshold according to Fig. 2.8. Thus, a UAV has a high probability of transmitting with probability β_1. Therefore, adjusting $\mathcal{R}_s$ can regulate the density of active UAVs and hence, the interference power due to UAV communications. More specifically, decreasing $\mathcal{R}_s$ leads to more severe interference

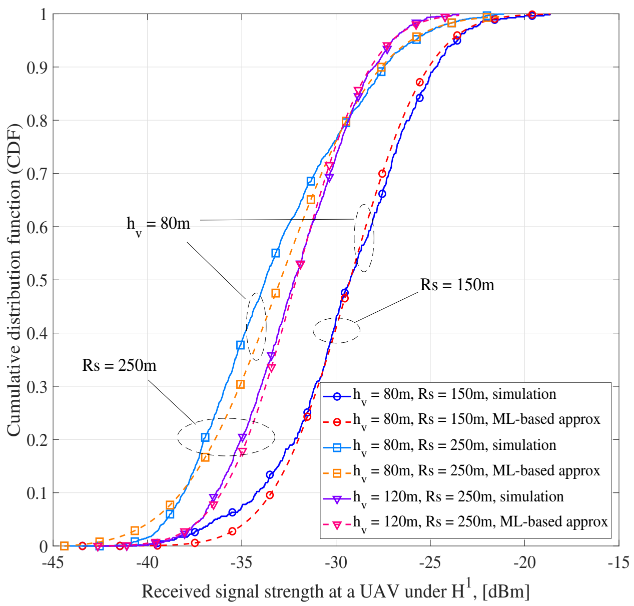


Figure 2.7: CDF of a UAV's received signal strength during spatial spectrum sensing in missed detection for different UAV flight heights and spatial spectrum sensing radiuses.

generated from UAVs. It can be observed from Fig. 2.9 that when the density of D2D-Txs becomes large, the UAV channel access probability reduces accordingly due to the incremental received signal strength from D2D-Txs at a UAV during spatial spectrum sensing. We can also find that, when $\mathcal{R}_s$ is small, increasing the UAV's flight height improves the UAV channel access probability. However, when $\mathcal{R}_s$ is large, increasing the UAV's height decreases its channel access probability. The rationale behind this is that when $\mathcal{R}_s$ is small, increasing the height largely increases the UAV's energy detection threshold. The energy detection thresholds for different UAV flight heights is not obvious when $\mathcal{R}_s$ is large, according to Fig. 2.8. On the other hand, based on the observation from the simulations, the values of total signal strength received from D2D-Txs are more concentrated around small values for a lower flight height than that for higher heights. Therefore, these can illustrate the behaviour of the crossover point of the UAV channel access probability for different flight height with

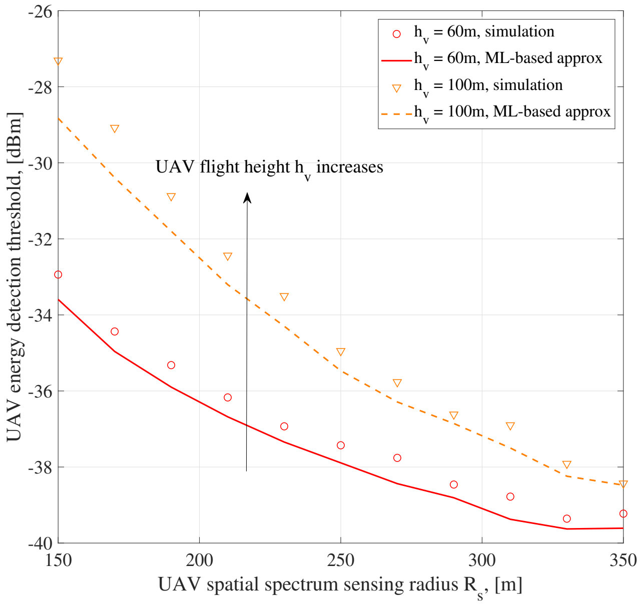


Figure 2.8: UAV energy detection threshold ε with respect to the UAV spatial spectrum sensing radius $\mathcal{R}_s$.

respect to the UAV's spatial spectrum sensing radius.

It is observed from Fig. 2.9 that when h_v is high and $\mathcal{R}_s$ is small, there are discrepancies between analytical results and simulation results. This can be explained as follows. When h_v increases, I^0 increases, due to the reduced intersection region $\mathcal{A}_{v_i}$ and the increased line-of-sight (LoS) probability between D2D-Txs and UAV. When I^0 increases, the UAV energy detection threshold ε increases accordingly to achieve the target spatial false alarm probability. When ε increases, the spatial missed detection probability (P_{md}) increases, and thus the probability P^1 increases. When h_v is high and $\mathcal{R}_s$ is small, the average channel access probability of UAV at height h_v (P_{UAV}^{act}) is dominated by P^1. Since P^1 depends on ε, a deviation in ε leads to a discrepancy in P_{UAV}^{act}. According to Fig. 2.8, we observe that when h_v is large and $\mathcal{R}_s$ is small, the discrepancy of ε between the approximate results and simulation results becomes large. This is because when h_v is high and $\mathcal{R}_s$ is small, the

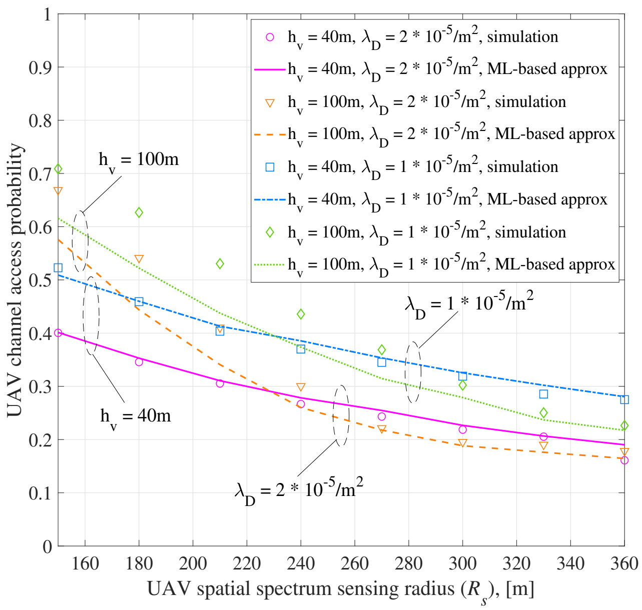


Figure 2.9: UAV channel access probability with respect to the UAV spatial spectrum sensing radius $\mathcal{R}_s$.

values of total received signal strength at a UAV are widely distributed. The approximate distribution may not accurately depict the distribution, and the statistical characteristics are not very accurate. Nevertheless, when h_v is small, the values of total received signal strength at a UAV concentrate around small values. Thus, the approximate distribution can characterize the distribution, and the statistical characteristics are satisfied. Therefore, we have the discrepancy between analytical results and simulation results when h_v is large and $\mathcal{R}_s$ is small. Moreover, this discrepancy does not have an obvious impact on the system-level performance, such as coverage probability and area spectrum efficiency.

In Fig. 2.10, the coverage probability of D2D communication is presented with respect to the UAV's spatial spectrum sensing radius $\mathcal{R}_s$. As can be seen, increasing $\mathcal{R}_s$ is beneficial to D2D communications. This is so because a larger value of $\mathcal{R}_s$ makes the UAV more sensitive of the radio environment and reduces its channel access probability, which reduces

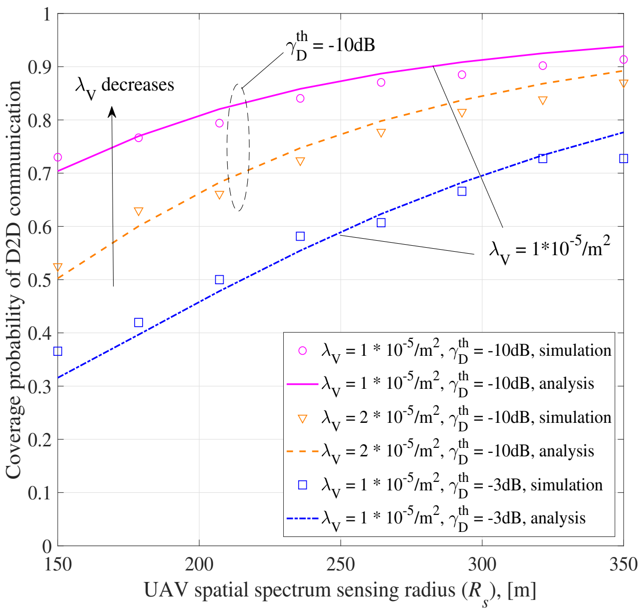


Figure 2.10: Coverage probability of a typical D2D communication vs. the UAV spatial spectrum sensing radius $\mathcal{R}_s$, where $l_k^d = 30m$.

the interference from UAVs to D2D communications. Besides, reducing the density of UAVs improves the coverage probability of D2D communications. Furthermore, from Fig. 2.10 we can observe that when $\mathcal{R}_s$ is small, the SINR threshold of D2D communications γ_D^{th} has a significant impact on the D2D communications coverage probability.

Fig. 2.11 plots the coverage probability of UAV communication as a function of the UAV's spatial spectrum sensing radius $\mathcal{R}_s$. It can be observed that the coverage probability of a typical UAV communication link increases with $\mathcal{R}_s$ due to the reduced co-channel interference generated from active UAVs. The coverage probability decreases with increasing density of UAVs λ_V and the increasing flight height h_V. The decrease of h_V results in the reduction of the serving link distance and thus enhances the desired signal strength at the UAV-Rx, despite the decreasing LOS occurrence probability. Compared with Fig. 2.9, when $\mathcal{R}_s$ is small, there is a trade-off between UAV channel access probability and coverage probability.

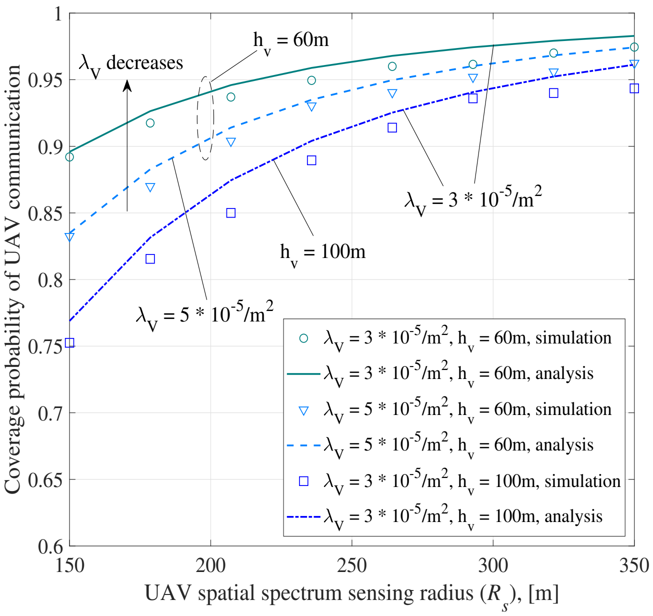


Figure 2.11: Coverage probability of a typical UAV communications vs. the UAV's spatial spectrum sensing radius $\mathcal{R}_s$, where $r_v = 30m$.

On the other hand, when $\mathcal{R}_s$ is large, it is advantageous for a UAV that has a lower flight height in terms of both coverage probability and channel access probability.

In Fig. 2.12, we compare the relate between the ASE of UAV networks ASE_V and the UAV's spatial spectrum sensing radius $\mathcal{R}_s$. The ASE_V decreases with $\mathcal{R}_s$ since the reduction of UAV channel access probability dominates the ASE of UAV networks. According to (2.40), when the ASE of D2D networks is guaranteed not less than a certain value ϑ, according to Fig. 2.10, we can obtain a minimum $\mathcal{R}_s$ which maximizes ASE_V. Besides, when increasing the UAV's transmit power P_V, we observe that the ASE_V increases. This is because of the improvement of the desired signal strength of UAV communications, while the impact of interference from D2D-Txs gradually diminished. However, increasing P_V deteriorates the coverage probability of D2D communications and thus the ASE of D2D networks. Therefore, there exists an optimal transmit power for UAVs to maximize the ASE of UAV networks

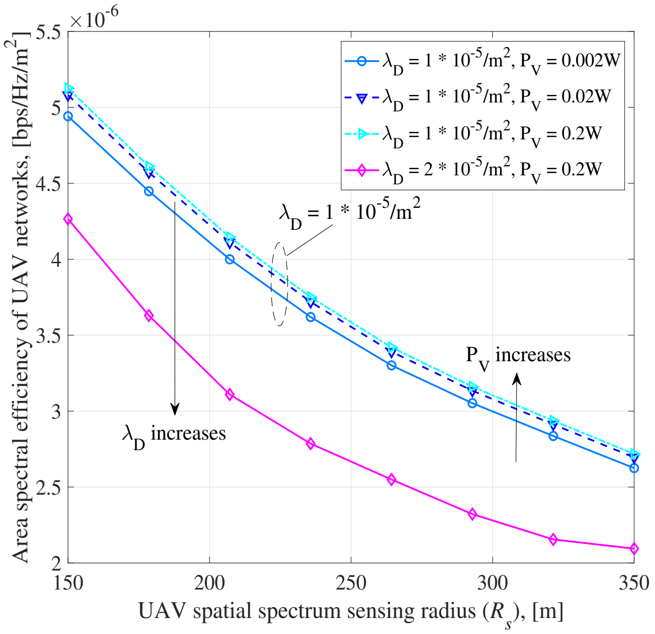


Figure 2.12: ASE of UAV networks vs. the UAV's spatial spectrum sensing radius $\mathcal{R}_s$, where $r_v = 30m$, $P_D = 0.2W$.

under the constraint of the minimum ASE_D. Also, when the density of D2D-Txs increases, ASE_V decreases due to the increased interference from D2D communications.

2.6 Summary

In this chapter, we develop a machine learning-assisted stochastic geometry framework for spectrum sharing between ground D2D and UAV communications. The D2D-Txs are regarded as the primary users while UAVs opportunistically access the licensed channel by implementing spatial spectrum sensing. We first analyze the spatial false alarm probability and the spatial missed detection probability of a typical UAV. Then, we derive the coverage probability of typical D2D-Rx and typical UAV-Rx, respectively. The ASEs of D2D and UAV networks are also characterized. The results show that a decrease in the spatial spectrum sensing radius of UAVs reduces the coverage probability of UAV communications, but it improves the ASE of UAV networks although the inter-UAV interference increases. The proposed tools allow obtaining the optimal spatial spectrum sensing radius of UAVs to maximize the ASE of UAV networks while guaranteeing the minimum ASE of D2D networks. In addition, the optimal transmit power of UAVs can be obtained to maximize the ASE of UAV networks while guaranteeing the performance of D2D communications.

Chapter 3

Energy Optimization for Air-Ground Integrated MEC Networks

3.1 Introduction

Mobile edge computing (MEC) allows mobile user equipment (UE) to offload computation tasks onto network edges such as cellular base stations (BSs), rather than computing locally, to reduce the computation latency and energy consumption of mobile device [44]. Different from traditional cloud computing, offloading computation tasks onto MEC servers in proximity can reduce the transmission delay [45]. However, the computation capacity of a MEC server is generally limited. How to efficiently allocate the limited computation resources of the MEC servers to UEs becomes a critical problem. Furthermore, due to the constrained battery capacity at the UE, new energy-saving methods are expected to be explored to prolong UE's lifetime [46]. Since the energy consumption for computing is usually higher than that for data transmission, offloading UEs' computation tasks to MEC servers can potentially reduce UEs' energy consumption.

In the meantime, unmanned aerial vehicle (UAV), owing to its desired channel conditions to ground UEs and the agility for convenient deployment [2], has been witnessed as a ground-breaking technique to provide extensive coverage, reliable line-of-site (LoS) connections and additional computation capabilities for UEs in future wireless networks [14]. Suppose that

UAVs can be equipped with MEC servers to implement ground UEs' computation tasks offloading. Compared with the ground infrastructure-based MEC, the UAV-assisted MEC support reliable LoS connections to UEs and can be deployed with flexible mobility. Therefore, UEs can transmit more data to UAVs and offload more computation tasks through the LoS channels within a certain time period, or UEs can transmit a certain amount of data with lower transmit power to UAVs, so as to save energy at UEs. In line with the obvious advantages of UAV-assisted MEC, researches were recently started and concentrated only on the UAV enabled MEC platform. However, the cooperation between the existing infrastructure and UAVs in MEC networks still remains unknown [47]. Such air-ground integrated MEC networks including ground computational access points (GCAPs) and UAVs need to be carefully designed to work cooperatively and efficiently. To be specific, the resources (including computation capacity, spectrum resources and power resources, etc) are well allocated in the air-ground integrated MEC networks. The location placement of UAVs can be managed to cooperatively offload computation tasks with GCAPs.

In this chapter, we aim to minimize the total energy consumption at UEs in air-ground integrated MEC networks which includes multiple GCAPs and multiple UAVs. An efficient resource allocation algorithm is proposed by jointly optimizing users' association, uplink power control, channel allocation, computation capacity allocation, and UAV placement. We consider the constraints of binary offloading, UEs latency, computation capacity, UAV energy consumption and available bandwidth. The primary energy consumption minimization problem is decomposed into several subproblems with block coordinate descent algorithm. To be specific, we apply Lagrangian dual method to solve the users' association and channel allocation subproblems. The uplink power control is obtained in cLoSed-form considering the constraints of UEs' latency requirements. The computation capacity allocation subproblem is decoupled into some parallel subproblems. In each parallel subproblem, forming a difference of convex (DC) functions programming, it is solved by convex-concave procedure and Karush–Kuhn–Tucker (KKT) conditions to achieve suboptimal solutions. In addition, UAV placement subproblem is proved to be a standard convex problem which can be solved efficiently. Different from existing works which only consider one UAV in MEC networks, our work extends to multiple UAVs and multiple GCAPs performing as MEC servers cooperatively towards multiple UEs while minimizing UEs' total energy consumption.

3.1.1 Related Works

Different from the ground MEC networks [48, 49, 50], UAV-assisted MEC possesses more reliable LoS connections for computation offloading, flexible mobility management and wide coverage connections [14, 47]. In [51], a mobile cloud computing system based on UAV-mounted cloudlet was studied to minimize the total mobile energy consumption, where both orthogonal and nonorthogonal access were considered. In [52], a UAV-assisted MEC network with stochastic computation tasks was investigated. The designed network only considered one UAV as MEC server to minimize the average weighted energy consumption of UEs and the UAV by designing the trajectory flying scheduling of the UAV. In [53], a one by one access scheme for UAV-aided MEC network was designed to minimize the total energy consumption of terminal devices, where only one UAV is served as MEC server. In [54], a UAV-enabled MEC network was studied where a UAV equipped with computing resources can server a number of ground users. The system in [54] aimed to minimize the sum of the maximal delay among all UEs in each time slot with joint offloading and trajectory design. However, the existing works related to the flying MEC only considered one UAV as the MEC server to provide computation and communication capabilities for UEs with reliable wireless connections [55]. Few works studied the multiple UAVs in MEC network and their cooperation with ground MEC networks. In [56], the MEC network with multiple UAVs was studied to minimize the network power consumption by optimizing the beamwidth and altitude of all UAVs. However, the objective of energy minimization was not considered, and the cooperation between UAVs and the existing ground infrastructure was not studied. In [57], a multi-UAV enabled system was established, where the deployment of UAVs and tasks scheduling were optimized to minimize the system energy consumption. However, the uplink power control and the channel allocation for ground-to-air connections were not considered. In [47, 58], the integration of air and ground for mobile edge networks was proposed to address the emerging computation-heavy Internet of Things (IoTs) applications. However, it lacks the practical algorithms design and detailed system analysis for such air-ground integrated MEC networks.

3.1.2 Contributions

We summarize the contributions of this chapter as follows:

- We consider a joint optimization problem with the aim of minimizing UEs' total energy consumption in air-ground integrated MEC networks while guaranteeing UEs' latency constraints, power constraints, computation capacity constraints and bandwidth constraint, which has not been investigated in the literature to the authors' best knowledge. Since the energy is usually limited at UEs, the considered energy minimization problem is practically appealing.
- An efficient iterative algorithm is proposed to solve the resulting problem by leveraging the block coordinate descent method. For computation capacity allocation, we first decompose the original problem into multiple small optimization problems and formulate a DC programming for each problem. Then, we utilize convex-concave procedure and KKT conditions to obtain suboptimal computation capacity allocation. The pro-

posed algorithm can be executed in a parallel manner. We also show that the proposed algorithm is guaranteed to converge.

- We investigate a cooperative computation offloading scheme in air-ground integrated MEC networks, where UAVs' three dimensional (3D) placement can be managed to cooperatively offload UEs' computation tasks with GCAPs. Intuitively, owing to the agility and flexible mobility, UAVs can be placed above the regions where there are few GCAPs and/or a large number of UEs. The height dependent path-loss channel model is considered for ground-to-UAV communications. Accordingly, the flight heights of UAVs are optimized together with UAVs horizontal coordinates.

The remainder of this chapter is organized as follows. In Chapter 3.2, the system model and problem formulation are illustrated. Chapter 3.3 studies the proposed algorithm, and after which the simulation results are given and investigated in Chapter 3.4. Finally, conclusions are drawn in Chapter 3.5.

3.2 System Model

3.2.1 Network Layout

We consider an integrated UAV and GCAP network as shown in Fig. 1, where multiple GCAPs and multiple UAVs equipped with MEC servers provide computation capabilities for ground UEs. The set of GCAPs is denoted by $\mathcal{M}^G = \left\{1, 2, \cdots, M^G\right\}$ where M^G is the number of GCAPs. The set of UAVs is given by $\mathcal{M}^A = \left\{M^G + 1, \cdots, M^G + M^A\right\}$ where M^A is the number of UAVs. K UEs are distributed on the ground and the set is denoted by $\mathcal{K} = \{1, 2, \cdots, K\}$. Here, we consider that GCAPs and UAVs are based on cellular networks

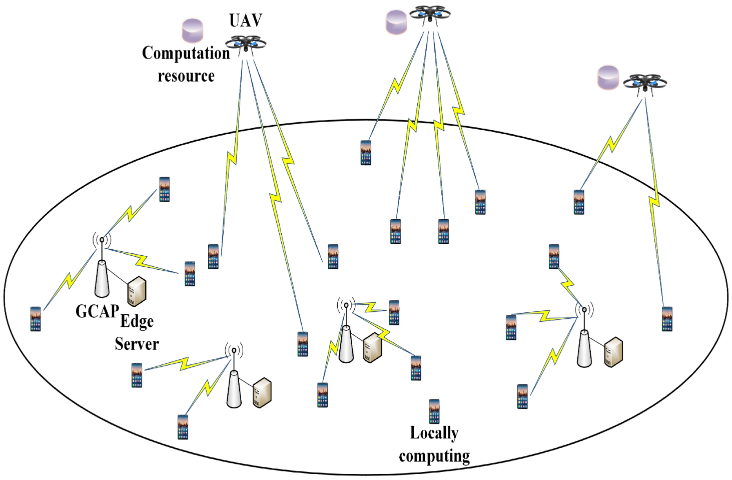


Figure 3.1: MEC in air-ground integrated networks.

in licensed band. The total bandwidth of the system is B Hz. The minimum and maximum allowable flight height of UAV are denoted by $H_{\min}$ and $H_{\max}$, respectively.

3.2.2 Offloading Protocol

In this subsection, we present users' association in a time slot. We consider the deterministic binary offloading scheme where the highly integrated or simple task has to be executed as a whole either at the mobile device or offloaded onto a GCAP or a UAV [45]. The set of possible servers is denoted by $\mathcal{M}' = \left\{0, 1, 2, \cdots, M^G, M^G + 1, \cdots, M^G + M^A\right\}$, where the first element in $\mathcal{M}'$ indicates the local server at UE itself. The association verctor of k^{th} UE is $\mathbf{a}_k = \left\{a_{k0}, \cdots, a_{k(M^G+M^A)}\right\}$ which is a $M^G + M^A + 1$ dimensional vector, where

$$a_{km} = \begin{cases} 1, \text{ if UE } k \text{ offloads onto server } m, & \forall k \in \mathcal{K}, \quad m \in \mathcal{M}' \\ 0, \text{ otherwise} \end{cases}. \tag{3.1}$$

For $a_{k0} = 1$, it represents that k^{th} UE locally processes its computation task. The system works in a time-slotted fashion. In each time slot, each UE decides to offload its computation task whether onto a UAV or a GCAP, or it processes the task locally. Therefore, we have

$$\sum_{m=0}^{M^G+M^A} a_{km} = 1, \quad \forall k \in \mathcal{K}. \tag{3.2}$$

We consider that m^{th} server has limited number of UEs associated with it, i.e.,

$$\sum_{k=1}^{K} a_{km} \leqslant U_m, \quad \forall m \in \mathcal{M}^E, \tag{3.3}$$

where $\mathcal{M}^E = \left\{\mathcal{M}^G \cup \mathcal{M}^A\right\}$ denotes the set of edge servers.

The computation task of UE k is denoted by $U_k = \left(D_k, T_k, F_k\right), \forall k \in \mathcal{K}$, where D_k is the computation offloading data size of UE k, T_k is the latency constraint of the task and F_k is the required number of central processing unit (CPU) cycles of the task.

3.2.3 Communication Model

UEs are assumed to operate in non-overlapping frequency bands such that the severe interference in LoS ground-to-air uplinks can be avoided. Denote the wireless channel bandwidth for UE k which offloads computation task onto MEC server m as B_{km} where $k \in \mathcal{K}, m \in \mathcal{M}^E$, and we have the bandwidth constraint as follows

$$\sum_{m=1}^{M^G+M^A} \sum_{k=1}^{K} a_{km} B_{km} \leqslant B. \tag{3.4}$$

Given the transmit power p_{km} of UE k offloaded onto GCAP m, the data rate of UE k is

$$R_{km}^G = B_{km} \log_2\left(1 + \frac{h_{km}^G p_{km}}{B_{km} n_0}\right), \quad \forall k \in \mathcal{K}, \quad m \in \mathcal{M}^G, \tag{3.5}$$

where n_0 represents the noise spectral density and h_{km}^G indicates the channel gain between UE k and GCAP m, and can be computed as

$$h_{km}^G = h_0^G g_{km}^G \left(L_{km}^G\right)^{-\alpha}, \quad \forall m \in \mathcal{M}^G, \tag{3.6}$$

where h_0^G is the channel power gain at reference distance, g_{km}^G accounts for possible fading, L_{km}^G is the distance between UE k and GCAP m and α is the path-loss exponent.

Denote the horizontal coordinates of UE k are (x_k, y_k) and the horizontal coordinates of UAV m are $\left(x_m^A, y_m^A\right)$. The horizontal distance between UE k and UAV m is given by

$$L_{km}^A = \sqrt{\left(x_m^A - x_k\right)^2 + \left(y_m^A - y_k\right)^2}, \quad \forall k \in \mathcal{K}, \quad m \in \mathcal{M}^A. \tag{3.7}$$

According to the height dependent path-loss channel model [59], the uplink channel gain between UE k and UAV m is

$$h_{km}^A = h_0^A \left(\sqrt{\left(z_m^A\right)^2 + \left(L_{km}^A\right)^2} \right)^{-\max\left(\alpha_1 - \alpha_2 \log_{10}\left(z_m^A\right), 2\right)}, \tag{3.8}$$

where h_0^A is the channel power gain at the reference distance, α_1 and α_2 are constant values depending on the communications environment, e.g. rural, urban, or dense urban.

If UE k decides to offload its task onto UAV m, the uplink data rate is calculated by

$$R_{km}^A = B_{km} \log_2 \left(1 + \frac{h_{km}^A p_{km}}{B_{km} n_0} \right). \tag{3.9}$$

3.2.4 Computation Model

The computation capacity of GCAP m allocated to UE k is f_{km}^G (the number of allocated CPU cycles per second), and the computation capacity constraint for GCAP m is

$$\sum_{k=1}^{K} a_{km} f_{km}^G \leqslant f_m^{G,\max}, \quad \forall m \in \mathcal{M}^G, \tag{3.10}$$

where $f_m^{G,\max}$ is the maximal computation capacity of GCAP m. In addition, the computation capacity of UAV m allocated to UE k is denoted by f_{km}^A, and the computation capacity constraint for UAV m is

$$\sum_{k=1}^{K} a_{km} f_{km}^A \leqslant f_m^{A,\max}, \quad \forall m \in \mathcal{M}^A, \tag{3.11}$$

where $f_m^{A,\max}$ is the computation capacity of UAV m.

Based on wireless transmissions, it makes sense for UEs to offload computation tasks onto proximate MEC servers as much as possible. If the wireless communications resources are sufficient, the more offloaded tasks, the less energy can be consumed at UEs and the earlier the tasks can be completed. The latency constraint of UE k is given by

$$a_{k0}\frac{F_k}{f_{k0}}+\sum_{m=1}^{M^G}a_{km}\left(\frac{D_k}{R_{km}^G}+\frac{F_k}{f_{km}^G}\right)+\sum_{m=M^G+1}^{M^G+M^A}a_{km}\left(\frac{D_k}{R_{km}^A}+\frac{F_k}{f_{km}^A}\right)\leqslant T_k,\quad \forall k\in\mathcal{K}, \tag{3.12}$$

where f_{k0} is the local computation capacity of UE k. In (3.12), the latency includes transmission latency and computation latency [45].

3.2.5 Energy Model

According to [60], the power consumption of GCAP m is

$$p_m^G=\rho_m\big(f_m^G\big)^{\varsigma},\quad \forall m\in\mathcal{M}^G, \tag{3.13}$$

where ρ_m is a constant depending on the circuit average switched capacitance, $\varsigma\geqslant 2$ is a constant representing the average activity factor, and $f_m^G=\sum\limits_{k=1}^{K}a_{km}f_{km}^G$ is the number of allocated CPU cycles at GCAP m per second. The maximal power of GCAP m is $p_m^{G,\max}$.

Similarly, the computation power consumption of UAV m is given by

$$p_m^{A,c}=\rho_m\big(f_m^A\big)^{\varsigma},\quad \forall m\in\mathcal{M}^A, \tag{3.14}$$

where $f_m^A=\sum\limits_{k=1}^{K}a_{km}f_{km}^A$ is the number of allocated CPU cycles at UAV m per second. Consider that the propulsion power consumption of UAV m is $p_m^{A,f}$. Then, the total power consumption of UAV m is given by

$$p_m^A=p_m^{A,c}+p_m^{A,f},\quad \forall m\in\mathcal{M}^A. \tag{3.15}$$

The maximal allowable power of UAV m is $p_m^{A,\max}$. If UE k executes its task locally, the power consumption of UE k is given by

$$p_k^{ue,c} = \rho_k (f_{k0})^{\varsigma}, \quad \forall k \in \mathcal{K}, \tag{3.16}$$

where ρ_k depends on the average switched capacitance.

3.2.6 Problem Formulation

We focus on the UEs' total energy consumption minimization problem, which is given by

$$\min_{\mathbf{a},\mathbf{p},\mathbf{f},\mathbf{q},\mathbf{b}} \sum_{k=1}^{K} \frac{F_k}{f_{k0}} a_{k0} p_k^{ue,c} + \sum_{k=1}^{K} \sum_{m=1}^{M^G} a_{km} p_{km} \frac{D_k}{R_{km}^G} + \sum_{k=1}^{K} \sum_{m=M^G+1}^{M^G+M^A} a_{km} p_{km} \frac{D_k}{R_{km}^A}, \tag{3.17}$$

s.t. (3.2), (3.3), (3.4), (3.10), (3.11), (3.12),

$$\rho_m \left(\sum_{k=1}^{K} a_{km} f_{km}^G \right)^{\varsigma} \leqslant p_m^{G,\max}, \quad \forall m \in \mathcal{M}^G \tag{3.17a}$$

$$\rho_m \left(\sum_{k=1}^{K} a_{km} f_{km}^A \right)^{\varsigma} + p_m^{A,f} \leqslant p_m^{A,\max}, \quad \forall m \in \mathcal{M}^A \tag{3.17b}$$

$$a_{km} = \{0, 1\}, \quad \forall k \in \mathcal{K}, m \in \mathcal{M}' \tag{3.17c}$$

$$0 \leqslant f_{km}^G, \quad \forall k \in \mathcal{K}, m \in \mathcal{M}^G \tag{3.17d}$$

$$0 \leqslant f_{km}^A, \quad \forall k \in \mathcal{K}, \forall m \in \mathcal{M}^A \tag{3.17e}$$

$$0 \leqslant p_{km} \leqslant p_k^{\max}, \quad \forall k \in \mathcal{K}, \quad m \in \mathcal{M}^E \tag{3.17f}$$

$$0 \leqslant B_{km}, \quad \forall k \in \mathcal{K}, \quad m \in \mathcal{M}^E \tag{3.17g}$$

$$H_{\min} \leqslant z_m^A \leqslant H_{\max}, \quad m \in \mathcal{M}^G, \tag{3.17h}$$

where $\mathbf{a} = \{a_{km}\}_{k\in\mathcal{K},m\in\mathcal{M}^E}$, $\mathbf{p} = \{p_{km}\}_{k\in\mathcal{K},m\in\mathcal{M}^E}$, $\mathbf{f} = \{\mathbf{f}^G, \mathbf{f}^A\}$, $\mathbf{f}^G = \{f_{km}^G\}_{k\in\mathcal{K},m\in\mathcal{M}^G}$, $\mathbf{f}^A = \{f_{km}^A\}_{k\in\mathcal{K},m\in\mathcal{M}^A}$, $\mathbf{q} = \{(x_m^A, y_m^A, z_m^A)\}_{m\in\mathcal{M}^A}$, $\mathbf{b} = \{B_{km}\}_{k\in\mathcal{K},m\in\mathcal{M}^E}$. In (3.17), the objective function contains the UEs' transmission energy and computation energy.

3.3 Proposed Algorithm

In this section, we study an efficient algorithm to solve the UEs' energy minimization problem (3.17). In the following, a coordinate descent optimization algorithm is introduced to obtain a suboptimal solution with an iterative mechanism.

3.3.1 Users' Association

By temporarily relaxing the constraints on $\mathbf{a}$, the user association problem is given by

$$\min_{\mathbf{a}} \sum_{k=1}^{K} \frac{F_k}{f_{k0}} a_{k0} p_k^{ue,c} + \sum_{k=1}^{K} \sum_{m=1}^{M^G} a_{km} p_{km} \frac{D_k}{R_{km}^G} + \sum_{k=1}^{K} \sum_{m=M^G+1}^{M^G+M^A} a_{km} p_{km} \frac{D_k}{R_{km}^A} \tag{3.18}$$

$$\text{s.t. } (3.2), (3.3), (3.10), (3.11), (3.12), (3.17a), (3.17b)$$

$$0 \leqslant a_{km} \leqslant 1, \forall k \in \mathcal{K}, m \in \mathcal{M}'. \tag{3.18a}$$

We use Lagrangian dual decomposition method to obtain the integer solutions of $\mathbf{a}$. The partial Lagrangian function of the users' association subproblem is given in (3.19),

$$\begin{aligned}
&\mathcal{L}\left(a_{km}, \beta_m, \gamma_m^G, \gamma_m^A, \mu_m, \lambda_m, \kappa_k\right) \\
&= \sum_{k=1}^{K} \frac{F_k}{f_{k0}} a_{k0} p_k^{ue,c} + \sum_{k=1}^{K} \sum_{m=1}^{M^G} a_{km} p_{km} \frac{D_k}{R_{km}^G} + \sum_{k=1}^{K} \sum_{m=M^G+1}^{M^G+M^A} a_{km} p_{km} \frac{D_k}{R_{km}^A} \\
&+ \sum_{m=1}^{M^G+M^A} \beta_m \left(\sum_{k=1}^{K} a_{km} - U_m\right) + \sum_{m=1}^{M^G} \gamma_m^G \left(\sum_{k=1}^{K} a_{km} f_{km}^G - f_m^{G,\max}\right) \\
&+ \sum_{m=M^G+1}^{M^G+M^A} \gamma_m^A \left(\sum_{k=1}^{K} a_{km} f_{km}^A - f_m^{A,\max}\right) + \sum_{m=1}^{M^G} \mu_m \left(\sum_{k=1}^{K} a_{km} f_{km}^G - \left(\frac{p_m^{G,\max}}{\rho_m}\right)^{\frac{1}{\varsigma}}\right) \\
&+ \sum_{m=M^G+1}^{M^G+M^A} \lambda_m \left(\sum_{k=1}^{K} a_{km} f_{km}^A - Q_m^A\right) + \sum_{k=1}^{K} \kappa_k \left(\sum_{m=1}^{M^G} a_{km} T_m^G + \sum_{m=M^G+1}^{M^G+M^A} a_{km} T_m^A + a_{k0} \frac{F_k}{f_{k0}} - T_k\right),
\end{aligned} \tag{3.19}$$

where $Q_m^A = {\rho_m}^{-\frac{1}{\varsigma}} \left(p_m^{A,\max} - p_m^{A,f}\right)^{\frac{1}{\varsigma}}$, $T_m^G = \frac{D_k}{R_{km}^G} + \frac{F_k}{f_{km}^G}$, $T_m^A = \frac{D_k}{R_{km}^A} + \frac{F_k}{f_{km}^A}$, $\{\beta_m\}_{m\in\mathcal{M}'}$, $\left\{\gamma_m^G\right\}_{m\in\mathcal{M}'}$, $\left\{\gamma_m^A\right\}_{m\in\mathcal{M}'}$, $\{\mu_m\}_{m\in\mathcal{M}'}$, $\{\lambda_m\}_{m\in\mathcal{M}'}$ and $\{\kappa_k\}_{k\in\mathcal{K}}$ are Lagrange multipliers associated with the constraints in problem (3.18).

The Lagrange dual function of the users' association subproblem is given by

$$g\left(\Omega\right) = \begin{cases} \min\limits_{\mathbf{a}} \mathcal{L}\left(a_{km}, \Omega\right) \\ \text{s.t.} \ \sum\limits_{m=0}^{M+1} a_{km} = 1, \forall k \in \mathcal{K} \\ \qquad 0 \leqslant a_{km} \leqslant 1, \forall k \in \mathcal{K}, m \in \mathcal{M}' \end{cases}, \tag{3.20}$$

where Ω is the set of Lagrange multipliers.

We now separate the users' association for GCAPs, UAVs and no association to MEC server. The following function $\mathcal{F}_{km}$ is defined to obtain the optimal integer users' association, shown as follows

$$\mathcal{F}_{km} = \begin{cases} p_{km}\dfrac{D_k}{R_{km}^G} + \beta_m + \gamma_m^G f_{km}^G + \mu_m f_{km}^G + \kappa_k T_m^G, \quad \forall k \in \mathcal{K}, \quad m \in \mathcal{M}^G \\ p_{km}\dfrac{D_k}{R_{km}^A} + \beta_m + \gamma_m^A f_{km}^A + \lambda_m f_{km}^A + \kappa_k T_m^A, \quad \forall k \in \mathcal{K}, \quad m \in \mathcal{M}^A \\ \dfrac{F_k}{f_{k0}} p_k^{ue,c} + \kappa_k \dfrac{F_k}{f_{k0}}, \quad \forall k \in \mathcal{K}, \quad m = 0 \end{cases}. \tag{3.21}$$

Thus, the users' association $\mathbf{a}$ can be expressed as

$$a_{km}^* = \begin{cases} 1, \ \text{if } m = \arg\min\limits_{m\in\mathcal{M}'} \mathcal{F}_{km} \\ 0, \ \text{otherwise} \end{cases}, \tag{3.22}$$

where the results have multiple linear calculations with low complexity.

After obtaining the optimal users' association $\mathbf{a}^* = \{a_{km}^*\}_{k\in\mathcal{K}, m\in\mathcal{M}'}$, we update the dual variables in each iteration. The dual problem of (3.18) is given by

$$\max_{\Omega} g\left(\Omega\right), \ \text{ s.t. } \Omega \geqslant 0. \tag{3.23}$$

Based on the sub-gradient of dual variables, the updating procedure is given as follows

$$\beta_m^{t+1} = \left[\beta_m^t + \varphi_1(t)\left(\sum_{k=1}^{K} a_{km} - U_m\right)\right]^+, \tag{3.24}$$

$$\gamma_m^{G,t+1} = \left[\gamma_m^{G,t} + \varphi_2(t)\left(\sum_{k=1}^{K} a_{km} f_{km}^G - f_m^{G,\max}\right)\right]^+, \tag{3.25}$$

$$\gamma_m^{A,t+1} = \left[\gamma_m^{A,t} + \varphi_3(t)\left(\sum_{k=1}^{K} a_{km} f_{km}^A - f_m^{A,\max}\right)\right]^+, \tag{3.26}$$

$$\mu_m^{t+1} = \left[\mu_m^t + \varphi_4(t)\left(\sum_{k=1}^{K} a_{km} f_{km}^G - \left(\frac{p_m^{G,\max}}{\rho_m}\right)^{\frac{1}{\varsigma}}\right)\right]^+, \tag{3.27}$$

$$\lambda_m^{t+1} = \left[\lambda_m^t + \varphi_5(t)\left(\sum_{k=1}^{K} a_{km} f_{km}^A - Q_m^A\right)\right]^+, \tag{3.28}$$

$$\kappa_k^{t+1} = \left[\kappa_k^t + \varphi_6(t)\left(\sum_{m=1}^{M^G} a_{km} T_m^G + \sum_{m=M^G+1}^{M^G+M^A} a_{km} T_m^A + a_{k0}\frac{F_k}{f_{k0}} - T_k\right)\right]^+, \tag{3.29}$$

where $[x]^+ = \max\{x, 0\}$ and $\varphi_i(t), i = \{1, \cdots, 6\}$ are the dynamically chosen step-size sequence.

3.3.2 Uplink Power Control

In this subsection, we obtain the closed-form power controls for UEs's data transmissions.

Proposition 3.1. *The energy consumption of UE k associated with GCAP m is non-decreasing with respect to the uplink transmit power p_{km}.*

Proof. Please refer to Appendix B.1 for the proof. □

Since the energy consumption of a UE increases with the UE's transmit power, the optimal transmit power of UEs can be obtained by satisfying the latency constraints.

Theorem 3.2. *The optimal transmit power of UE k to GCAP m is given by*

$$p^*_{km} = \min\left\{\xi^G_{km}\frac{B_{km}n_0}{h^G_{km}}, p^{\max}_k\right\}, \tag{3.30}$$

where $\xi^G_{km} = 2^{\frac{D_k}{B_{km}\left(T_k - \frac{F_k}{f^G_{km}}\right)}} - 1$. *The optimal transmit power of UE k to UAV m is given by*

$$p^*_{km} = \min\left\{\frac{\xi^A_{km}B_{km}n_0}{h^A_0\left((z^A_m)^2 + (L^A_{km})^2\right)^{-\frac{\alpha_{GA}\left(z^A_m\right)}{2}}}, p^{\max}_k\right\}, \tag{3.31}$$

where $\xi^A_{km} = 2^{\frac{D_k}{B_{km}\left(T_k - \frac{F_k}{f^A_{km}}\right)}} - 1$.

Proof. Please refer to Appendix B.2 for the proof. □

3.3.3 Channel Allocation

In this section, we solve the channel allocation subproblem. Owing to the nature of LoS connections between air and ground nodes, UAVs may suffer severe co-channel interference from other ground UEs, and the UEs connected to GCAPs can be interfered by UAVs in downlink computation results collection. In this chapter, we allocate channels to UEs to make the communications links orthogonal to avoid the severe interference in such air-ground integrated MEC networks.

The channel allocation subproblem is given by

$$\min_{\mathbf{b}} \sum_{m=1}^{M^G+M^A}\sum_{k=1}^{K} a^*_{km}p^*_{km}\frac{D_k}{R_{km}} \tag{3.32}$$

$$\text{s.t. } (3.4), (3.17g)$$

$$\sum_{m=1}^{M^G+M^A} a^*_{km}\frac{D_k}{R_{km}} + C^B_k \leqslant 0, \forall k \in \mathcal{K}, \tag{3.32a}$$

where $C_k^B = \sum_{m=1}^{M^G} \frac{a_{km}^* F_k}{f_{km}^G} + \sum_{m=M^G+1}^{M^G+M^A} \frac{a_{km}^* F_k}{f_{km}^A} + a_{k0}^* \frac{F_k}{f_{k0}} - T_k$, $R_{km} = R_{km}^G$ if $m \in \mathcal{M}^G$ and $R_{km} = R_{km}^A$ if $m \in \mathcal{M}^A$. The total bandwidth should be fully utilized to provide reliable data rate in order to minimize the UEs' total energy consumption.

Proposition 3.3. *Problem (3.32) is a convex problem.*

Proof. Please refer to Appendix B.3 for the proof. □

It's worth noting the convex problem (3.32) can be solved by the optimization tools such as CVX [61]. In this chapter, to achieve low complexity of the algorithm, we leverage Lagrangian dual method to solve the problem (3.32) efficiently by analytically exploring the form of solution. The Lagarange function of (3.32) is in (3.33),

$$\begin{aligned}\mathcal{L}(\mathbf{b}, \mathbf{v}, \lambda_b) = &\sum_{m=1}^{M^G+M^A} \sum_{k=1}^{K} a_{km}^* p_{km}^* \frac{D_k}{R_{km}} + \sum_{k=1}^{K} v_k \left[\sum_{m=1}^{M^G+M^A} a_{km}^* \frac{D_k}{R_{km}} + C_k^B \right] \\ &+ \lambda_b \left(\sum_{m=1}^{M^G+M^A} \sum_{k=1}^{K} a_{km}^* B_{km} - B \right) - \sum_{m=1}^{M^G+M^A} \sum_{k=1}^{K} a_{km}^* \omega_{km} B_{km},\end{aligned} \tag{3.33}$$

where $\mathbf{b} = \{B_{km}\}_{k \in \mathcal{K}, m \in \mathcal{M}^E}$, $\mathbf{v} = \{v_k\}_{k \in \mathcal{K}}$, $\mathbf{v}$ and λ_b are dual variables, C_k^B is given in (3.32a). It can be observed that $\mathcal{L}(\mathbf{b}, \upsilon, \lambda_b)$ is convex with respect to B_{km}.

The dual problem of (3.32) is expressed by

$$\begin{aligned}&\max_{\mathbf{v}, \lambda_b} \min_{\mathbf{b}} \mathcal{L}(\mathbf{b}, \mathbf{v}, \lambda_b) \\ &\text{s.t. } \mathbf{v} \geqslant 0, \lambda_b \geqslant 0.\end{aligned} \tag{3.34}$$

In the following theorem, we obtain the channel allocation strategy for each UE.

Theorem 3.4. *Given the dual variables* $\mathbf{v}$ *and* λ_b*, the channel allocation for UE* k *connected* *to MEC server m is given by*

$$B_{km}^{*} = \max\left\{\psi^{-1}\left(\frac{\lambda_b-\omega_{km}}{\left(p_{km}^{*}+v_k\right)D_k\ln 2}\right),0\right\},$$
$$\textit{where}\quad \psi\left(B_{km}\right)=\frac{\ln\left(1+\frac{h_{km}p_{km}^{*}}{B_{km}n_0}\right)-\frac{h_{km}p_{km}^{*}}{h_{km}p_{km}^{*}+B_{km}n_0}}{{B_{km}}^{2}\left[\ln\left(1+\frac{h_{km}p_{km}^{*}}{B_{km}n_0}\right)\right]^{2}}, \tag{3.35}$$

and $\psi^{-1}(\cdot)$ is the inverse function of $\psi(\cdot)$.

Proof. Please refer to Appendix B.4 for the proof. □

Proposition 3.5. *Function $\psi\left(B_{km}\right)$ decreases with B_{km}, and we can use bisection method to obtain the optimal B_{km}^{*}.*

Proof. Please refer to Appendix B.5 for the proof. □

In addition, the dual variables can be updated with subgradient method in each iteration as follows

$$v_k^{t+1}=\left[v_k^t+\phi_v(t)\left(\sum_{m=1}^{M^G+M^A}a_{km}^{*}\frac{D_k}{R_{km}}+C_k^B\right)\right]^{+}, \tag{3.36}$$

$$\lambda_b^{t+1}=\left[\lambda_b^t+\phi_b(t)\left(\sum_{m=1}^{M^G+M^A}\sum_{k=1}^{K}a_{km}^{*}B_{km}-B\right)\right]^{+}, \tag{3.37}$$

$$\omega_{km}^{t+1}=\left[\omega_{km}^t-\phi_\omega(t)\,B_{km}\right]^{+}, \tag{3.38}$$

where $\phi_k(t)$, $\phi_b(t)$ and $\phi_\omega(t)$ are the dynamically chosen step-size sequences in the iteration.

3.3.4 Computation Capacity Allocation

In this section, we optimize the computation capacity allocation for UEs given the optimized users' association, power control, bandwidth allocation. Define the set $\mathcal{K}_m =$

$\{k \in \mathcal{K} | \, a_{km} = 1\}$ as the set of UEs associated with MEC server m based on the results in Section III-A. The problem of computation capacity allocation can be reformulated as

$$\min_{\mathbf{f}^G,\mathbf{f}^A} \sum_{m=1}^{M^G} \sum_{k\in\mathcal{K}_m} \frac{p^*_{km} D_k}{R^{G}_{km}{}^{*}} + \sum_{m=M^G+1}^{M^G+M^A} \sum_{k\in\mathcal{K}_m} \frac{p^*_{km} D_k}{R^{A}_{km}{}^{*}} \tag{3.39}$$

$$\text{s.t. } (3.12), (3.17d), (3.17e),$$

$$\sum_{k\in\mathcal{K}_m} f^G_{km} \leqslant \min\left\{f^{G,\max}_m, f^{G,P,\max}_m\right\}, \forall m \in \mathcal{M}^G \tag{3.39a}$$

$$\sum_{k\in\mathcal{K}_m} f^A_{km} \leqslant \min\left\{f^{A,\max}_m, f^{A,P,\max}_m\right\}, \forall m \in \mathcal{M}^A, \tag{3.39b}$$

where $f^{G,P,\max}_m = \left(\frac{p^{G,\max}_m}{\rho_m}\right)^{\frac{1}{\varsigma}}$, $f^{A,P,\max}_m = \left(\frac{p^{A,\max}_m - p^{A,f}_m}{\rho_m}\right)^{\frac{1}{\varsigma}}$, and we integrate the MEC server's maximal computation capacity by considering its maximal power consumption.

By substituting the uplink transmit power in (3.39) with the optimal power consumption obtained in Section III-B and considering the latency constraint of each UE, we can further reformulate the computation capacity allocation problem as follows

$$\begin{aligned}
\min_{\mathbf{f}^G,\mathbf{f}^A} & \sum_{m=1}^{M^G} \sum_{k\in\mathcal{K}_m} \left(2^{\frac{D_k}{B^*_{km}\left(T_k - \frac{F_k}{f^G_{km}}\right)}} - 1\right) \frac{B^*_{km} n_0}{h^G_{km}} \left(T_k - \frac{F_k}{f^G_{km}}\right) \\
& + \sum_{m=M^G+1}^{M^G+M^A} \sum_{k\in\mathcal{K}_m} \left(2^{\frac{D_k}{B^*_{km}\left(T_k - \frac{F_k}{f^A_{km}}\right)}} - 1\right) \frac{B^*_{km} n_0 \left(T_k - \frac{F_k}{f^A_{km}}\right)}{h^A_0 \left(\left(z^A_m\right)^2 + \left(L^A_{km}\right)^2\right)^{-\frac{\alpha_{GA}\left(z^A_m\right)}{2}}}
\end{aligned} \tag{3.40}$$

$$\text{s.t. } \sum_{k\in\mathcal{K}_m} f^G_{km} \leqslant f^{G,prac}_m, \forall m \in \mathcal{M}^G \tag{3.40a}$$

$$\sum_{k\in\mathcal{K}_m} f^A_{km} \leqslant f^{A,prac}_m, \forall m \in \mathcal{M}^A \tag{3.40b}$$

$$f^{G,\min}_{km} \leqslant f^G_{km}, \forall k \in \mathcal{K}_m, m \in \mathcal{M}^G \tag{3.40c}$$

$$f^{A,\min}_{km} \leqslant f^A_{km}, \forall k \in \mathcal{K}_m, m \in \mathcal{M}^A \tag{3.40d}$$

where

$$
\begin{aligned}
f_m^{G,prac} &= \min\left\{ f_m^{G,\max}, \left(\frac{p_m^{G,\max}}{\rho_m}\right)^{\frac{1}{\varsigma}} \right\}, \\
f_m^{A,prac} &= \min\left\{ f^{A,\max}, \left(\frac{p_m^{A,\max} - p_m^{A,f}}{\rho_m}\right)^{\frac{1}{\varsigma}} \right\}, \\
f_{km}^{G,\min} &= \max\left\{ F_k \left[T_k - \frac{D_k}{R_{km}^{*G,\max}} \right]^{-1}, 0 \right\}, \\
f_{km}^{A,\min} &= \max\left\{ F_k \left[T_k - \frac{D_k}{R_{km}^{*A,\max}} \right]^{-1}, 0 \right\} \\
R_{km}^{*G,\max} &= B_{km}^{*} \log_2 \left(1 + \frac{h_{km}^{G} p_k^{\max}}{B_{km}^{*} n_0} \right), \\
R_{km}^{*A,\max} &= B_{km}^{*} \log_2 \left(1 + \frac{h_{km}^{A} p_k^{\max}}{B_{km}^{*} n_0} \right).
\end{aligned}
$$

B_{km}^{*} is the optimal bandwidth of UE k associated with MEC server m obtained in Section III-C.

Given the optimized users' association, power control, bandwidth allocation, the computation capacity allocation strategies are uncorrelated among different MEC servers. Therefore, problem (3.40) can be decomposed into $M^G + M^A$ parallel subproblems. For GCAP m, the computation capacity allocation problem is given by

$$
\min_{\mathbf{f}_m^G} \sum_{k \in \mathcal{K}_m} \left(2^{\frac{D_k}{B_{km}^{*}\left(T_k - \frac{F_k}{f_{km}^G}\right)}} - 1 \right) \frac{B_{km}^{*} n_0}{h_{km}^{G}} \left(T_k - \frac{F_k}{f_{km}^{G}} \right), \tag{3.41}
$$

$$
\text{s.t.} \quad \sum_{k \in \mathcal{K}_m} f_{km}^{G} \leqslant f_m^{G,prac}, \tag{3.41a}
$$

$$
f_{km}^{G,\min} \leqslant f_{km}^{G}, \forall k \in \mathcal{K}_m. \tag{3.41b}
$$

In addition, for UAV m, the computation capacity allocation problem is given by

$$\begin{aligned}\min_{\mathbf{f}_m^A} \sum_{k\in\mathcal{K}_m} & \left(2^{\frac{D_k}{B_{km}^*\left(T_k-\frac{F_k}{f_{km}^A}\right)}}-1\right)\left(T_k-\frac{F_k}{f_{km}^A}\right) \\ & \cdot \frac{B_{km}^* n_0 \left(\left(\left(z_m^A\right)^2+\left(L_{km}^A\right)^2\right)^{\frac{\alpha_{GA}\left(z_m^A\right)}{2}}\right)}{h_0^A},\end{aligned} \tag{3.42}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}_m} f_{km}^A \leqslant f_m^{A,prac}, \tag{3.42a}$$

$$f_{km}^{A,\min} \leqslant f_{km}^A, \forall k \in \mathcal{K}_m. \tag{3.42b}$$

Since (3.41) and (3.42) have the same form with respect to the computation capacity allocation (i.e., f_{km}^G and f_{km}^A), we will focus on the problem (3.41) in the sequel.

The objective function of (3.41) can be transformed as follows

$$\begin{aligned} & \min_{\mathbf{f}_m^G} \sum_{k\in\mathcal{K}_m} \left(2^{\frac{D_k}{B_{km}^*\left(T_k-\frac{F_k}{f_{km}^G}\right)}}-1\right)\frac{B_{km}^* n_0}{h_{km}^G}\left(T_k-\frac{F_k}{f_{km}^G}\right) \\ & = \min_{\mathbf{f}_m^G} \sum_{k\in\mathcal{K}_m} F\left(f_{km}^G\right) - G\left(f_{km}^G\right), \\ & \text{where} \\ & F\left(f_{km}^G\right) = C_2 2^{\frac{F_k D_k}{B_{km}^* T_k\left(T_k f_{km}^G - F_k\right)}} - \frac{B_{km}^* n_0}{h_{km}^G} T_k + \frac{B_{km}^* n_0 F_k}{h_{km}^G}\frac{1}{f_{km}^G}, \\ & G\left(f_{km}^G\right) = \frac{C_1 B_{km}^* n_0 F_k}{h_{km}^G f_{km}^G} 2^{\frac{F_k D_k}{B_{km}^* T_k\left(T_k f_{km}^G - F_k\right)}}, \end{aligned} \tag{3.43}$$

and $\mathbf{f}_m^G = \left\{f_{1m}^G, f_{2m}^G, \cdots, f_{K_m m}^G\right\}$, $C_1 = 2^{\frac{D_k}{B_{km}^* T_k}}$, $C_2 = 2^{\frac{D_k}{B_{km}^* T_k}} \frac{B_{km}^* n_0}{h_{km}^G} T_k$, and functions $F\left(f_{km}^G\right)$ and $G\left(f_{km}^G\right)$ are expressed in (3.43).

Proposition 3.6. *Problem (3.41) is a difference of convex (DC) functions programming.*

Proof. Define the function $h\left(x\right) = e^{\frac{1}{x}}$ and $g\left(x\right) = e^{\frac{1}{x}}\frac{1}{x}$. We have $h''\left(x\right) > 0$ and $g''\left(x\right) = e^{\frac{1}{x}}\left(\frac{1}{x^3} + \frac{4}{x^4} + \frac{1}{x^5}\right) > 0$. Therefore, functions $F\left(f^G_{km}\right)$ and $G\left(f^G_{km}\right)$ are convex with respect to f^G_{km}. In addition, the constraints in (3.41) are also convex. Since a summation of DC functions is still convex, problem (3.41) is a DC programming. □

In this chapter, the convex-concave procedure is used to solve the above DC programming [62]. In DC programming literature [63], the local optimization approach often yields the global optimum, and the starting-point choosing and regularization methods can also assist the local optimization approach to find global optimum. Since $G\left(f^G_{km}\right)$ is a convex function of f^G_{km}, $G\left(f^G_{km}\right)$ is lower bounded by its first order Taylor expansion [64] as follows

$$\begin{aligned}
&G\left(f^G_{km}\right) \geqslant G\left(f^G_{km}\left(t\right)\right) + G'\left(f^G_{km}\left(t\right)\right)\left(f^G_{km} - f^G_{km}\left(t\right)\right)\\
&\text{where}\\
&G'\left(f^G_{km}\left(t\right)\right)\\
&= \frac{C_1 B^*_{km} n_0 F_k}{h^G_{km} f^G_{km}\left(t\right)} 2^{\frac{F_k D_k}{B^*_{km} T_k\left(T_k f^G_{km}\left(t\right) - F_k\right)}}\left[-\frac{\ln 2 F_k D_k}{B^*_{km}\left(T_k f^G_{km}\left(t\right) - F_k\right)^2} - \frac{1}{f^G_{km}\left(t\right)}\right],
\end{aligned} \tag{3.44}$$

where $f^G_{km}\left(t\right)$ is the last optimal allocated computation capacity for UE k associated with GCAP m. Therefore, the objective function in (3.43) can be approximated $\mathcal{O}_F$ where

$$\begin{aligned}
&\mathcal{O}_F\\
&= \sum_{k\in\mathcal{K}_m} F\left(f^G_{km}\right) - \left[G\left(f^G_{km}\left(t\right)\right) + G'\left(f^G_{km}\left(t\right)\right)\left(f^G_{km} - f^G_{km}\left(t\right)\right)\right]\\
&= \sum_{k\in\mathcal{K}_m} C_2 2^{\frac{F_k D_k}{B^*_{km} T_k\left(T_k f^G_{km} - F_k\right)}} - \frac{B^*_{km} n_0}{h^G_{km}} T_k + \frac{B^*_{km} n_0 F_k}{h^G_{km}}\\
&\quad \cdot \frac{1}{f^G_{km}} - \frac{C_1 B^*_{km} n_0 F_k}{h^G_{km} f^G_{km}\left(t\right)} 2^{\frac{F_k D_k}{B^*_{km} T_k\left(T_k f^G_{km}\left(t\right) - F_k\right)}}\\
&\quad + \frac{C_1 B^*_{km} n_0 F_k}{h^G_{km} f^G_{km}\left(t\right)} 2^{\frac{F_k D_k}{B^*_{km} T_k\left(T_k f^G_{km}\left(t\right) - F_k\right)}}\left(f^G_{km} - f^G_{km}\left(t\right)\right)\\
&\quad \cdot \left[\frac{\ln 2 F_k D_k}{B^*_{km}\left(T_k f^G_{km}\left(t\right) - F_k\right)^2} + \frac{1}{f^G_{km}\left(t\right)}\right].
\end{aligned} \tag{3.45}$$

Problem (3.41) can be approximately reformulated by

$$\begin{aligned}
\min_{\mathbf{f}_m^G} \sum_{k\in\mathcal{K}_m} & C_2 2^{\frac{F_k D_k}{B_{km}^* T_k\left(T_k f_{km}^G - F_k\right)}} + \frac{B_{km}^* n_0}{h_{km}^G}\left(\frac{F_k}{f_{km}^G} - T_k\right) \\
& - \omega\left(f_{km}^G(t)\right) + \tau\left(f_{km}^G(t)\right)\left(f_{km}^G - f_{km}^G(t)\right) \qquad (3.46)
\end{aligned}$$

s.t. (49a), (49b)

where

$$\omega\left(f_{km}^G(t)\right) = \frac{C_1 B_{km}^* n_0 F_k}{h_{km}^G f_{km}^G(t)} 2^{\frac{F_k D_k}{B_{km}^* T_k\left(T_k f_{km}^G(t) - F_k\right)}}, \qquad (3.46a)$$

$$\tau\left(f_{km}^G(t)\right) = \omega\left(f_{km}^G(t)\right)\left[\frac{\ln 2 F_k D_k}{B_{km}^*\left(T_k f_{km}^G(t) - F_k\right)^2} + \frac{1}{f_{km}^G(t)}\right]. \qquad (3.46b)$$

It can be observed that problem (3.46) is a convex problem due to the fact that the sum of multiple convex functions is still a convex function. An iterative algorithm can be developed to update the variable $f_{km}^G(t)$ with a stopping criterion $\mathcal{O}_F\left(f_{km}^G(t)\right) - \mathcal{O}_F\left(f_{km}^G(t+1)\right) \leqslant \varepsilon_F$, where ε_F is set to be a small value. In each iteration, the optimal solution is obtained in the following theorem.

Theorem 3.7. *Given users' association, uplink optimal power control and optimal channel allocation, the computation capacity allocation for UE k associated with GCAP m is obtained as follows.*

(1) When $\sum_{k\in\mathcal{K}_m} \max\left({\Psi_{km}^G}^{-1}\left(\tau\left(f_{km}^G(t)\right)\right), f_{km}^{G,\min}\right) > f_m^{G,prac}$, *we have*

$$f_{km}^G = \max\left({\Psi_{km}^G}^{-1}\left(\lambda_f + \tau\left(f_{km}^G(t)\right)\right), f_{km}^{G,\min}\right), k \in \mathcal{K}_m, \qquad (3.47)$$

where λ_f *is calculated by*

$$\sum_{k\in\mathcal{K}_m} \max\left({\Psi_{km}^G}^{-1}\left(\lambda_f + \tau\left(f_{km}^G(t)\right)\right), f_{km}^{G,\min}\right) = f_m^{G,prac}, \qquad (3.48)$$

and

$$\Psi_{km}^{G}\left(y\right)=2^{\frac{F_kD_k}{B_{km}^{*}T_k\left(T_ky-F_k\right)}}\frac{C_2\ln 2F_kD_k}{B_{km}^{*}(T_ky-F_k)^2}+\frac{B_{km}^{*}n_0F_k}{h_{km}^{G}(y)^2}. \tag{3.49}$$

(2) When $\sum\limits_{k\in\mathcal{K}_m}\max\left({\Psi_{km}^{G}}^{-1}\left(\tau\left(f_{km}^{G}\left(t\right)\right)\right),f_{km}^{G,\min}\right)\leqslant f_{m}^{G,prac}$*, we have*

$$f_{km}^{G}=\max\left({\Psi_{km}^{G}}^{-1}\left(\tau\left(f_{km}^{G}\left(t\right)\right)\right),f_{km}^{G,\min}\right),k\in\mathcal{K}_m, \tag{3.50}$$

where $f_{km}^{G,\min}$ *is given in problem (3.40).*

Proof. Please refer to Appendix B.6 for the proof. □

For UAV m, the computation capacity allocation problem in (3.42) can be solved by the convex-concave procedure which is similar to that of problem (3.41) and it is omitted here to save space.

3.3.5 UAV Placement

UAVs' placement can be managed to cooperatively offload UEs' computation tasks with GCAPs [47]. According to the height dependent path-loss channel model, increasing UAV's flight height decreases the path-loss exponent, however, it also increases the communicating distance. Therefore, the 3D UAVs' positions need to be optimized. The UAV placement problem is presented in the following

$$\begin{aligned}\min_{\mathbf{q}}\sum_{m=M^G+1}^{M^G+M^A}\sum_{k\in\mathcal{K}_m}&\left(2^{\frac{D_k}{B_{km}^{*}\left(T_k-\frac{F_k}{f_{km}^{A*}}\right)}}-1\right)\frac{B_{km}^{*}n_0}{h_0^{A}}\\&\cdot\left[\left(z_m^{A}\right)^2+\left(L_{km}^{A}\right)^2\right]^{\frac{\max\left(\alpha_1-\alpha_2\log_{10}\left(z_m^{A}\right),2\right)}{2}}\left(T_k-\frac{F_k}{f_{km}^{A}}\right)\end{aligned} \tag{3.51}$$

$$\text{s.t. } H_{\min}\leqslant z_m^{A}\leqslant H_{\max},\forall m\in\mathcal{M}^{A}, \tag{3.51a}$$

where $H_{\min}$ and $H_{\max}$ denotes the minimum and maximum flight height of UAV, respectively.

The problem (3.51) can be decomposed into M^A subproblems and the subproblem for UAV m is given by

$$\min_{x_m^A, y_m^A, z_m^A} \sum_{k \in \mathcal{K}_m} C_{km}^q \left[\left(z_m^A\right)^2 + \left(x_m^A - x_k\right)^2 + \left(y_m^A - y_k\right)^2 \right]^{\frac{\max\left(\alpha_1 - \alpha_2 \log_{10}\left(z_m^A\right), 2\right)}{2}} \tag{3.52}$$

$$\text{s.t.}\ \ H_{\min} \leqslant z_m^A \leqslant H_{\max}, \tag{3.52a}$$

$$\text{where} \tag{3.52b}$$

$$C_{km}^q = \left(2^{\frac{D_k}{B_{km}^* \left(T_k - \frac{F_k}{f_{km}^{A*}}\right)}} - 1 \right) \frac{B_{km}^* n_0}{h_0^A} \left(T_k - \frac{F_k}{f_{km}^{A*}}\right). \tag{3.52c}$$

It can be verified that in (3.52), given horizontal coordinates $\left(x_m^A, y_m^A\right)$, the objective function is neither convex nor concave with respect to z_m^A. The problem (3.52) is a non-convex problem.

To solve the problem (3.52), we discrete the feasible set of UAV's flight height $[H_{\min}, H_{\max}]$ into $\{H_1, H_2, \cdots, H_{N_H}\}$, and we optimize the variables $\left(x_m^A, y_m^A\right)$ given a z_m^A, as follows

$$\min_{x_m^A, y_m^A} \sum_{k \in \mathcal{K}_m} C_{km}^q \left[\left(z_m^A\right)^2 + \left(x_m^A - x_k\right)^2 + \left(y_m^A - y_k\right)^2 \right]^{\frac{\max\left(\alpha_1 - \alpha_2 \log_{10}\left(z_m^A\right), 2\right)}{2}}. \tag{3.53}$$

Then, we repeat the above procedure over all discrete values of UAV's flight height. Finally, we select the optimal z_m^A with the minimum objective value in (3.53).

Note that the problem (3.53) is a convex problem and it can be solved by the optimization tools such as CVX [61]. To solve the problem (3.53) more efficiently, we use the bisection method as follows. According to the stationarity of KKT conditions of problem (3.53), we have

$$\begin{aligned}\frac{\partial \mathcal{O}_Q^m}{\partial x_m^A} = \sum_{k\in\mathcal{K}_m} & 2C_{km}^q \alpha_m^{AG}\left(x_m^A - x_k\right)\Big[\left(z_m^A\right)^2 \\ & +\left(x_m^A - x_k\right)^2 + \left(y_m^A - y_k\right)^2\Big]^{\alpha_m^{AG}-1} = 0,\end{aligned} \tag{3.54}$$

and

$$\begin{aligned}\frac{\partial \mathcal{O}_Q^m}{\partial y_m^A} = \sum_{k\in\mathcal{K}_m} & 2C_{km}^q \alpha_m^{AG}\left(y_m^A - y_k\right)\Big[\left(z_m^A\right)^2 \\ & +\left(x_m^A - x_k\right)^2 + \left(y_m^A - y_k\right)^2\Big]^{\alpha_m^{AG}-1} = 0,\end{aligned} \tag{3.55}$$

where $\mathcal{O}_Q^m$ is the objective function of problem (3.53) and $\alpha_m^{AG} = \frac{1}{2}\max\left(\alpha_1 - \alpha_2 \log_{10}\left(z_m^A\right), 2\right)$. By using the bisection method, we obtain the optimal values of x_m^A and y_m^A, respectively. In addition, due to the limited number of the elements in $\{H_1, H_2, \cdots, H_{N_H}\}$, the proposed approach is tractable and has low complexity.

3.3.6 Iterative Algorithm and convergence

The iterative algorithm for solving the problem (3.17) is given in Algorithm 1, where E_{UEs}^t represents the total energy consumption of UEs in t^{th} iteration.

Denote $\mathcal{O}(\mathbf{a}, \mathbf{b}, \mathbf{f}, \mathbf{q}, \mathbf{p})$ as the objective value of original problem (3.17). In step 3 of Algorithm 1, since $\mathbf{a}^{(t)}$ is one suboptimal users' association of problem (3.18) with the fixed $\mathbf{b}^{(t-1)}$, $\mathbf{f}^{(t-1)}$, $\mathbf{q}^{(t-1)}$ and $\mathbf{p}^{(t-1)}$, we have

$$\begin{aligned}&\mathcal{O}\left(\mathbf{a}^{(t-1)}, \mathbf{b}^{(t-1)}, \mathbf{f}^{(t-1)}, \mathbf{q}^{(t-1)}, \mathbf{p}^{(t-1)}\right) \\ &\geqslant \mathcal{O}\left(\mathbf{a}^{t}, \mathbf{b}^{(t-1)}, \mathbf{f}^{(t-1)}, \mathbf{q}^{(t-1)}, \mathbf{p}^{(t-1)}\right).\end{aligned} \tag{3.56}$$

In step 4 of Algorithm 1, for given $\mathbf{a}^{(t)}$, $\mathbf{f}^{(t-1)}$, $\mathbf{q}^{(t-1)}$ and $\mathbf{p}^{(t-1)}$, $\mathbf{b}^t$ is the optimal channel allocation of problem (3.32) since the strong duality holds for the dual problem, and we have

$$\begin{aligned}&\mathcal{O}\left(\mathbf{a}^{t}, \mathbf{b}^{(t-1)}, \mathbf{f}^{(t-1)}, \mathbf{q}^{(t-1)}, \mathbf{p}^{(t-1)}\right) \\ &\geqslant \mathcal{O}\left(\mathbf{a}^{t}, \mathbf{b}^{t}, \mathbf{f}^{(t-1)}, \mathbf{q}^{(t-1)}, \mathbf{p}^{(t-1)}\right).\end{aligned} \tag{3.57}$$

In step 5 of Algorithm 1, since $G\left(f_{km}^{G}\right)$ is lower bounded by its the first-order Taylor expansion as shown in (3.44), the objective value of convex problem (3.46) serves an upper bound of that of the DC programming (3.41). Denote $\mathcal{O}^{\text{ub},t}\left(\cdot\right)$ as an upper bound of UEs' total energy consumption which is calculated based on the optimization problem in (3.46) with respect to the computation capacity allocation $\mathbf{f}$. For given $\left(\mathbf{a}^{(t)},\mathbf{b}^{(t)},\mathbf{f}^{(t-1)},\mathbf{q}^{(t-1)}\right)$, it follows

$$\begin{aligned}
&\mathcal{O}\left(\mathbf{a}^{t},\mathbf{b}^{t},\mathbf{f}^{(t-1)},\mathbf{q}^{(t-1)},\mathbf{p}^{(t-1)}\right)\\
&\geqslant \mathcal{O}\left(\mathbf{a}^{t},\mathbf{b}^{t},\mathbf{f}^{(t-1)},\mathbf{q}^{(t-1)},\mathbf{p}^{*}\left(\mathbf{b}^{t},\mathbf{f}^{(t-1)},\mathbf{q}^{(t-1)}\right)\right)\\
&\overset{(a)}{=} \mathcal{O}^{\text{ub},t-1}\left(\mathbf{a}^{t},\mathbf{b}^{t},\mathbf{f}^{(t-1)},\mathbf{q}^{(t-1)},\mathbf{p}^{*}\left(\mathbf{b}^{t},\mathbf{f}^{(t-1)},\mathbf{q}^{(t-1)}\right)\right)\\
&\overset{(b)}{\geqslant} \mathcal{O}^{\text{ub},t}\left(\mathbf{a}^{t},\mathbf{b}^{t},\mathbf{f}^{t},\mathbf{q}^{(t-1)},\mathbf{p}^{*}\left(\mathbf{b}^{t},\mathbf{f}^{t},\mathbf{q}^{(t-1)}\right)\right)\\
&\overset{(c)}{\geqslant} \mathcal{O}\left(\mathbf{a}^{t},\mathbf{b}^{t},\mathbf{f}^{t},\mathbf{q}^{(t-1)},\mathbf{p}^{*}\left(\mathbf{b}^{t},\mathbf{f}^{t},\mathbf{q}^{(t-1)}\right)\right),
\end{aligned} \tag{3.58}$$

where $\mathbf{p}^{*}\left(\cdot\right)$ denotes the function of optimal power control as stated in Theorem 1. In (3.58), (a) holds since the first-order Taylor expansions in (3.44) are tight at the given local points, which means that the solution of problem (3.46) serves the same value of UEs' total energy consumption as that of problem (3.17); (b) holds since in step 5 of Algorithm 1 with given $\mathbf{a}^{t}$, $\mathbf{b}^{t}$ and $\mathbf{q}^{(t-1)}$, problem (3.46) is solved optimally with solution $\mathbf{f}^{t}$; (c) holds since $\mathcal{O}_{F}^{\text{ub},t}$ is the upper bound of the objective value of original problem (3.17) at $\mathbf{f}^{t}$. The inequality in (3.58) suggests that although only an approximate optimization problem is solved for obtaining the computation capacity allocation, the objective value of the original problem (3.17) is still non-increasing after each iteration.

In step 6 of Algorithm 1, given $\left(\mathbf{a}^{t},\mathbf{b}^{t},\mathbf{f}^{t},\mathbf{q}^{(t-1)}\right)$, we have

$$\begin{aligned}
&\mathcal{O}\left(\mathbf{a}^{t},\mathbf{b}^{t},\mathbf{f}^{t},\mathbf{q}^{(t-1)},\mathbf{p}^{*}\left(\mathbf{b}^{t},\mathbf{f}^{t},\mathbf{q}^{(t-1)}\right)\right)\\
&\geqslant \mathcal{O}\left(\mathbf{a}^{t},\mathbf{b}^{t},\mathbf{f}^{t},\mathbf{q}^{t},\mathbf{p}^{*}\left(\mathbf{b}^{t},\mathbf{f}^{t},\mathbf{q}^{t}\right)\right)\\
&= \mathcal{O}\left(\mathbf{a}^{t},\mathbf{b}^{t},\mathbf{f}^{t},\mathbf{q}^{t},\mathbf{p}^{t}\right).
\end{aligned} \tag{3.59}$$

Inequality (3.59) holds because $\mathbf{q}^{t}$ is the optimal 3D positions of UAVs regarding problem

Algorithm 1 Iterative resource allocation for air-ground integrated MEC networks.

```
Require:
    Set the initial solution $\left(\mathbf{a}^{(0)}, \mathbf{p}^{(0)}, \mathbf{b}^{(0)}, \mathbf{f}^{(0)}, \mathbf{q}^{(0)}\right)$, the iteration index $t = 0$, the tolerance of objective function $\varepsilon_{obj}$, the maximal iteration number $N_{itera}^{\max}$.
Ensure:
    The near-optimal resource allocation strategies for air-ground integrated MEC networks;
    The minimum UEs' total energy consumption.
while $\left(E_{UEs}^{t+1} - E_{UEs}^{t}\right) / E_{UEs}^{t} \geqslant \varepsilon_{obj}$ or $t \leqslant N_{itera}^{\max}$ do
    Updating iteration index $t = t + 1$;
    With fixed $\left(\mathbf{b}^{(t-1)}, \mathbf{f}^{(t-1)}, \mathbf{q}^{(t-1)}, \mathbf{p}^{(t-1)}\right)$, optimize $\mathbf{a}^{(t)}$ according to problem (3.18);
    With fixed $\left(\mathbf{a}^{(t)}, \mathbf{f}^{(t-1)}, \mathbf{q}^{(t-1)}, \mathbf{p}^{(t-1)}\right)$, optimize $\mathbf{b}^{(t)}$ from problem (3.32);
    With fixed $\left(\mathbf{a}^{(t)}, \mathbf{b}^{(t)}, \mathbf{f}^{(t-1)}, \mathbf{q}^{(t-1)}\right)$, optimize $\mathbf{f}^{(t)}$ from problem (3.41) and problem (3.42);
    With fixed $\left(\mathbf{a}^{(t)}, \mathbf{b}^{(t)}, \mathbf{f}^{(t)}\right)$, optimize $\mathbf{q}^{(t)}$ from problem (3.51);
    With fixed $\left(\mathbf{a}^{(t)}, \mathbf{b}^{(t)}, \mathbf{f}^{(t)}, \mathbf{q}^{(t)}\right)$, obtain $\mathbf{p}^{(t)}$ in closed-form based on Theorem 1;
    Obtain objective value $E_{UEs}^{t} = \mathcal{O}\left(\mathbf{a}^{t}, \mathbf{b}^{t}, \mathbf{f}^{t}, \mathbf{q}^{t}, \mathbf{p}^{*}\left(\mathbf{b}^{t}, \mathbf{f}^{t}, \mathbf{q}^{t}\right)\right)$;
end while
return The near-optimal resource allocation strategies for air-ground integrated MEC networks, $(\mathbf{a}^{*}, \mathbf{p}^{*}, \mathbf{b}^{*}, \mathbf{f}^{*}, \mathbf{q}^{*})$;
    The minimum UEs' total energy consumption, $E_{UEs}^{*}$.
```

(3.51) with fixed $\mathbf{a}^{t}$, $\mathbf{b}^{t}$ and $\mathbf{f}^{(t)}$.

Thus, the UEs' total energy consumption is non-increasing after the updates of users' association, channel allocation, computation capacity allocation, UAV placement and uplink power allocation. In addition, the UEs' total energy consumption is always positive. Since the UEs' total energy consumption is non-increasing in each iteration and is lower bounded by zero, Algorithm 1 converges.

3.3.7 Complexity Analysis

To solve the users' association problem (3.18), the complexity of updating users' association according to (3.22) is $O\left(K\left(M^{G} + M^{A}\right)\right)$. The complexity of updating dual variables in (3.24)-(3.29) is also $O\left(K\left(M^{G} + M^{A}\right)\right)$. Denote the number of iterations for dual method of problem (3.18) is N_A. The total complexity of problem (3.18) is $O\left(N_A K\left(M^G+M^A\right)\right)$.

For problem (3.32), denote the complexity of obtaining the inverse function as $O\left(\frac{1}{\varepsilon_1}\right)$. The complexity of updating dual variables in (3.36)-(3.37) is $O\left(K\right)$. Thus, the total complexity of problem (3.32) is $O\left(N_B K \log_2\left(\frac{1}{\varepsilon_1}\right)\right)$, where N_B is the number of iterations in problem (3.32).

For problem (3.39), denote the complexity of obtaining the inverse function (3.47) or (3.50) as $O\left(\frac{1}{\varepsilon_2}\right)$. The complexity of solving (3.48) is denoted by $O\left(\frac{1}{\varepsilon_3}\right)$. The total complexity of problem (3.39) is given by $O\left(\left(M^G+M^A\right)K\log_2\left(\frac{1}{\varepsilon_2}\right)\log_2\left(\frac{1}{\varepsilon_3}\right)\right)$.

Denote the complexity of solving (3.54) is $O\left(\frac{1}{\varepsilon_4}\right)$. The complexity of solving problem (3.53) is $O\left(2\log_2\left(\frac{1}{\varepsilon_1}\right)\right)$. The total complexity of problem (3.51) is given by $O\left(2M^A N_H \log_2\left(\frac{1}{\varepsilon_1}\right)\right)$, where N_H is the number of discrete values of UAV's flight height.

Thus, the total complexity of Algorithm 1 is given by

$$\begin{aligned} O\Bigg(N_0\Bigg(N_B K\log_2\left(\frac{1}{\varepsilon_1}\right)+\left(M^G+M^A\right)K\Bigg(N_A \\ +\log_2\left(\frac{1}{\varepsilon_2}\right)\log_2\left(\frac{1}{\varepsilon_3}\right)\Bigg)+2M^A N_H\log_2\left(\frac{1}{\varepsilon_1}\right)\Bigg)\Bigg), \end{aligned} \tag{3.60}$$

where N_0 is the number of outer iterations.

3.4 Simulation Results and Discussion

In this section, we provide numerical results based on our proposed resource allocation algorithm and evaluate the performance of the computation offloading in such air-ground integrated MEC networks. The simulated system works on the parameter settings shown in Table I, unless otherwise stated.

An illustration of the simulated scenario is presented in Fig. 3.2. UEs are uniformly dis-

Table 3.1: Default Parameter Setup

Parameter	**Value**
K, M_A, M_G	100, 10, 10
H_{min}, H_{max}	10 m, 120 m
UE max power P_k^{max}	0.5 W
$f_m^{G,max}$, $f_m^{A,max}$	1×10^{10} cycles/s, 1×10^{10} cycles/s
UE data size D_k	5×10^5 bits
Latency T_k	1000 ms [56]
Bit and cycle conversion c_{bc}	1×10^3 cycles/bit [54]
UE computation capacity f_k^{max}	2×10^9 cycles/s
Coefficients ρ, ς	1×10^{-27}, 3 [54]
Maximal connected UEs U_G, U_A	30, 10 [56]
Total bandwidth B	10 MHz [52]
Noise spectral density n_0	-104 dBm/Hz [52]
Path-loss exponent α	3

tributed on the ground within the region of interest form by 1000 m × 1000 m, where there are 100 UEs. There are 10 GCAPs which are uniformly distributed on the ground. The number of UAVs is 10. We consider the uniform distributed UAVs as the initial solutions of UAV placement, and the 3D positions of UAVs will be optimized based on the proposed algorithm in the sequel. In Fig. 3.2, we presented the optimized 3D positions of UAVs. Our simulations are run on Matlab using a computer with the CPU processor @ 1.8GHz and 1.99GHz. The small-scale power channel gain for ground wireless communications is assumed to be exponential distribution with mean value 1 [42]. We consider equal parameters for all GCAPs and UAVs (i.e., equal maximal computation capacity, equal maximal number of associated UEs). In simulations. we suppose that all UEs have equal computation tasks (i.e., equal computing data size, equal computation capacity).

In Fig. 3.3, we compare the total energy consumption of UEs with respect to the iteration number of the proposed algorithm. It can be observed that the objective value of the Problem (3.17), i.e., the total energy consumption of UEs, decreases rapidly within several iterations. The total energy consumption of UEs obtained by the proposed algorithm

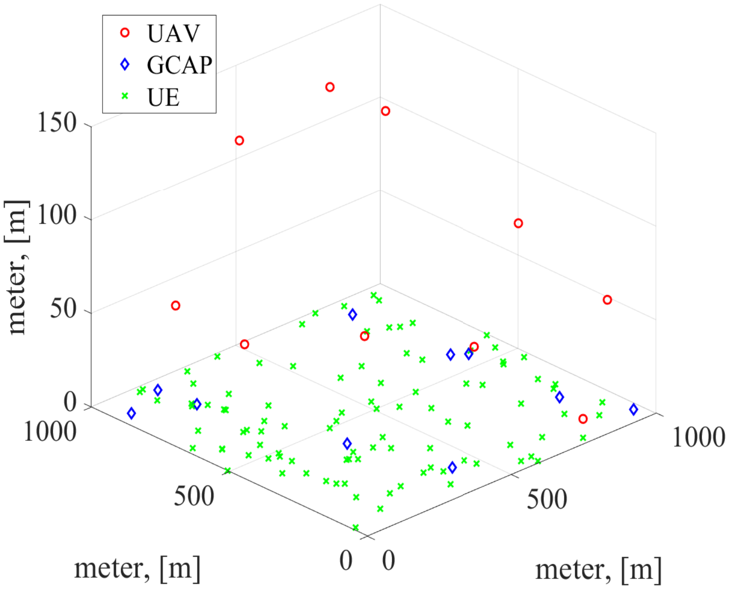


Figure 3.2: Air-ground integrated MEC networks.

outperforms that conducted by the original method where all UEs' computation tasks are executed locally at mobile devices denoted as iteration 0. In addition, according to Fig. 3.3, when the maximal computation capacity at MEC servers (f_{max}) increases, UEs can save more energy through computation offloading while guaranteeing their latency constraints. This is because, with the increase of the maximal computation capacity at MEC servers, more UEs can be offloaded onto UAVs and GCAPs, and thus UEs will only consume less energy for data transmissions than the great energy for locally computing. It can also be observed that the total energy consumption of UEs can be further reduced if the available system bandwidth increases. This is because the increased bandwidth compensate for the reduced transmit power at a UE to achieve a desired data rate. Therefore, the total energy consumption for transmissions at UEs reduces with the increase of available system bandwidth. Besides, the increased bandwidth allows more UEs to offload onto MEC servers through wireless communications while guaranteeing their latency constraints.

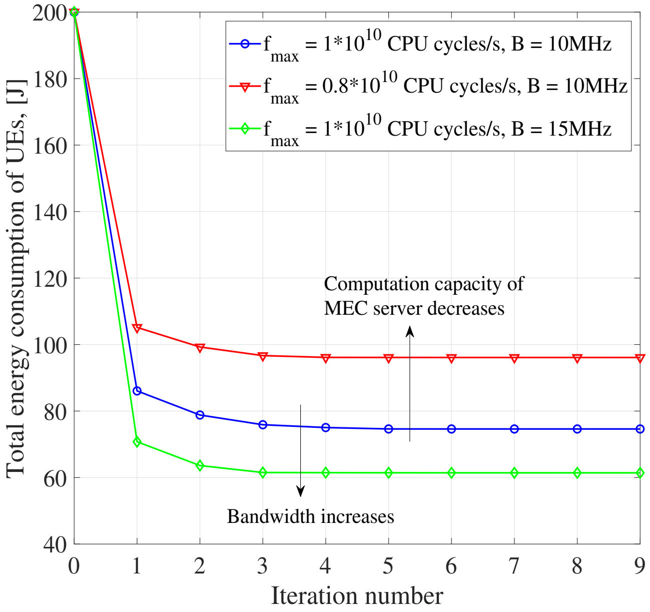


Figure 3.3: Energy consumption of UEs with respect to iteration numbers.

In Fig. 3.4, we compare the total energy consumption of UEs obtained by the proposed algorithm with respect to the computation capacity of UE. It can be observed that the total energy consumption of UEs increases with the UEs computation capacity. This can be explained as follows, when the UEs computation capacity increases, the execution time of computation tasks is reduced. However, the power consumption of computation increases more rapidly based on the average activity factor ς which is usually greater than 2 [60]. Therefore, the energy consumption for local computing at each UE increases with UE's computation capacity.

According to the above property, the MEC servers will consume more energy for the same tasks if MEC servers have higher computing capability. The motivation behind that UEs choose to offload computation tasks onto MEC servers is discussed as follows. Since energy is usually scarce at mobile devices, to prolong the battery life-time of mobile devices, UEs are willing to offload their computation tasks onto MEC servers to reduce their energy

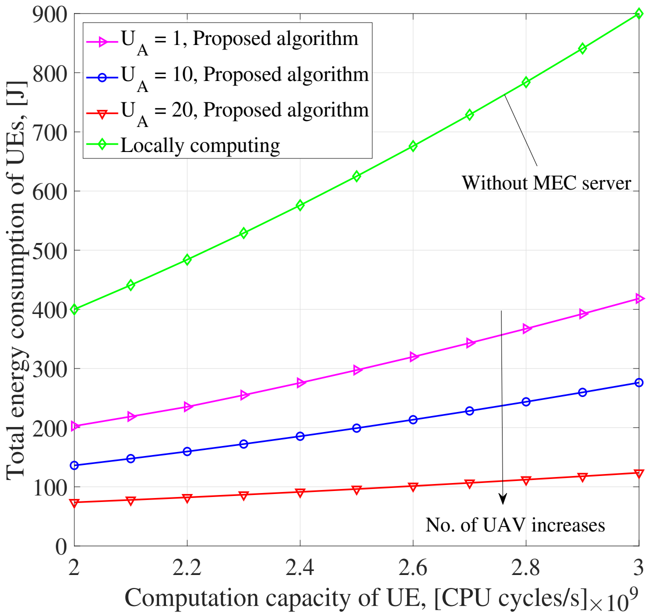


Figure 3.4: Energy consumption of UEs with respect to UE's computation capacity under different numbers of UAVs, where $D_k = 1 \times 10^6$ bits.

consumption at mobile devices. On the other hand, higher computing capability of MEC servers facilitates more computation tasks offloaded from UEs to MEC servers, which brings more income to service provider by charging from UEs while at the price of consuming more energy at MEC servers. Thus, both UEs and service provider would benefit from the computation offloading.

In addition, in Fig. 3.4, by comparing different numbers of UAVs, i.e, M_A, we observe that the total energy consumption of UEs decreases with M_A. This is because when M_A increases, more UEs can be offloaded onto UAVs for computation offloading with LOS connections, where some UEs transfer from GCAPs to UAVs to achieve lower transmit power, or get rid of locally computing, which decreases the total energy consumption of UEs. In Fig. 3.4, we also compare the performance of the proposed algorithm for air-ground integrated MEC networks and that of the method of local computing. It can be observed that much energy

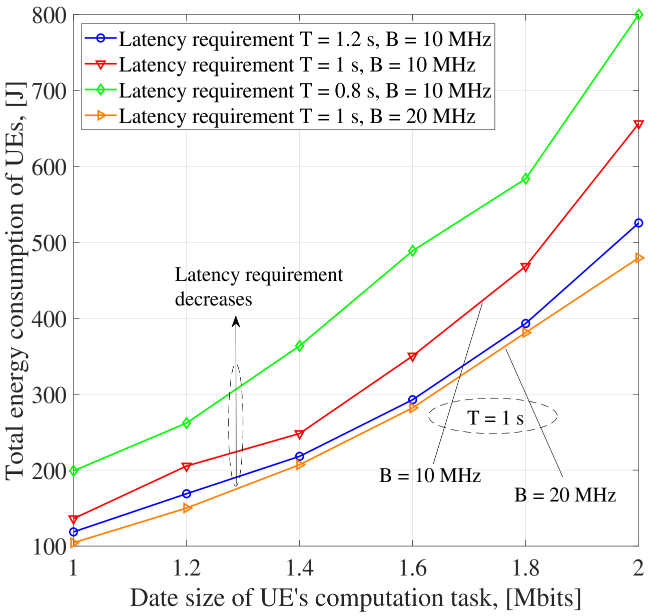


Figure 3.5: Energy consumption of UEs with respect to the data size of UE's computation task under different latency constraints.

can be saved at UEs based on the proposed algorithm in computation offloading.

In Fig. 3.5, we examine the total energy consumption of UEs with respect to the data size of UE's computation task. When the data size of UE's computation task increases, the total energy consumption of UEs improves. On one hand, more data needs to be transmitted from UEs to MEC serves and thus this consumes additional energy for data transmission. On the other hand, the limited computation capacity at MEC servers restricts some UEs to be offloaded onto MEC servers due to the unsatisfied latency constraints. The increased number of UEs which compute locally leads to the improvement of the total energy consumption of UEs. In addition, if the latency constraints of UEs increases, more UEs would prefer to wait for the computation results returned from MEC servers. Therefore, increasing the latency requirement of UEs reduces the total energy consumption of UEs. This makes UEs balance their computing data size and latency constraints in terms of their energy consumption in the MEC networks.

3.5 Summary

In this chapter, we developed an energy efficient resource allocation algorithm for an integrated UAV and ground mobile edge computing (MEC) network. We aimed to minimize the total energy consumption of UEs by optimizing users' association for computation offloading, uplink transmit power, allocated bandwidth, computation capacity and UAV placement. To solve the primary problem, we decomposed it into several manageable subproblems and alternatively optimizing each subproblem. Simulation results demonstrated the advantages of our proposed algorithm compared with the local computing method.

Chapter 4

Deep Learning-Assisted Energy-Efficient Task Offloading in Vehicular Edge Computing Systems

4.1 Introduction

In vehicular networks, vehicles need to process a significant amount of data in real-time for road traffic services such as safety, sensing, localization, and decision making [65, 66, 67]. In-vehicle users also require infotainment applications involving the execution of data. Due to the extensive workloads and the limited computation capacity at both vehicles and in-vehicle users, it is quite difficult to meet the latency requirements when they locally process their computation tasks [68]. Moreover, as vehicles and in-vehicle users are mobile, they are often subject to strict energy consumption restrictions. Besides, as the energy consumption of computation is usually greater than that of data transmission, it is expected to develop energy-efficient methods to reduce users' energy consumption and guarantee their latency requirements in computing. Vehicular edge computing (VEC) has become a promising computing architecture for vehicular networks, where users' computation tasks are offloaded to edge servers via roadside units (RSUs) which have wireless communications capability for receiving data from users and sending back computation results [69, 70]. Compared to locally executing computation tasks at users, VEC systems enjoy two main advantages of energy saving and latency guarantee, which benefit from the increased computation capability and the short communication distance.

In terrestrial mobile edge computing (MEC) systems [71], full channel state information (CSI) is assumed to be available at base stations [45, 72, 73]. This assumption does not hold anymore in VEC systems, since the channel varies fast due to the high speed of moving vehicles and it is quite difficult to estimate CSI and feed back to the RSUs [74, 75]. Therefore, it is expected to take into account the channel uncertainty in designing VEC systems. In many computation services and applications, the latency is required to be just below a given threshold, while users seek to achieve lower energy consumption as possible [9, 76]. Under the uncertainty of small-scale fading, it remains an open queation as to how to jointly allocate both communication and computation resources based on large-scale fading channel information in VEC systems to minimize the total energy consumption of users. The main differences of this work from the existing works related to VEC are presented as follows: a) the unknown small-scale fading in vehicular-to-RSU (V2R) channels is taken into account; b) multiple edge servers and users are considered; c) a practical partial offloading assumption is made for VEC systems; d) the objective is to minimize total energy consumption of users. Based on these differences and considerations, in this chapter, we minimize the total energy consumption of users by jointly optimizing the variables of users' association, data partition, transmit power, and computation resources at edge servers, subject to the constraints of partial offloading, the maximum transmit power, users' latency, outage performance, and computation capacity of edge servers. Such a computation offloading problem is practically appealing but has not been studied in the literature to the best of authors' knowledge.

4.1.1 Related Works

In existing works, various aspects of VEC have been investigated in the literature such as latency minimization in the integrated cloud and edge computing [77], VEC systems utility maximization in vehicle-assisted computation offloading ground users [78], federated learning in VEC systems [79], etc. However, there are few works addressing the energy minimization problem for VEC systems [80, 81]. In [80], the authors studied an energy-efficient workload offloading problem based on consensus alternating direction method of multipliers. However, [80] only considered the workload problem in one RSU's coverage and it lacks the cooperation between multiple RSUs and edge servers while the road traffic is dynamic and inhomogeneous along the road. Besides, the problem only optimized the portion of the workload, but did not investigate the joint communication and computation resource allocation problem. In [81], the authors studied an optimization problem to minimize the overall system energy consumption by optimizing the decisions between local and offloading and the number of allocated resource blocks; however, it did not consider the partial offloading in VEC systems as well as the joint communication and computation resource allocation among multiple edge servers. In addition, the above works did not consider the channel uncertainty in VEC systems caused by high-speed motion of vehicles.

4.1.2 Contributions

We summarize the contributions of this chapter as follows:

- We develop a computation offloading algorithm for VEC systems. Specifically, we minimize the total energy consumption of users by jointly optimizing the variables of user association, data partition, transmit power, and computation resources at edge servers, subject to the constraints of partial offloading, the maximum transmit power, user latency, outage performance, and computation capacity of edge servers.

- We consider the unknown small-scale fading in V2R channels and the outage performance in VEC systems. Moreover, our work is not restricted to a single RSU and an edge server. We consider multiple edge servers and users in VEC systems. Furthermore, we focus on the partial offloading model for computation tasks, making it possible to implement fine-grained computation offloading in VEC systems.

- We utilize deep learning method to obtain user association and integrate it with the developed optimization algorithm. Based on this approach, the computational complexity for obtaining the user association is very low, where the input network parameters simply go through a designed neural network model.

The remainder of this chapter is organized as follows. In Chapter 4.2 presents the system model. Chapter 4.3 studies the introduced algorithm. Numerical results are given and discussed in Chapter 4.4. Finally, conclusions are drawn in Chapter 4.5.

4.2 System Model

4.2.1 Network Layout

We consider a VEC system with M edge servers and K users as shown in Fig. 4.1. The set of edge servers is denoted by $\mathcal{M} = \{1, 2, \cdots, M\}$, and the set of users is indicated by $\mathcal{K} = \{1, 2, \cdots, K\}$. Each RSU has a wire-connected edge server that has a certain computation resource to process users' computation tasks.

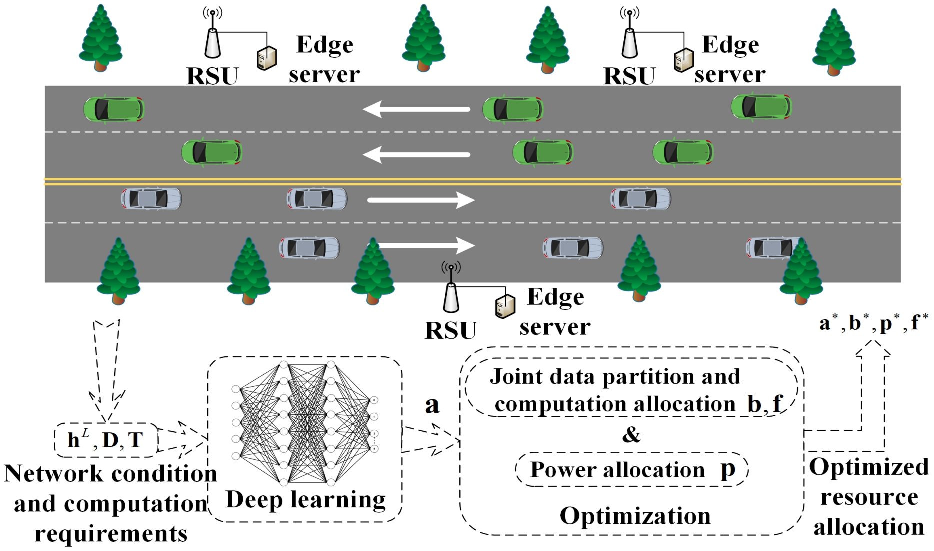


Figure 4.1: An architecture of vehicular edge computing systems.

4.2.2 Communication Model

Based on the channel modeling for vehicular communications in [74, 75], the V2R channel between k-th user and m-th RSU at time slot t is given by $h_{km}(t) = h_{km}^L g_{km}(t)$, where $h_{km}^L = h_{ref}\varsigma_{km}(L_{km})^{-\alpha}$ accounts for the large-scale fading component and $g_{km}(t)$ represents the small-scale fading component at time slot t. Specifically, h_{ref} is channel power gain at reference distance, ς_{km} is the shadowing component, L_{km} denotes the distance between k-th user and m-th RSU, and α is the pathloss exponent. The instantaneous uplink data rate $R_{km}(t)$ of the k-th user connected to the m-th RSU at the time slot t is given by $R_{km}(t) = B\log_2\left(1 + \frac{h_{km}(t)p_k}{N_0}\right)$, where B is the channel bandwidth, p_k indicates the transmit power of the k-th user and N_0 represents the noise power. We consider that users are allocated with orthogonal resources for uplink transmissions in a specific road segment. The adjacent road segments operate on different frequency bands. Thus, the interference from other road segments on the same frequency band is neglected. The maximum transmit power of the k-th user is denoted by $p_k^{\max}$.

In the time domain, time is equally divided by time slots of length T on the order of hundreds of microseconds. Many consecutive time slots construct a time block on the order of hundreds of milliseconds. The large-scale fading component is typically determined by users' locations which vary little within each time block. We assume that the large-scale fading component is known at RSUs, because the locations of vehicles are usually available at RSUs. However, the small-scale fading component varies rapidly during a time block due to the high mobility of vehicles, which is unavailable at RSUs, but its statistical characterization is assumed to be known. We assume that the small-scale fading component remains constant during one time slot but fluctuates as an independent and identically distributed (i.i.d.) random variable across different time slots. We consider the Rayleigh distribution for small-scale fading with parameter λ_g [74].

$$\begin{aligned} R_{km} &= \int_0^{\infty} B\log_2\left(1+\frac{h_{km}^L p_k}{N_0}x\right) f_g\left(x\right)dx \\ &= \frac{B\lambda_g}{\ln 2}\int_0^{\infty}\frac{e^{-\lambda_g x}}{x+\frac{N_0}{h_{km}^L p_k}}dx = \frac{B\lambda_g}{\ln 2}\Phi\left(\frac{N_0\lambda_g}{h_{km}^L p_k}\right), \end{aligned} \tag{4.1}$$

where $f_g\left(x\right) = \lambda_g e^{-\lambda_g x}$, $\Phi\left(x\right) = e^x E_1\left(x\right)$, $E_1\left(x\right) = \int_x^{\infty}\frac{e^{-y}}{y}dy$ $(x \geqslant 0)$ is the exponential integral function, and the exponential distribution of small-scale fading component (i.e., $f_g\left(x\right)$) is introduced.

4.2.3 Computation Offloading

In partial computation offloading, the k-th user offloads β_k ($\beta_k \in [0,1]$) portion of its data to an edge server, while the remaining $1-\beta_k$ portion of data is executed locally at the k-th user. The association vector of the k-th user is given by $\mathbf{a}_k = \{a_{k1}, \cdots, a_{kM}\}$, where $a_{km} = 1$ is defined as if partial task of the k-th user is offloaded on the m-th server, otherwise, $a_{km} = 0$.

Each user can offload partial data to only one edge server, obeying the constraint as follows

$$\sum_{m=1}^{M} a_{km} = 1, \quad \forall k \in \mathcal{K}. \tag{4.2}$$

Note that if the k-th user executes all of its data locally, we can set β_k to zero, regardless of the values of $\mathbf{a}_k$. In (4.2), each user is associated with one RSU, facilitating the control information exchange in VEC systems.

The computation task of the k-th user is expressed as $V_k = (D_k, T_k, F_k), \forall k \in \mathcal{K}$, where D_k is the data size of its computation task, T_k is the latency requirement of this task and F_k is the required number of central processing unit (CPU) cycles for processing this task. In general, the required number of CPU cycles is given by $F_k = c_{bc} D_k$, where c_{bc} is the coefficient for bit and cycle conversion [45].

4.2.4 Computation Model

Each edge server has limited computation capacity, upper bounded by $f_m^{\max}$, which indicates the maximum number of allocated CPU cycles per second at the m-th edge server. The allocated computation resource for the k-th user at m-th edge server is denoted by f_{km}, $\forall k, m$. The computation capacity constraint at the m-th edge server is given by

$$\sum_{k=1}^{K} a_{km} f_{km} \leqslant f_m^{\max}, \quad \forall m \in \mathcal{M}. \tag{4.3}$$

We consider that the k-th user transmit at R_{km} in a time block to reduce implementation complexity. Accordingly, the latency constraint of the k-th user is given by

$$\max \left\{ \sum_{m=1}^{M} a_{km} \beta_k \left(\frac{D_k}{R_{km}} + \frac{F_k}{f_{km}} \right), \frac{(1-\beta_k) F_k}{f_{k0}} \right\} \leqslant T_k, \tag{4.4}$$

where f_{k0} denotes the local computation capacity of the k-th user. The computation power consumption of k-th user is given by $p_k^C = \rho (f_{k0})^{\varsigma}$, $\forall k \in \mathcal{K}$, where ρ and ς are constants that depend on the average switched capacitance and the average activity factor, respectively [45].

4.2.5 Outage Probability

The outage probability of data transmission for the k-th user connected to the m-th RSU in a time slot is given by $\mathbb{P}^o_{km} = \mathbb{P}\left\{R_{km}(t) \leqslant R_{km}\right\}$. In addition, the required number of time slots for data transmission is denoted by N_{km}, where $N_{km} = \frac{\beta_k D_k}{T R_{km}}$. We consider that the user's average number of outage time slots is less than a certain threshold μ_o, i.e., $N_{km}\mathbb{P}\left\{R_{km}(t) \leqslant R_{km}\right\} \leqslant \mu_o, \forall k \in \mathcal{K}$, where μ_o denotes the threshold of the expected maximum outage time slots for transmitting β_k portion of data to the RSU.

Lemma 4.1. *The outage constraint of the k-th user connected to the m-th RSU is given by*

$$\sum_{m=1}^{M}\left\{1-\exp\left[-\left(2^{\frac{\lambda_g}{\ln 2}\Phi\left(\frac{N_0\lambda_g}{h^L_{km}p_k}\right)}-1\right)\frac{N_0\lambda_g}{h^L_{km}p_k}\right]\right\}\frac{a_{km}\beta_k D_k \ln 2}{TB\lambda_g\Phi\left(\frac{N_0\lambda_g}{h^L_{km}p_k}\right)} \leqslant \mu_o, \quad \forall k \in \mathcal{K}. \tag{4.5}$$

Proof. Please refer to Appendix C.1 for the proof. □

4.2.6 Problem Formulation

In this paper, we focus on minimizing the total energy consumption of users, as follows

$$\min_{\mathbf{a},\mathbf{b},\mathbf{p},\mathbf{f}} \sum_{k=1}^{K}\frac{(1-\beta_k)F_k p^C_k}{f_{k0}} + \sum_{k=1}^{K}\sum_{m=1}^{M}\frac{a_{km}p_k \ln 2\beta_k D_k}{B\lambda_g\Phi\left(\frac{N_0\lambda_g}{h^L_{km}p_k}\right)} \tag{4.6}$$

s.t. (4.2), (4.3), (4.4), (4.5),

$$a_{km} = \{0,1\}, \quad \forall k \in \mathcal{K}, \quad \forall m \in \mathcal{M}, \tag{4.6a}$$

$$0 \leqslant p_k \leqslant p^{\max}_k, \quad \forall k \in \mathcal{K}, \tag{4.6b}$$

$$0 \leqslant \beta_k \leqslant 1, \quad \forall k \in \mathcal{K}, \tag{4.6c}$$

$$0 \leqslant f_{km}, \quad \forall k \in \mathcal{K}, \quad \forall m \in \mathcal{M}, \tag{4.6d}$$

where $\mathbf{a} = \{a_{km}\}_{k\in\mathcal{K},m\in\mathcal{M}}$ is users' association, $\mathbf{b} = \{\beta_k\}_{k\in\mathcal{K}}$ indicates users' data partition, $\mathbf{p} = \{p_k\}_{k\in\mathcal{K}}$ denotes users' transmit power, and $\mathbf{f} = \{f_{km}\}_{k\in\mathcal{K},m\in\mathcal{M}}$ represents computation resource allocation. In the objective function of (4.6), the first term is the energy consumption for users' local computing, and the second term captures the energy consumption of users' data transmission. The constraints of problem (4.6) includes partial tasks offloading constraints, edge servers' computation capacity constraints, users' latency constraints, the outage performance constraints, the maximum transmit power of users, and feasible value constraints. Note that (4.6) is a non-convex problem due to the coupled variables in non-convex objective and constraints.

4.3 Proposed Algorithm

We study an efficient algorithm to solve the users' energy consumption minimization problem (4.6) in VEC systems.

4.3.1 Joint Data Partition and Computation Resource Allocation

Given the user's association $\mathbf{a}$ and power allocation $\mathbf{p}$, we jointly optimize users' data partition $\mathbf{b}$ and computation resource allocation $\mathbf{f}$, where the sub-problem is given by

$$\min_{\mathbf{b},\mathbf{f}} \quad \sum_{k=1}^{K} \frac{(1-\beta_k)\,F_k p_k^C}{f_{k0}} + \sum_{k=1}^{K} \frac{a_{km} p_k \beta_k D_k}{R_{km}} \tag{4.7}$$

$$\text{s.t. } (4.3), (4.5), (4.6c), (4.6d),$$

$$\frac{a_{km}\beta_k D_k}{R_{km}} + \frac{a_{km}F_k\beta_k}{f_{km}} - T_k \leqslant 0, \quad \forall k \in \mathcal{K}, \tag{4.7a}$$

$$\frac{(1-\beta_k)\,F_k}{f_{k0}} - T_k \leqslant 0, \quad \forall k \in \mathcal{K}. \tag{4.7b}$$

It can be verified that the left hand side (LHS) of constraint (4.7a) is not jointly convex for β_k and f_{km}. For ease of analysis, we introduce $\gamma_k = \sqrt{\beta_k}$ to transform (4.7) as follows

$$\min_{\mathbf{c},\mathbf{f}} - \sum_{k=1}^{K} {\gamma_k}^2 c_k \tag{4.8}$$

$$\text{s.t. } (4.3), (4.6d),$$

$$\frac{a_{km}{\gamma_k}^2 D_k}{R_{km}} + \frac{a_{km}{\gamma_k}^2 F_k}{f_{km}} - T_k \leqslant 0, \quad \forall k \in \mathcal{K}, \tag{4.8a}$$

$$\gamma_k^{lb} \leqslant \gamma_k \leqslant \gamma_k^{ub,1}, \forall k \in \mathcal{K}, \tag{4.8b}$$

where $c_k = \frac{F_k}{f_{k0}} p_k^C - \frac{a_{km} p_k \beta_k D_k}{R_{km}}$, $\gamma_k^{lb} = \max\left\{1 - \frac{T_k f_{k0}}{F_k}, 0\right\}^{\frac{1}{2}}$, $\gamma_k^{ub,1} = \min\left\{1, \sqrt{\frac{\mu_o T B \lambda_g}{\ln 2 D_k} \Theta\left(\frac{N_0 \lambda_g}{h_{km}^L p_k}\right)}\right\}$, $\Theta(\cdot)$ is given by

$$\Theta(x) = \Phi(x)\left[1 - \exp\left(1 - e^{\left(1 - 2^{\lambda_g \Phi(x)/\ln 2}\right)x}\right)\right]^{-1}. \tag{4.9}$$

Proposition 4.2. *The LHS of constraint (4.8a) is jointly convex for γ_k and f_{km}.*

Proof. Please refer to Appendix C.2 for the proof. □

However, the convexity of the objective function of problem (4.8) depends on $c_k, k \in \mathcal{K}$. We partition the set of users $\mathcal{K}$ into two subsets $\mathcal{K}_-$ and $\mathcal{K}_+$, where $\mathcal{K}_- = \{k | \, k \in \mathcal{K}, c_k \leqslant 0\}$, $\mathcal{K}_+ = \{k | \, k \in \mathcal{K}, c_k > 0\}$. With the successive convex optimization technique, in each iteration, the objective concave functions are approximated by more tractable functions at given local points. Recall that any concave function is globally upper-bounded by its first-order Taylor expansion at any point. For $k \in \mathcal{K}_+$, we have

$$-\sum_{k \in \mathcal{K}_+} {\gamma_k}^2 c_k \leqslant -\sum_{k \in \mathcal{K}_+} \left(\left(\gamma_k^i\right)^2 c_k + 2\gamma_k^i c_k \left(\gamma_k - \gamma_k^i\right)\right), \tag{4.10}$$

where γ_k^i is the value of γ_k in the ith iteration. Given γ_k^i, problem (4.8) is reformulated by

$$\begin{aligned} &\min_{\mathbf{c},\mathbf{f}} - \sum_{k\in\mathcal{K}_-} {\gamma_k}^2 c_k - \sum_{k\in\mathcal{K}_+} \gamma_k^i c_k \left(2\gamma_k - \gamma_k^i\right), \\ &\text{s.t. } (4.3), (4.6d), (4.8a), (4.8b), \end{aligned} \tag{4.11}$$

which is a convex problem. In the sequel, we apply the Lagrangian dual method to solve the problem in (4.11) efficiently by investigating the analytical form of solutions. The Lagarange function of (4.11) is given by

$$\min_{\mathbf{c},\mathbf{f}} L\left(\mathbf{c},\mathbf{f},\varpi,\vartheta\right), \tag{4.12}$$

where $L\left(\mathbf{c},\mathbf{f},\varpi,\vartheta\right)$ is given in (4.13) as follows

$$\begin{aligned} L\left(\mathbf{c},\mathbf{f},\varpi,\vartheta\right) = &- \sum_{k\in\mathcal{K}_-} {\gamma_k}^2 c_k - \sum_{k\in\mathcal{K}_+} \gamma_k^i c_k \left(2\gamma_k - \gamma_k^i\right) \\ &+ \sum_{m=1}^{M} \varpi_m \left(\sum_{k=1}^{K} a_{km}^* f_{km} - f_m^{\max}\right) + \sum_{k=1}^{K} \vartheta_k \left(\frac{a_{km}^* {\gamma_k}^2 D_k}{R_{km}^*} + \frac{a_{km}^* {\gamma_k}^2 F_k}{f_{km}} - T_k\right), \end{aligned} \tag{4.13}$$

where $\{\varpi_m\}_{m\in\mathcal{M}}$ and $\{\vartheta_k\}_{k\in\mathcal{K}}$ are the Lagrangian multipliers.

Since (4.12) is convex, we use the coordinate descent method to find the optimal solution to (4.12). Specifically, given $\mathbf{f}$, we first optimize $\mathbf{c}$; then, given the optimized $\mathbf{c}$, we optimize $\mathbf{f}$, which are shown as follows

$$\mathbf{c}^{j+1} = \arg\min_{\mathbf{c}} \; L\left(\mathbf{c},\mathbf{f}^j,\varpi,\vartheta\right), \tag{4.14}$$

$$\mathbf{f}^{j+1} = \arg\min_{\mathbf{f}} \; L\left(\mathbf{c}^{j+1},\mathbf{f},\varpi,\vartheta\right), \tag{4.15}$$

where $\mathbf{c}^{j+1}$ and $\mathbf{f}^{j+1}$ denote the optimized $\mathbf{c}$ and $\mathbf{f}$ in the $(j+1)$th iteration, respectively.

Theorem 4.3. *Given* $\mathbf{f}^j$*, for* k*-th user, if* $c_k \leqslant 0$*, we have* $\gamma_k^{j+1} = \gamma_k^{lb}$*; if* $c_k > 0$*, we have*

$$\gamma_k^{j+1} = \begin{cases} \gamma_k^{ub}, & \text{if } \gamma_k^{ub} \leqslant \gamma_k^{opt} \\ \gamma_k^{opt}, & \text{if } \gamma_k^{lb} \leqslant \gamma_k^{opt} < \gamma_k^{ub} \\ \gamma_k^{lb}, & \text{if } \gamma_k^{opt} < \gamma_k^{ub} \end{cases}, \tag{4.16}$$

where (4.17)

$$\gamma_k^{opt} = \gamma_k^i c_k \left(\vartheta_k \sum_{m=1}^{M} \frac{a_{km} D_k}{R_{km}} + \frac{a_{km} F_k}{f_{km}^j} \right)^{-1}, \tag{4.18}$$

$$\gamma_k^{ub} = \min \left\{ T_k \Big/ \sum_{m=1}^{M} \frac{a_{km} D_k}{R_{km}} + \frac{a_{km} F_k}{f_{km}^j}, \gamma_k^{ub,1} \right\}^{\frac{1}{2}}. \tag{4.19}$$

Given the optimized $\mathbf{c}^{j+1}$*, the optimal* f_{km}^{j+1} *is given by*

$$f_{km}^{j+1} = \gamma_k^{j+1} \left(\frac{\vartheta_k F_k}{\varpi_m} \right)^{\frac{1}{2}}. \tag{4.20}$$

Proof. Please refer to Appendix C.3 for the proof. □

Once obtaining the optimized $\mathbf{c}$ and $\mathbf{f}$, we update Lagrangian multipliers, as follows

$$\varpi_m^{i+1} = \left[\varpi_m^i + \pi_\varpi \left(\sum_{k=1}^{K} a_{km} f_{km} - f_m^{\max} \right) \right]^+, \tag{4.21}$$

$$\vartheta_k^{i+1} = \left[\vartheta_k^i + \pi_\vartheta \left({\gamma_k}^2 a_{km} \left(\frac{D_k}{R_{km}} + \frac{F_k}{f_{km}} \right) - T_k \right) \right]^+, \tag{4.22}$$

where π_ϖ and π_ϑ are the chosen step-sizes.

By optimizing $\mathbf{c}$, $\mathbf{f}$, and updating ϖ, ϑ, we obtain the optimal $\mathbf{c}$ and $\mathbf{f}$. Then, we calculate back $\mathbf{b}$ from $\mathbf{c}$.

4.3.2 Power Allocation

Given the user's association $\mathbf{a}$, data partition $\mathbf{b}$, and computation resource allocation $\mathbf{f}$, the power allocation sub-problem is given by

$$\min_{\mathbf{p}} \sum_{k=1}^{K} \sum_{m=1}^{M} \frac{a_{km} p_k \ln 2 \beta_k D_k}{B \lambda_g \Phi\left(\frac{N_0 \lambda_g}{h_{km}^L p_k}\right)} \tag{4.23}$$

$$\text{s.t. } (4.5), (4.6b),$$

$$\sum_{m=1}^{M} \frac{a_{km} \ln 2 \beta_k D_k}{B \lambda_g \Phi\left(\frac{N_0 \lambda_g}{h_{km}^L p_k}\right)} + \frac{a_{km} \beta_k F_k}{f_{km}} \leqslant T_k, \quad \forall k \in \mathcal{K}. \tag{4.23a}$$

Problem (4.23) can be decomposed into K sub-problems. For ease of analysis, we introduce the variable $\eta_k = \frac{N_0 \lambda_g}{h_{km}^L p_k}$. The power allocation sub-problem for the k-th user is given by

$$\min_{\eta_k} \frac{\ln 2 \beta_k D_k N_0}{B h_{km}^L \Phi\left(\eta_k\right) \eta_k} \tag{4.24}$$

$$\text{s.t. } \frac{\ln 2 \beta_k D_k}{B \lambda_g \Phi\left(\eta_k\right)} + \frac{\beta_k F_k}{f_{km}} \leqslant T_k, \tag{4.24a}$$

$$\left(1 - e^{\left(1 - 2^{\lambda_g \Phi(\eta_k)/\ln 2}\right)\eta_k}\right) \frac{\beta_k D_k \ln 2}{T B \lambda_g \Phi\left(\eta_k\right)} \leqslant \mu_o, \tag{4.24b}$$

$$\eta_k \geqslant \frac{N_0 \lambda_g}{h_{km}^L p_k^{\max}}. \tag{4.24c}$$

In the following, we provide the optimal expression of η_k regarding the problem (4.24).

Theorem 4.4. *The optimal η_k in (4.24) is given by*

$$\eta_k^* = \max\left\{\eta_k^{lb}, \min\left\{\eta_k^{ub,1}, \eta_k^{ub,2}\right\}\right\}, \tag{4.25}$$

where $\eta_k^{lb} = \frac{N_0 \lambda_g}{h_{km}^L p_k^{\max}}$, $\eta_k^{ub,1} = \Phi^{-1}\left(\frac{\ln 2 \beta_k D_k f_{km}}{B \lambda_g (T_k f_{km} - \beta_k F_k)}\right)$, and $\eta_k^{ub,2} = \Theta^{-1}\left(\frac{\beta_k \ln 2 D_k}{\mu_o T B \lambda_g}\right)$. $\Phi^{-1}(\cdot)$ and $\Theta^{-1}(\cdot)$ are the inverse functions of $\Phi(\cdot)$ and $\Theta(\cdot)$ (defined in (4.9)), respectively.

Proof. Please refer to Appendix C.4 for the proof. ☐

Then, we can reach the optimal p_k by using $p_k = \frac{N_0 \lambda_g}{h_{km}^L \eta_k}$.

4.3.3 Users' Association

We apply a deep neural network (DNN) to obtain user association schemes. The motivation is two-fold: a) this achieves a low computational complexity by simply going through a neural network model implemented in a real-time manner; b) we can enlarge and update the training dataset to obtain the desired system performance. The inputs of the DNN are the large-scale fading components $\overline{\mathbf{h}} = \left\{\overline{h}_{km}\right\}$, users' data size $\mathbf{D} = \{D_k\}$, and user latency requirements $\mathbf{T} = \{T_k\}$, while the output of the DNN is the user association scheme.

We utilize a exploitation and exploration policy to generate the dataset. Specifically, given $\overline{\mathbf{h}}, \mathbf{D}, \mathbf{T}$, we first obtain a user association scheme based on the nearest RSU association scheme. Then, we develop one-step exploration, where we change the association scheme of one of the K users while keeping other user associations the same. Since each user can access the other $M-1$ RSUs in the one-step exploration, there are $N_{one} = K\left(M-1\right)$ possible schemes in the whole one-step exploration. Next, N_{ran} user association schemes are generated with random exploration, where each user randomly selects one of M RSUs with probability $\frac{1}{M}$. Given each user association scheme, by optimizing $\mathbf{b}$, $\mathbf{f}$, and $\mathbf{p}$, we select the one with the lowest energy consumption from $1 + N_{one} + N_{ran}$ schemes as the output of the data pair.

In Fig. 4.3, we show the deep learning model for obtaining $\mathbf{a}$ in VEC systems. In the output layer, we obtain the output values with the value range of $[0, 1]$ by using Sigmoid functions. Then, we obtain the binary output values by selecting the RSU with the maximum output value for each user. Considering the difference of input values, we execute the data pre-

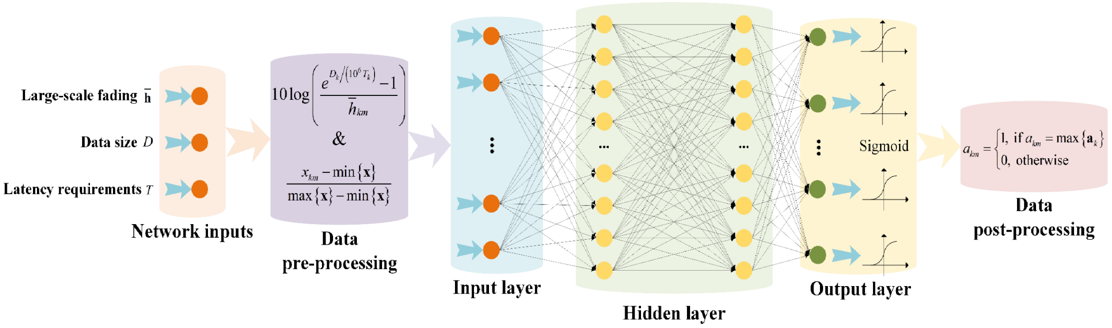


Fig 4.2: Deep learning model for obtaining user association **a** in VEC systems.

processing procedure, including the integration and normalization methods, as shown in Fig. 4.3. Specifically, for any k and m, we obtain $x_{km} = 10\log\left(\frac{e^{D_k/\left(10^6 T_k\right)}-1}{\bar{h}_{km}}\right)$ by considering the units of the variables. Then, we normalize the inputs based on $\frac{x_{km}-\min\{\mathbf{x}\}}{\max\{\mathbf{x}\}-\min\{\mathbf{x}\}}$ to scale input values between 0 and 1. The numbers of neurons in the input and output layers are equal to KM. After the training phase, we can use the trained DNN to calculate $\mathbf{a}$ for any $\bar{\mathbf{h}}, \mathbf{D}, \mathbf{T}$.

4.3.4 Algorithm, Convergence and Complexity

Algorithm

Algorithm 2 shows an iterative algorithm for solving problem (4.6), where $E^i = E\left(\mathbf{b}^i, \mathbf{p}^i, \mathbf{f}^i\right)$ represents the total energy consumption of users in the i-th iteration.

Algorithm 2 Energy-efficient RSU-assisted VEC algorithm.

```
Require:
    The tolerance $\varepsilon_{obj}$, the maximum iteration number $N_{itera}^{\max}$.
Ensure:
while $i \leqslant N_{itera}^{\max}$ do
    Updating iteration index $i = i + 1$;
    With fixed $\mathbf{p}^{(i-1)}$, jointly optimize $\mathbf{b}^{(i)}$ and $\mathbf{f}^{(i)}$ according to problem (4.7);
    With fixed $\mathbf{b}^{(i)}$ and $\mathbf{f}^{(i)}$, optimize $\mathbf{p}^{(i)}$ in problem (4.23);
    Obtain the objective value $E\left(\mathbf{b}^{i}, \mathbf{p}^{i}, \mathbf{f}^{i}\right)$;
    If $\left(E^{i-1} - E^{i}\right)/E^{i} \leqslant \varepsilon_{obj}$: Break; End If;
end while
return $(\mathbf{a}^{*}, \mathbf{b}^{*}, \mathbf{p}^{*}, \mathbf{f}^{*})$; The minimized total energy consumption of users $E^{*}$.
```

Given network conditions, obtain $\mathbf{a}$ with deep learning; — see step 1 above.

Convergence

Algorithm 2 has theoretical guarantee of convergence which is shown as follows

$$\begin{aligned} & E\left(\mathbf{b}^{i-1}, \mathbf{p}^{i-1}, \mathbf{f}^{i-1}\right) \overset{(a)}{=} E^{ub}\left(\mathbf{b}^{i-1}, \mathbf{p}^{i-1}, \mathbf{f}^{i-1}\right) \\ & \overset{(b)}{\geqslant} E^{ub}\left(\mathbf{b}^{i}, \mathbf{p}^{i-1}, \mathbf{f}^{i}\right) \overset{(c)}{\geqslant} E\left(\mathbf{b}^{i}, \mathbf{p}^{i-1}, \mathbf{f}^{i}\right) \overset{(d)}{\geqslant} E\left(\mathbf{b}^{i}, \mathbf{p}^{i}, \mathbf{f}^{i}\right), \end{aligned} \tag{4.26}$$

where E^{ub} denotes the total energy consumption of users based on problem (4.11), (a) holds since the first-order Taylor expansions in (4.10) is tight at given local points, (b) holds since $\mathbf{b}^{i}$ and $\mathbf{f}^{i}$ are jointly solved optimally in problem (4.11); (c) holds since E^{ub} is an upper bound of E with $\mathbf{b}^{i}$ and $\mathbf{f}^{i}$; (d) holds since $\mathbf{p}^{(i)}$ is the optimal solution to problem (4.23). Therefore, we have $E\left(\mathbf{b}^{i-1}, \mathbf{p}^{i-1}, \mathbf{f}^{i-1}\right) \geqslant E\left(\mathbf{b}^{i}, \mathbf{p}^{i}, \mathbf{f}^{i}\right)$. Since E is always positive, Algorithm 1 converges.

Complexity

Algorithm 2 incurs polynomial complexity in computation. The complexity to obtain $\mathbf{a}$ by DNN is $C_{DL} = \sum_{l=1}^{Layers}\left(n^{(l)}n^{(l-1)} + n^{(l)}\right) + KM$, where $n^{(l)}$ is the number of neurons including the bias unit in the l-th layer. To solve problem (4.7), the complexity of updating $\mathbf{c}$ and $\mathbf{f}$ in (4.12) is $O(K)$ due to the closed-form solutions based on Theorem 1. The

Table 4.1: Default Parameters Setup

Parameter	**Value**
K, M, p_k^{max}	10, 2, 0.5 W
Server's computation capacity f_m^{max}	1×10^{10} cycles/s
Data size D_k	[0.5,1.5] Mbits
Latency requirement T_k, time slot T	500 ms, 1 ms
Bit and cycle conversion c_{bc}	1×10^{3} cycles/bit
User's computation capacity f_{k0}	2×10^{9} cycles/s
Coefficients ρ, ς	1×10^{-27}, 3
Threshold of outage time slots μ_o	250
Noise spectral density n_0	-80 dBm
Path loss exponent α, bandwidth B	3, 0.5 MHz

complexity of updating dual variables is $O\left(K+M\right)$. The total complexity of problem (4.7) is $O\left(N_{\mathrm{bf}}\left(K+M\right)\right)$, where N_{bf} denotes the number of iterations of problem (4.7).

To solve problem (4.23), denote the complexity of inverse functions $\Phi^{-1}\left(\cdot\right)$ and $\Theta^{-1}\left(\cdot\right)$ as $O\left(\frac{1}{\varepsilon_1}\right)$ and $O\left(\frac{1}{\varepsilon_2}\right)$, respectively. Since the solution to (4.24) is in closed-form, the total complexity of problem (4.23) is $O\left(K\log_2\left(\frac{1}{\varepsilon_1\varepsilon_2}\right)\right)$.

The total complexity of Algorithm 2 is given by

$$O\left(N_{\mathrm{I}}\left(N_{\mathrm{bf}}\left(K+M\right)+K\log_2\left(\frac{1}{\varepsilon_1\varepsilon_2}\right)\right)+C_{DL}\right), \tag{4.27}$$

where N_{I} is the number of outer iterations of Algorithm 2.

4.4 Numerical Results and Discussions

In this section, numerical results are provided to assess the proposed algorithm for VEC systems. Default parameter settings are shown in Table I. The length of the road is 200 m and the number of lanes is 4. RSUs are equally spaced beside the road. Users are randomly distributed on the lanes. Given an average data size $D_{ave} = 1$ Mbits, user's data size randomly generated within $[0.5D_{ave}$,$1.5D_{ave}]$. For each setup, the result is averaged on

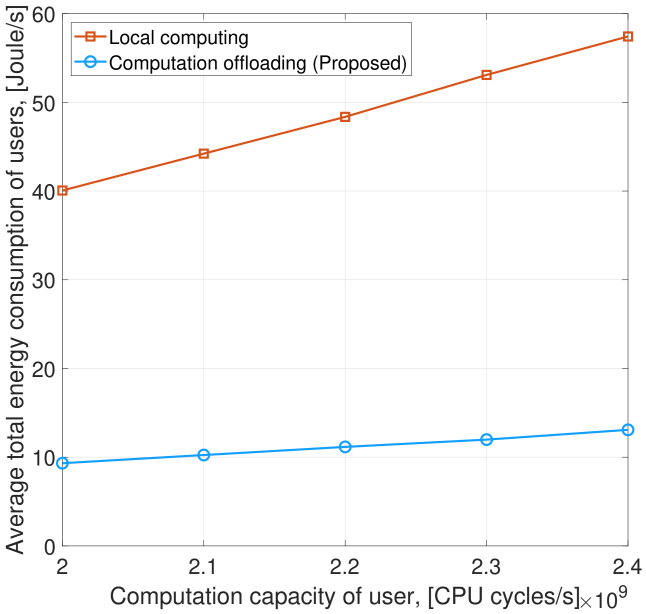


Figure 4.3: Average users' energy consumption versus f_{k0}.

1,000 tests. The learning rate is 0.01 with decay rate 0.1. The number of samples is 2,048. The number of neurons in hidden layer is 800. The size of mini-batch is 10. The number of random exploration N_{ran} is 50.

In Fig. 4.3, we compare the total energy consumption of users, i.e., E, versus user's computation capacity, i.e., f_{k0}, under various computing schemes. The developed energyefficient computation offloading scheme saves much energy for users compared to the traditional local computing scheme. Furthermore, it is observed that E increases with f_{k0} due to the dominated increased computation power consumption. Moreover, when f_{k0} becomes large, it is vital to offload computation tasks to edge servers to save users' energy.

In Fig. 4.4, we compare the average energy consumption of users, versus the number of users, under different methods. When K increases, users will consume more energy to execute their computation tasks due to the reduced allocated computation resource for each user. In addition, we compare the proposed deep learning method with other methods, in-

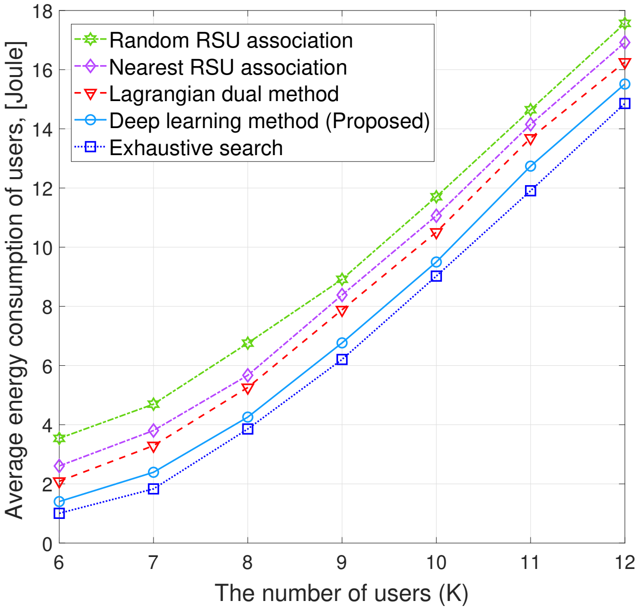


Figure 4.4: Average users' energy consumption versus K.

cluding exhaustive search, Lagrangian dual method [9], nearest RSU association, and random RSU association. It shows that the proposed deep learning-assisted computation offloading method approaches the optimal one (i.e., exhaustive search). Moreover, it achieves a smaller E than the other three methods. This is because the DNN model can capture the relationships between input and output data pairs, and thus it gives a near-optimal solution. Moreover, note that connecting to the nearest RSUs leads to reliable data rates, but the system neglects the cooperation among edge servers for load balancing.

In Fig. 4.5, we examine the different algorithms in terms of E. Specifically, we compare the proposed deep learning + joint $\mathbf{b}, \mathbf{p}, \mathbf{f}$ optimization (Opt) method (labeled as 'DL-Opt'), the deep learning + equal computation resource allocation (ECA) method (labeled as 'DL-ECA'), the Lagrangian dual + Opt method (labeled as 'Lag-Opt'), the Lagrangian dual + ECA method (labeled as 'Lag-ECA'), the random RSU association + Opt method (labeled as 'Ran-Opt'), the random RSU association + ECA method (labeled as 'Ran-ECA'). It is

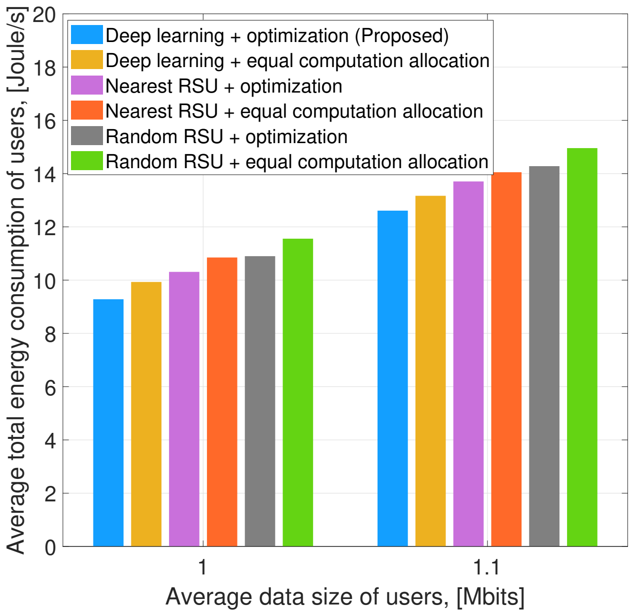


Figure 4.5: Average users' energy consumption versus D_{ave}.

observed that the proposed algorithm achieves the lowest energy consumption than other algorithms, which demonstrates the advantage of this work. Note that connecting to the nearest RSUs leads to reliable data rates, but the system neglects the cooperation among edge servers with load balancing. Moreover, it is observed that the proposed optimization algorithm saves much energy compared to the equal computation resource allocation algorithm, which validates the significance of this work.

4.5 Summary

In this chapter, we developed a deep learning-assisted energy-efficient computation offloading algorithm for VEC systems. The algorithm can solve the complex VEC problem and find a sub-optimal solution in a real-time manner with low complexity. Simulation results indicated the advantages of the proposed algorithm in substantially reducing users' total energy consumption compared with other methods.

Chapter 5

UAV Swarm-Enabled Aerial Reconfigurable Intelligent Surface

5.1 Introduction

Reconfigurable intelligent surface (RIS) has attracted significant attention in wireless networks, which is a planar surface comprising of a large number of low-cost passive reflecting elements. It improves communication performance by achieving fine-grained reflection-beamforming and reduces system energy consumption without requiring radio frequency (RF) chains [10, 13]. Moreover, with the full-duplex mode in operation without the addition of noise, RIS is more spectral-efficient than the conventional relay technology [11, 82, 83]. On the other hand, as a killer application in wireless networks, unmanned aerial vehicles (UAVs) can support various wireless services and are becoming more readily available nowadays [2, 84, 85, 86, 87]. UAVs have two main advantages, i.e., a) flexible placement and b) reliable air-ground communication links. As such, UAVs can be quickly placed in hotspots or during disasters to support reliable wireless communications.

Combining the above two promising technologies, i.e., RIS and UAVs, we can realize a new paradigm of aerial RIS (ARIS) to achieve efficient three-dimensional (3D) signal reflection, where RIS is mounted on a UAV. Compared to terrestrial RIS (TRIS), ARIS has the following three main advantages. First, ARIS is not restricted to the 180° half-space reflection such as TRIS, but instead, it provides a 360° panoramic full-angle reflection [12]. Second, with a UAV's agility and mobility in 3D space, ARIS is more flexible in placement than TRIS. Note that TRIS is usually placed on facades of a building or at a dedicated site requiring excessive site rental fees. Third, The path loss exponent in air-ground communications is usually smaller than that in terrestrial communications [35, 88]. Therefore, in most cases, the signal attenuation in ARIS's cascaded channel is usually smaller than that in TRIS's especially in the context of doubled path loss phenomena. To sum up, it is appealing from a practical standpoint to investigate the combination of UAVs and RIS, which shows the potential to improve system performance [89]. Although ARIS's location can be adjusted flexibly, there are some limitations for a single UAV-enabled ARIS in practical engineering. For example, due to the UAV's limited payload, ARIS usually has a restricted number of reflecting elements to guarantee flight stability and flexibility. In addition, deploying a single UAV usually incurs a low-rank channel matrix due to the dominated line-of-sight (LoS) air-ground connections. As a result, the significant aperture gain of a single UAV-enabled ARIS would not be guaranteed.

Compared to ARIS, UAV swarm-enabled ARIS (SARIS) system has the following advantages. First, SARIS increases the aperture gain by increasing the number of UAVs. Second, SARIS guarantees UAV flight stability and flexibility by allowing moderate-sized RIS on each UAV, especially under bad weather conditions or air turbulence. Third, SARIS supports spatial multiplexing for a large number of users by providing a rich scattering environment with different UAVs' positions. Forth, with the reduced RIS size on each UAV, the production cost of RIS can be decreased, and the flight time of the UAV can be prolonged.

In this chapter, we study a SARIS system, where multiple UAVs mounted with RIS assist the downlink transmissions between a base station (BS) and ground users. The beamforming and placement design for the SARIS system are investigated.

5.1.1 Related Works

For reflection in the sky, the combination between UAVs and RIS were investigated in [90, 91, 92, 93, 94, 95]. In general, these works fall into two categories. One is the TRIS-assisted UAV communications [90, 91, 92], and the other is the ARIS-assisted communications [93, 94, 95]. In [90], a joint UAV trajectory and TRIS passive beamforming design was investigated to maximize the ground user's average achievable rate. In [91], the authors leveraged the decaying deep Q-network to minimize ground users' energy consumption by jointly designing the UAV movement, the TRIS phase shift, and power allocation policy. In [92], a UAV was introduced to help the TRIS reflect its signals to the BS and enhance the UAV transmission by passive beamforming at the TRIS.

For ARIS-assisted communications, the transmitter and the receivers are on the ground. Thus, the channels (i.e., between transmitter and ARIS, and between ARIS to receivers) are air-ground channels. Furthermore, ARIS's position determines the array response in the LoS passive signal reflections. In [93], the authors utilized ARIS to maximize the worst-case signal-to-noise ratio (SNR) in a target area. However, the authors considered LoS air-ground connections. The NLOS connections influenced by ARIS's position were neglected. Moreover, the maximization of the worst-case SNR only considered the single-user beamforming. In [94], the authors investigated the ARIS-assisted uplink secure communications by optimizing the UAV trajectory, the ARIS phase shift, user association, and transmit power. However, ARIS's altitude was fixed in [94], and the NLOS air-ground connections and multiple antennas at the BS were not considered.

On the other hand, ARIS placement design plays a critical role in ARIS-assisted communications due to the significant far-field doubled path loss [96, 97]. The ARIS placement was studied in [93] where the ideal LoS air-ground channel was assumed. In [95], authors focused on the ARIS placement between fixed source and destination nodes. However, the random spatial distributions of ground users need to be considered in performance evaluation. Note that the existing works that integrated UAVs and RIS focused on single UAV networks. To the best of our knowledge, the multiple UAV-enabled ARIS systems have not been investigated. The formed UAV swarm network facilitates an increased aperture gain in passive signal reflections.

5.1.2 Contributions

The main contributions of our work are summarized in the following:

- *SARIS Beamforming Design:* A joint BS's transmit beamforming and SARIS's passive beamforming is designed. We aim to maximize the weighted sum-rate of ground users in the SARIS-assisted downlink communication system, subject to the maximum transmit power constraint and phase shift constraints. We utilize the alternating optimization technique to develop two beamforming schemes in SARIS systems, namely multi-user beamforming and weighted Round-Robin (WRR) single-user beamforming.

- *SARIS Placement Design:* We model and analyze the 3D SARIS system by considering the large-scale path loss and LoS/NLoS opportunistic connections. The randomness and correlations of UAVs' and users' positions are modeled by leveraging the tools from stochastic geometry. Specifically, the conditional probability density functions (PDFs) of the distance between the BS and a UAV, and the distance between a UAV and a user are derived, respectively. Moreover, we obtain the average channel power gain for an arbitrary user in SARIS-assisted downlink communication. The analytical model reflects the trade-off between the doubled path loss and the excessive NLoS path loss in the SARIS system, and the trade-off between the distance-dependent path loss and the excessive NLoS path loss.

- *Network Design Insights:* The analysis and simulation results provide network design insights. First, the multi-user beamforming scheme achieves a 1-30% performance gain compared to the WRR single-user beamforming scheme. However, the latter has a relatively lower computational complexity than the former. Second, the analytical model allows obtaining the optimal SARIS 3D position. With placement optimization, the SARIS system achieves a 20-500% performance gain compared to the system without placement optimization. Third, it is optimal to place SARIS above the BS when users are relatively near the BS. However, it is optimal to place SARIS "between" the BS and users when users are far from BS. Fourth, the channel estimation for NLoS components can be ignored to reduce system overhead at the cost of slight performance degradation.

5.1.3 Organization and Notations

The chapter is organized as follows. Chapter 5.2 presents the system model. Chapter 5.3 develops SARIS beamforming schemes. Chapter 5.4 derives an analytical framework for SARIS placement design. Simulation results are discussed in Chapter 5.5. Chapter 5.6 concludes the chapter.

The notations used in this chapter are listed in the following. For any matrix $\mathbf{A}$, $\mathbf{A}^{\dagger}$, $\mathbf{A}^{\mathrm{T}}$, and $\mathbf{A}^{\mathrm{H}}$ denote the conjugate, the transpose, and the conjugate transpose of $\mathbf{A}$, respectively. For a square matrix $\mathbf{B}$, $\mathbf{B}^{-1}$ denotes the inverse of $\mathbf{B}$. For any vector $\mathbf{x}$, $\|\mathbf{x}\|$ denotes the Euclidean norm. For any complex number x, $|x|$ and $\mathrm{Re}\left\{x\right\}$ denote the absolute value and the real part of x, respectively. $\mathbb{E}\left\{\cdot\right\}$ denotes statistical expectation.

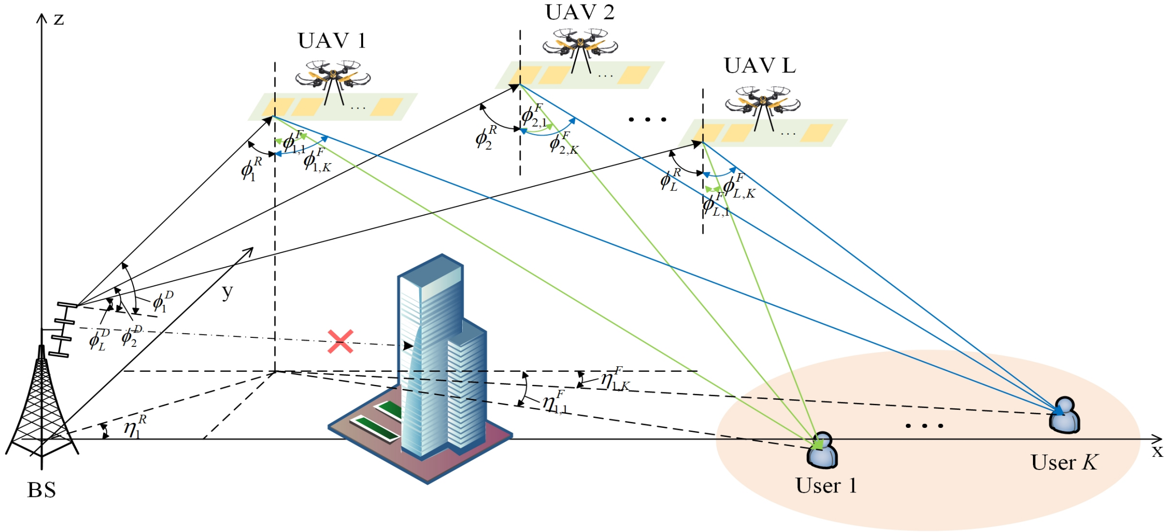


Fig. 5.1: An illustration of SARIS-assisted downlink communication system.

5.2 System Model

5.2.1 Network Layout

The network architecture is shown in Fig. 5.1. There are L ARIS forming a SARIS, and the set of ARISs is denoted by $\Psi_A = \{1, 2, \cdots, L\}$. Each ARIS comprises of a uniform linear array with N passive reflecting elements, separated by the distance $d_A < \lambda$, where λ is the carrier wavelength. The horizontal position of the l-th ARIS is $\mathbf{q}_l = [x_l^a, y_l^a]^{\mathrm{T}}$, where x_l^a is the coordinate of ARIS l on the x-axis and y_l^a is the coordinate of ARIS l on the y-axis. In this chapter, we consider the first reflection element of the AIRS as its reference point. The altitude of the SARIS is denoted by H. Considering the random locations of UAVs in a UAV swarm, we assume that the L ARISs are uniformly distributed in a circular region with the radius of R_A. The coordinate of the center of SARIS on the x-axis is denoted by x_A.

The BS is equipped with M antennas, and each antenna is separated by the distance $d_B < \lambda$. The BS is located at the origin of a 3D Cartesian coordinate system. The maximum allowable transmit power at the BS is $P_{\max}$. The distance between the BS and the l-th ARIS is given by

$$d_l^{\mathbf{G}} = \sqrt{\left\|\mathbf{q}_l\right\|^2 + H^2} = \sqrt{\left(x_l^a\right)^2 + \left(y_l^a\right)^2 + H^2}. \tag{5.1}$$

There are K single-antenna users randomly distributed in a ground circular region with the radius of R_U. The set of users is denoted by $\Psi_U = \{1, 2, \cdots, K\}$. The horizontal position of the k-th user is $\mathbf{u}_k = \left[x_k^u, y_k^u\right]^{\mathrm{T}}$, where x_k^u is the user's coordinate on x-axis and y_k^u is the user's coordinate on y-axis. The distance between the l-th ARIS and the k-th user is given by

$$d_{l,k}^{\mathbf{h}} = \sqrt{\left\|\mathbf{q}_l - \mathbf{u}_k\right\|^2 + H^2} = \sqrt{\left(x_l^a - x_k^u\right)^2 + \left(y_l^a - y_k^u\right)^2 + H^2}. \tag{5.2}$$

5.2.2 Channel Model

Let ϕ_l^R denote the zenith angle of arrival (AoA) of the signal transmitted from the BS to the l-th ARIS, i.e., the angle between the wave propagation direction and the positive z-axis, as illustrated in Fig. 5.1. Let η_l^R denote the azimuth AoA of the signal transmitted from the BS to the l-th ARIS, i.e., the angle between the horizontal projection of the wave propagation direction and the positive x-axis. Thus, the receive array response for the l-th ARIS is given by

$$\mathbf{a}_l^R = \left[1, e^{-j2\pi \frac{d_A}{\lambda} \overline{\phi}_l^R}, \cdots, e^{-j2\pi (N-1) \frac{d_A}{\lambda} \overline{\phi}_l^R}\right]^{\mathrm{T}}, \tag{5.3}$$

where $\overline{\phi}_l^R = \sin\left(\phi_l^R\right) \cos\left(\eta_l^R\right) = \frac{x_l^a}{d_l^{\mathbf{G}}}$, and $d_l^{\mathbf{G}}$ is given in (5.1).

Let ϕ_l^D denote the zenith angle of departure (AoD) of the signal transmitted from the BS to the l-th ARIS, as illustrated in Fig. 5.1, and we have $\sin\left(\phi_l^D\right) = \cos\left(\phi_l^R\right)$. The departure array response for the l-th ARIS is given by

$$\mathbf{a}_l^D = \left[1, e^{-j2\pi \frac{d_B}{\lambda} \overline{\phi}_l^D}, \cdots, e^{-j2\pi (M-1) \frac{d_B}{\lambda} \overline{\phi}_l^D}\right]^{\mathrm{T}}, \tag{5.4}$$

where $\overline{\phi}_l^D = \sin\left(\phi_l^D\right) = \cos\left(\phi_l^R\right) = \frac{H}{d_l^{\mathbf{G}}}$, and $d_l^{\mathbf{G}}$ is given in (5.1).

Denote the channel matrix between the BS and the l-th ARIS by $\mathbf{G}_l \in \mathbb{C}^{N\times M}$, and we have

$$\begin{aligned}\mathbf{G}_l &= \underbrace{\sqrt{\beta_l^{\mathbf{G}}}}_{\text{Path loss}}\left(\underbrace{\sqrt{\mu_{\text{LoS}}\mathbb{P}_l^{\text{LoS}}}\mathbf{G}_l^{\text{LoS}}}_{\text{LoS component}} + \underbrace{\sqrt{\mu_{\text{NLoS}}\mathbb{P}_l^{\text{NLoS}}}\mathbf{G}_l^{\text{NLoS}}}_{\text{NLoS component}}\right) \\ &= \sqrt{\beta_l^{\mathbf{G}}\mu_{\text{LoS}}\mathbb{P}_l^{\text{LoS}}}\mathbf{G}_l^{\text{LoS}} + \sqrt{\beta_l^{\mathbf{G}}\mu_{\text{NLoS}}\mathbb{P}_l^{\text{NLoS}}}\mathbf{G}_l^{\text{NLoS}}\end{aligned} \tag{5.5}$$

where $\beta_l^{\mathbf{G}} = \frac{\beta_0}{\|\mathbf{q}_l\|^2+H^2} = \frac{\beta_0}{\left(x_l^a\right)^2+\left(y_l^a\right)^2+H^2}$ is the large-scale distance-dependent path loss between the BS and the l-th ARIS, β_0 denotes the path loss at the reference distance, μ_{LoS} and μ_{NLoS} represent the excessive path loss in LoS and NLoS connections, respectively, since the signal attenuation in NLoS path is more severe than that in LoS path due to reflection, absorption, diffraction, and scattering. Moreover, $\mathbb{P}_l^{\text{LoS}}$ is the probability of the LoS connection between the BS and the l-th ARIS, which can be expressed as [35]

$$\mathbb{P}_l^{\text{LoS}} = \frac{1}{1+\eta_a \exp\left(-\eta_b\left(\phi_l^D - \eta_a\right)\right)}, \tag{5.6}$$

where η_a and η_b are constant values depending on the communications environment, e.g. rural, urban, or dense urban. In addition, the probability of the NLoS connection between the BS and the l-th ARIS is given by $\mathbb{P}_l^{\text{NLoS}} = 1-\mathbb{P}_l^{\text{LoS}}$. Note that both LoS and NLoS connections may exist in the air-ground propagation. We use the probabilities of occurrence of LoS and NLoS connections to obtain a spatial expectation of the path loss given a common elevation angle. Furthermore, in (5.5), $\mathbf{G}_l^{\text{LoS}}$ is the channel matrix for LoS connection between the BS and the l-th ARIS, which is given by $\mathbf{G}_l^{\text{LoS}} = e^{-j\frac{2\pi d_l^{\mathbf{G}}}{\lambda}}\mathbf{a}_l^R\left(\mathbf{a}_l^D\right)^{\text{H}}$ where $\mathbf{a}_l^R$ and $\mathbf{a}_l^D$ are given in (5.3) and (5.4), respectively. $\mathbf{G}_l^{\text{NLoS}}$ is the channel matrix for NLoS connection between the BS and the l-th ARIS, which has i.i.d. circularly symmetric complex Gaussian (CSCG) entries with zero mean and unit variance. This accounts for the small-scale fading which is assumed to be Rayleigh fading. In addition, the longer delay in NLoS path incurs a phase shift of the signal. When multiple independent NLoS signal paths combine at the receiver, the CSCG distribution captures the random phase shift. Note that $\mathbf{G}_l$ in (5.5) consists of the large-scale path loss, deterministic LoS component, uncertain NLoS component, and the occurrence probabilities of LoS and NLoS connections.

For the communication link between the l-th ARIS and the k-th user, let $\phi_{l,k}^F$ denote the zenith AoD of the signal and $\eta_{l,k}^F$ denote the azimuth AoD of the signal. Then, the reflect array response at the l-th ARIS is given by

$$\mathbf{a}_l^F = \left[1, e^{j2\pi\frac{d_A}{\lambda}\overline{\phi}_{l,k}^F}, \cdots, e^{j2\pi(N-1)\frac{d_A}{\lambda}\overline{\phi}_{l,k}^F}\right]^{\mathrm{T}}, \tag{5.7}$$

where $\overline{\phi}_{l,k}^F = \sin\left(\phi_{l,k}^F\right)\cos\left(\eta_{l,k}^F\right) = \frac{x_l^a - x_k^u}{d_{l,k}^{\mathbf{h}}}$, and $d_{l,k}^{\mathbf{h}}$ is given in (5.2).

Let $\mathbf{h}_{l,k}^{\mathrm{H}} \in \mathbb{C}^{1\times N}$ be the channel matrix between the l-th ARIS and the k-th user, and we have

$$\begin{aligned}\mathbf{h}_{l,k}^{\mathrm{H}} &= \underbrace{\sqrt{\beta_{l,k}^{\mathbf{h}}}}_{\text{Path loss}}\left(\underbrace{\sqrt{\mu_{\mathrm{LoS}}\mathbb{P}_{l,k}^{\mathrm{LoS}}}\mathbf{h}_{l,k}^{\mathrm{LoS}}}_{\text{LoS component}} + \underbrace{\sqrt{\mu_{\mathrm{NLoS}}\mathbb{P}_{l,k}^{\mathrm{NLoS}}}\mathbf{h}_{l,k}^{\mathrm{NLoS}}}_{\text{NLoS component}}\right)\\ &= \sqrt{\beta_{l,k}^{\mathbf{h}}\mu_{\mathrm{LoS}}\mathbb{P}_{l,k}^{\mathrm{LoS}}}\mathbf{h}_{l,k}^{\mathrm{LoS}} + \sqrt{\beta_{l,k}^{\mathbf{h}}\mu_{\mathrm{NLoS}}\mathbb{P}_{l,k}^{\mathrm{NLoS}}}\mathbf{h}_{l,k}^{\mathrm{NLoS}},\end{aligned} \tag{5.8}$$

where $\beta_{l,k}^{\mathbf{h}} = \frac{\beta_0}{\|\mathbf{q}_l - \mathbf{u}_k\|^2 + H^2} = \frac{\beta_0}{\left(x_l^a - x_k^u\right)^2 + \left(y_l^a - y_k^u\right)^2 + H^2}$ is the large-scale distance-dependent path loss between the l-th ARIS and the k-th user, and $\mathbb{P}_{l,k}^{\mathrm{LoS}}$ is the probability of the LoS connection between the l-th ARIS and the k-th user, which is given by

$$\mathbb{P}_{l,k}^{\mathrm{LoS}} = \frac{1}{1 + \eta_a \exp\left(-\eta_b\left(\frac{\pi}{2} - \phi_{l,k}^T - \eta_a\right)\right)}. \tag{5.9}$$

The probability of the NLoS connection between the l-th ARIS and the k-th user is expressed as $\mathbb{P}_{l,k}^{\mathrm{NLoS}} = 1 - \mathbb{P}_{l,k}^{\mathrm{LoS}}$. The channel matrix for the LoS connection in (5.8) is given by $\mathbf{h}_{l,k}^{\mathrm{LoS}} = e^{-j\frac{2\pi}{\lambda}d_{l,k}^{\mathbf{h}}}\left(\mathbf{a}_l^F\right)^{\mathrm{H}}$, while each element in $\mathbf{h}_{l,k}^{\mathrm{NLoS}}$ for the NLoS connection follows the CSCG distribution with zero mean and unit variance, i.e., $h_{l,k,n}^{\mathrm{NLoS}} \sim \mathcal{CN}(0,1)$.

The phase shift matrix of the l-th ARIS is $_l = \sqrt{\eta_R}diag\left[\theta_{l,1}, \cdots, \theta_{l,N}\right]$, which is a diagonal matrix, where η_R indicates the reflection efficiency and $\theta_{l,n} = e^{j\varphi_{l,n}}$ $(\varphi_{l,n} \in [0, 2\pi])$ [93, 98, 99]. The adopted RIS model assumes that each reflecting element individually alters the phase shift of the impinging electromagnetic wave. Other theoretical RIS models for wireless communications, such as the physics-based model and impedance network-based model [98], could also be integrated into SARIS systems.

5.2.3 Problem Formulation

Denote the data symbol for the k-th user by s_k. The transmitted signal at the BS is given by

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{w}_k s_k, \tag{5.10}$$

where $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$ is the transmit beamforming for the k-th user at the BS. The received signal at the k-th user is given by

$$y_k = \sum_{l=1}^{L} \mathbf{h}_{l,k}^{\mathrm{H}}\ {}_{l}\mathbf{G}_l \mathbf{x} + n_k, \tag{5.11}$$

where $y_k \in \mathbb{C}^{1\times 1}$, $n_k \sim \mathcal{CN}\left(0, {\sigma_0}^2\right)$ is the additive white Gaussian noise at the k-th user with the variance ${\sigma_0}^2$. The signal-to-interference-plus-noise ratio (SINR) of the k-th user is given by

$$\begin{aligned}\gamma_k &= \frac{\left|\sum\limits_{l=1}^{L} \mathbf{h}_{l,k}^{\mathrm{H}}\ {}_{l}\mathbf{G}_l \mathbf{w}_k\right|^2}{\sum\limits_{i=1, i\neq k}^{K} \left|\sum\limits_{l=1}^{L} \mathbf{h}_{l,k}^{\mathrm{H}}\ {}_{l}\mathbf{G}_l \mathbf{w}_i\right|^2 + {\sigma_0}^2} \\ &= \frac{\left|\overline{\mathbf{h}}_k^{\mathrm{H}} \mathbf{w}_k\right|^2}{\sum\limits_{i=1, i\neq k}^{K} \left|\overline{\mathbf{h}}_k^{\mathrm{H}} \mathbf{w}_i\right|^2 + {\sigma_0}^2},\end{aligned} \tag{5.12}$$

where $\overline{\mathbf{h}}_k^{\mathrm{H}} = \sum\limits_{l=1}^{L} \mathbf{h}_{l,k}^{\mathrm{H}}\ {}_{l}\mathbf{G}_l$ and $\overline{\mathbf{h}}_k^{\mathrm{H}} \in \mathbb{C}^{1\times M}$.

The users' weighted sum-rate maximization problem in SARIS system is shown as follows

$$\max_{\mathbf{W}, ,\mathbf{Q}} \sum_{k=1}^{K} \omega_k \log\left(1+\gamma_k\right) \tag{5.13}$$

$$\text{s.t. } \theta_{l,n} \in \mathcal{F},\ \forall l,\ \forall n, \tag{5.13a}$$

$$\sum_{k=1}^{K} \left\|\mathbf{w}_k\right\|^2 \leqslant P_{\max}, \tag{5.13b}$$

where $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_K] \in \mathbb{C}^{M\times K}$, $\ = \{\ _l\}_{\forall l \in \Psi_A}$, $\mathbf{Q} = \{\mathbf{q}_l\}_{\forall l \in \Psi_A}$, ω_k is the weight for the k-th user, and $\mathcal{F} = \{e^{j\varphi} | \, 0 \leqslant \varphi \leqslant 2\pi\}$. Note that in (5.13) beamforming provides the aperture gain and the SARIS placement impacts the large-scale fading. In the following, we study the beamforming and placement design for SARIS systems.

5.3 SARIS Beamforming Design

In this section, we utilize the alternating optimization approach to obtain the transmit beamforming at the BS and the passive beamforming at SARIS in downlink communications.

5.3.1 Problem Transformation

Based on [100], we can reformulate the problem (5.13) by introducing the auxiliary variables $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_K]^{\mathrm{T}}$, as follows

$$\max_{\mathbf{W}, ,\boldsymbol{\alpha}} \sum_{k=1}^{K} \omega_k \log\left(1+\alpha_k\right) \tag{5.14}$$

$$\text{s.t. } (M13a), (M13b),$$

$$\alpha_k \leqslant \gamma_k,\ \forall k. \tag{5.14a}$$

Problem (5.14) can be thought of as an outer optimization of $\{\mathbf{W},\ \}$ and an inner optimization of $\boldsymbol{\alpha}$ given fixed $\{\mathbf{W},\ \}$. Since the inner problem is a convex problem over $\boldsymbol{\alpha}$, the strong duality holds and it is equivalent to its dual problem as follows

$$\begin{aligned}&\min_{\boldsymbol{\lambda}} \max_{\boldsymbol{\alpha}} \ L(\boldsymbol{\lambda}, \boldsymbol{\alpha}) \\ &\text{where} \\ &L(\boldsymbol{\lambda}, \boldsymbol{\alpha}) = \sum_{k=1}^{K} \omega_k \log(1+\alpha_k) - \sum_{k=1}^{K} \lambda_k (\alpha_k - \gamma_k)\end{aligned} \tag{5.15}$$

is the Lagrangian function and $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \cdots, \lambda_K]^{\mathrm{T}}$ is the dual variable. Let $(\boldsymbol{\lambda}^*, \boldsymbol{\alpha}^*)$ be the saddle point of (5.15). We have $\alpha_k^* = \gamma_k$ and $\lambda_k^* = \frac{\omega_k}{1+\alpha_k^*} = \frac{\omega_k}{1+\gamma_k}$, by letting $\frac{\partial L(\boldsymbol{\lambda},\boldsymbol{\alpha})}{\partial \lambda_k} = 0$ and $\frac{\partial L(\boldsymbol{\lambda},\boldsymbol{\alpha})}{\partial \alpha_k} = 0$, respectively. Substituting λ_k^* into $L(\boldsymbol{\lambda}, \boldsymbol{\alpha})$ and combining with the outer optimization for $\{\mathbf{W},\ \}$, we have the reformulated objective function as follows

$$f_1(\mathbf{W},\ ,\boldsymbol{\alpha}) = \sum_{k=1}^{K} \omega_k \log(1+\alpha_k) - \sum_{k=1}^{K} \omega_k \alpha_k + \sum_{k=1}^{K} \frac{\omega_k (1+\alpha_k)\gamma_k}{1+\gamma_k}, \tag{5.16}$$

where $\alpha_k^* = \gamma_k,\ \forall k$. For a fixed $\boldsymbol{\alpha}$, optimizing $\mathbf{W}$ and   is equivalent to the following problem

$$\max_{\mathbf{W},} \ \sum_{k=1}^{K} \frac{\underline{\omega}_k \gamma_k}{1+\gamma_k} \tag{5.17}$$

$$\text{s.t. } (5.13a), (5.13b), \tag{5.17a}$$

where $\underline{\omega}_k = \omega_k (1+\alpha_k)$. Next, we apply the block coordinate method to optimize $\mathbf{W}$ and .

5.3.2 Transmit Beamforming

Given a fixed , the transmit beamforming problem is given by

$$\max_{\mathbf{W}} \sum_{k=1}^{K} \frac{\underline{\omega}_k \left| \overline{\mathbf{h}}_k^{\mathrm{H}} \mathbf{w}_k \right|^2}{\sum\limits_{i=1}^{K} \left| \overline{\mathbf{h}}_k^{\mathrm{H}} \mathbf{w}_i \right|^2 + {\sigma_0}^2} \tag{5.18}$$

$$\text{s.t. } (5.13b). \tag{5.18a}$$

Note that (5.18) is a multiple-ratio fractional programming problem. We utilize the quadratic transform proposed in [101] to reformulate the problem as

$$\max_{\mathbf{W},\boldsymbol{\beta}} \; f_2\left(\mathbf{W},\boldsymbol{\beta}\right) = \sum_{k=1}^{K} \left(2\sqrt{\underline{\omega}_k}\, \mathrm{Re}\left\{ \beta_k^{\dagger} \overline{\mathbf{h}}_k^{\mathrm{H}} \mathbf{w}_k \right\} - \left|\beta_k\right|^2 \left(\sum_{i=1}^{K} \left| \overline{\mathbf{h}}_k^{\mathrm{H}} \mathbf{w}_i \right|^2 + {\sigma_0}^2 \right) \right) \tag{5.19}$$

$$\text{s.t. } (5.13b). \tag{5.19a}$$

where $\boldsymbol{\beta} = \left[\beta_1, \beta_2, \cdots, \beta_K\right]^{\mathrm{T}}$ is a vector containing auxiliary variables. Considering that (5.19) is a biconvex optimization problem, we alternatively update $\mathbf{W}$ and $\boldsymbol{\beta}$.

It can be observed that for a given $\mathbf{W}$, the optimal $\boldsymbol{\beta}$ is given by.

$$\beta_k^{*} = \frac{\sqrt{\underline{\omega}_k}\, \overline{\mathbf{h}}_k^{\mathrm{H}} \mathbf{w}_k}{\sum\limits_{i=1}^{K} \left| \overline{\mathbf{h}}_k^{\mathrm{H}} \mathbf{w}_i \right|^2 + {\sigma_0}^2}, \;\; \forall k, \tag{5.20}$$

For a given $\boldsymbol{\beta}$, optimization problem for $\mathbf{W}$ is given by

$$\min_{\mathbf{W}} \sum_{k=1}^{K} \left(\left|\beta_k\right|^2 \left(\sum_{i=1}^{K} \left| \overline{\mathbf{h}}_k^{\mathrm{H}} \mathbf{w}_i \right|^2 + {\sigma_0}^2 \right) - 2\sqrt{\underline{\omega}_k}\, \mathrm{Re}\left\{ \beta_k^{\dagger} \overline{\mathbf{h}}_k^{\mathrm{H}} \mathbf{w}_k \right\} \right) \tag{5.21}$$

$$\text{s.t. } (5.13b). \tag{5.21a}$$

Problem (5.21) is a convex function. We use the Lagrangian dual method to obtain the optimal $\mathbf{W}$. The Lagrangian function of (5.21) is given by

$$\begin{aligned} L\left(\mathbf{W}, \mu_P\right) = & \sum_{k=1}^{K} \left(|\beta_k|^2 \left(\sum_{i=1}^{K} \left| \overline{\mathbf{h}}_k^{\mathrm{H}} \mathbf{w}_i \right|^2 + {\sigma_0}^2 \right) - 2\sqrt{\underline{\omega}_k}\, \mathrm{Re}\left\{ \beta_k^{\dagger} \overline{\mathbf{h}}_k^{\mathrm{H}} \mathbf{w}_k \right\} \right) \\ & + \mu_P \left(\sum_{k=1}^{K} \left\| \mathbf{w}_k \right\|^2 - P_{\max} \right), \end{aligned} \tag{5.22}$$

where μ_P is the dual variable for the power constraint. Based on (5.22), the dual problem of (5.21) is given by

$$\max_{\mu_P} \min_{\mathbf{W}} L\left(\mathbf{W}, \mu_P\right) \tag{5.23}$$

$$\text{s.t. } \mu_P \geqslant 0. \tag{5.23a}$$

Regarding the dual problem (5.23), given a μ_P, the optimal $\mathbf{w}_k$ is obtained by letting $\frac{\partial L(\mathbf{W}, \mu_P)}{\partial \mathbf{w}_k} = 0$, and we have

$$\mathbf{w}_k^{*} = \sqrt{\underline{\omega}_k} \beta_k \left(\mu_P \mathbf{I}_M + \sum_{i=1}^{K} |\beta_i|^2 \overline{\mathbf{h}}_i \overline{\mathbf{h}}_i^{\mathrm{H}} \right)^{-1} \overline{\mathbf{h}}_k, \tag{5.24}$$

where $\mathbf{I}_M$ denotes the $M \times M$ identity matrix. The updating procedure for μ_P is given by

$$\mu_P^{t_w+1} = \left[\mu_P^{t_w} + \lambda_P \left(\sum_{k=1}^{K} \left\| \mathbf{w}_k \right\|^2 - P_{\max} \right) \right]^{+}, \tag{5.25}$$

where $\mu_P^{t_w}$ denotes the value of μ_P in the t_w-th iteration in the optimization of $\mathbf{W}$, and λ_P is the updating step-size.

Based on (5.20), (5.24) and (5.25), we can obtain the optimal $\mathbf{W}$.

5.3.3 Passive Beamforming

Given fixed $\mathbf{W}$, the passive beamforming problem is given by

$$\max \sum_{k=1}^{K} \frac{\underline{\omega}_k \left| \sum\limits_{l=1}^{L} \boldsymbol{\theta}_l^{\mathrm{H}} \mathbf{v}_{l,k,k} \right|^2}{\sum\limits_{i=1}^{K} \left| \sum\limits_{l=1}^{L} \boldsymbol{\theta}_l^{\mathrm{H}} \mathbf{v}_{l,k,i} \right|^2 + {\sigma_0}^2} \tag{5.26}$$

$$\text{s.t. } (5.13a), \tag{5.26a}$$

where $\boldsymbol{\theta}_l = [\theta_{l,1}, \cdots, \theta_{l,N}]^{\mathrm{H}} \in \mathbb{C}^{N\times 1}$ and $\mathbf{v}_{l,k,i} = \sqrt{\eta_R} diag\left(\mathbf{h}_{l,k}^{\mathrm{H}}\right) \mathbf{G}_l \mathbf{w}_i \in \mathbb{C}^{N\times 1}$. Furthermore, (5.26) can be expressed as

$$\max_{\underline{\boldsymbol{\theta}}} \sum_{k=1}^{K} \frac{\underline{\omega}_k \left| \underline{\boldsymbol{\theta}}^{\mathrm{H}} \underline{\mathbf{v}}_{k,k} \right|^2}{\sum\limits_{i=1}^{K} \left| \underline{\boldsymbol{\theta}}^{\mathrm{H}} \underline{\mathbf{v}}_{k,i} \right|^2 + {\sigma_0}^2} \tag{5.27}$$

$$\text{s.t. } (5.13a), \tag{5.27a}$$

where $\underline{\boldsymbol{\theta}} = \mathrm{vec}\left(\widetilde{\boldsymbol{\theta}}\right) \in \mathbb{C}^{NL\times 1}$, $\widetilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_L] \in \mathbb{C}^{N\times L}$, $\mathrm{vec}\left(\cdot\right)$ denotes the vectorization manipulation, $\underline{\mathbf{v}}_{k,i} = \mathrm{vec}\left(\widetilde{\mathbf{v}}_{k,i}\right) \in \mathbb{C}^{NL\times 1}$, and $\widetilde{\mathbf{v}}_{k,i} = [\mathbf{v}_{1,k,i}, \cdots, \mathbf{v}_{L,k,i}] \in \mathbb{C}^{N\times L}$.

Note that (5.27) is a multiple-ratio fractional programming problem. Based on the quadratic transform, we reformulate the problem (5.27) as follows

$$\max_{\underline{\boldsymbol{\theta}}, \boldsymbol{\delta}} f_3\left(\underline{\boldsymbol{\theta}}, \boldsymbol{\delta}\right) = \sum_{k=1}^{K} \left(2\sqrt{\underline{\omega}_k}\, \mathrm{Re}\left\{ \delta_k^{\dagger} \underline{\boldsymbol{\theta}}^{\mathrm{H}} \underline{\mathbf{v}}_{k,k} \right\} - \left|\delta_k\right|^2 \left(\sum_{i=1}^{K} \left| \underline{\boldsymbol{\theta}}^{\mathrm{H}} \underline{\mathbf{v}}_{k,i} \right|^2 + {\sigma_0}^2 \right) \right) \tag{5.28}$$

$$\text{s.t. } (5.13a), \tag{5.28a}$$

where $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_K]^{\mathrm{T}}$ consists of auxiliary variables.

It can be observed from (5.28) that for a given $\underline{\boldsymbol{\theta}}$, the optimal $\boldsymbol{\delta}$ is obtained by

$$\delta_k^* = \frac{\sqrt{\underline{\omega}_k}\underline{\boldsymbol{\theta}}^{\mathrm{H}}\underline{\mathbf{v}}_{k,k}}{\sum\limits_{i=1}^{K}\left|\underline{\boldsymbol{\theta}}^{\mathrm{H}}\underline{\mathbf{v}}_{k,i}\right|^2 + {\sigma_0}^2}. \tag{5.29}$$

For a given $\boldsymbol{\delta}$, the problem for $\underline{\boldsymbol{\theta}}$ is expressed as

$$\max_{\underline{\boldsymbol{\theta}}} \quad -\underline{\boldsymbol{\theta}}^{\mathrm{H}}\mathbf{A}\underline{\boldsymbol{\theta}} + 2\,\mathrm{Re}\left\{\underline{\boldsymbol{\theta}}^{\mathrm{H}}\mathbf{b}\right\} + c \tag{5.30}$$

$$\text{s.t. } (5.13a), \tag{5.30a}$$

where $\mathbf{A} = \sum\limits_{k=1}^{K}|\delta_k|^2\sum\limits_{i=1}^{K}\underline{\mathbf{v}}_{k,i}\underline{\mathbf{v}}_{k,i}^{\mathrm{H}} \in \mathbb{C}^{NL\times NL}$, $\mathbf{b} = \sum\limits_{k=1}^{K}\sqrt{\underline{\omega}_k}\delta_k^{\dagger}\underline{\mathbf{v}}_{k,k} \in \mathbb{C}^{NL\times 1}$, and $c = -\sum\limits_{k=1}^{K}|\delta_k|^2{\sigma_0}^2$. In the following, we iteratively optimize one of the NL reflection coefficients in $\underline{\boldsymbol{\theta}}$ while keeping others fixed. With some linear manipulations, the subproblem for optimizing the n-th element in $\underline{\boldsymbol{\theta}}$, i.e., $\underline{\theta}_n$, is given by

$$\max_{\underline{\theta}_n} \quad -\underline{\theta}_n^{\dagger}\mathbf{A}_{n,n}\underline{\theta}_n + 2\,\mathrm{Re}\left\{\underline{\theta}_n^{\dagger}\mathbf{b}_n - \sum_{j=1,j\neq n}^{NL}\underline{\theta}_n^{\dagger}\mathbf{A}_{n,j}\underline{\theta}_j\right\} \tag{5.31}$$

$$\text{s.t. } \underline{\theta}_n \in \mathcal{F}, \tag{5.31a}$$

where $\mathbf{A}_{n,n}$ denotes the n-th row and m-th column element in $\mathbf{A}$, $\mathbf{b}_n$ is the n-th element in $\mathbf{b}$. It can be derived that the optimal $\underline{\theta}_n$ is given by

$$\underline{\theta}_n^* = \exp\left(j\angle\frac{\varpi_n}{|\varpi_n|}\right), \text{where } \varpi_n = \mathbf{b}_n - \sum_{j=1,j\neq n}^{NL}\mathbf{A}_{n,j}\underline{\theta}_j. \tag{5.32}$$

Based on (5.32), we can optimize all the reflection coefficients in SARIS iteratively. Note that this iterative updating method guarantees convergence and has a relatively low complexity compared to the conventional semidefinite relaxation (SDR) method [96]. Specifically, updating (5.32) has the complexity of $O\left(NL\right)$, and the complexity of overall updating is $O\left(N^2L^2\right)$. However, the complexity of SDR method is $O\left(N^6L^6\right)$.

Complexity: We use bid-O notation to represent the computational complexity of the AO method. The computational complexity of updating $\boldsymbol{\beta}$ is $O\left(KNLM\right)$. Since updating $\mathbf{w}_k$ requires the matrix inversion, the computational complexity of updating $\mathbf{w}_k$ is $O\left(KM^3 I_{\mathbf{w}}\right)$, where $I_{\mathbf{w}}$ is the number of iterations for updating $\mathbf{w}_k$. Moreover, the computational complexity of updating $\boldsymbol{\delta}$ is $O\left(KNLM\right)$. The complexity for obtaining $\underline{\theta}_n^*$ is $O\left(NL\right)$. Since we need to solve all the reflecting elements, the computational complexity of updating $\underline{\boldsymbol{\theta}}$ is $O\left(N^2 L^2 I_{\boldsymbol{\theta}}\right)$, where $I_{\boldsymbol{\theta}}$ is the iteration number of updating $\underline{\boldsymbol{\theta}}$. The total complexity of the AO method is $O\left(I_0\left(2KNLM + KM^3 I_{\mathbf{w}} + N^2 L^2 I_{\theta}\right)\right)$, where I_0 is the number of iterations.

Convergence: denote $f_1\left(\mathbf{W}^t,\ {}^t, \alpha^t\right)$ as the objective function in (M14) at the t-th iteration. Then, we have

$$f_1\left(\mathbf{W}^t,\ {}^t,\boldsymbol{\alpha}^t\right) \le f_1\left(\mathbf{W}^{t+1},\ {}^t,\boldsymbol{\alpha}^t\right) \le f_1\left(\mathbf{W}^{t+1},\ {}^{t+1},\boldsymbol{\alpha}^t\right) \le f_1\left(\mathbf{W}^{t+1},\ {}^{t+1},\boldsymbol{\alpha}^{t+1}\right), \tag{5.33}$$

since we optimize the block in each subproblem. Due to the maximum allowable power constraint, the users' SINRs are bounded. Thus, $f_1\left(\mathbf{W}^t,\ {}^t, \alpha^t\right)$ converges when t increases.

Optimality: Given a fixed , the subproblem (M17) for optimizing the variables $\mathbf{W}$ and $\boldsymbol{\beta}$ is a biconvex optimization problem, and each of variables is uniquely and optimally obtained. Since the function $f_2\left(\mathbf{W}, \boldsymbol{\beta}\right)$ is bounded from above and continuous, and the set of $\mathbf{W}$ is a compact set, the obtained solution $\left(\mathbf{W}^*, \boldsymbol{\beta}^*\right)$ is a partial optimum [102]. Similarly, given a fixed $\mathbf{W}$, $\left(\underline{\boldsymbol{\theta}}^*, \boldsymbol{\delta}^*\right)$ is a partial optimum for the subproblem (M26). Moreover, since $\mathbf{W}^*$ and * are uniquely attained in the subproblems, respectively, the converged solution $\left(\mathbf{W}^*,\ {}^*\right)$ is suboptimal and stationary [103].

5.3.4 Single-User Beamforming in SARIS Systems

In this subsection, we develop a WRR single-user beamforming scheme for SARIS systems. The single-user beamforming has closed-form solutions and achieves a low computational complexity.

In a single-user case, the transmitted signal for the k-th user at the BS is given by

$$\mathbf{x}_k = \sqrt{P_{\max}}\mathbf{w}_k s_k, \tag{5.34}$$

where $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$ is the transmit beamforming for the k-th user at the BS. For the k-th user, the optimization problem of $\mathbf{w}_k$ and is given by

$$\max_{\mathbf{w}_k,} \ \left|\overline{\mathbf{h}}_k^{\mathrm{H}}\mathbf{w}_k\right|^2 \tag{5.35}$$

$$\text{s.t. } (5.13a), \tag{5.35a}$$

$$\|\mathbf{w}_k\|^2 \leqslant 1, \tag{5.35b}$$

where $\overline{\mathbf{h}}_k^{\mathrm{H}} = \sum\limits_{l=1}^{L}\mathbf{h}_{l,k}^{\mathrm{H}}\ {}_l\mathbf{G}_l$.

The optimal transmit beamforming for the k-th user is given by

$$\mathbf{w}_k^* = \frac{\overline{\mathbf{h}}_k}{\left\|\overline{\mathbf{h}}_k^{\mathrm{H}}\right\|} = \frac{\sum\limits_{l=1}^{L}\mathbf{G}_l^{\mathrm{H}}\ {}_l^{\mathrm{H}}\mathbf{h}_{l,k}}{\left\|\sum\limits_{l=1}^{L}\mathbf{h}_{l,k}^{\mathrm{H}}\ {}_l\mathbf{G}_l\right\|}, \tag{5.36}$$

which is obtained by maximum ratio transmission (MRT) precoding. In addition, the passive beamforming problem is given by

$$\max_{\underline{\boldsymbol{\theta}}} \ \left|\underline{\boldsymbol{\theta}}^{\mathrm{H}}\underline{\mathbf{v}}_{k,k}\right|^2 \tag{5.37}$$

$$\text{s.t. } \left|\underline{\theta}_n\right| = 1, \ \forall n \in \{1,\cdots,NL\}, \tag{5.37a}$$

To solve (5.37), we iteratively optimize one of the NL reflection coefficients in $\underline{\boldsymbol{\theta}}$ while keeping others fixed. Thus, the optimal n-th phase shift on the l-th ARIS is given by

$$\theta_{l,n}^* = \exp\left(\varphi_0 - \angle\left(h_{l,k,n}^{\mathrm{H}}\mathbf{g}_{l,n}^{\mathrm{H}}\mathbf{w}_k\right)\right) = \exp\left(\varphi_0 - \angle\left(h_{l,k,n}^{\mathrm{H}}\right) - \angle\left(\mathbf{g}_{l,n}^{\mathrm{H}}\mathbf{w}_k\right)\right), \tag{5.38}$$

where $\varphi_0 = \angle\left(\underline{\theta}^{\mathrm{H}}\underline{\mathbf{v}}_{k,k}\right)$ is a constant, $h^{\mathrm{H}}_{l,k,n} \in \mathbb{C}^{1\times 1}$ is the n-th element of $\mathbf{h}^{H}_{l,k}$, $\mathbf{g}^{\mathrm{H}}_{l,n} \in \mathbb{C}^{1\times M}$ is the n-th row of $\mathbf{G}_l$.

Based on (5.36) and (5.38), we can obtain the optimal $\mathbf{W}$ and . Moreover, the users are scheduled based on the WRR policy, where the weight is ω_k for the k-th user. Note that updating (5.38) has the complexity of $O\left(M\right)$, and we need to solve NL times for all the NL reflection coefficients in $\underline{\boldsymbol{\theta}}$. Thus, the overall complexity of updating is $O\left(MNL\right)$. In addition, updating $\mathbf{W}$ is based on a closed-form solution, which has relatively low complexity compared to the iterative methods for multi-user beamforming.

5.4 SARIS Placement Design

In the previous section, we study two beamforming schemes for the SARIS system given the air-ground channel conditions. In practical engineering, the SARIS placement problem is also significant in the system design, which involves large-scale fading and thus transmission distance distributions. In this section, we develop a theoretical framework to analyze the impact of SARIS placement and nodes distributions on the overall system performance. The closed-form beamforming strategies in the previous section will be used to derive the average channel power gain for an arbitrary user in SARIS systems.

5.4.1 Intermediate Results

In this subsection, we obtain some intermediate results that will be used to derive the average channel power gain for an arbitrary user in the SARIS-assisted downlink communication system.

We consider a scenario that SARIS is placed to serve ground users in a circular region, as shown in Fig. 5.2. Users could be located in hotspots or disaster regions, whereas the direct links between the BS and users are blocked due to obstacles. We utilize a Poisson cluster process (PCP) to model the spatial distribution of users. The average number of users is denoted by n_U. In this chapter, we consider Matérn cluster process, which assumes symmetric uniform spatial distribution of users around a cluster center within a circular disc of radius R_U. The PDF of each user is given by

$$f_{\mathbf{U}}\left(\mathbf{u}\right)=\begin{cases}\frac{1}{\pi {R_U}^2},\ \text{if}\ \left\|\mathbf{u}\right\|\leqslant R_U\\ 0,\ \text{otherwise}\end{cases},\tag{5.39}$$

where $\mathbf{u}$ is a realization of random vector $\mathbf{U}$. As shown in Fig. 5.2, the coordinate of the users' cluster center is denoted by $c_U\left(x_U,0,0\right)$, where x_U is the distance between the BS and c_U.

Considering that a UAV swarm usually has a dynamic network topology and sometimes without flight formation, we use random spatial models to capture UAV locations. The PDF of each UAV is given by

$$f_{\mathbf{A}}\left(\mathbf{a}\right)=\begin{cases}\frac{1}{\pi {R_A}^2},\ \text{if}\ \left\|\mathbf{a}\right\|\leqslant R_A\\ 0,\ \text{otherwise}\end{cases},\tag{5.40}$$

where $\mathbf{a}$ is a realization of random vector $\mathbf{A}$, and R_A is UAV swarm radius. The SARIS coordinate on the x-axis is denoted by $c_A\left(x_A,0,0\right)$, where x_A is the distance between the BS and c_A.

Lemma 5.1. *Given the distance between the BS and SARIS coordinate on x-axis, i.e., x_A, and the SARIS flight height H, the conditional PDF of the distance between the BS and an*

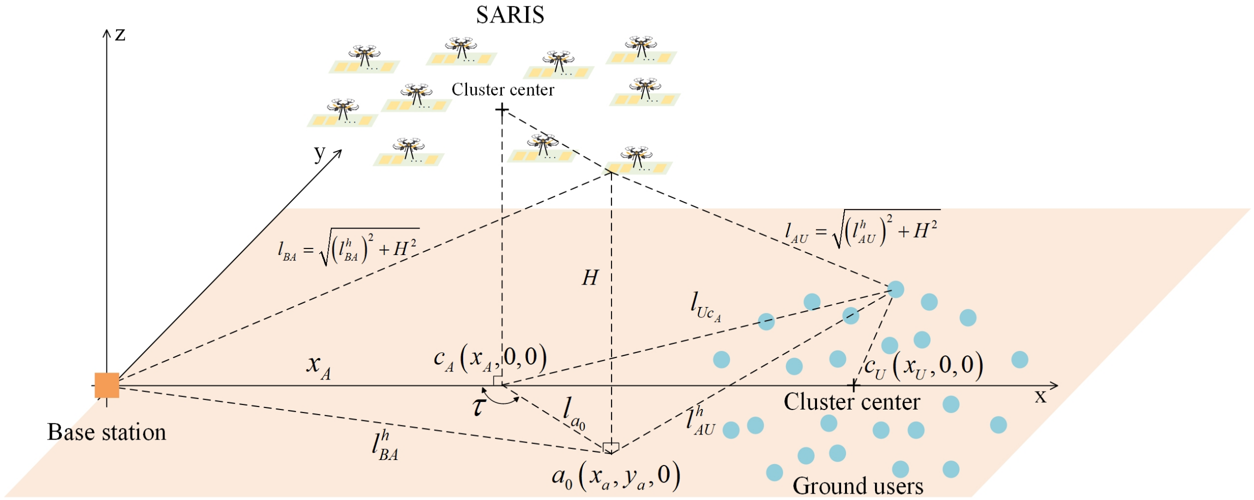


Fig 5.2: An illustration of SARIS 3D coordinate system.

arbitrary ARIS, i.e., l_{BA}*, is given by*

$$\begin{aligned} & f_{L_{BA}}\left(l_{BA}\right|x_A,H)=\frac{2l_{BA}}{\pi{R_A}^2}\Omega\left(x_A,\sqrt{{l_{BA}}^2-H^2},R_A\right), \\ & where\ \Omega\left(a,b,c\right)=\arccos\left(\frac{a^2+b^2-c^2}{2ab}\right), \\ & \sqrt{\left(\max\left\{0,x_A-R_A\right\}\right)^2+H^2}\leqslant l_{BA}\leqslant\sqrt{\left(x_A+R_A\right)^2+H^2}. \end{aligned} \tag{5.41}$$

Proof. Please refer to Appendix D.1 for the proof. □

Lemma 5.2. *Given the distance between the projection of the SARIS coordinate on the ground and the users' cluster center, i.e.,* $x_{AU}=x_U-x_A$*, and the SARIS flight height* H*, the conditional PDF of the distance between an arbitrary ARIS and an arbitrary user, i.e.,* l_{AU}*, is given by*

$$\begin{aligned} & if\ x_{AU}\geqslant R_U,\ then \\ & f_{L_{AU}}\left(l_{AU}\right|x_A,H)=\frac{4l_{AU}}{\pi^2{R_A}^2{R_U}^2}\int_{x_{AU}-R_U}^{x_{AU}+R_U}x\Omega\left(x,\sqrt{l_{AU}^2-H^2},R_A\right)\Omega\left(x_U-x_A,x,R_U\right)dx, \\ & where\ \ \sqrt{\left(x_{AU}-\left(R_A+R_U\right)\right)^2+H^2}\leqslant l_{AU}\leqslant\sqrt{\left(x_{AU}+R_A+R_U\right)^2+H^2}, \end{aligned} \tag{5.42}$$

$$\begin{aligned}
&\textit{and if } x_{AU} < R_U, \ \textit{then} \\
&f_{L_{AU}}\left(l_{AU} \middle| x_A, H\right) = \frac{4 l_{AU}}{\pi {R_A}^2 {R_U}^2} \int_0^{R_U - x_{AU}} x\Omega\left(x, \sqrt{l_{AU}^2 - H^2}, R_A\right) dx \\
&+ \frac{4 l_{AU}}{\pi^2 {R_A}^2 {R_U}^2} \int_{R_U - x_{AU}}^{x_{AU} + R_U} x\Omega\left(x, \sqrt{l_{AU}^2 - H^2}, R_A\right) \Omega\left(x_U - x_A, x, R_U\right) dx, \\
&\textit{where} \quad H \leqslant l_{AU} \leqslant \sqrt{\left(x_{AU} + R_A + R_U\right)^2 + H^2}.
\end{aligned} \tag{5.43}$$

Proof. Please refer to Appendix D.2 for the proof. ☐

5.4.2 Performance Evaluation

For ease of analysis, we consider the single-user case in the SARIS system for performance evaluation. The approximate absolute value of the k-th user's cascaded channel is given by

$$\begin{aligned}
\left| \sum_{l=1}^{L} \mathbf{h}_{l,k}^{\mathrm{H}} {}_{l} \mathbf{G}_l \mathbf{w}_k \right| &= \left| \sum_{l=1}^{L} \sum_{n=1}^{N} h_{l,k,n}^{\mathrm{H}} \theta_{l,n} \mathbf{g}_{l,n}^{\mathrm{H}} \mathbf{w}_k \right| \\
&\leqslant \sum_{l=1}^{L} \left| \sum_{n=1}^{N} h_{l,k,n}^{\mathrm{H}} \theta_{l,n} \mathbf{g}_{l,n}^{\mathrm{H}} \mathbf{w}_k \right| = \sum_{l=1}^{L} \sum_{n=1}^{N} \sqrt{\eta_R} \left| h_{l,k,n}^{\mathrm{H}} \right| \left| \mathbf{g}_{l,n}^{\mathrm{H}} \mathbf{w}_k \right| \\
&\overset{(a)}{\approx} \sum_{l=1}^{L} \sum_{n=1}^{N} \sqrt{\eta_R} \Upsilon_{l,k}^{AU} \left| \mathbf{g}_{l,n}^{\mathrm{H}} \mathbf{w}_k \right| = \sum_{l=1}^{L} \sqrt{\eta_R} \Upsilon_{l,k}^{AU} \sum_{n=1}^{N} \left| \mathbf{g}_{l,n}^{\mathrm{H}} \mathbf{w}_k \right| \\
&\overset{(b)}{\leqslant} \sum_{l=1}^{L} \sqrt{\eta_R} \Upsilon_{l,k}^{AU} \sum_{n=1}^{N} \left\| \mathbf{g}_{l,n}^{\mathrm{H}} \right\| \left\| \mathbf{w}_k \right\| \\
&\overset{(c)}{\approx} \sum_{l=1}^{L} \sqrt{\eta_R} \Upsilon_{l,k}^{AU} \sum_{n=1}^{N} \Upsilon_l^{BA} \left| \sum_{m=1}^{M} e^{j\omega_{n,m}^{G}} \right| \left\| \frac{\sum_{l=1}^{L} \mathbf{G}_l^{\mathrm{H}} {}_l^{\mathrm{H}} \mathbf{h}_{l,k}}{\left\| \sum_{l=1}^{L} \mathbf{h}_{l,k}^{\mathrm{H}} {}_{l} \mathbf{G}_l \right\|} \right\| \\
&= \sum_{l=1}^{L} \sqrt{\eta_R} \Upsilon_{l,k}^{AU} \Upsilon_l^{BA} N \sqrt{M},
\end{aligned} \tag{5.44}$$

$$\begin{aligned}
\text{where } \Upsilon_{l,k}^{AU} &= \sqrt{\beta_{l,k}^{\mathbf{h}}} \left(\sqrt{\mu_{\mathrm{LoS}} \mathbb{P}_{l,k}^{\mathrm{LoS}}} + \sqrt{\mu_{\mathrm{NLoS}} \mathbb{P}_{l,k}^{\mathrm{NLoS}}} \right), \\
\Upsilon_l^{BA} &= \sqrt{\beta_l^{\mathbf{G}}} \left(\sqrt{\mu_{\mathrm{LoS}} \mathbb{P}_l^{\mathrm{LoS}}} + \sqrt{\mu_{\mathrm{NLoS}} \mathbb{P}_l^{\mathrm{NLoS}}} \right).
\end{aligned}$$

In (5.44), (a) is obtained based on (5.8) where $\mathbf{h}_{l,k}^{\mathrm{H}} = \left\{\left|h_{l,k,n}^{\mathrm{H}}\right| e^{j\omega_n^h}\right\}_{n=1,\cdots,N}$. Denote $J_{l,k}^{\mathrm{LoS}} = \sqrt{\beta_{l,k}^{\mathbf{h}}\mu_{\mathrm{LoS}}\mathbb{P}_{l,k}^{\mathrm{LoS}}}$ and $J_{l,k}^{\mathrm{NLoS}} = \sqrt{\beta_{l,k}^{\mathbf{h}}\mu_{\mathrm{NLoS}}\mathbb{P}_{l,k}^{\mathrm{NLoS}}}$. Specifically, we have $\left|h_{l,k,n}^{\mathrm{H}}\right| \leqslant \left|J_{l,k}^{\mathrm{NLoS}}h_{l,k,n}^{\mathrm{NLoS}}\right| + \left|J_{l,k}^{\mathrm{LoS}}e^{-j\frac{2\pi}{\lambda}\left(d_{l,k}^{\mathbf{h}}+(n-1)d_A\bar{\phi}_{l,k}^F\right)}\right|$. Since $\mathbb{E}_h\left|J_{l,k}^{\mathrm{NLoS}}h_{l,k,n}^{\mathrm{NLoS}}\right| = J_{l,k}^{\mathrm{NLoS}}\mathbb{E}\left|h_{l,k,n}^{\mathrm{NLoS}}\right| > J_{l,k}^{\mathrm{NLoS}}$ and the modulus value $\left|e^{-j\frac{2\pi}{\lambda}\left(d_{l,k}^{\mathbf{h}}+(n-1)d_A\bar{\phi}_{l,k}^F\right)}\right| = 1$, we have $\left|h_{l,k,n}^{\mathrm{H}}\right| \approx J_{l,k}^{\mathrm{LoS}} + J_{l,k}^{\mathrm{NLoS}}$. (b) follows the Cauchy–Schwarz inequality. Moreover, (c) is obtained based on (5.5), where $\mathbf{G}_l = \left\{\left|G_{l,n,m}\right| e^{j\omega_{n,m}^G}\right\}$ for $n = 1,\cdots,N,\ m = 1,\cdots,M$ and $\left|G_{l,n,m}\right| \approx \sqrt{\beta_l^{\mathbf{G}}\mu_{\mathrm{LoS}}\mathbb{P}_l^{\mathrm{LoS}}} + \sqrt{\beta_l^{\mathbf{G}}\mu_{\mathrm{NLoS}}\mathbb{P}_l^{\mathrm{NLoS}}}$. The transmit beamforming $\mathbf{w}_k$ is obtained based on (5.36) for the k-th user.

The average channel power gain for an arbitrary user in SARIS systems is given by

$$\begin{aligned}
&\mathbb{E}\left\{\left|\sum_{l=1}^{L}\mathbf{h}_{l,k}^{\mathrm{H}}\ {}_l\mathbf{G}_l\mathbf{w}_k\right|^2\right\} \overset{(a)}{\leqslant} \mathbb{E}\left\{\left|\sum_{l=1}^{L}\sqrt{\eta_R}\Upsilon_{l,k}^{AU}\Upsilon_l^{BA}N\sqrt{M}\right|^2\right\} \\
&= \eta_R L^2N^2M\mathbb{E}_{\Psi_A,\Psi_U}\left\{\left|\Upsilon_{l,k}^{AU}\Upsilon_l^{BA}\right|^2\right\}, \\
&= \eta_R L^2N^2M\mathbb{E}_{\Psi_A,\Psi_U}\left\{\beta_{l,k}^{\mathbf{h}}\beta_l^{\mathbf{G}}\left(\mu_{\mathrm{LoS}}\mathbb{P}_{l,k}^{\mathrm{LoS}} + \mu_{\mathrm{NLoS}}\mathbb{P}_{l,k}^{\mathrm{NLoS}} + 2\sqrt{\mu_{\mathrm{LoS}}\mathbb{P}_{l,k}^{\mathrm{LoS}}\mu_{\mathrm{NLoS}}\mathbb{P}_{l,k}^{\mathrm{NLoS}}}\right)\right. \\
&\quad\left.\cdot\left(\mu_{\mathrm{LoS}}\mathbb{P}_l^{\mathrm{LoS}} + \mu_{\mathrm{NLoS}}\mathbb{P}_l^{\mathrm{NLoS}} + 2\sqrt{\mu_{\mathrm{LoS}}\mathbb{P}_l^{\mathrm{LoS}}\mu_{\mathrm{NLoS}}\mathbb{P}_l^{\mathrm{NLoS}}}\right)\right\} \\
&\overset{(b)}{\approx} \eta_R L^2N^2M\mathbb{E}_{\Psi_A,\Psi_U}\left\{\beta_{l,k}^{\mathbf{h}}\left(\mu_{\mathrm{LoS}}\mathbb{P}_{l,k}^{\mathrm{LoS}} + \mu_{\mathrm{NLoS}}\mathbb{P}_{l,k}^{\mathrm{NLoS}}\right)\beta_l^{\mathbf{G}}\left(\mu_{\mathrm{LoS}}\mathbb{P}_l^{\mathrm{LoS}} + \mu_{\mathrm{NLoS}}\mathbb{P}_l^{\mathrm{NLoS}}\right)\right\} \\
&= \eta_R L^2N^2M\mathbb{E}\left\{\frac{{\beta_0}^2}{l_{AU}^2 l_{BA}^2}\frac{\mu_{\mathrm{LoS}} + \mu_{\mathrm{NLoS}}\xi\left(l_{AU},H\right)}{1+\xi\left(l_{AU},H\right)}\frac{\mu_{\mathrm{LoS}} + \mu_{\mathrm{NLoS}}\xi\left(l_{BA},H\right)}{1+\xi\left(l_{BA},H\right)}\right\} \\
&\overset{(c)}{\approx} \eta_R L^2N^2M{\beta_0}^2\int_{l_{AU}^{\min}}^{l_{AU}^{\max}}\int_{l_{BA}^{\min}}^{l_{BA}^{\max}} f_{L_{BA}}\left(l_{BA}\middle|x_A,H\right)f_{L_{AU}}\left(l_{AU}\middle|x_A,H\right) \\
&\quad\cdot\frac{\left(\mu_{\mathrm{LoS}} + \mu_{\mathrm{NLoS}}\xi\left(l_{AU},H\right)\right)\left(\mu_{\mathrm{LoS}} + \mu_{\mathrm{NLoS}}\xi\left(l_{BA},H\right)\right)}{l_{AU}^2 l_{BA}^2\left(1+\xi\left(l_{AU},H\right)\right)\left(1+\xi\left(l_{BA},H\right)\right)}dl_{BA}dl_{AU}
\end{aligned} \tag{5.45}$$

where

$$\begin{aligned}
&l_{AU}^{\min} = \sqrt{\left(\max\left(0, x_U - x_A - \left(R_A + R_U\right)\right)\right)^2 + H^2}, \\
&l_{AU}^{\max} = \sqrt{\left(x_U - x_A + R_A + R_U\right)^2 + H^2}, \\
&l_{BA}^{\min} = \sqrt{\left(\max\left(0, x_A - R_A\right)\right)^2 + H^2},\ l_{BA}^{\max} = \sqrt{\left(x_A + R_A\right)^2 + H^2}, \\
&\xi\left(l_{AU},H\right) = \eta_a\exp\left(-\eta_b\left(\frac{180}{\pi}\arcsin\left(\frac{H}{l_{AU}}\right) - \eta_a\right)\right).
\end{aligned}$$

Note that $f_{L_{BA}}\left(l_{BA}\middle|\, x_A, H\right)$ and $f_{L_{AU}}\left(l_{AU}\middle|\, x_A, H\right)$ in (5.45) are given in Lemma 1 and Lemma 2, respectively. Specifically, (a) in (5.45) is obtained based on (5.44), (b) is derived by neglecting the root-related term since these are relatively small compared to other terms, and (c) is characterized by assuming the independence of Ψ_A and Ψ_U. It is worth noting that in (5.45), there exist two trade-offs in the large-scale fading resulted from the placement of SARIS. One is the trade-off between the doubled path loss and the excessive NLoS path loss. The other is the trade-off between the distance-dependent path loss and the excessive NLoS path loss.

Based on the above analysis, the optimal SARIS 3D position can be obtained by solving the following problem

$$\max_{x_A,H}\ \mathbb{E}\left\{\left|\sum_{l=1}^{L}\mathbf{h}_{l,k}^{\mathrm{H}}\ {}_l\mathbf{G}_l\mathbf{w}_k\right|^2\right\} \tag{5.46}$$

$$\text{s.t. } 0 \leqslant x_A \leqslant x_U, \tag{5.46a}$$

$$H_{\min} \leqslant H \leqslant H_{\max}, \tag{5.46b}$$

where $H_{\min}$ and $H_{\max}$ denote the minimum and maximum allowable SARIS flight height, respectively. Note that in (5.46), we can set the discrete feasible values for x_A and H to reduce the feasible region while achieving a desired performance. Note that in (5.46), we optimize the 3D position of SARIS cluster center for analytical tractability. The randomness in the cascaded large-scale fading, due to the random locations of ARISs in UAV swarm and users, is averaged in the objective function of (5.46).

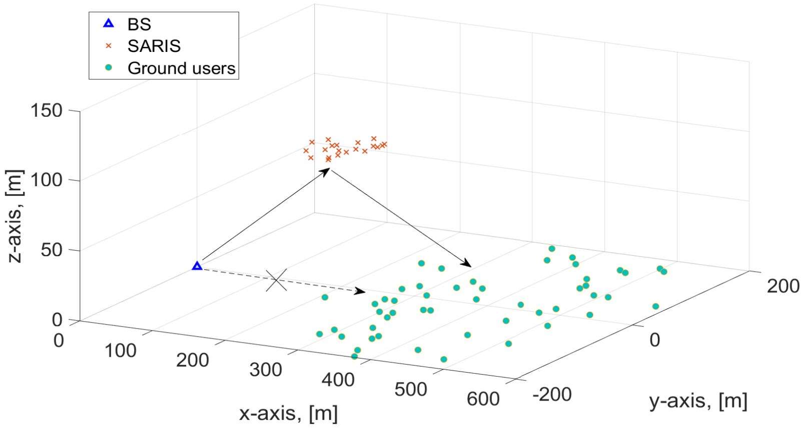
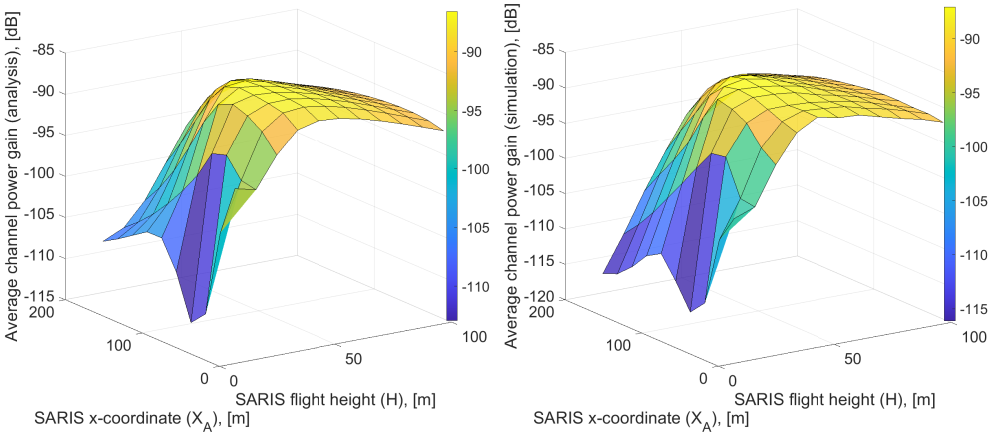


Fig 5.3: Simulation setup.

5.5 Simulation Results and Discussion

In this section, we do simulations to verify our analysis and evaluate the performance of the SARIS system. The simulation setup is shown in Fig. 5.3. We consider that the SARIS-assisted downlink transmissions from the BS to the ground users. The default parameters are given by $M = 16$, $N = 20$, $P_{\max} = 1$ W. We consider the dense urban environment [35], where $\eta_a = 12.1190$, $\eta_b = 0.5137$, $\mu_{LoS} = 0.6918$, $\mu_{LoS} = 0.005$. Simulation results are averaged over 100 experiments. The noise power is -80 dBm [13]. We set equal weights for all users. Other parameters are set as follows: $R_A = 10$ m, $R_U = 100$ m, $\beta_0 = -30$ dB, $H_{min} = 10$ m, $H_{max} = 200$ m, $\lambda = 0.1$ m, $d_A = d_B = \lambda/2$, $\eta_R = 0.9$, unless specified otherwise.

In Fig. 5.4, Fig. 5.5, Fig. 5.6, and Fig. 5.7, the user's average channel power gains are shown versus the x-axis and z-axis of the SARIS's center under $x_U = 200$ m, $x_U = 400$ m, $x_U = 600$ m, and $x_U = 1000$ m, respectively. The BS's position is $(0, 0, 0)$. The analytical results are obtained based on Section IV, and the simulation results are obtained based on Monte Carlo simulations. It is observed that the derived analytical results well approximate

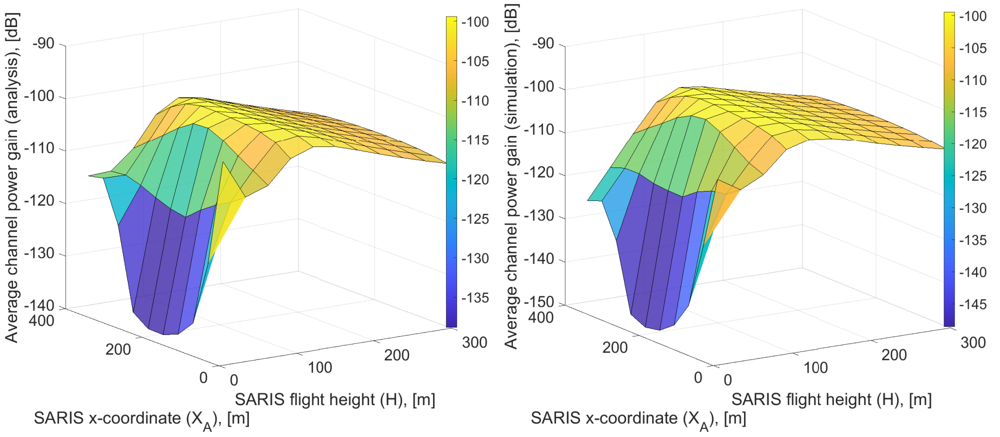


(a) Analytical results. (b) Monte Carlo simulation results.

Fig 5.4: Average channel power gain versus SARIS 3D position (analysis (left) and simulation (right)), where $x_U = 200\ m$.

the simulation results. In addition, there exists an optimal 3D position for SARIS given a x_U. When x_U is small, e.g., $x_U = 200$ m, it is optimal to place SARIS between the BS and users. When x_U is large, e.g., $x_U = 1000$ m, it is optimal to place SARIS above the BS. Intuitively, if SARIS is close to the BS or user, the doubled path loss is minimized, but the excessive path loss originated from NLoS connections becomes severe. Moreover, SARIS's higher altitude makes it easy to establish LoS connections between SARIS and BS/user, whereas this leads to an increased signal attenuation due to the increase of communication distance. Therefore, in SARIS placement, there is a trade-off between doubled path loss and excessive path loss. Furthermore, there is also a trade-off between distance-dependent path loss and NLoS probability, which is mainly due to the SARIS flight height. As such, SARIS placement is different from TRIS placement, where TRIS's optimal placement is to place the TRIS close to the BS or user. This is because the product of two path-loss models in negative exponential form.

In Fig. 5.8, we examine the weighted sum-rate versus x_U in the SARIS system with and without placement optimization, where $K = 4$, $M = 8$, $L = 10$. We evaluate the perfor-



(a) Analytical results. (b) Monte Carlo simulation results.

Fig 5.5: Average channel power gain versus SARIS 3D position (analysis (left) and simulation (right)), where $x_U = 400\ m$.

mance of the SARIS system with the following algorithms:

Scheme 1: Multi-user beamforming scheme in Section III.

Scheme 2: Weighted round-robin beamforming scheme in Section III-D.

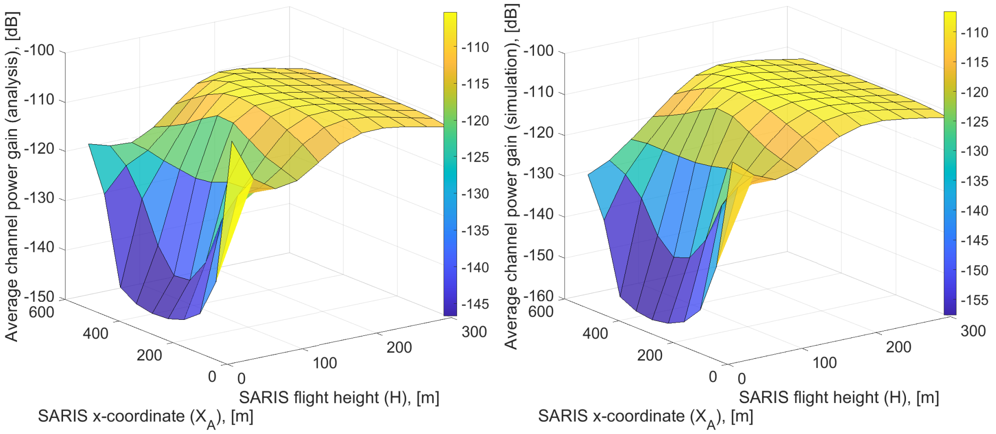
Analysis: Analytical results in Section IV-B with the SARIS placement optimization.

Random phase shifts: Random phase shifts at SARIS and WMMSE transmit beamforming at the BS [104].

WMMSE+SDR: WMMSE transmit beamforming at the BS and SDR technique for SARIS passive beamforming [96].

Ground RIS: Multi-user beamforming scheme in Section III, where RIS is near to the ground at the height of 10 m above the users.

The SARIS placement optimization is based on Section IV. In the case without SARIS placement optimization, we suppose that SARIS is at the altitude of $\frac{H_{\max}-H_{\min}}{2}$ above the ground users. Fig. 5.8 shows that the weighted sum-rate is significantly improved by optimizing SARIS placement. The weighted sum-rate improvement by optimizing the SARIS placement is notable when x_U becomes large, since the impacts of doubled path loss and excessive path

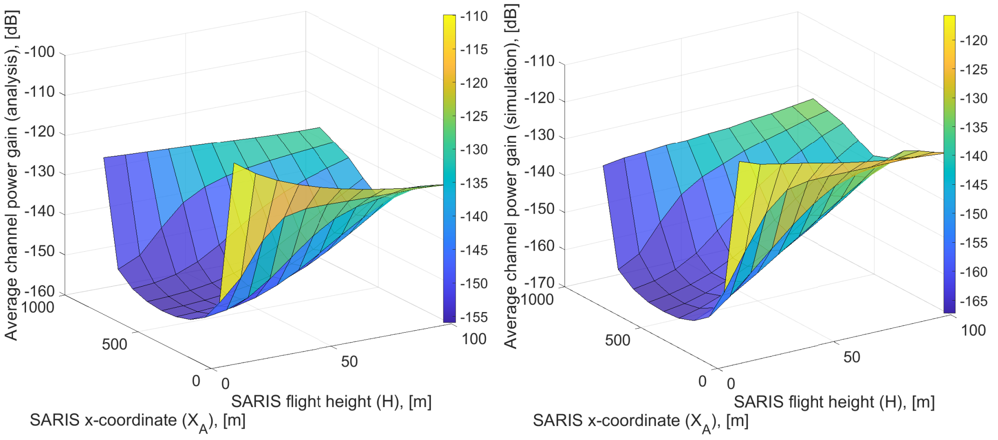


(a) Analytical results. (b) Monte Carlo simulation results.

Fig 5.6: Average channel power gain versus SARIS 3D position (analysis (left) and simulation (right)), where $x_U = 600\ m$.

loss on the system performance are obvious. Moreover, Fig. 5.8 shows that the weighted sum-rate decreases with x_U since the communication distance increases. Note that the performance gap between scheme 1 and scheme 2 decreases with x_U. This is because when x_U is large, the impact of path loss dominates the impact of beamforming gain on the system performance.

Regarding the beamforming algorithms, scheme 1 outperforms scheme 2, while scheme 2 achieves the closed-form solution which is tractable in calculation, and thus there is a trade-off between system performance and algorithmic complexity. Furthermore, it is observed that the performance of scheme 1 and scheme 2 are close when the SARIS placement is not optimized. As shown in Fig. 5.8, the analytical result in Section IV-B serves as an upper bound for scheme 2, which validates our analysis. As such, the analytical result can be used to evaluate the SARIS performance and determine the near optimal placement. In addition, for multi-user beamforming, the introduced scheme 2 achieves almost the same performance with the algorithm of WMMSE+SDR. However, the former has a lower algorithmic complexity. Last but not least, compared to the ground RIS, the SARIS system



(a) Analytical results. (b) Monte Carlo simulation results.

Fig 5.7: Average channel power gain versus SARIS 3D position (analysis (left) and simulation (right)), where $x_U = 1000$ m.

achieves significant weighted sum-rate improvement, which demonstrates the advantages of the introduced SARIS system.

In Fig. 5.9, we compare the weighted sum-rate versus R_A under different $P_{\max}$, where SARIS placement is optimized, $x_U = 400$ m, and $L = 10$. It is observed that the weighted sum-rate decreases with R_A. This is because the UAV swarm center is optimized in 3D space, and UAV's apparent deviation from the optimal position will impact the system performance. Thus, we have a design insight of reducing R_A in SARIS. In addition, one can increase $P_{\max}$ with scheme 1 to achieve the same weighted sum-rate with scheme 2. In this way, the system performance and algorithmic complexity are guaranteed with the cost of increased energy consumption at the BS.

In Fig. 5.10, we evaluate the weighted sum-rate versus the number of UAVs L, where $K = 4$, $M = 8$, $x_U = 400$ m. It is observed that the SARIS placement optimization plays a significant role in the system performance, which dominates the beamforming strategy. Furthermore, when L is large, the performance gap between scheme 1 and scheme 2 increases due to the increased aperture gain for multi-user spatial multiplexing. In Fig. 5.10, it is

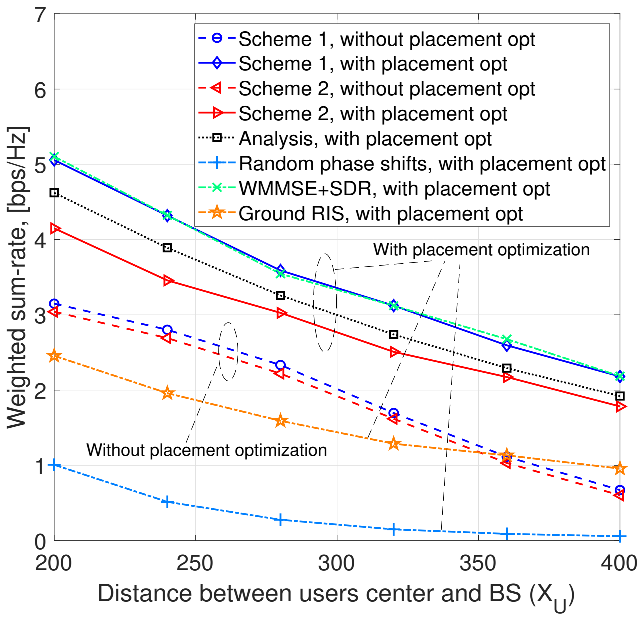


Fig 5.8: Weighted sum-rate versus X_U with and without placement optimization, where $K = 4$, $M = 8$, $L = 10$.

shown that under random phase shifts, although the SARIS placement is optimized enjoying a high probability of having LoS links, the weighted sum-rate is still undesirable compared to other schemes. This is because the random phase shifts incur severe mutual interference among users.

In Fig. 5.11, we examine the weighted sum-rate versus N, where the SARIS placement is optimized. It is observed that the weighted sum-rate increases with N, since a larger N provides a better aperture gain. Moreover, the increasing rate of the weighted sum-rate decreases with N, due to the logarithmic rate function.

In Fig. 5.12, we compare the impact of different channel state information (CSI) for SARIS system, where the SARIS placement is optimized, $K = 4$, $M = 8$, $x_U = 400$ m. We simulate the full CSI-based optimization and the LoS CSI-based optimization. The latter only considers the path loss and the LoS component determined by each ARIS's position,

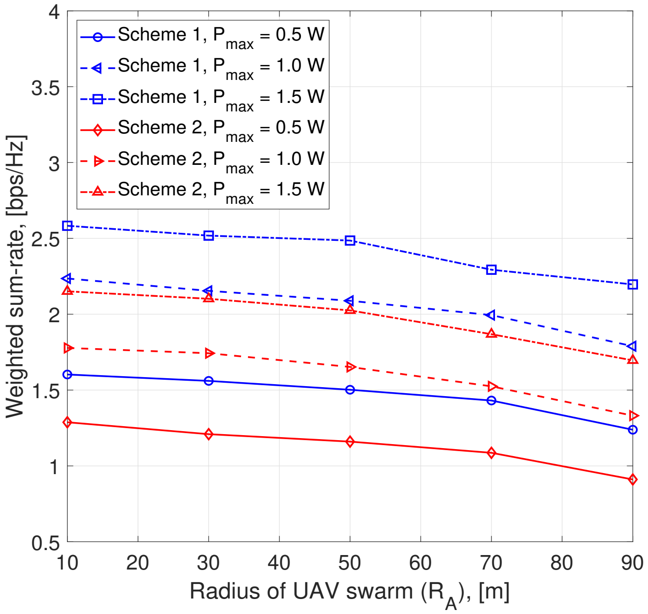


Fig 5.9: Weighted sum-rate versus R_A with placement optimization, where $x_U = 400$ m, $L = 10$.

and the NLoS component is omitted. It is observed from Fig. 5.12 that the performance has not decreased much when shifting from full CSI-based optimization to LoS CSI-based optimization. This is because the phase shifts of LoS components dominate that of NLoS components when SARIS placement is optimized. On the one hand, the NLoS probability is small given an optimized SARIS 3D position. On the other hand, the excessive NLoS path loss reduces the signal power strength of NLoS components. Therefore, we can utilize the ARISs' positions in the beamforming design and omit the channel estimation for small-scale fading, which reduces the system overhead. Moreover, to estimate the ARISs' positions, we can select a UAV leader to measure and store the mutual distances and directions from other UAVs to the UAV leader. As such, the UAVs' positions in the UAV swarm could be calculated based on the UAV leader's position. The estimation of signal parameters via rotational invariance techniques (ESPRIT) can be used to estimate the UAV leader's position [105]. In this way, the channel estimation overhead can be further reduced, where

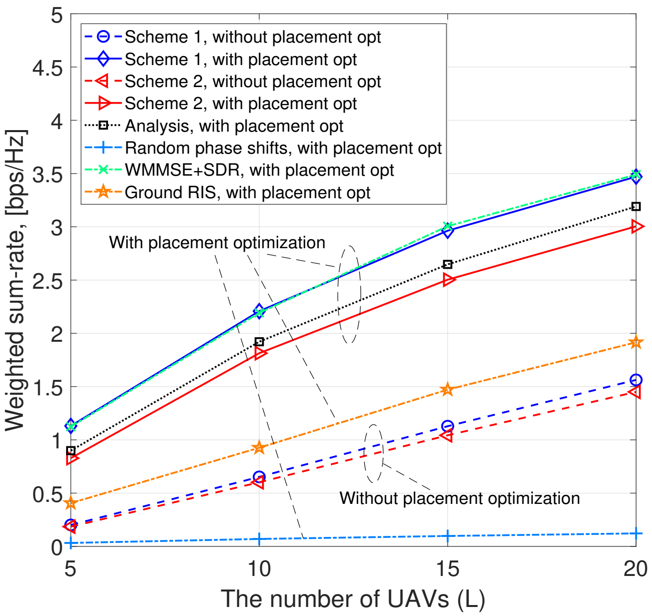


Fig 5.10: Weighted sum-rate versus L with and without placement optimization, where $K = 4$, $M = 8$, $x_U = 400$ m.

only the UAV leader's position needs to be measured.

5.6 Summary

We model and analyze a UAV swarm-enabled aerial reconfigurable intelligent surface (SARIS) system. Specifically, we consider a multi-user MISO system where SARIS is placed to assist the downlink transmissions for ground users. We introduce two beamforming schemes, namely multi-user beamforming and weighted WRR single-user beamforming. Furthermore, considering the random distributions of ground users and UAVs, we develop an analytical framework to characterize the average channel power gain of the SARIS system and optimize the SARIS 3D placement. The results show that when users are far from the BS, placing SARIS near the BS achieves a better performance. However, when users are near

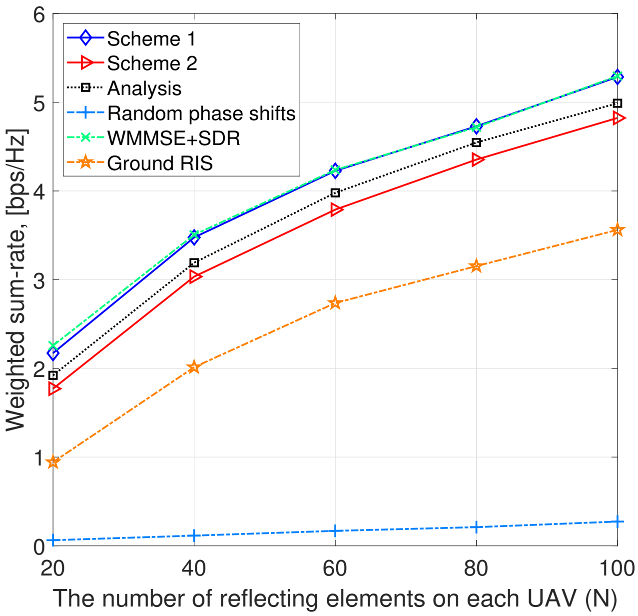


Fig 5.11: Weighted sum-rate versus N with placement optimization, where $L = 10$, $x_U = 400$ m.

the BS, placing SARIS between the BS and users can balance the doubled path loss and the excessive NLoS path loss. Moreover, simulation results demonstrate that the NLoS components can be ignored, which reduces the system overhead for channel estimation. In future work, the energy consumption of the SARIS system will be considered [106], where the energy supply technologies (e.g., wireless power transfer and energy harvesting [107]) and the energy-efficient communication protocols will be investigated to improve the operational time of the SARIS system.

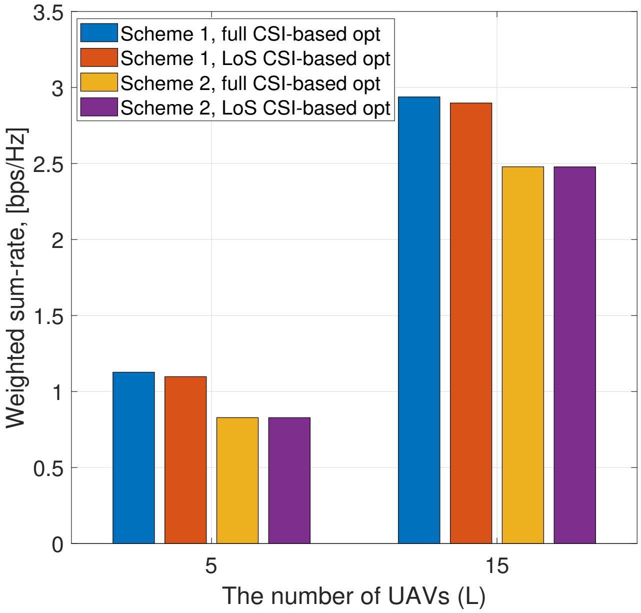


Fig 5.12: Weighted sum-rate versus L under full CSI-based and LoS CSI-based beamforming optimization, where $K = 4$, $M = 8$, $x_U = 400$ m, and the SARIS 3D position is optimized.

Chapter 6

Conclusions and Open Problems

In this chapter, we the main contributions of this dissertation and a few promising future research opportunities.

6.1 Summary of Contributions

In the last few years, UAVs have been widely developed from experiment to civilian use. Researchers started thinking about how to integrate UAVs into the wireless network for improving communication performance. In the meantime, users require intensive computation resources to execute their tasks in an energy-saving but low-latency way. Edge computing has been proposed to offload users' computation tasks to the edge servers. This dissertation focused on these two emerging technologies in wireless networks and explored a combination of their strengths in developing energy-and-spectrum-efficient cyber-physical systems.

Thus far, the main contributions of our work in this dissertation are summarized as follows. In Chapter 1, we presented the fundamentals and motivations when deploying UAVs, edge computing, and RIS. The limitations on the design of these communication networks were discussed. Then, we investigated the technical challenges and corresponding solutions of these emerging technologies in wireless networks.

In Chapter 2, we utilized machine learning and stochastic geometry to model and optimize the spectrum sharing policy among ground D2D and UAV users. In the considered scenario, the D2D-Txs are primary users, and UAVs are secondary users. UAVs opportunistically access the licensed spectrum by performing spatial spectrum sensing. To model such 3D spectrum sharing networks, we first obtain the probabilities of spatial false alarm and spatial missed detection of a typical UAV in the hybrid network. The coverage probability of a typical D2D-Rx and a typical UAV-Rx are derived based on the intermediate results. Finally, we characterize the ASEs of D2D and UAV networks, respectively. Simulation results show that reducing the spatial spectrum sensing radius of UAVs decreases the coverage probability of UAV communications. However, this improves the ASE of UAV networks despite the increasing inter-UAV interference. Based on the developed model, we can obtain the optimal spatial spectrum sensing radius of UAVs, which maximizes the ASE of UAV networks and guarantees the minimum ASE of D2D networks. Furthermore, the UAVs' optimal transmit power is optimized to maximize the ASE of UAV networks, given the requirement of minimum ASE of D2D networks.

In Chapter 3, we considered that UAV can provide computation resources for the ground users together with the GCAPs. In such air-ground integrated MEC systems, we developed an energy-efficient resource allocation algorithm. In the algorithm, we minimized the energy consumption of ground users by optimizing users' association, users' uplink transmit power, allocation bandwidth for data transmissions, assigned computation resources, and UAVs placement. We decomposed the primary energy consumption minimization problem into several manageable subproblems, and optimized each subproblem alternatively. Numerical results demonstrated the advantages of our proposed iterative algorithm compared with other computing method.

In Chapter 4, we focused on the VEC system, where the vehicle and in-vehicle users request computation offloading with the help of RSU and edge servers. Specifically, we developed a deep learning-assisted energy-efficient computation offloading algorithm for VEC systems. The developed algorithm aims to minimize the vehicle and in-vehicle users' energy consumption, while satisfying their latency constraints and outage constraints. By optimizing the users' association, data partition, transmit power, and computation resource allocation, we can efficiently leverage the computation resources at road-side edge servers. With the help of deep learning method for users' association, the algorithm can solve the complex VEC problem and find a near-optimal solution in a real-time manner with low complexity. Simulation results demonstrated the advantages of the proposed algorithm in substantially reducing users' total energy consumption compared with other methods.

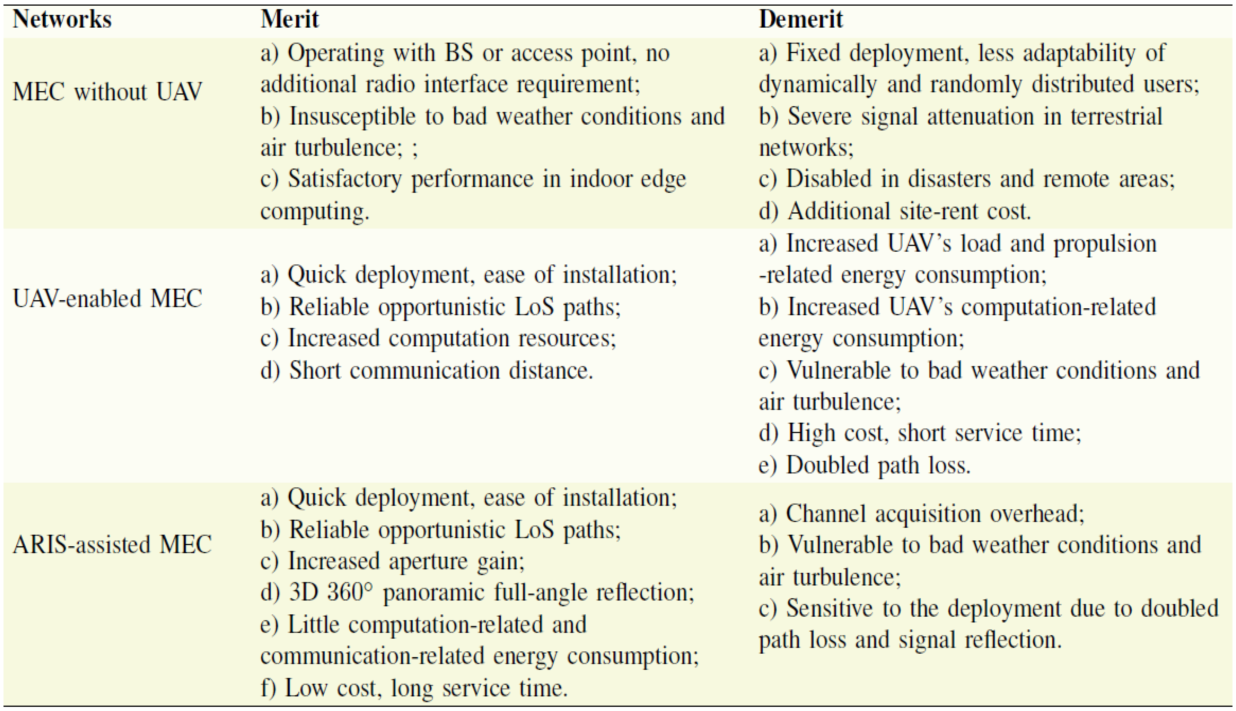
In Chapter 5, we focused on enhancing the communication performance with UAV and RIS. We modeled and analyzed a SARIS system where RIS are mounted on UAVs to assist the blocked cellular downlink communications. Specifically, we considered a multi-user MISO system where SARIS was placed to assist the downlink transmissions for ground users. We introduced two beamforming schemes, namely multi-user beamforming and weighted WRR single-user beamforming. Furthermore, considering the random distributions of ground users and UAVs, we developed an analytical framework to characterize the average channel power gain of the SARIS system and optimized the SARIS 3D placement. The results showed that when users are far from the BS, placing SARIS near the BS achieves a better performance. However, when users are near the BS, placing SARIS between the BS and users can balance the doubled path loss and the excessive NLoS path loss. Moreover, simulation results demonstrated that the NLoS components can be ignored, which reduces the system overhead for channel estimation.

6.2 Open Problems

There are numerous open research opportunities of the works presented in this dissertation. Few of the potential research directions on UAV and edge computing are listed below.

6.2.1 Efficient Aerial RIS-Assisted MEC Design

In previous chapters, we introduce aerial RIS (ARIS) mounted on UAVs and UAV-enabled MEC, respectively. Furthermore, we can bring ARIS to MEC. Specifically, RIS is mounted on UAV instead of the edge server to reflect users' signals to the BS for computation offloading. Such an ARIS-assisted MEC network enjoys many benefits. First, due to UAV's agility and mobility, ARIS can be quickly deployed in wireless networks. Its movement and placement can be designed to further improve the network performance. Second, the ARIS can provide LoS paths in air-ground channels to support reliable uplink transmissions for computation offloading, where the communication performance scales with the number of reflecting elements on ARIS. Third, different from the terrestrial RIS, ARIS achieves three-dimensional (3D) signal reflection, which is not restricted to the 180° half-space reflection, but instead, it provides a 360° panoramic full-angle reflection. Forth, ARIS does not have energy-hungry issues of the computation-related and communication-related energy consumption. This is because ARIS is light in weight and only reflects the signals rather than decoding, forwarding, and computing the information. Fifth, the cost of the ARIS-assisted MEC network is less than that of the UAV-enabled MEC network, resorting to the ease of RIS hardware implementation, which is made of two-dimensional digital metamaterial. Therefore, it is attractive but significant to investigate the orchestra of UAV, RIS, and MEC in both industry and academia. A comprehensive comparison of different MEC networks is given in Fig. 6.1, including the MEC without UAV, the UAV-enabled MEC, and the ARIS-assisted MEC.



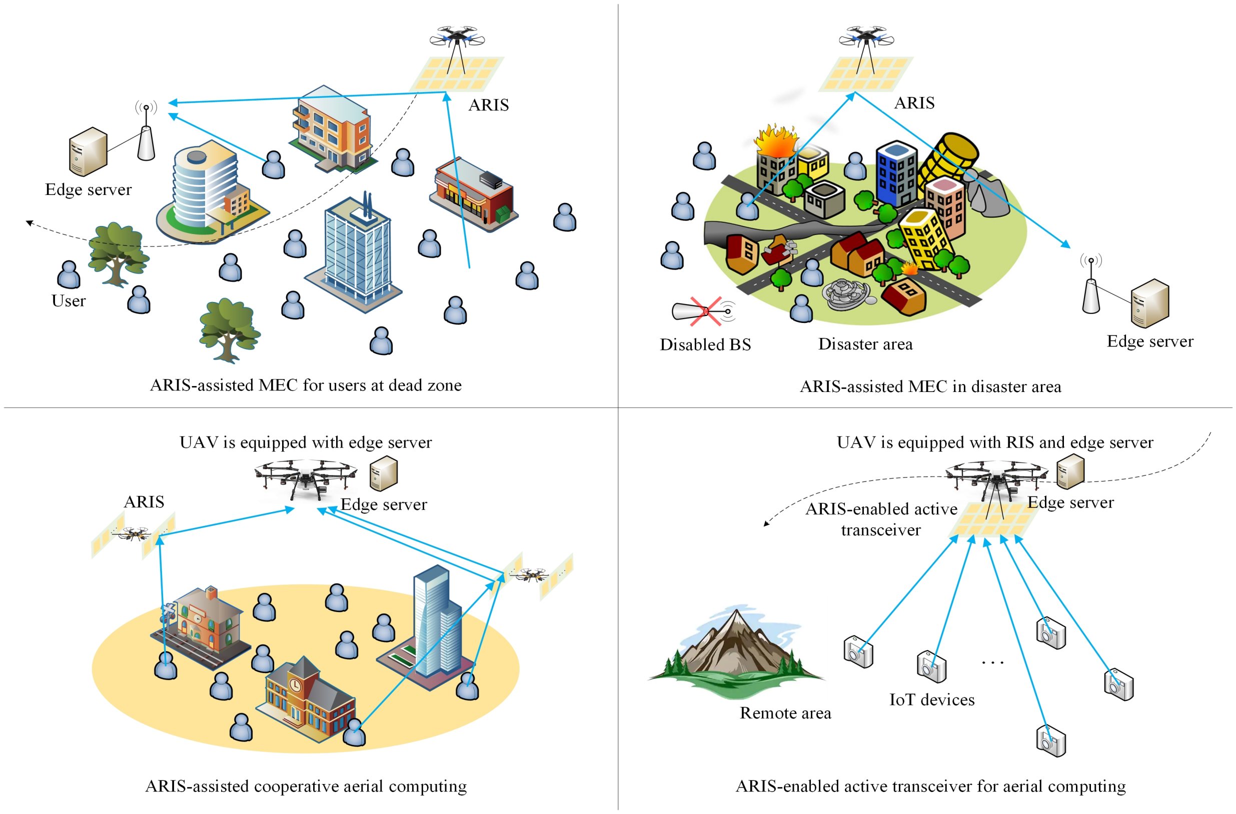
Networks	Merit	Demerit
MEC without UAV	a) Operating with BS or access point, no additional radio interface requirement; b) Insusceptible to bad weather conditions and air turbulence; ; c) Satisfactory performance in indoor edge computing.	a) Fixed deployment, less adaptability of dynamically and randomly distributed users; b) Severe signal attenuation in terrestrial networks; c) Disabled in disasters and remote areas; d) Additional site-rent cost.
UAV-enabled MEC	a) Quick deployment, ease of installation; b) Reliable opportunistic LoS paths; c) Increased computation resources; d) Short communication distance.	a) Increased UAV's load and propulsion -related energy consumption; b) Increased UAV's computation-related energy consumption; c) Vulnerable to bad weather conditions and air turbulence; d) High cost, short service time; e) Doubled path loss.
ARIS-assisted MEC	a) Quick deployment, ease of installation; b) Reliable opportunistic LoS paths; c) Increased aperture gain; d) 3D 360° panoramic full-angle reflection; e) Little computation-related and communication-related energy consumption; f) Low cost, long service time.	a) Channel acquisition overhead; b) Vulnerable to bad weather conditions and air turbulence; c) Sensitive to the deployment due to doubled path loss and signal reflection.

Figure 6.1: A comprehensive comparison of different MEC networks.

Notwithstanding its many benefits, the ARIS-assisted MEC network constitutes active (user, BS, UAV) and passive (ARIS) components. Moreover, it involves a joint movement and/or placement, communication, and computation design, thus differing significantly from the traditional MEC network. This motivates us to introduce ARIS into MEC network. Moreover, new transformative applications of ARIS in MEC networks are shown in Fig. 6.2.

In the following, we illustrate some potential but significant research opportunities for integrating ARIS into MEC networks.

Multi-Hop RIS in Air-Ground Cooperative Computing. RISs would be deployed in terrestrial networks as a part of the infrastructure in beyond-5G and the sixth-generation (6G) networks. Designing an efficient ARIS-assisted MEC network in the presence of TRISs is worth investigating. For example, multi-hop RIS-assisted cooperative computation offload-



ARIS-assisted MEC for users at dead zone

ARIS-assisted MEC in disaster area

ARIS-assisted cooperative aerial computing

ARIS-enabled active transceiver for aerial computing

Figure 6.2: Applications of ARIS in MEC.

ing through ARISs and TRISs can be designed, providing extended coverage and improved achievable data rate. TRISs can be deployed near BSs, while ARISs enjoy flexible mobility to provide reliable uplink data rates for ground users in their computation offloading. As such, the ARISs' movement/placement design, computation resource allocation, and the phase shifts of ARISs and TRISs can be jointly optimized to improve the network performance.

ARIS-Assisted Distributed Computing. Distributed computing serves as a privacy-preserving approach for multiple distributed devices to train a learning algorithm without exchanging their private data samples. Based on the partially observed data, the devices compute data locally and exchange their training parameters to improve learning accuracy. In the exchange of training parameters, a model aggregator is usually required. However, a ground model aggregator is sometimes challenging to be deployed in specific areas, such as disasters, and remote areas. Therefore, ARISs can be deployed to enable the parameters exchange and model aggregation, by providing reliable uplink and downlink data rates for devices with aperture gain and flexible mobility. Moreover, the trajectory and the frequency of model aggregation can be optimized to minimize UAV's energy consumption.

UAV Swarm-Enabled Cooperative ARIS for MEC. The number of reflecting elements on a single UAV is constrained resulting from the UAV's limited payload and flight stability. Note that an ARIS with a large size is sensitive to air turbulence. However, with moderate ARIS size, UAV swarm-enabled cooperative ARISs can be applied in MEC to increase the aperture gain for computation offloading [89]. In this scenario, UAV swarm deployment needs to be designed, given specific ground users' and BS's distributions. On the other hand, multiple ARISs can form a single-hop reflection together or realize multi-hop reflections cooperatively between users and the BS. The above two schemes are expected to be compared given specific network settings so that we can determine which one performs better under different circumstances.

Energy-Efficient ARIS Trajectory Design. In ARIS-assisted MEC networks, the propulsion-related energy consumption accounts for the vast majority of the ARIS's sum energy consumption, which is dissipated for supporting the hovering and mobility of the ARIS. For the rotary-wing ARIS, energy consumption is related to its velocity and acceleration. Thus, to maximize the ARIS's service time of computation offloading (i.e., minimize the ARIS energy consumption), we can jointly design the ARIS trajectory and phase shift matrix while guaranteeing users' latency requirements. As such, ARISs can form cascaded virtual LoS links between users and the BS, and decide whether to move or maintain hovering status for maximizing its long-term benefits. Moreover, in aerial computing networks with ARISs (e.g., Section II-C and II-D), it is worth to investigate the UAV's energy balance between the computation-related and propulsion-related energy consumption to maximize the service time of aerial computing.

ARIS-Assisted Wireless Powered MEC. Wireless power transfer (WPT) allows the BS to transmit energy by electromagnetic waves for energy-constrained devices. The convergence of WPT and MEC has fostered the rise of wireless powered MEC to prolong the battery's operation time. Traditional wireless powered MEC relies on the BS's multi-antenna beamforming technique, constrained to the BS's fixed deployment. By deploying an ARIS for both WPT and MEC, one can adjust the phase shift matrix and the ARIS movement to enhance the pointed beams' signal strength and form the cascaded virtual LoS paths. Users utilize their harvested energy of WPT to upload their computational tasks. In such a new 3D cascaded wireless channel, the doubly near-far problem in WPT needs to be revisited, and the computational tasks heterogeneity in MEC needs to be considered.

ARIS-Assisted Vehicular Edge Computing. Vehicular edge computing has become a promising computing architecture in vehicular networks [108]. Edge servers can offload vehicle users' computational tasks via roadside units. However, enormous vehicles sometimes dramatically appear on certain roads during the peak hours, and thus the computation resources of the edge servers on these saturated roads may not satisfy the users' intensive computation demands. In this case, ARISs can be quickly deployed on the saturated roads, offloading part of vehicle users' computational tasks to the edge servers on other roads that have available computation resources. The ARIS-assisted computation offloading and computation balancing for vehicular edge computing networks are worth to be investigated.

6.2.2 Computing in Space-Air-Ground Integrated Networks

The space-air-ground integrated network (SAGIN) is an exemplary network architecture in 6G networks to provide ubiquitous connectivity as shown in Fig. 6.3. In SAGINs, there are three network segments, namely terrestrial network, aerial network, and space network[109]. Aerial networks and space networks belong to non-terrestrial networks. In previous chapters, we study aerial network-enabled edge computing. In SAGINs, the space network can also help on offloading tasks for ground users. Assume that the geostationary orbit (GEO) satellites, medium earth orbit (MEO) satellites, and low earth orbit (LEO) satellites have the computing ability and/or act as relay nodes to forward the computation tasks from overloaded ground edge servers to available ground edge servers. However, the computing capability of satellites is usually limited due to their restricted size, weight, and onboard energy. Computing in SAGINs should be operated in a cooperative way where the three network segments collaboratively offload the computation tasks.

The cooperative computing over SAGINs can be utilized in many scenarios. In the following, we present some examples.

Computing in disasters. In disasters, the terrestrial base stations (BSs) are destroyed and disabled due to damage. The aerial network can be quickly deployed to support the computing services. In the meantime, the space network can assist the terrestrial network and aerial network in offloading computation tasks by forwarding the tasks to other available edge servers and/or the remote cloud servers. In this scenario, the system design should consider the impact of propagation delay resulting from space networks on edge computing performance.

Computing for remote Internet-of-Things (IoT). The current network can usually not cover remote areas such as forests, vast oceans, and other arduous environments due to

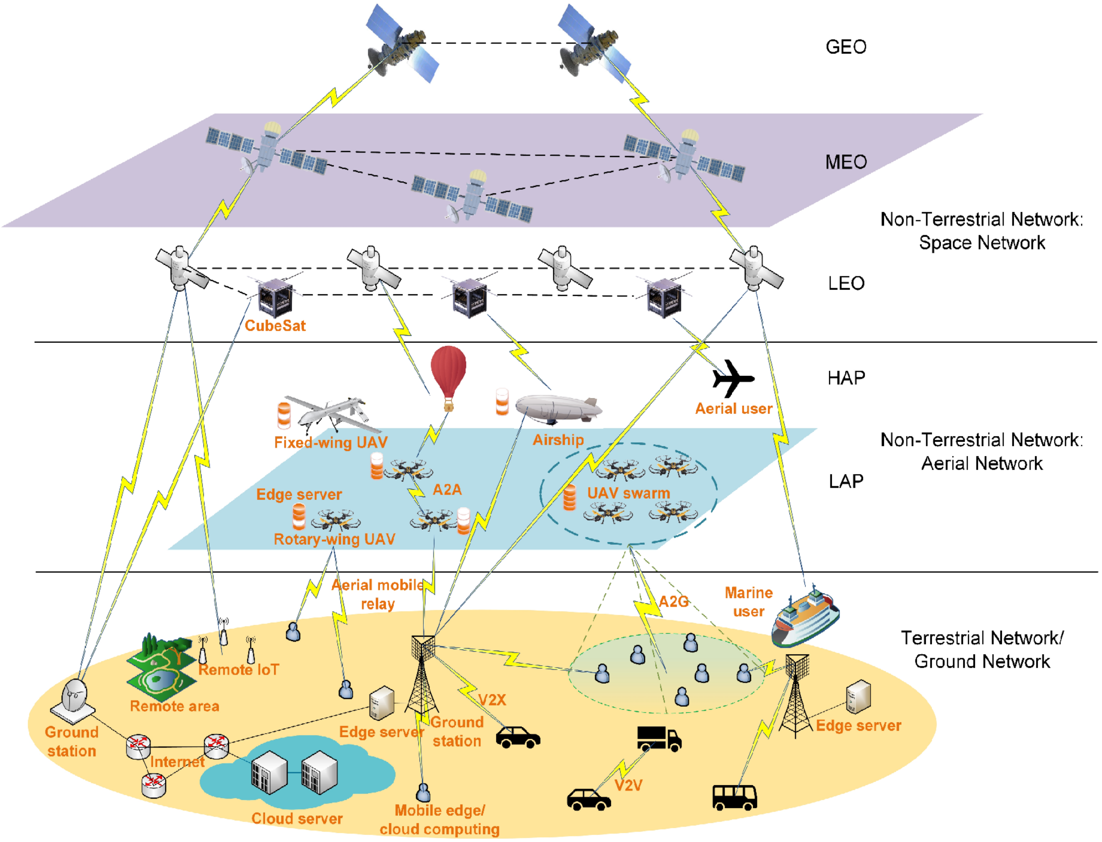


Figure 6.3: Computing architecture of space-air-ground integrated networks.

the limited network coverage capability. In these remote areas, IoT devices may be widely deployed to complete specific tasks requiring data computing. However, the battery capacity of a remote IoT device is limited, which restricts the IoT device from executing extensive computation tasks. To this end, SAGINs can be used to offload these remote IoT devices' computation tasks to edge/cloud servers to save their energy and provide globally scalable connectivity.

Computing for Internet-of-Space-Things (IoST). In non-terrestrial networks, satellites usually require computation resources to execute their intensive tasks. Due to the limited computation capacity and restricted onboard energy at satellites such as CubeSats, the tasks can be offloaded to other network segments to save energy consumption and reduce computation latency. The cooperative task offloading in SAGINs is worth to be investigated in the future. Note that the relative orbital velocity of satellites to other network components is usually significant, which generates the phase noise and frequency offset due to the Doppler frequency shift. As such, the orthogonality among sub-carriers would be destroyed. Therefore, new modulation schemes, such as orthogonal time-frequency space, can be utilized in information transmission.

Appendices

Appendix A

Proofs for Chapter 2

A.1 Proof of Lemma 2.4

Considering that there is at least one D2D-Tx in the $\mathcal{A}_{v_i}$ under the case of $\mathcal{H}^1$, we can straightforwardly obtain the test statistics Γ under $\mathcal{H}^1$ is greater than that under $\mathcal{H}^0$, as follows

$$\mathbb{P}\left(\Gamma > \varepsilon | \, \mathcal{H}^1, I^1\right) > \mathbb{P}\left(\Gamma > \varepsilon | \, \mathcal{H}^0, I^0\right) \Leftrightarrow (1 - P_{md}) > P_{fa}, \tag{A.1}$$

where ε is the energy detection threshold. In addition, we have

$$\begin{aligned} &P^0 - P^1 = (P_{fa} - 1 + P_{md})\,\beta_1 + (1 - P_{fa} - P_{md})\,\beta_0 \\ &= - \left[(1 - P_{md}) - P_{fa}\right]\beta_1 + \left[(1 - P_{md}) - P_{fa}\right]\beta_0 = \left[(1 - P_{md}) - P_{fa}\right](\beta_0 - \beta_1) > 0, \end{aligned} \tag{A.2}$$

where $\beta_0 > \beta_1$ in the system model.

Therefore, $P^0 > P^1$ holds, which completes the proof.

A.2 Proof of Theorem 2.5

According to (2.1), conditioned on the D2D serving distance $l_k^d = \left\| d_k - u_k^d \right\|$, the coverage probability of a typical D2D-Rx u_k^d is given by

$$\begin{aligned}
&\mathbb{P}_D^c | \, l_k^d \\
&= \mathbb{P}\left\{ SINR\left(u_k^d\right) \geqslant \gamma_D^{th} \right\} \\
&= \mathbb{P}\left\{ h_{d_k u_k^d} \geqslant \frac{\gamma_D^{th}\left(I_{u_k^d}^V + I_{u_k^d}^D + {\sigma_n}^2\right)}{P_D\left(l_k^d\right)^{-\alpha_{GG}}} \right| l_k^d = \left\| d_k - u_k^d \right\| \Bigg\} \\
&= \mathcal{L}_{I_{u_k^d}^D}\left(\frac{\gamma_D^{th}}{P_D\left(l_k^d\right)^{-\alpha_{GG}}} \right) \mathcal{L}_{I_{u_k^d}^V}\left(\frac{\gamma_D^{th}}{P_D\left(l_k^d\right)^{-\alpha_{GG}}} \right) \exp\left(-\frac{\gamma_D^{th}{\sigma_n}^2}{P_D\left(l_k^d\right)^{-\alpha_{GG}}} \right).
\end{aligned} \tag{A.3}$$

More specifically, we have

$$\begin{aligned}
&\mathcal{L}_{I_{u_k^d}^D}(s) \\
&= \mathbb{E}\left\{ \exp\left(-s \sum_{d_j \in \Phi_D, d_j \neq d_k} P_D h_{d_j u_k^d} \left\| d_j - u_k^d \right\|^{-\alpha_{GG}} \right) \right\} \\
&= \exp\left(-\frac{2\pi^2 \lambda_D (s P_D)^{\frac{2}{\alpha_{GG}}}}{\alpha_{GG} \sin\left(\frac{2\pi}{\alpha_{GG}}\right)} \right).
\end{aligned} \tag{A.4}$$

We define an integer N_h satisfying Φ_{V,H_n} will $H_{N_h} \approx \mathcal{R}_s$. The interference power generated from UAVs is obtained by a summation of each sub-region as follows

$$I_{u_k^d}^V = \sum_{n=1}^{N_h} I_{u_k^d}^{V,n} + \sum_{n=N_h+1}^{N} I_{u_k^d}^{V,n}, \tag{A.5}$$

The Laplace transform of the interference power from UAVs at a typical D2D-Rx u_k^d is

$$\mathcal{L}_{I_{u_k^d}^V}(s) = \bigcup_{n=1}^{N_h} \mathcal{L}_{I_{u_k^d}^{V,n}}(s) \bigcup_{n=N_h+1}^{\infty} \mathcal{L}_{I_{u_k^d}^{V,n}}(s). \tag{A.6}$$

In addition, we have

$$
\begin{aligned}
&\bigcup_{n=1}^{N_h} \mathcal{L}_{I^{V,n}_{u_k^d}}(s) \\
&= \mathbb{E}\left\{ \exp\left[-s \sum_{n=1}^{N_h} \sum_{v_j^{H_n} \in \Phi^L_{V,H_n}} \frac{P_{NLOS}\left(v_j^{H_n}, u_k^d\right) P_V \eta g_{v_j^{H_n} u_k^d}}{\left\| v_j^{H_n} - u_k^d \right\|^{\alpha_{AG}}} + \frac{P_{LOS}\left(v_j^{H_n}, u_k^d\right) P_V g_{v_j^{H_n} u_k^d}}{\left\| v_j^{H_n} - u_k^d \right\|^{\alpha_{AG}}} \right] \right\} \\
&\approx \exp\left\{ -\sum_{n=1}^{N_h} 2\pi P^{act,n}_{UAV} \lambda_{V,H_n} \int_{\sqrt{{\mathcal{R}_s}^2 - {H_n}^2}}^{\infty} x \left[1 - \left(1 + \xi'\left(x, P_V, H_n\right) + \Theta'\left(x, P_V, H_n\right)\right)^{-M} \right] dx \right\},
\end{aligned}
$$

and

$$
\begin{aligned}
\xi'\left(x, P_V, H_n\right) &= \frac{s\eta_M P_V \left(x^2 + {H_n}^2\right)^{-\frac{\alpha_{AG}}{2}} \Big/ M}{1 + C\exp\left[-B\left(E\arctan\left(H_n/x\right) - C\right)\right]} \\
\Theta'\left(x, P_V, H_n\right) &= \frac{\eta_M P_V \eta \left(x^2 + {H_n}^2\right)^{-\frac{\alpha_{AG}}{2}}}{M} - \eta\xi'\left(x, P_V\right),
\end{aligned}
\tag{A.7}
$$

where in (A.7) we approximately consider the interference at the typical D2D-Tx instead of at the typical D2D-Rx for analytical tractability.

When $n > N_h$, the flight heights of UAVs exceed the spatial spectrum sensing radius $\mathcal{R}_s$. Thus, based on the anlytical framework, UAVs will transmit with probability P^0. Therefore, the Laplace transform of $\sum_{n=N_h+1}^{N} I^{V,n}_{u_k^d}$ is obtained by substituting P^0 for $P^{act,n}_{UAV}$ in (A.7) and letting the lower limit of integral to be zero.

Combining (A.4), (A.6) into (A.3), we obtain the coverage probability of a typical D2D-Rx as in (2.36), which completes the proof.

A.3 Proof of Theorem 2.6

The serving distance of UAV u_i^v is denoted by l_i^v, where $l_i^{v2} = r_i^{v2} + {h_{v_i}}^2$.

According to (2.4), conditioned on r_i^v and h_{v_i}, the coverage probability of UAV-Rx u_i^v is

$$\begin{aligned}
\mathbb{P}_V^c \big| \, r_i^v, h_{v_i} &= \mathbb{P}\left\{ SINR\left(u_i^v\right) \geqslant \gamma_V^{th} \big| \, r_i^v, h_{v_i} \right\} \\
&= \mathbb{P}\left\{ g_{v_i u_i^v} \geqslant \frac{\gamma_V^{th} I_{u_i^v}^{agg}}{P_V L\left(v_i, u_i^v\right)\left(l_i^v\right)^{-\alpha_{AG}}} \,\middle|\, r_i^v, h_{v_i} \right\} \\
&\overset{(a)}{\approx} 1 - \mathbb{E}\left\{ \left[1 - \exp\left(- \frac{\eta_M \gamma_V^{th} I_{u_i^v}^{agg}}{P_V L\left(v_i, u_i^v\right)\left(l_i^v\right)^{-\alpha_{AG}}} \right) \right]^M \right\} \\
&\overset{(b)}{=} 1 - \sum_{n=0}^{M} \left(\begin{array}{c} M \\ n \end{array} \right) \mathbb{E}\left\{ - \exp\left(- \frac{\eta_M \gamma_V^{th} I_{u_i^v}^{agg}}{P_V L\left(v_i, u_i^v\right)\left(l_i^v\right)^{-\alpha_{AG}}} \right) \right\}^n \qquad \text{(A.8)} \\
&= \sum_{n=1}^{M} \left(\begin{array}{c} M \\ n \end{array} \right) (-1)^{n+1} \mathbb{E}\left\{ e^{-\frac{n\eta_M \gamma_V^{th}}{P_V L\left(v_i, u_i^v\right)\left(l_i^v\right)^{-\alpha_{AG}}} I_{u_i^v}^{agg}} \right\} \\
&= \sum_{n=1}^{M} \left(\begin{array}{c} M \\ n \end{array} \right) (-1)^{n+1} \mathcal{L}_{I_{u_i^v}^D}(s)\, \mathcal{L}_{I_{u_i^v}^V}(s)\, e^{-s{\sigma_n}^2},
\end{aligned}$$

where $I_{u_i^v}^{agg} = I_{u_i^v}^V + I_{u_i^v}^D + {\sigma_n}^2$ is the aggregated interference and noise power at u_i^v, $s = \frac{n\eta_M \gamma_V^{th}}{P_V L\left(v_i, u_i^v\right)\left(l_i^v\right)^{-\alpha_{AG}}}$ and $L\left(v_i, u_i^v\right) = P_{LOS}\left(v_i, u_i^v\right) + \eta P_{NLOS}\left(v_i, u_i^v\right)$ is given in (2.37), (a) is obtained by the approximation of normalized gamma distribution of $g_{v_i u_i^v}$ [110], (b) is obtained from Binomial theorem. Specifically, we have

$$\mathcal{L}_{I_{u_i^v}^D}(s) = \mathbb{E}\left\{ \exp\left(-s \sum_{d_j \in \Phi_D} P_D h_{d_j u_k^d} \| d_j - u_i^v \|^{-\alpha_{GG}} \right) \right\} \approx \exp\left(- \frac{2\pi^2 \lambda_D (s P_D)^{\frac{2}{\alpha_{GG}}}}{\alpha_{GG} \sin\left(\frac{2\pi}{\alpha_{GG}}\right)} \right). \qquad \text{(A.9)}$$

Similar to (A.5) and (A.6), the Laplace transform of the interference power from UAVs is

$$\mathcal{L}_{I_{u_i^v}^V}(s) = \bigcup_{n=1}^{N_h} \mathcal{L}_{I_{u_i^v}^{V,n}}(s) \bigcup_{n=N_h+1}^{N} \mathcal{L}_{I_{u_i^v}^{V,n}}(s). \qquad \text{(A.10)}$$

The results of (A.10) can be obtained with the similar methods to (A.7).

Finally, combining (A.9) and (A.10) into (A.8), we can obtain the desired results.

Appendix B

Proofs for Chapter 3

B.1 Proof of Proposition 3.1

The energy consumption Υ_{km}^{G} of UE k associated with GCAP m is given by

$$\Upsilon_{km}^{G} = \frac{p_{km} D_k \ln 2}{B_{km} \ln\left(1 + \frac{h_{km}^{G} p_{km}}{B_{km} n_0}\right)}. \tag{B.1}$$

The first order derivative of Υ_{km}^{G} with respect to P_{km} is

$$\frac{\partial \Upsilon_{km}^{G}}{\partial p_{km}} = \frac{D_k \ln 2}{B_{km}} \frac{\ln\left(\frac{B_{km} n_0 + h_{km}^{G} p_{km}}{B_{km} n_0}\right) - \frac{h_{km}^{G} p_{km}}{B_{km} n_0 + h_{km}^{G} p_{km}}}{\left[\ln\left(1 + \frac{h_{km}^{G} p_{km}}{B_{km} n_0}\right)\right]^2}. \tag{B.2}$$

Letting $x = \frac{B_{km} n_0 + h_{km}^{G} p_{km}}{B_{km} n_0}$, we have $x \geqslant 1$, $1 - \frac{1}{x} = \frac{h_{km}^{G} p_{km}}{B_{km} n_0 + h_{km}^{G} p_{km}}$. The numerator of (B.2) can be expressed as $g_p(x) = \ln(x) - \left(1 - \frac{1}{x}\right)$ and its first order derivative is given by $\frac{\partial g_p(x)}{\partial x} = \frac{1}{x} - \frac{1}{x^2} = \frac{x-1}{x^2} \geqslant 0$. Thus, we have $g_p(x) \geqslant 0$ and $\frac{\partial \Upsilon_{km}^{G}}{\partial p_{km}} \geqslant 0$, which indicates that Υ_{km}^{G} is non-decreasing with p_{km}.

The energy consumption of UE k associated with UAV m is given by

$$\Upsilon_{km}^{A} = \frac{p_{km} D_k \ln 2}{B_{km} \ln\left(1 + \frac{h_0^{A} p_{km}}{B_{km} n_0 \left(\left(z_m^{A}\right)^2 + \left(L_{km}^{A}\right)^2\right)^{\frac{\alpha_{GA}\left(z_m^{A}\right)}{2}}}\right)}. \tag{B.3}$$

where $\alpha_{GA}\left(z_m^{A}\right) = \max\left(\alpha_1 - \alpha_2 \log_{10}\left(z_m^{A}\right), 2\right)$. It can also be proved that $\frac{\partial \Upsilon_{km}^{A}}{\partial p_{km}} \geqslant 0$ which indicates that Υ_{km}^{A} is non-decreasing with the respect to P_{km}.

B.2 Proof of Theorem 3.2

To minimize the power consumption of UE k associated with GCAP m, the latency constraint should be satisfied, i.e., $\frac{D_k}{R_{km}^G} + \frac{F_k}{f_{km}^G} = T_k$, and we have

$$\log_2\left(1 + \frac{h_{km}^G p_{km}}{\sigma^2}\right) = \frac{D_k}{B_{km}\left(T_k - \frac{F_k}{f_{km}^G}\right)}. \tag{B.4}$$

Similarly, when UE k is connected to UAV m, the maximal latency constraint is expressed as $\frac{D_k}{R_{km}^A} + \frac{F_k}{f_{km}^A} = T_k$. In addition, we have

$$\frac{p_{km} D_k}{\Upsilon_{km}^A B_{km}} = \frac{D_k}{B_{km}\left(T_k - \frac{F_k}{f_{km}^A}\right)}. \tag{B.5}$$

By calculating the transmit power in (B.4) and (B.5), we obtain the closed-form expressions as shown in Theorem 1, which completes the proof.

B.3 Proof of Proposition 3.3

We define the function $f\left(x\right) = x \ln\left(1 + \frac{b}{ax}\right), x \geqslant 0, a > 0, b > 0$, and we have

$$f'\left(x\right) = \ln\left(1 + \frac{b}{ax}\right) - \frac{b}{b + ax} = \ln\left(y\right) + \frac{1}{y} - 1, \tag{B.6}$$

where $y = \frac{ax+b}{ax} \geqslant 1$, $1 - \frac{1}{y} = \frac{b}{ax+b}$. We define the function $h\left(y\right) = \ln\left(y\right) + \frac{1}{y}$, and we have $h'\left(y\right) = \frac{1}{y} - \frac{1}{y^2} \geqslant 0$ and $h\left(1\right) = 1$. Thus, we have $h\left(y\right) \geqslant 1$ and $f'\left(x\right) \geqslant 0$.

The second order derivative of $f\left(x\right)$ is given by

$$f''\left(x\right) = -\frac{b^2}{x{(b + ax)}^2} < 0, \tag{B.7}$$

which indicates that function $f\left(x\right)$ is a concave function with respect to x. Thus, R_{km} is a concave function with respect to B_{km}, and $\frac{1}{R_{km}}$ is a convex function of B_{km} accordingly which means the objective function of (3.32) is a convex function.

In addition, due to the fact that constraints of (3.32) are convex, problem (3.32) is a convex problem, which completes the proof.

B.4 Proof of Theorem 3.4

Since $\min_{\mathbf{b}} \mathcal{L}\left(\mathbf{b}, \mathbf{v}, \lambda_b\right)$ is a convex problem and the slater's condition is satisfied, we solve it by using KKT conditions and letting $\frac{\partial \mathcal{L}(\mathbf{b}, v, \lambda_b)}{\partial B_{km}} = 0$.

Then, based on KKT conditions, we have

$$\frac{\lambda_b - \omega_{km}}{\left(p_{km}^* + v_k\right) D_k \ln 2} = \frac{\ln\left(1 + \frac{h_{km} p_{km}^*}{B_{km} n_0}\right) - \frac{h_{km} p_{km}^*}{h_{km} p_{km}^* + B_{km} n_0}}{{B_{km}}^2 \left[\ln\left(1 + \frac{h_{km} p_{km}^*}{B_{km} n_0}\right)\right]^2}, \tag{B.8}$$

where $a_{km}^* = 1$, which completes the proof.

B.5 Proof of Proposition 3.5

By denoting $SNR_{km} = \frac{h_{km} p_{km}^*}{n_0}$, the function $\psi\left(x\right)$ is given by

$$\psi\left(x\right) = \frac{\ln\left(1 + \frac{SNR_{km}}{x}\right) - \frac{SNR_{km}}{SNR_{km} + x}}{x^2 \left[\ln\left(1 + \frac{SNR_{km}}{x}\right)\right]^2}. \tag{B.9}$$

In addition, the first order derivative of $\psi\left(x\right)$ is given by

$$\frac{\partial \psi\left(x\right)}{\partial x} = -\frac{\frac{SNR_{km}}{(SNR_{km} + x)}\left(\frac{1}{x} - \frac{1}{SNR_{km} + x}\right)}{x^2 \left[\ln\left(1 + \frac{SNR_{km}}{x}\right)\right]^2} - \frac{2\left[\ln\left(1 + \frac{SNR_{km}}{x}\right) - \frac{SNR_{km}}{SNR_{km} + x}\right]^2}{x^3 \left[\ln\left(1 + \frac{SNR_{km}}{x}\right)\right]^3}. \tag{B.10}$$

Since $\frac{1}{x} - \frac{1}{SNR_{km} + x} > 0$, we have $\frac{\partial \psi(x)}{\partial x} < 0$, which completes the proof.

B.6 Proof of Theorem 3.7

The Lagrangian function of problem (3.46) is given as follows

$$\begin{aligned}&\mathcal{L}\left(\mathbf{f}_m^G,\lambda_f\right)\\&=\sum_{k\in\mathcal{K}_m}\left[C_2 2^{\frac{F_kD_k}{B^*_{km}T_k\left(T_kf^G_{km}-F_k\right)}}+\frac{B^*_{km}n_0}{h^G_{km}}\left(\frac{F_k}{f^G_{km}}-T_k\right)\right.\\&\left.-\omega\left(f^G_{km}\left(t\right)\right)+\tau\left(f^G_{km}\left(t\right)\right)\left(f^G_{km}-f^G_{km}\left(t\right)\right)\right]+\lambda_f\left(\sum_{k\in\mathcal{K}_m}f^G_{km}-f_m^{G,prac}\right),\end{aligned}\tag{B.11}$$

where λ_f is the multiplier of (3.41a).

Since problem (3.46) is a convex problem and the Slater's conditions are satisfied, we can use KKT conditions to obtain the optimal solutions, which are expressed as follows

$$\begin{cases}\dfrac{\partial\mathcal{L}\left(\mathbf{f}_m^G,\lambda_f\right)}{\partial f^G_{km}}=0,k\in\mathcal{K}_m\\\lambda_f\left(\sum\limits_{k\in\mathcal{K}_m}f^G_{km}-f_m^{G,prac}\right)=0\\\sum\limits_{k\in\mathcal{K}_m}f^G_{km}\leqslant f_m^{G,prac},and f^{G,\min}_{km}\leqslant f^G_{km},k\in\mathcal{K}_m\\\lambda_f\geqslant 0\end{cases}.\tag{B.12}$$

More specifically, the derivative of (B.11) is given by

$$\begin{aligned}\frac{\partial\mathcal{L}\left(\mathbf{f}_m^G,\lambda_f\right)}{\partial f^G_{km}}=&-2^{\frac{F_kD_k}{B^*_{km}T_k\left(T_kf^G_{km}-F_k\right)}}\frac{C_2\ln 2F_kD_k}{B^*_{km}(T_kf^G_{km}-F_k)^2}\\&-\frac{B^*_{km}n_0F_k}{h^G_{km}(f^G_{km})^2}+\tau\left(f^G_{km}\left(t\right)\right)+\lambda_f,\quad k\in\mathcal{K}_m.\end{aligned}\tag{B.13}$$

It can be observed that $\frac{\partial\mathcal{L}\left(\mathbf{f}_m^G,\lambda_f\right)}{\partial f^G_{km}}$ increases with f^G_{km}. We consider the following two cases, i.e., $\lambda_f>0$ and $\lambda_f=0$.

Case 1: When $\lambda_f>0$, according to the fact that $\lambda_f\left(\sum\limits_{k\in\mathcal{K}_m}f^G_{km}-f_m^{G,prac}\right)=0$, we have $\sum\limits_{k\in\mathcal{K}_m}f^G_{km}=f_m^{G,prac}$.

In addition, since $\frac{\partial \mathcal{L}\left(\mathbf{f}_m^G, \lambda_f\right)}{\partial f_{km}^G} = 0, k \in \mathcal{K}_m$, we have

$$\begin{aligned} & 2^{\frac{F_k D_k}{B_{km}^* T_k \left(T_k f_{km}^G - F_k\right)}} \frac{C_2 \ln 2 F_k D_k}{B_{km}^* (T_k f_{km}^G - F_k)^2} + \frac{B_{km}^* n_0 F_k}{h_{km}^G (f_{km}^G)^2} \\ & = \lambda_f + \tau\left(f_{km}^G(t)\right), \quad k \in \mathcal{K}_m. \end{aligned} \tag{B.14}$$

Define the function $\Psi_{km}^G(y)$ as in (3.49). Then, we obtain the optimal f_{km}^G by (3.47) and λ_f is obtained by (3.48). Since $\Psi_{km}^G(y)$ decreases with y, we can determine that ${\Psi_{km}^G}^{-1}\left(\lambda_f + \tau\left(f_{km}^G(t)\right)\right)$ decreases with λ_f.

Owing to $\lambda_f > 0$, we have $\sum\limits_{k \in \mathcal{K}_m} \max\left({\Psi_{km}^G}^{-1}\left(\tau\left(f_{km}^G(t)\right)\right), f_{km}^{G,\min}\right) > f_m^{G,prac}$.

Case 2: When $\lambda_f = 0$, f_{km}^G is obtained by letting $\lambda_f = 0$ as shown in (3.50). According to the constraint $\sum\limits_{k \in \mathcal{K}_m} f_{km}^G \leqslant f_m^{G,prac}$, we have $\sum\limits_{k \in \mathcal{K}_m} \max\left({\Psi_{km}^G}^{-1}\left(\tau\left(f_{km}^G(t)\right)\right), f_{km}^{G,\min}\right) \leqslant f_m^{G,prac}$, which completes the proof.

Appendix C

Proofs for Chapter 4

C.1 Proof of Lemma 4.1

First, given a_{km}, we have $N_{km} = \frac{\beta_k D_k}{T R_{km}} = \frac{\beta_k D_k}{T \frac{B\lambda_g}{\ln 2}\Phi\left(\frac{N_0\lambda_g}{h_{km}^L p_k}\right)}$. According to the distribution of small-scale fading, we have

$$\begin{aligned}
&N_{km}\mathbb{P}\left\{R_{km}(t) \leqslant R_{km}\right\} \\
&= \frac{\beta_k D_k}{T\frac{B\lambda_g}{\ln 2}\Phi\left(\frac{N_0\lambda_g}{h_{km}^L p_k}\right)}\mathbb{P}\left\{\log_2\left(1+\frac{h_{km}^L p_k g_{km}(t)}{N_0}\right) \leqslant \frac{\lambda_g}{\ln 2}\Phi\left(\frac{N_0\lambda_g}{h_{km}^L p_k}\right)\right\} \\
&= \frac{\beta_k D_k \ln 2}{TB\lambda_g\Phi\left(\frac{N_0\lambda_g}{h_{km}^L p_k}\right)}\mathbb{P}\left\{g_{km}(t) \leqslant \left(2^{\frac{\lambda_g}{\ln 2}\Phi\left(\frac{N_0\lambda_g}{h_{km}^L p_k}\right)}-1\right)\frac{N_0}{h_{km}^L p_k}\right\} \\
&= \frac{\beta_k D_k \ln 2}{TB\lambda_g\Phi\left(\frac{N_0\lambda_g}{h_{km}^L p_k}\right)}\left\{1-\exp\left[-\left(2^{\frac{\lambda_g}{\ln 2}\Phi\left(\frac{N_0\lambda_g}{h_{km}^L p_k}\right)}-1\right)\frac{N_0\lambda_g}{h_{km}^L p_k}\right]\right\}.
\end{aligned} \tag{C.1}$$

Then, according to $N_{km}\mathbb{P}\left\{R_{km}(t) \leqslant R_{km}\right\} \leqslant \mu_o, \quad \forall k \in \mathcal{K}$, we have

$$\frac{1-\exp\left[-\left(2^{\frac{\lambda_g}{\ln 2}\Phi\left(\frac{N_0\lambda_g}{h_{km}^L p_k}\right)}-1\right)\frac{N_0\lambda_g}{h_{km}^L p_k}\right]}{\frac{TB\lambda_g}{\beta_k D_k \ln 2}\Phi\left(\frac{N_0\lambda_g}{h_{km}^L p_k}\right)} \leqslant \mu_o, \quad \forall k \in \mathcal{K}. \tag{C.2}$$

Considering the user association a_{km}, we have

$$\sum_{m=1}^{M} a_{km}\left\{1-\exp\left[-\left(2^{\frac{\lambda_g}{\ln 2}\Phi\left(\frac{N_0\lambda_g}{h_{km}^L p_k}\right)}-1\right)\frac{N_0\lambda_g}{h_{km}^L p_k}\right]\right\}\frac{\beta_k D_k \ln 2}{TB\lambda_g\Phi\left(\frac{N_0\lambda_g}{h_{km}^L p_k}\right)} \leqslant \mu_o, \quad \forall k \in \mathcal{K}. \tag{C.3}$$

C.2 Proof of Proposition 4.2

Define a function Z_k as follows

$$\mathrm{Z}_k = \frac{{\gamma_k}^2 D_k}{R^*_{km}} + \frac{{\gamma_k}^2 F_k}{f_{km}} - T_k. \tag{C.4}$$

Then, the Hessian matrix $\mathbf{H}_{\mathrm{Z}_k}$ of Z_k with respect to γ_k and f_{km} is given by

$$\mathbf{H}_{\mathrm{Z}_k} = \begin{bmatrix} \frac{2D_k}{R^*_{km}} + \frac{2F_k}{f_{km}} & -2\gamma_k \frac{F_k}{{f_{km}}^2} \\ -2\gamma_k \frac{F_k}{{f_{km}}^2} & 2{\gamma_k}^2 \frac{F_k}{{f_{km}}^3} \end{bmatrix}. \tag{C.5}$$

It can be observed that the first-order principal minors are non-negative, i.e., $\frac{\partial^2 \mathrm{Z}_k}{\partial {\gamma_k}^2} \geqslant 0$, $\frac{\partial^2 \mathrm{Z}_k}{\partial {f_{km}}^2} \geqslant 0$. In addition, the second-order principal minor, given by $|\mathbf{H}_{\mathrm{Z}_k}| = \frac{4D_k {\gamma_k}^2 F_k}{R^*_{km} {f_{km}}^3} \geqslant 0$, is also non-negative. Thus, according to Sylvester's criterion, the Hessian matrix $\mathbf{H}_{\mathrm{Z}_k}$ is a positive semi-definite matrix, which indicates that $\mathbf{H}_{\mathrm{Z}_k}$ is jointly convex for γ_k and f_{km}.

C.3 Proof of Theorem 4.3

If $c_k \leqslant 0$, $L(\mathbf{c}, \mathbf{f}, \varpi, \vartheta)$ is a quadratic function with respect to γ_k, and it is monotonically increasing with respect to γ_k when $\gamma_k \geqslant 0$. Therefore, the optimal γ_k is obtained at its lower bound value γ_k^{lb} which is obtained from (4.8b).

If $c_k > 0$, $L(\mathbf{c}, \mathbf{f}, \varpi, \vartheta)$ is a quadratic function and it is convex with respect to γ_k. By taking the first order derivative of $L(\mathbf{c}, \mathbf{f}, \varpi, \vartheta)$ with respect to γ_k, we have

$$\frac{\partial L(\mathbf{c}, \mathbf{f}, \varpi, \vartheta)}{\partial \gamma_k} = -2\gamma_k^i c_k + \vartheta_k 2\gamma_k a^*_{km} \left(\frac{D_k}{R^*_{km}} + \frac{F_k}{f_{km}} \right). \tag{C.6}$$

Letting $\frac{\partial L(\mathbf{c}, \mathbf{f}, \varpi, \vartheta)}{\partial \gamma_k} = 0$, we obtain γ_k^{opt} which minimizes $L(\mathbf{c}, \mathbf{f}, \varpi, \vartheta)$. From (4.8a), we obtain the upper bound of γ_k, i.e., γ_k^{ub}, which is given in (4.19), and the optimal γ_k is expressed in (4.16) by considering lower bound and upper bound of γ_k.

Given the optimized $\mathbf{c}^{j+1}$, for the kth user connected with the mth edge server, the first-order derivative of $L\left(\mathbf{c}^{j+1},\mathbf{f},\varpi,\vartheta\right)$ with respect to f_{km} is given by

$$\frac{\partial L\left(\mathbf{c}^{j+1},\mathbf{f},\varpi,\vartheta\right)}{\partial f_{km}}=\varpi_m-\vartheta_k\left(\gamma_k^{j+1}\right)^2\frac{F_k}{{f_{km}}^2}. \tag{C.7}$$

Letting $\frac{\partial L\left(\mathbf{c}^{j+1},\mathbf{f},\varpi,\vartheta\right)}{\partial f_{km}}=0$, we obtain the optimal f_{km}, i.e., f_{km}^{opt}, as follows

$$f_{km}^{opt}=\sqrt{\frac{\vartheta_k\left(\gamma_k^{j+1}\right)^2F_k}{\varpi_m}}. \tag{C.8}$$

In addition, the computation resource allocation needs to meet user's latency requirement. According to (4.8a), we have $f_{km}\geqslant f_{km}^{lb}=\frac{F_k}{\frac{T_k}{\left(\gamma_k^{j+1}\right)^2}-\frac{D_k}{R_{km}^*}}$. Combining (C.8) and f_{km}^{lb}, we obtain (4.20), which completes the proof.

C.4 Proof of Theorem 4.4

Due to the fact that $\frac{\partial\Phi(x)x}{\partial x}=\Phi\left(x\right)x+\Phi\left(x\right)-1\geqslant 0$, the function $\Phi\left(x\right)x$ increases with x. Thus, the objective value in (4.24) decreases with η_k, and the optimal η_k^* takes as large value as possible. Since $\frac{\partial\Phi(x)}{\partial x}=\Phi\left(x\right)-\frac{1}{x}\leqslant 0$, the LHS of (4.24a) increases with η_k. The upper bound of η_k in (4.24a), i.e., $\eta_k^{ub,1}$, is obtained by taking equality in (4.24a). In addition, the LHS of (4.24b) is an increasing function with respect to η_k. Thus, the upper bound of η_k in (4.24b), i.e., $\eta_k^{ub,2}$, is obtained by taking equality in (4.24b). Combining the upper bounds (i.e., $\eta_k^{ub,1}$, $\eta_k^{ub,2}$) and the lower bound of η_k (i.e., $\eta_k^{lb}=\frac{N_0\lambda_g}{h_{km}^Lp_k^{\max}}$), we obtain the desired results.

Appendix D

Proofs for Chapter 5

D.1 Proof of Lemma 5.1

In the Cartesian 3D coordinate system as shown in Fig. 5.2, the BS is located at the origin. Denote the projection of an arbitrary ARIS on the ground by $\mathbf{a_0} = (x_a, y_a, 0)$, where x_a is the ARIS coordinate on x-axis and y_a is the ARIS coordinate on y-axis. Then, the horizontal distance between the BS and $\mathbf{a_0}$ is given by $l_{BA}^h = \sqrt{{x_a}^2 + {y_a}^2}$. The conditional cumulative distribution function (CDF) of l_{BA}^h is given by

$$\begin{aligned}
&F_{L_{BA}^h}\left(l_{BA}^h \middle| x_A\right) \\
&= \iint_{\sqrt{{x_a}^2+{y_a}^2} \leqslant l_{BA}^h} f_{\mathbf{A}}\left(x_a - x_A, y_a\right) dx_a dy_a \\
&= \int_{x_a=-l_{BA}^h}^{x_a=l_{BA}^h} \int_{y_a=-\sqrt{\left(l_{BA}^h\right)^2 - {x_a}^2}}^{y_a=\sqrt{\left(l_{BA}^h\right)^2 - {x_a}^2}} f_{\mathbf{A}}\left(x_a - x_A, y_a\right) dx_a dy_a,
\end{aligned} \tag{D.1}$$

where $f_{\mathbf{A}}(\cdot)$ is given in (5.40).

Thus, the conditional PDF of L_{BA}^h is obtained by Leibniz's rule for differentiation, as follows

$$\begin{aligned}
f_{L_{BA}^h}\left(l_{BA}^h \middle| x_A\right) &= \frac{\partial F_{L_{BA}^h}\left(l_{BA}^h \middle| x_A\right)}{\partial l_{BA}^h} \\
&= \int_{-l_{BA}^h}^{l_{BA}^h} \frac{l_{BA}^h \left[f_{\mathbf{A}}\left(x_a - x_A, \sqrt{\left(l_{BA}^h\right)^2 - {x_a}^2}\right) + f_{\mathbf{A}}\left(x_a - x_A, -\sqrt{\left(l_{BA}^h\right)^2 - {x_a}^2}\right)\right]}{\sqrt{\left(l_{BA}^h\right)^2 - {x_a}^2}} dx_a.
\end{aligned} \tag{D.2}$$

By substituting (5.40) in (D.2), we have

$$\begin{aligned} f_{L_{BA}^h}\left(l_{BA}^h \middle| x_A\right) &= \frac{2}{\pi {R_A}^2}\int_{\max\left(-l_{BA}^h,\frac{{x_A}^2+\left(l_{BA}^h\right)^2-{R_A}^2}{2x_A}\right)}^{l_{BA}^h} \frac{l_{BA}^h}{\sqrt{\left(l_{BA}^h\right)^2-{x_a}^2}}dx_a \\ &\overset{(a)}{=} \frac{2l_{BA}^h}{\pi {R_A}^2}\int_{\frac{{x_A}^2+\left(l_{BA}^h\right)^2-{R_A}^2}{2x_A}}^{l_{BA}^h} \frac{1}{\sqrt{\left(l_{BA}^h\right)^2-{x_a}^2}}dx_a \\ &\overset{(b)}{=} \frac{2l_{BA}^h}{\pi {R_A}^2}\int_0^{\arccos\left(\frac{{x_A}^2+\left(l_{BA}^h\right)^2-{R_A}^2}{2x_A l_{BA}^h}\right)} 1 d\tau \\ &= \frac{2l_{BA}^h}{\pi {R_A}^2}\arccos\left(\frac{{x_A}^2+\left(l_{BA}^h\right)^2-{R_A}^2}{2x_A l_{BA}^h}\right), \end{aligned} \tag{D.3}$$

where (a) is obtained because if $x_A \geqslant R_A$, then $x_A - R_A \leqslant l_{BA}^h \leqslant x_A + R_A$, thus $\frac{{x_A}^2+\left(l_{BA}^h\right)^2-{R_A}^2}{2x_A} \geqslant -l_{BA}^h$; if $x_A < R_A$, then $R_A - x_A \leqslant l_{BA}^h \leqslant R_A + x_A$, thus $\frac{{x_A}^2+\left(l_{BA}^h\right)^2-{R_A}^2}{2x_A} \geqslant -l_{BA}^h$, in (b), we let $x_a = l_{BA}^h \cos\tau$.

The distance between the BS and an arbitrary ARIS is given by $l_{BA} = \sqrt{\left(l_{BA}^h\right)^2 + H^2}$. Therefore, the conditional PDF of the l_{BA} is given by

$$\begin{aligned} f_{L_{BA}}\left(l_{BA} \middle| x_A, H\right) &= f_{L_{BA}^h}\left(l_{BA}^h \middle| x_A\right)\left|\left|\frac{\partial l_{BA}^h}{\partial l_{BA}}\right|\right|_{l_{BA}^h=\sqrt{{l_{BA}}^2-H^2}} \\ &= \frac{2l_{BA}}{\pi {R_A}^2}\arccos\left(\frac{{x_A}^2+{l_{BA}}^2-H^2-{R_A}^2}{2x_A\sqrt{{l_{BA}}^2-H^2}}\right), \end{aligned} \tag{D.4}$$

which completes the proof.

D.2 Proof of Lemma 5.2

As shown in Fig. 5.2, for an arbitrary ground user, we denote the distance between the ground user and the projection of the SARIS coordinate on the ground by $l_{U c_A}$. Denote the horizontal distance between the ground user and the projection of an arbitrary ARIS on the ground by l_{AU}^h. Conditioned on l_{Uc_A}, the conditional PDF of l_{AU}^h is given by

$$f_{L_{AU}^h}\left(l_{AU}^h\middle|\, l_{Uc_A}\right) = \frac{2l_{AU}^h}{\pi {R_A}^2}\Omega\left(l_{Uc_A}, l_{AU}^h, R_A\right), \ \max\{0, l_{Uc_A} - R_A\} \leqslant l_{AU}^h \leqslant l_{Uc_A} + R_A. \quad \text{(D.5)}$$

Furthermore, if $x_{AU} \geqslant R_U$, the conditional PDF of l_{Uc_A} is given by

$$f_{L_{Uc_A}}\left(l_{Uc_A}\middle|\, x_A\right) = \frac{2l_{Uc_A}}{\pi {R_U}^2}\Omega\left(x_U - x_A, l_{Uc_A}, R_U\right), \ 0 \leqslant l_{Uc_A} \leqslant x_{AU} + R_U. \quad \text{(D.6)}$$

If $x_{AU} < R_U$, the conditional PDF of l_{Uc_A} is given by

$$f_{L_{Uc_A}}\left(l_{Uc_A}\middle|\, x_A\right) = \begin{cases} \frac{2l_{Uc_A}}{\pi {R_U}^2}\Omega\left(x_U - x_A, l_{Uc_A}, R_U\right), \ R_U - x_{AU} \leqslant l_{Uc_A} \leqslant x_{AU} + R_U \\ \frac{2l_{Uc_A}}{{R_U}^2}, \ 0 \leqslant l_{Uc_A} \leqslant R_U - x_{AU} \end{cases}, \quad \text{(D.7)}$$

Note that in (D.7), if $0 \leqslant l_{Uc_A} \leqslant R_U - x_{AU}$, we have $\mathbb{P}\{0 \leqslant l_{Uc_A} \leqslant R_U - x_{AU}\} = \frac{(R_U - x_U + x_A)^2}{{R_U}^2}$ and $f_{L_{Uc_A}}\left(l_{Uc_A}\middle|\, 0 \leqslant l_{Uc_A} \leqslant R_U - x_{AU}\right) = \frac{2l_{Uc_A}}{(R_U - x_U + x_A)^2}$. Therefore, we have

$$\begin{aligned} & f_{L_{Uc_A}}\left(l_{Uc_A}, 0 \leqslant l_{Uc_A} \leqslant R_U - x_{AU}\middle|\, x_A\right) \\ & = \mathbb{P}\{0 \leqslant l_{Uc_A} \leqslant R_U - x_{AU}\} f_{L_{Uc_A}}\left(l_{Uc_A}\middle|\, 0 \leqslant l_{Uc_A} \leqslant R_U - x_{AU}\right) \\ & = \frac{2l_{Uc_A}}{{R_U}^2}. \end{aligned} \quad \text{(D.8)}$$

Now, we are in the position of removing the condition of l_{Uc_A} from l_{AU}^h. The conditional PDF of l_{AU}^h, which is conditioned on x_A, is given by

$$\begin{aligned} & \text{if } x_{AU} \geqslant R_U, \text{then } f_{L_{AU}^h}\left(l_{AU}^h\middle|\, x_A\right) \\ & = \int_{x_{AU}-R_U}^{x_{AU}+R_U} \frac{2l_{AU}^h}{\pi {R_A}^2}\Omega\left(x, l_{AU}^h, R_A\right) \frac{2x}{\pi {R_U}^2}\Omega\left(x_U - x_A, x, R_U\right) dx, \end{aligned} \quad \text{(D.9)}$$

and

$$\begin{aligned} & \text{if } x_{AU} < R_U, \text{then } f_{L_{AU}^h}\left(l_{AU}^h\middle|\, x_A\right) \\ & = \frac{4l_{AU}^h}{\pi {R_A}^2 {R_U}^2} \int_{R_U - x_{AU}}^{x_{AU}+R_U} \frac{x}{\pi}\Omega\left(x, l_{AU}^h, R_A\right)\Omega\left(x_U - x_A, x, R_U\right) dx \\ & + \frac{4l_{AU}^h}{\pi {R_A}^2 {R_U}^2} \int_0^{R_U - x_{AU}} x\Omega\left(x, l_{AU}^h, R_A\right) dx. \end{aligned} \quad \text{(D.10)}$$

The distance between an arbitrary ARIS and an arbitrary user is given by $l_{AU} = \sqrt{\left(l_{AU}^h\right)^2 + H^2}$. Then, conditioned on x_A and H, the conditional PDF of l_{AU} can be obtained by $f_{L_{AU}}\left(l_{AU} \middle| x_A, H\right) = f_{L_{AU}^h}\left(l_{AU}^h \middle| x_{AU}\right) \left|\frac{\partial l_{AU}^h}{\partial l_{AU}}\right| \Big|_{l_{AU}^h = \sqrt{l_{AU}^2 - H^2}}$, which completes the proof.

Bibliography

[1] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, “5g: A tutorial overview of standards, trials, challenges, deployment, and practice,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[2] B. Shang, L. Liu, J. Ma, and P. Fan, “Unmanned aerial vehicle meets vehicle-to-everything in secure communications,” *IEEE Communications Magazine*, vol. 57, no. 10, pp. 98–103, 2019.

[3] Y. Saleem, M. H. Rehmani, and S. Zeadally, “Integration of cognitive radio technology with unmanned aerial vehicles: issues, opportunities, and future research challenges,” *Journal of Network and Computer Applications*, vol. 50, pp. 15–31, 2015.

[4] D. Lee, J. Lim, and H. Baek, “An airborne communication relay scheme for ieee 802.11 wlan based network,” in *2018 International Conference on Information Networking (ICOIN)*, pp. 426–431, 2018.

[5] N. Zhang, P. Yang, J. Ren, D. Chen, L. Yu, and X. Shen, “Synergy of big data and 5g wireless networks: Opportunities, approaches, and challenges,” *IEEE Wireless Communications*, vol. 25, no. 1, pp. 12–18, 2018.

[6] H. Guo, J. Liu, and J. Lv, “Toward intelligent task offloading at the edge,” *IEEE Network*, vol. 34, no. 2, pp. 128–134, 2020.

[7] X. Diao, J. Zheng, Y. Cai, Y. Wu, and A. Anpalagan, “Fair data allocation and tra-

jectory optimization for uav-assisted mobile edge computing," *IEEE Communications Letters*, vol. 23, no. 12, pp. 2357–2361, 2019.

[8] W. Zhang, L. Li, N. Zhang, T. Han, and S. Wang, "Air-ground integrated mobile edge networks: A survey," *IEEE Access*, vol. 8, pp. 125998–126018, 2020.

[9] B. Shang and L. Liu, "Mobile Edge Computing in the Sky: Energy Optimization for Air-Ground Integrated Networks," *IEEE Internet of Things Journal*, pp. 1–1, 2020.

[10] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng, "Coding metamaterials, digital metamaterials and programmable metamaterials," *Light: Science & Applications*, vol. 3, no. 10, pp. e218–e218, 2014.

[11] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[12] H. Lu, Y. Zeng, S. Jin, and R. Zhang, "Enabling panoramic full-angle reflection via aerial intelligent reflecting surface," 2020.

[13] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, 2020.

[14] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," *IEEE Communications Magazine*, vol. 54, pp. 36–42, May 2016.

[15] J. Kakar and V. Marojevic, "Waveform and spectrum management for unmanned aerial systems beyond 2025," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–5, Oct 2017.

[16] B. Shang, L. Zhao, K. Chen, and X. Chu, "Wireless-powered device-to-device-assisted offloading in cellular networks," *IEEE Transactions on Green Communications and Networking*, vol. 2, pp. 1012–1026, Dec 2018.

[17] S. Wu, R. Atat, N. Mastronarde, and L. Liu, "Improving the coverage and spectral efficiency of millimeter-wave cellular networks using device-to-device relays," *IEEE Transactions on Communications*, vol. 66, no. 5, pp. 2251–2265, 2018.

[18] B. Shang, L. Zhao, K. C. Chen, and X. Chu, "An economic aspect of device-to-device assisted offloading in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 17, pp. 2289–2304, April 2018.

[19] H. Chen, L. Liu, J. D. Matyjas, and M. J. Medley, "Optimal resource allocation for sensing-based spectrum sharing d2d networks," *Computers and Electrical Engineering*, vol. 44, pp. 107 – 121, 2015.

[20] B. Shang, L. Zhao, and K. C. Chen, "Enabling device-to-device communications in LTE-unlicensed spectrum," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, May 2017.

[21] X. Lin, J. G. Andrews, and A. Ghosh, "Spectrum sharing for device-to-device communication in cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 6727–6740, Dec 2014.

[22] Y. R. Ramadan, H. Minn, and Y. Dai, "A new paradigm for spectrum sharing between cellular wireless communications and radio astronomy systems," *IEEE Transactions on Communications*, vol. 65, pp. 3985–3999, Sep. 2017.

[23] Z. Chen and M. Kountouris, "Decentralized opportunistic access for D2D underlaid

cellular networks," *IEEE Transactions on Communications*, vol. 66, pp. 4842–4853, Oct 2018.

[24] H. Chen, L. Liu, T. Novlan, J. D. Matyjas, B. L. Ng, and J. Zhang, "Spatial spectrum sensing-based device-to-device cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, pp. 7299–7313, Nov 2016.

[25] R. Atat, L. Liu, H. Chen, J. Wu, H. Li, and Y. Yi, "Enabling cyber-physical communication in 5G cellular networks: Challenges, spatial spectrum sensing, and cyber-security," *IET Cyber-Physical Systems: Theory Applications*, vol. 2, no. 1, pp. 49–54, 2017.

[26] H. Chen, L. Liu, H. S. Dhillon, and Y. Yi, "Qos-aware d2d cellular networks with spatial spectrum sensing: A stochastic geometry view," *IEEE Transactions on Communications*, vol. 67, pp. 3651–3664, May 2019.

[27] B. Shang, L. Liu, H. Chen, J. Zhang, S. Pudlewski, E. S. Bentley, and J. Ashdown, "Spatial Spectrum Sensing-Based D2D Communications in User-Centric Deployed HetNets," in *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Dec 2019.

[28] C. Zhang and W. Zhang, "Spectrum sharing for drone networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, pp. 136–144, Jan 2017.

[29] H. Wang, J. Wang, G. Ding, J. Chen, Y. Li, and Z. Han, "Spectrum sharing planning for full-duplex UAV relaying systems with underlaid D2D communications," *IEEE Journal on Selected Areas in Communications*, vol. 36, pp. 1986–1999, Sep. 2018.

[30] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Unmanned aerial vehicle with

underlaid device-to-device communications: Performance and tradeoffs," *IEEE Transactions on Wireless Communications*, vol. 15, pp. 3949–3963, June 2016.

[31] M. M. Selim, M. Rihan, Y. Yang, L. Huang, Z. Quan, and J. Ma, "On the outage probability and power control of D2D underlaying NOMA UAV-assisted networks," *IEEE Access*, vol. 7, pp. 16525–16536, 2019.

[32] M. Rihan, M. M. Selim, C. Xu, and L. Huang, "D2D communication underlaying UAV on multiple bands in disaster area: Stochastic geometry analysis," *IEEE Access*, vol. 7, pp. 156646–156658, 2019.

[33] M. Bacha, Y. Wu, and B. Clerckx, "Downlink and uplink decoupling in two-tier heterogeneous networks with multi-antenna base stations," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 2760–2775, May 2017.

[34] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *2014 IEEE Global Communications Conference*, pp. 2898–2904, Dec 2014.

[35] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, pp. 569–572, Dec 2014.

[36] W. Khawaja, I. Guvenc, and D. Matolak, "UWB channel sounding and modeling for UAV air-to-ground propagation channels," in *2016 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–7, Dec 2016.

[37] E. Yanmaz, R. Kuschnig, and C. Bettstetter, "Achieving air-ground communications in 802.11 networks with three-dimensional aerial mobility," in *2013 Proceedings IEEE INFOCOM*, pp. 120–124, April 2013.

[38] Y. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 1326–1337, April 2008.

[39] X. Wu and Z. Ma, "Modeling and performance analysis of cellular and device-to-device heterogeneous networks," in *2017 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, Dec 2017.

[40] M. Haenggi and R. K. Ganti, "Interference in large wireless networks," *Foundations and Trends® in Networking*, vol. 3, no. 2, pp. 127–248, 2009.

[41] Z. Chen, C.-X. Wang, X. Hong, J. S. Thompson, S. A. Vorobyov, X. Ge, H. Xiao, and F. Zhao, "Aggregate interference modeling in cognitive radio networks with power and contention control," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 456–468, 2012.

[42] B. Shang and L. Liu, "Machine learning meets point process: Spatial spectrum sensing in user-centric networks," *IEEE Wireless Communications Letters*, vol. 9, pp. 34–37, Jan 2020.

[43] H. Qu, R. Tang, J. Zhao, and Z. Cao, "Performance tradeoff between user fairness and energy conservation in downlink NOMA systems," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, May 2018.

[44] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet of Things Journal*, vol. 5, pp. 450–465, Feb 2018.

[45] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Communications Surveys Tutorials*, vol. 19, pp. 2322–2358, Fourthquarter 2017.

[46] C. You, K. Huang, H. Chae, and B. Kim, “Energy-efficient resource allocation for mobile-edge computation offloading,” *IEEE Transactions on Wireless Communications*, vol. 16, pp. 1397–1411, March 2017.

[47] Z. Zhou, J. Feng, L. Tan, Y. He, and J. Gong, “An air-ground integration approach for mobile edge computing in iot,” *IEEE Communications Magazine*, vol. 56, pp. 40–47, August 2018.

[48] C. Wang, C. Liang, F. R. Yu, Q. Chen, and L. Tang, “Computation offloading and resource allocation in wireless cellular networks with mobile edge computing,” *IEEE Transactions on Wireless Communications*, vol. 16, pp. 4924–4938, Aug 2017.

[49] L. Yang, H. Zhang, X. Li, H. Ji, and V. C. M. Leung, “A distributed computation offloading strategy in small-cell networks integrated with mobile edge computing,” *IEEE/ACM Transactions on Networking*, vol. 26, pp. 2762–2773, Dec 2018.

[50] J. Feng, Q. Pei, F. R. Yu, X. Chu, and B. Shang, “Computation offloading and resource allocation for wireless powered mobile edge computing with latency constraint,” *IEEE Wireless Communications Letters*, vol. 8, pp. 1320–1323, Oct 2019.

[51] S. Jeong, O. Simeone, and J. Kang, “Mobile edge computing via a uav-mounted cloudlet: Optimization of bit allocation and path planning,” *IEEE Transactions on Vehicular Technology*, vol. 67, pp. 2049–2063, March 2018.

[52] J. Zhang, L. Zhou, Q. Tang, E. C. . Ngai, X. Hu, H. Zhao, and J. Wei, “Stochastic computation offloading and trajectory scheduling for uav-assisted mobile edge computing,” *IEEE Internet of Things Journal*, vol. 6, pp. 3688–3699, April 2019.

[53] M. Hua, Y. Wang, C. Li, Y. Huang, and L. Yang, “Uav-aided mobile edge computing

systems with one by one access scheme," *IEEE Transactions on Green Communications and Networking*, vol. 3, pp. 664–678, Sep. 2019.

[54] Q. Hu, Y. Cai, G. Yu, Z. Qin, M. Zhao, and G. Y. Li, "Joint offloading and trajectory design for uav-enabled mobile edge computing systems," *IEEE Internet of Things Journal*, vol. 6, pp. 1879–1892, April 2019.

[55] Y. Zhou, C. Pan, P. L. Yeoh, K. Wang, M. Elkashlan, B. Vucetic, and Y. Li, "Secure communications for uav-enabled mobile edge computing systems," *IEEE Transactions on Communications*, pp. 1–1, 2019.

[56] Z. Yang, C. Pan, K. Wang, and M. Shikh-Bahaei, "Energy efficient resource allocation in uav-enabled mobile edge computing networks," *IEEE Transactions on Wireless Communications*, vol. 18, pp. 4576–4589, Sep. 2019.

[57] Y. Wang, Z. Ru, K. Wang, and P. Huang, "Joint deployment and task scheduling optimization for large-scale mobile users in multi-uav-enabled mobile edge computing," *IEEE Transactions on Cybernetics*, pp. 1–14, 2019.

[58] N. Cheng, W. Xu, W. Shi, Y. Zhou, N. Lu, H. Zhou, and X. Shen, "Air-ground integrated mobile edge networks: Architecture, challenges, and opportunities," *IEEE Communications Magazine*, vol. 56, pp. 26–32, August 2018.

[59] R. Amorim, H. Nguyen, P. Mogensen, I. Z. Kovács, J. Wigard, and T. B. Sørensen, "Radio channel modeling for uav communication over cellular networks," *IEEE Wireless Communications Letters*, vol. 6, pp. 514–517, Aug 2017.

[60] M. E. T. Gerards, J. L. Hurink, and J. Kuper, "On the interplay between global dvfs and scheduling tasks with precedence constraints," *IEEE Transactions on Computers*, vol. 64, pp. 1742–1754, June 2015.

[61] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1." `http://cvxr.com/cvx`, Mar. 2014.

[62] X. Shen, S. Diamond, Y. Gu, and S. Boyd, "Disciplined convex-concave programming," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 1009–1014, IEEE, 2016.

[63] P. D. Tao *et al.*, "The dc (difference of convex functions) programming and dca revisited with dc models of real world nonconvex optimization problems," *Annals of operations research*, vol. 133, no. 1-4, pp. 23–46, 2005.

[64] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge university press, 2004.

[65] S. A. A. Shah, E. Ahmed, M. Imran, and S. Zeadally, "5G for Vehicular Communications," *IEEE Communications Magazine*, vol. 56, no. 1, pp. 111–117, 2018.

[66] C. Chen, L. Liu, T. Qiu, D. O. Wu, and Z. Ren, "Delay-Aware Grid-Based Geographic Routing in Urban VANETs: A Backbone Approach," *IEEE/ACM Transactions on Networking*, vol. 27, no. 6, pp. 2324–2337, 2019.

[67] H. Sami, A. Mourad, and W. El-Hajj, "Vehicular-OBUs-As-On-Demand-Fogs: Resource and Context Aware Deployment of Containerized Micro-Services," *IEEE/ACM Transactions on Networking*, vol. 28, no. 2, pp. 778–790, 2020.

[68] Z. Ning, J. Huang, and X. Wang, "Vehicular Fog Computing: Enabling Real-Time Traffic Management for Smart Cities," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 87–93, 2019.

[69] L. Liu, C. Chen, Q. Pei, S. Maharjan, and Y. Zhang, "Vehicular edge computing and networking: A survey," *arXiv preprint arXiv:1908.06849*, 2019.

[70] S. Liu, L. Liu, J. Tang, B. Yu, Y. Wang, and W. Shi, "Edge Computing for Autonomous Driving: Opportunities and Challenges," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1697–1716, 2019.

[71] S. Jošilo and G. Dán, "Computation Offloading Scheduling for Periodic Tasks in Mobile Edge Computing," *IEEE/ACM Transactions on Networking*, vol. 28, no. 2, pp. 667–680, 2020.

[72] L. Chen, S. Zhou, and J. Xu, "Computation Peer Offloading for Energy-Constrained Mobile Edge Computing in Small-Cell Networks," *IEEE/ACM Transactions on Networking*, vol. 26, no. 4, pp. 1619–1632, 2018.

[73] Y. Zhang, X. Lan, J. Ren, and L. Cai, "Efficient Computing Resource Sharing for Mobile Edge-Cloud Computing Networks," *IEEE/ACM Transactions on Networking*, pp. 1–14, 2020.

[74] L. Liang, G. Y. Li, and W. Xu, "Resource Allocation for D2D-Enabled Vehicular Communications," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 3186–3197, 2017.

[75] C. Guo, L. Liang, and G. Y. Li, "Resource Allocation for Vehicular Communications With Low Latency and High Reliability," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 3887–3902, 2019.

[76] J. Zhang, X. Hu, Z. Ning, E. C. . Ngai, L. Zhou, J. Wei, J. Cheng, and B. Hu, "Energy-Latency Tradeoff for Energy-Aware Offloading in Mobile Edge Computing Networks," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2633–2645, 2018.

[77] J. Zhao, Q. Li, Y. Gong, and K. Zhang, "Computation Offloading and Resource Al-

location For Cloud Assisted Mobile Edge Computing in Vehicular Networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7944–7956, 2019.

[78] Y. Liu, H. Yu, S. Xie, and Y. Zhang, "Deep Reinforcement Learning for Offloading and Resource Allocation in Vehicle Edge Computing and Networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 11158–11168, 2019.

[79] D. Ye, R. Yu, M. Pan, and Z. Han, "Federated Learning in Vehicular Edge Computing: A Selective Model Aggregation Approach," *IEEE Access*, vol. 8, pp. 23920–23935, 2020.

[80] Z. Zhou, J. Feng, Z. Chang, and X. Shen, "Energy-Efficient Edge Computing Service Provisioning for Vehicular Networks: A Consensus ADMM Approach," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 5087–5099, 2019.

[81] X. Li, Y. Dang, M. Aazam, X. Peng, T. Chen, and C. Chen, "Energy-Efficient Computation Offloading in Vehicular Edge Cloud Computing," *IEEE Access*, vol. 8, pp. 37632–37644, 2020.

[82] M. A. ElMossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, "Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 3, pp. 990–1002, 2020.

[83] Z. Yigit, E. Basar, and I. Altunbas, "Low complexity adaptation for reconfigurable intelligent surface-based MIMO systems," *IEEE Communications Letters*, vol. 24, no. 12, pp. 2946–2950, 2020.

[84] B. Shang, V. Marojevic, Y. Yi, A. S. Abdalla, and L. Liu, "Spectrum sharing for UAV communications: Spatial spectrum sensing and open issues," *IEEE Vehicular Technology Magazine*, vol. 15, no. 2, pp. 104–112, 2020.

[85] B. Shang, L. Liu, R. M. Rao, V. Marojevic, and J. H. Reed, “3D spectrum sharing for hybrid D2D and UAV networks,” *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5375–5389, 2020.

[86] M. Hua, L. Yang, Q. Wu, and A. L. Swindlehurst, “3D UAV trajectory and communication design for simultaneous uplink and downlink transmission,” *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5908–5923, 2020.

[87] Q. Wu, L. Liu, and R. Zhang, “Fundamental trade-offs in communication and trajectory design for UAV-enabled wireless network,” *IEEE Wireless Communications*, vol. 26, no. 1, pp. 36–44, 2019.

[88] *Technical Specification Group Radio Access Network; Study on Enhanced LTE Support for Aerial Vehicles (Release 15)*, document 3GPP TR 36.777 V15.0.0, 3rd Generation Partnership Project, Dec. 2017.

[89] B. Shang, R. Shafin, and L. Liu, “Uav swarm-enabled aerial reconfigurable intelligent surface (saris),” *IEEE Wireless Communications*, vol. 28, no. 5, pp. 156–163, 2021.

[90] S. Li, B. Duo, X. Yuan, Y. Liang, and M. Di Renzo, “Reconfigurable intelligent surface assisted UAV communication: Joint trajectory design and passive beamforming,” *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 716–720, 2020.

[91] X. Liu, Y. Liu, and Y. Chen, “Machine learning empowered trajectory and passive beamforming design in UAV-RIS wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 7, pp. 2042–2055, 2021.

[92] M. Hua, L. Yang, Q. Wu, C. Pan, C. Li, and A. L. Swindlehurst, “Uav-assisted intelligent reflecting surface symbiotic radio system,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 5769–5785, 2021.

[93] H. Lu, Y. Zeng, S. Jin, and R. Zhang, “Aerial intelligent reflecting surface: Joint placement and passive beamforming design with 3d beam flattening,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4128–4143, 2021.

[94] H. Long, M. Chen, Z. Yang, B. Wang, Z. Li, X. Yun, and M. Shikh-Bahaei, “Reflections in the sky: Joint trajectory and passive beamforming design for secure UAV networks with reconfigurable intelligent surface,” *arXiv preprint arXiv:2005.10559*, 2020.

[95] T. Shafique, H. Tabassum, and E. Hossain, “Optimization of wireless relaying with flexible UAV-borne reflecting surfaces,” *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 309–325, 2021.

[96] Q. Wu and R. Zhang, “Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.

[97] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, “Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 421–439, 2021.

[98] X. Yu, V. Jamali, D. Xu, D. W. K. Ng, and R. Schober, “Smart and reconfigurable wireless communications: From IRS modeling to algorithm design,” *arXiv preprint arXiv:2103.07046*, 2021.

[99] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, “Intelligent reflecting surface-aided wireless communications: A tutorial,” *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3313–3351, 2021.

[100] K. Shen and W. Yu, “Fractional programming for communication systems—part ii:

Uplink scheduling via matching," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2631–2644, 2018.

[101] K. Shen and W. Yu, "Fractional programming for communication systems—part i: Power control and beamforming," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, 2018.

[102] J. Gorski, F. Pfeuffer, and K. Klamroth, "Biconvex sets and optimization with biconvex functions: a survey and extensions," *Mathematical methods of operations research*, vol. 66, no. 3, pp. 373–407, 2007.

[103] D. P. Bertsekas, "Nonlinear programming," *Journal of the Operational Research Society*, vol. 48, no. 3, pp. 334–334, 1997.

[104] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted mmse approach to distributed sum-utility maximization for a mimo interfering broadcast channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.

[105] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 7, pp. 984–995, 1989.

[106] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing uav," *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2329–2345, 2019.

[107] S. Hu, Z. Wei, Y. Cai, C. Liu, D. W. K. Ng, and J. Yuan, "Robust and secure sum-rate maximization for multiuser miso downlink systems with self-sustainable irs," *IEEE Transactions on Communications*, pp. 1–1, 2021.

[108] B. Shang, L. Liu, and Z. Tian, “Deep learning-assisted energy-efficient task offloading in vehicular edge computing systems,” *IEEE Transactions on Vehicular Technology*, vol. 70, no. 9, pp. 9619–9624, 2021.

[109] B. Shang, Y. Yi, and L. Liu, “Computing over space-air-ground integrated networks: Challenges and opportunities,” *IEEE Network*, vol. 35, no. 4, pp. 302–309, 2021.

[110] H. Alzer, “On some inequalities for the incomplete gamma function,” *Mathematics of Computation of the American Mathematical Society*, vol. 66, no. 218, pp. 771–778, 1997.