

A Thesis for the Degree of Ph.D. in Engineering

AI-Powered Optimization Approaches for Antenna Beamforming in Multi-Cell HAPS Systems

September 2025

Graduate School of Science and Technology
Keio University

Wei Dai

Contents

List of Tables	iv
List of Figures	v
Abstract	ix
Acknowledgments	xiii
1 Introduction	1
1.1 Background	2
1.2 Motivation and Methodological Overview	6
1.2.1 Dynamic Beamforming	6
1.2.2 Clustering-Aided Beamforming Optimization	8
1.2.3 Neural Network-Based Beamforming Optimization	10
1.3 Outline and Contributions of the Dissertation	12
1.3.1 Outline of the Dissertation	12
1.3.2 Contributions	14
2 Related Work	19
2.1 Network Architecture and System Capabilities	20
2.2 Learning-Based Control and Optimization	22
2.3 Antenna Design and Beamforming Strategies	24
3 System Model	29
3.1 HAPS Model	30

3.2	Antenna Beamforming Model	32
4	A Clustering-Aided Optimization Algorithm for Antenna Beamforming in Multi-Cell HAPS Systems	37
4.1	Contributions	38
4.2	Limitations of Existing Work	40
4.3	Objective Function	42
4.4	Clustering-Aided Optimization Approach	45
4.4.1	PSO with Clustering-Aided Parameter Range Redefinition	45
4.4.2	Regulated K-means Algorithm	53
4.5	Evaluation	60
4.5.1	Experiment Specifications	60
4.5.2	Optimization Results	70
4.5.3	Throughput Performance	79
4.6	Conclusion	93
5	A Neural Network Approach for Beamforming Optimization in Multi-Cell HAPS Systems	95
5.1	Contributions	96
5.2	Limitations of Existing Work	97
5.3	Neural Network Approach	99
5.3.1	ConvNeXt V2-Based Neural Network Model	99
5.3.2	Antenna Parameter Normalization	104
5.3.3	Beam Assignment	107
5.4	Evaluation	109
5.4.1	Experiment Specifications	109
5.4.2	Optimization Performance	114
5.5	Conclusion	123
6	Conclusions and Future Work	125
6.1	Contributions	126

6.2	Future Work	127
Appendix A List of Author's Publications and Awards		131
A.1	Articles on Periodicals	131
A.2	Articles on International Conference Proceedings	131
A.3	Presentations at International Conferences	132
A.4	Articles on Domestic Conference Proceedings	132
A.5	Awards	132
References		133

List of Tables

1.1	Limitations of conventional multi-cell HAPS beamforming optimization approaches and the contributions of Chapter 4.	15
1.2	Limitations of conventional multi-cell HAPS beamforming optimization approaches and the contributions of Chapter 5.	16
2.1	Summary of the related work.	28
3.1	HAPS model parameters.	31
3.2	Antenna pattern parameters.	34
4.1	Simulation settings.	65
4.2	Fitness results across 100 UE distributions.	72
4.3	Fitness results in example areas.	74
4.4	Ranking analysis of how often each method outperforms the others across 100 UE distributions from 1st to 7th.	75
4.5	Ranking analysis of how often each method outperforms the others across 100 UE distributions from 8th to 14th.	76
4.6	Average final fitness score results of 100 UE distributions for additional exponent factors.	77
5.1	Training settings.	113

List of Figures

1.1	Illustration of a typical HAPS system functioning as a repeater.	3
1.2	Illustration of dynamic beam control based on UE distribution. HAPS continuously optimizes beams rather than relying on a fixed, uniform cell configuration.	7
1.3	Illustration of the structure of the dissertation.	12
1.4	Performance comparison of methods.	17
3.1	Illustration of the HAPS-enabled cellular networks.	30
3.2	Illustration of the four antenna pattern parameters, where h is the altitude of HAPS.	33
3.3	Antenna pattern of vertical or horizontal polarization for $\theta_{3\text{dB}} = 20^\circ$ and $\phi_{3\text{dB}} = 20^\circ$	35
4.1	Flowchart of the complete optimization process.	47
4.2	Illustration of the clustering-aided parameter range redefinition, where O is the location of HAPS, C_1, C_2, C_3 are the cluster centroids, d_1 is the maximum distance from centroid to UEs in cluster 1, and C'_1 is a candidate beam direction for beam 1. (a) Beamwidth search range (shaded area). (b) Direction angle search range (shaded area).	51
4.3	Illustration of the proposed clustering algorithm, where C is the cluster centroid and d is the maximum distance from centroid to objects. (a) K-means clustering results. (b) Regulated K-means clustering results. . .	56
4.4	Illustration of the complete regulated K-means algorithm iteration. . . .	57

4.5	Illustration of example HAPS coverage area 1 and coverage area 2 in the Kyushu region.	61
4.6	Illustration of example HAPS coverage area 3 in the Chugoku region. . .	62
4.7	Illustration of example HAPS coverage area 4 in the Ibaraki Prefecture. .	63
4.8	Illustration of example HAPS coverage area 5 in the Osaka Prefecture. . .	64
4.9	SINR distribution for the uniform cell configuration in the two example areas in Kyushu. (a) Example area 1. (b) Example area 2.	67
4.10	Comparison of multiple p values applied to the proposed method in terms of exponential weighting function.	69
4.11	Comparison of multiple p values applied to the proposed method in terms of convergence curve averaged across 100 UE distributions.	78
4.12	Throughput performance average across 100 UE distributions for $p = 4.265$. (a) CDF of UE throughput. (b) Distribution of throughput improvement over the uniform cell configuration from low throughput UEs to high throughput UEs.	80
4.13	Throughput performance in example area 1 for $p = 4.265$. (a) CDF of UE throughput. (b) Distribution of throughput improvement over the uniform cell configuration from low throughput UEs to high throughput UEs. . . .	82
4.14	Throughput performance in example area 2 for $p = 4.265$. (a) CDF of UE throughput. (b) Distribution of throughput improvement over the uniform cell configuration from low throughput UEs to high throughput UEs. . . .	83
4.15	Throughput performance in example area 3 for $p = 4.265$. (a) CDF of UE throughput. (b) Distribution of throughput improvement over the uniform cell configuration from low throughput UEs to high throughput UEs. . . .	84
4.16	Throughput performance in example area 4 for $p = 4.265$. (a) CDF of UE throughput. (b) Distribution of throughput improvement over the uniform cell configuration from low throughput UEs to high throughput UEs. . . .	85

4.17	Throughput performance in example area 5 for $p = 4.265$. (a) CDF of UE throughput. (b) Distribution of throughput improvement over the uniform cell configuration from low throughput UEs to high throughput UEs. . . .	86
4.18	Comparison of multiple p values applied to the proposed method in terms of throughput improvement distribution over the uniform cell configuration from low throughput UEs to high throughput UEs, averaged across 100 UE distributions.	87
4.19	Best SINR distribution achieved among all non-clustering-aided methods for $p = 4.265$. (a) Example area 1. (b) Example area 2.	88
4.20	SINR distribution achieved by K-means clustering-aided PSO for $p = 4.265$. (a) Example area 1. (b) Example area 2.	89
4.21	SINR distribution achieved by constrained K-means clustering-aided PSO for $p = 4.265$. (a) Example area 1. (b) Example area 2.	90
4.22	SINR distribution achieved by regulated K-means clustering-aided PSO for $p = 4.265$. (a) Example area 1. (b) Example area 2.	91
5.1	Illustration of the proposed neural network model for a 3-cell HAPS system.	101
5.2	Illustration of the beam-to-beam assignment between network outputs and ground truth data.	108
5.3	Illustration of two example HAPS coverage areas for 30 km coverage radius. (a) Coverage area 1 in the Mie Prefecture. (b) Coverage area 2 in the Chiba Prefecture.	111
5.4	Illustration of two example HAPS coverage areas for 50 km coverage radius. (a) Coverage area 1 in the Yamanashi Prefecture. (b) Coverage area 2 in the Nagano prefecture.	112
5.5	Average CDF of UE throughput. (a) 30 km coverage radius. (b) 50 km coverage radius.	115
5.6	CDF of UE throughput for 30 km coverage radius. (a) Example area 1. (b) Example area 2.	116

5.7	CDF of UE throughput for 50 km coverage radius. (a) Example area 1. (b) Example area 2.	117
5.8	SINR distribution in the two example areas for 30 km coverage radius. (a) Example area 1. (b) Example area 2.	118
5.9	SINR distribution in the two example areas for 50 km coverage radius. (a) Example area 1. (b) Example area 2.	119
5.10	SINR distribution in additional example areas for 30 km coverage radius.	121
5.11	SINR distribution in additional example areas for 50 km coverage radius.	121
5.12	CDF of UE throughput in additional example areas for 30 km coverage radius.	122
5.13	CDF of UE throughput in additional example areas for 50 km coverage radius.	122

AI-Powered Optimization Approaches for Antenna Beamforming in Multi-Cell HAPS Systems

Wei Dai
dai@ohtsuki.ics.keio.ac.jp.
Keio University, 2025

Supervisor: Prof. Tomoaki Ohtsuki, Ph.D.
ohtsuki@ics.keio.ac.jp

Abstract

High altitude platform station (HAPS) systems are recognized as a promising technology for the future evolution of wireless communications, particularly in the context of the Internet of Things (IoT) and the development of beyond 5G (B5G) and 6G networks. By operating at stratospheric altitudes, HAPS systems are able to provide extensive wide-area coverage, low latency, enhanced resilience, and cost-effective network access, particularly for remote and underserved regions where terrestrial or satellite infrastructure is insufficient. HAPS are expected to serve as essential components in space–air–ground integrated networks (SAGIN), contributing to massive IoT connectivity, emergency recovery, and seamless integration with terrestrial and non-terrestrial communication layers. The increasing importance of HAPS is underscored by their capability to enable line-of-sight (LoS) propagation, efficient multi-access edge computing, and stable connectivity in a variety of application scenarios, including real-time data acquisition and critical event response.

Nevertheless, the unique characteristics of HAPS, including their continuous movement, the variability in deployment environments, and the inherently dynamic and non-uniform distribution of user equipments (UEs), present significant technical challenges. A major issue lies in maintaining consistently high throughput for UEs across the coverage area, as the movement of HAPS and the mobility of UEs frequently result in sub-optimal cell configurations, uneven throughput distribution, and increased occurrence of low-throughput UEs. To maintain robust network performance and guarantee quality of

service, it is essential for HAPS systems to dynamically adjust antenna beamforming in real time, optimizing both beam direction and coverage to adapt to instantaneous UE distributions. However, conventional reactive or heuristic methods for dynamic beam control are insufficient in minimizing the number of low throughput UEs and generally exhibit high computational complexity, making them unsuitable for practical, large-scale, and latency-sensitive HAPS operations.

Chapter 1 provides a comprehensive introduction to the context and motivation for this research, including a detailed overview of HAPS technology, its integration in future wireless network architectures, and a summary of the technical challenges in dynamic cell configuration and antenna beamforming. The chapter also introduces the structure of the thesis, outlines the research objectives, and summarizes the primary contributions, including the sequential development of two novel approaches to HAPS beamforming optimization.

Chapter 2 presents a comprehensive review of existing research on the performance enhancement of HAPS networks. The chapter highlights how prior studies have addressed key challenges such as coverage reliability, dynamic control, and throughput optimization under practical constraints. This review provides the necessary technical context for the problem setting and establishes the foundation upon which the proposed optimization methods are developed.

Chapter 3 establishes the foundational system and antenna models used throughout this dissertation. By adopting standardized configurations and parameter settings, this chapter ensures that all subsequent evaluations are conducted under realistic and reproducible conditions, providing a consistent basis for the development and assessment of the proposed optimization frameworks.

To address the aforementioned challenges, Chapter 4 proposes a clustering-aided optimization algorithm that combines particle swarm optimization (PSO) with a novel regulated K-means clustering method. In this chapter, the regulated K-means algorithm is introduced to group UEs into balanced clusters, providing meaningful spatial structure

for initializing the optimization process. This enables the algorithm to precisely redefine the parameter search space for each cell, thereby reducing computational complexity and improving the ability to achieve the global optimum in antenna parameter optimization. By integrating regulated K-means with PSO, the proposed method dynamically determines the set of antenna pattern parameters for multi-cell HAPS systems based on the current UE distribution. Extensive simulations based on real-world UE distribution datasets demonstrate that the proposed approach outperforms existing methods in reducing the proportion of low throughput UEs, achieving a more balanced throughput distribution, and maintaining low computational complexity, thereby supporting the practical requirements of HAPS operations.

Recognizing the limitations of iterative optimization for real-time applications, Chapter 5 proposes a neural network-based beamforming optimization approach. While the clustering-aided PSO algorithm achieves state-of-the-art performance in determining antenna parameters, its reliance on iterative search introduces variability and constrains real-time applicability. To overcome these limitations, a deep learning model based on the ConvNeXt V2 architecture is developed to directly predict antenna parameters from UE distributions represented as spatial density maps. The network employs a normalization strategy to enable generalization across different coverage radii and incorporates a permutation-invariant loss function, using the Jonker–Volgenant (JV) assignment algorithm, to resolve the order ambiguity inherent in multi-cell configurations. The neural network model is trained using ground truth data generated by the PSO-based method and is rigorously evaluated using extensive real-world datasets. The results indicate that the neural network-based method achieves throughput performance close to that of the best iterative optimization while significantly reducing inference time and ensuring stability, thus meeting the demands for robust real-time beam control in HAPS systems.

Chapter 6 concludes the thesis by summarizing the major findings, analyzing the implications of the proposed methods, and discussing their significance for intelligent HAPS-enabled networks. The chapter also provides recommendations for future research

in both algorithmic and architectural aspects, with the aim of realizing intelligent networks that fulfill the requirements of an increasingly connected IoT ecosystem.

Acknowledgments

Completing this doctoral thesis has been a long and transformative journey, one made possible only through the support, guidance, and encouragement of many remarkable individuals.

First and foremost, I wish to express my deepest gratitude to my supervisor, Professor Tomoaki Ohtsuki. His expertise, insightful feedback, and steady encouragement have shaped not only this research, but also my growth as an independent researcher. Professor Ohtsuki's high standards, patience, and commitment to academic excellence have been a source of inspiration throughout my time at Keio University. I am especially appreciative of the freedom he provided to pursue my own ideas, while always being available for advice and direction when most needed.

I am also grateful to the members of my thesis committee: Professor Hiroshi Shigeno, Professor Yukitoshi Sanada, and Professor Guan Gui. Their thoughtful questions, critical comments, and constructive suggestions have greatly enriched this work and broadened my perspective.

Sincere thanks are also due to all the colleagues and friends in the Ohtsuki Laboratory. I have benefited greatly from our technical discussions, collaborative spirit, and mutual support—both inside and outside the lab.

To my friends—near and far, old and new—thank you for bringing balance, laughter, and understanding during times of both challenge and celebration.

Most importantly, I would like to thank my family for their unwavering support, love, and encouragement. My parents' steadfast belief in me has been the foundation of my education and personal growth. Their sacrifices, trust, and understanding have been essential to the completion of this thesis and every step of this journey.

To everyone who has helped, encouraged, or inspired me along the way, I offer my heartfelt thanks.

Wei Dai

Chapter 1

Introduction

1.1 Background

The evolution of wireless communication technologies toward beyond 5th-generation (B5G) and 6th-generation (6G) networks has positioned the Internet of Things (IoT) as a critical focus, driven by the increasing complexity of IoT applications and the need for intelligent connectivity solutions, such as edge intelligence [1–6]. To meet the increasing networking demands of IoT, high altitude platform station (HAPS) systems are emerging as a key candidate, offering wide-area coverage, enhanced network resilience, and improved service delivery, particularly in remote and inadequately connected regions [7–9].

The HAPS coverage area spans from a few dozen kilometers to a radius of up to 100 km, making it instrumental in extending communication services to regions that previously lacked network access, such as remote islands, mountainous areas, or any regions with no cellular network infrastructure. Traditional terrestrial based stations often encounter challenges in these regions due to difficult terrain, high costs, and limited accessibility. Furthermore, HAPS has also been considered for integrating simultaneous wireless information and power transfer (SWIPT), providing a sustainable solution to prolong the lifetime of energy-constrained devices in unpopulated regions [10].

Commercial HAPS mobile communication services propose the deployment of unmanned aerial vehicles (UAVs), including balloons, airships, and specially designed aircraft, operating in the stratosphere at altitudes of around 20 km and providing coverage with a radius ranging from tens to hundreds of kilometers [8]. For example, several companies have been actively developing the physical systems and establishing network connectivity for HAPS systems using solar-powered UAVs [11–13]. Fig. 1.1 illustrates a typical HAPS system functioning as a repeater. In this setup, the HAPS connects to the Internet through a feeder link via a gateway station and delivers services directly to user equipments (UEs) within its coverage area through service links.

Compared to satellite communications, another form of Non-Terrestrial Network (NTN), HAPS also offers cost efficiency and ease of operation, as well as low latency. HAPS system operates at much lower altitudes than communication satellites, leading to significantly different round trip times (RTT) for data transmission. For example, geostationary

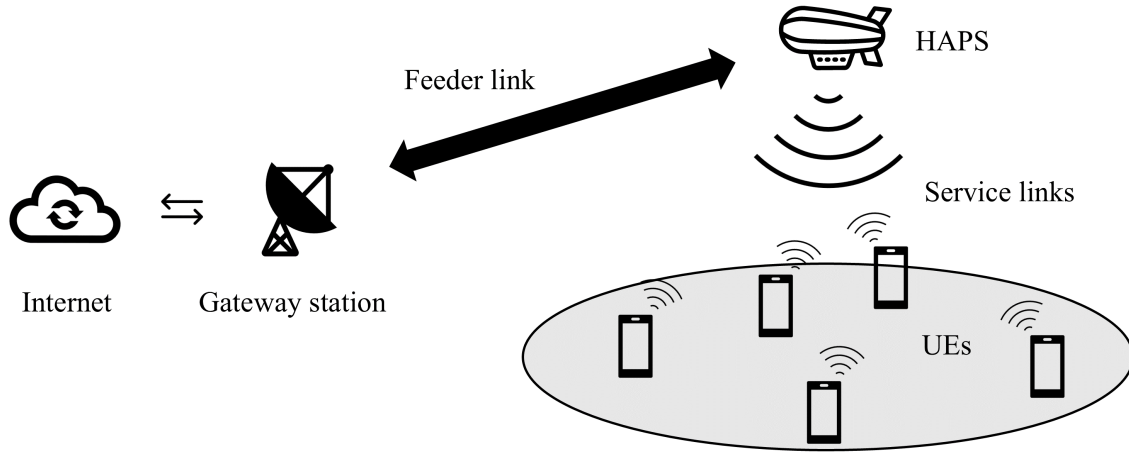


Figure 1.1. Illustration of a typical HAPS system functioning as a repeater.

earth orbit (GEO) satellites have an RTT of 240–280 ms, medium earth orbit (MEO) satellites have an RTT of 54–86 ms, and low earth orbit (LEO) satellites have an RTT of 6–30 ms [14]. In contrast, the RTT for HAPS is approximately $360 \mu\text{s}$ at a distance of 50 km and $680 \mu\text{s}$ at 100 km. This advantage makes HAPS an promising system for supporting low latency communications, such as Ultra-Reliable and Low Latency Communication (URLLC) where the 3GPP standard targets a packet failure rate of 10^{-5} within 1 ms over-the-air latency [15]. These advantages position HAPS as a key component in B5G and 6G mobile communication systems.

With the growing focus on space–air–ground integrated networks (SAGIN), HAPS systems have gained increasing attention for their ability to enhance IoT connectivity through seamless cooperation with satellites, UAVs, and terrestrial networks [16, 17]. HAPS systems enable massive access, low-latency, cost-effective, and energy-efficient solutions, supporting real-time, high-reliability IoT applications across wide areas [18, 19]. Additionally, HAPS systems can rapidly and seamlessly restore network services following natural disasters, allowing UEs to connect without the need for additional equipment [20, 21].

These advantages underscore the role of HAPS systems, positioned in proximity to UEs compared to satellites, as a crucial component in advancing research on novel SAGIN architectures designed to support IoT applications. For instance, HAPS operating between UEs and satellites can provide stable and seamless communication services for

IoT applications in remote areas by storing and forwarding data during intermittent connections, or by mitigating interference through null directivity control [22, 23].

Moreover, HAPS functioning as a multi-access edge computing (MEC) server within SAGIN can provide scalable, low-latency solutions for managing complex IoT data, offering additional computational capacity for resource-intensive IoT applications, or predicting massive IoT traffic demand to optimize dynamic communication resource control [24, 25]. Recent research has further integrated reconfigurable intelligent surfaces (RIS) and semantic communication into MEC-assisted integrated aerial-ground networks (MEC-IAGN), proposing a novel framework to enhance computational efficiency and network resilience against jamming attacks and imperfect channel state information (CSI) [26].

In parallel with the evolution of HAPS-based IoT solutions, the broader development of integrated non-terrestrial network architectures has highlighted the role of LEO satellite constellations, HAPS constellations, and airborne nodes as dynamic and reconfigurable components within SAGIN. Recent research emphasizes the importance of heterogeneous constellations combining CubeSats and HAPS, where multi-layer architectures can be optimized for region-specific coverage, low latency, and inter-layer task coordination [27–29]. To support this, design frameworks propose the joint use of satellite onboard processing, ground gateways, and MEC-enabled airborne platforms for real-time responsiveness and data compression [30–32]. Meanwhile, constellation design optimization for IoST and edge-aware LEO–HAPS–terrestrial coordination enables scalable and latency-aware connectivity across diverse geographic and mobility conditions [28, 33]. In these multi-tier networks, architectural models such as the satellite–terrestrial integrated edge computing network (STECN) and task-oriented intelligent networking approaches demonstrate the feasibility of coordinated MEC deployment across LEO, HAPS, and terrestrial nodes [29, 34].

Efforts to strengthen real-time responsiveness, interference mitigation, and energy efficiency have driven architectural innovations involving intelligent mobility management,

multi-node scheduling, and forecast-based handover in dynamic and mobile NTN environments [29,35–37]. For instance, flexible mobility control and relay-aided architectures have been proposed to reduce the burden of frequent LEO handovers and to mitigate uplink jamming or downlink eavesdropping through satellite beam hopping and selective relay switching [38,39]. HAPS systems are also well-suited to serve as computation-aware relays between user devices and satellites, using dynamic edge coordination to offload IoT or sensory data depending on priority, delay sensitivity, and power availability [32,40,41]. These advances align with broader studies in urban blockage modeling and interference-aware densification, which show the critical need for elevated platforms like HAPS in ensuring robust LoS links and managing spatially dense communication footprints [36,42]. Furthermore, emerging service frameworks leverage clustering, trajectory prediction, and multi-access strategies to optimize real-time computation delivery for high-mobility use cases such as vehicular networks or aerial traffic management [34,35,43].

The convergence of terrestrial, airborne, and satellite MEC platforms facilitates distributed task execution and cross-domain orchestration, enabling end-to-end services even under infrastructure constraints. HAPS platforms, with their quasi-stationary flight paths and flexible altitude control, offer stable mid-layer computation points between terrestrial base stations and rapidly moving LEO satellites [31,33,44]. Hybrid networking strategies employing free-space optics (FSO), mmWave, and multi-beam antennas further enhance HAPS’s role in vertical backhaul/fronthaul frameworks, especially in ultra-dense or disaster-recovery scenarios [31,41,45]. To ensure service availability, mobility-aware service migration and containerized orchestration schemes have been introduced for both satellite and airborne nodes, enabling low-latency edge support for delay-sensitive applications such as augmented reality (AR)/virtual reality (VR), or sensor fusion [29,32]. Additionally, developments in blockchain-enhanced security mechanisms and decentralized identity verification offer viable solutions to mitigate SIN vulnerabilities across satellite and HAPS communication layers [38]. Altogether, these advancements underscore the

importance of incorporating HAPS into the future design of globally distributed, energy-aware, and resilient NTN systems supporting massive-scale IoT and edge-intelligent applications [34, 43, 46].

1.2 Motivation and Methodological Overview

1.2.1 Dynamic Beamforming

To support a wide range of applications, it is crucial to ensure sufficient throughput between HAPS and UEs. Multi-user multiple-input multiple-output (MU-MIMO) and massive MIMO with beamforming techniques have been extensively developed to enhance communication services and are now explored for integration into HAPS systems [47–57].

However, implementing massive MIMO in HAPS communications presents two primary challenges. The first challenge is the practical implementation of a specially designed phased array antenna with a sufficient number of antenna elements to support a massive number of UEs over a very wide coverage area. This requirement is technically demanding and difficult to achieve in practice. The second challenge is the computational complexity associated with handling CSI feedback, which becomes excessively high due to the vast coverage area and the large number of UEs. Additionally, the pilot overhead required for CSI feedback can be substantial, further complicating the implementation.

In addition to massive MIMO, the multi-cell configuration with single-cell frequency reuse is also considered as a solution [58–65]. However, challenges arise from the irregular and constantly changing distribution of UEs within the coverage area. The continuous movement of HAPS, driven by varying coverage requirements and influenced by wind and turbulence, introduces instability that can affect UE throughput and lead to frequent handovers [66–68]. Additionally, the non-uniform and dynamic nature of UE distribution exacerbates these issues, leading to further degradation in network performance.

In this dissertation, to address these challenges and prevent network outages for UEs, we focus on methods that enable HAPS to continuously optimize cell configuration and quickly adapt to changing conditions. Given the altitude at which HAPS operates and

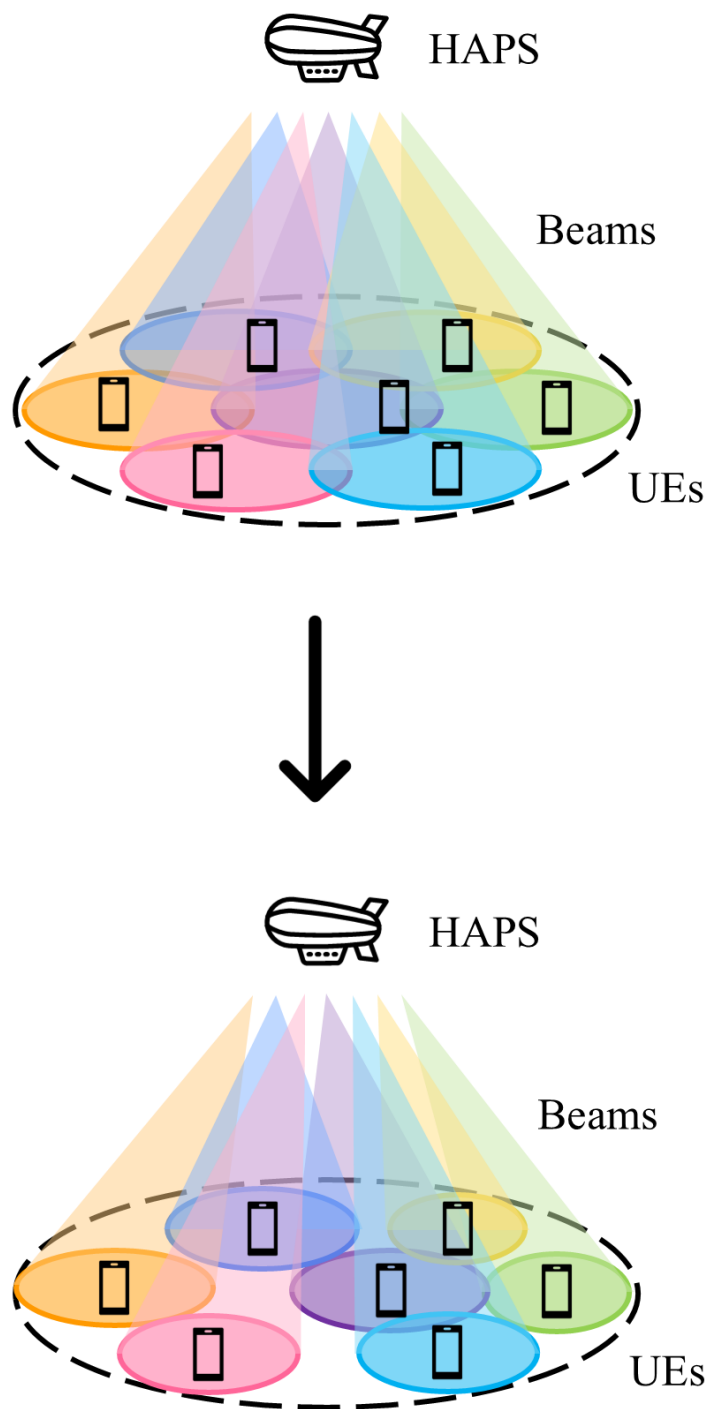


Figure 1.2. Illustration of dynamic beam control based on UE distribution. HAPS continuously optimizes beams rather than relying on a fixed, uniform cell configuration.

the scenarios in which it is deployed, HAPS is expected to provide strong line-of-sight (LoS) communication services to UEs. The LoS propagation allows for the assumption of free space path loss in data transmission. Consequently, CSI feedback can be simplified to angular information of UEs relative to the HAPS, which is essentially location-based data [69]. In this case, optimizing cell configuration based on UE distribution will be equivalent to optimizing a MU-MIMO system.

Moreover, the optimal solution should be applicable to both downlink (DL) and uplink (UL) communications. By regularly obtaining the optimal multi-cell configuration at small time intervals, we can dynamically control the antenna to maintain consistently strong throughput performance for UEs and reduce the occurrence of handovers, as illustrated in Fig. 1.2.

1.2.2 Clustering-Aided Beamforming Optimization

The challenges identified in the dynamic beamforming of HAPS systems reveal that conventional approaches are often inadequate for coping with the highly dynamic and non-uniform distribution of UEs over large coverage areas. Traditional methods, such as static cell configuration or uniform cell splitting, lack adaptability to frequent changes in UE spatial distribution and platform mobility [70–72]. As a result, these techniques frequently result in regions with persistently low throughput and fail to provide balanced or fair network performance.

To address these challenges, a variety of optimization approaches have been explored, including metaheuristics such as genetic algorithm (GA), as well as machine learning-based methods such as reinforcement learning [73–79]. Among these, iterative optimization algorithms like GA have shown potential for finding improved antenna configurations by dynamically optimizing parameters based on current UE distributions [73, 74]. However, such metaheuristic methods generally operate over a large, high-dimensional search space without explicit use of spatial information, which increases the risk of local optima and results in high computational cost. Furthermore, their performance can be unstable

and dependent on algorithm initialization, which makes them less suitable for real-time or large-scale deployment in practical HAPS systems.

Reinforcement learning approaches have also been investigated to enable HAPS to autonomously adapt beamforming strategies [75–79]. While reinforcement learning methods offer the possibility of online adaptation and learning from interaction with the environment, they often require extensive exploration and training time, and their real-time convergence and reliability in the context of highly dynamic multi-cell HAPS scenarios remain significant challenges.

To overcome the above limitations, this dissertation introduces a clustering-aided particle swarm optimization (PSO) algorithm for beamforming optimization in multi-cell HAPS systems. The key novelty of this approach lies in leveraging spatial clustering of UEs to explicitly guide the optimization process. A regulated K-means clustering algorithm is first applied to group UEs into clusters that are both spatially contiguous and balanced in size. This pre-clustering step provides meaningful spatial information that narrows and redefines the search space for PSO, enabling more efficient and robust global optimization of antenna parameters for each cell.

By integrating regulated K-means clustering with PSO, the proposed method achieves several important advantages. It significantly reduces computational complexity by focusing the search on relevant parameter regions, improves network fairness by balancing the allocation of resources among UEs, and enhances overall throughput by enabling dynamic adaptation to diverse and rapidly changing UE distributions. Simulation results confirm that the clustering-aided PSO algorithm consistently outperforms conventional heuristic, metaheuristic, and reinforcement learning-based approaches in terms of minimizing low throughput UEs, maximizing fairness, and supporting practical, scalable deployment for real-world HAPS operations.

In summary, the clustering-aided PSO method proposed in this thesis offers a novel and effective solution for dynamic beamforming in HAPS systems. By incorporating spatial structure into the optimization process, it directly addresses the limitations of existing methods and advances the state of the art in HAPS-enabled wireless communication.

1.2.3 Neural Network-Based Beamforming Optimization

The limitations of conventional and state-of-the-art methods for dynamic beamforming in HAPS systems motivate the need for a fundamentally different approach. As discussed previously, a range of iterative search-based algorithms, including metaheuristic techniques such as GA, reinforcement learning approaches, and the proposed clustering-aided PSO method, have been widely studied to determine optimal antenna parameters for multi-cell HAPS systems [73–79].

While the clustering-aided PSO achieves superior throughput and fairness performance compared to earlier approaches, all of these methods share two critical limitations as iterative search-based algorithms. First, their iterative nature often constrains real-time applicability, as practical device and resource limitations may prevent timely computation of optimal parameters for rapidly changing user distributions. Second, their inherent randomness and sensitivity to initialization introduce instability, which means that repeated runs on the same input distribution can yield different results, with some runs achieving higher performance and others significantly worse.

In addition to iterative optimization, a geometry-based approach has been considered for HAPS beamforming configuration [80]. This approach uses fixed geometric formulas and clustering algorithms to assign antenna parameters based on the spatial layout of UEs. The geometry-based approach is capable of extremely fast computation and can easily meet real-time operational constraints. However, its main drawback is a lack of throughput optimization, as the antenna parameters are not adapted to maximize UE throughput but are instead determined by static rules. This results in poor overall performance and suboptimal coverage for many UEs, making the geometry-based approach unsuitable for scenarios where quality of service (QoS) is critical.

Another line of research attempts to leverage neural network models, specifically a multilayer perceptron (MLP)-based approach, to directly predict antenna parameters. Earlier work has combined MLPs with classifier chains and ensemble strategies to accelerate computation for real-time requirements [81, 82]. Although the MLP-based approach can

offer significantly reduced inference time compared to iterative search algorithms, it remains fundamentally limited in its capacity to capture the complexity of the beamforming optimization task. In particular, the MLP-based approach is restricted to predicting only the horizontal beam direction for each beam, while other essential antenna parameters must be preset to fixed values. As a result, this approach is unable to optimize the full set of beamforming parameters, and its achievable throughput lags far behind the performance of advanced iterative approaches.

To overcome these limitations, this dissertation proposes a novel neural network-based beamforming optimization approach utilizing the ConvNeXt V2 architecture. ConvNeXt V2 is a modern convolutional neural network model that excels at capturing complex spatial patterns from structured data, making it well-suited for learning the relationship between UE distributions and antenna parameters in HAPS systems. The core innovation lies in replacing the iterative search process entirely with a supervised deep learning framework that directly predicts all antenna parameters from the input UE spatial distribution. This method addresses both stability and real-time applicability challenges inherent to conventional approaches. Since the ConvNeXt V2 model is trained in a supervised manner, it always produces stable and consistent output for the same input distribution, which ensures reliability in dynamic operational scenarios.

Moreover, this approach provides predictable and scalable computation time, as the model architecture can be adjusted in size to balance the trade-off between computational resources and prediction accuracy. Comprehensive evaluations demonstrate that the proposed ConvNeXt V2-based method is able to maintain UE throughput performance close to that of the best iterative optimization algorithms, while offering stable, repeatable results and guaranteed inference speed suitable for real-time deployment. This advancement highlights the superiority of the proposed approach for practical HAPS systems, enabling robust, scalable, and efficient beamforming control for future intelligent non-terrestrial networks.

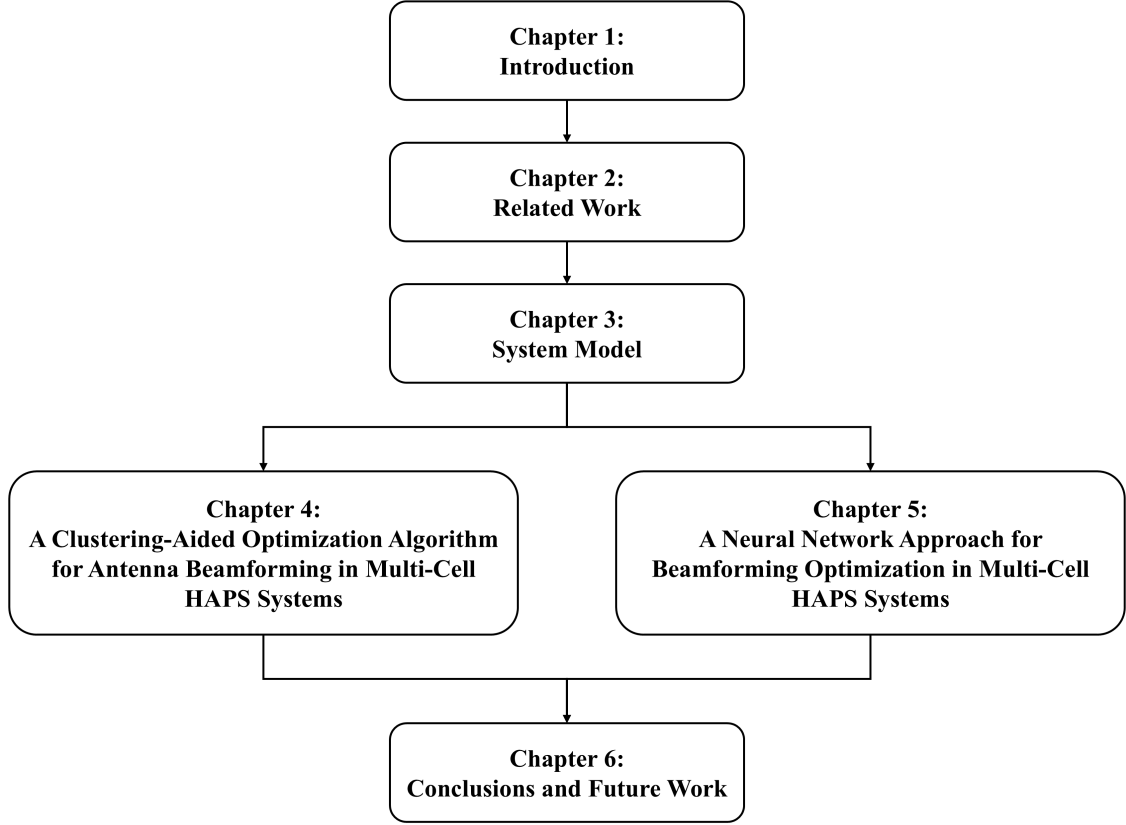


Figure 1.3. Illustration of the structure of the dissertation.

1.3 Outline and Contributions of the Dissertation

1.3.1 Outline of the Dissertation

This dissertation is organized as follows, with each chapter systematically addressing key objectives and challenges related to dynamic beamforming optimization in HAPS systems. The overall structure of this dissertation is illustrated in Fig. 1.3

Chapter 2 provides a comprehensive review of existing research related to performance enhancement in HAPS networks. The chapter systematically examines prior work on network architecture, resource management, optimization algorithms, and antenna beamforming techniques. The literature is organized to reflect a logical progression from general system-level advancements to more specialized methods for improving network efficiency and service quality. This chapter establishes the broader research context of the dissertation and provides the technical foundation for the development of the proposed optimization approaches in the following chapters.

Chapter 3 establishes the simulation framework used in this dissertation by detailing the HAPS system and antenna beamforming models. It specifies the network parameters, user equipment characteristics, and SINR-based throughput calculation used for performance evaluation. The chapter also presents the cylindrical phased array antenna model and defines the four antenna pattern parameters that serve as the optimization targets in subsequent chapters.

Chapter 4 presents the first main contribution of this research by developing a clustering-aided PSO algorithm for beamforming optimization in multi-cell HAPS environments. This chapter begins by highlighting the practical challenges that arise when traditional beamforming optimization approaches are applied to multi-cell HAPS systems. The limitations of multiple conventional methods are analyzed in detail to clarify the necessity of more advanced optimization frameworks. This chapter first defines the objective function for antenna beamforming optimization based on user throughput, providing a clear performance criterion for multi-cell HAPS configurations. It then presents the proposed clustering-aided PSO method, which integrates a novel regulated K-means algorithm to generate balanced and spatially contiguous user clusters. These clusters provide meaningful initialization for the PSO process, enabling more efficient search space reduction and guiding the optimization toward high-quality cell configurations. Comprehensive experimental results are presented, including performance evaluation under various realistic scenarios, quantitative comparison with existing state-of-the-art methods, and detailed analysis of throughput, fairness, and computational complexity, all using real-world user distribution datasets.

Chapter 5 introduces the second major contribution of this dissertation, which is a neural network-based beamforming optimization framework utilizing the ConvNeXt V2 architecture. The chapter begins with an analysis of the inherent challenges associated with iterative optimization, geometry-based approaches, and earlier MLP-based models, especially with respect to real-time applicability and stability in practical deployment. The design and construction of the ConvNeXt V2-based model are described in detail, including the representation of input user distribution, output parameter encoding, network

architecture, and the specific training strategies employed to ensure robust and generalizable performance. The methodology for benchmarking the proposed method is then provided, with experimental evaluation covering throughput performance, computational efficiency, result stability, and scalability under diverse operational scenarios. The chapter concludes with a discussion of the practical implications of the neural network-based approach, comparing its performance to the best iterative optimization results and demonstrating its suitability for real-time HAPS operation.

Chapter 6 provides an overall conclusion to the dissertation, summarizing the key findings and contributions of the proposed approaches. This chapter discusses the significance of the results for HAPS-enabled wireless networks and provides recommendations for future research based on the work presented in this dissertation.

1.3.2 Contributions

The main contributions of Chapter 4 and Chapter 5 in this dissertation are concisely summarized in Table 1.1 and Table 1.2, respectively. These tables systematically compare the conventional approaches, clarify their primary limitations, and highlight the advances introduced by the proposed methods in this work. By presenting the key objectives, technical challenges, and novel solutions addressed in each chapter, these summaries provide a clear and structured overview of the research advancements achieved in this dissertation. Furthermore, a visual comparison of different methods is shown in Fig. 1.4. This figure clearly states the strengths and weaknesses of both conventional and proposed methods in terms of two key dimensions: throughput performance and computation time.

TABLE 1.1
LIMITATIONS OF CONVENTIONAL MULTI-CELL HAPS BEAMFORMING OPTIMIZATION APPROACHES AND THE
CONTRIBUTIONS OF CHAPTER 4.

Objective	• Optimize HAPS beamforming for throughput and fairness based on UE distribution.
Conventional approaches	• Uniform cell splitting or fixed configuration divides coverage areas without considering UE spatial distribution [70–72]. • Metaheuristic methods such as GA search for optimal antenna parameters but do not utilize UE spatial structure [73, 74]. • Reinforcement learning methods adapt beamforming policies via throughput evaluation and rewards [75–79].
Conventional approaches limitations	• Uniform cell splitting cannot adapt to non-uniform or changing UE distributions, causing persistent low-throughput areas and poor balance. • Metaheuristic methods face high computational complexity and may get stuck in local optima due to large search spaces and lack of spatial awareness. • Reinforcement learning requires long training and may not converge reliably in diverse HAPS scenarios.
Contributions of Chapter 4	• Regulated K-means clustering forms spatially contiguous, balanced clusters of UEs for informed initialization. • The search space for PSO is narrowed using clustering, improving optimization efficiency and robustness. • The method achieves better throughput balance and reduces low-throughput UEs compared to conventional methods. • Computational complexity is kept low, enabling practical, real-time HAPS deployment.

TABLE 1.2
LIMITATIONS OF CONVENTIONAL MULTI-CELL HAPS BEAMFORMING OPTIMIZATION APPROACHES AND THE
CONTRIBUTIONS OF CHAPTER 5.

Objective	• Achieve stable and real-time HAPS beamforming with throughput comparable to advanced iterative methods.
Conventional approaches	• Iterative search approaches, including clustering-aided PSO, determine antenna parameters by repeated optimization [73–79]. • The geometry-based approach assigns antenna parameters using fixed formulas and clustering [80]. • The MLP-based approach predicts only partial antenna parameters with a neural network and preset values [81, 82].
Conventional approaches limitations	• Iterative search approaches do not guarantee real-time performance and may give unstable results for the same input. • The geometry-based approach does not optimize throughput and often results in poor performance. • The MLP-based approach cannot predict the full set of antenna parameters and fails to reach the best throughput.
Contributions of Chapter 5	• The ConvNeXt V2-based approach directly predicts all antenna parameters from UE distribution without iterative search. • The method guarantees stable and consistent results for each input. • Computation time is fixed and scalable to real-time requirements. • The method achieves throughput close to the best iterative optimization and supports practical deployment.

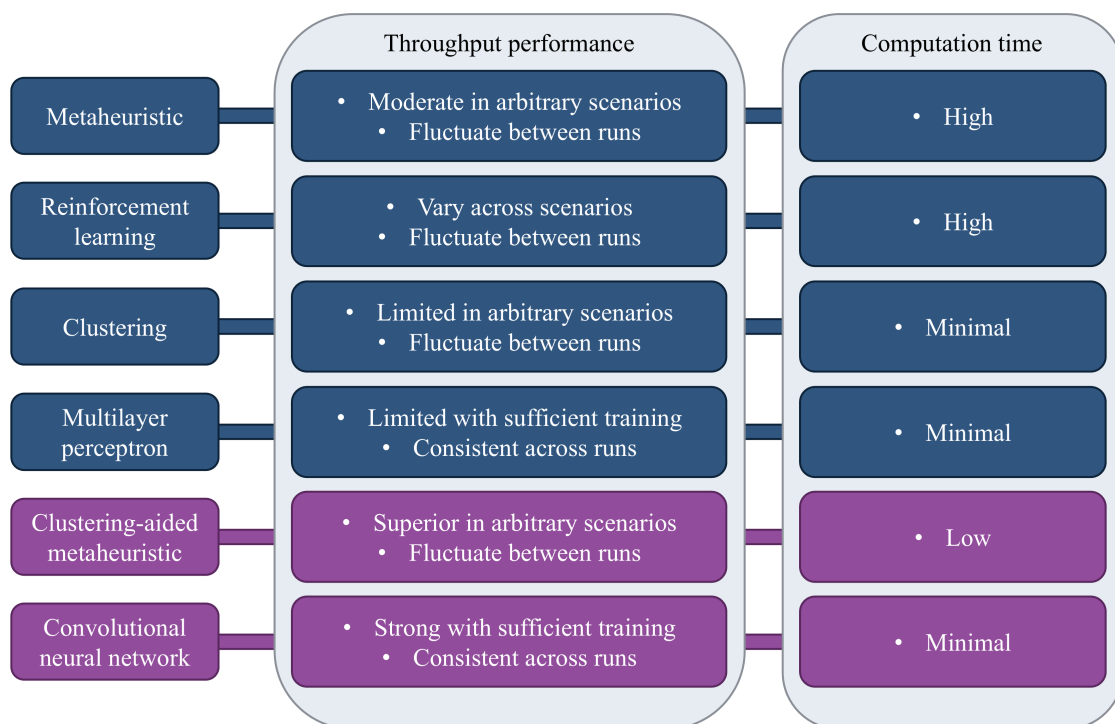


Figure 1.4. Performance comparison of methods.

Chapter 2

Related Work

2.1 Network Architecture and System Capabilities

HAPS have been increasingly recognized as a promising component in future wireless networks, offering flexible deployment, wide-area coverage, and rapid service recovery in both rural and urban settings [83–85]. Their integration into NTN enables enhanced support for diverse application scenarios, including broadband access, disaster recovery, and IoT connectivity [86, 87]. To unlock their full potential, considerable research has addressed fundamental challenges related to coverage extension, signal quality, mobility support, and link stability under various atmospheric and geometric constraints [88–91].

Early studies focused on the spatial deployment and antenna modeling of HAPS to ensure reliable link quality and minimal inter-beam interference across large coverage areas [90, 92]. Analytical frameworks were proposed to evaluate the influence of beamwidth, elevation angle, and inter-HAP distance on system capacity and signal-to-interference-plus-noise ratio (SINR), leading to optimized multi-cell planning strategies for static and quasi-static deployments [90, 91]. Meanwhile, user-side enhancements such as directional receive antennas and cooperative access techniques were also considered to increase overall throughput and reliability [92, 93].

The introduction of mobile and dynamic HAPS architectures further motivated the development of mobility-aware mechanisms, especially to support seamless handover in large-scale multi-cell environments [94–96]. Various handover optimization techniques have been studied, including predictive algorithms using motion estimation, cooperative handover schemes, and machine learning–based methods to ensure continuous connectivity [97–100].

As the demand for high-capacity HAPS networks grows, advanced interference mitigation strategies and multi-layered architectures have emerged to support scalability and system robustness. Research into heterogeneous HAPS constellations and hybrid space-air-ground integration has introduced layered topologies that enable efficient relay, aggregation, and backhaul mechanisms [88, 101, 102]. These configurations support both broadband and narrowband services and offer adaptive load balancing across service layers by leveraging technologies such as adaptive coding and multi-connectivity [85, 89].

Complementing these architectural advances, waveform design and spectral efficiency improvements have also been explored. Technologies such as non-orthogonal multiple access (NOMA), faster-than-Nyquist (FTN) signaling, and filter bank multicarrier (FBMC) have been proposed to enhance spectrum utilization in constrained HAPS environments [86, 103–108]. These waveforms exhibit resilience to Doppler shifts and channel dispersion and are suitable for high-altitude propagation conditions, providing a basis for supporting heterogeneous traffic patterns and high-throughput applications.

In parallel, significant work has been conducted on dynamic radio resource management (RRM) and call admission control (CAC) strategies tailored for HAPS [109–114]. These methods aim to efficiently allocate bandwidth and scheduling resources across multiple cells, accounting for user density fluctuations, QoS requirements, and inter-cell interference. Power-aware and fairness-aware admission schemes have demonstrated improved system utilization, especially in asymmetric user distributions and under constrained power budgets typical of airborne platforms [111, 113].

Beyond conventional RRM and waveform optimization, recent studies have investigated more comprehensive approaches to improving HAPS network performance through the integration of multiple technologies. Multi-hop architectures combining free-space optical and radio-frequency links have been proposed to enhance backhaul capacity and improve link reliability, particularly in environments affected by cloud coverage or radio interference [115–118]. These systems leverage the complementary characteristics of both technologies to support stable communication under varying atmospheric conditions.

In addition, HAPS-assisted mobile edge computing and IoT frameworks have been explored to reduce latency and improve task offloading efficiency, often through cooperative UAV relays or adaptive user association mechanisms [93, 119]. These architectures support location-aware processing in aerial networks and are suitable for deployment in remote or infrastructure-limited regions.

Recent works have also applied artificial intelligence to enhance HAPS performance in resource allocation, user association, and mobility adaptation. Reinforcement learning

methods have been integrated with spectrum access and dynamic reconfiguration strategies to enable adaptive decision-making under uncertainty [119, 120]. These approaches demonstrate the potential for self-organizing control in dynamic network environments and represent early steps toward intelligent, data-driven HAPS management.

Collectively, these developments highlight the growing sophistication of HAPS systems and their increasing integration with terrestrial and satellite networks. They provide a strong basis for continued research into control mechanisms, optimization techniques, and improvements in service reliability.

2.2 Learning-Based Control and Optimization

The increasing complexity and dynamic nature of HAPS deployments have prompted the use of intelligent algorithms to improve network performance through adaptive configuration and control. Traditional rule-based mechanisms are often insufficient to handle real-time variability in user distribution, mobility, and channel conditions. As a result, learning-based and bio-inspired optimization techniques have been explored to enable responsive and data-driven decision-making in HAPS networks [121, 122].

Machine learning approaches such as reinforcement learning have shown significant promise in HAPS resource management. These methods enable autonomous decision-making under partial observability and limited feedback, making them suitable for scenarios involving high-latency backhaul links and unpredictable atmospheric dynamics. Survey studies have detailed a wide range of reinforcement learning applications in wireless communications, including user association, spectrum access, and energy-efficient routing [121]. These approaches are particularly applicable to airborne and non-terrestrial platforms where centralized control may be infeasible.

In the context of HAPS, reinforcement learning has been applied to dynamic spectrum access and channel allocation to improve spectrum utilization without explicit coordination. For instance, two-stage learning frameworks have been introduced that combine temporal-difference updates with replay memory to stabilize policy learning and improve

convergence [123, 124]. These algorithms allow individual nodes to independently select transmission slots or channels based on observed interference and historical success, leading to more distributed and collision-resilient communication strategies.

Beyond channel access, reinforcement learning has also been explored for intelligent access network selection and RAT switching in environments where multiple connectivity options are available. In such cases, agents are trained to dynamically choose the most suitable link or access node based on performance metrics such as signal quality, congestion, and energy consumption [125]. These methods reduce the need for centralized signaling and offer adaptability in heterogeneous and time-varying network conditions.

Metaheuristic optimization techniques, including swarm intelligence and evolutionary algorithms, have been widely applied to support decentralized resource allocation and channel management in HAPS networks. PSO has been used to address access point selection and beamforming challenges in multi-user environments, demonstrating strong convergence properties and robustness to changing network topologies [126]. Similarly, reinforcement learning variants have been combined with swarm-based methods to optimize channel allocation under dynamic traffic loads and limited coordination, enabling scalable and fair spectrum distribution among aerial nodes [127].

In addition to online learning and metaheuristic methods, lightweight embedded learning models have been proposed to enable local intelligence within resource-constrained HAPS or UAV nodes. Such models leverage real-time sensor data to perform localized decision-making with minimal computational overhead, supporting delay-sensitive and autonomous operations in edge-based HAPS architectures [128]. This direction opens up possibilities for fully distributed network control without dependence on constant ground communication.

Task and resource allocation in HAPS-supported edge computing systems have also benefited from intelligent control frameworks. Federated learning has been proposed to manage computation and communication trade-offs while preserving data privacy across distributed airborne nodes [129]. These frameworks enable collaborative model training among HAPS, UAVs, and user terminals without the need for raw data exchange, which is

especially valuable in privacy-sensitive applications such as vehicular networks or remote healthcare services.

Multi-agent learning approaches have been introduced to optimize hierarchical control tasks such as user offloading, service migration, and energy balancing. For instance, Stackelberg game formulations combined with reinforcement learning allow leader–follower dynamics between aerial platforms and users to be modeled effectively under bandwidth and power constraints [130]. These techniques achieve near-optimal allocation outcomes in real time while accounting for user demand heterogeneity and system-level objectives.

Collectively, these intelligent control strategies demonstrate the growing importance of adaptive algorithms in managing the complexity of modern HAPS networks. By enabling decentralized optimization, situational awareness, and learning-driven adaptation, these techniques provide a foundation for scalable and autonomous aerial communication systems capable of supporting diverse and evolving service requirements [122].

2.3 Antenna Design and Beamforming Strategies

Antenna beamforming and cell configuration have become central topics in the design of HAPS-based wireless systems due to their direct impact on user throughput, signal quality, and overall system capacity. As HAPS platforms typically serve wide coverage areas with constrained onboard resources, effective management of directional antenna beams is essential for maintaining consistent link performance and minimizing interference. Early studies have proposed static and semi-dynamic beam configurations based on phased array antennas to extend service coverage while preserving carrier-to-interference-plus-noise ratio (CINR) at the cell edges [131, 132]. Analytical evaluations of multi-HAP deployments have shown that optimizing parameters such as beamwidth, pointing offset, and HAP spacing radius can significantly enhance coverage uniformity and reduce inter-cell interference [132, 133].

To further improve spectral efficiency and link reliability, several works have investigated advanced antenna array architectures, including concentric circular and distributed

subarray structures. These designs support narrow beamwidths and spatial diversity, enabling flexible control of signal patterns and directional gains [133, 134]. Diversity analysis in massive MIMO HAPS scenarios confirms that distributed antennas offer superior performance in fading environments, particularly when supporting multiple users or autonomous aerial coordination [134]. Complementary efforts have also explored the use of statistical channel properties to guide user grouping and outer beamformer design, resulting in reduced computational complexity and improved beam allocation strategies [135].

More recent studies have combined antenna optimization with system-level clustering or beam coordination to address performance bottlenecks caused by uneven user distribution. For example, clustering-assisted joint transmission schemes have been applied to mitigate cell-edge degradation and improve spectral utilization in densely populated coverage zones [131, 135]. These methods typically integrate phased array control with user-aware beam design to adaptively reconfigure beam directions and bandwidth allocations based on real-time user clustering.

In addition to system-level modeling, numerous studies have applied global optimization techniques to refine beam parameters and improve throughput in multi-cell HAPS configurations. PSO and GA have been employed to adjust beam direction and beamwidth in response to non-uniform user distributions, showing notable improvements in fairness and coverage quality across large service areas [72, 73]. To reduce the complexity inherent in large search spaces, enhanced optimization frameworks have introduced decomposition-based strategies and search space refinement, enabling more efficient parameter tuning without compromising system-level performance [74]. These approaches are especially relevant when balancing computational feasibility with the need for spatial adaptability in wide-area antenna deployments.

As the scale and dynamism of HAPS systems increase, learning-based methods have gained attention for their potential to support real-time beam adaptation under varying network conditions. Reinforcement learning and fuzzy logic frameworks have been applied to control key antenna parameters, such as tilt and beamwidth, enabling adaptive responses to platform motion and fluctuating user demand [75, 76]. These approaches

aim to reduce the proportion of underserved users while maintaining overall throughput, particularly in scenarios affected by HAPS drift or rotation. Recent developments further incorporate multi-agent structures and state abstraction techniques to improve scalability and learning stability in large multi-cell configurations [76, 77].

Hybrid frameworks that combine clustering with learning algorithms have also been introduced to improve decision granularity and reduce computational overhead. By leveraging user distribution patterns through clustering mechanisms, these methods enable more targeted beam control and enhance throughput for edge users without requiring exhaustive feedback or full channel knowledge [78, 80]. Such combinations of spatial preprocessing and learning-based adjustment reflect a broader trend toward modular, data-driven optimization schemes that can generalize across user environments.

Another line of research has focused on ensuring fairness and service continuity across all users, especially under complex spatial distributions and platform dynamics. Several methods have introduced objective functions that balance throughput maximization with equitable user experience, often by assigning greater importance to underserved or low-throughput users during optimization [73, 77, 78]. These strategies are particularly relevant in urban or post-disaster contexts where user density and priority can vary significantly across regions. Incorporating fairness into beam configuration enhances overall network utility and supports more resilient service delivery.

Meanwhile, neural network-based models have been proposed to approximate beam optimization outcomes with lower computational overhead. These methods are typically trained using solutions from iterative algorithms and are capable of providing near-instant predictions once deployed [81, 82]. Techniques such as classifier chaining and ensemble modeling have been adopted to address dependencies among antenna parameters across multiple cells, improving accuracy and constraint satisfaction. These learning-based predictors offer a scalable alternative to search-based approaches and are particularly useful for supporting real-time decision-making in large-scale systems.

Complementary work has examined the role of cell geometry and hierarchical architectures in supporting large-scale HAPS deployments. Directional beam-pointing schemes

and tessellated cell designs have been proposed to enhance coverage uniformity and reduce interference, particularly in wide-area or irregular terrain environments [70, 71]. These approaches are often evaluated using practical deployment constraints such as platform spacing, cell overlap, and elevation-dependent gain variation. Multi-tier architectures and coordinated beam reuse schemes have also been shown to improve throughput and fairness in edge regions while minimizing inter-cell conflicts [136, 137].

In parallel, advanced modeling frameworks have incorporated traffic demand and interference dynamics into beam control strategies. Efficiency evaluations based on game-theoretic formulations and AI-assisted scheduling have been developed to account for mutual interference and delay constraints, leading to more responsive and application-aware HAPS systems [138]. Attention has also been given to integrating user selection and actor–critic reinforcement learning for highly mobile scenarios, enabling beamforming strategies that adapt to vehicular traffic and dynamic IoT environments [79].

Together, these works highlight a wide range of algorithmic and learning-based techniques that support the evolving requirements of HAPS networks. From improving edge-user throughput to enabling fast and reliable multi-beam coordination, the ongoing development of antenna and beamforming strategies continues to play a central role in enhancing the performance of HAPS as a flexible and scalable network platform. The key themes and representative studies discussed throughout this chapter are concisely summarized in Table 2.1.

TABLE 2.1
SUMMARY OF THE RELATED WORK.

Network architecture and system capabilities	• System deployment, coverage extension, and integration with terrestrial and satellite networks [83–93, 101, 102]. • Mobility support and seamless handover in dynamic HAPS environments [94–100]. • Resource and spectrum management including waveform design and RRM/CAC strategies [86, 103–114]. • Robust connectivity and edge computing through hybrid architectures and UAV/HAPS cooperation [93, 115–119]. • Initial applications of AI for adaptive resource control and decision-making [119, 120].
Learning-based control and optimization	• Reinforcement learning for adaptive control in dynamic and uncertain environments [121–125]. • Hybrid learning and optimization approaches including swarm intelligence and metaheuristics [126, 127]. • Collaborative and lightweight learning frameworks for distributed HAPS operations [128–130].
Antenna design and beamforming strategies	• Beam configuration and phased array design for wide-area and high-throughput coverage [131–135]. • Optimization-based and heuristic methods for beam parameter tuning [72–74]. • Learning-driven beam control with clustering and fairness enhancements [73, 75–80]. • Neural network models for fast beamforming prediction and real-time operation [81, 82]. • Architectural strategies for scalable coverage and interference management [70, 71, 136–138].

Chapter 3

System Model

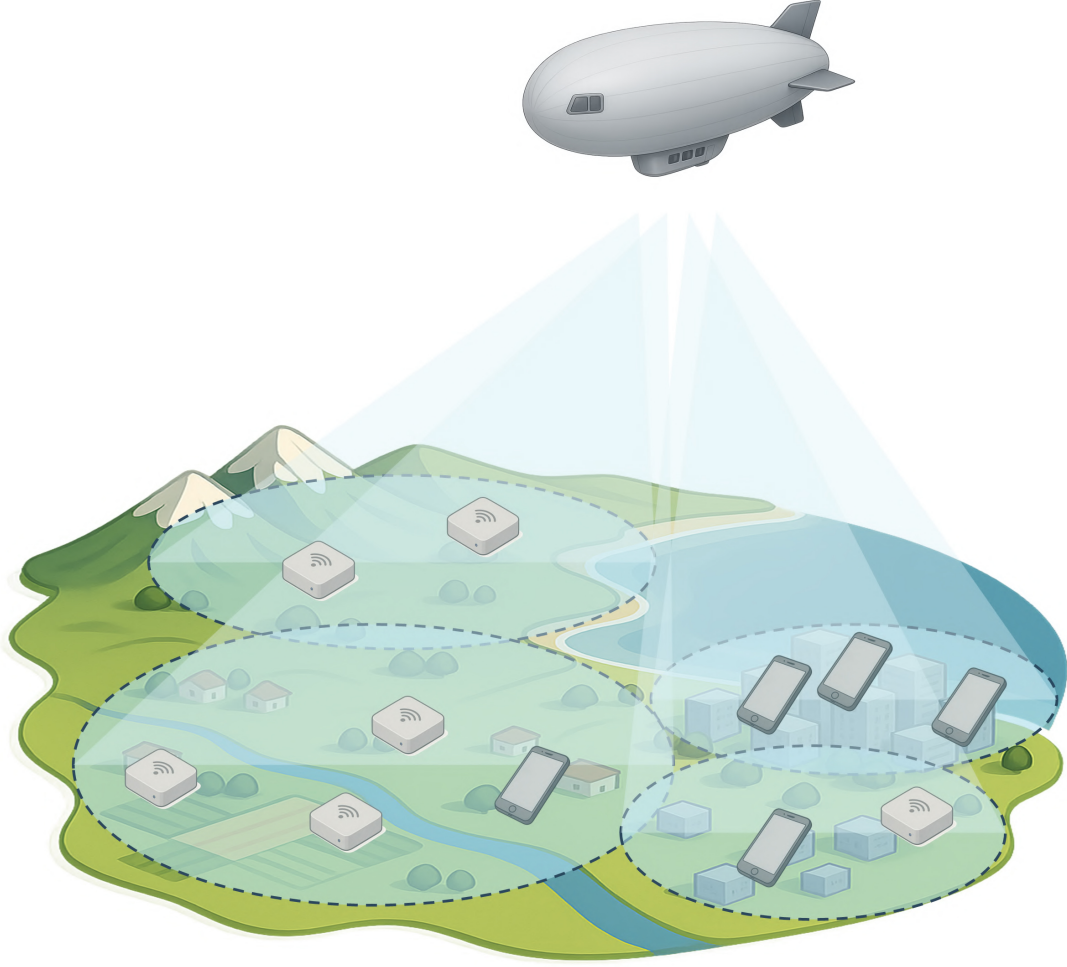


Figure 3.1. Illustration of the HAPS-enabled cellular networks.

3.1 HAPS Model

The HAPS cellular model developed in [59] serves as the foundation for our study. This model provides a comprehensive and well-established framework for conducting HAPS simulations, and it is designed to closely adhere to the International Telecommunication Union Radiocommunication Sector (ITU-R) recommendations [139]. By following this framework, we ensure that our simulations incorporate all essential parameters and considerations necessary for producing realistic and reliable results. An illustration of the HAPS-enabled cellular networks is shown in Fig. 3.1.

In this dissertation, we adopt this foundational model and specify a set of key system parameters to accurately characterize the HAPS environment. Specifically, the HAPS system is defined with a transmission power of 43 dBm per cell, which provides sufficient

Table 3.1. HAPS model parameters.

Parameter	Value
Service link transmission frequency	2 GHz
Transmission power	43 dBm/cell
Bandwidth	18 MHz
Propagation	Free space path loss
UE antenna gain	-3 dBi
UE noise figure	5 dB

radiated energy for service coverage. The service link transmission frequency is set to 2 GHz, reflecting a practical operational band for HAPS communications, while a bandwidth of 18 MHz is allocated for data transmission. We assume free space path loss in the propagation model, taking advantage of the LoS conditions that are characteristic of HAPS deployments at stratospheric altitudes.

To further ensure the realism of our simulation environment, device-level parameters for UEs are included in the model. The UE antenna gain is set at -3 dBi, and the UE noise figure is specified as 5 dB. These values are chosen to reflect commonly encountered device specifications and to provide a fair evaluation of signal quality under practical operating conditions [140, 141]. All of the system parameters used in our simulations are summarized in Table 3.1.

Within this framework, the DL SINR at the receiver, denoted as γ_{SINR} , is a critical metric for assessing the quality of the received signal. The SINR can be directly calculated based on the specified HAPS system parameters and the spatial configuration of the network. Once the DL SINR is determined, it is then possible to compute the UE throughput, which serves as a fundamental metric for evaluating the effectiveness of any proposed optimization algorithm. Throughput calculations based on SINR allow for a comprehensive performance assessment of the HAPS system, supporting detailed comparisons between

different optimization methods under consistent and realistic simulation conditions. Notably, based on this system model, we only focus on the optimization problem regarding spatial domain cell configuration. Further channel allocation is not considered in the scope of this work.

3.2 Antenna Beamforming Model

In this dissertation, we consider a beamforming model that utilizes a cylindrical phased array antenna, as originally proposed in [142]. The cylindrical phased array configuration is adopted for its suitability in supporting multi-cell coverage within HAPS systems. In order to implement effective beamforming, a weighting matrix for the antenna elements is required. This model specifically simplifies the beamforming process for the context of simulating multi-cell HAPS configurations, making it tractable for optimization and system-level evaluation.

The beamforming model is characterized by four principal antenna pattern parameters for each beam, which serve as the essential degrees of freedom in the design. These parameters are carefully selected to enable precise control over the shape and direction of each beam. The collective values of the four parameters for each beam are used to define the antenna weighting matrix, which is responsible for generating the desired radiation pattern. By reducing the optimization task to these four key parameters, the model significantly narrows the search space for the optimization algorithm, which in turn lowers computational complexity and expedites convergence toward optimal or near-optimal solutions.

The validity of this approach has been supported through experimental evaluations, which have demonstrated that the beam patterns generated by this model closely align with the measured results of actual cylindrical phased array antennas [143]. Although the model does not fully encapsulate every detail and complexity of real-world beamforming, it is considered a reasonable and practical approximation of the relevant antenna characteristics required for system-level simulation and optimization.

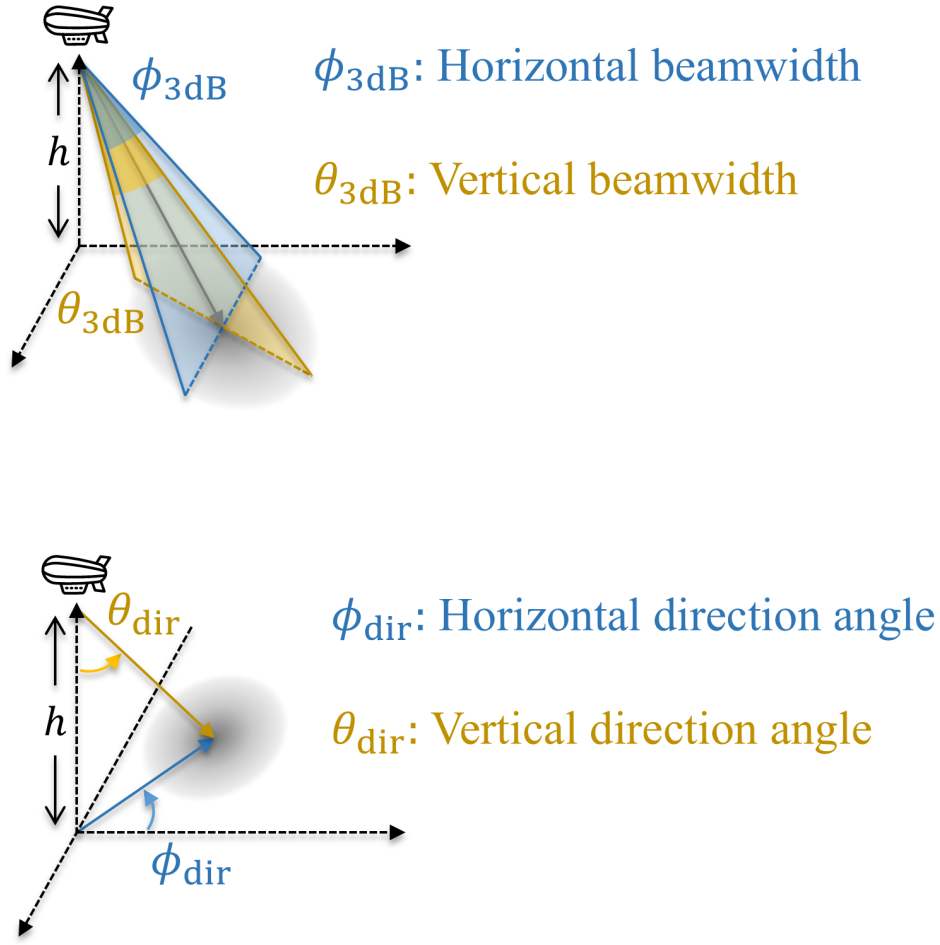


Figure 3.2. Illustration of the four antenna pattern parameters, where h is the altitude of HAPS.

The four key parameters in this model are as follows: the vertical direction angle θ_{dir} , the horizontal direction angle ϕ_{dir} , the vertical half-power beamwidth θ_{3dB} , and the horizontal half-power beamwidth ϕ_{3dB} . These parameters collectively determine the orientation, shape, and spatial selectivity of each beam, thereby facilitating targeted coverage and enhanced performance. The configuration and relationship of these parameters are illustrated in Fig. 3.2.

In practical HAPS deployments, the achievable ranges of these four parameters are constrained by the physical and design limitations that are inherently associated with the chosen cylindrical phased array antenna configuration. However, for the purpose of carrying out a thorough, in-depth algorithmic analysis, and also to avoid the risk of restricting

Table 3.2. Antenna pattern parameters.

Parameter	Range
Vertical direction angle	$0^\circ < \theta_{\text{dir}} < 90^\circ$
Horizontal direction angle	$0^\circ < \phi_{\text{dir}} < 360^\circ$
Vertical half-power beamwidth	$0^\circ < \theta_{3\text{dB}} < 180^\circ$
Horizontal half-power beamwidth	$0^\circ < \phi_{3\text{dB}} < 360^\circ$

or narrowing the scope of the optimization process, this dissertation adopts the maximum theoretical ranges defined in the model. These theoretical ranges provide a comprehensive basis for evaluating the flexibility, robustness, and overall performance of the proposed optimization algorithm in various scenarios. All relevant parameter ranges and associated values are summarized in Table 3.2. Similar to the parameter ranges, the maximum number of beams can be formed are also limited by the cylindrical phased array antenna. In this dissertation, we also assume that the number of cells to be configured is decided before the optimization process.

Modified from the ITU-R recommendations [139], the antenna gain as a function of the angular offset Ψ is defined in a piecewise form given by:

$$G(\Psi) = \begin{cases} -3 \left(\frac{\Psi}{\Psi_b} \right)^2, & 0 \leq \Psi \leq \Psi_1 \\ L_N, & \Psi_1 < \Psi \leq \Psi_2 \\ X - A \log_{10}(\Psi), & \Psi_2 < \Psi \leq \Psi_3 \\ L_F, & \Psi_3 < \Psi \end{cases} \quad (3.1)$$

where Ψ_b represents half of the 3dB beamwidth, $\Psi_1 = \Psi_b \sqrt{-L_N/3}$, $\Psi_2 = 3.745\Psi_b$, $\Psi_3 = 10^{(X-L_F)/A}$, and $X = L_N + A \log_{10}(\Psi_2)$. The near-in side-lobe level L_N and far-out side-lobe level L_F are specified as -20 dB and -30 dB, respectively, and the constant A are set to 20.

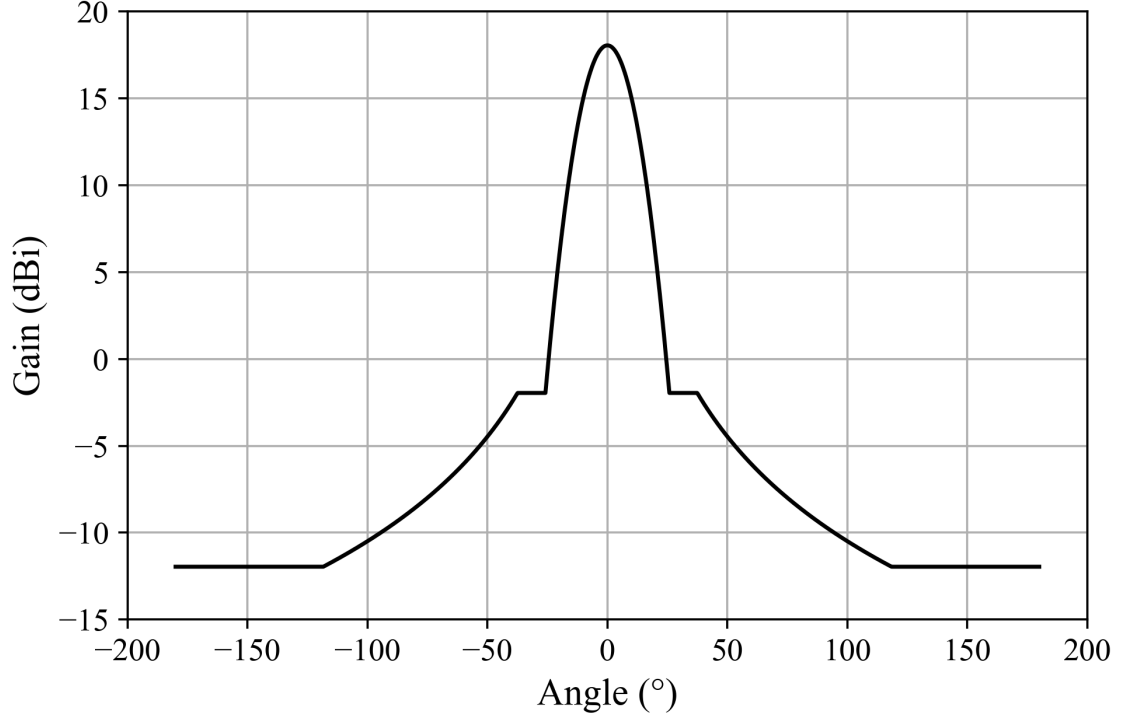


Figure 3.3. Antenna pattern of vertical or horizontal polarization for $\theta_{3\text{dB}} = 20^\circ$ and $\phi_{3\text{dB}} = 20^\circ$.

For each polarization component, the gain pattern is computed using Eq. (3.1). The combined antenna gain is then expressed as:

$$G = \max(G_v + G_h, L_F) + G_p, \quad (3.2)$$

where G_v and G_h are the vertical and horizontal gains, respectively, and G_p denotes the peak gain. Based on the beamwidths in the elevation and azimuth planes, the peak gain G_p is defined as:

$$G_p = 10 \log_{10} \left(\frac{80^2}{\theta_{3\text{dB}} \phi_{3\text{dB}}} \right) + 6. \quad (3.3)$$

This formula corresponds to an ideal antenna with a peak gain of 6 dBi and a beamwidth of 80° in both planes. An example for the antenna pattern of vertical or horizontal polarization for $\theta_{3\text{dB}} = 20^\circ$ and $\phi_{3\text{dB}} = 20^\circ$ is shown in Fig. 3.3.

Chapter 4

A Clustering-Aided Optimization

Algorithm for Antenna Beamforming in Multi-Cell HAPS Systems

4.1 Contributions

In this chapter, we present a comprehensive optimization framework to address the technical challenges identified in dynamic beamforming for multi-cell HAPS systems. Specifically, we introduce a solution that integrates a clustering-aided PSO algorithm with a regulated K-means clustering approach. The motivation for this framework is rooted in the need to systematically and efficiently improve throughput performance for user distributions that are often highly non-uniform and time-varying in realistic deployment scenarios.

Our methodology begins by formulating the cell configuration problem as an explicit optimization task. We define an objective function that mathematically characterizes the goal of improving throughput performance for HAPS, providing a clear and rigorous foundation for the optimization process. To solve this optimization problem, we employ PSO, leveraging its well-established advantages in terms of convergence speed, moderate computational complexity, and proven ability to search for global optimal solutions in high-dimensional spaces.

As an initial step, UEs are grouped into clusters according to the number of cells required in the HAPS system. These clusters serve as the critical foundation for deriving appropriate mathematical formulas that are used to redefine the search boundaries for each antenna pattern parameter. This redefinition of the search space directly reduces the dimensionality and complexity of the PSO search, while simultaneously improving the algorithm's capacity to find global optima. The carefully redefined search range is essential for reducing the risk of the optimization algorithm becoming trapped in local optima, all while maintaining sufficient flexibility for effective global optimization.

A key element of our framework is the selection and adaptation of the K-means clustering algorithm. K-means is considered highly suitable for our problem, given its efficiency in handling two-dimensional data and its low computational overhead. However, conventional K-means clustering, which partitions objects solely based on geometric proximity without considering the size of each cluster, can result in significant imbalance.

Such imbalances can be detrimental to the subsequent PSO optimization, as uneven clusters often yield suboptimal solutions and cause throughput degradation for a significant number of UEs.

To address this challenge, we introduce a regulated K-means algorithm that extends the classical approach by modifying the centroid updating rule to consider the evenness of cluster sizes. This regulated approach adaptively adjusts centroids so that clusters converge to more balanced sizes, thereby ensuring a fairer distribution of UEs across the cells. By incorporating this regulated mechanism, our method maintains the inherent low time complexity of standard K-means while substantially improving the initialization process for PSO, which in turn leads to a more effective and reliable optimization outcome. As a result, our framework not only enhances the average throughput experienced by UEs but also ensures a more balanced throughput distribution across the entire coverage area.

The effectiveness and practical value of our proposed framework are thoroughly evaluated through extensive simulations utilizing realistic UE distributions collected from various cities. The results from these simulations demonstrate that our regulated K-means-aided PSO algorithm consistently outperforms existing state-of-the-art methods, both in terms of achieving solutions closer to the global optimum and in reducing the occurrence of low throughput UEs. In addition, the approach delivers a significant improvement in the overall throughput performance of HAPS systems.

The main contributions of this chapter are summarized as follows:

- We propose a clustering-aided PSO algorithm to optimize the multi-cell configuration for HAPS, which effectively reduces the number of low throughput UEs, provides a more balanced throughput distribution, and thereby improves the throughput performance of HAPS systems.
- We derive search range redefinition formulas that convert UE distribution clustering information into the desired antenna parameter search range, reducing the computational complexity while improving the ability to find the globally optimal solution.

- We propose a novel regulated K-means algorithm that addresses the imbalanced clustering issue caused by conventional K-means, leading to more precise search range redefinition that enhances the ability to achieve global optimization.

4.2 Limitations of Existing Work

The optimization of multi-cell configuration based on UE distribution for HAPS systems is recognized as an NP-hard problem, presenting considerable challenges in finding an efficient solution. There are multiple previous studies that explore relevant approaches [70, 71, 73, 75, 76, 80, 144–148].

A beam-pointing algorithm proposed for HAPS ensures contiguous coverage over extended service areas, which leads to improved CINR for UEs by managing cell overlap [70]. Another study proposes a two-tier HAPS cellular architecture that is specifically designed to improve coverage and capacity by addressing inter-cell interference and uneven cell loading, particularly at the edge of coverage [71]. In addition, [72] presents a beamforming technique that utilizes concentric circular antenna arrays optimized by a comprehensive learning particle swarm optimizer (CLPSO) to reduce side-lobe levels (SLL) and enhance carrier-to-interference ratio (CIR) in HAPS systems. While these methods focus primarily on minimizing interference between adjacent cells, they fall short in addressing the challenges associated with HAPS movement and the dynamic distribution of UEs.

In terrestrial networks, the literature also demonstrates active research on dynamic beamforming optimization. A novel deep reinforcement learning algorithm has been developed for online antenna tuning, optimizing both the tilt angle and beamwidths to maximize the weighted sum-rate of UEs in heterogeneous networks (HetNets) [149]. Furthermore, a deep reinforcement learning framework has been designed to autonomously optimize MIMO broadcast beams based on user distribution, dynamically adjusting beam patterns to enhance coverage and network performance [150]. In the context of HAPS systems, reinforcement learning-based approaches have also been developed to optimize

beam control. For instance, fuzzy Q-learning is applied to dynamically control beam parameters in multi-cell configuration, which helps improve signal power and reduce the number of low throughput UEs by adjusting antenna settings based on received power thresholds [75]. Building on this approach, a deep Q-network (DQN)-based method has been proposed to further improve throughput by modeling the problem as a stochastic game equilibrium, thereby reducing the complexity of interactions among agents and achieving high convergence rates [76].

However, the proposed reinforcement learning-based methods rely on continuous learning processes to find the optimal solution for the current coverage area and UE distribution, which can result in high time complexity. This reliance on ongoing learning makes it difficult to meet the need for frequent updates at small time intervals to quickly respond to changes in the environment. Additionally, the performance of these methods is highly dependent on the design of effective reward functions, a task that is often challenging. While the proposed methods have successfully improved throughput with their designed reward functions, it is uncertain whether the solution they provide is optimal for various UE distributions or if there remains potential for further improvement.

In satellite-terrestrial integrated networks (STIN), a hybrid analog-digital beamforming approach has been proposed to maximize the secrecy-energy efficiency (SEE) under imperfect CSI by formulating and solving a constrained optimization problem [144]. In HAPS systems, a two-step GA-based approach is proposed to address the challenges of dynamic cell optimization, enabling optimization based on a designed objective function [73]. This objective function mathematically represents the throughput performance goal, accounting for UE distribution. However, the effectiveness of this method in finding the optimal solution is limited. Although it employs a co-evolutionary method to enhance the optimization process, this approach cannot consider all parameters simultaneously, leading to residual interference that cannot be fully mitigated during convergence. Furthermore, the co-evolutionary method, which optimizes different groups of parameters in successive iterations, increases the number of iterations and thus raises the time complexity.

Research beyond HAPS systems demonstrates that carefully eliminating redundant solution representations in GA can enhance both computational efficiency and solution quality [145]. In cellular networks, reducing the search range of antenna azimuth and tilt through solution space restructuring is shown to improve coverage optimization performance and accelerate convergence, where GA benefits from avoiding redundant parameter exploration [146].

In MIMO-based HAPS systems, K-means clustering is applied for user grouping based on measurable slow-varying channel parameters, which enhances beamforming while reducing CSI feedback overhead and computational complexity [147]. An extended approach for massive MIMO further improves beamforming by grouping users based on the chordal distance between statistical-eigenmodes, avoiding the complexity of singular value decomposition [148].

A K-means-based method is also proposed for multi-cell HAPS systems, which benefits from very low time complexity due to its reliance on straightforward geometric calculations [80]. It applies K-means clustering along with a fixed geometric formula to transform clustering information into cell configuration. However, this method does not incorporate throughput information in the optimization process and hence cannot address inter-cell interference issues, resulting in degraded throughput performance and relatively inferior cell configuration. Moreover, K-means clustering can lead to imbalanced cell sizes, where some cells cover significantly more UEs than others, which further prevents the even distribution of throughput among all UEs.

4.3 Objective Function

In proposing a meta-heuristic optimization-based approach for multi-cell HAPS configuration, it is essential to design an objective function that effectively addresses our underlying goal. The objective function serves as the mathematical foundation that directs the optimizer toward achieving an optimal cell configuration, ensuring that the antenna parameters selected in each iteration are evaluated against a rigorously defined criterion for performance.

To establish the objective function, we begin by defining the key variables involved in the optimization process. Let $\mathbf{a} = [\theta_{\text{dir}}, \phi_{\text{dir}}, \theta_{3\text{dB}}, \phi_{3\text{dB}}]$ represent a four-dimensional vector of antenna pattern parameters for a single beam, where each parameter governs an essential aspect of the beam's spatial orientation and coverage. The full candidate solution for the system, denoted as $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_K\}$, is composed of K such vectors, each corresponding to one cell in the HAPS system. The objective of the optimizer, therefore, is to determine the optimal set $\mathbf{A}$ that achieves the best possible network performance according to our defined criterion.

The optimization process relies fundamentally on the calculation of throughput for UEs located within the HAPS coverage area. It is important to note that, in the cell configuration problem, the only available information from UEs is their location. Using the position of the HAPS as the origin, we construct a Cartesian coordinate system, in which all UE locations are represented and stored within the matrix $\mathbf{U}$. This coordinate-based approach allows the system model to calculate the DL SINR for each UE, which then enables the derivation of the achievable throughput.

Given that HAPS systems typically serve wide coverage areas, the number of UEs in the region can be substantial. To make the optimization computationally tractable, it is practical to sample a sufficiently large number of UEs that accurately reflects the underlying distribution and density of the actual UE population. We denote the set of sampled UEs as $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_N\}$, where each $\mathbf{u}$ is a two-dimensional location vector in the Cartesian plane and N is the total number of sampled UEs. By adopting this sampling approach, we are able to maintain computational efficiency while preserving the fidelity of network performance evaluation.

Our central objective is to maximize UE throughput while also promoting a fair and balanced distribution of resources across all UEs. This dual objective aims to improve the overall service quality for every UE and, in particular, to reduce the risk of network outages for those with relatively low throughput. To realize this goal, we construct an objective function that seeks to maximize the weighted sum of UE throughputs, assigning

greater weight to UEs with lower throughput. This weighting mechanism is crucial for incentivizing the optimizer to prioritize improvements for the most disadvantaged users.

Formally, the objective function is defined as follows:

$$\begin{aligned}
& \min_{\mathbf{A}} \sum_{n=1}^N s w_n \frac{1}{\frac{B}{N_k} \log_2 (1 + \gamma_{\text{SINR},n})}, \\
& \text{s.t. } \theta_{\text{dir},\min} < \theta_{\text{dir}} < \theta_{\text{dir},\max} \\
& \quad \phi_{\text{dir},\min} < \phi_{\text{dir}} < \phi_{\text{dir},\max} \\
& \quad \theta_{\text{3dB},\min} < \theta_{\text{3dB}} < \theta_{\text{3dB},\max} \\
& \quad \phi_{\text{3dB},\min} < \phi_{\text{3dB}} < \phi_{\text{3dB},\max}
\end{aligned} \tag{4.1}$$

where s denotes a scaling factor for the objective, B is the system bandwidth, N_k represents the number of UEs in the k -th cell, and $(B/N_k) \log_2 (1 + \gamma_{\text{SINR},n})$ computes the throughput for the n -th UE. The term $w_n \in (0, 1]$ is a normalized weight for the n -th UE, specifically designed to emphasize lower throughput users in the optimization.

A key feature of the weighting strategy is the use of an exponential weighting function, formally expressed as:

$$w_n = \left(\frac{n}{N} \right)^p, \tag{4.2}$$

where $n = 1, 2, 3, \dots, N$ is the index of the sorted UE matrix, and $p \in [1, \infty)$ is the exponent factor. The set $\mathbf{U}^t = \{\mathbf{u}_1^t, \mathbf{u}_2^t, \mathbf{u}_3^t, \dots, \mathbf{u}_N^t\}$ denotes the UE matrix sorted in descending order of throughput at the t -th optimization iteration. Based on this ordering, we derive the weight vector $\mathbf{w} = [w_1, w_2, w_3, \dots, w_N]$ for all UEs.

The exponent factor p is an important tuning parameter that provides flexibility in the optimization objective. By adjusting p , we can control the balance between overall throughput maximization and fairness. A smaller value of p results in a more uniform weight distribution, treating all UEs similarly. In contrast, a larger p increases the emphasis on UEs with lower throughput by assigning them higher weights, thus directing the optimizer to allocate more resources to the most underserved users. The exponential weighting function is strictly monotonically increasing over the specified interval, ensuring that UEs with lower throughput are consistently prioritized.

This problem is formulated as a single-objective optimization, where the fitness score is defined as the value obtained from the objective function. Importantly, a lower fitness score indicates a more effective antenna parameter configuration for achieving our optimization goals. To facilitate straightforward comparison between different fitness scores, the function is constructed to minimize the sum of throughput reciprocals toward zero, as opposed to maximizing the sum of throughput, which would be unbounded. This approach provides a clearly defined numerical target for the optimizer and maintains comparability across different solution sets.

The scaling factor s is introduced solely for the purpose of analytical clarity, ensuring that the fitness score remains within a presentable range for interpretation and result analysis. Due to the properties of the exponential function, weights for low-throughput UEs can be nearly suppressed to zero for small index values, while increasing steadily to one for larger indices. Consequently, UEs assigned larger weights exert a greater influence on the objective function, which effectively drives the optimizer to prioritize throughput improvements for the lowest throughput users, sometimes at the expense of already well-served UEs.

Overall, this objective function formulation is critical for guiding the meta-heuristic optimizer to achieve balanced and high-performance multi-cell configurations in HAPS systems, supporting both aggregate throughput and fairness objectives in the evaluation of different antenna parameter solutions.

4.4 Clustering-Aided Optimization Approach

4.4.1 PSO with Clustering-Aided Parameter Range Redefinition

In our proposed method, we employ PSO as the optimizer to determine the antenna parameters based on our designed objective function. PSO is a widely used swarm intelligence-based meta-heuristic optimization algorithm, recognized for its efficiency in exploring the search space and achieving high-quality solutions. Using PSO, we begin by randomly initializing a population of particles according to the defined swarm size and

the allowed range of each antenna parameter. Each particle in the swarm represents a candidate solution described by $\mathbf{A}$, and its fitness score is evaluated through the objective function defined in (4.1).

At each iteration, the algorithm updates the position of each particle by assigning it a direction and velocity, allowing the swarm to collectively explore new regions of the solution space. The motion of each particle is governed by both its personal best fitness score and the best fitness score discovered by any particle in the entire swarm, referred to as the global best. Through this collaborative and iterative update process, PSO effectively guides all particles to converge towards the optimum solution.

However, as the number of parameters to be optimized increases fourfold with the number of cells, the dimensionality of the search space for PSO becomes extremely large in the case of multiple cells. The challenge is further compounded by the computational cost associated with calculating the fitness score from the antenna parameters, which involves intricate and time-consuming calculations. As a result, PSO becomes increasingly prone to getting stuck in local optima, which impedes its ability to discover the true globally optimal solution. Simply increasing the swarm size or the number of iterations does not fundamentally resolve this issue and leads to unacceptably high computational complexity, which is particularly undesirable for systems that require rapid response times.

To address this challenge, we introduce a pre-clustering step that utilizes the UE location matrix $\mathbf{U}$ to infer the ranges of antenna pattern parameters. The optimal cell configuration is inherently dependent on the underlying UE distribution. By incorporating this pre-clustering step, our method allows the optimizer to initialize the search space for each parameter with reduced and more meaningful bounds, thereby enhancing its ability to achieve global optimization. The complete process of our approach is illustrated in Fig. 4.1.

For the pre-clustering step to be truly effective in initializing appropriate search ranges, we propose a novel clustering algorithm based on K-means. In the context of 2-dimensional data such as $\mathbf{u}$, K-means is considered one of the fastest clustering algorithms while still providing strong clustering performance. We leverage the computational efficiency of

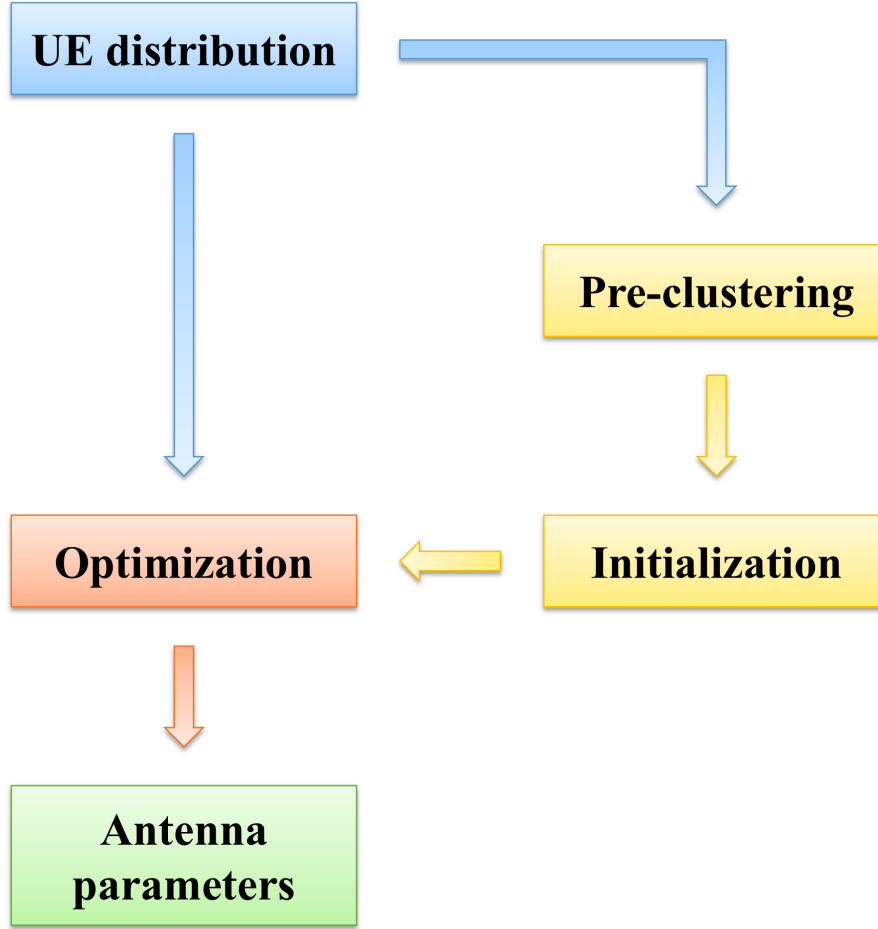


Figure 4.1. Flowchart of the complete optimization process.

K-means and adapt its updating rule to better fit the needs of our optimization problem, resulting in a new algorithm termed regulated K-means. Like the original K-means, the regulated K-means algorithm groups the 2-dimensional UE location data into K clusters based on K centroids in the Cartesian plane, with the grouping performed iteratively. The K resulting clusters can be regarded as K roughly separated cells, which are then used to define the search range initialization for the optimization. In this section, we focus on how our method leverages the clustering information to achieve effective search space reduction, while the specific details of the regulated K-means algorithm are provided in the following section.

The overall procedure of our proposed method is systematically described in **Algorithm 1**. The key innovation in this algorithm lies in the use of conversion formulas that redefine the parameter range for each antenna parameter to be optimized, making use of

the clustering-derived matrices $\mathbf{C}$ and $\mathbf{d}$. Here, $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \dots, \mathbf{c}_K\}$ records the centroid locations of clusters 1 to K , where each $\mathbf{c}_k = [c_{k,x}, c_{k,y}]$ is a 2-dimensional centroid location vector. The vector $\mathbf{d} = [d_1, d_2, d_3, \dots, d_K]$ contains the maximum Euclidean distance from each centroid to any UE in its respective cluster.

Utilizing $\mathbf{C}$, $\mathbf{d}$, and the HAPS altitude h , we derive the formulas that redefine the ranges of antenna pattern parameters $\theta_{3dB,k}$, $\phi_{3dB,k}$, $\theta_{dir,k}$, and $\phi_{dir,k}$ for each beam k . These conversion formulas, based on trigonometric relationships, are given as follows:

$$\begin{aligned} \theta_{3dB,k} &> 0^\circ, \\ \theta_{3dB,k} &< \tan^{-1} \left(\frac{\|\mathbf{c}_k\| + d_k}{h} \right) - \tan^{-1} \left(\frac{\|\mathbf{c}_k\| - d_k}{h} \right). \\ \phi_{3dB,k} &> 0^\circ, \\ \phi_{3dB,k} &< 2 \tan^{-1} \left(\frac{d_k}{\sqrt{\|\mathbf{c}_k\|^2 + h^2}} \right). \end{aligned} \tag{4.3}$$

$$\begin{aligned} \theta_{dir,k} &> \tan^{-1} \left(\frac{\|\mathbf{c}_k\|}{h} \right) - \frac{\theta_{3dB,k}}{4}, \\ \theta_{dir,k} &< \tan^{-1} \left(\frac{\|\mathbf{c}_k\|}{h} \right) + \frac{\theta_{3dB,k}}{4}. \\ \phi_{dir,k} &> \tan^{-1} \left(\frac{c_{k,y}}{c_{k,x}} \right) - \frac{\phi_{3dB,k}}{4}, \\ \phi_{dir,k} &< \tan^{-1} \left(\frac{c_{k,y}}{c_{k,x}} \right) + \frac{\phi_{3dB,k}}{4}. \end{aligned} \tag{4.4}$$

To clarify the underlying principles of the conversion formulas (4.3) and (4.4), we provide a visual illustration in Fig. 4.2. For the half-power beamwidth parameters $\theta_{3dB,k}$ and $\phi_{3dB,k}$, the redefined ranges specified by (4.3) for both the vertical and horizontal components ensure that each k -th beam can cover an area extending up to a circle of radius d_k . Fig. 4.2(a) highlights this search range, with the shaded area indicating the allowable beamwidth for an example beam. Through this approach, we achieve a significant reduction in the search range while still guaranteeing that all UEs within a cluster are covered by the beam.

Moreover, this strategy is robust to clusters with irregular shapes. The beamwidth range is chosen to be large enough to jointly cover the edge regions of adjacent cells,

Algorithm 1 Clustering-Aided PSO Algorithm for Multi-Cell Antenna Pattern Optimization

Input: $\mathbf{U}, K, P, T, c_1, c_2, w_{\min}, w_{\max}, V_{\max}, r_1, r_2$

Output: $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_K\}$

```

1: Group UEs  $\mathbf{U}$  into  $K$  clusters
2: return cluster centroids  $\mathbf{C}$  and maximum distance  $\mathbf{d}$ 
3: Redefine search range for each antenna parameter using  $\mathbf{C}, \mathbf{d}$  based on (4.3) and (4.4)
4: for  $i = 1$  to  $P$  do
5:   for  $j = 1$  to  $4K$  do
6:     Initialize swarm position  $x_{i,j}$  with random uniform value in  $[lb_j, ub_j]$  defined by
       search range
7:     Initialize velocity  $v_{i,j}$  with zero
8:   end for
9: end for
10: for  $t = 1$  to  $T$  do
11:   for  $i = 1$  to  $P$  do
12:     Clamp  $\mathbf{x}_i$  to ensure  $x_{i,j} \in [lb_j, ub_j]$ 
13:     Evaluate fitness  $f(\mathbf{x}_i)$  using objective function defined in (4.1)
14:     if  $f(\mathbf{x}_i) < f(\mathbf{pbest}_i)$  then
15:        $\mathbf{pbest}_i \leftarrow \mathbf{x}_i$ 
16:     end if
17:     if  $f(\mathbf{x}_i) < f(\mathbf{gbest})$  then
18:        $\mathbf{gbest} \leftarrow \mathbf{x}_i$ 
19:     end if
20:   end for
21:   Update inertia weight  $w \leftarrow w_{\max} - \frac{t(w_{\max} - w_{\min})}{T}$ 
22:   for  $i = 1$  to  $P$  do
23:     for  $j = 1$  to  $4K$  do
24:       Generate random factors  $r_1, r_2 \sim U(0, 1)$ 
25:        $v_{i,j} \leftarrow wv_{i,j} + c_1r_1(\mathbf{pbest}_{i,j} - x_{i,j})$ 
         $+ c_2r_2(\mathbf{gbest}_j - x_{i,j})$ 
26:       Clamp  $v_{i,j}$  to  $[-V_{\max}, V_{\max}]$ 
27:        $x_{i,j} \leftarrow x_{i,j} + v_{i,j}$ 
28:     end for
29:   end for
30: end for
31: return  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_K$  from  $\mathbf{gbest}$ 

```

thus enabling the algorithm to optimize cell boundaries more flexibly, reduce inter-cell interference, and minimize the number of UEs positioned at the edge of each cell. This definition ensures that the reduced search space encompasses the maximum meaningful range after pre-clustering.

It is equally important to properly adjust the beam direction angles $\theta_{\text{dir},k}$ and $\phi_{\text{dir},k}$. While the cluster centroid provides a good indication of the rough center of each beam, further refinement is necessary. Therefore, the optimizer is allowed a degree of freedom to search for the optimal beam direction within the bounds defined by (4.4), which are determined by the centroid locations. In Fig. 4.2(b), the shaded region represents the allowable search range for the beam direction of an example beam. By permitting the center of the k -th beam to shift up to $d_k/2$ away from the centroid both horizontally and vertically, the model accounts for the possibility that, in the worst case, the centroid is located at the edge of the cluster. In this way, the direction search range is maximized while still retaining relevance to the cluster structure.

To further illustrate the search space reduction, let the vector

$$\mathbf{x}_i = [x_{i,1}, x_{i,2}, x_{i,3}, \dots, x_{i,4K}] \in \mathbb{R}^{4K}$$

denote the swarm position of the i -th particle in the PSO algorithm, where $i = 1, 2, 3, \dots, P$. This vector maps to the flattened form of $\mathbf{A}$ with a dimension of $4K$, representing the exact data structure of a candidate set of antenna parameters for K beams during optimization. Each element $x_{i,j}$ in $\mathbf{x}_i$ corresponds to one of the beam parameters $\theta_{\text{dir},k}$, $\phi_{\text{dir},k}$, $\theta_{\text{3dB},k}$, or $\phi_{\text{3dB},k}$.

Without additional information, each $x_{i,j}$ originally lies within the theoretical range

$$x_{i,j} \in (x_{i,j,\text{min}}, x_{i,j,\text{max}}),$$

which corresponds to the full search space defined by the parameter ranges in Table 3.2. This full search space enables each beam to search across the entire coverage area geospatially.

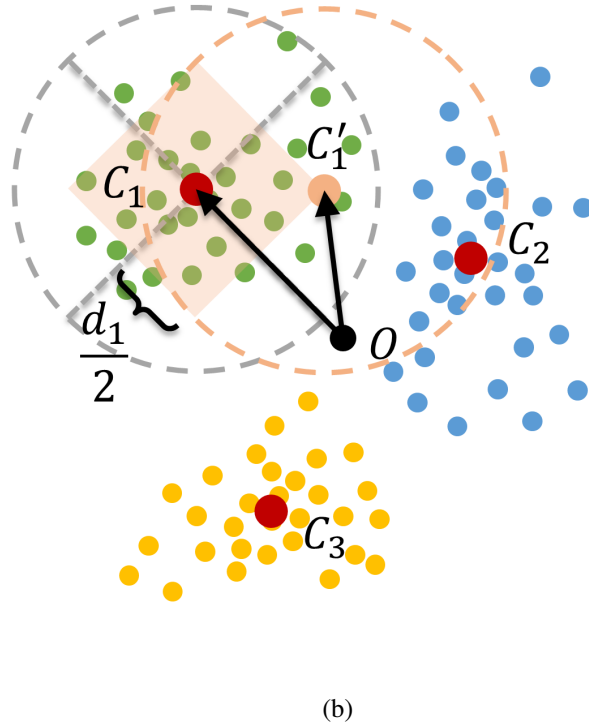
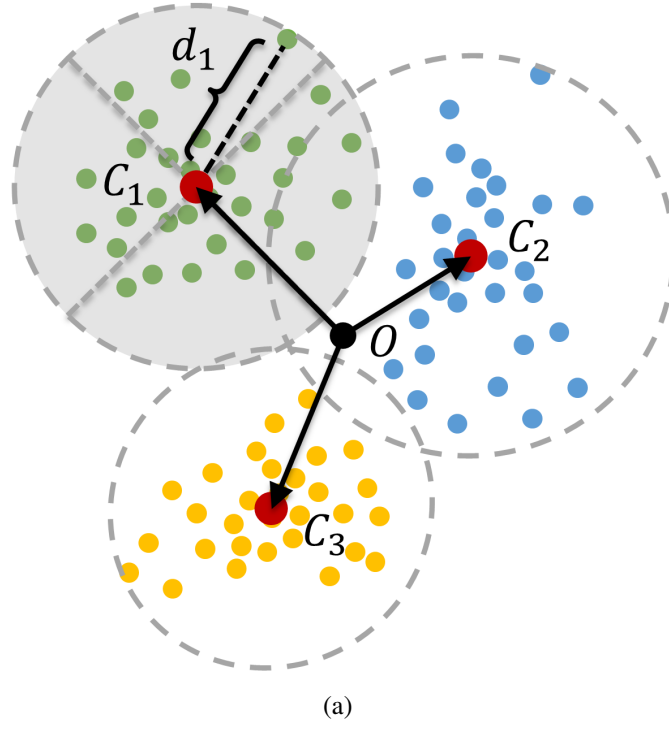


Figure 4.2. Illustration of the clustering-aided parameter range redefinition, where O is the location of HAPS, C_1 , C_2 , C_3 are the cluster centroids, d_1 is the maximum distance from centroid to UEs in cluster 1, and C'_1 is a candidate beam direction for beam 1. (a) Beamwidth search range (shaded area). (b) Direction angle search range (shaded area).

However, with our proposed method, the search range for each parameter is redefined as

$$x_{i,j} \in (x'_{i,j,\min}, x'_{i,j,\max})$$

using the clustering-aided algorithm. By leveraging geospatial information, our method divides the coverage area into K subareas, each corresponding to one beam. As illustrated in Fig. 4.2, each subarea is smaller than the full coverage area, restricting the beam direction $\theta_{\text{dir},k}$, $\phi_{\text{dir},k}$ and beamwidth $\theta_{3\text{dB},k}$, $\phi_{3\text{dB},k}$ to narrow ranges compared with their original bounds. For the k -th beam, the redefined search ranges eliminate the redundant search areas that belong to the subareas of the other beams.

As a result, the new search space for all K beams becomes a subset of the original search space, expressed as

$$\mathcal{S}' \subset \mathcal{S},$$

where

$$\begin{aligned}\mathcal{S} &= \prod_{j=1}^{4K} (x_{i,j,\min}, x_{i,j,\max}), \\ \mathcal{S}' &= \prod_{j=1}^{4K} (x'_{i,j,\min}, x'_{i,j,\max}).\end{aligned}$$

This new search space reduces the number of potential candidate solutions that need to be explored, transforming the search from a broad, unstructured domain to multiple localized and well-defined subdomains.

Overall, our proposed algorithm achieves an effective balance between search space reduction and solution flexibility, substantially improving the ability of PSO to find the globally optimal solution with lower computational complexity. The search space is narrowed in a principled manner, such that it remains large enough for effective optimization but is targeted enough to reduce the risk of getting stuck in local optima.

Finally, the PSO algorithm is configured with the following standard settings: swarm size P , iteration number T , cognitive parameter c_1 , social parameter c_2 , minimum inertia weight $w_{\min}$, maximum inertia weight $w_{\max}$, maximum velocity $V_{\max}$, and random factors r_1 and r_2 . These settings ensure both sufficient diversity and convergence properties in the

search, supporting the robust determination of antenna parameters for multi-cell HAPS optimization.

4.4.2 Regulated K-means Algorithm

To explain the proposed regulated K-means algorithm, it is essential to first identify the insufficiency of the conventional K-means algorithm in addressing our problem. In cell optimization, one of the most important objectives is to reduce inter-cell interference, which plays a significant role in maintaining network performance and service quality. Nevertheless, since UEs share the same frequency within each cell, distributing a similar number of UEs to each cell is also crucial for improving UE throughput and maximizing overall efficiency. Achieving such distribution enables more balanced resource allocation, which helps prevent network congestion in any particular cell and leads to enhanced fairness among users. When reducing the search range by pre-clustering, a balance between these two objectives, interference minimization and user balance, needs to be found to maintain the potential to reach the global optimum within the reduced search space. Failure to do so may result in suboptimal performance, making the design of the clustering step particularly important for the overall optimization framework.

By minimizing the within-cluster sum of squares (WCSS), K-means ensures that clusters are as compact and as distinct from one another as possible. The objective function of K-means clustering can be expressed as

$$\min_{\mathbf{c}_1, \dots, \mathbf{c}_K} \sum_{k=1}^K \sum_{n=1}^N \|\mathbf{u}_n^{(k)} - \mathbf{c}_k\|^2,$$

where $\mathbf{u}_n^{(k)}$ denotes the object, corresponding to the UE in our problem, that belongs to the k -th cluster. This mathematical formulation shows that the clustering process focuses on minimizing the spread of points within each cluster, which promotes the formation of spatially compact groupings.

The initialization performed by K-means pre-clustering increases the likelihood of interference reduction due to the proximity of users within each cluster. However, it does not

account for the effect of cluster size, which can result in certain clusters containing many more UEs than others. Such imbalance is problematic in wireless systems, as overloaded clusters can experience higher interference and resource contention, while underloaded clusters may not utilize their allocated resources efficiently.

As K-means is considered naturally suitable for balanced clustering with a given number of clusters, conventional algorithms have been developed based on K-means to explicitly address cluster size imbalance. These algorithms enforce cluster size balance by modifying the assignment step in K-means while keeping the centroid update step unchanged. The main goal is to achieve clusters that are not only compact but also uniform in size, which helps ensure more even distribution of throughput across all users. Addressing cluster size balance is essential for systems like HAPS, where user distribution can be highly variable and dynamic.

The constrained K-means algorithm formulates the clustering problem as a constrained optimization and solves it to complete the assignment step, where each cluster is required to satisfy a predefined size [151]. However, the method requires manually predefining the cluster size, which limits its adaptability across UE distributions with varying characteristics. This requirement for explicit cluster size specification introduces inflexibility, making the method less suitable for practical situations where user density and spatial patterns can change rapidly or unpredictably. Furthermore, its theoretical computational complexity is $O(K^{3.5}N^{3.5})$, which is significantly higher than that of K-means, failing to meet our requirement for low complexity. High computational complexity makes real-time optimization impractical and increases the computational burden for large-scale deployments.

Another algorithm, known as balanced K-means, replaces the assignment step in K-means with a linear assignment problem, enforcing strictly equal-size clusters [152]. Similar to the constrained K-means algorithm, the strict equal-size clustering also fails to provide well-suited initialization for the optimization step across varying UE distributions. More importantly, because this method solves the linear assignment problem without considering the spatial contiguity of clusters, each cluster can be split apart and

wrapped around others. This lack of spatial integrity leads to configurations where some clusters are fragmented or disconnected, which is not desirable for beamforming and cell coverage planning. This makes the algorithm inapplicable to our problem, as we need the pre-clustering step to robustly provide meaningful spatial information for initializing the optimizer. Moreover, the complexity of balanced K-means is $O(N^3)$, which is also much higher than that of K-means, further contributing to its impracticality for scenarios requiring fast computation or frequent updates. In summary, while both constrained and balanced K-means attempt to address cluster size, they introduce new issues related to flexibility, spatial structure, and computational cost.

In contrast, the minmax K-means algorithm provides an adaptive approach to balancing cluster size, while keeping the overall computational complexity the same as that of K-means. This method adds dynamically updated weights to the original K-means objective function, adjusting the assignment by penalizing clusters with larger intra-cluster variance [153]. The introduction of adaptive weighting enables some responsiveness to imbalance, but it comes at the cost of stability. Nevertheless, the method is highly sensitive to initialization and weight updating, which in some cases can result in low-quality clustering outcomes that are even worse than those produced by the conventional K-means algorithm. In particular, poor initialization or suboptimal adjustment schedules can cause the algorithm to converge to inferior local minima, limiting its reliability for robust HAPS cell configuration.

To address these problems and provide effective cluster information for initializing the optimizer, we propose a novel balanced clustering algorithm named regulated K-means. At each iteration, the algorithm allows extra movement of objects from oversized clusters to adjacent undersized clusters, resulting in a more balanced set of clusters than that produced by K-means, as illustrated in Fig. 4.3. This movement is performed according to spatial proximity, ensuring that UEs are only transferred to clusters in which they are reasonably located. The algorithm adaptively balances the trade-off between WCSS and cluster size while preserving the spatial contiguity of clusters, grouping UEs into appropriate clusters with computational complexity similar to the conventional K-means

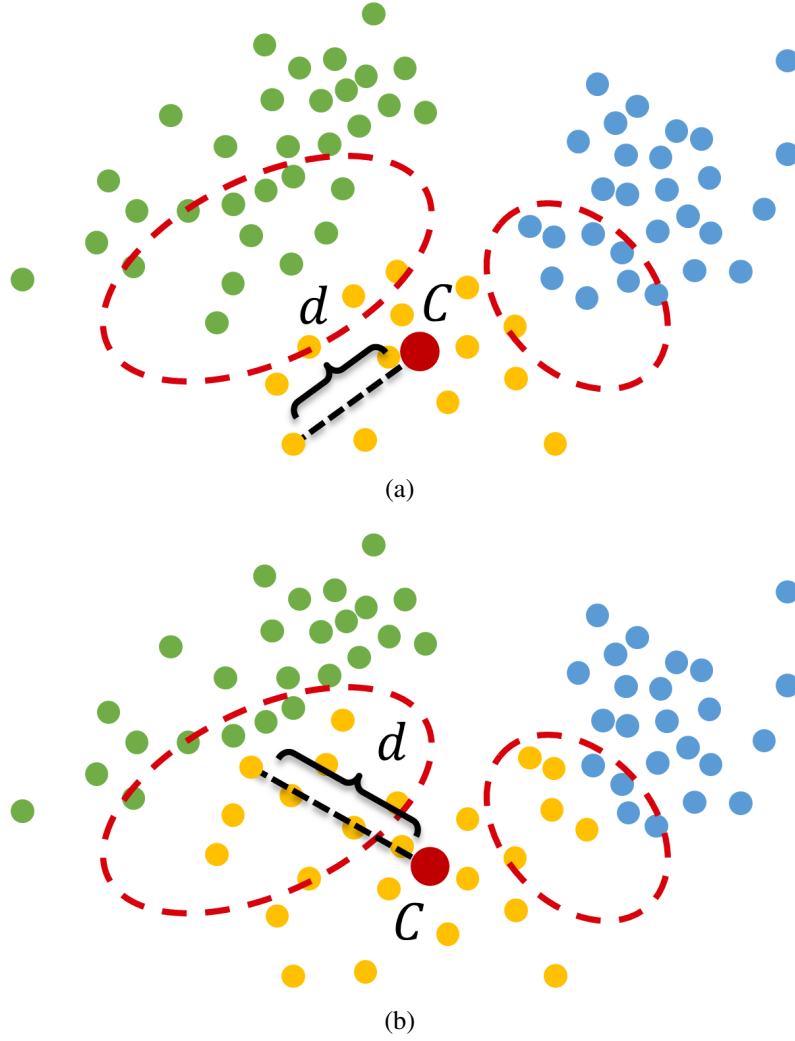


Figure 4.3. Illustration of the proposed clustering algorithm, where C is the cluster centroid and d is the maximum distance from centroid to objects. (a) K-means clustering results. (b) Regulated K-means clustering results.

algorithm. In other words, our approach strives to achieve the dual goals of compact, contiguous, and balanced clusters, without introducing prohibitive computation or fragmentation.

Therefore, our proposed regulated K-means pre-clustering can guide PSO to initialize parameters with a stronger ability to converge to the global optimum than conventional methods. By achieving more balanced and spatially meaningful clusters at the pre-clustering step, the search space for subsequent optimization is both reduced and better aligned with the actual requirements of multi-cell configuration, helping PSO avoid poor local optima and converge more efficiently.

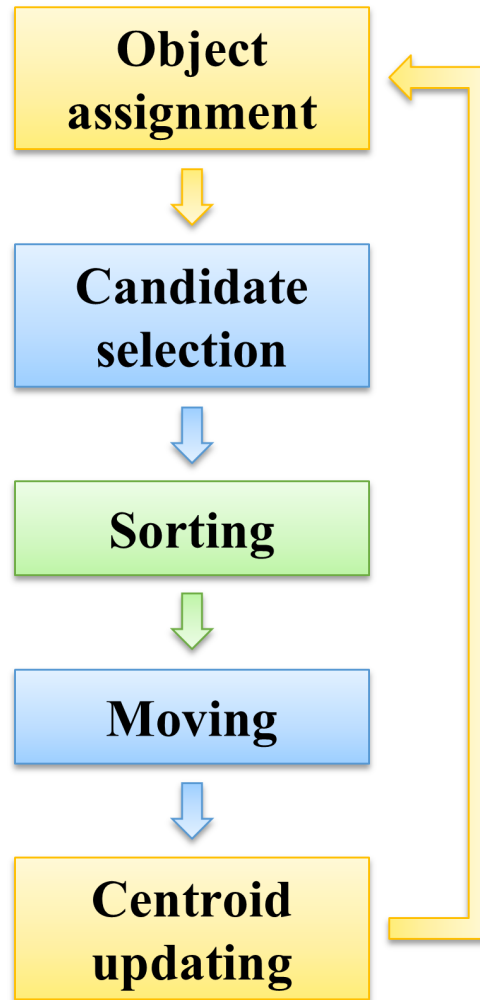


Figure 4.4. Illustration of the complete regulated K-means algorithm iteration.

The complete iteration process of the proposed algorithm can be separated into five steps: object assignment, candidate selection, sorting, moving, and centroid updating, as shown in Fig. 4.4. The algorithm begins by initializing centroids, which can be performed by K-means++. This initialization approach improves the chance of better convergence by providing a diverse set of starting points for the iterative process. The initial step involves calculating a standard cluster size S by dividing the total number of objects N by the desired number of clusters K . This standard serves as a target for balancing efforts during each subsequent iteration.

The algorithm then iterates through a process where each object is assigned to the nearest centroid, ensuring an initial grouping. During each iteration, the algorithm evaluates each cluster to determine if its size is below the standard size S . If so, it identifies

potential candidates for reallocation by calculating the Euclidean distance d_c between cluster centroids and considering objects in nearby clusters that are within this distance. This step leverages the spatial structure of the clusters, making sure that only objects which are suitably close are considered for movement, thereby preserving the overall geometric arrangement of the clusters.

These candidates are then sorted based on the absolute difference between their distance to the current cluster's centroid and to the potential new cluster's centroid. This sorting ensures that objects are reallocated in a way that minimizes overall disruption to the clusters' layout. By moving the best-matched objects first, the algorithm can incrementally improve balance with minimal negative impact on cluster compactness. The goal is to achieve a balance in cluster sizes while ensuring that objects are moved to where they fit spatially, thus maintaining the cohesion and effectiveness of the clustering. Such an approach ensures that both size and spatial proximity are considered throughout the process, addressing the shortcomings of earlier algorithms.

Following the sorted order of candidates, a meticulous selection process is undertaken. Objects identified for potential reallocation are moved to achieve cluster size balance, adhering to two key criteria. The move must not compromise the standard cluster size S , and each object can be moved only once to prevent multiple reassignments. This strict enforcement prevents oscillations or repeated moves, which can otherwise slow down convergence or degrade cluster integrity. This process continues iteratively, updating cluster centroids after each round, until the centroids stabilize without further movement or until a maximum number of iterations T_c is reached. This stopping rule ensures that the algorithm terminates within a reasonable number of steps while still providing an opportunity for convergence.

This regulated approach to K-means clustering ensures that clusters are not only defined by the distance of objects to their centroids but also by a deliberate consideration of cluster size uniformity, enhancing the algorithm's applicability to scenarios where balanced cluster distribution is critical. The complete proposed algorithm is illustrated in

Algorithm 2 Regulated K-Means Clustering Algorithm

Input: $\mathbf{U}, N, K, T_c$ **Output:** $\mathbf{C} = \{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \dots, \mathbf{c}_K\}$

- 1: Initialize K cluster centroids for N objects
 - 2: Calculate standard cluster size $S = N/K$
 - 3: **repeat**
 - 4: Assign each object in $\mathbf{U}$ to its closest centroid in $\mathbf{C}$
 - 5: **for** each cluster **do**
 - 6: **if** cluster size $< S$ **then**
 - 7: Calculate d_c as euclidean distance between its centroid and nearest adjacent centroid
 - 8: Record all objects belonging to other clusters within d_c as candidates to move in
 - 9: **end if**
 - 10: **end for**
 - 11: Sort all candidate objects in ascending order according to difference between their distance to current and destination cluster centroids
 - 12: **for** each candidate in order **do**
 - 13: **if** not yet moved **and** destination cluster size $< S$ **and** current cluster size $> S$ **then**
 - 14: Move to its destination cluster
 - 15: **end if**
 - 16: **end for**
 - 17: Update cluster centroids $\mathbf{C}$
 - 18: **until** cluster centroids do not move **or** maximum iterations T_c are reached
 - 19: **return** $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \dots, \mathbf{c}_K$
-

Algorithm 2. The use of both spatial and balancing mechanisms makes regulated K-means particularly well-suited to HAPS optimization, where both fairness and geometric regularity are essential.

The computational complexity of the algorithm can be determined based on K-means, which has a linear average complexity of $O(KNT_c)$ for 2-dimensional data. For the regulated K-means algorithm, the only difference from K-means lies in the sorting step, which introduces additional computational overhead in each iteration. Therefore, the time complexity is expressed as $O((KN + x \log x)T_c)$, where x is the number of candidate objects identified for potential reallocation in each iteration, and $O(x \log x)$ accounts for the additional cost introduced by the sorting algorithm. In practice, this additional sorting cost is usually modest, as the number of candidates is often much smaller than the total number of UEs, and the algorithm converges in a relatively small number of steps.

In particular, since the number of candidate objects to move is considerably less than the total number of objects, the complexity of our proposed algorithm remains close to that of the conventional K-means algorithm. This ensures that regulated K-means remains practical and scalable, even as the size of the network and the number of users grow.

Finally, the total complexity of our proposed optimization algorithm is the sum of the complexities of regulated K-means and PSO, expressed as $O((KN + x \log x)T_c + 4KPT)$, where $O(4KPT)$ denotes the complexity of the PSO algorithm. Notably, although the complexity of the PSO algorithm is expressed in terms of dimension $4K$, swarm size P , and iteration number T , the practical implementation time is also influenced by the cost of fitness evaluation using the HAPS model. Because the fitness must be evaluated for each swarm position at every iteration, the optimization step dominates the actual computational cost during implementation, while the complexity of the pre-clustering step in our proposed method is negligible. This further reinforces the motivation for proposing the clustering-aided PSO approach, as the pre-clustering step can improve optimization quality by reducing the search space when the maximum available swarm size and iteration number are constrained due to practical time limits.

4.5 Evaluation

4.5.1 Experiment Specifications

In the evaluation phase of this study, we conduct a comprehensive examination of cell optimization results using real-world UE distribution data to ensure the robustness and practicality of our proposed methods. The dataset utilized for our experiments is obtained from AGOOP Corp. [154], which provides extensive, high-quality spatial distribution data of UEs across Japan. From this dataset, we sample UE distribution data from 100 different coverage areas that collectively span the entire region of Japan, representing a wide variety of geographic environments and urbanization levels. In other words, the experimental results are derived from optimizing system performance across 100 distinct and independent UE distributions, thereby providing a diverse and statistically meaningful

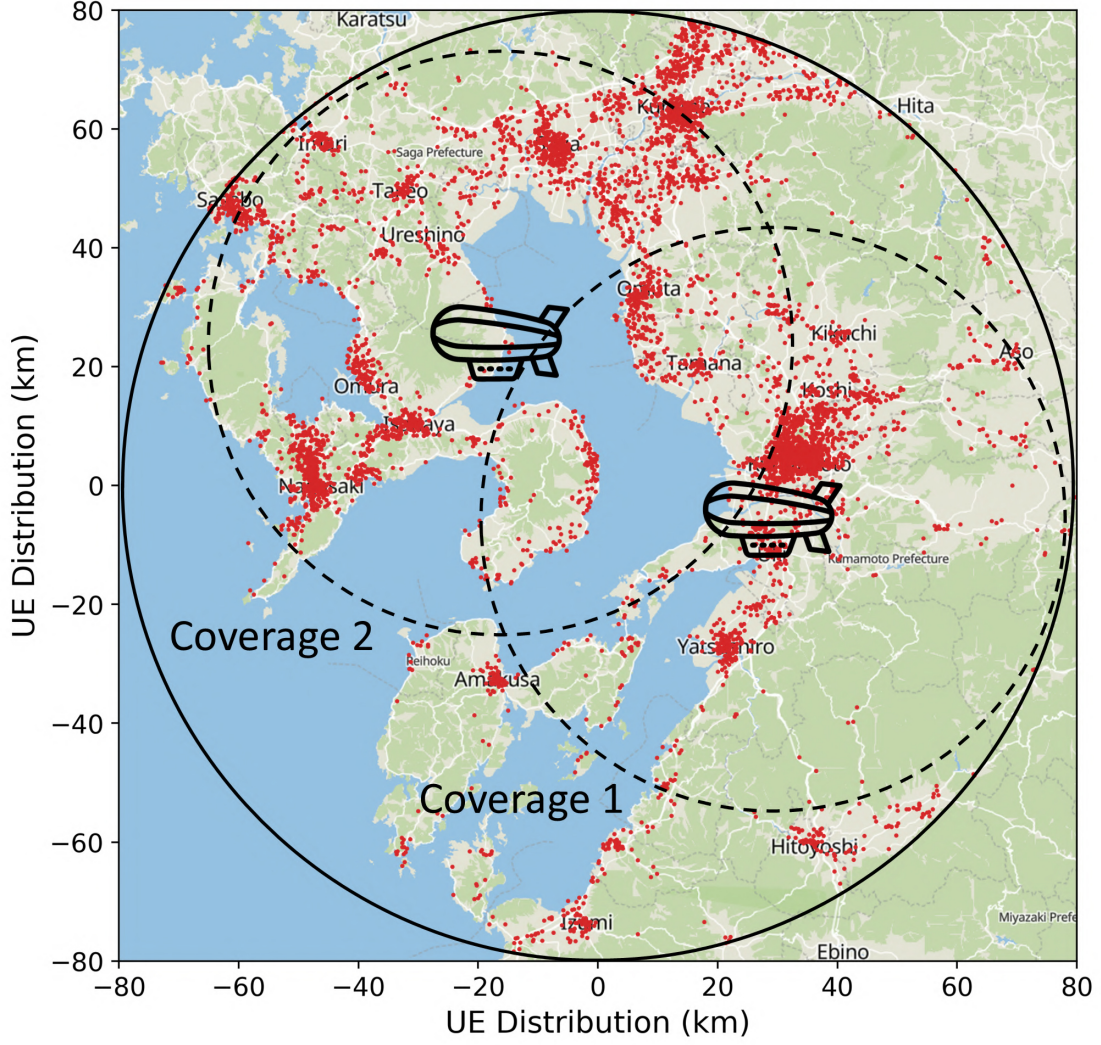


Figure 4.5. Illustration of example HAPS coverage area 1 and coverage area 2 in the Kyushu region.

evaluation. This large-scale sampling approach is intended to capture the broad variability that may be encountered in practical HAPS deployments, increasing the generalizability and credibility of the findings.

To facilitate detailed performance analysis and gain deeper insight into the system behavior under different deployment conditions, we further select five representative coverage areas for explicit presentation and discussion. These five areas are chosen based on their geographic diversity and relevance to typical HAPS deployment scenarios. Each area embodies a distinct UE distribution type, which can be attributed to unique geographic features, degrees of urbanization, and local environmental challenges that are highly relevant to HAPS network planning and operation. The locations of these areas,

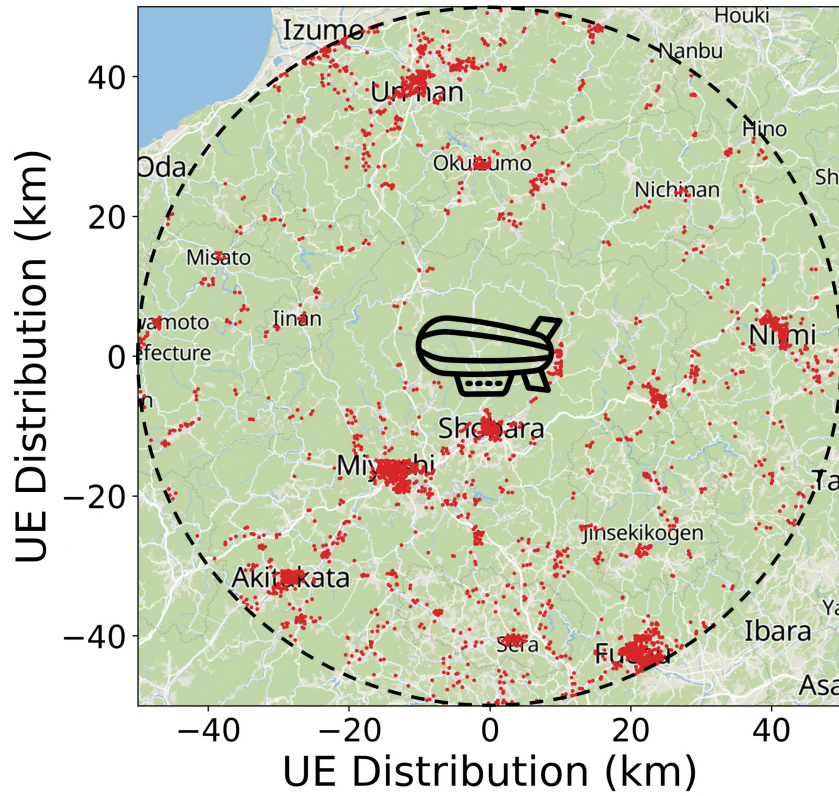


Figure 4.6. Illustration of example HAPS coverage area 3 in the Chugoku region.

along with the corresponding UE distributions, are visually illustrated in Figs. 4.5, 4.6, 4.7, and 4.8. These figures are generated based on OpenMapTiles and OpenStreetMap data [155], offering a clear geographical context for each case.

In addition to geographic variation, these areas are specifically selected for their potential vulnerability to various natural disasters, such as earthquakes, typhoons, floods, landslides, and tsunamis. The frequency and severity of such disasters depend on terrain, climate, and proximity to coastlines, making these regions particularly meaningful as representative scenarios for simulating the resilience and adaptability of HAPS deployments in real-world settings. The evaluation thus accounts for not only technical optimization under normal operating conditions, but also the critical role of HAPS in supporting emergency communication and rapid recovery in disaster-prone regions.

A closer examination of these example areas reveals their unique characteristics. As shown in Fig. 4.5, both area 1 and area 2 are situated in the Kyushu region of Japan, but they differ substantially in environmental context and UE spatial patterns. Area 1 encompasses an urban–suburban transition zone, incorporating mountainous terrain and

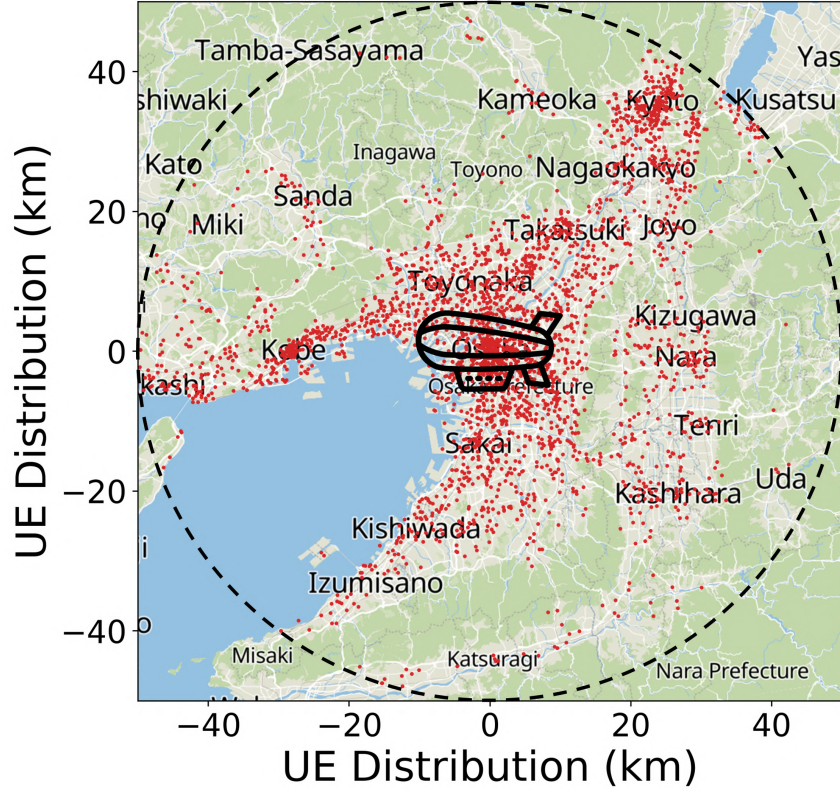


Figure 4.8. Illustration of example HAPS coverage area 5 in the Osaka Prefecture.

Area 4, located in Ibaraki Prefecture as shown in Fig. 4.7, comprises coastal plains, forests, and agricultural land in addition to both urban and suburban development. The result is a moderately dense UE distribution, with significant concentration in the urban and coastal regions, reflecting a mixture of population centers and rural expanse. Finally, area 5 is selected from Osaka Prefecture and is illustrated in Fig. 4.8. This region is a highly urbanized and industrialized metropolitan district, featuring continuous high-density UE distribution arranged in a radial pattern around a major urban and industrial core. This area serves as an archetype for dense metropolitan deployments where advanced cell optimization can yield substantial benefits.

The simulation settings used throughout the evaluation are systematically detailed in Table 4.1. In these experiments, we simulate a 9-cell HAPS system with a coverage radius of 50 km at a deployment altitude of 20 km, leveraging the cylindrical phased array antenna model described earlier in the system model section [142]. This configuration is selected to represent a typical operational scenario for HAPS, reflecting realistic deployment parameters encountered in practice. The choice of a 9-cell structure allows for a

Table 4.1. Simulation settings.

Parameter	Notation	Value
HAPS altitude	h	20 km
HAPS coverage radius	R	50 km
Number of cells	K	9
Number of sampled UEs	N	3000
Number of repeated runs	M	30
Cognitive parameter	c_1	2
Social parameter	c_2	2
Minimum inertia weight	$w_{\min}$	0.2
Maximum inertia weight	$w_{\max}$	0.9
Maximum velocity	$V_{\max}$	6
Random factors	r_1, r_2	[0, 1]
Swarm size	P	100
Iteration number	T	200
Maximum clustering iterations	T_c	100
Exponent factor	p	0.231, 1, 4.265, 13.513
Scaling factor	s	1000

thorough evaluation of multi-cell optimization while keeping computational requirements manageable.

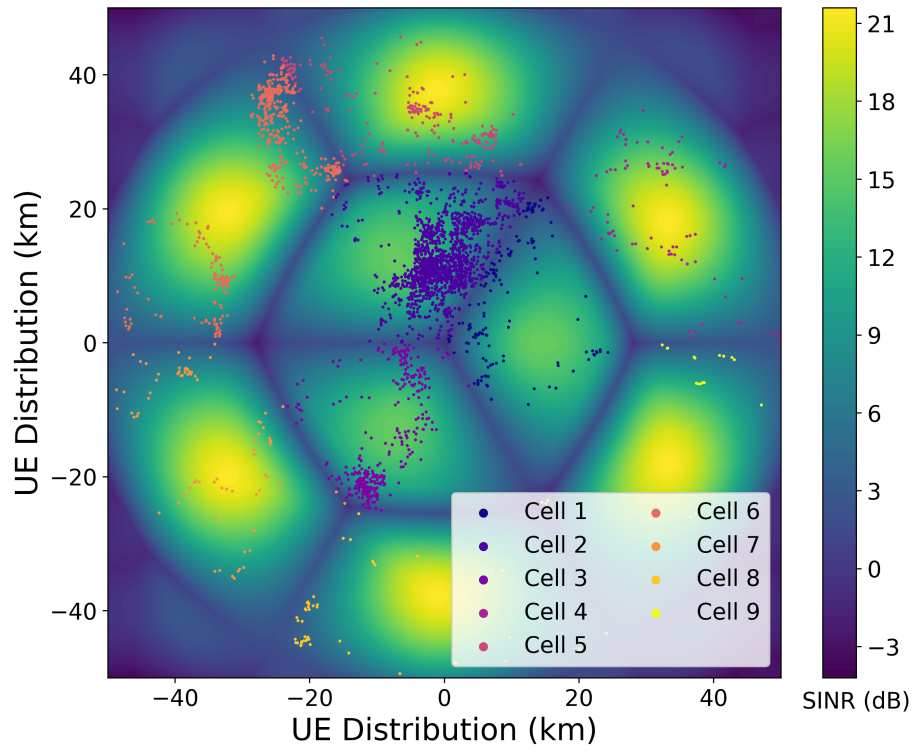
To maintain experimental consistency and statistical validity, we sample exactly 3,000 UEs within each distribution, a number determined to be sufficient for preserving essential spatial and density information without introducing unnecessary computational overhead. This approach ensures that the system's performance is assessed under comparable user loads across all evaluation scenarios. Regarding the meta-heuristic optimization process, we set the swarm size to 100 and the number of iterations to 200 for PSO in our proposed method. These values are chosen to represent an appropriately low complexity for practical PSO implementation, ensuring effective convergence to high-quality solutions

within a reasonable computational budget. More importantly, to highlight the comparative advantage of our method, we apply the same swarm size and iteration number as reference points for all competing methods, guaranteeing that all approaches are evaluated under an optimization effort and computational complexity that is at least as stringent as that used for our algorithm. This careful alignment is critical for fairness and validity in comparative benchmarking.

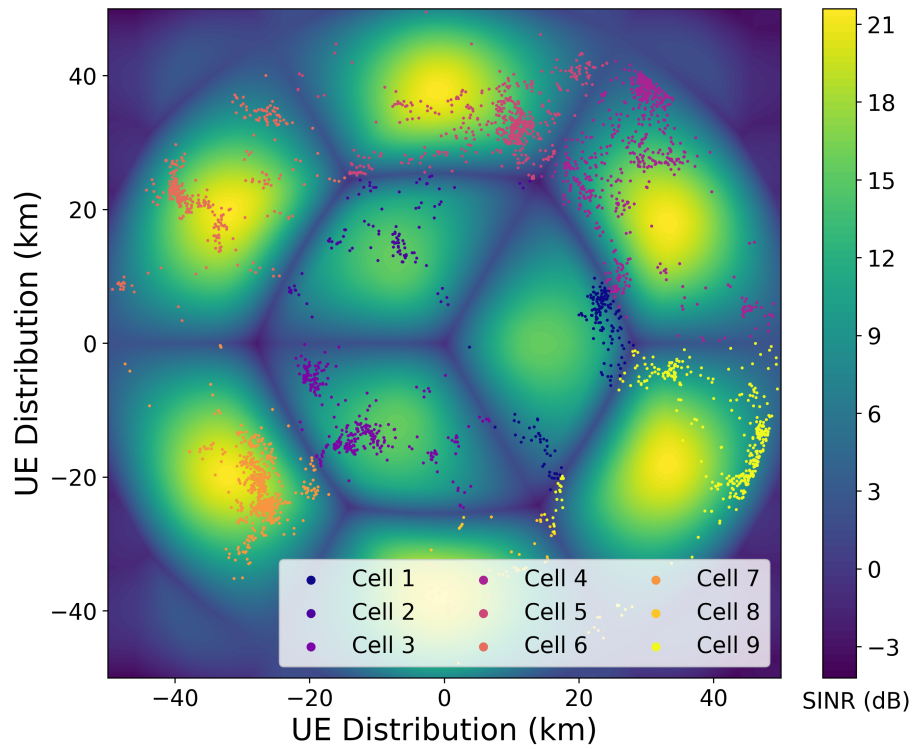
To establish comprehensive benchmarks, we evaluate multiple conventional methods alongside our proposed algorithm. The baseline for all comparisons is a uniform configuration, as developed in [59], where the 9 cells are uniformly configured regardless of actual UE distribution. This baseline represents the default system state prior to any UE distribution-based optimization and provides a clear reference for quantifying the improvements realized through more advanced methods. Fig. 4.9 presents the SINR distributions for this uniform configuration in the two example areas in Kyushu, serving as a visual and quantitative baseline.

Beyond the baseline, we examine the performance of conventional PSO and GA methods implemented alone under the same complexity settings as our approach. Additionally, we assess two modifications of PSO. The first is a sectorization-based method, which divides the circular coverage area into three equal sectors for the nine cells, thus roughly reducing the search range and simplifying the optimization process for PSO. The second is a heavy version of PSO, where the swarm size is increased to 500 and the number of iterations is increased to 1000, both of which are considered very large for this class of problems [156]. These alternative settings provide a means to evaluate the impact of computational effort on optimization performance.

Furthermore, our evaluation includes the proposed clustering-aided PSO method combined with four different clustering algorithms: conventional K-means, minmax K-means, constrained K-means with strictly equal-size clusters, and our newly proposed regulated K-means. To provide a comprehensive analysis, we also evaluate the GA method assisted



(a)



(b)

Figure 4.9. SINR distribution for the uniform cell configuration in the two example areas in Kyushu. (a) Example area 1. (b) Example area 2.

by our proposed pre-clustering algorithm. For all clustering-aided methods, the maximum number of clustering iterations T_c is consistently set to 100, ensuring comparability across different clustering strategies.

In addition to these, we assess several state-of-the-art methods from the literature on HAPS multi-cell optimization, as identified and discussed in the related work section. These include the geometry-based K-means method, the DQN method, and the co-evolutionary GA method [73, 76, 80]. In particular, since the DQN method can benefit from our proposed pre-clustering algorithm, we also evaluate the performance of the clustering-aided DQN approach to provide a thorough and contemporary comparison. The DQN method is evaluated with a step number of 100 and an episode number of 200, settings which, although resulting in higher actual complexity than our method due to network training overhead, are used to maintain a comparable level of optimization effort for fair benchmarking. For the co-evolutionary GA method, the population size is set to 100, and the iteration number to 50 per cell group. With a total of four cycles, this setup yields 200 iterations per parameter, resulting in an actual computational complexity three times greater than our proposed method, as the 9 cells are divided into three groups but the optimization effort per group is maintained at the same level.

All optimization-based methods in the comparison consistently utilize the objective function defined in (4.1) for the calculation of fitness. This ensures that differences in performance can be attributed solely to optimization effectiveness rather than to discrepancies in objective criteria. The main evaluation is performed using $p = \log_{0.85} 0.5 \approx 4.265$, which is chosen as an appropriate exponent to achieve our throughput improvement goal. This particular value of p ensures that the lowest 15% of UEs, as ranked by throughput, begin to receive weights exceeding 0.5, with the minimum throughput user receiving a weight of 1. This design places a significant emphasis on improving the throughput of the least served UEs.

To further demonstrate the consistency and robustness of our method under varying weight distributions, we also evaluate the performance for three additional values of p : 0.231, 1, and 13.513. These four values respectively correspond to the threshold for low

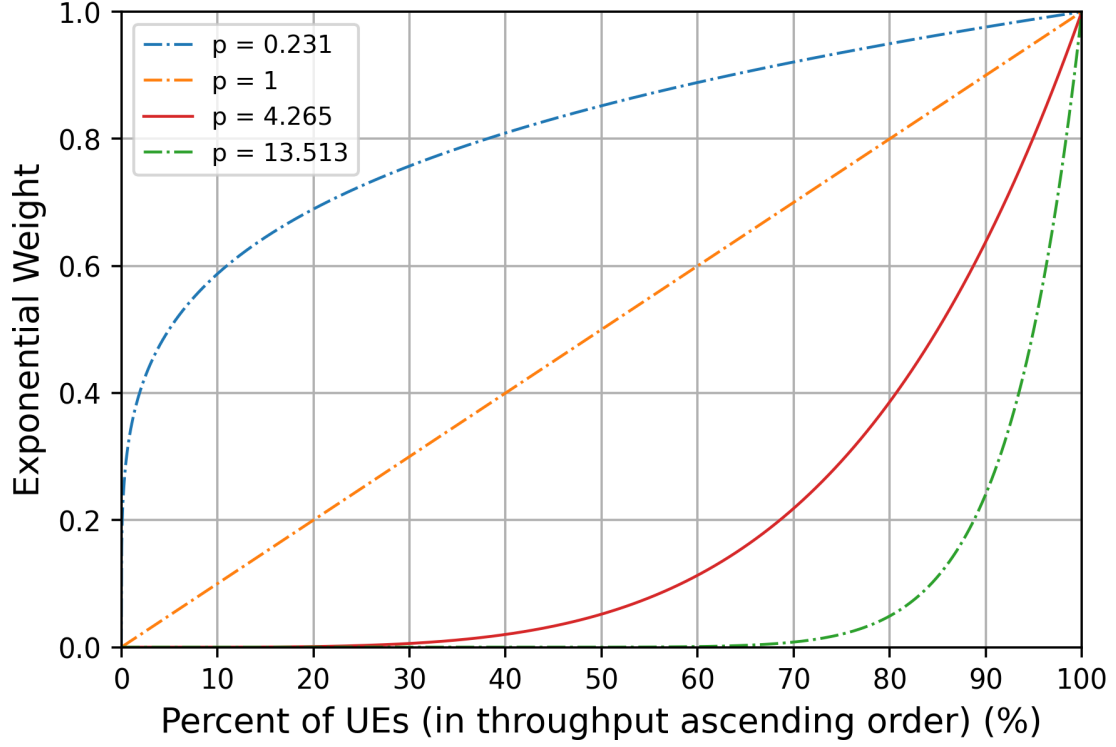


Figure 4.10. Comparison of multiple p values applied to the proposed method in terms of exponential weighting function.

throughput UEs to receive weights larger than 0.5 at 95%, 50%, 15%, and 5% of the sorted UEs, effectively varying the weight distribution from nearly uniform to heavily prioritized for the lowest throughput users. The characteristics of these weighting functions are graphically illustrated in Fig. 4.10. The scaling factor in the objective function is set to 1000 to maintain the numerical results within a readable and interpretable range.

In this simulation, we also make a practical and necessary assumption to constrain the beam directions within the coverage area, which limits the maximum vertical beam direction to $\arctan(R/h)$ and the maximum vertical and horizontal beamwidth to $2 \arctan(R/h)$. This adjustment is essential because the theoretical parameter ranges defined in Table 3.2 create an excessively large search space, posing substantial challenges for methods that do not employ search range reduction. Such a constraint ensures that all methods are compared fairly, as each is initially provided with a search space that aligns with the fundamental geometry and deployment scenario of the simulation. This, in turn, allows for a more meaningful evaluation of the optimization algorithms' ability to find high-quality solutions within practical bounds.

Overall, the experiment specifications are carefully designed to ensure both rigor and realism in performance assessment, reflecting the complex demands and variable scenarios encountered in real-world HAPS deployments. By thoroughly benchmarking across a diverse array of conditions and method types, we provide a robust foundation for analyzing and demonstrating the effectiveness and generalizability of the proposed clustering-aided PSO approach.

4.5.2 Optimization Results

In this section, we comprehensively evaluate the optimization results of the proposed clustering-aided PSO approach and a variety of benchmark algorithms by analyzing their performance using the fitness score metric. The fitness score provides a direct, quantitative measure of the quality of the cell configuration achieved by each method, making it a reliable indicator for comparing different optimization strategies. A lower fitness score signifies a more desirable solution, reflecting improved UE throughput and more balanced distribution of service across the network. The use of this clear and interpretable metric allows for rigorous assessment of the methods under diverse conditions and supports objective benchmarking.

To begin the evaluation, we conduct a detailed fitness analysis for $p = 4.265$, which, as established in the experiment specifications, is a well-matched choice for our optimization objective. This exponent is selected to emphasize improvements for UEs with the lowest throughput, supporting fairness and robustness in the optimized cell configuration. Table 4.2 presents the fitness results across 100 different UE distributions for each evaluated method. The column labeled “Final” summarizes the average final fitness scores over 100 distributions after cell optimization, with each distribution’s score computed as the average of 30 independent runs to ensure statistical significance and to account for inherent algorithm stochasticity.

The results in Table 4.2 clearly indicate that our proposed method consistently achieves the lowest average fitness score among all compared methods. This demonstrates the superior optimization capability of our clustering-aided PSO approach in adapting to

varying UE distributions and maximizing overall network performance. In contrast, the geometry-based K-means method is found to yield the highest fitness scores, with only limited improvement over the baseline. This is an expected outcome, as the geometry-based method lacks a mechanism to optimize antenna parameters based on actual UE throughput; it instead relies solely on fixed geometric rules for configuration, which are insufficient for capturing the dynamic and spatially heterogeneous nature of practical deployments.

Further analysis reveals that the DQN method performs worse than the other optimization-based methods, highlighting the advantages of meta-heuristic approaches in this particular problem setting. The meta-heuristic optimizers, such as PSO and GA, benefit from a clearly defined objective function, enabling them to explore the search space more systematically and converge toward high-quality solutions. Among these, standard PSO outperforms standard GA, reaffirming the suitability of PSO as a population-based optimizer for the complex and high-dimensional optimization landscape encountered in multi-cell HAPS configuration problems.

A key aspect of the evaluation is the comparison of different strategies for search space reduction and algorithmic enhancement. The results show that both the computationally heavier PSO variant and the simplified sectorization-based version achieve lower fitness scores than standard PSO, but the most significant improvements are observed with the four clustering-aided PSO methods. These include PSO guided by conventional K-means, minmax K-means, constrained K-means, and our proposed regulated K-means. All four variants yield notably lower fitness scores, underscoring the critical importance of carefully constraining the search space. By utilizing meaningful clustering information to redefine the search range, these methods guide the optimizer away from local optima and toward globally superior solutions.

Among the clustering-aided approaches, both the constrained K-means and regulated K-means methods consistently outperform the conventional K-means variant, validating the positive effect of balanced clustering on optimization quality. In contrast, the minmax

Table 4.2. Fitness results across 100 UE distributions.

Method	Average Fitness Score			
	1st Iteration	Final	Final (Best Run)	CV
Base (Uniform)	–	12.79	12.79	–
K-Means (Geometry)	–	10.34	8.77	9.27%
DQN	–	6.92	6.49	5.59%
GA	16.05	6.92	5.99	8.71%
GA (Co-Evolution)	–	6.69	5.70	9.44%
PSO	15.95	5.89	4.93	8.83%
PSO (Sectorization)	13.72	5.67	4.71	9.66%
PSO (Heavy)	13.82	5.38	4.58	7.73%
DQN (Regulated K-Means)	–	5.89	4.76	10.27%
GA (Regulated K-Means)	8.81	5.63	4.65	6.79%
PSO (MinMax K-Means)	9.90	5.32	4.54	8.12%
PSO (K-Means)	9.52	5.23	4.53	6.58%
PSO (Constrained K-Means)	8.75	5.03	4.46	5.46%
PSO (Regulated K-Means)	8.73	4.75	4.31	5.66%

^aThe results are presented for $p = 4.265$.

K-means-based approach, despite its adaptive balancing mechanism, delivers worse performance than even the conventional K-means, a result attributable to its instability and occasional poor clustering outcomes. The regulated K-means-based approach stands out, consistently achieving the lowest fitness score and thus the best performance among all methods tested. Because the only varying factor among the clustering-aided methods is the clustering algorithm, these results confirm that our proposed regulated K-means algorithm provides the most effective guidance for parameter search range redefinition and overall optimization.

The benefit of the pre-clustering step is not limited to PSO-based approaches. Table 4.2 also shows that both the clustering-aided DQN and clustering-aided GA methods achieve substantial improvements over their standard versions. This highlights the general applicability and effectiveness of our pre-clustering algorithm in facilitating better initialization and search range reduction for a variety of population-based and learning-based optimizers. On the other hand, although the co-evolutionary GA method also yields lower fitness scores than standard GA, the improvement is less pronounced compared to the clustering-aided strategy, emphasizing the importance of intelligent search space reduction.

To further assess robustness, we calculate the average coefficient of variation (CV) of the final fitness scores across repeated runs for each method. The CV provides insight into the consistency and reliability of each algorithm in achieving similar results under multiple runs, which is particularly relevant for stochastic optimizers. The results reveal that clustering-aided optimization approaches—including those employing K-means, minmax K-means, constrained K-means, and regulated K-means—achieve relatively low CV values, reflecting increased robustness and reduced variance in outcomes. Notably, although the proposed regulated K-means-based approach does not achieve the absolute lowest CV, its value is among the lowest, demonstrating stable and repeatable performance.

It is reasonable that the constrained K-means-based approach achieves the lowest CV, as its strict enforcement of cluster size produces highly similar cluster structures and search spaces across different runs. An interesting observation arises in the case of the DQN method, whose CV increases after applying the pre-clustering step. This is because the DQN approach, which searches by adjusting a single candidate solution, becomes sensitive to the initialization provided by pre-clustering, leading to greater run-to-run variability. In the original DQN method, fixed initial parameters ensure low CV, whereas the stochastic nature of the pre-clustering step introduces variability. In contrast, population-based methods such as PSO and GA naturally average over multiple candidate solutions, rendering them more compatible with the stochastic initialization provided by pre-clustering.

Table 4.3. Fitness results in example areas.

Method	Fitness Score (Example Area)				
	Area 1	Area 2	Area 3	Area 4	Area 5
Base (Uniform)	18.43	7.34	9.33	7.82	12.89
K-Means (Geometry)	14.27	5.94	7.30	6.88	9.71
DQN	8.00	5.11	6.58	6.35	6.04
GA	6.94	6.01	6.05	6.50	6.33
GA (Co-Evolution)	6.80	5.31	5.64	6.07	5.51
PSO	6.20	5.06	5.36	5.46	6.06
PSO (Sectorization)	5.87	4.70	4.72	5.11	5.15
PSO (Heavy)	5.68	4.56	5.00	5.01	5.72
DQN (Regulated K-Means)	6.33	4.15	5.41	5.15	5.44
GA (Regulated K-Means)	6.06	4.67	4.62	5.11	5.36
PSO (MinMax K-Means)	6.15	3.78	4.32	4.53	4.97
PSO (K-Means)	5.87	3.50	4.19	4.49	4.82
PSO (Constrained K-Means)	5.44	3.60	4.01	4.50	4.60
PSO (Regulated K-Means)	5.28	3.40	3.86	4.37	4.49

^aThe results are presented for $p = 4.265$.

Beyond final scores, we also record the average fitness score at the first iteration of optimization to assess the quality of the initialization provided by pre-clustering. This comparison is only meaningful for methods that optimize all cells simultaneously. The findings demonstrate that pre-clustering-based methods yield improved fitness at the outset, with the regulated K-means variant consistently providing the best initialization among all evaluated approaches. This strong initialization advantage translates into improved final results after iterative optimization. Additionally, by recording the lowest fitness scores obtained from 30 repeated runs per distribution, we can evaluate each method’s best-case potential. Although these scores are naturally lower, the relative performance rankings

Table 4.4. Ranking analysis of how often each method outperforms the others across 100 UE distributions from 1st to 7th.

Method	Occurrences by Fitness Score Rank Position						
	1st	2nd	3rd	4th	5th	6th	7th
PSO (Regulated K-Means)	75%	18%	5%	1%	1%	–	–
PSO (Constrained K-Means)	11%	39%	17%	21%	8%	4%	–
PSO (K-Means)	6%	14%	28%	13%	14%	9%	12%
PSO (MinMax K-Means)	3%	12%	16%	20%	11%	10%	7%
GA (Regulated K-Means)	–	–	7%	6%	23%	31%	26%
DQN (Regulated K-Means)	5%	7%	6%	9%	5%	5%	6%
PSO (Heavy)	–	9%	15%	16%	20%	15%	8%
PSO (Sectorization)	–	–	5%	9%	11%	15%	20%
PSO	–	–	–	3%	5%	7%	6%
GA (Co-Evolution)	–	–	–	1%	1%	2%	4%
GA	–	–	–	–	–	–	–
DQN	–	1%	1%	1%	1%	2%	11%
K-Means (Geometry)	–	–	–	–	–	–	–
Base (Uniform)	–	–	–	–	–	–	–

^aThe results are according to the fitness scores in the “Final” column of Table 4.2.

remain unchanged, further supporting the superiority and consistency of our proposed method.

A more granular analysis is presented in Table 4.3, which details the fitness results for the five example UE distributions previously introduced. Due to the unique features of each coverage area, the fitness scores exhibit distinct patterns. It is observed that denser distributions—such as those in areas 1, 4, and 5—produce higher fitness scores, as the clustering and optimization process must contend with increased inter-cell interference and greater complexity in balancing throughput among a larger number of edge UEs. Among these, area 1, which features a small and highly concentrated urban core, yields

Table 4.5. Ranking analysis of how often each method outperforms the others across 100 UE distributions from 8th to 14th.

Method	Occurrences by Fitness Score Rank Position						
	8th	9th	10th	11th	12th	13th	14th
PSO (Regulated K-Means)	–	–	–	–	–	–	–
PSO (Constrained K-Means)	–	–	–	–	–	–	–
PSO (K-Means)	3%	1%	–	–	–	–	–
PSO (MinMax K-Means)	9%	8%	4%	–	–	–	–
GA (Regulated K-Means)	7%	–	–	–	–	–	–
DQN (Regulated K-Means)	17%	19%	8%	6%	3%	4%	–
PSO (Heavy)	8%	6%	3%	–	–	–	–
PSO (Sectorization)	23%	10%	7%	–	–	–	–
PSO	19%	27%	25%	8%	–	–	–
GA (Co-Evolution)	7%	15%	17%	26%	22%	5%	–
GA	2%	5%	24%	29%	34%	6%	–
DQN	5%	9%	11%	27%	27%	4%	–
K-Means (Geometry)	–	–	1%	4%	14%	65%	16%
Base (Uniform)	–	–	–	–	–	16%	84%

^aThe results are according to the fitness scores in the “Final” column of Table 4.2.

the highest fitness scores. In contrast, sparser distributions, such as those in areas 2 and 3, generally lead to lower fitness scores, with area 2 presenting the most natural fit for a finely-tuned multi-cell configuration. These variations underscore the impact of both inter-cell interference and cluster balance on the optimization outcome. Importantly, despite these differences, our proposed method outperforms other approaches across all representative UE distribution types, demonstrating adaptability and effectiveness in a wide range of deployment scenarios.

To further evaluate each method’s ability to achieve optimal solutions across different UE distributions, we conduct a comprehensive ranking analysis. Tables 4.4 and 4.5

Table 4.6. Average final fitness score results of 100 UE distributions for additional exponent factors.

Method	$p = 0.231$	$p = 1$	$p = 13.513$
Base (Uniform)	31.10	23.96	5.99
K-Means (Geometry)	24.09	18.79	5.07
DQN	17.80	13.15	3.50
GA	18.38	13.74	3.10
GA (Co-Evolution)	18.39	13.45	3.12
PSO	15.71	11.48	2.73
PSO (Sectorization)	14.95	10.96	2.65
PSO (Heavy)	14.64	10.62	2.46
DQN (Regulated K-Means)	14.79	10.93	3.05
GA (Regulated K-Means)	14.53	10.70	2.61
PSO (MinMax K-Means)	14.27	10.52	2.59
PSO (K-Means)	13.60	10.05	2.50
PSO (Constrained K-Means)	13.40	9.70	2.29
PSO (Regulated K-Means)	12.72	9.25	2.25

present the number of times each method achieves the best, second-best, and subsequent rankings based on the average final fitness score for each distribution. The results show that the clustering-aided approaches, especially our proposed regulated K-means-based method, dominate the top rankings. Our method achieves the best performance on 75 out of 100 distributions, ranks second on 18, third on 5, and fourth or fifth on only 1 each, with no lower rankings. This not only confirms the method's superior average performance, but also its ability to consistently deliver top results across a diverse set of deployment scenarios.

Other clustering-aided variants, such as K-means, minmax K-means, and constrained K-means-based methods, also achieve best performance in some cases, reflecting the possibility that different clustering algorithms may produce similar clusters and search spaces

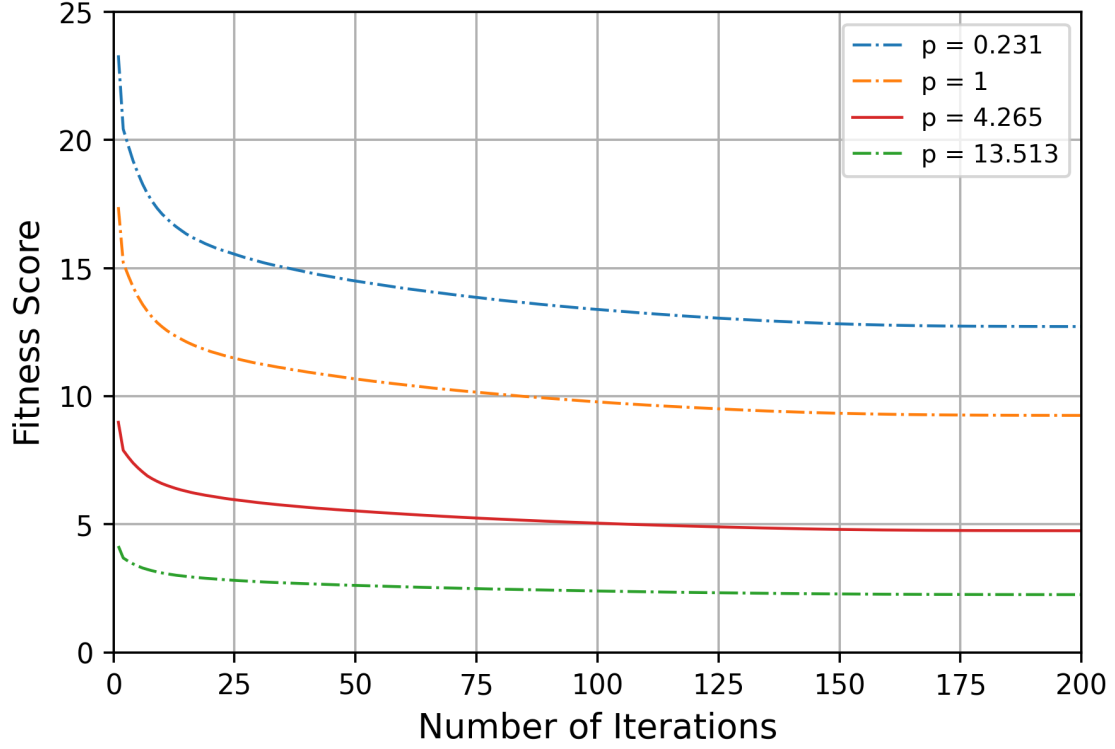


Figure 4.11. Comparison of multiple p values applied to the proposed method in terms of convergence curve averaged across 100 UE distributions.

under certain conditions. The DQN-based method, by contrast, exhibits high variability in ranking—from best in some scenarios to much lower ranks in others—highlighting the challenges of reward function design and robustness for learning-based optimizers in this context.

The consistency of our proposed method is further validated through evaluation under different weighting distributions. In Table 4.6, we present the average final fitness scores for the 100 UE distributions at three additional p values, supplementing the main evaluation at $p = 4.265$. Across all four p values, our regulated K-means-based approach achieves the lowest fitness score, demonstrating consistent superiority over alternative methods regardless of the chosen weighting emphasis. The numerical values of the fitness scores vary as a function of p , reflecting different prioritizations of low-throughput UEs, but the relative performance of each method remains clear and consistent. For example, at $p = 13.513$, the performance of the conventional K-means-based approach no longer exceeds that of the heavy PSO variant, whereas the regulated K-means-based method still

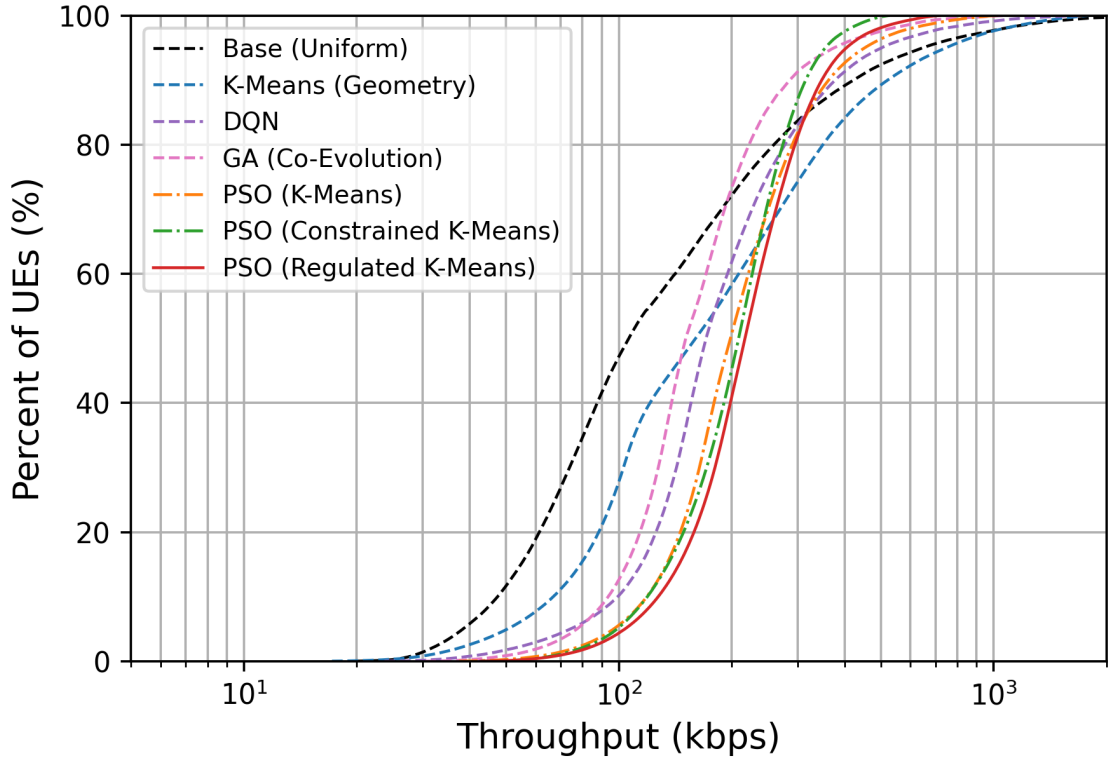
achieves the best results. This further confirms that the search ranges defined by regulated K-means are more precise and effective than those generated by conventional K-means.

Finally, we illustrate the convergence behavior of the proposed method by plotting the average convergence curves over 100 UE distributions for all four p values, as shown in Fig. 4.11. These results show that our method reliably achieves convergence within 200 iterations in every case, providing further evidence of both its efficiency and robustness across a wide spectrum of weighting strategies and distribution types.

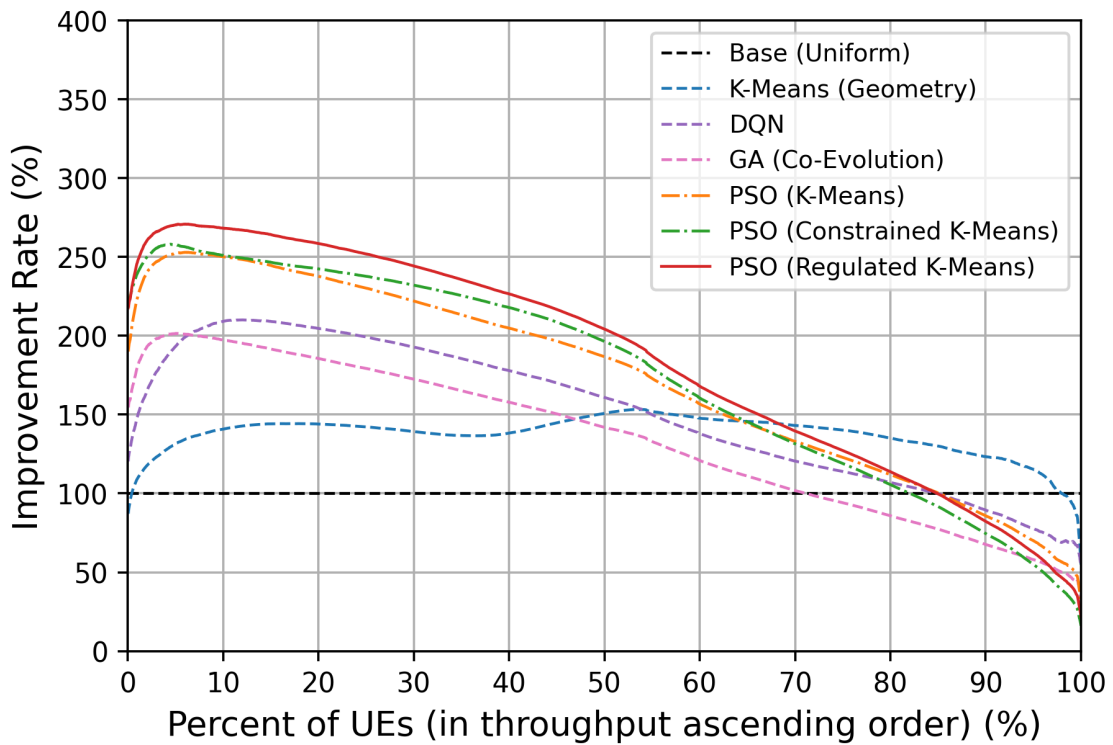
4.5.3 Throughput Performance

To provide a thorough and comprehensive comparison of the throughput performance under the optimized cell configurations, we analyze and visualize the distribution of UE throughput resulting from each evaluated method, as well as from the baseline scenario. The cumulative distribution function (CDF) of UE throughput is used as the primary metric for this comparison, as it provides an intuitive and detailed illustration of the throughput levels experienced by all UEs in the system. This approach enables the assessment of both the average performance and the fairness of throughput allocation across the entire user population.

We plot the optimized CDFs in Fig. 4.12(a), which are averaged over all 100 UE distributions, with each distribution result further averaged over 30 repeated optimization runs for $p = 4.265$. This plot allows for a direct comparison of the throughput distributions achieved by our proposed clustering-aided method, various conventional methods, and the uniform baseline configuration. The rightward shift of the CDF curve corresponding to our proposed method is especially prominent, signifying a substantial reduction in the number of UEs experiencing low throughput. This demonstrates that our approach consistently delivers superior throughput performance for the vast majority of UEs when compared to all alternative methods. This observation is strongly consistent with the previously discussed fitness score results, as both metrics are closely linked to the effectiveness of the underlying cell optimization.



(a)



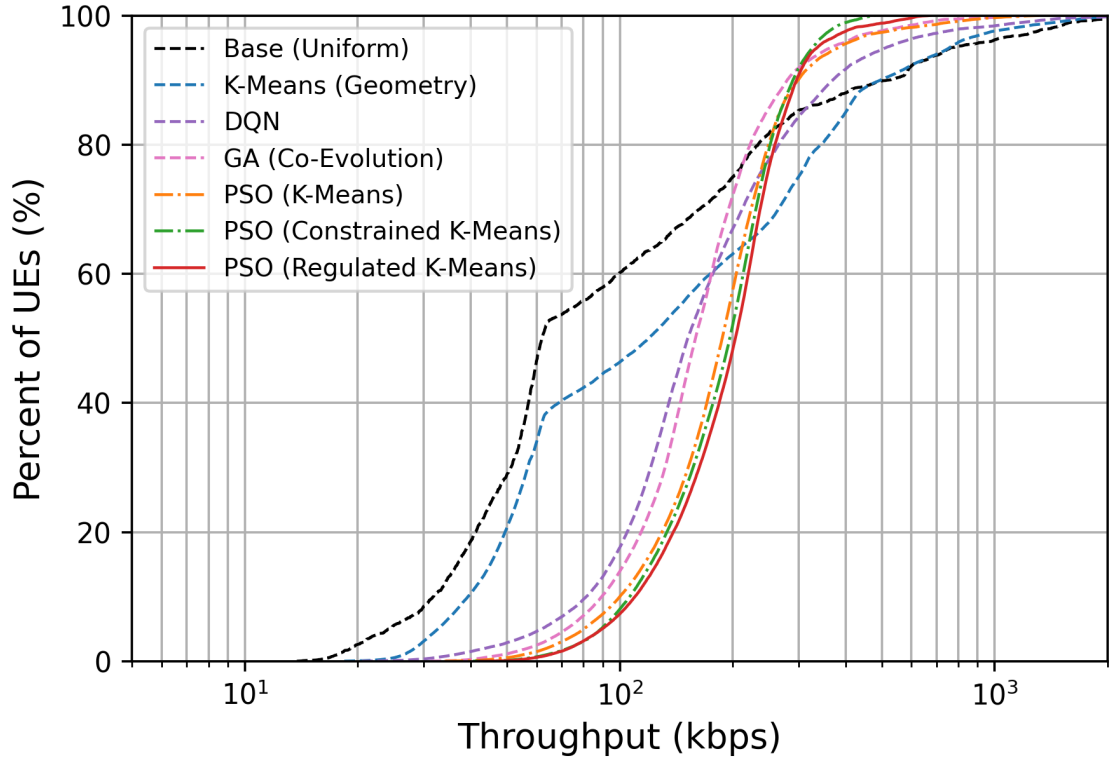
(b)

Figure 4.12. Throughput performance average across 100 UE distributions for $p = 4.265$. (a) CDF of UE throughput. (b) Distribution of throughput improvement over the uniform cell configuration from low throughput UEs to high throughput UEs.

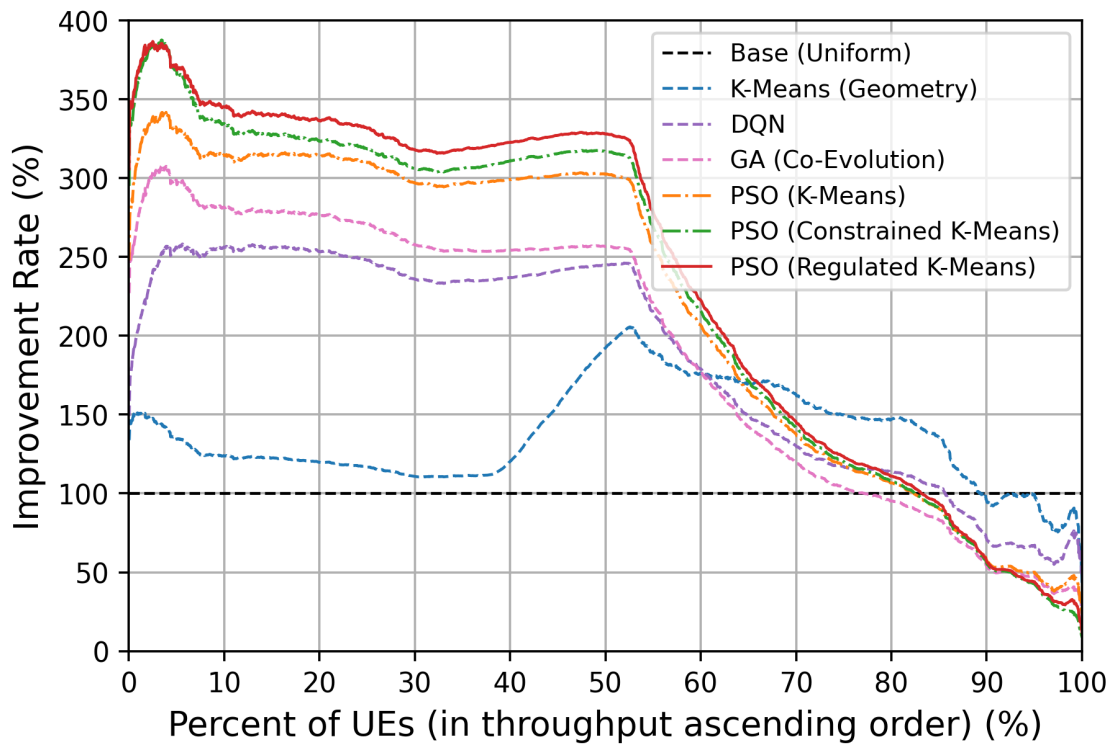
To gain further insights, we also examine the throughput improvement rate for each method, as shown in Fig. 4.12(b), where the improvement rate is computed relative to the uniform configuration, thereby isolating the gains attributable to the optimization process itself. This visualization provides a clear and quantitative picture of how much each method improves throughput across the spectrum of UEs, from those with the lowest to those with the highest throughput. Our results show that the proposed method is especially effective at improving the throughput of UEs at the lower end of the throughput distribution, which directly aligns with our stated optimization objective of prioritizing performance for UEs with poorer initial connectivity. Simultaneously, the method is designed to de-emphasize excessive gains for already well-served UEs, ensuring a fairer and more balanced network outcome. These findings are in full agreement with the theoretical goals of our objective function and validate the design of our approach.

In addition to the aggregate results, we present detailed throughput performance for $p = 4.265$ in the five example coverage areas introduced previously. The results are visualized in Figs. 4.13, 4.14, 4.15, 4.16, and 4.17. By analyzing these representative scenarios, we are able to demonstrate the robustness and generalizability of our proposed method across different deployment conditions and UE distribution patterns. Importantly, the performance trends observed in these examples are consistent with those in the overall fitness analysis. In particular, the reduction of low throughput UEs is noticeably more challenging in coverage areas associated with higher fitness scores, reaffirming the impact of challenging geographic and distributional features on optimization complexity. Nevertheless, our approach maintains its advantage across all scenarios.

To further explore the effects of the weighting function parameter p , we evaluate throughput performance across the full range of optimization objectives by presenting throughput improvement distributions for four distinct p values, as summarized in Fig. 4.18. By comparing the curves corresponding to different p values, we observe clear trends: smaller p values lead to less improvement in the low throughput region and greater improvement in the high throughput region, while higher p values have the opposite effect,

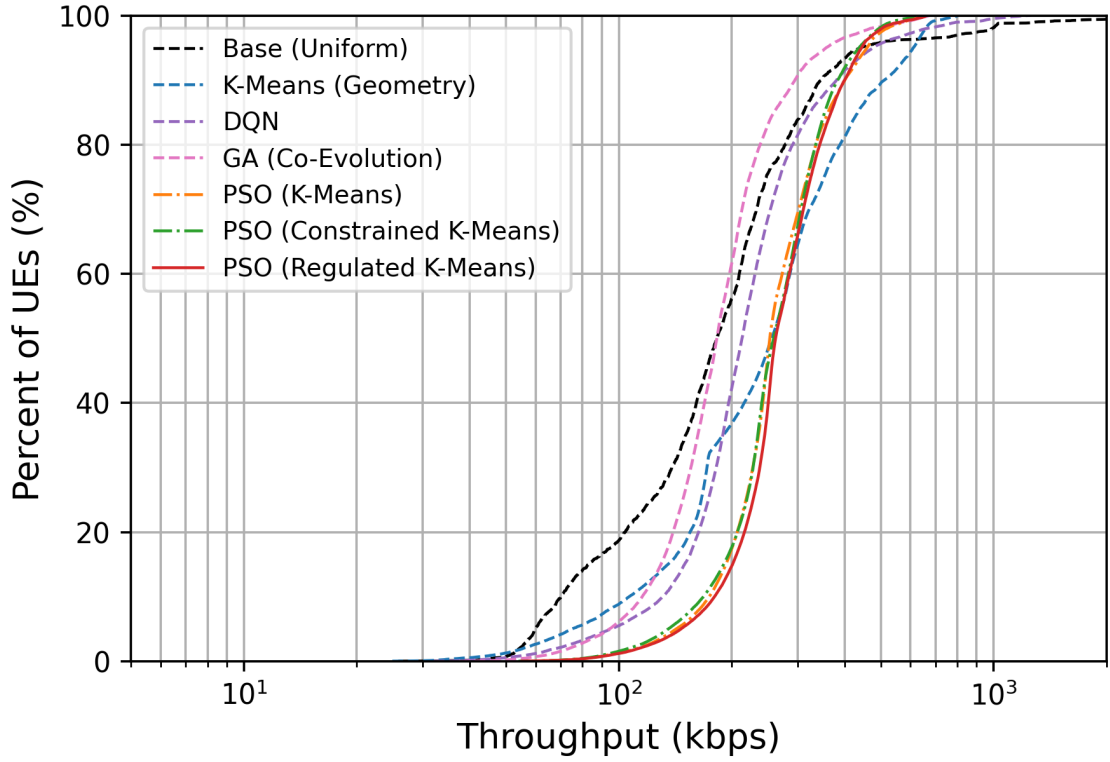


(a)

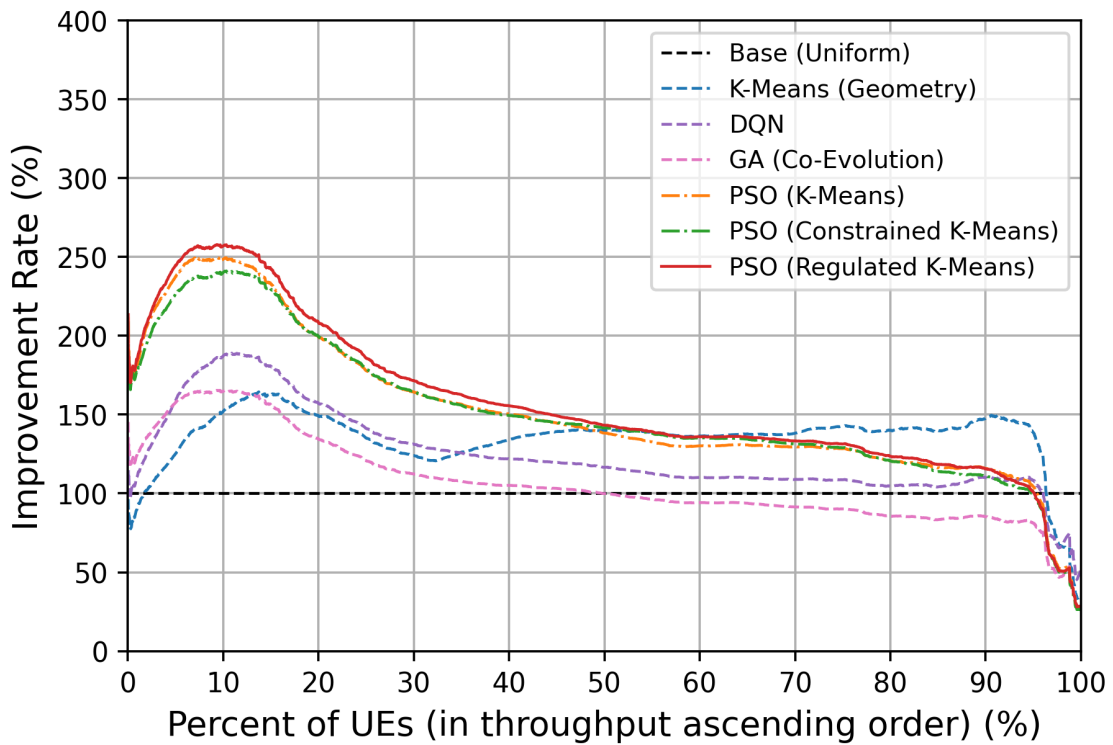


(b)

Figure 4.13. Throughput performance in example area 1 for $p = 4.265$. (a) CDF of UE throughput. (b) Distribution of throughput improvement over the uniform cell configuration from low throughput UEs to high throughput UEs.

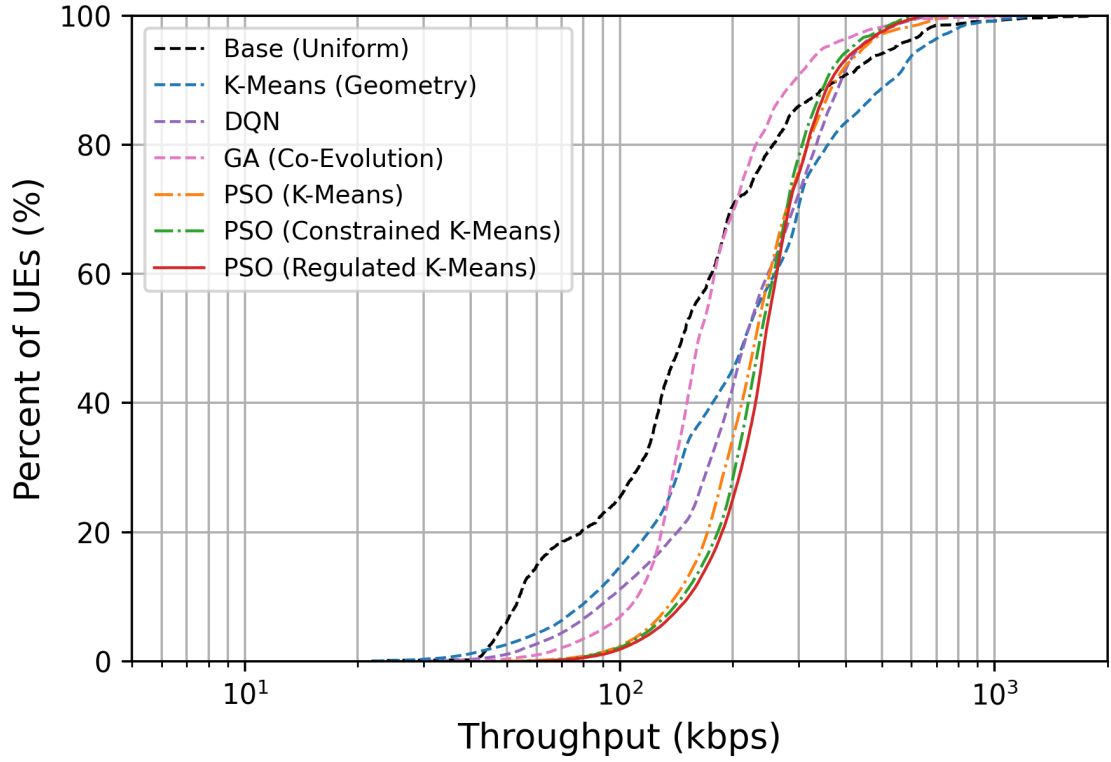


(a)

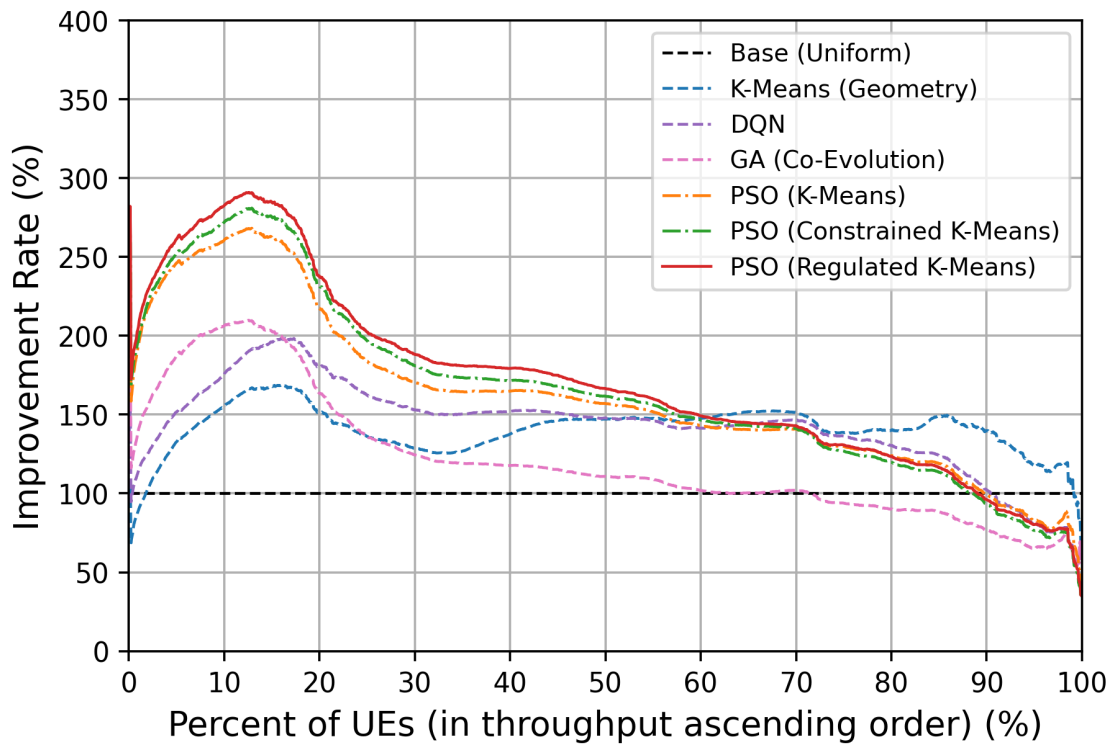


(b)

Figure 4.14. Throughput performance in example area 2 for $p = 4.265$. (a) CDF of UE throughput. (b) Distribution of throughput improvement over the uniform cell configuration from low throughput UEs to high throughput UEs.

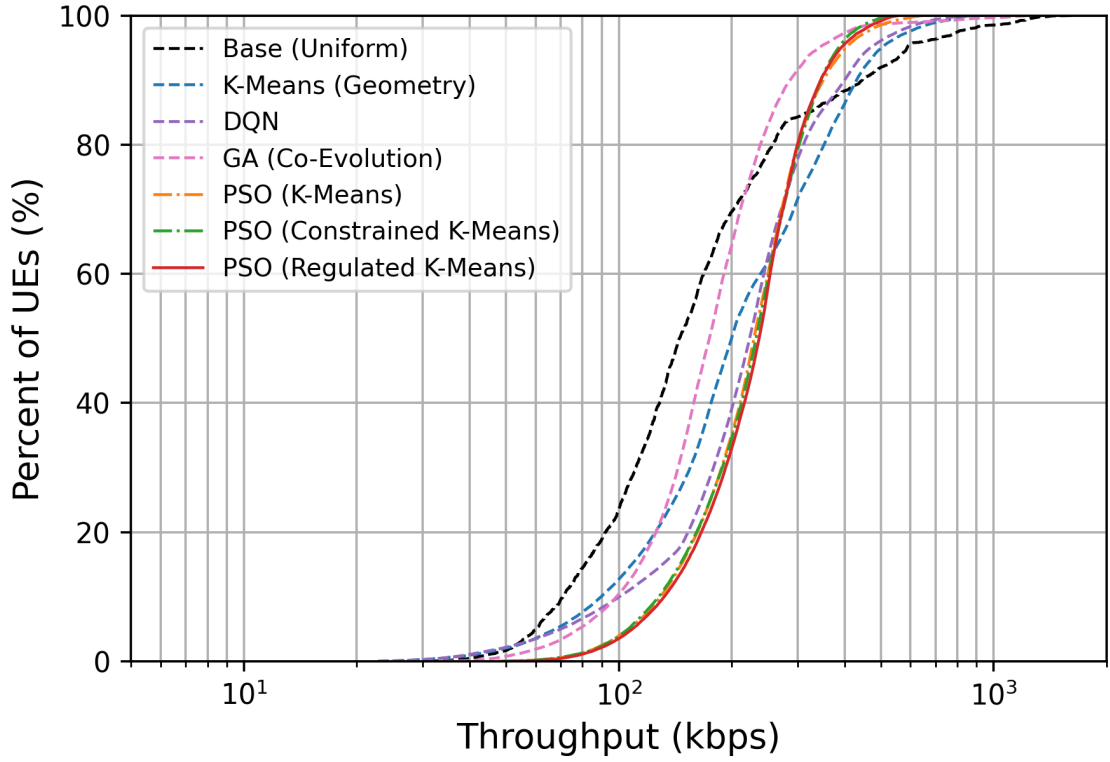


(a)

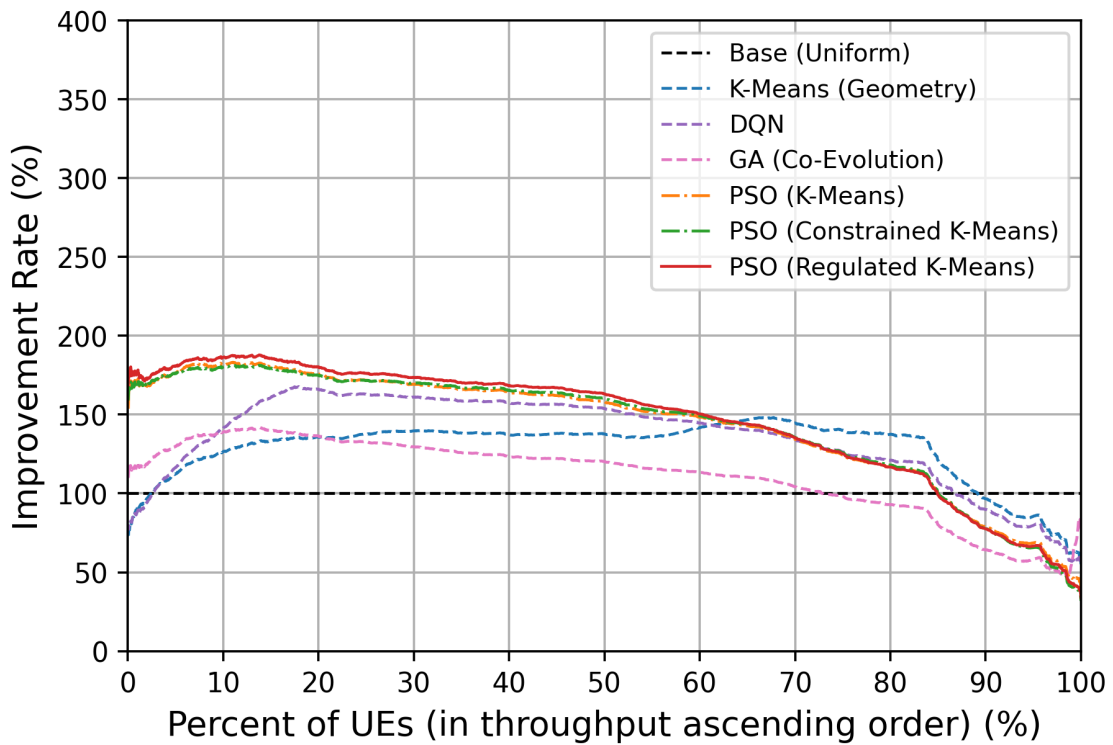


(b)

Figure 4.15. Throughput performance in example area 3 for $p = 4.265$. (a) CDF of UE throughput. (b) Distribution of throughput improvement over the uniform cell configuration from low throughput UEs to high throughput UEs.

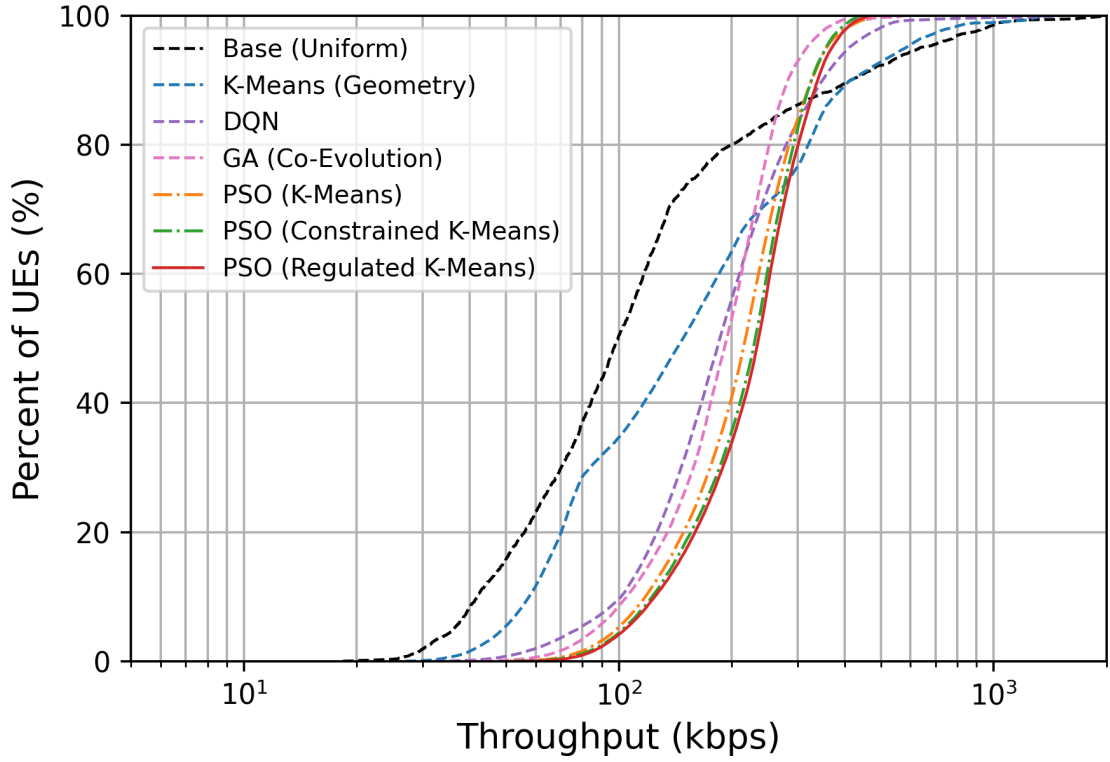


(a)

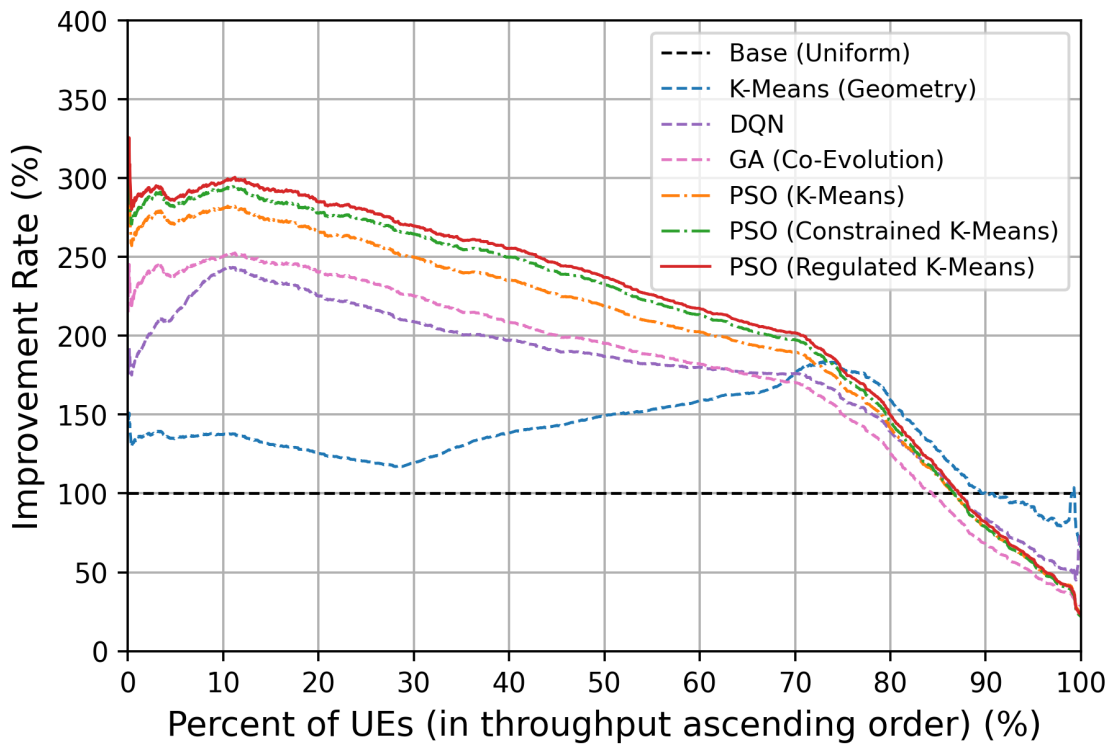


(b)

Figure 4.16. Throughput performance in example area 4 for $p = 4.265$. (a) CDF of UE throughput. (b) Distribution of throughput improvement over the uniform cell configuration from low throughput UEs to high throughput UEs.



(a)



(b)

Figure 4.17. Throughput performance in example area 5 for $p = 4.265$. (a) CDF of UE throughput. (b) Distribution of throughput improvement over the uniform cell configuration from low throughput UEs to high throughput UEs.

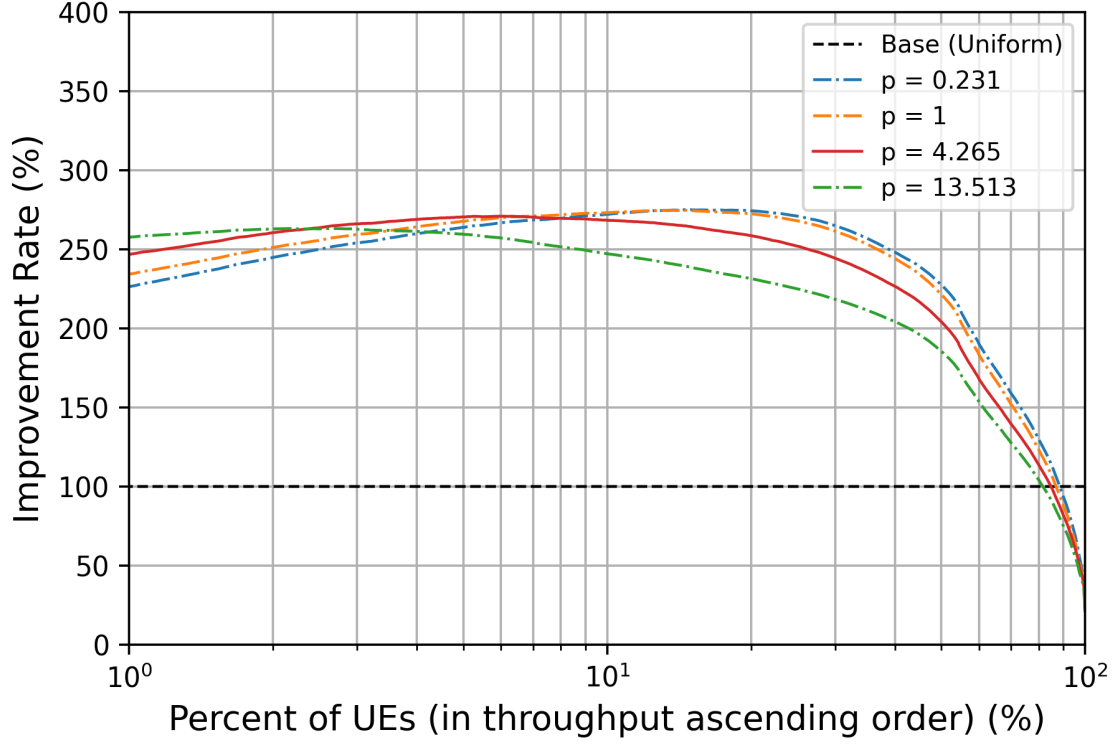
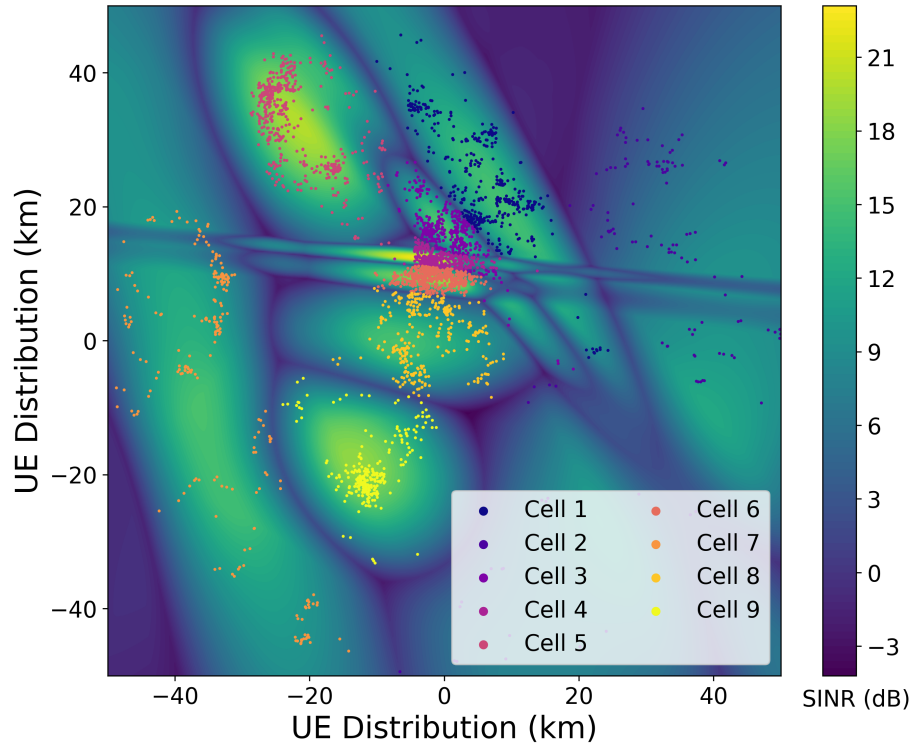


Figure 4.18. Comparison of multiple p values applied to the proposed method in terms of throughput improvement distribution over the uniform cell configuration from low throughput UEs to high throughput UEs, averaged across 100 UE distributions.

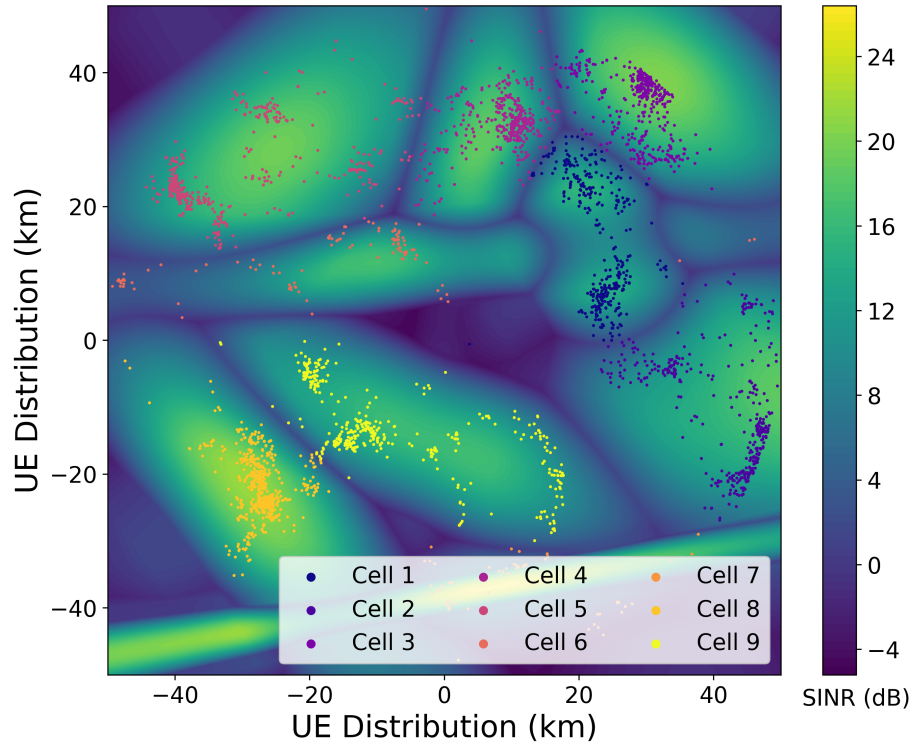
yielding substantial gains for low throughput UEs and smaller or even negative improvements for high throughput UEs. The results with different p values further highlight our goal of the throughput optimization, i.e., improving and distributing UE throughput based on the specific objective function.

This outcome directly reflects the behavior of the weighting functions plotted in Fig. 4.10, providing strong empirical validation that the weighting strategy effectively guides the optimization process according to the specified network objectives. Notably, our proposed method consistently achieves the most favorable improvement profile across the different values of p , demonstrating versatility and efficacy in meeting diverse system requirements.

Beyond aggregate statistics, we further illustrate the spatial impact of optimization by visualizing the SINR distributions and cell boundaries for selected example areas. Specifically, we focus on areas 1 and 2 in Kyushu, which are of particular interest as they correspond to the highest and lowest fitness scores, respectively, among the example

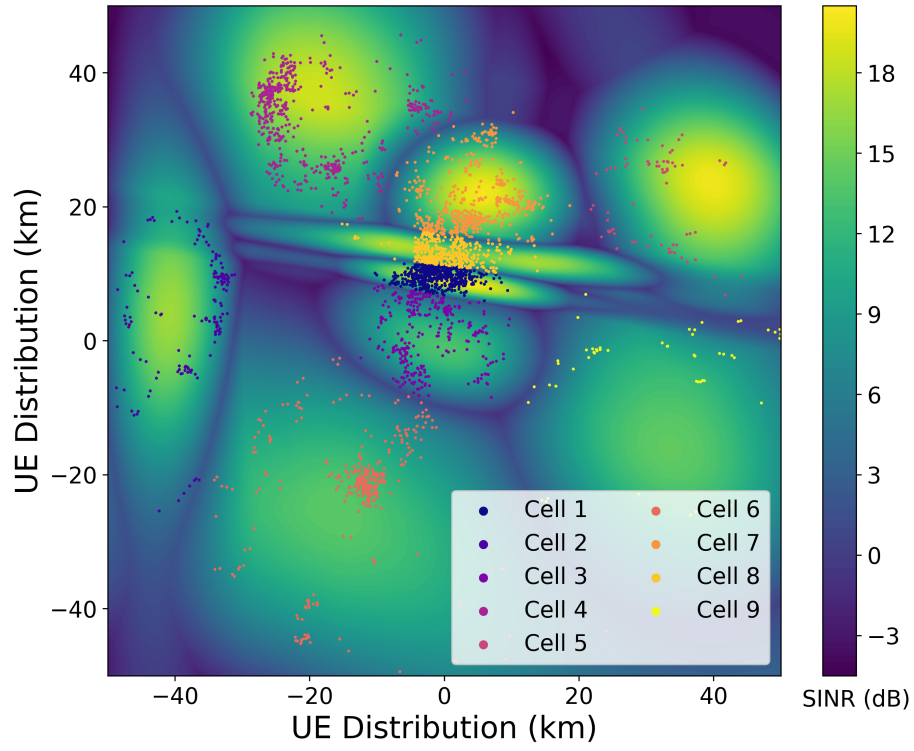


(a)

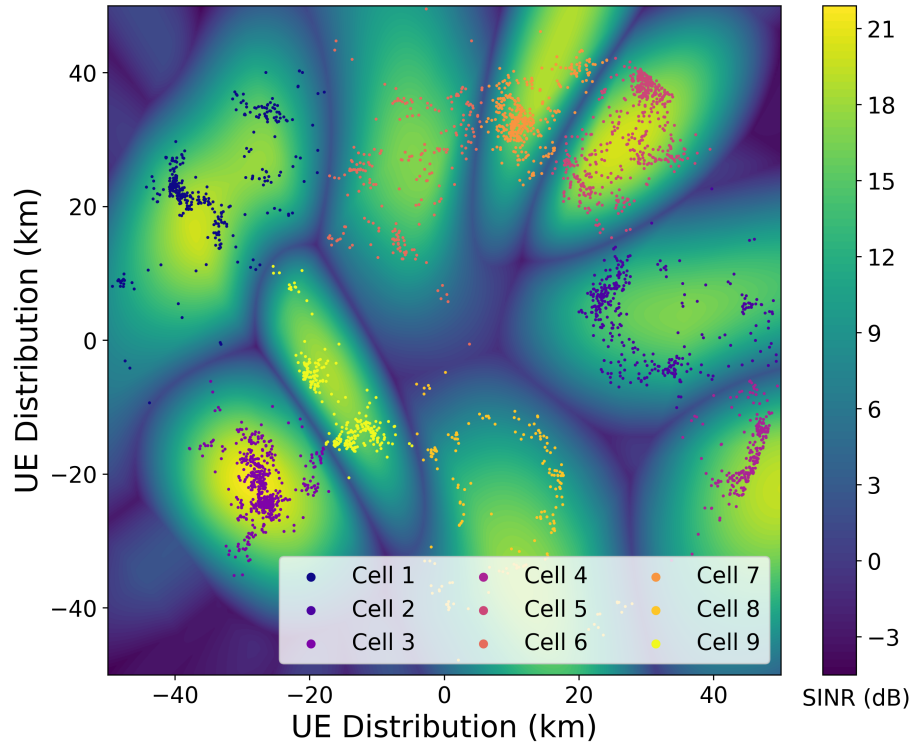


(b)

Figure 4.19. Best SINR distribution achieved among all non-clustering-aided methods for $p = 4.265$. (a) Example area 1. (b) Example area 2.

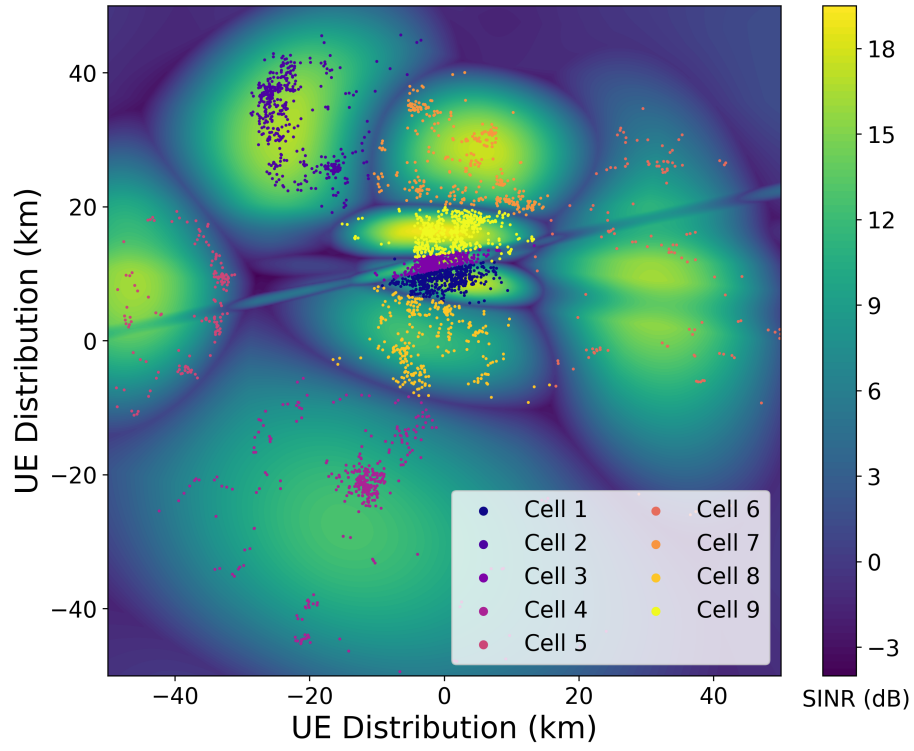


(a)

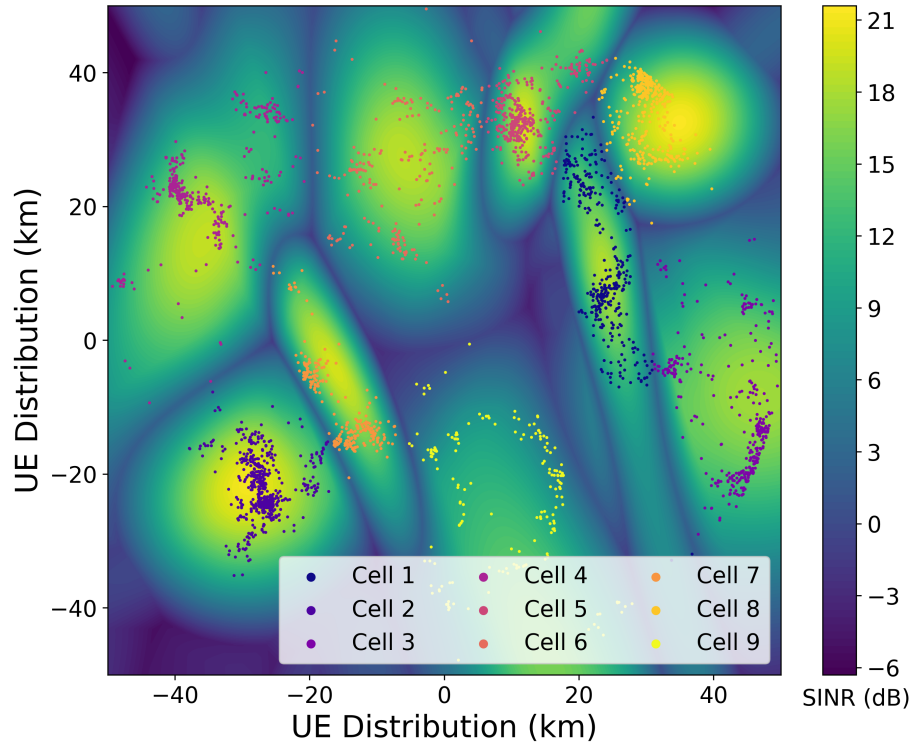


(b)

Figure 4.20. SINR distribution achieved by K-means clustering-aided PSO for $p = 4.265$. (a) Example area 1. (b) Example area 2.

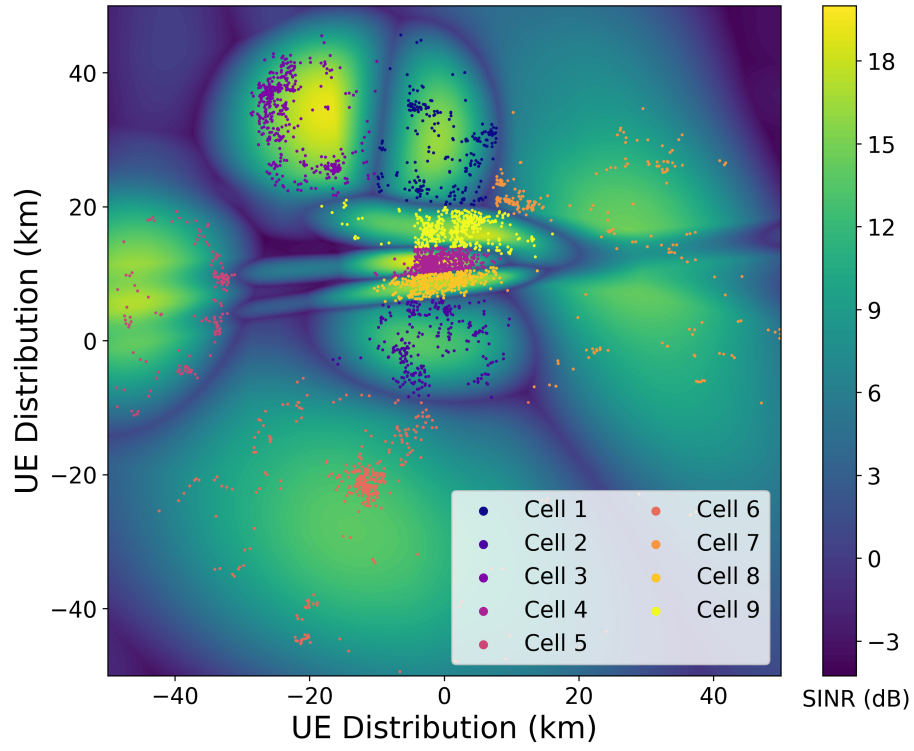


(a)

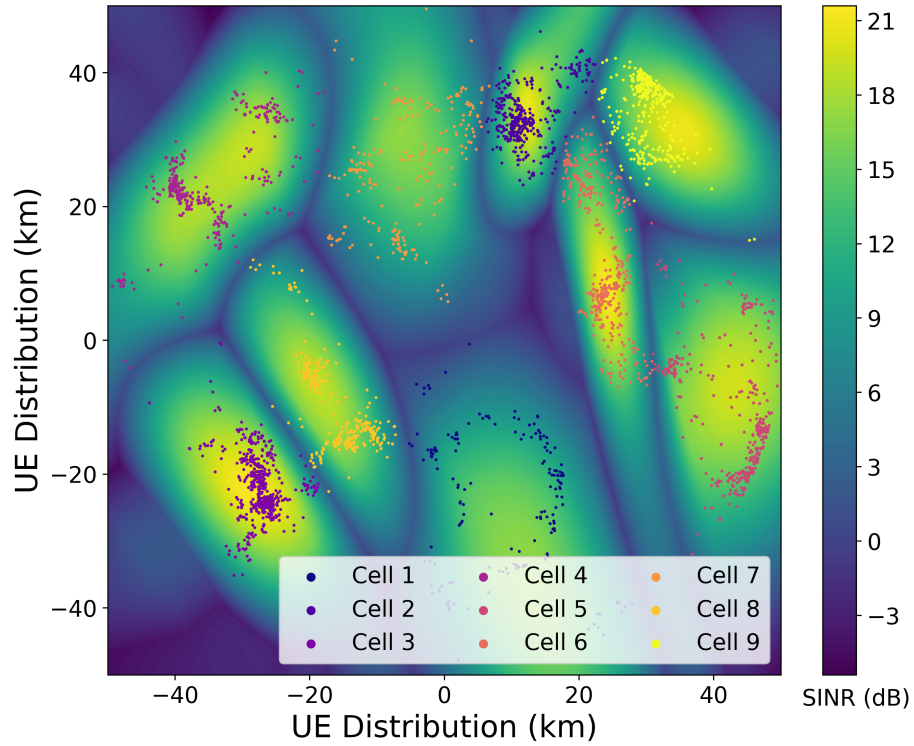


(b)

Figure 4.21. SINR distribution achieved by constrained K-means clustering-aided PSO for $p = 4.265$. (a) Example area 1. (b) Example area 2.



(a)



(b)

Figure 4.22. SINR distribution achieved by regulated K-means clustering-aided PSO for $p = 4.265$. (a) Example area 1. (b) Example area 2.

areas. The SINR is calculated using the optimized antenna parameters that yielded the best fitness scores across all runs. The results for these areas are presented in Figs. 4.19, 4.20, 4.21, and 4.22. In Fig. 4.19, we display the best results achieved among all non-clustering-aided methods, serving as a baseline for comparison. The subsequent figures present the outcomes of clustering-aided optimization, with each panel corresponding to a different clustering approach: conventional K-means, constrained K-means, and the proposed regulated K-means.

Through these visualizations, it is evident that the clustering-aided methods yield cell configurations that are better aligned with the spatial distribution of UEs, in stark contrast to the more rigid and less adaptive layouts observed for the baseline and non-clustering-aided methods. For all three clustering-aided approaches, cell boundaries are more intelligently positioned in regions with lower UE density, consistent with the goal of minimizing inter-cell interference and enhancing overall throughput.

However, key differences emerge in their effectiveness. The results for conventional K-means, presented in Fig. 4.20, show that this method may result in imbalanced allocation of UEs to cells, particularly in densely populated environments, which can limit its ability to achieve fair throughput. The results for constrained K-means, as illustrated in Fig. 4.21, demonstrate that although this method addresses the cluster size imbalance by enforcing strict uniformity, it can restrict flexibility and thus compromise the optimization of inter-cell interference.

In contrast, the proposed regulated K-means method, as shown in Fig. 4.22, strikes a desirable balance by adaptively adjusting cluster sizes according to the actual UE distribution, without rigidly enforcing equal sizes. This flexibility enables more precise and effective search range redefinition, resulting in improved alignment between the cell configuration and network performance objectives.

Taken together, these throughput performance evaluations, both quantitative and qualitative, clearly demonstrate the superiority of our proposed clustering-aided optimization approach. By effectively reducing the prevalence of low throughput UEs, aligning cell

boundaries with the underlying UE spatial distribution, and supporting flexible, objective-driven optimization through the weighting function, our method robustly achieves its intended performance gains across a wide range of practical deployment scenarios. These results provide strong support for the adoption of our approach in real-world HAPS multi-cell configuration problems.

4.6 Conclusion

In this chapter, we presented a comprehensive solution for dynamic multi-cell configuration in HAPS systems by proposing a clustering-aided PSO algorithm that optimizes antenna control based on the actual distribution of UEs within the coverage area. The fundamental concept underpinning our approach is the integration of clustering information derived from the spatial distribution of UEs, which is then utilized to inform the initialization of the optimization algorithm. By carefully redefining the search space for each set of antenna parameters, our method enhances the optimizer’s ability to explore promising regions of the solution domain, effectively reducing the tendency to become trapped in suboptimal local minima and thereby increasing the likelihood of identifying the global optimum. This strategy also addresses the significant computational complexity that naturally arises when simultaneously optimizing multiple beams in a large coverage area.

A key innovation introduced in our work is the regulated K-means clustering algorithm, which serves to overcome the limitations of conventional K-means in the context of HAPS cell optimization. While standard K-means clustering focuses solely on minimizing intra-cluster variance, it can often lead to severely imbalanced cluster sizes that negatively impact network performance by causing an uneven distribution of throughput among UEs. Our regulated K-means approach is specifically designed to address this challenge, adaptively guiding the clustering process to achieve a more balanced allocation of UEs across all clusters. This balanced clustering not only contributes to fairer throughput distribution but also ensures that the optimizer receives a more robust and meaningful initialization, further supporting the search for globally optimal antenna parameters.

The effectiveness of our proposed clustering-aided PSO algorithm has been validated through extensive simulations, leveraging real-world UE distribution datasets from a variety of representative coverage scenarios. The results clearly demonstrate that our approach consistently surpasses existing conventional and state-of-the-art optimization methods. In particular, the proposed method achieves a substantial reduction in the number of low throughput UEs, delivers a more equitable throughput distribution across the user population, and does so with a computational complexity that remains practical for real-time network operations.

Overall, this research marks a significant advancement toward the realization of intelligent and adaptive HAPS antenna control systems. By addressing the dual challenges of search space complexity and cluster size imbalance, our proposed solution provides a robust and effective framework for optimizing network coverage and quality in dynamically changing environments. The contributions of this chapter lay the groundwork for further exploration into learning-based and data-driven approaches, and highlight the importance of spatial clustering in future HAPS network design and optimization.

Chapter 5

A Neural Network Approach for Beamforming Optimization in Multi-Cell HAPS Systems

5.1 Contributions

In this chapter, we introduce a unified deep learning-based framework to address the challenges of real-time and scalable beamforming optimization in multi-cell HAPS systems. Motivated by the limitations of conventional iterative search and supervised MLP-based approaches, our method leverages advanced convolutional neural networks to directly predict the complete set of antenna parameters from raw UE spatial distributions. This approach aims to overcome the computational inefficiency, instability, and limited optimization scope that hinder existing methods, particularly under dynamic and diverse deployment conditions.

Our methodology formulates the multi-cell configuration task as an end-to-end supervised learning problem, where the objective is to infer optimal beamforming parameters for each cell, maximizing throughput performance while ensuring balanced user experience across the coverage area. To realize this, we design a neural network architecture based on ConvNeXt V2, a state-of-the-art vision model known for its scalability and strong feature extraction capabilities. The input to the model is constructed via a robust spatial density encoding of the UE distribution, effectively transforming irregular spatial data into a grid-based representation amenable to convolutional processing.

To enable consistent training and inference across varying HAPS coverage radii, we develop a scale-invariant antenna parameter normalization scheme. This procedure transforms beam directions and beamwidths into unit variables, allowing the model to generalize over different coverage scenarios without loss of geometric information. Furthermore, to address the order ambiguity inherent in ground truth beam assignments, we introduce a permutation-invariant, assignment-based loss function. By incorporating a linear assignment step into the learning process, the network reliably matches its output to the correct set of beam parameters, irrespective of the order present in the dataset.

The practical effectiveness of our approach is demonstrated through extensive experiments on a large-scale dataset of real-world UE distributions collected throughout Japan. Results show that the proposed method achieves throughput performance closely approaching the global optimum found by iterative optimization, while providing fast,

consistent predictions suitable for real-time control. The framework is further validated across a range of model sizes, confirming both its scalability and adaptability to deployment constraints.

The main contributions of this chapter are summarized as follows:

- We propose a unified deep learning framework based on ConvNeXt V2 that directly predicts the full set of cell configuration parameters from spatial UE distributions, enabling end-to-end, real-time beamforming optimization for HAPS systems.
- We introduce a robust spatial density encoding and a scale-invariant antenna parameter normalization scheme, allowing the model to generalize across diverse coverage radii while preserving the geometric relationship between UEs and antenna beams.
- We develop a permutation-invariant, assignment-based loss function that resolves ground truth order ambiguity, ensuring stable and reliable model training with order-independent supervision.
- We validate the proposed approach on a large real-world dataset, demonstrating its superior performance and deployment flexibility relative to conventional optimization and neural network baselines.

5.2 Limitations of Existing Work

Dynamic cell configuration and beamforming optimization in HAPS systems have been extensively investigated using a range of iterative search-based algorithms, including PSO, GA, and reinforcement learning methods such as DQN and fuzzy Q-learning [73–79]. As detailed in Chapter 4, these approaches leverage metaheuristic optimization and online learning processes to adapt antenna parameters in response to varying UE distributions, and our proposed method achieves the best improvement in UE throughput among them.

While these methods demonstrate significant improvements in throughput performance and adaptability, these iterative methods fundamentally rely on repeated search or learning procedures, resulting in non-negligible computational cost and inference variability. In practical scenarios, such limitations restrict their suitability for real-time operation, particularly in large-scale multi-cell HAPS deployments or when rapid response to dynamic UE distributions is required. Furthermore, the outcome of each optimization run may differ, introducing instability to the system's behavior.

In addition to these iterative search-based approaches, a geometry-based configuration method has also been explored for rapid cell assignment in HAPS systems [80]. This method assigns antenna parameters using fixed geometric formulas and clustering, enabling fast computation and simplified implementation. However, as discussed in Chapter 4, it does not optimize antenna parameters according to throughput performance, and therefore cannot achieve high-quality solutions, especially in the presence of complex or highly irregular UE distributions.

Building on these efforts, recent research proposes the use of supervised learning models to further accelerate cell configuration while preserving adaptability. In particular, a MLP-based approach is proposed to predict antenna parameters for HAPS cell configuration directly from the UE distribution [81, 82]. In this method, the UE distribution across the coverage area is encoded as sectorized density information and provided as input to the MLP, which is trained to predict a subset of the parameters necessary for cell configuration. The model leverages ground truth data obtained from iterative search optimizers and incorporates classifier chain and ensemble techniques to account for interdependencies among multiple cells, aiming to enable efficient prediction in multi-cell scenarios.

The primary advantage of this MLP-based approach lies in its computational efficiency. Once trained, the MLP can predict optimized configurations for new UE distributions in real time, substantially reducing the computational overhead compared to algorithms that require repeated iterative calculations for each new scenario. Despite its promise, the MLP-based method exhibits two key limitations. First, the model is only able to predict a partial set of antenna parameters, with the remaining parameters preset

to fixed values. This limitation restricts the optimization scope and prevents the model from adapting fully to varying UE distributions. Second, the use of sectorized density information as input compresses the original UE distribution data, resulting in information loss that limits the model's ability to capture the detailed spatial features essential for complex beamforming optimization.

In summary, while conventional iterative optimization methods and their variants achieve high throughput performance, they remain computationally intensive and unstable for real-time operations. Geometry-based and clustering-assisted approaches, although computationally efficient, fail to exploit the full optimization potential. Supervised learning models such as the MLP-based method provide computational efficiency and initial adaptability, but their architecture and input encoding restrict their capability to perform end-to-end optimization. These challenges motivate the development of a new learning-based method that can jointly optimize the complete set of antenna parameters, make efficient use of spatial distribution information, and deliver robust, real-time performance for HAPS multi-cell beamforming.

5.3 Neural Network Approach

5.3.1 ConvNeXt V2-Based Neural Network Model

In the optimization process defined in this dissertation, the only information available for determining the optimal antenna parameters is the spatial distribution of UEs, stored in the matrix $\mathbf{U}$ as two-dimensional location vectors within the HAPS coverage area. Since the optimal multi-cell configuration is highly dependent on the specific patterns and density of UE distributions, this spatial information can be naturally interpreted as structured data analogous to an image. Such a formulation motivates the application of neural network architectures developed for computer vision, as these models are highly effective at learning spatial patterns and extracting relevant features from dense grid-based representations.

To address the limitations of conventional iterative and MLP-based approaches, we propose a deep learning method that directly predicts the complete set of antenna parameters from the UE distribution. The complete process of the proposed neural network approach for an example 3-cell HAPS system is shown in Fig. 5.1. In our approach, we adopt ConvNeXt V2 as the backbone model [157, 158]. ConvNeXt V2 is a convolutional neural network architecture that modernizes the standard ResNet design by introducing large kernel convolutions, streamlined block structures, and advanced normalization and activation strategies, resulting in improved representation and training efficiency. The use of large kernel convolutions enables ConvNeXt V2 to effectively capture global spatial context and long-range dependencies in the input, fulfilling a similar role to attention mechanisms but implemented purely through convolutional operations.

Importantly, ConvNeXt V2 is known for achieving state-of-the-art top-1 accuracy on the ImageNet-1K dataset and demonstrating superior performance across a wide range of vision tasks. Its design incorporates key advances from both convolutional and transformer-based models, providing strong feature extraction capability and efficient inference. In addition, ConvNeXt V2 is a fully convolutional architecture that offers exceptional scalability not only in input size but also in model capacity, with variants ranging from lightweight, e.g., atto, nano, to large-scale, e.g., large, huge configurations. This flexibility enables practical adjustment of the model’s computational footprint according to available hardware resources and real-time constraints. These properties are considered essential for the practical operation of our beamforming optimization task in HAPS systems, making ConvNeXt V2 a suitable choice. In this work, the ConvNeXt V2 backbone is employed without modification to its hidden layers, preserving the original model architecture.

The process of our method begins by transforming the original UE distribution data into a mapped density matrix through a spatial density encoding step. This step transform the 2D UE locations stored in the matrix $\mathbf{U}$ into a fixed-size array that preserves the spatial density pattern of each distribution. This array serves as the input for the neural network,

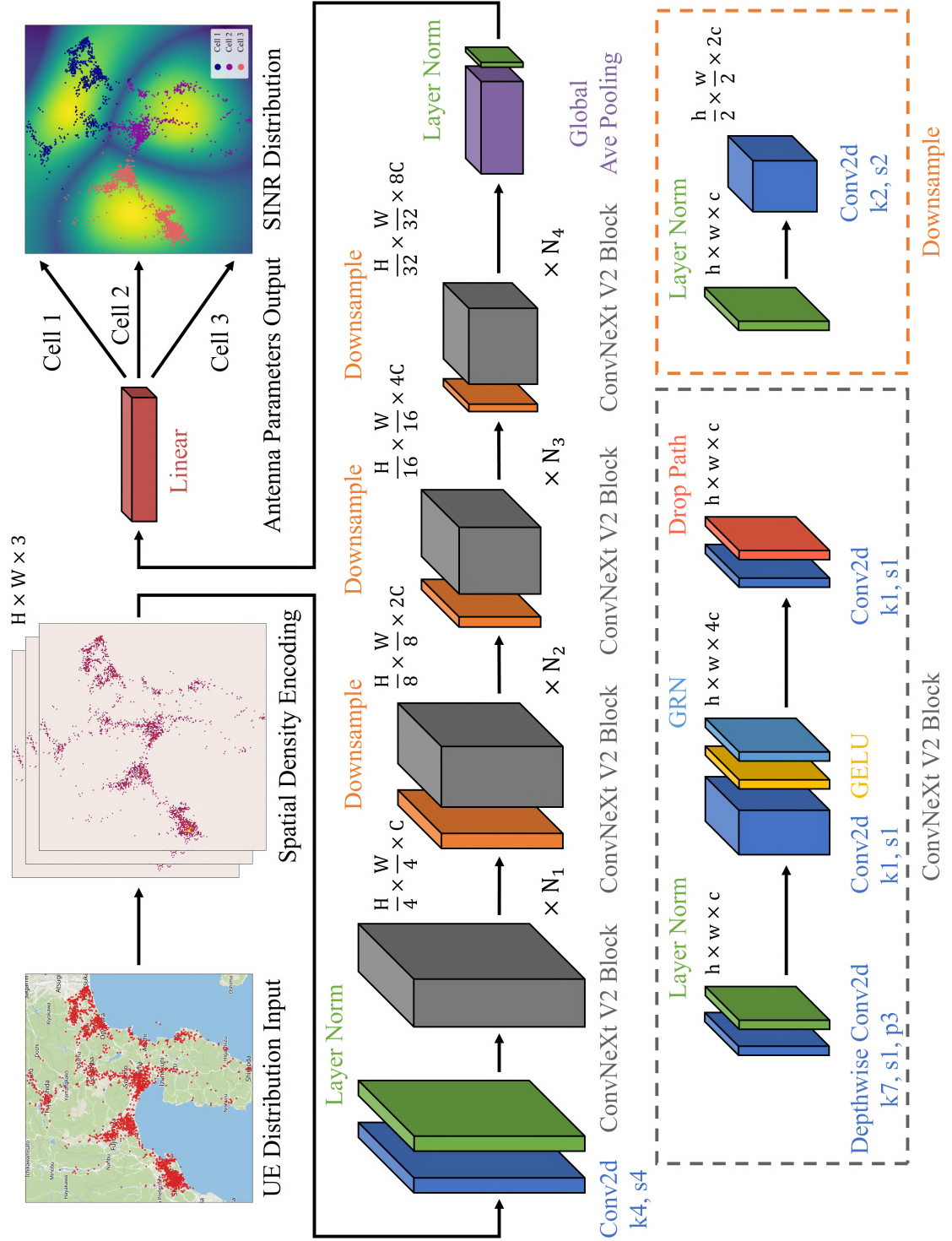


Figure 5.1. Illustration of the proposed neural network model for a 3-cell HAPS system.

enabling it to learn from the spatial distribution of UEs while being compatible with the standard architecture of ConvNeXt V2.

The HAPS coverage area is considered a circle of radius R centered at the origin of the Cartesian coordinate system. In this step, we aim to map this circular region onto a square array of shape $H \times W$, where H and W are the height and width, e.g., 224×224 . The mapping is defined so that the region $[-R, R] \times [-R, R]$ in real space corresponds to the $H \times W$ pixel array. Let $\mathbf{D} \in \mathbb{R}^{H \times W}$ denote the density matrix, where the index of each entry $d_{i,j}$ has the range $i \in \{0, 1, \dots, H-1\}$ and $j \in \{0, 1, \dots, W-1\}$. Let (x_n, y_n) represent the coordinates of the n -th UE relative to the coverage center. For each UE at position (x_n, y_n) , its corresponding index (i_n, j_n) in the mapped density matrix is defined as follows:

$$\begin{aligned} i_n &= (H-1) - \left\lfloor H \frac{(y_n + R)}{2R} \right\rfloor, \\ j_n &= \left\lfloor W \frac{(x_n + R)}{2R} \right\rfloor, \end{aligned} \quad (5.1)$$

where the floor function is used to ensure the consistency. Subsequently, the entry $d_{i,j}$ is computed as the total number of UEs whose mapped pixel indices are equal to (i, j) . The complete formula can be expressed using the indicator function, as follows:

$$d_{i,j} = \sum_{n=1}^N \mathbb{1}(i_n = i \wedge j_n = j). \quad (5.2)$$

To ensure numerical stability and match the input requirements of the neural network, we normalize the density matrix $\mathbf{D}$ for each distribution by its own maximum value. The normalization is expressed as follows:

$$\tilde{d}_{i,j} = \frac{d_{i,j}}{d_{\max}}, \quad (5.3)$$

where $d_{\max}$ represents the maximum entry in the density matrix $\mathbf{D}$. This normalization ensures that all the input values are in the range $[0, 1]$, where the pixel $d_{i,j} = 1$ indicates the highest density region and the pixel $d_{i,j} = 0$ indicates the area with no UE existing. The rationale is that the absolute UE count is less essential for the beamforming optimization

task than the spatial pattern of density of UEs. The optimization relies on recognizing and responding to the relative distribution and clustering of UEs, which this normalization preserves. Thus, the density map extracted through this encoding captures the relevant information while ensuring model robustness and generalization.

Finally, the resulting normalized 2D array $\tilde{\mathbf{D}}$ is replicated across three channels to form a $H \times W \times 3$ tensor. This shape matches the expected input for convolutional neural networks pre-trained on standard image datasets. The encoded spatial density map provides the neural network with an effective representation of UE distribution patterns, allowing the model to learn spatial features relevant to cell configuration. This transformation ensures compatibility with the ConvNeXt V2 architecture and enables robust prediction for various UE distribution scenarios.

The normalized input tensor is then fed into the neural network, which applies a series of convolutional layers to extract hierarchical spatial features, ultimately passing the learned feature representation to a fully connected output layer. At the output stage, a linear activation is used to produce continuous-valued antenna parameters. For a K -cell HAPS system, the output vector is partitioned into K groups, each containing four antenna parameters corresponding to vertical direction angle θ_{dir} , horizontal direction angle ϕ_{dir} , vertical half-power beamwidth $\theta_{3\text{dB}}$, and horizontal half-power beamwidth $\phi_{3\text{dB}}$ for each beam. In order to ensure the model can be applied to different HAPS coverage radii, we design the model to be invariant to the absolute coverage size, enabling unified learning and inference across a range of deployment scenarios. Therefore, this structure enables end-to-end inference from arbitrary spatial UE distribution to the full set of parameters for cell configuration.

A key consideration in the supervised training of this model is the choice of loss function. While the objective function defined in Chapter 4 is well-suited for metaheuristic optimizers, it is not suitable for direct use as a training loss due to its non-differentiability and non-convexity, which impede effective gradient-based optimization. Therefore, we employ the mean squared error (MSE) loss, which is standard for regression tasks and provides smooth gradients for stable and efficient learning. During training, the predicted

antenna parameters are supervised using ground truth solutions obtained from the state-of-the-art optimization method developed in Chapter 4, thereby ensuring the neural network learns to approximate efficient configurations.

5.3.2 Antenna Parameter Normalization

A fundamental challenge in learning-based HAPS beamforming optimization is that the optimal antenna parameters, i.e. beam directions and beamwidths, are intrinsically dependent on the coverage radius, due to the changing geometric relationships between the HAPS, the coverage area, and the UE distribution. This means that, for a given UE spatial pattern, the absolute values of optimal antenna parameters will differ according to the scale of the coverage region. Consequently, directly training a neural network to predict absolute antenna parameters across multiple coverage radii would lead to inconsistent supervision and suboptimal generalization.

To overcome this issue and enable a unified neural network model applicable to arbitrary coverage radii, we introduce a normalization framework for the antenna parameters. Instead of training the network to regress absolute parameter values, we derive a set of conversion formulas that transform the original antenna parameters into unit-normalized variables. These variables encode the spatial configuration of each beam as a proportion of the coverage radius, thereby eliminating direct dependence on the coverage scale. In this way, the network is trained to output these normalized variables, which can be efficiently converted back to the appropriate absolute parameter values for any specified coverage radius by applying the inverse transformation.

This normalization approach is grounded in the observation that the geometric relationships governing the interaction between UE spatial distribution and optimal cell configuration are fundamentally invariant to scale when expressed in unit-relative terms. By projecting beam direction angles and beamwidths onto the 2D UE distribution plane and representing them through trigonometric relationships as normalized distances, we create

a consistent and physically meaningful numerical representation that is suitable for learning and generalizes across deployment scenarios. The detailed theoretical formulation of these conversion steps is presented in this section.

To formalize this normalization procedure, we begin by recalling the set of four optimal antenna parameters for the k -th cell $\theta_{\text{dir}}, \phi_{\text{dir}}, \theta_{3\text{dB}}, \phi_{3\text{dB}}$, the HAPS altitude h , and the coverage radius R . The key idea is to transform these ground truth angular parameters into spatial variables that are expressed relative to the coverage radius. Therefore, we define four alternative ground truth variables $c_{k,x}^u, c_{k,y}^u, d_{k,\theta}^u, d_{k,\phi}^u$. These unit variables are derived by projecting the beam direction angle and beamwidth onto the 2-dimensional coverage plane, replacing the original absolute antenna parameters.

Specifically, let $\mathbf{c}_k = [c_{k,x}, c_{k,y}]$ denote the 2D vector representing the projected beam center location on the UE distribution plane, defined with respect to the HAPS altitude and the beam direction angle. Let $d_{k,\theta}$ represent the projected distance from the beam center to the far edge of the vertical beamwidth, and $d_{k,\phi}$ represent the projected distance from the beam center to the horizontal beamwidth edge. These intermediate values serve to preserve the spatial geometry of the beam structure and form the basis for the unit-variable representation. The transformation from the original antenna parameters to these intermediate geometric variables is given by the formulas as follows:

$$\begin{aligned}
c_{k,x} &= h \tan \theta_{\text{dir},k} \cos \phi_{\text{dir},k}, \\
c_{k,y} &= h \tan \theta_{\text{dir},k} \sin \phi_{\text{dir},k}, \\
d_{k,\theta} &= h \tan \left(\frac{\theta_{3\text{dB},k}}{2} + \theta_{\text{dir},k} \right) - \|\mathbf{c}_k\|, \\
d_{k,\phi} &= \sqrt{\|\mathbf{c}_k\|^2 + h^2} \tan \left(\frac{\phi_{3\text{dB},k}}{2} \right).
\end{aligned} \tag{5.4}$$

To normalize these variables, we divide each of those by the coverage radius R , resulting in the unit-normalized variables used as the following training ground truth:

$$\begin{aligned} c_{k,x}^u &= \frac{c_{k,x}}{R}, \\ c_{k,y}^u &= \frac{c_{k,y}}{R}, \\ d_{k,\theta}^u &= \frac{d_{k,\theta}}{R}, \\ d_{k,\phi}^u &= \frac{d_{k,\phi}}{R}. \end{aligned} \tag{5.5}$$

Therefore, during training, the neural network is supervised to predict the four unit-normalized variables $c_{k,x}^u, c_{k,y}^u, d_{k,\theta}^u, d_{k,\phi}^u$ for each beam.

At inference time, to obtain the absolute antenna parameters for any desired coverage radius R' , the unit variables are first rescaled by R' as follows:

$$\begin{aligned} c'_{k,x} &= R' c_{k,x}^u, \\ c'_{k,y} &= R' c_{k,y}^u, \\ d'_{k,\theta} &= R' d_{k,\theta}^u, \\ d'_{k,\phi} &= R' d_{k,\phi}^u. \end{aligned} \tag{5.6}$$

Finally, the absolute antenna parameters are reconstructed by inverting the geometric relationships, as expressed in the following formulas:

$$\begin{aligned} \theta'_{\text{3dB},k} &= 2 \left(\tan^{-1} \left(\frac{\|\mathbf{c}'_k\| + d'_{k,\theta}}{h} \right) - \tan^{-1} \left(\frac{\|\mathbf{c}'_k\|}{h} \right) \right), \\ \phi'_{\text{3dB},k} &= 2 \tan^{-1} \left(\frac{d'_{k,\phi}}{\sqrt{\|\mathbf{c}'_k\|^2 + h^2}} \right), \\ \theta'_{\text{dir},k} &= \tan^{-1} \left(\frac{\|\mathbf{c}'_k\|}{h} \right), \\ \phi'_{\text{dir},k} &= \tan^{-1} \left(\frac{c'_{k,y}}{c'_{k,x}} \right). \end{aligned} \tag{5.7}$$

This formulation enables the network to learn a scale-invariant, normalized representation of antenna configuration parameters, while ensuring accurate and lossless recovery

of the true beam directions and beamwidths for any specified HAPS deployment. By encoding the essential spatial relationships in this manner, we achieve robust generalization and practical deployment flexibility across arbitrary coverage radii.

5.3.3 Beam Assignment

A key challenge in training neural networks for multi-cell beamforming optimization is the inherent order ambiguity of ground truth data. In our problem, for each UE distribution, the set of antenna parameters, i.e., beam direction angles and beamwidths, corresponding to the K beams is fundamentally order-independent. The HAPS system can apply any permutation of these K parameter sets to the physical beams, as long as the resulting configuration is consistent. However, the neural network output structure imposes a fixed ordering, i.e., each output group is associated with a specific index. If the correspondence between output and ground truth is inconsistent, this creates ambiguity during training, hindering the network’s ability to learn the true relationship between input features and beam configurations.

This order ambiguity can lead to degraded performance, as the network may be penalized for predicting the correct set of parameters in an order different from the arbitrary ground truth order in the dataset. Importantly, there is no robust or general rule to consistently determine a canonical order for the beams in ground truth, since the optimized cell configuration is determined by complex, non-interpretable relationships between the UE distribution and the resulting beam parameters. As a result, enforcing a fixed ordering in the data is both impractical and unreliable.

To address this problem, we introduce an assignment-based loss function into the network training process, enabling the network to learn directly from order-independent ground truth. The central idea is to permit a flexible one-to-one mapping between the predicted beam parameters and the ground truth, selecting the assignment that yields the minimum loss at each training step. Concretely, for each training example, we solve a linear assignment problem between the K predicted parameter groups and the K ground truth groups, finding the pairing that minimizes the total MSE across all beams. This

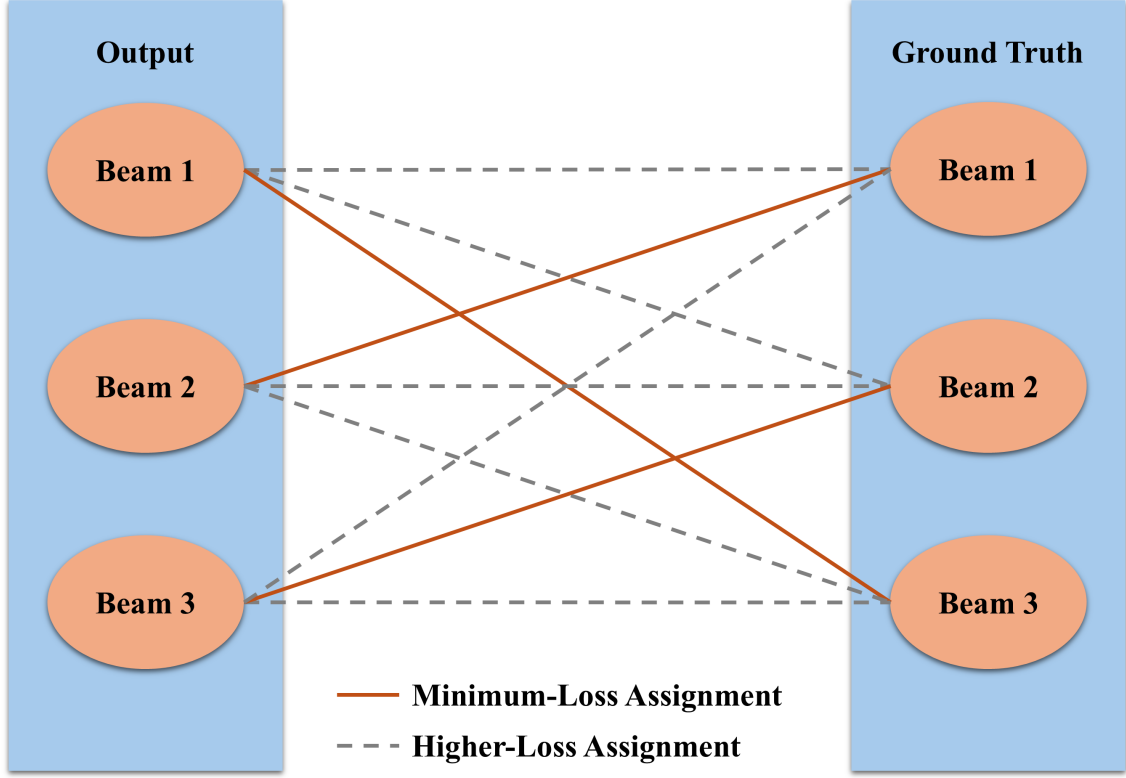


Figure 5.2. Illustration of the beam-to-beam assignment between network outputs and ground truth data.

procedure is visually illustrated in Fig. 5.2, where solid lines denote the minimum-loss assignment and dashed lines denote higher-loss alternatives.

Let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_K]$ denote the predicted antenna parameter groups for K beams, and $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_K]$ the corresponding ground truth groups. We define a $K \times K$ cost matrix $\mathbf{L}$, where each element l_{ij} is the MSE between predicted parameters $\mathbf{y}_i$ and ground truth $\mathbf{t}_j$:

$$l_{ij} = \frac{1}{D} \|\mathbf{y}_i - \mathbf{t}_j\|_2^2, \quad (5.8)$$

where D is the dimensionality of each parameter group. In our case, D equals 4 because we optimize four parameters for each beam. The minimum-loss assignment π^* is found by solving:

$$\pi^* = \underset{\pi \in S_K}{\operatorname{argmin}} \sum_{i=1}^K l_{i, \pi(i)}, \quad (5.9)$$

where S_K is the set of all permutations of $\{1, 2, \dots, K\}$. The final loss for this training instance is then given by:

$$\mathcal{L}_{\text{assign}} = \frac{1}{K} \sum_{i=1}^K l_{i, \pi^*(i)}. \quad (5.10)$$

This is a classical linear assignment problem, for which we employ the Jonker–Volgenant (JV) algorithm, a widely adopted method for fast assignment computation. The JV algorithm solves the assignment problem exactly in polynomial time, ensuring computational efficiency even as the number of beams increases. By minimizing the assignment-based loss $\mathcal{L}_{\text{assign}}$, the network is trained to output beam parameter sets that match the ground truth in a permutation-invariant manner, overcoming the order ambiguity and ensuring robust learning.

This approach eliminates the need for any rule-based sorting of beams and guarantees that, for every input UE distribution, the learned mapping is both correct and agnostic to the original ordering of the ground truth. As a result, the neural network achieves improved performance and reliability in predicting cell configurations for dynamic and complex UE distributions.

5.4 Evaluation

5.4.1 Experiment Specifications

In the evaluation of our proposed neural network-based beamforming approach, we employ a comprehensive dataset of real-world UE distributions, consistent with the data source used for the first method. Specifically, the UE location data are collected from AGOOP Corp. [154]. To facilitate supervised learning and rigorous performance assessment, we construct a large-scale dataset comprising 40,000 distinct UE distribution samples. This dataset is generated by first selecting 20,000 unique coverage area centers across the entire region of Japan. For each center, we sample UE locations from the real-world data under two different coverage radii: 30 km and 50 km. The UE distribution

in each coverage area is caused by the influence of distinct geographical features and human activities. As a result, our dataset contains 20,000 UE distributions for each coverage radius, providing a balanced and diverse representation of practical deployment scenarios.

The dataset is divided into training and testing subsets according to a standard 8:2 split, yielding 32,000 distributions for training and 8,000 for testing. Since each UE distribution sample is distinct in terms of both the coverage center and distribution characteristics, this ensures that the data in the test set remains diverse and unseen, thereby enabling a thorough evaluation of the model’s generalization capability. During evaluation, we report the average optimization performance on the test subset, which consists of 4,000 UE distributions for the 30 km radius and 4,000 for the 50 km radius, respectively. To provide detailed insights, we further present results for several representative example coverage areas. Two representative example areas are selected for each coverage radius and illustrated in Figs. 5.3 and 5.4, enabling visualization of the optimization results across different spatial scenarios and coverage sizes. As in the first method, the underlying maps are generated using OpenMapTiles and OpenStreetMap data [155].

In this evaluation, we focus on a 3-cell HAPS system, for which ground truth antenna parameters are generated using the PSO-based optimization approach presented in Chapter 4. For each UE distribution in the dataset, we uniformly sample 3,000 UEs to ensure statistical consistency with the simulation environment adopted in prior experiments. Recognizing that the PSO algorithm exhibits inherent performance variability across runs, we perform 30 independent optimization runs for each distribution and select the best-performing result, i.e., the parameter set yielding the lowest fitness score, as the ground truth solution. This strategy ensures that the training supervision is based on the most effective configurations discovered by the state-of-the-art optimization method.

The fitness score for each distribution is evaluated using the same objective function defined in Chapter 4, maintaining consistency with earlier chapters and ensuring a fair benchmark for both training and evaluation. Details of the training settings are summarized in Table 5.1. Specifically, the neural network is trained for 200 epochs using a

Table 5.1. Training settings.

Configuration	Value
Optimizer	AdamW
Base learning rate	2.5e-4
Weight decay	0.01
Optimizer momentum β_1, β_2	0.9, 0.999
Stability constant ϵ	1e-8
Learning rate schedule	One-cycle policy
Batch size	32
Epochs	200

batch size of 32, employing the AdamW optimizer with a one-cycle learning rate schedule [159, 160]. This combination is selected to facilitate stable convergence and robust generalization across the diverse UE distribution scenarios present in the dataset.

To comprehensively evaluate the effectiveness of our proposed neural network approach, we compare its performance with several baseline methods and model variants. As the primary benchmark, we consider the PSO-based optimization method from Chapter 4 under three scenarios: the best-achieved performance, i.e., the ground truth as previously described, the average performance across 30 independent optimization runs for each UE distribution, and the worst-case performance observed among these runs. This comprehensive assessment characterizes both the optimality and variability of iterative optimization.

In addition, we include the MLP-based neural network method proposed in [81], as discussed in the related work section, to provide direct comparison with the most relevant existing learning-based approach. To further analyze the contribution of our model design, we evaluate three variants of our ConvNeXt V2-based model. The first variant is the ConvNeXt V2 Large model without the proposed beam assignment mechanism. The second variant is a lightweight ConvNeXt V2 Atto model incorporating the beam assignment mechanism. The third variant is the ConvNeXt V2 Large model with beam assignment

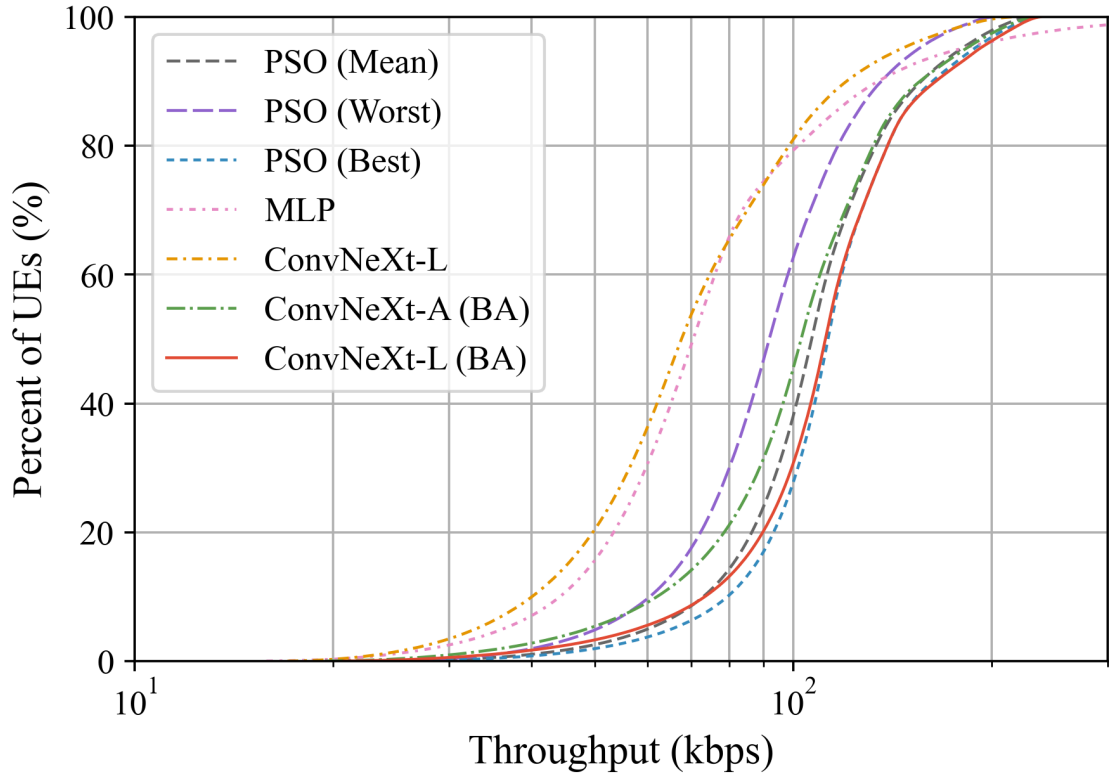
enabled. The computational complexities of the Large and Atto models are 34.4 GFLOPs and 0.55 GFLOPs, respectively. Since FLOPs are directly related to the computational resources required for inference, the evaluation of these variants assists in assessing model performance under deployment settings with varying computational constraints.

By including both optimization-based and learning-based baselines, as well as ablation variants of our model, we systematically examine the impact of model capacity and the beam assignment mechanism on performance. This comprehensive comparison enables a rigorous assessment of the advantages of our approach relative to established and state-of-the-art alternatives.

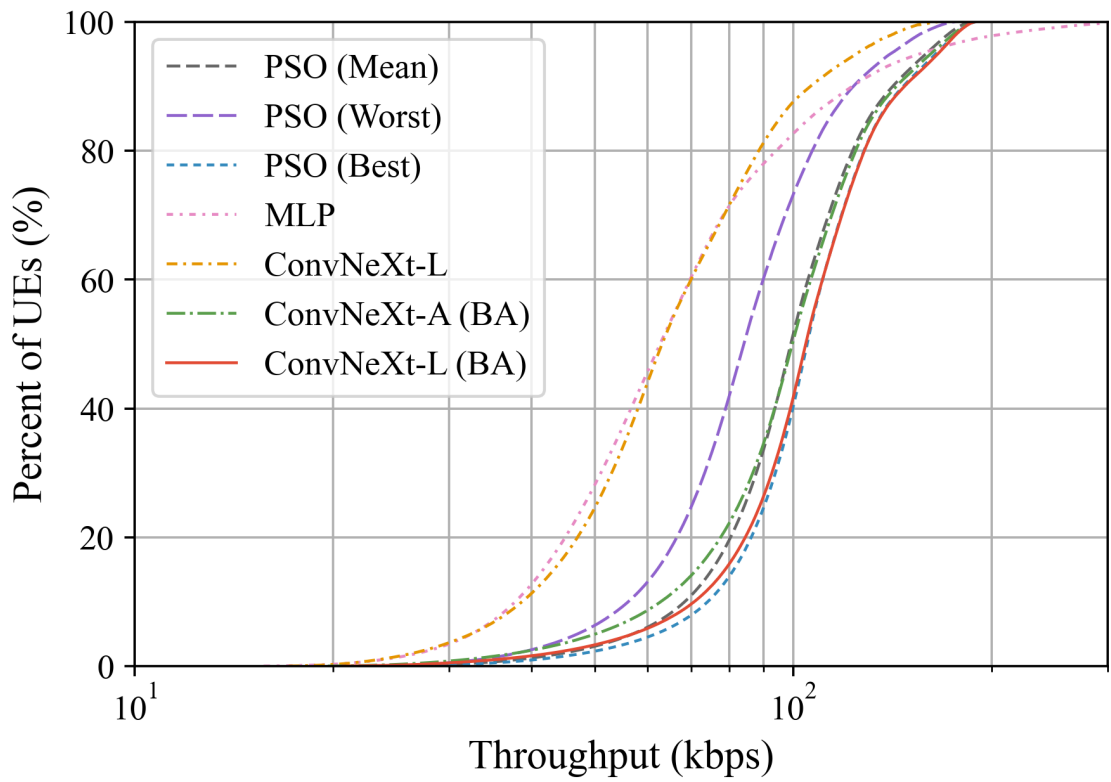
5.4.2 Optimization Performance

For a comprehensive comparison, we present the average CDF of UE throughput for the 4,000 UE distributions in the 30 km test set and the 4,000 distributions in the 50 km test set, as shown in Fig. 5.5(a) and Fig. 5.5(b), respectively. Through the CDF results, we can fairly compare the performance of different methods across all UEs within the coverage area, considering the entire range from lowest throughput to highest throughput levels. Overall, the CDF curves indicate that the proposed method approximates the throughput distribution achieved by the PSO-based ground truth, with the ConvNeXt V2 Large model most closely approaching the ground truth. Furthermore, the CDF results show that the relative performance among the compared methods remains consistent between the 30 km and 50 km cases.

The MLP-based method demonstrates significantly degraded performance, falling short even when compared to the worst-case PSO-based results. This outcome confirms that the MLP-based method fails to capture the complexity of the task, as it lacks the capacity to optimize the full set of antenna parameters. The ConvNeXt V2 Large model without beam assignment also shows poor throughput performance, demonstrating that, even with a state-of-the-art architecture, meaningful learning does not occur without beam assignment due to confusion from the order-independent ground truth data.

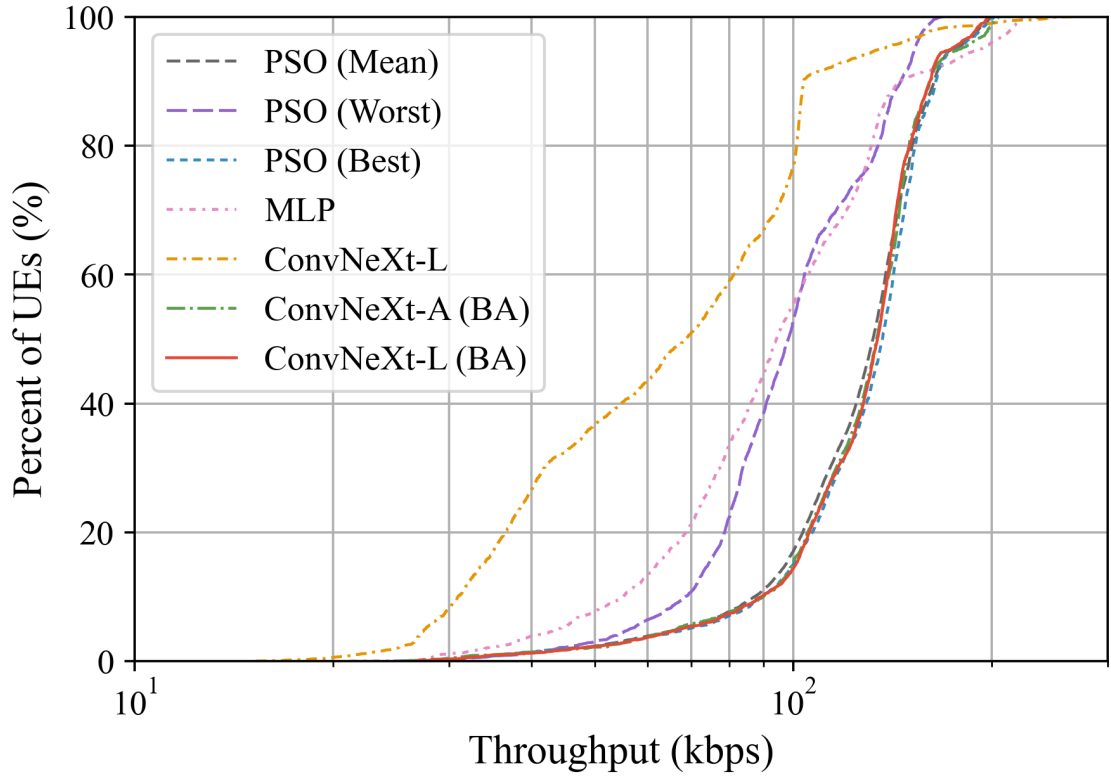


(a)

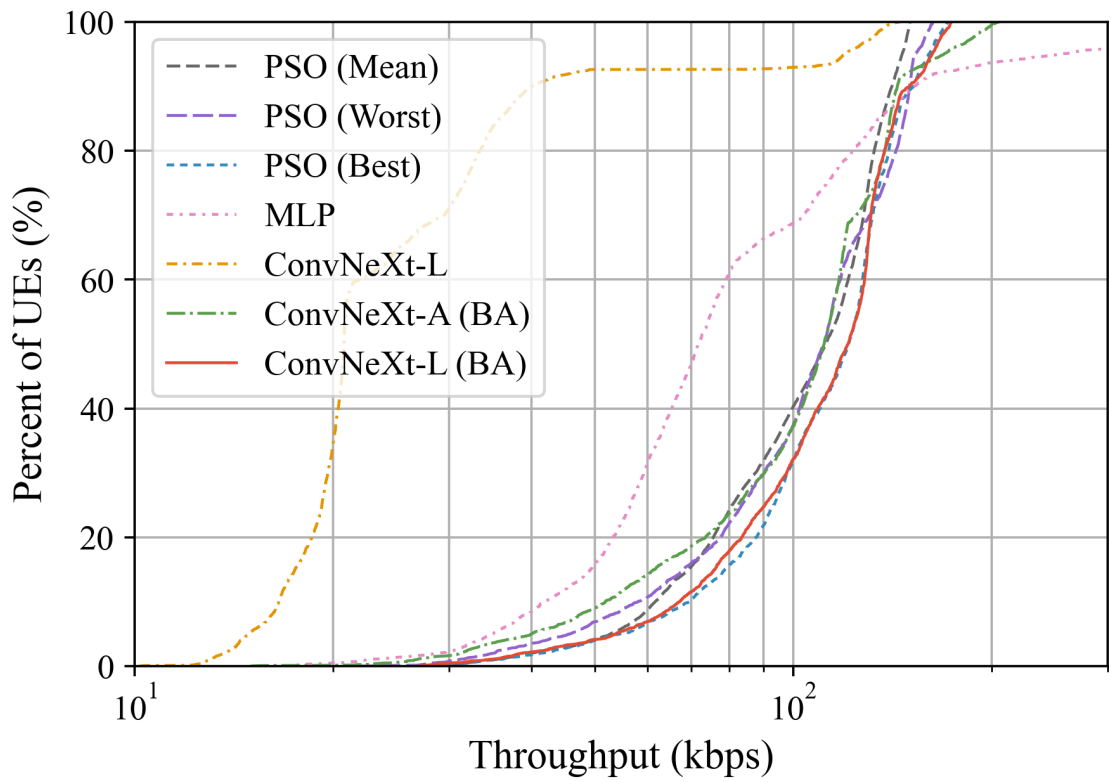


(b)

Figure 5.5. Average CDF of UE throughput. (a) 30 km coverage radius. (b) 50 km coverage radius.

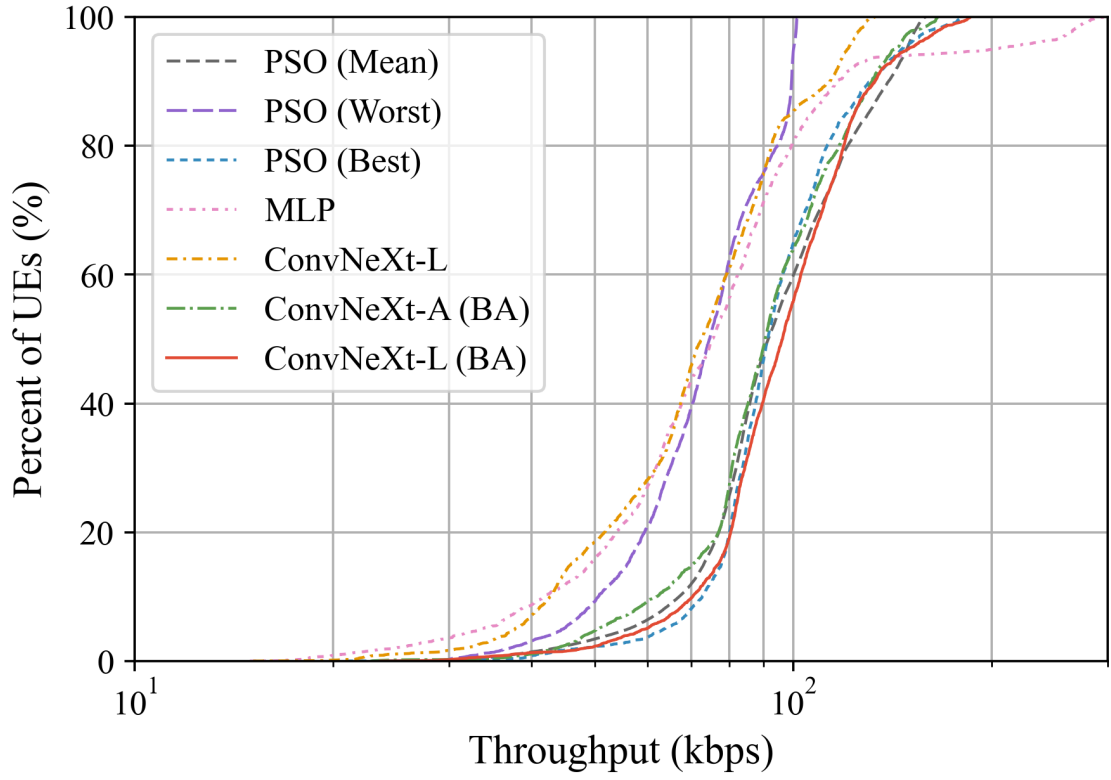


(a)

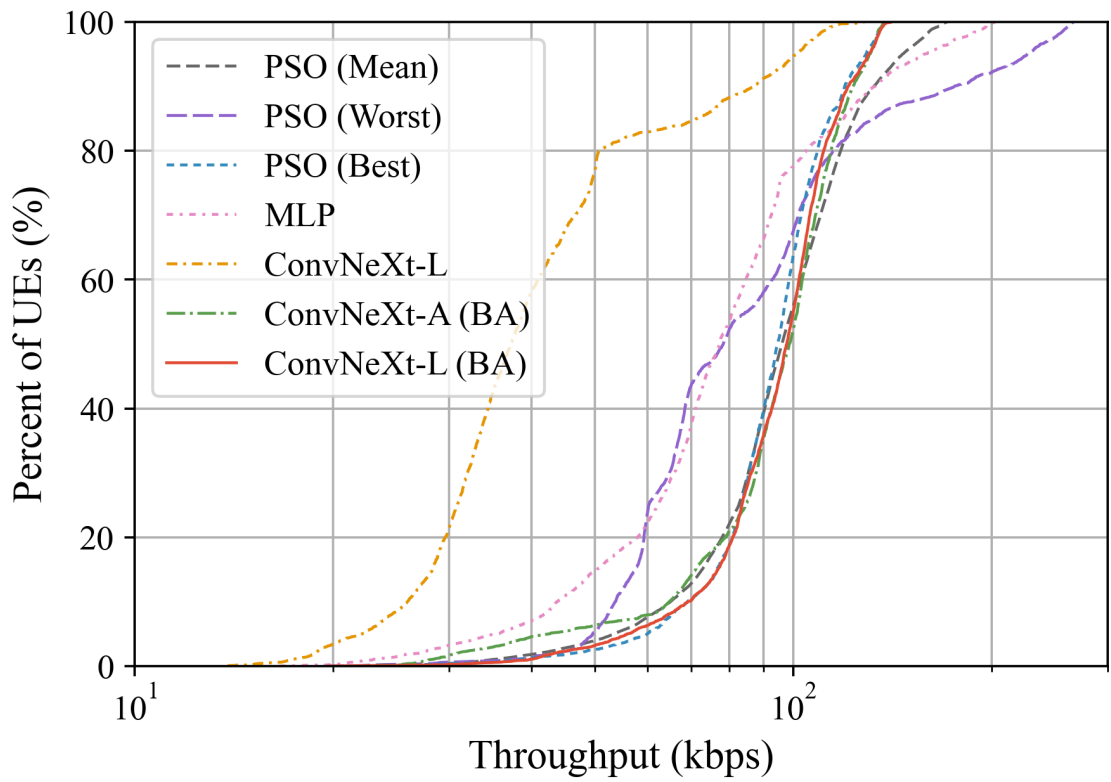


(b)

Figure 5.6. CDF of UE throughput for 30 km coverage radius. (a) Example area 1. (b) Example area 2.

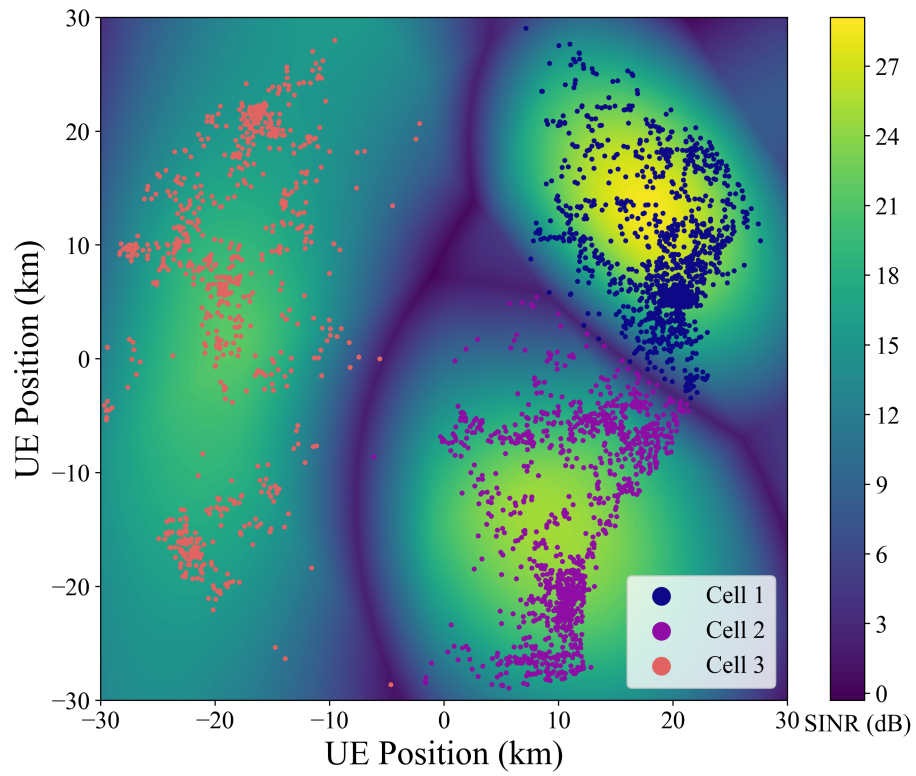


(a)

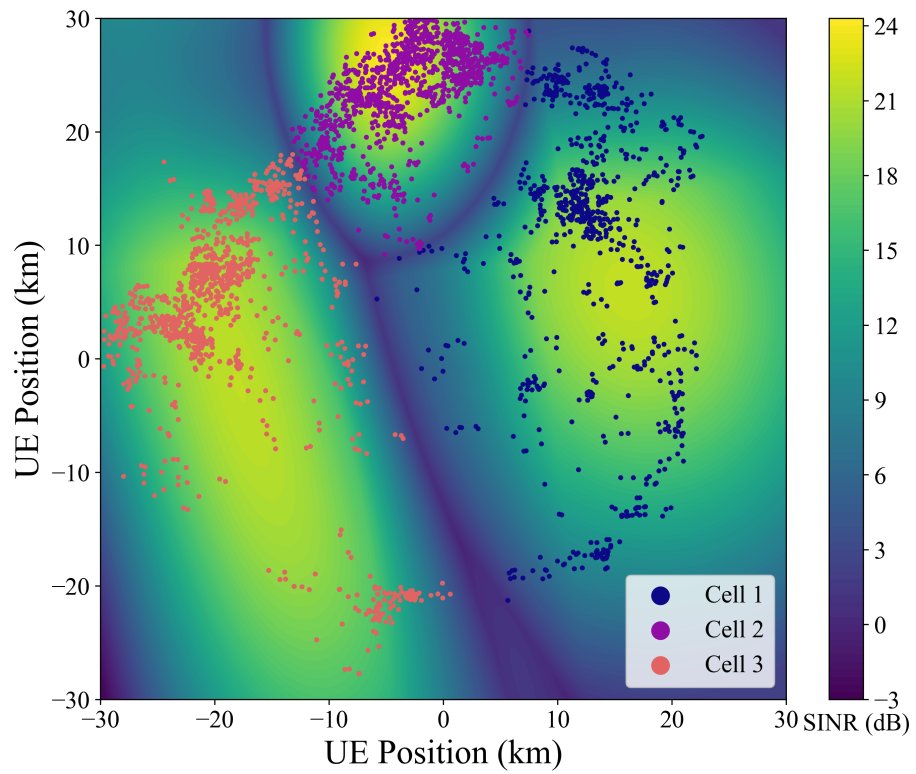


(b)

Figure 5.7. CDF of UE throughput for 50 km coverage radius. (a) Example area 1. (b) Example area 2.

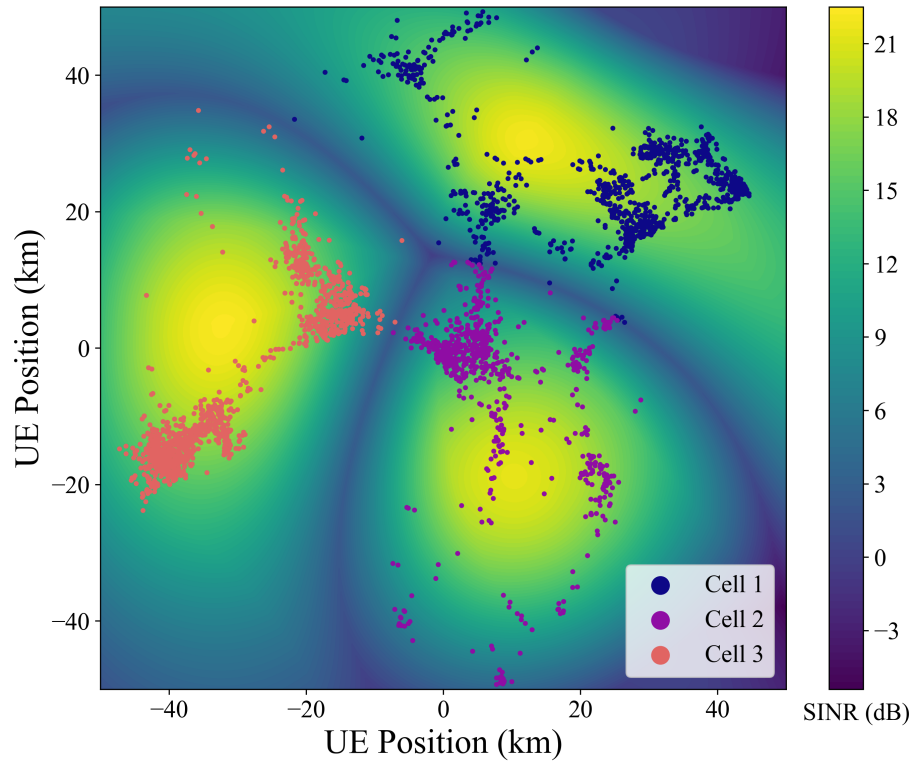


(a)

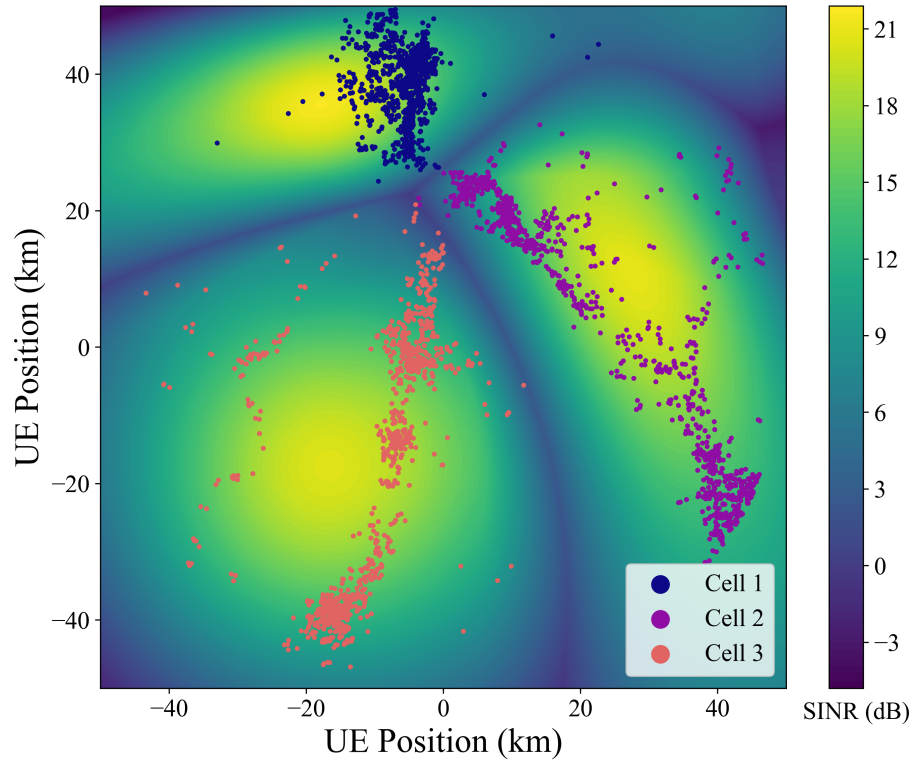


(b)

Figure 5.8. SINR distribution in the two example areas for 30 km coverage radius. (a) Example area 1. (b) Example area 2.



(a)



(b)

Figure 5.9. SINR distribution in the two example areas for 50 km coverage radius. (a) Example area 1. (b) Example area 2.

The ConvNeXt V2 Atto model with beam assignment outperforms these above two methods but is still worse than the average PSO-based performance. Specifically, the performance gap narrows as throughput increases, but widens in the lower throughput range. However, the Atto model outperforms the worst-case PSO results, confirming that even a lightweight version of the proposed method can deliver acceptable performance.

In contrast, the ConvNeXt V2 Large model with beam assignment surpasses the PSO average and closely matches the ground truth, particularly across most of the throughput range. A slight gap remains in the low-throughput tail, which is attributed to the sensitivity of throughput performance to minor deviations in the predicted antenna parameters. Even with well-trained models, exact numerical prediction is inherently limited by the regression formulation and the use of a linear output layer.

In addition, we explicitly plot the CDF curves for the example areas in both the 30 km and 50 km coverage scenarios, as shown in Figs. 5.6 and 5.7, respectively. Although the CDF curves show different patterns across example areas, the relative performance remains consistent with the average results, demonstrating the robust overall performance of our proposed method. For these example areas, we also plot the corresponding SINR distributions obtained by applying the beamforming results from the ConvNeXt V2 Large model with beam assignment, as shown in Figs. 5.8 and 5.9. These visualizations provide clear insight into cell configuration and further highlight the effectiveness of our cell optimization approach. In each example area, it is observed that cells are properly configured to match the UE distributions: cell boundaries are positioned in areas of relatively low UE density to reduce inter-cell interference, while UE numbers are balanced across cells to improve throughput fairness.

To further highlight the performance across diverse UE distribution scenarios, we explicitly present both the CDF of throughput and the SINR distribution for four additional coverage areas with radii of 30 km and 50 km, respectively. The UE distributions with SINR are shown in Figs. 5.10 and 5.11. The CDF results are shown in Figs. 5.12 and 5.13.

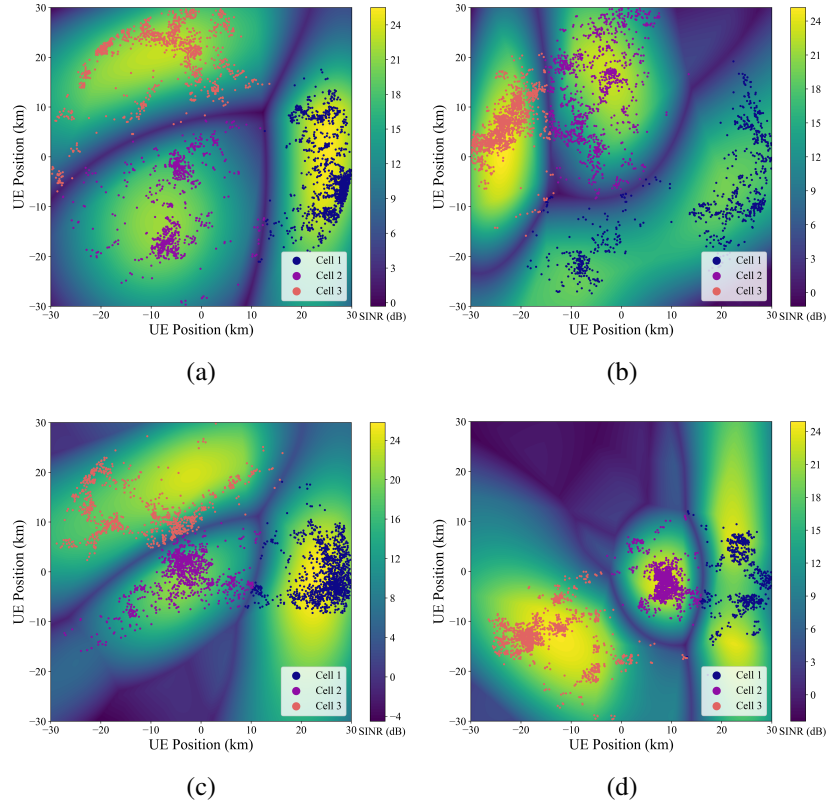


Figure 5.10. SINR distribution in additional example areas for 30 km coverage radius.

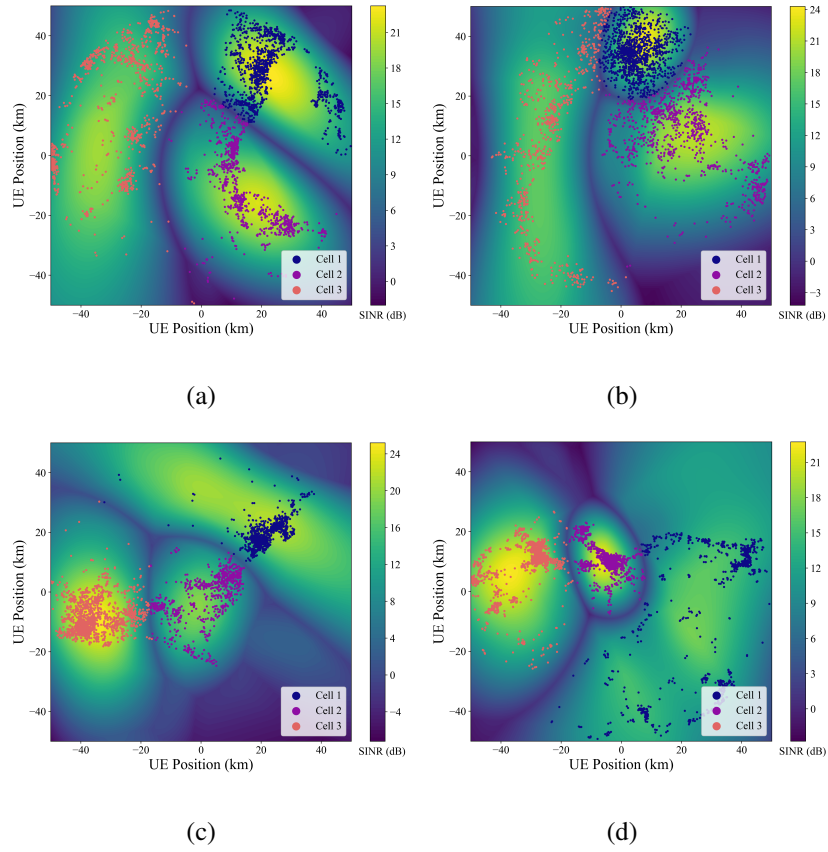


Figure 5.11. SINR distribution in additional example areas for 50 km coverage radius.

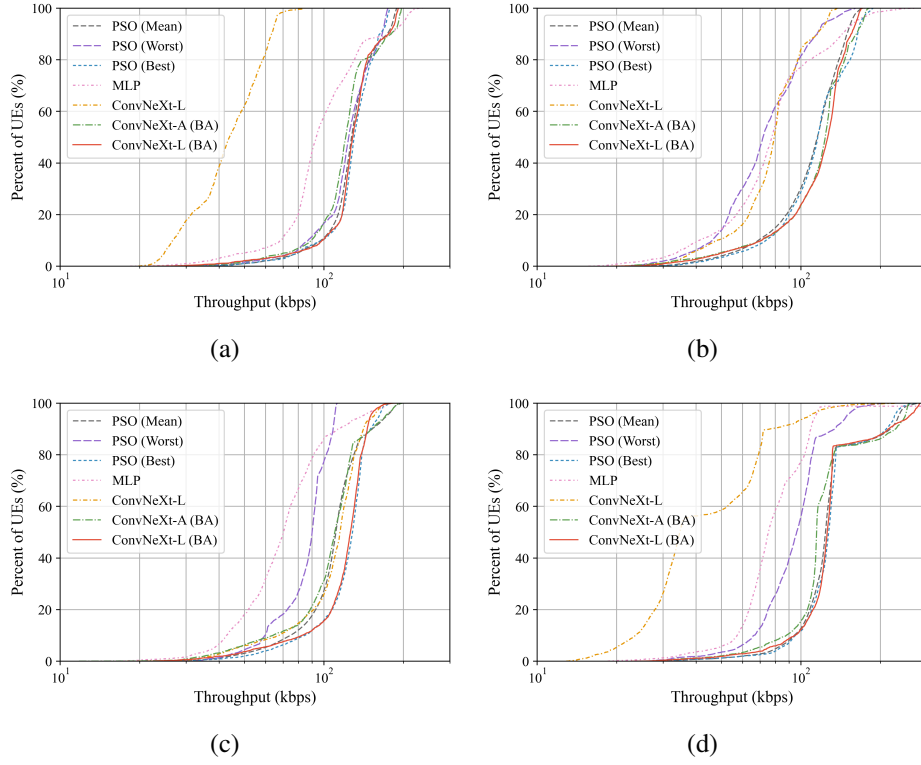


Figure 5.12. CDF of UE throughput in additional example areas for 30 km coverage radius.

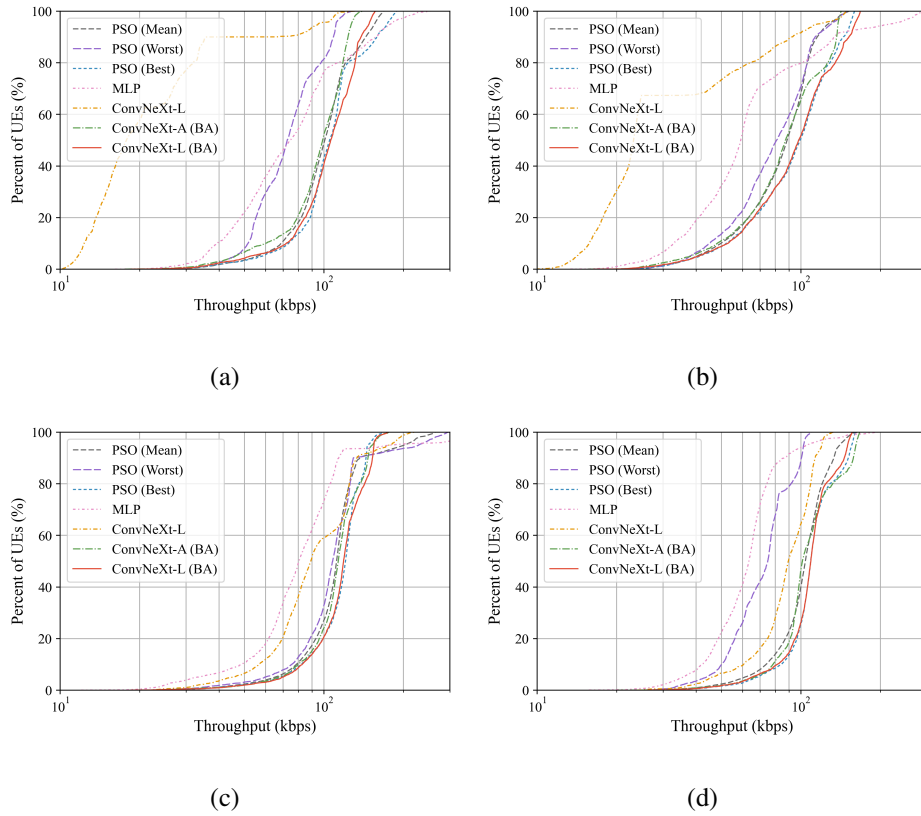


Figure 5.13. CDF of UE throughput in additional example areas for 50 km coverage radius.

Overall, the results demonstrate that our proposed method effectively addresses both the challenge of real-time optimization and the instability inherent in the PSO-based method. Unlike PSO, which suffers from run-to-run variability, the neural network model provides consistent predictions, with performance reliably exceeding the PSO average. Furthermore, the scalability of the ConvNeXt V2 architecture allows deployment flexibility: higher throughput performance can be achieved when HAPS system resources permit the use of larger models, while lighter models can still deliver robust results under resource-constrained conditions.

5.5 Conclusion

This chapter has presented a novel deep learning-based method for multi-cell beamforming optimization in HAPS systems. Building on the strengths of advanced convolutional architectures, we have established a unified framework that directly predicts the full set of antenna parameters from raw UE spatial distributions, efficiently addressing both real-time operation and deployment across wide-area coverage regions with diverse and dynamic UE distributions.

The proposed approach leverages a robust spatial density encoding to represent UE distributions, applies a scale-invariant normalization for antenna parameters to enable generalization across different coverage radii, and incorporates a permutation-invariant assignment-based loss function to address the order ambiguity of ground truth data.

Extensive experiments on real-world datasets demonstrate that the proposed approach achieves throughput performance comparable to state-of-the-art iterative optimization, while substantially reducing computational cost, instability, and inference time associated with online search-based methods. The results further confirm that the model architecture is flexible and scalable, adapting to various hardware and deployment constraints without compromising performance.

Overall, this work advances the frontier of learning-based beamforming optimization for HAPS and provides a practical solution for robust and efficient multi-cell coverage optimization across a wide range of deployment scenarios and user distribution patterns.

Chapter 6

Conclusions and Future Work

6.1 Contributions

This thesis has made substantial advances in the field of dynamic multi-cell beamforming optimization for HAPS systems, addressing both algorithmic challenges and practical requirements for intelligent network deployment. Two complementary frameworks, clustering-aided optimization and neural network approach, were proposed, thoroughly analyzed, and rigorously validated on large-scale, real-world datasets in Chapters 4 and 5. The clustering-aided optimization method reduces the computation time required to achieve a good optimization quality compared to other iterative search methods, while the neural network approach guarantees fast inference powered by GPU. These contributions solidify the foundations for the effective operation of HAPS systems in real-world environments.

At the algorithmic level, the first major contribution, presented in Chapter 4, is the design of a clustering-aided PSO optimization framework that incorporates a novel regulated K-means clustering method. This approach effectively overcomes the limitations of conventional clustering by ensuring balanced cluster sizes, enabling more precise search range definition for antenna parameters, and improving the global optimality and fairness of multi-cell configurations in HAPS deployments. Through extensive simulations, this method demonstrated consistent superiority over existing iterative search algorithms, particularly in handling highly heterogeneous and complex user distributions.

The second core contribution, detailed in Chapter 5, is the development of a deep learning-based solution leveraging the ConvNeXt V2 architecture. By introducing a robust spatial density encoding for UE distributions, a scale-invariant antenna parameter normalization technique, and a permutation-invariant assignment-based loss, this framework enables end-to-end prediction of all antenna parameters directly from raw spatial data. The neural model achieves throughput performance comparable to the best iterative methods while drastically reducing inference time and eliminating run-to-run variability, which is crucial for real-time and large-scale deployment scenarios. The adaptability and efficiency of this approach set a new standard for practical, intelligent control in HAPS systems.

Beyond the algorithmic innovations, this thesis advances the integration of HAPS as a pivotal platform within future wireless network architectures, with particular emphasis on enabling efficient, scalable, and flexible support for massive IoT connectivity. By bridging robust optimization methods with scalable learning-based inference, the work lays a technical foundation for the autonomous, adaptive, and resilient operation of HAPS in support of global coverage, high-capacity broadband, and the evolving demands of next-generation mobile networks. The frameworks and methodologies developed herein directly contribute to the vision of HAPS as a core enabler of intelligent NTN and a transformative component within the broader ecosystem of B5G and 6G wireless communications.

6.2 Future Work

While this thesis has made significant progress toward efficient multi-cell beamforming optimization for HAPS systems, several open challenges and limitations remain. The current methods, though effective at improving throughput and fairness, still face constraints in terms of fully capturing the diversity and non-stationarity of real-world UE distributions, as well as adapting to rapid environmental and network dynamics. In particular, ensuring robust generalization to highly irregular, previously unseen UE patterns and supporting seamless adaptation under time-varying or mobility-driven scenarios continue to pose difficulties. Moreover, further gains in efficiency and reliability are necessary for practical deployment at even larger scales or under stringent resource constraints.

Future work should focus on advancing the core algorithmic frameworks to address these challenges. Investigating more powerful neural architectures, such as those incorporating attention mechanisms, spatio-temporal modeling, or graph-based representations, may help the models better capture complex spatial and temporal dependencies among UEs. Exploring hybrid approaches that integrate deep learning with iterative optimization or domain-knowledge-driven constraints could enhance both solution quality and interpretability. In addition, the adoption of reinforcement learning or meta-learning frameworks could enable continual adaptation to evolving UE distributions and network

environments with minimal supervision. Such advancements would further align with the ultimate goal of enabling intelligent, autonomous, and scalable HAPS systems that can deliver robust, high-quality wireless connectivity across a wide range of deployment scenarios.

Future work should also consider RIS-enhanced beamforming, where reconfigurable intelligent surfaces are deployed within the coverage region to reflect and shape signal propagation in coordination with HAPS beams, potentially improving spatial coverage and adaptability in challenging deployment scenarios. Generative models are also a promising direction, particularly for synthesizing diverse UE distributions to enrich training and evaluation, and potentially for developing novel inference frameworks that directly generate beamforming configurations.

As HAPS deployments evolve toward broader and more complex configurations, future research should explore the potential of multi-HAPS collaboration, where multiple platforms jointly serve large and possibly overlapping coverage areas. While this thesis focuses on a single-HAPS scenario, practical implementations are likely to involve multiple HAPSs operating in parallel to support wider geographic regions, denser UE populations, and more diverse service requirements. This expanded setting introduces new challenges for beamforming optimization, including inter-HAPS interference coordination, collaborative cell configuration, and dynamic UE association across platform boundaries. Addressing these challenges will require scalable optimization frameworks that operate under decentralized conditions and adapt effectively to local variations.

To support such distributed operation, advanced learning-based approaches such as federated learning and multi-agent reinforcement learning (MARL) should be further explored. Federated learning enables each HAPS to train models locally on observed UE distributions while synchronizing model updates across platforms without sharing raw data, supporting communication efficiency and data privacy. MARL allows each HAPS to act as an autonomous agent that learns to optimize its own beamforming strategy while coordinating with other platforms to improve overall system performance. When multiple HAPS platforms operate as federated learning nodes, they form a distributed intelligent

layer that enables scalable deployment across large regions. This architecture provides a foundation for the integration of multi-cell HAPS systems into broader SAGIN environments, where the beamforming optimization framework developed in this thesis can serve as a core component to support intelligent and adaptive network-wide connectivity.

Beyond algorithmic improvements, future research should further explore the diverse roles of HAPS systems as they become an integral part of next-generation wireless network architectures. As a representative platform within the broader family of NTN, HAPS are expected to play a pivotal role in bridging coverage gaps, supporting rapid deployment, and enabling resilient connectivity in B5G and 6G systems. In particular, the integration of HAPS into advanced network paradigms such as SAGIN will create new opportunities and challenges for dynamic multi-cell configuration. These scenarios will demand optimization frameworks capable of addressing heterogeneous service requirements, supporting high user mobility, and adapting to rapidly changing environments across large spatial and temporal scales.

Furthermore, with the anticipated proliferation of massive IoT deployments and the convergence of terrestrial and non-terrestrial infrastructures, dynamic HAPS cell configuration will become increasingly important for meeting diverse application demands and ensuring efficient spectrum utilization. Continued research should therefore focus on the development, validation, and field demonstration of robust multi-cell optimization solutions for HAPS in these evolving scenarios. Semantic-aware optimization may also be considered, where future beamforming strategies take into account the contextual importance of transmitted data. This could enable multi-cell HAPS systems to dynamically allocate resources based on traffic semantics, helping address diverse QoS requirements and improving spectrum efficiency in IoT-dense environments. This will be essential for realizing the vision of intelligent, flexible, and universally accessible wireless networks in the coming generations of mobile communication.

Pursuing these research directions will enable the optimization and learning frameworks developed in this thesis to evolve alongside the rapidly advancing landscape of wireless communications. In doing so, they will support the increasing demands of

intelligent network management and help pave the way for robust, flexible, and high-performance connectivity in future wireless systems.

Appendix A

List of Author's Publications and Awards

A.1 Articles on Periodicals

1. **W. Dai**, S. Yang, Y. Yin, M. Bouazizi, and T. Ohtsuki, "A Clustering-Aided Optimization Algorithm for Antenna Beamforming in Multicell HAPS Systems," *IEEE Internet of Things Journal*, vol. 12, no. 15, pp. 30168-30184, Aug. 2025, doi: 10.1109/JIOT.2025.3570163.

A.2 Articles on International Conference Proceedings

1. **W. Dai**, S. Yang, M. Bouazizi, and T. Ohtsuki, "K-Means Clustering-Aided Dynamic Multi-Cell Optimization Algorithm for HAPS," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec. 2023, pp. 5635–5640, doi: 10.1109/GLOBECOM54140.2023.10437302.

A.3 Presentations at International Conferences

1. **W. Dai**, S. Yang, Z. Li, M. Bouazizi, and T. Ohtsuki, “A Neural Network Approach for HAPS Dynamic Multi-Cell Optimization,” *A3 Foresight Program Workshop*, Jan. 2024.

A.4 Articles on Domestic Conference Proceedings

1. **W. Dai**, S. Yang, M. Bouazizi, and T. Ohtsuki, “Improved Initial Clustering for Antenna Control in Multi-Cell HAPS System,” *IEICE General Conference*, Mar. 2024.
2. **W. Dai**, S. Yang, M. Bouazizi, and T. Ohtsuki, “K-Means-Aided Particle Swarm Optimization for Antenna Control in Multi-Cell HAPS System,” *IEICE Society Conference*, Sep. 2023.

A.5 Awards

1. A3 Foresight Program 2024 Workshop Best Presentation Award, Feb. 2024.

References

- [1] D. C. Nguyen *et al.*, “6G Internet of Things: A comprehensive survey,” *IEEE Internet Things J.*, vol. 9, no. 1, pp. 359–383, Jan. 2022.
- [2] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, “Internet of Things: A survey on enabling technologies, protocols, and applications,” *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 2015.
- [3] W. Saad, M. Bennis, and M. Chen, “A vision of 6G wireless systems: Applications, trends, technologies, and open research problems,” *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020.
- [4] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, “The roadmap to 6G: AI empowered wireless networks,” *IEEE Commun. Mag.*, vol. 57, no. 8, pp. 84–90, 2019.
- [5] F. Tariq, M. R. A. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, “A speculative study on 6G,” *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 118–125, 2020.
- [6] L. Chettri and R. Bera, “A comprehensive survey on Internet of Things (IoT) toward 5G wireless systems,” *IEEE Internet Things J.*, vol. 7, no. 1, pp. 16–32, 2020.
- [7] R. Miura and M. Oodo, “Wireless communications system using stratospheric platforms: R and D program on telecom and broadcasting system using high altitude platform stations,” *J. Commun. Res. Lab.*, vol. 48, no. 4, pp. 33–48, Dec. 2001.

- [8] S. Karapantazis and F.-N. Pavlidou, “Broadband communications via high-altitude platforms: A survey,” *IEEE Commun. Surveys Tuts.*, vol. 7, pp. 2–31, 1st Quart. 2005.
- [9] G. K. Kurt *et al.*, “A vision and framework for the high altitude platform station (HAPS) networks of the future,” *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 729–779, 2nd Quart. 2021.
- [10] K. An *et al.*, “Exploiting multi-layer refracting RIS-assisted receiver for HAP-SWIPT networks,” *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 12 638–12 657, Oct. 2024.
- [11] AALTO HAPS Ltd. Accessed: Aug. 2024. [Online]. Available: <https://www.aaltohaps.com/>
- [12] Space Compass Crop. Accessed: Aug. 2024. [Online]. Available: <https://space-compass.com/en/>
- [13] SoftBank Crop. Accessed: Aug. 2024. [Online]. Available: <https://www.softbank.jp/en/corp/philosophy/technology/special/ntn-solution/haps/>
- [14] *Study on Using Satellite Access in 5G*, 3GPP document TR22.822 V16.0.0, Jun. 2018.
- [15] *Study on Scenarios and Requirements for Next Generation Access Technologies*, 3GPP document TR 38.913 v15.0.0, Sep. 2018.
- [16] J. Liu, Y. Shi, Z. M. Fadlullah, and N. Kato, “Space-air-ground integrated network: A survey,” *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2714–2741, 4th Quart. 2018.
- [17] Z. Jia, M. Sheng, J. Li, D. Zhou, and Z. Han, “Joint HAP access and LEO satellite backhaul in 6G: Matching game-based approaches,” *IEEE J. Sel. Areas Comm.*, vol. 39, no. 4, pp. 1147–1159, Apr. 2021.

- [18] S. Xu, J. Liu, T. K. Rodrigues, and N. Kato, “Robust multiuser beamforming for IRS-enhanced near-space downlink communications coexisting with satellite system,” *IEEE Internet Things J.*, vol. 9, no. 16, pp. 14 900–14 912, Aug. 2022.
- [19] D. Zhou, M. Sheng, J. Li, and Z. Han, “Aerospace integrated networks innovation for empowering 6G: A survey and future challenges,” *IEEE Commun. Surveys Tuts.*, vol. 25, no. 2, pp. 975–1019, 2nd Quart. 2023.
- [20] Y. Shibata, W. Takabatake, K. Hoshino, A. Nagate, and T. Ohtsuki, “HAPS cell design method for coexistence on terrestrial mobile networks,” in *Proc. IEEE 97th Veh. Technol. Conf. (VTC-Spring)*, Jun. 2023, pp. 1–6.
- [21] Y. Shibata, W. Takabatake, K. Hoshino, A. Nagate, and T. Ohtsuki, “HAPS cell design method for coverage extension considering coexistence on terrestrial mobile networks,” *IEEE Access*, vol. 12, pp. 55 506–55 520, 2024.
- [22] Y. Kawamoto *et al.*, “HAPS-based interference suppression through null broadening with directivity control in space-air-ground integrated networks,” *IEEE Trans. Veh. Technol.*, vol. 72, no. 12, pp. 16 098–16 107, Dec. 2023.
- [23] Z. Jia, M. Sheng, J. Li, and Z. Han, “Toward data collection and transmission in 6G space–air–ground integrated networks: Cooperative HAP and LEO satellite schemes,” *IEEE Internet Things J.*, vol. 9, no. 13, pp. 10 516–10 528, Jul. 2022.
- [24] Y. Kawamoto, M. Takahashi, S. Verma, N. Kato, H. Tsuji, and A. Miura, “Traffic-prediction-based dynamic resource control strategy in HAPS-mounted MEC-assisted satellite communication systems,” *IEEE Internet Things J.*, vol. 11, no. 8, pp. 13 824–13 836, Apr. 2024.
- [25] Z. Jia, Q. Wu, C. Dong, C. Yuen, and Z. Han, “Hierarchical aerial computing for Internet of Things via cooperation of HAPs and UAVs,” *IEEE Internet Things J.*, vol. 10, no. 7, pp. 5676–5688, Oct. 2023.

- [26] Y. Sun *et al.*, “Multi-functional RIS-assisted semantic anti-jamming communication and computing in integrated aerial-ground networks,” *IEEE J. Sel. Areas Comm.*, vol. 42, no. 12, pp. 3597–3617, Dec. 2024.
- [27] M. Centenaro, C. E. Costa, F. Granelli, C. Sacchi, and L. Vangelista, “A survey on technologies, standards and open challenges in satellite IoT,” *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1693–1720, 3rd Quart. 2021.
- [28] A. Kak and I. F. Akyildiz, “Designing large-scale constellations for the internet of space things with CubeSats,” *IEEE Internet Things J.*, vol. 8, no. 3, pp. 1749–1768, Feb. 2021.
- [29] R. Xie, Q. Tang, Q. Wang, X. Liu, F. R. Yu, and T. Huang, “Satellite-terrestrial integrated edge computing networks: Architecture, challenges, and open issues,” *IEEE Netw.*, vol. 34, no. 3, pp. 224–231, May/Jun. 2020.
- [30] O. Kodheli *et al.*, “Satellite communications in the new space era: A survey and future challenges,” *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 70–109, 1st Quart. 2021.
- [31] M. S. Alam, G. K. Kurt, H. Yanikomeroglu, P. Zhu, and N. D. Đào, “High altitude platform station based super macro base station constellations,” *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 103–109, Jan. 2021.
- [32] C. Li, Y. Zhang, R. Xie, X. Hao, and T. Huang, “Integrating edge computing into low earth orbit satellite networks: Architecture and prototype,” *IEEE Access*, vol. 9, pp. 39 126–39 137, 2021.
- [33] A. Alsharoa and M.-S. Alouini, “Improvement of the global connectivity using integrated satellite-airborne-terrestrial networks with resource optimization,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5088–5100, Aug. 2020.
- [34] J. Liu, X. Du, J. Cui, M. Pan, and D. Wei, “Task-oriented intelligent networking architecture for the space–air–ground–aqua integrated network,” *IEEE Internet Things J.*, vol. 7, no. 6, pp. 5345–5358, Jun. 2020.

- [35] S. Ji, M. Sheng, D. Zhou, W. Bai, Q. Cao, and J. Li, “Flexible and distributed mobility management for integrated terrestrial-satellite networks: Challenges, architectures, and approaches,” *IEEE Netw.*, vol. 35, no. 4, pp. 73–81, Jul./Aug. 2021.
- [36] T. Bai, R. Vaze, and R. W. Heath, “Analysis of blockage effects on urban cellular networks,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 5070–5083, Sep. 2014.
- [37] Y. Li, W. Zhou, and S. Zhou, “Forecast based handover in an extensible multi-layer LEO mobile satellite system,” *IEEE Access*, vol. 8, pp. 42 768–42 783, 2020.
- [38] Z. Bao, M. Luo, H. Wang, K.-K. R. Choo, and D. He, “Blockchain-based secure communication for space information networks,” *IEEE Netw.*, vol. 35, no. 4, pp. 50–57, Jul./Aug. 2021.
- [39] R. Han, L. Bai, C. Jiang, J. Liu, and J. Choi, “A secure architecture of relay-aided space information networks,” *IEEE Netw.*, vol. 35, no. 4, pp. 88–94, Jul./Aug. 2021.
- [40] A. Baltaci, E. Dinc, M. Ozger, A. Alabbasi, C. Cavdar, and D. Schupke, “A survey of wireless networks for future aerial communications (FACOM),” *IEEE Commun. Surveys Tuts.*, vol. 23, no. 4, pp. 2833–2884, 4th Quart. 2021.
- [41] N. Bhushan *et al.*, “Network densification: the dominant theme for wireless evolution into 5G,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb 2014.
- [42] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta, “Are we approaching the fundamental limits of wireless network densification?” *IEEE Commun. Mag.*, vol. 54, no. 10, pp. 184–190, Oct. 2016.
- [43] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, “Data-driven intelligent transportation systems: A survey,” *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

- [44] J. Li, K. Xue, J. Liu, Y. Zhang, and Y. Fang, “An ICN/SDN-based network architecture and efficient content retrieval for future satellite-terrestrial integrated networks,” *IEEE Netw.*, vol. 34, no. 1, pp. 188–195, Jan./Feb. 2020.
- [45] M. Alzenad, M. Z. Shakir, H. Yanikomeroglu, and M.-S. Alouini, “FSO-based vertical backhaul/fronthaul framework for 5G+ wireless networks,” *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 218–224, Jan. 2018.
- [46] Y. Su, Y. Liu, Y. Zhou, J. Yuan, H. Cao, and J. Shi, “Broadband LEO satellite communications: Architectures and key technologies,” *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 55–61, Apr. 2019.
- [47] C. Ding, J.-B. Wang, H. Zhang, M. Lin, and G. Y. Li, “Joint optimization of transmission and computation resources for satellite and high altitude platform assisted edge computing,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 1362–1377, Feb. 2022.
- [48] K. Tashiro, K. Hoshino, and A. Nagate, “Cylindrical massive MIMO system for HAPS: Capacity enhancement and coverage extension,” in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC-Spring)*, Apr. 2021, pp. 1–6.
- [49] K. Tashiro, K. Hoshino, and A. Nagate, “Nullforming-based precoder for spectrum sharing between HAPS and terrestrial mobile networks,” *IEEE Access*, vol. 10, pp. 55 675–55 693, 2022.
- [50] O. Abbasi, H. Yanikomeroglu, and G. Kaddoum, “Hemispherical antenna array architecture for high-altitude platform stations (HAPS) for uniform capacity provision,” *IEEE Trans. Wireless Commun.*, vol. 23, no. 12, pp. 19 489–19 504, Dec. 2024.
- [51] K. Tashiro, K. Hoshino, and A. Nagate, “Three-stage precoding scheme based on nullforming and subarray partitioning for aerial–terrestrial spectrum sharing,” *IEEE Access*, vol. 13, pp. 53 640–53 659, 2025.

- [52] K. Tashiro, T. Ishikawa, K. Hoshino, and A. Nagate, “Field trials and evaluations of nullforming-based spectrum sharing between aerial platforms and terrestrial networks on 5G FDD bands,” in *IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2025, pp. 1–6.
- [53] K. Maki, K. Tashiro, T. Ishikawa, and K. Hoshino, “Downtilt optimization of array antennas based on envelope radiation patterns for HAPS service links with massive MIMO,” *IEEE Access*, vol. 13, pp. 34 307–34 320, 2025.
- [54] K. Tashiro, T. Ishikawa, K. Hoshino, and A. Nagate, “Implementation and experimental evaluation of nullforming-based interference reduction scheme for HAPS-terrestrial spectrum sharing,” in *Proc. IEEE 100th Veh. Technol. Conf. (VTC-Fall)*, 2024, pp. 1–6.
- [55] T. Ishikawa, K. Tashiro, K. Hoshino, and A. Nagate, “Coordinated nullforming with limited control parameters for spectrum sharing between high-altitude platforms and terrestrial networks,” in *Proc. IEEE 99th Veh. Technol. Conf. (VTC-Spring)*, 2024, pp. 1–6.
- [56] K. Tashiro, K. Hoshino, and A. Nagate, “Interference reduction between HAPSs using subarray grouping and nullforming techniques for cylindrical massive MIMO systems,” in *Proc. IEEE 97th Veh. Technol. Conf. (VTC-Spring)*, 2023, pp. 1–5.
- [57] T. Ishikawa, K. Tashiro, K. Hoshino, and A. Nagate, “Spectrum sharing between high-altitude platforms and terrestrial networks using interference coordination by null sweeping,” in *Proc. IEEE 98th Veh. Technol. Conf. (VTC-Fall)*, 2023, pp. 1–6.
- [58] B. El-Jabu and R. Steele, “Cellular communications using aerial platforms,” *IEEE Trans. Veh. Technol.*, vol. 50, no. 3, pp. 686–700, May 2001.
- [59] Y. Shibata, N. Kanazawa, K. Hoshino, Y. Ohta, and A. Nagate, “A study on cell configuration for HAPS mobile communications,” in *Proc. IEEE 89th Veh. Technol. Conf. (VTC-Spring)*, Apr. 2019, pp. 1–6.

- [60] Y. Shibata, N. Kanazawa, M. Konishi, K. Hoshino, Y. Ohta, and A. Nagate, “System design of gigabit HAPS mobile communications,” *IEEE Access*, vol. 8, pp. 157 995–158 007, 2020.
- [61] F. Hsieh, F. Jardel, E. Visotsky, F. Vook, A. Ghosh, and B. Picha, “UAV-based multi-cell HAPS communication: System design and performance evaluation,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6.
- [62] Y. Shibata, W. Takabatake, and K. Hoshino, “Field experiment of a dynamic area optimization algorithm using a cylindrical antenna for HAPS,” in *Proc. IEEE 29th Asia Pacific Conf. Commun. (APCC)*, 2024, pp. 503–506.
- [63] M. Konishi, T. Nishimaki, Y. Shibata, S. Nabatame, and A. Nagate, “An experimental study of uplink co-channel spectrum-sharing system between HAPS and terrestrial mobile communication networks,” in *Proc. IEEE 92nd Veh. Technol. Conf. (VTC-Fall)*, 2020, pp. 1–5.
- [64] M. Konishi, T. Nishimaki, Y. Shibata, S. Nabatame, and A. Nagate, “A study of co-channel spectrum-sharing system between HAPS and terrestrial mobile communication networks,” in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, 2020, pp. 1–5.
- [65] Z. Li, S. Yang, Y. Yin, M. Bouazizi, and T. Ohtsuki, “Beamforming design using UE positions and 3D terrain-building in HAPS system,” in *Proc. IEEE 100th Veh. Technol. Conf. (VTC-Fall)*, 2024, pp. 1–7.
- [66] J. Thornton and D. Grace, “Effect of lateral displacement of a high-altitude platform on cellular interference and handover,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1483–1490, Jul. 2005.
- [67] X. Wang, L. Li, and W. Zhou, “The effect of HAPS unstable movement on handover performance,” in *Proc. Wireless Opt. Commun. Conf. (WOCC)*, May 2019, pp. 1–5.

- [68] H. Panfeng, C. Naiping, and N. Shuyan, "Coverage model of multi beam antenna from high altitude platform in the swing state," *2016 7th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, pp. 750–753, 2016.
- [69] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Found. Trends Signal Process.*, vol. 11, no. 3-4, p. 154–655, 2017.
- [70] S. C. Arum, D. Grace, P. D. Mitchell, and M. D. Zakaria, "Beam-pointing algorithm for contiguous high-altitude platform cell formation for extended coverage," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, vol. Sep., 2019, pp. 1–5.
- [71] S. C. Arum, D. Grace, and P. D. Mitchell, "Extending coverage and capacity from high altitude platforms with a two-tier cellular architecture," *IEEE Trans. Mobile Comput.*, vol. 23, no. 2, pp. 1942–1953, Feb. 2024.
- [72] A. M. Ismaiel, E. Elsaidy, Y. Albagory, H. A. Atallah, A. B. Abdel-Rahman, and T. Sallam, "Performance improvement of high altitude platform using concentric circular antenna array based on particle swarm optimization," *AEU - Int. J. Electron. Commun.*, vol. 91, pp. 85–90, 2018.
- [73] Y. Shibata, W. Takabatake, K. Hoshino, A. Nagate, and T. Ohtsuki, "Co-evolutionary dynamic cell optimization algorithm for HAPS mobile communications," in *Proc. IEEE 95th Veh. Technol. Conf. (VTC-Spring)*, Jun. 2022, pp. 1–6.
- [74] Y. Shibata, W. Takabatake, K. Hoshino, A. Nagate, and T. Ohtsuki, "Two-step dynamic cell optimization algorithm for HAPS mobile communications," *IEEE Access*, vol. 10, pp. 68 085–68 098, 2022.
- [75] K. Wada *et al.*, "Dynamic antenna control for HAPS using fuzzy Q-learning in multi-cell configuration," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2022, pp. 1–6.

- [76] S. Yang, M. Bouazizi, and T. Ohtsuki, “Dynamic antenna control for HAPS using mean field reinforcement learning in multi-cell configuration,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2023, pp. 5670–5676.
- [77] S. Yang *et al.*, “Deep reinforcement learning evolution algorithm for dynamic antenna control in multi-cell configuration HAPS system,” *Future Internet*, vol. 15, no. 1, 2023.
- [78] M. Bouazizi, S. Yang, and T. Ohtsuki, “A low-complexity clustering-aided DQN method for dynamic antenna control in HAPS,” in *Proc. IEEE 99th Veh. Technol. Conf. (VTC-Spring)*, 2024, pp. 1–6.
- [79] S. Yang, T. Ohtsuki, and X. Jiang, “Joint angle-based user selection and multiagent reinforcement learning for dynamic beamforming in HAPS-assisted IoT vehicular networks,” *IEEE Internet Things J.*, early access, 2025.
- [80] S. Yang *et al.*, “Dynamic antenna control for HAPS using geometry-based method in multi-cell configuration,” in *Proc. IEEE 95th Veh. Technol. Conf. (VTC-Spring)*, Jun. 2022, pp. 1–6.
- [81] W. Takabatake, Y. Shibata, and K. Hoshino, “Neural-network-based dynamic area optimization algorithm for high-altitude platform station,” in *Proc. IEEE 98th Veh. Technol. Conf. (VTC-Fall)*, Oct. 2023, pp. 1–6.
- [82] W. Takabatake, Y. Shibata, K. Hoshino, and T. Ohtsuki, “Time-efficient neural-network-based dynamic area optimization algorithm for high-altitude platform station mobile communications,” *Future Internet*, vol. 16, no. 9, 2024.
- [83] F. A. d’Oliveira, F. C. L. D. Melo, and T. C. Devezas, “High-altitude platforms—Present situation and technology trends,” *J. Aerosp. Technol. Manag.*, vol. 8, no. 3, pp. 249–262, Jul.–Sep. 2016.
- [84] S. C. Arum, D. Grace, and P. D. Mitchell, “A review of wireless communication using high-altitude platforms for extended coverage and capacity,” *Comput. Commun.*, vol. 157, no. 1, pp. 232–256, May 2020.

- [85] S. Sibiya and O. O. Olugbara, “Reliable Internet of Things network architecture based on high altitude platforms,” in *Proc. IEEE Conf. Inf. Commun. Technol. Soc. (ICTAS)*, 2019, pp. 1–4.
- [86] A. Benjebbour, “An overview of non-orthogonal multiple access,” *ZTE Commun.*, vol. 15, no. 51, pp. 21–30, Jun. 2017.
- [87] J. G. Andrews *et al.*, “What will 5G be?” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [88] F. Dong, H. Han, X. Gong, J. Wang, and H. Li, “A constellation design methodology based on QoS and user demand in high-altitude platform broadband networks,” *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2384–2397, Dec. 2016.
- [89] D. Sarddar *et al.*, “A handover management in LEO satellite network using angular and distance based algorithm,” *Int. J. Comput. Appl.*, vol. 975, p. 8887, Oct. 2011.
- [90] J. Thornton, D. Grace, M. H. Capstick, and T. C. Tozer, “Optimizing an array of antennas for cellular coverage from a high altitude platform,” *IEEE Trans. Wireless Commun.*, vol. 2, no. 3, pp. 484–492, May 2003.
- [91] M. Alzenad and H. Yanikomeroglu, “Coverage and rate analysis for unmanned aerial vehicle base stations with LoS/NLoS propagation,” in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2018, pp. 1–7.
- [92] O. Said and A. Tolba, “Performance evaluation of a dual coverage system for Internet of Things environments,” *Mobile Inf. Syst.*, vol. 975, pp. 1–20, Dec. 2016.
- [93] S. O. Olatinwo and T.-H. Joubert, “Enabling communication networks for water quality monitoring applications: A survey,” *IEEE Access*, vol. 7, pp. 100 332–100 362, Dec. 2019.
- [94] P. He, N. Cheng, and J. Cui, “Handover performance analysis of cellular communication system from high altitude platform in the swing state,” in *Proc. IEEE Int. Conf. Signal Image Process. (ICSIP)*, 2016, pp. 407–411.

- [95] P. He, N. Cheng, and S. Ni, "Improved LMS predictive link triggering for handover in HAPS communication system," in *Proc. IEEE 8th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, 2016, pp. 1–5.
- [96] K. Li, Y. Li, Z. Qiu, Q. Wang, J. Lu, and W. Zhou, "Handover procedure design and performance optimization strategy in LEO-HAP system," in *Proc. IEEE 11th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, 2019, pp. 1–7.
- [97] S. Alsamhi and N. Rajput, "An intelligent hand-off algorithm to enhance quality of service in high altitude platforms using neural network," *Wireless Pers. Commun.*, vol. 82, no. 4, pp. 2059–2073, Jan. 2015.
- [98] Y. Albagory, M. Nofal, and A. Ghoneim, "Handover performance of unstable-yaw stratospheric high-altitude stations," *Wireless Pers. Commun.*, vol. 84, no. 4, pp. 2651–2663, May 2015.
- [99] P. He, N. Cheng, S. Ni, and C. Li, "An adaptive handover scheme based on cooperative transmission from high altitude platform stations," in *Proc. IEEE 2nd Adv. Inf. Technol. Electron. Autom. Control Conf. (IAEAC)*, 2017, pp. 1306–1310.
- [100] S. Li, D. Grace, J. Wei, and D. Ma, "A novel guaranteed handover scheme for HAP communications systems with adaptive modulation and coding," in *Proc. IEEE Veh. Technol. Conf. (VTC-Fall)*, 2010, pp. 1–5.
- [101] A. Alsharoa and M. Alouini, "Facilitating satellite-airborne-terrestrial integration for dynamic and infrastructure-less networks," *CoRR*, vol. abs/1912.03819, 2019. [Online]. Available: <http://arxiv.org/abs/1912.03819>
- [102] M. Mirahsan, R. Schoenen, and H. Yanikomeroglu, "Hethetnets: Heterogeneous traffic distribution in heterogeneous wireless cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2252–2265, Oct. 2015.
- [103] F. Rusek and J. B. Anderson, "Constrained capacities for faster-than-Nyquist signaling," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 764–775, Feb. 2009.

- [104] J. A. Prakash and G. R. Reddy, "Efficient prototype filter design for filter bank multicarrier (FBMC) system based on ambiguity function analysis of hermite polynomials," in *Proc. IEEE Int. Mutliconf. Autom. Comput. Commun. Control Compressed Sens. (iMac4s)*, 2013, pp. 580–585.
- [105] Z. E. Ankaralı, B. Peköz, and H. Arslan, "Enhanced OFDM for 5G RAN," *ZTE Commun.*, vol. 15, no. 51, pp. 11–20, Jun. 2020.
- [106] J. B. Anderson, F. Rusek, and V. Öwall, "Faster-than-Nyquist signaling," *ZTE Commun.*, vol. 101, no. 8, pp. 1817–1830, Mar. 2013.
- [107] J. E. Mazo, "Faster-than-Nyquist signaling," *Bell Syst. Techn. J.*, vol. 54, no. 8, pp. 1451–1462, Oct. 1975.
- [108] A. D. Liveris and C. N. Georghiades, "Exploiting faster-than-Nyquist signaling," *IEEE Trans. Commun.*, vol. 51, no. 9, pp. 1502–1511, Sep. 2003.
- [109] A. Abrardo and D. Sennati, "Centralized radio resource management strategies with heterogeneous traffics in HAPS WCMA cellular systems," *IEICE Trans. Commun.*, vol. 86, no. 3, pp. 1040–1049, Mar. 2003.
- [110] Y. C. Foo, W. L. Lim, and R. Tafazolli, "Centralized total received power based call admission control for high altitude platform station UMTS," in *Proc. 13th IEEE Int. Symp. Pers. Indoor Mobile Radio Commun.*, 2002, pp. 1596–1600.
- [111] Y. C. Foo, W. L. Lim, and R. Tafazolli, "Centralized downlink call admission control for high altitude platform station UMTS with onboard power resource sharing," in *IEEE 56th Veh. Technol. Conf.*, 2002, pp. 549–553.
- [112] Y. C. Foo, W. L. Lim, and R. Tafazolli, "Call admission control schemes for high altitude platform station and terrestrial tower-based hierarchical UMTS," in *Proc. IEEE 9th Int. Conf. Commun. Syst. (ICCS)*, 2004, pp. 531–536.

- [113] G. Araniti, A. Molinaro, and A. Iera, "Multicast in terrestrial-HAP systems: User number vs. user distribution oriented RRM policies," in *Proc. IEEE 66th Veh. Technol. Conf.*, 2007, pp. 154–158.
- [114] F. Adachi, D. Garg, S. Takaoka, and K. Takeda, "Broadband CDMA techniques," *IEEE Wireless Commun.*, vol. 12, no. 2, pp. 8–18, Apr. 2005.
- [115] C. K. Datsikas, K. P. Peppas, N. C. Sagias, and G. S. Tombras, "Serial free-space optical relaying communications over gamma-gamma atmospheric turbulence channels," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 2, no. 8, pp. 576–586, Aug. 2010.
- [116] M. Sharma, D. Chadha, and V. Chandra, "High-altitude platform for free-space optical communication: Performance evaluation and reliability analysis," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 8, no. 8, pp. 600–609, Aug. 2016.
- [117] A. M. Salhab, "A new scenario of triple-hop mixed RF/FSO/RF relay network with generalized order user scheduling and power allocation," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 2016, no. 1, pp. 1–20, Oct. 2016.
- [118] E. T. Michailidis, P. B. N. Nomikos, D. Vouyioukas, and A. G. Kanatas, "Outage probability of triple-hop mixed RF/FSO/RF stratospheric communication systems," in *Proc. IEEE 10th Int. Conf. Adv. Satellite Space Commun. (SPACOMM)*, 2018, pp. 1–6.
- [119] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for wireless resource management," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2017, pp. 1–6.
- [120] Z. Ding, X. Lei, G. K. Karagiannidis, J. Y. R. Schober, and V. K. Bhargava, "Broadband CDMA techniques," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

- [121] N. C. Luong *et al.*, “Applications of deep reinforcement learning in communications and networking: A survey,” *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3133–3174, 4th Quart. 2019.
- [122] L. Zhang, Y.-C. Liang, and D. Niyato, “6G visions: Mobile ultra-broadband, super Internet-of-Things, and artificial intelligence,” *China Commun.*, vol. 16, no. 8, pp. 1–14, Aug. 2019.
- [123] V. Raj, I. Dias, T. Tholeti, and S. Kalyani, “Spectrum access in cognitive radio using a two-stage reinforcement learning approach,” *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 20–34, Feb. 2018.
- [124] M. Qiao, H. Zhao, L. Zhou, C. Zhu, and S. Huang, “Topology-transparent scheduling based on reinforcement learning in self-organized wireless networks,” *IEEE Access*, vol. 6, pp. 20 221–20 230, 2018.
- [125] D. D. Nguyen, H. X. Nguyen, and L. B. White, “Reinforcement learning with network-assisted feedback for heterogeneous RAT selection,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6062–6076, Sep. 2017.
- [126] H. Ahmadi, K. Katzis, and M. Z. Shakir, “A novel airborne self-organising architecture for 5G+ networks,” in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, 2017, pp. 1–5.
- [127] M. Guan, “An intelligent wireless channel allocation in HAPS 5G communication system based on reinforcement learning,” *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 138, pp. 1–10, May 2019.
- [128] S. Branco, A. G. Ferreira, and J. Cabral, “Machine learning in resource-scarce embedded systems, FPGAs, and end-devices: A survey,” *Electronics*, vol. 8, no. 11, p. 1289, Nov. 2019.
- [129] S. Wang *et al.*, “Federated learning for task and resource allocation in wireless high-altitude balloon networks,” *IEEE Internet Things J.*, vol. 8, no. 24, pp. 17 460–17 475, Dec. 2021.

- [130] J. Liu *et al.*, “Minimization of offloading delay for two-tier UAV with mobile edge computing,” in *Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, 2019, pp. 1534–1538.
- [131] M. D. Zakaria, D. Grace, P. D. Mitchell, T. M. Shami, and N. Morozs, “Exploiting user-centric joint transmission-coordinated multipoint with a high altitude platform system architecture,” *IEEE Access*, vol. 7, pp. 38 957–38 972, 2019.
- [132] G. Chen, D. Grace, and T. C. Tozer, “Performance of multiple high altitude platforms using directive HAP and user antennas,” *Wireless Pers. Commun.*, vol. 32, no. 3-4, pp. 275–299, Feb. 2005.
- [133] F. Dong, M. Li, X. Gong, H. Li, and F. Gao, “Diversity performance analysis on multiple HAP networks,” *Sensors*, vol. 15, no. 7, pp. 15 398–15 418, Jun. 2015.
- [134] D.-W. Yue, S. Xu, and H. H. Nguyen, “Diversity gain of millimeter-wave massive MIMO systems with distributed antenna arrays,” *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 54, Mar. 2019.
- [135] Z. Lian, L. Jiang, C. He, and D. He, “User grouping and beamforming for HAP massive MIMO systems based on statistical-eigenmode,” *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 961–964, Jun. 2019.
- [136] D. Grace, J. Thornton, G. Chen, G. P. White, and T. C. Tozer, “Improving the system capacity of broadband services using multiple high-altitude platforms,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 700–709, Mar. 2005.
- [137] Z. Lin, M. Lin, Y. Huang, T. de Cola, and W.-P. Zhu, “Robust multi-objective beamforming for integrated satellite and high altitude platform network with imperfect channel state information,” *IEEE Trans. Signal Process.*, vol. 67, no. 24, pp. 6384–6396, Dec. 2019.
- [138] M. Guan *et al.*, “Efficiency evaluations based on artificial intelligence for 5G massive MIMO communication systems on high-altitude platform stations,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6632–6640, Oct. 2019.

- [139] *Minimum Performance Characteristics and Operational Conditions for High Altitude Platform Stations providing IMT-2000 in the Bands 1885-1980 MHz 2010-2025 MHz and 2110-2170 MHz in the Regions 1 and 3 and 1885-1980 MHz and 2110 2160 MHz in Region 2*, Rec. document ITU-R M.1456, 2000.
- [140] *Characteristics of Terrestrial IMT-Advanced Systems for Frequency Sharing/Interference Analyses*, document ITU-R M.2292-0, 2013.
- [141] D. Grace and M. Mohorcic, *Broadband communications via high altitude platforms*. John Wiley & Sons, 2011.
- [142] K. Hoshino, S. Sudo, and Y. Ohta, “A study on antenna beamforming method considering movement of solar plane in HAPS system,” in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2019, pp. 1–5.
- [143] S. Sudo, K. Hoshino, and Y. Ohta, “Basic evaluation of service link antenna for footprint fixation in HAPS system,” in *Proc. IEEE 92th Veh. Technol. Conf. (VTC-Fall)*, Nov. 2020, pp. 1–5.
- [144] Z. Lin, M. Lin, B. Champagne, W.-P. Zhu, and N. Al-Dhahir, “Secrecy-energy efficient hybrid beamforming for satellite-terrestrial integrated networks,” *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6345–6360, Sep. 2021.
- [145] Y. Yoon and Y.-H. Kim, “An efficient genetic algorithm for maximum coverage deployment in wireless sensor networks,” *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1473–1483, Oct. 2013.
- [146] Y. Qin, W. Huangfu, H. Zhang, and K. Long, “Boosting the cellular network coverage optimization in accordance with the metric structure of antenna variables,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8303–8314, Dec. 2020.
- [147] Q. Zhang, Q. Xi, C. He, and L. Jiang, “User clustered opportunistic beamforming for stratospheric communications,” *IEEE Commun. Lett.*, vol. 20, no. 9, pp. 1832–1835, Sep. 2016.

- [148] Z. Lian, L. Jiang, C. He, and D. He, “User grouping and beamforming for HAP massive MIMO systems based on statistical-eigenmode,” *IEEE Commun. Lett.*, vol. 8, no. 3, pp. 961–964, Jun. 2019.
- [149] E. Balevi and J. G. Andrews, “A novel deep reinforcement learning algorithm for online antenna tuning,” *2019 IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1–6, Dec. 2019.
- [150] R. Shafin *et al.*, “Self-tuning sectorization: Deep reinforcement learning meets broadcast beam optimization,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 4038–4053, Jun. 2020.
- [151] P. Bradley, K. Bennett, and A. Demiriz, *Constrained K-Means Clustering*, Microsoft Res., Redmond, WA, USA, 2000.
- [152] M. Malinen and P. Fränti, “Balanced k-means for clustering,” *Joint Int. Workshop Struct. Syntactic Stat. Pattern Recognit. (S+SSPR)*, pp. 32–41, Aug. 2014.
- [153] G. Tzortzis and A. Likas, “The minmax k-means clustering algorithm,” *Pattern Recognit.*, vol. 47, no. 7, pp. 2505–2516, Jul. 2014.
- [154] Agoop Corporation. *Dynamic Population Data*. Accessed: March. 2023. [Online]. Available: <https://agoop.co.jp/service/dynamic-population-data/>
- [155] OpenMapTiles. Accessed: Aug. 2024. [Online]. Available: <https://openmaptiles.org/>
- [156] A. P. Piotrowski, J. J. Napiorkowski, and A. E. Piotrowska, “Population size in particle swarm optimization,” *Swarm Evol. Comput.*, vol. 58, p. 100718, Nov. 2020.
- [157] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, “A ConvNet for the 2020s,” in *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11 966–11 976.

- [158] S. Woo *et al.*, “ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders,” in *IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 16 133–16 142.
- [159] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in *Int. Conf. Learn. Representations*, 2019.
- [160] L. N. Smith and N. Topin, “Super-convergence: Very fast training of neural networks using large learning rates,” in *Artif. Intell. Mach. Learn. Multi-Domain Oper. Appl.*, May 2019, p. 369–386.