IAB Network Planning: A Deep Learning Framework for 6G Wireless Systems

Jie Zhang

Doctor of Philosophy

University of York

Department of Physics, Engineering and Technology

September 2025

Abstract

Rapid growth in mobile data traffic and bandwidth-intensive applications is driving unprecedented densification in 5th Generation Mobile Networks (5G) and beyond. Integrated Access and Backhaul (IAB) operating in the Millimeter Wave (mmWave) spectrum offers abundant bandwidth and flexible wireless backhaul, but network planning remains NP-hard and mmWave links are vulnerable to environmental disruptions. This thesis presents a comprehensive framework for intelligent IAB network planning through a systematic evolution of optimisation and learning methods.

We first establish baselines using Mixed-Integer Linear Programming (MILP) formulations and greedy heuristics, achieving near-optimal solutions for networks with up to 30 nodes while exposing exponential computational growth for larger deployments. To improve scalability, we develop Deep Q-Network (DQN) variants with action elimination, reducing the required number of deployed nodes by 12–15% compared with heuristic baselines. Building on this, we propose discrete adaptations of Soft Actor-Critic (SAC) enhanced with transfer learning, achieving a 20% reduction in deployed nodes and reducing training time by 50% when adapting to new configurations.

Another contribution of our work is reformulating IAB planning as a graph decision process and introducing a graph-centric policy based on Graph Attention Network v2 (GATv2) with edge-conditioned attention. This enables resilience-aware deployment that maintains 87.1% coverage under 30% link failures (15.4% improvement over state-of-the-art) while reducing node requirements by 26.7%. The framework achieves linear computational complexity $\mathcal{O}(E \cdot d_h)$, making large scale deployments feasible. Finally, we discuss practical integration with Open Radio Access Network (O-RAN) via the Non-RT RIC, supporting deployability in next-generation disaggregated networks.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, Hamed Ahmadi and Paul Mitchell, for their continuous guidance, encouragement, and invaluable insights throughout my doctoral journey. Their expertise and patience have been instrumental in shaping both this thesis and my development as a researcher.

My heartfelt thanks also go to my colleagues and friends in the ISA and iTwins group. The countless discussions, collaborations, and shared struggles have not only improved my research but also made these years enjoyable and memorable. I would also like to acknowledge my own perseverance during this journey. The countless late nights, long days, and moments of doubt tested me deeply, yet they also taught me resilience, focus, and dedication. This thesis is a reflection not only of academic collaboration but also of the silent hours of effort and determination invested along the way. Keep calm and carry on!

On a personal note, I owe an immeasurable debt to my family. To my parents, thank you for your endless love, sacrifice, and encouragement—you have always been my strongest foundation. To my brother Wenlong Zhang, thank you for reminding me of the world beyond research and for standing by me when I needed it most.

Finally, to everyone who has walked alongside me in this long journey—whether through a kind word, a shared meal, or a moment of understanding—thank you. This thesis is as much yours as it is mine.

Jie Zhang,
York,
09 2025.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Publications arising from this thesis are acknowledged in the following section, entitled *Related Publications*.

Related publications

Published work

1. **Jie Zhang**, Swarna Bindu Chetty, Qiao Wang, Chenrui Sun, Paul Daniel Mitchell, David Grace, and Hamed Ahmadi. *Optimizing 6G Dense Network Deployment for the Metaverse Using Deep Reinforcement Learning.* Proc. IEEE INFOCOM 2025. **(Best Student Paper Award)**

2. **Jie Zhang**, Qiao Wang, Paul Daniel Mitchell, and Hamed Ahmadi. *An Integrated Access and Backhaul Approach to Sustainable Dense Small Cell Network Planning. Information*, 2023.

3. Chenrui Sun, Swarna Bindu Chetty, **Jie Zhang**, David Grace, and Hamed Ahmadi. *Energy Consumption Reduction for UAV Trajectory Training: A Transfer Learning Approach.* Proc. IEEE WCNC 2025.

4. **Jie Zhang**, Mostafa Rahmani Ghourtani, Swarna Bindu Chetty, Paul Daniel Mitchell, and Hamed Ahmadi. *Graph Neural Networks Approach for Joint IAB Network Planning and Resilience Challenges.* Proc. IEEE GLOBECOM 2025.

5. **Jie Zhang**, Swarna Bindu Chetty, Paul Daniel Mitchell, Mostafa Rahmani Ghourtani, Mohammad Shojafar, and Hamed Ahmadi. *Advanced Transferable Deep Reinforcement Learning for Multi-Hop Integrated Access and Backhaul Network Planning. IEEE Access*, 2026.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Background and Motivation

The telecommunications industry stands at a critical juncture as it transitions from 5G to 6G networks, driven by an unprecedented surge in mobile data traffic that is projected to reach 5,016 exabytes per month by 2030, representing a 10-fold increase from 2023 levels driven by bandwidth-intensive applications such as augmented reality, autonomous vehicles, and the metaverse [1]. Traditional approaches relying exclusively on fiber backhaul connections become economically prohibitive in dense urban environments, with installation costs ranging from $20,000 to $150,000 per kilometer [2].

Integrated Access and Backhaul (IAB) technology, standardized by the 3rd Generation Partnership Project (3GPP) in Release 16, emerges as a transformative solution to this deployment challenge [3]. By enabling base stations to simultaneously provide user access and wireless backhaul connectivity, IAB greatly reduces the need for fibre infrastructure while accelerating network densification. Specifically, IAB networks consist of two primary components: IAB-donor nodes, which maintain connectivity to the 5G core network via wired connections, and IAB-nodes, which act as wireless relay stations that extend coverage through multi-hop backhaul links.

This architectural design allows operators to flexibly deploy dense small-cell networks in urban environments without the prohibitive cost of extensive fiber rollout, while still ensuring reliable end-to-end connectivity. Field trials in Chicago's Lincoln Park have demonstrated that IAB can achieve up to 68× improvement in edge user throughput and 395× improvement in median user throughput compared to macro-only deployments, while requiring fiber connections for only 10–20% of base stations [4]. The adoption of mmWave frequencies provides the substantial bandwidth necessary for multi-gigabit backhaul links. However, mmWave propagation presents unique challenges that fundamentally alter network planning paradigms. Moreover, the quasi-optical propagation characteristics demand strict Line of Sight (LoS) alignment, with even human body blockage causing signal drops [5]. These physical constraints transform IAB network planning into a complex multi-dimensional optimization problem that must simultaneously balance coverage maximization, deployment cost minimization, capacity provisioning, and resilience assurance. Unlike traditional cellular planning where base stations operate independently, IAB networks exhibit intricate interdependencies through multi-hop backhaul paths, creating cascading effects where a single node failure can isolate entire network segments. This interconnected nature, combined with the vulnerability of mmWave links to environmental disruptions, elevates resilience from a desirable feature to a fundamental design requirement.

Despite these promising capabilities, the transition from IAB proof-of-concept to city-scale deployment reveals fundamental gaps between theoretical potential and practical implementation. The very characteristics

that make IAB attractive—wireless backhaul flexibility and mmWave capacity—introduce unprecedented planning complexities that traditional optimization frameworks cannot address. This paradox motivates our investigation into why existing methodologies fail and how intelligent approaches can bridge this gap.

1.2 Research Challenges and Objectives

The deployment of IAB networks at scale confronts some fundamental challenges that existing methodologies inadequately address, necessitating a comprehensive research framework to transform network planning from computationally intractable problems into intelligent, adaptive optimization solutions:

Computational complexity represents the primary barrier to practical deployment. IAB network planning constitutes a NP-hard complexity as the planning problem must jointly determine: (i) discrete node placement decisions grows exponentially with the number of candidate locations. For example, a modest urban scenario with 400 candidate locations and 20 required nodes, this translates to approximately 10^{38} possible combinations.; (ii) multi-hop topology formation respecting LoS constraint at mmWave frequencies [6].(iii) resource allocation across access and backhaul links with coupled capacity constraints where backhaul capacity directly limits access provisioning. The resulting formulation yields a complex non-convex problem, where the non-convex feasible region emerges from bilinear terms in flow conservation constraints and the discrete nature of placement decisions

Dynamic Adaptability challenges arise from constantly evolving envi-

ronments with temporal and spatial variations. In IAB networks, donor nodes require wired connectivity to the core network, meaning their deployment is constrained by the spatial availability of fiber infrastructure. Consequently, the set of feasible IAB donor and node locations can vary significantly between different deployment scenarios—for example, urban, suburban, or disaster-recovery environments—introducing additional spatial dynamics into the planning problem. Traditional static planning approaches fail to capture these dynamics, resulting in networks that perform poorly under real-world variations. The challenge extends beyond re-optimization, requiring learning mechanisms that generalize across configurations while maintaining computational efficiency. The inability to adapt dynamically becomes particularly critical when considering the fragility of mmWave links, leading to our third challenge.

Resilience requirements have evolved from optimization objectives to critical constraints due to mmWave link vulnerability. Field measurements reveal blockage events with mean inter-arrival time $\tau_b \in [10, 20]$s in urban environments, where single link failures can trigger cascading outages affecting 30–40% of network nodes through dependency chains [7]. The fundamental challenge lies in designing topologies that satisfy conflicting objectives: meeting a target coverage ratio through m-connected subgraphs ($m \geq 2$) for fault tolerance while minimizing deployed number of nodes. This fundamentally constraining the achievable resilience-cost trade-off in practical deployments.

To address these challenges, this thesis pursues a progressive research agenda with four primary objectives. First, we establish comprehensive baseline optimization frameworks that capture essential IAB deployment con-

straints, including mmWave propagation characteristics and multi-hop backhaul requirements, providing performance benchmarks and problem structure insights. Second, we develop learning-based algorithms leveraging DRL techniques to transform NP-hard optimization into tractable learning tasks through specialized state representations, constraint-aware exploration, and efficient training mechanisms capable of handling hundreds of candidate locations. Third, we enable rapid adaptation through transfer learning frameworks that allow pre-trained models to quickly adapt to new deployment scenarios with no, reducing computational overhead by 50% or more while maintaining near-optimal performance through transferable feature representations. Fourth, we achieve resilience-aware scalability by leveraging the inherent graph structure of IAB networks through Graph Neural Network (GNN) architectures that capture spatial dependencies and multi-hop relationships, reducing computational complexity from exponential to linear scaling while incorporating resilience as a first-class design constraint.

1.3 Structure and Main Contributions of thesis

This thesis presents a comprehensive framework for intelligent IAB network planning, progressing from traditional optimization methods to advanced machine learning approaches. The research addresses fundamental challenges in mmWave network deployment, including computational scalability, dynamic adaptability, and resilience requirements for next-generation wireless systems.

Chapter 2 establishes the theoretical foundations through a systematic literature review that adopts a comprehensive approach to exploring IAB network planning methodologies for 6G wireless networks, focusing on the evolution from standardization to practical deployment. The review employs strategic keyword searches using carefully selected terms across reputable sources, with citation tracking from technical specifications and high-impact field trials ensuring comprehensive coverage of both standardization efforts and cutting-edge research. Through critical analysis and synthesis, we establish a comprehensive framework spanning from protocol-level specifications to advanced optimization techniques, bridging the gap between IAB standardization and practical deployment by systematically analyzing architecture components, topology trade-offs, and planning methodologies. The synthesis reveals the evolution from traditional optimization approaches to learning-based methods, while identifying critical research gaps in scalability, resilience, and dynamic adaptation. Strategic visual representations including architectural diagrams, deployment parameter tables, and performance comparison matrices enhance comprehension by illustrating complex protocol stacks, providing deployment-ready configurations derived from [8], and quantifying topology trade-offs essential for practical IAB network planning.

Chapter 3 develops baseline optimization methodologies through heuristic-based planning frameworks. We formulate the IAB deployment problem into MILP to achieve optimal single-hop network architectures in dense donor scenarios, providing a robust mathematical foundation for addressing node placement challenges. For IAB donor distributions characteristic of realistic deployments, we introduce a greedy algorithm that effectively handles

the non-convex, NP-hard nature of multi-hop problems. A key innovation is our data rate constraint formulation that replaces traditional fixed hop-count limitations with dynamic capacity-based constraints, validated through comprehensive 60 GHz link budget analysis ensuring practical deployment feasibility. Through 100 simulation runs, we establish performance benchmarks that serve as baselines for evaluating the advanced optimization techniques developed in subsequent chapters.

Chapter 4 pioneers the application of DRL to discrete IAB deployment problems through an advanced framework utilizing DQN, Double Deep Q-Network (DDQN), and Dueling DQN architectures. We develop specialized state and action representations tailored for ultra-dense environments with up to 400 candidate locations, incorporating innovative action elimination strategies that reduce the search space by 60-80% through constraint-aware pruning of infeasible deployments. The Markov Decision Process (MDP) formulation effectively balances meeting a target coverage ratio and deployment minimization objectives in complex multi-hop scenarios. Experimental results demonstrate 12-15% reduction in required nodes compared to heuristic baselines while maintaining 98% coverage targets, validating the superiority of learning-based approaches for adaptive network planning.

Chapter 5 addresses computational efficiency challenges through the integration of advanced Reinforcement Learning (RL) algorithms with TL methodologies. We present the first adaptation of continuous control algorithms, specifically SAC and TD3, to discrete IAB planning domains, providing rigorous convergence guarantees with polynomial sample complexity of $\mathcal{O}(|S||A|/\varepsilon^2)$. The TL framework enables rapid adaptation to new

donor configurations, achieving 50% training time reduction while maintaining deployment efficiency through knowledge transfer from pre-trained models. The entropy-regularized exploration mechanism achieves 20% node reduction compared to baseline approaches, demonstrating the effectiveness of maximum entropy RL in high-dimensional discrete action spaces.

Chapter 6 leverages the inherent graph structure of IAB networks to achieve scalable resilience-aware optimization through GNN-based approaches. We formulate a resilience-aware MDP with resilience constraints integrated through penalty-based reward design, ensuring network robustness under failure conditions. The edge-conditioned GATv2 framework processes heterogeneous node types and dynamic link utilization patterns, dramatically reducing computational complexity from exponential $\mathcal{O}(2^N)$ to linear $\mathcal{O}(E \cdot d_h)$. Extensive evaluation demonstrates 87.1% coverage retention under 30% link failures—a 15.4% improvement over state-of-the-art methods—while achieving 26.7% reduction in the number of deployed IAB nodes across diverse urban scenarios. This graph-centric approach represents the culmination of our progressive methodology, combining theoretical rigor with practical scalability.

In summary, this thesis advances the state-of-the-art in IAB network planning through a systematic progression from traditional optimization to advanced machine learning techniques. Our theoretical contributions include formal MDP formulations with proven convergence guarantees and comprehensive complexity analysis. Algorithmic innovations encompass discrete adaptations of continuous RL algorithms, TL frameworks for rapid adaptation, and graph-based resilience modeling. Practical implementations

demonstrate significant performance improvements across multiple conditions. These contributions collectively enable intelligent, adaptive, and resilient IAB deployments essential for next-generation wireless networks, providing a complete framework that bridges theoretical foundations with practical deployment requirements.

2 Literature Review

This chapter establishes the groundwork for understanding the methodologies and technologies explored in this thesis, with a specific focus on IAB and network planning strategies for 6G wireless networks. It begins by introducing the core principles of IAB, emphasizing its significance in enabling cost-effective network densification through wireless backhaul solutions. Next, the chapter delves into optimization techniques tailored for IAB network planning, especially machine learning-driven, highlighting their capability to address dynamic network conditions and complex planning challenges. Following this, it examines topology design, focusing on the trade-offs between Spanning Tree and Directed Acyclic Graph configurations to enhance resilience and capacity in IAB deployments. Finally, the chapter explores resource allocation strategies, specifically designed to support IAB network planning by optimizing interference management and ensuring quality of service in multi-hop IAB environments. Collectively, these concepts establish the groundwork for the innovative, learning-based IAB network planning approaches developed in subsequent chapters.

2.1 Introduction of Integrated Access and Backhaul

IAB represents a key paradigm in 5G New Radio (NR) architecture, standardized by the 3GPP in Release 16 [3]. IAB technology addresses the fundamental challenge of cost-effective network densification by enabling wireless backhaul connectivity between base stations, particularly crucial for mmWave deployments where fiber infrastructure is economically prohibitive. Some of the more promising recent studies have shown that, under typical deployment scenarios, IAB can yield substantial performance gains—for example, improvements of up to 68× in edge user throughput and 395× in median user throughput compared to macro-only deployments [4]. Moreover, field experiments in Lincoln Park, Chicago, have demonstrated that IAB can extend coverage with only a limited number of fiber-connected donor base stations while still meeting Quality of Service (QoS) requirements [9].

Building on this donor–relay architecture, IAB enables rapid network deployment with reduced the need for fibre infrastructure while maintaining 5G performance requirements for Enhanced Mobile Broadband (eMBB), Ultra-Reliable Low-Latency Communications (uRLLC), and Massive Machine-Type Communications (mMTC) services. Unlike traditional base stations, IAB enables massive beamforming and uses mmWave technology to provide inexpensive wireless backhaul with substantial bandwidth capacity, supporting the "plug and play" design that allows dense and flexible deployment [10]. These performance advantages of IAB technology stem from its unique architectural design, which fundamentally differs from traditional cellular deploy-

ments. To develop effective planning strategies, it is essential to understand the underlying architectural components and protocol stack organization that enable multi-hop wireless connectivity.

2.2 IAB Architecture and Protocol Stack

2.2.1 IAB-Donor Components

The IAB-donor serves as the central control entity and traffic convergence point between the access network and 5G core. The donor implements a distributed architecture with clear separation of control and user plane functions as shown in 2.1. The IAB-Donor Central Unit (CU) hosts upper protocol layers including Packet Data Convergence Protocol (PDCP), Service Data Adaptation Protocol (SDAP), and Radio Resource Control (RRC). The CU further separates into CU-Control Plane (CU-CP) for signaling functions and CU-User Plane (CU-UP) for data forwarding optimization. Meanwhile, the IAB-Donor Distributed Unit (DU) manages lower protocol layers including Physical Layer (PHY), Medium Access Control (MAC), and Radio Link Control (RLC) layers, providing wireless backhaul connectivity to IAB-nodes within its radio coverage area.

The donor connects to the 5G core via the Next Generation (NG) interface and communicates with distributed units through the F1 interface, which separates control plane (F1-C) and user plane (F1-U) traffic for enhanced scalability and performance [11]. This split architecture enables coordinated performance features including load management and real-time performance optimization by placing protocols requiring high real-time performance in the

DU while placing PDCP and RRC layers that do not require high real-time performance in the CU. The baseband unit functionality from conventional base stations is effectively separated into two parts: the PDCP and RRC layers are placed in the CU, while the physical layer, MAC and RLC layers are positioned in the DU [12].

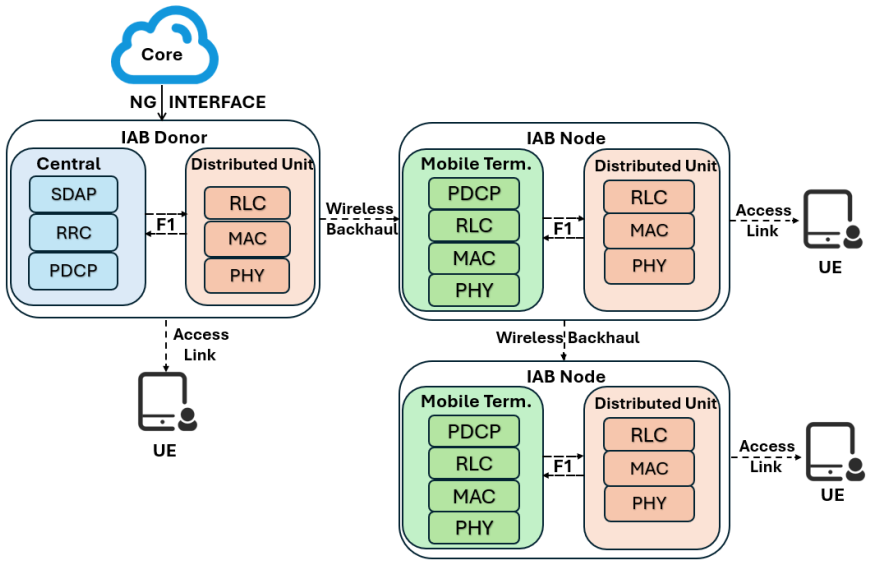

Figure 2.1: IAB network architecture showing protocol stack organization and Distributed Unit connects to other Distributed Unit via wireless backhaul.

2.2.2 IAB-Node Architecture

IAB-nodes function as Layer-2 regenerative relays, implementing a dual-functionality architecture that enables both access and backhaul services. The IAB-Mobile Termination (MT) provides User Equipment (UE) functionality enabling connection to parent nodes via the NR air interface. The MT handles upstream communication and appears as a standard UE to the serving parent node, establishing backhaul connections toward parent IAB-nodes or IAB-donors through RRC connection establishment procedures similar to

normal user equipment [4]. Simultaneously, the IAB-DU implements Next Generation NodeB (gNB)-DU functionality to serve downstream UEs and child IAB-nodes. The DU operates transparently to served UEs, appearing as a conventional base station while providing access connections to ordinary mobile users or downstream connections to child IAB-nodes.

This architecture enables decode-and-forward relaying where packets are fully decoded and re-encoded at each hop, providing error correction and signal regeneration while maintaining end-to-end QoS requirements. The F1 interface between CU and DU is standardized with F1's control plane (F1-C) responsible for signaling between the CU and DU, while F1's user plane (F1-U) handles the transmission of application data. The lower three protocols collectively form the NR-Uu interface, which connects the UE to the DU over the air [13]. This separation of CU and DU functionalities not only enhances scalability but also defines the constraints under which backhaul-aware network planning in this thesis is formulated.

2.2.3 Protocol Stack Organization

The IAB protocol stack enhances the standard 5G NR protocol stack with the Backhaul Adaptation Protocol (BAP) sublayer, which serves as the key enabler for multi-hop transmission capabilities across the IAB network [14]. Access links utilize the standard NR protocol stack (PHY/MAC/RLC/PDCP/RRC) for UE connectivity, while backhaul links employ enhanced NR protocols with BAP sublayer positioned above RLC for routing functionality. The F1 interface maintains logical separation between IAB-node-DU and IAB-donor-CU with dedicated control and user plane connections.

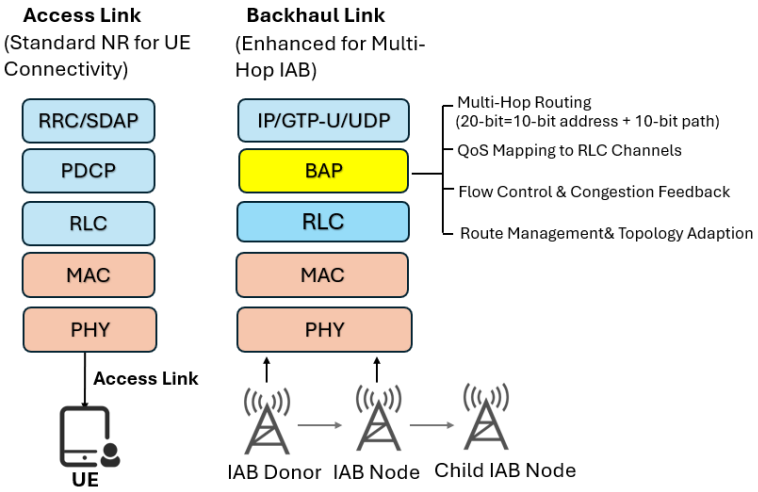

Figure 2.2: IAB protocol stack organization showing the integration of BAP sublayer for multi-hop routing. The diagram illustrates the layered architecture with BAP positioned above RLC in backhaul links, enabling hop-by-hop packet forwarding from IAB-donor to destination nodes.

The BAP operates as a specialized Layer-2 routing protocol that enables hop-by-hop packet forwarding from IAB-donor to destination nodes using BAP routing identifiers [15]. This multi-hop routing capability utilizes 20-bit identifiers consisting of a 10-bit BAP address and 10-bit BAP path identifier, enabling efficient packet forwarding across complex network topologies [13]. BAP supports dynamic routing table updates and provides QoS mapping functionality to maintain service differentiation throughout the multi-hop path, ensuring that different traffic types receive appropriate treatment as they traverse multiple IAB nodes.

This protocol organization ensures seamless integration with existing 5G infrastructure while providing specialized backhaul capabilities essential for multi-hop network operation as shown in 2.2. Any IP traffic can be forwarded

through the BAP, including F1 interface traffic and network management functions, providing comprehensive support for various services across the multi-hop topology.

2.2.4 IAB Resource Management and Parameters

Beside the BAP protocol, IAB systems require careful configuration of resource multiplexing strategies to support multi-hop transmission. Time Division Multiplexing (TDM) serves as the primary approach in 3GPP Release 16 with dynamic resource allocation, while Frequency Division Multiplexing (FDM) enables simultaneous access and backhaul operation in spectrum-rich deployments [10]. Spatial Division Multiplexing (SDM) leverages beamforming for spatial separation in complex interference environments.

Building on the evaluation methodology introduced in 3GPP TR 38.874, we consider the three canonical deployment classes—Urban Macro, Urban Micro and Indoor Hotspot and apply their multi-hop extensions for IAB. The configuration set in Table 2.1 is restricted to values that can be traced to peer-reviewed literature, public field-trial reports or openly available 3GPP-compliant tools. These standardized deployment parameters are not only important for aligning with 3GPP-compliant assumptions but also form the baseline configuration used in the simulation and optimization studies presented in this thesis.

Table 2.1: IAB Deployment Parameters by 3GPP Scenario Classification

Scenario	Parameter	Value	Reference
Urban Macro	**Power Configuration**		
	IAB-Donor Tx Power	40 dBm	[16]
	IAB-Node Tx Power	33 dBm	[16]
	UE Tx Power	23 dBm	[17]
	Radio Configuration		
	Frequency Band	28/39 GHz	[18]
	Channel Bandwidth	100 MHz	[18]
	Network Topology		
	Maximum Hops	2–4	[19]
Urban Micro	**Power Configuration**		
	IAB-Donor Tx Power	30 dBm	[20]
	IAB-Node Tx Power	26 dBm	[20]
	UE Tx Power	23 dBm	[20]
	Radio Configuration		
	Frequency Band	28 GHz	[18]
	Channel Bandwidth	100 MHz	[18]
Indoor Hotspot	**Power Configuration**		
	IAB-Donor Tx Power	21 dBm	[21]
	IAB-Node Tx Power	21 dBm	[21]
	UE Tx Power	23 dBm	[21]
	Radio Configuration		
	Frequency Band	4 GHz	[21]
	Channel Bandwidth	10 MHz	[21]

2.2.5 IAB Network Topologies

In IAB networks, the topology significantly influences both performance and the complexity of network planning. The 3GPP has standardized two primary topologies: the Spanning Tree (ST) and the Directed Acyclic Graph (DAG). These configurations, detailed in 3GPP Release 16 and 17 respectively [3], offer distinct characteristics that shape planning strategies, optimization algorithms, and deployment outcomes. This section introduces the foundational aspects of ST and DAG topologies before exploring their specific implications for network planning.

Spanning Tree Topology

The ST topology, introduced in 3GPP Release 16 [3], employs a hierarchical structure wherein each IAB-node connects to a single parent node. This single-path design simplifies network implementation and management, making it an attractive choice for initial 5G deployments where rapid rollout is a priority. The inherent simplicity of ST reduces the complexity of planning algorithms to linear time, typically $O(|V|)$, where $|V|$ denotes the number of nodes [22]. Such efficiency supports centralized optimization techniques, such as Mixed Integer Linear Programming (MILP), which can determine globally optimal solutions for node placement and resource allocation in moderately sized networks [23]. Despite these advantages, the ST topology's reliance on single-path connectivity introduces notable limitations. A key vulnerability is its susceptibility to single points of failure; the failure of an intermediate node disrupts service for all downstream nodes. Research by Islam et al. [24] suggests that while ST topologies achieve coverage probabilities of 92–95%,

they often require 30–40% more infrastructure than multi-path alternatives to ensure reliability. Consequently, planners must weigh the benefits of simplicity against the need for additional redundancy to maintain robust network performance.

Directed Acyclic Graph Topology

The DAG topology, enhanced in 3GPP Release 17 [3], allows each IAB-node to connect to multiple parent nodes, improving resilience and load distribution. This multi-path capability, however, increases the complexity of network planning. Optimization algorithms for DAG topologies must address the combinatorial nature of multiple routing paths, resulting in computational complexity that scales quadratically, i.e., $O(N^2)$, for optimal solutions[25]. This complexity often necessitates distributed planning approaches, which are better suited to manage the inter-node coordination required for resource allocation across multiple paths [26]. Moreover, the presence of redundant paths demands sophisticated interference management, often employing advanced beamforming and resource partitioning techniques [23].

The performance benefits of DAG topologies are significant. Studies by Simsek et al. [23] report up to a 26% improvement in overall network performance compared to ST configurations. Additionally, worst-case user rates improve by factors of 3 to 5 in dual-parent setups [22], making DAG topologies particularly suitable for dense urban environments where capacity and reliability are critical. These advantages, however, come at the cost of increased planning effort to balance performance gains with operational com-

plexity.

Table 2.2: Topology Performance Comparison

Metric	Spanning Tree	DAG (Pmax=2)	DAG (Pmax=3)	Reference
Network Throughput	Baseline	+20-26%	+25-30%	[23]
Worst-case User Rate	Baseline	3-5× improvement	4-6× improvement	[22]
Planning Complexity	O(V)	$O(V^2)$	$O(V^3)$	[25]
Failure Resilience	Low	High	Very High	[27]
Deployment Cost	Baseline	+15-20%	+25-35%	[28]

The choice between ST and DAG topologies profoundly impacts network planning. Table 2.2 collates the main findings reported in recent literature and field trials. The ST column serves as the single-parent baseline, whereas the two DAG columns illustrate results when each child may associate with up to two or three parents ($P_{\max} = 2, 3$). For ST topologies, planning focuses on hierarchical node placement to minimize backhaul distance to donor nodes while ensuring coverage. Genetic algorithms have proven effective, achieving near-optimal solutions with coverage rates up to 92% using greedy heuristics [22]. In contrast, DAG topologies require multi-objective optimization to balance coverage, capacity, and interference. Advances in GNN have shown 12–18% performance improvements over traditional methods in DAG planning [29]. Consequently, deployment goals also dictate topology selection. ST topologies suit initial or rural deployments due to their simplicity, while DAG topology are better suited for urban or mission-critical scenarios requiring enhanced resilience and performance.

2.3 Optimization Techniques

IAB planning couples IAB node placement, topology design, and resource scheduling. The joint problem can be formulated as a multi-commodity flow

with discrete placement variables; even under convex flow costs, the presence of binary decision variables renders it NP-hard. Optimization techniques for IAB network planning can be broadly categorized into three classes: traditional mathematical optimization, learning-based methods, and hybrid approaches. Each category offers distinct advantages and limitations, making them suitable for different deployment scenarios and planning objectives. Traditional methods, including Mixed-Integer Programming (MIP), and heuristics, provide foundational solutions for static or theoretical scenarios. Learning-based methods, such as DRL, GNN, and transfer learning, leverage machine learning for dynamic adaptability. Hybrid approaches integrate multiple paradigms to balance theoretical rigor with practical flexibility. The following subsections detail these methods, their applicable scenarios, limitations, and comparative analysis.

2.3.1 Traditional Optimization Methods

Mixed-integer programming approaches have formed the foundation of early IAB network planning research, providing globally optimal solutions for small to medium-scale deployment scenarios. Polese et al. [30] formulated IAB deployment as a comprehensive MIP problem, achieving global optimality for networks with fewer than 50 candidate locations while maintaining tractable computational complexity with exponential complexity $\mathcal{O}(2^N)$ that becomes intractable for realistic urban scenarios. The MIP formulation enables precise modeling of deployment constraints including interference relationships, capacity limitations, and regulatory requirements, making it particularly suitable for scenarios where optimal solutions are critical and computational

resources are available. Recent advances in MIP approaches include the integration of link budget analysis at 60 GHz to ensure feasibility of achieving reasonable data rates, as demonstrated in single-hop network optimization for scenarios with dense donor distribution [31].

Heuristic algorithms have emerged as practical alternatives to MIP approaches when dealing with large-scale networks where optimal solutions become computationally intractable. [32] proposed a greedy heuristic achieving 92-95% coverage in realistic deployment scenarios, though requiring 30-40% more infrastructure than optimal solutions due to myopic decision-making that neglects long-term optimization objectives and resilience considerations. Advanced heuristic methods [31] have evolved to address multi-hop network problems in scenarios with sparse donor distribution by Gurobi solver, with modified greedy algorithms demonstrating efficacy in handling the non-trivial, NP-hard nature of multi-hop IAB deployments while incorporating data rate constraints that depart from traditional fixed hop number limitations. While traditional optimization methods provide theoretical foundations and guaranteed optimality for small-scale problems, the increasing complexity and dynamic nature of modern IAB deployments have driven researchers toward adaptive, learning-based approaches that can handle uncertainty and scale effectively

2.3.2 Learning-Based Optimization Methods

Deep reinforcement learning has revolutionized IAB network planning by enabling adaptive optimization in dynamic environments where traditional approaches struggle with computational complexity and changing conditions.

DQN-based approaches have demonstrated particular effectiveness in discrete decision-making scenarios typical of node placement problems. Zhang et al.[33] applied DQN/Double/Dueling DQN for IAB node deployment in dense 6G networks, addressing discrete action spaces effectively for node placement optimization.

Advanced DRL architectures including TD3, and SAC have addressed limitations of basic DQN approaches in handling large state and action spaces characteristic of dense network planning scenarios. [34] formulate a joint site-selection and height-placement problem for tethered UAV relays in mmWave IAB and this work has extended these approaches to continuous action spaces with effective discretization of continuous outputs for actionable network deployment decisions.

Graph neural networks have emerged as powerful tools for modeling the inherent graph structure of IAB networks, where nodes and connections form natural graph representations. Yang et al. [29] developed GNN-based network optimization achieving 12-18% improvement over traditional heuristics through message-passing and node features including location and signal strength, though lacking resilience constraints and heterogeneous node modeling capabilities. Liu et al. [35] surveyed GNN applications in network management, focusing on virtual function orchestration that provides insights applicable to physical infrastructure deployment. Recent advances include multi-head GNN architectures for joint optimization problems and heterogeneous GNN models specifically designed for mmWave networks that capture the unique propagation characteristics and deployment constraints. Recognizing that neither traditional nor learning-based methods alone can address

all IAB planning challenges, recent research has explored hybrid methodologies that combine the theoretical rigor of mathematical optimization with the adaptability of machine learning techniques.

2.3.3 Hybrid Optimization Approaches

Hybrid methodologies that combine multiple optimization paradigms have emerged to address the limitations of individual approaches while leveraging their respective strengths. The integration of stochastic geometry with DRL enables networks to benefit from the mathematical rigor of geometric analysis while adapting to dynamic conditions through learning-based optimization[36]. These approaches typically employ stochastic geometry for initial network design and deployment planning, followed by DRL-based fine-tuning for operational optimization. Recent implementations have demonstrated the effectiveness of hybrid approaches in balancing theoretical foundations with practical adaptability requirements.

Graph-based approaches combined with reinforcement learning have shown particular promise in IAB topology optimization. Simsek et al. [23] developed a graph embedding and deep reinforcement learning approach for IAB topology design, demonstrating superior performance compared to individual methodologies. The hybrid approach leverages graph embedding to capture network structural relationships while utilizing reinforcement learning for dynamic adaptation to changing network conditions.

The optimization techniques presented establish a comprehensive toolkit for IAB network planning, evolving from traditional MIP approaches with optimality guarantees but limited scalability, to practical heuristics achiev-

ing near-optimal coverage, and ultimately to adaptive learning-based methods utilizing DRL and GNN for dynamic environments. Hybrid approaches combining multiple paradigms represent the current frontier, balancing theoretical rigor with practical flexibility. These diverse methodologies, each with distinct advantages and applicable scenarios, provide the algorithmic foundation for tackling real-world deployment challenges. The following section examines how these optimization tools have been applied in state-of-the-art research across various planning objectives, demonstrating their practical effectiveness in addressing coverage, capacity, QoS, and multi-objective optimization challenges in 5G and beyond networks.

2.4 IAB Network Planning and Deployment: State-of-the-Art Research

IAB network planning has emerged as a critical research domain driven by increasing demand for cost-effective network densification and the proliferation of mmWave technologies in 5G and beyond networks. The complexity of IAB deployment stems from the unique characteristics of wireless backhaul links, stringent LoS requirements, and the need to balance coverage, capacity, and the number of deployed nodes.

Research efforts have evolved significantly over time. Early approaches focused on traditional optimization methods, while recent advances have incorporated sophisticated learning-based methodologies. These developments address diverse deployment scenarios ranging from dense urban environments to rural coverage extension. The scope of IAB network planning encompasses

multiple interconnected challenges. These include node placement optimization, topology design, resource allocation, and resilience management. This section categorizes the planning objectives into three key themes: coverage optimization, QoS assurance, and multi-objective planning.

Due to the relatively limited number of studies, we also consider several complementary technologies that address similar objectives. These include relay nodes (particularly those operating in decode-and-forward mode) and small cell deployments. While distinct from IAB, these technologies share similar architectural goals and can serve as enablers or supplements to IAB-based deployments, especially in heterogeneous or high-density scenarios.

2.4.1 Coverage Optimization

IAB coverage optimization focuses primarily on strategic node placement algorithms that maximize service area while maintaining signal quality constraints across multi-hop wireless topologies. This area represents one of the most mature aspects of IAB research, yet significant challenges remain in addressing dynamic environmental factors.

Recent research demonstrates that genetic algorithm-based approaches achieve superior performance in IAB node placement. Madapatha *et al.* have developed efficient schemes for both node placement and backhaul link distribution, under various deployment constraints (e.g. antenna gain, blockage probability, geographical limitations) [32]. In their work, they report notable improvements in coverage and service reliability under constrained deployment scenarios. For instance, in some specific setups (e.g. with blockage-aware placements) reports notable improvements in coverage and service re-

liability under constrained deployments.

Complementary technologies have shown promising results in enhancing coverage objectives. Unmanned Aerial Vehicle (UAV)-based base stations provide dynamic coverage adaptation capabilities, achieving 50% reduction in communication failure rates and 58.3% round-trip time improvements compared to traditional terrestrial deployments [37]. This performance improvement is particularly valuable in emergency scenarios or temporary coverage extensions. However, existing coverage optimization research exhibits significant limitations. Most studies focus on static optimization problems that inadequately represent real-world deployment complexities. Critical gaps include insufficient consideration of temporal traffic variations, seasonal foliage changes affecting millimeter wave propagation, and adaptive interference management in dense deployment scenarios.

In summary, while coverage optimization in IAB networks has matured through diverse algorithmic strategies and complementary technologies, future research must move beyond static formulations to embrace adaptive, context-aware, and environmentally robust approaches that reflect real-world deployment dynamics.

2.4.2 Quality of Service Assurance

QoS assurance in IAB networks requires sophisticated management of end-to-end service quality across multi-hop wireless topologies. The challenge lies in maintaining service level agreements while accommodating diverse traffic characteristics and the inherent complexity of wireless backhaul links.

The hierarchical QoS framework represents a significant advancement in

IAB QoS management. This framework implements a two-stage downlink scheduler specifically designed for IAB networks [38]. Stage I performs access and backhaul link scheduling at individual base stations using Weighted Proportional Fair algorithms. Stage II manages traffic distribution across downstream links with congestion-aware allocation mechanisms. This approach achieves 10× improvement in 5th percentile throughput performance, with optimal performance occurring when child node limitation is set to 3 nodes per parent.

Advanced scheduling algorithms demonstrate substantial performance improvements through sophisticated resource allocation frameworks. The Coordinated Parallel Resource allocation (CPReal) framework achieves 67.3% throughput gain compared to traditional scheduling under bursty traffic conditions while reducing average end-to-end delay by 34.4% [39]. This improvement is particularly significant for applications requiring consistent quality of service.

Multi-hop latency characteristics present both challenges and opportunities for optimization. Research shows that latency exhibits linear increases with hop count, adding 4–6ms per network tier [40]. However, advanced scheduling techniques can mitigate these delays. Optimized TDM patterns and proactive bandwidth reporting achieve latency reductions of approximately 5ms through micro-phase adjustments. These optimizations enable sub-2ms delay requirements for Ultra-Reliable Low-Latency Communications (URLLC) services.

Reliability enhancement mechanisms focus on path redundancy and dynamic adaptation. Multi-connectivity support and dynamic traffic distribu-

tion across redundant paths provide robust service continuity [40]. Network coding techniques offer particularly promising results, providing proactive redundancy without retransmission overhead and achieving sub-second topology adaptation capabilities that maintain 99.999% service availability.

Cross-layer optimization approaches using Reinforcement Learning demonstrate significant performance improvements. These systems achieve 50% latency reduction while maintaining 99.9999% reliability [40]. The combination of intelligent routing decisions and adaptive resource allocation enables superior performance in dynamic network conditions.

However, existing QoS assurance research suffers from fundamental limitations. Critical gaps include insufficient scalability analysis for ultra-dense deployments, inadequate modeling of interference correlation across multi-hop paths, and limited consideration of energy-delay trade-offs in resource allocation decisions.

2.4.3 Multi-Objective Approaches

Multi-objective optimization in IAB networks is essential due to the inherent trade-offs among competing goals, such as maximizing coverage and capacity while minimizing deployment costs, energy consumption, and latency. While some earlier approaches have focused on single objectives (e.g., maximizing throughput or minimizing cost), such formulations only capture part of the deployment challenge. In practice, these objectives often conflict; for instance, enhancing coverage through denser node placement increases capacity but elevates costs and energy use, necessitating Pareto-optimal solutions that allow operators to select configurations based on priorities like economic

constraints or environmental impact. Several metaheuristic algorithms have been proposed to address these challenges in IAB network planning:

The Non-dominated Sorting Genetic Algorithm II (NSGA-II) is particularly effective for IAB optimization, as it efficiently generates diverse Pareto fronts for problems with non-convex objective spaces, making it suitable for handling the combinatorial nature of node placement and resource allocation in wireless environments. For example, a reinforcement learning-enhanced NSGA-II has been applied to optimize energy-delay trade-offs in mmWave IAB heterogeneous networks, achieving balanced solutions that reduce aggregate energy consumption and delay compared to single-objective methods [41].

Similarly, the Multi-Objective Particle Swarm Optimization (MOPSO) offers complementary strengths in dynamic scenarios, adapting to varying traffic and environmental conditions to produce well-distributed non-dominated solutions [42]. Other prominent algorithms include the Multi-Objective Evolutionary Algorithm based on Decomposition (Multiobjective Evolutionary Algorithm based on Decomposition (MOEA/D)), which decomposes the problem into scalar subproblems for scalable optimization in wireless routing [43], and the Strength Pareto Evolutionary Algorithm 2 (SPEA2), which uses fitness assignment based on dominance to improve diversity in sensor deployment and network lifetime extension [44].

Comparative analyses show NSGA-II outperforming MOEA/D in convergence for high-dimensional IAB problems, while SPEA2 excels in maintaining solution diversity under uncertainty. Economic considerations are integral to IAB multi-objective frameworks, often incorporated via Total Cost of Owner-

ship (TCO) models that account for Capital Expenditures (Capital Expenditure (CapEx)), including hardware and infrastructure, and Operational Expenditure (OpEx), such as energy and maintenance. Studies indicate that IAB can reduce TCO by up to 50% per cell site compared to traditional fixed backhaul, primarily by eliminating separate backhaul equipment and leveraging shared spectrum [45]. Industry estimates suggest 10–20% of 5G sites may adopt IAB for cost-effective deployment in challenging locations [46]. Hybrid approaches combining these algorithms further enhance performance by integrating exact methods like Mixed Integer Programming (MIP) for constraint handling, as seen in joint association and resource allocation models that maximize sum rates while respecting QoS constraints [47].

Performance evaluations demonstrate the efficacy of these approaches in real-world scenarios. For instance, multi-objective frameworks balancing data freshness and spectral efficiency in multi-hop IAB networks achieve significant throughput gains and reduced latency [48]. Energy efficiency optimizations, incorporating intelligent caching and bandwidth partitioning, yield improvements in energy-delay metrics [41]. Large-scale MIP-based deployments have shown near-optimal solutions with efficiency gains [49]. Despite these advances, multi-objective optimization in IAB networks faces limitations, including computational scalability for real-time decisions, inadequate handling of parameter uncertainties (e.g., blockages in mmWave), and insufficient integration of regulatory and site acquisition constraints [10]. Future research should focus on adaptive algorithms that incorporate machine learning for dynamic environments. A summary of key research contributions in IAB network planning is provided in Table 2.3, highlighting problems ad-

dressed, methods employed, performance metrics, and references.

2.5 Summary

This literature review establishes the foundational understanding for advancing IAB network planning methodologies. The analysis demonstrates that IAB technology offers transformative potential for cost-effective network densification, with field trials showing up to 68× improvement in edge user throughput compared to macro-only deployments [4]. The architectural examination reveals that BAP sublayer integration and the choice between ST and DAG topologies fundamentally impact network performance, with DAG configurations offering up to 26% performance improvements at increased optimization complexity.

The systematic survey reveals a clear evolution from traditional MIP formulations with theoretical optimality but exponential complexity $\mathcal{O}(2^N)$, through practical heuristic approaches achieving 92-95% coverage, to learning-based methods utilizing DRL and GNN for dynamic environments. State-of-the-art research demonstrates significant achievements: genetic algorithms achieve 200% capacity improvements, hierarchical QoS frameworks provide 10× throughput improvements, and multi-objective approaches reduce TCO by up to 50% per cell site. However, existing methods primarily address static scenarios and lack integration of scalability, adaptability, and resilience requirements for practical 6th Generation Mobile Networks (6G) deployments.

The convergence of IAB with mmWave frequencies introduces unique propagation challenges including path loss, atmospheric attenuation, and blockage susceptibility that fundamentally alter network planning paradigms.

Traditional cellular planning approaches are inadequate for IAB networks, where multi-hop backhaul dependencies create complex interdependencies requiring holistic optimization strategies.

Building on these foundations, this thesis develops a progressive optimization framework addressing the identified limitations. Beginning with baseline heuristic approaches to establish performance benchmarks, the research advances through intelligent learning-based methods that leverage the inherent structure of IAB networks to achieve scalable, adaptive, and resilient deployment strategies essential for next-generation wireless systems.

Table 2.3: Summary of IAB Network Planning Research

Problem	**Method/Model**	**Performance Metric**	**Reference**
Strategic node placement	Genetic algorithm	Up to 200% capacity improvement	[32]
Dynamic coverage adaptation	UAV-based base stations	50% reduction in failure rates, 58.3% RTT improvement	[37]
End-to-end QoS management	Hierarchical QoS framework	10× improvement in 5th percentile throughput	[38]
Resource allocation for QoS	CPReal framework	67.3% throughput gain, 34.4% delay reduction	[39]
Multi-hop latency and reliability	Advanced scheduling and network coding	5 ms latency reduction, 99.999% availability	[40]
Cross-layer QoS optimization	Reinforcement Learning	50% latency reduction, 99.9999% reliability	[40]
Multi-objective optimization	NSGA-II	Reduction in energy-delay aggregate	[41]
Dynamic multi-objective optimization	MOPSO	Well-distributed solutions	[42]
Economic analysis	TCO models	Up to 50% TCO reduction per cell site	[45]
Hybrid optimization	Hybrid approaches	Superior throughput and coverage	[47]
Large-scale deployment optimization	MIP with constraints	Near-optimal solutions, efficiency gain	[49]

3 Heuristic-Based IAB Network Planning

3.1 Introduction

In this chapter, IAB will provide cost-effective, high-bandwidth backhaul solutions by employing mmWave technology, making it a sustainable option for next-generation networks [50]. Figure 3.1 illustrates the IAB architecture from the definition of 3GPP Release-16 [51], comprising IAB-donors and IAB-nodes. IAB-donors connect to the core network via fiber, offering wireless access to mobile users and backhaul to other IAB-nodes, while IAB-nodes extend connectivity by providing both access and backhaul services wirelessly. This dual functionality will enable dense, flexible deployments, reducing the environmental impact compared to traditional infrastructure-heavy approaches.

The motivation for this chapter stems from the limitations of existing IAB network planning models, which often oversimplify location constraints. High-frequency mmWave bands, while enhancing backhaul performance, introduce significant pathloss and require precise node placement under LoS conditions. Furthermore, IAB-nodes must be deployed at predefined candidate locations, constrained by physical, regulatory, and performance factors. Traditional linear programming approaches fail to adequately address these

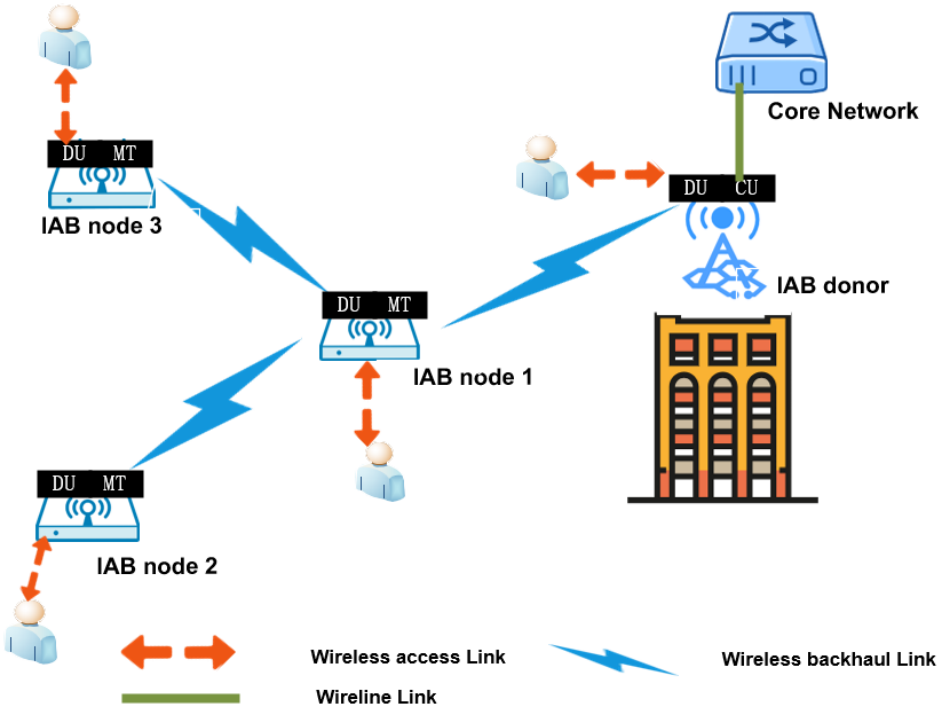

Figure 3.1: Integrated Access and Backhaul (IAB) architecture. The central unit (CU) in the IAB-donor connects to the core network, while its distributed unit (DU) and the DUs in IAB-nodes serve user equipment (UEs) and other nodes. Mobile terminals (MTs) link with the DU of a parent node or donor.

complexities, prompting the development of a location aware, heuristic-based optimization methodology in this study. Our work aims to minimize the number of IAB-nodes while ensuring coverage and backhaul requirements, contributing to the overarching thesis goal of designing sustainable and efficient 5G networks.

3.2 System Model

We consider an IAB network consisting of IAB nodes and donors, operating in the 60 GHz license-exempt frequency band. The system is designed to achieve

fiber-like data rates, reaching up to 1 Gbit/s. We specifically focus on a city-based deployment, incorporating a wireless mesh network, as illustrated in Figure 3.2.

Figure 3.2: A sample mmWave integrated access and backhaul (IAB) network model with directional antenna.

The candidate sites for IAB node deployment include urban infrastructures, such as lamp-posts and bus-stops. These locations are selected in a way to facilitate LoS transmission for user access whenever possible. Our design faces several challenges, particularly in identifying optimal installation locations for each of the IAB nodes while minimizing the total number of nodes. While potential multi-hop scenarios pose some limitations, this is

further compounded by the need to ensure that the data transfer rates are consistently served.

To guide our discussion, we outline the following key assumptions governing the system model:

- All nodes/donors are unified: We introduce this assumption to streamline the network design and analytical processes. Each node and donor is characterized by a uniform access and backhaul radius. An area is considered covered if it lies within the access radius of a deployed node or donor. Similarly, a node is deemed serviceable by a donor if it is located within the donor's backhaul radius. This consistency and standardiation, together with the requirement that all nodes and donors are at the same height, eliminates deviations that may muddy network performance evaluations, allowing us to focus on optimising deployment sites.

- LoS transmission is consistently maintained: Operating in the 60 GHz band necessitates a focus on ensuring LoS connections, owing to the mmWave characteristics, which are highly susceptible to blockages and attenuation. Consequently, the consistent maintenance of LoS transmissions is integral to attaining optimal data throughput and network performance. The nodes are, therefore, positioned to facilitate LoS, ensuring that the inherent propagation characteristics of the 60 GHz frequency are maximized to deliver robust connectivity.

- Interference management via directional antennas and orthogonal resource partitioning: Backhaul links employ narrow-beam directional

sector antennas, which provide strong spatial isolation between concurrent transmissions. Access and backhaul links are assigned orthogonal time slots following a Time Division Multiplexing (TDM) scheme, consistent with the primary resource management approach standardised in 3GPP Release 16 [52].

To facilitate analysis, we assume that a certain number of donors pre-exist within an area, which is composed of one or more plane rectangular zones (an example is shown in Figure 3.3). Therefore, the core of our work is to deploy a finite number of nodes within potential locations to ensure that both access and backhaul requirements are adequately met.

In our model, we aim to deploy IAB donors and nodes at designated candidate locations. Here, the candidate location of a node is symbolized as j, while i represents the exact coordinates where a donor is situated. Our primary objective is to determine the optimal deployment locations of the IAB nodes and, at the same time, to minimize the number of deployed nodes. We further define the communication model, ensuring that the power received at each grid center fulfills or surpasses a predefined Signal to Noise Ratio (SNR) threshold, SNR_0. This requirement ensures that users (on the access side) are guaranteed to be provided with their necessary data rate service.

3.2.1 Communication Model

The foundation of our analysis revolves around understanding the received power, P_r, of a particular node. Mathematically, this can be described by

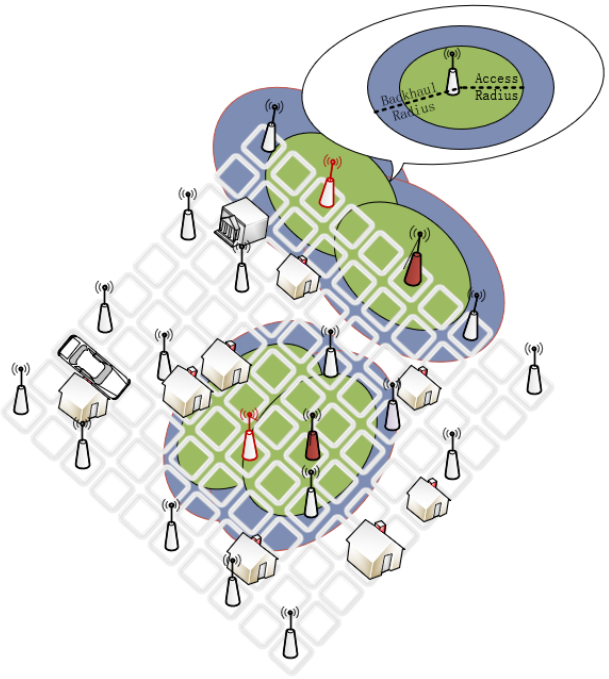

Figure 3.3: The grid indicates the area waiting to be covered and the location of deployment: solid red markers represent donors connected to the core network, red outlined markers denote deployed nodes, and white markers indicate potential node positions that remain undeployed. The green areas highlight the access coverage facilitated by the deployed nodes and donors, while the blue regions indicate the extent of backhaul coverage.

the following:

$$P_r = P_{node} + G_{all} - L_{all} - N_0, \tag{3.1}$$

$$G_{all} = G_t + G_r, \tag{3.2}$$

$$L_{all} = L_r + O_{tr} + L_{loss}, \tag{3.3}$$

$$N_0 = -174 + 10\log_{10}(W), \tag{3.4}$$

Here, P_r represents the received power in dB, while G_{all} signifies the cumulative antenna gains from the transmitting side G_t and the receiving side G_r. L_{all} embodies the total losses, which include the path loss L_{loss}, atmospheric attenuation O_{tr}, and rain-induced attenuation L_r, with N_0 denoting thermal noise.

Delving deeper into the path loss, it emerges as a function of the distance between nodes and encapsulates the consequences of signal propagation. Given the pronounced attenuation characteristics of 60 GHz mmWave, our parameter selections are crucial. We use the parameters 12.6 dB/km from ITU-R P.838 for rain attenuation if we consider an average annual rainfall rate of 35 mm/h, which corresponds to moderate convective rainfall typical of temperate climates [53]. To further strengthen our analysis, we also considered oxygen absorption of about 16 dB/km [54]. We undertook a comprehensive link budget analysis to ascertain the viability and performance of our proposed communication system and Figure 3.4 shows the results of the analysis.

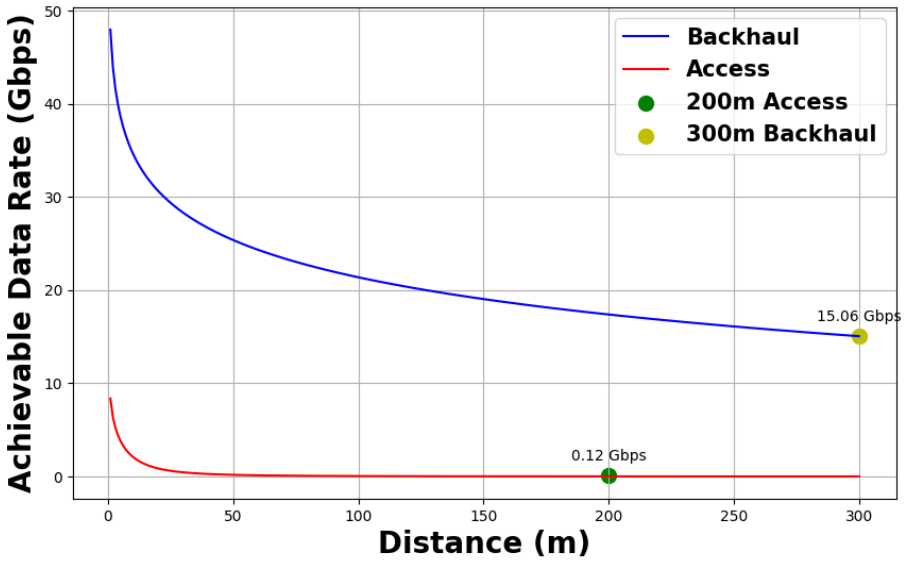

Figure 3.4: Achievable data rate for access and backhaul with transmit power 30 dbm, antenna gain for access side 5 dBi, directional antenna gain for backhaul side 30 dBi, and pathloss exponent 2.

Given the challenges posed by such pronounced path losses, the design aspiration for our system is heavily influenced by the advantages of antenna design. For the access side, the preference is towards omni-directional antennas. Their innate ability to ensure consistent radiation across all horizontal trajectories is encapsulated by the gain expression:

$$G_{\text{omni}} = G_0, \tag{3.5}$$

where G_0 symbolizes the invariant gain regardless of direction. On the flip side, backhaul links demand a more directional approach, specifically utilizing sector antennas. These are strategically chosen to counteract the challenges posed by atmospheric and rain attenuations. Their spatial filtering capabilities stand as a bulwark against interference, with the gain between two nodes

expressed as:

$$G_{tr} = \begin{cases} G_t \cdot G_r & \text{if } \frac{-\theta_{HPBW}}{2} \leq \psi \leq \frac{\theta_{HPBW}}{2} \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

We delineate our communication model with the distinct operational dynamics of access and backhaul links. The access link, characterized by connections between nodes/donors and user equipment, capitalizes on the ubiquitous coverage afforded by omni-directional antennas. In this domain, each node/donor produces a circular coverage footprint, ensuring extensive and uniform service delivery to user devices. In contrast, the backhaul link, tasked with establishing connections among nodes and between nodes and donors, employs fixed directional links. The employment of fixed directional links amplifies link reliability, curtails interference, and bolsters capacity. Therefore, we use circular representation for omni-directional coverage and use straight lines when considering backhaul links. Our primary goal remains to determine the minimum number of these conceptual circles required for full coverage. Such a modeling perspective finds its roots in prior scholarly research [55].

3.2.2 Data Rate Constraint

Figure 3.5 depicts a general K-hop IAB model. In this model, the link between two IAB nodes is the backhaul link, and the link between the node and the UE is the access link. Here, we only consider a typical backhaul link in a multi-hop IAB system where the source node α_k sends data to a destination donor through several relay nodes. In the analysis of communi-

cation network node performance, the management of data rates serves as a core metric. Specifically, the output data rate of each node must exceed its input data rate, due to 'overhead', i.e., the additional data load in the communication process. This overhead may include a range of factors, such as error correction, protocol handling, and data encryption. In practice, this relationship ensures the robustness and reliability of the system while placing specific requirements on node design and the overall network architecture. It is important to emphasize that the overhead factor modifies the total outflow rate of the IAB contributing node to include any additional data processing or transmission tasks it undertakes. For general IAB nodes, the inflow rate is sourced from a preceding donor or another node dispensing backhaul services. The outflow from such a node includes both its local access needs and the backhaul obligations of any downstream nodes, assuming that it acts as a backhaul provider. Considering that as the number of hops increases, the data rate decreases until it is unable to serve the subsequent hops, we can formulate a network without the risk of loop by incorporating the data rate constraint.

An IAB node's (including both donor and relay nodes) outgoing data rate can be categorized into:

1. Its access rate from UEs.

2. The backhaul rate for succeeding nodes, considering that this IAB node provides them with backhaul services.

Consequently, the data rate relationship between each node and the next

Table 3.1: Common radio/propagation/traffic assumptions used throughout the thesis.

Item	Setting
Carrier frequency f_c	60 GHz
Access antenna gain	5 dBi (omni)
Backhaul antenna gain	30 dBi (directional)
Transmit power	30 dBm
Pathloss	2 [56]
Oxygen absorption (60 GHz)	16 dB/km
Backhaul link capacity	15 Gbps
Access demand per node	$A_j \in [0.1, 1.5]$ Gbps

hop node can be mathematically represented as:

$$B_k = f(B_{k+1} + A_k) + C \tag{3.7}$$

In Equation (3.7), B_k represents the backhaul data rate at the k-th node. This rate is calculated as a function of both the backhaul data rate of the $k+1$-th node, denoted as B_{k+1}, and the access data rate at the k-th node, represented by A_k. The function $f(\cdot)$ denotes the application of a fixed overhead to the sum of B_{k+1} and A_k. C is a constant greater than 0.

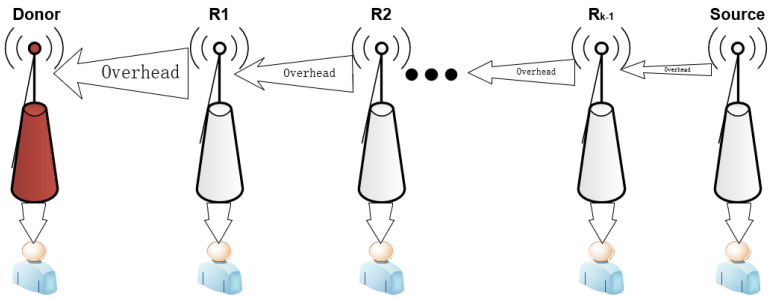

Figure 3.5: K-hop network's data rate constraint.

3.3 Problem Formulation

Given that the placement of IAB donors is predefined, which is mainly due to the availability of the fixed (optical) connections to the core network, our main objective is to minimise the number of nodes that need to be deployed while covering all the considered areas. This initiative lays the groundwork for broader network scalability. We first formulate and solve the problem for a one-hop scenario and then extend it to a multi-hop one.

Based on the presented system model, Figure 3.2 can be roughly translated into the two-dimensional map of Figure 3.3. We consider deploying IAB donors and nodes at candidate locations $j = (x_j, y_j)|\forall j \in J$ to provide coverage service. For simplicity, j represents the node's candidate location, while i denotes the location where a donor has been deployed. The grid to be covered is defined as $k = (x_k, y_k)|\forall k \in K$, ensuring that the power received at the grid center meets or exceeds the minimum SNR_0 requirement, thereby providing the necessary data rate service to users (access side).

3.3.1 One-Hop Problem Formulation

We define α_j as a binary indicator variable to show whether location i is selected to deploy a donor (1 for selected, 0 otherwise). Similarly, a binary variable α_j is defined to show node deployment in location j (1 for selected, 0 otherwise). $Y_{ij} \in \{0, 1\}$ is a backhaul service indicator when a donor deployed at i can provide the required backhaul service to a node placed at location j. Similarly, $C_{jk} \in \{0, 1\}$ and $C_{ik} \in \{0, 1\}$ are access service indicators when nodes placed at j and donors placed at i may serve k, respectively. The SNR is the basis for the indications of C_{ik}, C_{jk}. As the received power

P_r is a function of distance, the required access/backhaul service is ensured when nodes/donors are positioned appropriately and P_r exceeds the threshold SNR_0.

Objective

The objective of the optimization problem is to minimize the number of nodes required to provide access and backhaul services whilst aiming to cover the whole area. As $\alpha_j = 1$ implies that a node will be deployed at location (x_j, y_j), the optimization goal is to minimize α_j:

$$\min \sum \alpha_j \tag{3.8}$$

Access and Backhaul Constraints

To determine the location of α_j, we consider a set of constraints grounded in both access service provision across the entire spatial region and one-hop wireless backhaul's data rate limitations. The following constraints elucidate the service provision mechanism:

- Access Service Provision: As defined by expression (3.9), there is a coverage constraint, meaning that the nodes should be deployed in a way to cover all the considered area. Expressions (3.10) and (3.11) further delineate the conditions under which a donor i and node j can collectively ensure coverage service to k, contingent on the SNR.

$$\sum_{i=1}^{I} a_i C_{ik} + \sum_{j=1}^{J} a_j C_{jk} \geq 1, \quad \forall k \in K \tag{3.9}$$

$$C_{ik} = \begin{cases} 1 & \text{if } SNR(d_{i,k}) > SNR_0 \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

$$C_{jk} = \begin{cases} 1 & \text{if } SNR(d_{j,k}) > SNR_0 \\ 0 & \text{otherwise} \end{cases} \tag{3.11}$$

- Backhaul Service Provision: Expression (3.12) defines that each deployed node α_j must be supported by a donor i. Meanwhile, expression (6.9) ascertains that the donor's available data rate remains sufficient for service provision, even when catering for multiple nodes. This is premised on the assumption that the donor's capacity F_i is invariant, A_i solely depends on the donor's access data rate, and R_o is a fixed overhead. Hence, $\sum_{j \in J} Y_{ij} R_{ij}$ represents the data rate required for all nodes j connected to donor i:

$$\alpha_j - Y_{ij} \leq 0, \quad \forall i \in I, \quad \forall j \in J \tag{3.12}$$

$$F_i - R_o(A_i + (\sum_{j \in J} Y_{ij} R_j)) \geq 0, \quad \forall i \in I \tag{3.13}$$

- Data Rate Interpretation: The data rate R_{ij} between a donor at position i and a node at position j is formulated considering both the access data rate and the associated overhead:

$$R_{ij} \geq f(A_j) \tag{3.14}$$

where $f(x)$ is the linear function capturing the overhead and A_j is determined by the access data rate.

From the aforementioned analysis, it is evident that both the coverage and backhaul constraints must be satisfied while minimizing the number of nodes during the network planning process. We assume orthogonal access–backhaul resource partitioning via TDM, consistent with 3GPP Release 16 specifications [52]. This assumption is maintained consistently across all subsequent chapters. Consequently, we present the optimization problem as follows:

$$\begin{aligned} &\min \sum \alpha_j \\ \text{subject to:} \quad &(9)(10)(11)(12)(13)(14) \\ &C_7 : \alpha_j \in 1, 0, \forall j \in [J] \end{aligned}$$

All parameters are listed in Table 3.2. The formulated problem is a mixed-integer (binary) linear programming one which can be solved by solvers like CPLEX.

Table 3.2: One-hop and multi-hop formulation parameters.

One-hop and Multi-hop Formulation Parameters	
i	Pre-deployment donor location
j	Number of potential node locations
k	Number of grids needing to be covered
R_o	Overhead or required overhead
I	Set of all donor locations
J	Set of all potential node locations
K	Set of grid locations that need to be covered
u	Set of active users in the coverage of a donor or node
α_j	Indicates whether the candidate location is chosen to deploy a node
C_{ik}	Indicates whether grid $k \in K$ can be covered when a node is deployed at location $i \in I$
C_{jk}	Indicates whether grid $k \in K$ can be covered when a node is deployed at location $j \in J$
Y_{ij}	Indicates whether donor i can provide backhaul to node j, i.e., $i \in I$, $j \in J$
Y'_{jn}	Indicates whether node at $j \in J$ can provide backhaul to node at $n \in J$
Y'_{nm}	Indicates whether node at $n \in J$ can provide backhaul to node at $m \in J$
d_{ij}/d_{pq}	Distance between donor and node, or between any nodes p and q, where $p, q \in I \cup J \cup U$
R_{ij}/R_{pq}	Data rate between donor i and node j, or between nodes p and q, where $p, q \in I \cup J \cup U$
$A_i/A_j/A_x$	Access data rate when donor/node is deployed at i, j, or for a general node/donor x
F_i	Fixed data rate of donor $i \in I$

3.3.2 Multi-Hop Problem Formulation

The previous one-hop scenario makes it clear that it places strict requirements on the location and count of donor distributions. Recognizing that achieving a practical donor distribution is not always feasible, we propose a multi-hop problem model inspired by the single-hop problem scenario. The inherent multi-hop problem involves determining the most optimal node placement to ensure efficient data transfer via node-to-node communications whilst minimizing the number of nodes deployed.

Objective

The objective of this problem is to minimize the number of deployed nodes to cover all grid places:

$$\min \sum_{j \in J} \alpha_j \tag{3.15}$$

Coverage Constraint

Each grid place must be covered by at least one deployed node or one donor.

$$\sum_{j \in J} C_{jk} \alpha_j + \sum_{i \in I} C_{ik} \geq 1, \forall k \in K \tag{3.16}$$

Data Rate Constraint

For each deployed donor, the flow-in data rate should be a constant value determined by a fixed fiber data input rate. The data rate that flows out from a donor node comprises two parts: (i) the data rate required for its own access coverage, and (ii) the backhaul data rate required by the downstream nodes if the donor node provides backhaul service to them. It is important to note that the total flow-out data rate from a donor node is subject to an overhead multiplier, which accounts for any additional data processing and transmission overheads incurred in the donor node. For regular nodes, the flow-in data rate is contributed by the upstream donor or another node that provides backhaul service to them. The flow-out data rate from a regular node is constituted by its local coverage data rate requirements and the backhaul data rate for any downstream nodes, given the condition that this node provides a backhaul service to the next-hop node. Here, expression (3.17) describes the data rate limitation of the donor. Similar to the donor, the to-

tal data rate flow-out from a regular node is also multiplied by an overhead factor to account for additional data transmission overheads. It is also worth noting that a node will only function properly when it can secure a sufficient data rate service, or at least meet its own access consumption. Here, expression (3.17) describes the data rate limitation of a donor. Expressions (3.18) and (3.19) represent the data rate limitation of the node. The fundamental constraint in this optimization problem is that the total data rate flow-in for any node, whether it is a donor node or a regular node, must always exceed its data rate flow-out. This condition ensures the robustness and sustainability of the network's data rate distribution, allowing for uninterrupted data service across the network. Failure to meet this data rate condition would render the node inactive and disrupt the network deployment to the next hop:

$$F_i - R_o(A_i + (\sum_{j \in J} Y_{ij} R_{ij})) \geq 0 \quad \forall i \in I \tag{3.17}$$

$$R_{ij} Y_{ij} - R_o(A_j + \sum_{n \in J, n \neq j} Y'_{jn} R_{jn}) \geq 0 \quad \forall i \in I, \forall j \in J \tag{3.18}$$

$$R_{jn} Y'_{jn} - R_o(A_n + \sum_{m \in J, m \neq n} Y'_{nm} R_{nm}) \geq 0 \quad \forall j, n \in J \tag{3.19}$$

From the aforementioned analysis, it is evident that both access and backhaul constraints must be satisfied during the network deployment process. And the data rate constraint will automatically limit the number of the hops. Consequently, we present the optimization problem as follows:

$$\text{Minimize} \quad \sum_{j \in J} \alpha_j \tag{3.20}$$

$$\text{subject to:} \quad \sum_{i \in I} C_{ik} + \sum_{j \in J} C_{jk}\alpha_j \geq 1 \quad \forall k \in K \tag{3.21}$$

$$A_x = \sum_{u \in U} R_{xu} + C \quad \forall x \in I \bigcup J \tag{3.22}$$

$$0 \leq R_{pq} \leq SNR(d_{pq}) \tag{3.23}$$

$$(17)(18)(19)$$

$$Y_{ij} <= \alpha_j \tag{3.24}$$

$$Y'_{jn} < \alpha_j \tag{3.25}$$

3.3.3 Greedy Approach to Multi-Hop Optimization Problems

The multi-hop problem formulated in the last section is characterized by its non-convex and mixed-integer nature and the non-convexity of the problem arising from the constraints involving products of decision variables (e.g., $R_{ij}Y_{ij}$), and the constraints involving a function of the SNR with respect to distance, which could be non-linear depending on the specific form of $f(x)$. Non-convex problems are generally more difficult to solve due to the potential existence of multiple local optimal solutions. This class of problems can be solved efficiently using heuristic or metaheuristic approaches. It should be noted that the results of these methods are not guaranteed to be the optimal solutions but they will efficiently achieve good sub-optimums. To solve this complex problem, we consider using a greedy algorithm, an iterative algorithm that makes the locally optimal choice at each stage with the hope of finding the global optimum. The algorithm is divided into two stages: the

coverage stage and the backhaul stage.

In the coverage stage 1, we aim to cover every grid. The network is initialized with the deployed donor only. The potential node locations are sorted by their capacity to cover uncovered grid points and iteratively added to the network if they are within the backhaul radius of a donor or another node in the network, resulting in updated covered grid points.

Algorithm 1: Coverage Stage

Input : A: target area, P: potential nodes, N: current donor/decided nodes, R_c: coverage radius, R_b: backhaul radius
Output : Updated A, P, N

while *exists uncovered area in* A **do**
  $F \leftarrow \{\ p \in P \mid dist(p, N_{donor}) \leq R_b\ \}$;
  Sort F by potential coverage in uncovered area;
  if $F = \emptyset$ **then**
    break
  end
  $n^* \leftarrow$ top node in F;
  Add n^* to N; mark covered area in A;
  Remove n^* from P;
end
return A, P, N;

In the backhaul stage 2, the objective is to establish a coherent data flow topology. Taking into account the imperatives of reduced latency, augmented reliability, and enhanced throughput, a strategic inclination towards maximizing single-hop connections is promoted. This strategy facilitates a hierarchical structure where nodes, post their one-hop deployment, serve as anchor points for subsequent multi-hop connections, thereby ensuring comprehensive coverage, especially for nodes that lie beyond the donor's backhaul radius. Given the incorporation of directional antennas for the backhaul connections, interference is substantially minimized. Moreover, the rich spectrum afforded by mmWave technology provides the requisite bandwidth to

accommodate a profusion of direct, one-hop connections. To methodically realize this network topology, we initiate a graph with the donor nodes as root vertices. Subsequently, deployed nodes are based on their proximity to the donor or the closest deployed node. Each node is then tethered to the nearest donor which fulfills the data rate prerequisites, thereby engendering a potential network linkage. Nodes that are incompatible with these stringent constraints are relegated to a waiting list for subsequent analysis.

Algorithm 2: Backhaul Stage

Input : Area A, Potential nodes P, Network nodes N, Backhaul radius R_{bh}
Output : Updated grid G, P, N, Status

Identify non-donor nodes $\mathcal{ND}$;
foreach *donor* $d \in N$ **do**
  foreach $n \in \mathcal{ND}$ **do**
    if $dist(d, n) \leq R_{bh}$ *& rate OK* **then**
      Connect $n \rightarrow d$, update rates;
    else
      Add n to W if not in W;
    end
  end
end
Set H = first-hop nodes;
while $W \neq \emptyset$ **do**
  $H_{new} \leftarrow \emptyset$;
  foreach $h \in H$ **do**
    foreach $n \in W$ **do**
      if $dist(h, n) \leq R_{bh}$ *& rate OK* **then**
        Connect $n \rightarrow h$, update rates;
        Move n from W to H_{new};
      end
    end
  end
  if $H_{new} = \emptyset$ **then**
    Break;
  end
  $H \leftarrow H_{new}$;
end
return $G, P, N, Status$;

The algorithm persistently assesses the nodes within the waiting list,

striving to bridge them to the most proximal node or donor that aligns with the data rate specifications. In the event that a compatible connection is unattainable, the specific node is extricated from the network. This necessitates an update of the covered grid points, and consequentially, the algorithm reverts to the coverage stage to ensure that all grid points maintain their coverage integrity. The algorithm continues through these stages until all grid points are covered and all data rate constraints are satisfied, which indicates that a solution has been found. While the greedy algorithm does not guarantee to find the global optimum due to the non-convex nature of the problem, it is a practical choice for obtaining a feasible solution in a reasonable amount of time. Its performance can potentially be improved by incorporating additional strategies, such as local search or heuristics, depending on the specific characteristics of the problem at hand.

3.4 Simulation Results

3.4.1 Simulation Setup

The simulation is set in a 1000 $\times$ 1000 m grid, within which the potential nodes are systematically positioned every 10 m, echoing the real-world spacing of street lights and, thus, yielding 100 $\times$ 100 potential node sites. The donors, being the cornerstone of this simulation, are not predetermined but rather are introduced as variable elements. Their quantities and positions are randomly generated inputs, due to the in-the-real-world constraint of donors being tethered to the core network via wired connections. This introduces a level of unpredictability and variability into our simulations, ensuring that

our model is not just theoretical but also works under various predefined donor cases. Based on the donor locations and number characterized by randomness, we proceed to incorporate statistical methods to analysis and comprehend the resultant effects on the algorithm's performance and efficacy. Each area within the 1000×1000 grid is mandated to be under the coverage of either a node or a donor. The nodes, in turn, are linked to the donors ensuring a seamless backhaul communication, underpinned by our assumption of LoS communication. This dual conditionality of coverage and backhaul ensures that the deployment of nodes is not just about coverage but is the requirement for effective communication back to the core network via the donors. In this thesis, to encapsulate the variation in data rates for different nodes or donors, each donor and node is assigned an aggregate random access data rate ranging from 0.1 Gbps to 1.5 Gbps (100 Mbps to 1500 Mbps). The 1.5 Gbps setting can be interpreted as, for example, supporting multiple concurrent high-rate sessions within a hotspot region, such as a small number of multi-hundred-Mbps XR/ultra-HD streams or a larger number of moderate-rate eMBB users, depending on the assumed traffic mix [57, 58]. Also, each donor's capacity is at 15 Gbps, a specification that ensures the robustness and efficacy of the wireless backhaul communication. Also, the link budget analysis of Figure 3.4 ensures that such data rate assumptions are reasonable. An overhead of 1.2 is configured to account for various communication inefficiencies and to ensure a more realistic simulation scenario.

3.4.2 One-Hop Simulation Result

In the last section, we formulated the single-hop problem of IAB network planning for the random distribution of donors. To solve the formulated optimization problem, we adopted the Python-based Gurobi library, which has been shown to be efficient in previous location coverage problems [59].

Figure 3.6 is a testimonial to the simulation's effectiveness. With nine donors, randomly positioned, the simulation strategically orchestrated the deployment of a mere eight additional nodes to achieve full grid coverage.

To analyse the relationship between the number of donors and the number of feasible optimization models. We simultaneously monitor the accumulation of both feasible and infeasible optimization models until a total of 100 models is reached. In Figure 3.7, we illustrate the number of feasible models as the number of donors varies between 5 and 20. Feasible models representing the Gurobi optimizer can successfully find an optimal solution. The plot reveals a corresponding increase in the number of feasible models, highlighting the positive impact of the donor quantity on achieving practical solutions. As the number of donors increases, there is a greater likelihood of obtaining a feasible solution for the given coverage constraints and the number of donors required to cover the area, leading to a reduction in the marginal benefits delivered as the donor density increases further.

The number of IAB nodes deployed when the number of donors increases from 5 to 20 is shown in Figure 3.8. The figure demonstrates a decline in the average nodes deployed as the quantity of donors escalates. The standard deviation around the mean indicates that the number of nodes deployed may be affected by factors including donor location. Furthermore, a deeper

analysis of the results unveils a substantial increment in the total number of deployments, embodying both donors and nodes. Despite the surge in donors, the comprehensive number of deployments increases. This suggests that an increase in the number of donors does not necessarily deter the growth in the total deployments.

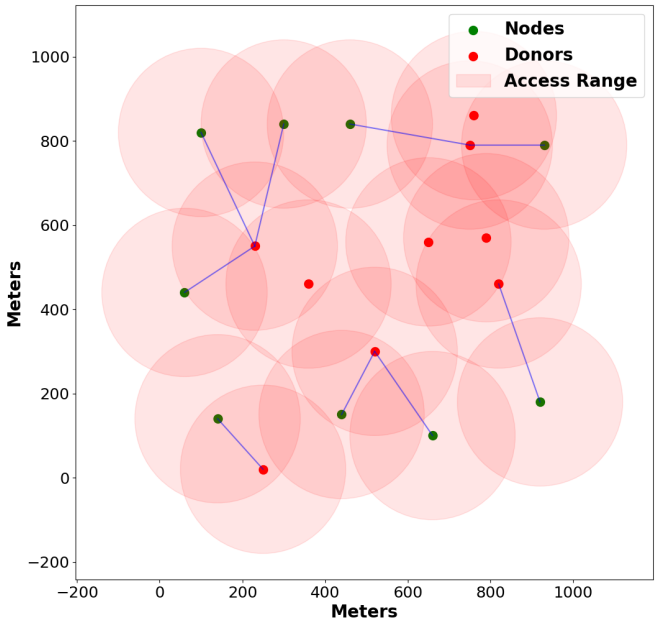

Figure 3.6: Optimization one-hop deployment result on a 1000 m $*$ 1000 m area with predefined donor positions.

3.4.3 Multi-Hop Simulation Results

We formulated the multi-hop problem of IAB network planning for random sparse distributions of donors. In our pursuit to establish the effectiveness of the proposed algorithm, we initially embarked on a one-hop greedy simulation. The results obtained from the simulation were compared to the outcome of the earlier mixed-integer linear programming (MILP) optimization problem. As shown in Figure 3.9, we found that as the number of donors swells,

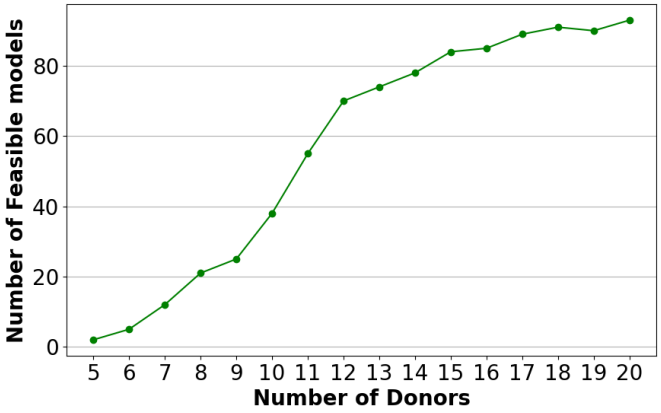

Figure 3.7: Number of feasible models of one-hop simulation for different numbers of donors.

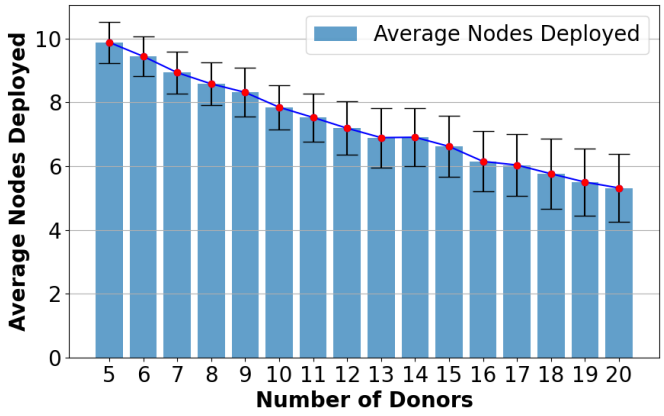

Figure 3.8: Average deployed nodes for one-hop optimal in 100 simulations with 200 m coverage radius and 300 m backhaul radius.

the one-hop greedy algorithm becomes more proficient, gradually approximating the globally optimal solution provided by MILP. For example, when the donor count escalates to 20, this difference dwindles to approximately 0.5 nodes. This performance of the single-hop greedy algorithm, in particular the gradual agreement with the MILP results (less than one node difference) as the number of donors increases, is concrete proof of the effectiveness of the algorithm.

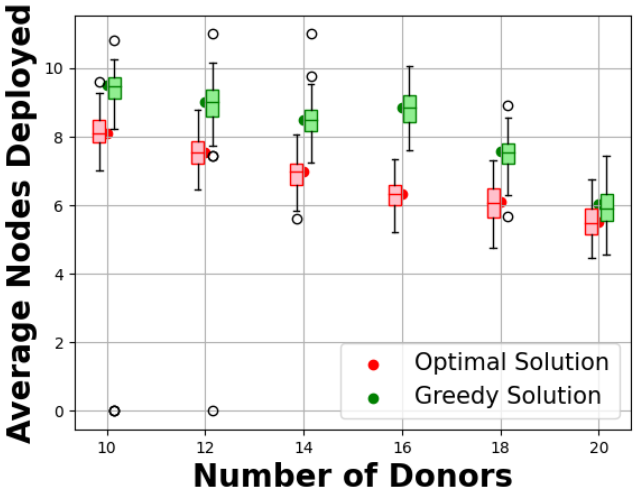

Figure 3.9: Comparison of average nodes deployed using one-hop optimal and one-hop greedy strategies as a function of the number of donors, with boxplots indicating variability around the means.

In the given multi-hop scenario, where the coverage radius and backhaul radius for each donor/node are stipulated at 200 and 300, respectively, a deployment comprising six donors resulted in the utilization of 16 nodes in one successful deployment, as shown in Figure 3.10. One noteworthy observation from the results is that, while a node possesses the capability to connect with multiple donors, it does not necessarily imply simultaneous connections with all of them. Instead, this potential for multi-hop connections can be attributed to the donor providing sufficient data transfer rates to satisfy the node's requirements. This suggests that following a multi-hop formulation can effectively limit the hop length by the data rate constraint.

Next, we conducted a comparative analysis of the greedy multi-hop method and the one-hop optimal strategy. For each specified donor, we conducted 100 simulations and documented the results. There are significant differences between the one-hop optimal and the multi-hop greedy approach, as shown

in Figure 3.11. Starting with only five donors, the one-hop yields two feasible models out of 100 simulations, compared to the multi-hop strategy's 12, which means that the multiple hops increase the likelihood of successful deployment. As donor nodes increase, the successful number for the multi-hop method grows more slowly than its optimal one-hop counterpart. At 11 donors, the one-hop method achieves 45 successful deployments, compared to the multi-hop's 62. And with the increment in donor nodes, the successful rates of both methods approach convergence, which means the advantage of the multi-hop strategy diminishes. Figure 3.12 investigates the influence of multi-hop on the average number of nodes deployed. The data reveal a trend of decrement in the average number of nodes concomitant with increment in the multi-hop. The values oscillate from an initial 16.916 with five donor nodes, attenuating to 7.914 when the donor nodes ascend to 20.

To examine the effect of the potential node location density on network coverage, we simulated various potential node distances ranging from 10 m to 50 m, with the results displayed in Figure 3.13. A critical observation from the early stages of network deployment emerged, particularly with fewer donors, such as five. The data suggest that as the gap between potential positions widens, the average number of nodes deployed incrementally increases. This implies that in environments with limited donors, larger separations might require deployment of more nodes for full coverage. Notably, in scenarios with five donors, choosing a separation of 10 m optimizes the deployment count, suggesting an efficient strategy for the initial stages. Conversely, in situations with a higher donor count, such as 20 donors, the influence of space on the node deployment diminishes, with values consistently ranging

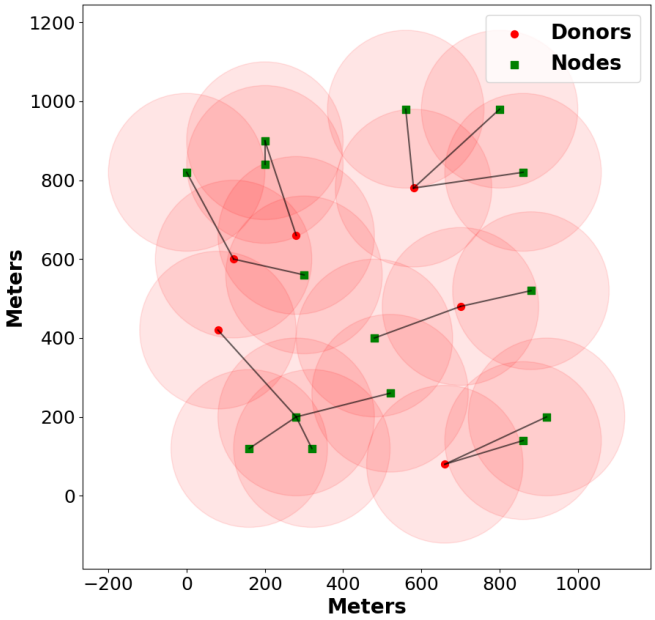

Figure 3.10: Deployment result of greedy multi-hop in a 1000 m $*$ 1000 m area with coverage radius 200 m and backhaul radius 300 m. The line shows that the remaining data rate of the current hop can provide enough backhaul service for the next hop.

between 7.5 to 8.5, regardless of the separation distance.

3.5 Conclusions

This chapter has presented a comprehensive heuristic-based approach to IAB network planning, establishing fundamental methodologies that address the practical challenges of deploying IAB networks under real-world geographical and resource constraints. Through rigorous mathematical formulation, we developed a MILP-based optimization framework for single-hop scenarios in dense donor environments, achieving optimal node placement solutions. For sparse donor distributions, our proposed greedy algorithm effectively handles the inherent complexity of multi-hop IAB deployments, demonstrating signif-

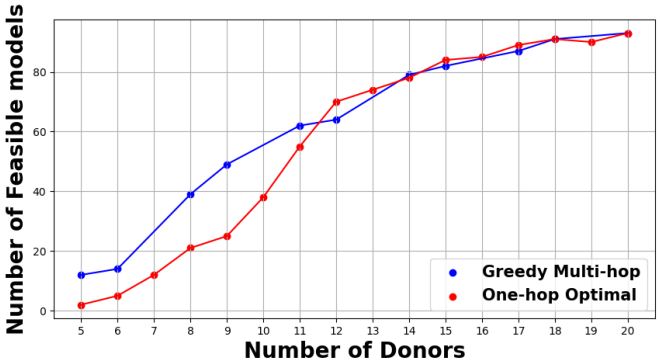

Figure 3.11: Number of feasible models out of 100 simulations between greedy multi-hop and one-hop optimal strategies across various donor counts.

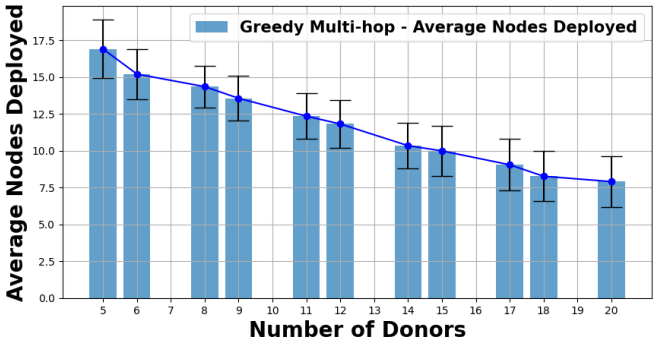

Figure 3.12: Average nodes deployed for the greedy multi-hop strategy with respective standard deviations across different donor counts

icant improvements in network feasibility compared to single-hop approaches. The integration of data rate constraints, validated through comprehensive link budget analysis at 60 GHz, provides a more realistic alternative to traditional fixed hop-count limitations. Our simulation results reveal that while the proposed heuristic methods achieve near-optimal performance in dense donor scenarios, they face computational scalability challenges in large-scale dynamic networks. The greedy algorithm's performance variability across different network configurations highlights the need for more adaptive and intelligent optimization approaches. These findings establish important baseline

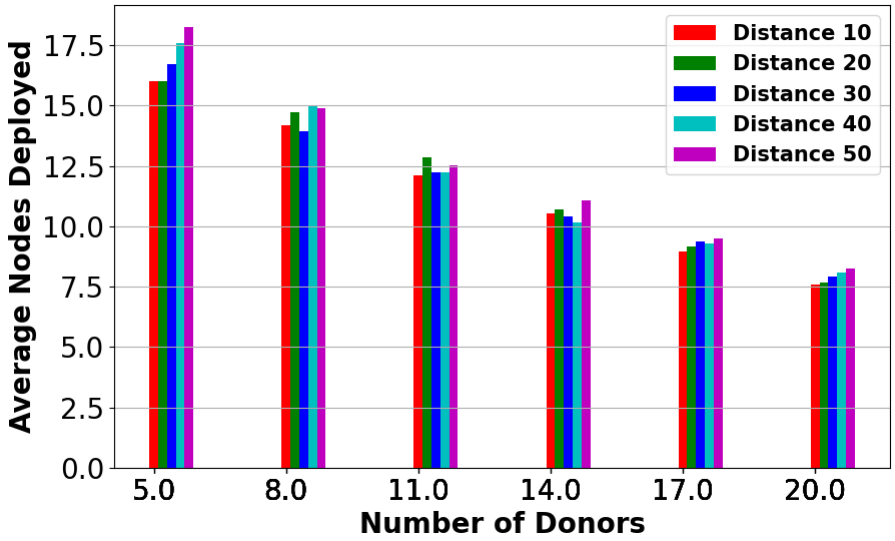

Figure 3.13: Comparison of the average nodes deployed against the number of donors, differentiated by potential node distance from 10 m to 50 m. Potential node locations are distributed equidistant in a $1000 * 1000$ area and each distance showcases a distinct trend in node deployment as the donor count varies.

performance metrics and identify key limitations of traditional optimization methods, particularly their inability to adapt to dynamic network conditions and learn from deployment experiences. The methodologies developed in this chapter provide essential benchmarks for evaluating advanced optimization techniques, setting the stage for the deep reinforcement learning-based approach presented in the following chapter, which addresses the scalability and adaptability challenges identified in this work.

4 Deep Q-Learning Approaches for IAB Network Planning

4.1 Introduction

Building upon the heuristic-based optimization approaches developed in Chapter 3, this chapter explores the application of advanced machine learning techniques to address the complex challenges of IAB network planning. While the previous chapter demonstrated the effectiveness of traditional optimization methods, including MILP formulations and greedy algorithms, these approaches face scalability limitations and struggle to adapt to dynamic network conditions. The computational complexity increases exponentially with network size, and the static nature of these methods makes them unsuitable for environments with changing demands and topological constraints.

To address these limitations, this chapter introduces DRL as a paradigm shift toward intelligent, adaptive network planning. DRL represents a powerful convergence of reinforcement learning principles with deep neural networks, enabling autonomous agents to learn optimal decision-making policies through interaction with complex environments [60]. Unlike traditional optimization approaches that require explicit mathematical formulations and predefined solution strategies, DRL agents can discover optimal policies through

trial-and-error learning, making them particularly well-suited for dynamic and uncertain environments.

4.1.1 Fundamentals of Reinforcement Learning

RL is a computational approach to learning whereby an agent learns to take actions in an environment to maximize cumulative reward [61]. The fundamental framework is based on the MDP, which provides a mathematical foundation for modeling decision-making problems. An MDP is formally defined by a tuple (S, A, P, R, γ), where S represents the state space, A the action space, P the transition probability function, R the reward function, and γ the discount factor.

The agent's goal is to learn an optimal policy $\pi^* : S \to A$ that maximizes the expected cumulative discounted reward. This is achieved through the action-value function $Q^\pi(s, a)$, which represents the expected return when taking action a in state s and following policy π thereafter. The Bellman equation provides the theoretical foundation for RL algorithms:

$$Q^\pi(s, a) = \mathbb{E}_\pi[R_{t+1} + \gamma Q^\pi(S_{t+1}, A_{t+1}) | S_t = s, A_t = a] \tag{4.1}$$

Traditional RL methods, such as Q-learning [62], maintain tabular representations of the action-value function. However, this approach becomes computationally intractable for problems with large or continuous state and action spaces, which is precisely the challenge faced in IAB network planning.

4.1.2 Deep Reinforcement Learning

DRL addresses the scalability limitations of traditional RL by employing deep neural networks as function approximators. This approach enables the handling of high-dimensional state spaces and complex decision-making problems that would be impossible to solve with tabular methods. The seminal work by Mnih et al. introduced the DQN algorithm [60], which combines Q-learning with deep neural networks to approximate the optimal action-value function.

The DQN algorithm addresses two critical challenges in applying neural networks to RL: the correlation between consecutive samples and the non-stationarity of the target values. These issues are resolved through two key innovations: experience replay and target networks. Experience replay breaks the correlation between consecutive experiences by storing transitions in a replay buffer and sampling random mini-batches for training. Target networks provide stable target values by maintaining a separate network with periodically updated parameters.

The loss function for DQN training is defined as:

$$L(\theta) = \mathbb{E}_{(s,a,r,s')\sim U(D)} \left[\left(r + \gamma \max_{a'} Q(s', a'; \theta^-) - Q(s, a; \theta) \right)^2 \right] \tag{4.2}$$

where θ represents the parameters of the main network, θ^- the parameters of the target network, and $U(D)$ a uniform distribution over the replay buffer D.

4.1.3 Advanced DQN Variants

Several variants of DQN have been developed to address specific limitations and improve performance. DDQN tackles the overestimation bias inherent in standard DQN by decoupling action selection from action evaluation [63]. The DDQN update rule is:

$$y_t^{\text{DDQN}} = r_t + \gamma Q(s_{t+1}, \arg\max_a Q(s_{t+1}, a; \theta_t); \theta_t^-) \tag{4.3}$$

Dueling DQN introduces an architectural innovation that separates the estimation of state values and action advantages [64]. This approach is particularly beneficial in environments where the value of most actions in a given state is similar, as it allows for more efficient learning of the state-value function while maintaining action-specific advantages.

4.1.4 Application to IAB Network Planning

The application of DRL to IAB network planning represents a natural evolution from the static optimization approaches explored in Chapter 3. The network planning problem exhibits several characteristics that make it well-suited for DRL approaches:

- **High-dimensional state space**: The network state includes deployment status, connectivity information, and data rate constraints across numerous potential locations.

- **Complex action dependencies**: The placement of each IAB node affects coverage, interference, and backhaul capacity for other nodes.

- **Non-convex optimization landscape**: Traditional optimization methods struggle with the non-convex nature of the multi-hop network planning problem.

- **Dynamic adaptability requirements**: Real-world deployments require adaptation to changing traffic patterns, environmental conditions, and performance requirements.

This chapter demonstrates how DRL techniques can overcome the limitations identified in Chapter 3, providing more scalable and adaptive solutions for IAB network planning. The proposed approach leverages DQN, DDQN, and Dueling DQN algorithms to learn optimal node placement strategies while satisfying coverage and backhaul constraints.

The remainder of this chapter is organized as follows: Section 4.2 presents the system model and problem formulation, building upon the framework established in Chapter 3. Section 4.3 formulates the IAB planning problem as an MDP and details the DRL solution approach. Section 4.4 presents comprehensive simulation results comparing different DRL algorithms with the heuristic baseline from Chapter 3 and 4.5 concludes the chapter and provides insights for future research directions.

4.2 System Model and Problem Formulation

The system model and problem formulation for IAB network deployment in this study build directly on the framework established in Chapter 3, Section 3.2. We consider an urban environment where IAB nodes, operating in the mmWave frequency band, are deployed at predefined locations (e.g.,

lamp-posts, traffic lights) to ensure efficient high-speed wireless coverage and backhaul connectivity. The key components, including IAB nodes and donor nodes, along with the communication model capturing power transmission, antenna gains, path loss, and atmospheric attenuation, are detailed in Chapter 3, Equations (1)–(4). These equations define the received power (P_r), total antenna gain (G_{all}), total signal attenuation (L_{all}), and thermal noise (N_0), tailored for mmWave propagation characteristics to meet SNRthresholds for user data rate requirements.

The problem formulation, also introduced in Chapter 3, Section 3.3, aims to minimize the number of deployed IAB nodes while ensuring full area coverage and sufficient backhaul data rates. The objective function and constraints, including coverage constraints (Equations (5)–(7)) and backhaul data rate constraints (Equations (8)–(10)), are adopted here without modification. All relevant parameters are listed in Chapter 3, Table 3.1, which defines symbols such as i, j, k, R_o, α_j, C_{ik}, C_{jk}, Y_{ij}, Y'_{jn}, R_{ij}, A_j, and F_i. This formulation ensures consistency with the heuristic approaches in Chapter 3 while providing the foundation for the DRL-based optimization developed in this chapter.

4.3 DRL and MDP formulation

DRL combines the decision-making capabilities of classical reinforcement learning with the representation learning power of deep neural networks. An agent in the DRL framework learns to make decisions by observing the state of the environment, taking actions, and receiving feedback in the form of rewards. These rewards guide the agent to discover policies that maximize

cumulative future rewards. More specifically, DRL is leveraged to find an optimal deployment strategy that minimizes node deployment while ensuring satisfactory SNRand adequate backhaul data rates.

We formulate the multi-hop network deployment as an MDP in the last part, aiming to minimize the number of nodes while maintaining satisfactory access and backhaul data rates. To fully define our MDP, we specify the following key components:

State (S)

The state captures the deployment status of nodes, data rates, and connectivity within the network:

- **Deployment Status of Nodes** (D):

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix}$$

  where d_{ij} is 1 if a node is deployed at location (i, j), and 0 otherwise.

- **Data Rates** (R):

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

  where r_{ij} is the data rate left of the node at location (i, j).

- **Node Connectivity Matrix** (N):

$$N = \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ n_{n1} & n_{n2} & \cdots & n_{nn} \end{bmatrix}$$

  where n_{ij} means the number of backhaul connections provided from node i to node j, and $n_{ij} = 0$ otherwise.

Actions (A)

The actions are defined for each potential node location. For each location j, the action can either be 1 for deploying a new node at a potential place or 0 to maintain no change. In this case, the action space size is J+1.

Reward Function (R)

The reward function is formulated to optimize network deployment by balancing coverage threshold, node deployment minimization, and network connectivity. It is defined as:

$$R = -\alpha A_{\text{uncov}} - \beta N_{\text{nodes}} + \delta(C) - \eta \max(0, N_{\text{nodes}} - N_{\text{ref}}) \tag{4.4}$$

In (4.4), R represents the total reward resulting from the agent's action. The term $-\alpha A_{\text{uncov}}$ penalizes the agent based on the total uncovered area $A_{\text{uncov}} = \sum_{i \in \mathcal{U}} a_i$, where $\mathcal{U}$ is the set of uncovered grid indices and a_i is the area of grid cell i. The coefficient α determines the weight of this penalty, emphasizing the importance of maximizing coverage.

The second term, $-\beta N_{\text{nodes}}$, introduces a penalty proportional to the number of deployed nodes $N_{\text{nodes}} = \sum_{i=1}^{n} d_i$, where d_i indicates the deploy-

ment status of node i ($d_i = 1$ if deployed; $d_i = 0$ otherwise). The coefficient β controls the significance of the number of deployed nodes, encouraging the agent to minimize the number of nodes used.

The function $\delta(C)$ provides a coverage reward or penalty based on the achieved coverage percentage C relative to a predefined threshold C_t:

$$\delta(C) = \begin{cases} -\lambda \left(1 - \dfrac{C}{C_t}\right), & \text{if } C < C_t \\ \gamma\, e^{(C - C_t)}, & \text{if } C \geq C_t \end{cases} \tag{4.5}$$

In (4.5), when the coverage C is below the threshold C_t, a penalty scaled by λ is applied, proportional to the shortfall $1 - C/C_t$. When C meets or exceeds C_t, a reward is granted, increasing exponentially with $C - C_t$ and scaled by γ. This structure motivates the agent to meet or exceed the coverage threshold.

The final term of (4.4), $-\eta \max(0,\ N_{\text{nodes}} - N_{\text{ref}})$, imposes a penalty if the number of deployed nodes exceeds a reference number N_{ref}, with η controlling the penalty's weight discouraging unnecessary node deployments.

Overall, the reward function guides the agent to achieve target coverage while minimizing the number of deployed nodes, promoting effective and efficient network deployment strategies. By balancing penalties and rewards through the coefficients α, β, λ, γ, and η, the agent is steered toward optimal configurations that meet coverage goals without excessive resource expenditure.

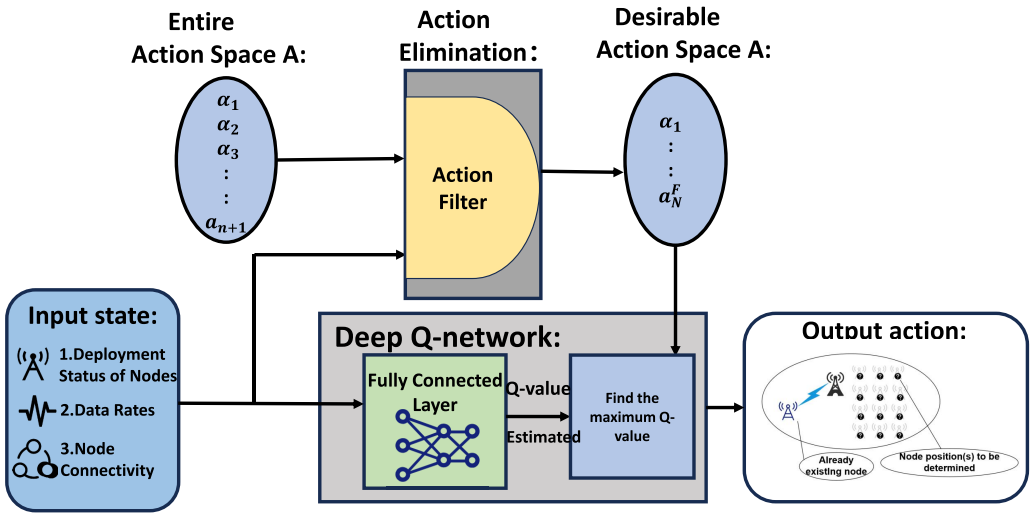

Figure 4.1: Deep Q Network with Action Elimination for IAB Network Planning

DQN and Action Elimination

DQN enhance Q-learning by employing neural networks to approximate the optimal action-value function $Q^*(s, a)$ in complex environments. In the IAB network planning, DQN are utilized to learn optimal node placement strategies. The overall system architecture, including the integration of DQN with action elimination, is illustrated in Figure 4.1.

The DQN approach is characterized by two key equations. The first equation is the DQN update rule, which is defined as $Q(s_t, a_t; \theta) = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-)$. In this equation, s_t and s_{t+1} represent the current and next states, a_t is the action, r_t is the reward, γ is the discount factor, and θ and θ^- are the parameters of the Q-network and target network, respectively. The second equation is the training objective, which is expressed as $L(\theta) = \mathbb{E}\left[(y_t - Q(s_t, a_t; \theta))^2\right]$. Here, $y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-)$ is the target Q-

value, which is used to calculate the loss function and update the parameters of the Q-network.

In IAB network planning, the action space can be extremely large, as each potential node location is a possible action. However, many actions may be invalid or unnecessary in a given network state, such as deploying nodes in already covered areas. Action elimination prunes these ineffective actions and reduces the search space by dynamically narrowing down the feasible action set based on the current state and constraints. This approach significantly accelerates the convergence of reinforcement learning algorithms and enhances training efficiency. This is implemented through two main algorithms:

Algorithm 3: DQN with Action Elimination for Network Deployment

```
for episode in episodes do
    Reset environment and initialize deployed nodes
    while not done do
        Obtain valid actions using Algorithm 2
        action = select_action(state, valid_actions)
        next_state, reward, done = step(action)
        store_transition(state, action, reward, next_state, done)
        agent.experience_replay()
        state = next_state
    end
    update_target_network()
    epsilon_decay()
end
```

Algorithm 1 provides a high-level overview of our DQN with action elimination approach, while Algorithm 2 details the action filtering process that is crucial for reducing the action space. This approach enables the DQN to learn efficient network deployment strategies by focusing on feasible and

Algorithm 4: Action filter

```
Input: All actions, Deployed nodes, taken actions, min distance
Output: filtered_actions
Initialize filtered actions
for action in All actions do
    if action is in deployed nodes then
        |
    end
    check whether action meet backhaul constraint
    for left actions do
        calculate distance between new action and taken actions)
        if distance > min distance then
            Action is not valid
        end
    end
    if is valid then
        Add action to filtered_actions
    end
end
return filtered actions
```

promising actions, leading to improved coverage with fewer nodes in our problem.

4.4 Simulation Results and Analysis

To evaluate the performance and adaptability of different reinforcement learning models, we first implement the algorithms described in Section II on DQN, And then expanding to DDQN, and Dueling DQN and compare them with a heuristic approach. To verify the robustness of our algorithms in different initial donor environments, we test three distinct initial donor placement patterns: a five-dice pattern for balanced distribution, a vertical pattern for linear arrangement, and a pentagon pattern for dispersed geometric distribution. By simulating these diverse scenarios, we aim to gain comprehensive

insights into how these models perform and adapt to various initial network configurations. This approach assesses the flexibility and effectiveness of our proposed methods in real-world urban environments with different donor placement constraints. Table 4.1 summarizes the key simulation parameters, network settings, and deep network settings used across all three initial donor configurations.

Table 4.1: DQN Architecture and Key Hyperparameters

Parameter	**Value**
Neural Network Architecture	
Hidden Layers	3 Layers
Neurons per Layer	1024, 512, 256
Activation Function	ReLU
Output Layer	Outputs n_{actions} Q-values
Layer Normalization	After each hidden layer
Training Hyperparameters	
Learning Rate (α)	0.001
Optimizer	Adam
Discount Factor (γ)	0.99
Batch Size	512
Replay Memory Size	20,000 transitions
Target Network Update Frequency	Every 64 steps
Environment Parameters	
Map Size	1000m $\times$ 1000m
Grid Size	50m
Node Coverage Radius	200m
Node Backhaul Radius	300m
Frequency Band	60 GHz
Overhead Factor	1.2
Node Data Rate	2 Gbps
Donor Data Rate	30 Gbps

4.4.1 Models Compared

In this study, we compare several algorithms, starting with a heuristic algorithm from our previous work [31], which uses a greedy strategy to achieve target coverage while ensuring network connectivity. The main model proposed is based on DQN. To address overestimation issues in DQN, we also explore DDQN, which uses separate networks for action selection and evaluation, leading to more stable learning. Additionally, the Dueling DQN model is introduced, which improves performance by separating state value and action advantage estimation.

1. Heuristic Algorithm: A greedy strategy from our previous work [31] serves as a baseline that iteratively selects locations providing maximum coverage increase while ensuring network connectivity.

2. DQN: The main model proposed in this study. Subsequent models are based on DQN.

3. DDQN: Addresses overestimation problem in DQN by using separate networks for action selection and evaluation, leading to more stable learning.

4. Dueling DQN: Separates state value and action advantage estimation in DQN, allowing better generalization and improved performance in states where action choice has less impact.

4.4.2 Simulation Result

To investigate the learning behavior of the reinforcement learning models, we examine the convergence of the reward over the training episodes. Figure 4.2

presents the reward convergence plot for DQN, DDQN, and Dueling DQN with 500 moving windows in five-dice environment. All three models demonstrate a consistent improvement in the reward as the training progresses, indicating their ability to learn effective deployment strategies. However, the rate of convergence and the final reward values differ among the models. Dueling DQN demonstrates the fastest convergence, followed by DDQN, and then DQN. This order of convergence speed can be attributed to the architectural differences among the models. Notably, while DQN initially shows slower convergence, it eventually achieves the same reward level as DDQN and Dueling DQN after approximately 29000 episodes. To test the per-

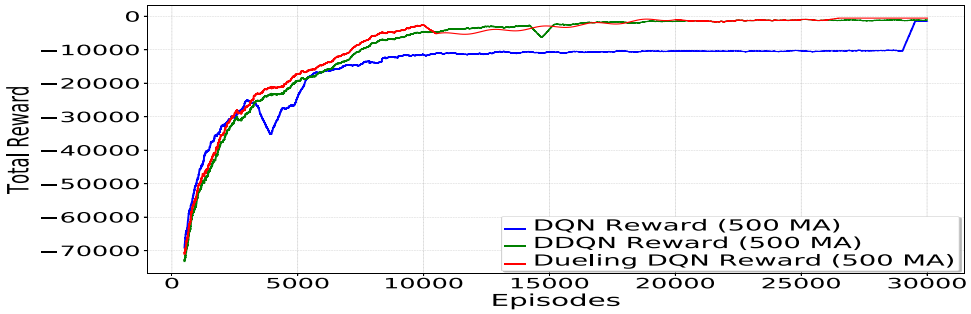

Figure 4.2: Reward vs Episodes Comparison in five-dice distribution donor environment.

formance of the proposed models in different initial donor environments, we conduct experiments with donors arranged in a vertical pattern. Figure 4.3 illustrates the best deployment test results from 100 tests for DQN, DDQN, and Dueling DQN in this setting. And both Dueling DQN and DDQN achieve full coverage with least nodes cost.

Figure 4.4 compares the average number of nodes deployed by a heuristic approach (Greedy) and three reinforcement learning models (DQN,DDQN, and Dueling DQN) across three environments. The Greedy approach consis-

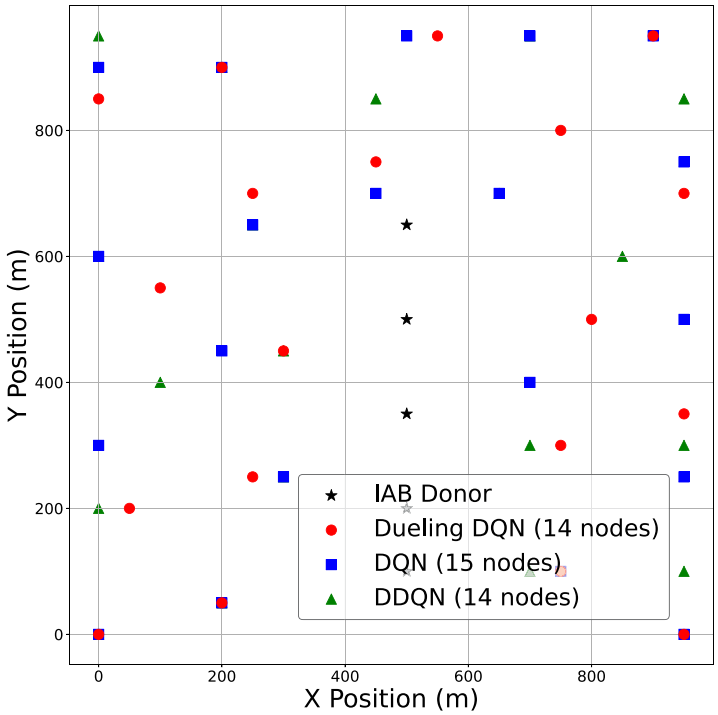

Figure 4.3: Final network planning for three models in vertical distribution donor environment.

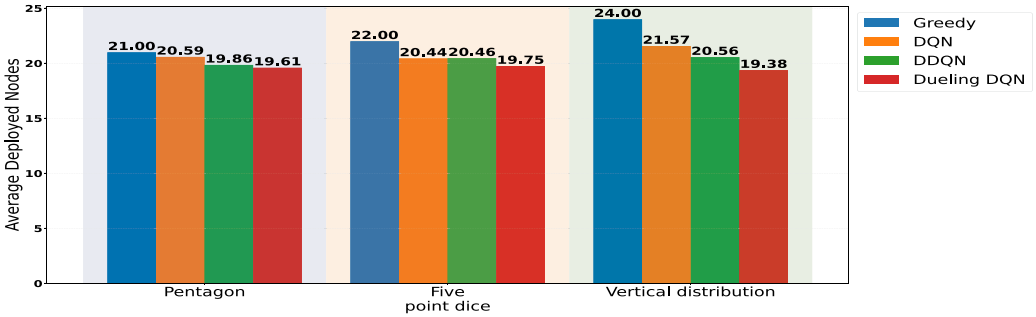

Figure 4.4: Deployed Nodes vs Different Initial Donor Environment.

tently requires the most nodes (mean 22.33) and shows the highest variance, indicating poor adaptability. In contrast, the DRL models demonstrate superior performance and flexibility. Dueling DQN emerges as the most efficient, deploying the fewest nodes on average (19.58) and maintaining the most consistent performance across environments, suggesting excellent generalization capabilities. DDQN (mean 20.29) and DQN (mean 20.87) also show significant improvements over Greedy, with DDQN generally outperforming DQN. These results suggest that Dueling DQN's architecture enables it to learn more efficient and generalizable deployment strategies, offering a promising solution for adaptive IAB network deployment in diverse urban settings.

4.5 Conclusions

This chapter has successfully demonstrated the transition from traditional heuristic optimization to advanced DRL techniques for IAB network planning, addressing the scalability and adaptability limitations identified in last Chapter. Through comprehensive formulation of the network planning problem as an MDP and implementation of specialized DQN algorithms with action elimination strategies, we achieved significant performance improvements over conventional approaches. The experimental results reveal that DRL methods require 12-15% fewer nodes than heuristic baselines while maintaining complete coverage and connectivity requirements across diverse donor configurations. Dueling DQN emerged as the most effective algorithm, demonstrating superior generalization capabilities and consistent performance across different network scenarios. However, the analysis also reveals a critical limitation inherent to learning-based approaches: the sub-

stantial training time requirements. Our experiments show that achieving convergence requires approximately 30,000 episodes, translating to considerable computational overhead compared to the immediate solutions provided by heuristic methods in Chapter 3. This training burden becomes particularly challenging when adapting to different initial donor configurations, as each new scenario necessitates extensive retraining to achieve optimal performance. While the DRL framework demonstrates excellent learning capabilities within specific environments, the lack of transferability across different network setups limits its practical deployment efficiency. These observations highlight the need for more sophisticated learning paradigms that can leverage knowledge acquired from previous deployments and adapt rapidly to new configurations. The findings establish important foundations for the advanced techniques explored in the following chapter, where SAC algorithms and transfer learning mechanisms address these limitations by enabling efficient knowledge transfer across diverse initial donor arrangements, significantly reducing training requirements while maintaining the superior performance characteristics demonstrated by learning-based approaches.

5 Transfer-Learning-Enhanced Soft Actor–Critic for IAB Planning under Dynamic Environments

5.1 Introduction

Building upon the DRL foundations established in Chapter 4, this chapter addresses the critical limitations identified in previous approaches, particularly the extensive training time requirements and the inability to efficiently adapt to varying network configurations. While Chapter 4 demonstrated the effectiveness of DQN-based approaches for IAB network planning, the results revealed that learning-based algorithms require substantial computational resources and training episodes to achieve convergence. Furthermore, when network conditions change—such as different initial donor configurations—these models necessitate complete retraining, making them impractical for dynamic deployment scenarios.

To overcome these challenges, this chapter introduces advanced DRL techniques combined with TL methodologies, specifically focusing on Soft Actor-Critic(SAC) algorithms and their adaptation to discrete action spaces. The integration of these technologies addresses both the scalability and adaptability requirements essential for next-generation IAB network planning in the context of 6G and Open RAN architectures.

5.1.1 Soft Actor-Critic: Principles and Advantages

SAC represents a significant advancement in the field of DRL, particularly for scenarios requiring robust exploration and stable learning [65]. Unlike traditional actor-critic methods that focus solely on reward maximization, SAC incorporates maximum entropy reinforcement learning principles, seeking to maximize both the expected return and the policy entropy simultaneously. This dual objective is mathematically formulated as:

$$J(\pi) = \sum_{t=0}^{\infty} \mathbb{E}_{(s_t,a_t)\sim\rho_\pi}[\gamma^t(r(s_t,a_t) + \alpha\mathcal{H}(\pi(\cdot|s_t)))] \tag{5.1}$$

where $\mathcal{H}(\pi(\cdot|s_t)) = -\log\pi(a_t|s_t)$ represents the policy entropy, and α is the temperature parameter that controls the trade-off between exploration and exploitation [66].

The key advantages of SAC include:

- **Enhanced Exploration**: The entropy regularization encourages the agent to explore diverse actions, preventing premature convergence to suboptimal policies—a critical advantage in complex network planning scenarios with large discrete action spaces.

- **Sample Efficiency**: SAC's off-policy learning capability allows it to reuse past experiences efficiently, significantly reducing the number of environment interactions required compared to on-policy methods [67].

- **Stability**: The automatic temperature adjustment mechanism in SAC provides more stable training dynamics compared to traditional policy gradient methods, particularly beneficial for high-dimensional optimization problems like IAB network planning.

The original SAC algorithm was designed for continuous action spaces. However, network planning problems inherently involve discrete decisions. Recent research has explored adaptations of SAC to discrete domains [68], demonstrating that the entropy regularization benefits can be preserved through categorical policy representations and appropriate sampling mechanisms.

5.1.2 Transfer Learning in Reinforcement Learning

TL has emerged as a crucial paradigm for addressing the data efficiency and adaptability challenges in RL applications. The fundamental premise of TL is to leverage knowledge acquired in one domain (source task) to accelerate learning in a related but different domain (target task) [69]. In the context of network planning, this translates to reusing models trained on specific network configurations to rapidly adapt to new deployment scenarios.

The mathematical foundation of TL in RL can be understood through the concept of task similarity and knowledge transferability. Given a source MDP $\mathcal{M}_s = (\mathcal{S}_s, \mathcal{A}_s, P_s, R_s, \gamma_s)$ and a target MDP $\mathcal{M}_t = (\mathcal{S}_t, \mathcal{A}_t, P_t, R_t, \gamma_t)$, the goal is to transfer knowledge from an optimal policy π_s^* learned in $\mathcal{M}_s$ to accelerate learning of π_t^* in $\mathcal{M}_t$ [70].

5.1.3 Integration with Open RAN Architectures

The evolution toward Open Radio Access Network (RAN) architectures introduces new opportunities and challenges for intelligent network planning. Open RAN's disaggregated architecture, with its separation of hardware and software components and standardized interfaces, enables more flexible and intelligent network management approaches [71].

Our proposed framework is specifically designed to integrate with Open RAN control architectures through the following mechanisms as shown in 5.1:

- **Non-RT RIC Integration**: The DRL-TL planner operates within the Non-RT RIC, generating optimal deployment policies based on long-term network optimization objectives.
- **Near-RT RIC Coordination**: Deployment policies are communicated to the Near-RT RIC via the A1 interface, enabling real-time control decisions through specialized xApps.
- **Closed-Loop Optimization**: Network performance metrics collected through the O1 interface provide continuous feedback for model refinement and adaptation.

This integration enables dynamic, intelligent network planning that can adapt to changing conditions while leveraging the flexibility and standardization benefits of Open RAN architectures.

5.1.4 Chapter Contributions and Organization

Building upon the insights gained from Chapters 3 and 4, this chapter makes the following key contributions:

- **Advanced DRL Adaptation**: Development of discrete-action adaptations for both SAC and TD3 algorithms, specifically tailored for IAB network planning with large action spaces and complex constraints.

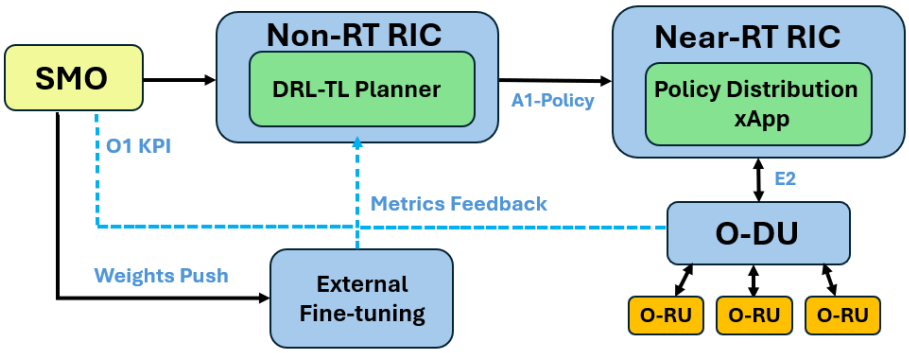

Figure 5.1: Functional placement of the proposed transferable-DRL deployment engine within the O-RAN control stack: the DRL-TL planner (Non-RT RIC) sends A1 policies to the Near-RT RIC, whose xApp drives O-DU/RUs over E2; KPIs return via O1 for continual fine-tuning of model weights.

- **Transfer Learning Framework**: Introduction of a comprehensive TL methodology that enables rapid adaptation of pre-trained models to new donor configurations, reducing training time by approximately 50% while maintaining deployment efficiency.

- **Multi-Algorithm Comparison**: Systematic evaluation of different DRL approaches (SAC, TD3, Dueling DQN) across various network scenarios, providing insights into their relative strengths and limitations. And convergence analysis and mathematical foundations for the proposed discrete SAC/TD3 algorithms, including the impact of TL on convergence properties.

- **Theoretical Foundations**: Rigorous convergence analysis demonstrating that our discrete SAC/TD3 adaptations maintain polynomial sample complexity and that transfer learning preserves convergence guarantees while accelerating training.

The remainder of this chapter is organized as follows: Section 5.2 presents the system model and problem formulation, building upon the foundations established in previous chapters. Section 5.3 details the proposed DRL algorithms and TL framework. Section 5.4 provides comprehensive experimental evaluation and analysis. Section 5.5 concludes the chapter with insights for future research directions.

5.2 System Model and Problem Formulation

5.2.1 System Model

The system model and problem formulation used in this chapter build directly on the baseline framework established in Chapter 3 (Sec. 3.2–3.2.2). We consider an urban IAB deployment operating in the 60 GHz mmWave band, where a fixed set of fibre-connected donor nodes provides wireless backhaul connectivity to deployed IAB nodes. Candidate IAB sites are restricted to a finite set of feasible urban locations (e.g., lamp-posts, bus-stops, or other street furniture), typically selected to facilitate LoS transmission whenever possible. Figure 5.2 illustrates an example deployment scenario.

The communication model follows Chapter 3, Sec. 3.2.1. In particular, the received power, total antenna gain, aggregate propagation loss, and thermal noise are computed using the same expressions as in Chapter 3 (see Eqs. (3.1)–(3.4)), which define the received power P_r, cumulative antenna gain G_{all}, total attenuation L_{all}, and thermal noise N_0, respectively. The access/backhaul antenna assumptions (omni-directional access gain and directional sector backhaul gain) are also adopted without change from Chapter 3.

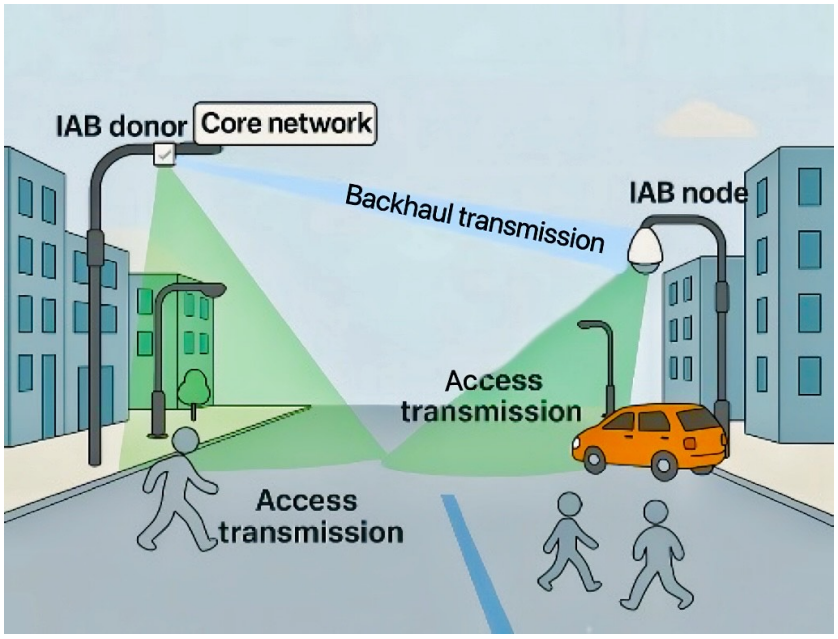

Figure 5.2: A sample mmWave integrated access and backhaul (IAB) network model.

Common parameter values (e.g., carrier frequency, antenna gains, transmit power, oxygen absorption, and traffic demand range) follow Table 3.1.

From Chapter 5 onward, we consider learning under dynamic environments where the deployment scenario may vary across episodes/tasks (e.g., different donor layouts), while preserving the same candidate-site representation and the same SNR-based feasibility modelling. Moreover, instead of enforcing full (100%) coverage at the grid level as in the strict baseline case, we adopt a more practical coverage-ratio target:

Let I denote the set of pre-deployed donors, J the set of candidate IAB sites, and K the set of grid cells, consistent with Chapter 3. Let $\alpha_j \in \{0, 1\}$ indicate whether a node is deployed at candidate site $j \in J$. Let C_{ik} and C_{jk} denote donor/node SNR-based coverage indicators for grid cell $k \in K$, as defined in Chapter 3 (cf. the SNR-based coverage indicators in Sec. 3.3).

We define the coverage ratio as

$$\mathcal{C} = \frac{1}{|K|} \sum_{k \in K} \mathbb{1} \left(\sum_{i \in I} C_{ik} + \sum_{j \in J} \alpha_j C_{jk} \geq 1 \right), \tag{5.2}$$

All remaining feasibility constraints (backhaul connectivity and data-rate feasibility with overhead) and the objective of minimising the number of deployed IAB nodes follow the baseline formulation in Chapter 3 (Sec. 3.3) and are adopted here without modification; the only chapter-specific change in the coverage requirement is the threshold-based constraint.

5.3 Proposed Solution Framework

In this section, we introduce our solution framework for this problem. Our goal is to minimize the number of deployed IAB nodes while keeping the coverage that the deployed network must cover. To tackle the inherent complexity of the deployment decision, we formulate the problem as a MDP with a discrete action space. We then present two state-of-the-art DRL algorithms: twin delayed deep TD3 and SAC, each adapted and enhanced for our specific problem domain.

5.3.1 MDP Formulation

In order to utilize DRL to derive the optimal solution, we formulate the MDP based on the problem formulation in section 5.2:

State Space $\mathcal{S}$: The state space $\mathcal{S}$ is represented as a tuple:

$$s = (\mathbf{D}, \mathbf{R}, \mathbf{M}),$$

where:

- **Deployment Status of Nodes** (D):

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix}$$

  where d_{ij} is 1 if a node is deployed at location (i, j), and 0 otherwise.

- **Data Rates** (R):

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

  where r_{ij} is the data rate left of the node at location (i, j).

- **Node Connectivity Matrix M:** This matrix captures the backhaul connectivity between nodes. Its element m_{ij} represents the number of backhaul connections provided from node i to node j. If no connection exists, then $m_{ij} = 0$:

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix}$$

Action Space $\mathcal{A}$: The discrete action space is defined as:

$$\mathcal{A} = \{a_0, a_1, \ldots, a_N\},$$

where for $j = 0, 1, \ldots, N-1$:

$$a_j : \quad \text{Deploy a node at candidate site } j,$$

and

$$a_N : \quad \text{Do nothing.}$$

Thus, action space is $|\mathcal{A}| = N + 1$. Unlike prior works that either ignore infeasible actions or apply crude penalties, we introduces a two-fold, end-to-end differentiable mechanism during training:

- **Action Masking**: we remove any action a violating $Y_{ij} \leq \alpha_j$ or data-rate constraints.
- **Hard Penalty**: any infeasible (s, a) incurs a large negative reward $r \leftarrow R_{\min} - \Delta$ (with $\Delta \gg R_{\max}$), ensuring violations are never selected in the limit.

It is worth noting that logits masking before softmax does not change the Bellman backup or gradient estimator, and the hard penalty only affects off-policy returns.

Transition Dynamics $P(s'|s, a)$: The state transition function $P : \mathcal{S} \times \mathcal{A} \to \mathcal{P}(\mathcal{S})$ governs how the state evolves when an action is taken. For a given state $s = (\mathbf{D}, \mathbf{R}, \mathbf{M})$ and action a_j (with $j < N$), the next state s' is

given by:

$$s' = (\mathbf{D}', \mathbf{R}', \mathbf{M}'),$$

and where $\mathbf{D}'$, $\mathbf{R}'$ and $\mathbf{M}'$ are updated according to the coverage map and connectivity. When the "do nothing" action a_N is chosen, the state remains unchanged.

Reward Function $r(s, a)$: The immediate reward function $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is designed to incentivize coverage improvements and penalize excessive node deployments based on our constraint analysis. It is defined as:

$$r(s, a) = B(s, a) - P(s, a),$$

where:

- $B(s, a)$ increases with the improvement in coverage percentage, which is defined as:

  $$B(s, a) = A_{\text{covered}} - A_{\text{total}}$$

  where A_{total} means the total area that needs to be covered and A_{covered} means the covered area after that action implementation.

- $P(S, a)$ (the penalty matrix) is given by:

  $$P(S, a) = \begin{bmatrix} \kappa_1 \cdot \text{B(s,a)} \\ \kappa_2 \cdot \left[\max(0, \sum_{j=1}^{N} D_j - N_{\text{ref}})\right]^p \end{bmatrix}^T \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

  where $\kappa_1, \kappa_2 > 0$ are scaling constants, N_{ref} is a reference deployment number from[31] to discourage excessive node deployment, and p ad-

justs the penalty's growth rate.

Our proposed solution framework models the network planning problem as an MDP with a discrete action space and ensures that every grid cell $k \in K$ is covered so that the signal quality meets a minimum service threshold, and that robust backhaul connectivity is established between IAB donors and IAB nodes (or among IAB nodes) with sufficient data rate to support local access and downstream traffic.

5.3.2 TD3 and SAC Adaptation for Discrete Actions

Our framework leverages two DRL methods originally developed for continuous control. However, network planning problems require decisions that are inherently discrete, such as whether or not to deploy a node at a candidate location. To address this, we first propose specific adaptations for TD3.

TD3 Adaptation:

TD3 is an actor-critic algorithm originally designed for continuous action spaces. It contains two main components: an actor network and two critic networks that provide more conservative estimates (prevent overestimations) of Q values and delayed policy updates to stabilize training. Although there are few studies on discrete actions, these algorithms cannot cope with large action spaces. Motivated by this observation, we propose a novel framework which attempts to solve the action adaption challenges.

Since our decision (deploy a node or not) is not "any number" but a clear choice from a list of options, we change the TD3 actor. Instead of outputting one continuous number, the actor now produces a vector of numbers

logits corresponding to one possible action in our discrete set $a \in \mathcal{A}$ (with $\mathcal{A} = \{0, 1, \ldots, N-1\}$). We modify the actor to output a vector of logits corresponding to each possible action, which are converted into a probability distribution using the softmax function:

$$P(a|s) = \frac{\exp\left(\pi_\theta(s)_a\right)}{\sum_{a' \in \mathcal{A}} \exp\left(\pi_\theta(s)_{a'}\right)}. \tag{5.3}$$

This equation essentially converts raw scores into probabilities that add up to 1. Unlike [72], which employs Gumbel-Softmax for continuous-to-discrete action mapping in general control tasks, our approach integrates an epsilon-greedy strategy with a constraint-driven action masking mechanism tailored for IAB network planning. Specifically, we incorporate hard constraints (e.g., backhaul and data rate requirements) directly into the action selection process by masking infeasible actions before softmax application and applying a hard penalty for constraint violations. This ensures that only feasible deployment decisions are considered, significantly improving exploration efficiency in high-dimensional discrete action spaces (up to 401 actions in our setup).

The twin critic networks, $Q_{\phi_1}(s, a)$ and $Q_{\phi_2}(s, a)$, parameterized by ϕ_1 and ϕ_2 respectively, estimate the expected return for each state-action pair. To mitigate overestimation bias, the target for the critic update is computed as:

SAC Adaptation:

SAC is an off-policy actor-critic algorithm based on the maximum entropy framework. Its objective is twofold: to maximize the expected return and to maximize the entropy of the policy, which leads to more robust exploration.

Algorithm 5: Discrete Action Selection in Adapted TD3 with Hard Constraints

Input: Current state s, actor network π, discrete action set $\mathcal{A}$, exploration parameter ϵ
Output: Selected action a

Compute logits vector: $\mathbf{a}_c \leftarrow \pi(s;\theta)$;
Generate action mask: $\mathbf{m}$ // $\mathbf{m}[a] = 0$ `if` a `valid,` $-\infty$ `otherwise`
Apply mask: $\mathbf{a}_c \leftarrow \mathbf{a}_c + \mathbf{m}$;
Convert masked logits to probabilities: $\mathbf{p} \leftarrow \text{softmax}(\mathbf{a}_c)$;
if *random number* $< \epsilon$ **then**
    Select a random action $a \in$ valid_actions;
else
    Sample discrete action $a \sim \text{Categorical}(\mathbf{p})$;
return the selected action a;

Traditionally, SAC models its policy as a Gaussian distribution for continuous action spaces. To apply SAC to our discrete-action domain, where the choice is whether or not to deploy a node at a candidate location, we adapt the algorithm:

1. **Policy Network Modification:** Similar to the TD3 adaptation, the SAC policy network is modified to output a vector of logits corresponding to each discrete action in $\mathcal{A}$.

2. **Softmax Conversion:** These logits are then passed through the softmax function (as shown in (5.3)) to convert them into a categorical probability distribution. This probability distribution represents the likelihood of each discrete action.

3. **Categorical Policy and Action Sampling:** The policy is then treated as a categorical distribution, from which we sample an action. This directly provides us with a discrete decision.

4. **Entropy-Regularized Objective:** The SAC objective function is adjusted to include an entropy term that encourages exploration:

$$J_\pi(\theta) = \mathbb{E}_{s\sim\mathcal{D}}\left[\sum_{a\in\mathcal{A}} P(a|s)\left(\alpha \log P(a|s) - Q_\phi(s,a)\right)\right], \tag{5.4}$$

   where α is the temperature parameter balancing exploration (via the entropy term) and exploitation (maximizing the Q-value). The Q-network $Q_\phi(s,a)$ estimates the value of taking action a in state s.

5. **Training Process:** As with TD3, transitions are stored in a replay buffer. In each training iteration, a mini-batch is sampled to update both the Q-networks and the policy network. Furthermore, a soft update mechanism is used to smoothly update the target networks, ensuring training stability.

Algorithm 6 is the pseudocode for the SAC training loop.

5.3.3 Motivation for Transfer Learning

In dynamic network planning scenarios, it is not hard to find varying donor distributions present significant challenges for model training. Training a DRL model from scratch for each new donor setting is computationally intensive and time-consuming, requiring the model to relearn fundamental patterns such as spatial relationships, connectivity dynamics, and resource allocation trade-offs. This inefficiency is particularly pronounced in complex environments where rapid adaptation is essential. TL offers a promising solution by adapting a pre-trained model to new donor configurations with minimal retraining. This approach leverages previously learned spatial features,

Algorithm 6: Discrete SAC Training

Input: Replay buffer B, policy network π, Q-networks Q_1 and Q_2, target networks Q'_1 and Q'_2, number of episodes N, steps per episode T, and soft-update parameter τ

Output: Trained policy network π

Initialize replay buffer B, policy network π, Q-networks Q_1 and Q_2 with random parameters, target networks Q'_1 and Q'_2;
for *episode* $= 1$ **to** N **do**
    Initialize state s_0;
    for *step* $= 1$ **to** T **do**
        Sample action $a_t \sim \pi(s_t)$;
        Execute action a_t; observe reward r_t and s_{t+1};
        Store transition (s_t, a_t, r_t, s_{t+1}) in B;
        Sample a mini-batch from B;
        Update Q-networks Q_1 and Q_2 using Q'_1 and Q'_2;
        Update policy network π;
        Soft-update target networks:;
            $Q'_i \leftarrow \tau Q_i + (1-\tau)Q'_i, \quad \forall\, i \in \{1, 2\}$;
    end
end
return *Trained policy network* π

link optimization, or coverage strategies, thereby accelerating convergence and enhancing overall adaptability in dynamic network environments. Our TL framework provides a possible methodology applicable to multiple DRL models, including SAC and DQN, across diverse network planning contexts:

- **Pre-Training Phase:** DRL agents are initially trained in a baseline environment with a standard configuration. This phase focuses on learning general features, such as spatial organization, connectivity patterns, and resource trade-offs, which are broadly applicable across environments. We call it source environment, where all weights are saved (θ^*) for future use.

- **Fine-Tuning Phase:** The pre-trained DRL models are transferred to a target environment with a different donor configuration. Fine-tuning involves adjusting key hyperparameters—such as learning rate, exploration rate, or update frequency—to suit the new setup. To preserve the general knowledge learned during pre-training, early network layers may be frozen, allowing higher-level layers to adapt to the new environment's specifics.

Figure 5.3 shows how to leverage TL to enable rapid adaptation to heterogeneous configurations.

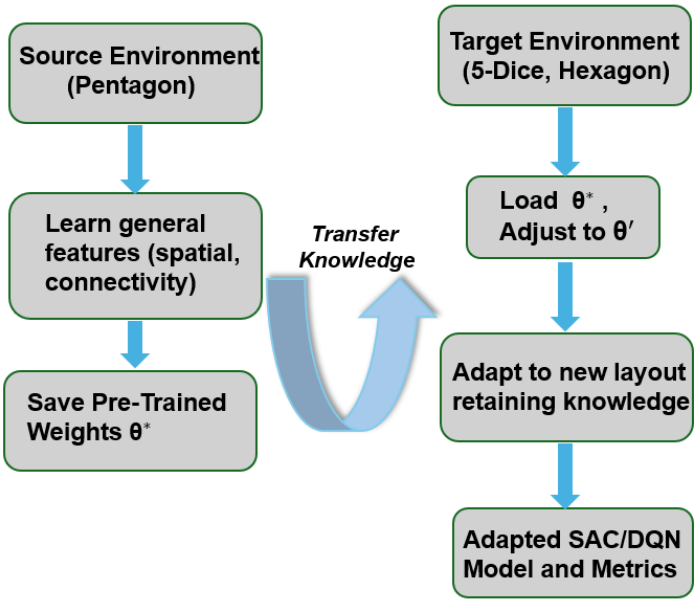

Figure 5.3: Systematic Flow Diagram of TL Framework for Adapting DRL Models in Heterogeneous Network Environments

To evaluate the effectiveness of our TL framework, we will perform experiments in different donor configurations, and the following performance metrics will be considered:

- Convergence speed (the number of episodes required for the transferred model to stabilize).
- Final achieved coverage percentage.
- Number of IAB nodes deployed.

5.3.4 Convergence Analysis and Theoretical Guarantees

The theoretical foundations of our proposed algorithms are critical for understanding their behavior in large-scale IAB deployment scenarios. This section establishes formal convergence guarantees for our discrete adaptations of SAC and TD3, and demonstrates that transfer learning preserves these properties while providing acceleration benefits.

Convergence of Discrete SAC/TD3

We first establish that our discrete action space adaptations maintain the convergence properties of their continuous counterparts. Under standard assumptions for finite MDPs (bounded rewards, sufficient exploration, and Robbins-Monro learning rates), we prove the following:

Theorem 3.1 (Main Convergence Result): The discrete SAC and TD3 algorithms for IAB deployment converge to an ε-optimal policy with sample complexity:

$$T = \widetilde{\mathcal{O}}\left(\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^2\varepsilon^2}\right) \tag{5.5}$$

This polynomial complexity ensures practical scalability for realistic IAB networks with hundreds of candidate locations. The proof follows from the

contraction property of the Bellman operator and careful analysis of the discretization error introduced by our softmax-based action selection mechanism.

Transfer Learning Acceleration

A key theoretical contribution is proving that transfer learning preserves convergence while providing quantifiable acceleration:

Theorem 3.2 (TL Acceleration): When transferring from a source to target IAB configuration, if the transferred Q-function initialization satisfies $\|\widehat{Q}_0^{\text{TL}} - Q^*_{\text{target}}\|_\infty \leq \|\widehat{Q}_0^{\text{random}} - Q^*_{\text{target}}\|_\infty$, then transfer learning reduces the number of required training episodes by:

$$\Delta T = \frac{1}{1-\gamma} \log \left(\frac{\|\widehat{Q}_0^{\text{random}} - Q^*\|_\infty}{\|\widehat{Q}_0^{\text{TL}} - Q^*\|_\infty} \right) \tag{5.6}$$

This theoretical result aligns with our empirical findings of approximately 50% training time reduction.

Practical Implications

These theoretical guarantees provide several important insights for practical deployment:

- The logarithmic dependence on the action space size $|\mathcal{A}|$ confirms scalability to networks with thousands of candidate positions
- SAC's entropy regularization provides an additional factor of $\mathcal{O}(\log |\mathcal{A}|)$ improvement in exploration efficiency

- Transfer learning offers provable benefits without compromising optimality guarantees

Based on the above, the results are discussed and analyzed in the next section.

5.4 Results and Analysis

5.4.1 Experimental Settings

In our experimental evaluation, we simulate a 60 GHz urban environment covering an area of 1000 m × 1000 m. The service area is partitioned into a grid with candidate IAB node locations spaced at 50 m intervals. The wireless propagation model follows key parameters from the 3GPP TR 38.901 standard: a 60 GHz center frequency, high free-space path loss and additional atmospheric attenuation, and directional beamforming gains on the backhaul links. Moreover, a grid cell is declared covered when the received power exceeds −80 dBm—ensuring reliable throughput.

To benchmark our DRL approaches, we compare four models:

- **Heuristic Model:** A greedy algorithm [31] that sequentially deploys nodes by maximizing immediate coverage improvement under data rate constraints and solved by Gurobi solver. Although fast and interpretable, its myopic nature leads to suboptimal long-term performance.

- **Dueling-DQN with Action Elimination:** This model leverages a dueling architecture that decomposes Q-values into state value and advantage components, combined with an action elimination mechanism to prune non-promising candidate placements [33].

- **Proposed TD3 Model:** TD3 being proposed in last section.
- **Proposed SAC Model:** SAC being proposed in last section.

These two approaches (Greedy and Dueling-DQN) form our set of benchmark models. The greedy method provides a baseline heuristic performance, while the three DRL methods allow us to evaluate learning-based optimization under different algorithmic paradigms. All learning models are trained using the same reward structure as shown in the last section: encouraging meeting coverage threshold with minimal nodes. Training is run until convergence (when additional episodes yield negligible improvement in coverage).

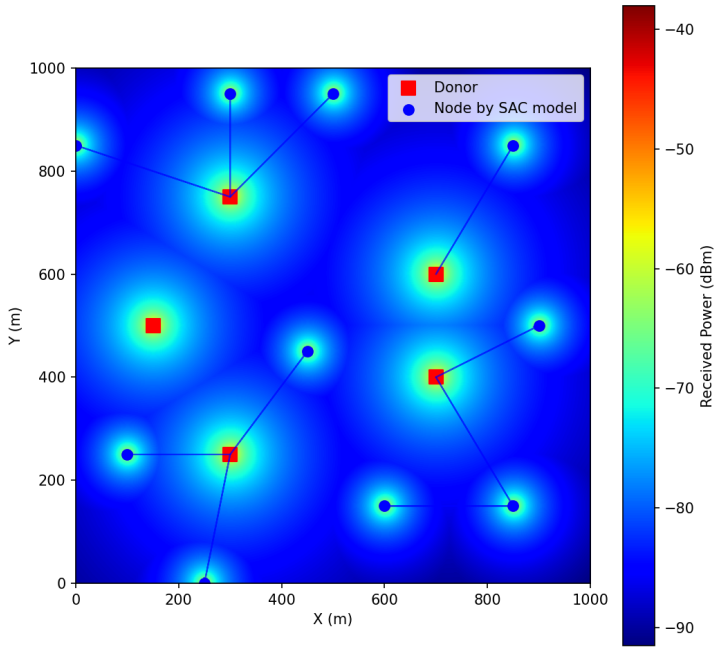

Figure 5.4: SAC-network planning result: 10 IAB nodes deployed achieving 98.93% coverage.

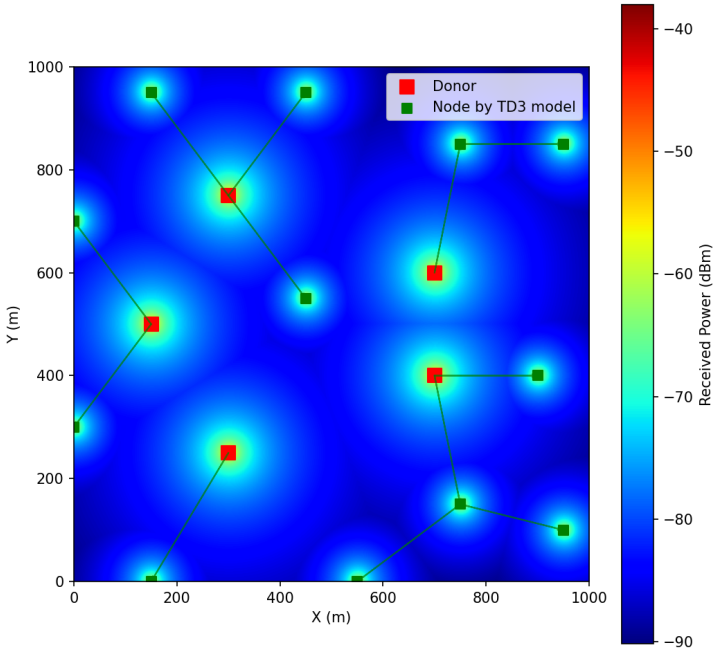

Figure 5.5: TD3-network planning result: 12 IAB nodes deployed with 98.00% coverage.

5.4.2 Synthesis of Findings

We evaluate the four approaches (SAC, TD3, Dueling-DQN, and Greedy), each trained to achieve a coverage threshold of 98%. Performance is measured by the final coverage percentage, the number of deployed IAB nodes, and the spatial distribution of received power. Figures 5.4, 5.5, and 5.6 display example deployment results. Table 5.1 summarizes the main hyperparameters and network configurations used in our experiments. Figure 5.4 illustrates the deployment result of the SAC model for the pentagon donor configuration. In this configuration, the SAC model requires the deployment

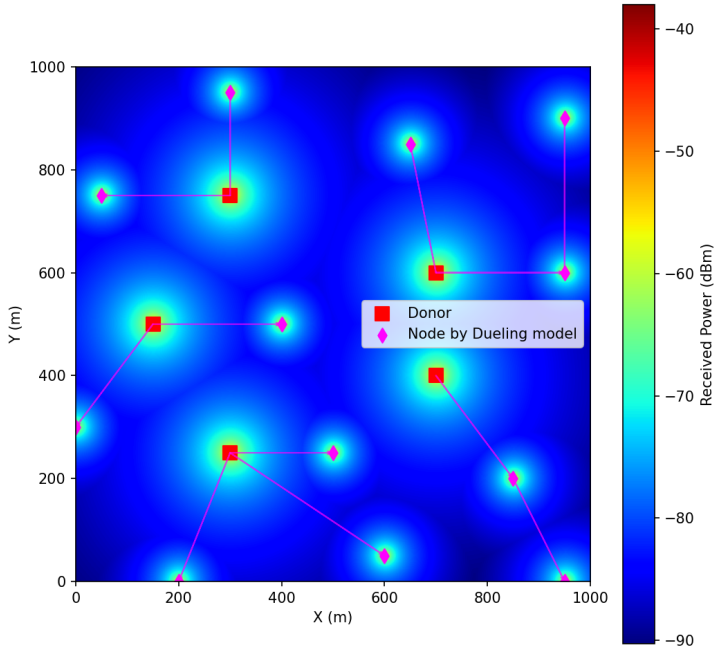

Figure 5.6: Dueling-DQN network planning result: 12 IAB nodes deployed achieving 98.88% coverage.

of an additional 10 IAB nodes to achieve nearly 99% coverage. In contrast, Figure 5.5 shows the outcome for the TD3 model under the same donor configuration, which deploys 12 nodes to reach 98.00% coverage, while Figure 5.6 demonstrates that the Dueling-DQN model also deploys 12 nodes, achieving 98.88% coverage. Training convergence curve will be shown in the next TL section.

From the above figures, we observe that the SAC model achieves the best balance between coverage and node minimization (with 10 IAB nodes reaching nearly 99% coverage), while TD3 and Dueling-DQN deploy slightly more nodes to meet the coverage requirement. In multi-hop IAB network plan-

Table 5.1: Hyperparameter Settings for SAC

Parameter	Value/Description
Training Configuration	
Training Episodes	6,000
Max Steps per Episode	50
Batch Size	512
Optimizer	Adam
Learning Rate (Policy & Q-Networks)	1×10^{-5}
Alpha Optimizer Learning Rate	1×10^{-5}
Network Architecture	
Policy Network (Discrete SAC)	[1024, 512, 256]
Q-Network (Q1 & Q2)	[1024, 512, 256]
SAC Algorithm Parameters	
Discount Factor γ	0.99
Soft Update Coefficient τ	0.005
Replay Buffer Size	50,000
Initial Entropy Coefficient	0.1
Environment Parameters	
Uncovered Area Penalty Weight	1.0
Deployment Penalty Weight	0.5
Coverage Threshold (for early stop)	98%

ning, each decision essentially involves choosing whether to deploy a node within a discrete set of candidate locations. The number and distribution of these donor positions directly determine the size of the action space. To investigate whether donor sparsity affects algorithmic performance, we considered three donor layouts (Five-Dice, Pentagon, and Hexagon) under three sparsity conditions (Sparse, Moderate, and Narrow).

5.4.3 Evaluation of Multi-Hop IAB Deployment Algorithms under Different Sparse Level Donor Configurations

Tables 5.2–5.4 summarize the performance of each algorithm across different donor configurations. The values represent the best results achieved in terms of the number of IAB nodes deployed and the coverage percentage from 100 test runs.

Performance Comparison in Sparse Donor Settings

In sparse donor settings (donors deployed at the corners), the uncovered area is large, making it challenging to achieve global optimality. SAC achieves superior performance across all configurations. More specifically, in the Sparse Five-Dice arrangement, SAC requires only 10 IAB nodes to reach coverage threshold, while it requires only 9 IAB nodes in the Sparse Pentagon and Hexagon configurations. This excellent performance is attributed to SAC's entropy-regularized policy optimization, which balances exploration and exploitation effectively within a high-dimensional action space. TD3's performance in sparse settings is constrained by its adaptation mechanism for discrete action spaces as shown in Algorithm 5, computing logits vectors and converting them to probabilities via softmax followed by categorical sampling creates precision losses in high-dimensional spaces, affecting optimal deployment location selection. These discretization issues limit TD3's ability to make optimal deployment decisions. Dueling-DQN faces challenges in sparse donor settings due to its distance-based action elimination method. Specifically, when candidate nodes are abundant, the distance-based filtering

mechanism may be overly restrictive, resulting in suboptimal performance compared to SAC.

Table 5.2: Nodes Deployed in Sparse Donor Settings (IAB Node Deployment and Donor Nodes Excluded)

Donor Geometry	SAC	TD3	Dueling-DQN	Greedy
Five-Dice	**10**	12	11	13
Pentagon	**9**	11	10	13
Hexagon	**9**	11	10	12

Performance Comparison in Moderate Donor Settings

In moderate donor settings (donors deployed centrally), the uncovered area is significantly reduced, and SAC continues to demonstrate good performance in all arrangements, requiring only 9 and 8 IAB nodes. TD3's performance improves in moderate settings but still falls short of SAC. Dueling-DQN performs better in moderate settings if minimum number of deployed nodes is main objective as its distance-based action elimination becomes more effective in more structured action spaces. When considering the comprehensive impact of all deployed nodes, Dueling-DQN's separation of state value and action advantage functions may not accurately reflect overall network performance.

Table 5.3: Nodes Deployed in Moderate Donor Settings (IAB Node Deployment and Donor Nodes Excluded)

Donor Geometry	SAC	TD3	Dueling-DQN	Greedy
Five-Dice	**9**	11	10	13
Pentagon	**8**	10	9	12
Hexagon	**8**	9	9	10

Performance Comparison in Narrow Donor Settings

In narrow donor settings (donors deployed near the center with significant overlap), certain models are able to leverage the overlapping regions more effectively. SAC maintains its superior performance in both the Five-Dice and Pentagon settings. However, Dueling-DQN achieves the best performance specifically in the narrow Hexagon configuration, deploying 9 IAB nodes. This unique advantage of Dueling-DQN in the narrow Hexagon setting can be attributed to its distance-based action elimination method. TD3's performance in narrow settings improves compared to sparse environments but still lags behind both SAC and Dueling-DQN.

Table 5.4: Nodes Deployed in Narrow Donor Settings (IAB Node Deployment and Donor Nodes Excluded)

Donor Geometry	SAC	TD3	Dueling-DQN	Greedy
Five-Dice	**9**	12	10	13
Pentagon	**8**	11	9	12
Hexagon	9	12	**9**	9

Training Time Comparison Across Donor Geometries

Table 5.5 summarizes the computational training time requirements across different donor geometries. SAC consistently demonstrates the most efficient training process, requiring only 4.8–5.5 hours across all donor geometries. TD3 requires moderately more training time (6.9–9.4 hours), while Dueling-DQN demands significantly longer training periods (10.5–24.8 hours). The Greedy approach is the most computationally intensive, requiring 18.0–29.2 hours across different configurations.

In summary, our analysis reveals that donor sparsity and geometric config-

Table 5.5: Approximate Training Time (hours) for Different Donor Geometries. The values are mean training times from different sparse-level donors.

Donor Geometry	SAC	TD3	Dueling-DQN	Greedy
Five-Dice	5.1	8.1	17.8	23.1
Pentagon	5.3	8.5	18.5	24.3
Hexagon	5.0	7.5	14.4	21.9

uration significantly influence both the absolute performance of node deployment algorithms and the relative advantages of different approaches. SAC offers the most robust performance across varied conditions while also providing the most computationally efficient training process for achieving near-optimal IAB node deployments. For scenarios with narrow hexagonal donor arrangements, Dueling-DQN presents a viable alternative that can achieve superior coverage, albeit at the cost of significantly increased training time. In certain highly structured environments, such as narrow Hexagon configurations, even simple Greedy algorithms can produce near-optimal results, providing additional algorithm choices for specific application scenarios.

5.4.4 Transfer Learning in Network Planning

Training a network planning model from scratch for each new IAB donor deployment scenario is computationally inefficient and time-consuming. Each donor configuration, defined by a fixed set of base station placements and densities, presents a distinct environment for IAB node deployment. Without prior knowledge, a model must rediscover optimal strategies for every scenario, leading to significant resource demands[73]. Our approach is motivated by the insight that certain features—such as propagation characteristics and deployment heuristics—are shared across different donor layouts. By reusing

pre-trained representations and strategies, we hypothesize that models can adapt more rapidly and efficiently to new configurations. To explore this, we conducted experiments to investigate how TL performs across varying shapes and numbers of donor settings, specifically transferring knowledge from a Pentagon donor layout to Hexagon and Five-Dice configurations.

We pre-trained SAC and Dueling DQN models on a Pentagon donor layout (5 donors), which served as the source environment, using 10,000 episodes to learn general features like spatial patterns and connectivity dynamics. For SAC, we set the entropy temperature α to 0.2 to encourage exploration, while for Dueling DQN, we used an ϵ-greedy exploration rate of 0.1. These models were then fine-tuned on two target settings: a Five-Dice donor layout (5 donors) and a Hexagon donor layout (6 donors), requiring only 5,000 episodes each due to transferred knowledge. During fine-tuning, the first two layers of both models were frozen to preserve pre-trained features, and exploration was reduced ($\alpha = 0.1$ with dynamic adjustment for SAC, $\epsilon = 0.05$ for Dueling DQN) to focus on exploiting learned strategies while adapting to new donor geometries and counts. The experiments aimed to assess whether strategies learned in the Pentagon layout could accelerate convergence and enhance efficiency in these distinct environments. The results are presented in two parts: the first two describe the transfer from Pentagon to Five-Dice, and the subsequent two detail the transfer from Pentagon to Hexagon. A noteworthy point is that in the number of nodes curves below, the IAB donors are also considered in the number of nodes deployed (Pentagon: 5 donors, Hexagon: 6), which leads to a larger number of convergences.

Transfer from Pentagon to Five-Dice Donor Setting

For the Five-Dice donor setting, we similarly compared SAC and Dueling DQN with and without TL. The performance is illustrated in Figures 5.7 and 5.8, where plot coverage percentage and deployed nodes are shown against training episodes, respectively.

- **SAC with TL**: Starting with 92% coverage and approximately 32 deployed nodes, SAC leverages pre-trained strategies from the Pentagon layout to rapidly achieve 98% coverage and reduce the node count to 15–16 within 2,000 episodes, eventually stabilizing at 99% coverage. This efficiency highlights SAC's ability to balance coverage maximization and node minimization through effective exploration and exploitation of transferred knowledge.

- **SAC without TL**: Beginning at 97% coverage with nearly 40 nodes, SAC requires thousands of episodes to exceed 98% coverage and reduce nodes to 16, stabilizing at 99%. The initial over-deployment and slower convergence emphasize the computational burden of learning without prior knowledge.

- **Dueling DQN with TL**: Initiating at 99.6% coverage and 30 nodes, Dueling DQN reaches 98% coverage with around 17 nodes within 5910 episodes, stabilizing at 98.6%. While TL accelerates adaptation, its value-based approach results in a slightly slower and less optimal adjustment compared to SAC.

- **Dueling DQN without TL**: Starting at 92% coverage with 25 nodes, it takes over 7800 episodes to approach 98% coverage and reduce nodes

to 16–17. This prolonged learning phase underscores the significant advantage provided by TL.

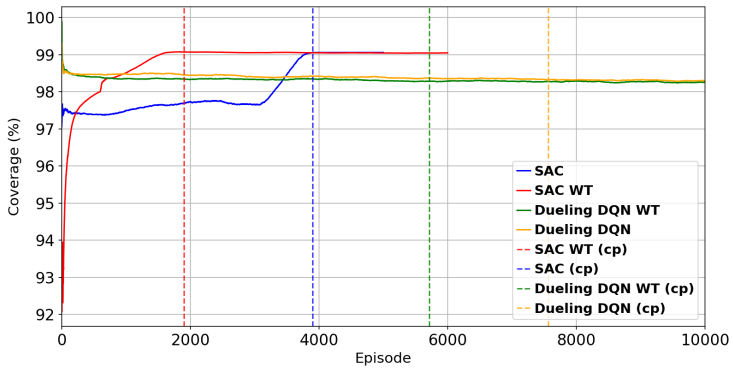

Figure 5.7: Pentagon Pre-trained model transferred to Five-Dice: Coverage vs. Training episodes (WT: With Transfer; cp: Convergence Point).

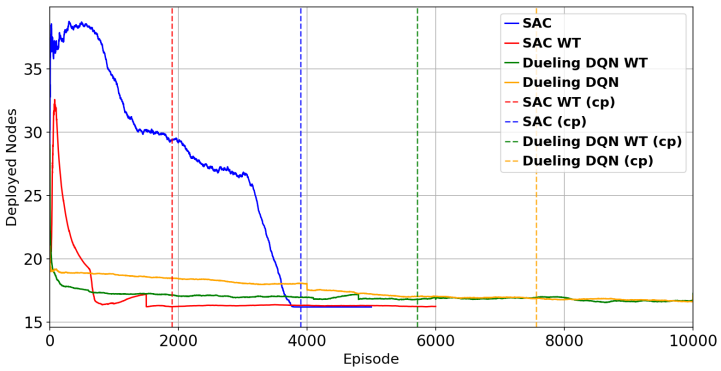

Figure 5.8: Pentagon Pre-trained model transferred to Five-Dice: Deployed nodes vs. Training episodes (WT: With Transfer; cp: Convergence Point).

TL markedly improves performance in the Five-Dice setting by enabling both algorithms to reuse Pentagon-derived features, reducing the need for extensive exploration. SAC with TL outperforms Dueling DQN, achieving higher coverage (99% vs. 98.2%) with almost the same nodes (16) in much

fewer episodes (4000 vs. 5910). SAC's policy-based optimization appears better suited to exploit pre-trained knowledge for adapting to shape differences, whereas Dueling DQN's cautious value estimation leads to a more gradual improvement. Overall, TL halves training time—roughly a 50% reduction for SAC in the number of episodes required to converge.

Transfer from Pentagon to Hexagon Donor Setting

The Hexagon donor setting introduces greater complexity, differing from the Pentagon in both shape and number of donors. We assessed how TL aids SAC and Dueling DQN in navigating these dual differences, with results shown in Figures 5.9 and 5.10.

- **SAC with TL**: Starting at 99% coverage with approximately 21 nodes, SAC quickly surpasses 98% coverage and reduces nodes to 16 within 5,000 episodes, stabilizing at 99.1%. TL effectively mitigates the challenges posed by shape and donor count differences, enabling rapid and efficient adaptation.

- **SAC without TL**: Beginning at 99.2% coverage with 40 nodes, SAC reaches 98% coverage and reduces nodes to 15 after 5,800 episodes. The slower progress reflects the difficulty of learning from scratch in a more complex environment.

- **Dueling DQN with TL**: Initiating at 99% coverage with around 18 nodes, Dueling DQN achieves 98% coverage with 15 nodes within 7100 episodes, stabilizing at 98.5%. TL provides a notable boost, though adaptation remains slower than SAC due to its value-based nature.

- **Dueling DQN without TL**: Starting at 98% coverage with 22 nodes, it requires nearly 9,000 episodes to reduce nodes to 17. This extended convergence highlights the critical role of TL in complex settings.

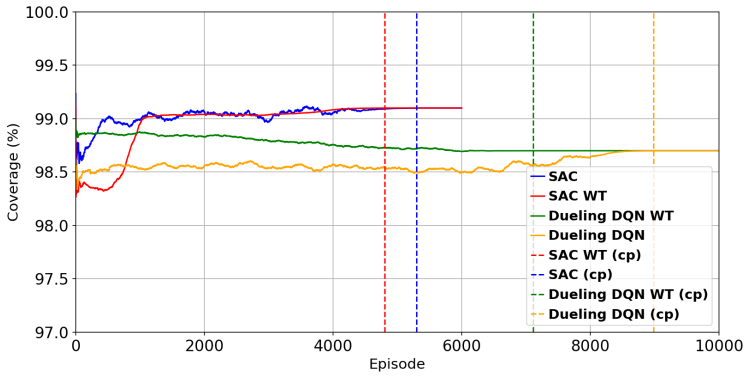

Figure 5.9: Pentagon pre-trained model transferred to Hexagon: Coverage vs. training episodes (WT: With Transfer; cp: Convergence Point).

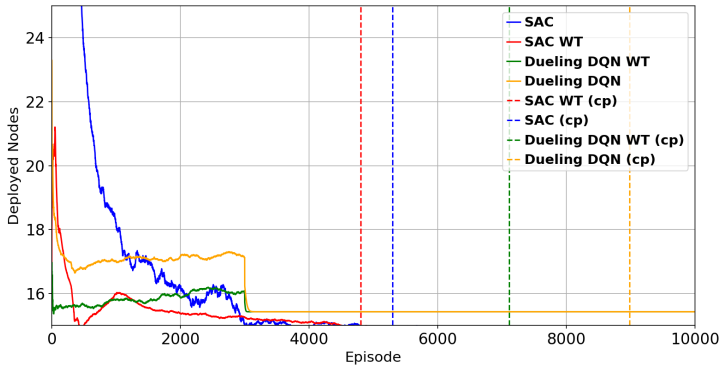

Figure 5.10: Pentagon pre-trained model transferred to Hexagon: Deployed nodes vs. training episodes (WT: With Transfer; cp: Convergence Point).

TL remains highly effective in the Hexagon setting, though the increased complexity slightly slows convergence compared to Five-Dice (e.g. 4800 vs. 2,000 episodes for SAC to reach convergence point). SAC with TL excels,

achieving 99% coverage with 16 nodes, outperforming Dueling DQN's 98.5% coverage with 17 nodes. The additional challenge of donor count differences reduces transfer efficiency slightly, but both algorithms leverage Pentagon knowledge to significantly outperform their no-transfer counterparts. SAC's faster and more robust adaptation suggests it better handles combined shape and number variations.

5.4.5 Algorithmic Trade-Offs and Practical Implications

Overall, SAC consistently minimizes node deployment and adapts fast to varying donor configurations, making it ideal when number of deployment is a primary concern. TD3, with its continuous control and smooth value estimates, performs competitively in some settings but may deploy extra nodes in densely covered regions. Dueling-DQN offers competitive performance under some conditions but can be less efficient in sparse settings. The Greedy algorithm, although conceptually simple, generally requires the highest number of nodes, reflecting its lack of foresight in balancing current benefits and future objective. From a practical standpoint, these findings imply that SAC—can further optimize deployments. Moreover, TL using SAC allows for rapid adaptation to new donor layouts, minimizing the need for extensive re-training in dynamic urban environments.

5.4.6 Limitations and failure modes of transfer learning.

Although TL accelerates convergence in the experiments above, its benefit can be reduced under two common conditions. First, a large domain gap (e.g., substantially different donor counts or topology statistics) can lead to negative transfer [69]. This is reflected by the Pentagon→Hexagon transfer (6 vs. 5 donors), which converges more slowly than Pentagon→Five-Dice (matched donor count): SAC requires roughly 4800 versus 2000 episodes, and the node-count gain is correspondingly smaller. Second, overly aggressive layer freezing during fine-tuning can prevent adaptation of spatial representations; in preliminary runs where all layers were frozen, training stagnated at about 94–95% coverage, indicating that higher-level layers should remain trainable when the target topology diverges from the source.

5.5 Conclusion

This chapter has presented a significant advancement in IAB network planning through the integration of advanced DRL algorithms with TL methodologies. Our discrete adaptations of SAC and TD3 algorithms, specifically tailored for IAB deployment constraints, demonstrate superior performance compared to traditional approaches, achieving up to 20% reduction in required IAB nodes while maintaining 98% coverage targets. The incorporation of entropy regularization in SAC proved particularly effective in balancing exploration and exploitation in high-dimensional discrete action spaces with up to 401 candidate locations.

The TL framework represents a crucial breakthrough in addressing computational efficiency challenges. By leveraging knowledge from pre-trained models, we achieved approximately 50% reduction in training episodes when adapting to new donor configurations, transforming IAB planning from a computationally prohibitive task to a practically deployable solution. Our theoretical analysis provides rigorous convergence guarantees with polynomial sample complexity $\mathcal{O}(|\mathcal{S}||\mathcal{A}|/\varepsilon^2)$, confirming the scalability of our approach. Furthermore, the seamless integration with Open RAN architectures through the Non-RT RIC ensures that our framework aligns with industry standards for next-generation network deployments.

However, our evaluation also revealed fundamental limitations of the current DRL-based approach. While effective for networks with hundreds of candidate locations, the polynomial complexity still poses challenges for city-scale deployments with thousands of potential sites. More critically, our MDP formulation treats each deployment decision independently, failing to fully exploit the inherent graph structure of IAB networks where connectivity patterns and multi-hop relationships fundamentally determine network performance.

These limitations motivate the exploration of GNN in the next chapter. The graph-structured nature of IAB topology—with nodes as vertices and backhaul links as edges—naturally aligns with GNN architectures. Specifically, Graph Attention Networks v2 can capture complex spatial dependencies and multi-hop propagation effects that are difficult to model in traditional DRL state representations. By learning directly on the network graph structure, GATv2 can potentially achieve better scalability through induc-

tive learning, enable zero-shot generalization to unseen network sizes, and provide interpretable attention mechanisms that reveal critical connectivity patterns. This graph-centric approach promises to overcome the scalability barriers identified in this chapter while maintaining the adaptability benefits we have demonstrated.

6 GATv2-Powered Graph Neural Networks for Large-Scale IAB Optimization

6.1 Introduction

This chapter addresses the other fundamental scalability limitations identified in previous methods by leveraging the inherent graph structure of IAB networks. While the transfer learning-enhanced SAC framework demonstrated significant improvements in training efficiency, its state representation still treats the network topology as a flat feature vector, failing to fully exploit the rich relational information encoded in the network's connectivity patterns. This chapter introduces a paradigm shift by reformulating the IAB deployment problem through the lens of graph learning, specifically employing GNN with attention mechanisms to capture complex spatial dependencies and multi-hop relationships.

6.1.1 Graph Neural Networks: Foundations and Advantages

GNNs represent a powerful class of deep learning models designed to operate directly on graph-structured data [74]. Unlike traditional neural networks

that assume Euclidean structure in data, GNNs can process arbitrary graph topologies through iterative message passing between connected nodes. The fundamental operation of a GNN can be expressed as:

$$\mathbf{h}_v^{(k+1)} = \sigma\left(\mathbf{W}_{\text{self}}^{(k)}\mathbf{h}_v^{(k)} + \sum_{u\in\mathcal{N}(v)} \mathbf{W}_{\text{neigh}}^{(k)}\mathbf{h}_u^{(k)}\right), \tag{6.1}$$

where $\mathbf{h}_v^{(k)}$ represents the hidden state of node v at layer k, $\mathcal{N}(v)$ denotes the neighborhood of node v, and $\mathbf{W}_{\text{self}}^{(k)}$, $\mathbf{W}_{\text{neigh}}^{(k)}$ are learnable weight matrices [75].

The key advantages of GNNs for network optimization include:

- **Permutation Invariance**: GNNs produce consistent outputs regardless of node ordering, essential for deployment scenarios where candidate locations have no inherent sequence.

- **Inductive Learning**: Once trained, GNNs can generalize to graphs of different sizes and topologies, enabling zero-shot deployment in new urban environments [76].

- **Spatial Dependency Modeling**: Through multi-hop message passing, GNNs naturally capture k-hop neighborhood influences, critical for modeling interference patterns and multi-hop backhaul paths in IAB networks.

6.1.2 Graph Attention Networks: Enhanced Representation Learning

While standard GNN treat all neighbors equally during aggregation, GATv2 introduce attention mechanisms that learn to weight neighbor contributions based on their relevance [77]. The attention mechanism in GATv2 computes coefficients:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T[\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_j]))}{\sum_{k \in \mathcal{N}(i)} \exp(\text{LeakyReLU}(\mathbf{a}^T[\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_k]))} \tag{6.2}$$

where $\|$ denotes concatenation, $\mathbf{a}$ is a learnable attention vector, and α_{ij} represents the importance of node j's features to node i.

The evolution to GATv2 addresses a critical limitation of the original Graph Attention Network (GAT) architecture. Brody et al. [78] demonstrated that GAT's attention mechanism can only compute static attention patterns, limiting its expressiveness. GATv2 modifies the attention computation to:

$$\alpha_{ij} = \mathbf{a}^T \text{LeakyReLU}(\mathbf{W}[\mathbf{h}_i \| \mathbf{h}_j]) \tag{6.3}$$

This reformulation enables dynamic attention that adapts based on the query-key relationship, crucial for IAB networks where the importance of connections varies with deployment state and traffic patterns.

6.1.3 Resilience Challenges in mmWave IAB Networks

Operating in high-frequency bands above 24 GHz, mmWave networks face severe propagation limitations including high path loss and extreme sus-

ceptibility to blockages [6]. These characteristics make mmWave networks inherently vulnerable to frequent link failures, where temporary obstacles or environmental changes can instantly disconnect users or isolate network segments. Traditional deployment approaches that optimize solely for coverage fail to address these resilience requirements, leading to networks that are efficient under ideal conditions but fragile in practice.

6.1.4 Chapter Contributions and Organization

This chapter makes the following key contributions to the field of intelligent IAB network planning:

- **Resilience-Aware Graph Formulation**: Development of a novel graph representation that explicitly encodes resilience constraints through heterogeneous node types and edge-conditioned features, enabling GNN-based reasoning about fault tolerance.
- **Edge-Conditioned GATv2 Architecture**: Design of a specialized GATv2 framework that processes both node and edge features, capturing the complex interplay between deployment decisions, link capacities, and traffic flows.
- **Scalable Optimization Framework**: Integration of GATv2 with Proximal Policy Optimization (PPO) reinforcement learning, achieving linear computational complexity $\mathcal{O}(E \cdot d_h)$ compared to exponential complexity $\mathcal{O}(2^N)$ of traditional MIP approaches.
- **Comprehensive Resilience Evaluation**: Extensive analysis demonstrating 87.1% coverage retention under 30% link failures—a 15.4%

improvement over state-of-the-art methods—while reducing deployed node count by up to 26.7%.

The remainder of this chapter is organized as follows: Section 6.2 presents the system model and problem formulation with explicit resilience constraints. Section 6.2.2 details the graph representation and GATv2 architecture. Section 6.3 provides comprehensive experimental evaluation across diverse urban scenarios. Section 6.4 concludes with insights and future research directions.

6.2 System Model and Problem Formulation

We consider an urban mmWave IAB deployment at 60 GHz over a discretised service area. The set of fibre-connected IAB donors $\mathcal{I}$ serves as gateways to the core network, and candidate IAB node locations $\mathcal{J}$ correspond to feasible urban infrastructures (e.g., lamp posts and utility poles). Unless explicitly stated otherwise, the underlying physical-layer assumptions, link-budget/SNR feasibility checks, and common radio/traffic settings follow the baseline system model in Chapter 3 (Sec. 3.2–3.2.2 and Table 3.1). Therefore, we do not repeat the received-power, gain/loss decomposition, antenna modelling, or thermal-noise expressions here.

From Chapter 6 onward, the model is extended to explicitly incorporate resilience-aware planning. In addition to meeting a target coverage ratio and end-to-end backhaul feasibility (as in Chapter 3), each deployed IAB node must maintain at least m independent active backhaul connections to improve fault tolerance. To capture backhaul congestion and failure-recovery headroom more explicitly, we introduce continuous link-flow variables and

a capacity reservation factor β, enabling evaluation of end-to-end service feasibility under disrupted or rerouted traffic.

6.2.1 Problem Formulation

We formulate the resilience-aware IAB deployment as a mixed-integer optimization problem that minimizes the number of deployed IAB nodes while ensuring coverage, connectivity, and resilience requirements. The problem is non-convex due to binary deployment variables and bilinear capacity constraints.

Decision Variables

Let $\mathcal{I}$ denote the set of fiber-connected IAB donors and $\mathcal{J}$ the set of candidate IAB node locations. The decision variables are:

- $\alpha_j \in \{0, 1\}$: deployment indicator for candidate IAB node location $j \in \mathcal{J}$
- $Y_{pq} \in \{0, 1\}$: backhaul link activation from node p to IAB node q
- $R_{pq} \geq 0$: traffic flow rate on link $p \to q$ (Mbps)
- $U_k \in \{0, 1\}$: coverage indicator for grid cell $k \in \mathcal{K}$

For all IAB donors $i \in \mathcal{I}$, we set $\alpha_i = 1$.

Coverage Constraints

The coverage requirement ensures adequate SNR levels across the service area:

$$U_k \leq \sum_{i \in \mathcal{I}} C_{ik} + \sum_{j \in \mathcal{J}} C_{jk} \alpha_j, \quad \forall k \in \mathcal{K} \tag{6.4}$$

where $C_{ik}, C_{jk} \in \{0, 1\}$ indicate whether IAB donor i or IAB node j provides coverage to cell k with SNR $\geq \text{SNR}_{threshold}$.

The aggregate coverage constraint mandates:

$$\sum_{k \in \mathcal{K}} U_k \geq \theta_{cov} |\mathcal{K}| \tag{6.5}$$

ensuring at least fraction θ_{cov} of grid cells receive adequate coverage.

Resilience and Connectivity Constraints

Link activation follows logical consistency:

$$Y_{pq} \leq L_{pq} \alpha_p \alpha_q, \quad \forall p \in \mathcal{I} \cup \mathcal{J}, q \in \mathcal{J}, p \neq q \tag{6.6}$$

where $L_{pq} \in \{0, 1\}$ indicates link feasibility based on SNR requirements from the communication model.

The resilience constraint ensures redundant connectivity:

$$\sum_{p \in \mathcal{I} \cup \mathcal{J}, p \neq j} Y_{pj} \geq m \cdot \alpha_j, \quad \forall j \in \mathcal{J} \tag{6.7}$$

requiring each deployed IAB node ($\alpha_j = 1$) to maintain at least m active backhaul connections for fault tolerance.

Capacity and Flow Constraints

Link capacity constraints reserve bandwidth for failure recovery:

$$R_{pq} \leq (1-\beta) C_{pq} Y_{pq}, \quad \forall p \in \mathcal{I} \cup \mathcal{J}, q \in \mathcal{J}, p \neq q \tag{6.8}$$

where C_{pq} is the physical link capacity (Mbps) and $\beta \in [0,1]$ reserves fraction β of link capacity for traffic rerouting during link failures. This ensures that when primary paths fail, backup routes have sufficient capacity.

IAB donor capacity constraints limit fiber backhaul usage:

$$\sum_{j \in \mathcal{J}} R_{ij} Y_{ij} + R_o A_i \leq F_i, \quad \forall i \in \mathcal{I} \tag{6.9}$$

where F_i is IAB donor i's fiber capacity, $A_i > 0$ represents local access demand, and $R_o > 1$ accounts for protocol overhead including MAC layer framing, PHY layer pilot symbols, routing protocol signaling, and retransmission mechanisms.

Flow conservation at IAB nodes ensures traffic balance:

$$\sum_{p \in \mathcal{I} \cup \mathcal{J}, p \neq j} R_{pj} Y_{pj} \geq R_o \left(A_j \alpha_j + \sum_{n \in \mathcal{J}, n \neq j} R_{jn} Y_{jn} \right), \quad \forall j \in \mathcal{J} \tag{6.10}$$

where inbound traffic (left side) must satisfy local access demand A_j and outbound forwarding requirements (right side), both scaled by overhead factor R_o.

Optimization Objective

The objective minimizes the number of deployed IAB nodes:

$$\min \sum_{j \in \mathcal{J}} \alpha_j \tag{6.11}$$

subject to constraints (6.4)-(6.10).

This formulation ensures optimal IAB node placement while maintaining network resilience through redundant connectivity and capacity reservation for failure recovery.

6.2.2 Graph Representation

To enable GNN-based optimization, we transform the mixed-integer formulation into a heterogeneous attributed digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X}, \mathbf{E}, \mathbf{g})$ that captures network topology and deployment state.

Vertices $\mathcal{V} = \mathcal{I} \cup \mathcal{J}$ comprise IAB donors ($i \in \mathcal{I}$) and candidate IAB node locations ($j \in \mathcal{J}$).

Directed edges $\mathcal{E} = \{(p, q) \mid L_{pq} = 1, p \in \mathcal{I} \cup \mathcal{J}, q \in \mathcal{J}, p \neq q\}$ represent feasible mmWave backhaul links satisfying SNR requirements.

Node Features $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}| \times d_v}$ Each vertex $v \in \mathcal{V}$ has feature vector:

$$\mathbf{x}_v = \left[\alpha_v, \frac{A_v}{A_{\max}}, \frac{N_v}{m}, \mathbb{1}_{\{v \in \mathcal{I}\}} \right]$$

where $\alpha_v \in \{0, 1\}$ is the deployment status, $\frac{A_v}{A_{\max}}$ is normalized access demand (with $A_{\max} = \max_{v \in \mathcal{V}} A_v$ from traffic analysis), $\frac{N_v}{m}$ represents resilience ratio

(current connections over requirement m), and $\mathbb{1}_{\{v\in\mathcal{I}\}}$ indicates IAB donor type. Normalization prevents gradient instability caused by different feature scales.

Edge Features $\mathbf{E} \in \mathbb{R}^{|\mathcal{E}|\times d_e}$ Each directed link $(p, q) \in \mathcal{E}$ carries:

$$\mathbf{e}_{pq} = \left[\frac{C_{pq}}{C_{\max}}, \frac{R_{pq}}{C_{pq}}, L_{pq}\right]$$

where $\frac{C_{pq}}{C_{\max}}$ is normalized link capacity (with $C_{\max} = \max_{(p,q)\in\mathcal{E}} C_{pq}$ from strongest feasible link), $\frac{R_{pq}}{C_{pq}}$ represents current utilization ratio, and $L_{pq} = 1$ confirms link feasibility.

Global Features $\mathbf{g} \in \mathbb{R}^4$ Graph-level parameters:

$$\mathbf{g} = [\theta_{cov}, m, \beta, R_o]$$

encoding coverage target, resilience requirement, capacity headroom, and protocol overhead.

This representation enables the GNN to learn spatial dependencies while reasoning about deployment constraints, resilience requirements, and capacity utilization during the optimization process.

6.2.3 Markov Decision Process Formulation

We formulate the IAB deployment problem as a MDP $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$ to enable reinforcement learning optimization:

State Space $\mathcal{S}$ The state $s_t \in \mathcal{S}$ at step t is represented by the attributed graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}, \mathbf{X}_t, \mathbf{E}_t, \mathbf{g})$, where node and edge features encode current deployment status, traffic flows, and connectivity patterns.

Action Space $\mathcal{A}$ At each step, the agent selects action $a_t \in \mathcal{A} = \{0\} \cup \{j \mid j \in \mathcal{J}, \alpha_j = 0\}$ to either deploy an IAB node at candidate location j ($a_t = j$) or maintain current configuration ($a_t = 0$).

Reward Function $\mathcal{R}$ The reward function balances coverage improvement against deployed node count and resilience violations:

$$r_t = \kappa_1 \Delta U_{cov} - \kappa_2 \alpha_{deploy} - \kappa_3 N_{vulnerable}$$

where ΔU_{cov} represents coverage percentage improvement, $\alpha_{deploy} \in \{0, 1\}$ indicates new IAB node deployment, and $N_{vulnerable}$ penalizes nodes violating the resilience constraint m. The coefficients $\kappa_1, \kappa_2, \kappa_3$ control the relative importance of coverage gains versus deployed node count and constraint violations.

GATv2-based Policy Architecture The policy $\pi_\theta(a|s)$ employs a GATv2 encoder with edge-conditioned attention to process the graph state:

$$\mathbf{h}_v^{(l+1)} = \text{GATv2}^{(l)}(\mathbf{h}_v^{(l)}, \{\mathbf{h}_u^{(l)}, \mathbf{e}_{uv}\}_{u \in \mathcal{N}(v)})$$

where $\mathbf{h}_v^{(l)}$ represents node embeddings at layer l, and edge features $\mathbf{e}_{uv}$ condition the attention mechanism. A pointer-based actor network computes deployment probabilities over valid candidate locations, while a critic net-

Algorithm 7: GATv2-PPO for Resilient IAB Deployment

Input: Donor configuration $\mathcal{D}$, candidate locations $\mathcal{J}$, training episodes E

Output: Optimized deployment policy π_θ^*

for $e = 1$ **to** E **do**
- Initialize environment with fixed donor topology $\mathcal{D}$;
- Reset deployment state $\mathbf{s}_0$ and coverage metrics;
- **while** *coverage threshold* θ_{cov} *not achieved* **do**
  - Construct heterogeneous graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}, \mathbf{X}_t, \mathbf{E}_t)$;
  - Apply GATv2 encoding: $\mathbf{H}^{(L)} = \text{GATv2}(\mathbf{X}_t, \mathbf{E}_t, \mathcal{E})$;
  - Sample deployment action: $a_t \sim \boldsymbol{\pi}_t$;
  - Execute action a_t and observe transition (s_t, a_t, r_t, s_{t+1});
  - Compute reward r_t with resilience penalties;
  - Store experience $(\mathcal{G}_t, a_t, r_t, \mathcal{G}_{t+1}, d_t)$ in replay buffer $\mathcal{B}$;
  - Update graph state $\mathcal{G}_{t+1}$ based on deployment changes;

return *Trained policy* π_θ^* *achieving resilient network deployment*;

Table 6.1: Key Symbols and Definitions

Symbol	Definition
Y_{pq}	Binary link activation variable
U_k	Binary coverage indicator for cell k
L_{pq}	Link feasibility indicator (0/1)
R_{pq}	Traffic flow on link $p \to q$ [Mbps]
m	Minimum resilience degree (inbound links)
R_o	Protocol overhead factor (> 1)
β	Backup capacity fraction for resilience
θ_{cov}	Target coverage fraction
N_v	Number of active inbound connections to node v

work estimates state values for PPO training.

6.3 Simulation and Results

6.3.1 Simulation setup

The simulation environment models urban mmWave IAB deployment across 1×1 km^2 service areas with 400 potential node locations distributed at 50m intervals to mimic realistic urban infrastructure density, corresponding to lamp posts and utility poles. This grid-based approach ensures fair algorithm comparison while maintaining practical relevance. Three deployment scenarios are used to evaluate algorithm performace: Pentagon (5 donors in pentagonal setting), Five-Dice (donors positioned as dice-5 pattern), and Vertical (linear arrangement). Table 6.2 summarizes the key simulation parameters used throughout the evaluation.

To address the resilience-aware IAB deployment challenge, we propose GATv2 with edge-conditioned attention mechanism detailed in Algorithm 7 to capture spatial dependencies and connectivity constraints inherent in IAB networks. The model's ability to process heterogeneous node types (IAB donors vs. candidate nodes) and dynamic edge features (link capacity, utilization) makes it particularly suited for modeling complex relationships between deployment decisions, network topology, and resilience requirements. The model is trained using Proximal Policy Optimization (PPO) within the MDP framework, enabling the agent to balance coverage objectives against deployed node count and constraint violations. Model performance is benchmarked against two established approaches:

- **Greedy Heuristic**: A greedy algorithm [31] that sequentially deploys nodes by maximizing immediate coverage improvement under

Table 6.2: Simulation Parameters and Model Configuration

Parameter	Value	Reference
Transmit power P_{tx}	30 dBm	[79]
Antenna gain G_t, G_r	25 dBi	[79]
Noise figure	7 dB	[79]
Operating frequency	60 GHz	[79]
SNR threshold	10 dB	[80]
Coverage threshold θ_{cov}	98%	–
Backup capacity fraction for resilience β	0.2	–
Overhead R_o	1.2	–
Resilience parameter m	2	–
Reward coefficients $\kappa_1, \kappa_2, \kappa_3$	4.0, 0.2, 0.5	–
Graph encoder	2-layer GATv2	–
Hidden units	64 per layer	–
Attention heads	8	–
Learning rate	3×10^{-4}	–
Batch size	32	–
Discount factor γ	0.99	–
Clip ratio	0.2	–
Training episodes	8000	–

data rate constraints and solved by Gurobi solver. Although fast and interpretable, its myopic nature leads to suboptimal long-term performance.

- **Dueling DQN with GCN**: Deep reinforcement learning approach using a Graph Convolutional Network backbone with experience replay and target networks [81].

6.3.2 Result analysis

Fig. 6.1 demonstrates the spatial deployment achieved by GATv2 in the Five-Dice configuration. The heat map shows coverage zones from -96 dBm to -10

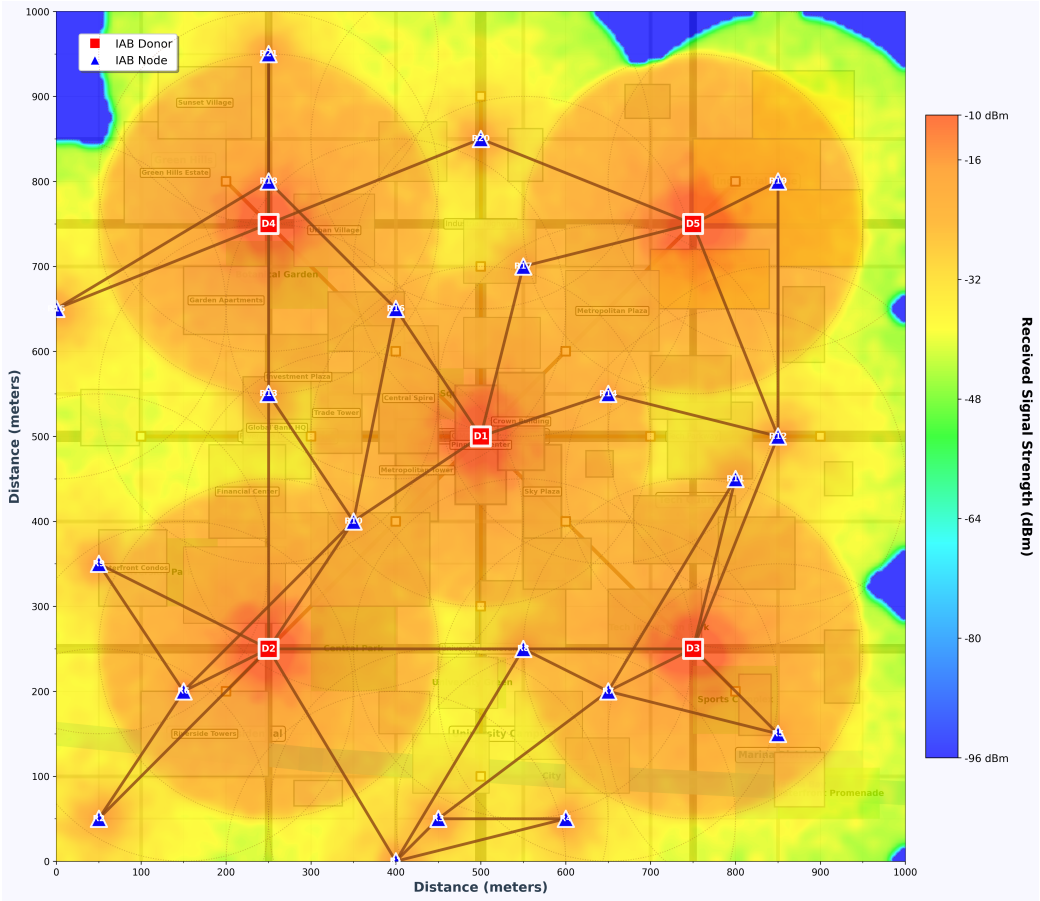

Figure 6.1: GATv2-optimized IAB network deployment showing Five-Dice donor configuration (red squares D1-D5), deployed nodes (blue triangles), and signal strength heat map. Black lines indicate backhaul connections forming a resilient mesh topology with redundant paths.

dBm, with deployed nodes (blue triangles) creating overlapping high-SNR areas that eliminate coverage gaps between donors D1-D5. The mesh connectivity pattern (black lines) ensures each node maintains multiple backhaul connections, satisfying the $m = 2$ redundancy constraint through GATv2's learned policy that balances coverage and resilience requirements. Building upon this spatial optimization capability, Fig. 6.2 quantifies deployment efficiency across three donor configurations. GATv2 consistently outperforms

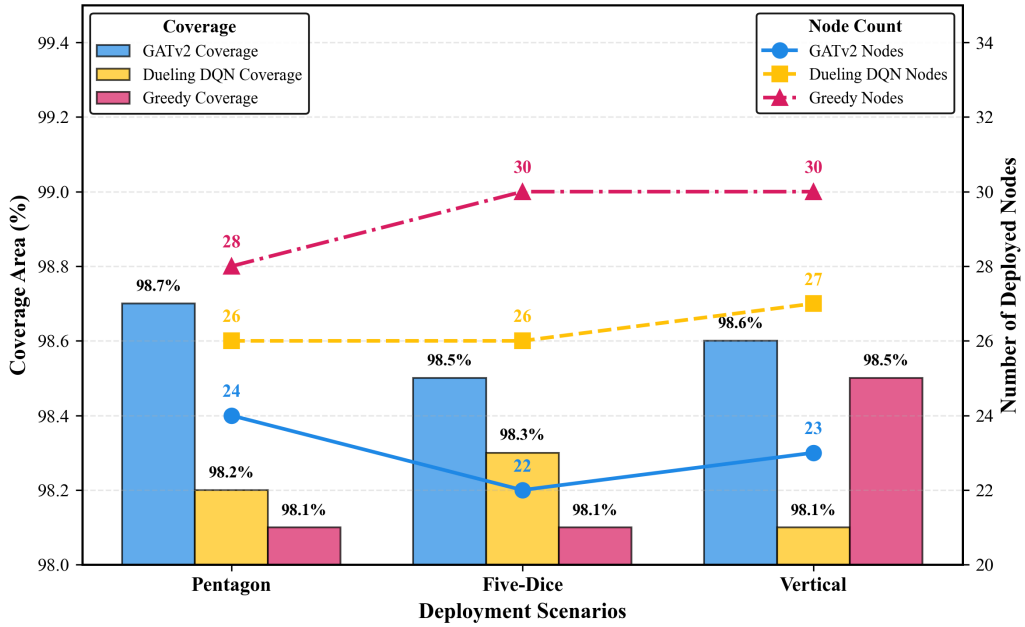

Figure 6.2: Multi-scenario deployment efficiency comparison across Pentagon, Five-Dice, and Vertical donor configurations.

baselines: Pentagon scenario achieves 98.7% coverage with 24 nodes versus 26 (DQN, 98.2%) and 28 (Greedy, 98.1%); Five-Dice requires only 22 nodes for 98.5% coverage, representing 26.7% reduction compared to Greedy's 30 nodes; Vertical uses 23 nodes versus 27 (DQN) and 30 (Greedy) for 98.6% coverage. These improvements result from GATv2's attention mechanism capturing multi-hop spatial dependencies for global topology optimization, surpassing greedy methods' myopic decisions and standard GCNs' limited graph representation.

Beyond deployment efficiency, the resilience-aware design proves critical under network stress conditions. To evaluate fault tolerance, we conduct link failure simulation using random failure models. The methodology randomly disables 10%, 20%, and 30% of backhaul links, with each failure rate tested across 100 independent trials. We measure coverage retention as the network

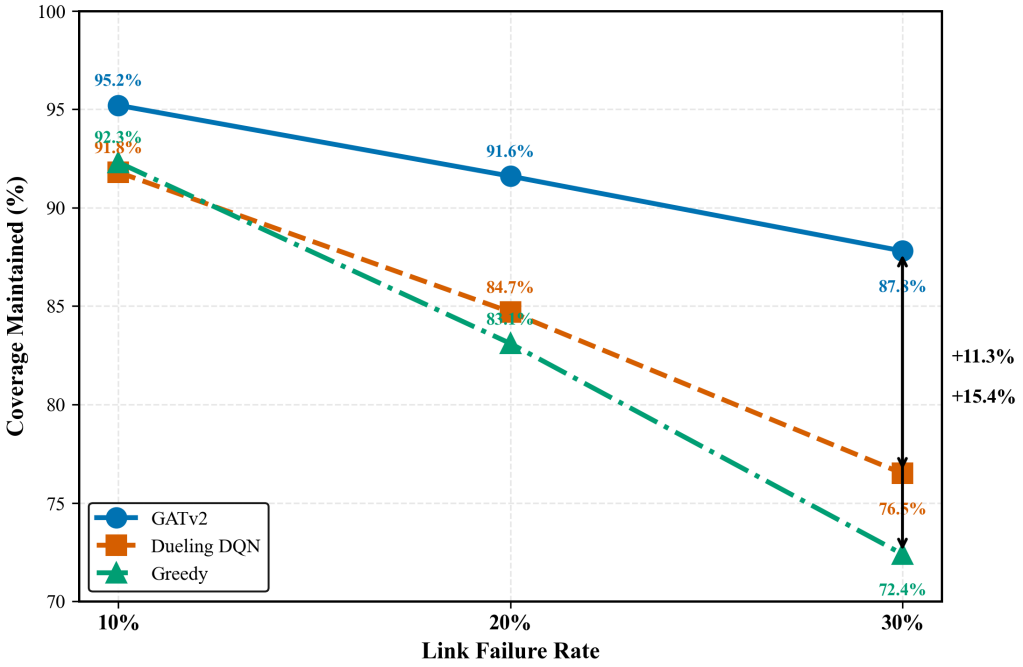

Figure 6.3: Network resilience under progressive link failures, and progressive failure tests randomly disable 10%, 20%, and 30% of backhaul links to simulate mmWave blockage events.

robustness indicator. Fig. 6.3 demonstrates that GATv2 maintains superior performance across all failure scenarios: 95.2% coverage retention at 10% failure versus 91.3% (DQN) and 92.3% (Greedy); 91.6% at 20% failure versus 84.7% (DQN) and 83.1% (Greedy); 87.1% at 30% failure versus 76.6% (DQN) and 72.4% (Greedy). The 11.3-15.4% performance advantage stems from the explicit m connectivity constraint in the MDP formulation, preventing single-connected deployments that cause cascading failures in baseline methods.

6.3.3 Complexity Analysis

The computational complexity of GATv2-based deployment is analyzed for a network with N candidate nodes and E potential backhaul links. The graph

attention mechanism requires $\mathcal{O}(E \cdot d_h)$ operations per attention head, where d_h is the hidden dimension. With 2 attention heads and 2 GATv2 layers, the encoding complexity is $\mathcal{O}(4E \cdot d_h)$. The actor network performs $\mathcal{O}(|\mathcal{A}| \cdot d_h)$ operations for action selection over valid action set $\mathcal{A}$.

For our urban scenarios with $N = 400$ nodes and average node degree 8, $E \approx 1600$. With $d_h = 32$, total complexity per episode is $\mathcal{O}(51200 + 32|\mathcal{A}|)$. This linear scaling contrasts favorably with mixed-integer programming approaches exhibiting exponential complexity $\mathcal{O}(2^N)$ for binary placement variables.

The PPO training complexity is $\mathcal{O}(B \cdot T \cdot C_{forward})$ per update, where $B = 32$ is batch size, T is trajectory length, and $C_{forward}$ is the forward pass cost. Memory requirements scale as $\mathcal{O}(N \cdot d_h + E \cdot d_e)$ for node and edge embeddings, remaining manageable for large-scale deployments.

6.4 Conclusion

This chapter has demonstrated the transformative potential of GNN-based approaches for resilient IAB network planning. By reformulating the deployment problem through a graph-centric lens and leveraging GATv2's dynamic attention mechanisms, we achieved significant improvements in both deployment efficiency and network resilience compared to the DRL-based methods explored in previous chapters.

The edge-conditioned GATv2 architecture successfully captures the complex spatial dependencies and multi-hop relationships inherent in IAB topologies, enabling deployment strategies that inherently consider fault tolerance. Our experimental results across diverse urban scenarios confirm three criti-

cal advantages: deployment efficiency with 14.3-26.7% reduction in required nodes, exceptional resilience maintaining 87.1% coverage under severe failure conditions, and linear computational scaling suitable for city-scale deployments.

7 Conclusions and future work

7.1 Conclusions

This thesis has presented a comprehensive framework for intelligent IAB network planning, progressing systematically from traditional optimization methods to advanced machine learning approaches. Through rigorous mathematical formulation, algorithmic innovation, and extensive experimental validation, we have addressed the fundamental challenges of scalability, adaptability, and resilience in mmWave network deployment for next-generation wireless systems.

Firstly, we established robust theoretical foundations and mathematical modeling frameworks for IAB network planning. The formulation of deployment problems as MILP optimization and subsequent MDP representations provided rigorous mathematical underpinnings for the research. A key theoretical innovation was the replacement of traditional fixed hop-count limitations with dynamic capacity-based constraints, validated through comprehensive 60 GHz link budget analysis. The systematic evolution from discrete optimization problems to continuous RL formulations, culminating in graph-based representations, demonstrates a coherent theoretical progression that bridges operations research and machine learning methodologies.

Secondly, significant algorithmic innovations were achieved across multiple dimensions of network planning optimization. The development of specialized action elimination strategies reduced search spaces by 60-80% in high-dimensional discrete environments. We pioneered the successful adaptations of continuous control algorithms—specifically SAC and TD3—to discrete IAB planning domains with proven convergence guarantees of polynomial sample complexity $O(|S||A|/\varepsilon^2)$. The introduction of edge-conditioned GATv2 frameworks enabled processing of heterogeneous node types and dynamic link utilization patterns, achieving computational complexity reduction from exponential $O(2^N)$ to linear $O(E \cdot d_h)$ scaling suitable for city-scale deployments.

Thirdly, comprehensive experimental validation demonstrated substantial performance improvements across all proposed methodologies. DRL-based approaches achieved 12-15% reduction in required nodes compared to heuristic baselines while maintaining 98% coverage targets. The TL framework realized 50% training time reduction when adapting to new donor configurations, transforming computationally prohibitive tasks into practically deployable solutions. Most significantly, the graph-based resilience-aware optimization maintained 87.1% coverage retention under 30% link failures—a 15.4% improvement over state-of-the-art methods—while simultaneously achieving 26.7% reduction in the number of deployed IAB nodes across diverse urban scenarios.

Lastly, the research contributions provide immediate practical impact for next-generation wireless network deployment. The seamless integration with Open RAN architectures through Non-RT RIC ensures compatibility with in-

dustry standards and deployment frameworks. The systematic methodology progression from traditional optimization to intelligent learning approaches offers network operators flexible deployment strategies adaptable to varying environmental conditions and performance requirements. These innovations collectively advance the state-of-the-art in network planning, providing a complete framework that bridges theoretical foundations with practical deployment requirements essential for the transition to 6G networks.

7.2 Future work

Several promising research directions emerge from our findings that warrant further investigation to address evolving challenges in 6G network deployment:

- **Reliable Network Enhancement**: Develop multi-path redundancy optimization algorithms that dynamically balance backup capacity reservation with network efficiency, incorporating GNN-based failure prediction models and graph attention mechanisms to identify critical bottlenecks before failures occur.

- **Joint optimisation of donor placement and IAB node count:** The present work assumes that donor locations are fixed by the availability of fibre infrastructure, and optimises solely over the placement of IAB relay nodes. A natural and practically important extension is the joint optimisation of donor placement and IAB node deployment, trading off the higher per-unit cost of fibre-connected donors against the reduction in IAB relay nodes they enable. The empirical results in

Tables 5.2–5.4 already demonstrate that donor geometry and density profoundly affect the number of IAB nodes required; a formal joint optimisation would allow network operators to select the most cost-effective balance between donor infrastructure and wireless relay density for a given coverage target.

- **Scalable donor-to-node ratio analysis for city-scale deployments:** The GATv2 framework developed in Chapter 6 achieves linear complexity $\mathcal{O}(E \cdot d_h)$ and is therefore well positioned to study how the ratio of IAB nodes to donor nodes scales with deployment area, traffic density, and resilience requirements. Such analysis would provide network planners with practically actionable guidelines—for example, the expected number of IAB relay nodes per donor node under different urban morphologies—extending the per-scenario observations of Chapter 3 to a general, data-driven planning tool.

These research directions build upon the foundations established in this thesis, extending our intelligent planning frameworks toward fully autonomous, self-optimizing 6G networks that can adapt to changing conditions while maintaining optimal performance and reliability.

Appendices

A

Donor Geometry Definitions

We use three representative donor geometries to generate controlled topology patterns for evaluation and transfer learning.

Five-dice Given an area of side length L and centre $(L/2, L/2)$, donors are placed at $(L/4, L/4)$, $(3L/4, L/4)$, $(L/2, L/2)$, $(L/4, 3L/4)$, $(3L/4, 3L/4)$.

Regular pentagon Let (x_c, y_c) be the centre and r be the radius. The k-th donor is $(x_c + r\cos\theta_k,\ y_c + r\sin\theta_k)$, where $\theta_k = \theta_0 + 2\pi(k-1)/5$.

Regular hexagon Similarly, $\theta_k = \theta_0 + 2\pi(k-1)/6$ with $k = 1, \ldots, 6$.

References

[1] Ericsson, "Ericsson mobility report: Mobile data traffic outlook," *Ericsson Mobility Report*, June 2024.

[2] Deloitte Consulting, "The cost of fiber deployment in urban environments," *Telecommunications Infrastructure Report*, 2023.

[3] 3GPP, "NR; NR and NG-RAN overall description," Tech. Rep. TS 38.300, 3rd Generation Partnership Project, Sept. 2021.

[4] C. Madapatha, B. Makki, C. Fischione, and T. Svensson, "On integrated access and backhaul networks: Current status and potentials," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1374–1389, 2020.

[5] G. R. MacCartney, T. S. Rappaport, and S. Rangan, "Rapid fading due to human blockage in pedestrian crowds at 5G millimeter-wave frequencies," *IEEE Access*, vol. 5, pp. 27674–27682, 2017.

[6] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, 2013.

[7] A. Gupta and R. K. Jha, "Resilience in 5G mobile networks: A cross-layer perspective," *IEEE Wireless Communications*, vol. 28, no. 4, pp. 178–184, 2021.

[8] 3GPP, "Study on integrated access and backhaul," Tech. Rep. TR 38.874 V16.0.0, 3rd Generation Partnership Project, Dec. 2018. Technical Report.

[9] M. Cudak, A. Ghosh, A. Ghosh, and J. G. Andrews, "A key enabler for 5G millimeter-wave deployments," *IEEE Communications Magazine*, vol. 59, pp. 88–94, Apr. 2021.

[10] Y. Zhang, M. A. Kishk, and M.-S. Alouini, "A survey on integrated access and backhaul networks," *Frontiers in Communications and Networks*, vol. 2, pp. 1–24, 2021.

[11] 3GPP, "NG-RAN; architecture description," Tech. Rep. TS 38.401, 3rd Generation Partnership Project, Dec. 2021.

[12] C. Saha and H. S. Dhillon, "Millimeter wave integrated access and backhaul in 5G," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 12, pp. 2669–2684, 2019.

[13] H. Ronkainen, A. Ericsson, and C. Östberg, "Integrated access and backhaul – a new type of wireless backhaul in 5G," *Ericsson Technology Review*, 2020.

[14] 3GPP, "NR; backhaul adaptation protocol (BAP) specification," Tech. Rep. TS 38.340, 3rd Generation Partnership Project, June 2021.

[15] Q. Liu, H. Tian, G. Nie, and H. Wu, "Joint uplink and downlink scheduling in hetnets with wireless self-backhaul," in *Proc. 24th Asia-Pacific Conference on Communications (APCC)*, pp. 292–297, IEEE, Nov. 2018.

[16] C. Madapatha, B. Makki, C. Fischione, and T. Svensson, "Optimal topology formation and adaptation of integrated access and backhaul networks," *Frontiers in Communications and Networks*, vol. 2, pp. 1–14, 2021.

[17] J. Sorsa, "Integrated access and backhaul in millimeter-wave cellular," Master's thesis, Tampere University, 2022.

[18] E. Dahlman, S. Parkvall, and J. Skold, "Integrated access and backhaul: New option for 5G," *Ericsson Technology Review*, vol. 2020, no. 6, pp. 25–35, 2020. Accessed: 2024-07-23.

[19] N. Polyakov and M. Platonova, "Assessing latency of packet delivery in the 5G 3GPP integrated access and backhaul architecture with half-duplex constraints," *Future Internet*, vol. 14, no. 11, p. 345, 2022.

[20] Y. Yamamoto, H. Tanaka, and K. Sato, "Field evaluation of 5G mmWave relays in various indoor scenarios," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, IEEE, 2021.

[21] MathWorks, "Evaluate 3GPP indoor reference scenario." MATLAB Documentation, 2024. Accessed: 2024-07-23.

[22] C. Madapatha, B. Makki, C. Fischione, and T. Svensson, "On integrated access and backhaul networks: Current status and potentials," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1374–1389, 2021.

[23] M. Simsek, O. Orhan, M. Nassar, O. Elibol, and H. Nikopour, "IAB topology design: A graph embedding and deep reinforcement learning approach," *IEEE Communications Letters*, vol. 25, no. 2, pp. 489–493, 2021.

[24] M. N. Islam, N. Abedini, G. Hampel, S. Subramanian, and J. Li, "Investigation of performance in integrated access and backhaul networks," in *Proc. IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 597–602, IEEE, May 2018.

[25] H. Alghafari and M. S. Haghighi, "Decentralized joint resource allocation and path selection in multi-hop integrated access backhaul 5G networks," *Computer Networks*, vol. 207, p. 108837, 2022.

[26] M. Pagin, T. Zugno, M. Polese, and M. Zorzi, "Resource management for 5G NR integrated access and backhaul: A semi-centralized approach," *IEEE Transactions on Wireless Communications*, vol. 21, no. 2, pp. 753–767, 2022.

[27] M. N. Islam, S. Subramanian, and A. Sampath, "Integrated access backhaul in millimeter wave networks," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2017.

[28] B. Zhai, M. Yu, A. Tang, and X. Wang, "Mesh architecture for efficient integrated access and backhaul networking," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, IEEE, May 2020.

[29] J.-H. Huh, T. Inagaki, J. Nakazato, M. Arai, K. Yano, and M. Hasegawa, "Graph neural network-based multi-metric performance modeling in urban multi-RAT wireless networks," *ICT Express*, 2025.

[30] M. Polese, M. Giordani, T. Zugno, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "Integrated access and backhaul in 5G mmWave networks: Potential and challenges," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 62–68, 2020.

[31] J. Zhang, Q. Wang, P. Mitchell, and H. Ahmadi, "An integrated access and backhaul approach to sustainable dense small cell network planning," *Information*, vol. 15, no. 1, p. 19, 2024.

[32] C. Madapatha, B. Makki, E. Dahlman, M. S. Alouini, and T. Svensson, "Constrained deployment optimization for IAB-enabled mmWave 5G networks," 2022. Manuscript / conference version as cited in thesis.

[33] J. Zhang, S. Chetty, Q. Wang, C. Sun, P. D. Mitchell, D. Grace, and H. Ahmadi, "Optimizing 6G dense network deployment for the metaverse using deep reinforcement learning," 2025.

[34] Y. Zhang, M. A. Kishk, and M.-S. Alouini, "Deployment optimization of tethered drone-assisted integrated access and backhaul networks," *IEEE Transactions on Wireless Communications*, vol. 23, no. 4, pp. 2668–2680, 2024.

[35] Z. Liu, X. Wang, Y. Zhang, and S. Chen, "Graph neural networks for network management: A survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, pp. 1856–1888, 2022.

[36] X. Li, Q. Wang, J. Liu, and W. Zhang, "3D deployment with machine learning and system performance analysis of UAV-enabled networks," in *2020 IEEE/CIC International Conference on Communications in China (ICCC)*, pp. 554–559, 2020.

[37] A. Rahman *et al.*, "5G-enabled UAVs for energy-efficient opportunistic networking," *Heliyon*, vol. 10, p. e31861, June 2024.

[38] A. S. Tan and E. D. Biyar, "QoS-aware autonomous IAB-node activation and access control," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2021.

[39] M. Yu, B. Zhai, A. Tang, and X. Wang, "Coordinated parallel resource allocation for integrated access and backhaul networks," *Computer Networks*, vol. 218, p. 109533, 2023.

[40] K. Liu, Y. Zhou, and X. Chen, "Latency optimization and reliability assurance in multi-hop IAB networks," *IEEE Transactions on Mobile Computing*, vol. 21, pp. 3821–3834, Nov. 2022.

[41] W. Shang and V. Friderikos, "Reinforcement learning based NSGA-II for energy-delay trade-off in IAB mmWave het-nets," *Proc. IEEE International Conference on Communications (ICC)*, 2023.

[42] M. Neema, E. Gopi, and P. S. Reddy, "Optimizing broadband access and network design in wireless mesh networks using multi-objective particle swarm optimization," vol. 230, pp. 275–286, 2023. 3rd International Conference on Evolutionary Computing and Mobile Sustainable Networks (ICECMSN 2023).

[43] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2007.

[44] J. M. Lanza-Gutiérrez, N. Caballé, J. A. Gómez-Pulido, B. Crawford, and R. Soto, "Toward a robust multi-objective metaheuristic for solving the relay node placement problem in wireless sensor networks," *Sensors*, vol. 19, no. 3, p. 677, 2019.

[45] GSMA, "Wireless backhaul evolution," Apr. 2022.

[46] VIAVI Solutions, "5G business case puzzle: Putting it all together," tech. rep., 2019.

[47] B. Lim, J.-H. Lee, J.-H. Kwon, K.-H. Kim, J.-M. Lee, H. Park, Y.-S. Ha, Y.-J. Han, and Y.-C. Ko, "Joint association and resource allocation for multi-hop integrated access and backhaul (IAB) network," *Journal of Communications and Networks*, vol. 25, no. 4, pp. 440–455, 2023.

[48] S. F. Abedin, A. Mahmood, Z. Han, and M. Gidlund, "Resource optimization in multi-hop iab networks: Balancing data freshness and spectral efficiency," *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 3, pp. 1287–1310, 2025.

[49] S. L. Fong, J. Bucheli, A. Sampath, A. M. Bedewy, M. D. Mare, O. Shental, and M. N. Islam, "A mixed-integer linear programming approach to deploying base stations and repeaters," 2023.

[50] Y. Zhang, M. A. Kishk, and M.-S. Alouini, "A survey on integrated access and backhaul networks," *Frontiers in Communications and Networks*, vol. 2, 2021.

[51] 3GPP, "Integrated access and backhaul for 3GPP release 16," tech. rep., 3rd Generation Partnership Project (3GPP), 2020. Technical Specification/Report, Release 16.

[52] 3GPP, "NR; NR and NG-RAN overall description; stage 2," Tech. Rep. TS 38.300 V17.1.0, 3rd Generation Partnership Project, Mar. 2022. Technical Specification.

[53] International Telecommunication Union, "ITU-R P.838: Specific attenuation model for rain for use in prediction methods," tech. rep., ITU Radiocommunication Sector, Geneva, Switzerland, 2003.

[54] G. A. Siles, J. M. Riera, and P. Garcia-del Pino, "Atmospheric attenuation in wireless communication systems at millimeter and THz frequencies [Wireless Corner]," *IEEE Antennas and Propagation Magazine*, vol. 57, no. 1, pp. 48–61, 2015.

[55] Y. Fang and D. Brown, "Base station deployment optimization in federated networks with multi-hop communication," in *MILCOM 2022 - 2022 IEEE Military Communications Conference (MILCOM)*, pp. 1030–1037, 2022.

[56] L. Correia and P. Frances, "A propagation model for the estimation of the average received power in an outdoor environment in the millimetre waveband," in *Proceedings of IEEE Vehicular Technology Conference (VTC)*, pp. 1785–1788 vol.3, 1994.

[57] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.

[58] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020.

[59] C. K. Anjinappa, F. Erden, and I. Güvenç, "Base station and passive reflectors placement for urban mmWave networks," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 3525–3539, 2021.

[60] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb. 2015.

[61] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, MIT Press, 2 ed., 2018.

[62] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992.

[63] H. van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-Learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, pp. 2094–2100, 2016.

[64] Z. Wang, T. Schaul, M. Hessel, H. van Hasselt, M. Lanctot, and N. de Freitas, "Dueling network architectures for deep reinforcement learning," in *Proceedings of the International Conference on Machine Learning (ICML)*, vol. 48, pp. 1995–2003, 2016.

[65] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proceedings of the 35th International Conference on Machine Learning*, vol. 80 of *Proceedings of Machine Learning Research*, pp. 1861–1870, PMLR, 2018.

[66] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel, and S. Levine, "Soft actor-critic algorithms and applications," 2019.

[67] T. Haarnoja, S. Ha, A. Zhou, J. Tan, G. Tucker, and S. Levine, "Learning to walk via deep reinforcement learning," in *Robotics: Science and Systems XV*, 2019.

[68] P. Christodoulou, "Soft actor-critic for discrete action settings," 2019.

[69] M. E. Taylor and P. Stone, "Transfer learning for reinforcement learning domains: A survey," *Journal of Machine Learning Research*, vol. 10, no. 56, pp. 1633–1685, 2009.

[70] A. Lazaric, M. Restelli, and A. Bonarini, "Transfer of samples in batch reinforcement learning," in *Proceedings of the 25th International Conference on Machine Learning*, (Helsinki, Finland), pp. 544–551, ACM Press, 2008.

[71] L. Bonati, S. D'Oro, M. Polese, S. Basagni, and T. Melodia, "Intelligence and learning in O-RAN for data-driven nextg cellular networks," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 21–27, 2021.

[72] T. Wang *et al.*, "Twin-delayed deep stochastic policy gradient: A gumbel-softmax TD3 for discrete actions," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, 2024.

[73] C. Sun *et al.*, "Energy-consumption reduction for UAV trajectory training: A transfer-learning approach," *arXiv preprint arXiv:2501.11243*, 2025.

[74] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2008.

[75] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[76] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems*, pp. 1024–1034, 2017.

[77] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[78] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?," in *International Conference on Learning Representations*, 2022.

[79] ITU-R, "Minimum requirements related to technical performance for IMT-2020 radio interface(s)," Tech. Rep. M.2410-0, ITU-R, Nov. 2017.

[80] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," Technical Report TR 38.901 V17.0.0, 3rd Generation Partnership Project, Sophia Antipolis, France, June 2022.

[81] Y. Yang, D. Zou, and X. He, "Graph neural network-based node deployment for throughput enhancement," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 10, pp. 14810–14824, 2024.