Received 5 August 2024; accepted 16 August 2024. Date of publication 23 August 2024; date of current version 5 September 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3449234

Analysis of Duplexing Patterns in Multi-Hop mmWave Integrated Access and Backhaul Systems

NIKITA TAFINTSEV [1] (Graduate Student Member, IEEE), DMITRI MOLTCHANOV [1], WEI MAO [2], HOSEIN NIKOPOUR [2], SHU-PING YEH [2], SHILPA TALWAR [2] (Senior Member, IEEE), MIKKO VALKAMA [1] (Fellow, IEEE), AND SERGEY ANDREEV [1,3] (Senior Member, IEEE)

[1]Tampere Wireless Research Center, Tampere University, 33014 Tampere, Finland

[2]Intel Labs, Intel Corporation, Santa Clara, CA 95054, USA

[3]Department of Telecommunications, Brno University of Technology, 60190 Brno, Czech Republic

CORRESPONDING AUTHOR: N. TAFINTSEV (e-mail: nikita.tafintsev@tuni.fi)

This work was supported in part by Intel Corporation, and in part by the Research Council of Finland (Projects RADIANT, ECO-NEWS, and SOLID).

ABSTRACT Integrated Access and Backhaul (IAB) technology promises to facilitate cost-effective deployments of 5G New Radio (NR) systems operating in both sub-6 GHz and millimeter-wave (mmWave) bands. As full-duplex wireless systems are in their infancy, initial deployments of IAB networks may need to rely on half-duplex operation to coordinate transmissions between access and backhaul links. However, the use of half-duplex operation not only makes the scheduling of links in the IAB networks interdependent, but also the number of their feasible combinations grows exponentially with the network size, thereby posing challenges to the efficient design of such systems. In this paper, by accounting for mmWave radio characteristics, we propose a joint resource allocation and link scheduling framework to enhance the user equipment (UE) throughput in multi-hop in-band IAB systems. We keep the problem in the form of linear programming type for the feasibility of the practical applications. We show that the increased number of uplink and downlink transmission time interval (TTI) configurations does not result in improved UE throughput as compared to two-TTI configuration. Further, we demonstrate that in-band IAB systems tend to be backhaul-limited, and the utilization of multi-beam functionality at the IAB-donor alleviates this limitation by doubling the average UE throughput. Finally, we show that the use of proportional-fair allocations allows the average UE throughput to be improved by around 10% as compared to the max-min allocations.

INDEX TERMS 5G mobile communication, half-duplex operation, integrated access and backhaul, scheduling, mmWave, resource allocation.

I. INTRODUCTION

AS THE standardization of fifth-generation (5G) New Radio (NR) technology continues with its 5G-Advanced phase, cellular network operators have already initiated commercial deployments of 5G NR cellular networks, which operate in both sub-6 GHz and millimeter-wave (mmWave) bands and promise to deliver gigabit data rates for enhanced mobile broadband (eMBB) services. However, the usage of mmWave radios is hampered by adverse propagation properties, including severe path loss caused by obstacles and diffuse scattering phenomenon [1]. As a result, ubiquitous coverage in the mmWave band requires the dense deployment of small cells, hence resulting in high capital expenditures for network operators [2].

To address the aforementioned challenges of mmWave-based 5G NR systems, 3GPP has recently proposed Integrated Access and Backhaul (IAB) technology [3]. Using wireless backhauling to base stations (BSs) without fiber connectivity, which are named IAB-nodes, IAB enables data relaying between user equipment (UE) and BSs that are connected to the 5G core (5GC) network, which are named IAB-donors. IAB promises a flexible solution for network

densification as well as reduction of capital and operational expenditures. Characterized by improved coverage and capacity with cost-efficient backhauling and spectrum efficiency [4], such systems are becoming an attractive option for mmWave 5G network rollouts, while also mitigating the dependency on the availability of a wired connection to the 5GC at each network access location [5], [6]. Several cellular operators have already demonstrated interest in implementing IAB systems within their 5G networks [7]. It is expected that IAB technology is to be employed in up to $10-20\%$ of 5G sites [8].

One of the essential constraints of IAB networks is the scheduling of wireless transmissions at IAB-nodes. Similar to conventional BSs, IAB-nodes are expected to be equipped with multiple antenna arrays covering different sectors and performing electronic beam steering to serve UEs as well as backhaul links. This raises several issues related to: (i) simultaneous transmission or reception and (ii) simultaneous transmission and reception, also known as full-duplex operation. While the former can be addressed by exploiting spatial multiplexing and scheduling redundancies, as shown in [9], the latter remains feasible only for specific deployment scenarios. Despite significant progress in full-duplex communications [10], [11], such systems are still in their early days. As a result, 3GPP in [3] strictly mandates the support of half-duplex operation in IAB systems.

The use of half-duplex regime in IAB deployments poses unique challenges to system designers. The reason is that the transmission time intervals (TTIs) utilized for data transfer in multi-hop half-duplex systems are no longer independent. In addition, 5G NR introduces flexibility in terms of TTI allocation in both downlink (DL) and uplink (UL) directions. To ensure that radio resource allocations can maximize the system throughput, one needs to consider numerous potential TTI configurations across the network in both DL and UL directions. As a result, the resource allocation problem becomes combinatorial in nature, the exact solution of which is computationally demanding for practical implementation [12], [13]. However, it remains unclear whether one needs to consider all the possible TTI combinations in practical network optimization.

Despite the expectation that IAB systems may be deployed in up to 20% of 5G sites and the requirement by 3GPP that IAB deployments should support half-duplex regime, most studies assume full-duplex operation mode. The rationale is that full-duplex assumption allows for streamlined optimization solutions, such as those based on conventional network flow control formalism [14], as all the links may be utilized in the network simultaneously. However, if IAB system operates in half-duplex mode, not all the links are available simultaneously, which cannot be easily accounted for in a conventional optimization formulation.

One of the approaches to address this constraint is to formalize it by utilizing the tools from queuing networks and stochastic control theory. This approach was employed in several works, such as [12] and [15]. The major shortcoming of this method is in that the resulting problem is formulated in terms of a Markov decision process (MDP) with a large state space and thus requires approximate solution methods, such as those based on reinforcement learning. While the latter is commonly utilized, it does not guarantee the existence of a solution and makes unclear how close the obtained result is to the optimal one. Another way to address the aforementioned problem is to enumerate all the duplexing patterns and formulate a network flow problem. A major advantage of this proposed technique is that it relies upon conventional solution tools for the specified problem and guarantees that the optimal solution is eventually found.

Our subject work aims to characterize the performance trade-offs associated with the IAB topology design, comprehensive TTI allocation schemes, and multi-beam operation. We consider an in-band time division multiplexing (TDM)-based IAB system with the half-duplex operation. We propose a performance optimization framework capable of determining radio resource allocations and the number of feasible link configurations in the system, henceforth referred to as link scheduling patterns [12], while also keeping the complexity of linear programming. The developed framework takes into account mmWave propagation conditions, multi-beam antenna design, and multi-hop communication characteristics.

More specifically, we propose an approach for optimizing the airtime of the duplexing patterns in half-duplex IAB systems to maximize the UE throughput in both UL and DL. The formulated method determines the exact optimized solution and does not necessitate approximations. The resulting formulation falls under the classification of linear programming, which is characterized by polynomial complexity. As the main metric of interest, we consider the end-to-end throughput of UEs under max-min and proportional fairness (PF) criteria. By using this framework, we (i) identify link scheduling patterns for improved IAB operation, (ii) analyze performance of single- and multi-hop topologies, (iii) characterize multi-beam performance gains, and (iv) compare max-min and PF policies.

The main contributions of our study are as follows.

- Using the decomposition approach, we propose a linear programming framework for UE resource allocation and link scheduling in IAB systems by accounting for multi-hop communication, flexible TTI durations, and half-duplex constraint.
- We demonstrate that for improved performance of mmWave 5G in-band IAB networks with simultaneous transmission or reception, it is sufficient to select only two TTIs, since increasing the number of TTIs does not lead to improved average UE throughput.
- We show that in-band IAB systems tend to be backhaul-limited by having underutilized resources at the access interfaces of IAB-nodes and that the usage of multi-beam capabilities at the IAB-donor can drastically improve the average UE throughput.

- We compare different UE resource allocation policies and establish that the PF policy allows for improved efficiency of resource utilization as compared to the max-min option, hence generally achieving higher average UE throughput at the expense of fairness degradation.

The rest of this text is organized as follows. First, in Section II, we provide essential background on the 3GPP IAB architecture and review the related literature. Then, we introduce our system model in Section III. Our performance optimization framework is developed in Section IV. Numerical results are presented in Section V. Finally, conclusions are drawn in the last section.

II. BACKGROUND AND RELATED WORK

In this section, we first outline certain technical details of the IAB technology and then proceed to review the related work.

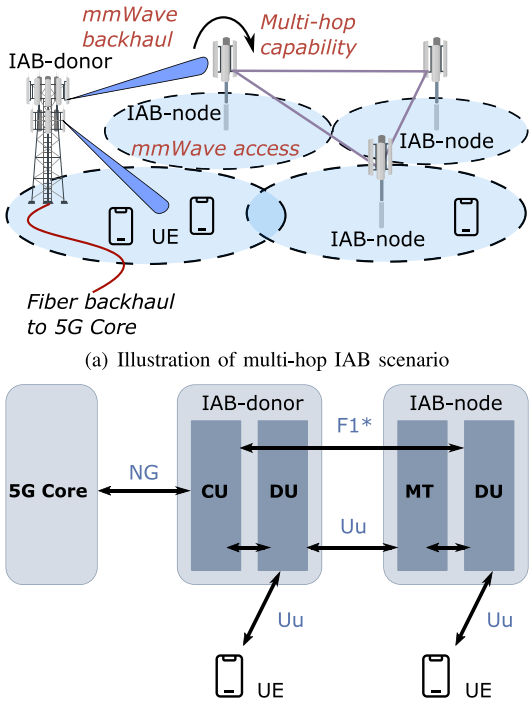
A. IAB NETWORKS

IAB considerations were introduced as part of 3GPP Release 16 of the 5G NR specifications [3]. As shown in Fig. 1(a), IAB design relies on the usage of wireless radio access interfaces not only for links between BSs and UEs but also for backhaul links to relay data traffic through BSs without fiber connectivity, known as IAB-nodes. Accordingly, BSs connected to the 5GC act as IAB-donors, while IAB-nodes wirelessly attach to them. IAB architecture facilitates network flexibility by enabling multi-hop architecture, adaptive topology, and dynamic resource allocation.

IAB-donor comprises a centralized unit (CU) and a distributed unit (DU), see Fig. 1(b). The CU is a logical node that arbitrates all the control and upper-layer functionalities, such as Radio Resource Control (RRC), mobility control, and session management. On the other hand, the DU supports lower-layer operations, such as Radio Link Control (RLC), MAC, and PHY functions. The ultimate goal is to encapsulate the time-critical functionalities, e.g., scheduling and retransmission so that they are performed at DU, while keeping other functions at CU.

Each IAB-node supports DU and mobile terminal (MT) functionalities. The MT part manages radio and protocol interface layers to parent IAB-nodes or IAB-donors, while the DU part provides connectivity to UEs and the MT parts of child IAB-nodes. The DU is also connected to the CU hosted by the IAB-donor using the NR F1* interface. For legacy compatibility, the MT part of an IAB-node acts as a regular UE from the perspective of its serving BS. Similarly, from a UE perspective, the DU part of an IAB-node manifests as the DU of a conventional BS.

IAB technology supports both sub-6 GHz and mmWave spectrum bands, and can operate in standalone (SA) or non-SA (NSA) mode. In practice, IAB is highly relevant for mmWave frequencies, where backhaul links can leverage a larger amount of spectrum and further benefit from massive beamforming. Also, 3GPP considers IAB networks under



(a) Illustration of multi-hop IAB scenario

(b) IAB architecture and functional split

FIGURE 1. Topological and technological features of IAB-based systems.

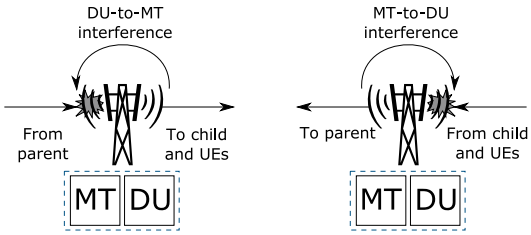
both in-band and out-of-band operational modes. In the in-band mode, access and backhaul functions are multiplexed within the same frequency band, whereas in the out-of-band mode, the access and the backhaul links reside in separate frequency bands. The in-band operational mode with dynamic partitioning of access and backhaul spectrum is currently being preferred by 3GPP, together with half-duplex operation of the IAB-nodes, since it allows for efficient utilization of scarce spectrum resources [3].

In-band mode requires the multiplexing of both the access and the backhaul transmissions within the same frequency band. Hence, the radio resources have to be orthogonally divided between the access and the backhaul components, either in time, frequency, or space, thereby requiring a centralized or decentralized scheduling coordination mechanism.

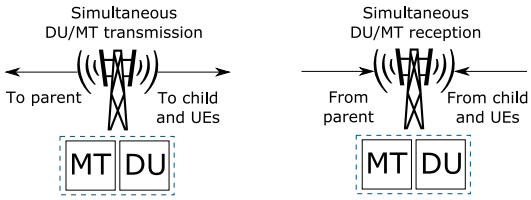
B. DU/MT COORDINATION

In a time division duplex (TDD)-based IAB network, where DU and MT parts of an IAB-node operate over the same frequency band, it is necessary to implement coordination of time-domain resources between the DU and MT. The resource coordination mechanisms and related signaling are explicitly supported by the IAB specifications [3], [16]. It is noteworthy that the time-domain coordination applies to both in-band backhauling, where backhaul and access links reside in the same frequency band, and multi-hop out-of-band backhauling, where MT and DU are used for backhaul links [17].

The time-domain coordination is required to avoid the full-duplex problem, where transmissions to be received by the MT are severely interfered by DU transmissions, as



(a) Illustration of intra-node interference



(b) Illustration of half-duplex operation in IAB system

FIGURE 2. Full-duplex and half-duplex transmissions.

well as DU reception is interfered by MT transmissions, see Fig. 2(a). One of the approaches to enable full-duplex DU/MT operation and avoid extreme intra-node interference is to propagation-wise isolate the DU and the MT parts of an IAB-node, which is feasible in certain deployment scenarios. For example, an IAB-node may provide outdoor-to-indoor service with the MT part mounted outside, while the DU part is located inside.

To avoid intra-node interference in half-duplex systems, one needs to ensure that the DU and MT transmissions or receptions are separated in the spatial domain, where they operate simultaneously on the same frequency band but within different antenna panels pointing in opposite directions. As illustrated in Fig. 2(b), one can envision simultaneous DU/MT operation, where the DU and the MT perform in different transmission directions. This applies to both simultaneous DU UL and MT DL, as well as simultaneous DU DL and MT UL transmissions. Coupled with mmWave directionality, this may result in limited self- and inter-cell interference.

Due to hardware limitations and self-interference, 3GPP initially assumed time separation between DU and MT. A more recent technical specification entitled "Integrated access and backhaul radio transmission and reception" has been approved for Release 18, which aims at improving efficiency and reducing latency [16]. Specifically, it incorporates simultaneous DU and MT operations within an IAB-node, where the DU and the MT perform in different transmission directions. While the full-duplex regime between the MT and the DU is still under development and remains feasible only for specific deployment scenarios, the half-duplex option is more feasible and may be deployed in various environments.

C. RELATED WORK

IAB networks have recently attracted increased interest from academic and industrial communities. Here, we examine the related literature and summarize the current state-of-the-art knowledge.

An overview of IAB systems is provided in [18], where their design and architecture options are discussed. Focusing on the impact of interference, the authors in [19] study the performance of moving IAB-nodes and propose a solution based on inserting silent slots over the TDD frame pattern. In [20], the scheduling of access and backhaul links is studied in mmWave IAB networks. The results show that the minimum achievable throughput increases with growing network densification and refined antenna parameters. In [21], a flexible TDD-based IAB network is considered, where IAB-donor and IAB-nodes operate in different UL/DL modes. The authors formulate and solve an optimization problem to minimize the weighted user queues during two time slots, while considering both UL and DL.

Resource allocation-focused studies have also proliferated recently. In [22], the authors develop a resource allocation scheme for IAB networks, which aims to improve access link reliability and ensure end-to-end quality of service (QoS) for the traffic flows in a multi-connectivity environment. By considering multi-connectivity together with multi-hop backhauling, that scheme addresses the joint route selection and resource allocation problem for differentiated services. The authors of [13] propose a semi-centralized resource allocation framework for IAB networks, which employs the Maximum Weighted Matching method on graphs. Their results show improved end-to-end throughput as well as decreased overall network congestion.

In [23], a resource allocation scheme based on the sequential convex programming approach for IAB mmWave-enabled cellular networks is proposed to increase the network capacity. The authors consider a system with one IAB-donor and multiple IAB-nodes, where users can reuse the backhaul bandwidth resources. Their simulation results confirm the effectiveness of the proposed algorithm, which can achieve a 25% gain in terms of the system throughput. Another resource allocation-related solution is provided in [24]. In that paper, the resource allocation problem is explored by considering fairness between UEs in IAB networks. A game-theoretic solution is formulated in both centralized and distributed ways to improve user experience and network throughput.

In [25], a joint resource allocation and routing optimization problem is studied to reduce the operational expenses for a mmWave-ready IAB network. A mixed-integer linear programming-based formulation is presented and solved with a branch-and-bound algorithm. By using ray-tracing tools under a realistic simulation setup, the authors show that IAB networks significantly reduce fiber deployment costs. Further, the authors of [26] present a joint resource allocation and the next hop selection optimization problem for a multi-hop IAB network to improve the average UE data rate. The proposed method focuses on an optimized mesh topology by maximizing a defined objective function. Their results suggest that IAB networks significantly improve the UE data rates as compared to traditional cellular networks.

Several studies adopt the use of machine learning techniques for IAB optimization. Specifically, in [27], the authors formulate a resource allocation problem in an IAB network as a non-convex mixed integer programming problem. To tackle it, they develop a framework based on deep reinforcement learning for dynamically allocating spectrum resources. The authors of [28] propose optimized resource allocation algorithms that maximize the weighted sum-rate performance with a nominal IAB-node placement and then find the preferred locations by using the proposed resource allocation solutions. In [29], the authors introduce a reinforcement learning-based resource allocation approach to increase throughput in mmWave IAB networks. For the multi-hop network architecture, they tackle flow allocation and link scheduling problems by jointly managing access and backhaul bandwidth.

Even though 3GPP currently focuses on half-duplex operation in IAB networks, the full-duplex mode is also under discussion. In [30], the authors consider full-duplex transmission in multi-hop IAB networks and derive a closed-form solution for ergodic capacity. They show an advantage of full-duplex operation over half-duplex regime. A closed-form expression for optimal transmit power in the in-band full-duplex IAB system is derived in [31] by aiming to maximize the network sum-rate of DL and UL. The authors evaluate the IAB network under realistic system settings and provide numerical results based on the obtained analytical expressions. The authors of [32] investigate the potential of full-duplex technology in mmWave-based IAB networks to address the latency and throughput requirements. They derive practical constraints by using queueing theory, formulate a network utility maximization problem, and characterize the improvements from upgrading to full-duplex IAB-nodes.

Despite the growing body of literature on IAB networks, several research areas remain for further improvement. First, most previous works assume fixed UL and DL TTI patterns, thereby disregarding the impact of flexible TTI allocations on the system performance. This limitation may result in suboptimal resource allocations, particularly under dynamic network conditions. Second, previous studies predominantly focus on single-hop communication, hence overlooking the potential benefits and challenges of multi-hop IAB networks, which can substantially affect the overall system performance.

Our present study addresses these gaps by proposing a comprehensive framework for UE resource allocation and link scheduling in mmWave IAB systems. This methodology incorporates multi-hop communication, flexible TTI patterns, and half-duplex constraints. By introducing these important considerations, our solution enhances the efficiency and the adaptability of IAB networks. Communications service providers can leverage this framework to optimize the network performance, especially in scenarios with varying traffic loads and dynamic topologies, thereby improving both service quality and operational efficiency.

TABLE 1. Notation used in this work.

Parameter	Definition
Main parameters	
$x_{tn}^{(m)}$	UE time allocation
$y_t^{(m)}$	Backhaul time allocation
ϵ_t	TTI t duration
z	Lower bound in max-min optimization problem
$d_n^{(m)D}$	UE DL data rate
$d_n^{(m)U}$	UE UL data rate
δ_{tn}	Access DL incidence variable
φ_{tn}	Access UL incidence variable
γ_t^m	Backhaul DL incidence variable
ζ_t^m	Backhaul UL incidence variable
Additional parameters	
M	Number of IAB-nodes
N_0	Number of UEs connected to IAB-donor
N_m	Number of UEs connected to IAB-node m
h_B	Antenna height at BS
h_U	Antenna height at UE
d_2, d_3	2D and 3D distances
R	Radius of macro-cell
p_B	Small-scale blockage probability
p_L	Large-scale LOS probability
λ_p	Blocker density
r_p	Blocker radius
h_p	Blocker height
κ_i	LOS/blockage state probability
B	System bandwidth
$s_n^{(m)D}$	UE DL spectral efficiency
$s_n^{(m)U}$	UE UL spectral efficiency
$s_B^{(m)D}$	Backhaul spectral efficiency in DL
$s_B^{(m)U}$	Backhaul spectral efficiency in UL
$S_n^{(m)D}$	SINR of access link in DL
$S_B^{(m)D}$	SINR of backhaul link in DL
P_B	Emitted power
G_B	BS antenna gain
G_U	UE antenna gain
I_M	Inter-cell interference
Ω_i	Log-normal shadow fading
$\mathbf{1}_R$	Indicator function of simultaneous reception at IAB-nodes
I_S	Self-interference
W	Number of interfering links
N_0	Thermal noise

III. SYSTEM MODEL

In this section, we introduce our system model by specifying the deployment, topology, antenna, propagation, and traffic considerations. Further, we introduce our metrics of interest.

A. DEPLOYMENT AND TOPOLOGY

By following the spatial configuration employed in 3GPP simulation assumptions for a typical outdoor deployment

scenario of a two-layer heterogeneous network [3], we consider a cellular deployment having IAB-donors associated with a circular macro-cell of radius R and M IAB-nodes deployed within the coverage of the IAB-donor. Our scenario has a constant number of active UEs uniformly distributed in the area. Assuming that each UE is served by only one node, the UE cell association is based on the maximum reference signal received power (RSRP). The number of UEs is then set for IAB-donor and each IAB-node, $N_0, N_1, \ldots, N_M$. We consider the in-band mode, where both IAB-donor and IAB-nodes operate over mmWave bands for backhaul and access transmission and reception. We also assume a TDD, Time Division Multiple Access (TDMA)-based scheduling, where the access and backhaul links are multiplexed in half-duplex manner. Similarly to [12], [13], we require that the topology is fixed as a result of the environment and network operator preferences. The topology formation approach is utilized as an example, and the framework proposed in this study can be applied to other topology formations, e.g., [33], [34].

B. TRANSMISSION AND TRAFFIC

At the IAB-donor, we consider single-lobe and multi-lobe antennas with planar antenna arrays. Each of the lobes (beams) may be steered in a dedicated direction, i.e., we assume hybrid analog-digital or digital beamforming techniques [35] to enable multi-beaming, where multiple beams can be used independently at the same time. At cost-limited IAB-nodes, we resort to single-lobe antennas only. The antenna radiation patterns are calculated by following the evaluation assumptions described in [36].

We examine a frame with the duration of 10 ms, as illustrated in Fig. 3. According to [3], in TDD half-duplex mode, the frame may be further divided into multiple TTIs, each of which can be designated for UL, DL, or remain silent. The 3GPP specifications do not determine the number or arrangement of TTIs; however, they ratify the numerology used to define the TTI and symbol durations. It is important to note that the selection of active TTIs and their directions is influenced not only by traffic conditions but also by the half-duplex constraint, hence potentially impacting the UE performance. To address this, the IAB-donor dynamically communicates the number of TTIs and their arrangement to the IAB-nodes. Our study aims to determine the preferred values for these variables.

In our work, we consider overloaded network conditions, where all the UEs always have data traffic in both DL and UL directions. To this aim, we employ the full-buffer traffic model as specified in [37].

C. CHANNEL CHARACTERISTICS

The propagation of mmWave signals is susceptible to blockage effects due to severe penetration loss as well as higher reflection and diffraction losses. We assume that the wireless backhaul side of the IAB network is designed so that small-scale blockage by smaller dynamic objects, such as vehicles, does not occur. However, on the access side

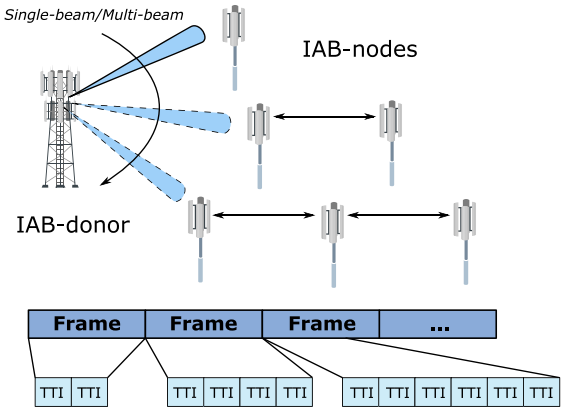


FIGURE 3. Frames and TTIs in the considered system.

of the IAB network, we assume that small-scale as well as large-scale blockage by larger stationary objects, such as buildings, can take place as described below.

First, to model small-scale blockage, we adopt a dynamic blockage model. Accordingly, blockers move by following the random direction model (RDM) [38]. Their density is λ_p units/m^2, r_p is the radius, and h_p is the height. The corresponding time-averaged blockage probability is given by [39]

$$p_B(d_3) = 1 - e^{-2\lambda_p r_p\left[\sqrt{d_3^2-(h_B-h_U)^2}\frac{h_p-h_U}{h_B-h_U}+r_p\right]}, \quad (1)$$

where d_3 is the 3D distance to UE, h_B is the height of serving BS, and h_U is the height of UE.

On top of the small-scale blockage, we account for blockage by large-scale stationary objects, which manifests in line-of-sight (LOS) and non-line-of-sight (NLOS) conditions [40]. We adopt the 3GPP Urban Micro (UMi) street canyon model with the following LOS probability as a function of 2D distance d_2 between BS and UE

$$p_L(d_2) = \begin{cases} 1, & d_2 \le 18 \text{ m} \\ 18d_2^{-1} + e^{-\frac{d_2}{36}}\left(1 - 18d_2^{-1}\right), & d_2 > 18 \text{ m}. \end{cases} \quad (2)$$

In what follows, we enumerate the possible blockage states as: 0 – (NLOS, blocked), 1 – (LOS, blocked), 2 – (NLOS, non-blocked), and 3 – (LOS, non-blocked). The corresponding state probabilities are given by

$$\kappa_0 = \left[1 - p_L\right]p_B,\ \kappa_1 = p_L p_B,$$
$$\kappa_2 = \left[1 - p_L\right]\left[1 - p_B\right],\ \kappa_3 = p_L\left[1 - p_B\right]. \quad (3)$$

The attenuation caused by small-scale objects is added to the LOS and NLOS states, which is assumed to be 20 dB [41].

By following the 3GPP IAB evaluation specifications [3], we employ the 3GPP Urban Macro (UMa) channel model for IAB-donor–to–UE and IAB-donor–to–IAB-node interfaces, and the 3GPP UMi street canyon channel model for IAB-node–to–UE and IAB-node–to–IAB-node interfaces [40]. In general, the value of the signal-to-interference-plus-noise ratio (SINR) at IAB-donor, IAB-node, or UE can be written

as a weighted function of SINR in different states, see (3), which yields

$$S(d_3) = \sum_{i=0}^{3} \frac{C_i(d_3)\Omega_i\kappa_i}{N_0 B + I_M}, \tag{4}$$

where $C_i(d_3) = P_B G_B G_U / L_i(d_3)$, P_B is the transmit power, G_B and G_U are the transmit and receive gains, $L_i(d_3)$ is the path loss in different states as defined above, Ω_i is the log-normal shadow fading available from [40], κ_i is the LOS/blockage state probability as defined in (3), N_0 is the thermal noise, B is the available bandwidth and I_M is the inter-cell interference.

D. MODELING CONSIDERATIONS

Our main metric of interest is the average UE throughput in both DL and UL directions across various IAB configurations. In the following section, we develop a performance optimization framework designed to improve this throughput under different objective functions, such as max-min fairness and PF criteria. This framework not only aims to enhance the overall network efficiency, but also addresses specific system requirements and constraints, thereby providing a robust solution for diverse IAB-based scenarios.

By focusing on the UE throughput, we conduct a comprehensive qualitative assessment of IAB deployments considering the half-duplex constraint. This analysis includes (i) identifying the optimized link scheduling patterns, (ii) examining both single-hop and multi-hop topologies, (iii) characterizing the gains from multi-beam configurations, and (iv) comparing the effectiveness of max-min and PF policies. The novelty of our work lies in its holistic approach, which integrates these diverse aspects to offer a detailed understanding of IAB network performance. Mobile operators and service providers can utilize these insights to implement more efficient IAB deployments, hence leading to enhanced network capacity, better resource utilization, and improved user experience.

IV. OPTIMIZATION FRAMEWORK

In this section, we introduce our optimization framework. Instead of formulating the problem at hand similarly to previous works [12], [15], which utilize the tools of control theory and MDP methods inherently leading to approximate solutions, we propose to employ elements of the conventional network flow formalism [14]. To avoid mixed integer programming (MIP) problem formulation characterized by the exponential time complexity, in this section, we decompose the problem into two parts: duplexing patterns enumeration and linear programming framework formalization.

As the linear programming framework involves multiple variables for both access and backhaul links, we introduce it gradually starting with two TTIs in Section IV-B. We (i) demonstrate how to introduce the flow variables representing the airtime provided to the UE at the access in (5) and to the traffic aggregates at the backhaul in (6), (ii) formalize the objective function in (7), and (iii) supplement the formulation with access, backhaul, and TTI constraints in (8), (11), (12), and (16), respectively, by accounting for radio-specific factors including interference.

Further on, we extend the two-TTI notion to the case of multiple TTIs. The key difference lies in the increased number of potential duplexing patterns in the network, which results in a higher number of equations for flow variables and constraints. Finally, extensions for the multi-chain topology and different objective functions are provided in Section IV-C.

A. DUPLEXING PATTERNS

To avoid formulating our challenge as a MIP problem, which would then require accounting for all possible combinations of duplexing patterns in a single optimization task, we divide the proposed framework into two components. In the first stage, we enumerate all possible combinations of duplexing patterns. In the second stage, we introduce the linear programming constructs that need to be applied to a certain number of TTIs utilized for planning the transmission cycle in our IAB system.

All possible combinations of DL and UL transmissions for a chain of two IAB-nodes are shown in Fig. 4. Note that the number of these combinations depends on the number of TTIs and can be calculated using conventional algorithms such as Maximum Independent Set Problem [42]. By counting the combinations, the total number of all feasible configurations for such a network is 18, while the number of configurations grows exponentially with the number of IAB-nodes in the chain. However, the overall problem remains linear in nature, which allows it to be solved in polynomial time.

B. LINEAR PROGRAMMING FRAMEWORK

To explain the introduced optimization framework gradually, we start with a special case being a two-TTI system, and then extend it to the case of multiple TTIs.

1) TWO-TTI SYSTEM

We begin by considering a single-chain topology as shown in Fig. 5. Here, we assume that the frame is divided into two TTIs with unknown durations, which we define with variables ϵ_1 and ϵ_2. Then, we introduce the variables

$$x_{tn}^{(m)},\ m = 0, 1, \ldots, M,\ t = 1, 2,\ n = 1, 2, \ldots, N_m, \tag{5}$$

which represent the fraction of time allocated to UE n connected to node m in TTI t, where M is the number of IAB-nodes, and N_m is the number of UEs associated with node m.

Similarly, we consider continuous variables responsible for the fraction of time for the backhaul links allocated to node m in TTI t

$$y_t^{(m)},\ m = 1, 2, \ldots, M,\ t = 1, 2. \tag{6}$$

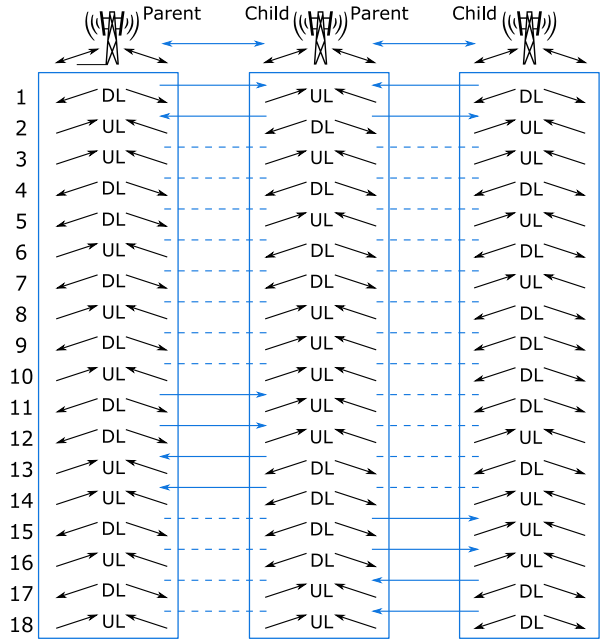


FIGURE 4. DL and UL (from the UE perspective) TDD configurations with half-duplex constraint. The blue arrows show the backhaul link transmission directions and the dash lines refer to no transmission between nodes.

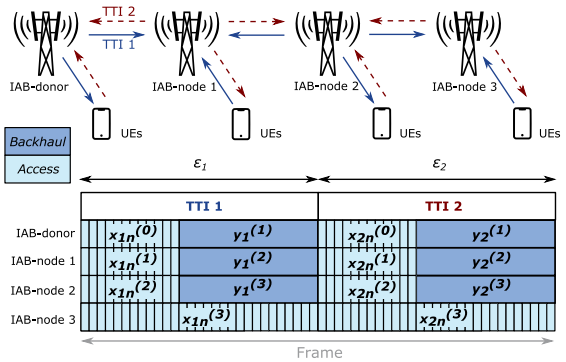


FIGURE 5. An illustration of single-chain two-TTI system.

For the given system, a network operator may be interested in different optimization functions. As an example, we utilize max-min fairness criterion, where the objective function takes the following form

$$\textbf{maximize} \min \left(d_n^{(m)D}, d_n^{(m)U}\right), \quad n = 1, 2, \ldots, N_m,\ m = 0, 1, \ldots, M, \quad (7)$$

where $d_n^{(m)D}$ and $d_n^{(m)U}$ are the DL and UL data rates provided to each UE n at every node m.

We now proceed by specifying the constraints for this optimization problem. The objective function in (7) can be alternatively expressed by maximizing an additional variable z that is a lower bound for each of the individual variables. Specifically, the first set of constraints represents the achievable data rates of each UE in the DL direction

$$\begin{cases} z \le d_n^{(0)D} = Bs_n^{(0)D} x_{1n}^{(0)}, & n = 1, \ldots, N_0, \\ z \le d_n^{(1)D} = Bs_n^{(1)D} x_{2n}^{(1)}, & n = 1, \ldots, N_1, \\ z \le d_n^{(2)D} = Bs_n^{(2)D} x_{1n}^{(2)}, & n = 1, \ldots, N_2, \\ z \le d_n^{(3)D} = Bs_n^{(3)D} x_{2n}^{(3)}, & n = 1, \ldots, N_3, \\ \ldots & \ldots \\ z \le d_n^{(M)D} = Bs_n^{(M)D} x_{tn}^{(M)}, & n = 1, \ldots, N_M, \end{cases} \quad (8)$$

where B is the available bandwidth, and $s_n^{(m)D}$ is the spectral efficiency of DL for UE n connected to node m. Note that depending on the number of IAB-nodes (odd/even), the TTI number for the last node will be different, see Fig. 5, i.e., if M is odd, then DL is in the second TTI $t = 2$; otherwise, DL is in the first TTI $t = 1$.

The spectral efficiency is calculated as follows

$$s_n^{(m)D} = \log_2\left(1 + S_n^{(m)D}\right), \quad (9)$$

where $S_n^{(m)D}$ is the SINR of the access link established as in (4). The only unknown in (4) is the interference component. Considering a cellular 3GPP deployment of the IAB system having six neighboring cells, the interference at a randomly chosen UE can be approximated by

$$I_M = \sum_{i=1}^{W} \sum_{j=0}^{3} C_{i,j}\left(d_{3,i}\right)\Omega_{i,j}\kappa_{i,j}, \quad (10)$$

where W is the number of interfering links and $d_{3,i}$ is the distance to i-th interferer. Note that the actual interference varies with the amount of allocated resources. To keep our model complexity reasonable, we assume the worst-case scenario, where all the potential interferers produce interference toward the node of interest.

The UE data rates in UL are defined similarly to (8) as

$$\begin{cases} z \le d_n^{(0)U} = Bs_n^{(0)U} x_{2n}^{(0)}, & n = 1, \ldots, N_0, \\ z \le d_n^{(1)U} = Bs_n^{(1)U} x_{1n}^{(1)}, & n = 1, \ldots, N_1, \\ \ldots & \ldots \\ z \le d_n^{(M)U} = Bs_n^{(M)U} x_{tn}^{(M)}, & n = 1, \ldots, N_M, \end{cases} \quad (11)$$

where $s_n^{(m)U}$ is the spectral efficiency of UL for UE n connected to node m that is computed similarly to (9) with SINR and interference expressed according to (4) and (10). Similarly to (8), the TTI number for the last node will be different depending on the total number of IAB-nodes, i.e., if M is odd, then UL is in the first TTI $t = 1$; otherwise, UL is in the second TTI $t = 2$.

The UE data rates in both directions may be limited not only by access, but by backhaul as well. Specifically, the sum of the data rates at the access interface of a particular IAB-node should be less than or equal to the capacity of the backhaul link connecting the considered IAB-node to the parent node. We assume that the channel is static, which means that it remains unchanged throughout a transmission. We also consider the pipelining of the data, where the time allocations are repeated periodically to ensure that

the inequalities hold and that the data rate calculations are accurate. The two inequalities for DL and UL are associated with each IAB-node. Since a frame is divided into two TTIs, the TTI number differs on either side of the inequality. With this in mind, we introduce backhaul data rate constraints as follows

$$\begin{cases} \sum_{n=1}^{N_1} s_n^{(1)D} x_{2n}^{(1)} + \sum_{n=1}^{N_2} s_n^{(2)D} x_{1n}^{(2)} + \ldots \le s_B^{(1)D} y_1^{(1)}, \\ \sum_{n=1}^{N_1} s_n^{(1)U} x_{1n}^{(1)} + \sum_{n=1}^{N_2} s_n^{(2)U} x_{2n}^{(2)} + \ldots \le s_B^{(1)U} y_2^{(1)}, \\ \sum_{n=1}^{N_2} s_n^{(2)D} x_{1n}^{(2)} + \sum_{n=1}^{N_3} s_n^{(3)D} x_{2n}^{(3)} + \ldots \le s_B^{(2)D} y_2^{(2)}, \\ \sum_{n=1}^{N_2} s_n^{(2)U} x_{2n}^{(2)} + \sum_{n=1}^{N_3} s_n^{(3)U} x_{1n}^{(3)} + \ldots \le s_B^{(2)U} y_1^{(2)}, \\ \ldots \\ \sum_{n=1}^{N_M} s_n^{(M)D} x_{1n}^{(M)} \le s_B^{(M)D} y_2^{(M)}, \\ \sum_{n=1}^{N_M} s_n^{(M)U} x_{2n}^{(M)} \le s_B^{(M)U} y_1^{(M)}, \end{cases} \tag{12}$$

where $s_B^{(m)D}$ and $s_B^{(m)U}$ are the spectral efficiencies of the backhaul link for node m in DL and UL, respectively. Note that similarly to (8) and (11), depending on the number of IAB-nodes (odd/even), the TTI numbers in the last two equations are different.

Note that SINR for a backhaul link in DL is computed as

$$S_B^{(m)D}(d_3) = \frac{C_{m,3}(d_3)\Omega_{m,3}}{N_0 B + I_M}, \tag{13}$$

where state 3 indicates that the link is in the LOS conditions and is not blocked, see (3).

When calculating the interference for the backhaul links, blockage by smaller obstacles such as humans and vehicles can be disregarded due to larger heights of the IAB-nodes and IAB-donors considered in the IAB system as per [3]. However, one still needs to account for large-scale blockage, i.e.,

$$I_M = \mathbf{1}_R I_S + \sum_{i=1}^{W} \sum_{j=2}^{3} C_{i,j}\big(d_{3,i}\big)\Omega_{i,j}\kappa_{i,j}, \tag{14}$$

where $\mathbf{1}_R$ is the indicator function of simultaneous reception and I_S is the self-interference. Note that I_S can be approximated by its upper bound of 3 dB, as reported in [9].

Finally, we ensure that the allocated time utilized for data rate calculations is less than unity, which is

$$\epsilon_1 + \epsilon_2 = 1, \tag{15}$$

while the sum of individual allocations at both access and backhaul obeys the TTI durations. Note that the superscripts are off by 1 for $x_{tn}^{(m)}$ and $y_t^{(m)}$, since the variables responsible for the time allocation on backhaul links are associated with the child node, as shown in Fig. 5. Also, recall that each IAB-node is equipped with two sets of radios, one of which is associated with the MT part and used for backhaul connection to the parent node, while the other one is associated with the DU part and shares the time between the access links and the backhaul to the child node. Keeping this in mind, the final TTI constraints are expressed as

$$\begin{cases} \sum_{n=1}^{N_0} x_{1n}^{(0)} + y_1^{(1)} \le \epsilon_1, \\ \sum_{n=1}^{N_1} x_{1n}^{(1)} + y_1^{(2)} \le \epsilon_1, \\ \ldots \\ \sum_{n=1}^{N_M} x_{1n}^{(M)} \le \epsilon_1, \\ \sum_{n=1}^{N_0} x_{2n}^{(0)} + y_2^{(1)} \le \epsilon_2, \\ \sum_{n=1}^{N_1} x_{2n}^{(1)} + y_2^{(2)} \le \epsilon_2, \\ \ldots \\ \sum_{n=1}^{N_M} x_{2n}^{(M)} \le \epsilon_2. \end{cases} \tag{16}$$

The overall problem is therefore formulated as

$$\max_{x,y,z,\epsilon} \; z \quad \text{s.t. (8), (11), (12), (16).} \tag{17}$$

Note that due to the nature of the constraints and the objective function, the optimization function in question is a linear programming problem that can conveniently be solved using conventional tools with polynomial time complexity. Therefore, even for a higher than typical number of IAB-nodes associated with the IAB-donor, a feasible solution can be obtained.

2) MULTI-TTI SYSTEM

Similar to the two-TTI system, we now proceed to formulate the multi-TTI optimization problem. Continuous variables representing the fraction of time for UE n associated with node m in TTI t can therefore be written as

$$x_{tn}^{(m)}, \quad m = 0, 1, \ldots, M, \quad t = 1, 2, \ldots, T, \; n = 1, 2, \ldots, N_m. \tag{18}$$

In contrast to the two-TTI system, the achievable data rate of each UE is calculated as a sum of the data rates in each TTI. To streamline the formulation and differentiate between DL and UL, we introduce link incidence variables to indicate the directions of the links, whose values are determined by the patterns in Fig. 4 for each TTI

$$\delta_{tn} = \begin{cases} 1, \text{ if link of UE } n \text{ is DL}, \\ 0, \text{ otherwise}. \end{cases} \tag{19}$$

$$\varphi_{tn} = \begin{cases} 1, \text{ if link of UE } n \text{ is UL}, \\ 0, \text{ otherwise}. \end{cases} \tag{20}$$

Similarly, for backhaul links

$$\gamma_t^m = \begin{cases} 1, \text{ if backhaul link of node } m \text{ is DL}, \\ 0, \text{ otherwise}. \end{cases} \tag{21}$$

$$\zeta_t^m = \begin{cases} 1, \text{ if backhaul link of node } m \text{ is UL}, \\ 0, \text{ otherwise}. \end{cases} \tag{22}$$

The DL and UL data rate constraints for all UEs in the system are defined as

$$z \le d_n^{(m)D} = Bs_n^{(m)D} \sum_{t=1}^{T} \delta_{tn} x_{tn}^{(m)},$$
$$z \le d_n^{(m)U} = Bs_n^{(m)U} \sum_{t=1}^{T} \varphi_{tn} x_{tn}^{(m)},\ m = 0, 1, \ldots, M. \quad (23)$$

The backhaul constraints take the following form

$$\sum_{n=1}^{N_m} \sum_{t=1}^{T} s_n^{(m)D} \delta_{tn} x_{tn}^{(m)} \le \sum_{t=1}^{T} s_B^{(m)D} \gamma_t^m y_t^{(m)},$$
$$\sum_{n=1}^{N_m} \sum_{t=1}^{T} s_n^{(m)U} \varphi_{tn} x_{tn}^{(m)} \le \sum_{t=1}^{T} s_B^{(m)U} \zeta_t^m y_t^{(m)},\ m = 1, \ldots, M. \quad (24)$$

In addition to the normalization constraint

$$\sum_{t=1}^{T} \epsilon_t = 1, \quad (25)$$

the following TTI constraints need to be satisfied

$$\sum_{n=1}^{N_m} x_{tn}^{(m)} + y_t^{(m+1)} \le \epsilon_t,\ m = 0, 1, \ldots, M-1,$$
$$\sum_{n=1}^{N_M} x_{tn}^{(M)} \le \epsilon_t, \quad (26)$$

which should hold for all the TTIs $t = 1, 2, \ldots, T$. Finally, the overall problem can be written similarly to (17) subject to the constraints in (23), (24), (25), and (26).

Note that the number of constraints depends on the number of TTIs considered in the system as illustrated in Fig. 4.

C. MULTI-CHAIN TOPOLOGY AND OTHER EXTENSIONS

To extend our formulation of the optimization problem to the case of a multi-chain multi-TTI system, it is necessary to introduce an additional set of constraints specified for the given number of chains. The difference in the formulation emerges for the TTI constraints at the IAB-donor, when it shares the time across child IAB-nodes, while the rest of the constraints remain the same. That is, by adding $\mu_t^{(c)}$, we have

$$\sum_{n=1}^{N_0} x_{tn}^{(0)} + \sum_{c=1}^{C} \mu_t^{(c)} \le \epsilon_t, \quad (27)$$

where C is the total number of child IAB-nodes at the first hop and $\mu_t^{(c)}$ is the fraction of time of their backhaul links.

The time-sharing limitation of the backhaul links can be alleviated by utilizing multi-beam technology at the IAB-donor, which can be implemented via hybrid/digital beamforming [35]. Particularly, multiple beams directed at child IAB-nodes allow for the total bandwidth to be allocated without time division, thus multiplying the amount of the available radio frequency resources. In these settings, one beam can be dedicated to the UEs directly attached to the IAB-donor, while each child IAB-node may employ a separate beam. In essence, this approach is similar to parallel transmissions for each of the IAB chains. Specifically, the TTI constraint for the IAB-donor has to then be modified as follows

$$\sum_{n=1}^{N_0} x_{tn}^{(0)} \le \epsilon_t,$$
$$\mu_t^{(c)} \le \epsilon_t,\ c = 1, \ldots, C. \quad (28)$$

Not limited to the max-min optimization function, another commonly used allocation criterion is PF, which can also be considered by adopting the following objective function

$$\max \sum_{m=0}^{M} \sum_{n=1}^{N_m} \left(\log d_n^{(m)D} + \log d_n^{(m)U} \right). \quad (29)$$

In (29), one can alternatively consider the logarithm of the sum of allocations in both directions if the channel is nearly symmetric across UL and DL. Note that one can also account for additional characteristics, such as asymmetry of traffic demands in the DL and UL directions, or priorities between data flows. Both considerations can be introduced via weighting coefficients in either the max-min (17) or the PF (29) formulation [14].

As compared to max-min fairness, PF is known to provide better results in terms of the overall system throughput while at the same time preventing from extremely large or small allocations. However, it does so at the cost of sacrificing some fairness among the UEs. Note that the resulting problem remains convex in nature and thus can still be solved in polynomial time [14]. Also, the range extension technique applied to spectral efficiency can be used to enforce the α-fairness objective function that captures the trade-offs between throughput and fairness, while keeping the solution within the class of linear programming [43], [44].

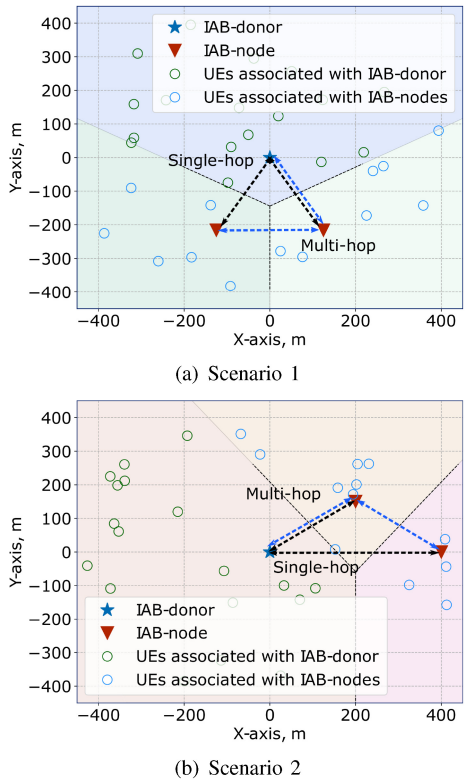
Formulating an optimization problem for IAB networks presents a challenging task in itself, due to the complexity of accommodating various link scheduling patterns and topology configurations. This complexity is further aggravated by the need to account for half-duplex constraints and dynamic network conditions, which significantly impact the feasibility and efficiency of potential solutions. As a solution algorithm for all the following numerical results, we utilize the advanced process optimizer (APOPT) of the GEKKO optimization suite as an internal solver [45]. APOPT is a software package for addressing large-scale optimization problems. It is an active-set sequential quadratic programming solver that employs a warm-start approach to speed up successive programming functions.

V. NUMERICAL RESULTS

In this section, we elaborate on our numerical results. First, we consider the benefits of multi-hop topologies in the IAB system by using single-hop topology as a baseline. We then assess the need for multiple TTIs to have more flexibility in resource allocation. Further, we address the use of multi-beam antennas, followed by a performance evaluation of

TABLE 2. Default system parameters for numerical assessment.

Parameter	Value
Cell radius, R	450 m
Carrier frequency, f_c	30 GHz
System bandwidth, B	400 MHz
IAB-donor height, h_D	25 m
IAB-node height, h_N	15 m
Effective environment height, h_E	1 m
UE height, h_U	1.5 m
Blocker height, h_p	1.7 m
Blocker radius, r_p	0.3 m
Blocker density, λ_p	0.1 m^{-2}
IAB-donor Tx power, Tx_D	40 dBm
IAB-node Tx power, Tx_N	33 dBm
UE Tx power, Tx_U	23 dBm
IAB-donor antenna gain	15 dBi
IAB-node antenna gain	15 dBi
UE antenna gain	10 dBi
Number of UEs, N	30
Number of IAB-nodes, M	4



(a) Scenario 1

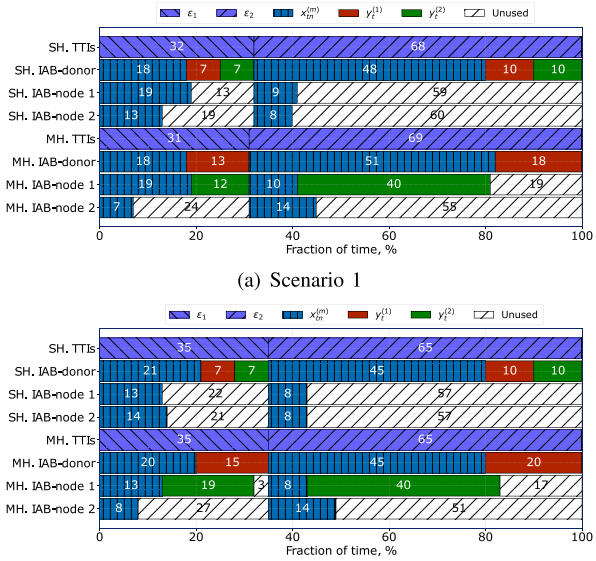
(b) Scenario 2

FIGURE 6. Two considered IAB system deployments.

the IAB system under different objective functions. The default system parameters used in this section are given in Table 2.

A. SINGLE-HOP AND MULTI-HOP TOPOLOGIES

The use of multi-hop transmission is an essential feature in IAB systems. Here, we target to address the scenarios, where



(a) Scenario 1

(b) Scenario 2

FIGURE 7. Time allocations at access and backhaul.

the multi-hop topology provides performance gains in terms of the average UE throughput. To this aim, we consider a two-TTI system with one IAB-donor and two IAB-nodes serving 30 UEs for two different IAB-node deployments within the IAB-donor coverage as shown in Fig. 6. In the first deployment, see Fig. 6(a), the IAB-donor and two IAB-nodes are placed so that their locations form an equilateral triangle, while in the second deployment, see Fig. 6(b), the IAB-nodes are placed at different distances from the IAB-donor.

The average UE throughput for the max-min objective function and under the considered IAB-node deployments operating in single-hop (SH) and multi-hop (MH) modes is illustrated in Fig. 8. Recall that for the max-min formulation, all UEs achieve the same data rates in both the DL and the UL directions. The associated TTI allocations, ϵ_1 ϵ_2, as well as the access and backhaul time allocations at all nodes are shown in Fig. 7(a) and 7(b). As one may observe, for the scenario given in Fig. 6(a), there are no gains enabled by the multi-hop topology, since the average UE throughput is the same for both multi-hop and single-hop options. The associated time allocations, see Fig. 7(a), suggest that the configurations are backhaul-limited in both single-hop and multi-hop cases, as there always remain unused resources at IAB-nodes. In both setups, the single-hop topology serves as a baseline.

In the scenario shown in Fig. 6(b), however, the multi-hop topology allows for slightly higher data rates to be achieved, with a difference between single-hop and multi-hop cases being 4 Mbps, which is about 10% of the UE throughput in the single-hop scenario. In terms of time allocations, we observe in Fig. 7(b) a higher fraction of backhaul time at the IAB-donor being assigned to the parent IAB-node. The use of the multi-hop configuration also changes the TTI durations, ϵ_1 and ϵ_2, which implies that the gain is mainly

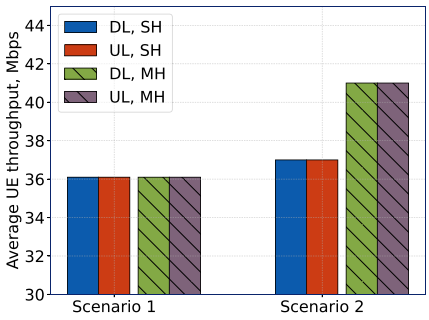


FIGURE 8. Average UE throughput for considered scenarios.

TABLE 3. Average UE throughput for different environmental settings.

Type	Fixed allocations [4], [47]	Proposed scheme
Suburban	33 Mbps	37 Mbps
Urban	35 Mbps	39 Mbps
Dense urban	38 Mbps	45 Mbps
Highrise urban	41 Mbps	55 Mbps

due to better spectral efficiency between IAB-nodes than between the IAB-donor and the IAB-node in the single-hop case.

Observe that the multi-hop configuration may cause additional performance degradation. At both the control and the data planes, this degradation results in higher latency, while the throughput gains remain marginal. In our experiments, we also note that this observation holds for scenarios with more than two IAB-nodes. Therefore, we can conclude that when designing an IAB system topology, **the use of a single-hop configuration should be preferred whenever possible**. This is beneficial unless the multi-hop topology is required due to, e.g., blockage by buildings at the IAB-donor–IAB-node interface or limited receive power at the IAB-node.

Furthermore, to assess the performance of the two-TTI system across diverse environmental settings and compare it with existing alternatives, we incorporate various IAB deployment areas as an additional parameter of interest. This entails the utilization of realistic radio deployment measurements from the ITU-R [46] to ensure accurate parameterization. Specifically, we categorize the deployment areas into four distinct types: suburban, urban, dense urban, and highrise urban. Our performance comparison results for equally divided and flexible radio resource allocation schemes are presented in Table 3, thereby facilitating a comparative assessment across different environmental settings. To evaluate the performance, we utilize the static 50/50 slot format for the two-TTI allocation scheme [4]. As compared to the fixed allocation scheme, where the division between UL and DL is static, our proposed framework optimizes the TTI durations as part of the overall procedure. It has been observed that our scheme demonstrates superior performance in terms of the average UE throughput.

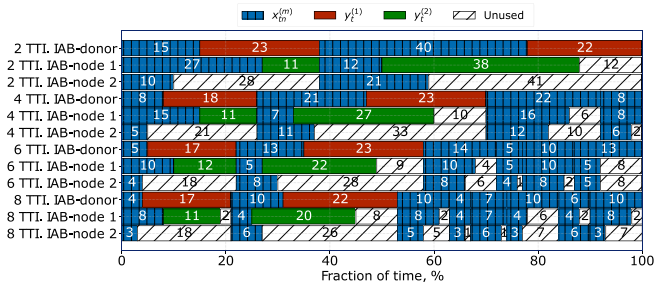


FIGURE 9. Time allocations for different numbers of TTIs.

TABLE 4. Average UE throughput, frame division, and solution time for different numbers of TTIs.

TTIs	UE throughput, Mbps	Frame division, %	Solution time
2	41.5	38, 62	Baseline
4	41.5	26, 44, 22, 8	+10.56%
6	41.5	22, 36, 14, 5, 10, 13	+13.32%
8	41.5	21, 32, 10, 4, 7, 10, 6, 10	+20.41%

B. NUMBER OF TTI ALLOCATIONS

We now proceed with assessing the effect of the number of TTIs into which a single frame can be divided. In general, a larger number of TTIs should provide additional flexibility in the allocation of access and backhaul resources. However, note that increasing the number of TTIs to be considered in the optimization problem leads to an exponential growth in the number of constraints, thus expanding the solution time and the associated packet scheduling complexity at the IAB-nodes. Therefore, in what follows, we find the number of TTIs that is sufficient to improve the operation of self-backhauled IAB systems. For this purpose, we consider the IAB network that comprises two IAB-nodes operating in multi-hop manner and compare different frame configurations featuring 2, 4, 6, and 8 TTIs for DL and UL transmission, while serving 30 UEs.

Table 4 reports the average UE throughput, frame division, and solution time for different numbers of TTIs, while Fig. 9 demonstrates the corresponding UE time allocations as well as the backhaul usage at IAB-donor and IAB-nodes. First, we notice that the **average end-to-end UE throughput does not increase with a larger number of TTIs**. It is important to note that as the number of TTIs grows, their usage pattern becomes more complex. Similar observations hold for larger numbers of TTIs and more complex topologies involving multiple chains of IAB-nodes and higher numbers of IAB-nodes in the chains. Hence, we may conclude that a two-TTI setup should be sufficient for the improved operation of the IAB system in terms of the UE throughput. We also note that this configuration results in a much shorter solution time. We additionally address the relative computational complexity by demonstrating that despite an increase in the number of equations, a solution remains feasible.

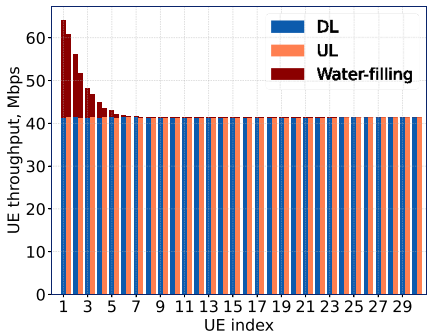


FIGURE 10. UE throughput with and without water-filling option.

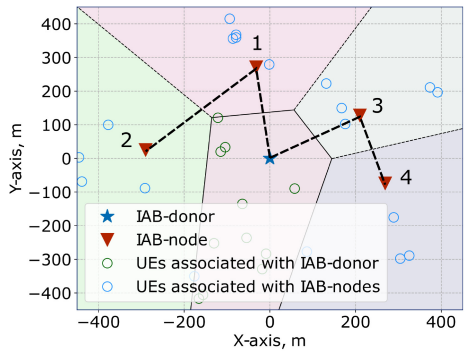
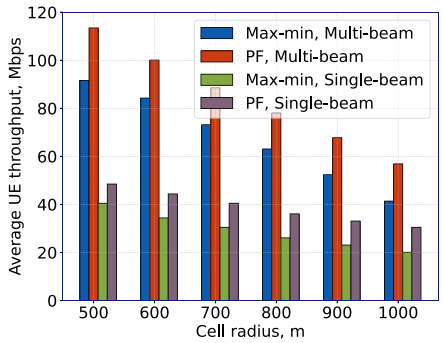
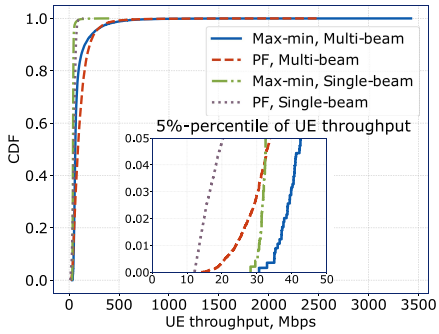


FIGURE 11. An illustration of two-chain IAB scenario.



(a) Average UE throughput as a function of cell radius



(b) CDF of UE throughput

FIGURE 12. UE throughput results.

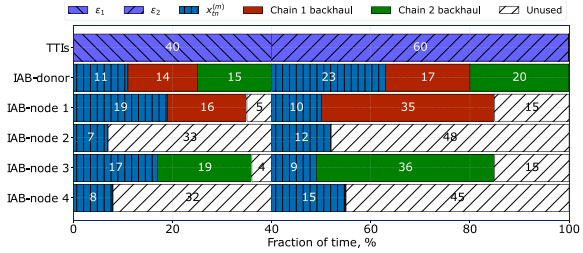
Finally, we highlight that the utilized max-min formulation allows for further improvements in the UE throughput by employing the water-filling algorithm [14], as illustrated in Fig. 10. Recall that if the backhaul data rate constraints are non-binding, the slack value associated with the constraints represents the difference between the actual value and the value that it would have if it were satisfied with an equality [14]. By using the water-filling algorithm, the available slack resources are allocated between the UEs to achieve capacity maximization. Specifically, since the max-min function seeks to maximize the minimum data rates, certain allocations can still be improved without violating the half-duplex constraint and thus affecting the allocations for other UEs. In the backhaul-limited operating regime, there are certain UEs that are directly served by the IAB-donor, whose allocations can be further improved on top of the max-min solution.

C. MULTI-BEAM OPERATION

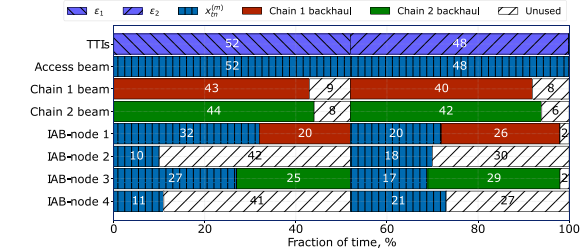
Previously, we observed that IAB systems with in-band backhauling are predominantly backhaul-limited, especially at the IAB-donor. Therefore, it is reasonable to evaluate the IAB system operation with multi-beam functionality employed at the IAB-donor and induced by the use of hybrid/digital beamforming [35]. We thus proceed with assessing this functionality for the scenario of two directly connected chains, each consisting of two IAB-nodes, as shown in Fig. 11.

A comparison of the average UE throughput for IAB-donor operating in single-beam and multi-beam regimes is presented in Fig. 12(a) for different cell radii. The corresponding access and backhaul allocations are shown in Fig. 13. Analyzing the data presented, we learn that the use of multi-beam operation effectively doubles the UE throughput. The reason for this is that despite splitting the power between two beams and thus reducing the spectral efficiency of the backhaul links, the overall backhaul throughput is still improved. This behavior is typical for throughput-constrained systems, see Fig. 13, as opposed to coverage-constrained deployments, where the backhaul link performance to the IAB-nodes may degrade and the gain from using multi-beam operation diminishes. Note that even under relatively large cell radii, the multi-beam system allows for higher average UE throughputs to be achieved, as confirmed by Fig. 12(a).

The above conclusions are also supported by the TTI allocations in Fig. 13. Here, for the single-beam system, all the backhaul resources are fully utilized, while the access resources at the IAB-nodes are underutilized. Operating with two beams allows us to efficiently overcome this limitation at the IAB-node. However, as one may notice, the bottleneck is then at the first IAB-nodes of the chains, even for a configuration with only one downstream IAB-node in each chain. This bottleneck, however, may not be eliminated by using more comprehensive antennas. To further address this challenge, dedicated UE association schemes and/or additional bandwidth on the backhaul links may be required.



(a) Single-beam time allocations



(b) Multi-beam time allocations

FIGURE 13. Time allocations for single- and multi-beam IAB-donor antennas.

D. IMPACT OF OBJECTIVE FUNCTIONS

It is known that the use of the max-min optimization criterion aims to ensure the highest possible fairness and may not efficiently utilize the system resources, even when the water-filling algorithm is applied [44]. However, as discussed in Section IV, our designed framework does not preclude the use of different objective functions, including the commonly employed PF criterion, which implicitly accounts for the channel states of the UEs [48]. Therefore, we now compare the application of the max-min and PF objective functions in terms of the UE throughput.

The average UE throughput as a function of cell radius, its cumulative distribution function (CDF), and 5%-percentile under 500 m cell radius are shown in Fig. 12 for max-min and PF allocations as well as single-beam and multi-beam IAB-donor operation. By analyzing the data presented in Fig. 12(a), it can be observed that the use of the PF criterion results in significantly higher average UE throughput for both the single-beam and the multi-beam cases. Specifically, in the multi-beam regime, an increase in the average UE throughput is approximately $15-22$ Mbps, while for the case of single-beam operation, the gain is $8-10$ Mbps. Note that the average gain of approximately 10% is visible for all of the cell radii considered. This growth is due to the fact that a logarithmically larger volume of resources is allocated to the UEs with more favorable channel conditions. Therefore, when applying the PF objective function, we trade fairness for the overall system throughput by utilizing the spectral resources more efficiently. The steeper nature of the max-min CDFs as compared to the PF CDFs notable in Fig. 12(b) further supports this observation.

VI. CONCLUSION

The recently proposed IAB technology for 5G NR systems represents a decisive paradigm shift in the way that future cellular systems are to be built. Operating over multi-hop topologies, IAB architecture induces the half-duplex constraint, which limits the choice of simultaneously feasible active links and complicates network performance optimization. In this study, we first proposed a framework for resource management in IAB systems with the half-duplex constraint and flexible TTI allocations, by taking into account the radio design and propagation characteristics of the mmWave band. We then utilized this framework to characterize the performance trade-offs associated with the system topology design, the use of comprehensive TTI allocation schemes, multi-beam operation, and objective functions.

The presented numerical results demonstrated that for the maximization of the average UE throughput, only the absolute UL and DL allocation parameters are essential, while the DL/UL pattern may alternate for a lower delay. Increasing the number of TTIs does not necessarily lead to improved UE throughput and significantly complicates the schedule that needs to be produced in real-time at the IAB-donor. Even though the latter is a linear programming task, the number of constraints grows exponentially.

Further, we showed that single-hop IAB configurations should be used whenever possible, unless a multi-hop topology is required, e.g., due to backhaul blockage or unsatisfactory SINR values. We also observed that in-band IAB systems tend to be backhaul-limited due to underutilized resources at the access interfaces of IAB-nodes. One way to considerably improve the UE throughput is to enable multi-beam operation at the IAB-donor. Finally, we established that the PF objective function allows for improved efficiency of resource utilization as compared to the max-min criterion, thus achieving higher average UE throughput at the expense of fairness.

REFERENCES

[1] H. Holma, A. Toskala, and T. Nakamura, *5G Technology: 3GPP New Radio*. Hoboken, NJ, USA: Wiley, 2020.

[2] (GSMA Co., London, U.K.). *The Economics of mmWave 5G. An Assessment of Total Cost of Ownership in the Period to 2025*. Accessed: Aug. 21, 2024. [Online]. Available: https://data.gsmaintelligence.com/research/research/research-2021/the-economics-of-mmwave-5g

[3] "Study on integrated access and backhaul; (Release 16)," 3GPP, Sophia Antipolis, France, Rep. TR 38.874, Jan. 2019.

[4] Y. Sadovaya et al., "Integrated access and backhaul in millimeter-wave cellular: Benefits and challenges," *IEEE Commun. Mag.*, vol. 60, no. 9, pp. 81–86, Sep. 2022.

[5] (Nokia, Espoo, Finland). *Integrated Access and Backhaul: Why it is Essential for mmWave Deployments*. Accessed, Aug. 21, 2024. [Online]. Available: https://www.nokia.com/blog/integrated-access-and-backhaul-why-it-is-essential-for-mmwave-deployments/

[6] (Qualcomm, San Diego, CA, USA). *What's in the Future of 5G Millimeter Wave*. Accessed: Aug. 21, 2024. [Online]. Available: https://www.qualcomm.com/news/onq/2021/01/whats-future-5g-millimeter-wave

[7] (IEEE ComSoc., New York, NY, USA). *AT&T and Verizon to Use Integrated Access and Backhaul for 2021 5G Networks*. Accessed: Aug. 21, 2024. [Online]. Available: https://techblog.comsoc.org/2019/12/16/att-and-verizon-to-use-integrated-access-and-backhaul-for-2021-5g-networks/

[8] (Light Reading Co., New York, NY, USA). *Verizon to Use 'Integrated Access Backhaul' for Fiber-Less 5G*. Accessed: Aug. 21, 2024.

[Online]. Available: https://www.lightreading.com/5g/verizon-to-use-integrated-access-backhaul-for-fiber-less-5g
[9] Y. Sadovaya et al., "Self-interference assessment and mitigation in 3GPP IAB deployments," in *Proc. IEEE Int. Conf. Commun.*, 2021, pp. 1–6.
[10] M. Dibaei and A. Ghaffari, "Full-duplex medium access control protocols in wireless networks: A survey," *Wireless Netw.*, vol. 26, no. 4, pp. 2825–2843, 2020.
[11] K. E. Kolodziej, B. T. Perry, and J. S. Herd, "In-band full-duplex technology: Techniques and systems survey," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 3025–3041, Jul. 2019.
[12] M. Gupta, A. Rao, E. Visotsky, A. Ghosh, and J. G. Andrews, "Learning link schedules in self-backhauled millimeter wave cellular networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8024–8038, Dec. 2020.
[13] M. Pagin, T. Zugno, M. Polese, and M. Zorzi, "Resource management for 5G NR integrated access and backhaul: A semi-centralized approach," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 753–767, Feb. 2022.
[14] M. Pióro and D. Medhi, *Routing, Flow, and Capacity Design in Communication and Computer Networks*. Amsterdam, The Netherlands: Elsevier, 2004.
[15] N. Yarkina, D. Moltchanov, and Y. Koucheryavy, "Counter waves link activation policy for latency control in in-band IAB systems," *IEEE Commun. Lett.*, vol. 27, no. 11, pp. 3108–3112, Nov. 2023.
[16] *Integrated Access and Backhaul (IAB) Radio Transmission and Reception, Release 18*, 3GPP Standard TS 38.174, Jan. 2024.
[17] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*. London, U.K.: Academic, 2020.
[18] M. Polese et al., "Integrated access and backhaul in 5G mmWave networks: Potential and challenges," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 62–68, Mar. 2020.
[19] V. F. Monteiro et al., "TDD frame design for interference handling in mobile IAB networks," in *Proc. IEEE Glob. Commun. Conf.*, 2022, pp. 5153–5158.
[20] C. Fang, C. Madapatha, B. Makki, and T. Svensson, "Joint scheduling and throughput maximization in self-backhauled millimeter wave cellular networks," in *Proc. 17th Int. Symp. Wireless Commun. Syst.*, 2021, pp. 1–6.
[21] P. Jayasinghe, A. Tölli, J. Kaleva, and M. Latva-aho, "Traffic aware beamformer design for integrated access and backhaul with flexible TDD," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2020, pp. 1–6.
[22] C. Huang, X. Wang, and X. Wang, "Effective-capacity-based resource allocation for end-to-end multi-connectivity in 5G IAB networks," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6302–6316, Aug. 2022.
[23] S. Zhang, X. Xu, M. Sun, X. Tao, and C. Liu, "Joint spectrum and power allocation in 5G integrated access and backhaul networks at mmWave band," in *Proc. 31st Annu. IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, 2020, pp. 1–7.
[24] Y. Liu, A. Tang, and X. Wang, "Joint incentive and resource allocation design for user provided network under 5G integrated access and backhaul networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 673–685, Apr.–Jun. 2020.
[25] M. N. Islam, S. Subramanian, and A. Sampath, "Integrated access backhaul in millimeter wave networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2017, pp. 1–6.
[26] M. N. Islam, N. Abedini, G. Hampel, S. Subramanian, and J. Li, "Investigation of performance in integrated access and backhaul networks," in *Proc. IEEE Conf. Comput. Commun. Workshops*, 2018, pp. 597–602.
[27] W. Lei, Y. Ye, and M. Xiao, "Deep reinforcement learning-based spectrum allocation in integrated access and backhaul networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 3, pp. 970–979, Sep. 2020.
[28] J. Y. Lai, W.-H. Wu, and Y. T. Su, "Resource allocation and node placement in multi-hop heterogeneous integrated-access-and-backhaul networks," *IEEE Access*, vol. 8, pp. 122937–122958, 2020.
[29] B. Zhang, F. Devoti, and I. Filippini, "RL-based resource allocation in mmWave 5G IAB networks," in *Proc. Mediterranean Commun. Comput. Netw. Conf.*, 2020, pp. 1–8.
[30] X. Zhang, F. Liu, and H. Xia, "Ergodic capacity analysis for full-duplex integrated access and backhaul system," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2019, pp. 1–5.
[31] D. Korpi, T. Riihonen, and M. Valkama, "Self-backhauling full-duplex access node with massive antenna arrays: Power allocation and achievable sum-rate," in *Proc. 24th Eur. Signal Process. Conf.*, 2016, pp. 1618–1622.
[32] M. Gupta, I. P. Roberts, and J. G. Andrews, "System-level analysis of full-duplex self-backhauled millimeter wave networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1130–1144, Feb. 2023.
[33] C. Huang and X. Wang, "A Bayesian approach to the design of backhauling topology for 5G IAB networks," *IEEE Trans. Mobile Comput.*, vol. 22, no. 4, pp. 1867–1879, Apr. 2023.
[34] M. Simsek, O. Orhan, M. Nassar, O. Elibol, and H. Nikopour, "IAB topology design: A graph embedding and deep reinforcement learning approach," *IEEE Commun. Lett.*, vol. 25, no. 2, pp. 489–493, Feb. 2021.
[35] F. Sohrabi and W. Yu, "Hybrid analog and digital beamforming for mmWave OFDM large-scale antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1432–1443, Jul. 2017.
[36] "Universal terrestrial radio access (UTRA) and evolved universal terrestrial radio access (E-UTRA); verification of radiated multi-antenna reception performance of user equipment (UE); (Release 17)," 3GPP, Sophia Antipolis, France, Rep. TR 37.977, Apr. 2022.
[37] "Study on scenarios and requirements for next generation access technologies; (Release 17)," 3GPP, Sophia Antipolis, France, Rep. TR 38.913, Apr. 2022.
[38] P. Nain, D. Towsley, B. Liu, and Z. Liu, "Properties of random direction models," in *Proc. 24th Annu. IEEE Joint Conf. Comput. Commun. Soc.*, 2005, pp. 1897–1907.
[39] M. Gapeyenko et al., "Analysis of human-body blockage in urban millimeter-wave cellular communications," in *Proc. IEEE Int. Conf. Commun.*, 2016, pp. 1–7.
[40] "Study on channel model for frequencies from 0.5 to 100 GHz; (Release 17)," 3GPP, Sophia Antipolis, France, Rep. TR 38.901, Jan. 2024.
[41] G. R. MacCartney, S. Deng, S. Sun, and T. S. Rappaport, "Millimeter-wave human blockage at 73 GHz with a simple double knife-edge diffraction model and extension for directional antennas," in *Proc. IEEE Veh. Technol. Conf.*, 2016, pp. 1–6.
[42] D. V. Andrade, M. G. Resende, and R. F. Werneck, "Fast local search for the maximum independent set problem," *J. Heuristics*, vol. 18, pp. 525–547, Feb. 2012.
[43] M. Gerasimenko et al., "Adaptive resource management strategy in practical multi-radio heterogeneous networks," *IEEE Access*, vol. 5, pp. 219–235, 2017.
[44] M. Gerasimenko et al., "Cooperative radio resource management in heterogeneous cloud radio access networks," *IEEE Access*, vol. 3, pp. 397–406, 2015.
[45] L. D. Beal, D. C. Hill, R. A. Martin, and J. D. Hedengren, "GEKKO optimization suite," *Processes*, vol. 6, no. 8, p. 106, 2018.
[46] *Propagation Data and Prediction Methods Required for the Design of Terrestrial Broadband Radio Access Systems Operating in a Frequency Range from 3 to 60 GHz*, ITU-Rec. P.1410, Int. Telecommun. Union, Geneva, Switzerland, Feb. 2012.
[47] C. Saha, M. Afshang, and H. S. Dhillon, "Bandwidth partitioning and downlink analysis in Millimeter wave integrated access and backhaul for 5G," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8195–8210, Dec. 2018.
[48] F. Capozzi, G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in LTE cellular networks: Key design issues and a survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 678–700, 2nd Quart., 2013.

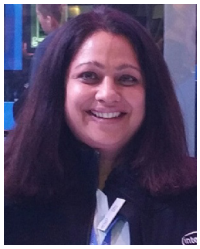
NIKITA TAFINTSEV (Graduate Student Member, IEEE) received the B.Sc. degree (Hons.) in radio engineering, electronics and telecommunication systems from the Peter The Great St. Petersburg Polytechnic University, St. Petersburg, Russia, in 2017 and the M.Sc. (Tech.) degree (Hons.) in information technology from Tampere University, Finland, in 2019. He is currently pursuing the Ph.D. degree with the Unit of Electrical Engineering, Tampere University. His research interests include performance evaluation and optimization methods for 5G/6G wireless networks, integrated access and backhaul systems, and applications of machine learning.

DMITRI MOLTCHANOV received the M.Sc. and Cand.Sc. degrees from the Saint Petersburg State University of Telecommunications, Russia, in 2000 and 2003, respectively, and the Ph.D. degree from the Tampere University of Technology in 2006. He is currently an University Lecturer with the Faculty of Information Technology and Communication Sciences, Tampere University, Finland. He has co-authored over 150 publications. His current research interests include 5G/5G+ systems, ultra-reliable low-latency service, industrial IoT applications, mission-critical V2V/V2X systems, and blockchain technologies.

WEI MAO received the B.E. and M.E. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2004 and 2007, respectively, and the Ph.D. degree in electrical engineering from the California Institute of Technology, in 2015. From 2015 to 2017 he was a Postdoctoral Scholar with the Department of Electrical Engineering, University of California at Los Angeles. In 2017 he joined as a Research Scientist with Intel Corporation, Santa Clara, CA, USA. His research interests include reinforcement learning, communications, signal processing, information theory, and coding.

HOSEIN NIKOPOUR is a Principal Engineer, an Applied AI Research Scientist, and the Manager at Intel Labs, with a focus on intelligent and resilient next generation wireless networks, hierarchical learning (H-learning), graph neural networks, integrated access and backhaul networks, high frequency bands, and network coding. Prior to his current role, he contributed to wireless solutions for 3GPP cellular standards during his tenure at Huawei and Nortel, Canada. He is widely recognized for his work on Sparse Code Multiple Access for massive machine-type communications. He holds a portfolio of over 150 Patents and has authored numerous publications, which have received more than 6400 citations.

SHU-PING YEH received the M.S. and Ph.D. degrees in electrical engineering from Stanford University in 2005 and 2010, respectively. She is an AI-Applied Principal Engineer with the Wireless System Research Lab, Intel, where she conducts research on wireless broadband technologies. She has over ten years of research and development experience in wireless industry and holds more than 30 U.S. Patents. Her current research focus includes open RAN architecture, AI/ML for RAN control and management, network slicing, and interworking of multiple radio access technologies within a network.



SHILPA TALWAR (Senior Member, IEEE) received the M.S. degree in electrical engineering and the Ph.D. degree in applied mathematics from Stanford University in 1996. She is an Intel Fellow and the Director of Wireless Systems Research, Intel Labs organization, Intel Corporation. She leads a research team focused on advancements in ultra-dense multi-radio network architectures and applications of machine learning and artificial intelligence techniques to wireless networks. While at Intel, she has contributed to IEEE/3GPP/ORAN standard bodies, including 4G/5G, and Open RAN Intelligence Control Specifications. She is a Co-Editor of book on 5G *Towards 5G: Applications, Requirements and Candidate Technologies*. Prior to Intel, she held several senior technical positions in wireless industry working on a wide range of projects, including algorithm design for 3G/4G & WLAN chips, satellite communications, and GPS. She is an author of 70+ technical publications and holds 65+ Patents.

MIKKO VALKAMA (Fellow, IEEE) received the M.Sc. (Tech.) and D.Sc. (Tech.) degrees (Hons.) in electrical engineering from the Tampere University of Technology, Tampere, Finland, in 2000 and 2001, respectively, where he is currently a Full Professor and the Unit Head of electrical engineering. His Ph.D. Dissertation was focused on advanced I/Q signal processing for wideband receivers: models and algorithms. In 2003, he was a Visiting Postdoctoral Research Fellow with the Communications Systems and Signal Processing Institute, San Diego State University, San Diego, CA, USA. His current research interests include radio communications, radio localization, and radio-based sensing, with particular emphasis on 5G and 6G mobile radio networks. He was a recipient of the Best Ph.D. Thesis Award of the Finnish Academy of Science and Letters for his Ph.D. Dissertation.

SERGEY ANDREEV (Senior Member, IEEE) received the Ph.D. degree from Tampere University of Technology, and the Cand.Sc., and Dr.Habil. degrees from SUAI. He is a Professor of Wireless Communications and an Academy Research Fellow with Tampere University, Finland. He is also a Research Specialist with Brno University of Technology, Brno, Czech Republic. He has been a Visiting Senior Research Fellow with King's College London, U.K. and a Visiting Postdoctoral Researcher with the University of California at Los Angeles, USA. He co-authored more than 300 published research works on intelligent IoT, mobile communications, and heterogeneous networking.