Università degli Studi di Padova

Department of Information Engineering

Ph.D. Course in Information Engineering

Science and Information Technology Curriculum

XXXVI Series

Resilient communication in mmWave and THz communication based on IAB and RIS

Supervisor

Prof. Michele Zorzi

Co-supervisor

Prof. Andrea Zanella

Ph.D. Candidate

Amir Ashtari Gargari

Academic Year 2023/2024

Words cannot express my gratitude to my wife for her invaluable patience and support. Her belief in me has kept my spirits and motivation high during this process. I also could not have undertaken my academic journey without the support of my parents, who kindly helped me in all parts. Additionally, this endeavor would not have been possible without the generous support from my supervisor, advisors, and all my colleagues. I am grateful to them for their editing help, late-night feedback sessions, and moral support.

Abstract

5th generation (5G) and 6th generation (6G) are significant wireless technology advancements that are expected to operate on millimeter wave (mmWave) and sub-Terahertz (THz) spectrum, respectively. While 5G significantly improved data speeds and connectivity, 6G is anticipated to go even further by enabling even higher frequencies and superior performance, making it a key enabler for meeting the requirements of future digital applications for higher data rates and bandwidth. In the context of 5G and future generations like 6G, resiliency becomes even more critical as these networks support a wide range of applications, including mission-critical services, IoT, and autonomous systems, where uninterrupted communication is essential.

In the first phase of the thesis, first we summarize State Of Art (SOA) channel modeling approaches in sub-THz. Then we introduce two novel channel modeling approaches -THz-GAN channel and Simplified Stochastic Channel Model in THz (SSCH-THz)- at 140 GHz, based on statistical and Machine learning based to cope with the usefulness of conventional channel modeling approaches in the THz spectrum and the absence of general channel modeling at this spectrum. We implement proposed channel models and benchmarks in ns-3 in order to validate channels and enable full-stack performance of applications like UDP and TCP based on KPI performance metrics. The results indicate that with 1000 fewer training steps and a smaller required dataset size, the THz-GAN channel and the benchmark channel models KPI are extremely similar, with a maximum CIR difference of 10 dBm. On the other hand, Integrated Access and Backhaul (IAB) and RIS are flexible and low-cost solutions to deploy ultra-dense fifth and sixth-generation (5G and 6G) systems, as they enable wireless backhaul links based on the same technology and specifications as for the access link. In the second phase, we introduce a new system-level simulator on top of sionna to support the IAB network and self-backhauling and SAFEHAUL: novel RL-based backhaul schedular. In addition, we proposed ns-3 implementation of RIS, then, we profile RIS performance by implementing the RIS signal model in ns-3 mmwave. Then, we consider the deployment of mixed mmWave and sub-terahertz links to increase the capacity of the backhaul network, and provide the first performance evaluation of the potential of sub-terahertz frequencies for 5G and 6G Integrated Access and Backhaul (IAB). Our results show that IAB with sub-terahertz links can outperform a mmWave-only deployment with improvements of 4× for average user throughput and a reduction of up to 50% for median latency.

Contents

Listing of figures

Listing of tables

Listing of acronyms

3GPP . . . . . 3rd Generation Partnership Project

5G . . . . . 5th generation

AoA . . . . . Angle of Arrival

AoD . . . . . Angle of Departure

BLER . . . . . Block Error Rate

BS . . . . . Base Stations

CB . . . . . Code Block

CDF . . . . . Cumulative Distribution Function

CQI . . . . . Channel Quality Indicator

CRT . . . . . Complete Report Table

CSI . . . . . Channel Side Information

DBF . . . . . Digital Beamforming

DC . . . . . Dual Connectivity

DL . . . . . Downlink

eNB . . . . . evolved Node Base

EPC . . . . . Evolved Packet Core

EPS . . . . . Evolved Packet System

FDD . . . . . Frequency Division Duplexing

FS . . . . . Fast Switching

HBF . . . . . Hybrid Beamforming

IA . . . . . Initial Access

LOS . . . . . Line of Sight

MAC . . . . . Medium Access Control

MCS . . . . . Modulation and Coding Scheme

MIMO . . . . . Multiple Input Multiple Output

MME . . . . . Mobility Management Entity

MMIB . . . . . Mean Mutual Information per coded Bit

MTU . . . . . Maximum Transfer Unit

NLOS . . . . . Non Line of Sight

OFDM . . . . . Orthogonal Frequency Division Multiplexing

OFDMA . . . . . Orthogonal Frequency-Division Multiple Access

PDCP . . . . . Packet Data Convergence Protocol

PDU . . . . . Packet Data Unit

PHY . . . . . Physical

QoE . . . . . Quality of Experience

RA . . . . . Random Access

RAN . . . . . Radio Access Network

RAT . . . . . Radio Access Technology

RLC . . . . . Radio Link Control

RMS . . . . . Root Mean-Squared

RRC . . . . . Radio Resource Control

SCG . . . . . Secondary Cell Group

SDN . . . . . Software Defined Networking

SDU . . . . . Service Data Unit

SI . . . . . System Information

SIB . . . . . System Information Block

SRB . . . . . Signalling Radio Bearer

TB . . . . . Transport Block

TDD . . . . . Time Division Duplexing

TDMA . . . . . Time Division Multiple Access

UE . . . . . User Equipment

UL . . . . . Uplink

ULA . . . . . Uniform Linear Array

DSRC . . . . . Dedicated Short-Range Communication

PRR . . . . . Packet Receipt Rate

UPA . . . . . Uniform Planar Array

API . . . . . Application Programming Interface

CDF . . . . . Cumulative Distribution Function

Inverse CDF . . . . Inverse Cumulative Distribution Function

MLE . . . . . Maximum Likelihood Estimation

p2p-PSNR . . . . . Point-to-Plane PSNR

LiDAR . . . . . Light Detection and Ranging

V2X . . . . . Vehicle-To-Everything

HSC . . . . . Hybrid Semantic Compression

SCAROS . . . . . Scalable and Robust Self-backhauling Solution

CVaR . . . . . Conditional Value at Risk

BAP . . . . . Backhaul Adaptation Protocol

RL . . . . . Reinforcement Learning

GAN . . . . . Generative Adversarial Network

cGAN . . . . . Conditional Generative Adversarial Network

MNO . . . . . Mobile Network Operator

IAB . . . . . Integrated Access and Backhaul

RIS . . . . . Reconfigurable Intelligent Surface

SSCH-THz . . . . . Simplified Stochastic Channel Model in THz

VR . . . . . Virtual Reality

ReLU . . . . . Rectified Linear Unit

KPI . . . . . Key Performance Indicator

KS . . . . . Kolmogorov–Smirnov

E2E . . . . . End-to-End

P2P . . . . . peer-to-peer

XR . . . . . eXtended Reality

HMD . . . . . Head Mounted Device

1 Introduction

Increasing demand for higher data rates, lower latency, and more connected devices is driving the evolution of wireless communication technologies from 4G to 5G and beyond. 5G and 6G, along with mmWave and sub-terahertz communication, play an essential part in fulfilling these requirements. 5G and 6G promise much higher data transfer rates than their predecessors. They enable 4K video transmission in real time, augmented and virtual reality, and IoT device support [3]. 5G reduces latency to milliseconds, making mission-critical applications such as autonomous vehicles and remote surgery possible. 6G seeks to reduce latency even further, opening the door to new applications. Both 5G and 6G are designed to connect a large number of devices concurrently, enabling smart cities, industrial automation, and the Internet of Things (IoT). 5G and 6G employ mmWave and sub-terahertz frequencies to attain these high data rates [4]. However, these frequencies have a limited range and are susceptible to interference from structures. Signals at mmWave and sub-terahertz frequencies are susceptible to higher path loss, necessitating the use of sophisticated beamforming and antenna technologies to maintain signal strength. Spectrum allocation for these technologies presents a significant challenge, particularly in congested frequency bands [5].

1.1 Motivation and Contribution

During my doctoral studies, we focused on both mmwave and sub-THz frequency challenges and attempted to solve the problem. The main contributions of my research are listed

as follows:

- First, we review channel modeling approaches for the spectrum above 100 GHz (Sec. 3). Moreover, it is a challenge to collect datasets with the scarce THz channel measurements. So, we generate a dataset based on two channel models (HB channel and FS) for indoor meeting room scenarios, which we used in developing channel models.

- we introduce Simplified Stochastic Channel Model in THz (SSCH-THz) , using simplified mathematic formulas to model the channel in the indoor scenario at 140 Gigahertz (GHz) (Sec. 4). By using this approach we reach highly accurate channel modeling with much lower complexity. However the issue with this is not a generality, so we introduce an alternative channel model to cope with this.

- For sub-THz communication, we introduce THz-GAN channel, a cGAN-based channel modeling approach to cope with not generality channel modeling issue (Sec. 4). The THz-GAN channel core is cGAN, which uses transfer learning to adapt to new scenarios. For the first time at this spectrum, thanks to generality and ease of use this channel model can be used as generalized channel modeling in all deployment scenarios.

- In addition, we proposed ns-3 Implementation of THz-GAN channel and SSCH-THz . Then profile the full stack system performance based on Key Performance Indicator (KPI) such as throughput, and latency using TCP, and User Datagram Protocol (UDP) applications in ns-3 simulator. we use proposed and benchmark channel modeling to investigate the effect of channel modeling in full-stack analysis. End-to-end performance measurement of the system is made possible by full stack analysis, which is essential for validating the performance of algorithms implemented at each layer of the system.

- We model the scheduling and path selection problem in IAB mmWave networks as a multi-agent multi-armed bandit problem (Sec. 6.3). We consider multiple fiber base stations, simultaneously supporting many self-backhauled mmWave base stations. In our model, the self-backhauled base stations independently decide the links to activate. The consensus among the base stations is reached via standard-defined procedures (Section 6.4.3).

- We present the first solution to provide reliable performance in IAB-enabled networks (Sec. 6.4). Specifically, we investigate the joint minimization of the average end-to-end latency and its expected tail loss. To this aim, we propose Safehaul, a learning approach that leverages the coherent risk measure CVaR[6]. CVaR measures the tail average of the end-to-end latency distribution that exceeds the maximum permitted latency, thus ensuring the network's reliability.

- We analytically bound the regret of Safehaul, i.e., we bound the loss of Safehaul compared to the case when the delays associated to all end-to-end paths between self-backhauled base stations and fiber base stations are known a priori. We show that, for the case when there are no conflicts between the decision of the self-backhauled base stations, the average regret of Safehaul tends to zero as the time increases. This regret bound characterizes the learning speed and proves that Safehaul converges to the optimal scheduling and path selection solution that jointly minimizes the average end-to-end latency and its expected tail loss.

- We provide a new means of simulating multi-hop IAB networks by extending NVIDA's GPU-accelerated simulator Sionna [1] (Sec. 6.5). Specifically, we add codebook-based analog beamforming capabilities for both uplink and downlink communications. In addition, we add internal Ray Tracer (RT) of Sionna in order to generate Channel Impulse Response (CIR). Further, we extend Sionna by implementing system-level components such as layer-2 schedulers and buffers and Backhaul Adaptation Protocol (BAP)-like routing across the IAB network. We believe our IAB extensions will be instrumental for the open-source evaluation of future research on self-backhauled mmWave networks.

- Exploiting the above simulator, we evaluate and benchmark Safehaul against two state-of-the-art algorithms [7, 8] based on deployment in two different locations (Manhattan and Padova). The results confirm that Safehaul is significantly more reliable than the considered benchmarks, as it exhibits much tighter variance in terms of both latency (up to 71.4%) and packet drop rate (at least 39.1%). Further, Safehaul achieves up to 43.2% lower average latency and 11.7% higher average throughput than the reference schemes.

- We develop a greedy path generation algorithm that automatically selects the frequency band of an IAB link (between 28 GHz and 140 GHz) and assigns routes so that each

IAB node can reach the IAB donor **(Sec. ??Chapter:I . The frequency selection aims at avoiding bottlenecks, i.e., the algorithm selects the band that provides the highest capacity when accounting for the congestion that may arise in the proximity of the IAB donor. In addition, we consider and compare different ratios of sub-terahertz and mmWave links, which can be mapped to licensing constraints for out-of-band backhaul, and two different bandwidths for the sub-terahertz links (10 GHz and 32 GHz), which consider exclusive licensing or sharing with other services, respectively [9].**

- **We model the IAB network in a custom-developed 3rd Generation Partnership Project (3GPP) Release 17 simulator based on the open-source tool Sionna [1], with 3GPP and state-of-the-art mmWave and sub-terahertz channel models, and realistic and detailed 3GPP-based physical and Medium Access Control (MAC) layers. Our results quantify for the first time the performance improvement that sub-terahertz links can introduce in IAB networks, which can push beyond the limits of the in-band mmWave backhaul and support more than 50 users with 120 Mbps streams and a single donor without congestion (compared to about 33 Mbps for in-band mmWaves).**

1.2 Thesis Structure

The remainder of this thesis is split into five major sections. The first described in Chapter 2, provides basic material and tools for the research. The second, comprising Chapter 3 and Chapter 4, describes the SOA and proposed sub-thz channel models. The third section starts with a short introduction of 5G networks (Chapter 5 and follows with a description of the research on mmWave frequency reliability using IAB (Chapter 6) and RIS (Chapter 7) to improve communication range and dependability. The fourth section (Chapter 8) illustrates the possibility of combining sub-THz and mmWave in self-bakhclung, bringing together two previous sections. The fifth section describes the Statistical Analysis and End-to-End Performance Evaluation of Traffic Models for Automotive Data within Chapter 9. Chapter 10 concludes the thesis and suggests directions for future research.

2 Background

In this chapter, we will provide a concise description of the materials utilized in our analysis of the subject matter. You can skip this chapter and refer back to it while studying the thesis.

2.1 Deep Neural Networks (DNN)

Deep learning is a subfield of artificial intelligence (AI) and machine learning that focuses on training artificial neural networks to execute tasks that ordinarily require human intelligence. These neural networks are composed of multiple layers that are interconnected, allowing them to learn and represent complex data patterns. A Deep Neural Network (DNN) is comprised of different layers; these layers, known as *hidden layers*, enable DNNs to model complex relationships within data. The *input layer* receives unprocessed data, such as images, text, or numeric values. The *hidden layers* transform input data utilizing weighted connections and nonlinear activation functions. With their many *hidden layers*, deep networks are able to detect intricate patterns. The *output Layer* is the last layer of a neural network that provides predictions or classifications [10]. Forward propagation is utilized by DNNs to make predictions. The data flows from the input layer to the hidden layers to the output layer.

During training, they use backpropagation to minimize prediction errors by adjusting the network's weights and biases. Training deep neural networks can be difficult due to vanishing and exploding gradient issues. DNNs have enabled advancements in numerous fields,

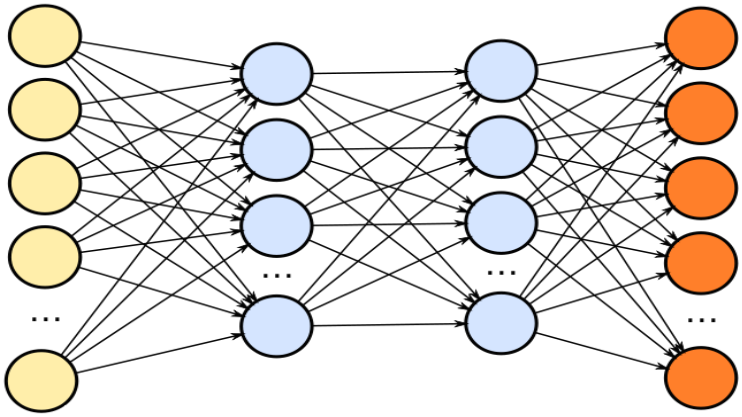


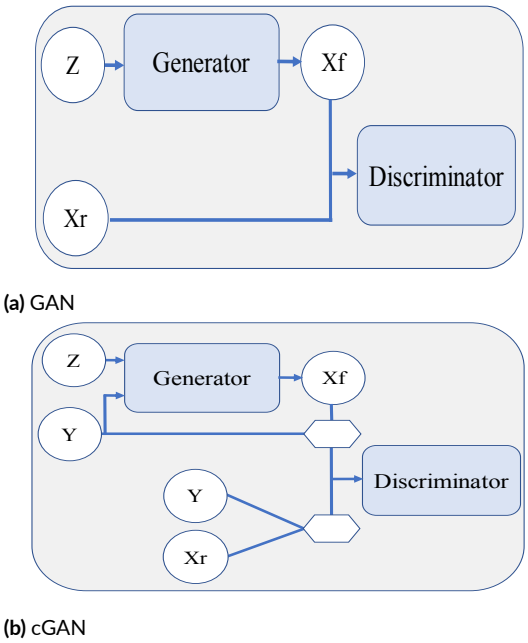
Figure 2.1: Architecture of DNNs including input layer (yellow circles), hidden layer(s) (blue circles), output layer (orange circles)

from healthcare to autonomous vehicles, by enabling computers to comprehend and process complex data, making them a cornerstone of contemporary artificial intelligence. Recent DNN applications in communication conducted by a couple of researchers will be discussed in the thesis. In my research, I focus on the GAN and RL models of DNNs, so I will provide a brief introduction to them in the following sections.

2.2 Generative Adversarial Network (GAN)

In machine learning, a GAN is a type of artificial neural network used to generate new data samples that resemble a given dataset. GANs have two major components: a generator and a discriminator. The generator attempts to generate data samples, such as images or text, that are similar to the real data it was trained on by using random noise as input. Over time, it learns to generate increasingly realistic data. On the other side, the discriminator's function is to differentiate between real data and data generated by the generator. It is trained to become more adept at distinguishing genuine from fake data. A GAN is trained through a competition between the generator and the discriminator. Here is how it operates:

Starting with random noise, the generator generates fictitious data. The discriminator assesses both real and fake examples. In an effort to fool the discriminator, the generator improves its ability to generate more convincing fake data over time. Simultaneously, the discriminator improves its ability to differentiate between real and fake data. This cycle is repeated until the generator produces data that is nearly indistinguishable from actual data. GANs have a variety of applications, including image generation, data augmentation, and



(a) GAN

(b) cGAN

Figure 2.2: (a)GAN architecture, (b) cGAN architecture where additional y controls the process of learning

style transfer. They have played a significant role in advancing the field of deep learning and have yielded impressive results in a variety of domains.

2.2.1 Conditional Generative Adversarial Network (GAN)

In machine learning, a cGAN is a type of generative model. It is an extension of the standard framework for GAN. cGANs augment the GAN architecture with conditions, such as class labels or other auxiliary information. This conditioning enables the generation of data samples that not only resemble the underlying distribution of the training data but also satisfy a set of conditions.

Same as GAN's, there are two neural networks within a cGAN: a generator and a discriminator. As inputs, the generator receives both random noise and conditioning information and produces samples designed to match the desired conditions. Under consideration of the conditions, the discriminator attempts to distinguish between real data and generated data.

There are numerous applications for cGANs, including image-to-image translation, text-to-image synthesis, and style transfer, where conditional information guides the generation process. In computer vision and natural language processing tasks, they have been utilized extensively to generate data that satisfies specific constraints or criteria. In wireless commu-

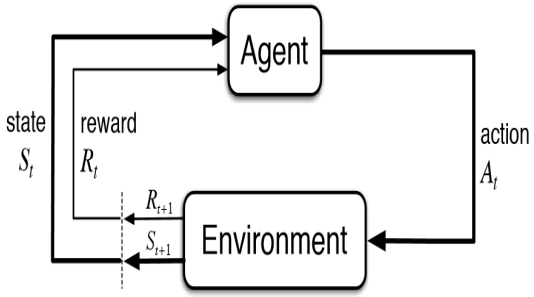


Figure 2.3: fundamental process of RL

nication, GANs have a variety of applications that enhance and improve various aspects of the field. Here are several prominent applications and trends: GANs can be used to generate fake wireless signals for security testing and training wireless communication systems to identify and defend against spoofed signals [11]. GANs are utilized to improve wireless channel recognition in aerospace communication scenarios. Neural networks, including GANs, can improve the precision of wireless communication channel recognition and adaptation [12]. In addition, GANs have the ability to learn high-dimensional and complex wireless environments, thereby assisting in the design and optimization of wireless communication systems [13].

2.3 Reinforcement learning (RL)

Reinforcement learning (RL) is a subset of machine learning that focuses on maximizing cumulative rewards through environment-based decision-making. Due to its capacity to optimize network performance, it has gained prominence in wireless communication. In this section, the main features of RL are described.

The fundamental Principles of Reinforcement Learning are as follows:

Agent: The decision-maker who interacts with the surroundings. The environment entails the external system with which an agent interacts. **State** (S): A snapshot of the environment at a specific time. **Action** (A): The options available to the agent. **Policy**: A strategy that determines the agent's action choice. **Reward** (R): A numeric value that represents an action's immediate benefit.

To optimize various aspects of network management and performance, RL has been increasingly utilized in the field of wireless communication. An actor-critic spectrum allocation scheme and a DNN were integrated in a study to achieve real-time spectrum alloca-

tion with DRL methods. This method utilized RL to dynamically allocate spectrum resources in wireless communication, thereby enhancing efficiency and flexibility [14]. This research [15] aims to increase network intelligence, optimize resource allocation, and enhance communication quality in wireless networks of the next generation. The use of RL in wireless communication is not limited to academia. Successful industry applications are emerging in which RL is utilized to optimize network management, improve quality of service, and boost the overall performance of wireless networks [16]. RL is viewed as an effective method for resolving complex decision-making issues within the context of 6G networks. It can address the enormous and intricate state and action spaces of these advanced wireless systems [17]. RL is beneficial for wireless communication because it enables systems to learn and build knowledge about radio channels without prior knowledge of channel conditions [18]. Extensive applications of RL to optimize the physical layer of wireless communications are available. This requires the definition of RL strategies to improve the fundamental aspects of wireless signal transmission [19].

2.4 Sionna: Empowering Link-Level Simulations in Wireless Communication

Sionna [1] is an open-source GPU-accelerated library for link-level simulations intended primarily for wireless communication systems. It provides researchers and engineers working on the physical layer of wireless networks with a valuable resource. In this section, I will explore the most important aspects of Sionna, its characteristics, and its role in link-level simulations:

Sionna is primarily designed for conducting simulations at the link level, which involve modeling and analyzing the physical layer of communication systems. Sionna's technological foundation is comprised of popular open-source software libraries, such as TensorFlow and Keras, and fully written in *python*. It uses these libraries to implement various components as Keras layers, thereby simplifying the process of constructing and customizing communication system models. Sionna is extremely versatile and can be used to simulate a variety of communication scenarios. It supports the implementation of point-to-point links and is compliant with 5G NR (New Radio) and 3GPP (3rd Generation Partnership Project) standards, making it suitable for both current and future wireless technologies [2]. Sionna is an open-source project, meaning that anyone is free to use, modify, and contribute to it. Its open nature encourages research community collaboration and innovation. Sionna incorpo-

rates deep learning techniques, specifically neural networks, into link-level simulations. In addition, Sionna provides tutorials and documentation to assist users in getting started with link-level simulations. Recent studies on Sionna include, among others, the following: The authors in [20] propose deep-unfolded interleaved detection and decoding (DUIDD), a new paradigm that reduces the complexity of IDD while achieving even lower error rates. Using NVIDIA's Sionna link-level simulator, they demonstrate the effectiveness of DUIDD in a 5G-near multi-user MIMO-OFDM wireless system with a novel low-complexity soft-input soft-output data detector, an optimized low-density parity-check decoder, and channel vectors from a commercial ray-tracer. In this paper [21], the authors demonstrate that the binary cross-entropy (BCE) loss is a sensible choice in uncoded systems, such as for training machine learning (ML)-assisted data detectors, but may not be optimal in coded systems. They propose a new loss functions aimed at minimizing the block error rate and SNR deweighting, a novel method for optimizing the performance of communication systems across a spectrum of signal-to-noise ratios. Through simulations in NVIDIA Sionna, the usefulness of the proposed loss functions and SNR deweighting are demonstrated.

In conclusion, Sionna is a powerful tool for wireless communication researchers and engineers. It combines the capabilities of deep learning with the adaptability of open-source software, making it a valuable asset for conducting advanced link-level simulations and driving communication technology innovations.

2.5 QD-Relization Ray Tracing Tool

The NIST Q-D Channel Realization Software* is a Matlab-based Three-Dimensional (3D) ray tracing tool. It offers a flexible, scalable, and realistic channel model based on measurement campaigns [22] to facilitate the design of next-generation wireless communication and sensing systems for mm-wave frequencies. The primary characteristics of the NIST Q-D Channel Realization Software are as follows:

- Support for a configurable scenario with multiple nodes.
- Accurate channel description thanks to geometrical ray tracing dependent on the input Computer-Aided Design (CAD) model of the environment and thanks to extensive measurement-based diffuse dispersion models.
- Facilitate node mobility.

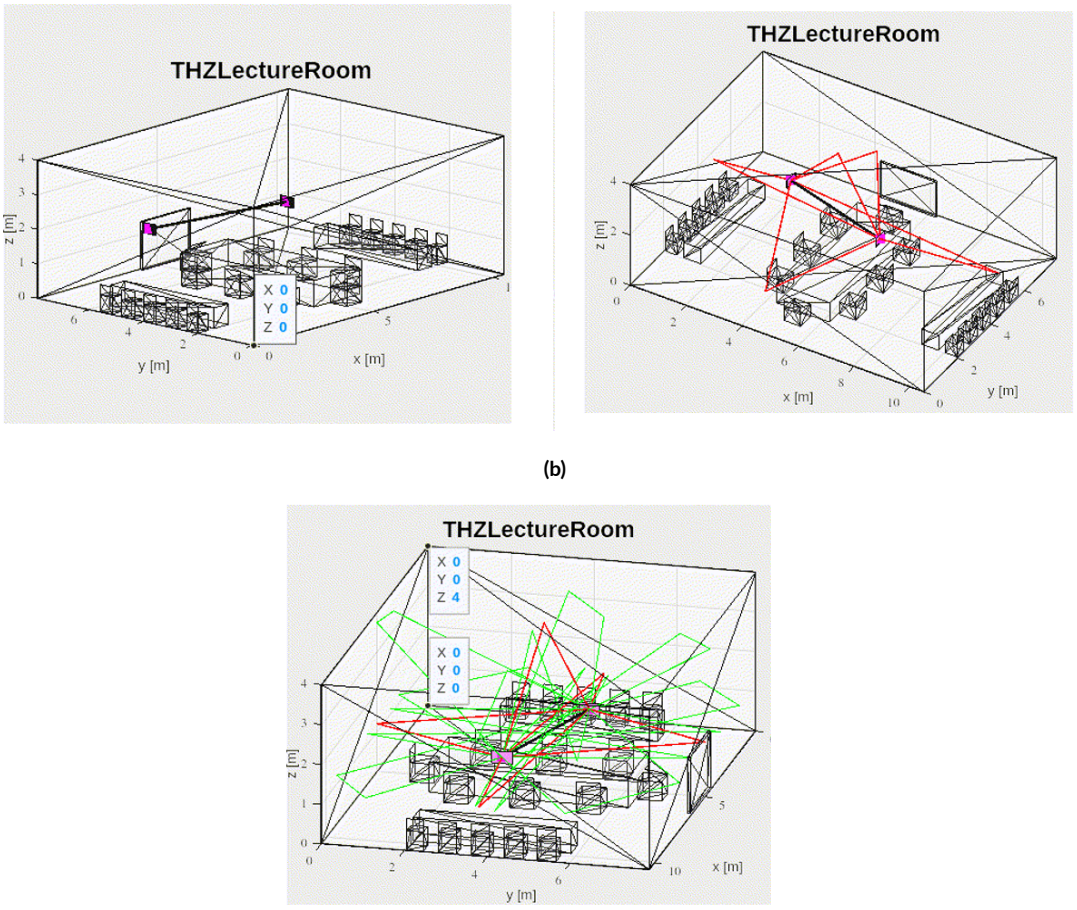
*https://github.com/signetlabdei/qd-realization

- Description of propagation in indoor and outdoor environments utilizing National Institute of Standards and Technology (NIST), 802.11ay Task Group (TGay), or custom material libraries Support for Multiple-Input Multiple-Output (MIMO) technologies: scaling the propagation description to nodes with multiple antenna arrays in order to circumvent the increased computational complexity.
- Configurable orientation of antennas.
- Configurable orientation at discrete points in time for each device.
- Support for dynamic targets that can be configured for sensing applications. Using geometrical ray tracing, ensure the spatial and temporal consistency of the target-related channel.

This tool's official version is designed to support only mmWave band frequency. Therefore, I incorporate the following steps to adopt the tool that supports 140 GHz for my research: Initially, I define new scenarios that are described in detail in Section 4.2. Then I define new material library files containing the 140 GHz coefficient of various materials [†] such as limestone, glass, and wood. In addition, the tool only supports reflection up to the second order, so I modified the equations to support reflection up to the third order. The updated version of the tool can be found at this link[‡].

[†] https://github.com/AmirAshtariG/qd-realization/tree/TeraSupport/src/material_libraries

[‡] https://github.com/AmirAshtariG/qd-realization/tree/TeraSupport



(a) (b)

(c)

Figure 2.4: Example of graphical output of QD-Realization tool in the THz scenario for a) one order of reflection, a) up to second orders of reflection c) up to third order of reflection

3

Sub-THz Communication

This chapter begins with a brief overview of the sub-THz band communication possibility and available tools. The IEEE 802.15.3d standard will then be described to illustrate the potential of this band for 6G communication. The necessity of system-level simulators for sub-THz 6G networks will then be discussed. By analyzing the ns-3 module of Terasim, I demonstrate how end-to-end performance analysis will be possible. This section will conclude by introducing SOA channel modeling in this spectrum as a benchmark for my work, which will be explained in the following chapter.

3.1 Introduction on sub-THz communication: potential and challenges

Globally connecting people, devices, and systems, wireless communication has become an essential component of contemporary life. Researchers and engineers are focusing on the Sub-THz frequency range as the demand for higher data rates, lower latency, and more efficient spectrum utilization increases. This relatively unexplored spectrum, ranging from approximately 100 GHz to 1 THz, offers the potential for revolutionary wireless network communication advancements. However, this spectrum's complexity necessitates an in-depth understanding of its characteristics and the development of innovative technologies. The available bandwidth at Sub-THz frequencies is significantly greater than at conventional microwave frequencies. This spectrum abundance enables the transmission of significantly larger data volumes, which could revolutionize applications like ultra-HD video streaming,

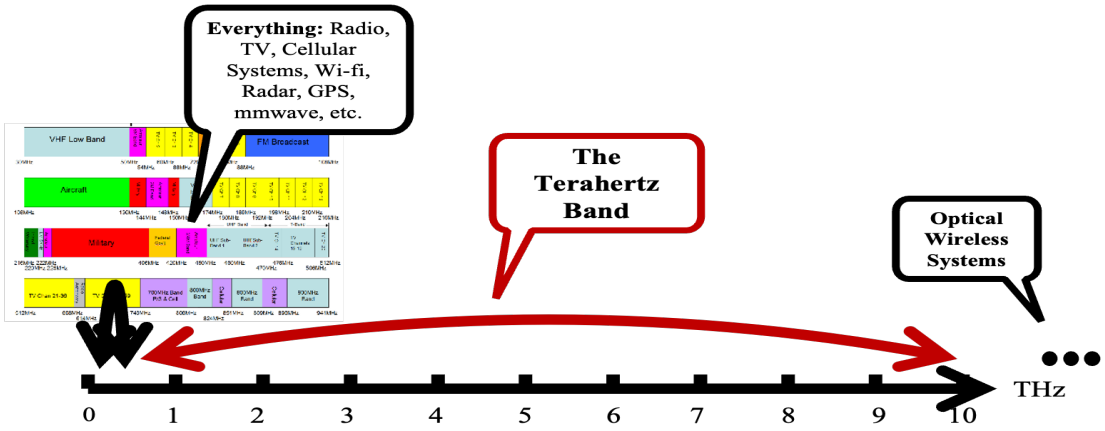


Figure 3.1: THz band versus mmWave and Sub-6G

augmented reality, and remote medical diagnostics. Due to its greater bandwidth and shorter wavelengths, the Sub-THz spectrum has the potential to support an enormous number of devices simultaneously. This enables 6G to reach new heights, with smart cities, industrial automation, and sensor networks benefiting from seamless connectivity and minimal interference. Sub-THz wavelengths enable the creation of highly focused beams, which in turn reduces propagation delays. This feature is essential for applications that require ultra-low latency, such as remote surgery, autonomous vehicles, and virtual reality environments. Simulations at the system level help optimize beamforming techniques and network architectures to achieve the desired latency levels. While the Sub-THz spectrum presents exciting opportunities, it also presents inherent propagation difficulties. These signals are extremely sensitive to atmospheric absorption, scattering, and diffraction at these frequencies. Simulations at the system level must accurately model these effects in order to predict signal coverage, link quality, and network performance as a whole. The proliferation of Sub-THz spectrum-operating wireless devices raises interference concerns. It is necessary to develop effective interference mitigation techniques, such as dynamic spectrum sharing and intelligent resource allocation. Simulations at the system level provide a controlled environment for testing these strategies prior to their deployment in the real world. Sub-THz communication-capable devices require specialized components and manufacturing processes. The initial cost of these devices may be prohibitive, and their integration with existing network infrastructures presents engineering challenges. Simulations aid in the evaluation of device complexity, performance, and cost-effectiveness.

3.1.1 IEEE 802.15.3d Standard

In this subsection, we describe IEEE 802.15.3d as an example of an IEEE standard for sub-THz communication in order to highlight the practical effectiveness and potential of this band. The IEEE 802.15.3d standard is part of the IEEE 802.15 family of wireless personal area networks (WPAN)-focused standards [23]. IEEE 802.15.3d focuses specifically on wireless communication in the sub-THz frequency band. Important aspects of IEEE 802.15.3d are as follows. This standard addresses wireless communication in the sub-THz frequency range, which has the potential for high data rates and is an emerging area for wireless communication. The IEEE 802.15.3d standard defines the physical layer (PHY) and medium access control (MAC) sublayer specifications. These specifications are essential for enabling sub-THz wireless connectivity with low data rates. IEEE 802.15.3d is an important step towards standardizing consumer wireless communications in the sub-THz frequency band, which has the potential to support a variety of applications, including high-speed data transmission. Researchers and engineers have been diligently developing IEEE 802.15.3d-compliant waveforms and technologies to exploit the benefits of the sub-THz band for wireless data transmission. IEEE 802.15.3d is an IEEE standard that defines PHY and MAC specifications for wireless communication in the sub-THz frequency band. As part of ongoing efforts to explore and standardize the use of this frequency band for high-speed wireless communication, which holds promise for a variety of applications, this frequency band is being investigated and standardized.

3.2 System-level simulations of sub-THz network

6G networks are still largely theoretical and are expected to be developed in the coming decades. Practical tools for this network are limited and particularly focused on physical layer deployments, so the development and testing of technologies for these advanced generations of wireless communication networks require simulators. System-level simulation in wireless communication engineering refers to the simulation and comprehensive evaluation of the performance of an entire wireless communication system, including all its components and interactions. Simulations are essential for designing, analyzing, and optimizing wireless networks and systems. Here are some important aspects of simulation at the system level in wireless communication engineering:

- **Comprehensive Network Modeling**: The simulation at the system level models the entire wireless communication network, including base stations, mobile devices, user

terminals, and network infrastructure. It takes into account the spatial distribution, characteristics, and interactions of these elements.

- **propagation model**: Realistic radio propagation models are used to simulate the propagation of electromagnetic signals through the environment, taking into account factors such as path loss, shadowing, fading, and interference. These models aid in the evaluation of signal quality and coverage.

- **Network Protocol**: The simulation incorporates the behavior of various network protocols, such as MAC (Medium Access Control) protocols, routing algorithms, and higher-layer communication protocols such as TCP/IP. It evaluates the effect that these protocols have on network performance.

- **Traffic Modeling**: The simulation includes models for simulating and creating different types of traffic. This helps evaluate how the network handles diverse data and application types.

- **Resource Allocation**: Simulations at the system level analyze strategies for resource allocation, such as frequency and time slot allocation, power control, and beamforming techniques. These strategies are crucial for optimizing the capacity and effectiveness of a network.

- **Interference Analysis**: In wireless communication, interference from neighboring cells or devices is a significant factor. Simulations at the system level assess interference effects and evaluate interference mitigation techniques.

- **Mobility and Handover**: Simulations of mobile wireless systems take user mobility patterns and handover mechanisms between base stations or cells into account. This is essential for evaluating the service quality during handovers.

- **Quality of Service (QoS) Evaluation**: Simulations at the system level evaluate network performance metrics such as throughput, latency, packet loss, and QoS metrics for various services and applications.

- **Optimization**: By running simulations with various parameters and configurations, engineers can optimize the network design, antenna placement, and resource allocation to achieve specific performance objectives.

- **Spectrum Management**: As the demand for wireless communication increases, spectrum management becomes increasingly important. Simulations at the system level can aid in evaluating the impact of spectrum-sharing techniques and regulatory policies.

- **Cost Analysis**: Simulations can reveal the cost-effectiveness of various network deployment strategies and technologies, enabling network operators to make informed investment decisions.

Overall, system-level simulation in wireless communication engineering is crucial for designing and optimizing wireless networks, ensuring they meet performance requirements and effectively utilizing available resources while taking into account the complexity of real-world scenarios. In this field, simulation tools such as NS-3, OPNET, MATLAB/Simulink, and custom-built software are frequently used to conduct these exhaustive analyses.

3.3 Network Simulator 3 (ns-3):Advancing Wireless Communication Research

ns-3 is a popular open-source network simulator that can be used for modeling and simulating wireless communication networks, including sub-THz channels. ns-3 provides a flexible and extensible platform for the development of wireless communication models and algorithms and can be used to study a wide range of network configurations and communication scenarios. It also provides a large number of tools and modules that can be used to study various aspects of wireless communication, including the physical layer, the MAC layer, and the network layer. ns-3 is a discrete-event network simulator which is a highly effective tool for communication and networking researchers to develop novel protocols and analyze complex systems. It is the successor to ns–2, which had been utilized by the networking community for over a decade to design and validate network protocols. The ns–3 source code can be downloaded from the project's website. ns–3 can now be used to simulate a wide variety of wireless and wireline networks, protocols, and algorithms, thanks to the addition of several modules by an active community of researchers from both industry and academia. There is comprehensive documentation on the models on the ns–3 website, regarding the both models' structure and what a user can do with them. In addition, the documentation provides a comprehensive tutorial on how to install ns–3, configure ns–3 scenarios and topologies, manage the collection of statistics, and log valuable messages. The tutorial is an excellent starting point for researchers who are new to ns–3.

Multiple folders constitute the ns–3 simulator. The src directory contains a collection of C++ classes that implement a variety of modular simulation models and network protocols. ns-3 is particularly useful for cross-layer design and analysis because the modules can be aggregated and instantiated to create diverse simulated network scenarios. The modularity and use of object-oriented design patterns also make it simple to incorporate and experiment with new algorithms in the network architecture. Each module is organized into numerous subfolders that contain the model's documentation and source code, as well as its helpers, examples, and tests. The helpers affiliated with every model play a crucial role. They are classes that conceal from the end user the complexity of setting up a complete scenario, for instance by automatically assigning IP addresses or by connecting the various classes of a protocol stack. The build subdirectory contains the simulator's binaries. Lastly, the scratch folder is a special folder where scripts containing examples and scenarios can be generated on the run.

- **Modularity and Extensibility**: ns-3 is renowned for its modular architecture, which enables researchers to simulate a vast array of network protocols, including those unique to wireless communication. Its extensibility permits the incorporation of new wireless models, mobility scenarios, and custom protocols in order to accurately simulate the behavior of real-world networks.

- **Realistic Channel Models**: ns-3 provides a variety of realistic channel models for a variety of wireless technologies, including Wi-Fi, LTE, mmWave, and sub-THz*, allowing researchers to analyze and optimize the performance of wireless communication systems in various environments and scenarios.

- **Protocol Development**: Using ns-3, researchers can design, test, and evaluate novel wireless communication protocols. This is especially useful in the context of 6G, where new protocols are anticipated to be developed to support ultra-low latency, massive connectivity, and high data rates.

- **Integration with External Tools**: ns-3 can be integrated with other tools and libraries, thereby enhancing its capabilities and allowing researchers to utilize additional features for their simulations. This integration is useful for modeling complex situations and incorporating actual data. As an example, the QD-realization † raytracer,

*Sub-THz channel models are limited and are developing

†https://github.com/signetlabdei/qd-realization

which is written in Matlab, has the capability of exporting rays to be easily integrated in ns-3.

- **Scalability**: From small-scale scenarios to large-scale network deployments, ns-3 can simulate networks of varying sizes. This scalability is essential for assessing network performance under varying conditions.

3.4 ns-3-TeraSim module

TeraSim is the first simulation platform for THz communication networks that supports THz devices' capabilities, from the protocol stack to the modeling of directional antenna patterns [24]. TeraSim considers two types of application scenarios, i.e., nano-scale scenarios, for short-range communications, and macro-scale, for traditional macro (e.g., cellular) scenarios. The module features common channel, antenna, and energy models for nano and macro applications, and separate MAC and Physical (PHY) models. A summary of the simulator's key features are listed below:

- **THz Band**: TeraSim is dedicated to simulating communication networks operating in the 0.1 to 10 THz frequency range.
- **THz Device Capabilities**: Captures the capabilities of THz devices, making it an all-inclusive platform for simulating THz-band communication.
- **Open source**: TeraSim is an open-source platform developed as an extension for ns-3 which is available at this link[‡].
- **Nanoscale and Macroscale Scenarios**: TeraSim is versatile, as it supports simulations for both nanoscale and macroscale scenarios in the THz communication domain.

The antenna model of TeraSim is based on the directional communication scheme introduced in [24], where alignment between mobile nodes is achieved with a rotating antenna. The model is implemented in the *THzDirectionalAntenna* class, which extends the ns-3 cosine antenna with the possibility to rotate. The *THzDirectionalAntenna* class calculates the antenna gain based on the mobility and the RX orientation. It is possible to configure different parameters for the antenna model, including whether the antenna is static or rotates,

[‡] https://apps.nsnam.org/app/thz/

the rotation speed and initial phase, the maximum gain and beamwidth. A beamwidth of 360 degrees models an omnidirectional antenna.

The already implemented channel model (based on [25]) accounts for the frequency selectivity based on different molecular absorption patterns at different frequencies of the terahertz spectrum. However, the current implementation of TeraSim only accounts for LoS propagation (more details Terasim can be found in [25]).

3.5 SOA CHANNEL MODELING ABOVE 100 GHZ

In the paragraphs that follow, we go over various channel models and modeling techniques for the sub-THz, mmWave, and sub-6 GHz spectrums. Additionally, we describe Terasim, a ns-3 module for simulating THz communication, at the end of the section.

3.5.1 Physics-based Channel Modeling

To assess the THz communication potential and to proceed with a realistic and accurate description of the actual capacity and channel model in the THz band, it is important to study a physics-based channel model [26]. The purpose of physics-based channel modeling is to describe the wireless communication devices in the electromagnetic (EM) spectrum. In [25], researchers focus on electromagnetic communications between nano devices and develop a physical channel model for wireless communication in the THz Band (0.1-10.0 THz). Physics-based channel models allow us to compute the signal path loss and the molecular absorption noise per each frequency and, finally, the channel capacity of EM nano networks [27]. The most remarkable example of this sort of channel model is [25], which takes the first step toward clearly defining the communication model for devices communicating wirelessly in the THz spectrum. This study offers a channel model for EM communications in the THz Band by addressing molecular absorption and providing formulations for total path loss and molecular noise. The CIR is represented as

$$h_C(f,d) = \left(\frac{c}{4\pi f d}\right) \exp\left(-\frac{k_{abs}(f)d}{2}\right), \tag{3.1}$$

where c, f, d, and k_{abs} refer to the speed of light, frequency, the distance between receiver and transmitter, and the molecular absorption coefficient of the medium, respectively. k_{abs} is determined by the transmission medium's molecular composition, i.e., the type and concentration of molecules present in the channel. However, physics-based channel modeling

approaches are limited to LoS scenarios and cannot describe the NLoS or fading effects, both of which are critical in the THz spectrum.

3.5.2 Stochastic Channel Modeling

Stochastic Channel Model (SCM) is a mathematical approach used in wireless communication to predict the behavior of a wireless channel based on probabilistic models that captures the statistical properties of the wireless channel. The SCMs are employed to assess the effectiveness of communication and localization systems by simulating a radio channel behavior. The SCM method relies on datasets derived from empirical channel measurements. This is less computationally complex than alternative channel modeling techniques, e.g., ray tracing. Several SCM models use fitted distributions to describe the temporal and spatial channel features, such as Angle of Arrival (AoA), Angle of Departure (AoD), Time of Arrival (ToA), and complex amplitudes. Clusters of multipath components can be described using the Saleh-Valenzuela (S-V) model with the tapped delay line [28]. [29] and [30] provide examples of such channel modeling applied to 140 GHz and 300 GHz signals, respectively. In comparison, other SCMs quantitatively explain the impulse response of channel and antenna parameters without explicitly including propagation. [31] reports on indoor wide-band propagation and penetration measurements for a typical building using wideband channel sounder equipment working at 140 GHz. The same group provides indoor measurements and models for reflection, scattering, transmission, and large-scale path loss at mmWave and sub-THz frequencies [32]. [33] represents the statistical characterization of three bands between 300 and 400 GHz which is based on a broad set of measurements in LoS and NLoS environments and includes spatial and temporal variations. [30] introduces stochastic indoor 300 GHz spatial-temporal channel model that considers parameters such as polarization, ray amplitudes, ToAs, AoAs and AoDs, and path specific frequency dispersion. Furthermore, THz channel modeling via a mixture of gamma distributions is proposed in [34]. A geometric-based stochastic time-varying model at 110 GHz is proposed in [35] for THz vehicle-to-infrastructure communications. The next paragraph is an example of a detailed explanation of SCM channel model operating at 140 GHz, which we will refer to it in upcoming sections.

Authors in [29] introduced a 3GPP-like indoor spatial SCM, which we will refer to it as The FS channel model in the rest of paper. This model is based on an experimental channel measurement campaign at 28 GHz and 140 GHz, in an office environment. The channel model is based on the concept of Time Cluster (TC) and Spatial Lobe (SL), which represent

the temporal and spatial statistics of the channel model, respectively. TCs comprise Multi Path Components (MPCs) propagating adjacent in time, which may come from different AoA (AoD), whereas SLs define the main direction of arrival (departure). The directional CIR is

$$h_{\mathrm{FS}}(\tau,\phi,\theta) = \sum_{n=1}^{N}\sum_{m=1}^{M_n} a_{n,m} e^{j\varphi_{n,m}} \delta(\tau-\tau_{n,m}) \\ g_{\mathrm{TX}}(\phi-\phi_{n,m}) g_{\mathrm{RX}}(\theta-\theta_{n,m}), \tag{3.2}$$

where a, τ, ϕ, and θ are the magnitude, absolute propagation delay, AoD vector, and AoA vector, respectively. Also, g_{TX} and g_{RX} are the TX and RX complex amplitude antenna patterns, respectively.

3.5.3 Ray tracing Channel Models

The stochastic nature of the SCM (3.5.2) channel models makes them generic, i.e., they can represent a general rural or urban scenario but without details. As a result, they do not precisely describe the THz signal's interactions in a particular deployment and hence are not very efficient to be utilized for thorough planning and capacity analysis in real-world scenarios [36]. RT can be used to accurately simulate the propagation of mmWave signals in a variety scenarios [37]. RTs is based on the scenario's geometry and characterizes the various propagation aspects of each MPC, such as ToA, Doppler shift, polarization, AoD at the TX, and AoA at the RX, offering higher accuracy of rays propagation in channel modeling [38]. Additionally, simulators can perform ray tracing to model the channel's temporal and spatial evolution, a significant point for appropriate wireless system planning. Recently, numerous RT modeling for THz band was done in order to evaluate the RT's suitability for THz waves. For example, [39] presents a ray tracing channel model at 300 GHz for close proximity THz communication, such as kiosk downloading. Besides that, [40] employs a ray tracing simulator with validated electromagnetic parameters for vehicle-to-infrastructure THz communication. Similarly, [41] describes how to characterize sub-THz channels operating at 90-200 GHz using deterministic simulations in indoor office and outdoor in-street scenarios.

3.5.4 Quasi-Deterministic Modeling

As described in 3.5.2 and 3.5.3, RT channel modeling enables high accuracy at the cost of time and resource, whereas, stochastic channel modeling is computationally lightweight but sacrifices accuracy. Thus, by strategically integrating two or more approaches, the development of Quasi-deterministic channel modeling methods for resolving complicated THz channel models with low latency is becoming interesting and promising [42]. Quasi-deterministic channel models properly represent the propagation of waves as described by the EM wave theory by modeling the propagation of radio waves in a given environment. This scheme is become interesting in THz communications due to its high time efficiency and accuracy compared to the individual methods. The hybrid method has been utilized in [43] for sub-THz communication channel modeling. In the following paragraphs, we will discuss a recent example of these channel modelings. The HB model [44] describes indoor sub-THz propagation by integrating RT and stochastic approaches. This channel model is based on wideband channel measurements obtained in a typical conference room operating at frequencies between 130 and 143 GHz. To investigate cluster behavior and wave propagation in the THz band, the HB model integrates MPC clustering and matching processes with RT methodologies [45]. The RT component enables the accurate representation of realistic scattering conditions in a given situation, whereas the stochastic component models random scatterers (e.g., wall texture or small objects) that may be difficult to capture correctly within the RT scenario [46]. The CIR as a function of the frequency f is given by

$$h_{HB}(\tau, \theta, f) = h_{RT}(\tau, \theta, f) + h_S(\tau, \theta, f), \tag{3.3}$$

where τ and θ indicate the delay and azimuth AoA, and $h_{RT}(\cdot)$ and $h_S(\cdot)$ are the channel components modeled through RT and stochastic methods. h_{RT} combines the Fresnel equations with the geometrical data to estimate the LoS and the reflection losses from the scattering of the walls (central sub-path in RT clusters). h_S, on the other hand, stochastically models the diffraction of extra sub-paths within the RT clusters, as well as the scattering of reflections from additional obstacles (non-RT clusters). The CIR generated by the RT, $h_{RT}(\tau, \theta, f)$, is represented as

$$\begin{aligned} h_{RT}(\tau, \theta, f) = {} & A_t(\phi_{LoS})\alpha_{LoS}(f)\delta(\tau - \tau_{LoS})\delta(\theta - \theta_{LoS}) \\ & + \sum_{l=1}^{L_{RT}} A_t t(\phi_{l,0})\alpha_{l,0}(f)\delta(\tau - \tau_{l,0})\delta(\theta - \theta_{l,0}), \end{aligned} \tag{3.4}$$

where the subscript $l, 0$ indicates the central path in the l^{th} sub-path, $A_t(\cdot)$ represents the antenna pattern at the TX, and $\alpha_{l,0}$, $\tau_{l,0}$, $\theta_{l,0}$ and $\phi_{l,0}$ represent amplitude gain, ToA, azimuth AoA, and AoD vectors of the sub-path, respectively. L_{RT} is the number of RT clusters. The CIR of the stochastic component, $h_S(\tau, \theta, f)$, is represented as

$$\begin{aligned} h_S(\tau,\theta,f) = &\sum_{l=1}^{L_{RT}} \sum_{\substack{p=-Q_l, \\ p\neq 0}}^{P_l} A_t(\phi_{l,p})\alpha_{l,p}(f)\delta(\tau-\tau_{l,p})\delta(\theta-\theta_{l,p}) \\ &+ \sum_{q=1}^{L_s} \sum_{s=-T_q}^{S_q} \alpha_{q,s}(f)\delta(\tau-\tau_{q,s})\delta(\theta-\theta_{q,s}), \end{aligned} \tag{3.5}$$

where subscripts l, p and s, q indicates the p^{th} sub-path in the l^{th} RT cluster, and the s^{th} sub-path in the q^{th} non-RT cluster, respectively. Notice that the antenna pattern of the TX is not involved in the statistical CIR in (3.5). The reason for this is that using a directional antenna at TX, all statistical parameters for the hybrid model are derived from channel measurements. As a result, the produced amplitudes contain the impact of the antenna pattern at TX.

3.5.5 Machine Learning Aided Channel Modeling

In recent years, the rapid advance of Artificial Intelligence (AI) has shown how DNN-based methods such as deep belief networks, deep convolutional neural networks, and recurrent neural networks, can be effective in a variety of research fields and demonstrated remarkable performance in wireless systems [47, 48]. Traditional channel modeling aforementioned above needs extensive domain understanding and technical skill in radio signal propagation over electromagnetic fields. To overcome the constraints associated with traditional channel modeling methodologies, neural networks, and machine learning techniques are being investigated as possible universal solutions for a variety of 6G application and communication environments [49, 50, 51]. Incorporating deep learning techniques into the channel modeling could solve high-complexity knowledge requirements in higher frequency propagation, particularly sub THz communication. As a result, it motivates researchers in the wireless communication field communication to describe wireless propagation using a DNN [52, 53]. Although machine learning might open up new possibilities for wireless channel modeling, its usefulness in practical applications may be limited by the need to gather large amounts of data, tune hyperparameters, and undergo a lengthy training process. DNNs are usually used for supervised learning tasks where the desired output is known, while GANs

are used for unsupervised learning tasks where the goal is to generate new data that is similar to existing data and more suitable choice for channel modeling. [54] is a recent illustration of the use of transfer learning in THz channel modeling. In this research, authors use GAN as the core of modeling to capture the distributions. However GAN are well-known to generate data, they cannot control and specify channel conditions in producing channel response which results in a lack of generality. On the other hand, to resolve issue of high dataset required for machine learning approaches transfer learning concept has been introduced, which try to transfer the acquired information from the source domain (subject) to the target domain (object) in order to reduce the performance loss of the models in the target domain. This strategy involves applying the knowledge acquired from one problem to a distinct but related problem. These methods should address the fundamental transfer issues of What, How, and From Where. When prior experiences (from other sources of knowledge) are insufficiently relevant to the new domain, a brute-force transfer may degrade learning performance, resulting in negative knowledge transfer phenomena. [55] is one of the examples of mitigating transfer learning in wireless communication. They address the problem of joint user-cell association and the selection of a number of beams to maximize the aggregate network capacity. They show that in stationary scenarios, the transfer learning-based approach achieves about 29% convergence speedup, and under mobility, scenarios improves 12% traffic load. In [56] authors proposed low resolution Multiple Input, Multiple Output (MIMO) channel estimation in mmWave which takes advantage of transfer learning to speed up new environment adaptation. [57] utilize DNN based transfer learning in channel modeling operating at mmWave using RT dataset that they reach acceptable accuracy.

4

Proposed channel modeling in Sub-THz Communication

It is anticipated that 6G mobile networks will switch to carrier frequencies in the spectrum above 100 GHz to fulfill the demand for higher data rates and bandwidth. The effective utilization of THz frequencies requires a thorough understanding of wireless channel characteristics. An accurate performance evaluation that highly relies on channel models provides essential input for defining and developing wireless network components above 100 GHz. The harsh propagation environment, as well as the requirement for directional communications and signal processing at high data rates, present challenges for the development of networking solutions at such high frequencies and in particular leads traditional channel modeling approaches to be less effective in this spectrum. In this chapter, we introduce THz-GAN channel, an ML-based channel model in order to deal with the usefulness of conventional channel modeling approaches in the THz spectrum and the absence of general channel modeling approaches. We validate the generality of THz-GAN channel by utilizing transfer learning and test approach in two indoor scenarios. In addition, we introduce the SSCH-THz which is based on fitted distributions and works based on simplified mathematics formulas. Also, we propose channel model implementation in ns-3 in order to validate by KPI performance metrics. We compare the full-stack performance of UDP and TCP using the proposed channel models versus benchmarks in different scenarios. The results indicate that with 1000× lower training steps and dataset size, THz-GAN channel and the

benchmark channel models generate similarly with a maximum 10 dBm mismatch. In addition, UDP and TCP application KPIs in the ns-3 of different channel models follow very closely.

4.1 Introduction

The development of 6G solutions has become the industry and research community's primary focus as 5G is now being commercially deployed [58]. As wireless communication systems progress toward 6G, there is a growing need to address the exponentially increasing demand for faster data rates, ultra-low latency, and massive connectivity In order to satisfy the demand for higher bandwidth, it is anticipated that 6G networks will support carrier frequency in spectrum above 100 GHz, in the lower part of the THz band. Exploring the THz frequency band has shown promise, as it offers the potential for order-of-magnitude higher data rates and ultra-low latency wireless communication systems which could be also used for high-capacity backhaul and sensing [59, 60]. For instance, the IEEE 802.15.3d standard (described in Sec. 3.1.1) provides up to 69 GHz bandwidth in sub-THz spectrum to support cutting-edge applications like Virtual Reality (VR) and telepresence using 3D holograms [23]. Notably, the spectrum that is available above 100 GHz is anticipated to enable terabit-per-second links and to make it easier to integrate new communications and sensing paradigms.

Developing new approaches, algorithms, and protocols in the spectrum above 100 GHz poses challenges since it requires tools that are extremely accurate and dependable. Due to the requirement for specialized Radio Frequency (RF) components, wireless testbeds for these frequency bands are currently very expensive. This limits them to a small number of nodes [61], and mostly physical layer focused evaluation. So, simulations can serve as a platform for end-to-end, full-stack performance evaluation. However, as discussed in [38], the accuracy of simulators is highly dependent on the channel model accuracy and its interaction with higher layers and protocols.

Channel modeling is a fundamental task of theoretical studies and practical implementations of wireless communication systems. The effective utilization of THz frequencies requires a thorough understanding of wireless channel characteristics and the development of accurate channel models. Accurate channel models enable us to comprehend the precise physical effects of wireless channels on transmitted radio signals, which is critical for designing and deploying effective and feasible communication technologies in real-world applica-

tion environments. Modeling THz channel in terms of propagation and fading has become critical for advancing investigation at the physical layer and higher [62]. The current channel models for lower frequencies lack key modeling elements for the THz spectrum; in particular, significant propagation loss and frequency-dependent molecule absorption complicate traditional channel modeling. Thus, researchers conducted several measurement campaigns and used different channel modeling approaches to represent the THz spectrum's characteristics in a variety of scenarios and environments, resulting in different channel model formulations with distinct levels of complexity and precision. The traditional channel modelings at the sub-THz is ineffective because it requires extensive channel modeling knowledge, a dependence on scenarios, a lack of a generic model, and an excess of modeling time.

All of these complexities affect the channel modeling development at higher frequencies, resulting in the lack of a comprehensive and precise channel model in sub-THz communication, which highlighted the requirement for novel channel modeling approaches in this spectrum. The GANs have the advantage of learning and modeling complex datasets without making any statistical assumptions, which makes them superior to other types of DNN methods. The problem of generality and accuracy of channel modeling can be resolved by using cGANs, which enables the flexibility to define various scenarios and combine them with transfer learning to quicken the learning process, by reducing the required datasets in learning and initiation of the model based on previously learned channels. By placing the burden of channel modeling complexity on the algorithm, this approach makes generality in channel modeling possible at higher frequencies.

4.2 Dataset generation

In this section, we describe the data generation process, which we need to model channels at 140 GHz. First, we describe deployment scenarios, then the process of data gathering in ns-3 is discussed. A well-curated dataset is required to validate the models and algorithms used to simulate the sub-THz wireless channel. This ensures that the models accurately reflect the behavior of the real-world channel, which is essential for the development of dependable wireless communication systems in the real-world. Building a real-world test bed for sub-THz communication is very expensive and time-consuming, especially it requires the construction of specialized test facilities. In contrast, simulators provide a more cost-effective alternative and allowing to quickly generate large amounts of data at a lower cost. However, the accuracy of the datasets generated by simulators can be affected by assumptions made in

the model, and there is always a trade-off between model complexity and computation time. RT is a great tool for generating accurate dataset of propagation of THz signals. RT allows for the creation of a controlled and reproducible environment for channel modeling. In addition, RT provides a great deal of flexibility in terms of studied scenarios. Ease of varying parameters in the simulators, such as the distance between the transmitter and receiver, the number of obstacles, and the type of environment, in order to study the impact of these factors on the sub-THz channel is required. For example, it is possible to study the impact of specific channel impairments, such as multipath fading or atmospheric absorption, on the performance of sub-THz communication systems. In the following subsections, we describe the simulation scenarios and the full process of data generation.

4.2.1 Deployment Scenarios

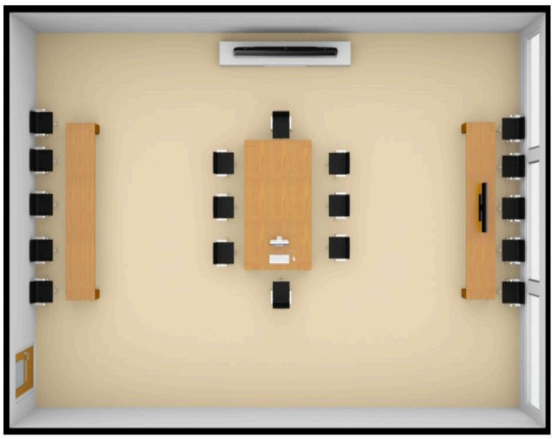
We define three distinct indoor scenarios: the Lecture room (*SC1*), conference room (*SC2*), and L-shape (*SC3*) room that are shown in Fig 4.1. In order to be able to examine different channel modeling performances we consider different geometry and furniture in each scenario. We consider *SC1*, which has the same size as that used for channel measurements in [44], as the baseline scenario. For all scenarios, we consider a mobility pattern that covers both NLoS and LoS conditions. In the following sub-sections geometrical details of deployment scenarios are provided.

SC1

In order to be able to use the proposed channel model in [44], the lecture room's floor plan is closely aligned with the proposed channel model, with the dimension of 7.9 m $\times$ 10.5 m. Three wooden tables and 18 wooden chairs are in the room. North of the floor plan is where the television is located. The material of the walls, including the ceiling and roof walls are limestones, whereas the right wall is made of glass.

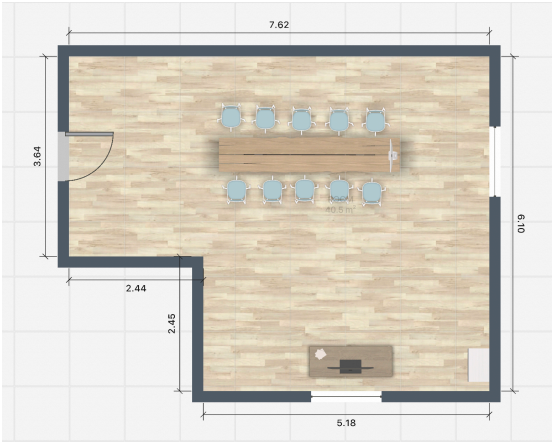
SC2

The dimensions of the Conference Room are 5 m $\times$ 8 m. The room is equipped with eight wooden chairs and two wooden tables in the center of the room, as well as four additional wooden chairs on the room's right. The presentation area is located to the left of the floor plan and contains one wooden chair, a wooden table, and a television. Notice that the material of all walls is limestone.



(a) Lecture room (*SC1*)

(b) conference room (*SC2*)



(c) L-shape (*SC3*)

Figure 4.1: Top view of the floor plan of the deployment scenarios including furniture

SC3

The dimensions of the L-shape room are 5.18 m × 6.10 m plus 2.44 m× 3.64 m The room is equipped with ten wooden chairs and a wooden table in the center of the room and a television set in the south of the floor plan. Notice that the material of all walls is limestone.

4.2.2 Dataset generation through ns-3

To collect data we used two different channel model representations, the HB channel model [44] is a quasi-deterministic channel model that contains ray tracing and statistics modeling, and the FS channel model [29], which is a fully stochastic channel modeling approach. We choose the portion of HB and FS in the total dataset is 70% and 30% respectively due to

reason of higher specific detailed consideration in RT. We used two channels to provide a more complete and accurate representation of the wireless THz channel due to the lack of general approaches to channel modeling at 140 GHz frequency. This approach helps to account for the different propagation characteristics of the direct and scattered signals, and provides insights into the behavior of the channel in different environments. Ray tracing is a more accurate and flexible method for generating datasets for sub-THz communication compared to statistical models. We used the following steps to gather the channel dataset for each scenario from ns-3.

- Set up the simulation scenario: Define the simulation scenario, including the network topology, the communication parameters, and the channel conditions.
- Run the simulation multiple times with different parameters or channel conditions to generate a large dataset.
- Collect and store the channel data.

We described the ns-3 implementation of HB and FS channel models here [63].

4.3 Stochastic channel model

In this section, we introduce the SSCH-THz, as an example of a novel simplified stochastic channel modeling concept for the sub-terahertz spectrum. The purpose of developing SSCH-THz is to keep the channel model accurate while reducing its complexity compared to traditional channel modeling approaches. The SSCH-THz model captures the statistical properties of the channel such as fading, scattering, and attenuation, and represents them by utilizing random distributions. The SSCH-THz channel model produces the channel response using a statistical method. The SSCH-THz approach works in all frequency bands and in particular we utilize it to represent the channel model for the Lecture Room (Fig. 4.1a) at 140 GHz. In the first step, as described in Section 4.2, we utilize the ns-3 implementations [63] of two-channel models [29, 44] to generate the dataset for the Lecture room scenario (Fig. 4.1a). Then we fit the received power based on three parameters: distance from TX to RX (d), RX antenna beamwidth (β), and channel condition (δ) -LoS/NLoS-. In order to fit we examine different random distributions and choose the distribution with minimum distance to dataset. Finally, we present a fully stochastic channel model based on the statistical distribution that could be utilized efficiently in industry and research. Notably, the SSCH-THz channel model supports 180° RX β and d up to 15 m.

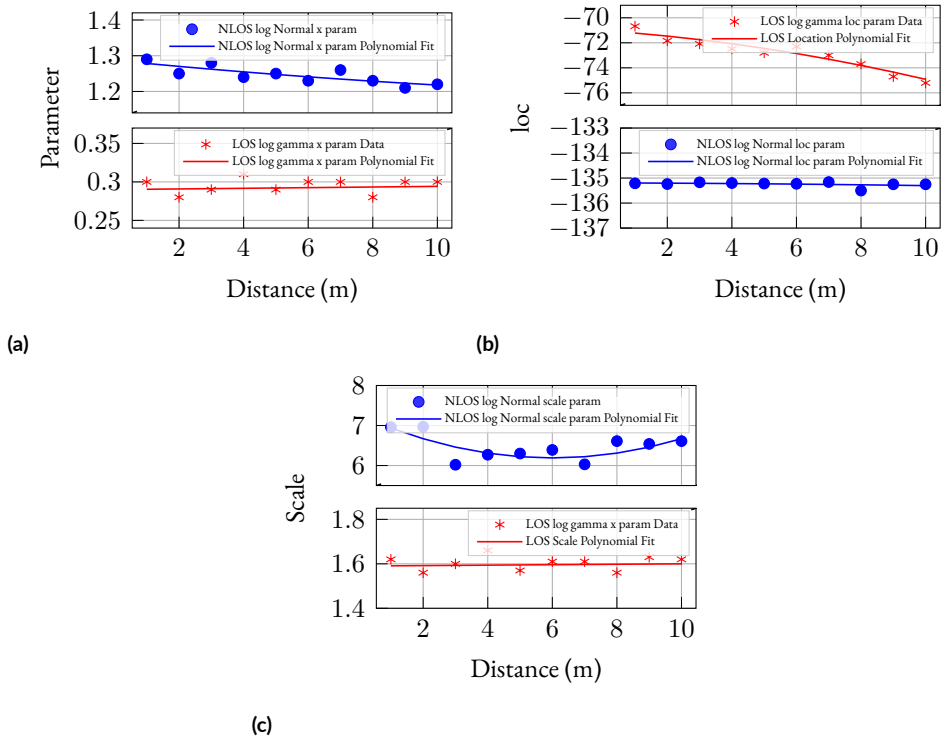


Figure 4.2: Different distribution parameter learning, (a) x param fitting for LoS and NLoS of SSCH-THz, (b) loc param fitting for LoS and NLoS of SSCH-THz, (c) $scale$ param fitting for LoS and NLoS of SSCH-THz.

4.3.1 Formulation

In this subsection, we elaborate on the distribution fitting of statistical channel modeling. To do the fitting, we compare the empirical Cumulative Distribution Function (CDF) of the collected dataset derived from the dataset, CDF_{data}, with different random distributions by applying Kolmogorov-Smirnov (KS)test. The KS test is an important performance metric for fitting distributions because it provides a way to compare the goodness-of-fit of a sample distribution to a theoretical distribution. The KS test measures the maximum distance between the Empirical Distribution Function (EDF) of the sample and the CDF of the theoretical distribution being tested. By using the KS test, we derive the optimal parameter values for minimizing the error between CDF_{data} and random distributions and quantify the divergence of the random distributions from the CDF_{data}. In the proposed method, the LoS and NLoS components are expected to fluctuate differently and follow different distributions, as is common in the real world. So, we proposed separate formulas to quantify CIR of the LoS and NLoS. The SSCH-THz channel CIR is denoted by the formula (4.1) where

log-gamma and log-normal distributions are fitted to LoS and NLoS channel models, respectively. In Fig. 4.2a the values per different distances of distribution and the fitted lines are presented. In the proposed channel model the value of δ, d, and β are required and sufficient to generate the channel response.

$$H(\delta, d, \beta) = \begin{cases} Log - Gamma(x_{los}, l_{los}, s_{los}) & \delta = los \\ Log - Normal(x_{nlos}, l_{nlos}, s_{nlos}) & \delta = nlos \end{cases} \tag{4.1}$$

Where x, l, and s indicate the shape, location, and scale of distributions, respectively, which are described below. All parameters of the channel model in the following equations are summarized in Table 4.1. In the LoS scenario, x_{los} and s_{los} are fitted with polynomials of degree 2, which are represented by Equations(4.2) and (4.3) respectively. For l_{los}, the equation shown in Eq. (4.4), there are two case statements: if β is greater than 30 degrees, the distribution is fitted by a polynomial of degree 2, otherwise by a polynomial of degree 3. the reason for using a higher order model for the field of 30 is that for β greater than 30 the majority of MPCs are received, whereas the number of received MPCs is restricted when the β is less than 30.

$$x_{los}(d) = a_1 * d + a_2 * d^2 + a_3, \tag{4.2}$$

$$s_{los}(d) = b_1 * d + b_2 * d^2 + b_3, \tag{4.3}$$

$$l_{los}(d, \beta) = \begin{cases} c_1 * d + c_2 * d^2 + c_3, & \beta \geq 30 \\ d_1 * d + d_2 * \beta + d_3 * d^2 + d_4 * d * \beta & \\ +d_5 * \beta^2 + d_6 * d^3 + d_7 * d^2 * \beta & \\ +d_8 * d * \beta^2 + d_9 * \beta^3 + d_{10}, & otherwise \end{cases} \tag{4.4}$$

In contrast, l_{nlos} and s_{nlos} are fitted with polynomials of degree 2, which are represented by Equations (4.5) and (4.6), respectively, in the NLoS scenario. Equation (4.7) contains two case statements for x_{nlos}: we fitted by a polynomial of degree 2 if β is greater than 30 degrees, otherwise we formulated a polynomial of degree 3 for β less than 30 degrees.

$$l_{nlos}(d) = e_1 * d + e_2 * d^2 + e_3, \tag{4.5}$$

$$s_{nlos}(d) = f_1 * d + f_2 * d^2 + f_3, \tag{4.6}$$

$$x_{nlos}(d,\beta) = \begin{cases} g_1 * d + g_2 * d^2 + g_3, & \beta \geq 30 \\ h_1 * d + h_2 * \beta + h_3 * d^2 + h_4 * d * \beta & \\ +h_5 * \beta^2 + h_6 * d^3 + h_7 * d^2 * \beta & \\ +h_8 * d * \beta^2 + h_9 * \beta^3 + h_{10}, & otherwise \end{cases} \tag{4.7}$$

In summary, in this section, we describe how to model the channel using simplified mathe-

Table 4.1: Parameters for the channel coefficient generation procedure

Parameter	Value	Parameter	Value	Parameter	Value
a_1	0.00042	a_2	−0.000003	a_3	0.292
b_1	−0.0042	b_2	0.0004	b_3	1.605
c_1	−0.18	c_2	−0.021	c_3	−71.02
d_1	−1.1	d_2	3.9	d_3	0.74
d_4	−0.2	d_5	−0.18	d_6	−0.058
d_7	0.0078	d_8	0.0023	d_9	0.0029
d_{10}	−98.23				
e_1	−0.0042	e_2	−0.0007	e_3	−135.19
f_1	−0.36	f_2	0.031	f_3	1.608
g_1	−0.0088	g_2	0.00189	g_3	1.28
h_1	−0.5	h_2	−0.045	h_3	0.061
h_4	0.01	h_5	−0.004	h_6	−0.0023
h_7	−0.00078	h_8	−0.000013	h_9	0.000097
h_{10}	1.17				

matics equations and distribution which works at 140 GHz frequency. To model the channel, we fit the β, and d, in both δ (LoS or NLoS) to find the distribution with the best KS score among more than 80 distributions [*]. Figure. 4.2 illustrates the regression line to find the best parameters for distribution at different distances, whereas we summarize the coefficients of parameters in Table 4.1.

4.4 THz-GAN channel

The THz-GAN channel is a sub THz model-free indoor channel modeling approach that relies on cGAN. GANs are novel sorts of DNN widely recognized for their ability to produce data based on the exact statistics of a dataset through indirect learning. Channel conditions

[*]Distributions list can be found https://fitter.readthedocs.io/

as a controlling procedure are an integral part of channel modeling, determining the requisite cGAN architecture in this context. In the THz-GAN channel approach, we employ a conventional cGAN architecture with a customized architecture, which is described in detail in subsection 4.4.3. The cGAN-based channel modeling framework completely avoids the critical analysis and complex processing of raw measurement data in traditional channel modeling methods. So we introduced THz-GAN channel as an alternative to traditional channel modeling and SSCH-THz to get beyond the need for considerable prior knowledge of channel modeling and complex processing of raw measurements in traditional channel modeling approaches. The architecture of the THz-GAN channel is trained based on the 140 GHz indoor channel model of *SC1* to act as a baseline model, which allows for rapid and precise channel modeling. We define the adaptability feature, which employs this baseline model and adapts a new dataset through transfer learning in order to model new channels for *SC2* and *SC3* scenarios. The adaptability feature of the THz-GAN channel enables generality and a detailed channel model approach, which reduces the time and cost associated with developing novel channel models for each scenario. In the next paragraphs, first, the GAN and cGAN will be described, then the architecture of cGAN, and then the parameter learning method will be emphasized. Finally, the transfer learning approach will be introduced.

4.4.1 GAN

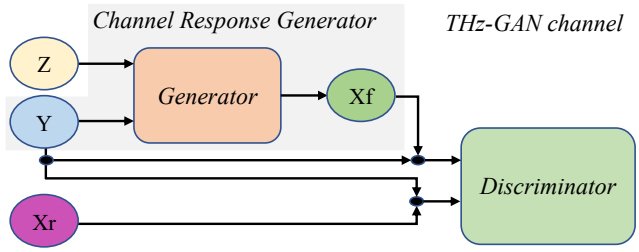
GAN are a unique kind of DNN known for their capacity to generate data based on a dataset's precise statistics using indirect learning. GANs can be trained on the limited available data to generate synthetic data that can help in modeling the THz channel. As discussed earlier in Sec. 3.5, the THz spectrum has unique channel characteristics such as high attenuation, directional propagation, and susceptibility to atmospheric absorption. These complex characteristics make it difficult to model the THz channel accurately using traditional approaches [64]. GANs can capture the complex channel characteristics of the THz spectrum and generate synthetic data that can be used for channel modeling [54]. These features lead GAN to fit very well into the concept of channel modeling to generate channel responses based on previously collected data. GAN are a class of data distribution modeling methods that consist of two functions: the generator G, which converts a sample from a random uniform distribution to the data distribution, and the discriminator D, which determines whether a given sample corresponds to the data distribution or not. Both G and D may be non-linear mapping functions, for example, a DNN. Adversarial networks have the advantages of using just backpropagation to acquire gradients, requiring no inference during learn-

ing, and efficiently including a large range of inputs and interactions into the model, which renders GANs applicable to a variety of applications, including image processing [65, 66], medicine [67, 68], text transmission [69, 70], etc. Additionally, as proved in [71], GAN is capable of producing cutting-edge log-likelihood estimates and realistic samples. To learn a generator distribution p_g over the dataset, the generator constructs a mapping function $G(z)$ from a noise distribution $p_z(z)$ where z is input noise. The discriminator, $D(x)$, returns a single scalar indicating the likelihood that x originated from training data (Xr) rather than p_g (Xf) which indicates that data is real or fake. Both G and D are trained concurrently, parameters are adjusted for G and D to minimize $\log_{10}(1 - D(G(z)))$ and $\log_{10} D(x)$, respectively, by following min-max value function $V(D, G)$:

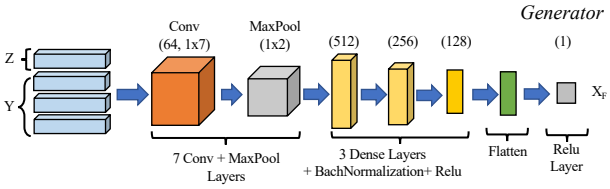
$$\begin{aligned} \underbrace{\min}_{G}\underbrace{\max}_{D} V(D,G) = \mathbb{E}_{x\sim_{|data}(x)}[\log_{10} D(x)] \\ + \mathbb{E}_{z\sim_{|z}(z)}[\log_{10}(1 - D(G(z)))] \end{aligned} \tag{4.8}$$

4.4.2 Conditional generative adversarial network (cGAN)

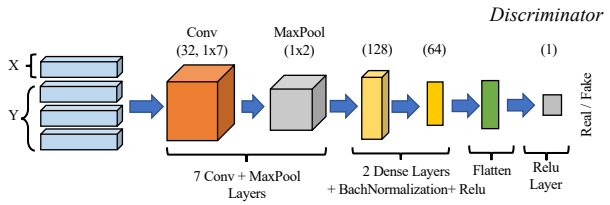
cGAN is a subtype of GAN that applies the Y space (channel conditions) to the Generator and Discriminator, resulting in a controllable training approach. In wireless communication systems, the channel conditions such as distance and receiver bandwidth can change in different scenarios due to various factors such as the mobility of the TX or RX, changes in the environment, and interference from other signals. Thus integrating these conditions as a controlling strategy establishes the necessary cGAN architecture compared to GAN. If both the generator and discriminator are conditioned on some additional information y, such as class labels or input from other domains, generative adversarial networks can be expanded to a conditional model. As we are modeling the channel based on cGAN, y is a channel condition,e.g, d, β and LoS/NLoS. Conditioning can be accomplished by feeding y into both the discriminator and generator as an extra input layer. The generator integrates the previous input noise $p_z(z)$ and y into a joint hidden representation, and the adversarial training framework enables substantial flexibility in the composition of this hidden representation. The discriminator accepts x and y as inputs and applies a discriminative function on them [72].



(a) THz-GAN channel architecture based on cGAN



(b) Generator of cGAN



(c) Discriminator of cGAN

Figure 4.3: Structure of the THz-GAN channel including architecture

$$\underbrace{min}_{G}\underbrace{max}_{D} V(D,G) = \mathbb{E}_{x\sim_{data}(x)}[\log D(x \mid y)] + \mathbb{E}_{z\sim_{z}(z)}[\log(1 - D(G(z \mid y)))] \tag{4.9}$$

4.4.3 Architecture

In this section, we define the THz-GAN channel architecture, based on cGAN (described in Subsection 4.4.2). Figure 4.3 illustrates the architecture of the THz-GAN channel, whereas Z, Y, Xr, and Xf represent the input noise, channel conditions, original path loss, and generated channel response, respectively. The generator (G) adapts to channel conditions and noise, whereas the Discriminator (D) is responsible for identifying real and fake data. Fig-

ure 4.3a represents the architecture of G and D whereas G and D take as input and $\{Xr, Y\}$ & $\{Xf, Y\}$, respectively. Figure 4.3b depicts G 's architecture that comprises twelve layers: $\{Z, Y\}$ feeds as input to the network seven convolutions (size 64), and MaxPool layers (size 1×2) are in charge of feature extraction and reducing size followed by three Dense layers (sizes: 512, 256, and 128) and batch normalization. Finally, we apply one flatten layer with Rectified Linear Unit ($ReLU$) Activation function of size 1 as output. The architecture of D is detailed in Figure 4.3c, which comprises eleven layers: combination of $\{Xr, Y\}$ or $\{Xf, Y\}$ feeds as input to the network. Seven convolutions (size 32), and MaxPool (size 1×2) layers are in charge of feature extraction and reducing size followed by two Dense layers (sizes: 128, 64) and batch normalization. Finally, we apply one flatten layer with $ReLU$ Activation function of size 1 as output. The dense layer implements the operation:Output = activation(dot(input, Kernel weight)), where activation is the passed-in element-wise activation function argument for activation kernel weight which determines the matrix of weights of the output. Specifically, in all dense layers, we utilized the $ReLU$ function in our implementation as an activation function. The main reason for employing the $ReLU$ activation function is that it works well in practice [73] and could reduce the time and complexity required for training and operating steps THz-GAN channel in real-time applications. Batch Normalization is the application of a transformation that preserves the mean near to 0 output and close to 1 output standard deviation. A Flatten layer modifies the form of the row data, which equals the number of elements included in the row data, equivalent to creating a 1d array of elements. Notice that, after training, G is used as a channel response generator (depicted as gray in Figure 4.3), however both G and D are necessary to adapt and learn the new channel models.

Finding the best parameters for training a cGAN is a challenging task as it requires optimizing multiple hyperparameters, including the learning rate, batch size, number of epochs, number of generator and discriminator layers, activation functions, regularization techniques, etc. We trained the THz-GAN channel with different combinations of hyperparameters and evaluated the performance using the evaluation metric and keep the hyperparameters that achieve the best results which are as follows. We utilize $ReLU$ as an activation function, 80000 epochs, batch size of 32 with Stochastic gradient descent as optimizer with 0.001 learning rate. For loss function for G and D we utilize mean absolute error and binary cross-entropy, respectively.

4.4.4 Adaptability Feature

In this section, we introduce the adaptability feature of the proposed channel modeling approach, which enables re-usability of the THz-GAN channel in new scenarios with a lower dataset and higher speed. The most important advantages of using transfer learning in novel channel modeling are listed below:

- **Reduced required amount of data:** The gathering of data for each new scenario for training machine learning models for THz wireless communication is expensive and time-consuming. Transfer learning can be used to leverage pre-trained models on other related scenarios, to extract useful features. So these pre-trained models have already learned features that are generalized to other similar scenarios and can be applied to the new scenario and adopted by limited data.

- **Faster training:** Training deep learning models from scratch on limited data can be computationally expensive and time-consuming. Transfer learning allows for faster training of models as pre-trained models can be used as a starting point, and only the model needs to be adopted in the new scenario. This can significantly reduce the time and resources required for training a deep learning model for THz wireless communication.

- **Improved Accuracy:** Transfer learning improves the accuracy of deep learning models for THz wireless communication by leveraging the knowledge learned from pre-trained models in related scenarios. This can lead to better generalization and more accurate predictions, especially when the amount of available data for training the new scenario is limited.

The procedure of transfer-learning-based channel modeling for new scenarios is illustrated in Figure 4.4. The required steps to generate a new channel model are as follows.

Pre-training the discriminator and generator

Pre-training the discriminator and generator on $SC1$ provides a starting point for indoor channel modeling. We utilize the weights of the Discriminator and Generator of $SC1$ model as initial weights of the THz-GAN channel neural networks for $SC2$ and $SC3$.

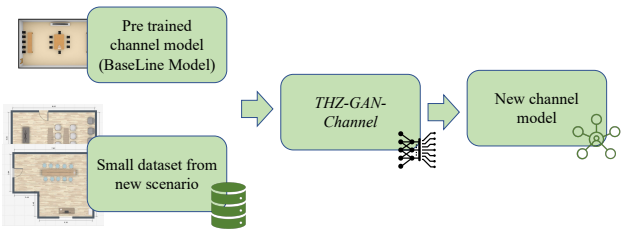


Figure 4.4: Transfer Learning

Fine-tuning the discriminator and generator

After pre-training the THz-GAN channel for $SC1$, we fine-tune the THz-GAN channel model based on the dataset from $SC2$ or $SC3$ to adapt to the specific characteristics of the particular model to represent the new channel model.

4.5 ns-3 integration

To assess the channel modeling performance and how higher layers perform in the next generation of communication a full-stack study is required. So we integrated the proposed channel models in the ns-3 Terasim module to compare the efficiency of channel modeling approaches by full-stack performance analysis. In the *THzChannel* class of *TeraSim*, waveform propagation through a wireless channel is modeled. A waveform object is created by the *THzSpectrumValueFactory* class and provided to the channel object. The *THzDirectionalAntenna* module is used to first determine the direction of the devices and then to acquire the antenna gains. The received power is then computed by the *THzSpectrumPropagationLoss* class using the estimated antenna gain and CIR. The CIR is frequency-selective and dependent on both frequency and distance. The *THzChannel* object then transmits the payload and the received power to the recipients' physical layer.

4.5.1 Integration of HB and FS Channels

TeraSim models the waveform propagation through the wireless channel in the *THzChannel* class, as shown in Figure 4.5. The *THzSpectrumValueFactory* class generates an object representing a waveform and passes it to the channel object. This first checks the orientation of the devices, and obtains the antenna gains through the *THzDirectionalAntenna* module. Then, the *THzSpectrumPropagationLoss* class calculates the received power based on the calculated antenna gain and the CIR. The CIR is frequency selective and a function of the distance and frequency. Finally, the *THzChannel* object passes the packet along with the received power to the physical layer of receivers.

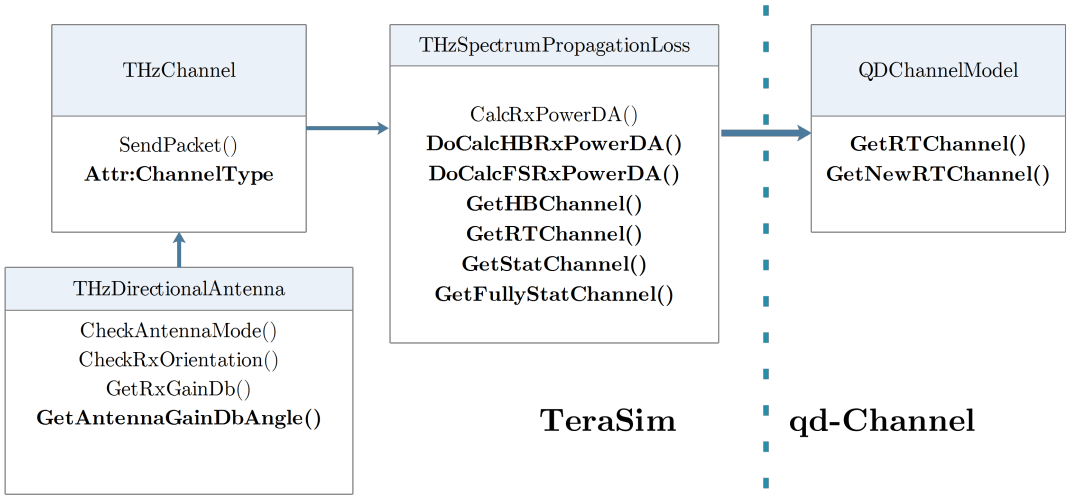


Figure 4.5: TeraSim classes and new channel modeling code.

Figure 4.5 summarizes the integration that we have made to support the HB and FS in TeraSim. Most of the updates have been introduced in the *THzSpectrumPropagationLoss* class. Two new functions are responsible for calculating the received power for HB and FS channels, and the type of channel can be selected through the *ChannelType* attribute in the *THzChannel* class. Additionally, while the CIR for FS is generated at run-time in ns-3, the CIR for the HB requires the generation of the RT MPCs offline. For this, we use the Q-D Channel RT tool [74, 46], an open-source MATLAB-based RT tool for the mmWave spectrum. We updated the parameters (e.g., permittivity) in the material library and the path loss equation with those introduced in [44, 75]. The MPCs generated by the RT are then loaded at run-time using the ns-3 qd-channel module.[†]

4.5.2 Integration of the Simplified Stochastic Channel Model in THz (SSCH-THz) in ns-3

The CIR of StatChannel relies on a random distribution which we implement in ns-3. So the CIR for the channel model of SSCH-THz is produced at run-time based on the fully stochastic channel model represented in Eq. (4.1). We integrate the channel model in the ns-3 Terasim module whereas, the *THzSpectrumPropagationLoss* class has been the host of the majority of the integrations. The received power for the channel is determined by a novel function implemented in *THzSpectrumPropagationLoss*.

[†] https://github.com/signetlabdei/qd-channel

4.5.3 Integration of THz-GAN channel in ns-3

Integration of THz-GAN channel is a bit more challenging. We follow these steps: First, in order to train, we implement the THz-GAN channel model in Python language and Keras toolbox. Then, we compile the code in C++ by utilizing the Frugally-deep[‡] library. To serve as a channel model, we then incorporate the code into the *THzSpectrumPropagationLoss* class. It should be noted that thanks to ns3-ai[§] real-time ns-3 to Python is available, but the simulation run time decreased drastically in this manner.

4.6 Performance Evaluation

In this section, we describe the performance analysis. First, we describe a statistical analysis in which we validate the proposed channel modeling approaches based on statistical tests and validation. Then we show full-stack analysis based on ns-3, in order to highlight the advantages of proposed channel modeling approaches. It is worth mentioning that, implement all the channel models in $SC1$, whereas in $SC2$ and the $SC3$ scenarios we implemented the THz-GAN channel by transfer learning approaches in which weights are derived from the THz-GAN channel in $SC1$'s channel model. In order to validate the THz-GAN channel in $SC2$ and $SC3$ evaluation we developed the HB as a benchmark.

4.6.1 Statistical Results

It is crucial to validate the chosen distribution by the appropriate performance metric after we fit data into a distribution. KS is an important metric in statistical approaches that compares the distributions of two datasets in order to determine if a dataset follows a particular distribution or not [76]. The test yields a *P-value* that represents the likelihood of obtaining the observed difference in distributions by chance. So, a higher *P-value* indicates that the two datasets are likely drawn from the same distribution. The *D-value*, also known as the KS statistic, is a numerical measure of the maximum difference between the CDFs of the two datasets being compared. It represents the largest vertical distance between the CDFs and quantifies the discrepancy between the distributions. Therefore, the datasets are more similar if the *D-value* is small. We performe the KS test in SSCH-THz in $SC1$, and for the THz-GAN channel in $SC1$, $SC2$, and $SC3$, which we summarized in the Table 4.2. It is

[‡] https://github.com/Dobiasd/frugally-deep
[§] https://github.com/hust-diangroup/ns3-ai

shown that all channel models pass the test with acceptable outcomes (high *P-value* and low *D-value*).

Table 4.2: A goodness of Fitting (two sample Kolmogorov Smirnov Test) for $SC1$, $SC2$, $SC3$ proposed channel models compare to HB.

Channel Model	*P-value*	*D-value*
SSCH-THz	0.292	0.071
THz-GAN channel ($SC1$)	0.451	0.053
THz-GAN channel ($SC2$)	0.419	0.091
THz-GAN channel ($SC3$)	0.389	0.012

The reduction of learning time and data volume is the aim of transfer learning channel modeling. So to assess this, we summarize the training Mean Absolute Error (MAE) for three scenarios of the THz-GAN channel in Fig. 4.6. Notice that for $SC1$, $SC2$, and $SC3$ we collected 1000000, 1000 and 1000 samples, respectively. From the figure, it can be observed that fine-tuned models in $SC2$ and $SC3$ reach less than 0.1 MAE in less than 80 epochs while the THz-GAN channel in $SC1$ required 80000 epochs. This graph shows how fine-tuning and transfer learning can increase learning speed while also requiring less data.

4.6.2 Numerical Results (ns-3)

Channel CIR Analysis

Figure 4.7 depicts the ECDF of the received power (including antenna gain) for the HB, SSCH-THz, and THz-GAN channel, including LoS and NLoS, for three scenarios. In this simulation campaign, we measure the received power of the path in scenarios including both LoS and NLoS, in which results are collected from 10 simulation runs. The simulation's primary objective is to examine and contrast received power as a physical layer performance analysis of novel channel modeling approaches. It is obvious that the THz-GAN channel and HB model channels gain comparable performance in $SC1$(in which the main THz-GAN channel is trained), and the $SC2$ and $SC3$ (in which the THz-GAN channel is trained based on transfer learning) scenarios with a maximum difference of less than 10 dB (Fig. 4.8). On the other other hand, we assess the SSCH-THz channel performance in $SC1$ as well. It is evident that, the received power of SSCH-THz follows precisely other channels responses. From a physical layer perspective, a noteworthy aspect of this simulation campaign is the similarity of the results obtained by channel models in various scenarios, which emphasizes the high caliber of the suggested channel modeling approaches.

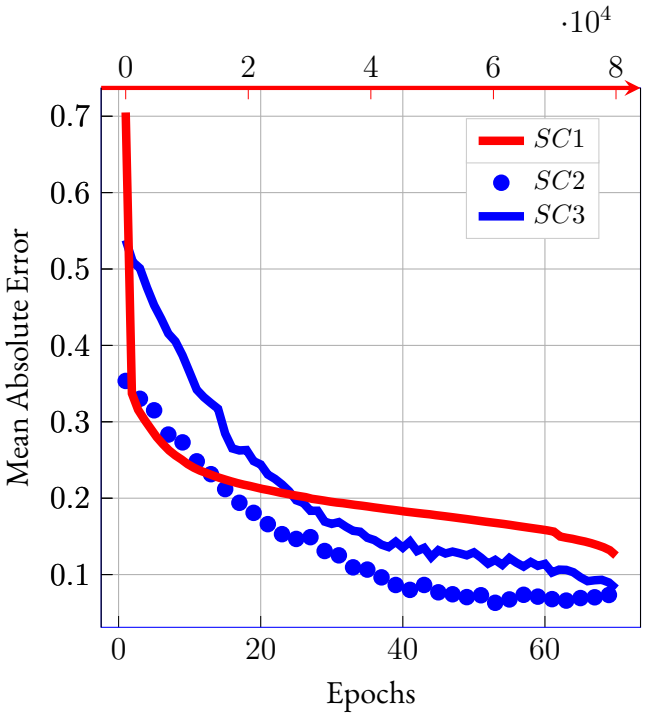


Figure 4.6: MAE of learning of the THz-GAN channel during training of various Scenarios; $SC2$ and $SC3$ are fine-tuned by transfer learning, while $SC1$ is direct learning.

Channels Run Time Analysis

It is crucial to set up a large number of runs when using simulators in order to produce fair results. In a large-scale simulation campaign using simulators, e.g., ns-3, channel execution time is one of the bottlenecks of the system simulation run time and and it can constrain the scope of simulation. For simulation campaigns, the channel response generation process can vary depending on the implementation method and be time-consuming. Figure 4.9 shows the average run times for generating one-circle channel responses for the various channel models in the three aforementioned scenarios. For this, we ran ten simulations for each configuration and averaged the results. In $SC1$ we measure run times per Terasim, FS, HB, THz-GAN channel, and SSCH-THz, whereas, Terasim hits the minimum average run time which is consistent with the LoS condition and only depends on the frequency of the channel model. Despite this, in $SC1$, among channel models that contain NLoS, SSCH-THz achieves the minimal average run time anticipated given that it adheres to simple stochastic channel modeling. With a small difference, the THz-GAN channel model achieves minimum execution time, which is comparable with FS and HB. HB channel model

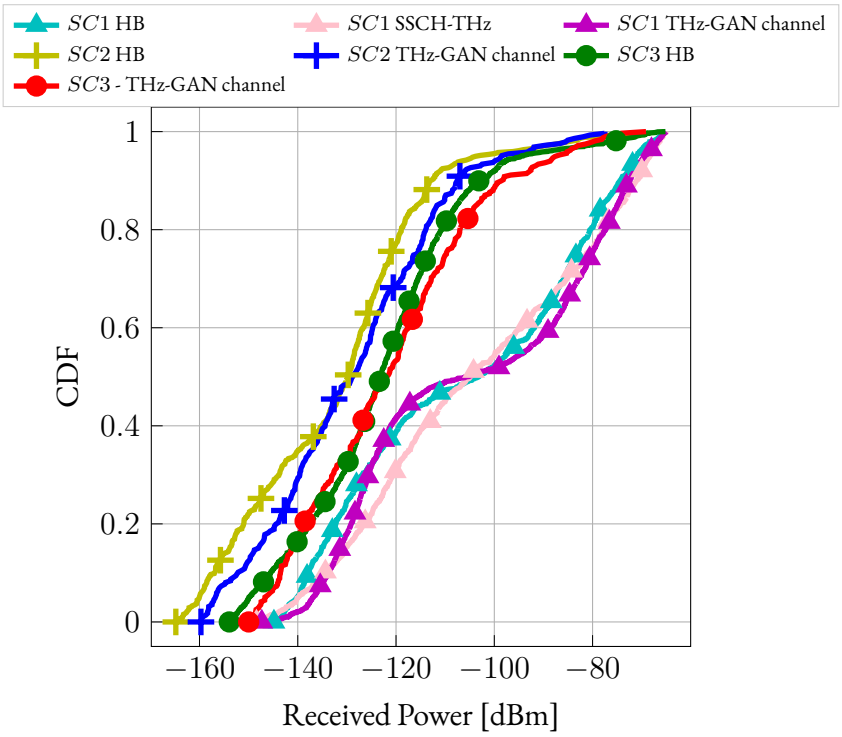


Figure 4.7: ECDF of HB, THz-GAN channel, and SSCH-THz channel responses in terms of received power including antenna gain in $SC1$, $SC2$, and $SC3$.

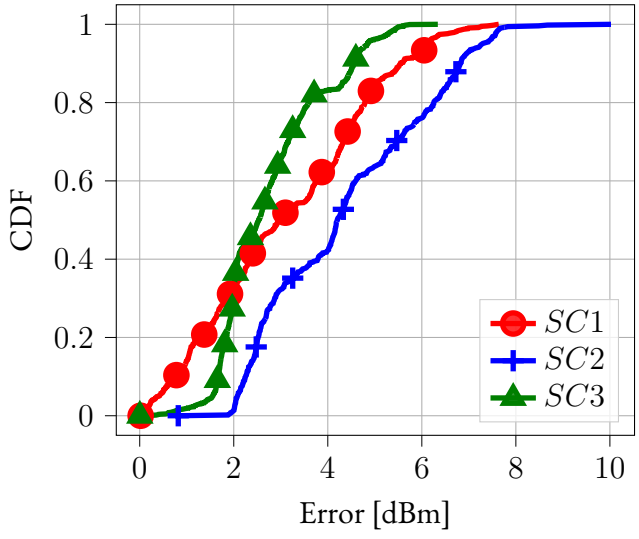


Figure 4.8: ECDF of different HB versus THz-GAN channel channel responses in terms of received power including antenna gain in $SC1$, $SC2$, and $SC3$.

requires the longest time to generate each channel response, combining two ray tracing and stochastic parts. On the other hand, In $SC2$ and $SC3$ in line with $SC1$, THz-GAN channel achieves a comparable lower run time average than HB. For instance in $SC2$ and $SC3$ THz-GAN channel is 9.3 and 11.3 times faster than HB, respectively.

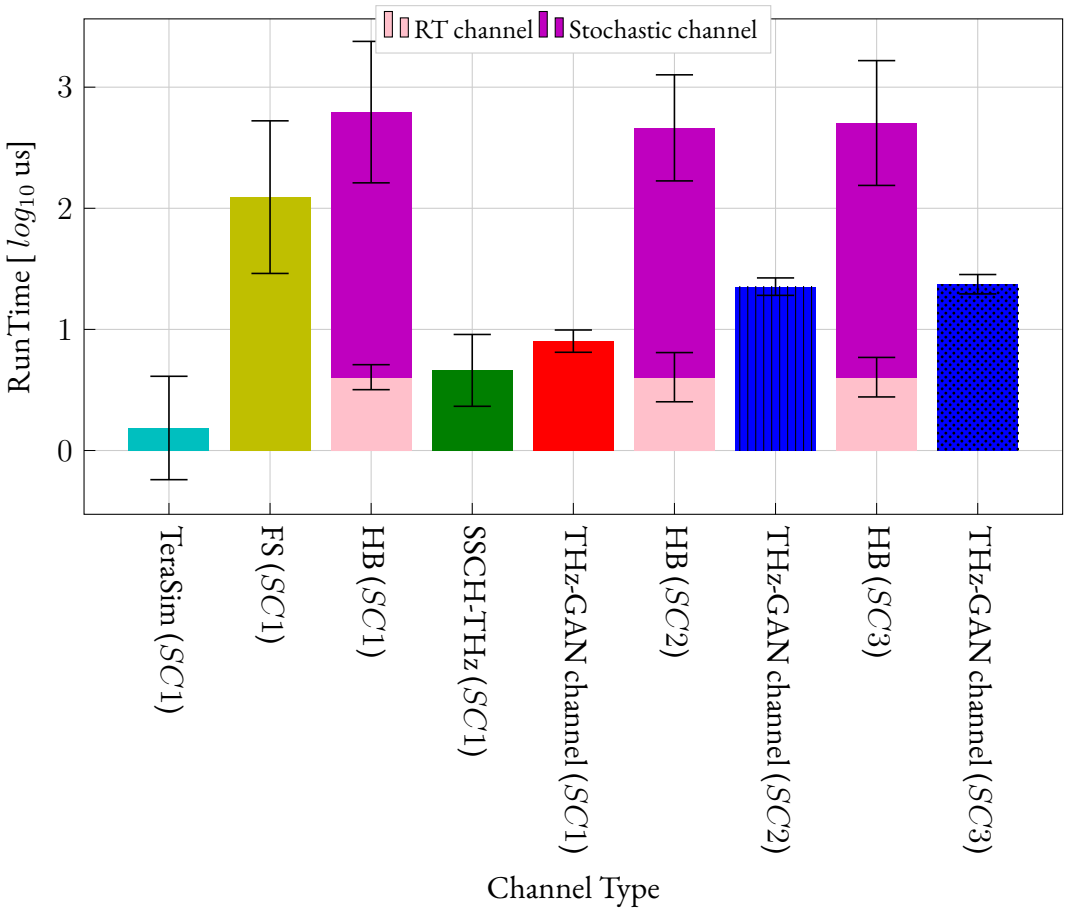


Figure 4.9: Average channel run times in $SC1$, $SC2$, and $SC3$ scenarios using Terasim, FS, HB, SSCH-THz, and THz-GAN channel.

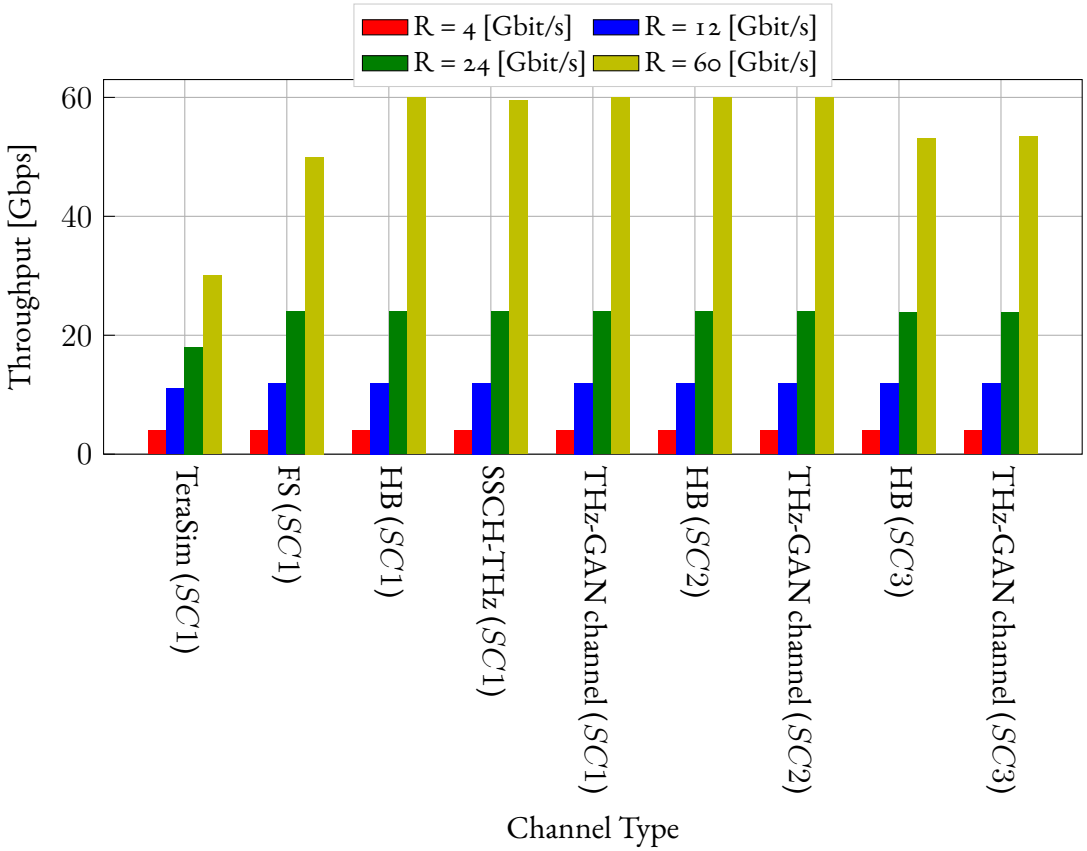
UDP application Performance Analysis

We investigate UDP-based simulation for various source rates in order to assess full-stack system analysis based on novel channel modeling approaches. Figures 4.10a and 4.10b show the throughput and latency of Terasim, HB, FS SSCH-THz, and THz-GAN channel, for various system source rates, ranging from 4 $[Gbit/s]$ to 60 $[Gbit/s]$, respectively.

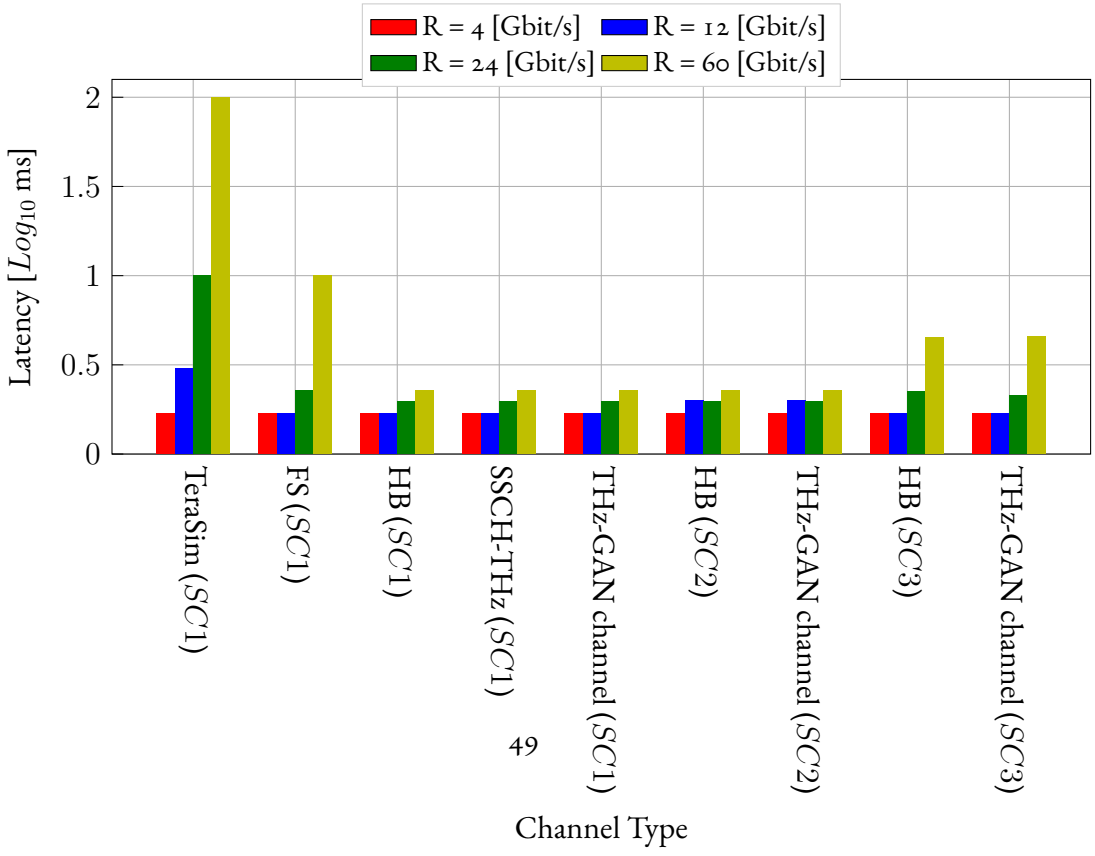
From the results, can be seen that for lower source rate values, the performance of the five-channel models in $SC1$ scenario are similar. The TeraSim channel results indicate a slightly higher latency in comparison to FS, HB, SSCH-THz, and THz-GAN channel in which performance metrics are almost constant. Increasing source rate to 60 $[Gbit/s]$ results in 30.3 and 3.1 times higher latency and 2.5 and 1.2 times lower throughput of the Terasim and FS compared to other channel models, respectively. HB and THz-GAN channel catches similar results in terms of throughput and latency in $SC2$ and $SC3$ which are in line with other performance analysis and demonstrate the benefits of transfer learning based channel modeling.

TCP application Performance Analysis

However, UDP analysis is valuable, especially for real-time applications where low latency is critical. UDP analysis often focuses on measuring packet loss rates, and monitoring application-specific data within the UDP payloads. TCP provides more built-in features for analyzing network traffic due to its reliable and connection-oriented nature. This makes TCP analysis more comprehensive and detailed for understanding network behavior and troubleshooting. Consequently, we run simulations based on various channel modeling techniques and using TCP to extend full-stack analysis. We summarized the TCP throughput as the performance metric, for different channel modeling techniques, in Fig. 4.11. by comparing the throughput evaluation of full-stack analyses of various channel modeling techniques in $SC1$, $SC2$, and $SC3$. We note that the novel channel models adhere to existing techniques.



(a) Throughput



(b) Latency

Figure 4.10: UDP performance of different channel models in various deployment scenarios ($SC1$, $SC2$, and $SC3$).

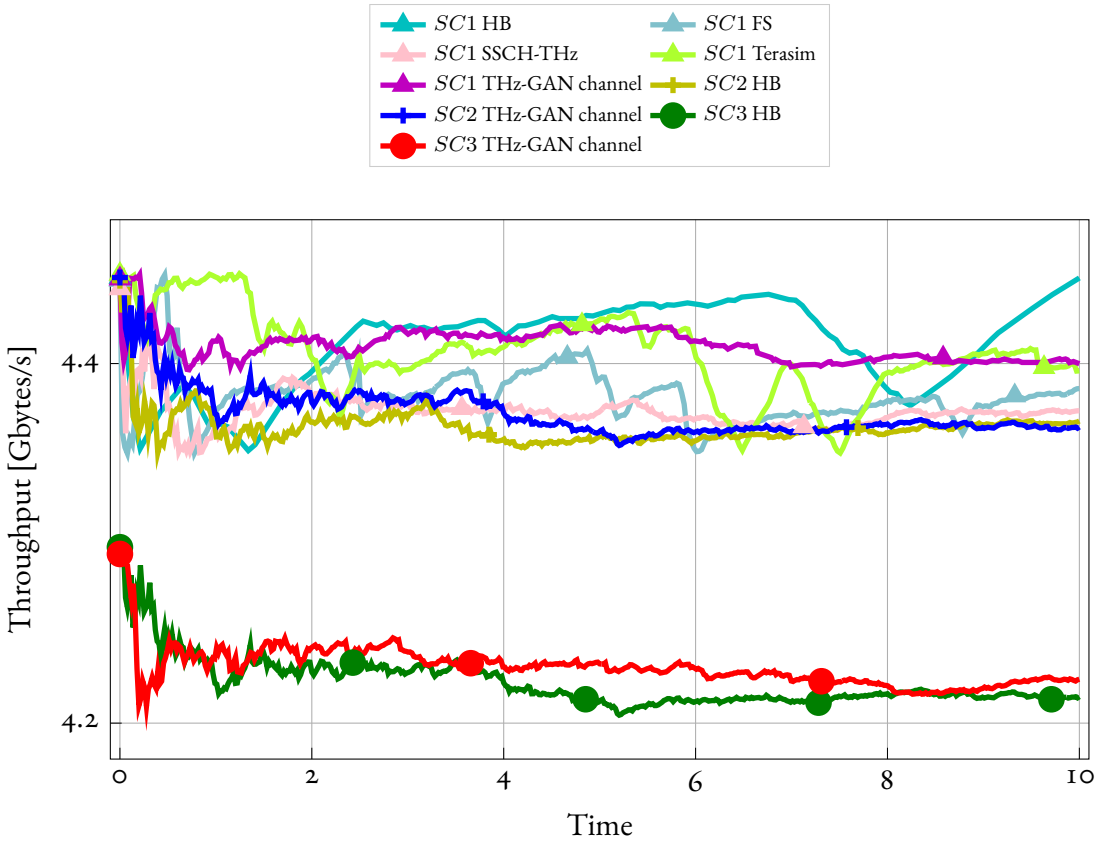


Figure 4.11: Instance of TCP throughput of different channel models with 20 degree of β.

5

Integrated and Access (IAB) in 5G Networks

The topic of 5G networks will be covered in some detail in the following chapter. Here, we provide an explanation of the 3gpp standard that makes 5g NR possible. After that, a discussion of the difficulties and opportunities presented by wireless communication at mmwave frequencies follows. In conclusion, we present the RIS and IAB as possible solutions to extend the communication range. These potential solutions will be discussed in further depth in the following chapters.

5.1 Introduction

5G, an abbreviation for the fifth generation of wireless communication technology, represents a substantial advancement in the field of connectivity. It is designed to deliver remarkable improvements in wireless network speed, responsiveness, and capabilities. 5G is intended to reach transmission speeds of up to 20 Gbps, making it significantly faster than its predecessors [77]. This speed is a game-changer, as it allows for rapid downloads, seamless streaming, and support for emerging technologies such as augmented reality (AR) and the Internet of Things (IoT). A key characteristic of 5G is its reduced latency. This enables real-time applications, such as autonomous vehicles and remote surgery, to be implemented [78]. In essence, 5G is not just about faster smartphones; it's the foundation for a more connected

world. It will enable previously unimaginable smart cities, advanced healthcare, and transformative experiences. 5G is poised to drive the next wave of innovation and connectivity on a global scale, given its potential to transform industries and daily life. 5G utilizes multiple frequency bands, each with their own characteristics and use cases [79]:

- **Low-Band 5G** (Sub-1 GHz): This band encompasses frequencies between 600 MHz and 1 GHz, with some sources extending it up to 2.3 GHz. Low-band 5G offers a larger service area, making it suitable for rural and suburban areas. It provides a modest increase in data transfer speeds over 4G.
- **Mid-Band 5G** (1 GHz to 6 GHz): These frequencies are utilized for mid-band 5G. They include frequencies such as 2.1 GHz and 2.3 GHz. Mid-band 5G provides a balance between coverage and data speeds and is suitable for urban and suburban environments.
- **High-Band 5G** (Above 24 GHz): Also referred to as mmWave (millimeter wave), high-band 5G operates between 24 GHz and 71 GHz. It offers incredibly fast data transfer rates but has limited coverage and is primarily deployed in dense urban areas.

5.2 3GPP NR: the Set of Specifications for 5G Networks

3GPP NR (Third Generation Partnership Project New Radio) is a set of technical standards that define 5G networks [80]. These specifications are crucial for ensuring interoperability and compatibility between equipment and devices from various manufacturers within the 5G ecosystem. 3GPP introduced 5G NR due to the limitations of LTE (Long-Term Evolution) networks [81]. These constraints include:

1. In terms of data speeds, LTE networks had reached their practical limits, despite offering faster data transfer rates than earlier technologies. The demand for faster connections and applications requiring more bandwidth necessitated a new solution.
2. As the number of connected devices and data-intensive applications such as streaming and IoT grew, LTE networks faced capacity challenges. 5G New Radio (NR) was created to support a significantly larger number of devices and to provide greater capacity.
3. LTE networks had relatively low latency, but 5G aimed to reduce latency even further to meet the needs of autonomous vehicles and real-time remote control applications.

4. LTE networks operated within a limited range of frequency bands and coverage areas. 5G NR introduced mmWave frequencies and a broader spectrum, allowing for improved coverage and performance, particularly in dense urban areas.

5. 5G NR was designed to be more energy-efficient than LTE, making it suitable for a wider range of IoT and battery-powered devices.

6. 5G NR introduced network slicing, which enables operators to create multiple virtual networks within a single physical network. This customization was not possible on the LTE network.s.

In summary, 3GPP introduced 5G NR to address the limitations of LTE networks, such as the need for faster speeds, increased capacity, lower latency, broader coverage, greater energy efficiency, and advanced features such as network slicing. These modifications were necessary to meet the requirements of emerging technologies and applications. Key aspects of 3GPP 5G NR network specifications are as follow [2]: The specifications cover a variety of technical aspects of 5G technology, such as network architecture, radio frequencies, and protocols. The 3GPP periodically releases new versions of these specifications. Release 15, for example, introduced the initial 5G NR specifications. 3GPP NR ensures that 5G devices from various manufacturers are compatible with 5G networks, fostering a competitive and diverse market. These specifications enable network operators to deploy 5G networks that can serve a variety of user devices, including smartphones and IoT devices.

5.3 Potential and Challenges of mmWave band communication

In 5G networks, mmWave is an essential technology component. mmWave operates between 24GHz and 100GHz in the high-frequency range. This frequency band has a significantly larger bandwidth than lower frequency bands, allowing for faster data transmission [82]. mmWave technology offers high capacity and gigabit-plus throughput. This is essential for the delivery of high-speed and high-capacity data services, which are fundamental to the 5G experience. While mmWave technology does not inherently reduce latency, its high capacity and data transfer speeds contribute to lower latency in 5G networks, enabling real-time applications such as autonomous vehicles and augmented reality. mmWave is essential to the realization of 5G's potential, enabling new applications and services that require massive data transfer capabilities [83]. It complements other frequency bands used in 5G networks to provide extensive capacity and coverage. mmWave technology enables innovative use cases, such

as Fixed Wireless Access (FWA), where 5G is used to bypass traditional wired connections and deliver broadband internet to homes and businesses [84]. In summary, the ability of mmWave technology to provide high capacity, broad bandwidth, and low latency makes it a crucial component of 5G networks. It enables the delivery of high-speed data services and supports a wide range of innovative applications that drive the advancement of 5G technology. Despite significant aforementioned advantages, mmWave and 5G network deployment comes with numerous challenges [85]:

- mmWave signals have a limited range and are easily blocked by obstacles such as buildings and vegetation. Effective coverage requires a dense network of small cells.
- mmWave signals require a direct line of sight between the transmitter and receiver, making them susceptible to signal interruptions in urban environments with tall buildings.
- Interference from a variety of sources degrade mmWave signal quality and impact network performance.
- Building the necessary infrastructure for mmWave and 5G networks, including the installation of small cells, can be costly, particularly in urban areas.
- mmWave signals are susceptible to atmospheric losses as a result of path loss, atmospheric absorption, rain, ally, and alluvium losses.
- Existing devices may not be compatible with mmWave frequencies, requiring users to upgrade to 5G-capable devices.
- Regulatory Difficulties Spectrum allocation for mmWave frequencies and navigating regulatory requirements can be challenging.

5.4 Relays

Relays are an essential component of 5G mmWave networks, addressing the unique challenges posed by high-frequency mmWave signals and ensuring that the network delivers on its promises of high data rates, low latency, and reliable connectivity in a variety of environments [86]. Relays play an important role in 5G mmWave radio access networks that mains are listed below [87].

- **Coverage Enhancement**: mmWave signals have limited coverage because of their high attenuation. By acting as intermediaries between base stations and user devices, relays can extend coverage by bridging coverage gaps.
- **Path Loss Compensation**: mmWave frequencies are especially susceptible to path loss in urban environments. This loss can be compensated for through the strategic placement of relays, ensuring dependable connectivity.
- **Improved Signal Quality**: Relays can improve signal quality by reducing interference and ensuring a direct line-of-sight connection between the base station and the user equipment. This is essential for achieving the high data rates promised by 5G.
- **Enhanced Capacity**: By offloading traffic from the base station, relays help distribute network load more evenly, thereby increasing network capacity and decreasing congestion.
- **Diversity and Reliability**: Relays introduce diversity into the network, which increases its reliability. They can retransmit signals to overcome signal loss or obstructions, ensuring a strong connection.
- **Low Latency Services**: Relays can reduce latency for ultra-low latency applications such as autonomous vehicles and remote surgery by providing a shorter path for data transmission.

In mmWave communication, both active and passive relays are essential for signal propagation. Using Reflecting Intelligent Surfaces (RIS) and Integrated Access and Backhaul (IAB) as examples, here is a list of the distinctions between them, which I describe in more detail in the next sections.

1. **Reflecting Intelligent Surface - Passive Relay:** RIS is composed of passive elements (reflectors) that modify the propagation of incoming signals [88]. It operates by controlling signal reflections and does not require active amplification equipment. RIS is excellent for optimizing beamforming, signal steering, and mitigating path loss in mmWave settings.
2. **IAB - Integrated Access and Backhaul - Active Relay:** Similar to conventional relays, the IAB amplifies and retransmits signals actively. Components: Power sources

and active amplification apparatus are required [89]. IAB is suitable for extending coverage and enhancing signal quality in weak signal locations. IAB configurations are typically more complex and require meticulous network planning.

In terms of hardware and power requirements, RIS is less complex than IAB. IAB is an active relay that amplifies and retransmits signals, whereas RIS is a passive relay that manipulates signal propagation using reflecting surfaces. The choice between them is dependent on the particular use case and network requirements, taking into account factors such as complexity and energy consumption [90].

5.4.1 Reconfigurable Intelligent Surface (RIS)

Wireless communications rely on the transmission and reception of electromagnetic pulses, which are frequently produced and detected by antennas. Each type of antenna is designed to operate over a particular frequency range and to deliver a particular level of performance. As antennas are responsible for converting electrical signals into electromagnetic waves and vice versa, their design and placement have a significant impact on the performance of a wireless communication system. RISs are a new form of technology that can be used to transmit and receive electromagnetic waves in lieu of conventional antennas. These surfaces are made up of a large number of small, individually controllable elements, each of which can be used to manipulate the phase and amplitude of the electromagnetic waves that pass through it. By altering the shape and orientation of these elements, RISs can reflect, absorb, or transmit electromagnetic radiation, enabling them to serve a variety of purposes. The fundamental principles underlying RISs stem from the concept of metasurfaces. A metasurface is a two-dimensional structure made up of a large number of subwavelength-sized elements, each of which can be individually controlled to modify the phase and amplitude of electromagnetic waves passing through it [91]. These elements can be constructed from a variety of materials, such as metals, dielectrics, and semiconductors, and arranged in a variety of configurations to accomplish specific electromagnetic properties [92]. Metasurfaces have been utilized in numerous applications, such as antennas, filters, polarizers, and beamformers [93]. However, the concept of RISs adds the capacity to control the geometry and orientation of the elements in real time to the concept of metasurfaces. This enables RISs to modify their shape and configuration to the particular requirements of the wireless communication system, making them more versatile and efficient than conventional antennas [94]. In addition to their adaptable shape and configuration, RISs have a number of additional advan-

tages over conventional antennas. These surfaces can be made significantly smaller and more compact than conventional antennas, making them better suited for use in devices with limited space [95]. Recently, reconfigurable intelligent surfaces have attracted a great deal of interest as a potentially effective means of enhancing the performance of wireless communication systems. RISs are passive, reconfigurable metasurfaces that can control the phase, amplitude, and polarization of electromagnetic waves that impinge upon them. The ability to intelligently control the reflection of impinging waves enables the RIS to function as a programmable reflector, allowing the wireless propagation environment to be tailored to enhance the performance of wireless communication systems. Prior to now, research on RISs has concentrated on three primary application areas [96]:

- **Spectrum efficiency**: By judiciously regulating the phase shifts of the RIS elements, the interference level between various transmission channels can be minimized, resulting in increased spectrum efficiency.
- **Energy efficiency**: Wireless communication systems can increase energy efficiency by directing signal energy to the intended receiver.
- **Physical layer security**: By producing a favorable propagation environment for authorized users and a hostile environment for eavesdroppers, the physical layer security of wireless communication systems can be improved.

Beamforming is a technique used to direct electromagnetic radiation in a particular direction, and it is frequently employed in wireless communication systems to expand the network's range and capacity. Intelligent surfaces that are reconfigurable can be used to shape the beam of electromagnetic waves, allowing them to target specific devices or avoid obstacles [97]. This is particularly useful in environments where signal interference is prevalent or where multiple devices need to be connected. Cloaking is the process of rendering an object or region invisible to electromagnetic radiation; this technique has been of interest to scientists for decades. Intelligent surfaces that can be reconfigured can be used to construct cloaking devices that render an object or region invisible to electromagnetic radiation over a specific frequency range [98]. This could have numerous applications in the real world, including military, security, and medical applications. In addition, it enables the deployment of machine learning algorithms to the device at the data's source.

RISs have several advantages over conventional antennas, including their size, versatility, and adaptability. These surfaces can be made significantly smaller and more compact than

conventional antennas, making them better suited for use in devices with limited space [99]. In addition, these surfaces can alter their shape and configuration to the particular requirements of the wireless communication system, allowing them to perform a broader range of functions than conventional antennas [100]. Despite these benefits, the widespread implementation of reconfigurable intelligent surfaces in wireless communications confronts a number of obstacles. These surfaces require a large number of individually controllable elements, which can be expensive to produce and maintain [101]. Another challenge is the complexity of these surfaces, as they require sophisticated control algorithms to maximize their performance and adapt to shifting conditions [102]. In addition, regulatory and standards concerns must be resolved before these surfaces can be implemented on a significant scale [103]. The implementation of a RIS-enabled smart radio network [104] must resolve a number of challenges that may arise including the following:

- **Channel Modeling**: The behavior of signal propagation in the presence of an RIS is poorly understood, making it difficult to model the channel accurately.
- **Protocol Design**: Integration of RISs into a radio network necessitates the development of new protocols to govern and coordinate the RIS's operation with other network elements.
- **Complexity**: the control and management of a large number of RIS elements can be computationally complex and may require advanced algorithms to optimize the network's performance;
- **Synchronization**: The RIS elements must be synchronized to work together efficiently, which can be difficult when the network is highly dynamic;
- **Energy Efficiency**: Because the operation of a RIS-enabled network can be power-hungry, it is crucial to design energy-efficient algorithms for managing the RIS;
- **Security and Privacy**: The use of a large number of RIS components can pose security risks, and it is crucial to employ robust security mechanisms to safeguard the network from malicious attacks;
- **Deployment and Maintenance**: The deployment and maintenance of a RIS-enabled network can be expensive and requires careful planning to ensure that the network is deployed as efficiently and effectively as possible;

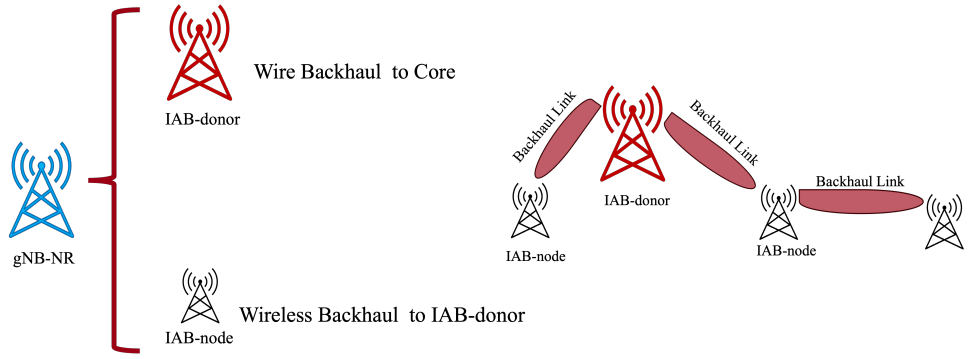


Figure 5.1: Different types of BS in IAB network and backhaul connection

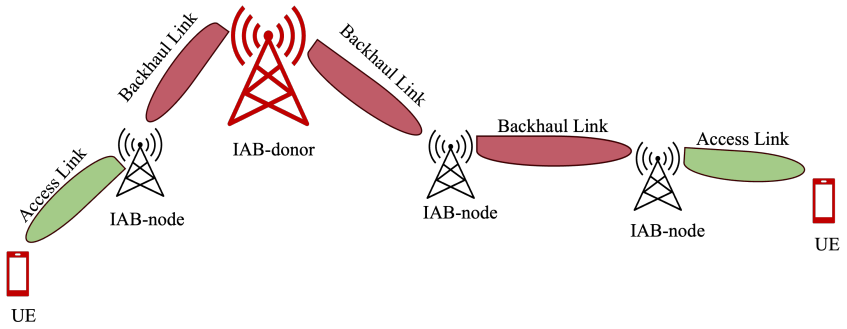


Figure 5.2: illustration of an example of UE connection to the network (Uplink) in the IAB Network.

- **Interference**: Due to the reflection of the signals, introducing a RIS can increase interference in the network, which can make its management more difficult.

5.4.2 Integrated Access and Backhaul (IAB)

As discussed earlier, operating in the mmWave spectrum presents a unique set of challenges, including severe path and penetration losses [105, 106]. Using high-gain antennas to help close the link, thereby introducing directionality in the communication, and electronic beamforming to support mobile users is a promising strategy for overcoming such limitations. Network densification is also used to enhance performance by decreasing inter-site distance to establish more robust access channels. However, an ultra-dense deployment incurs high capital and operational expenses (capex and opex) for network operators [107] due to the need to provide high-capacity backhaul connections to a greater number of cellular base stations than in networks operating at lower frequencies. Network disaggregation (the separation of the layers of the protocol stack into different physical equipments) [108] and virtualization (the use of software-based rather than hardware-based protocol stack implementa-

tions) [107] can lower capex and opex by reducing the complexity of individual base stations. Some researchers have also begun investigating the viability of Integrated Access and Backhaul (IAB), in which only a subset of gNB connect to traditional fiber-like infrastructures and the rest wirelessly relay the backhaul traffic, possibly via multiple hops and at mmWave frequencies [109]. The 3GPP has acknowledged the significance of the IAB framework as a cost-effective alternative to wired backhaul. In fact, it has recently concluded a Study Item for 3GPP NR Release 16 [81] that investigates the architectures, radio protocols, and physical layer aspects of sharing radio resources between access and backhaul links. Although the 3GPP LTE and LTE-Advanced standards already provide specifications for base stations with wireless backhauling capabilities, the Study Item on IAB envisions a more advanced and flexible solution, including the support of multi-hop communications, dynamic multiplexing of the resources, and a plug-and-play design to reduce deployment complexity. However, despite widespread agreement regarding IAB's ability to reduce costs, the design of a cost-effective and high-performance IAB network remains an open research problem. Research on wireless backhaul solutions has spanned the last two decades in an effort to replace expensive fixed links with more flexible wireless connections. For example, mesh and multihop wireless backhaul architectures for IEEE 802.11 networks have been extensively studied [110, 111]. However, integrated solutions that provide both access and backhaul functionality have not yet been extensively adopted in the cellular domain. There is relay functionality built into the LTE specifications, but it has not been widely deployed due to its limited flexibility [111]: the resource configuration is fixed, it only supports single-hop relaying, and there is a fixed association between the relay and the parent base station that connects it to the wired core network. In contrast, the wireless backhaul links used to supplement fiber optic cables for backhauling traffic in sub-6 GHz cellular networks are typically point-to-point solutions that are not integrated with the RAN. In spite of this, wireless backhaul integration with radio access is viewed as a promising solution for 5G cellular networks. Preliminary results on wireless backhaul for 5G using mmWave links were presented in [112, 113] and demonstrated that such solutions can meet the anticipated increase in mobile traffic demands. However, they did not contemplate a close integration between the access and the backhaul, which is instead the focus of the more recent 3GPP SI on IAB for NR [2], recently finalized by the 3GPP. In this instance, the primary goal was to evaluate the viability of integrated access and wireless backhaul over NR (i.e., the 5G radio interface) and to propose potential solutions to assure efficient backhauling operations. This Study Item resulted in a Work Item and is anticipated to be incorporated into future 3GPP specification

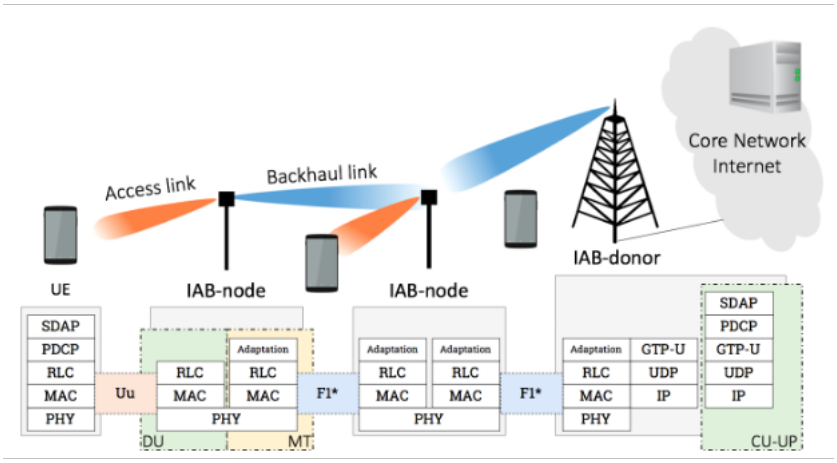


Figure 5.3: Protocol stack and basic architecture of an IAB network.

releases. The Study Item examined fixed wireless relays with both in-band and out-of-band backhauling capabilities, with a focus on the former, which is more difficult in terms of network design and management but maximizes spectrum utilization. Half-duplex relaying will be supported for the in-band scenario, although the 3GPP Specification does not rule out support for full-duplex relaying [2]. According to [2], IAB operations are spectrum agnostic, so relays can be deployed in a plug-and-play fashion in either the above-6 GHz or sub-6 GHz spectrum, and can operate in both SA (connected to the 5G core network) and NSA (connected to the 4G EPC) modes. (i) a Spanning Tree (ST) in which each IAB-node is connected to a single parent, and (ii) a Directed Acyclic Graph (DAG) in which each IAB-node may be connected to multiple upstream nodes. In addition, IAB relays will provide greater network deployment and configuration flexibility compared to LTE. As stated in [2, 114], 5G IAB relays will be used in both outdoor and indoor scenarios, with multiple wireless steps, to extend coverage, and should be able to reconfigure the topology autonomously to prevent service outages. In addition, a flexible allocation of access and transmission resources is envisioned in order to improve resource allocation efficiency.

5.4.3 Architecture

As depicted in Figure 5.3, the logical architecture of an IAB network consists of multiple IAB-nodes, which have wireless backhauling capabilities and can serve both UEs and other IAB-nodes, and IAB-donors, which have fiber connectivity to the core network and can serve both UEs and IAB-nodes. Initially, the Study Item proposed five distinct configuration options for the architecture, with varying degrees of decentralization of network capabilities

and backhauling solutions. The final version, which was chosen for future standardization, was favored due to its minimal impact on the core network specifications, lower relay complexity and processing requirements, and reduced signaling overhead. According to the selected architecture, each IAB-node hosts two NR functions: (i) a Mobile Termination (MT), used to maintain the wireless backhaul connection to an upstream IAB-node or IAB-donor, and (ii) a Distribution Unit (DU), used to provide access connection to UEs or the downstream MTs of other IAB-nodes. The DU connects to a CU hosted by the IAB-donor by means of the NR F1□ interface operating over the wireless backhaul link. In the access of IAB-nodes and donors, two interfaces coexist, namely the Uu interface (between the UEs and the DU of the gNBs) and the previously mentioned F1 interface. With this option, it is possible to exploit the functional separation of the radio protocol stack: the CU at the IAB-donor retains all control and upper layer functionality, while the lower layer operations are delegated to the DUs at the IAB-nodes. RRC, Service Data Adaptation Protocol (SDAP), and PDCP layers reside in the CU, whereas RLC, MAC, and PHY layers are hosted by the DUs. On top of RLC, an additional adaptation layer is added to route data across the IAB network topology, thereby permitting an end-to-end connection between DUs and the CU.

5.4.4 Network Procedures and Topology Management

In an IAB deployment, the establishment and management of the network topology are crucial factors to consider. This is due to the fact that the end-to-end performance of the entire network is highly dependent on the number of steps between the donor and the end relay, the number of relays the donor must support, and the strategies employed for network formation, route selection, and resource allocation. It is necessary to optimize the performance of various network procedures involving topology and resource management in order to ensure efficient IAB operations. During the IAB-node configuration, the topology is established, which is an essential phase. When an IAB-node becomes active, it chooses the upstream node to which it will attach. To achieve this, the MT executes the same initial access procedure as a user equipment (UE), i.e., it uses the synchronization signals conveyed by the available cells (officially referred to as synchronization signal block (SSB) in NR) to estimate the channel and select the parent. In addition, although not currently supported by the specifications, we argue that it would be advantageous if the MT could retrieve additional information (e.g., the number of hops to reach the donor, the cell load, etc.) and then choose the cell to attach to based on more advanced path selection metrics [115] than just the Received Signal Strength (RSS). The IAB-node then configures its DU, establishes the F1 connection to the

CU in the remote IAB-donor, and is prepared to offer services to UEs and other IAB-nodes. During this initial phase, the IAB-node may transmit its topological location within the IAB network to the IAB-donor. The topology management function then dynamically adapts the IAB topology in order to maintain service continuity (e.g., when a backhaul link is degraded or lost), or for load balancing purposes (e.g., to avoid congestion). The IAB-nodes may transmit periodic information about traffic volume and backhaul link quality in addition to the information provided during the initial setup procedure. This enables the CU to be aware of the overall IAB topology, determine the optimal configuration, and alter it by modifying network connectivity (i.e., the associations between IAB-nodes). It is also possible to provide increased redundancy and load balancing if the IAB-nodes support a DAG topology with multiple connections to multiple upstream nodes. In this instance, the CU manages the addition/removal of redundant channels based on propagation conditions and traffic demand for each wireless backhaul link.

5.4.5 Scheduling and Resource Multiplexing

The requirement to multiplex access and backhaul traffic within the same frequency band necessitates half-duplex operations for in-band IAB operations. This limitation is acknowledged in the 3GPP Study Item report [81], but full-duplex solutions are not ruled out. Consequently, the radio resources must be partitioned orthogonally between the access and the backhaul, either in time (Time Division Multiplexing (TDM), which is the preferred solution in [2]), frequency (Frequency Division Multiplexing (FDM)), or space (Space Division Multiplexing (SDM)), using a centralized or decentralized scheduling coordination mechanism across the IAB-nodes and the IAB-donor. Regardless of the limitations imposed by the half-duplex constraint, the IAB network must meet the access traffic needs of all users. Therefore, the available resources should be allocated equitably, taking into consideration channel measurements and topology-related data that may be exchanged between IAB-nodes. In addition, both hop-by-hop and end-to-end flow control mechanisms should be implemented to mitigate the risk of congestion on intermediate hops, which may occur under poor propagation conditions.

6

Research study of 'IAB': Safehaul: Risk-Averse Learning for Reliable mmWave Self-Backhauling in 6G Networks

Wireless backhauling at millimeter-wave frequencies (mmWave) in static scenarios is a well-established practice in cellular networks. However, highly directional and adaptive beamforming in today's mmWave systems have opened new possibilities for self-backhauling. Tapping into this potential, 3GPP has standardized IAB allowing the same base station to serve both access and backhaul traffic. Although much more cost-effective and flexible, resource allocation and path selection in IAB mmWave networks is a formidable task. To date, prior works have addressed this challenge through a plethora of classic optimization and learning methods, generally optimizing KPIs such as throughput, latency, and fairness, and little attention has been paid to the reliability of the KPI. We propose Safehaul, a risk-averse learning-based solution for IAB mmWave networks. In addition to optimizing average performance, Safehaul ensures reliability by minimizing the losses in the tail of the performance distribution. We develop a novel simulator and show via extensive simulations that Safehaul not only reduces the latency by up to 43.2% compared to the benchmarks, but also exhibits significantly more reliable performance, e.g., 71.4% less variance in latency.

6.1 Introduction

The emergence of mmWave cellular systems created a unique opportunity for Mobile Network Operators (MNOs) to leverage a scalable and cost-effective approach to deal with network densification. The fact that mmWave base stations can support fiber-like data rates facilitates the use of the same base station for both access and backhaul traffic, a solution which in 3GPP parlance is referred to as IAB. Consequently, 3GPP has included IAB in the standard [116, 80] covering the details on architecture, higher layer protocols, and the radio. Although Release 17 of 5G-NR defines the interfaces, architectures, and certain system parameters, the actual configuration and resource allocation is left to MNOs.

Traditional self-backhauled networks featured fixed-wireless links decoupled from access networks with static configurations. In contrast, IAB should account for the dynamic nature of the backhaul links (particularly in mmWave deployments) and their integration with the access network. Further, IAB allows the traffic to traverse several hops (i.e., base stations) to reach its destination, thus increasing the problem's complexity. *In addition to the scheduling problem, an IAB network should:* (i) *solve the problem of path selection and link activation at the backhaul while considering inter-cell interference and* (ii) *decide on serving access or backhaul traffic depending on the access load and the ingress backhaul traffic from neighboring base stations.*

Prior work. Methodologically, the majority of the existing works [117, 118, 119, 120, 121, 122, 123, 124, 125] focus on classic optimization techniques to solve the above-mentioned problem. However, given the large number of parameters involved, such formulations often result in non-convex problems that are too complex for real-time operations, but are nonetheless valuable indicators as performance upper bounds. Recently, some works focus on more practical solutions which can be deployed in real networks[126, 127, 128]. Specifically, these works leverage RL to tackle both resource allocation and/or path selection in IAB mmWave networks and demonstrate that RL-based solutions achieve real-time performance.

Regardless of the methodology, prior works mostly aim at maximizing the network capacity [117, 118, 119, 120, 121, 122, 123, 124], minimizing latency [129, 7] and improving throughput fairness [118, 130]. Although these approaches successfully improve the network performance, MNOs are often more *concerned about their reliability*. For this reason many commercial products rely on *simplified* but reliable algorithms for resource allocation, despite their sub-optimal performance. In this article, we propose Safehaul, a reinforcement learning-based solution for reliable scheduling and path selection in IAB mmWave systems

under network dynamics. We use the concept of risk aversion, commonly used in economics [6, 131], to measure and enhance the reliability of Safehaul. The following summarizes our contributions:

- We model the scheduling and path selection problem in IAB mmWave networks as a multi-agent multi-armed bandit problem (Section 6.3). We consider multiple fiber base stations, simultaneously supporting many self-backhauled mmWave base stations. In our model, the self-backhauled base stations independently decide the links to activate. The consensus among the base stations is reached via standard-defined procedures (Section 6.4.3).

- We present the first solution to provide reliable performance in IAB-enabled networks (Section 6.4). Specifically, we investigate the joint minimization of the average end-to-end latency and its expected tail loss. To this aim, we propose Safehaul, a learning approach that leverages the coherent risk measure CVaR[6]. CVaR measures the tail average of the end-to-end latency distribution that exceeds the maximum permitted latency, thus ensuring the network's reliability.

- We analytically bound the regret of Safehaul, i.e., we bound the loss of Safehaul compared to the case when the delays associated to all end-to-end paths between self-backhauled base stations and fiber base stations are known a priori. We show that, for the case when there are no conflicts between the decision of the self-backhauled base stations, the average regret of Safehaul tends to zero as the time increases. This regret bound characterizes the learning speed and proves that Safehaul converges to the optimal scheduling and path selection solution that jointly minimizes the average end-to-end latency and its expected tail loss.

- We provide a new means of simulating multi-hop IAB networks by extending NVIDA's GPU-accelerated simulator Sionna [1] (Section 6.5). Specifically, we add codebook-based analog beamforming capabilities for both uplink and downlink communications. Further, we extend Sionna by implementing system-level components such as layer-2 schedulers and buffers and BAP-like routing across the IAB network. We believe our IAB extensions will be instrumental for the open-source evaluation of future research on self-backhauled mmWave networks.

- Exploiting the above simulator, we evaluate and benchmark Safehaul against two state-of-the-art algorithms [7, 8] based on deployment in two different locations (Manhat-

tan and Padova). The results confirm that Safehaul is significantly more reliable than the considered benchmarks, as it exhibits much tighter variance in terms of both latency (up to 71.4%) and packet drop rate (at least 39.1%). Further, Safehaul achieves up to 43.2% lower average latency and 11.7% higher average throughput than the reference schemes.

6.2 System Model

We consider a cellular system with N base stations capable of self-backhauling and D base station with a fiber connection to the core network. Following 3GPP terminology, we refer to self-backhauled base stations as IAB-nodes (BS-nodes) and the fiber base stations as IAB-donors (BS-donors)*. Each BS-node connects to the core network via a (multi-hop) wireless link to a BS-donor. The sets of all BS-nodes and BS-donors are denoted by $\mathcal{N} = \{1, \ldots, N\}$ and $\mathcal{D} = \{1, \ldots, D\}$, respectively. The system works in a time-slotted fashion starting from time slot $i = 1$ until a finite time horizon I. All the time slots $i = 1, \ldots, I$ have the same duration. The BS-nodes are equipped with two RF chains. One RF chain is used exclusively for the communication with cellular users (access network), while the second RF chain is used for self-backhauling. In line with the 3GPP specification [114], we assume half-duplex self-backhauling, i.e., in each time slot i a BS-node can either transmit, receive or remain idle.

We model the connections between the base stations in slot i as a graph $\mathcal{G}_i = \{\mathcal{V}, \mathcal{E}_i\}$, see Fig. 6.1. The set $\mathcal{V} = \mathcal{N} \cup \mathcal{D}$ of vertices is formed by all the BS-nodes and BS-donors in the system. The set $\mathcal{E}_i$ of edges is composed of the available wireless links (n, l) between a BS-node $n \in \mathcal{N}$ and any BS (BS-donor or BS-node) $l \in \mathcal{V}$ in time slot i. Note that $\mathcal{G}_i$ is not static. In a given time slot i, some links may be unavailable due to failure, blockage, or interference. Thus, only feasible wireless links are considered in the set $\mathcal{E}_i$. The path $X_{n,d}$ from BS-node n to any BS-donor d is a sequence of intermediate links (n, l). $X_{n,d}$ changes over time according to the traffic loads of the intermediate BS-nodes and to the channel conditions. We model the activation of link (n, l) with the binary variable $x_{n,l,i}$. When $x_{n,l,i} = 1$, the link is activated and BS-node n transmits to BS $l \in \mathcal{V}$ in time slot i. $x_{n,l,i} = 0$ indicates that the link is deactivated.

Each BS-node n has a finite data buffer with capacity $B_n^{\max}$ to store the backhaul data to be transmitted to any of the BS-donors. In each time slot i, BS-node n is characterized by

*Please note that throughout the paper we will use interchangeably BS-nodes and IAB-nodes (and similarly for BS-donors and IAB-donors)"

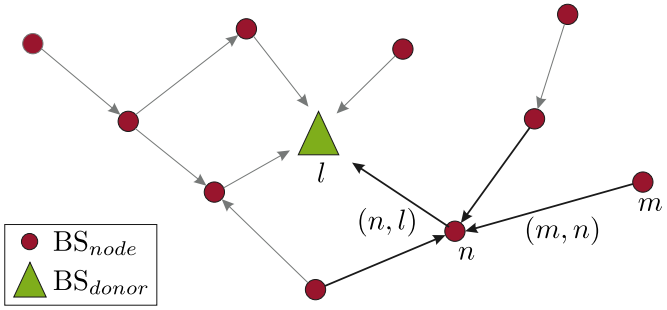


Figure 6.1: Example of a graph $\mathcal{G}_i$

its load and average queuing time. The load, denoted by $B_{n,i} \in \mathbb{N}$, indicates the number of data packets stored in the buffer at the beginning of time slot i. The average queuing time $t^{\mathrm{q}}_{n,i} \in \mathbb{R}^+$ is the average number of time slots the current packets in the data buffer have been stored. Additionally, we denote by $t^{\mathrm{tx}}_{n,l,i} \in \mathbb{R}^+$ and $M_{n,l,i} \in \mathbb{R}^+$ the transmission time from BS-node n to BS l in time slot i, and the amount of successfully transmitted data on the link, respectively. We define the maximum tolerable latency $T_{\max}$ as the maximum time a packet can take from its source BS-node to any BS-donor. Any packet that is not delivered before $T_{\max}$ milliseconds will be dropped. The average maximum end-to-end latency $\bar{T}_{n,d}$ from BS-node n to BS-donor d is the average, over the complete time horizon I, of the maximum delay a packet originating from BS-node n takes to reach any BS-donor d in time slot i. This is calculated as $\bar{T}_{n,d} = \frac{1}{I}\sum_{i=1}^{I} T_{n,d,i}$, where $T_{n,d,i}$ is the maximum end-to-end latency among all the packets originating in BS-node n which reach BS-donor d in time slot i. $T_{n,d,i}$ is a sample of the random variable $T_{n,d}$ drawn from an unknown stationary probability distribution P that depends on the links $x_{n,l,i'}$, $n \in \mathcal{N}$, $l \in \mathcal{V}$, $i' = 1, \ldots, i$, activated up to time i, the user's mobility, the location of the BS-node n, the interference in the system, and the queue dynamics. Accordingly, we define the average maximum end-to-end latency in the system $\bar{T}$ as

$$\bar{T} = \frac{1}{ND}\sum_{n=1}^{N}\sum_{d=1}^{D}\bar{T}_{n,d}. \tag{6.1}$$

6.3 Problem Formulation

The joint minimization of the average maximum end-to-end latency and the expected value of its tail loss in IAB-enabled networks is formulated in this section. We first introduce CVaR, the risk metric accounting for minimizing the events in which the end-to-end latency is higher than $T_{\max}$. Next, we formulate the optimization problem in the complete network.

6.3.1 Preliminaries on CVaR

Traditionally, latency minimization in IAB-enabled networks has focused on optimizing the expected value of a latency function [129, 7]. However, such an approach fails to capture the time variability of the latency distribution, thus leading to unreliable systems in which $T_{n,d,i} > T_{\max}$, for any $i = 1, ..., I$, $n \in \mathcal{N}$ and $d \in \mathcal{D}$. *For this purpose, we consider not only the average end-to-end latency $\bar{T}$ in the system, but also its expected tail loss based on the CVaR*[6].

Having in mind that $T_{n,d}$ is a random variable, we assume it has a bounded mean on a probability space $(\Omega, \mathcal{F}, P)$, with Ω and $\mathcal{F}$ being the sample and event space, respectively. Using a risk level $\alpha \in (0, 1]$, the $\mathrm{CVaR}_\alpha(T_{n,d})$ of $T_{n,d}$ at risk level α quantifies the losses that might be encountered in the α-tail. More specifically, it is the expected value of $T_{n,d}$ in its α-tail distribution [**?**]. Formally, $\mathrm{CVaR}_\alpha(T_{n,d})$ is defined as [6]

$$\mathrm{CVaR}_\alpha(T_{n,d}) = \min_{q \in \mathbb{R}} \left\{ q + \frac{1}{\alpha} \mathbb{E}\left[\max\{T_{n,d} - q, 0\}\right] \right\}, \tag{6.2}$$

where the expectation in (6.2) is taken over the probability distribution P. Note that lower $\mathrm{CVaR}_\alpha(T_{n,d})$ results in higher system reliability because the expected end-to-end latency in the α-worst cases is low. Moreover, note that α is a risk aversion parameter. For $\alpha = 1$, $\mathrm{CVaR}_\alpha(T_{n,d}) = \mathbb{E}[T_{n,d}]$ which represents the traditional risk-neutral case. Conversely, $\lim_{\alpha \to 0} \mathrm{CVaR}_\alpha(T_{n,d}) = \sup\{T_{n,d}\}$. CVaR has been shown to be a coherent risk measure, i.e., it fulfills monotonicity, subadditivity, translation invariance, and positive homogeneity properties [132].

6.3.2 Optimization Problem

We jointly minimize the average maximum end-to-end latency and its expected tail loss for each BS-node. For this purpose, we decide which of the (n, l) links to activate in each time slot i during the finite time horizon I. In the following, we formulate the optimization problem from the network perspective and consider the sum over all BS-nodes in the system. The latency minimization problem should consider three different aspects: (i) link activation is constrained by the half-duplex nature of self-backhauling, (ii) only data stored in the data buffers can be transmitted, and (iii) packet drop due to buffer overflow should be avoided. Formally, the problem is written as:

$$\underset{\{x_{n,l,i}\}}{\text{minimize}} \quad \sum_{n\in\mathcal{N}}\left(\sum_{d\in\mathcal{D}}\left(\frac{1}{I}\sum_{i=1}^{I}T_{n,d,i}\right)+\eta\text{CVaR}_{\alpha}(T_{n,f})\right) \tag{6.3a}$$

subject to

$$\sum_{l\in\mathcal{V},l\neq n}x_{n,l,i}+\sum_{l\in\mathcal{N}}x_{l,n,i}=1,\ \ n\in\mathcal{N},\ i=1,\dots,I, \tag{6.3b}$$

$$\sum_{j=1}^{i}B_{n,i}\geq\sum_{j=1}^{i}M_{n,l,i,} \qquad n\in\mathcal{N},l\in\mathcal{V},\ i=1,\dots,I, \tag{6.3c}$$

$$\sum_{j=1}^{i}B_{l,j}+\sum_{j=1}^{i}M_{n,l,j}\leq B_{l}^{\max},n\in\mathcal{N},l\in\mathcal{V},\ i=1,\dots,I, \tag{6.3d}$$

$$x_{n,l,i}\in\{0,1\}, \qquad n\in\mathcal{N},l\in\mathcal{V},\ i=1,\dots,I. \tag{6.3e}$$

In (6.3a), $\eta\in[0,1]$ is a weighting parameter to control the trade-off between minimizing the average maximum end-to-end latency $\bar{T}_{n,d}$ and the expected loss of its tail. As the considered scenario is not static, solving (6.3) would require complete non-causal knowledge of the system dynamics during the complete time horizon I. However, in practical scenarios, knowledge about the underlying random processes is not available in advance. For example, the BS-node's loads $B_{n,i}$ depend not only on the transmitted and received backhaul data, but also on the randomly arriving data from its users. Similarly, the amounts of transmitted data $M_{n,l,i}$ depend on the varying channel conditions of both BS n and l. As a result, the exact values of $T_{n,l,i}$, $B_{n,i}$ and $M_{n,l,i}$ are not known beforehand. For this reason, we present in Sec. 6.4 Safehaul, a multi-agent learning approach to minimize in each BS-node the average maximum end-to-end latency and the expected value of the tail of its loss.

6.4 Our proposed solution: Safehaul

In this section, we describe Safehaul, a multi-agent learning approach for the joint minimization of the average maximum end-to-end latency and its expected tail loss in IAB mmWave networks. In Safehaul, each BS-node independently decides which links (n,l) to activate in every time slot i by leveraging a multi-armed bandit formulation. The consensus among the BS-nodes is reached by exploiting the centralized resource coordination and topology management role of IAB-donors [116, Sec. 4.7.1].

6.4.1 Multi-Armed Bandit Formulation

Multi-armed bandit is a tool well suited to problems in which an agent makes sequential decisions in an unknown environment[133]. In our scenario, each BS-node n decides, in each time slot i, which of the links (n, l) to activate without requiring prior knowledge about the system dynamics. The multi-armed bandit problem at BS-node n can be characterized by a set $\mathcal{A}_n$ of actions and a set $\mathcal{R}_n$ of possible rewards. The rewards $r_{n,i} \in \mathcal{R}_n$ are obtained in each time slot i as a response to the selected action $a_{n,i} \in \mathcal{A}_n$. The actions are the links that BS-node n can activate, and the rewards are a function of the observed latency. We define $\mathcal{A}_n$ as $\mathcal{A}_n = \{(n, l), (m, n) | n, m \in \mathcal{N},\ l \in \mathcal{V}\}$, where link (n, n) indicates that BS-node n remains idle. As blockages, overloads, or failures might render certain links (n, l) temporarily unavailable, we define the set $\mathcal{A}_{n,i} \subseteq \mathcal{A}_n$ of available actions in time slot i as $\mathcal{A}_{n,i} = \{(n, l), (l, n) | (n, l), (l, n) \in \mathcal{E}_i\}$. Selecting action $a_i = (n, l)$ in time slot i implies $x_{n,l,i} = 1$.

The rewards $r_{n,i}$ are a function of the end-to-end latencies $T_{n,d,i}$ and depend on whether at BS-node n a link (n, l) or (l, n) is activated. BS-node n is connected to the BS-donor via multi-hop wireless links. Consequently, $T_{n,d,i}$ cannot be immediately observed when a link (n, l), with $l \notin \mathcal{D}$ is activated. In fact, the destination BS-donor d might not even be known to BS-node n in time slot i. To overcome this limitation, we define the rewards $r_{n,i}$ as a function of the next-hop's estimated end-to-end latency $\hat{T}_{l,d,i}$ as

$$r_{n,i} = \begin{cases} t^{\mathrm{q}}_{l,i} + t^{\mathrm{tx}}_{n,l,i} + \hat{T}_{l,d,i}, & \text{for link } (n, l) \\ t^{\mathrm{q}}_{n,i} + \hat{T}_{n,d,i}, & \text{for link } (l, n), \end{cases} \tag{6.4}$$

where $\hat{T}_{l,d,i}$ is calculated as $\hat{T}_{l,d,i} = \min_{(l,m) \in \mathcal{E}_i} \hat{T}_{l,m,i}$.

6.4.2 Latency and CVaR Estimation

As given in (6.4), BS-node n learns which links (n, l) to activate by building estimates of the expected latency $\hat{T}_{n,l}$ associated to each of them. Let $K_{n,l,i} = \sum_{j=1}^{i} x_{n,l,i}$ be the number of times link (n, l) has been activated up to time slot i. The estimated $\hat{T}_{n,l}$ is updated using the sample mean as

$$\hat{T}_{n,l,i+1} = \frac{K_{n,l,i} \hat{T}_{n,l,i} + r_{n,i}}{K_{n,l,i} + 1}, \tag{6.5}$$

where the subindex i is introduced to emphasize that the estimate is built over time.

The CVaR definition given in (6.2) requires $T_{n,d}$ which, as discussed before, is known a priori. Hence, we leverage the CVaR estimator derived in [134] to calculate the estimated CVaR of a link (n, l). Let $\tilde{r}_n^1, \ldots, \tilde{r}_n^{K_{n,l,i}}$ be all the rewards received up to time i. The estimated $\widehat{\mathrm{CVaR}}_i(n, l)$ in time slot i is calculated as [134]

$$\widehat{\mathrm{CVaR}}_i(n,l) := \inf_{t \in \mathbb{R}} \left(t + \frac{1}{\alpha \cdot K_{n,l,i}} \sum_{k=1}^{K_{n,l,i}} [\tilde{r}_n^k - t]^+ \right). \tag{6.6}$$

Using the estimates in (6.5) and (6.6), BS-node computes the value $Q_n(a_{n,i} = (n, l))$ associated to the selected action $a_n \in \mathcal{A}_n$, and defined as

$$Q_n(a_{n,i}) = \hat{T}_{n.l,i} + \eta \widehat{\mathrm{CVaR}}_i(n, l). \tag{6.7}$$

Note that (6.7) is aligned with the objective function in (6.3a). Actions with an associated low value $Q_n(a_{n,i})$ lead to lower end-to-end latency and a low expected value on its tail.

6.4.3 Consensus

All the BS-nodes independently decide which links to activate based on their estimates of the end-to-end latency. As a consequence, conflicting actions may be encountered. A conflict occurs when two or more BS-nodes n and m aim at activating a link to a common BS $l, l \in \mathcal{V}$, i.e., $x_{n,l,i} = x_{m,l,i} = 1$. We reach consensus by first retrieving the buffer and congestion status of the various IAB-nodes, leveraging the related BAP layer functionality [116, Sec. 4.7.3]. With this information at hand, conflicts are resolved by prioritizing the transmission of the BS-node with the larger queuing times $t_{n,i}^{\mathrm{q}}$ and loads $B_{n,i}$. Then, we let the IAB-donor mark as *unavailable* the time resources of the remaining base stations with conflicting scheduling decisions [116, Sec. 10.9]. Note that as the learning is performed at each BS-node, only the link activation decision and the weighted sum of $t_{n,i}^{\mathrm{q}}$ and $B_{n,i}$ are transmitted. Thus, low communication overhead is achieved.

6.4.4 Implementation of Safehaul

Here, we describe how the above-mentioned solution can be implemented in a real system. Specifically, we elaborate on the required inputs and the interactions among the different entities as well as the pseudo-code of Safehaul, see Alg. 6.1.

Algorithm 6.1 Safehaul algorithm at each BS-node

Require: α, η, $\mathcal{A}_n$

1: Initialize $\hat{T}_{n,l}$, $\widehat{\mathrm{CVaR}}(n,l)$, and Q_n for all $(n,l) \in \mathcal{E}_1$
2: Set counters $K_{n,l} = 0$ and initial action $a_{n,1} = (n,n)$
3: **for** every time slot $i = 1, ..., I$
4: perform action $a_{n,i}$, observe reward $r_{n,i}$ and increase counter $K_{n,l}$ by one ▷ Eq. (6.4)
5: update latency estimate $\hat{T}_{n,l}$ ▷ Eq: (6.5)
6: update CVaR estimate $\widehat{\mathrm{CVaR}}(n,l)$ ▷ Eq: (6.6)
7: update $Q_n(a_{n,i})$ ▷ Eq: (6.7)
8: select next action $a_{n,i+1}$ using ϵ-greedy ▷ Eq. (6.8)
9: share $a_{n,i+1}$, $t^{\mathrm{q}}_{n,i}$ and $B_{n,i}$ with the other BS-nodes
10: if required, update $a_{n,i+1}$ to reach consensus ▷ Sec. 6.4.3

Safehaul is executed at each BS-node n. For its implementation, the MNO provides α, η and $\mathcal{A}_n$ as an input. α is the risk level parameter that influences the level of reliability achieved in the system. Similarly, η controls the impact the minimization of the latency in the α-worst cases has on the overall performance. Both parameters, α and η, are set by the MNO depending on its own reliability requirements. The set $\mathcal{A}_n$ depends on the considered network topology, which is perfectly known by the MNO. The set $\mathcal{A}_n$ includes all links (n,l) and (l,n) to and from the first-hop neighbors of BS-node n.

The execution of Safehaul begins with the initialization of the latency and CVaR estimates, and the values Q of the actions in $\mathcal{A}_n$. Additionally, the counters $K_{n,l}$, that support the calculations of $\hat{T}_{n,l}$ and $\widehat{\mathrm{CVaR}}(n,l)$, are initialized for all links in $\mathcal{A}_n$ (lines 1-2). These parameters are updated and learnt throughout the execution of Safehaul. At time slot $t = 0$, no transmission has occurred and $B_{n,0} = 0$. Hence, BS-node n remains idle for the first time slot $i = 1$, i.e., $a_{n,1} = (n,n)$ (line 2). Next, and in each of the subsequent time slots $i \in \{1, \ldots, I\}$, the selected action is performed and the corresponding reward is obtained (line 4). If BS-node n transmits in time slot i, i.e., $a_{n,i} = (n,l)$, the reward $r_{n,i}$ is sent by the receiving BS l through the control channel. If $a_{n,i} = (l,n)$, the reward $r_{n,i}$ depends, as given in (6.4), only on the current estimates at BS-node n and the status of its buffer $B_{n,i}$. With the observed reward $r_{n,i}$, the counter for action $a_{n,i}$ is increased and the latency and CVaR estimates are updated (lines 4-6). Using the new estimates (lines 5 and 6), the value $Q(a_{n,i})$ of the performed action $a_{n,i}$ is updated (line 7). The next action $a_{n,i+1}$ is then selected according to ϵ-greedy (line 8), which is a well-known method to balance the exploitation of links with estimated low latency, and the exploration of unknown but potentially better ones. In ϵ-greedy, a random action $a_{n,i+1}$ from the set $\mathcal{A}_{n,i+1}$ is selected with probability $\epsilon \in [0,1]$. With probability $(1-\epsilon)$, instead, the action that yields the estimated lowest value is chosen,

i.e.,

$$a_{n,i+1} = \begin{cases} \text{randomly selected action from } \mathcal{A}_{n,i+1}, & \text{if } x \leq \epsilon \\ \underset{b_n \in \mathcal{A}_{n,i+1}}{\operatorname{argmax}} Q_n(b_n), & \text{if } x > \epsilon, \end{cases} \tag{6.8}$$

where x is a sample taken from a uniform distribution in the interval $[0, 1]$. Once the action $a_{n,i+1}$ is selected, it is shared with other BS-nodes in the network along with $t^{\mathrm{q}}_{n,i}$ and $B_{n,i}$ (line 9). As described in Section 6.4.3, this goes through the control channel. If conflicts arise, consensus is reached by prioritizing the transmission of the BS-node with the largest loads and queuing times (line 10).

6.4.5 Regret Analysis

The regret ζ is defined as the expected loss caused by the fact that the optimal action is not always selected [135]. Let $\bar{T}^*$ and $\bar{T}_{a_n}$ be the expected delay associated to the optimal action $a^* \in \mathcal{A}_n$ and the non-optimal action $a_n \in \mathcal{A}_n$, respectively. Similarly, let CVaR^* and CVaR_{a_n} be the CVaR of the optimal action $a^* \in \mathcal{A}_n$ and the non-optimal action $a_n \in \mathcal{A}_n$, respectively. Formally, the regret ζ_i after i time slots is defined as

$$\begin{aligned} \zeta_i &= \sum_{a_n \in \mathcal{A}_n} \left(\left(\bar{T}^* + \eta \mathrm{CVaR}^* \right) - \left(\bar{T}_{a_n} + \eta \mathrm{CVaR}_{a_n} \right) \right) \mathbb{E}[K_{a_n,i}] \\ &= \sum_{a_n \in \mathcal{A}_n} \Delta_{a_n} \mathbb{E}[K_{a_n,i}], \end{aligned} \tag{6.9}$$

where $K_{a_n,i}$ is the number of times action a_n has been selected up to time slot i.

Proposition 1. *For a network $\mathcal{G}$ in which the independent decisions of the BS-nodes do not lead to conflicts, let $A_n = |\mathcal{A}_n|$ be the number of available actions for BS-node n. Additionally, let $c > 0$, $0 < d \leq 1$, and $\epsilon_i := \min(1, \frac{cA_n}{d^2 i})$. Then, there exists a positive constant $C > 1$, such that the probability that Safehaul chooses a non-optimal action a_n after $i \geq cA_n/d$ time slots is upper bounded as*

$$\begin{aligned} \mathbb{P}[a_{n,i} = a_n] \leq & \frac{c}{d^2 i} + \frac{4e}{d^2} B_i^{\frac{c}{2}} + \frac{2Cd^2}{c \ln\left(\frac{(i-1)d^2 e^{0.5}}{cA_n}\right)} \\ & + 4C \left(\frac{c}{d^2} \ln\left(\frac{(i-1)d^2 e^{0.5}}{cA_n} \right) \right) B_i^{\frac{c}{5d^2}}, \end{aligned}$$

with $B_i = \frac{cA_n}{(i-1)d^2 e^{0.5}}$.

Proof. See the Appendix (Sec. 6.6.6). □

Theorem 1. *For a network $\mathcal{G}$ in which the independent decisions of the BS-nodes do not lead to conflicts, the regret ζ_i of Safehaul after i time slots is upper bounded by*

$$\zeta_i \leq \sum_{a_n \in \mathcal{A}_n} \Delta_{a_n} \left(1 + \sum_{i'=2}^{i} \left[\frac{c}{d^2 i'} + \frac{4e}{d^2} B_{i'}^{\frac{c}{2}} + \frac{2Cd^2}{c \ln\left(\frac{(i'-1)d^2 e^{0.5}}{cA_n}\right)} 4C \left(\frac{c}{d^2} \ln\left(\frac{(i'-1)d^2 e^{0.5}}{cA_n} \right) \right) B_{i'}^{\frac{c}{5d^2}} \right] \right),$$

where $c > 0$ and $0 < d \leq 1$. The leading order of the regret ζ_i is $O\left(Cid^2 \left(c \ln\left(\frac{(i-1)d^2 e^{0.5}}{cA_n}\right)\right)^{-1} \right)$.

Proof. From the definition in (6.9), the regret can be upper bounded as

$$\zeta_i \leq \sum_{a_n \in \mathcal{A}_n} \Delta_{a_n} \left(1 + \sum_{i'=2}^{i} \mathbb{P}[a_{n,i'} = a_n] \right), \tag{6.10}$$

by considering that $\mathbb{E}[K_{a_n,i}] \leq 1 + \sum_{i'=2}^{i} \mathbb{P}[a_{n,i'} = a_n]$. The bound is obtained by including the result of Proposition 1 in (6.10) as

$$\zeta_i \leq \sum_{a_n \in \mathcal{A}_n} \Delta_{a_n} \left(1 + \sum_{i'=2}^{i} \left[\frac{c}{d^2 i'} + \frac{4e}{d^2} B_{i'}^{\frac{c}{2}} + \frac{2Cd^2}{c \ln\left(\frac{(i'-1)d^2 e^{0.5}}{cA_n}\right)} 4C \left(\frac{c}{d^2} \ln\left(\frac{(i'-1)d^2 e^{0.5}}{cA_n} \right) \right) B_{i'}^{\frac{c}{5d^2}} \right] \right), \tag{6.11}$$

It is now easy to see that the leading order is given by the fourth summand in the brackets, i.e.,

$$O\left(\frac{Cid^2}{c \ln\left(\frac{(i-1)d^2 e^{0.5}}{cA_n}\right)} \right), \tag{6.12}$$

which means the average regret $\zeta_i / i \to 0$ when $i \to \infty$. □

6.5 Simulation setup

Given the lack of access to actual 5G (and beyond) network deployments, prior works mostly rely on *home-grown* simulators for performance evaluation. Although this is a valid approach,

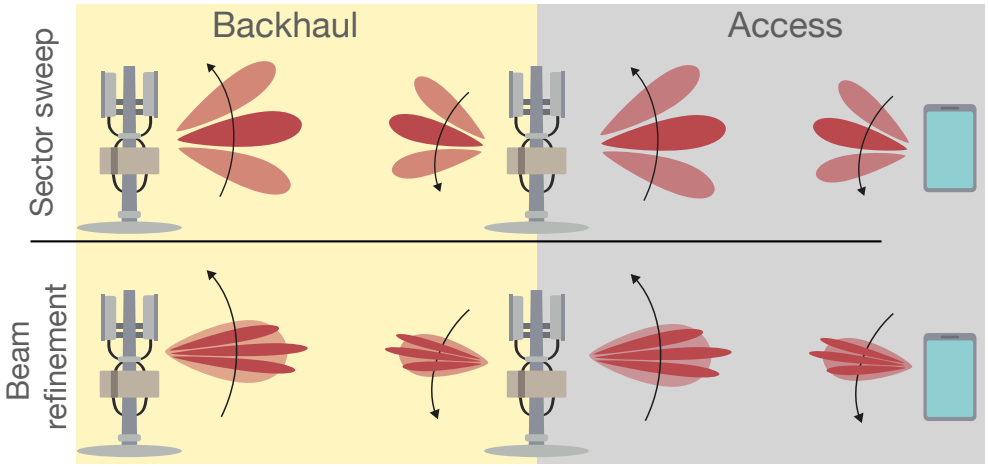


Figure 6.2: Schematic of the hierarchical beam management procedure. First, the general direction is estimated using wide beams (top). Then, the search is refined using the narrow beams codebook.

these simulators often cannot fully capture the real network dynamics, introducing strong assumptions in the physical and/or the upper layers of the protocol stack. Until very recently, the most complete simulator for IAB networks was a system-level simulator [136] developed as an extension of the ns-3 *mmWave* module [137]. However, despite accurate modeling of the IAB protocol stack, it is currently behind the latest IAB specifications[†]. Moreover, the ns-3 IAB extension is unsuitable for large simulations with hundreds of nodes due to reliance on an older version of the *mmWave* module. Therefore, in our work we opt for Sionna [1], which is an open-source GPU-accelerated toolkit based on TensorFlow.

However, unlike the aforementioned ns-3 module, Sionna is a physical layer-focused simulator that does not explicitly model 5G networks, thus lacking the characterization of the 5G-NR upper-layer protocol stack. Hence, we extend Sionna by including the system-level functionalities such as MAC-level scheduling and RLC-level buffering. Furthermore, since Sionna exhibits slight differences compared to the 5G-NR physical layer, we extend Sionna's physical layer model [1] with the 5G-NR procedures In the following, we describe the details of our extensions, which we make publicly available[‡].

6.5.1 Extensions to Sionna's physical layer module

In this section, we describe the physical layer modification that were necessary to evaluate IAB scenarios using Sionna.

[†]For instance due to the assumption of L-3 (instead of L-2) relaying at the IAB-nodes which was based on a draft version of the TR 38.874 [114].

[‡]https://github.com/TUDA-wise/safehaul_infocom2023

Codebook-based Beamforming

Sionna's native beamforming only supports Zero-Forcing (ZF) pre-coding in downlink. Therefore, as a first step, we extend Sionna by implementing an NR-like codebook-based analog beamforming both at the transmitter and at the receiver. Specifically, we assume that the beamforming vectors at the transmitter w_{tx} and at the receiver w_{rx} are a pair of codewords selected from a predefined codebook. The codebook is computed by defining a set of beam directions $\{\omega_{n,m}\}$ which scans a given angular sector with a fixed beamwidth. The steering vector $a_{n,m}$ corresponding to direction $\omega_{n,m}$ can be computed as:

$$a_{n,m} = \Big[1, \ldots, e^{j\frac{2\pi}{\lambda}d(i_H \sin\alpha_p \sin\beta_q + i_V \cos\beta_q)}, \ldots, e^{j\frac{2\pi}{\lambda}d((N_H-1)\sin\alpha_p \sin\beta_q + (N_V-1)\cos\beta_q)}\Big]^T, \tag{6.13}$$

where N_H and N_V are the number of horizontal and vertical antenna elements, respectively. The horizontal and vertical index of a radiating element is denoted by $i_H \in \{0, \ldots, N_H - 1\}$ and $i_V \in \{0, \ldots, N_V - 1\}$, respectively. α_p and β_q represent the azimuth and elevation angles of $\omega_{p,q}$. Next, we define the codebook as the set $\{\left(\sqrt{N_H N_V}\right)^{-1} w_{p,q}\}$.

In line with the 5G-NR beam management procedure [138], we assume the lack of complete channel knowledge, i.e., the communication endpoints do not know the corresponding channel matrix. Accordingly, an exhaustive search is conducted to identify the best pair of codewords resulting in the highest SINR. We leverage a hierarchical search, in which the communication pairs first perform a wide-beam search in which the transmitter and the receiver approximate the direction of communication, see Fig. 6.2. Next, the beamforming direction is fine-tuned through a beam refinement procedure going through a codebook with narrow beams. Consequently, we employ two types of codebooks, one with wide beams for sector sweep and another with narrow beams for beam refinement.

SINR Computations

Since Sionna does not natively calculate the SINR, we add this functionality to the simulator to better model the impact of interference in our simulations. We compute the SINR experienced by Transport Blocks (TBs) by combining the power of the intended signal with that of the interferers and of the thermal noise. Specifically, we first compute the power P_n of the intended signal at receiver n over frequency f and at time slot i. Then, we obtain the overall interference power by leveraging the superposition principle and summing the received

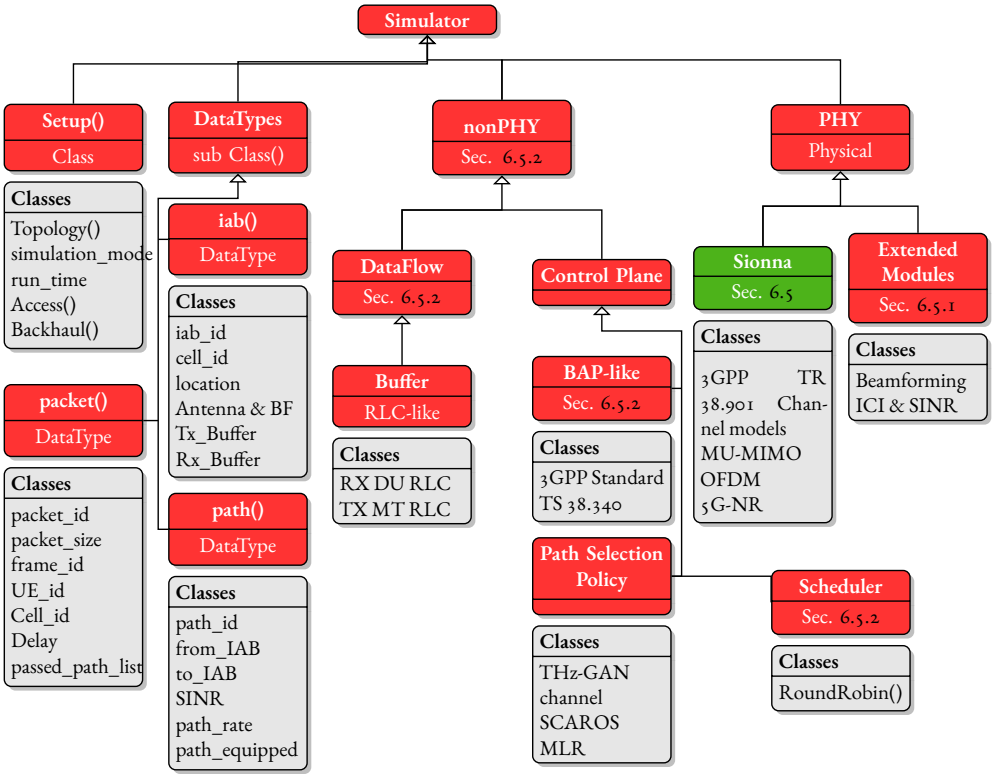


Figure 6.3: Overall design of our Sionna's extension. The red blocks represent our additions to the baseline simulator, i.e., Sionna [1].

power from all other interfering base stations $P_m(i,\ f)$ where $m \neq n$. For the purposes of this computation, we assume that each interferer employs the beamforming vector yielding the highest Signal-to-Noise Ratio (SNR) towards its intended destination. Similarly, the transmitter and the receiver use the beamforming configuration estimated via the hierarchical search procedure. Finally, the SINR is $\gamma_n(i,\ f) = \frac{P_n(i,f)}{\sum\limits_{m \neq n} P_m(i,f) + \sigma(i,f)}$ where $\sigma(t,\ f)$ is the thermal noise at the receiver.

6.5.2 System-level extensions to Sionna

As mentioned, Sionna is mainly a physical layer simulator. However, to get closer to IAB networks as specified in Rel. 17, we have extended Sionna by implementing a selection of system-level features. To such end, we introduced a discrete-event network simulator for modeling IAB networks. This system-level extension operates on top of Sionna and provides basic functionalities such as a MAC-level scheduler, layer-2 buffers, and data flow and path selection mechanisms. Our simulator, as depicted in Fig. 6.3, generates a variety of system-

level KPIs such as latency, throughput, and packet drop rate.

Data Flow and buffer

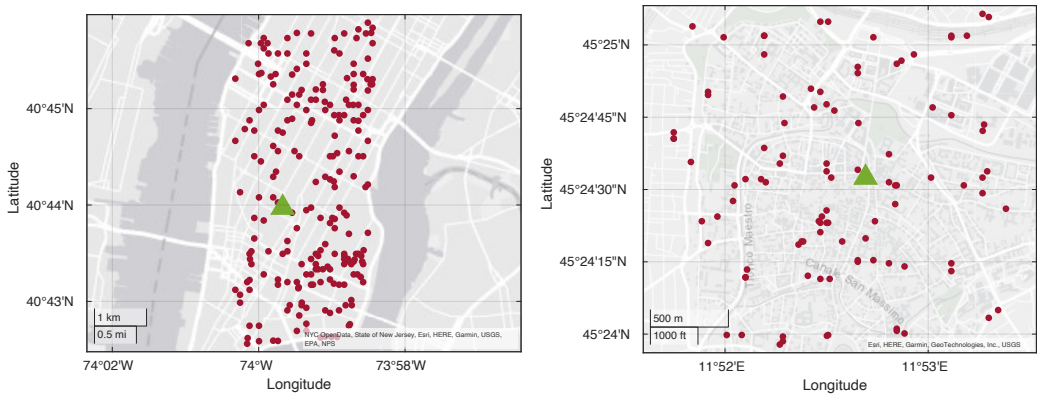
3GPP has opted for a layer-2 relaying architecture for BS-nodes where hop-by-hop Radio Link Control (RLC) channels are established. This enables retransmissions to take place just over the afflicted hops, thus preventing the need for traversing again the whole route from the BS-donor whenever a physical layer TB is not decoded. This design results in a more efficient recovery from transmission failures and reduces buffering at the communication endpoints [139]. To mimic this architecture, we have implemented RLC-like buffers at each base station. Specifically, each BS-node features layer-2 buffers for both received and transmitted packets. For instance, the data flow for an uplink packet is the following. The UE generates packets and sends a transmission request to the base station. Consequently, the scheduler allocates OFDM symbols for this transmission, which is eventually received and stored at the RX buffer of its Distributed Unit (DU). Next, the packet is placed into the TX buffer to be forwarded to the suitable next hop BS-node. This procedure is repeated until the packet crosses all the wireless-backhaul hops and reaches the BS-donor. Note that the packet can be dropped due to latency constraints or to interference.

BAP

To manage routing within the wireless-backhauled network, the 3GPP introduced BAP, i.e., an adaptation layer above RLC which is responsible for packet forwarding between the BS-donor and the access BS-nodes [81]. Our simulator mimics this by associating each BS-node to a unique BAP ID. Moreover, we append a BAP routing ID to each packet at its entry point in the Radio Access Network (RAN) (i.e., the BS-donorand the UEs for DL and UL data, respectively). Then, this identifier is used to discern the (possibly multiple) routes toward the packet's intended destination [81]. The choice of the specific route is managed by Safehaul.

Scheduler

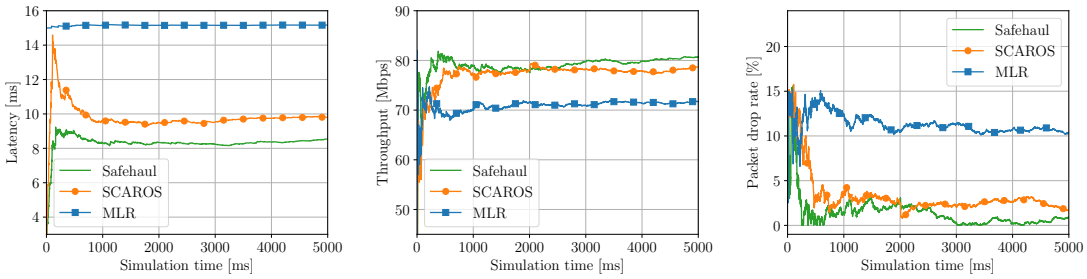
Finally, we implemented a MAC-level scheduler which operates in a Time Division Multiple Access (TDMA) mode. The scheduler periodically allocates the time resources to backhaul or access transmissions in a Round-Robin fashion. Specifically, each cell first estimates the number of OFDM symbols needed by each data flow by examining the corresponding buffer. Then, the subframe's OFDM symbols are equally allocated to the users. If a user requires



(a) Locations of the 223 BS-nodes and BS-donors in Manhattan, NYC.

(b) Locations of the 100 BS-nodes and BS-donors in Padova.

Figure 6.4: Location of BS-nodes in Manhattan and Padova topologies.



(a) Average per-UE end-to-end latency

(b) Average per-UE throughput

(c) Average per-UE packet drop rate

Figure 6.5: Average network performance for 50 UEs and 80 Mbps per-UE source rate (Scenario 1).

fewer symbols to transmit its complete buffer, the excess symbols (the difference between the available slot length and the needed slot length) are dispersed to the other active users.

6.6 Performance Evaluation

In our simulations, we consider realistic cellular base station deployments in Manhattan, New York City[§] and in the historical city center of Padova. Specifically, for the former we collect the locations of $N = 223$ 5G-NR base stations in an area of 15 Km^2 as depicted in Fig. 6.4a. On the other hand, in the Padova topology we combine locations of $N = 100$ 4G-LTE BS of different MNOs (WINDTRE, TIM, and Vodafone) in an area of 10 Km^2 as depicted in Fig. 6.4b, due to the lack of 5G-NR base station deployment at the time of

[§]The locations correspond to the network of T-Mobile, as it has the largest deployment among the MNOs.

Table 6.1: Simulation parameters.

Parameter	Value
Carrier frequency and bandwidth	28 GHz and 400 MHz
IAB RF chains	2 (1 access + 1 backhaul)
Pathloss model	UMi-Street Canyon [2]
Number of BS-nodes N	{223 NY, 100 Padova}
Source rate	{40, 80} Mbps
IAB Backhaul and access antenna array	8H×8V and 4H×4V
UE antenna array	4H×4V
IAB and UE height	15 m and 1.5 m
IAB antenna gain	33 dB
Noise power	10 dB
Risk level α	0.1
Reliability weight factor η	1

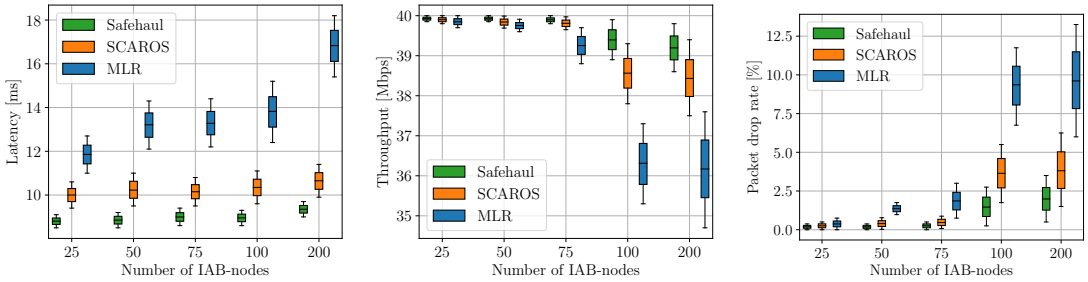
writing this paper. The detailed simulation parameters are provided in Table 9.5. We used the channel model outlined by 3GPP in TR 38.901 [2], which provides a statistical channel model for 0.5-100 GHz, and analyzed the "Urban Micro (UMi)-StreetCanyon" scenario.

Benchmarks. To provide better insights on the performance of Safehaul, we replicate two approaches from the state of the art: (i) Scalable and Robust Self-backhauling Solution (SCAROS), a learning-based approach that minimizes the average latency in the network [7], and (ii) Maximum-local-rate (MLR), a greedy approach aiming to maximize throughput by selecting the links with the highest data rate.

Our evaluation consists of six scenarios, in which we study the convergence of the algorithms to a steady state, the number of BS-nodes, the number of BS-donors, and the impact of risk aversion. When demonstrating the results, we show the average throughput, latency, and packet drop rate per UE. Furthermore, we show the statistical variance of the obtained results using candlesticks which include the corresponding max, min, mean, and 10 and 90 percentiles.

6.6.1 Scenario 1: Average Network Performance

Analyzing the performance of the algorithms as a function of time is crucial to determine the convergence speed of the learning-based techniques, i.e., Safehaul and SCAROS. Hence, in Fig. 6.5 we show the average network performance over time for three metrics: latency,



(a) Per-UE end-to-end Latency **(b)** Per-UE throughput **(c)** Per-UE packet drop rate

Figure 6.6: Network performance for $\{25,\ 50,\ 75,\ 100,\ 200\}$ BS-node, 2 UEs per BS-nodes on average, and 40 Mbps per-UE source rate (Scenario 2).

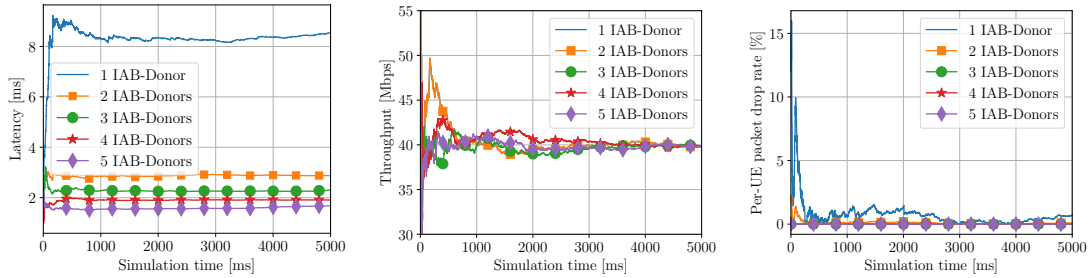
throughput, and packet drop rate.

In Fig. 6.5a, we can observe that Safehaul rapidly converges to an average latency of approximately 8.6 ms which is 12.2% and 43.4% lower than the latency of SCAROS and MLR, respectively. The high performance of Safehaul stems from the joint minimization of the average latency and the expected value of its tail loss, which results in avoiding risky situations where latency goes beyond $T_{\max}$. This is not the case for SCAROS where we observe a high peak in the latency before convergence, i.e., between zero and 1000 ms. *It is exactly the avoidance of such transients in Safehaul that leads to higher reliability in the system.* The reliability offered by Safehaul allows MNOs to deploy self-backhauling in an online fashion and without disrupting the network operation. The performance of MLR is constant throughout the simulation, as it is not designed as an adaptive algorithm.

Figure 6.5b shows that the risk-aversion capabilities of Safehaul have no negative impact on the average throughput of the network. The performance of Safehaul is comparable to that of SCAROS, approximately 79.3 Mbps, and 11.7% larger than the performance of MLR. The performance shown in Figure 6.5c is consistent with the behavior observed in Figure 6.5a. As Safehaul additionally minimizes the α-worst latency, it achieves the lowest packet drop rate compared to the reference schemes, namely, 30.1% (84.0%) lower than SCAROS (MLR).

6.6.2 Scenario 2: Impact of the Network Size

In Fig. 6.6 we evaluate the reliability of the three considered approaches for different network sizes. Specifically, we vary the number of BS-nodes starting from 25 up to 200. At the same time, we increase the load in the network by increasing the number of UEs. From the figures, we can clearly see that Safehaul consistently achieves a lower variation compared to



(a) Average per-UE end-to-end Latency **(b)** Average per-UE throughput **(c)** Average per-UE packet drop rate

Figure 6.7: Network performance for 50 UEs and 40 Mbps per-UE source rate, versus the number of BS-donors (Scenario 3).

the reference schemes. This verifies that Safehaul achieves the intended optimization goal, i.e., the joint minimization of the average end-to-end delay and its expected tail loss.

Fig. 6.6a shows that Safehaul is able to maintain an almost constant latency as the number of BS-nodes increases. Specifically, the variation of latency with Safehaul is 56.1% and 71.4% less than with SCAROS and MLR, respectively. Furthermore, Safehaul achieves 11.1% and 43.2% lower latency compared to SCAROS and MLR, where the high variance exhibited by the latter is due to a lack of adaptation capabilities.

As shown in Fig. 6.6b, the average throughput of the learning-based approaches Safehaul and SCAROS remains constant for the different values of network size. However, the lowest variation in the throughput is achieved by Safehaul, i.e., only 0.90 compared to 1.9 and 2.8 in the benchmark schemes. Such behavior corroborates Safehaul's reliability capabilities.

The packet drop rate for different numbers of BS-nodes is shown in Fig. 6.6c. Safehaul not only consistently outperforms the reference schemes, but also has the minimum variation in the results (by at least 47.3% compared to benchmarks). Considering the largest network size and load, i.e., 200 BS-nodes and 400 UEs, Safehaul achieves 49.3% and 81.2% lower packet drop rate compared to SCAROS and MLR, respectively.

6.6.3 Scenario 3: Impact of the number of BS-donors

Although the benchmark schemes do not support multiple BS-donors, Safehaul is designed to accommodate such scenarios. In Fig. 6.7, we investigate the impact of the number of BS-nodes on Safehaul. To this end, we keep the number of UEs and their data rate constant. We observe in Fig. 6.7a that the highest latency is experienced when only one BS-donor is present in the network. This stems from the tributary effect of self-backhauling where the traffic flows towards a central entity which itself can become a bottleneck. As the number of

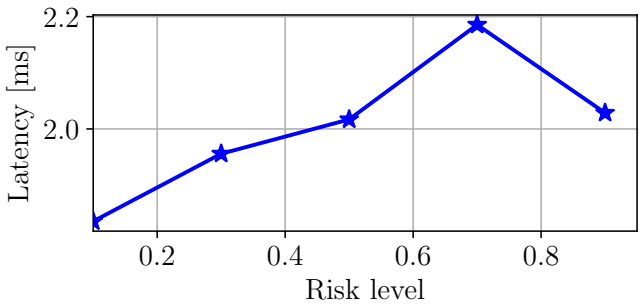
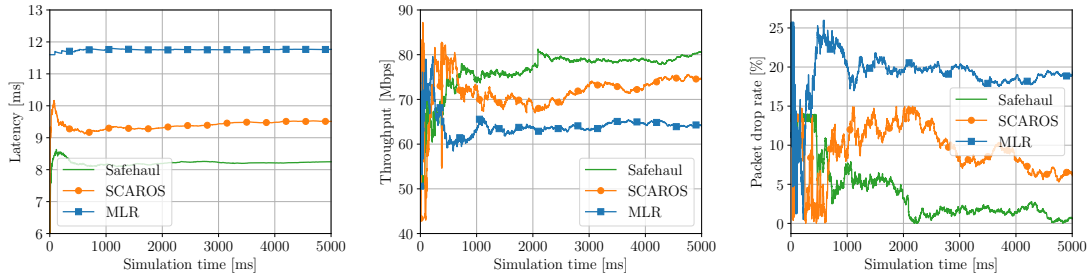


Figure 6.8: Average latency for 50 UEs and 20 Mbps per-UE source rate, versus the risk level α (Scenario 4)

BS-donors increases, the traffic is more evenly distributed, resulting in lower latency. Specifically, the average latency decreases from 8.2 ms for $D = 1$, to 1.7 ms when $D = 5$. As mentioned, since the load is constant in this scenario, the average throughput also remains constant for all different numbers of BS-donors, see Fig. 6.7b. Notably, Safehaul's learning speed is maintained for the different values of D. This is an important design feature of Safehaul because having more BS-donors means that the number of paths a BS-node has to the core network increases exponentially. From a learning perspective, such increment implies a larger action set and a lower learning speed. Safehaul avoids this problem by learning the average latency based on the estimates of its neighbors and not on the complete paths to the BS-donors. Finally, Fig. 6.7c shows that increasing the number of BS-donors significantly reduces the packet drops, which also stems from a better distribution of traffic flows in the network, as observed in Fig. 6.7a.

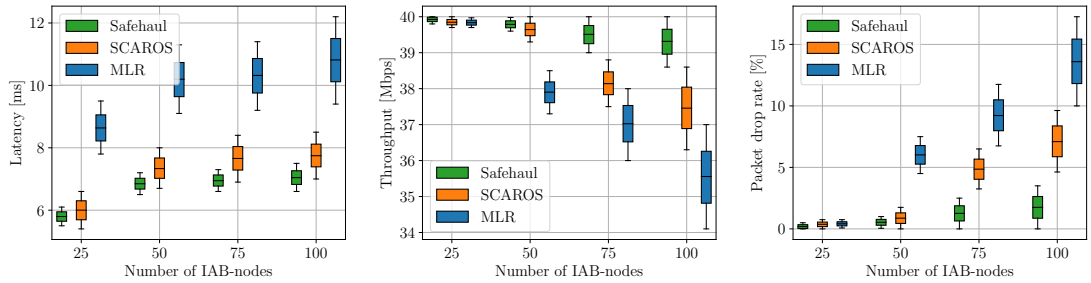
6.6.4 Scenario 4: Impact of the risk parameter α

The definition of losses in the tail of the latency distribution is controlled by the risk level parameter α. Its impact on the average latency is shown in Fig. 6.8, where an increasing behavior is observed for $\alpha \leq 0.7$. The lowest latency is achieved for $\alpha = 0.1$, which corresponds to the most risk-averse, and therefore the most reliable, case out of all the considered ones. The non-monotonic behavior of the average latency versus α can be explained by the so-called exploration-exploitation trade-off: the higher α, the higher the level of risk, which in turn leads Safehaul to learn more about the environment and choose a more reliable action. Eventually, as α grows beyond approximately 0.7, the performance of Safehaul tends to that of the risk-neutral case. As a consequence, the algorithm undertakes excessive exploration, which causes a degradation of the average latency performance.



(a) Average per-UE end-to-end latency **(b)** Average per-UE throughput **(c)** Average per-UE packet drop rate

Figure 6.9: Average network performance for 50 UEs and 80 Mbps per-UE source rate (Scenario 1) in Padova.



(a) Per-UE end-to-end Latency **(b)** Per-UE throughput **(c)** Per-UE packet drop rate

Figure 6.10: Network performance for $\{25,\ 50,\ 75,\ 100\}$ BS-nodes,2 UEs per BS-node on average, and 40 Mbps per-UE source rate (Scenario 2) in Padova.

6.6.5 Scenario 5: Performance in different topologies

To verify the generality of the proposed algorithms, it is essential to examine how they perform in different topologies, and consider both typical network performance metrics (i.e., along the lines of Scenario 1) and their stability with respect to the number of BS-nodes and BS-donors (Scenarios 2 and 3). To this end, we ran additional simulations in the deployment depicted in Fig. 6.4b, which mimics the BS-nodes locations of the historic center of Padova. We report the average network performance over time, in terms of end-to-end packet drop rate, throughput, and latency in Fig. 6.9. Overall, the outcomes of this simulation campaign are in line with those obtained in Scenario 1. Specifically, as seen in Fig. 6.9a, Safehaul quickly converges to an average latency of approximately 8 ms, which is 14% and 31% lower than SCAROS and MLR's latency. Fig. 6.9b shows the average per-UE throughput, in terms of which Safehaul achieves about 4% and 17% better performance than SCAROS and MLR, respectively. Similarly, the performance depicted in Fig. 6.9c is in line with the one reported in Figs. 6.9a and 6.9b, with Safehaul achieving approximately a 24% and 38% smaller packet

drop rate than SCAROS and MLR, respectively.

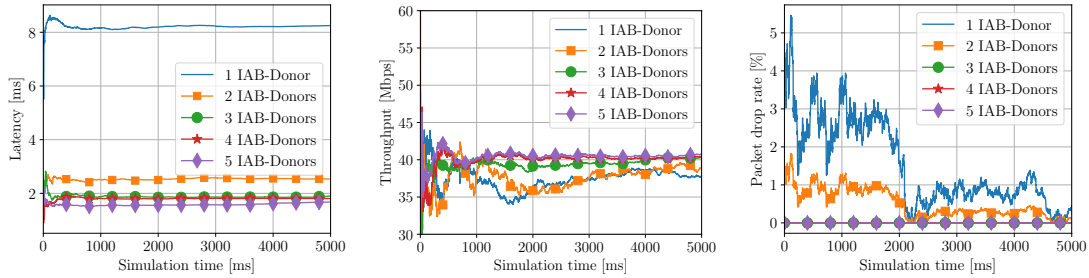
In Fig. 6.10, we compare the consistency of the performance of the three algorithms with respect to the network size. In particular, we change the number of BS-nodes from 25 to 100, keeping fixed the number of UEs per BS-node and thus effectively increasing the network load on the BS-donor. Results show that Safehaul, when compared to other schemes, exhibits minimal performance degradation when introducing additional BS-nodes and UEs. As can be seen in Fig. 6.10a, the latency achieved by Safehaul increases by at most 16% in the case of 100 BS-nodes, while SCAROS and MLR lead to a latency which is consistently higher and increases up to 27% and 25% when deploying additional nodes, respectively. Similar trends can be observed in Figs. 6.10b and 6.10c, which report throughput and packet loss versus the network size, respectively. Indeed, Safehaul is the best performer across the whole range of BS-nodes which have been considered. Furthermore, Safehaul loses 20% more packets with the denser network deployment (i.e., 100 BS-nodes), while reference schemes exhibit an increase in the packet loss of up to 33%.

We complete this analysis by examining how the number of donors affects the performance achieved by Safehaul in the Padova-like topology. As can be seen in Fig. 6.11, increasing the number of fiber-backhauled base stations progressively reduces the latency. Similarly, and in line with the results obtained in Scenario 3 and reported in Fig. 6.11c, the packet drop rate varies from approximately 0.08% when considering a single BS-donor, to approximately 0.003% in the presence of five BS-donors. The performance improvements introduced by additional fiber links saturate after 3 donors, thanks to the efficient routing and scheduling performed by Safehaul.

In summary, the results obtained in the additional topology mimicking the historical center of Padova are well aligned with those obtained in the Manhattan, NYC topology. Although the specific values of the network metrics achieved by the considered schemes in the two topologies are different (for instance, SCAROS achieves a 66% lower packet drop rate in Scenario 1 compared to Scenario 5), the trends among the various schemes are the same. Specifically, we observed that Safehaul consistently achieves the best performance in comparison to SCAROS and MLR across different metrics, which supports the claim that the proposed scheduler is capable of learning how to optimize arbitrary deployment topologies.

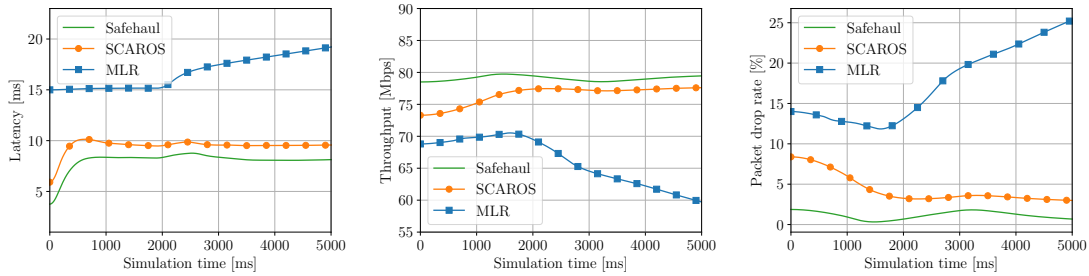
6.6.6 Scenario 6: Network resilience

In networking, resilience refers to the ability of a network to recover in a quick and effective fashion from disruptions, thus providing reliable and high-quality communication services



(a) Average per-UE end-to-end Latency **(b)** Average per-UE throughput **(c)** Average per-UE packet drop rate

Figure 6.11: Network performance for 50 UEs and 40 Mbps per-UE source rate, versus the number of BS-donors in Padova (Scenario 3).

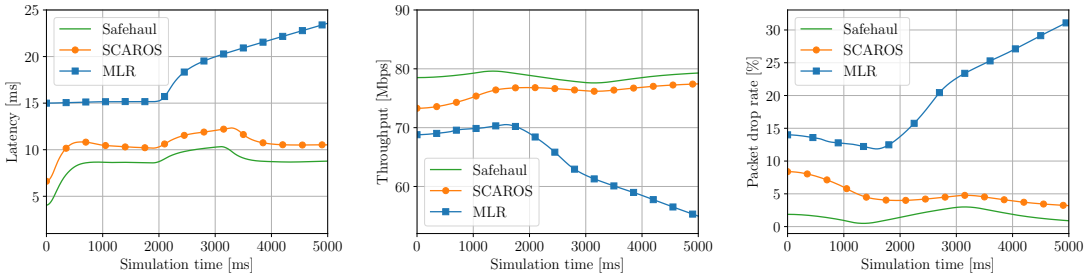


(a) Average per-UE end-to-end latency **(b)** Average per-UE throughput **(c)** Average per-UE packet drop rate

Figure 6.12: Average network performance for 50 UEs and 80 Mbps per-UE source rate where 1 random BS-node is shut down.

to its users. Specifically, the ability to recover from link failures is particularly important in IAB networks, where backhaul links are susceptible to the typical disruptions which plague the RAN due to its mobile and wireless nature. For instance, the links among BS-nodes can be degraded by adverse environmental conditions such as heavy rain and monsoons, physical obstacles and network congestion. These disruptions can cause temporary or permanent communication failures, which in turn result in degraded performance and/or loss of connectivity for the end users. To prevent and/or recover from these undesired events, a backhaul scheduler must detect, mitigate, and recover from various types of disruptions and failures, and must maintain the required level of service availability and performance despite the time-varying channel conditions.

We benchmark the resilience of the proposed algorithm by mimicking radio link failures, which we simulate by stopping BS-nodes at a fixed time instant (2000 s), and inspecting the resulting performance degradation. Since the failed node(s) is (are) chosen at random, we run multiple simulations to estimate the average network performance, as shown in Figs. 6.12



(a) Average per-UE end-to-end latency **(b)** Average per-UE throughput **(c)** Average per-UE packet drop rate

Figure 6.13: Average network performance for 50 UEs and 80 Mbps per-UE source rate where 3 random BS-node are shut down.

and 6.13 for the case of one and three link failures, respectively.

Results show that MLR is unable to react to the link failure(s) due to its static and myopic policy. Specifically, the disruption causes an increase of 33% (60%) in the latency, and a decrease of up to 15% (23%) in the throughput when considering one (three) link failure(s). On the other hand, both Safehaul and SCAROS are capable of adapting the scheduling to the new topology. Indeed, both schemes show a transitory region where performance is slightly degraded since the algorithms are learning new routes and resource partitions to account for the lost link. Nevertheless, Safehaul and SCAROS eventually converge to a solution which provides approximately the same network performance as before the failures, both in the case of one and three lost links.

Appendix

For the proof of Proposition 1, Theorem 3 in [134] is needed. For completeness, we first present the theorem for the special case in which the considered random variables are independent. Next, we present the proof of Proposition 1.

Theorem 2. *Let* $T_{a_n,i}$ *be independent random variables where* $\max_{1\leq j\leq i} T_{a_n,j} = T_{\max}$*, with* $i \in \{1, 2, ...\}$*. Then, for any* $0 < \delta \leq 1/2$*,* $\xi > 0$ *and* $\gamma > 0$*, there exists a positive constant* C *which only depends on* ξ *and* γ*, such that the probability of the event* $|\widehat{\mathrm{CVaR}}_{a_n,i} - \mathrm{CVaR}_{a_n,i}| \geq 2\xi\alpha^{-1}T_{\max}i^{-\delta}(\ln \ln i)^{1/2} \ln i$ *is smaller or equal than* $Ce^{-(1+\gamma)\ln i}$*.*

Proof. See Theorem 3 in [134]. □

Proof of Proposition 1

From the system model and Proposition 1, we have that $c > 0$, $0 < d \leq 1$, and $\epsilon_n := \min(1, \frac{cA_n}{d^2 i})$. Moreover, $a_{n,i}$ is the action chosen by ϵ-greedy in time slot i and $K_{a_n,i}$ is the number of times, up to time slot i, in which Safehaul chose action a_n *at random*. Similarly, we use K_i^* for the counter of the optimal action. $T_{a_n,i}$ are independent random variables distributed according to the rewards linked to action a_n. We use T_i^* for the optimal action, and $\hat{T}_{a_n,i}$ is the estimated mean of the probability distribution of the rewards linked to action a_n using $K_{a_n,i}$ samples. As before, we use $\hat{T}_i^*$ for the optimal action. $\widehat{\mathrm{CVaR}}_{a_n,i}$ is the estimated CVaR of action a_n up to time slot i and $\widehat{\mathrm{CVaR}}_i^*$ is the estimated CVaR of the optimal action up to time slot i. Then, the probability that action a_n is chosen in time slot i is upper bounded as

$$\mathbb{P}[a_{n,i} = a_n] \leq \mathbb{P}\left[\delta_{a_n,i-1} \leq \delta_{i-1}^*\right]\left(1 - \frac{\epsilon_i}{A_n}\right) + \frac{\epsilon_i}{A_n}, \tag{6.14}$$

with $\delta_{a_n,i-1} = \hat{T}_{a_n,i-1} + \eta\widehat{\mathrm{CVaR}}_{a_n,i-1}$ and $\delta_{i-1}^* = \hat{T}_{i-1}^* + \eta\widehat{\mathrm{CVaR}}_{i-1}^*$. The first term in (6.14) is the probability of exploitation and the second term to the probability of exploration. Using the mean $\bar{T}_{a_n}$ and CVaR_{a_n} of action a_n, and the likewise defined $\bar{T}^*$ and CVaR^* for the optimal action, we set $\Delta_{a_n}^{\mathrm{mean}} := \bar{T}_{a_n} - \bar{T}^*$ and $\Delta_{a_n}^{\mathrm{cvar}} := \mathrm{CVaR}_{a_n} - \mathrm{CVaR}^*$. Using these definitions in (6.14) we conclude

$$\begin{aligned}
&\mathbb{P}\left[\delta_{a_n,i-1} \leq \delta_{i-1}^*\right] \leq \\
&\mathbb{P}\left[\delta_{a_n,i-1} \leq \eta\mathrm{CVaR}_{a_n} - \frac{\Delta_{a_n}^{\mathrm{mean}}}{2} + \bar{T}_{a_n} - \eta\frac{\Delta_{a_n}^{\mathrm{cvar}}}{2}\right] + \\
&\mathbb{P}\left[\bar{T}^* + \frac{\Delta_{a_n}^{\mathrm{mean}}}{2} + \eta\mathrm{CVaR}^* + \eta\frac{\Delta_{a_n}^{\mathrm{cvar}}}{2} \leq \delta_{i-1}^*\right] \qquad (6.15) \\
&\mathbb{P}\left[\hat{T}_{a_n,i-1} \leq \bar{T}_{a_n} - \frac{\Delta_{a_n}^{\mathrm{mean}}}{2}\right] + \mathbb{P}\left[\bar{T}^* + \frac{\Delta_{a_n}^{\mathrm{mean}}}{2} \leq \hat{T}_{i-1}^*\right] + \\
&\mathbb{P}\left[\widehat{\mathrm{CVaR}}_{a_n,i-1} \leq \mathrm{CVaR}_{a_n} - \frac{\Delta_{a_n}^{\mathrm{cvar}}}{2}\right] + \\
&\mathbb{P}\left[\mathrm{CVaR}^* + \frac{\Delta_{a_n}^{\mathrm{cvar}}}{2} \leq \widehat{\mathrm{CVaR}}_{i-1}^*\right]. \qquad (6.16)
\end{aligned}$$

Similar to [135], we use the Chernoff-Hoeffding bound for the first two terms in (6.16). For the last two summands, it remains to find a bound for the difference between the CVaR and its estimate $\widehat{\mathrm{CVaR}}$. From Theorem 2, we set $\xi := \Delta_{a_n}^{\mathrm{cvar}}\alpha/4T_{max}$, $\delta = 0.5$ and by using

the limit $\gamma \to 0$, we obtain

$$\mathbb{P}\left[|\widehat{\mathrm{CVaR}}_{a_n,i} - \mathrm{CVaR}_{a_n,i}| \geq \frac{\Delta^{\mathrm{cvar}}_{a_n}}{2} i^{-0.5} (\ln \ln i)^{0.5} \ln i\right] \leq \frac{C}{i}. \tag{6.17}$$

As $\max_i i^{-0.5}(\ln \ln i)^{0.5} \ln i < 1$, the condition $(\Delta^{\mathrm{cvar}}_{a_n}/2) i^{-0.5}(\ln \ln i)^{0.5} \ln i \leq \frac{\Delta^{\mathrm{cvar}}_{a_n}}{2}$ holds for all i. Therefore, considering the last two summands in (6.16), we conclude that there exists a positive constant C that satisfies

$$\mathbb{P}\left[|\widehat{\mathrm{CVaR}}_{a_n,i} - \mathrm{CVaR}_{a_n,i}| \geq \frac{\Delta^{\mathrm{cvar}}_{a_n}}{2}\right] \leq \frac{C}{i}. \tag{6.18}$$

The number of times action a_n has been selected up to time slot i is smaller than or equal to i, i.e., $K_{a_n,i} \leq i$. Using (6.18) we write the last two summands in (6.16) as

$$\mathbb{P}\left[\widehat{\mathrm{CVaR}}_{a_n,i-1} \leq \mathrm{CVaR}_{a_n} - \frac{\Delta^{\mathrm{cvar}}_{a_n}}{2}\right] \leq \frac{C}{K_{a_n,i-1}}, \tag{6.19}$$

and

$$\mathbb{P}\left[\mathrm{CVaR}^* + \frac{\Delta^{\mathrm{cvar}}_{a_n}}{2} \leq \widehat{\mathrm{CVaR}}^*_{i-1}\right] \leq \frac{C}{K^*_{i-1}}. \tag{6.20}$$

As in [135], we use Bernstein's inequality to get an estimate for $K_{a_n,i-1}$. Defining $x_0 := 1/2A_n \sum_{j=1}^{i-1} \epsilon_j$ for $i - 1 \geq \frac{cA_n}{d^2}$ we get $P(K_{a_n,i-1} \leq x_0) \leq e^{-\frac{x_0}{5}}$. Additionally, from [135]:

$$x_0 \geq \frac{c}{d^2} \ln\left(\frac{(i-1)d^2 e^{0.5}}{cA_n}\right) =: C'(i). \tag{6.21}$$

The same holds for the optimal action and K^*_{i-1}. Using these estimations for x_0, we can conclude that for $i - 1 \geq cA_n/d^2$

$$\begin{aligned}
&\mathbb{P}\left[\widehat{\mathrm{CVaR}}_{a_n,i-1} \leq \mathrm{CVaR}_{a_n} - \frac{\Delta^{\mathrm{cvar}}_{a_n}}{2}\right] && (6.22)\\
&\leq \sum_{j=1}^{i-1} \mathbb{P}[K_{a_n,i-1} = j] \frac{C}{j} \\
&= \sum_{j=1}^{\lfloor x_0 \rfloor} \mathbb{P}[K_{a_n,i-1} = j] \frac{C}{j} + \sum_{j=\lfloor x_0 \rfloor + 1}^{i-1} \mathbb{P}[K_{a_n,i-1} = j] \frac{C}{j} \\
&\leq C x_0 e^{-\frac{x_0}{5}} + \frac{C}{x_0} \leq C x_0 e^{-\frac{x_0}{5}} + \frac{C}{C'(i)}. && (6.23)
\end{aligned}$$

The same holds again for the optimal action

$$\mathbb{P}\left[\mathrm{CVaR}^* + \frac{\Delta_{a_n}^{\mathrm{cvar}}}{2} \leq \widehat{\mathrm{CVaR}}_{i-1}^*\right] \leq Cx_0 e^{-\frac{x_0}{5}} + \frac{C}{C'(i)}. \tag{6.24}$$

Together with the bounds from Theorem 3 in [135] it follows that for $C \geq 1$:

$$\begin{aligned}
&\mathbb{P}\left[a_{n,i} = a_n\right] \\
&\leq \frac{\epsilon_i}{A_n} + 4Cx_0 e^{-\frac{x_0}{5}} + \frac{4}{\left(\Delta_{a_n}^{\mathrm{mean}}\right)^2} e^{\frac{-\left(\Delta_{a_n}^{\mathrm{mean}}\right)^2 \lfloor x_0 \rfloor}{2}} + 2\frac{C}{C'(n)} \\
&\leq \frac{c}{d^2 i} + 2\frac{Cd^2}{c \ln\left(\frac{(i-1)d^2 e^{0.5}}{cA_n}\right)} + \frac{4e}{d^2}\left(\frac{cA_n}{(i-1)d^2 e^{0.5}}\right)^{\frac{c}{2}} + \\
&\quad 4C\frac{c}{d^2} \ln\left(\frac{(i-1)d^2 e^{0.5}}{cA_n}\right)\left(\frac{cA_n}{(i-1)d^2 e^{0.5}}\right)^{\frac{c}{5d^2}}.
\end{aligned}$$

□

7

Study of 'RIS': End-to-End Simulation of 5G Networks Assisted by RIS and AF Relays

High propagation and penetration loss at millimeter wave (mmWave) frequencies necessitate ultradense deployments of 5th generation (5G) base stations, which may be impractical and expensive for network operators. Integrated Access and Backhaul (IAB) has been proposed as a partial solution to this problem, despite concerns regarding power consumption and scalability. Recently, the research community has been examining Reconfigurable Intelligent Surface (RIS) and Amplify-and-Forward (AF) relays as more energy-efficient solutions to coverage problems in 5G scenarios. In this chapter we introduce a new simulation framework, based on ns-3, to simulate RIS/AF systems from a full-stack, end-to-end perspective, taking into account the impact of the channel model and protocol stack of 5G NR networks. Our objective is to determine if these technologies can be used to relay 5G traffic requests and, if so, how to dimension RIS/AF nodes based on the number of end users.

7.1 Introduction

5G networks are being carried out globally in order to provide 20 times greater peak throughput and 10 times lower latency than previous generations. In order to achieve this, the 3GPP

has released a new set of innovations for 5G networks [140], including support for network operations in the mmWave spectrum in conjunction with massiveMIMO (m-MIMO) technologies. In turn, mmWave transmissions introduce several propagation issues, most notably severe path and penetration losses, which force short-range communication [141]. Denser deployment of 5G mmWave base stations could be a solution, but it would be costly for network operators, particularly in terms of sites acquisition campaigns, rental fees, and fiber architecture to provide wired backhauling [142]. As part of its 5G NR Rel-16 [143] specifications, the 3GPP approved Integrated Access and Backhaul (IAB) as a new paradigm to supplant fiber-like infrastructures with self-configuring relays operating over wireless (mmWave) backhaul links. Despite its potential, m-MIMO-assisted IAB still requires sophisticated signal processing as well as hardware that is both expensive and energy-intensive [144]. This issue is exacerbated in rural/remote areas, where harsh weather and terrain, and the lack of a powerful electrical infrastructure in many instances, may further confound IAB installation [145]. Consequently, new technologies based on RIS and Amplify-and-Forward (AF) relays have been proposed as promising energy-efficient alternatives to surmount the coverage issues of mmWave networks [146]. An RIS is a metasurface that can be programmed to alter an electromagnetic (EM) field in a favorable direction. RISs are nodes that passively beamform the impinging signal without amplification, allowing them to guarantee minimum capacity requirements in inactive areas while consuming less power than IAB [147]. Instead, it is envisioned that AF relays will capture an incident electromagnetic wave emanating from a base station, actively amplify the received signal, and re-radiate it towards a target area to be served. They are candidates for attaining higher capacity compared to RIS nodes, albeit at a higher cost and at the expense of noise amplification [148]. Whether or not these technologies will be able to meet 5G (and beyond) service requirements, and if so, how to appropriately dimension RIS/AF systems, remain unanswered questions of critical importance. For testing and calibrating RIS/AF deployments, computer-based simulations are a viable alternative to field experiments with actual hardware due to scalability and flexibility concerns, as well as the high cost of testbed components. Prior works, such as [149], [150], have addressed this task, albeit with a focus on link-level analyses, which typically employ conservative system architecture assumptions and should be considered a lower bound for more representative end-to-end performance studies.

To address this deficiency, we present a more comprehensive system-level performance evaluation of RIS/AF deployments using a new simulation framework that operates end-to-end, thus incorporating the interaction with the 5G NR protocol stack and relative control

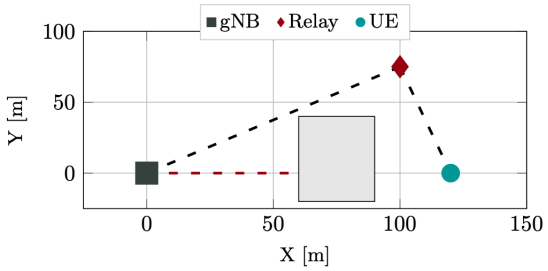
tasks, as well as the impact of the upper (transport and application) layers. ns-3, an open-source discrete-event simulator for wireless networks, serves as the foundation for our framework. Specifically, we describe our ns-3 implementation of the RIS/AF channel, which is based on the current 3GPP channel model for 5G networks as specified in [2] and implemented, for example, in the ns3-mmwave module [151], which models the Physical (PHY) and Medium Access Control (MAC) layers of the 5G NR protocol stack. On this basis, we undertake an extensive simulation campaign to compare the efficacy of RIS/AF nodes for relaying connectivity requests from end users to a baseline solution in which relays are not deployed. RISs and AF relays are valid solutions, particularly in small networks, despite the requirement for highEIRP AF relays to support more aggressive traffic applications. We provide guidelines for the optimal dimensioning of RIS and AF configurations, in terms of a number of antenna elements and amplification power, based on our simulations.

7.2 Implementation of the RIS Signal Model in ns-3

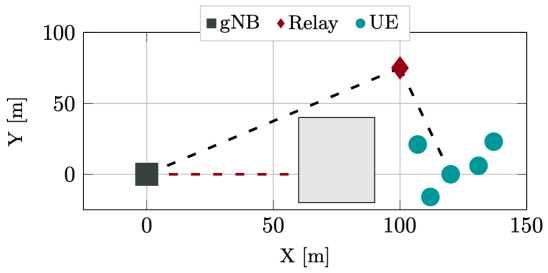
In Section III-A of [152], we described how our simulator calculates the channel (in terms of PSD) for RIS/AF relays, which is then used to compute the end-to-end SINR at the destination D. SINR can refer to either the SINR relative to the entire bandwidth for narrowband signals transmitted over frequency-flat channels or the SINR experienced over a single subcarrier for wideband signals transmitted over frequency-selective channels. In the second case, the SINRs corresponding to the different frequency chunks are mapped into a singular SINR value using additional maps derived from link-level simulations [153]. Based on this, our simulator defines a Link-to-System Mapping (L2SM), i.e. a table that associates a given SINR with a MAC-layer Transport Block (TB) error rate [154], which is then used to determine whether or not the TB has been received accurately. The upper layers of the 5G NR protocol hierarchy are modeled based on the *ns3-mmwave* module [151]. It employs a custom PHY layer that supports the NR frame structures and numerologies, as well as a MAC layer with ad hoc beamforming and scheduling policies. The Radio Link Control (RLC) and Packet Data Convergence Protocol (PDCP) layers implement network functions including segmentation, retransmissions, and/or reassembly of packets.

7.3 Performance Evaluation

In this section, we describe our simulation setup and parameters and evaluate the performance of RISs and AF relays as a function of various antenna array configurations.



(a) Scenario 1, with $N_{\mathrm{U}} = 1$.



(b) Scenario 2, with $N_{\mathrm{U}} = 5$.

Figure 7.1: Simulation scenarios, where we deploy one gNB, NU UEs and, possibly, a relay. A building (the gray rectangle) blocks the direct link (dashed red line) from the gNB to the UEs. In turn, the relay guarantees a LoS link (dashed black line) to all the devices.

7.3.1 Simulation Setup

In our simulations, we investigate two simple yet realistic urban canyon scenarios in which we deploy a single Next Generation Node Base (gNB), NU user equipments (UEs), with NU = 1 (5) in Scenario 1 (2), and a single relay, which may be either an RIS or an AF relay. Modeling the wireless channel as an Urban Macro (UMa) link [2]. The LoS/NLoS condition depends on the scenario's geometry. Specifically, we presume that the direct wireless link between the UEs and the gNB is obstructed by a building, as depicted in Fig. 7.1, which introduces an additional penetration loss modeled according to [[2], Sec. 7.4.3.1]. Through the relay, the end nodes can continue to communicate in LoS.

The parameters for our simulation are listed in Table 7.1 Specifically, UEs obtain User datagram protocol (UDP) data from a remote server. At each transmission opportunity towards the generic k-th UE, we presume that both AF and RIS relays can use their optimal configuration, i.e. the codeword yielding the highest end-to-end SINR towards UE k. The system operates at 28 GHz with a total bandwidth of 100 MHz, which will be shared by all devices using TDMA. The gNB is equipped with a 64-element antenna array and operates

TABLE I: Simulation parameters.

Parameter	Value
Carrier frequency	28 GHz
Total bandwidth	100 MHz
Number of UEs (N_U)	{1, 5}
gNB antenna array	8H×8V
gNB max RF power	33 dBm
UE antenna array	2H×1V
IRS antenna array	{10H×20V, 20H×40V,40H×80V, 60H×120V}
AF antenna array	{4H ×4V, 8H ×8V, 16H ×16V}
AF amplification	40 dB
Antenna radiation pattern	[13, Table 7.3-1]
UDP source rate	50 Mbps

Table 7.1: Simulation parameters

at 33 dBm. We evaluate between 200 and 7,200 reflecting elements for RIS purposes. We consider antenna arrays with 16 to 256 elements for the AF relay.

7.3.2 Numerical Results

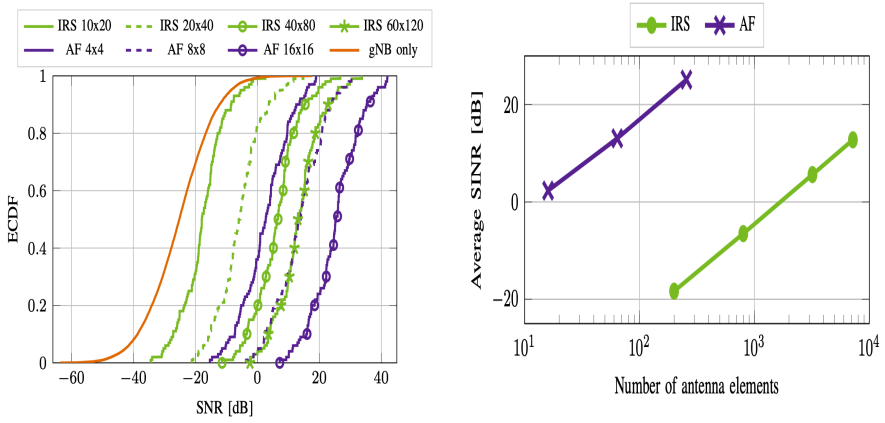
Now we compare the end-to-end performance of RIS- and AF-relay assisted networks:

- **SINR**: It is a measurement of the channel's excellence. It is dependent upon PHY-layer characteristics, such as the relative distance between the transmitter, receiver, and relay (if applicable), the operating frequency, propagation conditions, and channel bandwidth.
- **End-to-end throughput**: It is measured as the total number of received bytes per user divided by the total.
- **End-to-end latency**: It is measured from the time each packet is generated at the application layer to when it is successfully received. Accordingly, it accounts for both transmission and queuing times.
- **Packet Error Rate (PER)**. It is the ratio of the number of packets delivered with errors to the total number of packets transmitted.

The efficacy of the RIS/AF will be compared to a baseline scenario (referred to as "gNB-only") in which there is no intermediate relay.

SINR

Our analysis begins with the SINR statistics depicted in Figure 7.2 for Scenario 1 with NU = 1. First, we observe in Fig. 7.2a that the relay enhances the SINR (on average by up to +55 dB)



(a) (b)

Figure 7.2: SINR statistics for Scenario 1.

compared to the "gNB only" baseline, in which the UE communicates in NLoS. Notably, as illustrated in Figure 7.2b, both RIS and AF relays offer an end-to-end SINR gain that scales proportionally to the number of radiating elements at the relay. This effect is produced by the RIS beamforming gain and the fact that the power collected by the RIS is proportional to its surface area, which is in turn proportional to the number of radiating elements [155]. The AF-assisted configurations consistently outperform the RIS-assisted configurations in terms of SINR (on average up to +40 dB, with the same number of antennas): this is expected because the AF relay amplifies the signal, obtaining a greater end-to-end gain. SINR is less than 0 dB when the RIS has fewer than 800 elements, which justifies the use of very large RIS panels. In fact, a 60 120 element RIS panel has an average SINR of 13 dB, which is sufficient to support reliable transmissions as long as communication requirements are not excessively stringent, as we will demonstrate in the following paragraphs.

End-to-end throughput

In Fig. 7.3, we illustrate the end-to-end throughput at the application layer, taking into account the impact of the entire 5G NR protocol stack. When NU equals one (Scenario 1), the average throughput reveals the ergodic capacity. Given the extremely low SINR (below the sensitivity threshold for the majority of commercial receivers) at the physical layer, the throughput for the "gNB only" baseline is zero. Interestingly, even though the AF relay with 16 × 16 antennas guarantees, on average, 15 dB higher SINR than an RIS with 60 × 120 elements (from Fig. 7.2a), we see that the end-to-end throughput of the two configura-

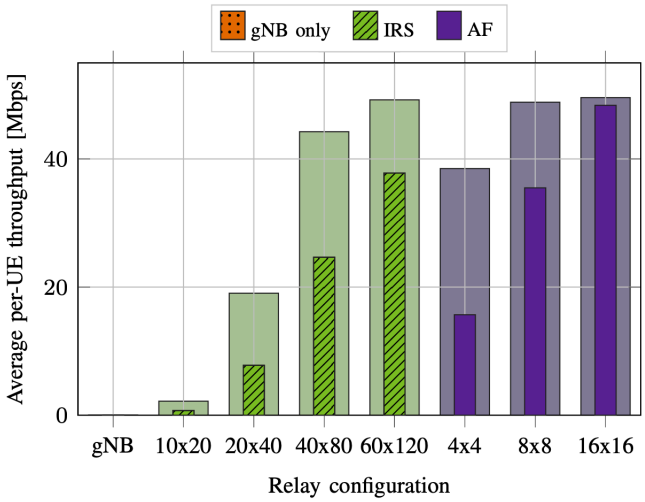


Figure 7.3: End-to-end per-UE throughput at the application layer in Scenario 1 (wide bars) and Scenario 2 (narrow bars) for different relay configurations.

tions is comparable. This demonstrates that in a straightforward scenario with a single UE, a SINR average of 15 dB is sufficient to meet all traffic demands. Due to its simplicity, the RIS is preferable to an AF relay in this situation. In addition, it is impractical to increase the RIS size further because the throughput is already maximal and equal to the UDP source rate (50 Mbps in our simulations). When NU = 5 (Scenario 2), the average per-UE throughput is substantially lower than in Scenario 1 because, in a multi-user scenario, radio resources must be shared among UEs, which can result in channel congestion. This result verifies the precision and veracity of our ns-3 architecture. However, this effect is diminished for extremely massive antenna panels. For instance, for an AF relay with 44 antennas, the per-UE throughput decreases by nearly 60%, whereas for a 16×16 element array, the per-UE throughput decreases by only 2%. Even in Scenario 2, AF assisted networks can continue to support the application source rate if at least 16 16 antennas are utilized. RISs, on the other hand, are unable to attain the full source rate offered by the application due to the limited SINR available at the PHY layer. Maximum achievable throughput is approximately 40 Mbps for 60 120 elements, which is 20% greater than in the case where NU = 1.

End-to-end latency

The 95th percentile of the end-to-end latency encountered at the application layer is plotted in Figure 7.4. Even in the simplest scenario with a single user equipment (UE) deployed

(Scenario 1), the performance is generally poor, as the latency is greater than 100 ms for the majority of relay configurations, indicating system instability. In reality, the use of relays with small antenna panels results in extremely high levels of queuing and buffering, which degrades latency. Despite the increased system complexity, this problem is solvable by configuring larger RIS and AF relays. For instance, an RIS with 60 120 elements and an AF relay with 8×8 elements can guarantee an end-to-end latency of less than 10 milliseconds, which satisfies the requirements of the majority of 5G applications. Notice that the latency for the configuration "gNB only" is not particularly representative, as it is relative to only the accurately received packets. In fact, without the relay, transmissions are in NLoS and several packets are lost (see the PER in Fig. 7.5), which makes the system less congested; the (few) packets that make it to the application layer are then transmitted with less latency. In spite of this, the latency is still greater than two orders of magnitude above the best RIS and AF configurations, indicating that relays are preferable for these types of networks. When NU equals 5 (Scenario 2), the latency is typically greater than when NU equals 1. This is to be anticipated given that UEs compete for available resources. In addition, using UDP as a transport protocol, thus with a full buffer source traffic model, each end-to-end flow does not self-regulate to the actual network conditions, thus congestion arises. Considering non-UDP traffic could improve performance: for instance, the congestion control mechanism available in Transmission Control Protocol (TCP) would regulate source traffic and prevent network congestion and buffer overflow. Even with the most aggressive RIS architecture of 60120 elements, the average latency is greater than 1000 ms, compared to 6.5 ms in Scenario 1. This is because, in this scenario, more than 20% of the packets are lost and retransmitted (see Fig. 7.5), thereby increasing the packet latency. In contrast, for an AF relay with 16×16 antennas, the latency is approximately 130 ms on average, with a PER as low as 3%, which can still support some essential target communication requirements. Even though RIS-assisted networks consume less energy, they are inappropriate in this scenario unless extremely large RIS panels are utilized.

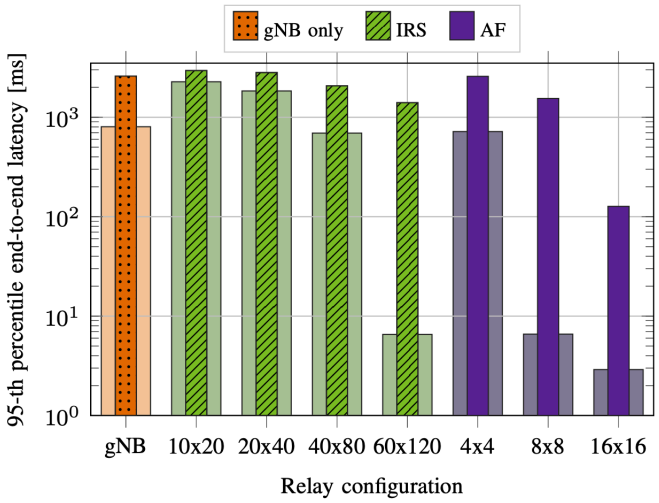


Figure 7.4: 95-th percentile of the end-to-end latency at the application layer in Scenario 1 (wide bars) and Scenario 2 (narrow bars) for different relay configurations.

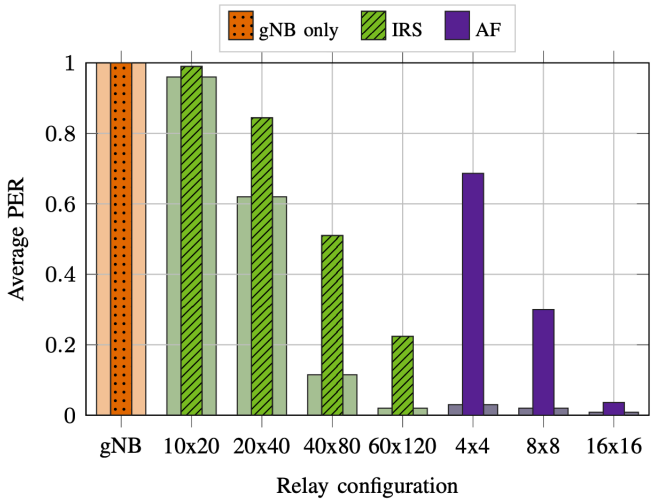


Figure 7.5: Average PER at the application layer in Scenario 1 (wide bars) and Scenario 2 (narrow bars) for different relay configurations

8

Potential of Sub-THz in Backhauling

IAB is a flexible and low-cost solution to deploy ultra-dense fifth and sixth generation (5G and 6G) systems, as it enables wireless backhaul links based on the same technology and specifications as for the access links. In this chapter, we consider the deployment of mixed mmWave and sub-terahertz links to increase the capacity of the backhaul network, and provide the first performance evaluation of the potential of sub-terahertz frequencies for 6G IAB. To do so, we develop a greedy algorithm that allocates frequency bands to the backhaul links (considering constraints on spectrum licenses, sharing, and congestion) and generates the wireless network mesh. We then profile the performance through a custom extension of the open-source system-level simulator Sionna that supports Release 17 IAB specifications and channel models up to 140 GHz. Our results show that IAB with sub-terahertz links can outperform a mmWave-only deployment with improvements of 4× for average user throughput and a reduction of up to 50% for median latency.

8.1 Introduction

Future wireless networks will accommodate data-rate intensive use cases which include untethered VR and mobile metaverse applications. This will further exacerbate the congestion on mobile access networks and backhaul systems [156]. For this reason, 5G cellular systems have pushed into the mmWave band, with typical deployments in the spectrum around 28 GHz and 39 GHz [157], and sub-terahertz mobile links are being considered for 6G applica-

tions [158, 159].

Wireless networks operating at such high frequencies will be deployed with extremely high density, to improve the probability of LoS coverage and mitigate the impact of the harsh propagation environment. To make ultra-dense deployments viable, the 3GPP has standardized an extension of 5G NR, i.e., IAB, which exploits the same waveform and protocol stack to provide access to mobile users and wireless backhaul for Next Generation Node Bases (gNBs) (i.e., the IAB nodes) thus limiting the need for fiber drops. The wireless backhaul topology terminates at a gNB with fiber connectivity to the data core, the IAB donor [160, 161, 80]. IAB also simplifies the deployment of cellular networks in on-demand or ad hoc contexts, as it removes the need for part of the wired backhaul.

IAB networks in 5G systems are a natural application for mmWave deployments, as telecom operators can easily fit carriers with 400 MHz of bandwidth in this spectrum. In addition, the directionality that mmWave arrays introduce helps reduce the interference. Nonetheless, studies have shown that bottlenecks can emerge at IAB donors, creating congestion, high latency, and degraded Quality of Service (QoS) for the end users, especially when backhaul links are constrained to re-using the same spectrum of the access (as in in-band IAB) [85].

In this context, out-of-band IAB with sub-terahertz links is seen as a solution to support immersive multimedia data-hungry streams. Specifically, the spectrum above 100 GHz has several sub-bands that could provide bandwidths wider than 10 GHz, thus potentially data rates in the excess of tens of Gbps [162]. Backhaul—a static deployment—is a promising use case for sub-terahertz links, which need pencil-sharp beams to close the link budget and are thus less resilient to mobility compared to traditional sub-6 GHz or mmWave frequencies.

In recent years, the literature has closed several gaps in terms of circuit, antenna design [163] and physical and MAC layer solutions for sub-terahertz systems [164]. When it comes to IAB with mixed sub-terahertz and mmWave links,* however, there are still several open questions in terms of network design and path selection. In this paper, we consider the problem of identifying a viable topology between IAB nodes and the IAB donors, including the carrier frequency of the backhaul links, and profile the performance that network planners can expect when mixing sub-terahertz and mmWave IAB links.

To this end, we develop a greedy path generation algorithm that automatically selects the frequency band of an IAB link (between 28 GHz and 140 GHz) and assigns routes so that each IAB node can reach the IAB donor. The frequency selection aims at avoiding bottlenecks, i.e., the algorithm selects the band that provides the highest capacity when accounting

*In this research, we consider the FR2 range of 3GPP NR (24.25 GHz to 71 GHz) as mmWaves.

for the congestion that may arise in the proximity of the IAB donor. In addition, we consider and compare different ratios of sub-terahertz and mmWave links, which can be mapped to licensing constraints for out-of-band backhaul, and two different bandwidths for the sub-terahertz links (10 GHz and 32 GHz), which consider exclusive licensing or sharing with other services, respectively [9].

We model the IAB network in a custom-developed 3GPP Release 17 simulator based on the open-source tool Sionna [1], with 3GPP and state-of-the-art mmWave and sub-terahertz channel models, and realistic and detailed 3GPP-based physical and MAC layers. Our results quantify for the first time the performance improvement that sub-terahertz links can introduce in IAB networks, which can push beyond the limits of the in-band mmWave backhaul and support more than 50 users with 120 Mbps streams and a single donor without congestion (compared to about 33 Mbps for in-band mmWaves).

This is the one of the first researches that provides a numerical evaluation of the potential associated with sub-terahertz links for IAB. Notably, [165] evaluates the sub-THz potential in backhaul networks from a physical layer perspective. This research demonstrates that sub-THz spectrum links can achieve multi-Gbps ratios in outdoor backhaul scenarios. [166] proposed unmanned aerial vehicles (UAV)-assisted backhaul solution to improve network coverage and data rate in heterogeneous networks with multiple tiers composed of sub-6 GHz, THz and UAV layers. In addition, the authors of [167] successfully adopted concurrent scheduling to increase system throughput in dense THz backhaul scheduling. Finally, [168] considers a multi-band IAB deployment, but with a bandwidth that is more limited than those considered in future 6G scenarios.

The rest of the paper is organized as follows. Sec. 8.2 introduces the system model. Sec. 8.3 describes the algorithm for frequency and path selection, which is then numerically evaluated in Sec. 8.4.

8.2 System Model

We consider a TDMA system in which a single IAB donor, featuring a fiber connectivity towards the Core Network (CN) and the Internet, exchanges data with N_{U} UEs. Without loss of generality, we consider uplink traffic only. To achieve uniform coverage, the donor is aided by N_{I} IAB nodes, which can be either connected to the former, or to neighboring base stations thus possibly realizing a multi-hop wireless backhaul.

We partition the time resources in T radio subframes of duration $T_{sub} = 1$ ms, and we

equip all nodes with buffers. Accordingly, the data that node i transmits to gNB k during subframe t is stored in its buffer $B_k(t)$, and represents either successfully received packets, in the case of the donor, or data to be relayed to the next hop along the path during subframe $t+1$, in the case of IAB nodes.

We assume that the backhaul links operate *either in the mmWave or in the THz bands* and that each IAB node features two RF chains, which are used for the backhaul and the fronthaul communications, respectively. In both cases, gNBs are equipped with directional antennas.

When gNB $k = 0, \ldots, N_{\mathrm{I}}$, with index 0 denoting the IAB donor, receives data from node j, packets experience a SINR $\gamma_{s,d}$ which can be expressed as

$$\gamma_{s,d} = \frac{|h^l_{s,d}|^2 \sigma_x^2}{\sigma_n^2 + \sum_{i \in \mathcal{I}} \sigma_i^2}, \tag{8.1}$$

where $h^l_{s,d}, \;\; l \in \{mW, \, sT\}$ represents the equivalent channel response between the communication endpoints when using mmWave or sub-THz links, respectively. $\mathcal{I}$ denotes the set of interferers, σ_x^2, σ_i^2 and σ_n^2 are the powers of the transmitted signal, the i-th received interfering signal, and the thermal noise at the receiver, respectively.

The corresponding access (backhaul) throughput $R^{\mathrm{A}}_{j,k}(t)$ ($R^{\mathrm{B}}_{j,k}(t)$) reads

$$R^{\mathrm{A}}_{j,k}(t) = \frac{1}{T_{sub}} \sum_{l=1}^{B^t_j} \mathbb{1} \left\{ \hat{b}_l(\gamma_{j,k}) = b_l \right\}, \tag{8.2}$$

where B^t_j denotes the number of bits transmitted from user (IAB node) j to gNB k during subframe t and $\hat{b}_l(\gamma_{j,k})$ is the l-th decoded bit at the receiver, as a function of $\gamma_{j,k}$.

Our goal is to maximize the average system sum-rate, defined as

$$\bar{R} \doteq \frac{1}{T} \sum_{j=1}^{N_{\mathrm{I}}} \sum_{t=1}^{T} R^{\mathrm{B}}_{j,0}(t), \tag{8.3}$$

by tuning the carrier frequency (either mmWave or THz) of each backhaul link. We remark that in this metric we take into account only the packets which are received at their final destination, i.e., the IAB donor.

$$\underset{\boldsymbol{P},\{\boldsymbol{S}(t)\}_t,\boldsymbol{T}}{\operatorname{argmax}} \ \bar{R}, \tag{8.5a}$$

$$\text{s.t. C1: } R^{\mathrm{B}}_{j,k}(t)T_{sub} \leq B_j(t) \ \ \forall\ j, \ \ \forall\ t \tag{8.5b}$$

$$\text{C2: } B_j(t+1) = \ B_j(t) + T_{sub}\left(\sum_{k=1}^{N_{\mathrm{U}}} R^{\mathrm{A}}_{k,j}(t) + \sum_{k=1}^{N_{\mathrm{I}}} R^{\mathrm{B}}_{k,j}(t) - \sum_{k=0}^{N_{\mathrm{I}}} R^{\mathrm{B}}_{j,k}(t)\right) \ \ \forall\ j, \ \ \forall t \tag{8.5c}$$

$$\text{C3: } \sum_{k=0}^{N_{\mathrm{I}}} \boldsymbol{S}\ [j,\ k]\ (t) + \sum_{k=1}^{N_{\mathrm{I}}} \boldsymbol{S}\ [k,\ j]\ (t) \leq 1 \ \ \forall\ j, \ \ \forall t \tag{8.5d}$$

$$\text{C4: } R^{\mathrm{B}}_{j,k}(t)\boldsymbol{S}\ [j,\ k]\ (t) = R^{\mathrm{B}}_{j,k}(t) \ \ \forall\ j, \ \ \forall k, \ \ \forall t \tag{8.5e}$$

$$\text{C5: } \sum_{j,k=0}^{N_{\mathrm{I}}} \boldsymbol{T} \leq \rho_{max} \sum_{j,k=0}^{N_{\mathrm{I}}} \boldsymbol{P} \tag{8.5f}$$

8.2.1 Channel Models

mmWave channel model

For the mmWave links, we consider the 3GPP 38.901 SCM [2], which models MIMO wireless channels for frequencies between 0.5 and 100 GHz.

In particular, [2] outlines the procedures for generating a channel matrix $\boldsymbol{H}_{s,d}$ whose entries $h^{j,k}_{s,d}$ correspond to the impulse response of the channel between the j-th radiating element of the antenna array of the transmitter (S), and the k-th radiating element of the antenna array of the receiver (D). Then, the channel matrix entries are combined with a frequency-flat path loss term PL.

When considering analog beamforming at both the transmitter and the receiver, the equivalent channel response $h^{mW}_{s,d}$ can be evaluated as

$$h^{mW}_{s,d} = \sqrt{10^{PL/10}} \cdot \boldsymbol{w}_d \boldsymbol{H}_{s,d} \boldsymbol{w}_s, \tag{8.4}$$

with $\boldsymbol{w}_s$ and $\boldsymbol{w}_d$ the beamforming vectors used at S and D, respectively.

THz channel model

For sub-THz, we use the physics-based channel modeling approach from [169], which includes molecular absorption and path loss. At THz-band frequencies, molecular absorption,

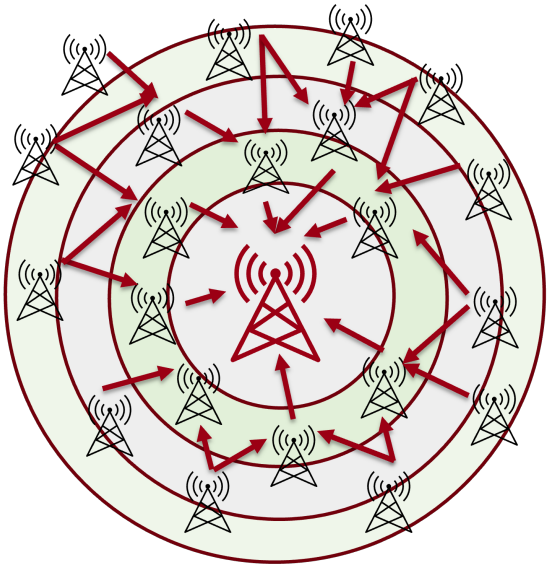


Figure 8.1: THZ link Selection based on tier and load

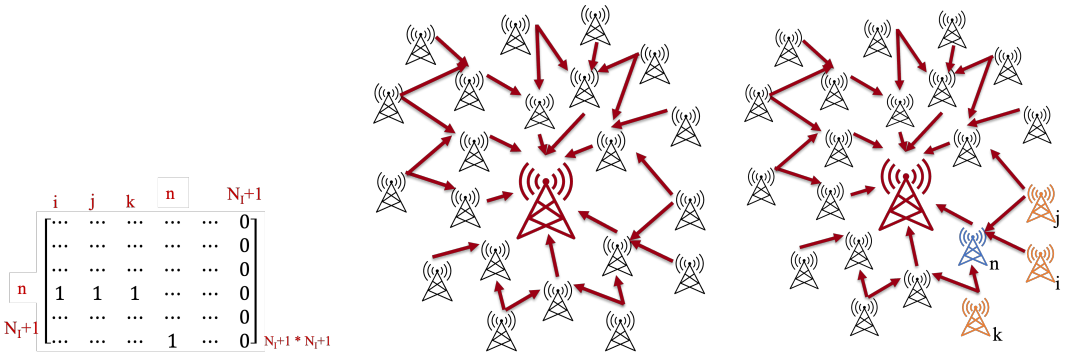
which causes both molecular absorption loss and molecular absorption noise, is the principal factor affecting electromagnetic wave propagation. $h_{s,d}^{tH}$ is the THz-band channel model introduced in [169], with additional transmit and receive antenna gains G_S and G_D, and is given by

$$h_{s,d}^{tH}(f,\, d) = \frac{c}{4\pi f d} \exp\left(-\frac{k_{abs}(f)d}{2}\right) G_S G_D, \tag{8.6}$$

where c stands for the speed of light and k_{abs} for the medium's molecular absorption coefficient, based on the type and composition of molecules [170].

8.3 Sum-rate optimization via THz Link Selection

We define $\boldsymbol{P} \in \{0,\, 1\}^{N_{\mathrm{I}}+1 \times N_{\mathrm{I}}+1}$ as the matrix which represents the possible active links among gNBs, i.e., $\boldsymbol{P}[i,\, j] = 1$ if and only if the wireless backhaul link between gNBs i and j is a feasible link; index 0 refers to the donor. Similarly, $\boldsymbol{S}(t) \in \{0,\, 1\}^{N_{\mathrm{I}}+1 \times N_{\mathrm{I}}+1}$ and $\boldsymbol{T} \in \{0,\, 1\}^{N_{\mathrm{I}}+1 \times N_{\mathrm{I}}+1}$ represent the links which are active during subframe t, and whether they use THz spectrum or not, respectively. Our objective is to maximize the average system sum-rate, by choosing whether each link is operating in the THz or the mmWave bands and the active links at each subframe. We perform the choice of $\boldsymbol{T}$ and $\boldsymbol{P}$ only once, with the goal of reducing the computational complexity of the algorithm.



(a) Example Matrix (Mx) of possible path deployment **(b)** Primary Network Map **(c)** Network Map associated with Mx

Figure 8.2: Example of Distance Aware Path Generation process and Matrix generation

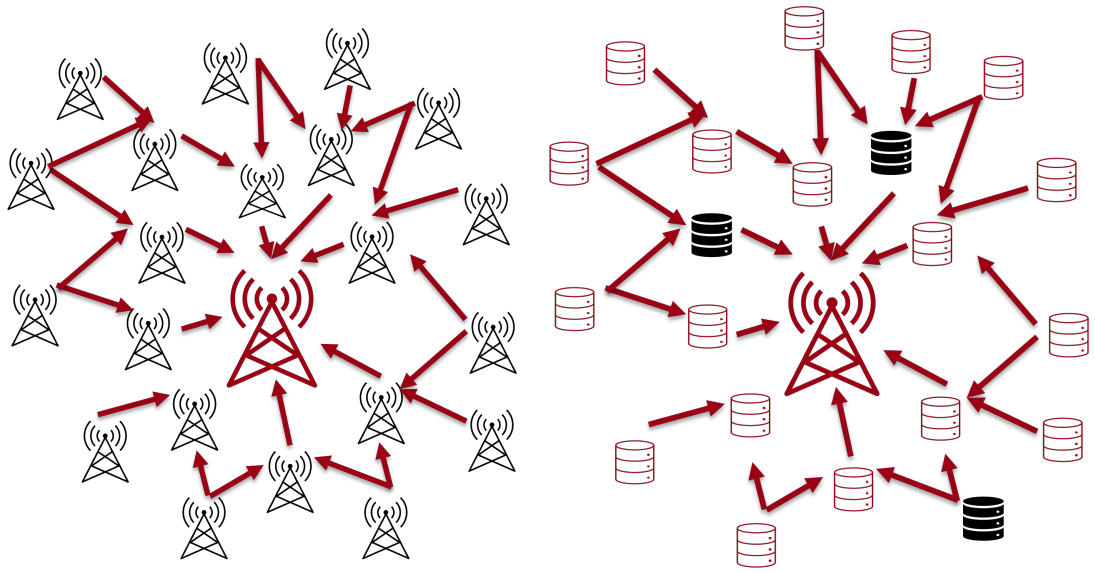
The optimization problem is thus formulated as (8.5a). Constraint C1 ensures that nodes do not transmit more data than available in their buffer. C2 enforces the proper evolution over time of the buffers occupancy, i.e., the buffer occupancy at time t must be equal to the one at subframe $t-1$, minus (plus) the outgoing (incoming) traffic from other nodes. Constraint C3 relates to the TDMA mode of operation, and ensures that each backhaul RF chain is used at most for one transmission/reception at any given subframe, while C4 imposes that only active links can exhibit a positive rate. Finally, with C5 we set an upper bound ρ_{max} on the maximum percentage of THz links.

8.3.1 Backhaul Scheduler

We remark that due to the binary nature of the $\boldsymbol{P}$, $\boldsymbol{S}(t)$ and $\boldsymbol{T}$ optimization variables, (8.5a) is an Integer Linear Program (ILP), thus NP-hard and not solvable in polynomial time. Therefore, in this section, we present a set of algorithms that solve the path selection and configuration problem heuristically and with low complexity.

Specifically, we first describe the pre-processing steps, referred to as *distance-aware path generation* (Alg. 8.1) and *THz-link selection* (Alg. 8.2), which prune the set of possible links established among gNBs and decide which of them are to operate in the THz bands, respectively. Then, we describe the *SINR-based scheduler* (Alg. 8.3), which differs from the former procedures as it is executed in each subframe to track the dynamic nature of the backhaul network.

The distance-aware path generation algorithm computes the $\boldsymbol{P}$ matrix, which encodes the potential connections between IAB nodes. $\boldsymbol{P}$ reduces the system complexity by restricting



(a) Primary Network Map

(b) Sorted nodes respect to their loads and SINR scheduler

Figure 8.3: Example of SINR-based Scheduler

possible paths from each IAB node and by avoiding looping. Specifically, Alg. 8.1 iterates over each IAB node n_j, establishing a connection towards the donor whenever the distance between them is smaller than d_{max}, i.e., a scenario- and frequency-dependent distance that guarantees a link performance above a certain threshold. In our case, the considered scenario involves a small and dense deployment of IAB nodes, so the path loss distance can be compensated by antenna gain, and d_{max} for THz and mmWave are assumed to have the same value. Moreover, the proposed pre-processing step performs additional attachments from neighboring nodes to n_j, as long as the resulting link exhibits a lower length compared to a direct connection toward the donor. In the opposite case, i.e., whenever a node n_k, neighbour of n_j, is closer to the donor than to the former, an attachment from n_j towards n_k is performed. Even though the latter may be topologically redundant, it can provide an alternative route for load balancing purposes, while still avoiding the creation of cycles.

The THz link selection policy identifies bottleneck links based on two heuristics: 1) links involving IAB nodes which are closer to the donor are more likely to be congested since they are usually used also for relaying traffic of subtending nodes; and 2) the average buffer occupancy provides an estimate of the loads incurred on each link. Accordingly, Alg. 8.2 partitions the IAB nodes into disjoint sets, referred to as *tiers*. Nodes are assigned to tiers based on their distance with respect to the donor, with tier 0 indicating the closest level to the donor. Then, the various backhaul links are marked as THz in descending order with

respect to the tier of the corresponding transmitting node, until the maximum ratio of non-mmWave links ρ_{max} is reached. Note that the algorithm may eventually reach a tier whose IAB nodes can not be all set as THz. In this case, ties within the same tier are broken by sorting its nodes with respect to their average traffic load, which we estimate by measuring the respective buffers. That is to say, nodes with higher buffer occupancy are given priority and thus they are set as THz before nodes exhibiting a lower traffic load. Note that this procedure can be based on long-term statistics, thus averaging the load of the nodes over multiple frames.

Finally, the SINR-based scheduler dynamically allocates resources, with the objective of maximizing the average sum rate by choosing a list of paths to be activated in each subframe. The rationale behind the proposed scheme is to schedule links based on their load. Specifically, in Alg. 8.3 we assign a transmission resource allocation priority which is directly proportional to the buffer occupancy of the transmitting node. Once the first endpoint is chosen, we determine the outgoing link by selecting the one with the highest SINR among those calculated in Alg. 8.1. Then, we set all links involving the corresponding transmitting and receiving nodes as infeasible (assigning zero to the corresponding transmitting (n) and receiving node (p_n^*) indexes in $\boldsymbol{P}_{temp}$), and repeat the procedure by considering the remaining nodes and links only, thus ensuring that the TDMA constraint is satisfied.

8.4 Performance Evaluation

This section introduces a performance evaluation based on a novel simulation setup (Sec. 8.4.1) in a dense cellular network (Sec. 8.4.2), with a comparison between different results of THz and mmWave networks (Sec. 8.4.3).

8.4.1 Simulation Setup

We have developed a system-level simulator that runs on top of Sionna [1], an open-source TensorFlow-based GPU-accelerated toolbox, and that includes the IAB networks described in Rel. 17. The proposed simulator, which is written in Python, is a system-level simulator that features 3GPP-compliant channel modeling and lower layers of the protocol stack. However, it lacks the implementation of 5G NR higher layers. Therefore, we added system-level functions like MAC-level scheduling and RLC-level buffering [51]. In addition, in this research we use the Terasim channel simulator [170] to generate channel responses and integrate them into Sionna. To accomplish this, we generate traces for each IAB node's channel

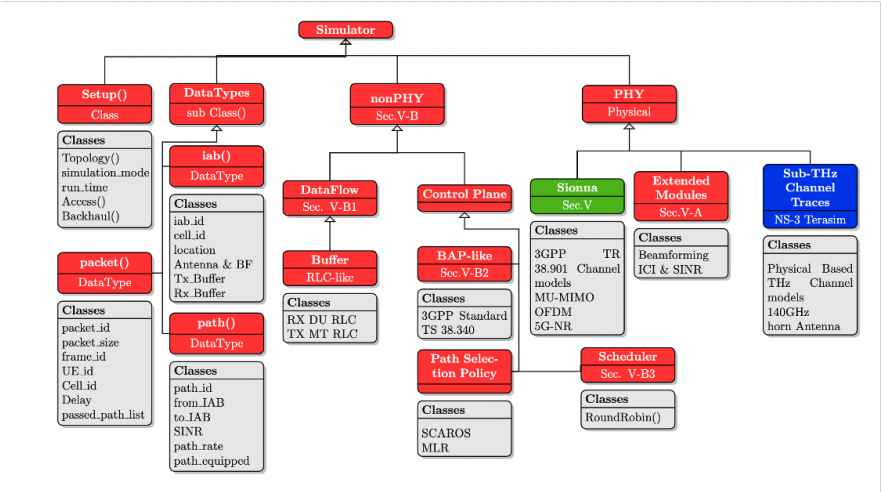


Figure 8.4: Network Simulator Extended to support sub-THz links

response and load them into Sionna. Terasim channel model integration allows us to generate channels up to 10 THz; in this simulation campaign, the sub-THz carrier frequency is 140 GHz. Several system-level KPIs, including latency, throughput, and packet loss rate, are produced by our simulator.

8.4.2 Simulation Scenario

We take a dense cellular base station deployment into account in our models. As shown in Fig. 8.5, we place IAB nodes at a density of 150 gNB/km^2, thus with an average intersite distance of 40 m. In Table 9.5, the specific simulation settings are displayed. For mmWave, we used the channel model outlined by 3GPP in TR 38.901 [2], a statistical 3GPP channel model for 0.5-100 GHz, while for sub-THz we used the THz-band channel model introduced in [169] and detailed in Sec. 8.2.1. The range of the user rate is 20 Mbps to 500 Mbps. We used a phased array antenna for mmWave and a horn antenna for THz, respectively. In mmWave we do beamforming based on a pre-generated codebook, in order to find the best beam pair for connection. For the purposes of SINR calculation, we assume that each interfering device utilizes the beamforming vector with the greatest SINR towards its intended target. In a similar fashion, both the transmitter and the receiver utilize the beamforming configuration calculated by the hierarchical search technique. We consider a scenario with a single donor to focus on the issues related to the bottleneck in the air interface of the donor itself, while extension to multiple donors is left for future work. We also set $d_{max} = 70$ m, as it has been experimentally shown that sub-THz links can operate in this range also in adverse

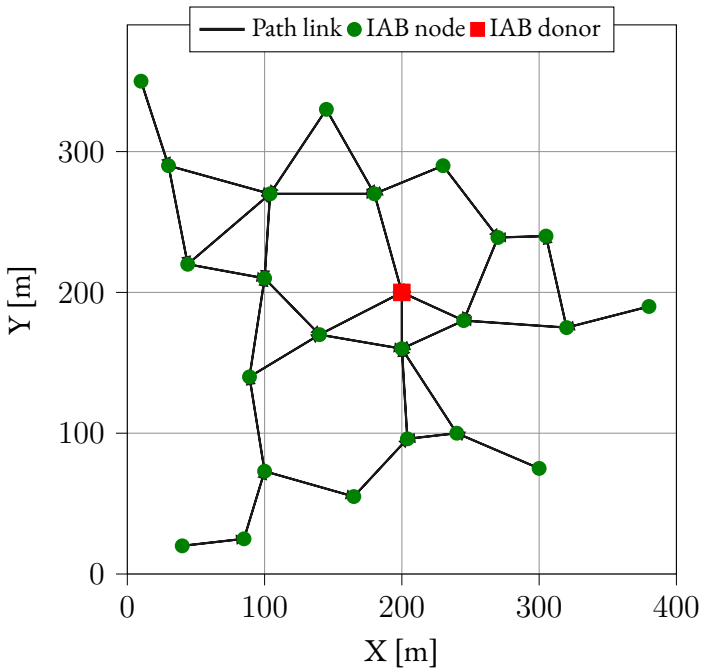


Figure 8.5: Simulation Scenario

weather conditions [171].

8.4.3 Results

In this section we report the outcomes of our numerical evaluation, focusing on end-to-end metrics measured at the IAB donor. We compare the performance achieved by different backhaul configurations, i.e., different maximum ratios of THz links and bandwidth, in terms of throughput, latency and packet drop ratio. We consider two baselines: *Random Scheduler (RS)* and *Random Links (RL)*. The former uses Alg. 8.2 and chooses at random a feasible set of active links during each subframe. On the contrary, RL randomly picks which links to set as THz, and uses Alg. 8.3 for scheduling. 10 simulations per configuration are executed, to obtain estimates which are averaged over the realizations of the wireless channels.

Fig. 8.6 reports the UE throughput achieved by the proposed solution, versus that achieved by RS and RL. Focusing on the former, it can be seen that Alg. 8.3 leads to a throughput increase of up to 40% compared to a random scheduling policy, thanks to the prioritization of the backhaul links incurring a higher load and exhibiting a higher number of subtending IAB nodes. Moreover, Alg. 8.2 introduces an additional throughput increase of up to 15% compared to RL.

Table 8.1: Simulation parameters.

Parameter	Value
Carrier frequency (mmWave)	28 GHz
Bandwidth (mmWave)	400 MHz
Carrier frequency (THz)	140 GHz
Bandwidth (THz)	{10, 32} GHz
IAB RF Chains	2 (1 access + 1 backhaul)
Pathloss model (mmWave)	UMi-Street Canyon [2]
Pathloss model (THz)	Physics-based [169]
Number of IAB nodes N_{I}	23
Number of users N_{U}	50
Per-UE source rate	{40, 80, 100, 200} Mbps
ρ_{max}	{0, 0.1, 0.3, 0.5, 0.7, 1}
gNB antenna array	8H × 8V
UE antenna array	4H × 4V
gNB and UE height	15 m and 1.5 m
gNB antenna gain (mmWave)	30 dB
gNB antenna gain (THz)	38 dB
Noise power	10 dBm

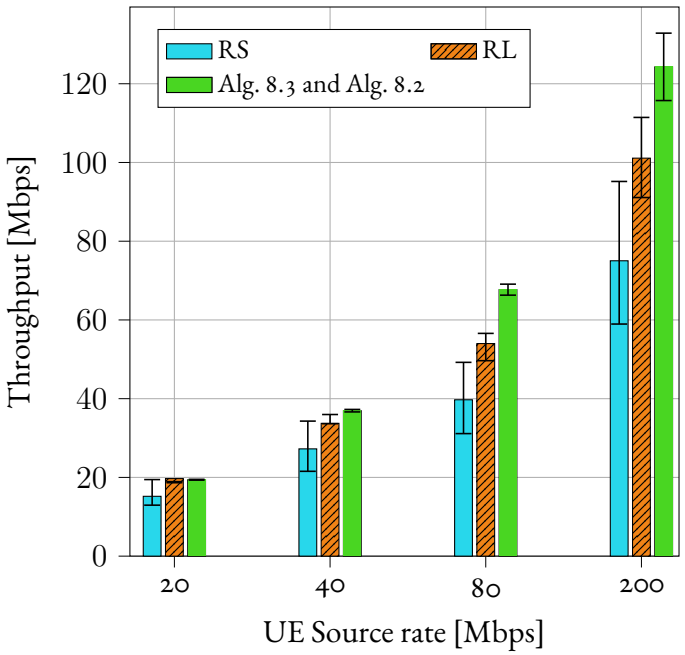


Figure 8.6: Throughput per UE for different schedulers and THz link selection policies, for THz bandwidth 32 GHz and $\rho_{max} = 0.3$.

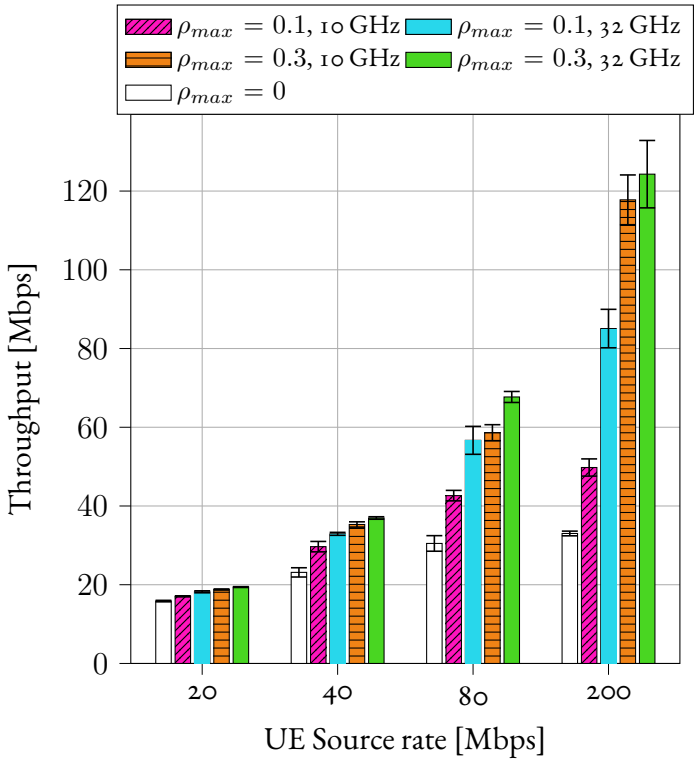


Figure 8.7: Throughput per UE for different configuration.

Fig. 8.7 illustrates the UE throughput for various configurations of sub-THz backhauling links and different UE source rates. The performance always improves by adding more bandwidth to the system through sub-THz links, despite the harsher propagation environment at higher frequencies. The performance gap increases with the user source rate. Indeed, mmWaves successfully sustain a source system rate of 1 Gbps (20 Mbps for 50 UE), but cannot match higher source rates, as the capacity saturates. The configuration with sub-THz links achieves a higher throughput in all scenarios and in particular for $\rho_{max} = 0.3$, 32 GHz achieves the highest throughput for all source rates. It is obvious that increasing the bandwidth improves the performance; nevertheless, increasing the percentage of the THz links from $\rho_{max} = 0.1$ to $\rho_{max} = 0.3$ has a more significant impact on throughput. This may be explained by considering the effects of replacing bottleneck backhauling mmWave links with THz links with higher bandwidth.

Similar considerations can be drawn from the results shown in Fig. 8.8, which reports the packet drop percentages for various backhaul configurations. The highest and lowest packet drop percentages across all UE source rates are achieved when using the mmWave

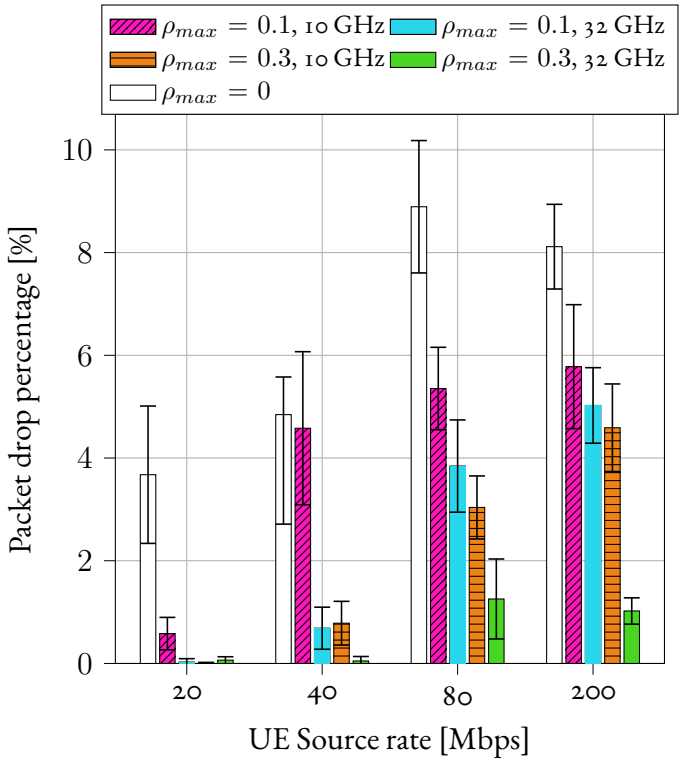


Figure 8.8: Backhaul packet drop percentage for different configurations.

and $\rho_{max} = 0.3$, 32 GHz configurations, respectively. Packet drop percentages at 20 Mbps source rates are close to zero for all configurations. The highest packet drop percentages among configurations including THz is $\rho_{max} = 0.1$, 10 GHz. It is noteworthy that the system performance is influenced directly by both the THz bandwidth and the link ratio, as seen in Fig 8.7.

Fig. 8.9 depicts the ECDF of the E2E latency experienced by packets which reach the donor, for different bachkaul configurations. Accordingly, both latencies accumulated over the fronthaul and backhaul links are taken into account, from the time packets are generated at the UE until they eventually reach the IAB donor. The plot shows that packet latency decreases as more sub-THz links are added to the network. In accordance with the aforementioned observations (Fig. 8.7 and Fig. 8.8), $\rho_{max} = 0.3$, 32 GHz has the lowest latency, whereas mmWave has the highest latency. The average latency for $\rho_{max} = 0.3$, 32 GHz, $\rho_{max} = 0.3$, 10 GHz, $\rho_{max} = 0.1$, 32 GHz, and $\rho_{max} = 0.1$, 10 GHz is approximately 51%, 24%, 24%, and 18% less than in mmWave.

Finally, the average system throughput for different ratios ρ_{max} of THz link is shown in

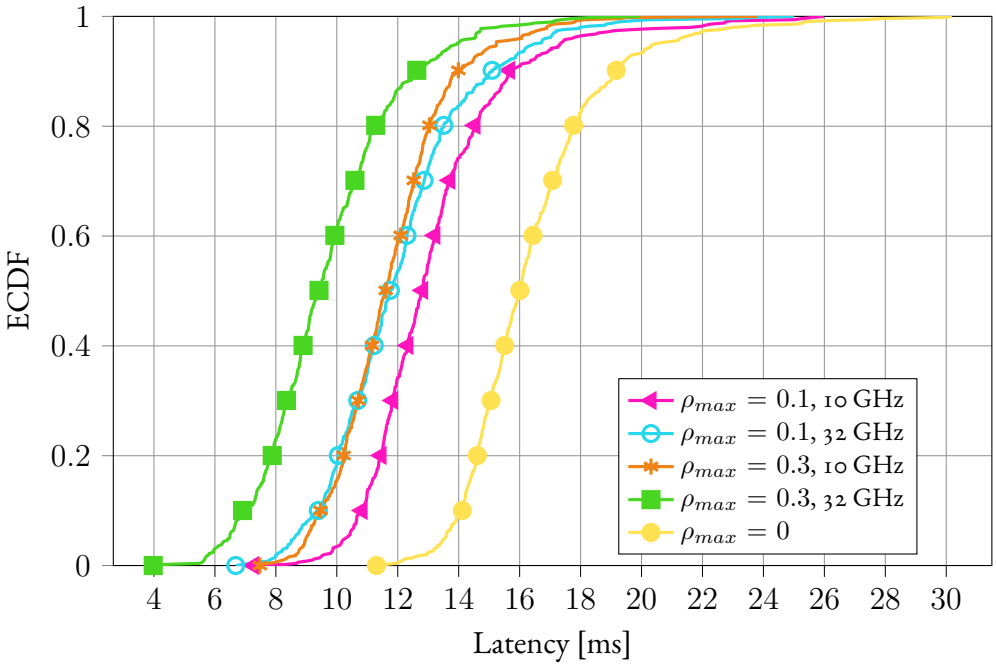


Figure 8.9: E2E latency ECDF for different configurations for 80 Mbps user rate.

Fig. 8.10. The system throughput increases with the inclusion of additional THz links. The figure also shows that system source rates of 2 Gbps, 4 Gbps, and 10 Gbps can be satisfied by a single donor when ρ_{max} is properly set. However, the larger demand of the 25 Gbps system source rate still cannot be satisfied, as the system becomes saturated. $\rho_{max} = 0.1$ and 1 can increase system throughput by up to four times and twelve times, respectively.

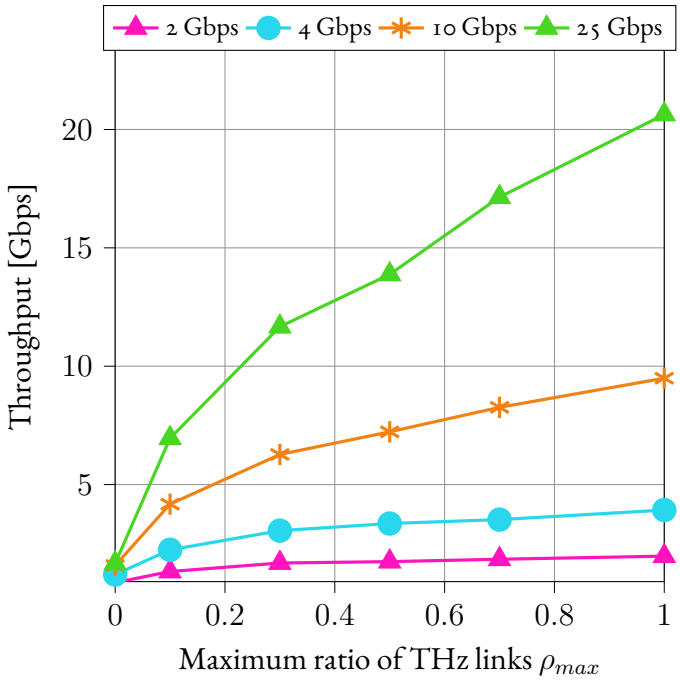


Figure 8.10: System throughput for different source rate and ratio of THz links.

Algorithm 8.1 Distance Aware Path Generation

$\boldsymbol{P} = [0]_{N_\mathrm{I}+1 \times N_\mathrm{I}+1}$
for $n_i = 1, 2, \ldots, N_\mathrm{I}$
  $d_i \leftarrow$ 3D distance between n_i and IAB donor
  if $d_i < d_{max}$
    $\boldsymbol{P}[n_i, 0] = 1$
  for $n_j = n_1 + 1, \ldots, N_\mathrm{I}$
    $d_{i,j} \leftarrow$ 3D distance between n_i and n_j
    if $d_{i,j} < d_{max}$
      $d_j \leftarrow$ 3D distance between n_j and IAB donor
      if $d_i < d_j$
        $\boldsymbol{P}[n_j, n_i] = 1$
      else
        $\boldsymbol{P}[n_i, n_j] = 1$

Algorithm 8.2 THz Link Selection

$\boldsymbol{N}_{\mathrm{T}}$ = Vector of IAB nodes tier index
$\boldsymbol{N}_{sort}$ = Vector of IAB node indexes, sorted with respect to their load
$\boldsymbol{T} \leftarrow [0]_{N_{\mathrm{I}}+1 \times N_{\mathrm{I}}+1}$
for $n = 1, 2, \ldots, N_{\mathrm{I}}$
    $d \leftarrow$ 3D distance between n and IAB donor
    $\boldsymbol{N}_{\mathrm{T}}[n] \leftarrow d \,//\, d_{tier}$
for $i = 1, \ldots, \max(\boldsymbol{N}_{\mathrm{T}})$
    $\boldsymbol{N}_T^i \leftarrow \{j \mid \boldsymbol{N}_T[j] == i\}$
    $\boldsymbol{L}_i \leftarrow$ links in $\boldsymbol{P}$ where nodes of $\boldsymbol{N}_T^i$ are the transmitting node
    if $\sum_{j,k} \boldsymbol{T}[j,\, k] + \dim(\boldsymbol{L}_i) < \rho_{max} \sum_{j,k} \boldsymbol{P}[j,\, k]$
        $\boldsymbol{T}[j,\, k] \leftarrow 1 \; \forall \; (j,\, k) \in \boldsymbol{L}_i$
    else
        while $\sum_{j,k} \boldsymbol{T}[j,\, k] < \rho_{max} \sum_{j,k} \boldsymbol{P}[j,\, k]$
            $n^* \leftarrow \min_n \; \mid \boldsymbol{N}_{sort}[n] \cap \boldsymbol{N}_T^i \neq \emptyset$
            $(n^*,\, k) \leftarrow$ link $\in \boldsymbol{L}_i \mid n^*$ is the transmitting node
            $\boldsymbol{T}[n^*,\, k] \leftarrow 1; \boldsymbol{L}_i \leftarrow \boldsymbol{L}_i \setminus (n^*,\, k)$

Algorithm 8.3 SINR-based Scheduler

$\boldsymbol{N}_{sort}$ = Vector of IAB nodes, sorted with respect to their load
$\boldsymbol{P}_{temp} = \boldsymbol{P}$
$\boldsymbol{S}(t) \leftarrow [0]_{N_{\mathrm{I}}+1 \times N_{\mathrm{I}}+1}$
for n in $\boldsymbol{N}_{sort}$
    $\gamma_{max} \leftarrow -\infty$
    for i in $0, \ldots, N_{\mathrm{I}}$
        if $\gamma_{n,i} > \gamma_{max}$
            $\gamma_{max} \leftarrow \gamma_{n,i}$
            $p_n^* \leftarrow i$
    $\boldsymbol{S}(t)[n,\, p_n^*] \leftarrow 1$
    $\boldsymbol{P}_{temp}[:,\, n],\, [n,\, :] \leftarrow [0]; \boldsymbol{P}_{temp}[:,\, p_n^*],\, [p_n^*,\, :] \leftarrow [0]$

9

Statistical Analysis and End-to-End Performance Evaluation of Traffic Models for Automotive Data

In this chapter, we will describe recent research on end-to-end lidar communication studies. First, we will provide a brief introduction to recent works and research motivation. After that, we will propose a statistical model of HSC models. In addition, the models were implemented in ns-3 to assess full-stack analysis. The section concludes with numerical and statistical results.

9.1 Introduction

Autonomous driving will redefine the way we perceive, interact with, and utilize vehicles on the roads, and support the transition from human-dependent to machine-driven transportation. This approach will reduce accidents, enhance traffic, reduce fuel consumption, and provide newfound mobility options for individuals with disabilities and the elderly [172].

Unlike conventional vehicles where a human driver takes control, autonomous vehicles will be equipped with a complete suite of sensors and powerful computing units to perceive the environment, make driving decisions, and navigate autonomously [173]. Besides cameras, Light Detection and Ranging (LiDAR) sensors are often used, as they are the most pre-

cise systems to measure range, and robust under almost all lighting and weather conditions with or without glare and shadows [174].

Notably, autonomous vehicles will implement object detection algorithms for detecting and classifying objects and obstacles in the surroundings [175, 176], including cars, pedestrians, and road signs, which allow the vehicles to make real-time decisions on how to navigate safely and avoid collisions. In this context, more robust scene understanding could be achieved if vehicles exchanged their sensor data to other vehicles and/or roadside infrastructures, a concept that is usually referred to as cooperative perception [177, 178], thereby enhancing the perception range of the vehicles beyond the capabilities of on-board instrumentation. However, object detection and cooperative perception come with at least two main concerns.

1) Availability of data for training. Object detection usually consists of different subtasks such as clustering, blob analysis, image segmentation, and feature extraction [179], which involve machine learning or deep learning to produce meaningful results. This approach requires the availability of massive amounts of labeled data for proper training, which are often expensive and time-consuming to generate [180]. Among many options, SemanticKITTI [181] is a popular open-source dataset providing multimodal (stereo RGB and LiDAR) data for road scenes. The dataset has more than 23 201 scenes for training, and 20 351 for testing, which makes it one of the largest publicly available datasets for autonomous driving research.

Overall, SemantiKITTI consists of around 80 GBytes of data, which is not easy to store, handle, and process. In particular, transmitting sensor data, e.g., for cooperative perception, can be challenging with standard Vehicle-To-Everything (V2X) communication technologies [182]. For example, a raw LiDAR perception frame in SemantiKITTI is, on average, 18 Mb: with a frame rate of 10 fps, it would produce a data rate of around 130 Mbps. For comparison, 3GPP C-V2X and IEEE 802.11p, i.e., the de-facto standards for V2X communication, can only offer a nominal data rate a few tens of Mb/s [183]. One possible method to solve capacity issues is by compressing data before transmission, which in turn introduces additional complications. First, compression may sacrifice accuracy to reduce the file size, with severe implications for operations that rely on data such as object detection. Second, data compression requires point-level processing of data, which may not be implemented in real time. For example geometry-based point cloud (G-PCC) [184], a possible standard for point cloud compression, can compress only 440k points/s [185] (for comparison, the HDL-32 LiDAR sensor used in SemantiKITTI captures 1.3M points/s). Moreover, compression involves expensive and energy-consuming hardware (especially high-end graphics

processing unit (GPUs)), that may not be available onboard the vehicles. Finally, while compressing data from cameras is relatively straightforward, there is no accepted standard for compression of LiDAR data.

2) Performance validation. To establish the accuracy and functionality of the algorithms developed for autonomous driving, it is imperative to subject them to a rigorous process of validation. In this sense, validation with real testbed is impractical due to limitations in scalability, flexibility, and the high cost of hardware components. Theoretical analyses, in turn, often introduce conservative and/or unrealistic assumptions on the system model, and may lead to wrong or misleading conclusions. On the contrary, simulations in a sandbox environment has the advantage to reduce costs and time consumption for validation, and would facilitate the research process. In particular, system-level simulators enables end-to-end simulations with considerations related to the full protocol stack. Notably, ns-3 [186] has gained great popularity within the network simulation community. It consists of a large set of predefined, ready-to-use, open-access modules to simulate different parts of the network, as well as different types of (wireless) systems, thus enabling scalable simulations. It also comes with modules to simulate V2X networks based on the most recent 3GPP specification for NR V2X [187], mobility traces using SUMO [188], and a simulation pipeline able to simulate and test machine learning algorithms within the network [189]. As such, it stands out as one of the most complete 5G-oriented tools to perform accurate simulations in the context of vehicular networks.

However, computer simulations need accurate modeling of the different components of the network at all the layers. As long as the application layer is concerned, autonomous driving requires sending automotive data, e.g., from the SemantiKITTI dataset. This requires: (i) the dataset to be available and stored on the simulator machine; (ii) the dataset to be accessed, saved, and processed on the RAM as the simulation progresses, which is source of a delay proportional to the size of the dataset; and (iii) the data to be converted into a stream of packets to be sent between two simulated network nodes. Another alternative is to consider synthetic traffic according to a statistical model of how the application operates. This is in contrast to a network simulation using packet-level traffic, since now the traffic is simulated based on the sole arrival process of packets (especially the packet size and the inter-packet time) according to some mathematical process. Our early results indicate that a single ns-3 simulation of 15 s takes on average around 605 s (10 min) to complete using automotive data from SemantiKITTI, vs. only 20 s using the corresponding statistical model. In the literature, statistical methods have been proposed to model, for example, the propagation of the

signal [74], the application (e.g., for web browsing [190] or, more recently, XR traffic [191]) or, in the vehicular domain, automotive radar reflections [192] or multi-sensor data fusions systems [193].

However, to date, there is no universal model for automotive data generated from sensors, especially from LiDARs. Ideally, this can be modeled as periodic traffic with a fixed frame rate and constant frame size proportional to the resolution of the sensor. However, automotive data may be compressed before transmission to reduce the burden on the channel, resulting in frames of variable size. Moreover, data may be possibly correlated both in space and time, which makes the model aperiodic. Therefore, automotive data should be modeled as a combination of random variables, after proper fitting via statistical methods.

9.1.1 Contributions

Based on the above introduction, the contributions of this paper can be summarized as follows.

- We provide a realistic statistical characterization of automotive data, specifically, of the size of LiDAR frames. To the best of our knowledge, this is the first model for automotive traffic, and is based on 4 Gb of data from the SemanticKITTI dataset.Given the importance of compression in the automotive scenario, we provide seven different models to characterize raw data and six representative compression configurations. Specifically, data is compressed using the state-of-the-art HSC algorithm, first proposed in [194], which supports different levels of compression to trade off efficiency against speed. We claim that the availability of stochastic models for automotive data brings several advantages compared to using actual data, including faster simulations and processing at the application layer, and no to limited storage of data.

- We quantify the accuracy of our statistical models. Specifically, we test different distributions, and identify the corresponding fitting parameters. The measure of accuracy is assessed via a custom statistical test based on the MLE, the Kolmogorov-Smirnoff, and the Bootstrap Resampling schemes [195]. Our results show that the uncompressed dataset can be accurately modeled with to a tLocationScale random variable, while the compressed datasets can be modeled based on the tLocationScale, Nakagami, Logicstic or Gamma distributions according to the level of compression. Interestingly, six of the seven models pass the test.

- We validate the accuracy of our statistical models via ns-3 simulations. First, we extend the ns-3 code base with new custom methods to generate random variables according to the tLocationScale, Nakagami, and Logicstic distributions, which are not natively available in ns-3. Then, we simulate the transmission of automotive data between two vehicles as a function of their distance. At the application layer, we consider the transmission of both packet-level (raw or compressed) data from the SemanticKITTI dataset, or synthetic data from our statistical traffic models, and evaluate the impact of the accuracy of the latter on some network metrics. We observe that, even though only six of the seven models pass the test, all the models provide similar, if not the same, results than using real data in terms of end-to-end latency and throughput. This shows that statistical models are a valid alternative to simplify network simulations while ensuring that results are statistically strong and accurate.

The rest of the chapter is organized as follows. In Sec. 9.2 we present some related works. In Sec. 9.3 we formalize our statistical analysis to identify accurate stochastic models for automotive data. Moreover, we evaluate the accuracy of those models based on the Kolmogorov-Smirnov test based on the Bootstrap method. In Sec. 9.4 we evaluate the impact of the accuracy of the models on the network under several metrics.

9.2 Related Work

Communication among vehicles is one of the key elements of the autonomous driving revolution. The combination of the highly dynamic and heterogeneous road environment and of the strict requirements imposed by the envisioned use cases [196] makes traditional communication systems and protocols unsuitable for V2X systems. Although it is clear that the traffic load generated by autonomous vehicles calls for novel network solutions [183, 197], and the 3GPP has already outlined the requirements, the data stream of the sensors for autonomous cars has yet to be fully characterized. A complete traffic characterization is necessary to choose the appropriate communication technology and to perform accurate network dimensioning. Furthermore, as network functions become increasingly virtualized, traffic modeling has become central for the assignment of the network resources [198]. Finally, a generative traffic model can be used in network simulations, greatly reducing the economic costs and the time required to design and test novel solutions and architectures. For these reasons, modeling the network traffic has been a central research topic for the last decades.

Authors in [199] present an extensive list of envisioned 5G applications and their corresponding traffic models, with limited references to the autonomous driving scenario. Several works model the data stream for Internet of Things (IoT) applications and Machine to Machine (M2M) traffic. Specifically, authors in [200] model the aggregated M2M traffic and the corresponding message delivery delay in a 5G network. On the contrary, [201] and [202] focus on the modeling of a single traffic source, deriving stochastic models based on Markov Chains. The survey in [203] reports an overview of traffic models used for peer-to-peer (P2P) information exchange in IoT blockchain networks. The traffic generated by video streaming has also been extensively investigated [204]. [205] proposes using nonlinear autoregressive models to predict the future frame size. Besides classic video applications, with the emergence of eXtended Reality (XR), the first efforts in characterizing the interactive video traffic have started [206, 207]. In addition to the more classical models based on state machine and autoregression techniques [201, 202, 208], more recently Machine and Deep Learning techniques were employed to model complex traffic streams. For instance, authors in [209] combine a deep-belief network with a compressed-sensing approach to predict the fast and the slow varying components of the traffic in a wireless mesh network. Other works in [210, 211] introduce long short-term memorys (LSTMs) to capture and replicate different traffic flow patterns.

For the specific case of automotive sensor data, few works exist which give a complete characterization of the network traffic or even of its magnitude. [208] introduced a machine-type communication traffic model specific to automotive traffic, that is spatially and temporally correlated. However, due to the limited number of available automotive applications, the model is quite general to guarantee the flexibility needed to reflect any resulting arrival process. [197] reports possible ranges of the datarate for different sensors, obtained from the specifications of commercial products and from conversations with industrial partners. Authors in [212] measure the mean frame size, frame rate, data bandwidth, and delay of the experimental demo system presented in the paper. Similarly, the datarates for the commercial Velodyne HDL-64 are reported in [213]. 9.1 summarizes the datarates reported in these works. However, this information is only partial, and a complete characterization of the traffic generated by sensor data remains vague, making it difficult to evaluate and design new systems.

	[197]	[213]	[212]
LiDAR	$10 - 100$	9.75 (5 Hz) 39 (20 Hz)	167 (9.95 Hz)
Camera	100700 (raw), 1090 (compressed)	-	10.87 (4.36 Hz)
Radar	< 1	-	-

Table 9.1: Data stream characterization in existing literature. All values are in Mbps.

9.2.1 Hybrid Semantic Compression (HSC)

The compression pipeline described in Sec. 9.2.1 exploits the semantic understanding of the 3D scene to further reduce its size. The deep learning module (RangeNet++) contributes to prioritize safety-critical over less relevant information, while the compression module (Draco) exploits the 3D spatial correlation of the data to represent it efficiently. The high number of parameters involved results in a rich set of compression levels and models, which make HSC a flexible solution for vehicle-to-vehicle and vehicle-to-infrastructure scenarios. However, the complexity required by a system to dynamically control such models significantly increases with their numerosity: approximately five hundreds different models can be identified combining the Draco parameters and HSC priority levels. Moreover, to quantify the benefits of the proposed framework from a network perspective, end-to-end system-level simulations are required. In general, such simulations can be time-consuming and computationally heavy, considering that HSC needs the inference processing of a large DL model (e.g., [214]) to be performed for the semantic analysis. Thus, hardware requirements to run those simulations become prohibitive, and often deep learning accelerators, like GPUs or TPUs, are needed.

To enable a more efficient scheduling system and to reduce the amount of resources at simulation time needed to compute network performance, we propose:

- A refinement of the all possible models by inspecting the rendered point-clouds after Draco compression, and clustering similar models together based on the average encoding and decoding time, and the average file size produced (Sec. 9.2.2);

- A statistical-based analysis of the simulated network traffic generated by individual HSC models. Our goal is to generate statistically representative traffic patterns with-

out requiring the complete pipeline computation. In this way, we can significantly save time and reduce energy consumption during the simulation process (Sec. 9.3).

9.2.2 Compression Model Refinement

Semantic understanding of data is a key component for the development of reliable autonomous vehicles [194]. The context of the sensed/received data, such as the location and orientation of objects in the environment, as well as the relationships between different obstacles in the scene, play a vital role in enabling a safe and efficient navigation. In the recent years, the adoption of Deep Learning models to address data semantic understanding has achieved remarkable results [215, 216, 217, 218, 219]. However, the inference performance of these statistical models can decrease as the input data deviates from the training set statistics, for example due to lossy compression of the input [220, 221, 222]. This poses a challenge in the considered scenario, as the lossy compression performed by Draco in the HSC paradigm can significantly alter the data distribution, potentially compromising the semantic performance. In this regard, we perform a qualitative analysis to assess the influence of Draco compression on the degradation of the point-clouds*.

Draco primary relies on two parameters: the Quantization Parameter (QP) and the Compression Level (CL). While QP specifies the number of bits used for quantizing input values, CL trades-off compression with encoding/decoding speed. Assuming that in a lossy compression pipeline the main source of distortion is quantization, we investigate the effect of QP on the compressed point, and we study the encoding/decoding time as function of CL. Thereby, we select a subset of parameters as *qualitatively acceptable and most relevant* for LiDAR automotive data.

Quantization Analysis

We perform a quantitative and qualitative analysis. In both cases, point-clouds are (i) encoded at a default CL value of 7, for different quantization parameters, $\mathrm{QP} \in \{0, 1, \ldots, 14\} \subset \mathbb{N}$, and (ii) decoded. Specifically, for the quantitative analysis, we compute

- *the number U of unique points in the compressed point-cloud*: Draco preserves the number of points between input and output, but as a result of the compression process,

*In general, quantifying the impact of quantization on the DL model inference performance is challenging, and requires the definition of proper metrics, which are out of the scope of this work.

it condenses multiple points into a single one, reducing the number of unique points in the compressed point-cloud;

- *the Point-to-Plane PSNR (p2p-PSNR)* [223] to quantify the geometric distortion of point-cloud compression.

Both the quantities are computed and averaged over a sample of 550 point clouds uniformly selected from SemanticKITTI dataset. In the qualitative analysis we render[†] the point-cloud for visual inspection. Fig.9.1 concisely summarizes the results of our analysis, showing U and the p2p-PSNR as function of QP, and the rendered point-clouds for QP $= \{8, 9, 10, 12\}$. We observe that employing compression settings with QP values below than 10 proves to be excessively aggressive, significantly impairing the visual perception of rendered point clouds, thus leading to considerable challenges in object identification. Furthermore, at QP $= 10$ we empirically notice a change in the slope of the p2p-PSNR curve corresponding to ≈ 50 dB, and a reduction of the number of unique points post-compression by a factor of six. Hence, based on these observations we selected values for Draco QP to be $\{10, 11, 12, 13, 14, 0\}$, where QP $= 0$ means no quantization. It is worth noting that our analysis is consistent with the algorithm description in the Draco repository[‡], where it is stated that *"most projects can set quantization values of about* 11 *without any noticeable difference in quality"*.

Compression Level Analysis

In order to gain a deeper understanding of the impact of compression on point clouds, a similar methodology is applied to the CL. Specifically, for each QP previously identified as qualitatively acceptable, the file size, encoding and decoding time are calculated for CL values in $\{0, 1, \ldots, 10\}$ using a sample of 550 point clouds uniformly selected from the SemanticKITTI dataset. The metrics are computed as the average over the sample and are depicted in Fig.9.2. Consistently with the parameters' definition, for a fixed QP, the file size exhibits relatively minimal changes as function of CL. A more relevant impact can be noticed for encoding/decoding time: in fact, for a fixed QP, we can observe a growing trend with increasing CL. In particular, the encoding/decoding curves related to QP $\in \{12, 13, 14\}$ have a similar behaviour. For this reason, we cluster them together and choose the corner case QP $= 14$ as cluster representative. From a CL perspective, we notice a relevant increment in the

[†]The point-clouds are rendered using Open3D [224].
[‡]https://github.com/google/draco

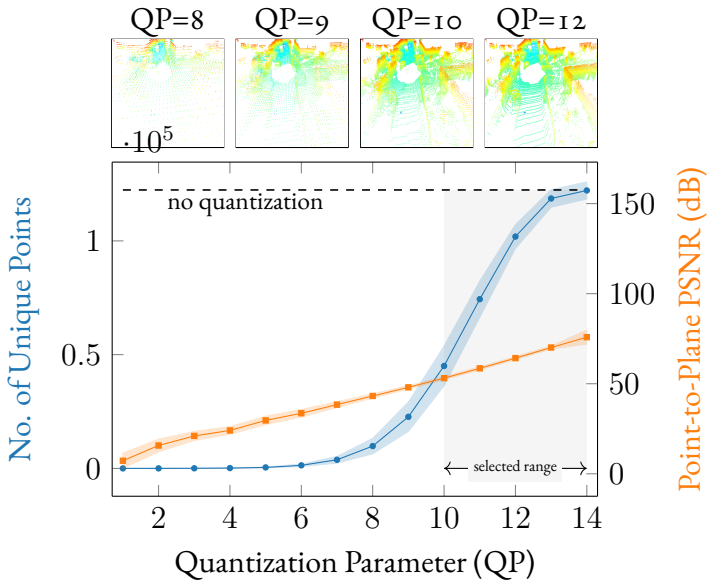


Figure 9.1: QP refinement.

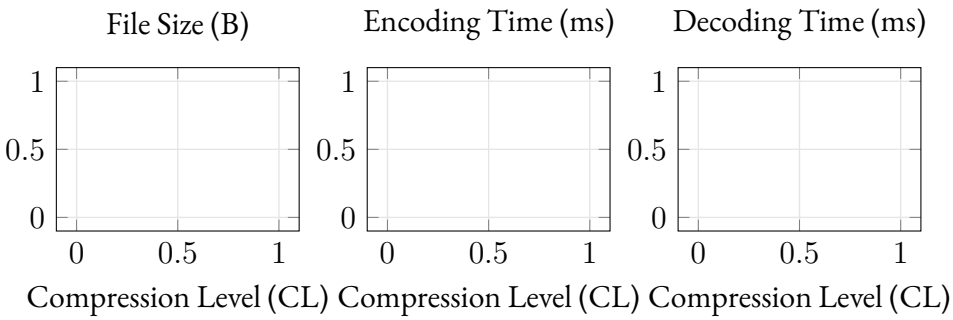


Figure 9.2: Compression Level analysis

encoding/decoding curves at CL $= 5$ without excessively affecting the file size. Therefore, we fix CL $= 5$ as it represents a sweet-spot between file size and encoding/decoding time.

In summary, we restrict the Draco parameter set to those that are deemed *qualitatively acceptable and most relevant* for LiDAR automotive data, specifically QP $\in \{0, 10, 11, 14\}$, CL $\in \{5\}$. Combined with the semantic priority levels of HSC {HSC-0, HSC-1, HSC-2}, we define 12 models, each of them identified with the following syntax: DX/SY, where X $\in \{11, 14\}$ represents the quantization parameter used by (D)raco, and Y $\in \{0, 1, 2\}$ the (S)emantic priority level of HSC. Finally, to simplify the traffic modelling at simulation time, we further reduce the models considered. Specifically, we focus on a entire HSC exploration for QP $\in \{0, 14\}$, leading to D0/S0, D0/S1, D0/S2, D14/S0, D14/S1, D14/S2, while for

QP $\in \{10, 11\}$ we consider D11/S0, which represents a pure Draco compression with no Downlink (DL) semantic module. In this way, the model space is effectively reduced to 7, allowing us to investigate whether the elimination of non-critical points by the semantic DL module in D14/S1 and D14/S2 yields comparable results to the scenario in which all points are retained but with reduced bit allocation during the quantization phase, i.e., D11/S0.

9.3 Statistical Models for Automotive Data

Considering that network-related metrics (e.g, channel occupancy, transmission latency) rely on the size of the transmitted file after compression, we conduct a statistical analysis of the traffic generated by HSC when semanticKITTI [181] is used as input data. We provide a CDF of the size of the transmitted file after compression, for each of the model defined in Sec. 9.2.2. In this way, according to the model selected in the running simulation, the size in Mb of the compressed 3D scene can be statistically generated, thus saving computation and time resources. To do so, we consider a rich set $\mathbb{P}$ of theoretical CDF families (see Tab.**??**) as hypothesis about the target unknown CDF F. Then, we test whether our hypothesis agrees with the CDF F_N of a random sample of N observations drawn from F. If it is not the case, this is the evidence that our hypothesis should be rejected. This is a well-known procedure referred to as goodness-of-fitting test. In particular, we perform a KS goodness-of-fitting test [**?**], whose statistics compare the hypothetical CDF $F_i(x)$ and the observed one $F_N(x)$ as follows

$$D_N = \max_x |F_N(x) - F_i(x)|. \tag{9.1}$$

In general, for continuous CDF and assuming that the sample comes from F_i (null hypothesis), $F_N \to F_i$ as $N \to \infty$ (strong law of large numbers), thereby D_N converges to zero almost surely. Moreover, under the same hypothesis, the limiting distribution of $\sqrt{N}D_N$ converges to a Kolmogorov distribution, regardless of F_i. As a result, under the null hypothesis and for N sufficiently large, (9.1) has a known distribution, and tabular critical values can be used.

We denote the set of 3D point-clouds in SemanticKITTI as

$$\mathcal{D} = \{p_j : p_j \in \mathbb{R}^{3\times n_j},\ j = 1, \ldots, N\}, \tag{9.2}$$

where n_j is the number of points in the point cloud $p_j \in \mathcal{D}$. We consider each HSC model (denoted by DX/SY) as a function $m : \mathcal{D} \to \hat{\mathcal{D}}$ that receives as input a point-cloud $p \in \mathcal{D}$

and outputs its compressed version $\hat{p} \in \hat{\mathcal{D}}$. Let's also denote the set of compressed file sizes through the model m as

$$B_m(\hat{\mathcal{D}}) = \{b_j \in \mathbb{R} : \; b_j = g(\hat{p}_j), \tag{9.3}$$

$$\hat{p}_j = m(p_j) \in \hat{\mathcal{D}}, \; j = 1, \ldots, N\}, \tag{9.4}$$

where $g(p)$ corresponds to the number of bits used to encode p (namely, the file size of p). The goal of our analysis is to find, for each HSC model m, the class of functions $\mathcal{P}_i = \{F_i(\,\cdot\,;\theta) : \theta \in \Theta_i\} \in \mathbb{P}$, and the correspondent CDF $F_i(\cdot, \theta_i)$, $i \in \{1, 2, \ldots, |\mathbb{P}|\}$ (our hypothesis) that best fit $B_m(\hat{\mathcal{D}})$ (sample of observation) according to some evaluation metric $h_m : \mathcal{P}_i \to \mathbb{R}$. In other words, our goal is performing a goodness-of-fit test for each class $\mathcal{P}_i$ in the presence of unknown parameters $\theta \in \Theta_i$, and comparing the evaluation metric obtained for each class. For example, if $\mathcal{P}_i$ is the family of Normal distributions, $\theta_i = (\mu, \sigma^2) \in \mathbb{R}^2$.

We construct the statistical test by first estimating the parameters $\hat{\theta}_i$ within each class $\mathcal{P}_i$, thus identifying a representative $CDF \; F_i(\cdot, \hat{\theta}_i)$, and then performing a KS goodness-of-fitting test across the representatives. However, estimating the parameters θ from the observed data may alter the limiting distribution of the test statistic in (9.1), which will become dependent on such parameters and, even under the null hypothesis, will drift from the Kolmogorov distribution, thereby causing critical values shift [**?**]. As a consequence, using tabular values in this scenario may lead to statistical inconsistent conclusions. For this reason, we adopt a parametric bootstrap re-sampling scheme as proposed in [**? ?**]. We now outline the mathematics underlying these techniques, and carefully describe each step.

Parameter Estimation

We perform, for each class $\mathcal{P}_i \in \mathbb{P}$, a MLE from the observed data $B_m(\hat{\mathcal{D}})$ to estimate $\hat{\theta}_i = \theta_i^{ML}$. We select the CDF $F_i(\cdot\,;\hat{\theta}_i) \in \mathcal{P}_i$ as representative for the class $\mathcal{P}_i$, thus as the model being tested.

Target KS statistic

We define the composite null hypothesis of the statistical test as

$$H_0^i : data are drawn from F_i(\cdot\,;\hat{\theta}_i), \; \forall i,$$

and we compute the KS statistic as follows

$$D_N(i) = \max_{x \in B_m(\hat{\mathcal{D}})} |F_N(x) - F_i(x; \hat{\theta}_i)|, \tag{9.5}$$

where $F_N(\cdot)$ is the empirical CDF of the observed data $B_m(\hat{\mathcal{D}})$. It is worth noticing that the hypothetical CDF $F_i(\cdot; \hat{\theta}_i)$ depends on $B_m(\hat{\mathcal{D}})$, and $D_N(i)$ is no longer distribution-free. Therefore, a bootstrapping re-sampling scheme is necessary.

Parametric Bootstrap Re-sampling Scheme

We construct L independent bootstrap resamples $\{B^*_{m,l}(\hat{\mathcal{D}})\}_{l=1}^{L}$ from the estimated population $F_i(\cdot; \hat{\theta}_i)$, and for each $B^*_{m,l}(\hat{\mathcal{D}})$, MLE is performed to estimate $\hat{\theta}^*_{i,l}$. Our goal is to compute, for each m and i, L KS statistics and compare them in a statistical sense with the target KS statistic in (9.5). Let's denote the empirical CDF of $B^*_{m,l}(\hat{\mathcal{D}})$ as $F^*_{N,l}(x)$, then the KS statistic computed on the l-th bootstrap resample is

$$D^*_{N,l}(i) = \max_{x \in B^*_{m,l}(\hat{\mathcal{D}})} |F^*_{N,l}(x) - F_i(x, ; \hat{\theta}^*_{i,l})|, \tag{9.6}$$

$\forall i \in \{1, 2, \ldots, |\mathbb{P}|\}$. Since $\sqrt{N}\big(F_N(x) - F_i(x; \hat{\theta}_i)\big)$ and $\sqrt{N}\big(F^*_{N,l}(x) - F_i(x, ; \hat{\theta}^*_{i,l})\big)$ converge to the same Gaussian process, then $\sqrt{N}D_N(i)$ and $\sqrt{N}D^*_{N,l}(i)$ have the same limiting distribution [**?**]. Therefore, the corrected critical values to test (9.5) at a significance level α can be obtained computing the $(1-\alpha)$-percentile of $\big\{\sqrt{N}D^*_{N,l}(i)\big\}_{l=1}^{L}$. Equivalently, one can compute the bootstrap p-value $\pi^*_{i,m}$ as

$$\pi^*_{i,m} = \frac{1}{L}\sum_{l=1}^{L} 1_{\{D^*_{N,l}(i) > D_N(i)\}} \tag{9.7}$$

$$= 1 - \hat{F}_{D^*_N(i)}(D_N(i)) \approx \text{Prob}\big(D^*_N(i) > D_N(i)\big), \tag{9.8}$$

where $\hat{F}_{D^*_N(i)}(D_N(i))$ is the empirical CDF of $\big\{D^*_{N,l}(i)\big\}_{l=1}^{L}$. Therefore, H^i_0 can be rejected at significance level α if $\pi^*_{i,m} < \alpha$, i.e., $\text{Prob}\big(D^*_N(i) \leq D_N(i)\big) > 1 - \alpha$. In simple words, for small value of α, we reject the null hypothesis if the probability of observing high values

for KS statistic is greater than $1 - \alpha$[§].

Model Selection

The KS statistical test with the bootstrap re-sampling scheme described in Sec.9.3 provides, for each model m, $|\mathbb{P}|$ statistical test outcomes, one for each distribution family. We define the set of indexes of distribution families for which the null hypothesis cannot be rejected as $\mathcal{J}_m$. For each $i \in \mathcal{J}_m$, we select as best fit the CDF family $\mathcal{P}_{i^\star_m}$ and $F_{i^\star_m}(\cdot, \hat{\theta}_{i^\star_m})$ such that $i^\star_m = \arg\max_{i \in \mathcal{J}_m} \pi^*_{i,m}$. If $\mathcal{J}_m = \emptyset$, we select as best fitting model the one that minimize the Normalized Root Mean Square Error (NRMSE), computed as

$$\text{NRMSE} = \sqrt{\frac{\sum_{k=1}^{N} \left(F_N(x_k) - F_i(x_k;\, \hat{\theta}_i)\right)^2}{\sum_{k=1}^{N} \left(F_N(x_k) - \bar{F}_i(\hat{\theta}_i)\right)^2}}, \tag{9.9}$$

where $\bar{F}_i(\hat{\theta}_i) = \frac{1}{N} \sum_{k=1}^{N} F_i(x_k;\, \hat{\theta}_i)$.

9.3.1 Results

In this section we present and discuss the results obtained from the statistical analysis. Tab. **??** shows the bootstrap p-values $\pi^*_{i,m}$ computed according to (9.7) for each HSC model m and CDF family i. The statistical analysis is carried with a significance level $\alpha = 0.01$, and for $L = 1000$. For each model m, we represent in bold the p-values that pass the test, i.e., $\pi^*_{i,m} \in \mathcal{I}_m$, and we highlight in grey the entries selected as best fit, that is $i^\star_m$. It is worth noticing that the model Do/So does not pass the test ($\mathcal{I}_m = \emptyset$), meaning that the KS statistic will assume high values almost surely. However, for consistency, we choose as best fitting the distribution with the minimum NRMSE, i.e., the tLocationScale distribution achieving a NRMSE $= 7.4e^{-3}$. In Tab. **??** we report the MLE parameters $\hat{\theta}_{i^\star}$ for the best fitting models. These values will be adopted for ns-3 implementation (Sec. 9.4) and the end-to-end perfomance evaluation (Sec. 9.5). In order to provide a comprehensive overview of the statistical results achieved, we complement the numerical analysis with fitting plots. Specifically, for each model, the empirical CDF of the observed data $F_N(\cdot)$ and the validated theoretical CDF $F_{i^\star}(\cdot;\, \hat{\theta}_{i^\star})$ are depicted in Fig. 9.3, together with the qq-plot (quantile-quantile plot). The latter represents a graphical diagnostic method to visually assess the numerical results ob-

[§]The reader should think the KS statistic as a "distance" between the the hypothetical distribution and the one of the observed data.

$\mathbb{P}$	D0/S0	D0/S1	D0/S2	D11/S0	D14/S0	D14/S1	D14/S2
BirnbaumSaunders	0	0.0	**0.061**	**0.013**	0.0	0.0	0.0
ExtremeValue	0	0.0	0.0	0.0	0.0	0.003	0.0
Gamma	0	0.0	**0.375**	**0.114**	0.0	0.0	**0.012**
GeneralizedExtremeValue	0	0.003	**0.039**	0.006	0.0	0.0	0.0
HalfNormal	0	0.0	0.003	0.0	0.0	0.0	0.0
InverseGaussian	0	0.0	0.002	**0.018**	0.0	0.0	0.0
Logistic	0	0.009	0.0	0.0	**0.036**	**0.024**	0.0
Loglogistic	0	0.0	**0.018**	**0.037**	0.01	0.0	0.0
Lognormal	0	0.0	**0.065**	**0.02**	0.0	0.0	0.0
Nakagami	0	0.001	0.0	**0.128**	0.0	0.0	0.004
Normal	0	**0.109**	0.0	**0.012**	0.0	0.0	0.0
Poisson	0	0.0	0.0	0.0	0.007	0.0	0.0
Rayleigh	0	0.0	0.0	0.0	0.0	0.0	0.0
tLocationScale	0	**0.029**	0.0	**0.014**	**0.026**	**0.026**	0.0
Weibull	0	**0.044**	**0.15**	0.0	0.003	0.0	0.01

Table 9.2: P-values $\pi^*_{i,m}$ computed according to (9.7), where $i \in \mathbb{P}$ represents the row index, while m is the column index and identify the HSC model under test. For each model m, we represent in bold the p-values that pass the test, i.e., $\pi^*_{i,m} \in \mathcal{I}_m$, and we highlight in grey the entries selected as best fit, that is $i^\star_m$.

tained by hypothesis testing. It plots on the x-axis the quantile function of one distribution and on the y-axis the quantile function of the other distribution. Therefore, if two distributions are identical, the qq-plot follows the bisector of the first quadrant. Although in Fig. 9.3 the Probability Density Function (PDF) and CDF of all the models fit the empirical data overall, the qq-plots visually highlight the local discrepancies between the theoretical model and the empirical observations. In fact, consistent with our previous numerical findings, the qq-plot of the model D0/S0 attributes the failure of the hypothesis testing to the heavy left tail of the empirical data with respect to the theoretical model. Conversly, for the rest of the models, only a light tail is observed, further diminishing with higher HSC semantic priority level. We interpret this trend as a consequence of the inherent data structure: as we increasingly leverage semantic information for compression, the qq-plot tail becomes progressively lighter.

m	Distribution $\mathbb{P}$	Parameters Θ	Values
D0/S0	tLocationScale	μ, σ, ν	3172.74, 64.41, 1.49
D0/S1	Normal	μ, σ	1458.7, 455.36
D0/S2	Gamma	a, b	1.87, 131.97
D11/S0	Nakagami	μ, ω	9.31, 4914.06
D14/S0	Logistic	μ, σ	197.54, 8.96
D14/S1	tLocationScale	μ, σ, ν	98.11, 16.83, 4.08
D14/S2	Gamma	a, b	2.81, 6.06

Table 9.3: MLE parameters for the statistical distributions selected.

9.4 ns-3 Implementation

9.4.1 Intro on ns-3

To address the need for a realistic performance evaluation of the statistical traffic model at mmWave spectrum, we develop the *StatisticalTraffic* module for popular ns-3 and make it publicly available[¶]. Integration with ns-3 makes it possible to study complex, end-to-end performance analysis of scenarios with several kinds of applications and mobility patterns. *StatisticalTraffic* module is a bursty traffic generator for HSC models which works on top of *BurstyApplication* [225]. *StatisticalTraffic* enables the full stack analysis and validation of different HSC models in various scenarios. In the following paragraphs, we will briefly describe the *BurstyApplication* module, and then discuss the integration of the *StatisticalTraffic* module.

Bursty Application

BurstyApplication is one of the first traffic model simulation platforms for VR applications which uses traffic traces obtained from industrial VR streaming software [225]. The *BurstyApplication* module enables generating and processing large bursts of packets; it continuously transmits data bursts, divided into smaller fragments of a given size. This framework enables a TX to send a packet that bursts fragmented into multiple packets, later re-aggregated in the RX. *BurstyApplication* can operate in different policies. For instance, it could send packets

[¶] https://github.com/signetlabdei/kitti-statistical-dataset

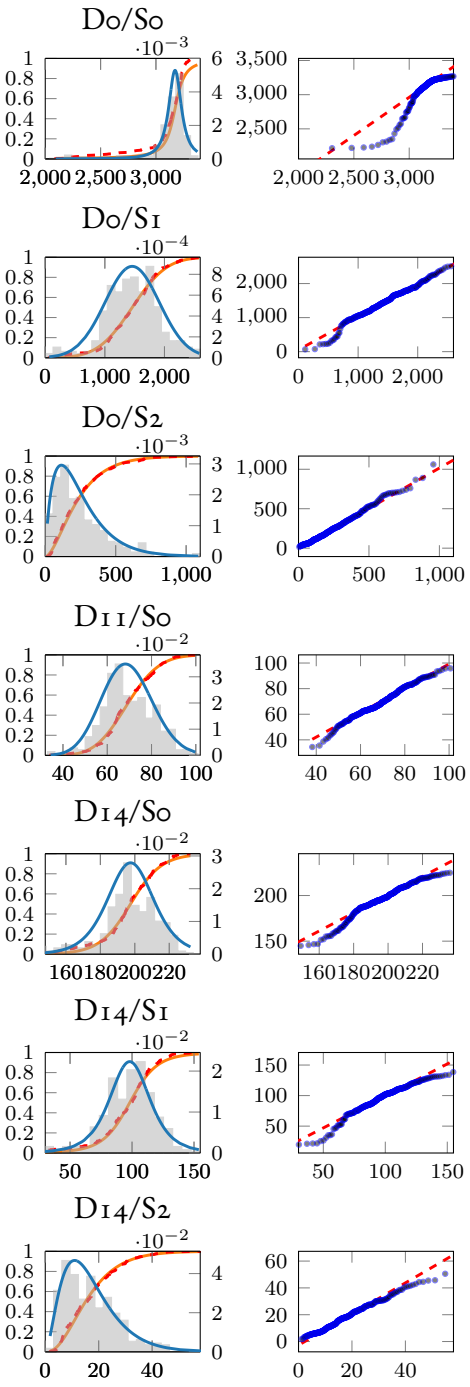


Figure 9.3: Fitting plots

in a deterministic or statistical fashion, indicating burst size and period time. A variety of

applications are made possible by the *BurstGenrator*, interface of *BurstyApplication*, which allows users to control the burst size and frame period.

Statistical Traffic Application

we developed *StatisticalTraffic* on top of *BurstyApplication* which is Application Programming Interface (API) for generating statistical traffics model in application layer. *StatisticalTraffic* periodically updates the *BurstSize* and *FramePeriod* based on prefer statistic traffics. *StatisticalTraffic* supports seven HSC traffic models in which the favored traffic model is determined by the *GetModel* interface, as depicted in Fig. 9.4. *StatisticModelInit* is in charge of initializing the random distribution parameters based on Table **??** (described in Sec. 9.3). Afterward, in each timestep *BurstSize* and *FramePeriod* is generated based on random distributions and feed to *BurstyApplication* by *BurstGenarator* API. Figure 9.4 summarizes the integration that we have made to support statistical traffic in ns-3. The remainder of the 5G NR protocol stack is emulated using the ns3-mmwave module [226]. This module implements a customized PHY layer that supports NR frame formats and numerologies, as well as a MAC layer that supports ad hoc beamforming and scheduling strategies. The RLC and Packet Data Convergence Protocol (PDCP) layers are based on the ns-3 Lena module and provide network tasks like as packet segmentation, re transmissions, and/or reassembly. Additionally, the module enables non-standalone deployments, handover, and mobility management through dual connection, as well as Carrier Aggregation (CA) at the MAC layer.

9.4.2 ns-3 implementation of the statistical distributions

Due to the statistical nature of *StatisticalTraffic*, random statistical distributions must be implemented directly in ns-3 in order to assess the full-stack performance of proposed traffic models. Although ns-3 has built-in support for many well-known distributions, including Normal, Exponential, Gamma, etc., not all random distributions related to proposed traffic methods are included. So, as a first step in implementing ns-3, we created a random number generator of tLocationScale, Nakagami, Logistic distributions. To achieve this, we use two techniques to create the necessary random distributions which are described in the following.

1) The built-in distributions of ns-3 implicitly exclude random distributions like Nakagami. We try to create new distributions for these kinds of distributions using built-in random distributions since this distribution can be explained by combining built-in distributions. 2) Since built-in distributions are unable to accurately express all random distribu-

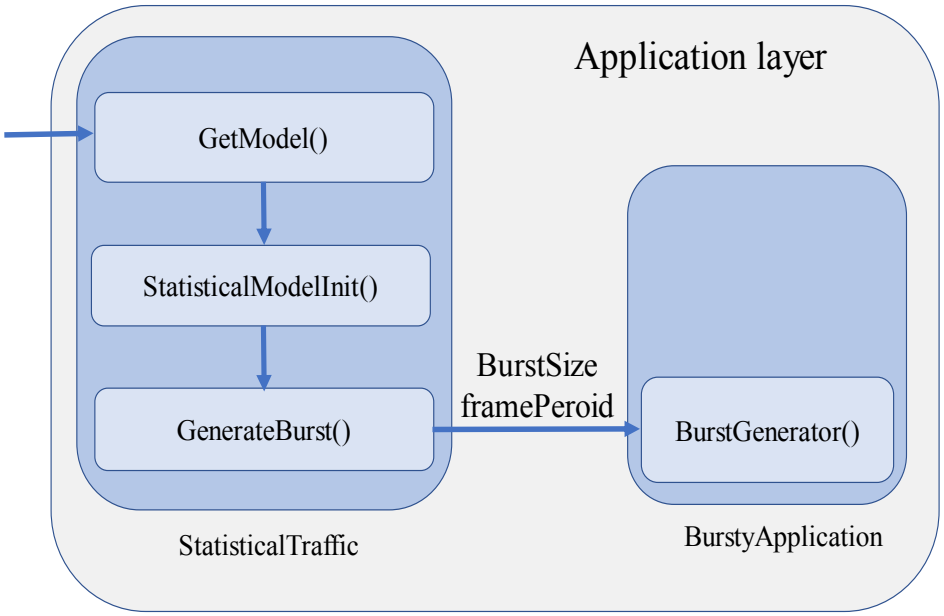


Figure 9.4: Application layer. how the *StatisticalTraffic* integrated to *BurstyApplication*

tions, we must implement them from scratch. In order to implement a random number generator, we used The Inverse Cumulative Distribution Function (Inverse CDF) sampling approach which is well-known for generating random variables from a specified distribution for its generality and simplicity. If the CDF of a probability distribution is known, the Inverse CDF creates a random sample of distribution using the procedures below. F is the mapping of the domain CDF to the interval on a one-to-one basis (0,1). If U is a uniform random variable on the interval $(0,\ 1)$, then $X = F^{-1}(U)$ has the same distribution as F. So to feed traffic models in ns-3, we generate random values using $F^{-1}(U)$ distributions with this approach.

1. tLocationScale [Inverse CDF] Direct implementation of Inverse CDF of tLocationScale is Np hard, so variant approaches for implementation of Inverse CDF of tLocationScale is proposed [227]. We implement central power series for the Inverse CDF which introduces in [227]. Equation 9.10 represents Inverse CDF of tLocationScale, where v is calculated in eq. 9.11 and $coef$ are coefficient of series which shown in Table **??**.

$$F_{tLoc}^{-1}(\mu,\, \sigma,\, v) = (\sum_{i=1}^{\infty} v^{2\times i+1} \times coef(i,\, v) + v) * \sigma + \mu \tag{9.10}$$

$$v = (u - \frac{1}{2})\sqrt{n \times \pi}\frac{\Gamma[n/2]}{\Gamma[(n+1)/2]} \tag{9.11}$$

2. *Logistic* [Inverse CDF]: In order to implement Inverse CDF of *Logistic* distribution, we adhere to the process outlined in [228]. Equation. 9.12 represents the CDF of the Logistic distribution with the input parameters μ and σ that indicates location and scale of distribution, respectively.

$$F_{logistic}(\mu,\, \sigma) = \frac{1}{1 + \exp\left\{-\frac{(x-\mu)}{\sigma}\right\}} \tag{9.12}$$

The Pseudo-version is obtained by solving the Eq. 9.13, where x is the unknown.

$$p = \frac{1}{1 + \exp\left\{-\frac{(x-\mu)}{\sigma}\right\}} \tag{9.13}$$

Equation. 9.14 represents the solution which obtained by solving Eq. 9.13.

$$x = \mu - \ln\,(\frac{1}{p} - 1) \times \sigma \tag{9.14}$$

Thus a sample X of Logistic distribution is obtained by formula in Eq.9.15.

$$X = \mu - \ln\,(\frac{1}{U} - 1) \times \sigma \tag{9.15}$$

where the U is output of the random generator. so the Inverse CDF of logistic distribution is obtained as Eq.9.16.

$$F_{logistic}^{-1}(\mu,\, \sigma) = \mu - \ln\,(\frac{1}{u} - 1) \times \sigma \tag{9.16}$$

3. Nakagami: Nakagami PDF can be explained explicitly based on Gamma PDF, which is built-in in ns-3. Equation 9.17 illustrates the relation between aforementioned dis-

Table 9.4: p-values of implemented distribution in ns-3

$\mathbb{P}$	p-value
tLocationScale	0.296
Logistic	0.880
Nakagami	0.702

tributions. So to generate a random variable for the Nakagami we can use a random number generator of the Gamma distribution.

$$F_{Nakagami}(\mu,\, w) = \sqrt{F_{Gamma}(\mu,\, \frac{w}{\mu})} \tag{9.17}$$

Where μ, w represents xx and xx, respectively.So we implemented (9.17) in ns-3 to generate random numbers for the Nakagami distribution.

9.4.3 ns-3 Implementation Evaluation

In order to assess the implementation of statistical distributions in the ns-3, an evaluation methodology was devised. The evaluation aimed to measure the accuracy and performance of the implemented distributions in ns-3. So the following steps were followed: To measure how well-implemented distributions' CDFs fit with the primary distribution, Kolmogorov-Smirnoff test statistics were developed. Kolmogorov-Smirnoff test metric allowed for a comprehensive assessment of the accuracy of the implemented distributions. Required data were gathered using the recently added random generator in ns-3(tLocationScale: Eq. 9.10, Logistic: Eq. 9.16, Nakagami: Eq. 9.17), and as a comparison, expected values were gathered using the *Scipy* module in Python. For validation, the generated data from the applied statistical distributions were compared to expected values or actual data. The collected data were statistically compared to evaluate the performance of the implemented distributions. The Kolmogorov-Smirnoff test statistics were computed to determine the level of similarity between the simulated results and the expected or real-world values, which are summarized in Table 9.4. This analysis provided insights into the accuracy and fidelity of the implemented statistical distributions. As it is depicted in Table 9.4, all implementations pass the test very well. As a result, the evaluation shows the strengths of the implemented statistical distributions, which show the reliability and validity of the ns-3 implementation of distributions.

Table 9.5: Network parameters used for the ns-3 simulation.

Parameter	Value
Carrier frequency	28 GHz
Bandwidth	200 MHz
Transmit Power	30 dBm
Channel Model	3GPP TR 38.901 (UMi-Street Canyon) [2]
LiDAR Inter Burst Interval	100 ms
Buffer Size	12 MB
Data direction	Uplink

9.5 Performance Evaluation

This section presents the simulation setup and the network metrics used for the validation of the ns-3 *StatisticalTraffic* Application, and the corresponding results.

9.5.1 Simulation Setup

To validate the *StatisticalTraffic* application, we consider a simple yet realistic urban scenario, designed to isolate and highlight the effects of the different statistical models: a vehicle transmits the LiDAR data to a gNB through a mmWave link. The two nodes have the same height, and the distance d between them is increased from 15 to 300 m as the vehicle moves away from the gNB.

The network simulation is run using ns-3, with the parameters reported in 9.5. We consider 28 GHz as the central frequency and 200 MHz of bandwidth. The channel is simulated according to the 3GPP standard channel model for frequencies ranging from 0.5 to 100 GHz [2]. Both the gNB and the vehicle are equipped with a 8×8 Uniform Planar Array (UPA). The beamforming vector at the TX and at the RX is based on the Singular Value Decomposition (SVD) of the channel matrix. The transmission power is 30 dBm.

At the application layer, the LiDAR bursts are generated every 100 ms. The data is compressed using HSC at the PDCP layer according to 7 representative the compression levels introduced in 9.3 (D0/S0, D0/S1, D0/S2, D11/S0, D14/S0, D14/S1, D14/S2) [194]. We measure the following E2E metrics at PDCP layer, which provides the first entry point for the packet stream:

- *E2E throughput*, measured as the ratio between the number of bytes received throughout the entire simulation and the total simulation time.
- *E2E latency*, measured as the difference between the time each packet is generated at

Table 9.6: Median source rate and throughput of the considered HSC compression levels. All values are measured in Mbps.

HSC Model	D0/S0	D0/S1	D0/S2	D11/S0	D14/S0	D14/S1	D14/S2
Source Rate	260.3	122.1	14.5	5.7	16.4	8	1.2
Throughput ($d < 50$ m)	259.4	124.3	20.2	5.6	17.3	8.6	1.4
Throughput ($d > 50$ m)	151	72.2	11.3	2.9	10	5	0.7

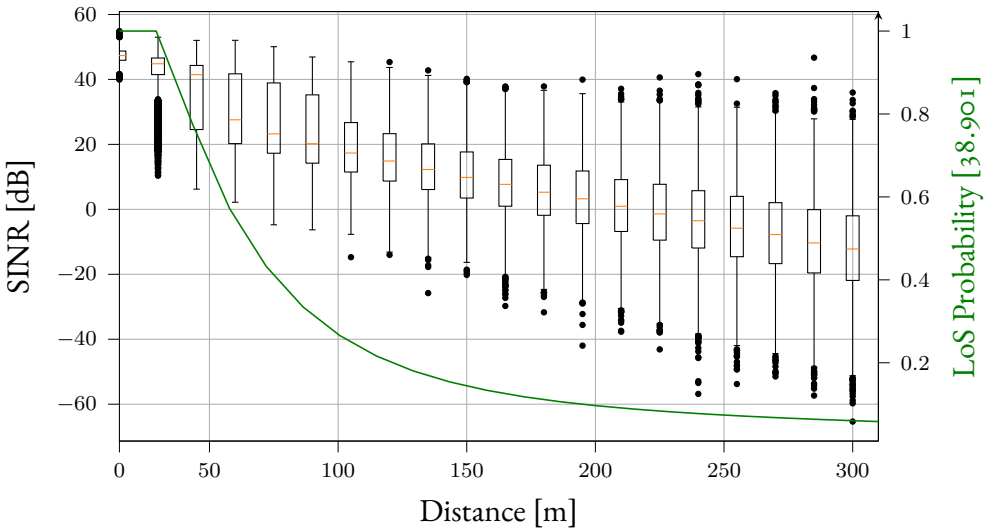


Figure 9.5: SINR at the receiver and LoS probability of the mmWave channel [2] at different TX-RX distances.

the application layer and when it is successfully received. Accordingly, it accounts for the transmission, compression, queuing, and decompression time.

9.5.2 Network Simulation results

In this section, we analyze the network performance obtained using Semantic Kitti [181]. The file size of the individual point clouds is extracted from the dataset and used as input to the *BurstyApplication* in ns-3, which in turn generates the packet stream according to the procedure outlined in 9.4.

In 9.5 we report the SINR at the RX as measured throughout the simulation with the settings presented in 9.5.1. We can observe that, as expected, at short distances the SINR presents very large values (around 50 dB) consistently, as the channel is mainly in LoS conditions. As the vehicle moves away from the gNB, the LoS probability (green line) decreases

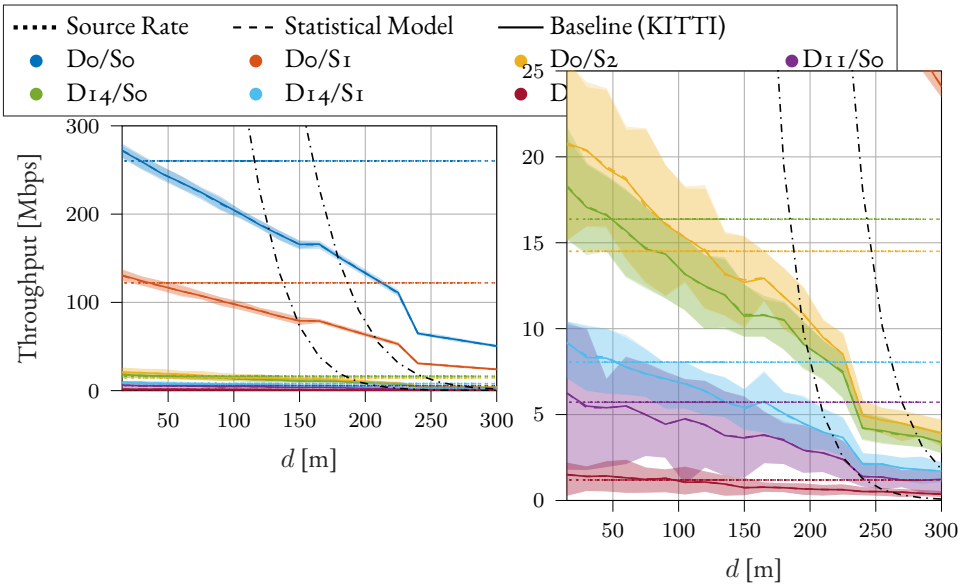


Figure 9.6: Median throughput (solid for KITTI and dashed for the statistical model) and corresponding confidence intervals (shaded areas) for different TX-RX distances and HSC compression levels.

d [m]	15-45	60	75	90	105	120	135	150	165	180	195	210	225	240	255	270	285
PRR	100.0	87.0	82.8	78.4	74.5	69.6	65.4	60.6	48.3	45.3	41.2	37.5	34.3	18.0	16.8	16.4	15.2

Table 9.7: Median PRR at the considered TX-RX distances.

according to [2], and the path loss increases. Accordingly, the SINR values have greater variance and lower values, down to -12 dB at 300 m.

This translates to a slow degradation of the theoretical capacity of the channel, that, considering the median (25% quantile) SINR, varies between 3.14 (3) Gbps, 2.75 (1.63) Gbps, and 16.42 (1.8) Mbps at 15, 45, and 300 m, respectively.

The first line of 9.6 reports the median source rate of the application, obtained by multiplying the median file size of the data compressed with the different HSC configurations by the frame rate of the LiDAR (10 Hz).[‖] We can observe that the current technologies for vehicular communications, like the Dedicated Short-Range Communication (DSRC) protocol [229] which has a 27 Mbps peak throughput, can not support the source rate of raw data (D0/S0) or of the lowest compression level (D0/S1). On the contrary, the mmWave channel can sustain the datarate of all the considered HSC levels. Indeed, the source rate

[‖]Note that, differently from the throughput, the source rate does not include the headers nor the communication overhead introduced by the protocol stack, which might lead to a throughput larger than the source rate in some cases (D0/S2, D14/S0–2).

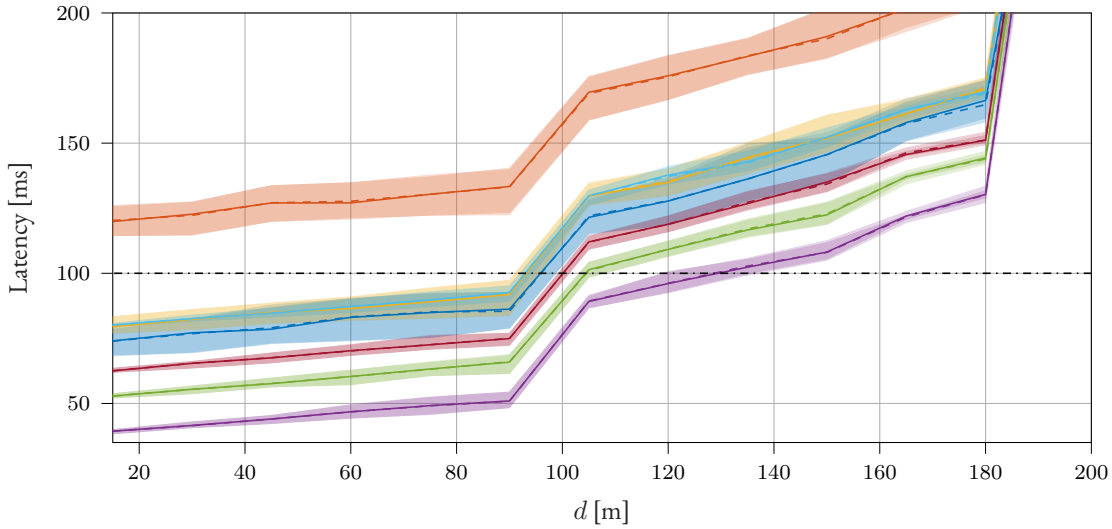


Figure 9.7: Median latency (solid for KITTI and dashed for the statistical model) and corresponding confidence intervals (shaded areas) for different TX-RX distances and HSC compression levels.

of all the HSC compression levels is well below the theoretical capacity, thus excluding a capacity-limited scenario. The steady decline of the E2E throughput reported in 9.6 is rather caused by the channel instability due to the greater NLoS probability at large distances, as further confirmed by the PRR values reported in 9.7. For short distances ($d < 45$ m), all the packets are successfully delivered. As the TX moves further away from the RX, the channel conditions worsen and a larger number of packets is dropped, leading to an increasing number of retransmissions and a corresponding throughput degradation. Correspondingly, from 9.7 we can observe that the latency for all the models steadily increases with the distance, up to 170 m, when it sharply exceeds 200 ms. Note that, in general, the requirement for the latency in vehicular networks is set to 100 ms [230]. The delay associated to all the HSC levels remains below the threshold for distances shorter than 100 m, except for D0/S1. This is counter-intuitive, as D0/S1 reduces the size of the pointclouds with respect to D0/S0 (raw), and thus the transmission requires less time. However, in this case, the compression ratio is not large enough to compensate for the de/compression delay, well-exemplifying the trade-off between the two metrics. On the contrary, D0/S2 is shown to compress effectively, with a reduction in the file size large enough to keep the latency below the threshold. Finally, the D11/S0 latency does not exceed 50 ms for distances up to 75 m, showing how reducing the number of quantization bits alone can be beneficial both in terms of source rate and delay.

However, D14/S2 is more effective in terms of rate reduction, at the cost of a larger latency, still below the 100 ms threshold, due to the semantic inference. Thus, the choice of the best HSC compression level should be made taking into account the availability of the network resources and the communication constraints. Furthermore, a more aggressive compression policy might lead to the degradation of the pointcloud quality, which in turn might affect the performance of the inference algorithms at the receiver. However, this analysis, partially presented in [231], is left for future works.

Thus, we can conclude that

- only the mmWave channel can support the streaming of raw lidar data, for communication links up to 100 m, where the LoS is dominant;
- the packet loss when the mmWave channel is in NLoS conditions leads to higher latency and lower throughput;
- compressing the pointclouds with HSC can effectively reduce the latency;
- a more aggressive compression policy is required if more traditional protocols, e.g. DSRC, are employed;
- the trade-off between compression ratio and delay is not trivial and should be taken into account when analyzing and designing communication protocols.

9.5.3 Statistical Model Evaluation

In this section, we assess whether using the statistical models derived in 9.4 and in 9.3 for the file size of the individual pointclouds instead of that of Kitti results in a valid distribution of the E2E metrics. Namely, we compared the throughput and the latency obtained with the *BurstyApplication* and the size of Kitti pointclouds as input, used as the baseline, with those obtained with the *StatisticalTraffic* application.

In 9.6 and 9.7, the dashed and the solid lines represent the median values of the metrics obtained with Kitti and the statistical models, respectively, thus offering a visual comparison between the two methods. We can observe that there is an almost perfect overlap between the two, for all HSC compression levels and TX-RX distances.

The numerical evaluation was carried out by repeating the KS test on the E2E metrics. 9.8 reports whether the throughput distribution passes the test. As expected, this is verified for all compression levels and distances, with the exception of D0/S0, which did not pass the

d [m]	D0/S0	D0/S1	D0/S2	D11/S0	D14/S0	D14/S1	D14/S2
15	✓	✓	✓	✓	✓	✓	✓
30	✓	✓	✓	✓	✓	✓	✓
45	✓	✓	✓	✓	✓	✓	✓
60	✓	✓	✓	✓	✓	✓	✓
75	✓	✓	✓	✓	✓	✓	✓
90	✓	✓	✓	✓	✓	✓	✓
105	✓	✓	✓	✓	✓	✓	✓
120	✓	✓	✓	✓	✓	✓	✓
135	✓	✓	✓	✓	✓	✓	✓
150	✓	✓	✓	✓	✓	✓	✓
165	✗	✓	✓	✓	✓	✓	✓
180	✗	✓	✓	✓	✓	✓	✓
195	✗	✓	✓	✓	✓	✓	✓
210	✗	✓	✓	✓	✓	✓	✓
225	✗	✓	✓	✓	✓	✓	✓
240	✗	✓	✓	✓	✓	✓	✓
255	✗	✓	✓	✓	✓	✓	✓
270	✓	✓	✓	✓	✓	✓	✓
285	✓	✓	✓	✓	✓	✓	✓
300	✗	✓	✓	✓	✓	✓	✓

Table 9.8: The distribution of the E2E throughput obtained using the statistical model is compared with that obtained using the KITTI dataset. All the distributions pass the KS test, with the exception of D0/S0 at larger distances, when the channel fluctuations are more significant.

test for the input distribution either. Notably, we can notice that the test fails at distances greater than 150 m, where the NLoS introduces additional randomness into the distribution. Nevertheless, in all the other cases, the distribution matches the baseline.

Finally, in 9.8 we report the run time of the simulations run with the *BurstyApplication* and the Kitti data (narrow bars) and with the *StatisticalTraffic* (wide bars). As expected, using the statistical models consistently reduces the run time of the simulations, with an average speed-up factor of $\times 18$. The largest gain is obtained when considering the raw data (D0/S0), with a $\times 26$ gain, bringing the simulation time from more than 10 minutes down to about 22 seconds.

We can thus conclude that the *StatisticalTraffic* application can effectively be used in place of the *BurstyApplication*, with a huge boost in terms of simulation run time while preserving

d [m]	D0/S0	D0/S1	D0/S2	D11/S0	D14/S0	D14/S1	D14/S2
15	✓	✓	✓	✓	✓	✓	✓
30	✓	✓	✓	✓	✓	✓	✓
45	✓	✓	✓	✓	✓	✓	✓
60	✓	✓	✓	✓	✓	✓	✓
75	✓	✓	✓	✓	✓	✓	✓
90	✓	✓	✓	✓	✓	✓	✓
105	✓	✓	✓	✓	✓	✓	✓
120	✓	✓	✓	✓	✓	✓	✓
135	✓	✓	✓	✓	✓	✓	✓
150	✓	✓	✓	✓	✓	✓	✓
165	✓	✓	✓	✗	✓	✗	✗
180	✓	✓	✓	✓	✓	✗	✓
195	✓	✓	✓	✓	✓	✓	✓
210	✓	✓	✓	✓	✓	✓	✓
225	✓	✓	✓	✓	✗	✓	✓
240	✓	✓	✓	✓	✓	✓	✓
255	✓	✓	✓	✓	✗	✓	✓
270	✓	✓	✓	✓	✓	✓	✓
285	✓	✓	✓	✓	✓	✓	✓
300	✓	✓	✓	✓	✓	✓	✓

Table 9.9: The distribution of the E2E latency obtained using the statistical model is compared with that obtained using the KITTI dataset. All the distributions pass the KS test, with the exception of some compression levels at larger distances, when the channel fluctuations are more significant.

good accuracy. The results were shown to be less statistically consistent for distances greater than 150 m when the channel becomes less stable. However, we showed that the transmission of LiDAR data at longer distances is not feasible under the current latency constraints. Thus, the *Statistical Traffic* application is effective in generating a realistic stream of LiDAR pointclouds for the valid range of operation.

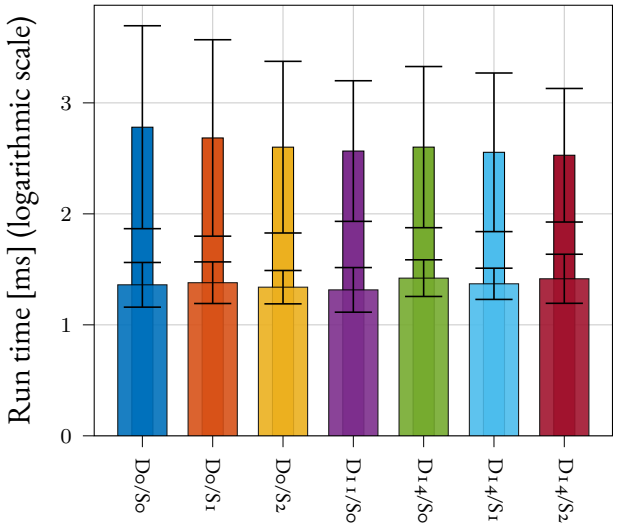


Figure 9.8: Average run time of HSC models measured in ns-3 based on KITTI (narrow bars) and Statistical (wide bars).

10 Conclusion

This thesis has investigated the design and performance of mmWave and sub-THz of 5G and 6G cellular networks from an end-to-end and system-level perspective. The main conclusions are as follows

- In the first part of the thesis we introduced the implementation of two SCMs for the ns-3 module TeraSim, based on the 140 GHz models from [29, 44]. This makes it possible to simulate LoS and NLoS scenarios for future 6G networks, including fading and the possibility of interacting with antenna array models. We also compared the two-channel models (based on a fully stochastic and an RT-based modeling) with simulations in an indoor scenario, which have highlighted how the two modeling strategies differ in their interaction with the directional antenna model of TeraSim.

- In Chapter 4, we proposed two novel channel modeling approaches to reduce the complexity of channel molding approaches in sub-THz spectrum. SSCH-THz is the fully stochastic approach. We also introduce THz-GAN channel, cGAN based channel modeling approach that supports generality based on transfer learning. We obtained the baseline dataset to feed our models from the ns-3 implementation of benchmark channel models, and we also implemented our own channel models in the ns-3. We showed that new channel models follow benchmark channel modeling very closely, and network performances based on different channel modelings are compara-

ble. Based on its simplicity of use and adaptability to all scenarios, THz-GAN channel outperforms all channel modeling techniques.

- In Chapter 6, we proposed the first reliability-focused scheduling and path selection algorithm for IAB mmWave networks. We illustrated that our RL-based solution can cope with the network dynamics including channel, interference, and load. Furthermore, we demonstrated that THz-GAN channel not only exhibits highly reliable performance in the presence of the above-mentioned network dynamics, but also outperforms the benchmark schemes in terms of throughput, latency and packet-drop rate. The reliability of THz-GAN channel stems from the joint minimization of the average latency, and the expected value of its tail losses, by leveraging CVaR as a risk metric.

- RISs and AF relays are among the most promising 6G network-enabling technologies. Not only can these elements enhance wireless device communication and coverage, but they can also reduce energy consumption in comparison to IAB systems. In Chapter 7, we proposed a signal model for RISs and AF relays based on the 3GPP TR 38.901 channel for 5G New Radio (NR) networks and described the methodology we used to execute network-level simulations of 5G scenarios with IRS and AF relay nodes. On the basis of this framework, simulations were conducted to provide numerical guidelines for sizing IRS/AF-assisted networks.

  - Both IRS and AF relays can increase the throughput, latency, and PER of end users relative to a baseline scenario in which relays are not deployed.
  - RISs are preferable to AF relays in small networks due to their inherent simplicity and energy efficiency.
  - AF relay are more appropriate in congested networks, while IRSs should be large, despite the increased system complexity

- We conclude Chapter 8 by providing the first performance evaluation of the possibilities of sub-terahertz frequencies for 6G IAB using the customized extension to the open-source Sionna simulator. This permits the use of greedy algorithms to evaluate the deployment of mixed mmWave and sub-terahertz lines to boost the backhaul network's capacity. We will broaden the analysis of the network's performance to cover a range of source traffic patterns, scenarios (including multi-donor instances, deploy-

ments with lower node density, or more realistic map-based scenarios as in [232, 233]), and protocol stack implementations as future work.

10.1 Future Work

Here we listed the future works.

- We will utilize the suggested channel models and a real dataset in the sub-THz domain to examine and authenticate proposed approaches to actual scenarios.
- We intended to augment the functionality and performance metrics of the novel IAB Simulator by incorporating additional modules. In addition, we intend to incorporate real-world datasets that I obtained from Telefonica during my internship as baseline scenarios for the simulator.

References

[1] J. Hoydis, S. Cammerer, F. A. Aoudia, A. Vem, N. Binder, G. Marcus, and A. Keller, "Sionna: An Open-Source Library for Next-Generation Physical Layer Research," *arXiv preprint arXiv:2203.11854*, Mar. 2022.

[2] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," Technical Report (TR) 38.901, Jun. 2018, version 15.0.0.

[3] I. Tomkos, D. Klonidis, E. Pikasis, and S. Theodoridis, "Toward the 6g network era: Opportunities and challenges," *IT Professional*, vol. 22, no. 1, pp. 34–38, 2020.

[4] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6g wireless communication systems: Applications, requirements, technologies, challenges, and research directions," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2020.

[5] Y. Lu and X. Zheng, "6g: A survey on technologies, scenarios, challenges, and the related issues," *Journal of Industrial Information Integration*, vol. 19, p. 100158, 2020.

[6] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *Journal of Risk*, vol. 2, pp. 21–41, 2000.

[7] A. Ortiz, A. Asadi, G. H. Sim, D. Steinmetzer, and M. Hollick, "SCAROS: A Scalable and Robust Self-Backhauling Solution for Highly Dynamic Millimeter-Wave Networks," $IEEE_JJSAC, vol.\ 37, no.\ 12, pp.2685 - -2698, Oct.2019.$

[8] M. Polese, M. Giordani, A. Roy, D. Castor, and M. Zorzi, "Distributed path selection strategies for integrated access and backhaul at mmWaves," in *Proc. IEEE Globecom*, Dec. 2018.

[9] M. Polese, V. Ariyarathna, P. Sen, J. V. Siles, F. Restuccia, T. Melodia, and J. M. Jornet, "Dynamic spectrum sharing between active and passive users above 100 GHz," *Communications Engineering*, vol. 1, no. 1, pp. 1–9, 2022.

[10] N. Abuzainab, M. Alrabeiah, A. Alkhateeb, and Y. E. Sagduyu, "Deep learning for thz drones with flying intelligent surfaces: Beam and handoff prediction," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2021, pp. 1–6.

[11] Y. Shi, K. Davaslioglu, and Y. E. Sagduyu, "Generative adversarial network for wireless signal spoofing," in *Proceedings of the ACM Workshop on Wireless Security and Machine Learning*, 2019, pp. 55–60.

[12] Y. Wang, L. Zhang *et al.*, "Gan-based wireless channel recognition enhancement in aerospace communication system," in *Journal of Physics: Conference Series*, vol. 1856, no. 1. IOP Publishing, 2021, p. 012036.

[13] C. Zou, F. Yang, J. Song, and Z. Han, "Generative adversarial network for wireless communication: Principle, application, and trends," *IEEE Communications Magazine*, 2023.

[14] Y. Huang, C. Xu, C. Zhang, M. Hua, and Z. Zhang, "An overview of intelligent wireless communications using deep reinforcement learning," *Journal of Communications and Information Networks*, vol. 4, no. 2, pp. 15–29, 2019.

[15] V. P. Rekkas, S. Sotiroudis, P. Sarigiannidis, S. Wan, G. K. Karagiannidis, and S. K. Goudos, "Machine learning in beyond 5g/6g networks—state-of-the-art and future trends," *Electronics*, vol. 10, no. 22, p. 2786, 2021.

[16] S. S. Mousavi, M. Schukat, and E. Howley, "Deep reinforcement learning: an overview," in *Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016: Volume 2*. Springer, 2018, pp. 426–440.

[17] A. Mekrache, A. Bradai, E. Moulay, and S. Dawaliby, "Deep reinforcement learning techniques for vehicular networks: Recent advances and future trends towards 6g," *Vehicular Communications*, vol. 33, p. 100398, 2022.

[18] A. A. Puspitasari and B. M. Lee, "A survey on reinforcement learning for reconfigurable intelligent surfaces in wireless communications," *Sensors*, vol. 23, no. 5, p. 2554, 2023.

[19] P. Mary, V. Koivunen, and C. Moy, "Reinforcement learning for physical layer communications," 2021.

[20] R. Wiesmayr, G. Marti, C. Dick, H. Song, and C. Studer, "Bit error and block error rate training for ml-assisted communication," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[21] R. Wiesmayr, C. Dick, J. Hoydis, and C. Studer, "Duidd: Deep-unfolded interleaved detection and decoding for mimo wireless systems," in *2022 56th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2022, pp. 181–188.

[22] C. Lai, R. Sun, C. Gentile, P. B. Papazian, J. Wang, and J. Senic, "Methodology for multipath-component tracking in millimeter-wave channel modeling," *IEEE Transactions on Antennas and Propagation*, vol. 67, no. 3, pp. 1826–1836, 2018.

[23] V. Petrov, T. Kurner, and I. Hosako, "Ieee 802.15. 3d: First standardization efforts for sub-terahertz band communications toward 6g," *IEEE Communications Magazine*, vol. 58, no. 11, pp. 28–33, Nov 2020.

[24] Q. Xia, Z. Hossain, M. Medley, and J. M. Jornet, "A link-layer synchronization and medium access control protocol for terahertz-band communication networks," *IEEE Transactions on Mobile Computing*, vol. 20, no. 1, pp. 2–18, Jan 2021.

[25] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," *IEEE Transactions on Wireless Communications*, vol. 10, no. 10, pp. 3211–3221, Aug 2011.

[26] A. Shafie, N. Yang, C. Han, J. M. Jornet, M. Juntti, and T. Kurner, "Terahertz communications for 6g and beyond wireless networks: Challenges, key advancements, and opportunities," *arXiv preprint arXiv:2207.11021*, Jul 2022.

[27] C. Han, W. Gao, N. Yang, and J. M. Jornet, "Molecular absorption effect: A double-edged sword of terahertz communications," *arXiv preprint arXiv:2205.12520*, May 2022.

[28] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE journal on selected areas in communications*, vol. 32, no. 6, pp. 1164–1179, Jun 2014.

[29] S. Ju, Y. Xing, O. Kanhere, and T. S. Rappaport, "Millimeter wave and sub-terahertz spatial statistical channel model for an indoor office building," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1561–1575, Apr 2021.

[30] S. Priebe and T. Kurner, "Stochastic modeling of thz indoor radio channels," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4445–4455, Aug 2013.

[31] Y. Xing and T. S. Rappaport, "Propagation measurement system and approach at 140 ghz-moving to 6g and above 100 ghz," in *2018 IEEE global communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.

[32] Y. Xing, O. Kanhere, S. Ju, and T. S. Rappaport, "Indoor wireless channel properties at millimeter wave and sub-terahertz frequencies," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[33] N. Khalid, N. A. Abbasi, and O. B. Akan, "Statistical characterization and analysis of low-thz communication channel for 5g internet of things," *Nano Communication Networks*, vol. 22, p. 100258, Dec 2019.

[34] K. Tekbıyık, A. R. Ekti, G. K. Kurt, A. Görçin, and S. Yarkan, "Modeling and analysis of short distance sub-terahertz communication channel via mixture of gamma distribution," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 4, pp. 2945–2954, march 2021.

[35] Y. Chen and C. Han, "Time-varying channel modeling for low-terahertz urban vehicle-to-infrastructure communications," in *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2019, pp. 1–6.

[36] Q. Hu, Y. Cai, A. Liu, G. Yu, and G. Y. Li, "Low-complexity joint resource allocation and trajectory design for uav-aided relay networks with the segmented ray-tracing channel model," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 6179–6195, June 2020.

[37] K. Qiu, S. Bakirtzis, H. Song, J. Zhang, and I. Wassell, "Pseudo ray-tracing: Deep leaning assisted outdoor mm-wave path loss prediction," *IEEE Wireless Communications Letters*, vol. 11, no. 8, pp. 1699–1702, May 2022.

[38] M. Lecci, P. Testolina, M. Polese, M. Giordani, and M. Zorzi, "Accuracy versus complexity for mmwave ray-tracing: A full stack perspective," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 7826–7841, June 2021.

[39] D. He, K. Guan, B. Ai, A. Fricke, R. He, Z. Zhong, A. Kasamatsu, I. Hosako, and T. Kürner, "Channel modeling for kiosk downloading communication system at 300 ghz," in *2017 11th European Conference on Antennas and Propagation (EUCAP)*. IEEE, 2017, pp. 1331–1335.

[40] H. Yi, K. Guan, D. He, B. Ai, J. Dou, and J. Kim, "Characterization for the vehicle-to-infrastructure channel in urban and highway scenarios at the terahertz band," *IEEE Access*, vol. 7, pp. 166 984–166 996, Nov 2019.

[41] G. Gougeon, Y. Corre, and M. Z. Aslam, "Sub-thz channel characterization ray-based deterministic simulations," *ITU Journal: ICT Discoveries*, vol. 2, no. 1, Nov 2019.

[42] Y. Chen, L. Yan, and C. Han, "Hybrid spherical- and planar-wave modeling and dcnn-powered estimation of terahertz ultra-massive mimo channels," *IEEE Transactions on Communications*, vol. 69, no. 10, pp. 7063–7076, July 2021.

[43] Y. Li, N. Li, and C. Han, "Ray-tracing simulation and hybrid channel modeling for low-terahertz uav communications," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[44] Y. Chen, Y. Li, C. Han, Z. Yu, and G. Wang, "Channel measurement and ray-tracing-statistical hybrid modeling for low-terahertz indoor communications," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 8163–8176, Dec 2021.

[45] G. E. Athanasiadou and A. R. Nix, "A novel 3-d indoor ray-tracing propagation model: The path generator and evaluation of narrow-band and wide-band predictions," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 4, pp. 1152–1168, July 2000.

[46] M. Lecci, M. Polese, C. Lai, J. Wang, C. Gentile, N. Golmie, and M. Zorzi, "Quasi-deterministic channel model for mmwaves: Mathematical formalization and validation," in *IEEE Global Communications Conference (GLOBECOM)*, Dec 2020.

[47] H. Lee, S. H. Lee, and T. Q. S. Quek, "Learning autonomy in management of wireless random networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 12, pp. 8039–8053, June 2021.

[48] F. Sohrabi, K. M. Attiah, and W. Yu, "Deep learning for distributed channel feedback and multiuser precoding in fdd massive mimo," *IEEE Transactions on Wireless Communications*, vol. 20, no. 7, pp. 4044–4057, Feb 2021.

[49] W. Guo, "Explainable artificial intelligence for 6g: Improving trust between human and machine," *IEEE Communications Magazine*, vol. 58, no. 6, pp. 39–45, June 2020.

[50] B. Yang, X. Cao, K. Xiong, C. Yuen, Y. L. Guan, S. Leng, L. Qian, and Z. Han, "Edge intelligence for autonomous driving in 6g wireless system: Design challenges and solutions," *IEEE Wireless Communications*, vol. 28, no. 2, pp. 40–47, April 2021.

[51] A. Ashtari Gargari, A. Ortiz, M. Pagin, A. Klein, M. Hollick, M. Zorzi, and A. Asadi, "Safehaul: Risk-averse learning for reliable mmWave self-backhauling in 6G networks," in *IEEE Conference on Computer Communications (INFOCOM)*, New York, USA, 2023.

[52] T. Wen, G. Xie, Y. Cao, and B. Cai, "A dnn-based channel model for network planning in train control systems," *IEEE Transactions on Intelligent Transportation Systems*, Jul 2021.

[53] X. Zheng and V. K. Lau, "Simultaneous learning and inferencing of dnn-based mmwave massive mimo channel estimation in iot systems with unknown nonlinear distortion," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 783–799, Jun 2021.

[54] Z. Hu, Y. Li, and C. Han, "Transfer generative adversarial networks (t-gan)-based terahertz channel modeling," *arXiv preprint arXiv:2301.00981*, 2023.

[55] M. Elsayed, M. Erol-Kantarci, and H. Yanikomeroglu, "Transfer reinforcement learning for 5g new radio mmwave networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 5, pp. 2838–2849, Dec 2021.

[56] W. Alves, I. Correa, N. González-Prelcic, and A. Klautau, "Deep transfer learning for site-specific channel estimation in low-resolution mmwave mimo," *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1424–1428, March 2021.

[57] W. Alves, I. Correa, N. González-Prelcic, and A. Klautau, "Deep transfer learning for site-specific channel estimation in low-resolution mmwave mimo," *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1424–1428, Mar. 2021.

[58] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G Networks: Use Cases and Technologies," *IEEE Commun. Mag.*, vol. 58, no. 3, pp. 55–61, March 2020.

[59] A. A. Gargari, M. Pagin, M. Polese, and M. Zorzi, "6g integrated access and backhaul networks with sub-terahertz links," in *2023 18th Wireless On-Demand Network Systems and Services Conference (WONS)*, 2023, pp. 13–19.

[60] H. Sarieddeen, N. Saeed, T. Y. Al-Naffouri, and M.-S. Alouini, "Next Generation Terahertz Communications: A Rendezvous of Sensing, Imaging, and Localization," *IEEE Communications Magazine*, vol. 58, no. 5, pp. 69–75, May 2020.

[61] P. Sen, V. Ariyarathna, A. Madanayake, and J. M. Jornet, "Experimental Wireless Testbed for Ultrabroadband Terahertz Networks," in *Proceedings of the 14th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, ser. WiNTECH'20, London, United Kingdom, 2020, p. 48–55.

[62] Q. Xia and J. M. Jornet, "Multi-hop relaying distribution strategies for terahertz-band communication networks: A cross-layer analysis," *IEEE Transactions on Wireless Communications*, vol. 21, no. 7, pp. 5075–5089, July 2022.

[63] A. A. Gargari, M. Polese, and M. Zorzi, "Full-Stack Comparison of Channel Models for Networks above 100 GHz in an Indoor Scenario," in *Proceedings of the 5th ACM Workshop on Millimeter-Wave and Terahertz Networks and Sensing Systems*, ser. mmNets '21. New Orleans, Louisiana: Association for Computing Machinery, 2021, p. 43–48. [Online]. Available: https://doi.org/10.1145/3477081.3481677

[64] Y. Yang, Y. Li, W. Zhang, F. Qin, P. Zhu, and C.-X. Wang, "Generative-adversarial-network-based wireless channel modeling: Challenges and opportunities," *IEEE Communications Magazine*, vol. 57, no. 3, pp. 22–27, Mar 2019.

[65] J. Zhao, L. Xiong, J. Li, J. Xing, S. Yan, and J. Feng, "3d-aided dual-agent gans for unconstrained face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 10, pp. 2380–2394, Jul 2018.

[66] T. Go, S. Lee, D. You, and S. J. Lee, "Deep learning-based hologram generation using a white light source," *Scientific reports*, vol. 10, no. 1, pp. 1–12, 2020.

[67] S. Baek, K. I. Kim, and T.-K. Kim, "Weakly-supervised domain adaptation via gan and mesh model for estimating 3d hand poses interacting objects," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6121–6131.

[68] M. D. Cirillo, D. Abramian, and A. Eklund, "Vox2vox: 3d-gan for brain tumour segmentation," in *International MICCAI Brainlesion Workshop*. Springer, oct 2020, pp. 274–284.

[69] L. Sixt, B. Wild, and T. Landgraf, "Rendergan: Generating realistic labeled data," *Frontiers in Robotics and AI*, vol. 5, p. 66, 2018.

[70] R. Spick, S. Demediuk, and J. Alfred Walker, "Naive mesh-to-mesh coloured model generation using 3d gans," in *Proceedings of the Australasian Computer Science Week Multiconference*, 2020, pp. 1–6.

[71]

[72] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, Nov 2014.

[73] C. Banerjee, T. Mukherjee, and E. Pasiliao Jr, "An empirical study on generalizations of the relu activation function," in *Proceedings of the 2019 ACM Southeast Conference*, Apr 2019, pp. 164–167.

[74] M. Lecci, P. Testolina, M. Polese, M. Giordani, and M. Zorzi, "Accuracy vs. complexity for mmwave ray-tracing: A full stack perspective," *IEEE Transactions on Wireless Communications*, Jun 2021.

[75] R. Piesiewicz, C. Jansen, S. Wietzke, D. Mittleman, M. Koch, and T. Kürner, "Properties of building and plastic materials in the thz range," *International Journal of Infrared and Millimeter Waves*, vol. 28, no. 5, pp. 363–371, May 2007.

[76] V. W. Berger and Y. Zhou, "Kolmogorov–smirnov test: Overview," *Wiley statsref: Statistics reference online*, 2014.

[77] Q. Zhang, X. Wang, Z. Li, and Z. Wei, "Design and performance evaluation of joint sensing and communication integrated system for 5g mmwave enabled cavs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 6, pp. 1500–1514, 2021.

[78] N. K. Mallat, M. Ishtiaq, A. Ur Rehman, and A. Iqbal, "Millimeter-wave in the face of 5g communication potential applications," *IETE journal of research*, vol. 68, no. 4, pp. 2522–2530, 2022.

[79] M. S. Sim, Y.-G. Lim, S. H. Park, L. Dai, and C.-B. Chae, "Deep learning-based mmwave beam selection for 5g nr/6g with sub-6 ghz channel information: Algorithms and prototype validation," *IEEE Access*, vol. 8, pp. 51 634–51 646, 2020.

[80] 3GPP, "NR; Integrated Access and Backhaul (IAB) radio transmission and reception," Technical Specification (TS) 38.174, Jun. 2022, v.17.1.0.

[81] ——, "NR; Backhaul Adaptation Protocol (BAP) specification," Technical Specification (TS) 38.340, May 2022, v.17.0.0.

[82] G. Kalfas, C. Vagionas, A. Antonopoulos, E. Kartsakli, A. Mesodiakaki, S. Papaioannou, P. Maniotis, J. S. Vardakas, C. Verikoukis, and N. Pleros, "Next generation fiber-wireless fronthaul for 5g mmwave networks," *IEEE Communications Magazine*, vol. 57, no. 3, pp. 138–144, 2019.

[83] L. Zhang, H. Zhao, S. Hou, Z. Zhao, H. Xu, X. Wu, Q. Wu, and R. Zhang, "A survey on 5g millimeter wave communications for uav-assisted wireless networks," *IEEE Access*, vol. 7, pp. 117 460–117 504, 2019.

[84] D. Ivanova, E. Markova, D. Moltchanov, R. Pirmagomedov, Y. Koucheryavy, and K. Samouylov, "Performance of priority-based traffic coexistence strategies in 5g mmwave industrial deployments," *IEEE Access*, vol. 10, pp. 9241–9256, 2022.

[85] 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 Commun. Mag.*, vol. 58, no. 3, pp. 62–68, Mar 2020.

[86] A. Fascista, A. Coluccia, H. Wymeersch, and G. Seco-Granados, "Ris-aided joint localization and synchronization with a single-antenna mmwave receiver," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 4455–4459.

[87] Y. Xanthos, W. Lyu, S. Yang, C. Assi, X. Zou, and N. Wei, "Joint localization and beamforming for reconfigurable intelligent surface aided 5g mmwave communication systems," *arXiv preprint arXiv:2210.17530*, 2022.

[88] J.-B. Gros, L. Santamaria, V. Popov, M. Odit, V. Lenets, X. Lleshi, A. Toubal, and Y. Nasser, "Design of reconfigurable intelligent surfaces at mmwave with application to 5g/6g," in *2023 17th European Conference on Antennas and Propagation (EuCAP)*. IEEE, 2023, pp. 1–4.

[89] P. Fiore, E. Moro, I. Filippini, A. Capone, and D. De Donno, "Boosting 5g mm-wave iab reliability with reconfigurable intelligent surfaces," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2022, pp. 758–763.

[90] B. Zhang, F. Devoti, and I. Filippini, "Rl-based resource allocation in mmwave 5g iab networks," in *2020 Mediterranean Communication and Computer Networking Conference (MedComNet)*. IEEE, 2020, pp. 1–8.

[91] G. C. Alexandropoulos, N. Shlezinger, I. Alamzadeh, M. F. Imani, H. Zhang, and Y. C. Eldar, "Hybrid reconfigurable intelligent metasurfaces: Enabling simultaneous tunable reflections and sensing for 6g wireless communications," *arXiv preprint arXiv:2104.04690*, 2021.

[92] J.-B. Gros, V. Popov, M. A. Odit, V. Lenets, and G. Lerosey, "A reconfigurable intelligent surface at mmwave based on a binary phase tunable metasurface," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1055–1064, 2021.

[93] H. Alakoca, M. Namdar, S. Aldirmaz-Colak, M. Basaran, A. Basgumus, L. Durak-Ata, and H. Yanikomeroglu, "Metasurface manipulation attacks: Potential security threats of ris-aided 6g communications," *IEEE Communications Magazine*, vol. 61, no. 1, pp. 24–30, 2022.

[94] X. Qian, M. Di Renzo, V. Sciancalepore, and X. Costa-Pérez, "Joint optimization of reconfigurable intelligent surfaces and dynamic metasurface antennas for massive mimo communications," in *2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2022, pp. 450–454.

[95] E. Ayanoglu, F. Capolino, and A. L. Swindlehurst, "Wave-controlled metasurface-based reconfigurable intelligent surfaces," *IEEE Wireless Communications*, vol. 29, no. 4, pp. 86–92, 2022.

[96] P. Wang, Z. Wei, T. Wu, W. Jiang, T. Hong, G. F. Pedersen, and M. Shen, "A machine learning framework for the design of stcdme structures in ris applications," *IEEE Transactions on Microwave Theory and Techniques*, 2023.

[97] X. Pei, H. Yin, L. Tan, L. Cao, Z. Li, K. Wang, K. Zhang, and E. Björnson, "Ris-aided wireless communications: Prototyping, adaptive beamforming, and indoor/outdoor field trials," *IEEE Transactions on Communications*, vol. 69, no. 12, pp. 8627–8640, 2021.

[98] Y. He, Y. Cai, H. Mao, and G. Yu, "Ris-assisted communication radar coexistence: Joint beamforming design and analysis," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 7, pp. 2131–2145, 2022.

[99] A. Fascista, M. F. Keskin, A. Coluccia, H. Wymeersch, and G. Seco-Granados, "Ris-aided joint localization and synchronization with a single-antenna receiver: Beamforming design and low-complexity estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 5, pp. 1141–1156, 2022.

[100] X. Liu, Y. Liu, Y. Chen, and H. V. Poor, "Ris enhanced massive non-orthogonal multiple access networks: Deployment and passive beamforming design," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 4, pp. 1057–1071, 2020.

[101] J. An, C. Xu, L. Gan, and L. Hanzo, "Low-complexity channel estimation and passive beamforming for ris-assisted mimo systems relying on discrete phase shifts," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1245–1260, 2021.

[102] M. Almekhlafi, M. A. Arfaoui, C. Assi, and A. Ghrayeb, "A low complexity passive beamforming design for reconfigurable intelligent surface (ris) in 6g networks," *IEEE Transactions on Vehicular Technology*, 2023.

[103] K. Keykhosravi, M. F. Keskin, G. Seco-Granados, and H. Wymeersch, "Siso ris-enabled joint 3d downlink localization and synchronization," in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp. 1–6.

[104] H. Kim, H. Chen, M. F. Keskin, Y. Ge, K. Keykhosravi, G. C. Alexandropoulos, S. Kim, and H. Wymeersch, "Ris-enabled and access-point-free simultaneous radio localization and mapping," *IEEE Transactions on Wireless Communications*, 2023.

[105] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5g: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE journal on selected areas in communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[106] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, 2014.

[107] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 gbps/ue in cellular systems: Understanding ultra-dense small cell deployments," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2078–2101, 2015.

[108] N. Makris, C. Zarafetas, P. Basaras, T. Korakis, N. Nikaein, and L. Tassiulas, "Cloud-based convergence of heterogeneous rans in 5g disaggregated architectures," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–6.

[109] H. S. Dhillon and G. Caire, "Wireless backhaul networks: Capacity bound, scalability analysis and design guidelines," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6043–6056, 2015.

[110] V. Gambiroza, B. Sadeghi, and E. W. Knightly, "End-to-end performance and fairness in multihop wireless backhaul networks," in *Proceedings of the 10th annual international conference on Mobile computing and networking*, 2004, pp. 287–301.

[111] M. Alicherry, R. Bhatia, and L. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *Proceedings of the 11th annual international conference on Mobile computing and networking*, 2005, pp. 58–72.

[112] C. Dehos, J. L. González, A. De Domenico, D. Ktenas, and L. Dussopt, "Millimeter-wave access and backhauling: The solution to the exponential data traffic increase in 5g mobile communications systems?" *IEEE communications magazine*, vol. 52, no. 9, pp. 88–95, 2014.

[113] X. Ge, H. Cheng, M. Guizani, and T. Han, "5g wireless backhaul networks: challenges and research advances," *IEEE network*, vol. 28, no. 6, pp. 6–11, 2014.

[114] 3GPP, "NR; Study on integrated access and backhaul," Technical Specification (TS) 38.874, 2018, v.0.6.0.

[115] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3gpp nr at mmwave frequencies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 173–196, 2018.

[116] 3GPP, "NR; Overall description; Stage-2," Technical Specification (TS) 38.300, 01 2018, version 15.0.0.

[117] C. Pan, H. Zhu, N. J. Gomes, and J. Wang, "Joint precoding and RRH selection for user-centric green MIMO C-RAN," $IEEE_{JW}COM,\ vol.16,\ no.5,\ pp.2891 - -2906,\ Mar.2017.$

[118] A. Alizadeh and M. Vu, "Load balancing user association in millimeter wave MIMO networks," $IEEE_{JW}COM,\ vol.18,\ no.6,\ pp.2932 - -2945,\ Mar.2019.$

[119] X. Huang, G. Xue, R. Yu, and S. Leng, "Joint scheduling and beamforming coordination in cloud radio access networks with QoS guarantees," $IEEE_{JV}T,\ vol.65,\ no.7,\ pp.5449 - -5460,\ Aug.2015.$

[120] H. T. Nguyen, H. D. Tuan, T. Q. Duong, H. V. Poor, and W.-J. Hwang, "Nonsmooth optimization algorithms for multicast beamforming in content-centric fog radio access networks," $IEEE_{JS}P,\ vol.68,\ pp.1455 - -1469,\ Jan.2020.$

[121] M. E. Rasekh, D. Guo, and U. Madhow, "Interference-aware routing and spectrum allocation for millimeter wave backhaul in urban picocells," in *Proc. IEEE Allerton*, Sep. 2015.

[122] G. Kwon and H. Park, "Joint user association and beamforming design for millimeter wave UDN with wireless backhaul," $IEEE_{JJ}SAC,\ vol.37,\ no.12,\ pp.2653 - -2668,\ Oct.2019.$

[123] A. Pizzo and L. Sanguinetti, "Optimal design of energy-efficient millimeter wave hybrid transceivers for wireless backhaul," in *Proc. IEEE WiOpt*, Jun. 2017.

[124] M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "Max-min rates in self-backhauled millimeter wave cellular networks," *arXiv preprint arXiv:1805.01040*, 2018.

[125] L. F. Abanto-Leon, A. Asadi, A. Garcia-Saavedra, G. H. Sim, and M. Hollick, "RadiOrchestra: Proactive management of millimeter-wave self-backhauled small cells via joint optimization of beamforming, user association, rate selection, and admission control," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 153–173, Jan. 2023.

[126] W. Lei, Y. Ye, and M. Xiao, "Deep reinforcement learning-based spectrum allocation in integrated access and backhaul networks," $IEEE_{JC}CN,\ vol.\ 6,\ no.\ 3,\ pp.970 - -979,\ May2020.$

[127] B. Zhang, F. Devoti, I. Filippini, and D. De Donno, "Resource allocation in mmWave 5G IAB networks: A reinforcement learning approach based on column generation," *Computer Networks*, vol. 196, p. 108248, Jun. 2021.

[128] Q. Cheng, Z. Wei, and J. Yuan, "Deep Reinforcement Learning-based Spectrum Allocation and Power Management for IAB Networks," in *Proc. IEEE ICC Workshops*, Jul. 2021.

[129] T. K. Vu, C.-F. Liu, M. Bennis, M. Debbah, and M. Latva-Aho, "Path selection and rate allocation in self-backhauled mmwave networks," in *Proc. IEEE WCNC*, Jun. 2018.

[130] M. Pagin, T. Zugno, M. Polese, and M. Zorzi, "Resource Management for 5G NR Integrated Access and Backhaul: A Semi-Centralized Approach," $IEEE_{JW}COM,\ vol.\ 21,\ no.\ 2,\ pp.753 - -767,\ Jul.2022.$

[131] H. Levy, *Stochastic Dominance*. Springer Science+Business Media New York, 1998.

[132] G. C. Pflug, *Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk*. Boston, MA: Springer US, 2000, pp. 272–281. [Online]. Available: https://doi.org/10.1007/978-1-4757-3150-7_15

[133] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018.

[134] Z. Luo and S. Ou, "The almost sure convergence rate of the estimator of optimized certainty equivalent risk measure under α-mixing sequences," *Communications in Statistics - Theory and Methods*, vol. 46, no. 16, pp. 8166–8177, Apr. 2017.

[135] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, no. 2–3, p. 235–256, May 2002.

[136] M. Polese, M. Giordani, A. Roy, S. Goyal, D. Castor, and M. Zorzi, "End-to-End Simulation of Integrated Access and Backhaul at mmWaves," in *Proc. IEEE CAMAD*, Oct. 2018.

[137] M. Mezzavilla, M. Zhang, M. Polese, R. Ford, S. Dutta, S. Rangan, and M. Zorzi, "End-to-End Simulation of 5G mmWave Networks," $'IEEE_{OC}STO', vol.\,20,\ no.\,3,\ pp.2237 - -2263,\ Apr.2018.$

[138] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 173–196, Sept. 2019.

[139] C. Madapatha, B. Makki, C. Fang, O. Teyeb, E. Dahlman, M.-S. Alouini, and T. Svensson, "On integrated access and backhaul networks: Current status and potentials," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1374–1389, Sep. 2020.

[140] 3GPP, "NR; NG-RAN overall description," Technical Specification (TS) 38.30, 2018.

[141] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, 2014.

[142] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 gbps/ue in cellular systems: Understanding ultra-dense small cell deployments," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2078–2101, 2015.

[143] 3GPP, "NR; Study on integrated access and backhaul," TR 38.874, 2018.

[144] 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.

[145] A. Chaoub, M. Giordani, B. Lall, V. Bhatia, A. Kliks, L. Mendes, K. Rabie, H. Saarnisaari, A. Singhal, N. Zhang *et al.*, "6g for bridging the digital divide: Wireless connectivity to remote areas," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 160–168, 2021.

[146] R. Flamini, D. De Donno, J. Gambini, F. Giuppi, C. Mazzucco, A. Milani, and L. Resteghini, "Toward a heterogeneous smart electromagnetic environment for millimeter-wave communications: An industrial viewpoint," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 10, pp. 8898–8910, 2022.

[147] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, 2019.

[148] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE transactions on wireless communications*, vol. 18, no. 8, pp. 4157–4170, 2019.

[149] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.

[150] E. Basar and I. Yildirim, "Simris channel simulator for reconfigurable intelligent surface-empowered communication systems," in *2020 IEEE Latin-American Conference on Communications (LATINCOM)*. IEEE, 2020, pp. 1–6.

[151] M. Mezzavilla, M. Zhang, M. Polese, R. Ford, S. Dutta, S. Rangan, and M. Zorzi, "End-to-end simulation of 5g mmwave networks," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2237–2263, 2018.

[152] M. Pagin, M. Giordani, A. A. Gargari, A. Rech, F. Moretto, S. Tomasin, J. Gambini, and M. Zorzi, "End-to-end simulation of 5g networks assisted by irs and af relays," in *2022 20th Mediterranean Communication and Computer Networking Conference (MedComNet)*, 2022, pp. 150–157.

[153] S. Lagen, K. Wanuga, H. Elkotby, S. Goyal, N. Patriciello, and L. Giupponi, "New radio physical layer abstraction for system-level simulations of 5g networks," in *ICC*

2020-2020 IEEE International Conference on Communications (ICC). IEEE, 2020, pp. 1–7.

[154] M. Mezzavilla, M. Miozzo, M. Rossi, N. Baldo, and M. Zorzi, "A lightweight and accurate link abstraction model for the simulation of lte networks in ns-3," in *Proceedings of the 15th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*, 2012, pp. 55–60.

[155] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Reconfigurable intelligent surfaces: Three myths and two critical questions," *IEEE Communications Magazine*, vol. 58, no. 12, pp. 90–96, 2020.

[156] X. Xu, Y. Pan, P. P. M. Y. Lwin, and X. Liang, "3D holographic display and its data transmission requirement," in *International Conference on Information Photonics and Optical Communications*, 2011, pp. 1–4.

[157] M. Shafi, A. F. Molisch, P. J. Smith, P. Z. T. Haustein, P. D. Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice," vol. 35, no. 6, pp. 1201–1221, June 2017.

[158] M. Polese, J. M. Jornet, T. Melodia, and M. Zorzi, "Toward End-to-End, Full-Stack 6G Terahertz Networks," *IEEE Communications Magazine*, vol. 58, no. 11, pp. 48–54, November 2020.

[159] 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, May 2020.

[160] C. Madapatha, B. Makki, C. Fang, O. Teyeb, E. Dahlman, M.-S. Alouini, 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, Sep. 2020.

[161] C. Saha and H. S. Dhillon, "Millimeter Wave Integrated Access and Backhaul in 5G: Performance Analysis and Design Insights," vol. 37, no. 12, pp. 2669–2684, Dec. 2019.

[162] I. F. Akyildiz, J. M. Jornet, and C. Han, "Terahertz band: Next frontier for wireless communications," *Physical Communication*, vol. 12, pp. 16–32, Sep 2014.

[163] A. Singh, M. Andrello, N. Thawdar, and J. M. Jornet, "Design and operation of a graphene-based plasmonic nano-antenna array for communication in the terahertz band," *'IEEE$_{JJ}$SAC'*, *vol*. 38, *no*. 9, *pp*.2104 − −2117, *Sep*.2020.

[164] S. Ghafoor, N. Boujnah, M. H. Rehmani, and A. Davy, "MAC Protocols for Terahertz Communication: A Comprehensive Survey," *'IEEE$_{OC}$STO'*, *vol*. 22, *no*. 4, *pp*.2236 − −2282, *FourthQuarter*2020.

[165] G. Gougeon, Y. Corre, M. Z. Aslam, S. Bicaïs, and J.-B. Doré, "Assessment of sub-THz mesh backhaul capabilities from realistic modelling at the PHY layer," in *14th European Conference on Antennas and Propagation (EuCAP)*, 2020, pp. 1–5.

[166] A. A. Raja, M. A. Jamshed, H. Pervaiz, and S. A. Hassan, "Performance analysis of UAV-assisted backhaul solutions in THz enabled hybrid heterogeneous network," in *Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2020, pp. 628–633.

[167] H. Jiang, Y. Niu, B. Ai, Z. Zhong, and S. Mao, "QoS-aware bandwidth allocation and concurrent scheduling for terahertz wireless backhaul networks," *IEEE Access*, vol. 8, pp. 125 814–125 825, July 2020.

[168] C. Saha, M. Afshang, and H. S. Dhillon, "Integrated mmWave Access and Backhaul in 5G: Bandwidth Partitioning and Downlink Analysis," in *IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.

[169] J. M. Jornet and I. F. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," vol. 10, no. 10, pp. 3211–3221, August 2011.

[170] Z. Hossain, Q. Xia, and J. M. Jornet, "Terasim: An ns-3 extension to simulate terahertz-band communication networks," *Nano Communication Networks*, vol. 17, pp. 36–44, Sep 2018.

[171] P. Sen, J. Hall, M. Polese, V. Petrov, D. Bodet, F. Restuccia, T. Melodia, and J. M. Jornet, "Terahertz Communications Can Work in Rain and Snow: Impact of Adverse

Weather Conditions on Channels at 140 GHz," in *Proceedings of the 6th ACM Workshop on Millimeter-Wave and Terahertz Networks and Sensing Systems*, ser. mmNets '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 13–18.

[172] D. J. Fagnant and K. Kockelman, "Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations," *Transportation Research Part A: Policy and Practice*, vol. 77, pp. 167–181, Jul. 2015.

[173] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE internet of things journal*, vol. 1, no. 4, pp. 289–299, Aug. 2014.

[174] Y. Li and J. Ibanez-Guzman, "LiDAR for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 50–61, 2020.

[175] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1341–1360, Mar. 2020.

[176] V. Rossi, P. Testolina, M. Giordani, and M. Zorzi, "On the role of sensor fusion for object detection in future vehicular networks," in *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, 2021, pp. 247–252.

[177] L. Hobert, A. Festag, I. Llatser, L. Altomare, F. Visintainer, and A. Kovacs, "Enhancements of v2x communication in support of cooperative autonomous driving," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 64–70, 2015.

[178] T. Higuchi, M. Giordani, A. Zanella, M. Zorzi, and O. Altintas, "Value-anticipating V2V communications for cooperative perception," in *IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2019.

[179] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[180] P. Testolina, F. Barbato, U. Michieli, M. Giordani, P. Zanuttigh, and M. Zorzi, "SELMA: SEmantic Large-Scale Multimodal Acquisitions in Variable Weather, Daytime and Viewpoints," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 7012–7024, Jul. 2023.

[181] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "Semantickitti: A dataset for semantic scene understanding of lidar sequences," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9297–9307.

[182] C. Cao, M. Preda, and T. Zaharia, "3D point cloud compression: A survey," in *24th International Conference on 3D Web Technology*, 2019.

[183] M. Giordani, A. Zanella, and M. Zorzi, "Millimeter wave communication in vehicular networks: Challenges and opportunities," in *2017 6th International Conference on Modern Circuits and Systems Technologies (MOCAST)*. IEEE, 2017, pp. 1–6.

[184] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, "An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC)," *APSIPA Transactions on Signal and Information Processing*, vol. 9, p. e13, 2020.

[185] F. Nardo, D. Peressoni, P. Testolina, M. Giordani, and A. Zanella, "Point cloud compression for efficient data broadcasting: A performance comparison," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2022.

[186] T. R. Henderson, M. Lacage, G. F. Riley, C. Dowell, and J. Kopena, "Network simulations with the ns-3 simulator," *SIGCOMM*, 2008.

[187] M. Drago, T. Zugno, M. Polese, M. Giordani, and M. Zorzi, "MilliCar: An ns-3 module for mmWave NR V2X networks," in *Workshop on ns-3*, 2020.

[188] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO-Simulation of Urban MObility," *International journal on advances in systems and measurements*, vol. 5, no. 3&4, 2012.

[189] M. Drago, T. Zugno, F. Mason, M. Giordani, M. Boban, and M. Zorzi, "Artificial Intelligence in Vehicular Wireless Networks: A Case Study Using ns-3," in *Workshop on ns-3*, 2022.

[190] J. Cao, W. S. Cleveland, Y. Gao, K. Jeffay, F. D. Smith, and M. Weigle, "Stochastic models for generating synthetic HTTP source traffic," in *IEEE IEEE Conference on Computer Communications (INFOCOM)*, 2004.

[191] M. Lecci, A. Zanella, and M. Zorzi, "An ns-3 implementation of a bursty traffic framework for virtual reality sources," in *Workshop on ns-3*, 2021.

[192] W. Buller, B. Wilson, L. van Nieuwstadt, and J. Ebling, "Statistical modelling of measured automotive radar reflections," in *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2013.

[193] M. Ahmadi-Pour, T. Ludwig, and C. Olaverri-Monreal, "Statistical modelling of multi-sensor data fusion," in *IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, 2017.

[194] A. Varischio, F. Mandruzzato, M. Bullo, M. Giordani, P. Testolina, and M. Zorzi, "Hybrid point cloud semantic compression for automotive sensors: A performance evaluation," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[195] C. Wang, B. Zeng, and J. Shao, "Application of bootstrap method in Kolmogorov-Smirnov test," in *International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*, 2011.

[196] 3GPP, "Service requirements for V2X services," Technical Specifications (TS) 22.186, Apr. 2022, version 17.0.0.

[197] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, "Millimeter-wave vehicular communication to support massive automotive sensing," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 160–167, 2016.

[198] L. Huo, D. Jiang, X. Zhu, Y. Wang, Z. Lv, and S. Singh, "A sdn-based fine-grained measurement and modeling approach to vehicular communication network traffic," *International Journal of Communication Systems*, vol. 35, no. 12, p. e4092, 2022, e4092 IJCS-19-0147.R2. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/dac.4092

[199] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A survey on 5g usage scenarios and traffic models," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 905–929, 2020.

[200] A. Bulashenko, S. Piltyay, A. Polishchuk, and O. Bulashenko, "New traffic model of m2m technology in 5g wireless sensor networks," in *2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT)*, 2020, pp. 125–131.

[201] O. N. Østerbø, D. Zucchetto, K. Mahmood, A. Zanella, and O. Grøndalen, "State modulated traffic models for machine type communications," in *2017 29th International Teletraffic Congress (ITC 29)*, vol. 1. IEEE, 2017, pp. 90–98.

[202] M. Sansoni, G. Ravagnani, D. Zucchetto, C. Pielli, A. Zanella, and K. Mahmood, "Comparison of m2m traffic models against real world data sets," in *2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. IEEE, 2018, pp. 1–6.

[203] L. Lao, Z. Li, S. Hou, B. Xiao, S. Guo, and Y. Yang, "A survey of iot applications in blockchain systems: Architecture, consensus, and traffic modeling," vol. 53, no. 1, feb 2020. [Online]. Available: https://doi.org/10.1145/3372136

[204] S. Tanwir and H. Perros, "A survey of vbr video traffic models," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1778–1802, 2013.

[205] H. Kalbkhani, M. G. Shayesteh, and N. Haghighat, "Adaptive lstar model for long-range variable bit rate video traffic prediction," *IEEE Transactions on Multimedia*, vol. 19, no. 5, pp. 999–1014, 2017.

[206] M. Lecci, F. Chiariotti, M. Drago, A. Zanella, and M. Zorzi, "Temporal characterization of xr traffic with application to predictive network slicing," in *2022 IEEE 23rd International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 2022, pp. 406–415.

[207] F. Chiariotti, M. Drago, P. Testolina, M. Lecci, A. Zanella, and M. Zorzi, "Temporal characterization and prediction of vr traffic: A network slicing use case," *IEEE Transactions on Mobile Computing*, pp. 1–18, 2023.

[208] E. Grigoreva, M. Laurer, M. Vilgelm, T. Gehrsitz, and W. Kellerer, "Coupled markovian arrival process for automotive machine type communication traffic modeling," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.

[209] L. Nie, X. Wang, L. Wan, S. Yu, H. Song, D. Jiang, and K. Zhang, "Network traffic prediction based on deep belief network and spatiotemporal compressive sensing in wireless mesh backbone networks," *Wirel. Commun. Mob. Comput.*, vol. 2018, jan 2018. [Online]. Available: https://doi.org/10.1155/2018/1260860

[210] Q. Zhuo, Q. Li, H. Yan, and Y. Qi, "Long short-term memory neural network for network traffic prediction," in *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, 2017, pp. 1–6.

[211] L. Zhang, H. Zhang, Q. Tang, P. Dong, Z. Zhao, Y. Wei, J. Mei, and K. Xue, "Lntp: An end-to-end online prediction model for network traffic," *IEEE Network*, vol. 35, no. 1, pp. 226–233, 2021.

[212] C.-H. Wang, T. Shimizu, H. Muralidharan, and A. Yamamuro, "A real-time high-definition vehicular sensor data sharing system using millimeter wave v2v communications," in *2020 IEEE Vehicular Networking Conference (VNC)*. IEEE, 2020, pp. 1–2.

[213] Q. Chen, S. Tang, J. Hochstetler, J. Guo, Y. Li, J. Xiong, Q. Yang, and S. Fu, "Low-latency high-level data sharing for connected and autonomous vehicular networks," in *2019 IEEE International Conference on Industrial Internet (ICII)*. IEEE, 2019, pp. 287–296.

[214] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "Rangenet ++: Fast and accurate lidar semantic segmentation," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 4213–4220.

[215] L. Wang, Y. Huang, Y. Hou, S. Zhang, and J. Shan, "Graph attention convolution for point cloud semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[216] L. Tchapmi, C. Choy, I. Armeni, J. Gwak, and S. Savarese, "Segcloud: Semantic segmentation of 3d point clouds," in *2017 International Conference on 3D Vision (3DV)*, 2017, pp. 537–547.

[217] Y. Zhang, Y. Qu, Y. Xie, Z. Li, S. Zheng, and C. Li, "Perturbed self-distillation: Weakly supervised large-scale point cloud semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 15 520–15 528.

[218] Y. Zhang, Z. Zhou, P. David, X. Yue, Z. Xi, B. Gong, and H. Foroosh, "Polarnet: An improved grid representation for online lidar point clouds semantic segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[219] B. Wu, X. Zhou, S. Zhao, X. Yue, and K. Keutzer, "Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 4376–4382.

[220] W. Lau, Z. Li, and K. W. Lam, "Effects of jpeg compression on image classification," *International Journal of Remote Sensing*, vol. 24, no. 7, pp. 1535–1544, 2003. [Online]. Available: https://doi.org/10.1080/0143116021014284 2

[221] T. Gandor and J. Nalepa, "First gradually, then suddenly: Understanding the impact of image compression on object detection using deep learning," *Sensors*, vol. 22, no. 3, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/3/1104

[222] Y.-Y. Jo, Y. S. Choi, H. W. Park, J. H. Lee, H. Jung, H.-E. Kim, K. Ko, C. W. Lee, H. S. Cha, and Y. Hwangbo, "Impact of image compression on deep learning-based mammogram classification," *Scientific Reports*, vol. 11, no. 1, p. 7924, 2021.

[223] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3460–3464.

[224] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *arXiv:1801.09847*, 2018.

[225] M. Lecci, M. Drago, A. Zanella, and M. Zorzi, "An open framework for analyzing and modeling xr network traffic," *IEEE Access*, vol. 9, pp. 129 782–129 795, sep 2021.

[226] M. Mezzavilla, M. Zhang, M. Polese, R. Ford, S. Dutta, S. Rangan, and M. Zorzi, "End-to-end simulation of 5g mmwave networks," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2237–2263, April 2018.

[227] W. T. Shaw, "Sampling student's t distribution-use of the inverse cumulative distribution function," *Journal of Computational Finance*, vol. 9, no. 4, p. 37, Jan 2006.

[228] J.-Y. Le Boudec, *Performance evaluation of computer and communication systems*. Epfl Press Lausanne, 2010, vol. 2.

[229] K. Abboud, H. A. Omar, and W. Zhuang, "Interworking of dsrc and cellular network technologies for v2x communications: A survey," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 9457–9470, Dec 2016.

[230] 3GPP, "Service requirements for enhanced V2X scenarios," Technical Specifications (TS) 22.186, Apr. 2022, version 17.0.0.

[231] F. Mason, M. Drago, T. Zugno, M. Giordani, M. Boban, and M. Zorzi, "A reinforcement learning framework for pqos in a teleoperated driving scenario," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2022, pp. 114–119.

[232] G. Gemmi, R. L. Cigno, and L. Maccari, "On the Properties of Next Generation Wireless Backhaul," *IEEE Transactions on Network Science and Engineering*, vol. 10, no. 1, pp. 166–177, Jan 2023.

[233] G. Gemmi, R. Lo Cigno, and L. Maccari, "On Cost-Effective, Reliable Coverage for LoS Communications in Urban Areas," *IEEE Transactions on Network and Service Management*, vol. 19, no. 3, pp. 2767–2779, Sep. 2022.

Acknowledgments

This thesis is supported by the EU MSCA ITN project MINTS “MIllimeter-wave Networking and Sensing for Beyond 5G” (grant no. 861222)