UPLINK CLUSTER-BASED RADIO RESOURCE SCHEDULING AND ALLOCATION FOR HETNET MMTC SCENARIOS

By

ABDELRAHMAN M. RAMADAN

A thesis submitted to the Graduate Program in Department of Electrical and Computer Engineering in conformity with the requirements for the Degree of Master of Applied Science

Queen's University

Kingston, Ontario, Canada

August, 2022

Dedication

FOR THE SAKE OF ALLAH THE OMNISCIENT

To my parents,

to my teachers and mentors,

and to my siblings.

Abstract

Current telecommunication networks face a surge in the number of connected Machine-type Communication (MTC) devices, creating an unprecedented disproportionate demand for existing resources, especially when working with a Heterogeneous Network (HetNet). This demand cannot be addressed adequately as the infrastructure's transition process between different generations is slow. Fourth Generation (4G) relies on Orthogonal Multiple Access (OMA), where a single user occupies its assigned sub-channel, and orthogonality offers interference-free communication for normally loaded scenarios but underperforms in overloaded scenarios. Whereas Fifth Generation (5G) is targeting more spectral efficiency by using the Non-orthogonal Multiple Access (NOMA) which is proposed to be used in future releases, allowing MTC devices to share the same resources in frequency and time. However, NOMA medium access techniques, in general, have a complex scheduler design as they group users/devices with aligned correlations, and then a challenging process is needed at the receiving side to decode messages from different devices. In this study, we formulate and simulate a 4G/5G Uplink scheduler that is based on dual NOMA-OMA. The objective is to achieve a tangible improvement in the spectral and scheduling efficiency of the network. We are able to optimize the system under HetNet objectives and clustering constraints in overloaded scenarios, to examine the limitations of both NOMA and OMA.

List of Publications

- A. Ramadan, N. Zorba, and H. S. Hassanein, "Uplink Cluster-Based Radio Resource Scheduling for HetNet mMTC Scenarios," IEEE Global Communications Conference (Globecom), 2022.

Acknowledgment

In the name of Allah, the Most Gracious, the Most Merciful. Praises and thanks are due to Allah who bestowed upon us endless blessings, and the faculties of seeing, thinking, and learning. My endeavors and plans were seen through only with His guidance and sustenance.

First I dedicate this thesis to my late supervisor Prof. Mohieddine Benammar and my late teacher Mr. Ibrahim Khadr, they continue to inspire me with their professionalism and their love for knowledge, may Allah bless their beautiful souls.

I would like to express my utmost gratitude to Prof. Nizar Zorba and Prof. Hossam Hassanein for their unmatched support, supervision, mentorship, and patience in teaching me the know-how of research and their valuable advice.

I'm infinitely indebted to my parents for their unrelenting support, their love, and for being my anchors. I'm what I'm because of them. To my siblings who filled my childhood with laughter and good times.

I would like to give special thanks to my very dear friend Sherif Azmy, who has been the one person I can share any kind of news with him, and for being a true friend, I'm indebted to you for life. And to my friend Mhd Saria for he was the best lab partner I could have asked for, Mohammed Anas, Ibrahim Amer, Ziad Mansour, and all the good people of Kingston who made me feel a bit at home. And lastly to my dear friends back in Qatar.

Table of Contents

List of Figures

List of Abbreviations

1D	One dimension
1G	First-Generation communication technology
2D	Two dimensions
2G	Second-Generation communication technology
3D	Three dimensions
3G	Third-Generation communication technology
3GPP	Third-Generation Partnership Project
4G	Fourth-Generation communication technology
5G	Fifth-Generation communication technology
6G	Sixth-Generation communication technology
AWGN	Additive White Gaussian Noise
BP	Belief Propagation
BS	Base Station
CSI	Channel-State Information
D2D	Device-to-Device
DFT	Discrete Fourier Transform
DL	Downlink
eMBB	Enhanced MBB
eNB	eNodeB
eNodeB	E-UTRAN NodeB
FCC	Federal Communications Commission
FDD	Frequency Division Duplex
FDMA	Frequency-Division Multiple Access
FDS	Frequency Domain Spreading
FEC	Forward Error Correction
FFT	Fast Fourier Transform
GB	Guard Band
GPS	Global Positioning System

GSM	Global System for Mobile communications
HSPA	High-speed Packet Access
ICIC	Inter-cell Interference Coordination
IDFT	Inverse Discrete Fourier Transform
IGMA	Interleave-Grid Multiple Access
IFFT	Inverse Fast Fourier Transform
IoT	Internet of Things
IoE	Internet of Everything
IP	Internet Protocol
ITU	International Telecommunications Union
KPI	Key Performance Indicator
LAN	Local Area Network
LSSA	Low Code Rate and Signature-Based Shared Access
LTE	Long-Term Evolution
MA	Multiple Access
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MIMO	Multiple Inputs Multiple Outputs
mMTC	Massive MTC
MTC	Machine-type Communication
MTD	Machine-type Device
NB-IoT	Narrow-band Internet of Things
NOMA	Non-Orthogonal Multiple Access
OFDM	Orthogonal Frequency-Division Multiplexing
PD-NOMA	Power Domain Non-Orthogonal Multiple Access
PHY	Physical layer
PRB	Physical Resource Block
QoS	Quality-of-Service
RAB	Radio-Access Bearer
RAN	Radio-Access Network

RB	Resource Block
RDMA	Repetition Division Multiple Access
RE	Resource Element
RLAN	Radio Local Area Networks
RRC	Radio-Resource Control
RRM	Radio Resource Management
RRS	Radio Resource Scheduling
RSMA	Resource Spread Multiple Access
Rx	Receiver
S1	Interface between eNodeB and the evolved packet core
SBS	Small Base Station
SDMA	Spatial Division Multiple Access
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference-and-Noise ratio
SIR	Signal-to-Interference ratio
SR	Scheduling Request
TDMA	Time-Division Multiple Access
TD-SCDMA	Time-division-Synchronous Code-Division Multiple Access
Tx	Transmitter
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telecommunications System
URLLC	Ultra-Reliable Low-Latency Communication
VoIP	Voice-over-IP
WLAN	Wireless Local Area Network
X2	Interface between eNodeBs.

List of Symbols

$\mathbb{R}^d$	d-dimensional coordinate space
f	frequency (unless specified otherwise)
L_T	Taps in a tapped delay line
E_n	Encoder
Bl	Transmission block length
C	Code Library (or book, unless specified otherwise)
δ_i	Total transmission power fraction of each device
cov	Covariance Matrix
l_x	Lifetime distribution function
T_n	Occurrence time
O_n	Inter occurrence time
W	Bounded region (unless specified otherwise)
$\lambda_2(W)$	Area of of W
β	Uniform intensity
μ	Mean (unless specified otherwise)
ρ_u	Problem parent point process
O_u^{Θ}	Problem daughter process
$\mathcal{M}$	The set of MTD

$\mathcal{S}$	The set of sub-channels in the system (unless specified otherwise)
s	Sub-channel s in the set S
$\mathcal{C}$	($\mathcal{U}$) The set of groups (the set of ranks in each group)
$\mathcal{B}$	The set of small base stations in the network
u_{max}	The maximum number of MTD in one group
$\gamma_{s,c,b}$	The binary indicator whether to allocate the s^{th} sub-channel to the c^{th} group to b^{th} SBS
$p_{s,m}$	The transmit power of the m^{th} MTD over the s^{th} sub-channel
$\mathcal{N}_0$	AWGN
$\alpha_{m,u,c,b}$	The binary indicator whether to assign the m^{th} MTD to b^{th} SBS to the u^{th} rank of group c
$R_{m,b}$	The total transmission rate of the m^{th} MTD in the b^{th} SBS
W	The bandwidth of a single sub-channel in one PRB
$h_{s,m}$	The channel gain of the m^{th} MTD over the s^{th} sub-channel
R_m^{th}	The minimum transmission rate of the m^{th} MTD
P_m^{max}	The maximum power threshold of the m^{th} MTD
ω	Linear scalarization weight (unless specified otherwise)

Chapter 1

Introduction

1.1 Overview and Motivation

The densification of an entire network is one of the primary motivations for improving next-gen networks' Spectral Efficiency (SE), including 5G and already established standards like 4G and its variants. This densification entails a dynamic change in the number of base stations and users and is not limited to only human users but also Machine-type Devices (MTDs). We define Radio Resource Management (RRM) as a network's ability to utilize radio resources (higher spectral efficiency) efficiently. RRM in Radio Access Networks (RAN), in particular, provides a way to manage radio resources in single and multi-cell scenarios (e.g., assign, reassign, and release). Moreover, with each generation, the demand and challenges vary drastically. In 4G the core components of Long Term Evolution (LTE) and subsequent releases were highly specialized in providing a seamless "super-fast" connection between users and the entertainment service providers (e.g., Netflix, YouTube) [1,2]. In comparison, 5G infrastructure is designed to provide a greater coverage and connection capacity to human and machine users alike. The unprecedented surge in the numbers of connected IoT devices [3], introduces a new set of paradigms currently analyzed by research and development communities. Therefore, the huge number of MTDs asking for connections and service, in turn, led to the emergence of paradigms that try to tackle problems related to their service either over 4G, 5G or through a hybrid approach. congenital to densification in communication networks, whether it is in time or frequency domains.

1.2 Challenges

Currently deployed systems lack flexibility and adaptability, and are in ever-consistent need of upgrades and fixes, the primary challenge most 4G/5G networks face is to provide acceptable Quality of Service (QoS) and efficiency for a large number of MTDs. That is where

dynamic Radio Resource Scheduling (RRS) becomes essential [3], as we can assign spectral resources to MTDs while achieving the minimum QoS requirements, whether in fairness, delay, or throughput. Challenges in dense networks can be many-sided. For example, some MTDs have bursty traffic patterns while others require higher rates and lower latencies, as we will elaborate further in Chapter 2. How we assign resources to the MTDs is a key consideration as it plays a core role in the performance of each device with respect to its neighboring devices and their relation with the base station, the channel conditions, and scheduling decisions.

Supporting a larger density of connections (e.g., bursty traffic from proximity sensors [3, 4]) per cell also means putting a larger load on 4G/5G channels [5, 6]. Some of the indispensable facilitators of next-gen networks are energy harvesting and saving, but its application to MTDs remains an open problem [7, 8]. Moreover, heterogeneity paradigms remain a big challenge [9, 10], which hold a significant impact on how we conceptualize the state of future networks, such as the Internet of Everything (IoE) [11]. We define heterogeneity in a network where users access the medium using two different techniques: Power-Domain Non-Orthogonal Multiple Access (PD-NOMA) and Orthogonal Frequency-Division Multiple Access (OFDMA), and we try to optimize their resources to satisfy the QoS requirements for the MTDs.

1.3 Objective and Contribution

The objective is to achieve tangible improvement in the spectral and scheduling efficiency in the network for MTDs while considering a HetNet setting; hence, the scheduling will be subject to heterogeneity constraints. We build an uplink (UL) scheduler for 4G/5G networks by investigating a multi-cell scenario, the users are grouped into ranked n-groups, some devices operate using 5G proposed PD-NOMA, and other devices operate using 4G OFDMA. We then solve the relaxed and non-relaxed NP-Hard optimization problem formulated for this scenario. Including the constraints that we are going to formulate in Chapter 4. The

proposed uplink scheduler can be illustrated as shown in Fig. 1.

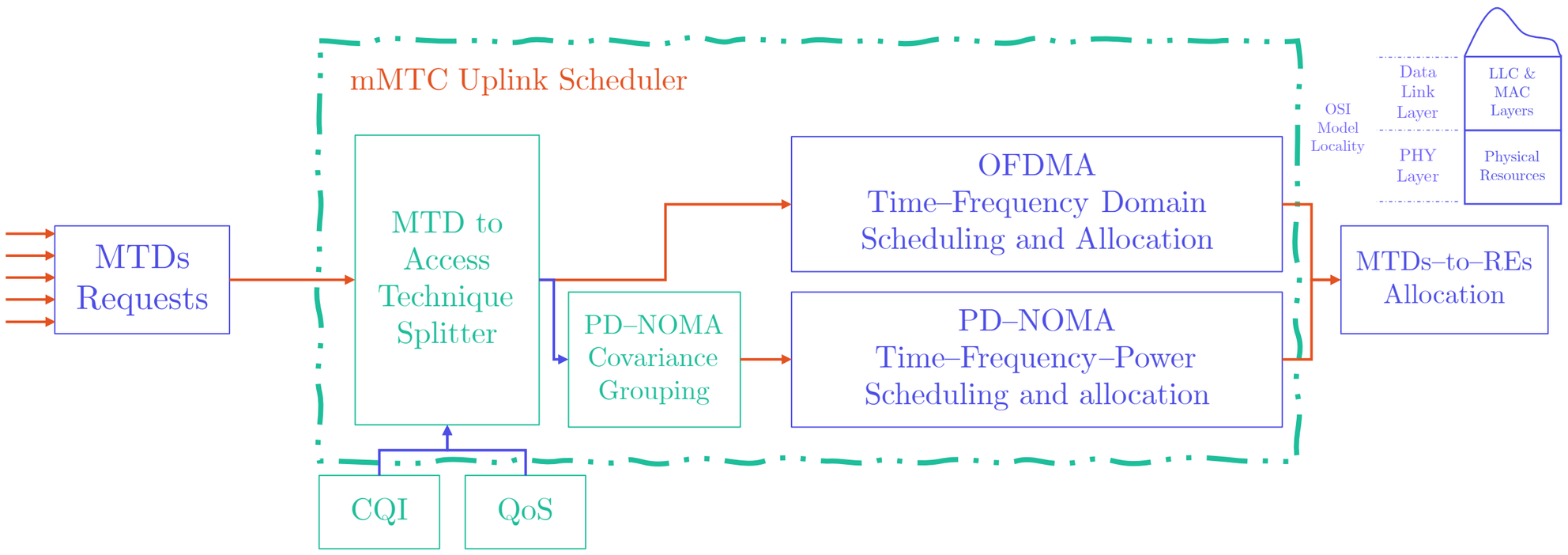


Figure 1: Proposed mMTC Uplink Scheduler

In this work, we have a multi-cell scenario and we want to provide a resource allocation and scheduling solution for 4G/5G connected MTDs. While 4G is based on OFDMA that guarantees interference-free service and higher QoS satisfaction, 5G in future releases is prospected to use PD-NOMA which is more spectrally efficient and allocates more users, but interference is allowed and thus less QoS satisfaction. The optimal point in the allocation of 4G to 5G users will be analyzed.

For example, we will reallocate some or all the resources in the 4G domain to users with particular QoS requirements, where they will suffer less from uncertainties affecting the channel, like interference from other MTDs, as they are using OFDMA instead of NOMA, so the orthogonality would alleviate the problems with channel conditions. MTDs with higher rate requirements will be scheduled on 5G network, where they will have to rely on increased technological advances in interference cancellation and processing power of Small Base Stations (SBSs).

The novelty of this work is that we explore the possibility of deploying a dynamic RRS algorithm as a solution to the relaxed optimization problem. We also give an insight into network trends under heterogeneity constraints in a Narrow-Band Internet of Things (NB-IoT) framework. The problem at hand incorporates orthogonal and non-orthogonal

multiple access, which will provide a perspective on how integration is realized in future networks. The main contributions of this thesis are as follows:

1. We set up a scenario that couples orthogonal and non-orthogonal multiple access techniques, while considering QoS requirements for MTDs.

2. We formulate a constrained Optimization Problem (OP) to maximize the sumrate of OFDMA and PD-NOMA connected MTDs.

3. We provide a Pareto optimal solution to a relaxed OP through a heuristic algorithm, where the formulated non-relaxed OP is NP-hard.

4. We develop a simulator that considers many features of standardized systems; we then test the simulator's performance and limitations under various scenarios.

1.4 Thesis Organization

This thesis is organized as follows: Chapter 2 introduces RRM paradigms under the umbrella of the Internet of Things (IoT) and specifically NB-IoT, with a comprehensive examination of current literature trends; Chapter 3 defines NOMA and OMA techniques and gives a formal introduction to MTD and SBS configuration schemes; Chapter 4 discusses dynamic radio resource management optimization problem formulation, together with its proposed solution; Chapter 5 showcases simulations and discusses the results; Chapter 6 presents conclusions and an overview of future work.

Chapter 2

Radio Resource Management Paradigms within the Context of IoT

Telecommunication networking became an established industry [12] during the 70s and 80s, with the introduction of the first commercial handheld mobile devices in 1983. A boom in demand for new devices necessitated a standardization effort to tackle fundamental challenges with the increasing new and future demand over limited radio resources. Consequently, one of the first standardization organizations formed was the 3rd Generation Partnership Project (3GPP)[1]. We shall briefly examine how RRM paradigms evolved to gain an insight into past solutions, and how the current state of standards compares to earlier stages of standards' development.

2.1 On the Evolution of RRM Paradigms

2.1.1 2G - 3G

2G was revolutionary in terms of capacity, bandwidth, and coverage, improved security, and the creation of standardized complex subsystems [13]. This revolution was possible with the support of many new medium access techniques, and algorithms. 2G marked the first move towards digital communication. 2G[2] relied primitively on Time Division Multiple Access (TDMA) and in later releases, Frequency Division Multiple Access (FDMA) [14–16]. RRM systems in this generation were the first to address some of the most common problems that accompany surging demand for bandwidth in a modern telecommunication system, all while trying to meet the QoS of that era. In the quest for more bandwidth and ever-increasing demand for higher rates, 3G was the first to use Wideband-Code-Division Mulitple

[1] which was preceded by the Advanced Mobile Phone System (AMPS) standard family developed in Bell Labs in 1973

[2] standardized by European Telecommunications Standards Institute (ETSI) and deployed in Finland by Nokia

Access (W-CDMA) access technology which offered, in return, unprecedented speeds [17]. From the RRM perspective, the Radio Resource Control (RRC), Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC) and Medium Access Control (MAC) protocols comprise the User Equipment (UE) interface of RAN. We will focus on the RRC protocol as other parts of the UE interface with RAN are slightly out of scope. RRC protocol was first introduced for 3G and was later enhanced and upgraded on several domains in the subsequent releases [18–20]. The RRM function in 3G core architecture includes Power Control (PC), Handover Control (HC), Load Control (LC), and Packet Data Scheduling (PS), as well as the Resource Manager (RM), [19, 21]. PS controls all Non-Real Time (NRT) traffic, including allocating suitable bit rates and scheduling packet data transmission while ensuring proper QoS in terms of packet delivery ratio and latency.

2.1.2 4G

4G was promised to be the second big revolution since 2G, whose true potential was only realized with releases 10 and 11 of 4G. The key features of 4G, spread spectrum medium access techniques employed in 3G systems were abandoned and replaced with OFDMA in all 4G/+ candidate systems. We will cover OFDMA in detail in Chapter 3 [4, 22–24]. 4G came with plenty of adequate changes to 3G – 3.9G established technologies, with the introduction of OFDMA and the stringent requirement for an All-Internet Protocol (IP) packet-switched networking, 4G was the first standard to be adopted globally for cellular communication. 4G in later releases was able to support practical rates of up to 1 Gbps Downlink (DL) speeds. 4G was the first to introduce support for mMTC and Device-to-Device (D2D) connections. From the RRM perspective, there have been many advances on many levels of the general RRM function. RRM function falls under the jurisdiction of Radio Protocol Architecture within RAN. RAN is homogeneous in terms of core architecture design as it was designed to use only one type of node, eNodeB, which by extension is responsible for all radio-centric functionalities of a single cell or multiple, eNodeb is connected to the core

network through the S1 user and control planes links and the X2 interface links to each other for the purposes of RRM and Inter-Cell Interference Coordination (ICIC). There are four main entities of RAN, Physical layer (PHY) [3], Medium-access control (MAC), RLC, and Packet data convergence protocol (PDCP). The main functions of these protocol entities are detailed in Fig. 2. The scheduler is a component of the MAC layer that manages the

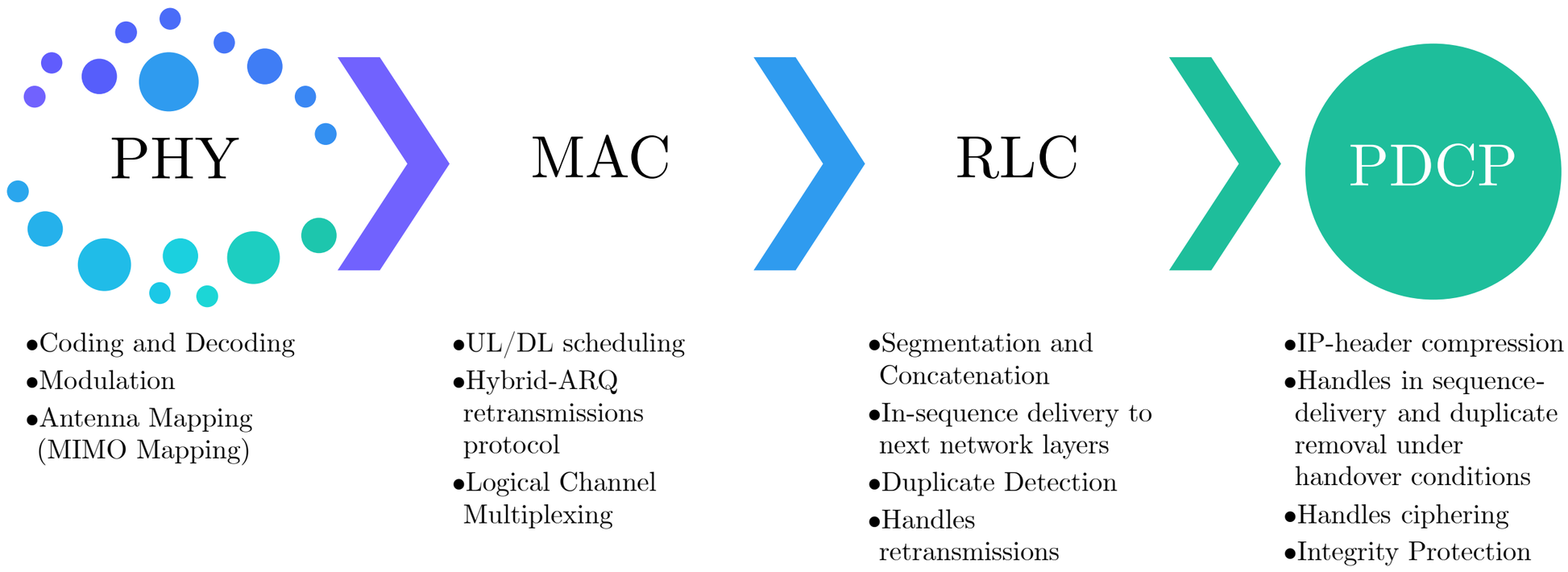


Figure 2: 4G RAN Protocols

allocation of UL and DL resources in terms of resource-block (RB) pairs. RB pairs are time-frequency frames of size 1 ms $\times$ 180 kHz. The design of the scheduler is not detailed in any of the standard releases and is left for the vendor or the operator to design its protocols as long as it complies with the standard's QoS requirements. It is also important to point out that scheduling is a function not only native to the MAC layer but to all the layers of RAN; nevertheless, it is conventionally referred to as an entity of the MAC layer. Thenceforth, we will focus on UL. Scheduling UL and DL transmissions are separate although similar; UL scheduling is done per UE, not per Radio Access Bearer (RAB). As a result, while the eNodeB scheduler regulates the UL load of a scheduled UE, the UE is still responsible for deciding which RAB/s to transmit over.

[3]Since we are working from the physical layer perspective, it is considered the highest layer

2.1.3 5G and Future Networks

With 3GPP Release 15, 5G was promised to be a game changer for telecommunication standards; however, its full potential remains unrealized. This subsection discusses its core architectural components following the works [3, 23, 25–27]. 5G was designed to achieve much higher rates, spectral efficiency, lower delay, and massive coverage compared to preceding generations. Typical use cases would include users of three types under two categories human users (enhanced Mobile Broadband (eMBB)), machine users (Ultra-Reliable Low-Latency Communication (URLLC), and massive Machine-type Communication (mMTC)). 5G theoretically can reach speeds of 20 Gbps in DL and up to 10 Gbps for UL connections. Perhaps latency is the most demanding requirement of providing latency of 1 ms in the user plane with a reliability of 99.999% probability of successful packet transmissions and the ability to cover 1 million devices per km^2 which is a specification tailored for mMTC connections with constrained energy consumption. OFDMA was introduced in 5G as the default medium access technique (we will discuss this in detail in Chapter 3) with the possibility of including NOMA access techniques in future releases of 5G.

Flexible numerology, Bandwidth Parts (BWPs), service multiplexing and mini-slotting, streamlined frame structure, Massive MIMO, high and low-band inter-networking, and ultra-lean transmission are key enablers to satisfy the requirements of this new era. The BWP allows for greater flexibility in how resources are distributed to a given subcarrier. BWPs are just about having versatility so that in a given bandwidth, many, distinct signal types can be sent. Most base stations will use the larger bandwidths available in 5G. However, UE capacities will vary, and it will be more difficult for some UEs to use the greater bandwidths available. BWPs allow multiplexing of various signals and signal types for better spectrum and UE power utilization and adaptation [28]. With BWPs, for various purposes, we can divide and use the subcarrier. Every 5G NR BWP has its own numerology, which ensures that we can configure each BWP differently with its particular signal characteristic, allowing for more efficient use of the spectrum and power. This role of flexible configuration is good for

integrating signals with various specifications. OFDMA offers an even further partitioning of resources in time (as granular as symbol times) and frequency (subcarrier-wise) domains, as well as support for massive MIMO. Moreover, 5G is designed to coexist with earlier standards and future ones. First, reserved resources which were not accessible for transmission can be configured by NR. Reservation leaves resources empty, allowing them to be exploited for future extensions. Secondly, physical signals and channels are contained within radio resources that we can configure and use. These properties provide future flexibility while remaining backward compatible. Third, NR reduces the number of always-on broadcasts. 5G core network introduces the service-based architecture, which focuses on facilitating connections between services rather than nodes. 5G RAN structure is similar to that of 4G as of Release 17, with specifications geared towards standardizing Non-Terrestrial Networks (NTNs) and satellite communication integration with 5G/NR.

Future Networks refer to 6G and beyond, although 6G is still in its infant stage of development, we touch briefly on the vision and possible enabling technologies of 6G. 6G is expected to provide data rates in the region of Tbps with latency as low as 100 μs, and up to 10 million connections per km^2. Typically every generation is supposed to offer, on average, an order of magnitude enhancement over the capabilities of the preceding generation. Many enabling technologies are being investigated for viability and integration; among those, we are concerned with the PHY impact of such technologies as Terahertz communication, which enables data rates at least ten times the current 5G capabilities, simply due to the large bandwidth availability in a spectrum range of 95 GHz to 3 THz. Nonetheless, operating in sub-THz bands can be problematic, as propagation can be highly direction-dependent, and signals suffer from compounded channel response due to molecular absorption or path blocking, where these effects are still under investigation. Ultra-Narrow beams, also known as Pencil beams, introduce a complex challenge to interference management, impacting medium access control and handover. Terahertz communications will add another layer of complexity to physical layer designs. Hence, solutions for modulation, coding problems under RF

hardware limitations, and low power AD/DA conversion circuit limits will be among the prioritized. In the class of new promising technologies, is the convergence of communication, sensing, and computing networks resulting in what the literature would refer to as the IoE, which will enable a level of QoS that is unprecedented, and by extension gives rise to context-aware communication.

2.2 An Introduction to NB-IoT Standard

In this section, we introduce the main components of the NB-IoT standard along with this work's consideration to comply with the standard. Our primary focus is NB-IoT physical layer design components from a UL perspective [29–35]. The majority of NB-IoT's functionality, as well as its core design of channels and signals, are inherited from 4G. However, to meet the low-cost and low-power requirements of NB-IoT-connected devices, the complexity of these features was reduced. The number of channels and signals was decreased and modified to meet the new NB-IoT frame topology. The system was intended to operate with a frequency bandwidth of 180 kHz (equivalent to one resource block in the 4G system) and to handle a large number of repetitions to enable long-distance broadcasts and deep interior penetration. The NB-IoT system was developed with substantial reuse of the 4G architecture. This approach enables quick and flexible deployment across older 4G cellular network infrastructures while maintaining the compatibility of the two technologies. NB-IoT system reuses the modulation techniques for DL and UL transmissions, namely OFDMA and Single-Carrier FDMA (SC-FDMA).

2.2.1 Operation Mechanisms

NB-IoT was designed with air interface in mind, to ensure that the system can coexist with legacy 4G carriers while maintaining the reliability of 4G systems. There are three main operation modes that are defined by 3GPP. As shown in Fig. 3, the three mechanisms are in-band, guard-band, and standalone.

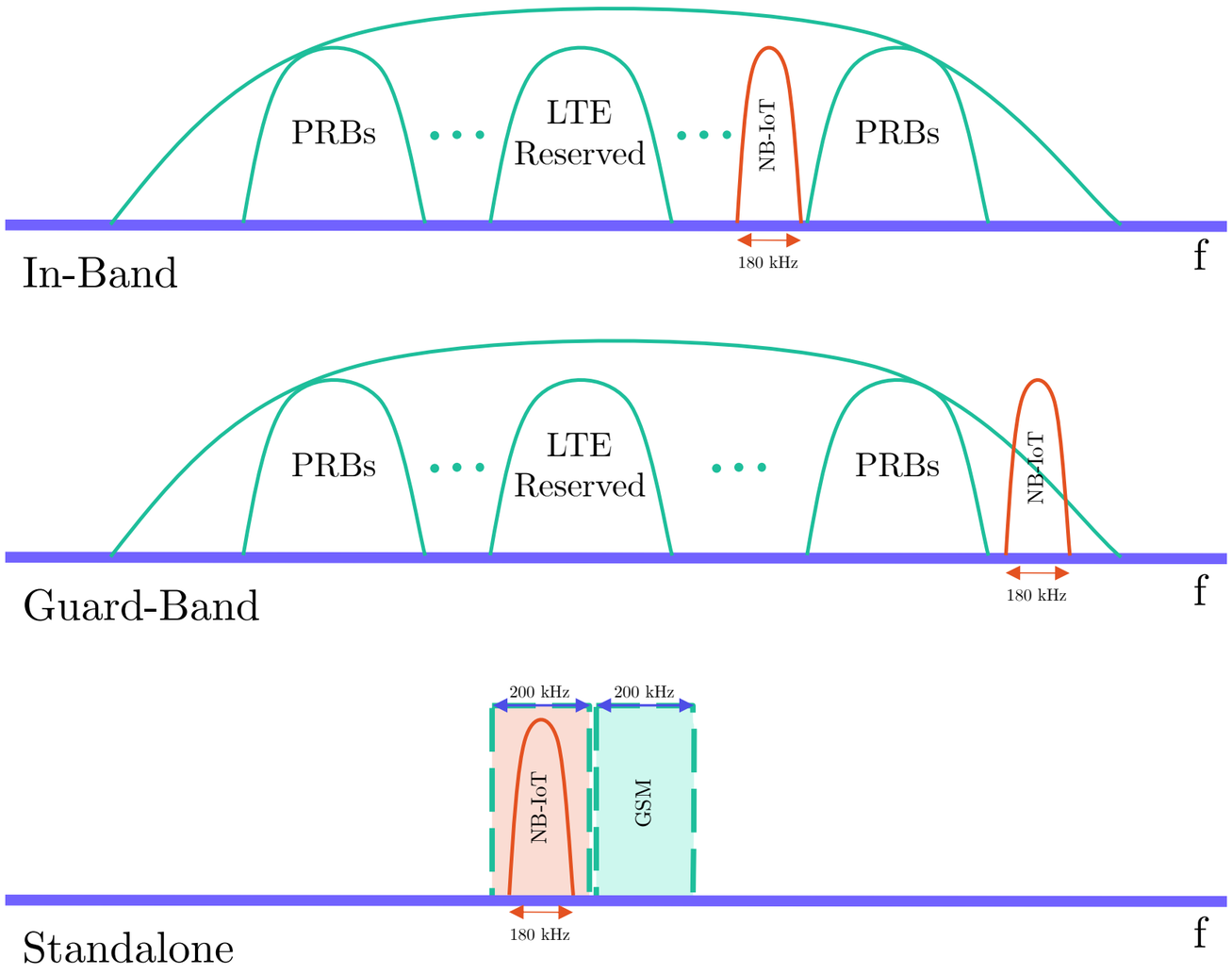


Figure 3: NB-IoT operation mechanisms

The In-band mode is where the NB-IoT Physical Resource Block (PRB) overtakes a single PRB from the bandwidth. Guard-band mode occupies a PRB of the full bandwidth typically PRB used for band-guarding from other carriers hence the name. In Standalone mode, the NB-IoT SubFrame (SF) overtakes a free spectrum of a 2G carrier, the NB-IoT still has a bandwidth size of 180 kHz, but it also has a band-guard of 10 kHz on both sides of the liberated spectrum. To minimize conflicts between NB-IoT transmissions and some key 4G channels and signals, such as the physical broadcast channel and synchronization signals, NB-IoT transmissions are prohibited for the six middle PRBs of the 4G system. Any PRB can be utilized for NB-IoT broadcasts in the guard-band mode. In this work, we will assume all our transmissions are in the in-band operation mode because it offers the most prominent advantages in resource utilization and simplicity of integration over older 4G networks.

2.2.2 Transmission Mechanism

With 3GPP Release 13 NB-IoT was devised to operate using Frequency-Division Duplexing (FDD). Using FDD implies that DL and UL communications occur in distinct frequency

ranges. To put it another way, the eNB and UE will transmit in one frequency range while receiving in another. However, NB-IoT UE modules' power, complexity, and battery life constrain the system. On the UE side, UEs transceive in a type B Half-Duplex-FDD (HD-FDD) mode. HD-FDD mode indicates that the UE can either send or receive data, but not both, with a time-guard between transmissions. Release 15 introduced the use of Time-Division Duplexing (TDD) for UL and DL transmissions.

2.2.3 NB-IoT Frame Structure

NB-IoT was conceived as a legacy 4G descendant; thus, it retains the same frame structural design as 4G, where the UL and DL transmission alike form 10 ms physical frames. The differences lie mainly in the signal and channel maps. NB-IoT frames consist of 10 SFs; the SFs are composed of two-time slots of 0.5 ms each for a total of 1 ms. System Frame Indexing (SFN) is then used to index frames with maximum allowable indexes of 1024, hence it occupies 10.24 s total frames time.

Physical Signal Frame Structure

On a larger scale (signal-wise) NB-IoT UL SF is structured as 12×14 REs where we have 12 subcarriers (if we choose 15 kHz spacing or 48 subcarriers if we choose 3.75 kHz) and 14 OFDMA symbols (each is 8.33 μs). However, if 3.75 kHz spacing is chosen, the radio frame type and REs representation will change. As a result, with 3.75 kHz spacing, the slot duration is four times greater, resulting in a slot length of 2 ms. This configuration of each frame means that a UL physical frame will consist of five 2 ms slots, where the duration of each Cyclic Prefix (CP) is 266.67 μs. In contrast, if a 15 kHz spacing is used the CP of the first symbol only lasts 5.2 μs, while the remaining six symbols have a CP that lasts 4.7 μs. Fig. 4 illustrates the NB-IoT UL framing system with a subcarrier spacing of 3.75 kHz.

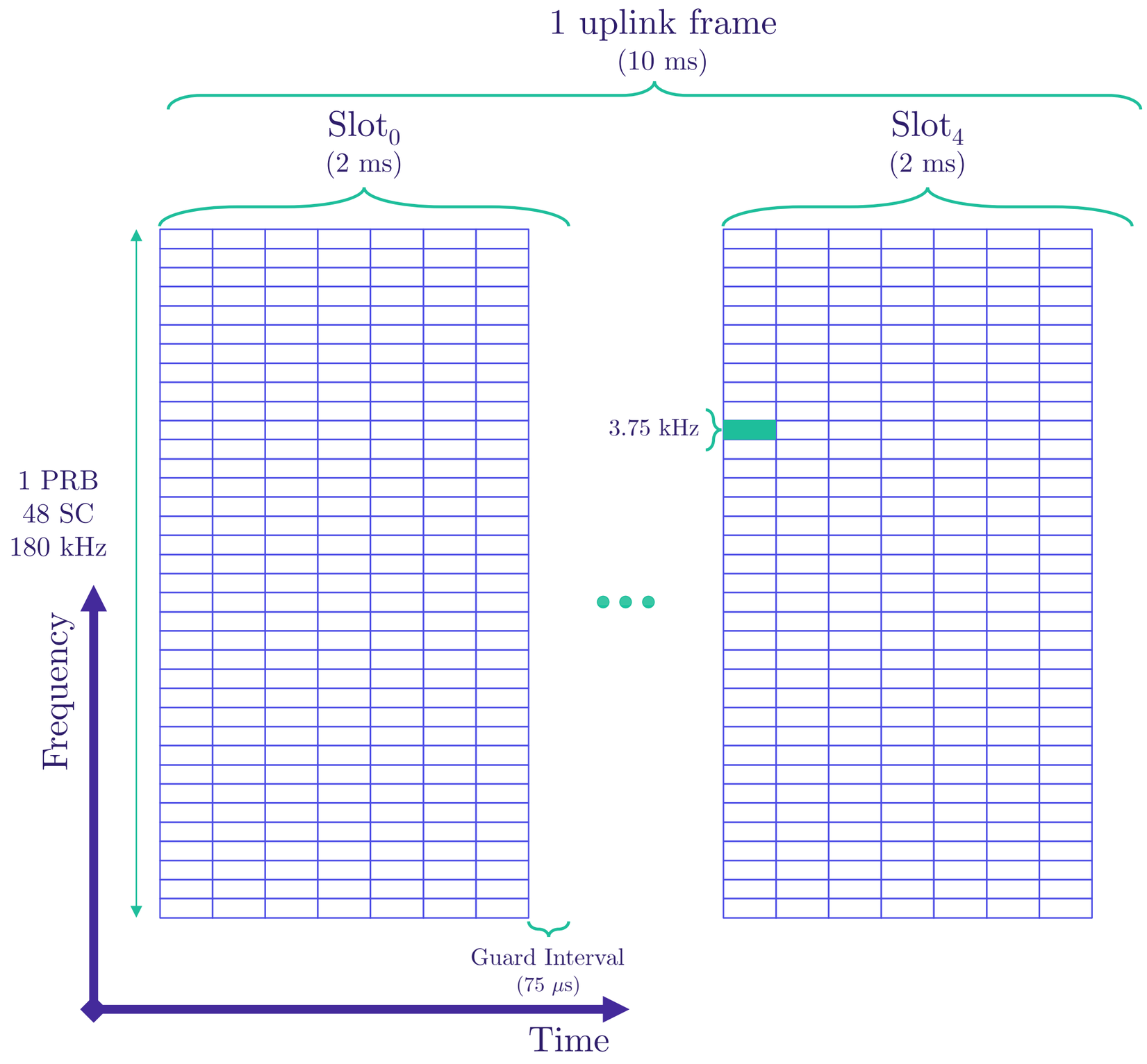


Figure 4: NB-IoT UL frame structure with subcarrier spacing = 3.75 kHz

2.2.4 Machine-type Communication in HetNet Settings

This subsection briefly introduces some of the main challenges and considerations of MTC, when designing a communication framework under heterogeneity constraints [3,36–39]. Conventionally, mMTC challenges differ depending on the type of communication constraints of the type of machine devices. Hence, MTC literature would fall under three categories mentioned in Chapter 2, mMTC, URLLC, and eMBB machine devices. URLLC and eMBB fall outside our scope; thus, we will not cover their challenges in depth here. Ultra reliability and low latency are very evident of the main requirements of URLLC type of devices usually robotics in a massive factory, or remote surgery robots, where their control signals and

transmissions are mission critical, thus they fall under the URLLC category. However, for example, 8K streaming and VR/AR offloading to an edge server would naturally require much higher throughputs and low latencies, but not so much reliability, henceforth they fall under eMBB type of devices.

mMTC focuses mainly on providing coverage and acceptable data rates for millions of connected devices (typically sensors and actuators) uploading their readings and reports to the cloud or an edge server. We detail how NOMA schemes work, but not their importance as access techniques, NOMA promises a solution to the problem of massive connectivity in Ultra-Dense Networks (UDN), as it tackles many challenges posed by mMTC constraints. However, the non-orthogonality of NOMA comes at the cost of more sophisticated receivers but offers much better coverage and low latencies for connected MTDs.

2.3 Related Work

In this section, we carefully examine the literature contribution and related works to our scope and scenarios. **Classical Formulation** the paradigm of radio resource allocation/scheduling is generally a non-convex NP-hard problem [40] and therefore computationally complex, particularly as the density of the network increases. Researchers have developed a range of centralized and distributed radio resource allocation algorithms (User-Pairing as in [41], or Game-based algorithms [42–44]), using various techniques such as weighted minimum mean square optimization, information-theoretic [45], and fractional programming [46]. **Clustering** is a fundamental building block in this work, usually, it is considered a process to generate points in a 2D space with little regard to the deep impact on performance which we will showcase in Chapter 3 and 4, some works employ techniques of general and cluster point processes and solve their problems with respect to the properties of the point processes [47] formulated an optimization problem constrained by properties of cluster process, they focused on maximizing performance for mMTC, and URLLC devices, however, they considered a single cell scenario and relied upon PD-NOMA for transmission. Thus, clus-

tering was useful as it can mitigate some challenges introduced by the full-collision property of NOMA. [48] proposed another use of the clustering process in Wireless Sensor Networks (WSNs) to balance energy burden, for example by picking a sensor node with sufficient energy from a cluster, to compensate for sensor nodes with inadequate energy. Thus, it will improve the life expectancy of WSNs. [49] was an important inspiration for the investigation of cluster processes in general, in this paper they were trying to bridge the gap between Poisson point processes and tier-based HetNets. They developed models for HetNets for different types of BS and UE configurations, which resulted in a unified model for simulating the non-uniformity of BSs and UEs locations throughout the HetNets. **Wireless Sensor Networks** is a direct application of the MTC model of communication since we are dealing with devices that have limited energy supplies and transmit in a bursty manner. [50] proposed a similar use of clustering based routing protocol to [48] to prolong WSNs life expectancy.

In the problem formulation, we relax the energy constraint (refer to Chapter 4) as it introduces a complexity of the same hardness as the joint resource scheduling and allocation, making our problem multifaceted with strenuous constraints. However, the literature and we believe it is of utmost priority to focus on **Energy Aware** communication due to its deep impact on the network as a whole, for example, that MTDs have limited energy supplies, and PD-NOMA works by ranking devices by predetermining their transmission power. Thus, energy-aware constraints are integral to any resource allocation problem formulation in current and future networks. [51] proposed the use of Deep Reinforcement Learning (DRL) in a D2D setting to solve a joint resource allocation and power control problem, to improve spectrum utilization and system capacity. [52] discussed the use of Intelligent Reflecting Surface (IRS)-aided Wireless-Powered Communication Network (WPCN) to minimize power consumption and maximize total throughput, they used a penalty-based algorithm to solve their non-convex optimization problem. [53] studies the consequences of facilitating extreme coverage over NB-IoT networks, where they offer channel-scheduling-based solutions. A

tractable analytical framework has been proposed to investigate the impact of control and data channel scheduling, as well as the coexistence of coverage classes, on the power consumption and latency of IoT devices. Their results demonstrate a considerable association between allowing extreme coverage for devices experiencing enormous path loss and degradation of battery life and latency for IoT devices experiencing lower path loss. **Reinforcement Learning (RL)** promises to be a solver for everything, and we have seen early attempts by researchers to use RL for RRM paradigms, where the authors in [54] used RL to control a Fuzzy-Neural model, to ensure QoS specification. The above-mentioned algorithm was analyzed in a scenario involving 3G, GSM/EDGE RAN (GERAN), and WLAN radio access technologies. Single-cell deployment was considered to easily test the actions of the algorithm. The best RAT and the assigned bit rate are given to each user both during the admission process and during the length of the session.

Authors in [55] and [56] approached the RMM paradigm in a formally similar manner but for different applications. [55] proposed the "Learn to schedule (LEASCH)" algorithm, a DRL-inspired approach to solve RRS in the MAC layer, LEASCH trains Dueling Deep Q Network (DDQN) agents. LEASCH can learn scheduling tasks from the ground up with no prior knowledge of the RRS framework, which is a typical advantage of DRL algorithms over AI-Based solutions [57–59]. While [56] mainly presented the use of basic RL algorithms like Q-Learning, and Actor-Critic networks, although, under the scope of vehicular networks. Furthermore, the Internet of Vehicles (IoV) has an intrinsic property, which is that vehicles have partial observability. Hence, proactive AI-based measures, sometimes are not enough, especially in overloaded scenarios, as long-term dependencies tend to form in stochastic networks, hence DRL is perfect for such applications. [60] introduces the Multi-Agent concept to resource allocation, in which each transmitter is treated as an agent. The neat property of Multi-Agent deployments is that it allows agents to take actions concurrently and in a distributed way while remaining oblivious of the simultaneous decisions of certain other agents, thus called partial observability [61, 62].

Chapter 3

Medium Access Techniques and Methods of Configuring SBSs and MTDs

In this chapter, we shall lay the bases on which we build our problem formulation, in the first section we discuss in detail the origins and mathematical properties of orthogonal multiple access and non-orthogonal multiple access techniques. We will focus on OFDMA and PD-NOMA. In addition, we will briefly discuss other techniques (e.g., CD-NOMA, RSMA).

3.1 Orthogonal Frequency-Division Multiplexing

OFDM has been the driving force of modern telecommunication as it is the most universally deployed medium access technique over the past decade, due to its robustness against different types of co-channel interference (with various distributions), multipath, and frequency selective fading [63–71]. Below we list some of the main fundamental characteristics of OFDM (advantages and disadvantages):

- It eliminates the need for sophisticated equalization filters. Therefore, channel equalization is simplified because OFDM uses numerous slowly modulated narrowband signals in contrast to using wind-banded signals with rapid modulation.
- Its robustness against Inter-Symbol Interference (ISI).
- It enables the adoption of Single Frequency networks, which are particularly appealing for the broadcasting type of applications.
- It has a much higher spectral efficiency over traditional Frequency Division Multiplexing (FDM) systems.
- However, it is highly sensitive to Doppler shifting. Hence the need for very precise time and frequency synchronization.

- The main drawback that affects spectral efficiency is the use of Cyclic Prefix (CP) and guard bands.
- A non-linear power amplifier transmission would be detrimental to the whole bandwidth range, Hence, a considerable out-of-band power leakage occurs, resulting in inter-carrier interference.

Moving forward we will center on **UL transmissions** and the considerations entailed under that scope. Another important assumption that will be integral to the system design, as well as the problem formulation, is that MTDs can transmit on both 4G and 5G channels but are only permitted to use one at a time as conventionally used in practical systems. The choice of which MA to use will be specified in the following sections.

3.1.1 Operation Mechanisms

The basic principle of OFDM system operation is that the spectrum is partitioned into narrowbanded controllable sub-carriers, hence they almost have flat fading. The high spectral efficiency (as shown in Fig. 5) is manifested when we realize that the bandwidth is allocated more efficiently with orthogonal and overlapping sub-carriers.

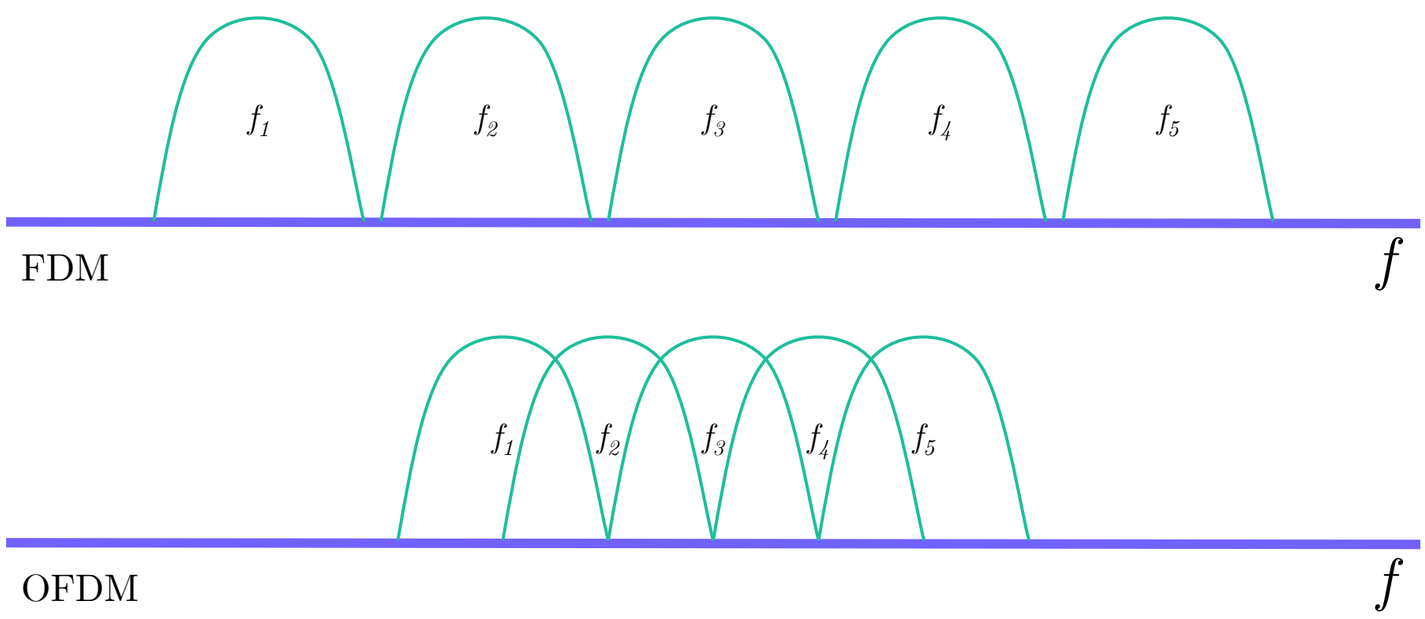


Figure 5: Spectral efficiency comparison between FDM and OFDM systems using a full cosine OFDM spectrum

The orthogonality is retained even with a time-dispersive channel. Shaping of the frequency spectrum allows the subcarrier to be sampled only at its peaks, thus the remaining

subcarriers have zero crossings at the peak, hence the sampled subcarrier experiences no interference. If this sampling occurs during the off-peak period, there may be interference from nearby subcarriers. Furthermore, by not truncating the spectrum of each subcarrier, the demands on resources are reduced, and the symbols can be time-limited. However, typically data symbols transmitted on distinct subcarriers are received with no interference. To elaborate on the implementation of OFDM, the symbols must first be analyzed in terms of frequency. Time-domain representations are obtained by computing the Inverse Fast Fourier Transformation (IFFT) of the data symbols. This temporal representation is given a CP, which is formed by blending it with an interval of the time representation of the symbols. After adding the CP, this data is broadcast across a frequency selective channel. The CP is omitted at the receiver, and the symbols are provided by the Fast Fourier Transformation (FFT) of the remainder. It is vital to choose a CP with a duration larger than the minimum delay spread interval.

Discrete–Time OFDM Model

Discrete-time modeling is a useful tool to view OFDM transmission from a signals and systems perspective. We have three types of OFDM block-based transmission techniques, block pre-coded, zero-padded, and cyclic prefix. We adopt block pre-coded in our formulation. Block-based transmissions are effective in mitigating channel-caused ISI. In block transmission (as shown in Fig. 6), we denote block lengths Bl where Bl are formed from input symbols $s_p[n]$ and $Bl >> L_T$, L_T denotes taps in Tapped Delay Line (TDL) [4] frequency selective channel model. We denote the to-be transmit i^{th} blocks with $\boldsymbol{s}_p[i]$ where

$$\boldsymbol{s}_p[i] = [\boldsymbol{s}_p[(i-1)Bl], \boldsymbol{s}_p[(i-1)Bl+1], \cdots, \boldsymbol{s}_p[(i-1)Bl+Bl-1]]^T \tag{1}$$

[4] A TDL is a delay line that has at least one "tap" in it. A delay-line tap collects a signal output from anywhere along the delay line, scales it if necessary, and commonly sums it with other taps to generate an output signal

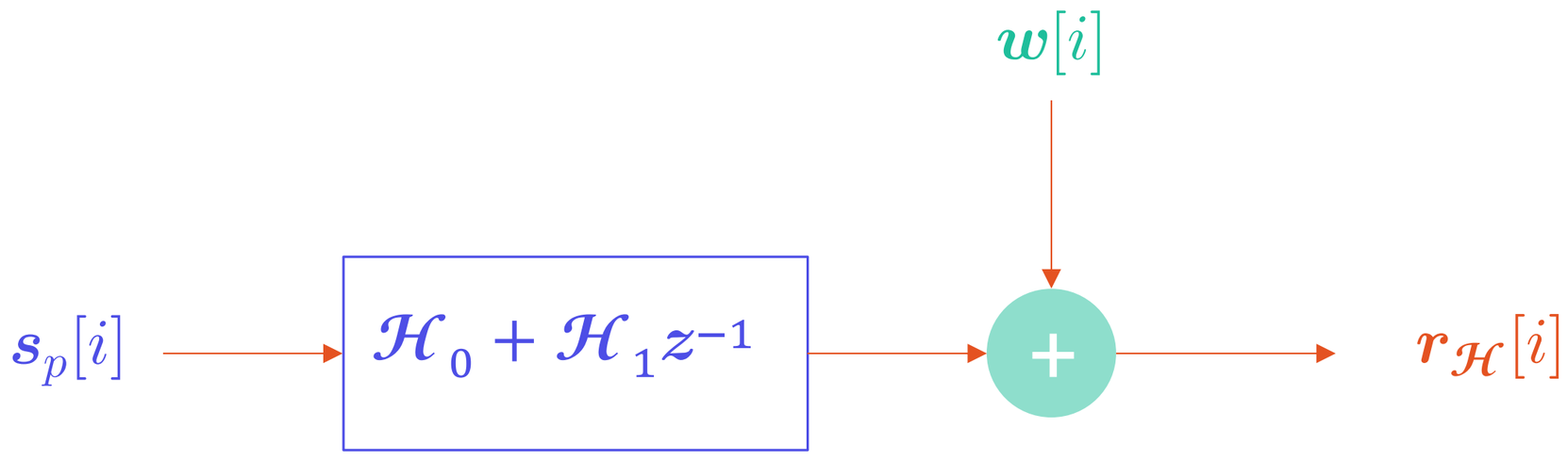


Figure 6: Block-Based transmissions

from Fig. 6 we can obtain

$$\boldsymbol{r}_{\boldsymbol{s}p}[i] = \boldsymbol{\mathcal{H}}_0 \boldsymbol{s}_p[i] + \boldsymbol{\mathcal{H}}_1 \boldsymbol{s}_p[i-1] + \boldsymbol{w}[i], \tag{2}$$

where $\boldsymbol{r}_{\boldsymbol{s}p}[i]$ and $\boldsymbol{w}[i]$ are received signal and noise (or interference) respectively. $\boldsymbol{\mathcal{H}}_0$ and $\boldsymbol{\mathcal{H}}_1$ are channel matrices of size $Bl \times Bl$

$$\boldsymbol{\mathcal{H}}_0 = \begin{bmatrix} h[0] & 0 & 0 & \cdots & 0 \\ \vdots & h[0] & 0 & \cdots & 0 \\ h[L_T-1] & \cdots & \ddots & \cdots & 0 \\ \vdots & \ddots & \cdots & \ddots & 0 \\ 0 & \cdots & h[L_T-1] & \cdots & h[0] \end{bmatrix}_{Bl \times Bl} \tag{3}$$

$$\boldsymbol{\mathcal{H}}_1 = \begin{bmatrix} 0 & \cdots & h[L_T-1] & \cdots & h[1] \\ \vdots & \ddots & 0 & \ddots & \vdots \\ 0 & \cdots & \ddots & \cdots & h[L_T-1] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \vdots & 0 \end{bmatrix}_{Bl \times Bl} \tag{4}$$

we stated that block transmission is robust against ISI, nevertheless, it introduces Inter-Block Interference (IBI) for block $\boldsymbol{s}_p[i]$ solely from the previous block $\boldsymbol{s}_p[i-1]$, for the mere fact of causality as a result of picking $Bl >> L_T$. This general framework is what we build the

OFDM discrete-time model on, as shown in Fig. 7 (please note we are only considering the OFDM Discrete-time model from an UL perspective). And by applying the same principle in Eq. 2, the received signal can be obtained as

$$\boldsymbol{r}[i] = \boldsymbol{\mathcal{H}}_0 \boldsymbol{T}_{cp} \boldsymbol{F}_N^H \boldsymbol{s}_p[i] + \boldsymbol{w}[i] \tag{5}$$

where $\boldsymbol{F}_N^H$ denotes the Hermitian of Inverse Discrete Fourier Transform (IDFT) matrix, and $\boldsymbol{T}_{cp}$ denotes the cyclic prefix inserted matrix.

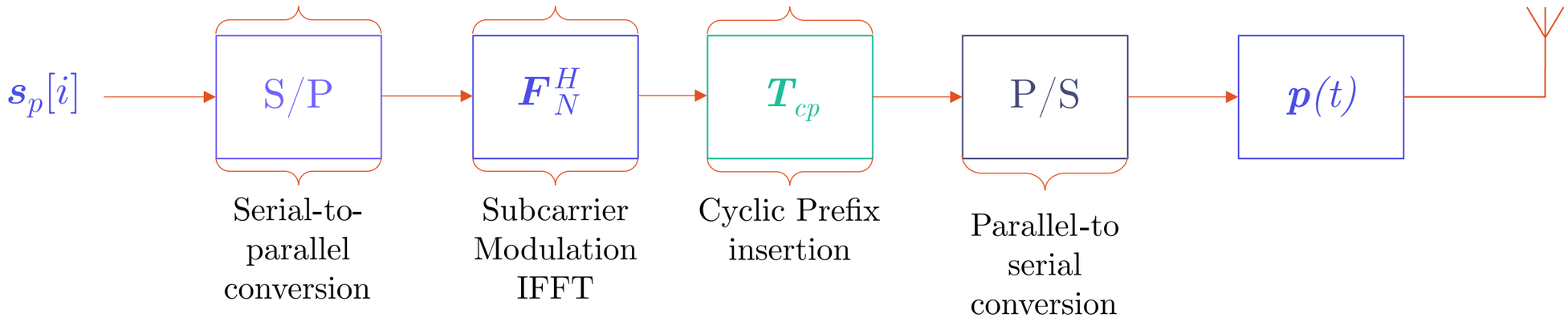


Figure 7: Discrete-Time OFDM block model

3.1.2 Multiple Access Extension–OFDMA

OFDMA was conceived to combine the use of OFDM and FDMA. The available subcarriers in OFDMA systems are separated across many mutually exclusive subcarriers (as shown in Fig. 8) that are allotted to different users for simultaneous transmission. The orthogonality of subcarriers ensures robustness against co-channel interference, while the use of a dynamic subcarrier assignment technique provides the system with improved resource management flexibility. Additionally, OFDMA retains from OFDM the capacity to adapt to the frequency domain channel distortions, without the use of complex time domain equalizers. Nonetheless, it still inherits the same drawbacks of legacy OFDM e.g., the strict requirement for very precise time and frequency synchronization.

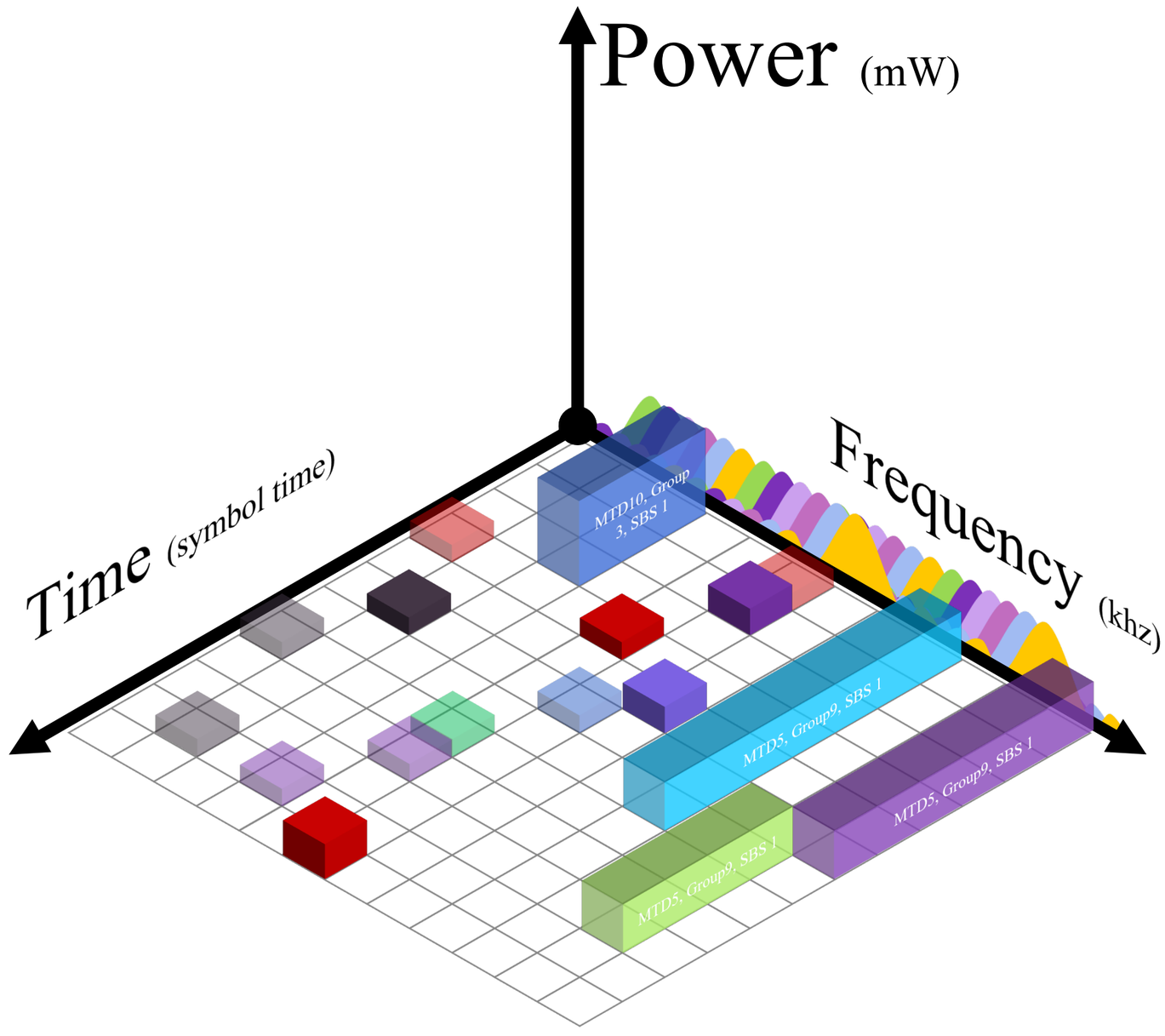


Figure 8: Illustration of OFDMA Technique

3.1.3 UL Operation Mechanisms

The primary distinction with DL transmission using OFDMA is when we consider a block-wise style of transmission, where $\mathcal{B}_m$ of Ds data symbols are passed into a carrier assignment scheme (CAS) block, and then a map of user-subcarrier is constructed. This operation results in a mapping d-dimensional vector of frequency domain data samples. Thenceforth, an IDFT operation is done on the mapping vector elements along with the addition of d-points CP. Finally, the mapping vector is converted into a block of time domain samples. We omit the analysis of synchronization schemes as we assume no frequency or time synchronization is required from the base station of the received MTDs signals.

3.2 Non-Orthogonal Multiple Access

NOMA is a relatively new multiple access technique, and the literature recently has

concluded that a standardized approach is easily formulated when dealing with different schemes of NOMA. We direct the reader to the seminal works on PD-NOMA [72] and [73–76]. PD-NOMA was presented back in the highly influential work by Saito, et al. [77] which sparked the interest in NOMA schemes. In the years following [77], there was a flux of new NOMA schemes in various domains each with different implementation purposes, improvements over existing schemes, and more generalizable characteristics. Fig. 9 lists the most commonly cited NOMA schemes proposed in the literature. Our scope remains focused on the PD-NOMA scheme so we considered only the techniques relying on SIC on the receiver side. As illustrated in Fig. 9, there are four main domains of non-orthogonality, Power Domain, Bit Level Interleaving Domain, Symbol Level Scrambling, and Spreading. We will briefly discuss some of the major techniques that assume SIC receivers as candidate receivers.

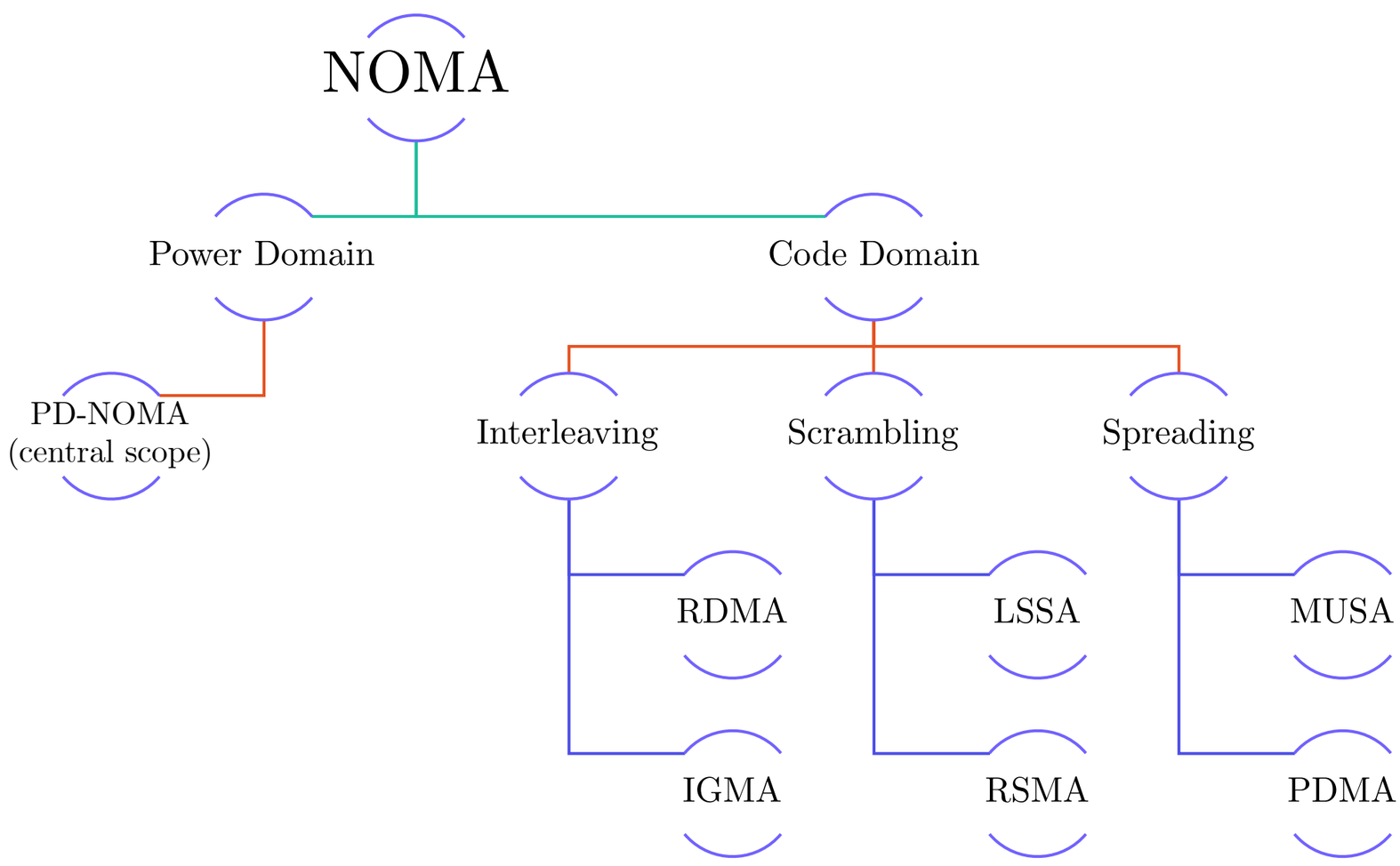


Figure 9: Breakdown of widely used NOMA techniques

3.2.1 Successive Interference Cancellation

We first need to introduce SIC as an interference mitigation technique [72, 78, 79]. To

process superposed data (e.g., power/spread domain) at a single receiver (e.g., SBS) SIC works by decoding MTDs received signals in the case of spread (code) domain. We make the assumption that the receiver has full knowledge of the spread sequence of each device, and in this case, the receiver is not concerned with their energy levels.

Decoding the most powerful MTD signal is the core concept of SIC followed by canceling the first decoded MTD signal's effect. Nevertheless, the user with the highest order is not known in advance; instead, it is determined by the strength of the correlations between each user's chip sequence and the signal received. A selector receives the correlation values and chooses the MTD to decode based on the strongest correlation. The estimation of signal power and the preservation of the cancellations order are both based on these correlative measures. The process is looped until all MTDs have been successfully decoded. The selector in our case is a covariance-based selector or more formally defined as a sampler there are various methods to sample MTDs for Transmissions, and probabilistic-based sampling methods such as simple random, stratified, and cluster-based samplings. Moreover, the literature also deploys non-probabilistic based sampling methods i.e., maximum variation, homogeneous, and Snowball samplings [80].

In the case of the power domain, it is much easier to determine the strongest device, as MTDs are arranged according to their signal strengths. We can represent this mathematically, first, we assume a n-user UL scheme where the SBS needs to decode the data received from two MTDs, in the superposition encoding part of the transmission process, we have two point-to-point encoders, $E_1 : \{0,1\}^{\lfloor 2^{LR_1} \rfloor} \rightarrow C^L$ up until $E_n : \{0,1\}^{\lfloor 2^{LR_n} \rfloor} \rightarrow C^L$ translates the input signal bits to n output coded sequences $S_1(k), ..., S_n(k)$. We denote the length of each code block by L, R represents the total throughput of each device[5]. Finally, C denotes the code library. The resulting signal is shown below

$$\mathcal{X}(k) = \sum_{i=1}^{n} \sqrt{P\delta_i} S_i(k) \tag{6}$$

[5]the floor function is represented as $\lfloor \cdot \rfloor$

δ_i denotes the total transmission power fraction of each device, where $\sum_{i=1}^{n} \delta_i = 1$. At the SBS the following operation is repeated until all devices data has been decoded (we assume additive noise is negligible and the received signal $\mathcal{Y}(k) = \mathcal{X}(k)$),

1. Decode highest ranked message using a single user decoder we denote as $F_1 : \{0,1\}^{\lfloor 2^{LR_1} \rfloor} \rightarrow C$.

2. We then deduct $\sqrt{P\delta_1}S_1(k)$ from the total received signal $\mathcal{Y}_{n-i}(k)$ hence

$$\mathcal{Y}'_{n-i}(k) = \mathcal{Y}_{n-i}(k) - \sqrt{P\delta_1}S_1(k) \tag{7}$$

3. Repeat 1 and 2 until all messages have been decoded successfully, wherein the first iteration $i = 1$.

Furthermore, there are different candidate reciever types that deploy various techniques to decode messages as well as mitigate interference some which are Message Passing Algorithm (MPA) [81], Parallel Interference Cancellation (PIC) [82], and other variants, Minimum Mean Square Error (MMSE) and Elementary Signal estimator (ESE) based SIC and PIC, we direct the reader to [74, 75] for more detailed overview on the uses of each technique. Many works [75, 83, 84] speculated the integration of massive MIMO with NOMA as an essential tool to enable the systems to exploit the benefits of both massive MIMO and NOMA. Massive MIMO-OMA systems with widespread linear processing at the BS are known to attain the highest SE in under-loaded settings, i.e., with fewer users than available antennas. As a result, such integration may be incapable of supporting huge connection in overcrowded networks where the number of MTDs vastly exceeds the number of antennas at the BS. To that aim, massive MIMO-NOMA has showed significant promise in addressing the connection requirements of overloaded systems. The high number of antennas at the BS in massive MIMO-NOMA may be leveraged to produce multiple beams for dividing users in the space domain, resulting in Spatial Division Multiple Access (SDMA). While there

is extensive research efforts on DL massive MIMO-NOMA, the UL case has received less attention. Furthermore, practically all previous work focuses on grant-based transmissions, in which user clusters and various NOMA-related transmission parameters are specified or pre-configured. As a result, the promise of massive MIMO-NOMA in providing grant-free UL transmissions has yet to be realized.

3.2.2 NOMA schemes

Power Domain

In the power domain, the most widely used type is PD-NOMA. The main principle of PD-NOMA is that we have the users partitioned physically not only in the time and frequency domain but also in the power domain as shown in Fig. 10. PD-NOMA is considered a Full collision scheme, as it allows concurrent transmission of all users using the same spectral resources. Consequently, we perform user multiplexing at the SBS in the power domain. The users are ranked based on their Channel State Information (CSI) and signal strengths; the ranking process is detailed in Algorithm 1. We can illustrate how PD-NOMA principally works in Fig. 10, as shown in the figure the users are spread over time/frequency domains, which is similar to the OFDMA scheme's RB design. The main difference is that users can transmit over the same time/frequency resources, whereas receivers rely on SIC to decode messages from each user. Moreover, we mathematically analyze intra-cluster interference solely and its effects (as we work with the assumption that there is no inter-cell interference for simplicity purposes) in the following section.

Algorithm 1: PD-NOMA Ranking scheme for n-MTD system

while True do
  Set $\boldsymbol{\mathcal{C}}$, R_m^{th}, P_m^{max}, and h_m^s
  $\forall m \in \boldsymbol{\mathcal{M}}$, $\forall b \in \boldsymbol{\mathcal{B}}$, $\forall u \in \mathcal{U}$, $\forall t \in T$, $\forall s \in \boldsymbol{\mathcal{S}}$
  Calculate Channel Covariance

  $cov_{h_m,h_n} = \frac{\sum_{i=1}^{N}(h_{m_i}-\bar{h_n})(h_{n_i}-\bar{h_n})}{N-1}$ $\forall m \in \boldsymbol{\mathcal{M}}$: $\tilde{h}_1 \geqslant \tilde{h}_2 \geqslant \cdots \geqslant \tilde{h}_M$
  Using k-means algorithm Group MTD based on zero covariance as a feature.
  for all $u \in \mathcal{U}$ **do**
    if $\mathcal{U}_{\text{size}} < \mathcal{C}_{\text{size}}$ **then**
      Assign mMTC $\{1, ..., (\mathcal{C} - \mathcal{U})\}$ to the lowest order $(u = 1)$ of $\{(\mathcal{U}+1), ..., \mathcal{C}\}$ clusters.
    else
      Assign mMTC $\{1, ..., (\mathcal{C} - \mathcal{U})\}$ to the next available order in the cluster. $\forall \boldsymbol{\mathcal{C}}$
    end if
  end for
end while=0

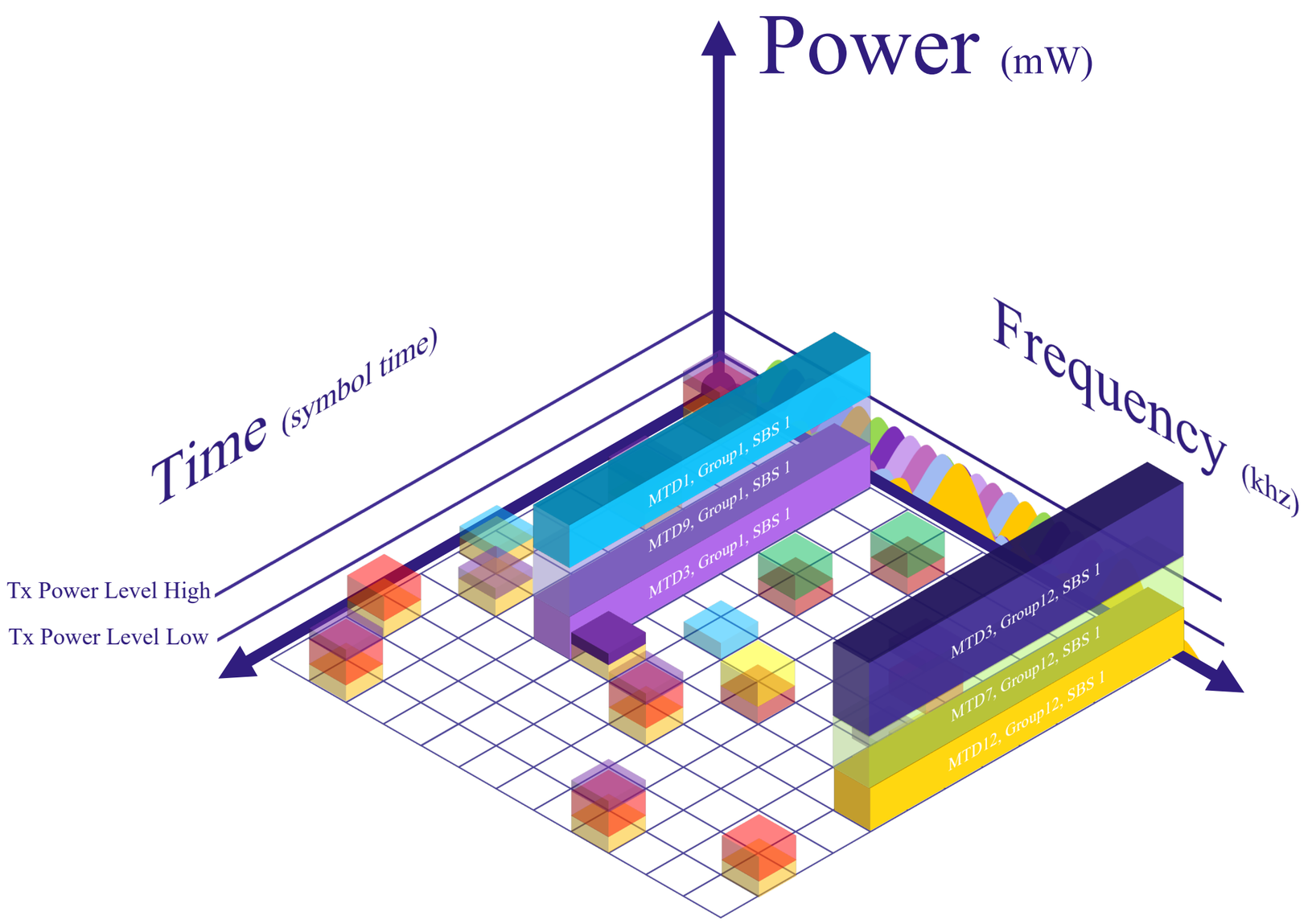


Figure 10: Illustration of PD-NOMA Technique

Code Domain

Similar to CDMA, different users are issued distinct codes and then multiplexed across the same time-frequency resources in code domain multiplexing. The distinction between power domain and code domain multiplexing is that code domain multiplexing can provide specific spreading and shaping gains at the expense of greater signal bandwidth [73]. In this subsection, we will discuss some techniques used in the code domain that would give us a glimpse into where academia and the industry are headed, as some of the techniques were proposed by industry (e.g., MediaTek, 3GPP) or in academia.

Spreading Consider standard SCMA [85] as an example. Developed by utilizing LDS-CDMA, each user's original bit stream is directly translated to a code-word and unique codebook. SCMA code-words are few, with just a few entries being non-zero. The main advantage SCMA has over traditional LDS-CDMA is that it uses multidimensional constellations. All SCMA code-words have a distinct position of non-zero elements. Fig. 11 depicts an example of SCMA resource mapping using four users, four resources, and a sparsity of two (each user transmits data over 2 out of 4 resources). The maximum combinations possible is given by $CB = \binom{N}{K}$, where CB is a binomial coefficient. **Pattern Division MA (PDMA)** [86] was designed based on SCMA the main distinctions are that PDMA has variable controllability of the sparsity of its spreading sequence and that it can be spread over numerous domains (power, code) hence it is not domain restrictive as SCMA. Fig. 12 depicts PDMA resources spreading over four users sharing four Resource elements (RE), where a binary code matrix 8 below

$$\boldsymbol{CM}^{[4,4]}_{\text{Code}} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{bmatrix} \tag{8}$$

defines a sparse spread map from load to a RE group, where each element maps to a specific group, and a '1' maps a RE to load data. A Belief Propagation (BP) [87] algorithm is used at the receiver to do MUD and decode the received signal messages.

Because lengthy spreading sequences employed in classical CDMA have very low cross-correlation, combining these codes with SIC results in increased processing complexity, latency, and error propagation at the receiver. As a result, short spread codes with minimal cross-correlation are ideal for grant-free UL Multi-User Shared Access (MUSA) [75]. The family of complicated spreading codes is an appropriate choice in this situation since it is short, owing to the design freedom with real and imaginary components. These spreading sequences are specifically intended to handle significant user overloading and to enable a simpler processing SIC design at the receiver.

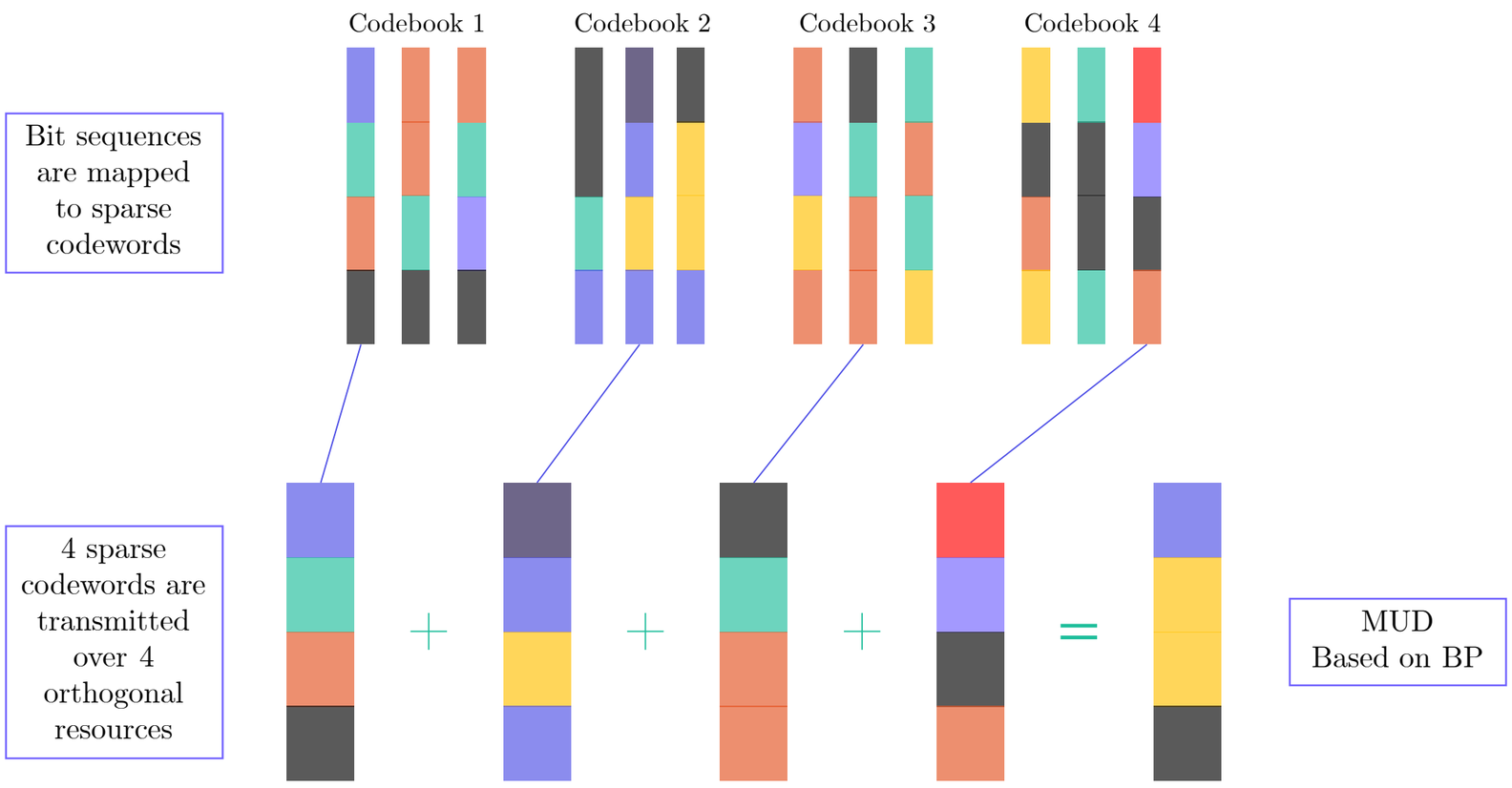


Figure 11: SCMA scheme: Resource spreading across 4 MTDs, 4 subcarriers

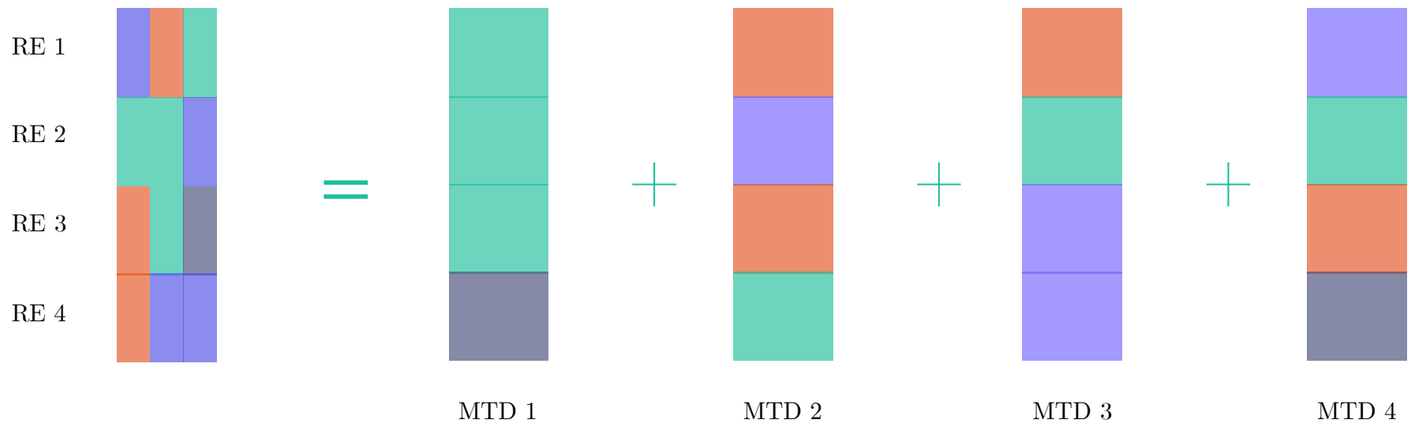


Figure 12: PDMA where REs are spread across n-users

3.3 A Primer on the Mathematics of SBSs and MTDs Configurations

Previous work on Point Processes (PP) [88–90] investigated random groupings of point occurrences where the points are located along time or space axes. However, in this subsection, we only pivot on 2D spatial processes as this is the main building block of our work, Nevertheless, we touch on the fundamentals of PP in the time axis as well.

3.3.1 Fundamentals of Point Process

A point process is a fundamentally stochastic process and has been studied throughout the Renaissance up until the early twentieth century through Life Tables and Counting Problems. We will briefly touch on each concept in the temporal domain, then we will move on to define point processes in higher dimensions (spatial).

Life Tables

Life Tables were first used to describe the sequential occurrences of successive events (replacements) of an object that is prone to failure and can be replaced by another object at each failure. The above definition formed the basis for Renewal Theory which will be discussed later in this section. A Life Table is essentially a list of all people who survive beyond a certain age in a given population, and are usually of size 1000 for proportional purposes. The most significant variables are l_x the number of people who live to age x, d_x, those who died between age x and $x+1$. Hence $d_x = l_x - l_{x+1}$, and finally q_x those who die before reaching age $x+1$, where $q_x = \frac{d_x}{l_x}$. Life Tables were constructed for discrete ages. However, to transition into how the Poisson process is formulated from this scope we will define it in continuous time and replace the ages with probabilities for each individual.

Survivor function

$$l_x = S(x) = Pr\{\text{lifetime} > x\} \tag{9}$$

Lifetime distribution function

$$f(x)\ d_x = Pr\{\text{life ends between } x\ and\ x + dx\}, \tag{10}$$

Hazard function

$$q(x)\ d_x = Pr\{\text{life ends between } \ x\ and\ x + dx \mid \text{it doesn't not end before } x\}. \tag{11}$$

Following the relations between these functions detailed above we arrive at a mathematical conclusion that the Life Table problem distribution function is an exponential function, where the hazard is a constant independent of age for $x > 0$, hence we get

$$f(x)\ dx = \lambda e^{-\lambda x},\ q(x) = \lambda,\ S(x) = e^{-\lambda x},\ F(x) = 1 - e^{-\lambda x}. \tag{12}$$

the above distribution could be realized in different context models as emphasized in [89], that lifetime distribution is another form of gamma, Weibull, and log-normal distributions.

Counting Problems

Counting Problems is an alternative and more standard approach to defining the point process paradigm in single dimension as well as higher dimensions. By definition, it is used to count the number of occurrences in-between intervals (or higher-dimensional regions) of time or space. The point process can be formulated as a counting process [90]. First, we will consider a one-dimensional point process where the occurrence times $T_1 < T_2 < ...$, this is illustrated in Fig. 13. However, studying these events proved to be complicated due to the occurrence times being strongly dependent since $T_i < T_{i+1}$.

We can instead study the inter-occurrence times $O_i = T_{i+1} - T_i$ as they have the

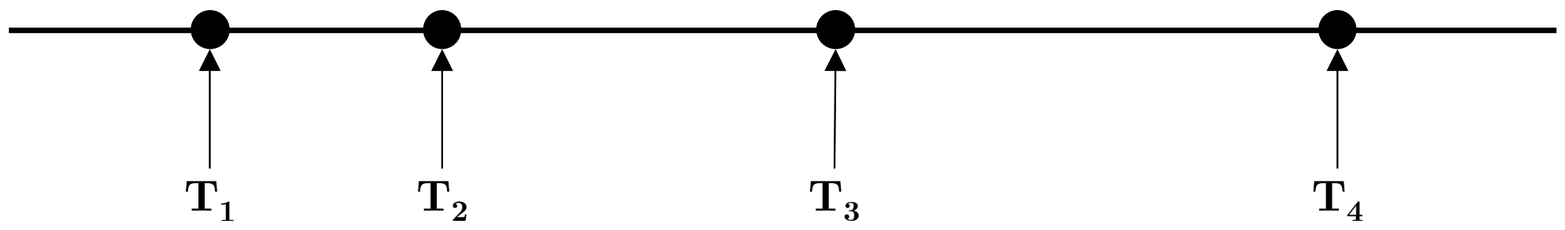


Figure 13: Occurrence times T_i

advantage of being independent as shown in Fig. 14. It is more common to define the point process as a counting process as shown below

$$N_t = \sum_{i=1}^{\infty} \mathbb{I}\{T_i \leq t\}, \tag{13}$$

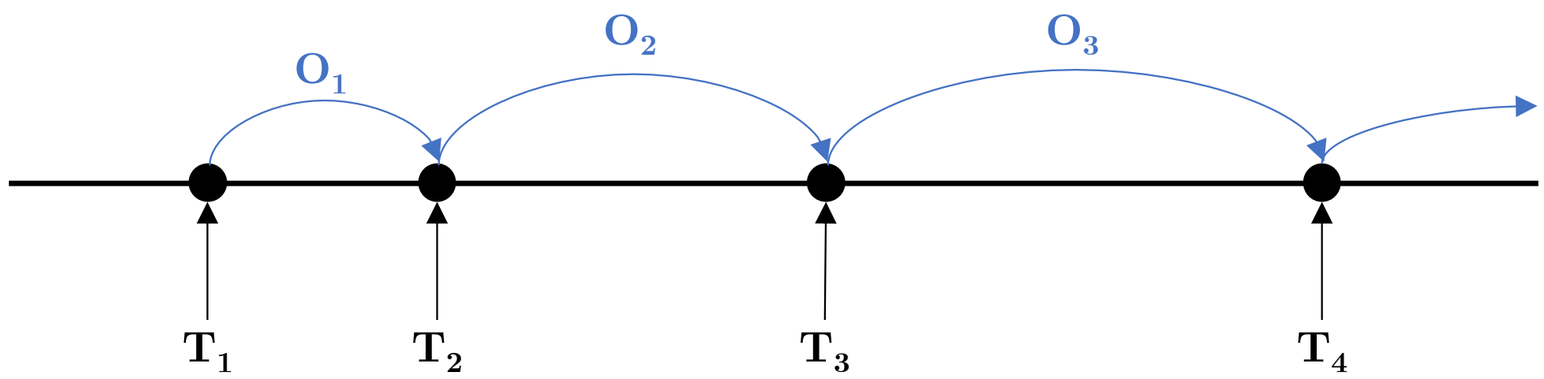


Figure 14: Inter-Occurrence times O_i

where N_t denotes the number of occurrences up until time t, and $\mathbb{I}$ is a binary indicator function that a certain statement (e.g., MTDs transmitting) of occurrence at time t is true or false. As stated above we know that the number of occurrence times are highly dependent. Consequently, we use interval counts $N(a, b] = N_b - N_a$ as they are disjoint in nature and are stochastically independent. This approach is more analogous while working in higher dimensions, as points in a 2D space cannot be ordered naturally, hence it would be more convenient to generalize Inter-occurrence interval counts to region counts. **Binomial Distribution as an alternative**, various problems have presented with a behavior that is not satisfactory by Poisson distribution characteristics (where the variance/mean ratio is identically unity), it has been discovered in ecology [91, 92] among other fields, that the distribution of inter-interval counts had a higher variance for a given mean. As an alternative, negative binomial distribution has been widely adopted [89]. To elaborate further, let us

generate a number n of points in the bounded region $W \subset \mathbb{R}^2$ at random locations. We then denote $X_1, ..., X_n$ to be Independent and Identically Distributed (i.i.d) uniformly distributed random points in W. The probability density of each X_i is

$$g(x) = \begin{cases} \frac{1}{\lambda_2(W)} & \text{if } x \in W \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

$\lambda_2(W)$ is denoted as the area of the region W, since the generated random points are uniformly distributed, therefore for any bounded set of points B in $\mathbb{R}^2$

$$Pr\{X_i \in B\} = \frac{\lambda_2(B \cap W)}{\lambda_2(W)}, \tag{15}$$

we denote the set of points that falls within the bounded region W as $N(B)$ and the points that do not as $V(B)$, both sets can be represented in terms of bounded set B as

$$\begin{aligned} N(B) &= \sum_{i=1}^{n} \mathbb{I}\{X_i \in B\} \\ V(B) &= \min_{i=1}^{n} \mathbb{I}\{X_i \notin B\}, \end{aligned} \tag{16}$$

Naturally $N(B)$ follows a binomial distribution with parameters n and $p = \lambda_2(B \cap W)/\lambda_2(W)$, therefore it is widely defined as a binomial process.

3.3.2 Basics of Poisson Process

Stationary One-Dimensional PPP

We define the PPP on a line in which $N(a_i, b_i]$ denotes the number of occurrences of a process falling in a half-closed interval $(a_i, b_i]$ where $a_i < b_i \leq a_{i+1}$, by the following equation

$$Pr\{N(a_i, b_i] = n_i, i = 1, ..., k\} = \prod_{i=1}^{k} \frac{[\lambda(b_i - a_i)]^{n_i}}{n_i!} e^{-\lambda(b_i - a_i)}, \tag{17}$$

The above definition has three distinct characteristics:

1. The numbers of generated points in all half-closed intervals are Poisson distributed.

2. We can say that all the number of points generated in discontinuous intervals are independent Random Variables (RVs).

3. Finally the distributions of the Intervals are declared to be stationary.

Stationarity refers to the fact that the distribution of points within a finite interval depends not on the location but on the interval's length.

Definition 1. *A point process is deemed stationary if, for any shifting vector* $v \in \mathbb{R}^d$ *the distribution of the shifted point process with vector* v *remains equivalent to the distribution of the original point process.*

Following the characteristics of the Poisson process from Eq. 17 we know that $\mu(a, b] = \lambda(b - a) = \sigma(a, b]$ where μ is the mean density of the points and σ is the variance. We can naturally observe that the mean and variance are equivalent and that they are proportional to the size of the half-closed interval, which is a further evidence of the stationarity of the Poisson process. Moreover, we can define the Poisson distribution as

$$Pr\{N = k\} = e^{-\mu}\frac{\mu^k}{k!}, k = 0, 1, 2, ... \tag{18}$$

Furthermore, if we split the interval $(a_i, b_i]$ into a large number n of smaller intervals, the number of inter-occurrence in each small interval is binary, except for an occurrence with a low probability. Since $N(a_i, b_i]$ is the summation of these occurrences, it follows a binomial distribution. As $n \to \infty$, we conclude that $N(a_i, b_i]$ have a Poisson distribution. There are other properties that can be drawn from the conclusion above specifically for the one-dimensional PPP,

1. The inter-occurrence times O_i are exponentially distributed with intensity β.

$$Pr\{O_i \leq o\} = 1 - e^{-\beta o}, o > 0. \tag{19}$$

2. The inter-occurrence times are independent (remember this property from the formulation of Life Tables).

These two extra properties, may give us a leeway into generating sequences of highly independent, exponentially distributed RVs $O_1, O_2, ..., O_j$, where the occurrence times are defined as $T_i = \sum_{1 \leq j \leq i} O_j$. Next we review **in-homogeneous PPP** in which the estimate of number of occurrences $(a_i, b_i]$ is

$$\mathbb{E}N(a, b] = \int_a^b \beta(t)dt, \beta(t) > 0 \tag{20}$$

Where $\beta(t) > 0$ denotes the intensity function or the probability that there will be a point within minuscule inter-occurrence interval $[t, t + dt]$. Hence, occurrences in discontinuous time intervals are independent in nature. **Conditional Property** Given a PPP defined in $\mathbb{R}^2$ that has a uniform intensity of $\beta > 0$, and that Region $W \subset \mathbb{R}^2$ that has surface area $0 < \lambda_2(W) < \infty$. $N(B)$ where $B \subseteq W$ is considered to have a binomial distribution

$$Pr\big(N(B) = k | N(W) = n\big) = \binom{n}{k} p^k (1 - p)^{n-k} \tag{21}$$

The proof can be found in section 1.5 of [90] and is omitted from this work.

Spatial (2D) PPP

It is possible to generalize the PPP to $\mathbb{R}^2$ as defined below.

Definition 2. *Poisson Process is defined in two-dimensional space* $\mathbb{R}^2$ *with uniform intensity* $\beta > 0$ *with properties such that*

Property 1. $\forall$ *divided regions of space* $B_1, ..., B_j$, *then* $N(B_1), ..., N(B_j)$ *are independent.*

Property 2. *Second $\forall$ sets B which are closed and bounded sets, $N(B)$ (count of sets) is Poisson distributed with $\mu = \beta\lambda_2(B)$.*

Remember that the binomial process is a Poisson process in disguise, the realization of the binomial process in Fig. 15 gives us a rather interesting disparity between the two processes, the fact that the generated points in a unit square region of space W are precisely 100 points for the binomial process while different realizations of the Poisson process could produce a different number of points.

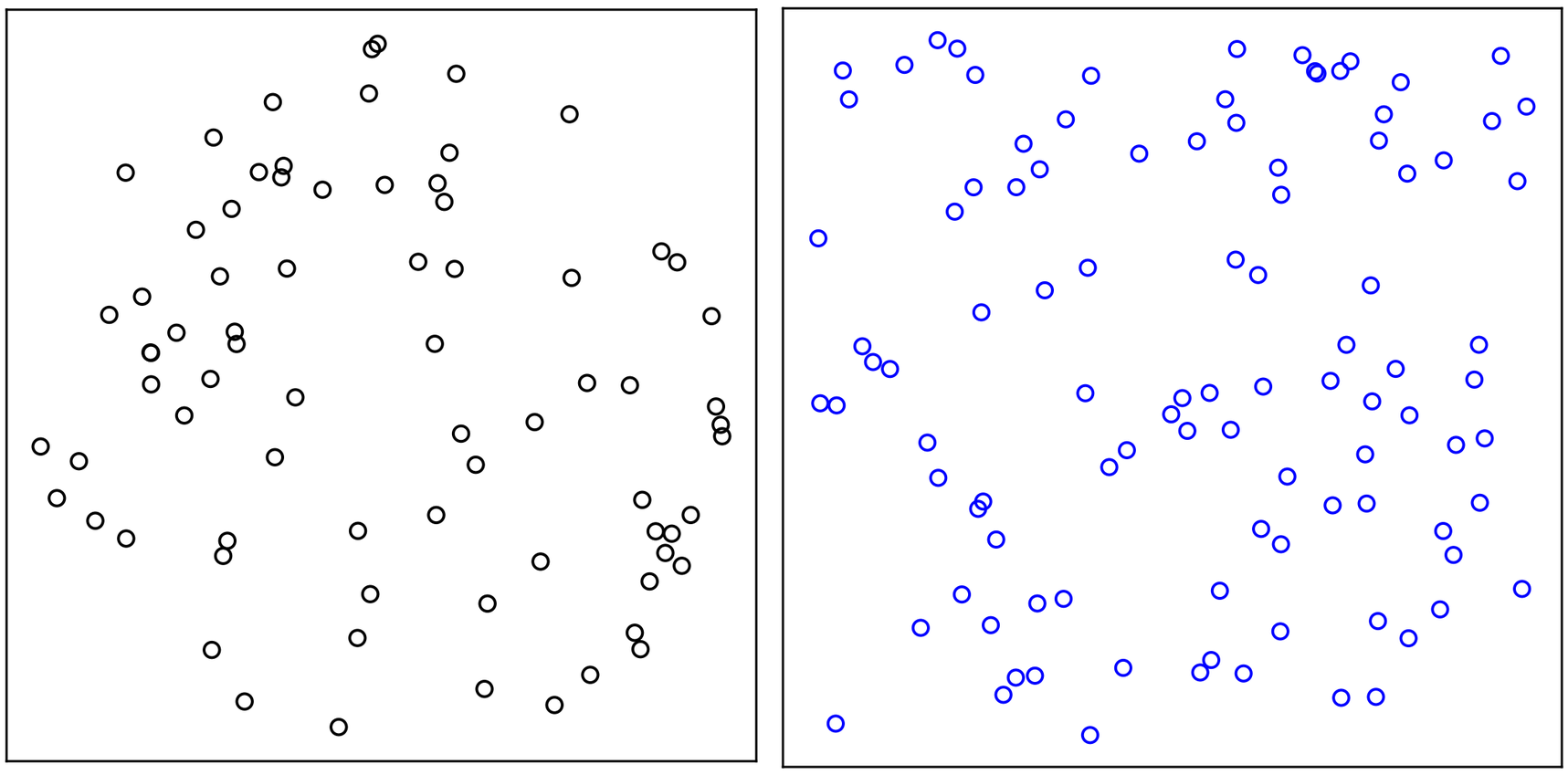


Figure 15: Realization of Poisson (left) Binomial (right) point processes $\beta = 100$

General Poisson Process

In this subsection we set the assumptions on which we define a general Poisson point process, first, we assume the process points are generated in higher dimensional spaces (Euclidean) $\mathbb{R}^d$ we shall denote any higher dimensional space as S, we denote Λ as a measure on S (Borel σ-field of disjoint compact spaces) and $N(A)$ as count of points in the space S. Given that there exists a bounded finite Λ, therefore we define the general Poisson process as

Definition 3. *A general PPP is defined on S with intensity Λ such that*

Property 1. *$\forall$ compact disjoint Borel sets $N(A)$ is Poisson distributed with a mean of $\lambda(A)$.*

Property 2. *$\forall$ Borel sets (that are bounded and disjoint), the counts of points of these sets $N(A_1), ..., N(A_i)$ are independent*

An example of such generalization, let there be a S sphere with a unit radius, our intensity measure Λ would be the area of the unit sphere in 3D thus $\Lambda = \beta\mu$ where μ is the area measure on the unit sphere 4π and β as always is the mean density. We can realize this generalization as shown in Fig. 16. However, while generating sphere bounded PPP, some considerations need to be taken into account while working with spherical coordinates (ρ, θ, Φ). Transitioning to the Cartesian coordinate system we have

$$\begin{aligned} x &= \rho\sin(\theta)\cos(\Phi) \\ y &= \rho\sin(\theta)\sin(\Phi) \\ z &= \rho\cos(\Phi) \end{aligned} \tag{22}$$

It ultimately depends on the purpose of generalization, it could be on the surface of the sphere, or inside the sphere. Both cases differ in the RV that needs to be uniformly distributed. In the First case, we use **Proposition 1.1** from [93], we simulate two new uniform RVs namely Ξ and Θ, where $-1 \leq \Xi \leq 1$ and $0 \leq \Theta \leq 2\pi$. Then the new Cartesian coordinates are

$$\begin{aligned} X &= r\sqrt{1-\Xi^2}\cos(\Theta) \\ Y &= r\sqrt{1-\Xi^2}\sin(\Theta) \\ Z &= r\Xi. \end{aligned} \tag{23}$$

where the radius of the sphere is denoted as r, this method works to ensure that the points generated are uniformly distributed over the surface. To elaborate, consider the RV $\Phi = \arccos(\Xi)$ is the Φ-coordinate of uniform point, by substituting with the trigonometric identity [6] where $\sin(\Phi) = \sqrt{1-\Xi^2}$. Recall that the surface area element under polar

[6] $\cos(\arcsin(x)) = \sin(\arccos(x)) = \sqrt{1-x^2}$

coordinates is $dA = \rho^2 \sin(\Phi) d\Phi d\theta$, integrating with respect to θ we get a constant, once we integrate with respect to Φ we get $\Xi = \cos(\Phi)$, thus instead of ensuring that Φ is uniformly distributed we need Ξ to be so. Please note to generate the process inside the sphere you only need Φ to be uniformly distributed.

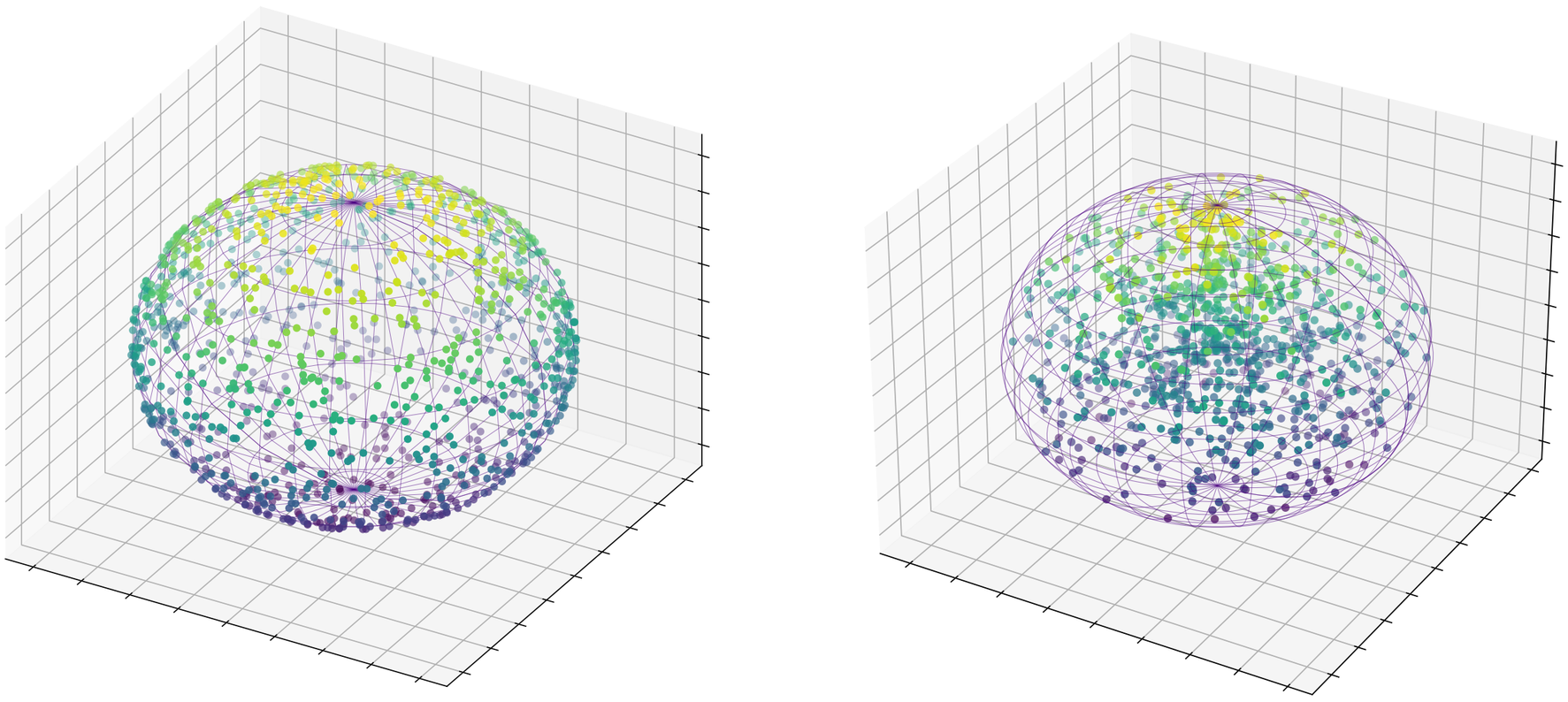


Figure 16: Realization of a homogeneous PPP orthogonal projected over a unit sphere $\beta = 1000$, on the sphere (left), inside the sphere (right)

3.3.3 Model Construction through Conditioning

PPP Transformations

There are numerous methods through which we can construct new point processes, for example, by changing or transforming certain characteristics or properties of an existing point process. We will focus on **Superpositioning** and **Clustering** which will generally lead to the introduction of the Marked Process. **Superpositioning** is defined by the union of two independent point processes X and Y sets of points, we denote the count of X process by $N_X(W)$ and of Y with $N_Y(W)$ where the counts of both processes lie in the region $W \subset \mathbb{R}^2$. Hence we can refer to the superposition of X and Y as the sum of the random measures of both processes, the realization is shown in Fig. 17. Moreover, if both

processes are independent, then we can say that the superpositioning of two uniform Poisson processes with mean densities μ and υ, is a uniform Poisson process with a mean density of the summation of both densities μ and υ.

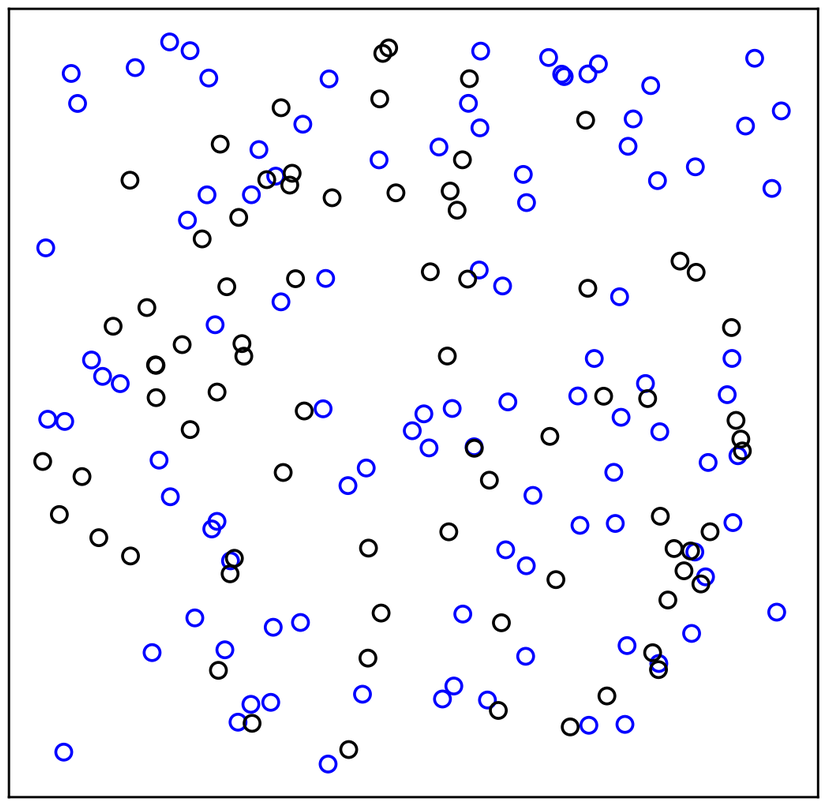


Figure 17: Superposition Realization of two PPPs

Cluster Process

From the work in [91] we will adopt some of the authors' mathematical findings into our elaboration on Matérn Cluster Process. As well as the works [49, 94–96] which are more aligned with our scope. Clustering processes have been an indispensable type to model the locations of objects or more precisely the occurrences of locations in spatial domain meaning $\mathbb{R}^d$ where $d \geq 2$. Cluster processes have been adopted in a wide range of applications, in ecology, stars formation, and departures in communication queuing systems. We informally define the cluster point process as a point process of cluster centers (cc) or Parent points and each cc is correlated with a random number of points with count $N(A)$, creating together secondary process points (daughters). The daughter points are spread around the cc in a defined fashion. The clustering process is, therefore, considered a superposition of all subsidiary daughter clusters. The clustering process then involves superimposing all of the various clusters, with points from the same cluster not being identified as such.

Definition 4. *N(A) is a Poisson cluster process on a separable spatial region W, with the cc process denoted as* $N_c(A)$ *on another separable space* E *and subsidiary point processes (measurable group)* $\{N(\cdot|e) : e \in E\}$*, where for every finite bounded region* $A \in \mathcal{Q}_W$*,*

$$N(A) = \int_E N(A|e)N_c(de) = \sum_{e_i \in N_c(\cdot)} N(A|e_i) < \infty \ a.s. \tag{24}$$

As mentioned the definition entails that the superposition of all clusters almost surely is finitely bounded. Although this is not necessarily the case for each of the clusters as they are not constrained by such requirements. We bypass the need to prove the mutual independency of the subsidiary (daughter) processes as it was already proved for Lemma 6.3.II in [89], and we assume the necessity of mutual independence of daughter processes to define the independent Poisson cluster process. Hence, by calculating the conditional expected value on the cc we get

$$\mathrm{E}[N(A)|N_c] = \sum_{e_i \in N_c} M_1(A|e_i) = \int_E M_1(A|e)N_c(de) \tag{25}$$

we denote $M_1(\cdot|e_i)$ as the mean of the cluster's daughter process centred at e. And since by definition we know that the daughter processes form a measurable group, then we can find the expected value of the measurable kernel (if $M_1(\cdot|e)$ exists) taking into account the cc process we find

$$\mathrm{E}[N(A)] = \int_E M_1(A|e)M^c(de), \tag{26}$$

we denote M^c as the expectation of the random measure of the process of cc. From the expectation above we can conclude that the first moment of a resulting superposition process exists $\iff$ the integral is finite $\forall$ Borel sets A. We need to point out that this result holds for higher-order moment measures. Moreover, if we consider the second-moment measure of the resultant process, we have to examine if two different daughter points from the superposition of clusters could fall into the product of the two sets A and B where $(A, B \in \mathcal{B}_x)$, we can

obtain the second moment of the resultant process

$$M_{[2]}(A \times B) = \int_E M_{[2]}(A \times B|e)M^c(de) + \int_E M_1(A|e_1)M_1(A|e_2)M^c_{[2]}(de_1 \times de_2) \tag{27}$$

One widely known example of cluster processes is **Matérn Cluster Process** in which parent points are uniformly distributed over a two-dimensional region $\mathbb{R}^2$ we can define the Joint cc process and subsidiary processes in simpler terms as detailed in Eq. 28, along with a realization of Matérn cluster process shown in Fig. 19. Applying principles introduced in this chapter we can generalize the Poisson process on a sphere to a 3D shape for instance. Fig. 18 illustrates a Matérn cluster process where the parent points engulfed n-uniformly distributed points inside a spherical region of space, this generalization could be used in, for example, users placement in a 3D environment as well as UAV positioning in regions of space to maximize coverage, the potential is this generalization using some of the techniques described above could be case-specific, for example, we can generalize it to different shapes with a uniform cross-section.

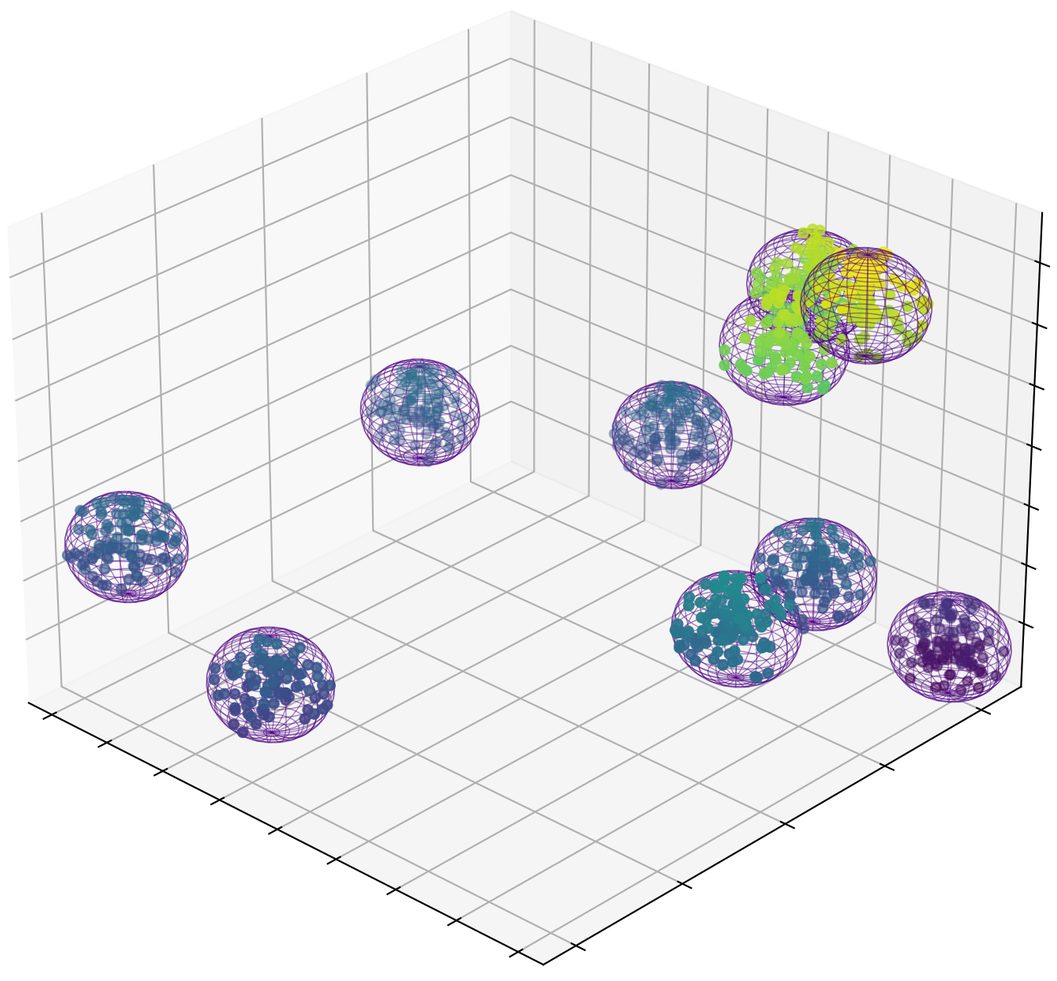


Figure 18: Realization of 3D Spherical Matérn cluster process $\beta = 100$/sphere

Chapter 4

Dynamic RRS Problem Formulation

4.1 System Model

We investigate a multi-cell scenario with one SBS in each cell center (cec) that supports mMTC and is following NB-IoT standard specifications. We assume MTDs are stationary or with very limited mobility. Denote $\boldsymbol{\mathcal{M}} = \{1, ..., M\}$ as the set of MTDs. For uplink data transfer in a single Transmission Time Interval (TTI), where active devices share a single PRB every TTI. Each PRB has a bandwidth that is partitioned into a series of sub-channels $\boldsymbol{\mathcal{S}} = \{1, ..., S\}$, where a bandwidth of W will be allocated to each sub-channel. To solve our formulated OPs We rely on Gurobi [97] for both the heuristic solution and the near-optimal solution.

We propose HetNet based Poisson cluster process (PCP) scheme by clustering MTDs in an NB-IoT network, where clusters are generated in a simulation environment according to Matérn cluster process [49], where the daughter points are distributed uniformly within a circle centered around a parent point as shown in Fig. 19. For simplicity, we shall denote the Matérn cluster process as PCP. The PCP process could be defined as

$$\rho_u = \bigcup_{\Theta \in \rho_{\mathrm{p}u}} \Theta + O_u^{\Theta}, \tag{28}$$

where ρ_{p_u} denotes the parent PPP of density λ_{p_u} and O_u^{Θ} represents the daughter point process, where each point at $s \in O_u^{\Theta}$ is i.i.d. around the cc $\Theta \in \rho_{\mathrm{p}_u}$, where the density of daughter point process is $\mathcal{D}_u$.

The MTDs (according to the PCP scheme), share each sub-channel resources and transceive non-orthogonally i.e., multiple MTDs can share the same sub-channel resources or in an orthogonal manner in case they are accessing the medium using OFDMA. As a result, the devices are dispersed into PCP groups, which we denote as $\boldsymbol{\mathcal{C}} = \{1, ..., C\}$ cluster sets. $\gamma_{s,c,b}$

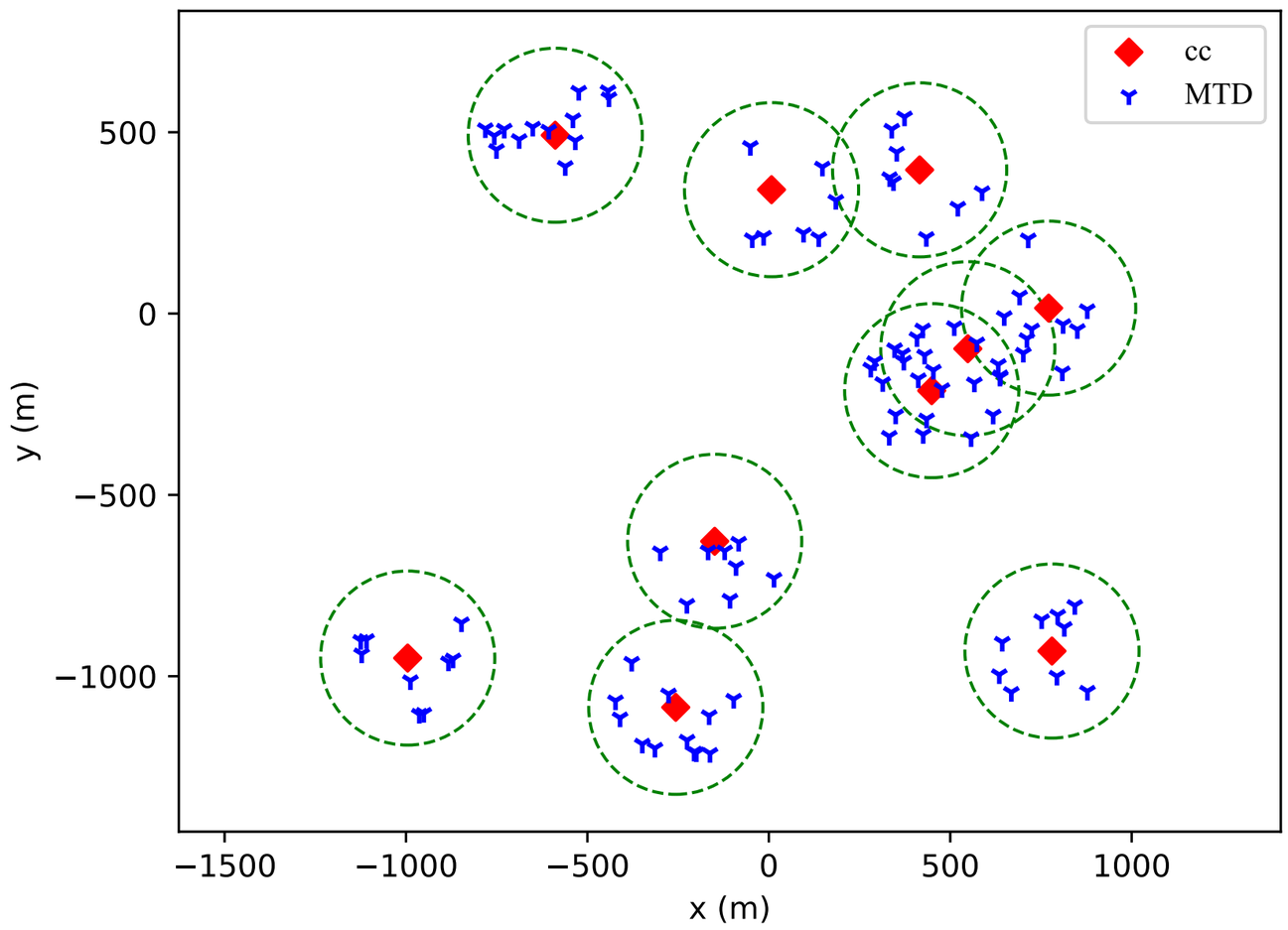


Figure 19: Simulation of a Matérn cluster process

is a binary indicator where sub-channel $s \in \boldsymbol{\mathcal{S}}$ is dedicated to group $c \in \boldsymbol{\mathcal{C}}$, hence $\gamma_{s,c,b} = 1$ if sub-channel s is allotted and zero otherwise. The MTDs use the same sub-channel for transmission, with transmitting powers of $p_{s,m}$, respectively. As a result, the SBS receives a combined message from MTDs with additive noise $\mathcal{N}_0$ and interference in case the MTDs are using PD-NOMA. We denote the point process of the b^{th} SBS tier as ρ_u, where ρ_u is the PCP ($\forall u \in \mathcal{U}_1$), where $\mathcal{U}_1$ is the index set of the SBS tiers being modeled as PCP. Defining the set of ranks (levels) in each PCP group as $\mathcal{U} = \{1, ..., u_{max}\}$, where u_{max} defines the maximum number of MTDs that can be in the same group and thereby utilize the allotted sub-channels We assume that $\boldsymbol{\mathcal{C}}_{\text{Size}} \times u_{max}$ is greater than the population of MTDs. It is worth noting that the MTD with the highest rank in each group is immune to other MTDs' interference, while the other MTDs are subjected to interference from MTDs with higher ranks ($u = 2, ..., u_{\max}$). Furthermore, transmit power and QoS requirements are taken into account while analyzing intra-group interference. PCP clustering uses the average channel gain of MTDs, $\tilde{h}_m = \sum_{s \in \boldsymbol{\mathcal{S}}} \frac{h_{m,s}}{\boldsymbol{\mathcal{S}}}$. Then based on their $\tilde{h}_m$, they are assigned to the group with the next highest group rank.

To simplify what we mean by clustering process, and to be methodically accurate. The term clustering applies not only to how we group random sets of objects located in constrained space as shown in Fig. 20 (based on their proximity for example).

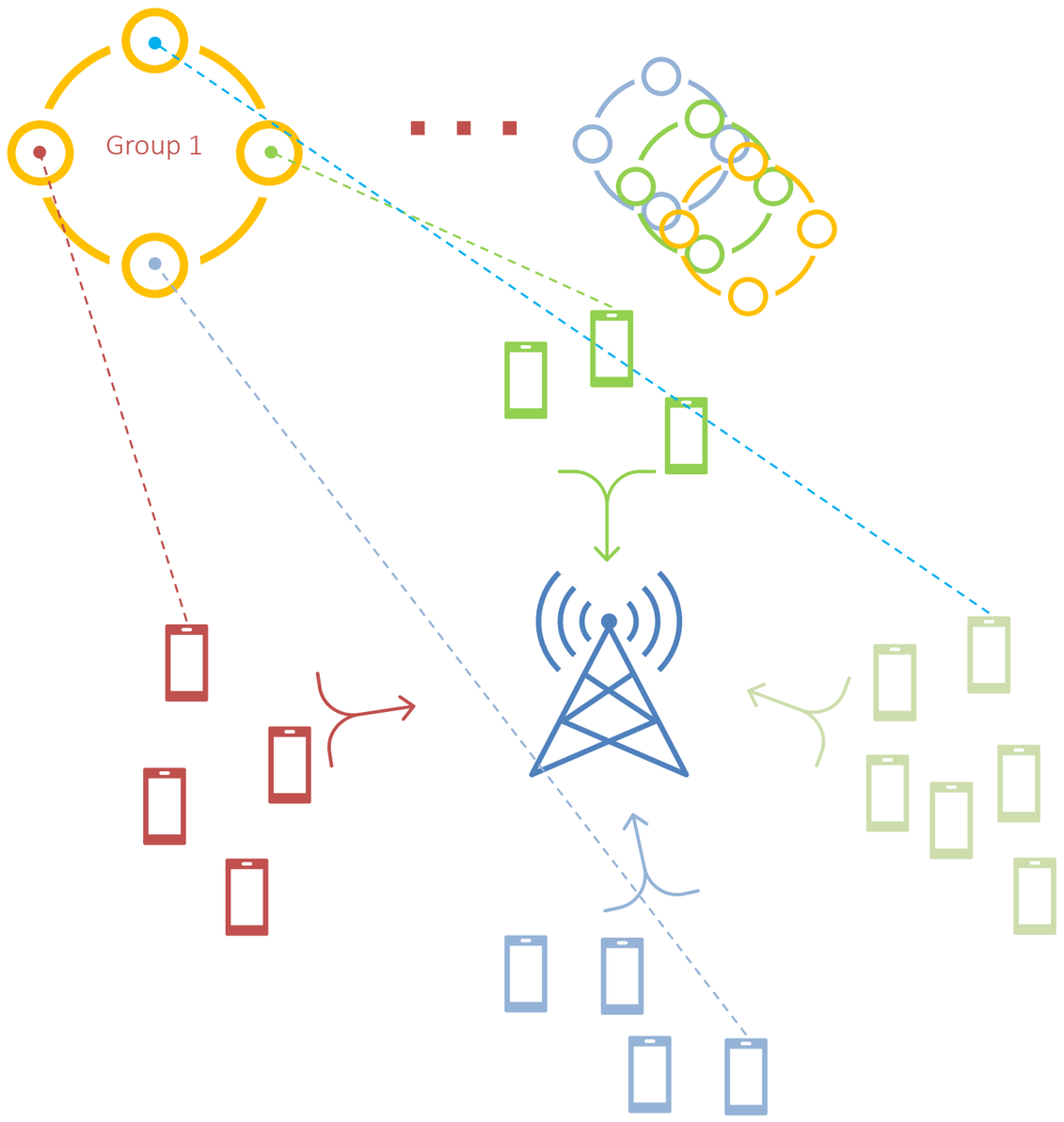


Figure 20: Grouping MTDs based on their covariance

But also to higher levels of correlation in certain metrics. Hence, clustering in our study could mean two things, either how we initialize the location of MTDs based on PCP models as mentioned above. The other definition would be how to cluster MTDs for transmission. For example, suppose we have a group of MTDs that want to transmit on s sub-channel of $\boldsymbol{\mathcal{C}}_{\text{PD-NOMA}}$, the MTDs then will be clustered in groups that show the least correlation in their CSI. Henceforth, the MTDs with the least similar CSI estimates will be clustered together, which will usually translate to different geographical positions. As demonstrated in [98], for MTDs' positioning, the CSI signature should be both generally consistent in the same area and differentiable at various locations.

4.2 Quality of Service Constraints

We start by defining $p_{m,s}$ as the m^{th} MTDs' transmit power over the s^{th} sub-channel, and $\alpha_{m,c,u,b}$ as the binary variable used to allocate the m^{th} MTD to the u^{th} rank of group c. If there is scheduled, $\alpha_{m,c,u,b} = 1$, and zero otherwise. As a result, the MTDs' achievable data rate $\boldsymbol{R}$, The aggregate rate over the designated sub-channels for PD-NOMA groups' overall SBSs are shown in Eq.29. Consequently, with the assumption of no interference in OFDMA groups due to the nature of the access technique, which is robust against co-channel interference. The aggregate rate of all OFDMA groups for all SBSs is shown below in Eq. 30

$$\begin{aligned} \boldsymbol{R}_{\text{PD-NOMA}} = \sum_{b \in \boldsymbol{\mathcal{B}}} \boldsymbol{R}_b = \sum_{c \in \atop \boldsymbol{\mathcal{C}}_{\text{PD-NOMA}}} \sum_{u \in \mathcal{U}} \alpha_{m,c,u,b} \\ \sum_{s \in \boldsymbol{\mathcal{S}}} \gamma_{s,c,b} W_1 \log_2 \left(1 + \frac{|h_{m,s}|^2 p_{m,s}}{\boldsymbol{\mathcal{N}}_0 W_1 + \sum_{h=u+1}^{u_{\max}} \alpha_d^{c,h} h_{d,s}^5 p_d^s} \right) \end{aligned} \tag{29}$$

$$\begin{aligned} \boldsymbol{R}_{\text{OFDMA}} = \sum_{b \in \mathcal{B}} \mathbf{R}_b = \sum_{c \in \boldsymbol{\mathcal{C}}_{\text{OFDMA}}} \alpha_{m,c,b} \\ \sum_{s \in \boldsymbol{\mathcal{S}}} \gamma_{s,c,b} W_2 \log_2 \left(1 + \frac{|h_{m,s}|^2 p_{m,s}}{\boldsymbol{\mathcal{N}}_0 W_2} \right) \end{aligned} \tag{30}$$

where $h_{m,s}$ is the channel gain on sub-channel s between the m_{th} MTDs and the SBS and $\mathcal{N}_0$ is the noise power spectral density. The m^{th} MTD experiences interference only from other higher-ranked MTDs in the same group using the sub-channel.

To ensure that we have a higher data rate than $\boldsymbol{R}_m^{th}$'s minimum data rate, the following constraint is required

$$R_m^b \geqslant R_m^{th}, \quad \forall m \in \boldsymbol{\mathcal{M}} \ \& \ \forall b \in \boldsymbol{\mathcal{B}}. \tag{31}$$

Furthermore, to discern between the minimum required data rates for devices accessing the

medium either using PD-NOMA or OFDMA, the following constraints are considered to make sure the QoS requirements of either are met.

$$\begin{aligned} R_m{}^{\text{PD-NOMA}} &\geqslant R_m^{\text{PD-NOMA}_{th}}, \quad \forall m \in \boldsymbol{\mathcal{M}}, \\ R_m^{\text{OFDMA}} &\geqslant R_m^{\text{OFDMA}_{th}}, \quad \forall m \in \boldsymbol{\mathcal{M}}, \end{aligned} \tag{32}$$

The m^{th} MTDs' overall transmit power is limited by its maximum power budget R_m^{th}, i.e.,

$$p_{m,s} \leqslant P_m^{\max}, \quad \forall m \in \boldsymbol{\mathcal{M}}. \tag{33}$$

4.3 Optimization Problem Formulation

The clustering-based OP for NB-IoT is described in this section as a sum rate maximization problem of MTDs. In addition to the QoS requirements in Eq. (31), and Eq. (33), we should impose additional limitations on the PD-NOMA clustering process. Each MTD, in particular, should be assigned to only one group with a single rank, i.e.,

$$\sum_{c \in \boldsymbol{\mathcal{C}}} \sum_{u \in \mathcal{U}} \alpha_{m,c,u,b} = 1, \quad \forall m \in \boldsymbol{\mathcal{M}}, \tag{34}$$

Because the purpose of PD-NOMA is to share spectral resources among multiple MTDs, the PD-NOMA grouping enforces the presence of more than one MTD in each group, i.e., the same applies to OFDMA groups as we consider inter-cell interference, but in this equation, we chose to disregard inter-cell interference

$$\sum_{m \in \boldsymbol{\mathcal{M}}} \sum_{u \in \mathcal{U}} \alpha_{m,c,u,b} \geqslant 2, \quad \forall c \in \boldsymbol{\mathcal{C}}_{\text{PD-NOMA}}. \tag{35}$$

$$\sum_{m \in \boldsymbol{\mathcal{M}}} \sum_{u \in \mathcal{U}} \alpha_{m,c,u,b} \geqslant 2, \quad \forall c \in \boldsymbol{\mathcal{C}}_{\text{OFDMA}}. \tag{36}$$

and we guarantee the priority of rank assignment in each group by starting with the lowest

rank in each group ($u = 1$), i.e.,

$$\alpha_{m,c,u,b} \leqslant \alpha_{m,c,u-1},\ \forall m \in \boldsymbol{\mathcal{M}},\ \forall c \in \boldsymbol{\mathcal{C}},\ 2 \leqslant u \leqslant u_{\max}, \tag{37}$$

Finally, the sumrate maximization of MTDs transceiving over OFDMA and PD-NOMA channels can be formulated as joint scheduling and sub-channel allocation multi-objective OP below

$$\underset{p_{m,s},\, \alpha_{m,c,u,b},\, \gamma_{s,c,b}}{\text{maximize}} \quad (\boldsymbol{R}_N, \boldsymbol{R}_O) \tag{38a}$$

subject to

$$R_m \geqslant R_m^{th},\ \ \forall m \in \boldsymbol{\mathcal{M}}, \tag{38b}$$

$$R_m^{\text{PD-NOMA}} \geqslant R_m^{\text{PD-NOMA}_{th}},\ \ \forall m \in \boldsymbol{\mathcal{M}}, \tag{38c}$$

$$R_m^{\text{OFDMA}} \geqslant R_m^{\text{OFDMA}_{th}},\ \ \forall m \in \boldsymbol{\mathcal{M}}, \tag{38d}$$

$$p_{m,s} \leqslant P_m^{\max},\ \ \forall m \in \boldsymbol{\mathcal{M}}, \tag{38e}$$

$$p_{m,s} > 0,\ \ \forall m \in \boldsymbol{\mathcal{M}},\ \ \forall s \in \boldsymbol{\mathcal{S}}, \tag{38f}$$

$$\alpha_{m,c,u,b} \leqslant \alpha_{m,c,u-1},\ \forall m \in \boldsymbol{\mathcal{M}},\ \forall c \in \boldsymbol{\mathcal{C}},\ 2 \leqslant u \leqslant u_{\max}, \tag{38g}$$

$$\sum_{c \in \boldsymbol{\mathcal{C}}} \sum_{u \in \mathcal{U}} \alpha_{m,c,u,b} = 1,\ \ \forall m \in \boldsymbol{\mathcal{M}}, \tag{38h}$$

$$\sum_{m \in \boldsymbol{\mathcal{M}}} \sum_{u \in \mathcal{U}} \alpha_{m,c,u,b} \geqslant 2,\ \ \forall c \in \boldsymbol{\mathcal{C}}_{\text{PD-NOMA}}, \tag{38i}$$

$$\sum_{m \in \boldsymbol{\mathcal{M}}} \sum_{u \in \mathcal{U}} \alpha_{m,c,u,b} \geqslant 2,\ \ \forall c \in \boldsymbol{\mathcal{C}}_{\text{OFDMA}}, \tag{38j}$$

$$\sum_{s \in \boldsymbol{\mathcal{S}}} \sum_{c \in \boldsymbol{\mathcal{C}}} \gamma_{s,c,b} \boldsymbol{\mathcal{W}}_{s,c,b} \leqslant W^{RB},\ \ \forall c \in \boldsymbol{\mathcal{C}},\ \ \forall s \in \boldsymbol{\mathcal{S}}, \tag{38k}$$

$$\gamma_{s,c,b} \in \{0, 1\},\ \ \forall c \in \boldsymbol{\mathcal{C}},\ \ \forall s \in \boldsymbol{\mathcal{S}}, \tag{38l}$$

$$\alpha_{m,c,u,b} \in \{0, 1\},\ \ \forall m \in \boldsymbol{\mathcal{M}},\ \ \forall c \in \boldsymbol{\mathcal{C}},\ \ \forall u \in \mathcal{U} \tag{38m}$$

where the constraints are considered as follows, (38b) requires MTDs to have data rates

greater than a data rate threshold requirement. (38c) requires MTDs using PD-NOMA to have a data rate greater than the minimum threshold PD-NOMA data rate. The same applies for (38d) which instead specifies a minimum data rate requirement for MTDs using OFDMA for transmissions.

The m^{th} MTDs total transmit power is restricted to the max power allocation by (38e), P_m^{max} available. (38f) restricts MTDs transmit powers to positive values. (38g) suggest that MTDs can be assigned to the u^{th} rank of the c^{th} group if all lower ranks have been assigned to other MTDs. (38h) is designed to guarantee that each device (MTD) is allocated to just one group and one rank within that group. (38i) and (38j) are to ensure that there is more than one member in each PD-NOMA and OFDMA group. (38k) ensures that the total bandwidth allotted to all PCP groups does not exceed one RB (the bandwidth of one RB in NB-IoT is 180 kHz). (38l) and (38m) verify that the variables $\gamma_{s,c,b}$, $\alpha_{m,c,u,b}$, and are only permitted to have binary values.

4.3.1 NP-Hardness of OP(38)

The general Eq. (38) is an NP-Hard Problem. The OP(38) scheduling part is identical to the Makespan job shop scheduling problem where we want to execute n jobs (data transmission $\alpha_{m,c,u,b}$) by i identical ($\boldsymbol{\mathcal{S}}_b^c$) (sub-channels), which is by definition a combinatorial NP-Hard problem [99].

Similarly, the resource allocation part is analogous to a Multi-Choice Multiple Knapsack Problem (MCMKP), as we are attempting to fit $\boldsymbol{\mathcal{M}}$ items into $\boldsymbol{\mathcal{S}}_c^b$ knapsacks (sub-channels), which is also an NP-Hard problem.

4.3.2 Solution to OP(38)

WLOG, we relax the power constraints for both PD-NOMA and OFDMA MTDs. We assume the MTDs transmit using a predetermined powers $p_{s,m} = P_{s,m}^{max} - \frac{u_{m,c,b}}{P_{s,m}^{step}}$, thus the power-related constraints can be relaxed. In order to simplify the problem further, we apply

linear scalarization, which is a priori method, where solving the single-objective OP formed from a multi-objective problem means that the optimal solutions of the single OP are Pareto optimal solutions to the multi-objective OP, the Pareto optimality could be attained by adjusting the weight of scalarization ω_N [100]. Finally, we propose a heuristic solution in the form of a Distributed alternating Convex Optimization Problem (DACOP) with two secondary optimization problems. Where the goal for the the first SOP39 is to maximize $\boldsymbol{R}$ subject to scheduling constraints of $\alpha_{m,c,u,b}$, which will feed its optimal scheduling to SOP40 whose goal is to maximize $\boldsymbol{R}$ subject to sub-channel $\gamma_{m,s,u,b}$ allocation constraints. We denote the resource distribution with ω_N, Therefore, SOP1 is formulated as follows:

$$\underset{\alpha_{m,c,u,b}}{\text{maximize}} \quad \sum_{i=1}^{k} (\omega_{N_i} \boldsymbol{R}_{N_i} + (1 - \omega_{N_i}) \boldsymbol{R}_{O_i}) \tag{39a}$$

$$\text{subject to} \quad 38b - 38d, 38g - 38j, 38m, \tag{39b}$$

$$\omega_N \in (0, 1], \forall \omega_N \in \Omega, \tag{39c}$$

(39c) bounds the weight of allocated resources of $\boldsymbol{R}_{\text{PD-NOMA}}$ over $\boldsymbol{R}_{\text{OFDMA}}$ and vice versa, ω_N, therefore, is between 0 and 1. Respectively, SOP2 is formulated as follows,

$$\underset{\gamma_{s,c,b}}{\text{maximize}} \quad \sum_{i=1}^{k} (\omega_{N_i} \boldsymbol{R}_{N_i} + (1 - \omega_{N_i}) \boldsymbol{R}_{O_i}) \tag{40a}$$

$$\text{subject to} \quad \sum_{s \in \boldsymbol{\mathcal{S}}} \gamma_{s,c,b} = 1, \quad \forall c \in \boldsymbol{\mathcal{C}}, \tag{40b}$$

$$38k, 38l, \tag{40c}$$

$$\omega_N \in (0, 1], \forall \omega_N \in \Omega, \tag{40d}$$

where each sub-channel cannot be assigned to more than one group, according to (40b). The formulated OP1 is an NP-Hard non-convex mixed integer non-linear program (MINLP), which is combinatorial in nature, accordingly, the heuristic solution (DACOP) proposed above can be used to solve OP(38). DACOP is first initialized by feeding the scheduling

OP with a randomized resource allocation $\gamma_{s,c,p}$ matrix, and then it optimizes for the best scheduling of all connected MTDs which is fed to the sub-channel allocation OP and so on until it converges to the most optimal scheduling and resource allocation matrices. The convergence of the DACOP is achieved on average in three alternating cycles.

Chapter 5

Simulation Results and Discussion

In this section, we present our simulation results the results accuracy is validated using Monte Carlo simulations with up 10^3 iterations for each scenario with 5×10^4 different scenarios. Henceforth, we will discuss which hyper-parameters of the simulation had the most consequential impact on the sumrate maximization. To construct the Pareto dominance rank plot, the main parameter to tune is ω_N (which represents, in our case, the resource allocations given to MTDs using OFDMA or PD-NOMA). The effect of changing ω_N results

Table 1: Simulation Hyper-parameters

Definition	**Value**
Parent points density	$\{ \lambda_{\mathrm{p}_u} \colon 10 \leqslant \lambda_{\mathrm{p}_u} \leqslant 200\}$
The cluster radius of Parent point process	$\{ r_c \colon 1000 \text{ m} \geqslant \mathrm{r}_c \geqslant 200 \text{ m}\}$
Daughter points density	$\{ \mathcal{D}_u : 100/\mathrm{km}^2 \leqslant \mathrm{D}_u \leqslant 20000/\mathrm{km}^2\}$
Path Loss Model (Macro-cell propagation model)	$L = 128.1 + 37.6 \log_{10}(d)$, d in km
Transmit power step	$P_{s,m}^{step} = 10000$
The maximum power threshold of the m^{th} MTD	$P_m^{max} = 0.1\ mW$
The bandwidth of a single sub-channel in one PRB	$W = 15\ kHz$
The MTD antenna gain of the m^{th} MTD over the s^{th} sub-channel	$\{G_{s,m} \colon 0 \text{ dB} \leqslant \mathrm{G}_{s,m} \leqslant 10 \text{ dB}\}$
The noise figure of the m^{th} MTD over the s^{th} sub-channel	$NF_{m,s} = 9\ dB$
Sample size (per experiment step)	$10 \leqslant \mathrm{n} \leqslant 1000$

in a Pareto frontier for each scenario, from which we can draw the following conclusions. The maximum achievable sumrate (where PD-NOMA MTDs are given the same resources as OFDMA MTDs) occurs at $\omega_N = 0.5$, as shown in Fig. 21 (100 MTDs/group). In the same figure, the total sumrate of both PD-NOMA and OFDMA curves is shown, which clearly translates the effect of varying ω_N over the total sumrate of all devices in that particular scenario. Furthermore, using a Monte Carlo method, we arrive empirically at the conclusion that $\omega_N = 0.566$. Therefore, PD-NOMA MTDs' sumrate will cross with the OFDM's at a much higher ω_N as we increase the group size.

Subsequently, we can observe in PD-NOMA vs. OFDMA Pareto dominance rank plots as shown in Fig. 24 we can see that as we reduce the interference level, we achieve a much larger capacity region. The above conclusion remains consistent even with different-sized groups. Similarly, Fig. 23, showcases that as we decrease the cluster radius (from 1000

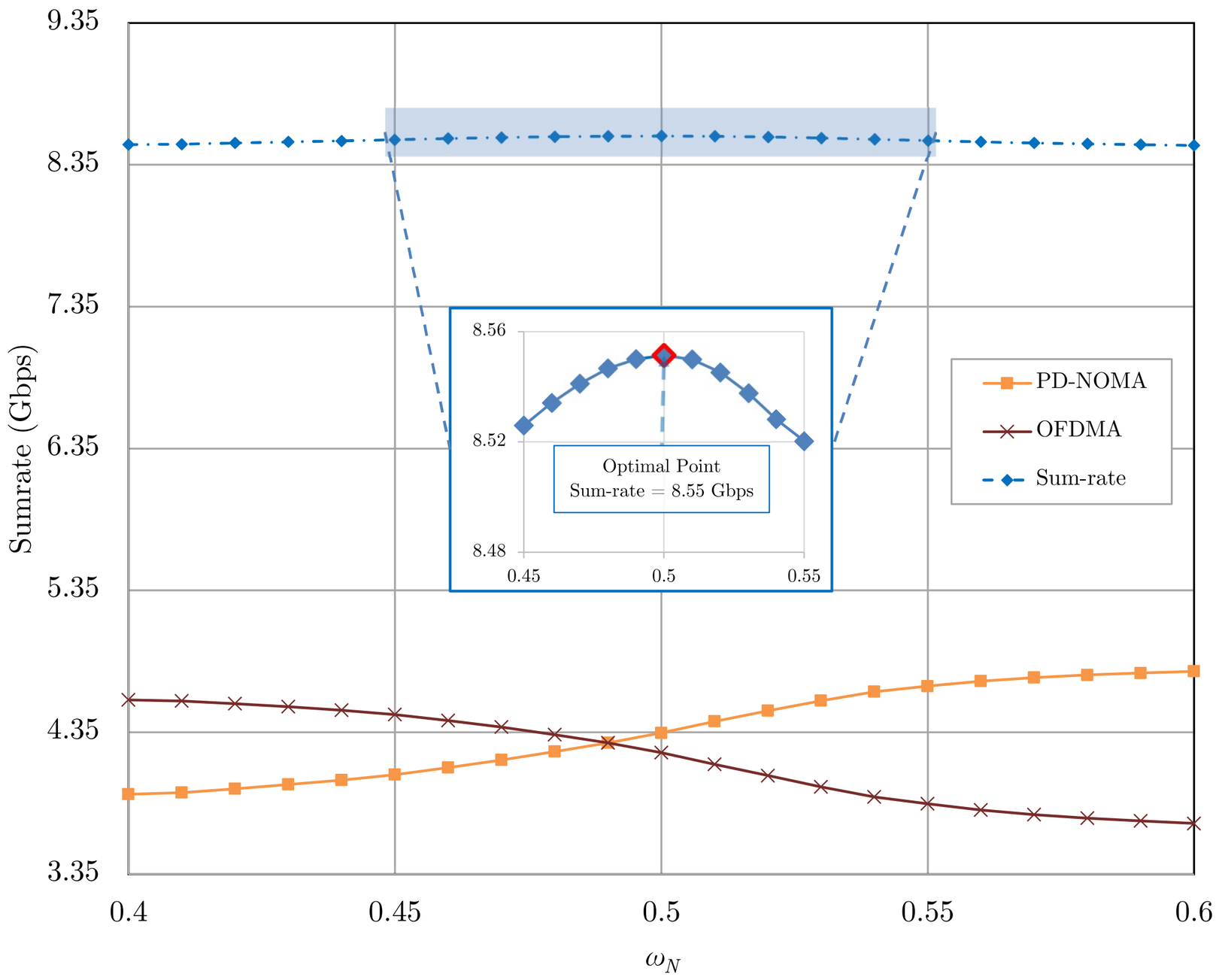


Figure 21: sumrate performance for variable ω_N (PD-NOMA & OFDMA sumrate vs Total sumrate)

m to 200 m), as illustrated in Fig. 22, we achieve a much higher sumrate for both PD-NOMA and OFDMA.

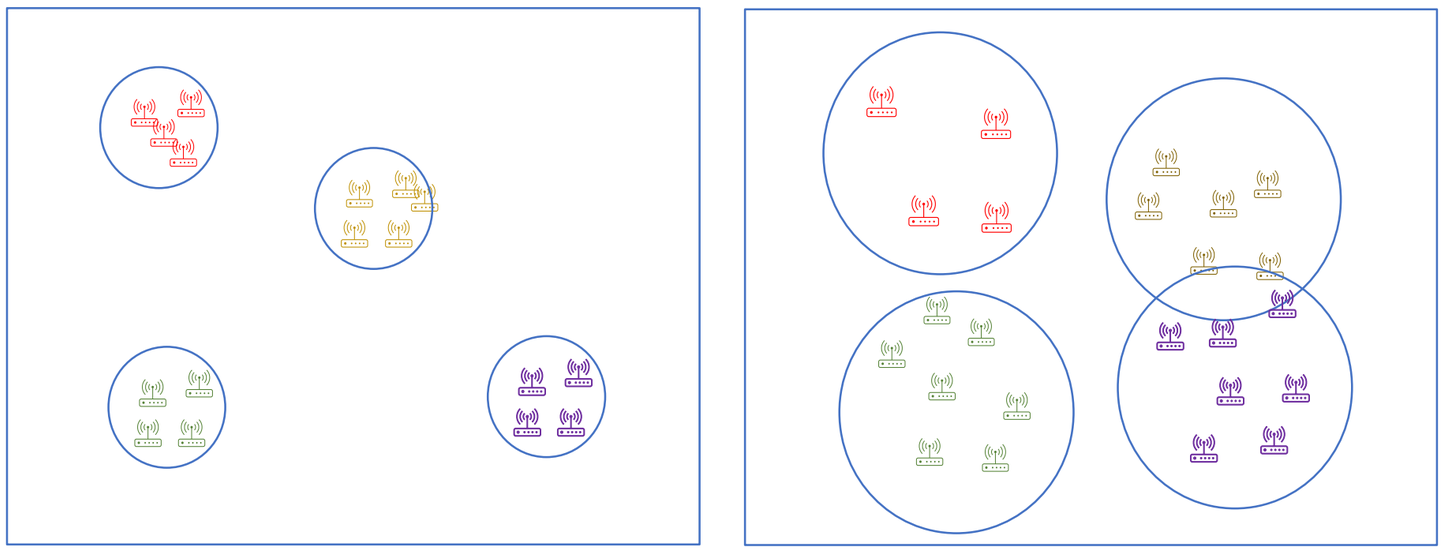


Figure 22: Varying cluster Radius

The above-mentioned observation is enriched as we look at the capacity region plots of

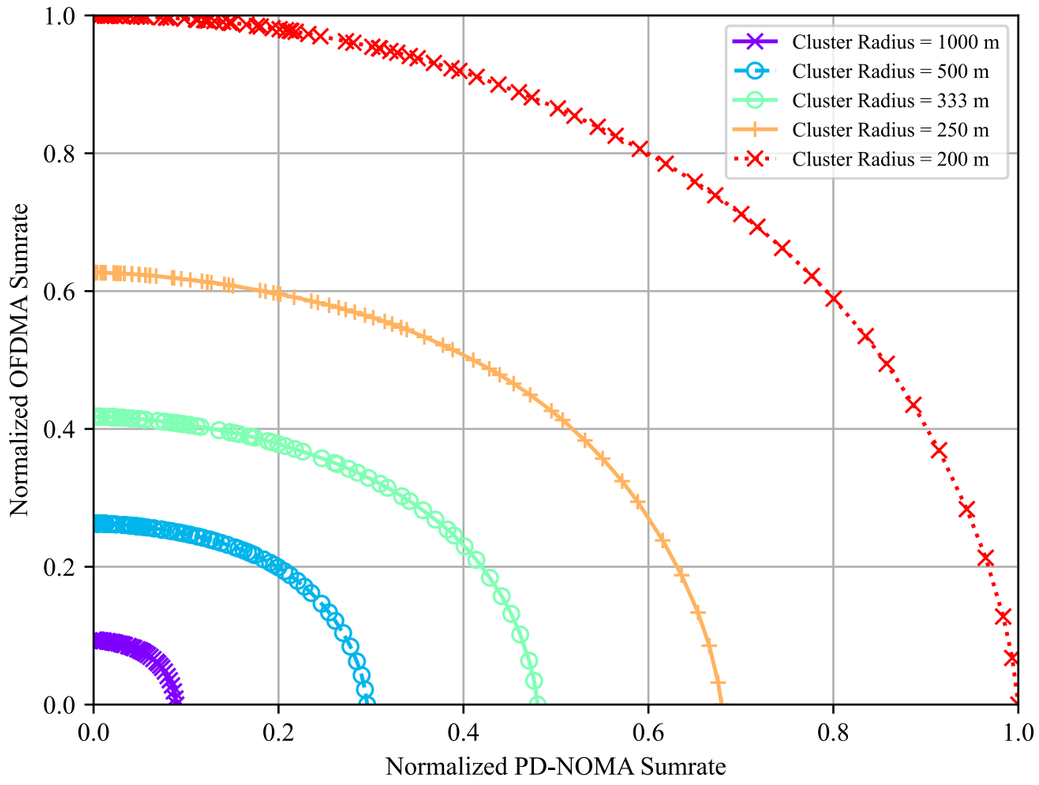


Figure 23: Relaxed MIP Pareto dominance rank plot (Normalized cluster radius effect on total sumrate)

PD-NOMA vs. OFDMA, where we can observe a clear trend toward reducing the generated MTDs' cluster radius. This result is due to how the algorithm is designed. The way it works, as we increase the cluster radius, MTDs are still going to be generated in a limited scaled space, meaning that their channel correlation will get higher as we increase the cluster radius, and even if we decide to pick the users with the least correlation in the same group, we will observe that this decision, which is one of the prime reasons for improving the overall spectral efficiency of the MTDs in the same group, is minimal. We think this result (slight improvement in sumrate) is important in determining where our logic should shift. We believe that as we increase the cluster radius, we need to re-cluster MTDs again based on their generated distributions. One easy way is to use the k-means clustering technique to regroup users based on their new surroundings, not their generated positions. In Fig. 25, we can still observe some of the drawn conclusions with the relaxed MIP. We can equivalently observe that as we increase cluster radius, the maximum achievable sumrate decreases. Furthermore, we can observe a great disparity in the number of scheduled MTDs between both access techniques, This difference in the scheduled MTDs is due to the fact that, as we increase the number of MTDs in a group the intra-group interference increases. Hence, most of the MTDs

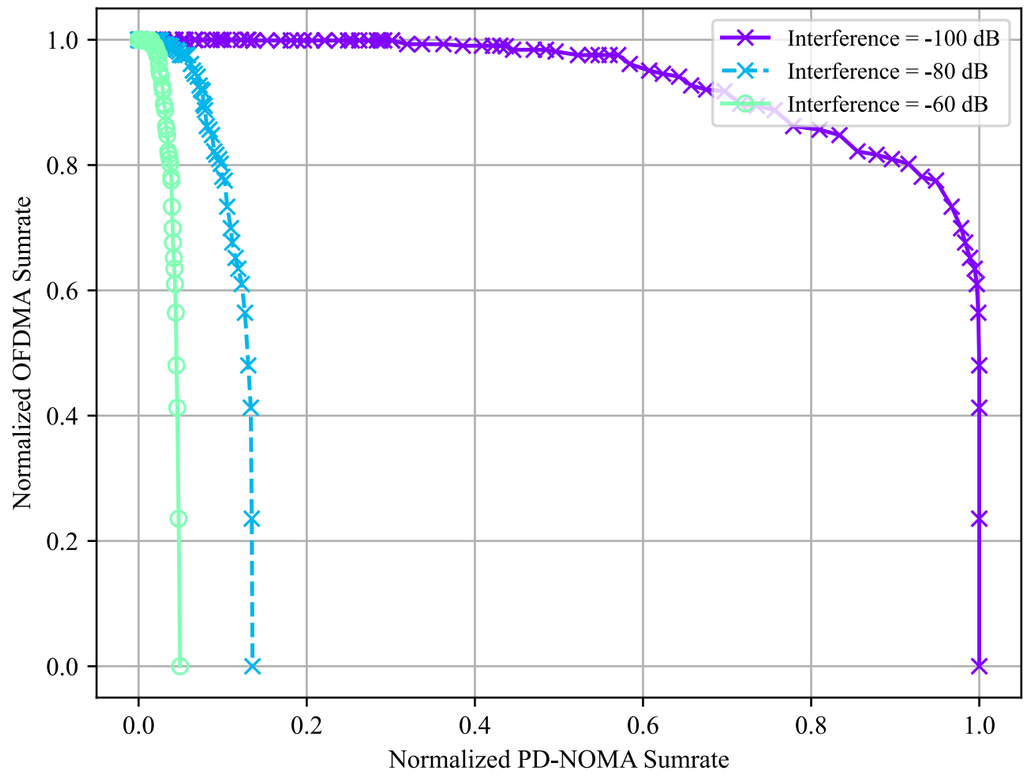


Figure 24: Relaxed MIP Pareto dominance rank plot (Normalized Intra-Interference effect on total sumrate)

will be able to transmit but at a lower throughput, which doesn't violate the constraint of the maximum Throughput threshold. The previous finding is apparent in the OFDMA figure since there is no intra-interference between members of the same group, their throughput is only affected by the channel. Therefore, the sumrate is constant as we increase the group size since we will only schedule the best possible out of all MTDs connected.

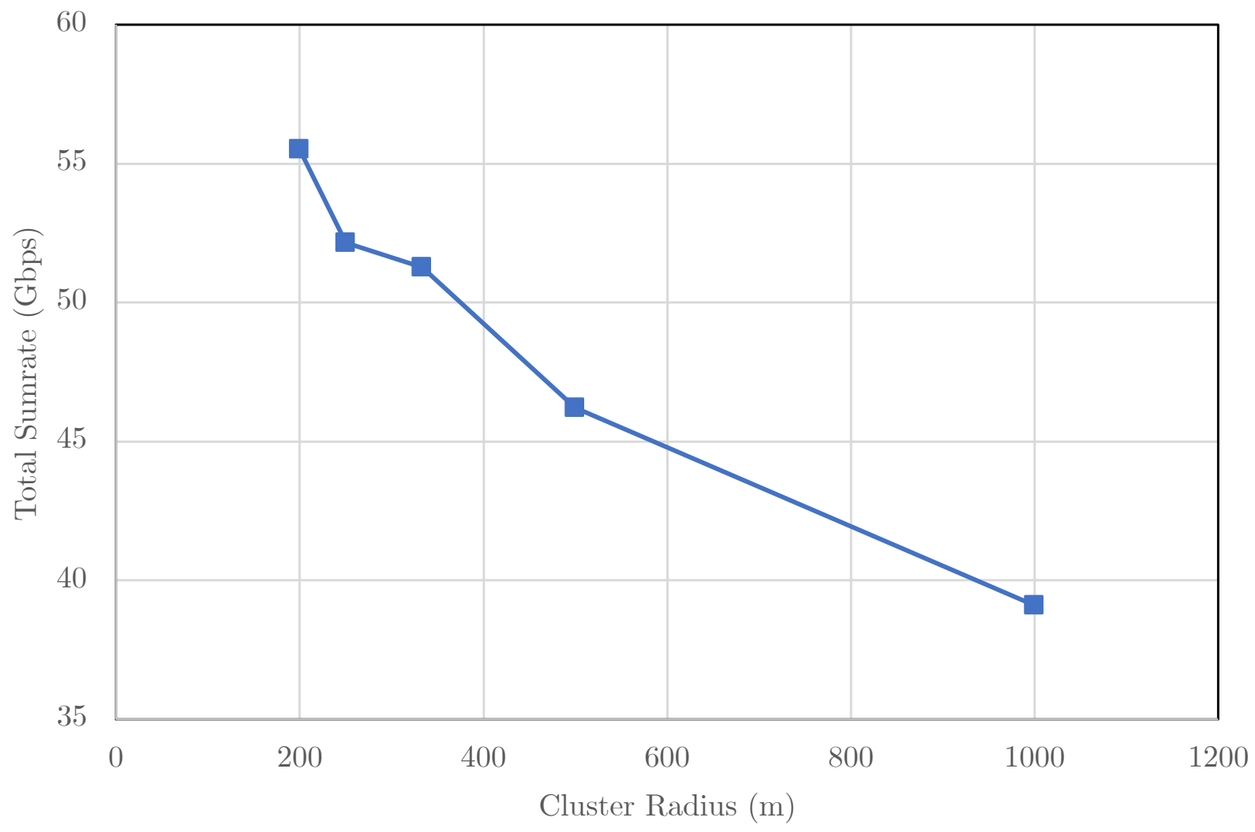


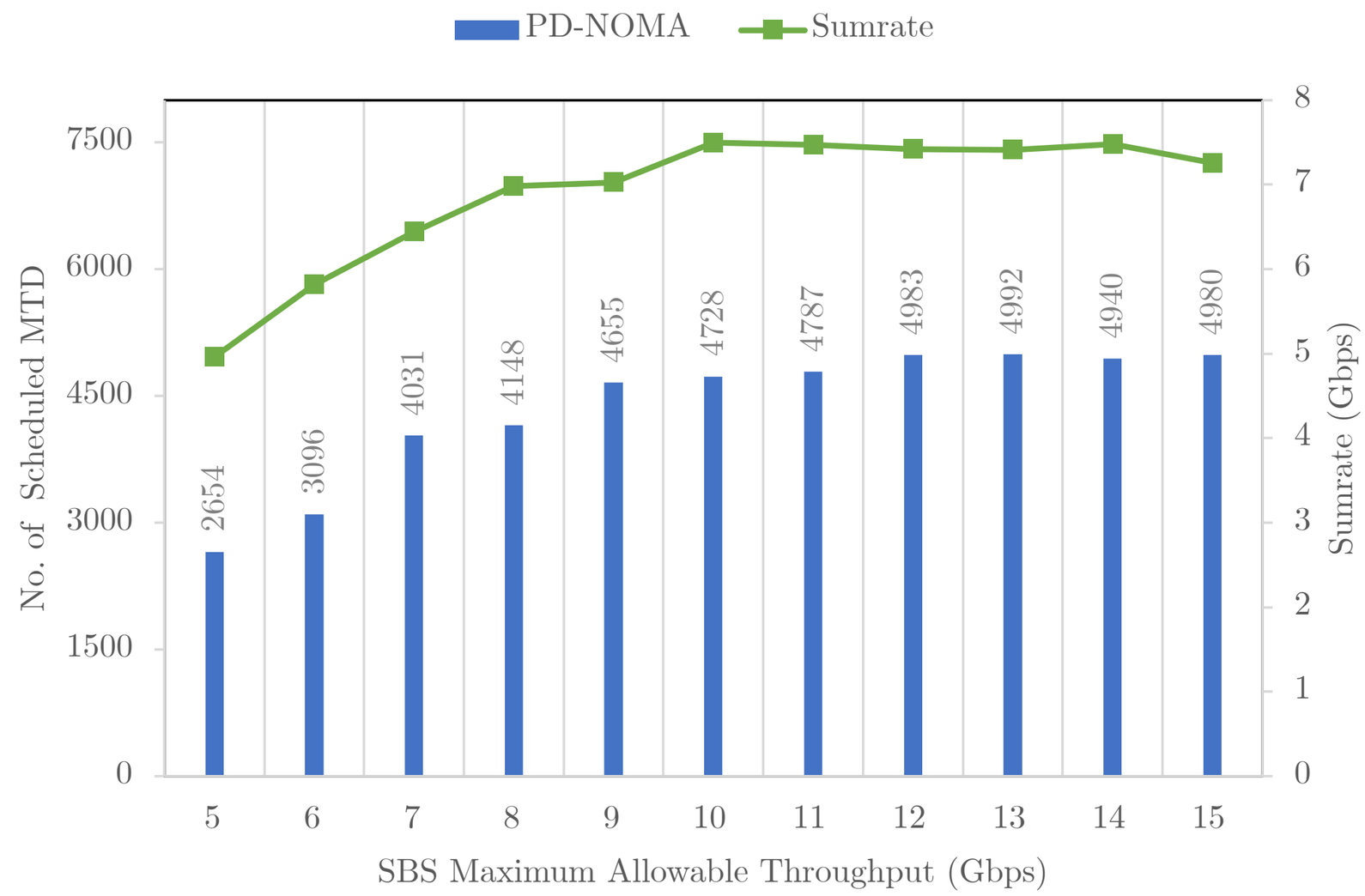
Figure 25: MINLP radius effect on sumrate maximization

5.1 Throughput Threshold Plots

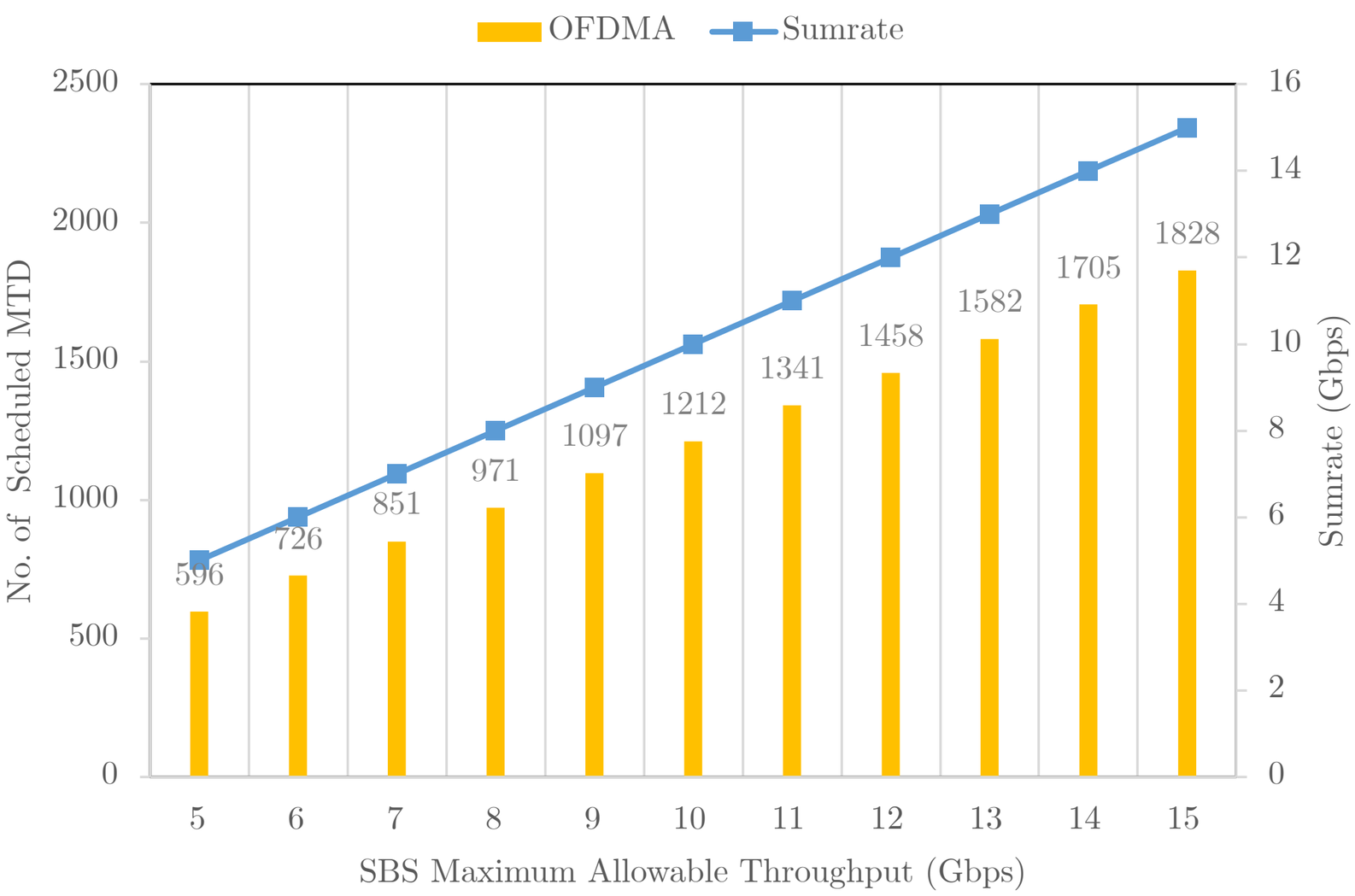
In this subsection, we discuss the effect of changing the SBSs' throughput threshold, meaning the maximum allowable throughput for all connected MTDs for each multiple access technique. We can deduce that as we increase the throughput capacity of OFDMA we will linearly increase the number of scheduled MTDs using OFDMA as illustrated in Fig. 26(b). While in Fig. 26(a), PD-NOMA-connected MTDs are unaffected by that increase in the throughput threshold, due to the high levels of intra-group interference per group. As a result, the scheduler will schedule all MTDs at higher throughput caps.

5.2 Scheduling Threshold Plots

In this subsection, we discuss the effect of changing the SBSs' scheduling capacity meaning there is no throughput threshold only the number of MTDs scheduled per TTI is limited. Thus, we can draw similar conclusions to the results from the previous subsection but on a different metric, which is the maximum achievable sumrate for both access techniques. OFDMA achievable throughput rates are shown in Fig. 27 (b); as stated, we can observe a linear growth of the sumrate of OFDMA-connected MTDs as the scheduling threshold increases. Moreover, we can conclude from Fig. 27 (a) that as we increase the maximum allowable scheduling threshold we can observe a stable sumrate for PD-NOMA connected MTDs since these devices are transmitting with very low rates due to intra-group interference. To emphasize the drawn conclusion above, in Fig. 27 (a), the best 500 MTDs have a total sumrate of 5.91 Gbps, while if we schedule all MTDs we only achieve a 7.60 Gbps on average, which is a mere improvement of 28.59 % as we allow for ten times more MTDs to transmit to the SBS. We can say that, on average, 500 MTDs (which are the highest ranks in their respective group) and 10% of all MTDs in the simulation achieve 77.76% of the maximum achievable rate of the connected MTDs.

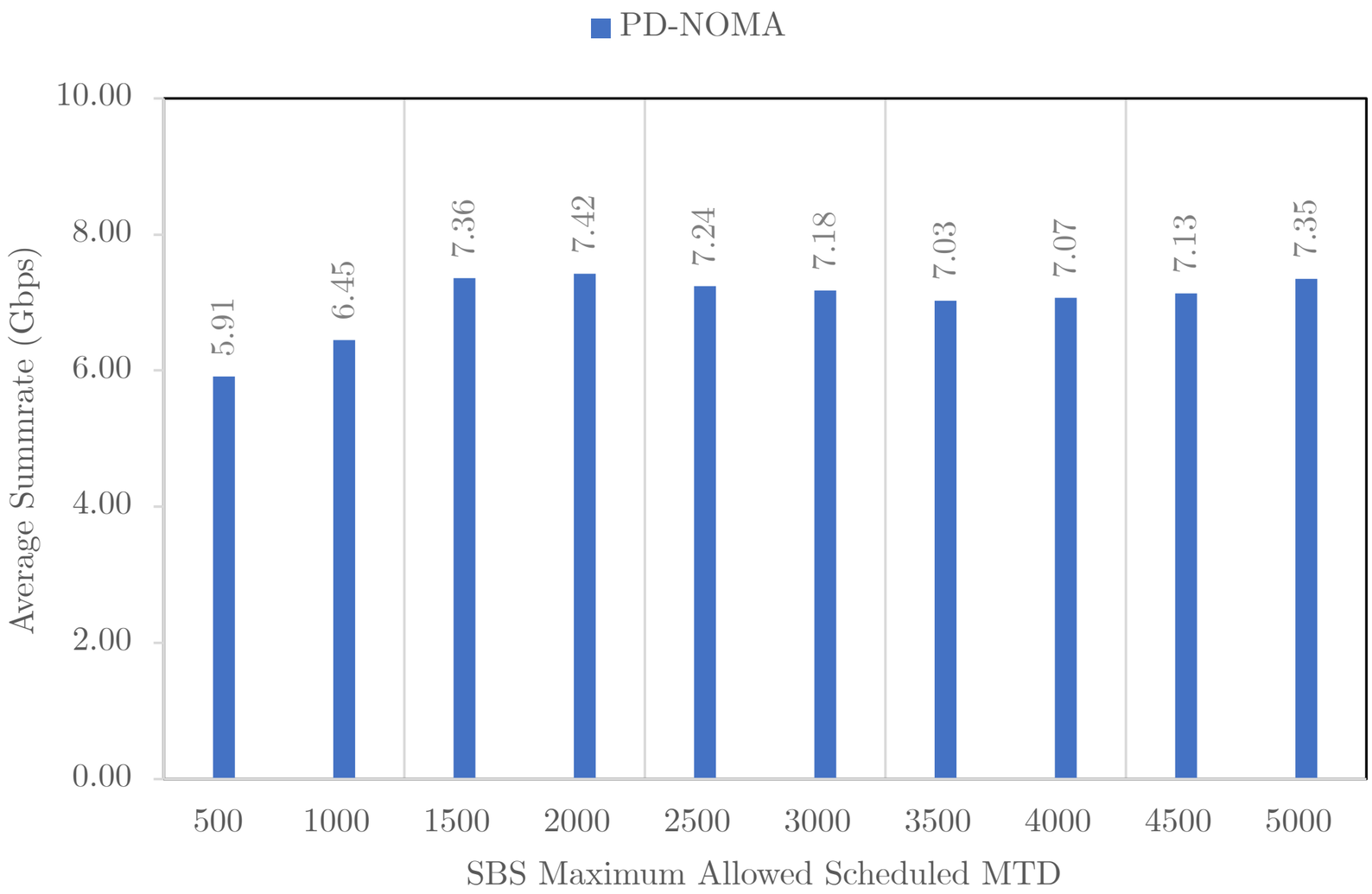


(a)

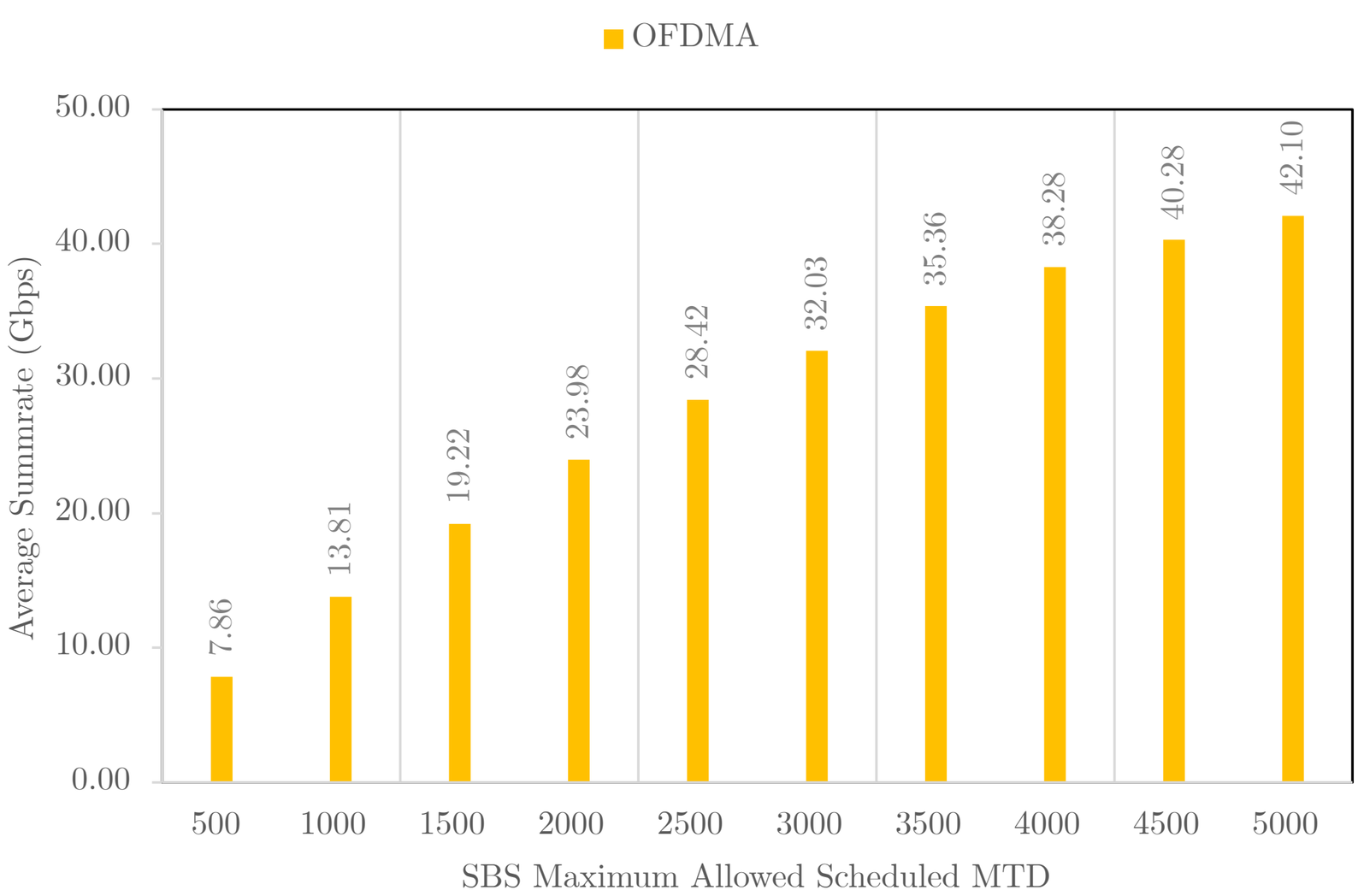


(b)

Figure 26: Number of scheduled MTDs with different throughput thresholds



(a)



(b)

Figure 27: Achievable sumrate with different total allowable scheduled devices

Chapter 6

Conclusions

6.1 Summary and Conclusion

This work presents our perspective on how backward compatibility could look in the near future. With the explosive demand for radio resources, especially with the stochastic demand from billions of machine-type devices, this accelerating burden could damage the current infrastructure with more dropouts and worsening QoS conditions. However, the standardization process is slow and the adoption of updated standards is more gradual than the growth of the number of users of all traffic types. Hence we are concerned with providing a blueprint on which backward compatibility is an important factor to consider when researching paradigms in this scope. We set up a simulation environment where MTDs are clustered following the Matérn cluster process, and then we group them based on their channel covariance. The grouping is designed to select m-MTDs per group for n-groups in b-SBSs, and then each group or number of groups is allocated a subcarrier, wherein each TTI we work under overloading conditions for all our experiments. We then formulated a joint scheduling and resource allocation optimization problem for the above-case scenario, which proved to be NP-Hard. Hence, we propose a heuristic solution to a relaxed OP and we solve our formulated optimization problems with Gurobi. We conclude that hyper-parameters like intra-group interference, varying cluster radius, and SBS maximum allowable throughput and scheduling capacity; have a tangible impact on the maximum achievable sumrate and/or scheduling capacity.

6.2 Recommendations and Future Work

- We explored the possibility of generalizing the Matérn cluster process to $\mathbb{R}^3$ as shown in Fig. 18, which could be a robust tool to enhance the simulation of telecommunication networks without the need to generate UEs by using, for example, a licensed software.

Henceforth, moving to 3D may introduce an order of magnitude of complexities on many levels of the problem (e.g., our scenario), which we believe is a ripe area for research.

- In our scenario, the scheduling and resource allocation is executed once every TTI and per RE, most of the deployed practical systems are still operating at SF level of scheduling and allocation, in an extension to this work we will focus on SF resource scheduling and allocation to perhaps give a more insightful results relative to the industry's requirements. This move to a higher level of resource scheduling and allocation will enable a less complex (processing complexity) scheduler and allows for the integration of practical 5G/4G communication system functionalities and processes.

- Mobility of the devices is perhaps one of the most important paradigms as many MTDs are mobile in nature and models like (Constant Acceleration, Gauss Markov, Random Walk 2D, and Random Waypoint) are commonly used models to simulate mobility in a telecommunication network. There is also the possibility of extending our work on PPP by adopting the method of spatial-temporal point processes.

- A promising extension to this work is by using DRL (e.g., actor-critic and policy-based methods) to solve the non-relaxed problem. We hypothesize that DRL will be a primary solving tool for current and next-gen telecommunication networks' paradigms.

References

[1] A. H. Khan, M. A. Qadeer, J. A. Ansari, and S. Waheed, "4G as a next generation wireless network," in *2009 International conference on future computer and communication*. IEEE, 2009, pp. 334–338.

[2] Z. Shen, A. Papasakellariou, J. Montojo, D. Gerstenberger, and F. Xu, "Overview of 3GPP LTE-advanced carrier aggregation for 4G wireless communications," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 122–130, 2012.

[3] N. H. Mahmood *et al.*, "Machine type communications: key drivers and enablers towards the 6G era," *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, p. 134, Jun. 2021. [Online]. Available: https://doi.org/10.1186/s13638-021-02010-5

[4] N. Abu-Ali, A.-E. M. Taha, M. Salah, and H. Hassanein, "Uplink Scheduling in LTE and LTE-Advanced: Tutorial, Survey and Evaluation Framework," *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1239–1265, 2014.

[5] K. Sridevi and M. A. Saifulla, "Control Plane Efficiency by Load Adjustment in SDN," in *Smart Trends in Computing and Communications*, ser. Lecture Notes in Networks and Systems, Y.-D. Zhang, T. Senjyu, C. So-In, and A. Joshi, Eds. Singapore: Springer, 2022, pp. 515–524.

[6] C.-W. Huang, S.-C. Tseng, P. Lin, and Y. Kawamoto, "Radio Resource Scheduling for Narrowband Internet of Things Systems: A Performance Study," *IEEE Network*, vol. 33, no. 3, pp. 108–115, 2019.

[7] F. Wang and X. Zhang, "Joint Optimization for Traffic-Offloading and Resource-Allocation Over RF-Powered Backscatter Mobile Wireless Networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 5, pp. 1127–1142, 2021.

[8] M. Peng, C. Wang, J. Li, H. Xiang, and V. Lau, "Recent advances in underlay heterogeneous networks: Interference control, resource allocation, and self-organization," *IEEE Communications Surveys Tutorials*, vol. 17, no. 2, pp. 700–729, 2015.

[9] Y. Xu, G. Gui, H. Gacanin, and F. Adachi, "A Survey on Resource Allocation for 5G Heterogeneous Networks: Current Research, Future Trends, and Challenges," *IEEE Communications Surveys Tutorials*, vol. 23, no. 2, pp. 668–695, 2021, conference Name: IEEE Communications Surveys Tutorials.

[10] P. Trakas, F. Adelantado, N. Zorba, and C. Verikoukis, "A QoE-Aware Joint Resource Allocation and Dynamic Pricing Algorithm for Heterogeneous Networks," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6.

[11] D. J. Langley *et al.*, "The internet of everything: Smart things and their impact on business models," *Journal of Business Research*, vol. 122, pp. 853–863, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S014829631930801X

[12] M. Fransman, "Evolution of the telecommunications industry into the internet age," *Jets Paper-University of Edinburgh Institute for Japanese European Technology Studies*, 2000.

[13] S. A. Kyriazakos and G. T. Karetsos, *Practical radio resource management in wireless systems*. Artech House, 2004.

[14] J. Zander, "Radio resource management an overview," in *Proceedings of Vehicular Technology Conference - VTC*, vol. 1, 1996, pp. 16–20 vol.1.

[15] ——, "Radio resource management in future wireless networks: Requirements and limitations," *IEEE Communications magazine*, vol. 35, no. 8, pp. 30–36, 1997.

[16] D. N. Knisely, Q. Li, and N. S. Ramesh, "cdma2000: A third-generation radio transmission technology," *Bell Labs Technical Journal*, vol. 3, no. 3, pp. 63–78, 1998.

[17] H. Kaaranen, A. Ahtiainen, L. Laitinen, S. Naghian, and V. Niemi, *UMTS networks: architecture, mobility and services*. John Wiley & Sons, 2005.

[18] O. I. Adu, B. O. Oshin, and A. A. Alatishe, "VoIP on 3GPP LTE network: A survey," *Journal of Information Engineering and Applications*, vol. 3, no. 11, 2013.

[19] R. B. Ali, S. Pierre, and Y. Lemieux, "UMTS-to-IP QoS mapping for voice and video telephony services," *IEEE network*, vol. 19, no. 2, pp. 26–32, 2005.

[20] L. Bos and S. Leroy, "Toward an all-IP-based UMTS system architecture," *IEEE Network*, vol. 15, no. 1, pp. 36–45, 2001.

[21] J. Laiho, A. Wacker, and T. Novosad, *Radio network planning and optimisation for UMTS*. John Wiley & Sons, 2006.

[22] E. Dahlman, S. Parkvall, and J. Skold, *4G, LTE-advanced Pro and the Road to 5G*. Academic Press, 2016.

[23] T. Akhtar, C. Tselios, and I. Politis, "Radio resource management: approaches and implementations from 4G to 5G and beyond," *Wireless Networks*, vol. 27, no. 1, pp. 693–734, 2021.

[24] H. Lee, S. Vahid, and K. Moessner, "A Survey of Radio Resource Management for Spectrum Aggregation in LTE-Advanced," *IEEE Communications Surveys Tutorials*, vol. 16, no. 2, pp. 745–760, 2014, conference Name: IEEE Communications Surveys Tutorials.

[25] O. Dizdar, Y. Mao, W. Han, and B. Clerckx, "Rate-Splitting Multiple Access: A New Frontier for the PHY Layer of 6G," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, 2020, pp. 1–7.

[26] X. Lin and N. Lee, *5G and Beyond*. Springer, 2021.

[27] S. Alraih *et al.*, "Revolution or Evolution? Technical Requirements and Considerations towards 6G Mobile Communications," *Sensors*, vol. 22, no. 3, p. 762, Jan. 2022, number: 3 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/1424-8220/22/3/762

[28] F. Abinader, A. Marcano, K. Schober, R. Nurminen, T. Henttonen, H. Onozawa, and E. Virtej, "Impact of bandwidth part (BWP) switching on 5G NR system performance." IEEE, 2019, p. 161–166.

[29] M. Kanj, V. Savaux, and M. Le Guen, "A Tutorial on NB-IoT Physical Layer Design," *IEEE Communications Surveys Tutorials*, vol. 22, no. 4, pp. 2408–2446, 2020.

[30] H. Fattah, *5G LTE Narrowband Internet of Things (NB-IoT)*. CRC Press, 2018.

[31] C. B. Mwakwata, H. Malik, M. Mahtab Alam, Y. Le Moullec, S. Parand, and S. Mumtaz, "Narrowband Internet of Things (NB-IoT): From Physical (PHY) and Media Access Control (MAC) Layers Perspectives," *Sensors*, vol. 19, no. 11, p. 2613, Jan. 2019, number: 11 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/1424-8220/19/11/2613

[32] Y. D. Beyene, R. Jantti, K. Ruttik, and S. Iraji, "On the Performance of Narrow-Band Internet of Things (NB-IoT)," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2017, pp. 1–6, iSSN: 1558-2612.

[33] X. Chen, Z. Li, Y. Chen, and X. Wang, "Performance Analysis and Uplink Scheduling for QoS-Aware NB-IoT Networks in Mobile Computing," *IEEE Access*, vol. 7, pp. 44 404–44 415, 2019, conference Name: IEEE Access.

[34] A. D. Zayas and P. Merino, "The 3GPP NB-IoT system architecture for the Internet of Things," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. Paris, France: IEEE, May 2017, pp. 277–282. [Online]. Available: http://ieeexplore.ieee.org/document/7962670/

[35] H. Malik, M. M. Alam, H. Pervaiz, Y. L. Moullec, A. Al-Dulaimi, S. Parand, and L. Reggiani, "Radio Resource Management in NB-IoT Systems: Empowered by Interference Prediction and Flexible Duplexing," *IEEE Network*, vol. 34, no. 1, pp. 144–151, Jan. 2020, conference Name: IEEE Network.

[36] M. El-Tanab and W. Hamouda, "An overview of uplink access techniques in machine-type communications," *IEEE Network*, vol. 35, no. 3, pp. 246–251, 2020.

[37] M. Mohammadkarimi, M. A. Raza, and O. A. Dobre, "Signature-based nonorthogonal massive multiple access for future wireless networks: Uplink massive connectivity for machine-type communications," *IEEE Vehicular Technology Magazine*, vol. 13, no. 4, pp. 40–50, 2018.

[38] M. Elbayoumi, M. Kamel, W. Hamouda, and A. Youssef, "NOMA-assisted machine-type communications in UDN: State-of-the-art and challenges," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1276–1304, 2020.

[39] N. Xia, H.-H. Chen, and C.-S. Yang, "Radio Resource Management in Machine-to-Machine Communications: A Survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 1, pp. 791–828, 2018, conference Name: IEEE Communications Surveys Tutorials.

[40] P. Jain and P. Kar, "Non-convex optimization for machine learning," *Foundations and Trends® in Machine Learning*, vol. 10, no. 3-4, p. 142–336, 2017, publisher: Now Publishers.

[41] E. C. Cejudo, H. Zhu, and J. Wang, "Resource Allocation in BER-Constrained Multicarrier NOMA Based on Optimal Channel Gain Ratios," in *2021 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2021, pp. 1–6.

[42] P. Semasinghe, S. Maghsudi, and E. Hossain, "Game Theoretic Mechanisms for Resource Management in Massive Wireless IoT Systems," *IEEE Communications Maga-*

zine, vol. 55, no. 2, pp. 121–127, Feb. 2017, conference Name: IEEE Communications Magazine.

[43] Y. Chen, B. Ai, Y. Niu, K. Guan, and Z. Han, "Resource Allocation for Device-to-Device Communications Underlaying Heterogeneous Cellular Networks Using Coalitional Games," *IEEE Transactions on Wireless Communications*, vol. 17, no. 6, pp. 4163–4176, Jun. 2018, conference Name: IEEE Transactions on Wireless Communications.

[44] U. Singh, A. Ramaswamy, A. Dua, N. Kumar, S. Tanwar, G. Sharma, I. E. Davidson, and R. Sharma, "Coalition Games for Performance Evaluation in 5G and Beyond Networks: A Survey," *IEEE Access*, vol. 10, pp. 15 393–15 420, 2022.

[45] A. Celik, A. Chaaban, B. Shihada, and M.-S. Alouini, "Topology Optimization for 6G Networks: A Network Information-Theoretic Approach," *IEEE Vehicular Technology Magazine*, vol. 15, no. 4, pp. 83–92, 2020.

[46] H. Al-Obiedollah, H. B. Salameh, S. Abdel-Razeq, A. Hayajneh, K. Cumanan, and Y. Jararweh, "Energy-efficient opportunistic multi-carrier NOMA-based resource allocation for beyond 5G (B5G) networks," *Simulation Modelling Practice and Theory*, vol. 116, p. 102452, 2022.

[47] A. Shahini and N. Ansari, "NOMA aided narrowband IoT for machine type communications with user clustering," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 7183–7191, 2019.

[48] I. E. Agbehadji, R. C. Millham, A. Abayomi, J. J. Jung, S. J. Fong, and S. O. Frimpong, "Clustering algorithm based on nature-inspired approach for energy optimization in heterogeneous wireless sensor network," *Applied Soft Computing*, vol. 104, p. 107171, Jun. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494621000946

[49] C. Saha, M. Afshang, and H. S. Dhillon, "Poisson cluster process: Bridging the gap between PPP and 3GPP HetNet models," in *2017 Information Theory and Applications Workshop (ITA)*, 2017, pp. 1–9.

[50] P. Rawat, S. Chauhan, and R. Priyadarshi, "A Novel Heterogeneous Clustering Protocol for Lifetime Maximization of Wireless Sensor Network," *Wireless Pers Commun*, vol. 117, no. 2, pp. 825–841, Mar. 2021. [Online]. Available: https://doi.org/10.1007/s11277-020-07898-8

[51] D. Wang, H. Qin, B. Song, K. Xu, X. Du, and M. Guizani, "Joint resource allocation and power control for D2D communication with deep reinforcement learning in MCC," *Physical Communication*, vol. 45, p. 101262, Apr. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1874490720303396

[52] M. Hua and Q. Wu, "Joint Dynamic Beamforming Design and Resource Allocation for IRS-Aided FD-WPCN," in *2021 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2021, pp. 01–06.

[53] A. Azari, Stefanović, P. Popovski, and C. Cavdar, "On the Latency-Energy Performance of NB-IoT Systems in Providing Wide-Area IoT Connectivity," *IEEE Transactions on Green Communications and Networking*, vol. 4, no. 1, pp. 57–68, Mar. 2020, conference Name: IEEE Transactions on Green Communications and Networking.

[54] L. Giupponi, R. Agusti, J. Pérez-Romero, and O. Sallent, "A novel joint radio resource management approach with reinforcement learning mechanisms." IEEE, 2005, p. 621–626.

[55] F. Al-Tam, N. Correia, and J. Rodriguez, "Learn to Schedule (LEASCH): A Deep Reinforcement Learning Approach for Radio Resource Scheduling in the 5G MAC Layer," *IEEE Access*, vol. 8, pp. 108 088–108 101, 2020.

[56] H. Yang, X. Xie, and M. Kadoch, "Intelligent Resource Management Based on Reinforcement Learning for Ultra-Reliable and Low-Latency IoV Communication Networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4157–4169, May 2019, conference Name: IEEE Transactions on Vehicular Technology.

[57] A. Azari, M. Ozger, and C. Cavdar, "Risk-Aware Resource Allocation for URLLC: Challenges and Strategies with Machine Learning," *IEEE Communications Magazine*, vol. 57, no. 3, pp. 42–48, Mar. 2019, conference Name: IEEE Communications Magazine.

[58] F. Hussain, S. A. Hassan, R. Hussain, and E. Hossain, "Machine Learning for Resource Management in Cellular and IoT Networks: Potentials, Current Solutions, and Open Challenges," *IEEE Communications Surveys Tutorials*, vol. 22, no. 2, pp. 1251–1275, 2020, conference Name: IEEE Communications Surveys Tutorials.

[59] Q. Zeng, Y. Du, K. Huang, and K. K. Leung, "Energy-Efficient Radio Resource Allocation for Federated Edge Learning," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6, iSSN: 2474-9133.

[60] N. Naderializadeh, J. Sydir, M. Simsek, and H. Nikopour, "Resource management in wireless networks via multi-agent deep reinforcement learning," 2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). Atlanta, GA, USA: IEEE, 5 2020, pp. 1–5, [Online; accessed 2021-01-22]. [Online]. Available: https://ieeexplore.ieee.org/document/9154250/

[61] M. Elsayed and M. Erol-Kantarci, "Radio Resource and Beam Management in 5G mmWave Using Clustering and Deep Reinforcement Learning," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[62] R. Li, Z. Zhao, Q. Sun, C.-L. I, C. Yang, X. Chen, M. Zhao, and H. Zhang, "Deep Reinforcement Learning for Resource Management in Network Slicing," *IEEE Access*, vol. 6, pp. 74 429–74 441, 2018, conference Name: IEEE Access.

[63] O. Edfors, M. Sandell, J. van de Beek, D. Landström, and F. Sjöberg, *An introduction to orthogonal frequency-division multiplexing*. Luleå tekniska universitet, 1997.

[64] S. B. Weinstein, "The history of orthogonal frequency-division multiplexing [history of communications]," *IEEE Communications Magazine*, vol. 47, no. 11, pp. 26–35, 2009.

[65] T. Jiang, L. Song, and Y. Zhang, *Orthogonal frequency division multiple access fundamentals and applications*. CRC Press, 2010.

[66] S. Srikanth, P. A. Murugesa Pandian, and X. Fernando, "Orthogonal frequency division multiple access in WiMAX and LTE: a comparison," *IEEE Communications Magazine*, vol. 50, no. 9, pp. 153–161, 2012.

[67] S. Barbarossa, M. Pompili, and G. B. Giannakis, "Channel-independent synchronization of orthogonal frequency division multiple access systems," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 2, pp. 474–486, 2002.

[68] M. Morelli, C.-C. J. Kuo, and M.-O. Pun, "Synchronization techniques for orthogonal frequency division multiple access (OFDMA): A tutorial review," *Proceedings of the IEEE*, vol. 95, no. 7, pp. 1394–1427, 2007.

[69] E. Yaacoub, A. M. El-Hajj, and Z. Dawy, "Weighted ergodic sum-rate maximisation in uplink orthogonal frequency division multiple access and its achievable rate region," *IET communications*, vol. 4, no. 18, pp. 2217–2229, 2010.

[70] S. Srikanth, V. Kumaran, C. Manikandan, and P. Murugesa Pandian, "Orthogonal frequency division multiple access: Is it the multiple access system of the future?" *AU-KBC Research Center*, 2007.

[71] A. B. Narasimhamurthy, M. K. Banavar, and C. Tepedelenliouglu, "OFDM systems for wireless communications," *Synthesis Lectures on Algorithms and Software in Engineering*, vol. 2, no. 1, pp. 1–78, 2010.

[72] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-s. Kwak, "Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges," *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 721–742, 2017.

[73] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, 2015.

[74] L. Dai, B. Wang, Z. Ding, Z. Wang, S. Chen, and L. Hanzo, "A survey of non-orthogonal multiple access for 5G," *IEEE communications surveys & tutorials*, vol. 20, no. 3, pp. 2294–2323, 2018.

[75] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for IoT: A survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 3, pp. 1805–1838, 2020.

[76] S. Gurugopinath, Y. Al-Hammadi, P. C. Sofotasios, S. Muhaidat, and O. A. Dobre, "Non-orthogonal multiple access with wireless caching for 5G-enabled vehicular networks," *IEEE Network*, vol. 34, no. 5, pp. 127–133, 2020.

[77] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *2013 IEEE 77th vehicular technology conference (VTC Spring)*. IEEE, 2013, pp. 1–5.

[78] P. Patel and J. Holtzman, "Analysis of a Simple Successive interference cancellation scheme in a DS/CDMA system," *IEEE journal on selected areas in communications*, vol. 12, no. 5, pp. 796–807, 1994.

[79] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Transactions on Communications*, vol. 98, no. 3, pp. 403–414, 2015.

[80] G. Sharma, "Pros and cons of different sampling techniques," *International journal of applied research*, vol. 3, no. 7, pp. 749–752, 2017.

[81] P. Raviteja, K. T. Phan, Y. Hong, and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE transactions on wireless communications*, vol. 17, no. 10, pp. 6501–6515, 2018.

[82] Q. Zhou, S. Shen, C.-W. Hsu, Y.-W. Chen, J. Finkelstein, and G.-K. Chang, "Novel parallel interference cancellation scheme for non-orthogonal multiple access in millimeter-wave ran using convolutional neural network," in *Optoelectronics and Communications Conference*. Optica Publishing Group, 2021, pp. W4A–6.

[83] A. S. de Sena, F. R. M. Lima, D. B. da Costa, Z. Ding, P. H. Nardelli, U. S. Dias, and C. B. Papadias, "Massive MIMO-NOMA networks with imperfect SIC: Design and fairness enhancement," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 6100–6115, 2020.

[84] A. Mishra, Y. Mao, L. Sanguinetti, and B. Clerckx, "Rate-splitting assisted massive machine-type communications in cell-free massive mimo," *IEEE Communications Letters*, 2022.

[85] H. Nikopour and H. Baligh, "Sparse code multiple access," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE, 2013, pp. 332–336.

[86] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern Division Multiple Access—A Novel Nonorthogonal Multiple Access for Fifth-Generation Radio Networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3185–3196, 2016.

[87] Y. Zhang, Z. Yuan, Q. Guo, Z. Wang, J. Xi, and Y. Li, "Bayesian receiver design for grant-free NOMA with message passing based structured signal estimation," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8643–8656, 2020.

[88] D. R. Cox and V. Isham, *Point processes*. CRC Press, 1980, vol. 12.

[89] D. J. Daley, D. Vere-Jones *et al.*, *An introduction to the theory of point processes: Volume I: Elementary theory and methods*. Springer, 2003.

[90] A. Baddeley, I. Bárány, and R. Schneider, "Spatial point processes and their applications," *Stochastic Geometry: Lectures Given at the CIME Summer School Held in Martina Franca, Italy, September 13–18, 2004*, pp. 1–75, 2007.

[91] F. Baccelli and C. Bordenave, "The radial spanning tree of a poisson point process," *The Annals of Applied Probability*, vol. 17, no. 1, pp. 305–359, 2007.

[92] D. I. Warton and L. C. Shepherd, "Poisson point process models solve the" pseudo-absence problem" for presence-only data in ecology," *The Annals of Applied Statistics*, pp. 1383–1402, 2010.

[93] J. Pitman and N. Ross, "Archimedes, gauss, and stein," *Notices AMS*, vol. 59, pp. 1416–1421, 2012.

[94] C. Saha, H. S. Dhillon, N. Miyoshi, and J. G. Andrews, "Unified Analysis of HetNets Using Poisson Cluster Processes Under Max-Power Association," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 3797–3812, Aug. 2019, conference Name: IEEE Transactions on Wireless Communications.

[95] M. Afshang and H. S. Dhillon, "Poisson Cluster Process Based Analysis of HetNets With Correlated User and Base Station Locations," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2417–2431, Apr. 2018, conference Name: IEEE Transactions on Wireless Communications.

[96] L. Yang, T. J. Lim, J. Zhao, and M. Motani, "Modeling and Analysis of HetNets With Interference Management Using Poisson Cluster Process," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 12 039–12 054, Nov. 2021, conference Name: IEEE Transactions on Vehicular Technology.

[97] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2022. [Online]. Available: https://www.gurobi.com

[98] Z. Wu *et al.*, "Accurate indoor localization based on CSI and visibility graph," *Sensors*, vol. 18, no. 8, p. 2549, 2018.

[99] D. Applegate and W. Cook, "A computational study of the job-shop scheduling problem," *ORSA Journal on computing*, vol. 3, no. 2, pp. 149–156, 1991.

[100] A. Ghane-Kanafi and E. Khorram, "A new scalarization method for finding the efficient frontier in non-convex multi-objective problems," *Applied Mathematical Modelling*, vol. 39, no. 23, pp. 7483–7498, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0307904X15001742