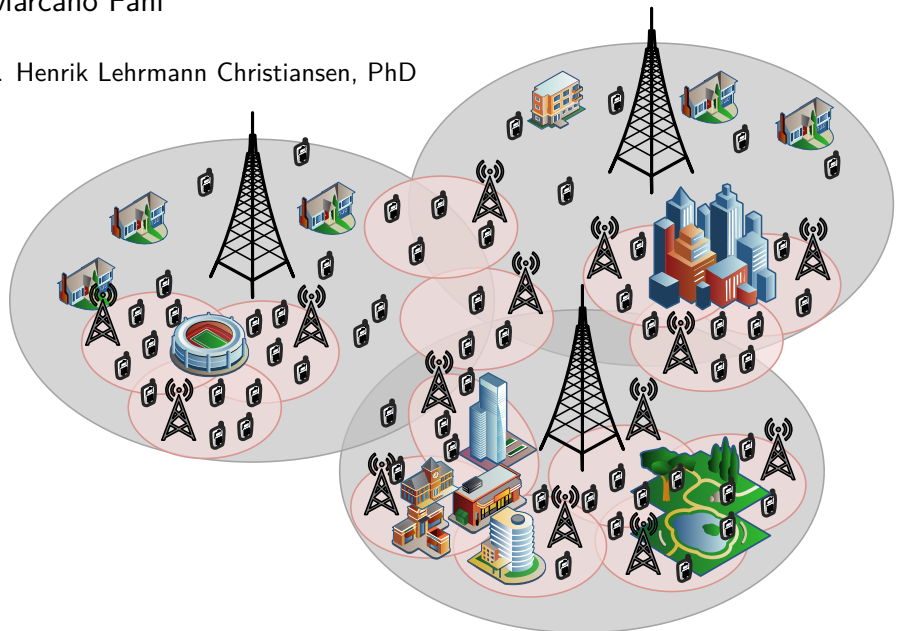


# Capacity Dimensioning for 5G Mobile Heterogeneous Networks

Andrea Sabrina Marcano Fani

Supervised by: Prof. Henrik Lehrmann Christiansen, PhD



Technical University of Denmark  
Kgs. Lyngby, Denmark 2018

*To my parents, Mara and Diógenes, for their unconditional support and encouragement. To my siblings Ángela, Diógenes and Manuel, for giving me joy. And to my dear nephew Uriel, because he holds a special place in my heart.*

**Technical University of Denmark**  
**DTU Fotonik, Department of Photonics Engineering**  
**Networks Technology and Service Platforms**  
Ørstedss Plads 343  
2800 Kgs. Lyngby  
DENMARK  
Tel: (+45) 4525 6352  
Web: [www.fotonik.dtu.dk/](http://www.fotonik.dtu.dk/)

$$f(x+\Delta x)=\sum_{i=0}^{\infty}\frac{(\Delta x)^i}{i!}f^{(i)}(x)$$

$$\int_a^b \varepsilon \Theta + \Omega \int \delta e^{i\pi} = -1$$
$$\sqrt{17}$$
$$\{2.7182818284\}$$
$$\chi^2$$
$$\Sigma$$
$$\gg$$
$$\approx$$
$$\lambda$$

οφείντοπισδφγηέκλ

# Contents

<b>Abstract</b>	v
<b>Resumé</b>	vii
<b>List of Acronyms</b>	ix
<b>Ph.D. Publications</b>	xiii
<b>Acknowledgements</b>	xv
<b>1 Introduction</b>	1
1.1 The evolution of mobile communications . . . . .	1
1.2 The need for 5G . . . . .	2
1.3 Thesis outline . . . . .	3
1.4 Research contributions . . . . .	4
<b>2 Introduction to 5G new air interface</b>	7
2.1 Introduction . . . . .	7
2.2 5G network architecture . . . . .	7
2.3 Key technology components for the 5G radio access . . . . .	15
2.4 Chapter summary . . . . .	25
<b>3 The theory behind NOMA</b>	27
3.1 Introduction . . . . .	27
3.2 Superposition transmission . . . . .	27
3.3 Successive interference cancellation . . . . .	28
3.4 Data rates . . . . .	29
3.5 NOMA advantages . . . . .	32
3.6 NOMA challenges . . . . .	32
3.7 Hybrid multiple access . . . . .	33
3.8 Chapter summary . . . . .	33

<b>4</b>	<b>5G NR and mmWave</b>	<b>35</b>
4.1	Introduction . . . . .	35
4.2	Cell discovery for mmWave small cells . . . . .	35
4.3	Network deployment: microwave vs mmWave . . . . .	54
4.4	Chapter summary . . . . .	57
<b>5</b>	<b>5G NR and NOMA</b>	<b>59</b>
5.1	Introduction . . . . .	59
5.2	QoS challenges . . . . .	59
5.3	Power allocation and MCS adjustment . . . . .	78
5.4	User pairing methods evaluation . . . . .	88
5.5	Chapter summary . . . . .	99
<b>6</b>	<b>5G HetNet with NOMA and mmWave</b>	<b>101</b>
6.1	Introduction . . . . .	101
6.2	Proposed load balancing technique: LB-NOMA . . . . .	102
6.3	Influence of mmWave on user pairing in NOMA . . . . .	110
6.4	HetNet capacity dimensioning . . . . .	118
6.5	Chapter summary . . . . .	133
<b>7</b>	<b>Conclusion and outlook</b>	<b>135</b>
	<b>Combined references</b>	<b>139</b>
	References from Chapter 1 . . . . .	139
	References from Chapter 2 . . . . .	139
	References from Chapter 3 . . . . .	142
	References from Chapter 4 . . . . .	143
	References from Chapter 5 . . . . .	145
	References from Chapter 6 . . . . .	147



# Abstract

Increasing the network capacity has historically been the main driver for the evolution of mobile communications, and that is no exception for the fifth generation (5G) of mobile networks. Therefore, enhancing the mobile broadband communications is one of the purposes of 5G.

This thesis analyses the performance and interworking of three solutions that have arisen as key pieces for boosting the capacity in 5G: heterogeneous networks (HetNets), millimeter wave (mmWave), and non-orthogonal multiple access (NOMA). Through diverse network scenarios and simulation models some of the challenges that are encountered when implementing mmWave and NOMA in HetNets are addressed, and methods to approach such challenges are proposed and evaluated.

Performance analyses provided for single cell and network scenarios show that the use of mmWave and NOMA can help 5G networks reach multi-Gbps rates and high spectral efficiency. Nevertheless, both technologies require rethinking how some of the network functions are done, and their implementation is not trivial. For mmWave, the challenges associated with the initial cell search in small cells are addressed; showing that new search methods are required to adjust to the highly directional transmissions that characterize communications at these frequencies. Approaches for HetNets where the macro and small cells cooperate to improve the performance of said methods are proven to be beneficial from both the access delay and a capacity dimensioning perspective.

For NOMA, special emphasis is done on how to mitigate the effects of the intra-cell interference. From the simulations results, it is shown that such interference can diminish the capacity benefits of NOMA if adequate methods are not used to account for it, especially in scenarios where NOMA coexist with other multiple access schemes. An approach considering transmission parameters adjustments done at the base station for NOMA is proposed; showing that with such type of adjustments the network can significantly benefit from the high capacity that NOMA offers. The importance of the user pairing method in the performance of NOMA is also addressed, confirming that the increase in the network capacity is directly related to the user-pairing algorithm. Moreover, it is shown that a flexible selection of the pairing algorithm subject to the load conditions in the network is preferred, since it can provide a balance between the algorithm complexity and efficiency and the network capacity. Therefore, this is a topic

that requires special planning when considering the use of NOMA, and even more so when mmWave is used.

For HetNets deployments, it is shown that the capacity can be further increased with NOMA when techniques that aim at maximizing the NOMA UEs and minimizing the OMA UEs in highly loaded cells are used; in this regard, the implementation of load balancing techniques designed for NOMA is suggested, since they can result in a better utilization of the available resources in the network.

Furthermore, through capacity dimensioning calculations, it is shown HetNets with NOMA and mmWave are a significant part of the solution for coping with the growing demand for data services and help 5G network reach their capacity goals.

# Resumé

Forøgelse af netværkskapaciteten har historisk set været hoveddrivkraften bag udviklingen af mobilkommunikation, og femte generations (5G) mobilnet er ingen undtagelse. Derfor er udbygelse af mobil bredbåndskommunikation et af formålene med 5G.

Denne afhandling analyserer ydeevnen og samspillet af tre løsninger som er blevet foreslået som vigtige brikker til kapacitetsudvidelse i 5G: Heterogene net (HetNets), Millimeter Waves (mmWaves) og Non-orthogonal Multiple Access (NOMA). Ved hjælp af en række netværks scenarier og simuleringsmodeller behandles nogle af de udfordringer der er ved at implementere mmWaves og NOMA i Hetnets og metoder til at imødegå disse udfordringer foreslås.

Performanceanalyse for enkeltceller og netværksscenarier bruges til at vise at brug af mmWaves og NOMA er et skridt på vejen til at 5G kan opnå multi-Gbps datarater og høj spektraleffektivitet. Ikke desto mindre kræver begge teknologer en gentænkning af hvordan visse netværksfunktioner udføres. For mmWaves bliver udfordringerne forbundet med initial cell search i small cell netværk belyst. Metoder til HetNets, i hvilke macro- og smallcells samarbejder om at forbedre performance af 5G, vises at være hensigtsmæssige både fra et access delay og kapacitetsdimensionerings synspunkt.

I NOMA er der speciel fokus på hvordan man kan afbøde effekterne af inter-celle interferens. Simuleringsresultater viser at sådan interferens kan mindske kapacitetsfordelene ved NOMA, medmindre der tages de nødvendige forholdsregler, specielt i de tilfælde hvor NOMA anvendes sammen med andre multiple access teknikker. Der foreslås en metode til at justere på sendeeffekten af basestationen i forbindelse med NOMA og det vises, at det kan forøge kapaciteten af NOMA væsentligt. Vigtigheden af parring af brugere og dets indflydelse på kapaciteten af NOMA undersøges, og dette bekræfter at måden brugeres parres på (parringsalgoritmen) har direkte indflydelse på netværkskapaciteten. Det vises desuden, at det er fordelagtigt med en fleksibel udvælgelse af parringsalgoritmer afhængig af netværksbelastningen, da det kan give en god balance mellem algoritmens kompleksitet, effektivitet og netværkskapacitet.

For NOMA udrulninger kan man opnå en yderligere forøgelse af kapaciteten med NOMA, når der anvendes teknikker, der forsøger at maksimere antallet af NOMA brugere og minimere OMA brugere; I denne sammenhæng foreslås der teknikker til at implementere load-balancing mekanismer, da de kan resultere i en bedre udnyttelse af netværk-

sressourcer.

Det vises endvidere gennem kapacitetsberegninger, at HetNets kombineret med NOMA og mmWaves er en vigtig brik i den samlede løsning til at håndtere de øgede krav til data tjenester og på den måde kan være et skridt på vejen til at nå kapacitetsmålene for 5G.

# List of Acronyms

<b>1G</b>	First Generation Mobile Networks
<b>2D</b>	Two-Dimensional
<b>2G</b>	Second Generation Mobile Networks
<b>3D</b>	Three-Dimensional
<b>3G</b>	Third Generation Mobile Networks
<b>3GPP</b>	Third Generation Partnership Project
<b>4G</b>	Fourth Generation Mobile Networks
<b>5G</b>	Fifth Generation Mobile Networks
<b>AiP</b>	Antenna-in-Package
<b>AoC</b>	Antenna-on-Chip
<b>BBU</b>	Base Band Unit
<b>BLER</b>	Block Error Rate
<b>BS</b>	Base Station
<b>BSM</b>	Binary Search Method
<b>C-RAN</b>	Centralized Radio Access Network
<b>CA</b>	Carrier Aggregation
<b>CDMA</b>	Code Division Multiple Access
<b>CDMA2000</b>	Code Division Multiple Access 2000
<b>CMOS</b>	Complementary Metal–Oxide–Semiconductor
<b>CN</b>	Core Network
<b>CoMP</b>	Coordinated Multipoint
<b>CPRI</b>	Common Public Radio Interface
<b>CQI</b>	Channel Quality Indicator
<b>D-RAN</b>	Distributed Radio Access Network
<b>DES</b>	Discrete Event Simulation
<b>DL</b>	Downlink
<b>DoA</b>	Direction of Arrival
<b>eCoMP</b>	Enhanced Coordinated Multipoint
<b>EDGE</b>	Enhanced Data Rates for GSM Evolution
<b>eICIC</b>	Enhanced Inter-cell Interference Coordination

<b>ESM</b>	Echaustive Search Method
<b>FD-MIMO</b>	Full Dimension Multiple-Input Multiple-Output
<b>FDMA</b>	Frequency Time Division Multiple Access
<b>feICIC</b>	Furhter Enhanced Inter-cell Interference Coordina- tion
<b>FTP</b>	File Transfer Protocol
<b>GPRS</b>	General Packet Radio Service
<b>GPS</b>	Global Positioning System
<b>GSM</b>	Global System for Mobile Communications
<b>HetNet</b>	Heterogeneous Network
<b>HSPA</b>	High-Speed Packet Access
<b>IaT</b>	Inter-Arrival Time
<b>ICIC</b>	Inter-cell Interference Coordination
<b>IMT-2020</b>	International Mobile Telecommunication system for 2020 and beyond
<b>IoT</b>	Internet of Things
<b>IP</b>	Internet Protocol
<b>ISD</b>	Inter-Site Distance
<b>ISI</b>	Intersymbol Interference
<b>JSM</b>	Joint Search Method
<b>LB-NOMA</b>	Load Balancing Non-Orthogonal Multiple Access
<b>LOS</b>	Line-of-sight
<b>LTE</b>	Long-Term Evolution
<b>M2M</b>	Machine-to-machine
<b>MA</b>	Multiple Access
<b>MAC</b>	Medium Control Access
<b>MCS</b>	Modulation and Coding Scheme
<b>MIMO</b>	Multiple-Input Multiple-Output
<b>mMIMO</b>	Massive Multiple-Input Multiple-Output
<b>mmWave</b>	Millimeter Waves
<b>MU-MIMO</b>	Multi-User Multiple-Input Multiple-Output
<b>MUST</b>	Multiuser Superposition Transmission
<b>NFV</b>	Network Function Virtualization
<b>NLOS</b>	No-Line-of-Sight
<b>NOMA</b>	Non-Orthogonal Multiple Access
<b>NR</b>	New Radio
<b>OBSAI</b>	Open Base Station Architecture Initiative
<b>OFDMA</b>	Orthogonal Frequency Division Multiple Access
<b>OMA</b>	Orthogonal Multiple Access
<b>OTN</b>	Optical Transmission Network
<b>PHY</b>	Physical Layer
<b>PON</b>	Passive Optical Network

---

<b>PRB</b>	Physical Resource Block
<b>QoS</b>	Quality of Service
<b>RAN</b>	Radio Access Network
<b>RAT</b>	Radio Access Technology
<b>RRH</b>	Radio Remote Head
<b>SC-FDMA</b>	Single-Carrier Frequency Division Multiple Access
<b>SDN</b>	Software Defined Networking
<b>SIC</b>	Successive Interference Cancellation
<b>SiGe</b>	Silicon-Germanium
<b>SINR</b>	Signal-to-Interference-plus-Noise Ratio
<b>SU-MIMO</b>	Single-User Multiple-Input Multiple-Output
<b>TBS</b>	Transport Block Size
<b>TDMA</b>	Time Division Multiple Access
<b>UDN</b>	Ultra-Dense Network
<b>UE</b>	User Equipment
<b>UL</b>	Uplink
<b>UMTS</b>	Universal Mobile Telecommunications System
<b>WCDMA</b>	Wideband Code Division Multiple Access
<b>WDM</b>	Wavelength Division Multiplexing
<b>Wi-Fi</b>	Wireless Fidelity





## Ph.D. Publications

### Paper A

M. Artuso, **A. Marcano**, H. Christiansen.

“Cloudification of mmWave-Based and Packet-Based Fronthaul for Future Heterogeneous Mobile Networks”.

Published in: *IEEE Wireless Communications*, vol. 22, no. 5, pp. 76-82, October 2015.

### Paper B

**A. S. Marcano**, H. L. Christiansen.

“Macro Cell Assisted Cell Discovery Method for 5G Mobile Networks”.

Published in: *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, Nanjing, China, 2016, pp. 1-5.

### Paper C

**A. S. Marcano**, H. L. Christiansen.

“Performance of Non-Orthogonal Multiple Access (NOMA) in mmWave wireless communications for 5G networks”.

Published in: *2017 International Conference on Computing, Networking and Communications (ICNC)*, Santa Clara, CA, US, 2017, pp. 969-974.

### Paper D

**A. S. Marcano**, H. L. Christiansen.

“A novel method for improving the capacity in 5G mobile networks combining NOMA and OMA”.

Published in: *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, Sydney, Australia, 2017.

### Paper E

**A. S. Marcano**, H. L. Christiansen.

“System-level performance of C-NOMA: a cooperative scheme for capacity enhancements in 5G mobile networks”.

Published in: *2017 IEEE 86th Vehicular Technology Conference (VTC Fall)*, Toronto, Canada, 2017.

**Paper F**

**A. S. Marciano**, H. L. Christiansen.

“Impact of NOMA on network capacity dimensioning for 5G mobile HetNets”.

Accepted for publication in: *IEEE Access*.

**Paper G**

Kishor Chandra, **Andrea S. Marciano**, Shahid Mumtaz, R. Venkatesha Prasad, Henrik L. Christiansen

“Combining Millimeter Wave and Non-Orthogonal Multiple Access for Power-Efficient 5G Ultra-Dense Networks: A System-Level Performance Evaluation”.

Accepted for publication in: *IEEE Vehicular Technology Magazine*.

# Acknowledgements

To me, it seems like it was yesterday when I got offered the opportunity to come to Denmark to pursue my Ph.D. I remember I had a lot of problems with my slow internet connection (oh, the irony!) during the interviews, and I clearly remember being warned about the Danish weather. Nonetheless, nothing could take away my excitement when I got the news. For that, I want to thank Lars Dittmann and Henrik Christiansen for trusting in me and giving me the chance to enroll myself in this very challenging, yet exciting journey. This has been a life-changing experience for me both professionally and personally, so I will be forever grateful to you.

To my supervisor Henrik Christiansen, I cannot thank you enough for your guidance and support, feedback, patience, all the chocolate and official DTU water, your effort in helping me learn some Danish (those vowels though...), and most of all for your good advice on how to improve my work and become a better researcher. Also, thank you very much for translating the abstract for me; I could have given it a try, but after seeing your translation I am happy I didn't. *Tusind tak!*

I want to thank my fellow Ph.D. students for making my days at DTU brighter and funnier. Special thanks to Matteo Artuso for being a great friend and for helping me review this thesis. Thanks to Artur Pilimon for your friendship and for all the long and meaningful talks. To Eder Ollora, thank you very much for all the talks, and coffees, and candies, and laughs; your friendship helped me get through the thesis writing period! Thanks to Aleksandra Checko, who so nicely provided me the information to apply for the Ph.D. position; I am sure I would not be here if it wasn't for that email.

To my family, thank you so much for all the support and love you constantly give to me. It is not easy being so far away from you, but as always, you help me get through the hardest times. *¡Los amo!*

To my mom Mara, thanks for your encouragement and friendship; you know how to give me that push to keep me going when I so badly need it. *¡Te amo!*

And last but not least, a huge special thank you to my beloved father Diógenes. You are and will always be my favorite professor. I cannot imagine getting to this level without your help, dedication, and all your lessons on telecommunications. You always have time and patience for me; so this treasure is for you! *Este reto fue difícil, pero tú me lo hiciste más fácil. ¡Te amo!*



# CHAPTER 1

## Introduction

---

As we found ourselves in an era where mobile communications have become an essential part of our everyday life, now more than ever it gives us the feeling that we can almost instantaneously have access to anything at anytime from anywhere. And to certain extent we can. Instant messaging and voice over Internet Protocol (IP) apps allow to easily communicate across the world. Mobile intranets allow for seamless communications in work environments, giving the employees access to their work-related tasks even when they are not at their desks. Video streaming and live video applications allow us to enjoy on-demand/live multimedia content through our mobile devices. These are only a few examples that reflect the impact that mobile communications have had in our society. All this mobile data being shared continuously all across the world has created the need for enhanced mobile networks that can handle the demand.

### 1.1 The evolution of mobile communications

Since the commercial deployment of the first mobile networks in the early 1980s, the mobile communications world has seen great advancements in relatively short time. Analog voice services were offered with the first generation (1G), introducing seamless connectivity for the first time and paving the way for mobile communications. With the second generation (2G) introduced in the late 1980s, digital voice services along with simple data (e.g., short text messages) were introduced. The increase voice capacity offered by 2G made it massively successful, with its more popular technology, Global System for Mobile Communications (GSM), still accounting for a significant share of the mobile subscriptions in developing markets. Nonetheless, GSM is currently the technology with the highest decline in subscriptions worldwide [Eri17b]. In its earlier versions, GSM used General Packet Radio Service (GPRS) to provide data services with peak rates of 0.5 Mbps; nowadays, after the introduction of Enhanced Data Rates for GSM Evolution (EDGE) and Evolved EDGE in the 2000s, Downlink (DL) peak rates up to 1.2 Mbps.

A big breakthrough in the mobile communications happened with the introduction of the third generation (3G) in the early 2000s. With 3G, mobile broadband services were offered, as well as faster and more reliable connectivity with higher Quality of Service (QoS); it was then possible to have access from everywhere to high-speed internet connections, multimedia services like video calling, and navigation services like Global Positioning System (GPS). As the successor of GSM, Universal Mobile Telecommuni-

cations System (UMTS) networks were widely deployed and today they cover around 80% of the world population and its coverage is estimated to grow to over 95% by 2023 [Eri17b]. In its more popular configuration, UMTS uses Wideband Code Division Multiple Access (WCDMA) for its radio interface and allows for DL peak rates up to 384 kbps. To cope with the increasing demand for data services, High-Speed Packet Access (HSPA) was introduced as an upgrade of WCDMA networks, allowing for DL peak rates of 14.4 Mbps. Later with the evolution of HSPA, known as HSPA+, DL peak rates of up to 42 Mbps, are supported.

The success and fast market penetration of 3G network created an insatiable demand for internet access and data services, forcing to rethink the design of mobile networks. In this respect, the fourth generation (4G) was introduced, with its first commercial deployment taking place in late 2009. 4G is based on the Long-Term Evolution (LTE) technology, and unlike the previous generations that relied on circuit-switched connections for real-time services and packet switched connections for data services, 4G LTE is a purely IP technology based only on packet switching. The main driver for 4G LTE was to deliver faster and better mobile broadband experiences while maintaining its compatibility with 3G. Orthogonal Frequency Division Multiple Access (OFDMA), high order modulation (up to 64QAM), large bandwidths (up to 20 MHz), Multiple-Input Multiple-Output (MIMO), and Carrier Aggregation (CA) are used in LTE for offering data rates that get can as high as 300 Mbps for the DL. Further enhancements have been done to the LTE standard since its beginnings; LTE Advance was launched in 2015 and relies on enhanced features like higher-order MIMO, enhanced CA, and interference cancellation to offer DL peak rates of 1 Gbps. The latest LTE version, known as LTE Advanced Pro, is in its very early stages and will offer DL data rates beyond 3 Gbps, building so the foundation for 5G networks and targeting the support of diverse new services including Internet of Things (IoT), smart cities, and autonomous driving. As the LTE subscriptions keep growing rapidly; by 2019, when the commercial launch of 5G is planned, LTE will account for more than 60% of all mobile subscriptions [Eri17b], making it the dominant technology. Then, a smooth migration from LTE to 5G networks is anticipated.

## 1.2 The need for 5G

The new fifth generation (5G) of mobile networks represents the future of mobile communications. 5G promises to deliver ubiquitous, ultra-fast, reliable mobile connectivity, offering new services and interconnecting not only people, but also objects, machines, and vehicles. The need for 5G comes mainly from the exponentially increasing demand for higher data rates (led mainly by video applications); furthermore, with the proliferation of the Internet of Things, billions of devices are expected to be connected to the mobile network by 2020 [Eri17b].

Machine-to-machine (M2M) communications, mobile broadband connectivity, and cloud services are some of the key enablers that have forced industry sectors such as

agriculture, automotive, constructions, health, retail, and transport, to redesign their business models and bring in innovation. 5G can help the industry achieve their goals by adjusting to the required services. Therefore, a greater flexibility and service versatility will be two of the major improvements introduced by 5G.

5G must be designed to fulfill the requirements of the International Mobile Telecommunication system for 2020 and beyond (IMT-2020) and to adapt to the diverse requirements of three main categories of services [Eri17a; Qua16]: enhanced mobile broadband; mission-critical services, such as autonomous vehicles, cellular-connected drones, and real-time health care; and massive IoT. Some of the main improvements that 5G will deliver are [Qua16]:

- Extreme data rates, with multi-Gbps peak rates and 100+ Mbps user experienced rates.
- Ultra-high density, with up to 1 million nodes per km<sup>2</sup>.
- Higher spectral efficiency, achieving more bits per Hz with new multiple access schemes and antenna techniques.
- Extreme capacity, with up to 10 Tbps per km<sup>2</sup>.
- Ultra-low latency, with values as low as 1 ms.
- Ultra-low energy, allowing 10+ years of life battery for IoT devices and optimizing network energy consumption.

### 1.3 Thesis outline

With the firsts deployments of 5G network expected to go live as soon as 2019, the standardization efforts for 5G are at their peak. As mentioned previously, enhanced mobile broadband will be one of the main services addressed by 5G. Therefore, capacity enhancements for this type of services are highly anticipated.

In this thesis, three solutions that have been identified as capacity boosters for the 5G Radio Access Network (RAN) are considered: Heterogeneous Network (HetNet)s, Non-Orthogonal Multiple Access (NOMA), and Millimeter Waves (mmWave). The purpose is to evaluate their performance, emphasizing the DL capacity improvements that can be expected by combining them, while addressing the challenges involved and proposing solutions. By doing so, the results obtained can be used for network capacity dimensioning; thus, providing an overview of some of the capabilities that 5G network can deliver.

Capacity dimensioning is a fundamental step in the process of deploying a network. Estimations of the number of sites needed, number of users expected to be served, monthly traffic, data plans offered, among others, are crucial when it comes to network planning process. An under-dimensioning of the network will compromise the QoS delivered, whereas an over-dimensioning will result in a sub-utilization of the network and unnecessary deployment costs for the network operators.

Considering this, the work presented in this thesis can be used as guidelines for the future deployments of 5G networks.

### 1.3.1 Structure

The research work done for this thesis is divided into six chapters. In Chapter 2, an introduction to the 5G air interface is provided, aiming at building the base for the following chapters. Chapter 3 is dedicated to the theory of NOMA. As it will be noticed from the research work, special emphasis is done on the implementation of NOMA; hence, the need for a comprehensive overview of the technology.

Two analyses related to the use of mmWave and their influence on the capacity dimensioning for 5G are presented in Chapter 4. The first is related to the initial cell search and the challenges that arise in this process when using the typical highly directional links with mmWave. The second shows a deployment comparison between a HetNet with microwave small cells and one with mmWave small cells. Similarly, Chapter 5 deals with the implementation of NOMA and how it affects the capacity for 5G; only single site scenarios are considered. Challenges related to the intra-cell interference and user pairing method are addressed, and proposed solutions are evaluated.

For the capacity dimensioning of a mmWave and NOMA based HetNet, Chapter 6 presents results from a network level perspective. The relationship between directional mmWave communication and the capacity benefits of NOMA are evaluated; load balancing techniques designed for NOMA are also analyzed. Dimensioning results based on the number of served users, number of needed sites, and network revenue are provided as the last analysis of the thesis. Finally, conclusions and outlook are available in Chapter 7.

## 1.4 Research contributions

The contributions of the research work presented in this thesis comprise the proposal of approaches for addressing some of the challenges directly related to the capacity improvements delivered by mmWave and NOMA for 5G. All the results were obtained through computational simulations of a 5G implementation. In Figure 1.1 the contributions of this work are summarized.

**The JSM cell search method for mmWave:** the Joint Search Method (JSM) is proposed as a solution for the cell discovery and initial access processes in HetNets with mmWave small cells. The basics of the JSM is to have a joint cell search where the macro and small cell cooperate. These processes are designed for microwave frequencies that do not require narrow beams for the signal propagation as it is the case with mmWave. The JSM method outperforms generic search methods evaluated, showing its benefits related to search overhead and initial access delay.

**The intra-cell interference with NOMA:** the impact that the intra-cell interference



in NOMA has on the quality of the received signal in comparison to Orthogonal Multiple Access (OMA) has been characterized. Through the results, it is shown that an average penalty of 12 dB must be accounted for in hybrid Multiple Access (MA) system for the NOMA UEs. This lead to the need of redesigning the propagation parameters and resource allocation for NOMA.

**MCS adjustments and extra transmission power for NOMA:** a method for mitigating the intra-cell interference in NOMA has been proposed. The method is based on Modulation and Coding Scheme (MCS) adjustments and extra transmission power allocation for the NOMA UEs with the purpose of improving their channel gains. This proposed method allows for a capacity gain in hybrid MA implementations, offering a solution for the co-existence of NOMA and OMA in 5G.

**Pairing methods for NOMA:** four pairing methods have been evaluated for NOMA and the use of a cost matrix to tag the each possible UEs pair based on the capacity gain it offers is proposed. The results confirm how the selection of the pairing method affects the capacity gain in the network. Furthermore, the results suggest that instead of considering a single pairing method, a pool of methods could be available so the network can flexibly select that that better adjust to the load conditions in the cells.

**LB-NOMA:** based on load balancing techniques, LB-NOMA is a scheme designed for NOMA implementation to be applied in scenarios where some cells are highly loaded while their direct neighbors have available resources. Forced handovers of the OMA UEs at the cell edges are performed in congested cells with the purpose of maximizing the NOMA UEs and minimizing the OMA UEs. With LB-NOMA significant capacity gains in the network can be achieved.

**mmWave and user pairing in NOMA:** the relationship between the beamwidth in directional mmWave transmissions and the pairing probability in NOMA has been evaluated. The beamforming gain of mmWave improves the performance of NOMA from a channel gain perspective, at the expenses of reducing chances of pairing UEs. Then, a balance between both performance metrics must be found in order to boost the capacity and without degrading the QoS when combining mmWave and NOMA.

**5G HetNet with mmWave and NOMA:** capacity evaluations were performed in a HetNet that combines mmWave and NOMA. It is shown trough UEs capacity dimensioning and network revenue estimations that this type of implementation is a fundamental part of the solution for the capacity requirements of 5G.

# 5G

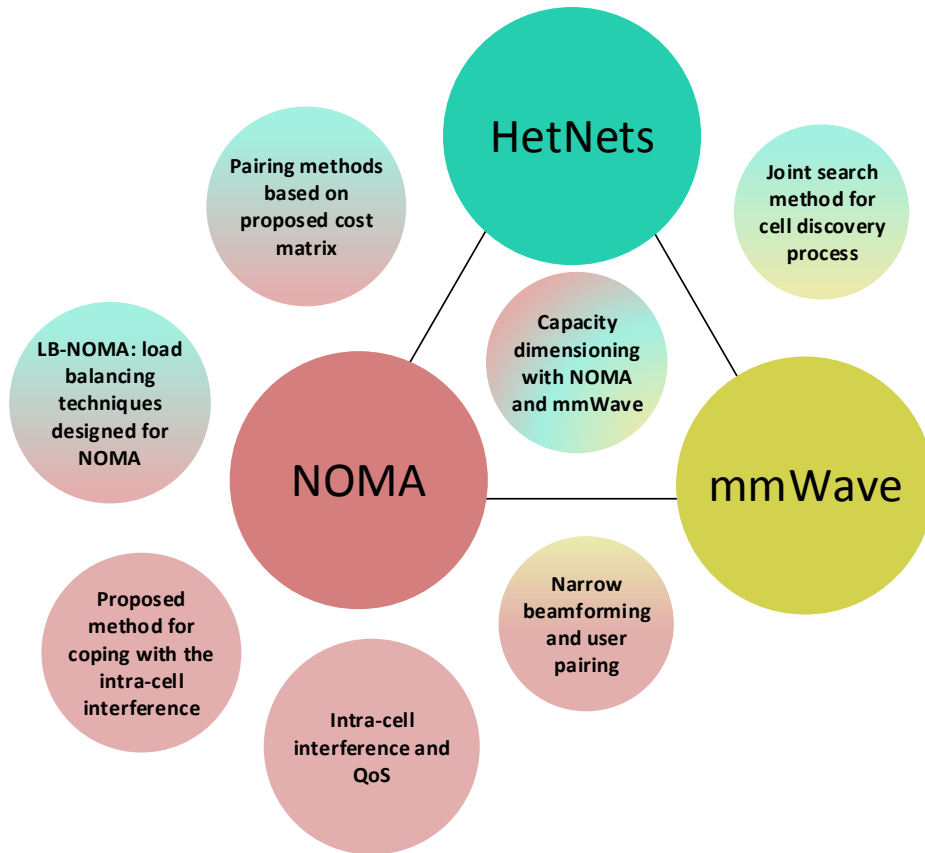


Figure 1.1: Research contributions from the thesis, comprising three solutions for 5G: HetNets, NOMA, and mmWave.

## CHAPTER 2

# Introduction to 5G new air interface

---

## 2.1 Introduction

When considering the requirements for Fifth Generation Mobile Networks (5G), it becomes apparent that an evolution of the radio interface is needed. The wireless air interface for the new radio access network RAN will comprise two tracks: the evolution of the LTE and what is known as New Radio (NR). A tight interworking of these two tracks will play a key role in achieving the goals of the IMT-2020 [ITU].

The LTE track for 5G networks will continue to undergo many enhancements to support and enable the IMT-2020 requirements; new features keep being added to the LTE standard, with the latest version, Release 14 [3GP17c], completed in June 2017 and ongoing research for Release 15. Unlike LTE for 5G, which has to provide backward compatibility, 5G NR is free of this and can introduce more fundamental changes, like new frequencies. In this regard, 5G NR will support different types of spectrum such as licensed, unlicensed, and shared, and will operate at low ( $< 1$  GHz), mid (1-6 GHz), and high ( $> 24$  GHz) frequency bands. Thanks to the scalability considered for 5G NR, it will eventually migrate to the frequencies currently used for LTE [Tey+17].

To cope with the extreme capacity demand and ultra-high density, 5G networks will have to adapt to diverse deployment topologies; massive deployment of small cells and advance antenna technologies will be used. They will be thereby characterized by a flexible deployment, by an interworking of Radio Access Technology (RAT)s, (e.g., 5G, 4G, and Wi-Fi) and by a versatility and adaptability of new key technologies that improve throughput, capacity, spectral efficiency, power consumption and device density.

This chapter focuses on the technologies for the 5G enhanced mobile broadband. First, in Section 2.2 the expected topology for the 5G RAN is presented. Second, an overview of mmWave, Massive Multiple-Input Multiple-Output (mMIMO) and Three-Dimensional (3D) beamforming, and NOMA, which are three of the promising technologies for 5G NR, is given in Sections 2.3.1, 2.3.2, and 2.3.3, respectively. The chapter is summarized in Section 2.4.

## 2.2 5G network architecture

### 2.2.1 Air interface

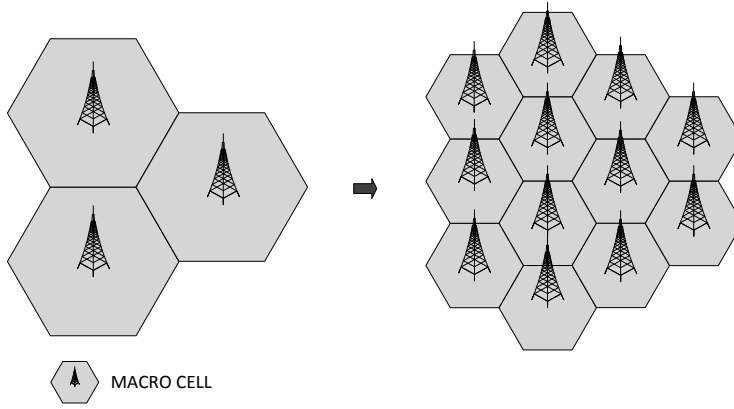
When it comes to increasing the capacity of a mobile network, there are three primary ways of doing so: adding more spectrum, increasing the spectral efficiency, and adding

more cell sites. The latter is known as network densification and through it many cell sites are deployed to support the high traffic demand, especially in densely populated urban areas and hotspots like e.g., shopping malls and stadiums.

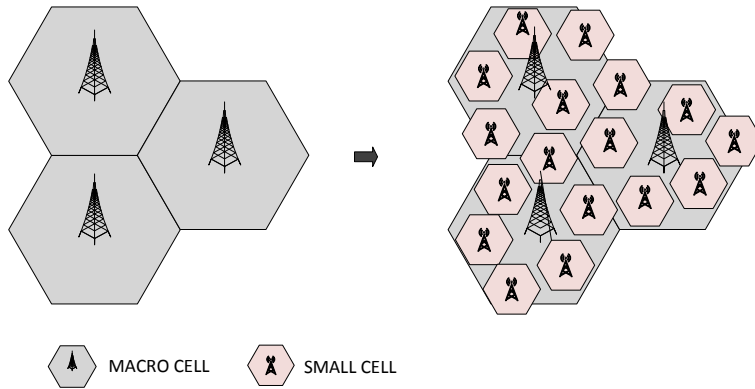
Because of the continuous traffic growth in mobile broadband, mobile network deployments must be adapted to cope with the high demand. Two of the requirements for 5G networks are to provide extreme network capacity and extreme data rates for the users; for both requirements, the densification of the RAN will play a key role. Network densification can be done by deploying more macro cell sites and maintaining a homogeneous network; this approach will imply reducing the Inter-Site Distance (ISD), thus increasing the traffic per km<sup>2</sup> without a corresponding increase in the traffic that needs to be handled by each macro cell. However, deploying more macro sites can only be considered until certain extent, since finding places for deploying this type of cell sites can become increasingly difficult and cost prohibited. Alternatively, the densification can be done through the deployment of lower-power network nodes, i.e., small cell sites, inside the coverage area of the already-deployed macro sites, creating a multi-layer or multi-tier network. A comparison of these two options for network densification is shown in Figure 2.1.

Due to their reduced size, site acquisition is easier and cheaper for small cell sites, as they are suitable for deployment in places such as lamp posts, traffic lights, and buildings facades. Moreover, the deployment of small cell sites has become simpler as features as interference, mobility, and Software Defined Networking (SDN) have been defined by the Third Generation Partnership Project (3GPP) for small cells [HTR16]. New wireless backhaul solutions have also emerged for small cell sites, facilitating their rollout. Small cell sites can be mainly added in hot spots where the data demand is high, by the edge of the coverage area of the macro cells to benefit the users susceptible to low QoS, and in area gaps not covered by the macro cells (both outdoor and indoor); thus, improving also the network coverage. The result of this type of deployment is a HetNet, where the small cells do not provide full-area coverage because of their short range but help boosting the capacity and also offloading traffic from the macro cells, thus improving the QoS for all the users in the network. Moreover, with the deployment of small cell sites the distance between the users and the base stations is reduced, which results in lower propagation losses, and higher data rates and energy efficiency [BL16; Rap+13b]. In Figure 2.2, an example of a typical architecture of a HetNet is shown.

Therefore, a major component of 5G will be an Ultra-Dense Network (UDN) configuration, with small cells being part of the foundation for efficiently increasing the network capacity. The concept of HetNets, however, is not a new one; Third Generation Mobile Networks (3G) and Fourth Generation Mobile Networks (4G) LTE networks already used the deployment of cell sites of different sizes within the same network. In 3G the macro and small cells were separated through the use of different frequencies. In 4G LTE, because of the frequency reuse of one, high interference due to the different cell sizes operating at the same frequency (in-band deployment), is a risk. Nevertheless, techniques such as Inter-cell Interference Coordination (ICIC), enhanced ICIC (eICIC), and further



(a) Homogeneous: deployment of more macro cell sites with their coverage area reduced.



(b) Heterogeneous: deployment of small cell sites on top of already deployed macro cell sites.

Figure 2.1: Network densification.

enhanced ICIC (feICIC) allow managing the inter-cell interference so that the network capacity can scale as more cells are added. The use of these interference mitigation techniques is a key feature for in-band deployments since they allow fully maximizing the benefits of HetNets. Furthermore, with feICIC the User Equipment (UE)s in the small cells have the capability of canceling the interference from the macro cell, thus experiencing higher data rates, (especially at the cell edges) and increasing the overall network capacity; this technique is especially beneficial for UDNs. Coordination techniques also help enhancing the performance of in-band HetNets; Coordinated Multipoint

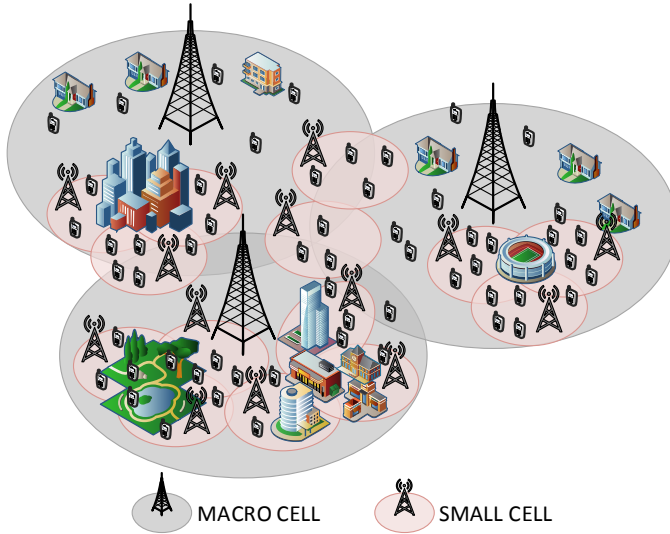


Figure 2.2: HetNet example with hot spots and cell edges covered by small cell sites.

(CoMP) and enhanced CoMP (eCoMP) allow to dynamically coordinate the scheduling and beamforming between neighboring cells. For deployments where the macro cells and the small cells operate at different frequencies (out-of-band deployment) the inter-cell interference can be handled with simple interference mitigation methods.

However, for UDNs like the ones expected for 5G, the interference management will be more challenging; the co-tier interference is rather high in this type of deployments. In a UDN, because of the large number of small cell sites and their short ISD (typically around 100 m), the users can be affected by many interfering signals without a clear dominant interference contribution. As a solution, two the following techniques can be considered: proactive time-domain inter-cell interference coordination and reactive carrier-based inter-cell interference coordination. In [HTR16], a detailed inside on the optimization of small cells can be found.

In [3GP17a], target scenarios for small cells enhancements have been defined for 5G HetNets. The out-of-band implementations in [3GP17a] represent one of the biggest advantages/changes for 5G; not only they allow exploring new frequency bands (e.g., mmWave bands) to enjoy more and wider spectrum, but also with them the decoupling of the control and user plane (C/U plane split) is possible.

In a conventional HetNet where macro cells and small cells operate at the same frequency, coverage and data services are simultaneously provided by both types of cells, with the control and data plane coupled. This architecture allows for ubiquitous coverage, at the expenses of having all the cells constantly working, even under low load

conditions, resulting in a sub-optimal use of resources and energy [Mus+16]. On the contrary, in a C/U plane split architecture, the macro cell is in charge of the control plane, and hence it provides ubiquitous coverage and manages the mobility using the lower frequency bands; it also provides data services for the UEs not covered by small cells. Moreover, the macro cells provide data services to high-speed UEs to avoid frequent handovers in the small cells. The small cells are in charge of the user plane, boosting the capacity by providing high-speed data connections, and more flexible/cost-energy efficient operations in higher frequencies [Moh+16; Mus+16; NTT14]; since the propagation losses increase as the frequency increases, high frequencies offer smaller coverage area, thus making them suitable for small cells. With the C/U plane split architecture, the UEs will be simultaneously connected to the macro and the small cells; this dual-connectivity allows for a fast handover of the UE to the macro cell in case that the connection to the small cell fails. With this architecture, a new interface will be required through which the macro cell can manage the small cells; this interface will allow the macro cell to activate/deactivate the small cells for energy saving purposes and to participate in the radio resource management to help mitigate the interference [Deh+14; Mus+16]. Figure 2.3 illustrates a HetNet with C/U plane split architecture.

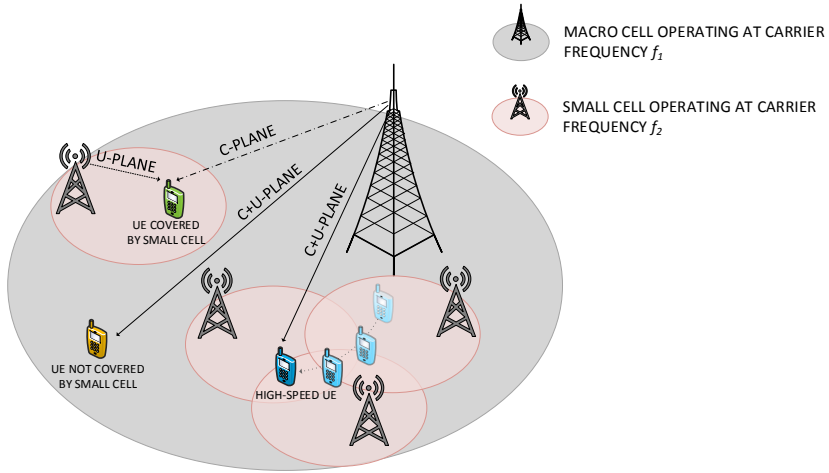


Figure 2.3: HetNet with C/U plane split architecture. High-speed UEs and those not covered by the small cells receive data and control from the macro cell. UEs covered by the small cells receive control from the macro cells and data from the small cells.

Dual-connectivity was first introduced for LTE, and 5G NR will continue to exploit its benefits; this will allow supporting simultaneous connections across 5G, 4G LTE, 3G, and Wi-Fi in a multi-connectivity deployment, that besides enabling a seamless and better user experience, will play an important role in allowing a high level of interworking between

LTE and NR for 5G. This is necessary in order to introduce 5G smoothly, simplify its deployment, and allow for a long transition period [Eri17b; Qua16]. LTE is continuing to evolve along with the standardization of NR, and pioneering many of the technologies that will be the foundation for 5G such as Gigabit LTE, advanced MIMO, higher-modulation order, and the use of unlicensed spectrum. With this tight interaction between LTE and NR, a user could, for example, be simultaneously connected to a low-frequency LTE layer that provides ubiquitous coverage, and to a high-frequency NR layer that provides high-data rates. Moreover, user plane aggregation between LTE and NR can also be done when both technologies are tightly integrated.

Even when the implementation of NR will require a new radio bearer, it is important to highlight that from a system perspective, LTE and 5G NR will be fully integrated [Eri17b]. NR can then be designed as a non-standalone system, using LTE as a mobility anchor and being an evolution of the existing networks, or as a stand-alone system for industrial applications and requiring a new 5G core network [Eri17b; Qua16]. In Figure 2.4 an example of a 5G air interface deployment illustrating such technology integration is shown.

### 2.2.2 Beyond the air interface

Naturally, not only the air interface will evolve towards 5G. Novel enablers such as Network Function Virtualization (NFV) and SDN provide the architectural flexibility needed to fulfill the diverse services requirements. The usage of both technologies is already happening for LTE, mainly for the Core Network (CN) [OMM16]. 5G networks, however, will explore these two technologies since the beginning, requiring a rethinking of some of the aspects of the traditional network design. The characteristic flexibility of 5G will extend to the RAN with the options of Centralized RAN (C-RAN) and Distributed RAN (D-RAN). The latter mainly operates in legacy architectures and is characterized by having all the Base Station (BS) radio components located in each cell. The BS radio consist of a Radio Remote Head (RRH) (i.e., a transceiver or an antenna) and a Base Band Unit (BBU) interconnected using Common Public Radio Interface (CPRI) or Open Base Station Architecture Initiative (OBSAI). In C-RAN the BBUs are relocated from each individual cell to a centralized unit refer to as a BBU pool; this centralization helps to simplify the network architecture, reducing operational costs, and speeding up deployments and scalability. The connection between the RRH and the BBU pool is known as fronthaul. C-RAN results highly interesting for UDNs since it relieves pressure on the network by having simplified cells with fewer equipments, less power, and less space.

Furthermore, with C-RAN the network performance can be improved by interconnecting the X2 and S1 interfaces coming from the multiple collocated BBUs, resulting in multiplexing gains. Nevertheless, C-RAN also comes with some implementation challenges. Since the digitized antenna data need to be carried over longer distances (i.e., from the cell to the BBU pool) it introduces latency, jitter and synchronization requirements



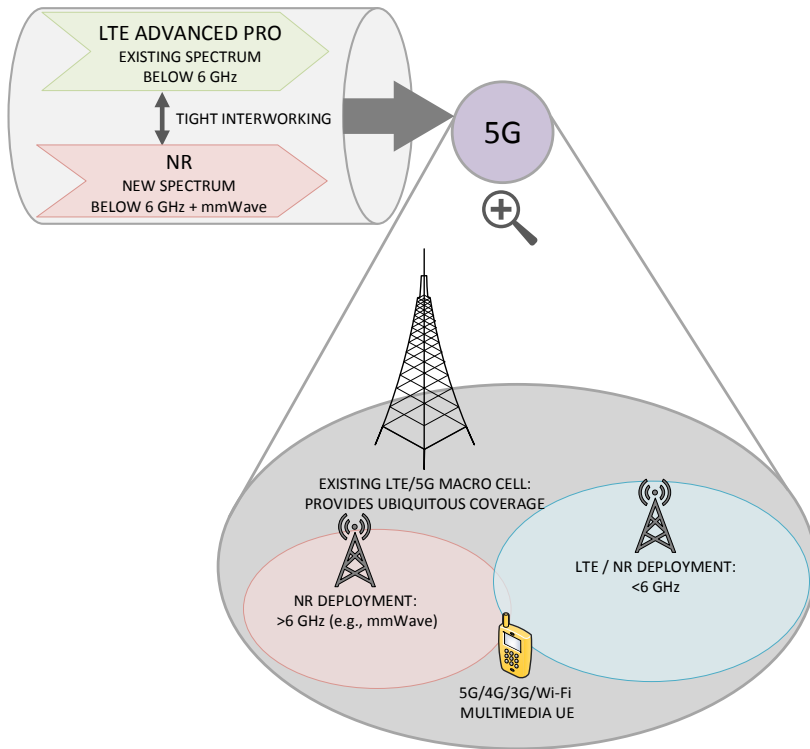


Figure 2.4: 5G air interface deployment

to the CPRI transport. A way of approaching this challenges is by using optical fiber connections from each site to the BBU pool; however, this significantly increases the deployment costs. Instead, potential alternatives candidates are: Optical Transmission Network (OTN); Passive Optical Network (PON); Wavelength Division Multiplexing (WDM); CPRI over Ethernet; and CPRI over wireless (e.g., microwave). This type of architecture also facilitates the use of NFV and SDN. Once the BBUs are centralized the baseband processing signal hardware can be replaced by a high volume server, virtualizing the BBU functionality (Cloud RAN) and further increasing the flexibility and performance of the network, as well as lowering the costs. C-RAN and Cloud RAN both represent fundamental changes for the mobile networks architectures.

Moving along the backhaul, the metro network will undergo changes towards the optimization of the cloud services. Nowadays, the metro network traffic surpasses that of the core network, making the former the center of the traffic growth. This pressures services providers to increase the CN capacity in order to avoid bottlenecks; the

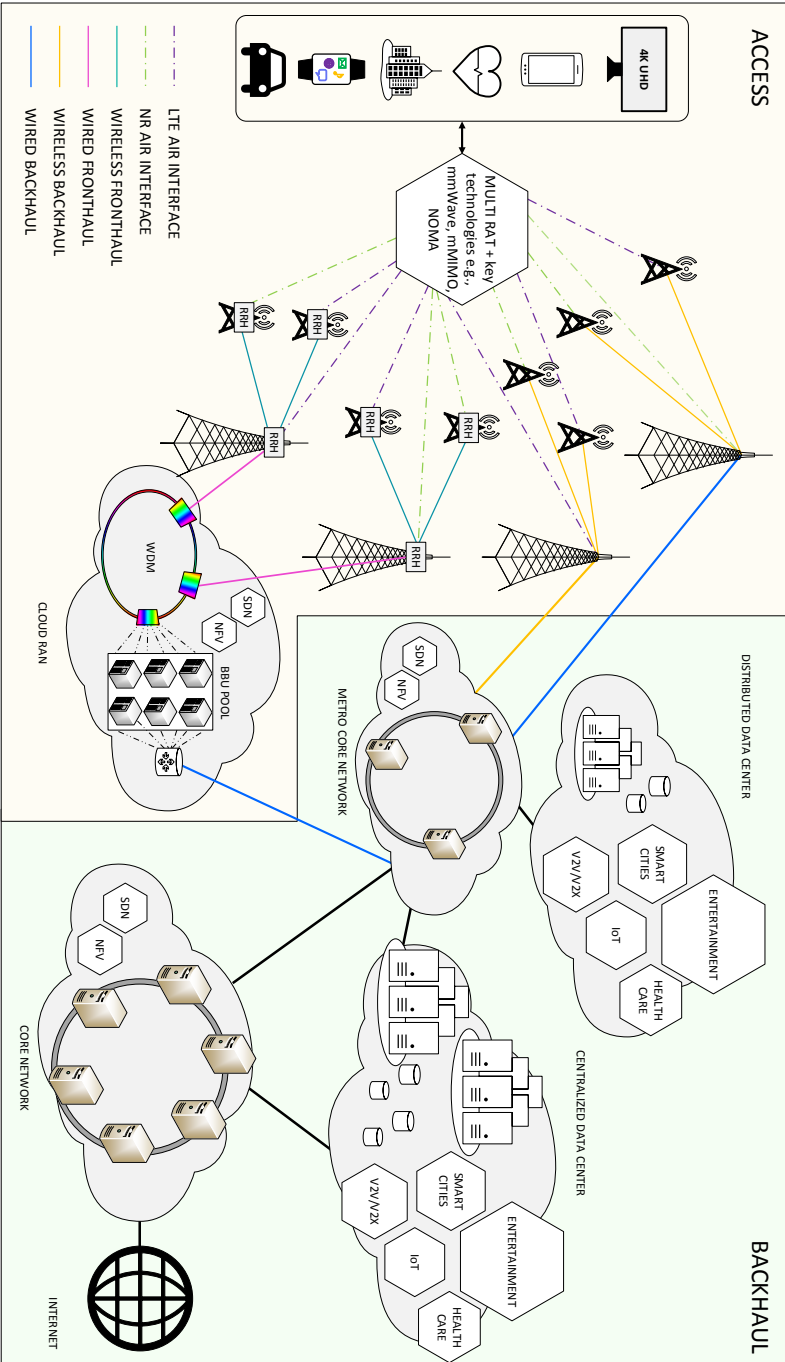


Figure 2.5: Example of a 5G network architecture

data centers, in particular, are specific challenges. Moreover, with edge computing the content storage and data center capacity are being decentralized, which creates substantial new traffic flows in the metro network. The metro network must thereby evolve to accommodate this shift. Since a big portion of the increasing metro network traffic will remain local, introducing a metro core into the metro network brings many benefits to the network. A metro core acts like a regional CN, providing an aggregation and distribution hub for subscribers and services traffic in the metro network. The growing traffic in the metro network can then be offloaded from the CN, allowing the latter to focus on switching the inter-metro and core transit traffic. More capacity, lower costs, and greater service flexibility can be achieved with the implementation of a metro core.

The combination of these innovative architectures and the redesign of the mobile network are key steps towards the success of 5G, allowing to fulfill the services requirements and deliver high QoS. In Figure 2.5 an example of a 5G network architecture is shown, which illustrates the interworking of LTE and NR, along with Cloud RAN, NFV, SDN, the converged metro and core network, and the decentralization of the data centers for providing the scalability and flexibility required for the main services for 5G.

## 2.3 Key technology components for the 5G radio access

The radio access for 5G will play a key role in addressing the exponential growth in the traffic demand; it must be designed to support a massive amount of devices connected to the network and offer highly reliable communications. Ubiquitous connectivity must be provided by 5G to any kind of device and any kind of application; therefore, the 5G radio access will not be based on a single specific technology, rather it will rely on a integration of several access and connectivity solutions that will fulfill the demand and address the requirements of the IMT-2020.

The capabilities of 5G will extend far beyond those of the current 4G LTE networks, therefore new technologies must be considered primarily targeting new spectrum and its efficient use, as well as that of the already available spectrum.

### 2.3.1 Millimeter waves

One of the major challenges that the current 4G LTE networks have had to deal with is the scarcity of spectrum in the typical microwave frequencies used for the wireless access and backhaul links. Although the introduction of techniques such as enhanced MIMO, eCoMP, HetNets, and CA help increase the capacity for LTE-Advanced implementations, they are not enough to cope with the huge high-speed traffic demand expected in the near future. For this, mmWave frequencies offer an attractive solution.

Mobile networks today mostly operate at frequency ranges from 300 MHz to 3 GHz with much less than 1 GHz of allocated spectrum [Rap+13b]. At these frequencies, the spectrum is highly congested due to the operation of many different services; AM and FM radio broadcasting, TV broadcasting, Wi-Fi, and current 3G/4G LTE cellular

communications are some examples of such services. This leads to the need to consider mobile communications in frequency bands that are available and can offer a high amount of spectrum. In the mmWave spectrum though, ranging from 30-300 GHz, the spectrum utilization is much lower. Military transmissions, vehicular radar, and backhaul connections are some of the services operating at mmWave frequencies. Nevertheless, a vast amount of spectrum is available at these frequencies. The 60 GHz band, also known as the V-Band, ranging from 57-66 GHz is characterized by a continuous block of 9 GHz of unlicensed available spectrum. The E-Band, comprised of two blocks of 5 GHz ranging from 71-76 GHz and 81-86 GHz, offers a total of 10 GHz of available spectrum. Other promising mmWave bands for 5G are at 28 GHz and 38 GHz, which combined roughly offer 4 GHz of spectrum. Yet the regulations on the use of these mmWave bands vary from country to country.

In the USA, 7 GHz of spectrum between 57-64 GHz [GF15] are designated for unlicensed mobile broadband communications; moreover, a 7 GHz adjacent expansion of this band, ranging from 64-71 GHz, has been made available for 5G unlicensed mobile broadband [Qual17b], creating a 14 GHz band of contiguous unlicensed spectrum. Other allocations for 5G in the USA comprise licensed spectrum between 27.5-28.35 GHz (2x425 MHz), 37.6-38.6 GHz (5x200 MHz), and 38.6-40 GHz (7x200 MHz); and light-licensed spectrum at 37-37.6 GHz (3x200 MHz). In Europe, the frequencies from 59-66 GHz have been allocated for mobile applications [Rap+14], although the main focus has been in the 26 GHz (24.25 GHz-27.5) [Eri17a; Qual17b]; further recommendations for the authorization of the spectrum between 31.8-33.4 GHz and 40.5-43.5 are agreed [Qual17b]. Similarly, in Japan, the 26 GHz and 28 GHz (with a maximum of 2 GHz) is considered for 5G [Eri17a; Qual17b], and allocation and technical rules are expected in 2018; in the 60 GHz band, the range from 59-66 GHz is available for mobile broadband. Korea have designated the bands between 57-66 GHz [Rap+14] and is considering allocating up to 2 GHz of spectrum between 26.5-29.5 GHz and 39 GHz no later than in 2021 [Eri17a; Qual17b]; Australia has designated the bands between 57-66 GHz and 59.3-62.9 GHz [Rap+14], and has announced that by 2018 trials will begin in the 26 GHz, 28 GHz, and 39 GHz bands [Eri17a; Qual17b]. An illustration of the wireless radio spectrum and available bandwidth for frequencies up to 100 GHz is shown in Figure 2.6; although this illustration is based on the USA spectrum, it is remarkably similar for most countries. For frequencies above 100 GHz hundreds of GHz are also available, which could be used for mobile communications.

This high amount of free spectrum is what makes mmWave bands so attractive for 5G networks. With more spectrum it is possible to provide higher data rates while using the same modulation techniques and it also provides more resources that can be shared among multiple users, allowing for more users to be served at the same time. Since one of the main aspects that has been present in every evolution of the mobile network is higher capacity, it is natural to question why mmWave bands had not been considered before for mobile communications; the reason being their challenging propagation characteristics.

Because of their short wavelength (less than 1 cm) most of the objects that mmWave

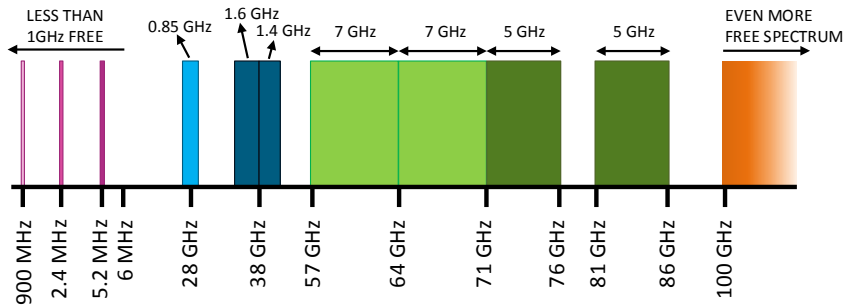


Figure 2.6: Illustration of the radio spectrum and available bandwidth.

frequencies encounter during their propagation path are much larger than their wavelengths, which makes signals at these frequencies fade rather quickly because of the high path losses, thus reaching short distances, typically within 100 m. Such unfavorable propagation characteristics were seen as a limitation when considering the deployment of mobile networks, especially when networks were homogeneous and consisted mainly of macro cells covering distances larger than 1 Km. However, as mobile networks continue to become denser the coverage area of the macro cells has started to shrink, resulting in dense deployments where the typical macro cell range is less than 1 Km. Furthermore, as network densification is moving towards heterogeneous networks (Section 2.2), there is the need to deploy small cells. In this regard, mmWave frequencies are now considered as good candidates for mobile networks, and the characteristics that were once seen as a limitation can be used as advantages for future deployments.

Nevertheless, the use of mmWave comes with many challenges. Besides the high path losses due to the small wavelengths, rain attenuation and  $O_2$  absorption play a major role in the limited distance achievable with mmWaves. Figure 2.7 illustrates the atmospheric attenuation due to the absorption of gaseous molecules such as  $O_2$  and water vapor for frequencies up to 200 GHz; the 60 GHz band suffers from significantly higher attenuation than the other mmWave frequencies that have been considered for 5G (e.g., 28 GHz, 38 GHz, E-band). At 180 GHz high attenuation is also present. Moreover, it can be seen that for the typical frequencies used in today's mobile networks ( $< 6$  GHz) the atmospheric attenuation is negligible in comparison to higher frequencies. As for the rain attenuation, an illustration of it is shown in Figure 2.8. Although water vapor does not significantly attenuate the signals at normal concentration, once this vapor turns into droplets the attenuation becomes higher; especially as the rain intensifies and the droplets become raindrops reaching up to 4 mm in diameter, which is in the order or even bigger than the wavelengths for mmWave. Similarly as with the atmospheric attenuation, it can be seen that propagation at mmWave frequencies can be highly affected by the rain.

These attenuations due to weather conditions add up to the propagation losses, making

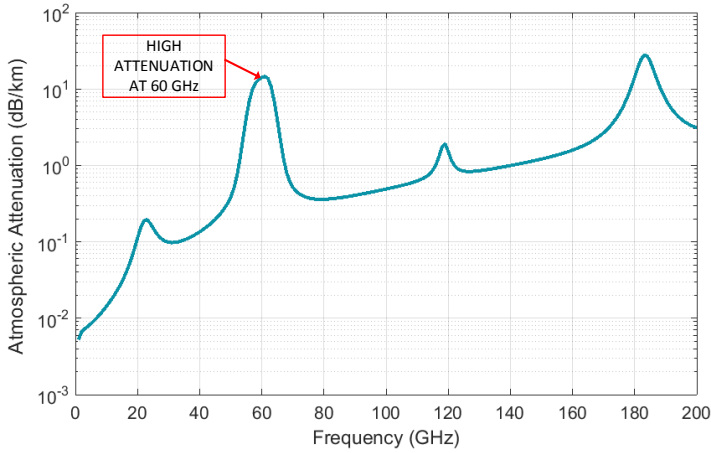


Figure 2.7: Atmospheric attenuation for frequencies up to 200 GHz.

outdoor communications at mmWave challenging since they need to overcome high losses. To address this, highly directional steerable antennas arrays can be used and rely on beamforming techniques to achieve high gain, making the mmWave signal more robust to weather conditions and penetration through obstacles. With the help of such antennas hundreds of meters [Qual7a] or even a few kilometers (e.g., 1 or 2) [Deh+14; Roh+14; Sam13] can be covered by high-speed mmWave connections.

The narrow beams and short distances of mmWave make them suitable for unlicensed and light-licensed transmissions; since the signal will have its power concentrated in a specific area avoiding power spillage and it will fade rapidly, the interference with other nearby signals will be minimal. Furthermore, higher security and privacy is offered with this type of transmissions.

With the high-gain steerable antennas, it is possible to have mmWave transmissions not only in Line-of-sight (LOS) scenarios but also in No-Line-of-Sight (NLOS) scenarios where the path between the transmitter and receiver is blocked by obstacles. The reflections and scattering allow wireless NLOS mmWave links to be established as long as the antennas are steered to find the objects that reflect the signals and scatter the energy, instead of having the transmitter and receiver antennas pointing at each other [Qual7a; Rap+13a]. The many reflective paths that can be experienced over large operating bandwidths generate high Intersymbol Interference (ISI), also known as frequency selectivity, due to the successive symbols arriving at the receiver that overlap and interfere with one another. This would be the case of omnidirectional mmWave transmissions, where the multipath components can generate tens or hundreds of nanoseconds of delay [Rap+14]. The highly directive steerable antennas can be used to mitigate the ISI by reducing the delay spread [MSR15].

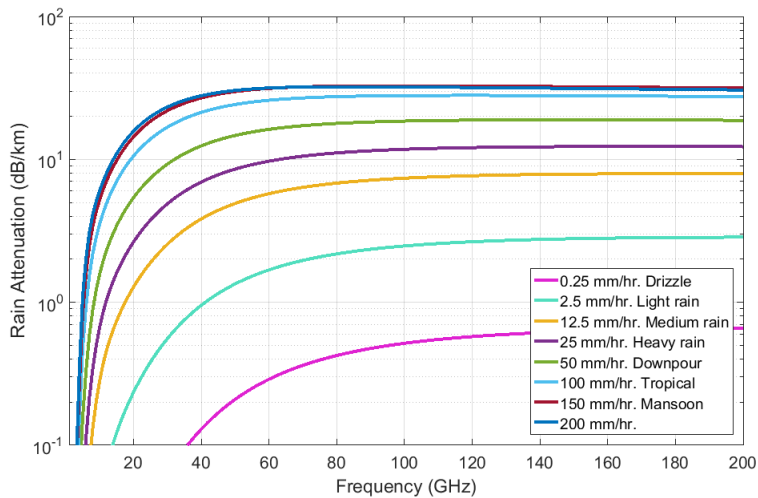


Figure 2.8: Rain attenuation for frequencies up to 200 GHz.

The challenges that come with the implementation of mmWave extend to many layers of the communications stack. At a Physical Layer (PHY) level, besides the propagation challenges addressed above, mmWave communications come with major hardware constraints related to the antennas design, power amplifiers with large dynamic ranges, and Complementary Metal–Oxide–Semiconductor (CMOS) technology with good reliability [Zha+17]. Highly dense antenna arrays with high-gain antenna elements must be designed for mmWave, and whether they will be integrated as Antenna-on-Chip (AoC) or Antenna-in-Package (AiP) remains a vendor preference. With AoC solutions the antenna array is directly integrated with the front-end circuit on the same chip using technologies like CMOS or Silicon-Germanium (SiGe); this solution is simpler and more affordable because of the short wavelength and reduce size of the antenna elements [Rap+14; Tit+12; ZL09]. With AiP the antenna array is fabricated in a separate chip-scale substrate and/or package and then integrated with the front-end circuit. This solution has demonstrated to achieve high speed and high gain for mmWave transmissions [Tit+12; ZL09], arising as a promising solution for 5G; however, AiP solutions are more complex than AoC and suffer from losses in the interconnections to the front-end circuit. Furthermore, the steering capabilities needed in the mmWave antennas require protocol modifications at the signal processing level of the PHY and data link layers to be able to direct the links [Rap+14].

New algorithms for channel equalization, modulation and beam steering must be considered for a smooth design that takes into account the mmWave wireless propagation characteristics and the hardware requirements. A trade-off between the linearity of

the equalization and the equalization period must be achieved in order to reduce the complexity of the mmWave devices, and although beam steering techniques help to facilitate the channel equalization process, they come with computational complexities.

Most of these complexities affect the processes of the Medium Control Access (MAC) layer. Not only is a tight integration between beam steering and modulation algorithms needed, but also beam steering, along with narrow beams, directly impacts on the cell discovery process and contributes to the hidden node problem in wireless networks; the protocols used to address these aspects rely on the MAC layer. The cell discovery process is linked to the protocol in charge of activating the link between the transmitter and the receiver, so that both elements are synchronized and can communicate; this is particularly challenging when directive steering beams are used. The hidden node problem occurs when a node, said node A, is unable to detect an ongoing transmission between two other nodes, one of which is outside the transmission range of A (area inside which other nodes are able to correctly receive A's packets) but inside A's carrier sense range (area encompassing those nodes whose transmission A can perceive but not necessarily receive the transmitted packets) [RG06]; in a situation like this, A can inadvertently interfere with the ongoing transmission between the two other nodes. This problem is already challenging to deal with in microwave system where omnidirectional transmissions are used, and it is even more challenging in mmWave systems with highly directional transmissions.

In overall, the architecture of mobile networks that integrate mmWave frequencies will likely be much different than today's system based on microwave communications. An illustration of a mobile network with mmWave is shown in Figure 2.9. With the reduced inter-cell interference and high gains thanks to beamforming techniques, high spatial reuse is possible. This can be highly beneficial for backhaul links, since many nodes can share the same spectrum, allowing for a rapid deployment of mmWave networks as well as tight interconnectivity and cooperation among the BSs. Such tight integration could also occur among microwave BSs and mmWave BSs through dual connectivity, having heterogeneous deployments with the mmWave BSs in charge of boosting the network capacity and offloading the traffic from the microwave BSs.

The deployment scenarios for mmWave communications include outdoor and indoor environments, at LOS and NLOS conditions, for both the access and backhaul links; although outdoor-to-indoor communications are not feasible with mmWave, due to their difficulty to penetrate through obstacles (like building facade), outdoor mmWave deployments can still help improving the indoor coverage by freeing up resources from the microwave BSs. This resources can then be allocated to the indoor users [Qual7a]. Furthermore, mmWave outdoor deployments can be complemented with mmWave indoor deployments.

A merging of mmWave and Wi-Fi services is also possible, which would allow for high data rates in the order of Gbps for many types of devices and services, thus offering an enhanced user experience to fulfill the requirements for 5G networks.



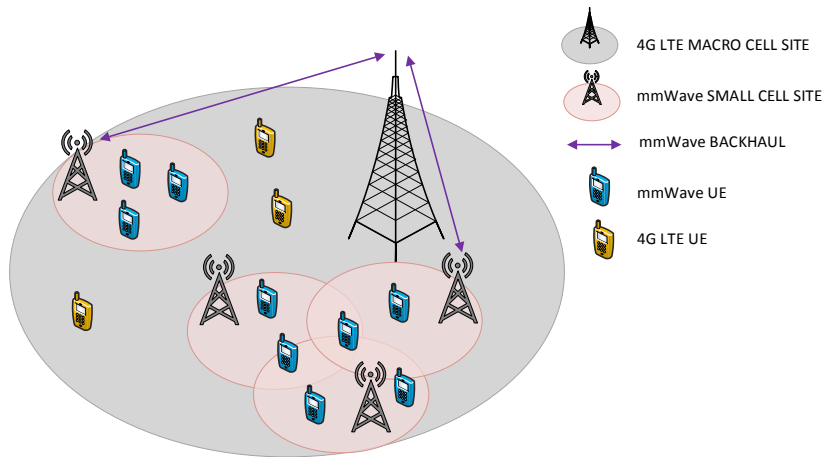


Figure 2.9: Example of a mmWave network.

### 2.3.2 Massive MIMO and beamforming

MIMO systems have gained increased attention due to their ability to enhance the spectral efficiency and improve the network capacity. In general, the more antennas the transmitter/receiver is equipped with, the more signal paths and the better the reliability of the link and the data rate. MIMO techniques were introduced since the beginnings of LTE in the 3GPP Release 8 as one of the major innovations to improve the network performance, and have since then evolved, supporting now many operation modes. Two MIMO schemes have been supported since Release 8: Single-User MIMO (SU-MIMO), where all the transmitted streams are received by one device; and multi-user MIMO (MU-MIMO) where the transmitted streams are received in parallel by more than one device using spatial separation. However, the number of antenna ports supported by MIMO have considerably changed throughout the different LTE versions.

In Release 8, MIMO was available only at the DL with a configuration of up to 4x4; the support of MIMO in the Uplink (UL) was introduced later in Release 10 (LTE Advanced) with a 4x4 configuration and DL MIMO 8x8 was also introduced. The evolution of MIMO in mobile networks continued even further with the announcement of LTE Advanced Pro in Release 13; full dimension MIMO (FD-MIMO) was introduced in this release with the support of 8, 12 and 16 antenna ports [3GPP17b] to increase the rates for MU-MIMO. Unlike conventional MIMO that only considers signal propagation in the horizontal (azimuth) plane, FD-MIMO systems use Two-Dimensional (2D) rectangular planar arrays to consider propagation in both the horizontal and the vertical (elevation) planes. To do so, FD-MIMO relies on 3D beamforming to be able to control the beam in a 3D space [Ji+17; San+16]. Multiple narrow beams can be created horizontally and vertically towards end

users with 3D beamforming, increasing the number of users served and benefiting those who are located in places that are hard to reach such as high buildings. With FD-MIMO not only is the capacity and coverage of the network drastically improved in comparison to conventional MIMO, but also larger number of antennas can be accommodated without increasing the deployment space since the antennas are placed in a square array rather than in a one-dimensional array [Ji+17]; this facilitates the deployment of large antenna arrays.

The next step with MIMO is to keep increasing the number of antennas supported to create even larger antennas arrays; this is commonly known as mMIMO. The first efforts of mMIMO were introduced in Release 14 with the support of up to 32 antenna ports [3GP17c] and improvements to the beamforming accuracy and efficiency for MU-MIMO. Further enhancements evolving to mMIMO will be part of the Release 15 and beyond, which are going to lead the way to 5G.

Hundreds of antennas are expected to be supported at the BS with mMIMO, allowing to serve many more users at the same time and with the same frequency resources. Most of the studies up to date considering MIMO implementations have been focused on microwave frequencies below 6 GHz [Boc+14]. Nevertheless, as mmWave frequencies are a promising implementation for 5G networks, the combination of mMIMO and mmWave could offer huge gains in spectral efficiency and network capacity. Furthermore, because of the extremely short wavelength, mmWave antennas are much reduced in size which makes it easier to pack hundreds of them into a small area, facilitating the implementation of mMIMO in both the BSs and the UEs.

As mentioned in Section 2.3.1, mmWave implementations will rely on highly directive beamforming to overcome propagation losses; thus, 3D beamforming can also be used for mMIMO with mmWave. This will allow for a seamless integration of all three technologies into 5G. Microwave mMIMO could then be realized in the macro BSs and in transportation systems such as trains and busses [BL16], where the space for deployment is not the biggest limitation. In Figure 2.10 an example of a network architecture combining 3D beamforming, mMIMO, and mmWave is shown.

### 2.3.3 Non-Orthogonal Multiple Access

Increasing the spectral efficiency is one of the main solutions to improve the capacity of a mobile network. Since the first generation of mobile networks, MA techniques have been used for the purpose of sharing the available resources among a large number of UEs in the most effective way. As one of the most limited resources in a mobile network is the spectrum, in an MA system different UEs get to simultaneously use the available bandwidth, thus increasing the spectral efficiency.

MA schemes can be broadly classified into two categories: OMA and NOMA [WXP06]. OMA schemes have the advantage of avoiding intra-cell interference, but they require careful cell planning to reduce inter-cell interference. The latter can be achieved by having sufficient distance between the re-used channels, which results in a low spectral

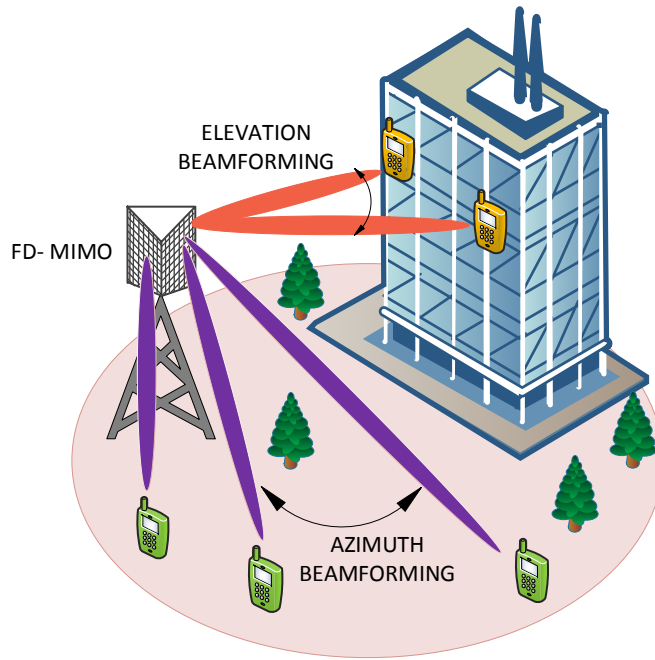


Figure 2.10: Network architecture combining mMIMO, mmWave and 3D beamforming.

efficiency. On the contrary, NOMA schemes are prone to high intra-cell interference but are robust against fading and inter-cell interference. MA techniques have been thoroughly studied in the literature, and have been implemented in mobile networks since their beginnings. 1G and 2G were dominated by OMA schemes: Frequency Time Division Multiple Access (FDMA) and Time Division Multiple Access (TDMA), respectively. However, in 2G, the cellular system IS-95 used Code Division Multiple Access (CDMA) in its asynchronous form (i.e. non-orthogonal) for the UL. The same trend continued in 3G with systems such as Code Division Multiple Access 2000 (CDMA2000) and WCDMA, where UL transmissions were also non-orthogonal.

In 4G LTE, OFDMA and Single-Carrier Frequency Division Multiple Access (SC-FDMA) were chosen for the DL and UL, respectively. The selection of these MA schemes was a key step for increasing the capacity and improving the performance in 4G LTE. Despite the significant enhancements that the MA schemes in 4G LTE offer, they might not be sufficient to cope with the expected traffic demands for 5G. Therefore, new MA schemes aiming at further increasing the spectral efficiency are highly anticipated.

NOMA has gained a lot of attention as an MA technique that can boost the capacity of 5G networks, because of its ability to increase the spectral efficiency [Din+17; Sai+13;

Yua+16]. Other benefits of using NOMA include higher cell-edge throughput, relaxed channel feedback, and low transmission latency. Furthermore, with NOMA, a good operating point where both spectrum efficiency and energy efficiency become optimum, can be achieved [Ria+17].

Two main categories of NOMA have been broadly defined in the literature: power-domain NOMA and code-domain NOMA. In the former, the signal of each multiplexed UE is separated in the power domain; in the latter, user-specific spreading codes are used to differentiate the multiplexed signals. The work in [Ria+17] presents an inside to the most relevant NOMA techniques. The work on this thesis is focused on the power-domain NOMA in the downlink, so from now on we refer to this scheme simply as NOMA.

In NOMA, besides the multiplexing in time and frequency domains, UEs are also multiplexed in the power domain, which is not sufficiently exploited in the MA schemes used in previous generations (e.g., FDMA, TDMA, CDMA, and OFDMA).

In the OMA schemes used in 4G LTE in the DL, transmissions are done with a full power but with a split bandwidth. This approach does not take advantage of the quantization rates in the system. Aiming at reducing the overhead associated with the channel state feedback between the UE and the BS, the UE reports a quantized Channel Quality Indicator (CQI) value. Each CQI corresponds to an MCS, and therefore to a discrete rate. The CQI reported by the UE depends on the Signal-to-Interference-plus-Noise Ratio (SINR) measured, and the higher the CQI the higher the channel gain, and vice versa. Because of these quantization rates, for some of the scheduled UEs, the power allocated might be enough to support the reported CQI, but not enough for the next higher CQI. In these cases, if OMA is used, a part of the allocated power can be considered as excess power and underutilization of the resources. With NOMA, however, this can be overcome.

Unlike the OMA schemes used in 4G LTE, orthogonality in the resources (e.g., frequency, time, spreading codes) is no longer needed with NOMA. The principle of NOMA is to select UEs with a high difference in their channel conditions and multiplex them in the same time/frequency resources, but with different levels of transmission power. This allows UEs with high channel conditions to access the resources assigned to UEs with poor channel conditions, hence increasing the spectral efficiency and the system capacity [Din+17]; here is where the advantage of NOMA over OMA schemes used in 4G LTE relies upon.

Figure 2.11 shows a multiplexing comparison between OMA and NOMA for four UEs. In the transmitter, signals from the multiplexed UEs are superposed and adaptive power allocation techniques are used to define the power for each UE. The power allocated depends on the channel conditions, the higher the channel gain the higher the power, and vice versa. Although power-sharing reduces the power allocated to each multiplexed UE, they benefit from being scheduled more often and having access to more bandwidth [Ben+14], as shown in Figure 2.11. In the receiver side, Successive Interference Cancellation (SIC) techniques are used to mitigate the intra-cell interference. The number of UEs that can be multiplexed in the same resources with NOMA is not restricted;

however, the intra-cell interference is proportional to the number of UEs. Moreover, the constellation of the superposed signal in the transmitter becomes more complex as the number of multiplexed UEs increases, posing great challenges on the decoding side and compromising the network performance.

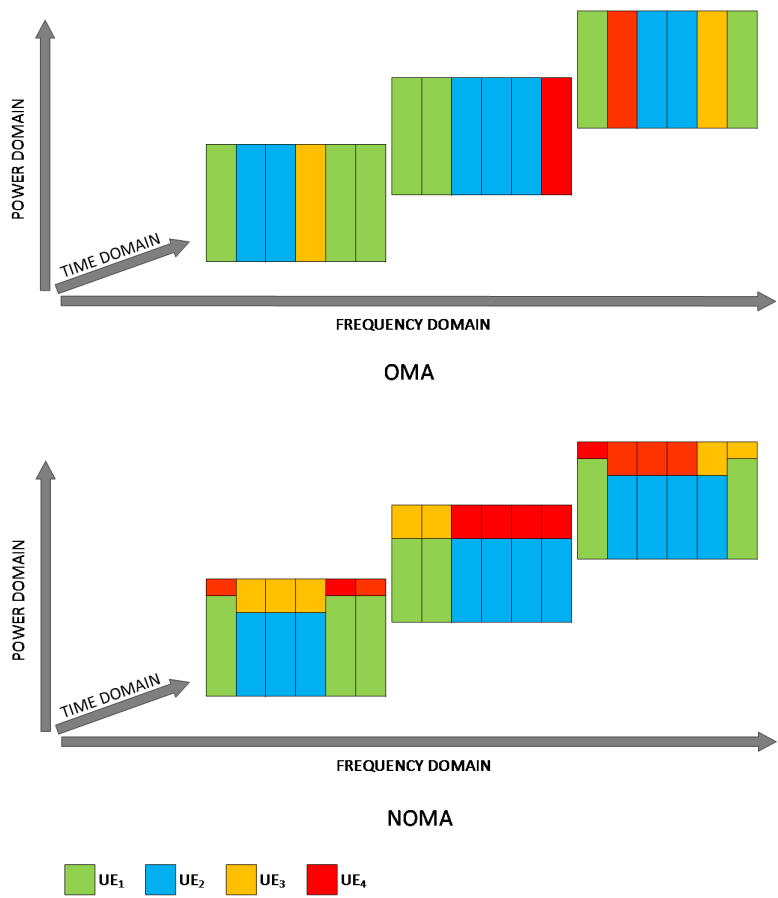


Figure 2.11: Users multiplexing differences between OMA and NOMA.

## 2.4 Chapter summary

Enhanced mobile broadband is one of the three main categories that will be targeted with the improvements of 5G NR. A flexible multi-RAT deployment with a versatility of new

technologies will characterize the future mobile network deployments, aiming at fulfilling the requirements of the IMT-2020.

The topology of 5G networks will consist of HetNets with massive deployment of small cell sites on top of the already existing macro cell sites, thus creating a two-tier architecture where the macro cells will be in charge of the data plane and the small cells will boost the network capacity. Since network densification by itself might not be enough to cope with the expected traffic demand, technologies such as mmWave, massive MIMO and 3D beamforming, and NOMA are promising solutions to increase the spectrum and use it in a more efficient way.

A seamless integration of these key technologies into ultra-dense HetNets is expected to contribute to the delivery of extreme data rates and significantly enhanced the network capacity.

## CHAPTER 3

# The theory behind NOMA

---

### 3.1 Introduction

The capacity benefits offered by NOMA has made it a promising MA scheme for 5G networks. As NOMA is one of the pillars of the work presented in this thesis, this chapter focuses on the theory behind NOMA.

In Section 3.2 superposition transmission is explained, followed by the explanation of SIC in Section 3.3. The data rates achievable with NOMA are explained in Section 3.4, whereas the advantages and challenges to consider for the implementation of NOMA are presented in Sections 3.5 and 3.6, respectively.

### 3.2 Superposition transmission

Superposition transmission is a physical layer technique, first proposed in [Cov72] that allows a single transmitter to simultaneously send a combination of independent signals to several UEs. The transmitted signal after applying superposition techniques for  $n$  UEs would be as:

$$X = \sum_{i=1}^n \sqrt{P_i} X_i \quad (3.1)$$

with

$$\sum_{i=1}^n P_i = 1 \quad (3.2)$$

where  $X_i$  is the signal corresponding to the UE <sub>$i$</sub> 's message,  $M_i$ ; and  $P_i$  is the power ratio for UE <sub>$i$</sub> . The difference between the values of  $P_i$  for each UE <sub>$i$</sub>  should be large enough to guarantee a successful decoding of the superposed signal. The waveform used for the transmissions could be based on orthogonal frequency division multiplexing (OFDM), same as in 4G LTE. In [3GP16], three superposition transmission schemes are defined for a maximum of two multiplexed UEs. In the first category, independent modulation of coded bits is performed: symbols are scaled according to the power ratio and then summed to obtain a composite constellation. This has been the category broadly used in the NOMA studies done up to date. Figure 3.1 illustrates an example of a NOMA

transmission for two UEs,  $UE_1$  and  $UE_2$ . The messages  $M_1$  and  $M_2$  are mapped to the signals  $X_1$  and  $X_2$ , respectively. These signals are then scaled according to the values of  $P_1$  and  $P_2$ , and summed to generate the superpositioned signal  $X = \sqrt{P_1}X_1 + \sqrt{P_2}X_2$  that is sent to both UEs, with  $P_1 + P_2 = 1$ . During the transmission, each signal is affected by the channel conditions of its respective receiver.

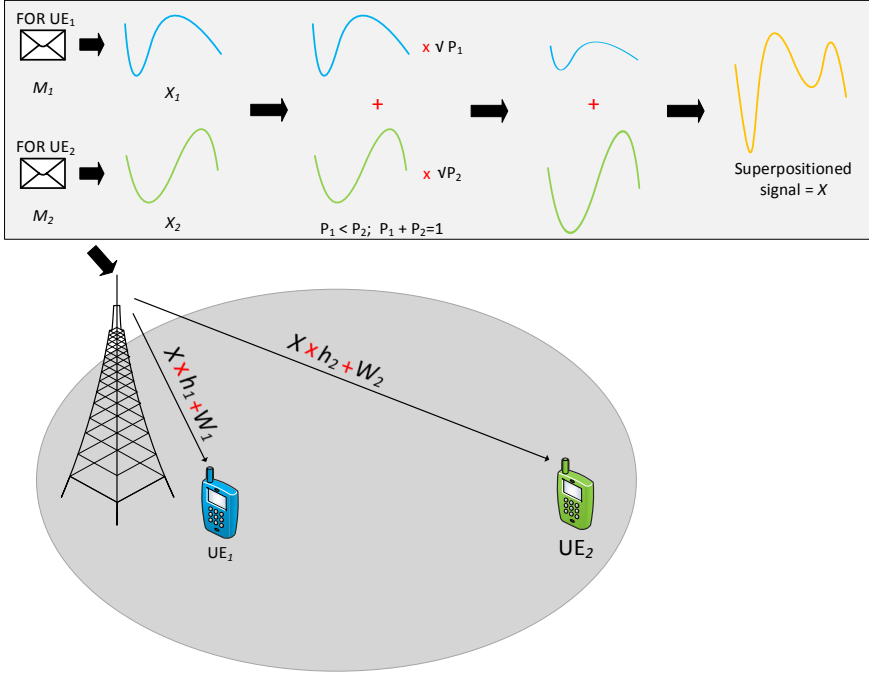


Figure 3.1: Transmission of signals in NOMA for two users.

### 3.3 Successive interference cancellation

Because of the non-orthogonality of NOMA, interference in the power domain is intentionally added in the transmitter. To mitigate this interference, SIC can be applied [Cov72]. The received signal by  $UE_i$  would be as:

$$Y_i = h_i X + W_i \quad (3.3)$$

where  $h_i$  represents the complex channel coefficients between  $UE_i$  and the BS, and  $W_i$  represents the Gaussian noise plus inter-cell interference experienced by  $UE_i$ . The optimal



order for decoding the received signal is in the order of the increasing signal strength (i.e. the channel gain normalized by the noise and inter-cell interference) [Sai+13]. Therefore, UEs are organized based on their signal strength so that any  $UE_n$  first decodes the strongest signal and removes it from the received combined signal, isolating the desired signal. To better exemplify SIC, let us assume that we have two UEs,  $UE_1$  and  $UE_2$ , and that  $UE_2$  is first in the decoding order, hence its signal is the strongest (with more power). In the  $UE_2$  receiver, the decoding will go as follows [Van+12]:

1. The message  $M_2$  is decoded from  $Y_2$ , treating  $X_1$  as noise. The interference caused by  $UE_1$  on  $UE_2$  should not significantly affect the performance of  $UE_2$ , as the power from such interference is likely to be much smaller than the desired signal. This is valid as long as an effective power allocation was performed in the transmitter.

For  $UE_1$  the decoding process is more complex and here is where SIC is applied:

1. The message  $M_2$  is decoded from  $Y_1$ , treating  $X_1$  as noise. This step is possible because of the fact that the channel gain of  $UE_1$  is higher than that of  $UE_2$ , so as long as the rate of  $UE_2$  is within the Shannon limits of its receiver, it will also be within the limits of the  $UE_1$  receiver.
2.  $X_2$  is regenerated by using an encoder, and with the knowledge of  $h_1$  and  $P_2$ ,  $h_1\sqrt{P_2}X_2$  is subtracted from  $Y_1$ , obtaining:

$$Y_1' = Y_1 - h_1\sqrt{P_2}X_2 = h_1\sqrt{P_1}X_1 + W_i \quad (3.4)$$

3. The message  $M_1$  is decoded from  $Y_1'$ .

Figure 3.2 shows an example of reception of NOMA for the same two UEs from Figure 3.1. Once the signal is received, the far UE,  $UE_2$ , simply decodes the stronger signal, whereas the near UE,  $UE_1$ , applies SIC before decoding its signal.

### 3.4 Data rates

Theoretically, it is known that NOMA offers a bigger capacity region than OMA [Ben+15; Ria+17] as shown in Figure 3.3; therefore NOMA is highly effective for increasing the system capacity when the difference in the channel gain of the multiplexed UEs is large. This difference is translated into a multiplexing gain. Assuming a successful decoding and no error propagation, the data rates with NOMA for  $UE_1$  and  $UE_2$ , can be represented by Equations 3.5 and 3.6, respectively.

$$R_1 = \beta \log_2 \left( 1 + \frac{P_1 |h_1|^2}{N_{o,1}} \right) \quad (3.5)$$

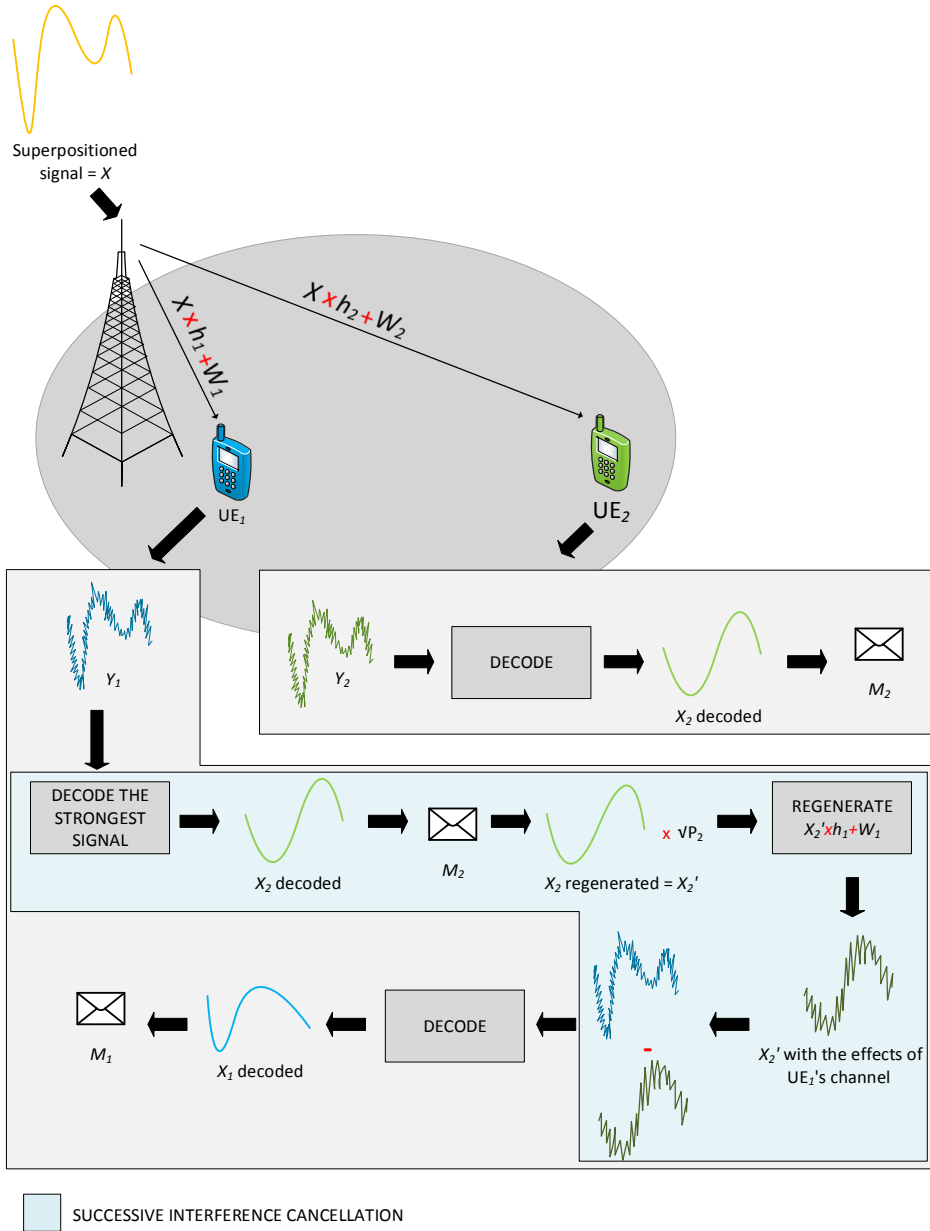


Figure 3.2: Reception of signals in NOMA for two users.

$$R_2 = \beta \log_2 \left( 1 + \frac{P_2 |h_2|^2}{P_1 |h_2|^2 + N_{o,2}} \right) \quad (3.6)$$

where  $\beta$  represents the available bandwidth, to which both UEs have access to at the same time;  $N_{o,i}$  is the power spectral density of  $W_i$ . As the values of  $R_1$  and  $R_2$  depend on the power allocation ratio  $P_1/P_2$ , the overall throughput gain of NOMA is tightly related to the power allocation scheme selected. In comparison, for an OMA transmission, the data rates of UE<sub>1</sub> and UE<sub>2</sub> are given by Equations 3.7 and 3.8, respectively:

$$R_1 = \alpha \log_2 \left( 1 + \frac{P_1 |h_1|^2}{\alpha N_{o,1}} \right) \quad (3.7)$$

$$R_2 = (1 - \alpha) \log_2 \left( 1 + \frac{P_2 |h_2|^2}{(1 - \alpha) N_{o,2}} \right) \quad (3.8)$$

where  $\alpha$  represents the bandwidth assigned to UE<sub>1</sub>, with the remaining bandwidth being assigned to UE<sub>2</sub>. The reason for the performance gain of NOMA is that the effect of the multiplicative factor (e.g.,  $\alpha$ ) outside of the logarithm in the OMA rate expressions - which reflects the splitting of the bandwidth resources among the users - is more damaging than the changes in the factors inside of the logarithm in the NOMA rate expressions [Din+17b] - which are due to the power allocation and the extra bandwidth.

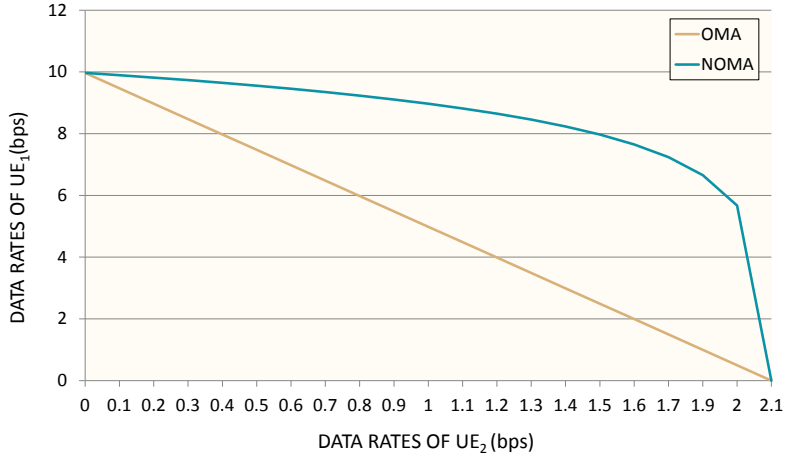


Figure 3.3: Capacity region comparison for OMA and NOMA with two UEs. The SINR for UE<sub>1</sub> is 30 dB, whereas for UE<sub>2</sub> is 5 dB.

### 3.5 NOMA advantages

The implementation of NOMA in mobile networks comes with several advantages beyond higher data rates, one of the most promising being the higher spectral efficiency. With NOMA it is possible to accommodate more UEs in the same resources than with OMA, this being particularly interesting for overloaded scenarios, which are highly expected in 5G networks. This reuse of resources allows using the available spectrum more efficiently.

Providing support for massive connectivity is another advantage of using NOMA. By implementing NOMA the number of scheduled UEs is not strictly limited by the number of available resources [Dai+15]; meaning that NOMA can allocate more UEs than OMA within the same time frame, hence increasing the number of served UEs within a subframe without requiring more resources.

Furthermore, NOMA takes advantage over the characteristic heterogeneity of the wireless channels and translates this difference into multiplexing gain [DFP16]. As mentioned previously, NOMA allows UEs with high channel gain to access the resources of UEs with low channel gain [Din+17b]; in this way, the channel conditions heterogeneity, which usually represents a challenge when managing the available resources in mobile networks, is used as an asset in NOMA.

### 3.6 NOMA challenges

Because of the loss of orthogonality and the UEs multiplexation, the implementation of NOMA comes with implementation challenges that need to be addressed when considering its integration to 5G networks.

The user pairing process is one of the aspects that poses a big challenge for NOMA; the UEs to be paired need to be carefully selected through dynamic pairing schemes to guarantee an increase in the system capacity. The achievable throughput in NOMA is also dependent on the power allocation; therefore, dynamic power allocation methods that adjust to the UEs channel conditions are also preferred to achieve higher capacity. This can increase the computational complexity.

Moreover, these algorithms for user pairing and power allocation in NOMA usually rely on highly accurate channel state information (CSI); this results challenging for high-speed environments, limiting the scenarios for which NOMA can be considered.

For the decoding of the signal, the process becomes more complex for the multiplexed UEs with high channel gains since they need to apply SIC in order to decode their message correctly. For the SIC extra information is (e.g., power allocation, SIC order), therefore an increase in the signaling is expected.

In terms of security, certain issues are raised with NOMA since the UEs with the higher channel gains have to decode the messages from the UEs with poor channel conditions. This can be, however, resolved through encryption techniques [Din+17a].

### 3.7 Hybrid multiple access

Many research works have already been done regarding NOMA and its performance, challenges, node cooperation, and user pairing. The works in [Ben+15; Dai+15; Din+17a; Ria+17; Shi+17; Yua+16] present a comprehensive approach to NOMA. Moreover, a NOMA version for the DL referred to as Multiuser Superposition Transmission (MUST) has been proposed by the 3GPP [3GP16] to be implemented in the 4G LTE networks. However, the implementation of NOMA in 5G does not mean that it will replace the OMA schemes used nowadays. Depending on the load and the UEs channel conditions, the system might decide to use either OMA or NOMA for each UE. This leads to having a hybrid MA system in 5G, where OMA and NOMA coexist [Din+17b; Ria+17]. This coexistence is in accordance with the characteristic mixture of key technologies of 5G. Figure 3.4 shows an example of UEs multiplexing in a hybrid MA system.

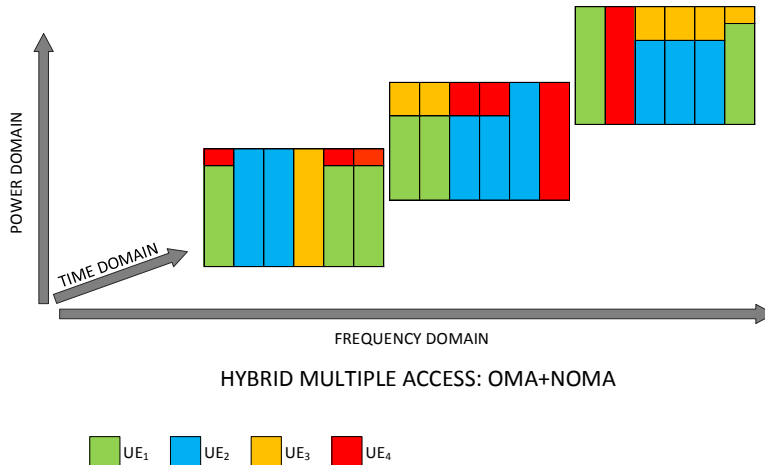


Figure 3.4: UEs multiplexing in a hybrid MA system, combining OMA and NOMA.

### 3.8 Chapter summary

NOMA has emerged as one of the key technologies to help increase the capacity of 5G networks. Higher spectral efficiency can be achieved with NOMA, in comparison to OMA, since the former allows reusing the available resources. For the transmission, NOMA multiplexes UEs with different channel gains in the same time/frequency resources and does the signal separation in the power domain. SIC techniques are applied in the receiver to decode the signals; this step adds complexity to the decoding process of the UE with higher channel gain, being this one of the NOMA implementation challenges.

Although many other challenges - related mainly to the power allocation, user pairing method, security - come with the use of NOMA, the high spectral efficiency that it offers enables the massive connectivity and high data rates. This motivates the consideration of NOMA for 5G. However, NOMA will not replace OMA; rather both schemes will coexist, leading to hybrid MA systems in 5G.

## CHAPTER 4

# 5G NR and mmWave

---

### 4.1 Introduction

With the implementation of mmWave in 5G NR, high data rates in the order of Gbps can be offered thanks to the vast amount of available spectrum at these frequencies. Several challenges come with the use of mmWave, one of the most noticeable being their short range due to high propagation losses; this makes mmWave suitable for small cell deployments. However, highly directional antennas and beamforming techniques should be used with mmWave to make their implementation feasible; thus, imposing more challenges to the mmWave deployments.

This chapter presents two studies done in relation to the implementation of mmWave frequencies in HetNets for 5G NR. First, in Section 4.2 three search methods for the cell discovery process of small cells in HetNets are analyzed; simulations were used to obtain the results. Second, in Section 4.3, a capacity dimensioning comparison can be found for two scenarios of spectrum usage in HetNets: one where the small cells operate at microwave frequencies, and other where they operate mmWave frequencies. The results for this section were obtained from a theoretical analysis. The chapter is summarized in Section 4.4.

### 4.2 Cell discovery for mmWave small cells

The use of mmWave in 5G NR comes with many implementation challenges that affect the layers of communications stack involved - that is the PHY layer and the MAC layer. Although the challenges in the former have been addressed in several studies, focusing mainly on the antenna design, circuit technology, and propagation, the challenges in the latter have not been as thoroughly dealt with. Beam steering, for example, is one of the promising techniques to be implemented for mmWave in 5G NR in order to reduce propagation losses and reach further distances; most of the functions related to the beam steering process rely on the MAC layer [Rap+14]. Other processes such as synchronization, cell discovery, and initial access also have to be adapted to operate at mmWave frequencies. As explained in Section 2.3.1, mmWave systems will rely on highly-directional large antenna arrays for maximum beamforming gain; this impacts on the cell discovery process, making it differ significantly from how it is done in nowadays mobile networks. When a user becomes available for the very first time in the network (i.e., the user turns on its radio), the cell discovery process starts, during which the UE

looks for the synchronization signals that are periodically sent by the cell to synchronize in time and frequency. Then, the UE extracts the system information (e.g., cell identity, random access procedure, frequency band) and continues to the random access procedure, in which the network registers the UE as active so it can be connected to the data plane and be able to receive and transmit actual data. In Figure 4.1 an example of the cell discovery process is illustrated.

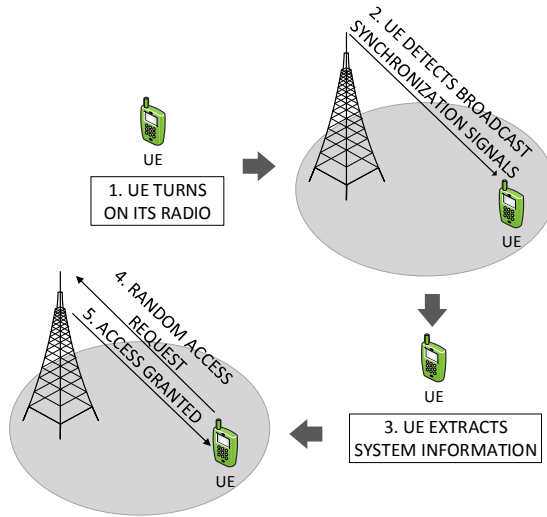


Figure 4.1: Example of the cell discovery process.

Typically, this initial search is made through omnidirectional signals that can be detected in the whole coverage area of the cell, and only after the UE has found a close-enough cell and access has been granted, beamforming techniques can be applied for the data transmissions intended for said UE.

However, in mmWave transmissions, the same principle cannot be applied. Because of the short reach of mmWaves, if beamforming is not used during the cell discovery process there will be a mismatch between the distance at which the UE can discover the cell and that at which a data connection with reasonable high rates can be established, as shown in Figure 4.2. This would lead to areas in which the UE is not capable of detecting the signals sent by the mmWave BS to synchronize and request access, although at this same distance high data rate connections could be established with the same undiscovered cell. Therefore, directional beamforming should be applied during the cell discovery process in the mmWave cells. The challenge then is selecting the transmission parameters and the initial cell search method that will offer the best performance, considering that with directional synchronization signals and initial access procedure, the cell has to perform a spatial search that will cause extra delays.



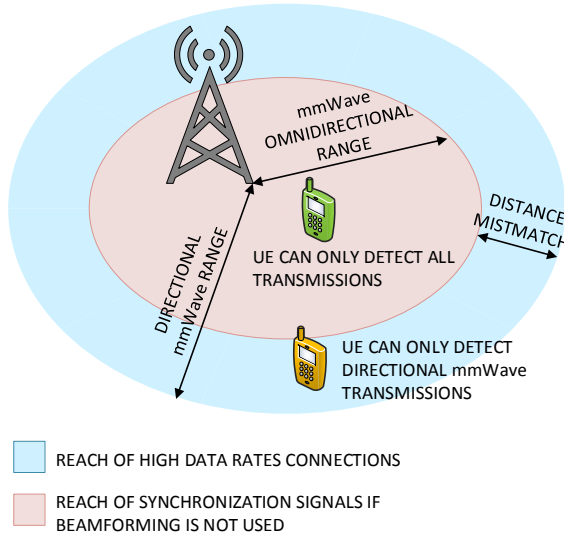


Figure 4.2: Example of the different ranges for omnidirectional and directional mmWave transmissions.

#### 4.2.1 State of the art

Up to the date, there are a few studies that have focused on the cell discovery process for mmWave. The work in [Sho+15] presents a two-step synchronization and initial access process from a MAC layer perspective for a HetNet with mmWave. During the first step, time-frequency synchronization is achieved for all the entities in the macro cell, including the small cells and the UEs. During the second step, the small cell performs a periodic spatial search using directional mmWave pilot transmissions; upon receiving a pilot, the UEs, continue with the initial access process. The results showed that a relatively small number of pilot transmissions guarantees discovery of a UE with high probability, and that this number increases with the directionality level, introducing a trade-off between boosting link budget and reducing synchronization overhead.

The authors in [Bar+15] focus on the cell discovery process for mmWave and consider two cases: a periodic transmission of synchronization signals in random directions to scan the angular space and an omnidirectional transmission of synchronization signals. The results show that the latter performs much better than the former by offering a lower misdetection probability. Sequential beam searching is suggested for future research. However, a real implementation of an omnidirectional mmWave transmission will require higher gain than a directional transmission; thus, some of the benefits of the omnidirectional search might be reduced. A hierarchical search for mmWave is proposed in [Des+14], where the UE can simultaneously process several narrow beams from the BS during a

set number of transmission stages; for each stage, the UE can process narrower signals. The results show that this search strategy can provide fast discovery but refinements are needed to improve its performance.

In [Pra+16], a mmWave discovery mechanism is proposed focusing on potential energy savings in UDNs rather than the reliability of the search method. A hybrid search, combining strengths of the exhaustive (i.e., high probability of detection) and the iterative (i.e., low delay) search methods, is proposed in [Hab+17]; the results show that a good balance between misdetection probability and discovery delay can be achieved with the proposed method. Finally, in [PZ17] a cell discovery method based on historical location of UEs is proposed; a learning algorithm is used to decrease the delay of finding UEs by autonomously prioritizing those sectors where more users are expected to be found according to previous experience. The results show an increment of the number of UEs found in the first scan by over 19% and a delay reduction by over 84% in comparison to the exhaustive search.

All of these referenced works assume that the search of the UE is only performed by the mmWave BS. Search methods that take advantage of a C/U plane split architecture, where the context information related to the UE location can be provided by the macro cell through the control plane, thus helping in the cell discovery process and minimizing the associated delay, are scarce. Although this solution is suggested in [Deh+14; Li+13], its performance is not evaluated. In [CFS15] two cell discovery methods for mmWave small cells are proposed, both relying on user location information provided by the macro cell; their performance is evaluated for different user distributions. The results show that the proposed methods outperform the random search.

In our work published in Paper B, the performance evaluation of a cell discovery process where the macro and the small cell work together to enhance the search process in mmWave small cells is presented. Furthermore, the proposed method is compared with two generic methods based on exhaustive search and binary search [Gio+16]. An exhaustive search for mmWave small cells is suggested in [JPY15] for the initial access procedure, whereas a performance comparison between the exhaustive search and the iterative search (similar to the binary search) is presented in [Gio+16]. An expansion of the latter including context information-based search method is presented in [GMZ16] along with a survey of several proposed techniques for the initial access procedure in mmWave cells; the analysis is done based on misdetection probability and discovery delay for a dense, urban, multi-path scenario.

#### 4.2.2 Joint search method

Because the mmWave small cells in 5G NR will operate tightly integrated with the macro cells, the cell discovery process for the former can take advantage of the ubiquitous coverage of the latter. Since all the UEs in the network will have a connection to the macro cell, which will be in charge of the control plane, an approach where the macro cell helps in the small cell discovery process is fairly valid. In this regard, the performance

of a macro cell assisted search method, to which we will refer to as JSM, is analyzed; as mentioned above, this method was introduced in our work published in Paper B.

For the JSM, it is assumed that the UE continuously reports to the macro cell the channel conditions, and through these reports, the macro cell can keep track of the location of the UE by estimating the Direction of Arrival (DoA) of the signal and by using triangulation processes. When a UE starts approaching one or more small cells, the macro BS broadcasts the location of the UE to the corresponding small BSs. Upon receiving this signal, each small BS selects a search area of predefined width (much smaller than  $360^\circ$ ) and divides it into several sectors depending on the beamwidth  $\theta$  value configured. The selected search area is centered around the location of the UE; for example, if the location corresponds to an angle of  $20^\circ$  (e.g., with respect to the magnetic north) and the width of the search area is set to  $10^\circ$ , then said area starts at  $15^\circ$  and finishes at  $25^\circ$ . A pilot signal is then sent to the middle sector (most likely where the UE is located); if the UE is not found there, pilot signals are randomly sent to the remaining sectors. If the UE is not found in any sector, the small BS assumes that it is unreachable; if after a predefined time window the UE is not connected to a small cell, the macro BS starts the process all over again of reporting the UE location to the small BS.

In Figure 4.3 the workflow for the JSM is shown whereas in Figure 4.4 illustrates the principle of the JSM with one macro cell, one small cell, and two new UEs,  $UE_1$  and  $UE_2$ ; the width of the search area is assumed to be  $45^\circ$  and  $\theta = 15^\circ$ . As it is assumed that the width of the search area is three times the beamwidth, such area is divided into three sectors. For  $UE_1$  only one pilot transmission was necessary. For  $UE_2$  the location estimation done by the macro cell was not accurate enough, so the small cell centered the search area to the erroneous location; this caused an extra pilot transmission since  $UE_2$  was not located in the middle sector. Two pilot transmissions were required to find  $UE_2$ .

With this method, the cell discovery process can be sped up significantly by reducing the number of pilot transmissions needed to find the UE. However, the biggest disadvantage of this method is that it requires highly precise location estimations by the macro cell to prevent the small cell from sending too many pilots or even not finding the UE. Moreover, highly precise synchronization between all the devices within the coverage area of the macro cell, as well as a highly precise channel estimation mechanisms to select the proper beamforming are needed; both requirements can be very challenging to achieve in mmWave communications. Nevertheless, this method allows using directive beamforming throughout all the search process, overcoming the coverage mismatch between omnidirectional and directional transmissions in mmWave; furthermore, it also helps to decrease the discovery delay, and thus the initial access delay, for the UEs in comparison to other generic search methods, like the exhaustive.

### 4.2.3 Generic search methods

To compare the performance of the JSM, we rely on two generic searching methods for mmWave that will be used as benchmarks: exhaustive search and binary search; in both



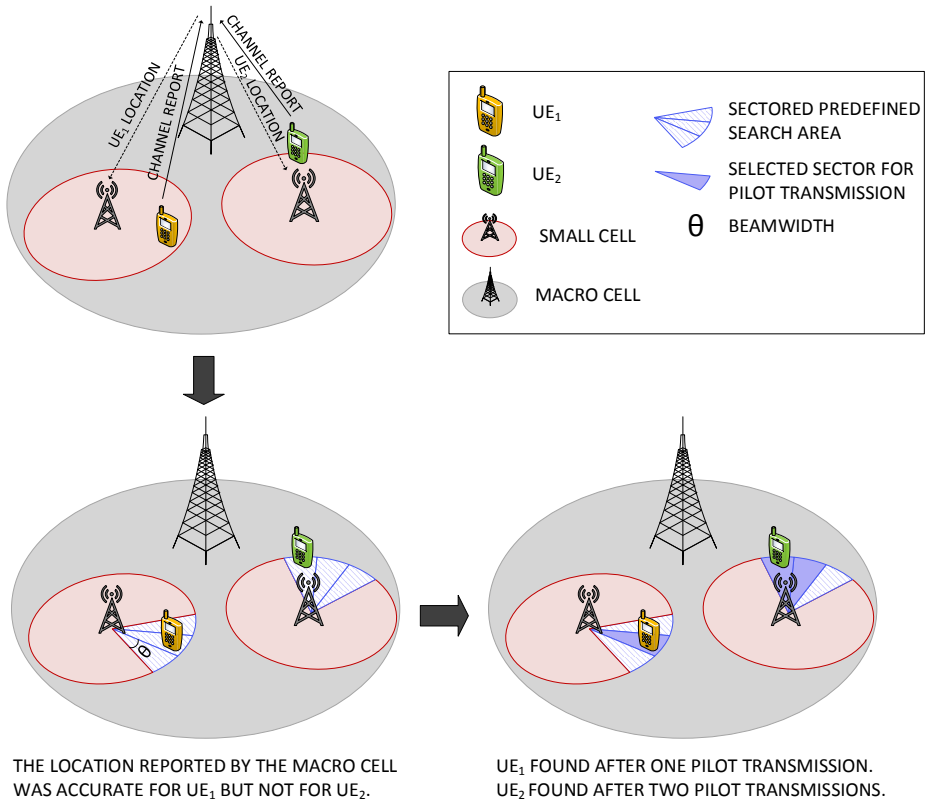


Figure 4.4: Example of the joint search method for the cell discovery process in mmWave cells.

methods the search is done only by the small cells and beamforming is used.

### Exhaustive search method (ESM)

The principle behind this method, as its name indicates, consists on exhaustively sent pilot signals to all the sectors to cover the whole coverage area and allow all the UEs to discover the cell. The  $360^\circ$  coverage area is divided into  $N$  sectors according to the beamwidth  $\theta$  defined; that is  $N = \frac{360^\circ}{\theta}$ . Pilot signals are then sent randomly to each sector direction until all sectors are covered. Once the BS sends a pilot signal, it waits for a predefined time window for a response from a UE before continuing the search in the other sectors. Upon receiving one of the pilot signals, the UE synchronizes with the cell and responds in the same DoA to proceed with the initial access and the request for a data

connection. At this point, both the BS and the UE know the best direction to reach each other.

Although this method allows using directive beamforming for all the pilot signals, thus overcoming the range mismatch problem mentioned above, the high discovery delay that can be experienced becomes an issue. In the worst case, a UE will have to wait until all sectors are covered before receiving the synchronization signals; which means that some UEs will be found considerably faster than others. This problem worsens as  $\theta$  decreases and, hence,  $N$  increases. Figure 4.5 illustrates the steps for the exhaustive search for  $N=8$  and a new UE. In this example, the UE is found after five pilot transmissions.

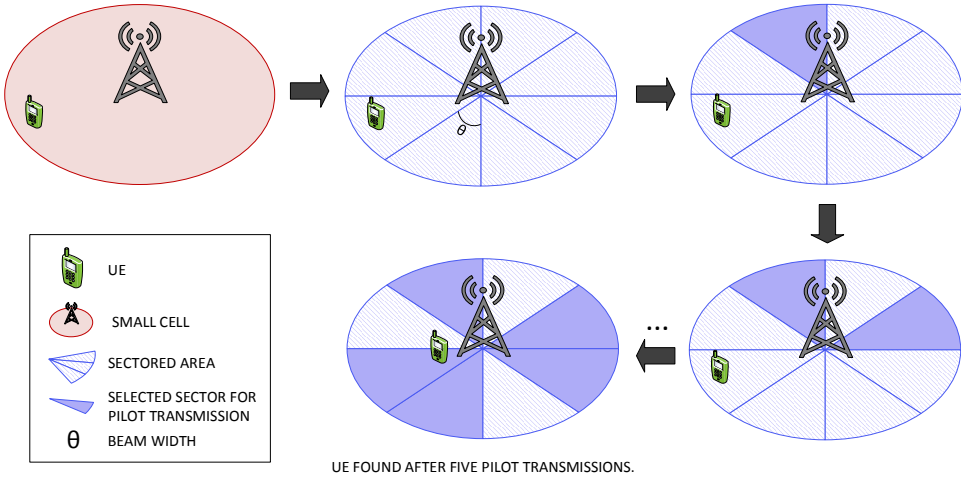


Figure 4.5: Example of the exhaustive search method for the cell discovery process in mmWave cells

### Binary search method (BSM)

This BSM is based on the iterative binary division of a spatial area into sectors, and it consists of two steps, as shown in Figure 4.6. In the first step, the  $360^\circ$  area is divided in each iteration  $i$  into  $N$  sectors; pilot signals are randomly sent to each sector. In each iteration, the value of  $N$  is increased by two, for a total of  $N = 2^i$  sectors per iteration. If no UE is found during an iteration after the BS waits for a response during a predefined time window, the BS continues to the next iteration and modifies the beamwidth to send the pilot signals to narrower sectors; this goes on until reaching the maximum number of iterations  $i_{max}$  with the minimum beamwidth  $\theta_{min}$ . However, if during one of the iterations a UE is found, for example, in sector  $N_j$ , with  $j \leq i_{max}$ , the BS continues to the next step. In the second step, the BS proceeds to refine the search by repeating the binary sectoring

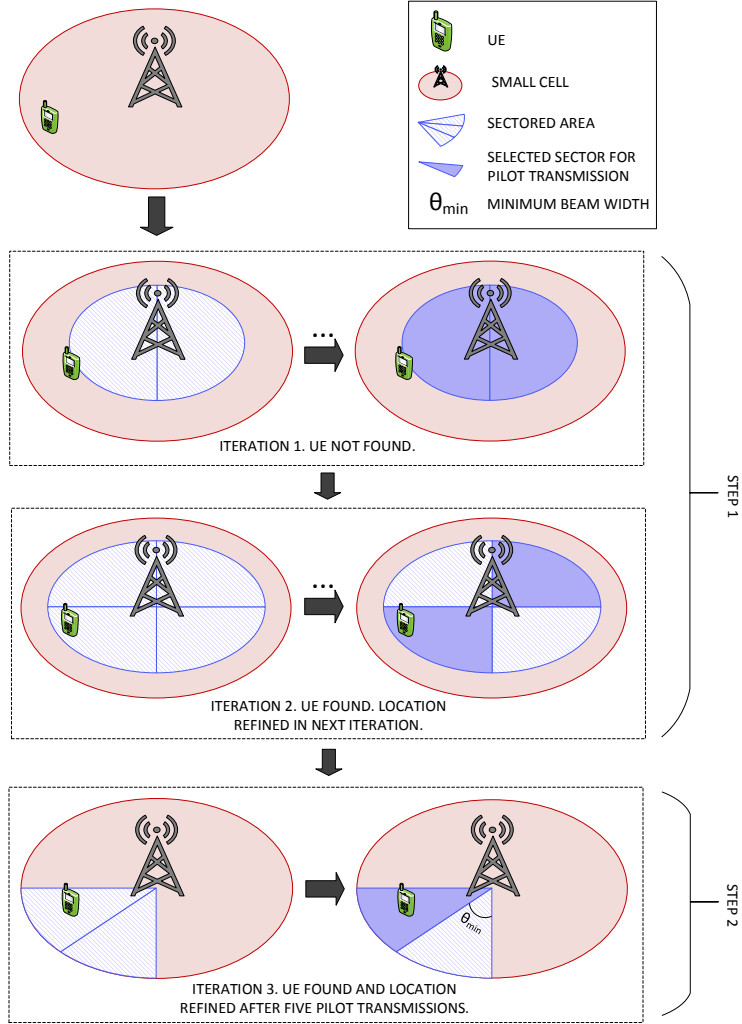


Figure 4.6: Example of the binary search method for the cell discovery process in mmWave cells

only in sector  $N_j$ ; pilot signals are then randomly sent to the sectors until the UE is found again. This step is repeated every time the UE is found until the minimum beamwidth  $\theta_{min}$  is reached. This refined search is done to improve the direction between the BS and the UE, so they can directionally reach each other in the best possible way. After the second

step, the BS goes back to the first step, continuing from the sector  $N_{j+1}$ , with  $j+1 \leq i_{max}$ .

With this method, the cells go through a “breathing” process by constantly changing its coverage range; the wider the beamwidth the smaller the coverage area of the cell and vice versa. This results beneficial for the UEs that are located close to the BS, since they can be found with a wide beam (e.g.,  $180^\circ$  or  $90^\circ$ ); therefore, fewer pilot transmissions are needed to determine the best direction to communicate with said UEs. On the contrary, for the UEs that are far from the BS, the latter might have to go through several processes of dividing the spatial area until one of such UEs is found, thus generating high discovery delays. The steps for the binary method are shown in Figure 4.6 for a new UE, a maximum of  $N=8$  for step 1 and  $\theta_{min} = 45^\circ$ . The UE in this example is found after five pilot transmissions.

#### 4.2.4 Network model

For comparing the performance of the search methods introduced in Sections 4.2.2 and 4.2.3 a HetNet was modeled consisting of a macro cell site with ten small cell sites deployed inside its coverage area. OPNET Modeler (now Riverbed Modeler [Riv]) was used for the simulations; OPNET is a Discrete Event Simulation (DES) modeler used for networking algorithms and protocols. In Figure 4.7 the network model is shown. As illustrated in Figure 4.7a the connection links between the network elements are considered wireless, however in the OPNET model in Figure 4.7b said wireless links were modeled as wired links for simplicity reasons; nevertheless, the wireless propagation characteristics such as the beamwidth and the area covered by each sector are considered in the model so that the model approximates to the real scenarios.

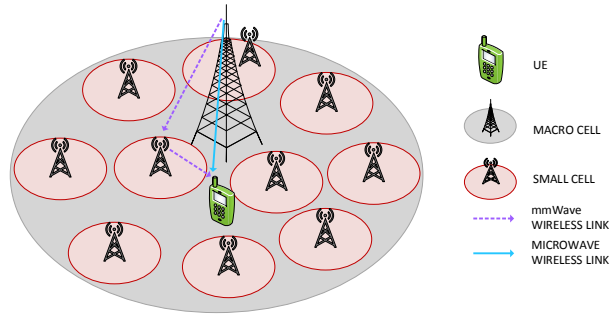
The coverage area of the cells is modeled as hexagons, and all cells are assumed to be in LOS. For the macro cell, microwave frequencies are considered and a maximum reach of 650 m is set. For the small cells, the 60 GHz mmWave band is used and it is assumed that the wireless link is established in a semi-directional communication mode, with the small BS using directional transmission/reception and the UE using omnidirectional. With this setup, and following the research work in [Sho+15], the maximum reach of the mmWave small cells,  $d_{max}$ , can be estimated through a simple distance-dependent propagation model as:

$$d_{max} = \frac{\lambda}{4\pi} \left( \frac{2\pi p}{\sigma\beta\theta} \right)^{1/\alpha} \quad (4.1)$$

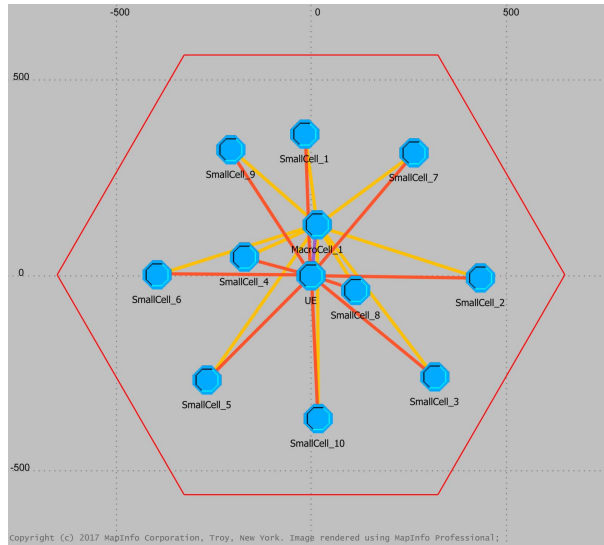
where  $\lambda$  is the wavelength,  $p$  is the transmission power of the small BS,  $\sigma$  represents the noise power,  $\beta$  is the UE receiver sensitivity,  $\theta$  is the beamwidth and  $\alpha$  is the path-loss exponent. In Table 4.1 the propagation parameters used in the model are summarized.

The five beamwidths configured corresponded to those used for a total of six iterations with the BSM, and for comparison reason the same beamwidths were used for the ESM and JSM. With the propagation parameters selected, the maximum reach for the small cells is 200 m. For the UE location, uniformly random distributed positions were used for a total of 1000 positions to cover the whole area of the macro cell; the three search





(a) Illustration



(b) OPNET model

Figure 4.7: Network model for performance evaluation of cell discovery methods in mmWave small cells

methods were evaluated for the same UE positions. If the UE was not located inside the coverage area of a small cell, a new position was used. In Figure 4.8 the percentage of the macro cell area covered by the small cells as a function of the number of deployed small cells is shown for five beamwidth values, whereas in Figure 4.9 the probability of the UE connecting to a small cell is shown for ten small cells; for the latter, 500 UE positions were considered for obtaining the results. It can be seen that the wider the beamwidth, the lower the probability that a UE will be inside the coverage area of a small cell since

Table 4.1: Propagation parameters for the performance evaluation of cell discovery methods in mmWave small cells.

CARRIER FREQUENCY (GHz)	60
WAVELENGTH (mm)	0.005
TRANSMISSION POWER (dBm)	20
TRANSMITTER ANTENNA GAIN (dBi)	15
RECEIVER ANTENNA GAIN (dBi)	0
NOISE POWER (dBm)	-127
BEAMWIDTH (°)	5.625, 11.5, 22.5, 46, 90
PATH-LOSS EXPONENT	3
RECEIVER SENSITIVITY (dBm)	10

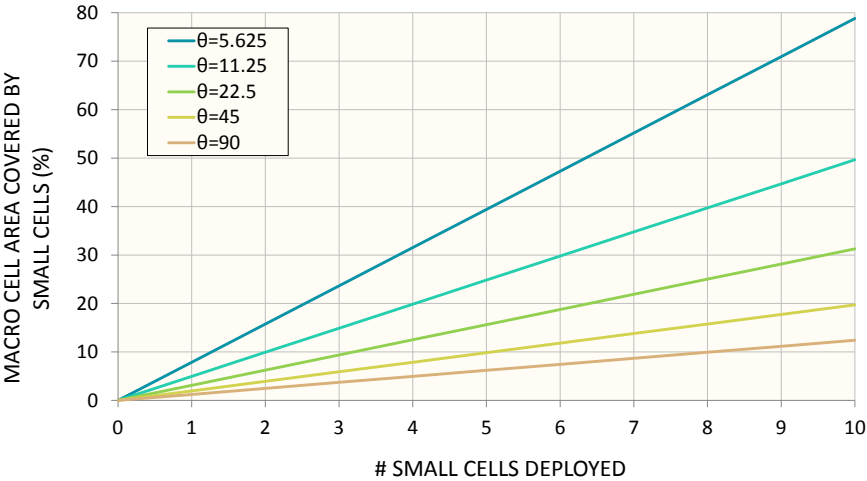


Figure 4.8: Percentage of the macro cell covered by the small cells.

less of the macro cell area is covered by the small cells. For the JSM, an erroneous UE location estimation by the macro cell was introduced; it was modeled as two random variables,  $x$  and  $y$ , following both a normal distribution with mean  $\mu = 20\text{m}$  and standard deviation  $\sigma = 5\text{m}$ . The work in [Med+09] was used as a reference for selecting the values of the distribution. The value of  $x$  represents the position error in the x-axis, and similarly  $y$  is the location error in the y-axis. The introduction of this error directly affects the performance of the JSM and helps establish more realistic conditions for the simulations.

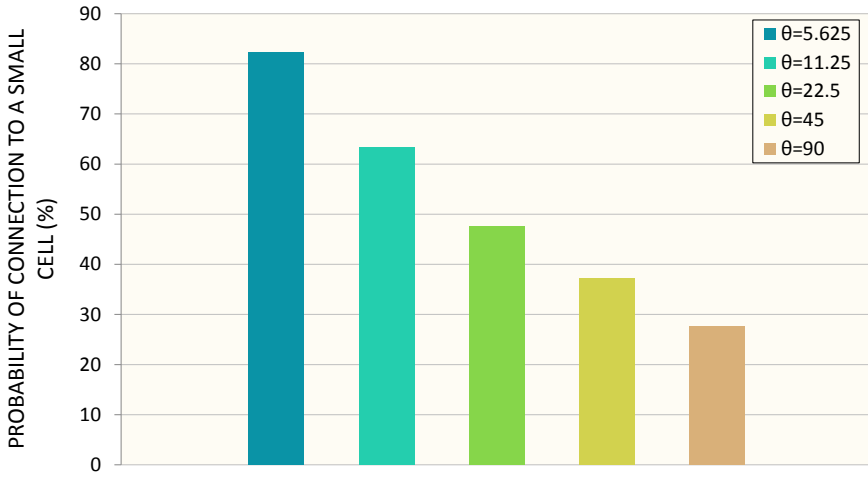


Figure 4.9: Probability of connection to a small cell.

#### 4.2.5 Network performance

For the performance analysis, the average number of pilot transmissions needed for each search method was considered; since the UE could be in the overlapping area of up to two small cells at the same time, the analysis was done for one and two small cells covering the UE. In Figure 4.10 the results for the ESM are shown; here the relationship between the number of pilot transmissions and the beamwidth can be clearly seen. An average of 32.5 pilot transmissions per small cell were experienced for the smallest beamwidth  $\theta = 5.625$  and with one small cell covering the UE; for the same beamwidth, 21 pilot transmissions per cell were needed on average to find the UE. This behavior is due to the fact that more than one cell participates in the initial search process, and since sectors are chosen randomly, there is a higher chance of finding the UE faster. The influence of the number of cells covering the UE on the number of pilot transmissions needed is less noticeable as the beamwidth becomes wider. The reason for this is that the area covered by each beam also becomes wider, making it easier to find the UE; thus, cell cooperation

for finding the UE is no longer needed. For a beamwidth  $\theta = 90^\circ$ , 2.4 pilot transmissions were needed with one small cell, and 1.3 were needed with two small cells. Even when the difference in the number of pilot transmissions considerably decreases as the beamwidth gets wider with 92-94% less transmissions needed, it is important to consider how likely it is for a UE to be connected to a small cell for wide values of beamwidth (Figure 4.9). For the wider beamwidth, i.e.,  $90^\circ$ , a UE has approximately 28% probability of connecting to a small cell, versus 82% probability for a beamwidth of  $5.625^\circ$ . Moreover, it must be considered that as the small cells are in charge of boosting the system capacity and helping offloading the macro cell, a connection to the small cells will be preferred when possible.

The same trend was experienced with the BSM, as can be seen in Figure 4.11. For all cases the pilot transmissions needed with the BSM are higher than with the ESM. The reason for this is that as the UEs locate farther away from the BS, more iterations are needed to find the UE and hence, more pilot transmissions are needed. Furthermore, the number of UEs located close to the BS is lower than those located on the outside, because of the cell geometry. An average of 47 pilot transmissions were needed for a beamwidth of  $5.625^\circ$  and with one cell, whereas 26 pilot transmissions were needed per cell with two cells. With a  $90^\circ$  beamwidth, an average of 3.2 and 1.6 pilot transmissions were experienced per cell for one and two cells, respectively. A decrease of up to 93-94% in the number of pilot transmissions was experienced by making the beamwidth wider.

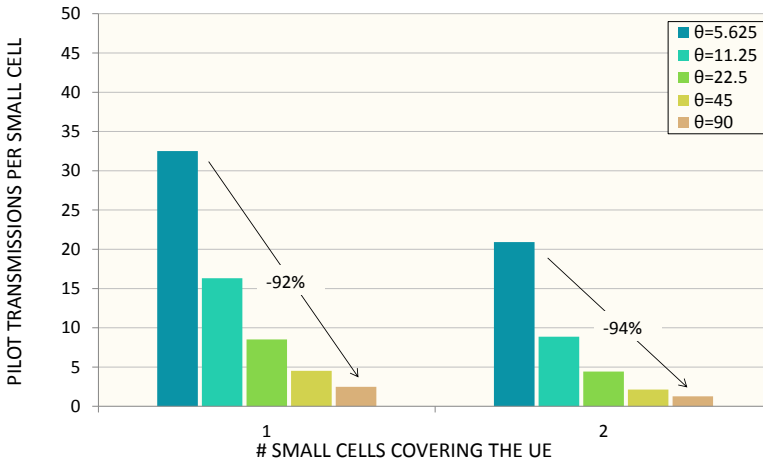


Figure 4.10: Pilot transmission needed to discover the UE for the ESM.

The performance of the JSM was significantly better than for the ESM and BSM, even considering the location estimation error introduced. In Figure 4.12 the results are shown. Three widths for the search area were defined: 1xbeamwidth, 3xbeamwidth, and

6xbeamwidth. This was done to evaluate the influence of said width on the number of pilot transmissions.

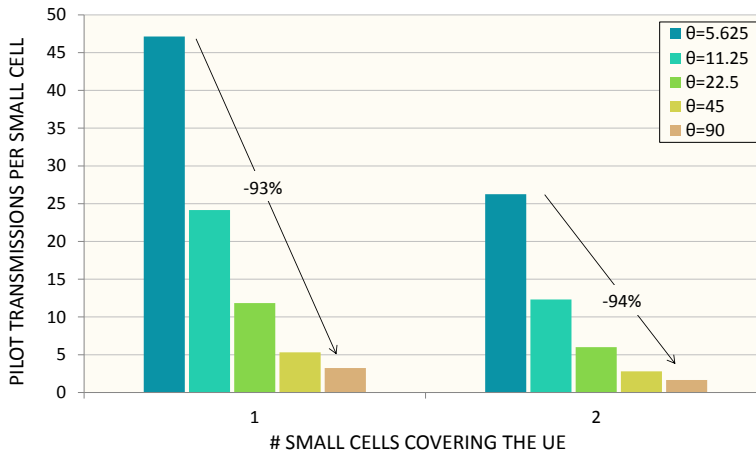
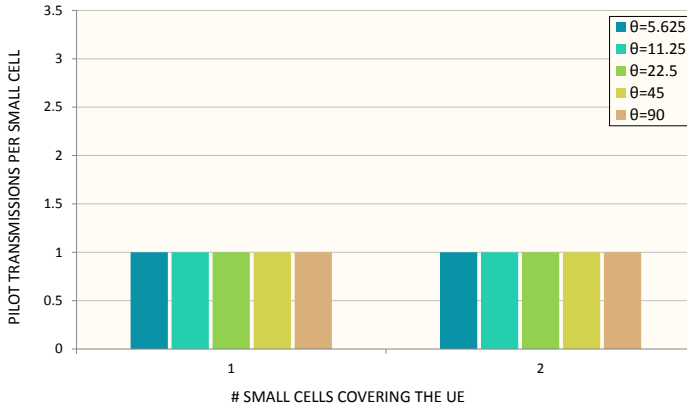


Figure 4.11: Pilot transmission needed to discover the UE for the BSM.

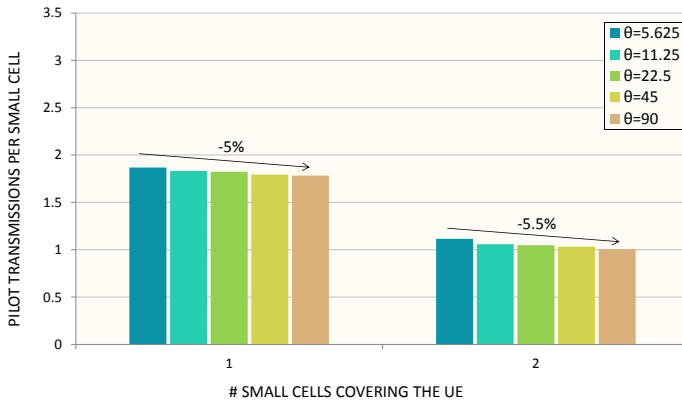
As expected, when the width of the search area is set to 1xbeamwidth, only one pilot transmission is needed to find the UE regardless of the beamwidth and the number of small cells covering the UE as can be seen in Figure 4.12a; for these results it is important to remember that there is only one sector where to look for it and that if not found, the JSM process starts all over again (Figure 4.3). For wider search areas, such as 3xbeamwidth and 6xbeamwidth, the average number of pilot transmission do vary depending on the beamwidth; however, the variation is very subtle in comparison to the ESM and BSM.

With a search area of 3xbeamwidth, an average of 1.9 and 1.1 pilot transmissions were needed for one and two small cells, respectively, when the narrowest beamwidth is selected as shown in Figure 4.12b; for the widest beamwidth, 1.78 and 1 pilot transmissions were needed with one and two small cells, respectively. The gain in number of pilot transmissions due to the beamwidth was between 5-5.5%, which is considerably lower than for the other search methods. With such low variation in the number of pilot transmissions, and hence in the discovery delay, the JSM offers a stable reference for such delay regardless of the beamwidth value selected; this is particularly important for network dimensioning purposes.

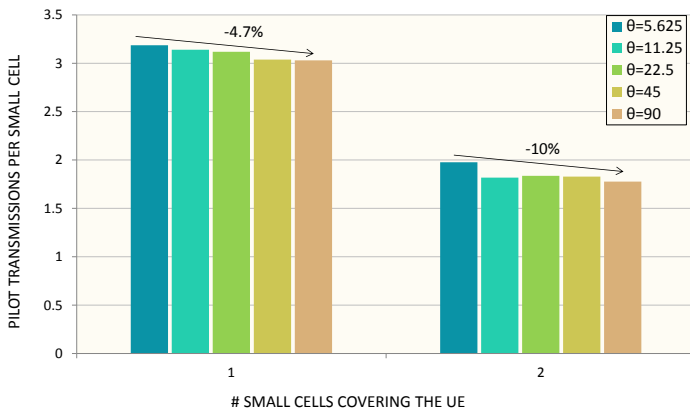
For a search area width of 6xbeamwidth, the results are shown in Figure 4.12c. For this case, higher number of pilot transmission are needed since more sectors where to look for the UE are defined. The variation in the number of pilot transmissions is higher for the case with two cells, with a 10% difference, whereas a 4.7% difference was experienced when only one small cell covers the UE. For the narrowest beamwidth, 3.2



(a) Search area width = 1xbeamwidth



(b) Search area width = 3xbeamwidth



(c) Search area width = 6xbeamwidth

Figure 4.12: Pilot transmission needed to discover the UE for the JSM for three values of search area width.

and 1.9 pilot transmissions were needed on average; for the widest beamwidth, 3 and 1.8 pilot transmissions were needed. As it can be seen from these results for the JSM, the selection of the width for the search area has a high influence on the performance of the cell discovery process; although at a first glance it seems like choosing a narrow search area, e.g., of the same width as the beamwidth, is the best solution, that will require the assumption that the location of the UE, sent to the small cell by the macro cell, is accurate enough to guarantee that the UE will indeed be located in such narrow search area. If this is not a case, a high probability of misdetection can be expected as the search area becomes narrower.

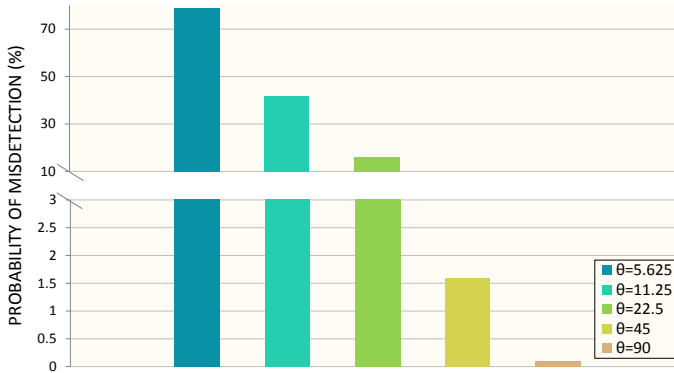
In Figure 4.13 this misdetection probability is shown for the JSM for the three widths of search area considered. For a search area of 1xbeamwidth (Figure 4.13a and for the lowest beamwidth of  $5.625^\circ$ , a 79% probability of misdetection was experienced, leaving this a highly low probability that the small cell is able to find the UE, since only 21% of the times the UE is located inside the sector where the small cell is sending the pilot signals. Therefore, is such a narrow search area is to be used, then a highly accurate location estimation mechanism must be implemented in the macro cell to avoid high discovery delays. Using wider beamwidth for this case will considerably decrease the misdetection probability, with values as low as 1.6% and 0.03% for beamwidths of  $45^\circ$  and  $90^\circ$ , respectively. However, this will solve the discovery delay issue but the UE will have a low probability of connecting to a small cell; thus using search areas with a width equal to the beamwidth does not offer a good performance for the JSM.

The use of search areas wider than the beamwidth should then be preferred. It can be seen from Figures 4.13b and 4.13c that the probability of misdetection is significantly reduced for the cases where the search area is 3xbeamwidth and 6xbeamwidth. For the former, only a 1.97% misdetection probability was experienced for the narrowest beamwidth of  $5.625^\circ$ ; for the other beamwidth values, said probability was below 1%. When the search area was 6xbeamwidth, the highest misdetection probability was 1.7% for a beamwidth of  $5.625^\circ$ .

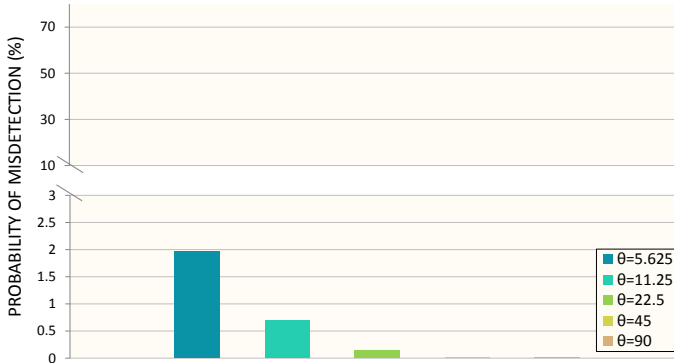
Since the difference between the misdetection probability between search areas of 3xbeamwidth and 6xbeamwidth is on average only 0.16%, but with a wider search area more pilot transmissions are needed to find the UE, then a search area of 3xbeamwidth should be the preferred configuration for the JSM. This would allow keeping a low cell discovery delay while finding the UE in approximately 98% of the cases, even for the cases where a narrow beamwidth such a  $5.625^\circ$  is used.

To estimate the improvement of the JSM over the ESM and the BSM, Figure 4.14 shows the gain of the JSM in terms of number of pilot transmission needed for the five values of beamwidth defined. A search area width of 3xbeamwidth was selected for JSM for this analysis, since as explained above is the option that offers the best trade-off between discovery delay and misdetection probability; the ESM and BSM were used as benchmarks.

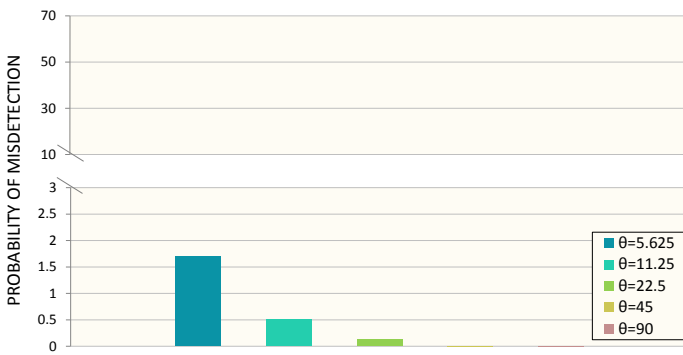
The benefits of implementing the JSM are clear from these results, where it can be seen that up to 94% and 96% of improvement in the pilot transmissions needed can be



(a) Search area width = 1xbeamwidth



(b) Search area width = 3xbeamwidth



(c) Search area width = 6xbeamwidth

Figure 4.13: Misdetection probability for the JSM for three values of search area width.



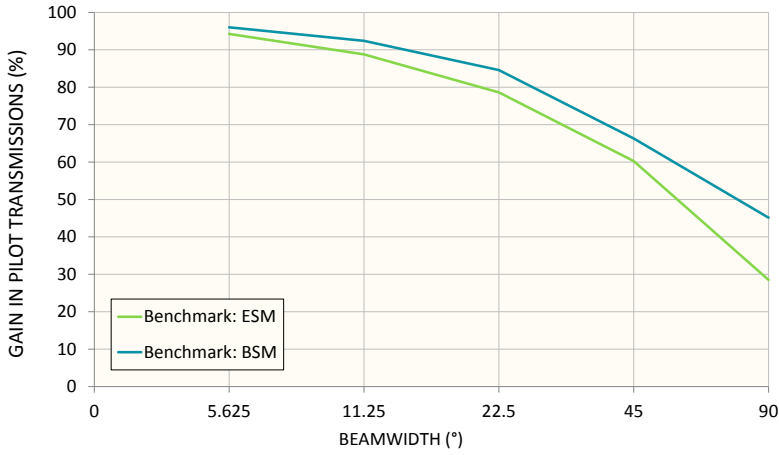


Figure 4.14: Gain in the pilot transmissions needed offered by the JSM, using the ESM and BSM as benchmarks.

achieved over the ESM and BSM, respectively. This improvement gets lower as the beamwidth gets wider; nevertheless, for a beamwidth of  $90^\circ$  28.5% and 45% fewer pilot transmissions are needed with the JSM in comparison to the ESM and BSM, respectively. This results directly affect the dimensioning of a HetNet as the one modeled, since the selection of the beamwidth to be used will not only determine the discovery delay, but also the coverage of the small cells. Therefore, depending on how many small cells can be deployed and the transmission parameters, the right beamwidth can be selected to offer a low discovery delay, a high probability of being discovered by the small cell, and a high probability of being inside the coverage area of a small cell. If a high number of small cells can be deployed, then wider beamwidths can be used along with the JSM with a search area equal to the beamwidth, since with this configuration the precision of the location estimation done by the macro cell does not need to be highly accurate; furthermore, the beamforming can be done in a more relaxed way since no highly precise alignment is needed between the small cell and the UE. On the contrary, if the number of small cells to be deployed is a limitation, then the use of narrow beams along with the JSM with a search area wider than the beamwidth is preferred. This will allow covering a higher percentage of the macro cell with a low number of small cells, while maintaining a low discovery delay, a low misdetection probability, and a high probability of the UE connecting to a small cell.

As the JSM outperforms both the ESM and the BSM, cell search methods for HetNets with mmWave where the macro cells and the small cells cooperate are recommended. This kind of approach would speed up the search process and reduce the initial access delay, while maintaining the directional transmissions that characterize mmWave. Further

work related to aspects such as the overhead caused by the extra information related to the UE location is recommended since it would offer an interesting insight at the trade-off between achieving high data rates and reducing said overhead.

### 4.3 Network deployment: microwave vs mmWave

As previously mentioned, one of the main advantages offered by the mmWave frequency bands is the large amount of available spectrum; up to 10 GHz of free spectrum can be found at these frequencies, allowing to reach high data rates with peak values of 10-25 Gbps [Wei+14]. This brings significant improvements when it comes to the network dimensioning, since either more capacity can be offered with the same number of deployed cells, or fewer cells are needed for the same capacity, in comparison to the use of microwave frequencies. The implementation of mmWave in 5G NR, has been considered for both the access and backhaul links, remembering that in a correctly planned network, the backhaul needs to be properly dimensioned to avoid bottlenecks.

Since one of the major challenges in HetNets is the inter-cell interference for in-band deployments, if the typical microwave frequencies (i.e., <6 GHz) are used for the macro and the small cells, fractional frequency reuse is an attractive solution for limiting said interference, especially in UDNs. In such implementation, the lower frequencies of the available spectrum can be used for the macro cells to guarantee ubiquitous coverage, while the higher frequencies can be used for the small cells, limiting their reach and interference by using beamforming techniques. Nevertheless, the capacity-boosting that small cells can offer with this type of deployment is rather limited, since the available spectrum in the sub-6 GHz band is scarce.

Out-of-band deployments can then be more attractive to reduce the inter-cell interference and to have more spectrum for each tier of the HetNet. When mmWaves are used for the small cells, then it is possible for them to work as capacity boosters and help reach the rates required for 5G NR.

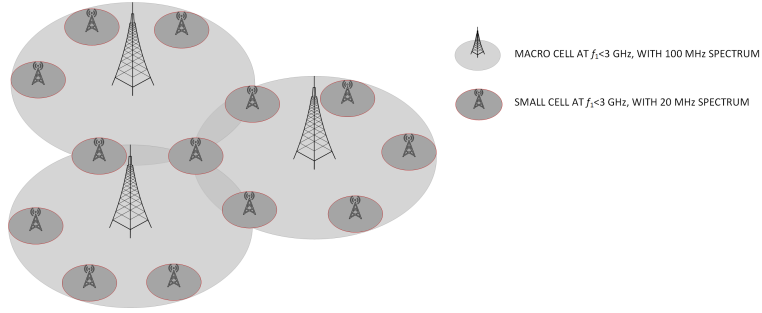
#### 4.3.1 Network model

Aiming at analyzing how the usage of the spectrum in a HetNet affects capacity dimension, two 5G deployment scenarios are considered and estimations are done for the access links; this analysis was presented as part of our work in Paper A. For both scenarios, a tier of macro cells is used in the sub-3 GHz frequency with an available spectrum of 100 MHz per cell, assuming carrier aggregation. A second tier of small cells is considered with sub-3 GHz frequencies and 20 MHz of spectrum per cell for the first scenario, and mmWave frequencies, in the 71-76 GHz band, and 250 MHz of spectrum per cell for the second scenario. In Figure 4.15 an example of the two scenarios is shown. For the macro cells the maximum reach was set to 850 m, whereas for the small cells it was set to 200 m; both types of cells were modeled as hexagons and a test area of 500 km<sup>2</sup> was used. The monthly traffic volume,  $T$ , that each cell can handle in GB/month/km<sup>2</sup>, assuming that

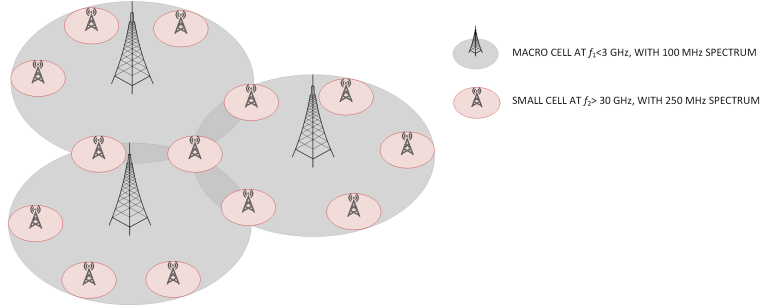
during the busy hour the average load of the cell is 50% and that the busy hour carries 15% of the daily traffic, was estimated following Equation 4.2, where  $CC$  is the cell capacity and is dependent on the bandwidth,  $BW$ , and the spectral efficiency,  $SE$ ; it was calculated following Equation 4.3, where  $SE = 1.74\text{bps/Hz/cell}$ .

$$T = \frac{CC[\text{Gbps}] \cdot 3600[\text{s}] \cdot 50\%}{15\% \cdot \text{cell area}[\text{km}^2]} \cdot 30[\text{days}] \quad (4.2)$$

$$CC[\text{Mbps}] = BW[\text{MHz}] \cdot SE[\text{bps/Hz/cell}] \quad (4.3)$$



(a) Scenario 1



(b) Scenario 2

Figure 4.15: HetNets deployment scenarios for two spectrum usage examples.

For the macro cell tier, a total of 266 cells are considered, hence covering 99.8% of the test area. These 266 macro cells are needed for both scenarios studied to provide the basic coverage of the network. The small cells are then deployed on top of the macro cells to increase the capacity.

### 4.3.2 Network dimensioning

A comparison of the deployment estimations for both scenarios is shown in Figure 4.16; here a reduction of up to 12% can be seen in the number of deployed cells needed to handle the same network traffic volume for the scenario 2 in comparison to scenario 1. These results reflect the benefits of the use of mmWave frequencies for the small cells in HetNets; each small cell in scenario 2 can support 12.5 times the traffic volume of a small cell in scenario 1. Such benefits increase as the network traffic volume increases, which indicates that for UDNs like the ones expected for 5G, the use of mmWave is an attractive solution to cope with the high traffic demand and the enhanced mobile broadband requirements of 5G.

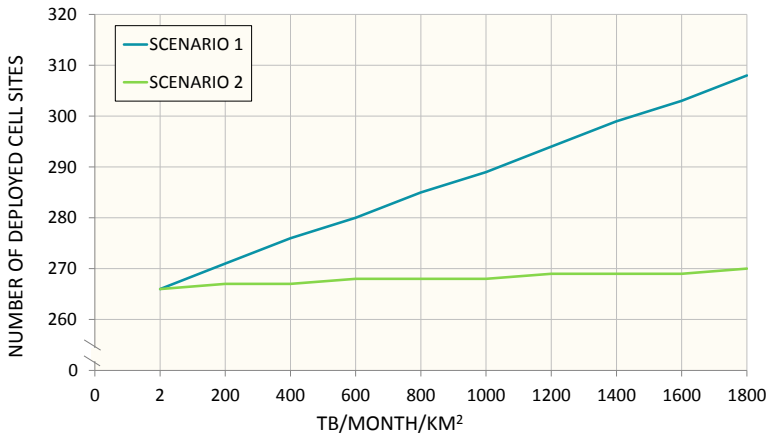


Figure 4.16: Capacity dimensioning comparison for two HetNets deployment scenarios.

Although for these calculations the first portion of the E-band was considered, it remains an open issue whether the lower mmWave bands (i.e., <60 GHz) will offer a better performance than the higher frequencies (i.e., >60 GHz). The 60 GHz band offers free license along with interference protection thanks to its atmospheric attenuation and narrow beamwidth; on the contrary, the E-band uses a "light licensing" scheme in most countries and offers higher reach since higher transmit power is allowed, and its atmospheric attenuation is lower. Furthermore, the E-band has more spectrum available with 10 GHz, enabling higher frequency reuse and higher capacity; these are the key reasons to choose the E-band as a viable solution for fixed and highly directional for both the access and the backhaul links. Deciding which portion of the E-band to use for these links is currently an open regulatory issue; a possibility is to use the lower part of the E-band for the access links, as it has been suggested in [Rap+13], while using the higher part of the E-band for the backhaul. This type of implementation would help to avoid interference within the same portion of the E-band in the proximity of the small cells.

Moreover, it is suggested to rely on the higher portion of the band for the backhaul and, due to the slightly better propagation of the radio waves, use the lower portion for the communication with the UEs.

Through the results presented in this section give an overview of the capacity benefits that can be expected for 5G deployments; nevertheless, calculations considering parameters taken from field measurements or network statistics will provide a more accurate capacity dimensioning adjusted to the scenario characteristics. Furthermore, other mmWave frequencies should also be considered for a better perspective on their influence on the cell capacity.

## 4.4 Chapter summary

Due to the challenges that mmWaves impose on the signal propagation, and the need to use beamforming techniques to benefit from the large spectrum that such frequencies offer, the cell discovery process of small cells operating at mmWave differs from that used in microwave small cells. To tackle this issue, the performance of a proposed search method for the cell discovery, named JSM and based on cooperation between the macro cell and the small cell, was compared to generic search methods. The results show that the JSM can offer significant improvements with a reduction of up to 96% in the number of pilot signal transmissions that the small cell need to send to find the UE. Furthermore, this improvement comes along with the use of narrow beamwidths, allowing to have a network that offers a high probability of connection to small cells with few cells deployed, low discovery delay, and high data rates.

The benefits of using mmWaves are further evaluated, by comparing two HetNet deployments; in one, microwave frequencies are used for all elements in the network, while in the other, mmWaves are used for the small cells. The results showed that with the latter a 12% improvement can be achieved in the number of cells needed to be deployed to handle the same network traffic volume. Thus, confirming the advantages that mmWave brings to dimensioning and deployment of HetNets for 5G.



## CHAPTER 5

# 5G NR and NOMA

---

### 5.1 Introduction

NOMA is a promising implementation for 5G networks, and it has gained a lot of attention recently in the research community. The high spectral efficiency that can be achieved with NOMA makes it an appealing MA scheme for cases when the resources are scarce [Sai+13], which are becoming more and more frequent nowadays. Nevertheless, such benefits come at the expenses of more complex signal transmission and reception, extra intra-cell interference, and higher complexity in the UEs scheduling and resource allocation, as explained in Chapter 3. In this chapter, research work related to some implementation challenges of NOMA for hybrid MA 5G networks is presented. The analyses are done for the DL since it is in this direction where the highest data demand typically occurs, and the results are obtained through simulations. For some scenarios, the combination of NOMA with mmWave frequencies is considered; these frequencies are used for coverage purposes and to show that by combining them with NOMA, huge gains in the data rates can be achieved. However, the challenges of combining mmWave with NOMA are outside the scope of this chapter and are addressed further in Chapter 6.

The structure of this chapter is as follows. Section 5.2 analyzes the implications that using NOMA has on the QoS delivered in comparison to OMA; a proposed solution to help NOMA overcome such QoS challenges is presented and analyzed in Section 5.3. In Section 5.4 the performance of four generic pairing methods for NOMA is compared, focusing on the capacity gain they offer versus the computational complexity; the reason for choosing generic methods is that they can offer an easy solution for addressing the user pairing challenge in NOMA. Finally, Section 5.5 summarizes the chapter.

As a side note, all the gains in the results analyses are calculated as the ratio between the parameter being evaluated and the benchmark; such gains are therefore dimensionless. The simulation model used for all the analyses is a 5G model developed in MATLAB using the LTE Toolbox; a subcarrier spacing  $\Delta_f = 2^\mu 15$  [kHz] with  $\mu = 0$  [3GP17b] is used, offering backward compatibility with 4G LTE.

### 5.2 QoS challenges

The benefits of using NOMA on the network capacity have been shown in several research works; [Ben+15; Din+17; Sai+13] are some examples. As illustrated in Figure 3.3 for two users, the data rate boundary for NOMA is outside that of OMA; therefore, NOMA

allows for higher data rates and hence higher network capacity, provided that the channel gains of the users in question are different.

Nevertheless, with the implementation of NOMA intentional intra-cell interference is introduced at the transmitter during the superposition process. This extra interference must be considered when analyzing the performance of NOMA; especially if a comparison with OMA is to be done. Under the same propagation and location conditions, a user will experience lower SINR from a NOMA signal than from an OMA signal, because of the extra intra-cell interference in NOMA. This means that the CQI reported by the UE to the BS can be expected to be lower when NOMA is used, thus reducing or even eliminating (in the worst cases) the data rate gain of NOMA over OMA. If this difference in the CQI reported is not accounted for, and the same CQI is used regardless of the MA scheme, an increase in the Block Error Rate (BLER) is expected; this will compromise the QoS delivered.

### 5.2.1 Single cell site model

To evaluate the QoS challenges of using NOMA instead of OMA, a comparison of both MA schemes based on the sum rate, SINR, and targeted BLER, is done; part of this work was presented in our Paper C. A single outdoor cell site was considered for the simulation; since the NOMA is expected to be smoothly integrated with the new technologies for 5G, such cell site is a mmWave small cell site operating at 73 GHz with a maximum reach of 200 m. For this study, the use of mmWave determines the cell size and the number of carriers available for transmission, this will help to emphasize the high data rates that can be achieved for 5G.

For the purpose of the simulations, a mapping of the SINR experienced by the UE into a CQI value is needed; the CQI is reported to the BS and it indicates which MCS should be used, and hence influences the data rate of the UE. Neither the SINR to CQI nor the CQI to MCS mappings are standardized, and in real implementations they are vendor/operator dependent. Therefore, as the first step in the simulations, an SINR to CQI mapping is derived for a target BLER of 10% and for four modulations: QPSK, 16QAM, 64QAM, and 256 QAM.

Typically, the UEs located closer to the BS will experience a higher SINR because the path losses are less than for the UEs located farther away; therefore, the coverage area of a cell could be roughly divided in concentric rings with the center being the location of the BS, each ring representing the area where certain modulation is supported. Therefore, UEs at the edge of the cell could support up to QPSK, whereas the UEs close to the BS could support up to 256QAM, as shown in Figure 5.1. Considering this, UE positions at variable distances from the BS should be evaluated for deriving accurate mappings of SINR to CQI. As a baseline in the simulations, this mapping is derived assuming OMA and is then used for NOMA as well so the penalty of the intentional intra-cell interference can be characterized. The CQI to MCS mapping is done following the proposal in [3GP14]. The flowchart in Figure 5.2 illustrates the process for the SINR to CQI mapping; the size of



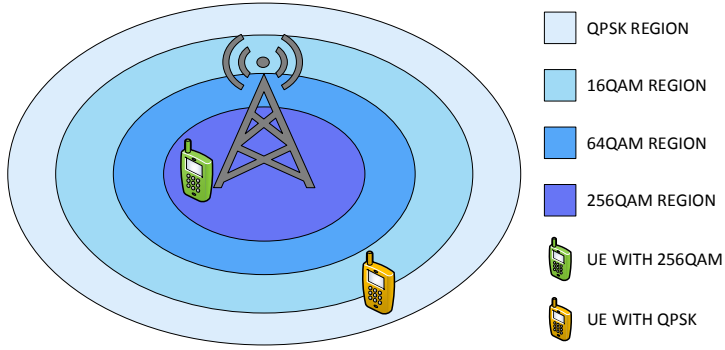


Figure 5.1: Example of a typical distribution of modulations in a mobile network.

the transmitted packets used for the mappings was selected according to the Transport Block Size (TBS) defined in the LTE standard [3GP17a] for 100 PRBs, since a carrier bandwidth of 20 MHz is considered. Table 5.1 shows the results of the mappings for OMA for a 10% BLER; a total of 500 positions covering the whole area of the cell were evaluated for these results.

### 5.2.2 Performance evaluation

Once the mappings are available, we proceed with the performance evaluation of NOMA over OMA. For this, a simulation scenario with two UEs is considered and it is assumed that both UEs need to be served at the same time and that the available resources must be shared between them; thus, the number of multiplexed UEs for NOMA is set to two. As explained in Section 2.3.3, the data rate gain with NOMA depends on the channel gain difference,  $\Delta_{\text{SINR}}$ , between the multiplexed UEs. To guarantee that  $\Delta_{\text{SINR}}$  will be large enough to have a gain in the data rates of each UE, the SINR to CQI mapping derived is used and different values of  $\Delta_{\text{SINR}}$  are evaluated. The location of the UEs is estimated according to the desired SINR while guaranteeing that there is a difference in the CQIs reported by the UEs, that is that  $\text{CQI}_1 > \text{CQI}_2$  provided that  $\text{UE}_1$  is the one with higher channel conditions. For example, if a  $\text{CQI}=2$  needs to be reported, then the SINR should be greater than -5.75 dB and lower than -2.7 dB. For this, a set of allowed CQI values is defined for each UE; for  $\text{UE}_1$  the set is  $S_1=\{3,4,5,\dots, 15\}$ , while for  $\text{UE}_2$  the set is  $S_2=\{2,3,4,\dots, 14\}$ . During each iteration of the model a value of  $\text{CQI}_2 \in S_2$ , and all the values of  $\text{CQI}_1 \in S_1$  with  $\text{CQI}_1 > \text{CQI}_2$  are evaluated one to one, that makes for a total of 13 iterations per run, with the number of CQI combinations evaluated being reduced by one in each iteration. For example, in the first iteration  $\text{CQI}_2=2$  and  $\text{CQI}_1=\{3,4,5,\dots, 15\}$ ; in the second iteration  $\text{CQI}_2=3$  and  $\text{CQI}_1=\{4,5,6, \dots, 15\}$ , and so on until all the values in  $S_2$  have been chosen. Figure 5.3 shows the simulation scenario, whereas in Table 5.2

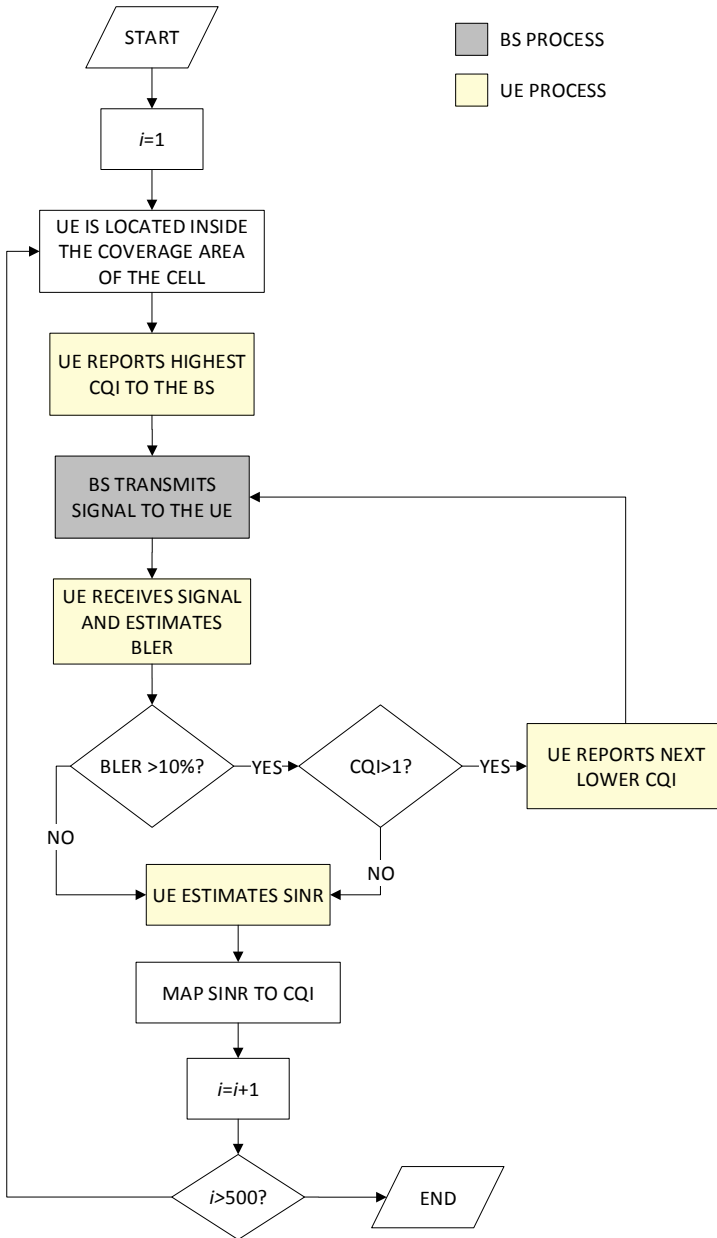


Figure 5.2: Logical process used for the SINR to CQI mapping for OMA.

Table 5.1: SINR to CQI mapping for OMA derived from simulations.

SINR (FROM SIMULATIONS)	CQI	MCS INDEX	MODULATION	MODULATION ORDER	CODE RATE ( $\times 1024$ )
-7	1	0	QPSK	2	78
-5.75	2	1	QPSK	2	193
-2.7	3	3	QPSK	2	449
1.4	4	5	16QAM	4	378
2.83	5	7	16QAM	4	490
4.48	6	9	16QAM	4	616
6.45	7	11	64QAM	6	466
8.14	8	13	64QAM	6	567
9.47	9	15	64QAM	6	666
11.43	10	17	64QAM	6	772
12.82	11	19	64QAM	6	873
14.47	12	21	256QAM	8	711
16.16	13	23	256QAM	8	797
17.53	14	25	256QAM	8	885
20.55	15	27	256QAM	8	948

the difference between  $CQI_1$  and  $CQI_2$ ,  $\Delta_{CQI}$ , is shown for all the CQI combinations considered. The logical process of the simulations is illustrated in Figure 5.4. To calculate the distance between the BS and the UEs based on the desired SINR/CQI, the following steps are applied:

1. Calculate the received power,  $P_{RX}$ , as:

$$P_{RX} = \text{SINR} - (P_N + P_I) \quad (5.1)$$

where  $P_I$  is the interference power and  $P_N$  is the noise power at the output of the receiver's low noise amplifier (LNA) and it is calculated as:

$$P_N = -174[\text{dBm}] + 10\log(B) + NF_{RX} + G_{RX} \quad (5.2)$$

where  $B$  is the bandwidth in Hz,  $NF_{RX}$  represents the noise figure of the receiver in dB, and  $G_{RX}$  is the receiver antenna gain in dBi.

2. Calculate the path losses,  $P_L$ , as:

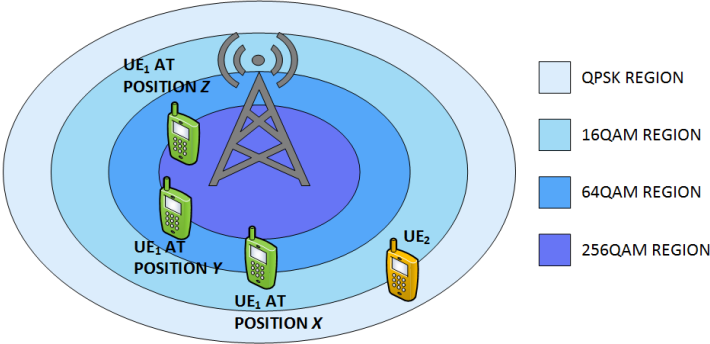


Figure 5.3: Simulation scenario used for comparing the performance of NOMA and OMA.

Table 5.2: CQI combinations considered for the simulations.

		CQI <sub>1</sub>													
		3	4	5	6	7	8	9	10	11	12	13	14	15	
CQI <sub>2</sub>	2	1	2	3	4	5	6	7	8	9	10	11	12	13	
	3		1	2	3	4	5	6	7	8	9	10	11	12	
	4			1	2	3	4	5	6	7	8	9	10	11	
	5				1	2	3	4	5	6	7	8	9	10	
	6					1	2	3	4	5	6	7	8	9	
	7						1	2	3	4	5	6	7	8	
	8							1	2	3	4	5	6	7	
	9								1	2	3	4	5	6	
	10									1	2	3	4	5	
	11										1	2	3	4	
	12											1	2	3	
	13												1	2	
	14													1	

CQI COMBINATION NOT CONSIDERED

$$P_L = P_{TX} - P_{RX} + G_{TX} + G_{RX} \quad (5.3)$$

where  $P_{TX}$  is the transmission power in dBm,  $G_{TX}$  is the transmitter antenna gain in dBi.

3. Calculate the distance between the BS and the UE,  $d$ . The following propagation model for mmWave outdoor channels at 73 GHz is used [Rap+14], and it is assume that the UE is in NLOS; this assumption was done to evaluate the model in the most challenging propagation conditions:

$$P_L = 69.8\text{dB} + 33\log(d) + x_\sigma \quad (5.4)$$

where  $x_\sigma$  represents the shadowing factor and it is a random Gaussian variable with mean zero and standard deviation  $\sigma = 7.6\text{dB}$ . The distance  $d$  is expressed in meters.

The same simulation scenario was used for NOMA and OMA, with the difference relying on the algorithm for the scheduler at the MAC layer. For NOMA, the scheduler assigns the whole bandwidth to both UEs and splits the power according to their channel gains. The power allocation process for each multiplexed UE in NOMA is a challenging task, and the ratios must be carefully chosen to allow the correct decoding of the signals on the receiver side. For this, the approach suggested in [Zhu+15] is followed, which is based on the estimation of the value of  $P_I$  that satisfies the following equation:

$$\frac{1}{2}\log_2(1 + \gamma_2) = \log_2 \left( 1 + \frac{(1 - P_1)\gamma_2}{\left(\frac{P_1}{r}\right)\gamma_2 + 1} \right) \quad (5.5)$$

where  $\gamma_2 > 0$  is the  $\text{SINR}_2$  and  $r$  is the number of transmitted layers. By satisfying Equation 5.5 the channel capacity of UE<sub>2</sub> with NOMA is equal to the one achieved with OMA when full power and half of the bandwidth is assigned to UE<sub>2</sub>. Since this is the case in this simulations, the use of this power allocation strategy is valid. Then,  $P_I$  can be calculated as follows for  $r=1$ :

$$P_1 = \frac{\sqrt{1 + \gamma_2} - 1}{\gamma_2} \quad (5.6)$$

For UE<sub>2</sub>, the value of  $P_2$  can be calculated as the remaining power:

$$P_2 = P_{TX} - P_1 \quad (5.7)$$

For OMA, half of the bandwidth is assigned to each UE in the scheduler. The size of the packet to be transmitted to the UEs is determined is selected according to

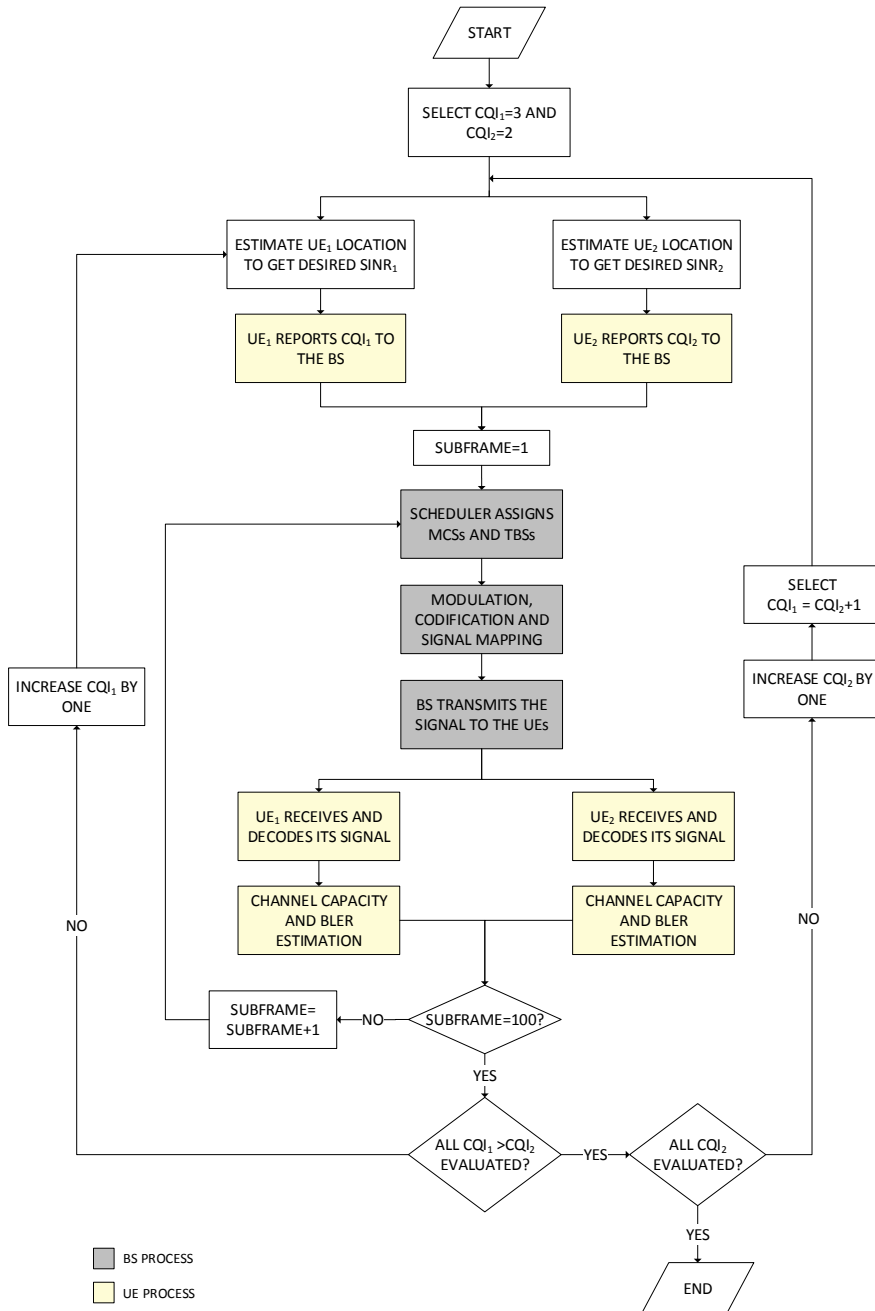


Figure 5.4: Logical process used for comparing the performance of NOMA and OMA.

the TBSs defined in the LTE standard [3GP17a] for 100 and 50 PRBs for NOMA and OMA, respectively. As the model operates at mmWave frequencies, a bandwidth of 800 MHz is assumed to be available for transmission. Such bandwidth is divided into eight channels of 100 MHz each; this is done so the transmission can be compatible with LTE, which supports a maximum of 100 MHz of channel bandwidth. Furthermore, orthogonal frequency-division multiplexing (OFDM) is used for the carrier modulation. For the receiver types in NOMA,  $UE_1$  implements symbol-level interference cancellation (SLIC), while  $UE_2$  uses linear minimum mean square error (LMMSE); for OMA, both UEs use a LMMSE receiver. The parameters for the simulations are summarized in Table 5.3.

Table 5.3: Simulation parameters for comparing the performance of NOMA and OMA.

	OMA	NOMA
CARRIER FREQUENCY (GHz)	73	
CARRIER BANDWIDTH (MHz)	20	
CARRIER COMPONENTS	40	
CARRIER BANDWIDTH PER UE (MHz)	10	20
CODING/DECODING	TURBO CODING	
MODULATION SCHEME	QPSK, 16QAM, 64QAM, 256QAM	
TRANSMISSION POWER (dBm)	15	
POWER ALLOCATION PER UE	1	ADAPTIVE (*)
WAVEFORM	OFDM	
TRANSMISSION MODE	SU-MIMO (2x2)	
TRANSMITTER ANTENNA GAIN (dBi)	37	
PATH LOSS MODEL	mmWave OUTDOOR FOR 73 GHz ACCESS (**)	
CHANNEL ESTIMATION	MMSE	
RECEIVER ANTENNA GAIN (dBi)	0	
RECEIVER NOISE FIGURE (dB)	6	
RECEIVER SCHEME	LMMSE	$UE_1$ = SLIC $UE_2$ = LMMSE

\* SEE EQUATION 4.6

\*\* SEE EQUATION 4.4

For the first comparison, Figure 5.5 shows the channel capacity gain of NOMA over OMA as a function of  $\Delta_{CQI}$  for all the CQI combinations shown in Table 5.2. It can be seen that such capacity gain varies considerably from 13% to 71.5%, with the lower values corresponding to the smallest  $\Delta_{CQI}$ , and vice versa. Thus, confirming that the capacity benefits of NOMA over OMA are greater as the channel gain difference between the multiplexed UEs increases. Nevertheless, even for the lowest  $\Delta_{CQI}$  an average gain

of 40% can be expected; for the higher values of  $CQI_1$  and  $CQI_2$  higher data rates and channel capacity can be achieved because of the higher modulation order and lower coding rate (Table 5.1).

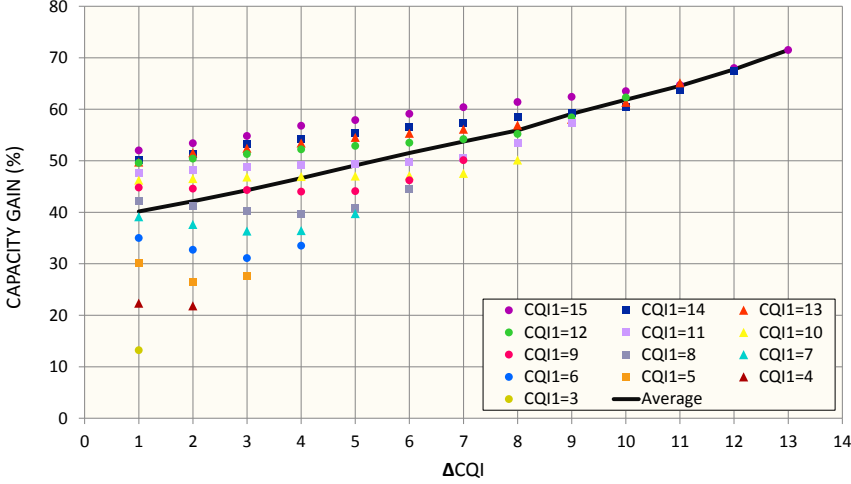
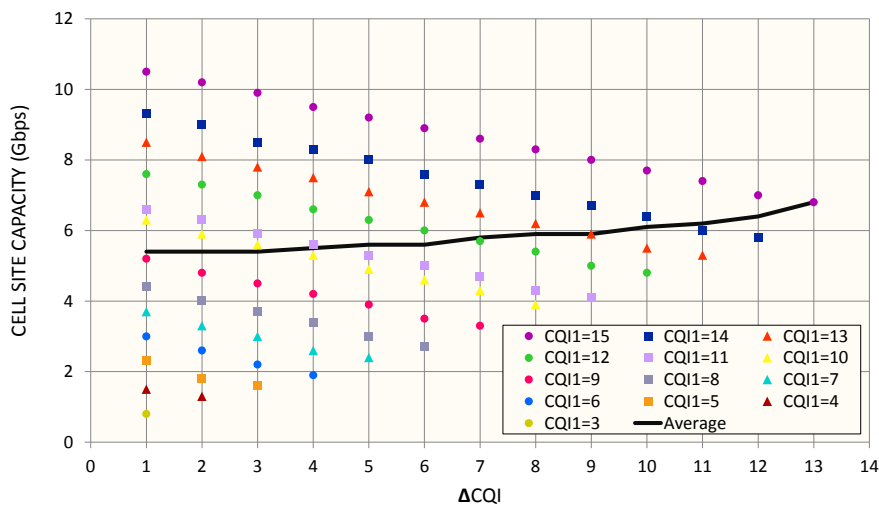


Figure 5.5: Channel capacity gain of NOMA using OMA as a benchmark.

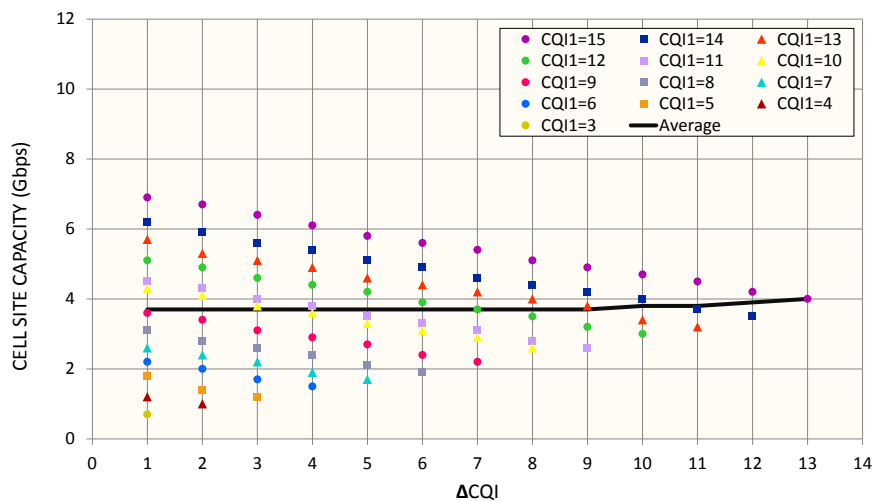
In Figure 5.6 the channel capacity for both MA schemes is shown; values of up to 21 Gbps and 13.8 Gbps were achieved with NOMA and OMA, respectively. These high values are possible thanks to the large channel bandwidth that is available at mmWave frequencies, allowing for a high spectral efficiency, up to 26.25 bps/Hz/cell with NOMA and 17.25 bps/Hz/cell with OMA. Furthermore, as the number of antenna ports supported by MIMO increases, higher channel capacity and spectral efficiency can be achieved.

Even though from a channel capacity perspective the benefits of using NOMA are clear, the implications that the intentional intra-cell interference added at the transmitter with NOMA have on the network performance must also be evaluated. For this, Figure 5.7 shows the BLER for NOMA. It can be seen that the trend is the opposite as for the channel capacity, with the BLER being lower as  $\Delta CQI$  increases; there are two aspects that play a role in this behavior: the modulation order and the power allocation algorithm. To understand this, let us assume that both UEs report a CQI that corresponds to a QPSK modulation, say  $CQI_1=3$  and  $CQI_2=2$ . In this case, although the composite constellation will not be very complex as shown in Figure 5.15d, the power ratio assigned to each UE will most likely not be sufficient to correctly decode the signal, especially for UE<sub>2</sub>; that is because both UEs need a high power ratio (e.g.,  $>0.5$ ) due to their low channel gain. If we then assume that both UEs have high channel gains and they report CQIs corresponding to a 256QAM modulation, say  $CQI_1=15$  and  $CQI_2=14$ , the composite constellation becomes





(a) NOMA



(b) OMA

Figure 5.6: Channel capacity for NOMA and OMA.

more complex as shown in Figure 5.15e, therefore making it quite challenging to decode correctly even with high channel gains. On the contrary, if the channel gains are very different, then  $UE_2$  can get most of the power (i.e.,  $P_2 > 0.5$ ) and the remaining power can be assigned to  $UE_1$  whose channel gain is high and does not need a high power ratio.

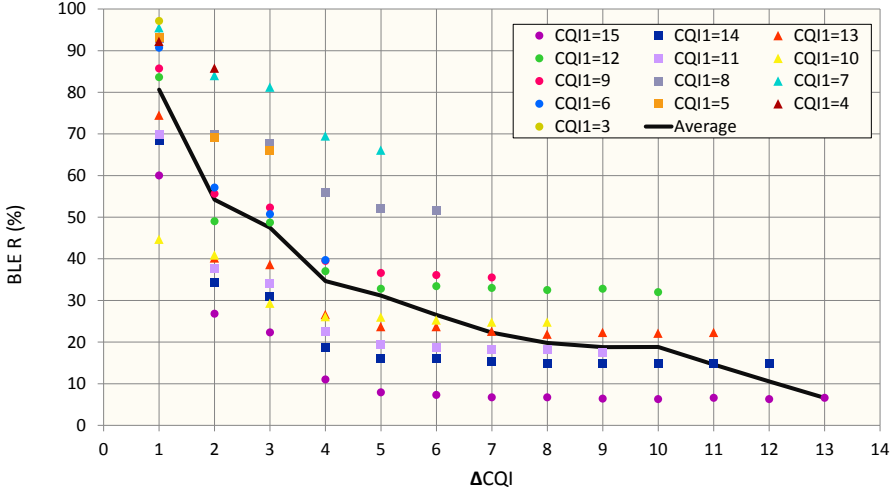


Figure 5.7: BLER for NOMA when using the SINR to CQI mapping derived for OMA.

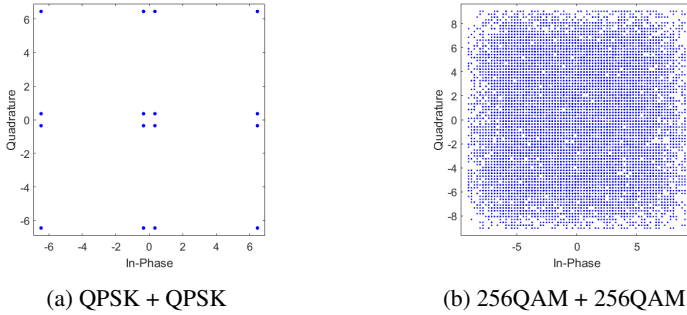


Figure 5.8: Composite constellation diagrams for two combinations of modulation schemes.

Nevertheless, it can also be seen from Figure 5.7 that for most of the cases the BLER is above the targeted 10%; in fact, only when  $CQI_1=15$  and  $\Delta_{CQI} \geq 5$  the BLER was kept below 10%. This indicates that the MCS assigned to the UEs was not the best match

for the channel gains. However, the MCS was assigned according to the CQI to MCS mapping shown in Table 5.1, which was derived using the same simulations for the OMA UEs and fulfilling the 10% BLER requirement. Through this results, we can confirm that NOMA UEs require higher levels of SINR than OMA UEs to successfully decode the signals under the same propagation conditions; this is due to the extra intra-cell interference in NOMA.

To see how this extra interference affects the decoding process, we can refer to Figures 5.9 and 5.10 where the constellation diagrams for 256QAM ( $\text{CQI}_1=15$ ) and QPSK ( $\text{CQI}_2=2$ ) are shown for NOMA and OMA, respectively.

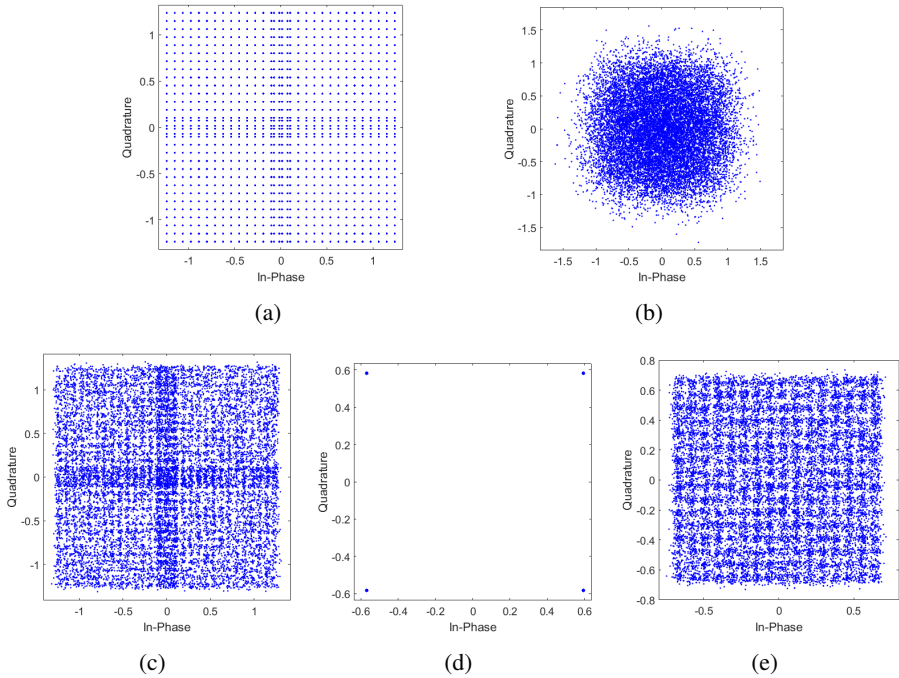


Figure 5.9: Constellation diagrams during the transmission and reception of signals with NOMA for QPSK and 256QAM: (a) Composite constellation sent to  $\text{UE}_1$  and  $\text{UE}_2$ ; (b) Equalized constellation received by  $\text{UE}_2$ ; (c) Equalized constellation received by  $\text{UE}_1$ ; (d) Regenerated QPSK constellation by  $\text{UE}_1$  corresponding to  $\text{UE}_2$ 's signal; (e) Recovered 256QAM constellation by  $\text{UE}_1$ .

From Figures 5.9 and 5.10 it can be seen that because of the complexity of the composite constellation, the NOMA UEs receive a more degraded constellation than the OMA UEs, since the same MCS is used regardless of the MA scheme. Particularly for  $\text{UE}_2$  the

decoding process is quite challenging; for this example the both NOMA UEs the BLER was one, while for the OMA UEs it was zero. This issue intensifies as the modulation of the multiplexed UEs become more complex. In Figure 5.11 the constellation diagrams are shown for 256QAM ( $CQI_1=15$ ) and 16QAM ( $CQI_2=4$ ). The same trend can be seen in this example, with the  $UE_2$  receiving an even more degraded constellation.

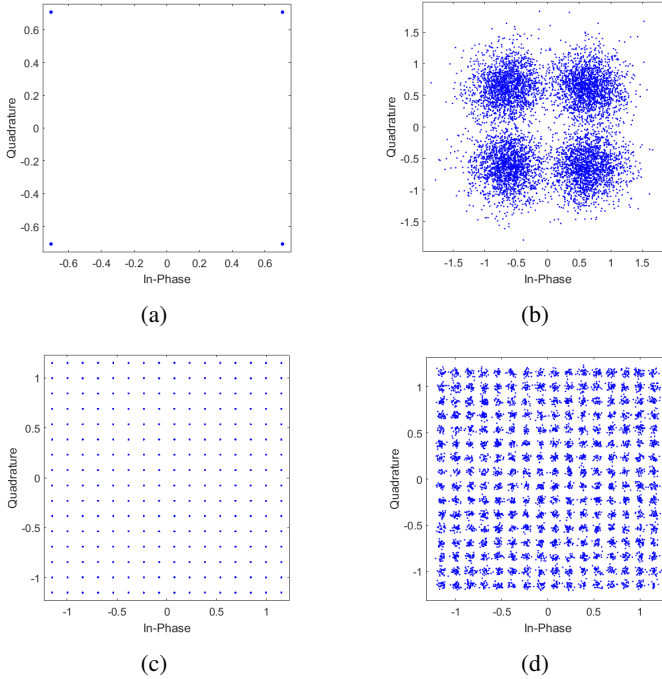


Figure 5.10: Constellation diagrams during the transmission and reception of signals with NOMA for QPSK and 256QAM: (a) Constellation for QPSK sent to  $UE_2$ ; (b) Equalized constellation received by  $UE_2$ ; (c) Constellation for 256QAM sent to  $UE_1$ ; (d) Equalized constellation received by  $UE_1$ .

To estimate how much is the penalty in the SINR when using NOMA instead of OMA, iterative simulations were performed following the process shown in Figure 5.12. For  $UE_2$  only the  $CQI=2$  corresponding to a QPSK modulation was considered; the reason for this is that for higher modulation orders for  $UE_2$  the SINR penalty was too high and the SINR needed significantly exceeded the values typically experienced in real implementations (e.g.,  $\gg 30$  dB). For simplicity reasons the location of the  $UE_2$  was fixed, and that of the  $UE_1$  varied to achieve the desired SINR.

The results are shown in Figure 5.13 for four values of  $CQI_1$  for NOMA and OMA. It

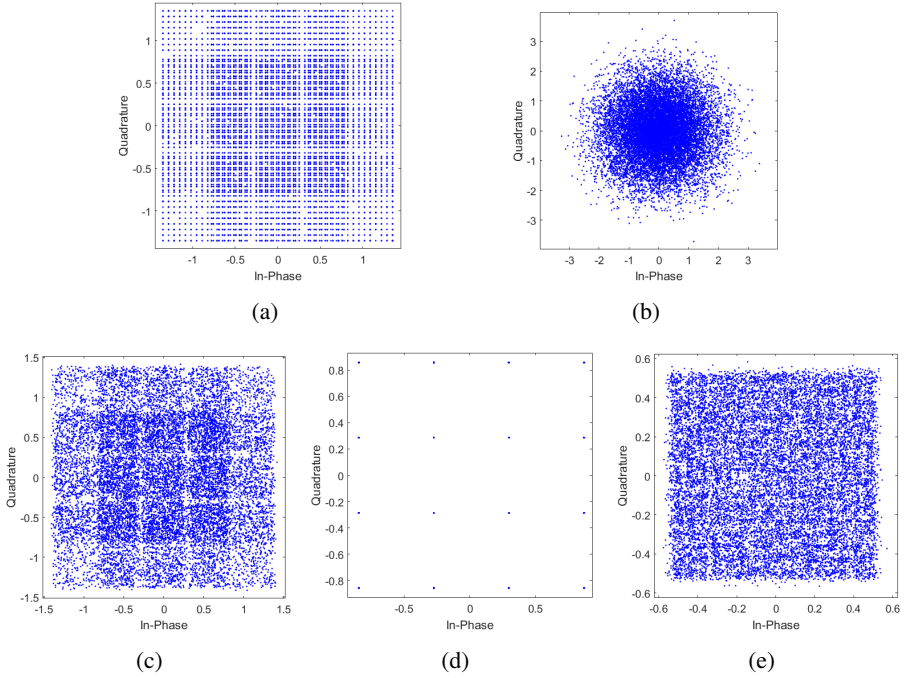


Figure 5.11: Constellation diagrams during the transmission and reception of signals with NOMA for 16QAM and 64QAM: (a) Composite constellation sent to UE<sub>I</sub> and UE<sub>2</sub>; (b) Equalized constellation received by UE<sub>2</sub>; (c) Equalized constellation received by UE<sub>I</sub>; (d) Regenerated 16QAM constellation by UE<sub>I</sub> corresponding to UE<sub>2</sub>'s signal; (e) Recovered 256QAM constellation by UE<sub>I</sub>.

can be seen that, for example, for the UE<sub>I</sub> to report a CQI<sub>I</sub> of 15 with OMA, it requires an SINR<sub>I</sub> of at least 20.55 dB for a 10% BLER. With NOMA, the same UE<sub>I</sub> in the same location requires an SINR<sub>I</sub> of at least 33.3 dB to report a CQI<sub>I</sub>=15 for a 10% BLER. That is an additional 12.75 dB in the SINR needed for NOMA. Moreover, for a 10% BLER the SINR<sub>2</sub> varied from -5.75 dB with OMA to 5 dB with NOMA, which corresponds to a difference of 10.75 dB. In Figure 5.14 the average penalty on the SINR for the UE<sub>2</sub> and a CQI<sub>2</sub>=2 is shown; here the values varied from 9.8 dB to 11.2 dB, with an average of 10.45 dB depending on the BLER. For a 10% BLER, the lowest penalty was 10.75 dB. After analyzing all the CQI combinations (Table 5.2) with CQI<sub>2</sub>= $\{1, 2\}$ , an average SINR penalty of 12 dB is present when choosing NOMA over OMA. When adjusting the SINR values for NOMA in the simulations and adding 12 dB extra, the signals were decoded correctly and the BLER was kept below 10%. For the same examples as in Figure 5.9, the

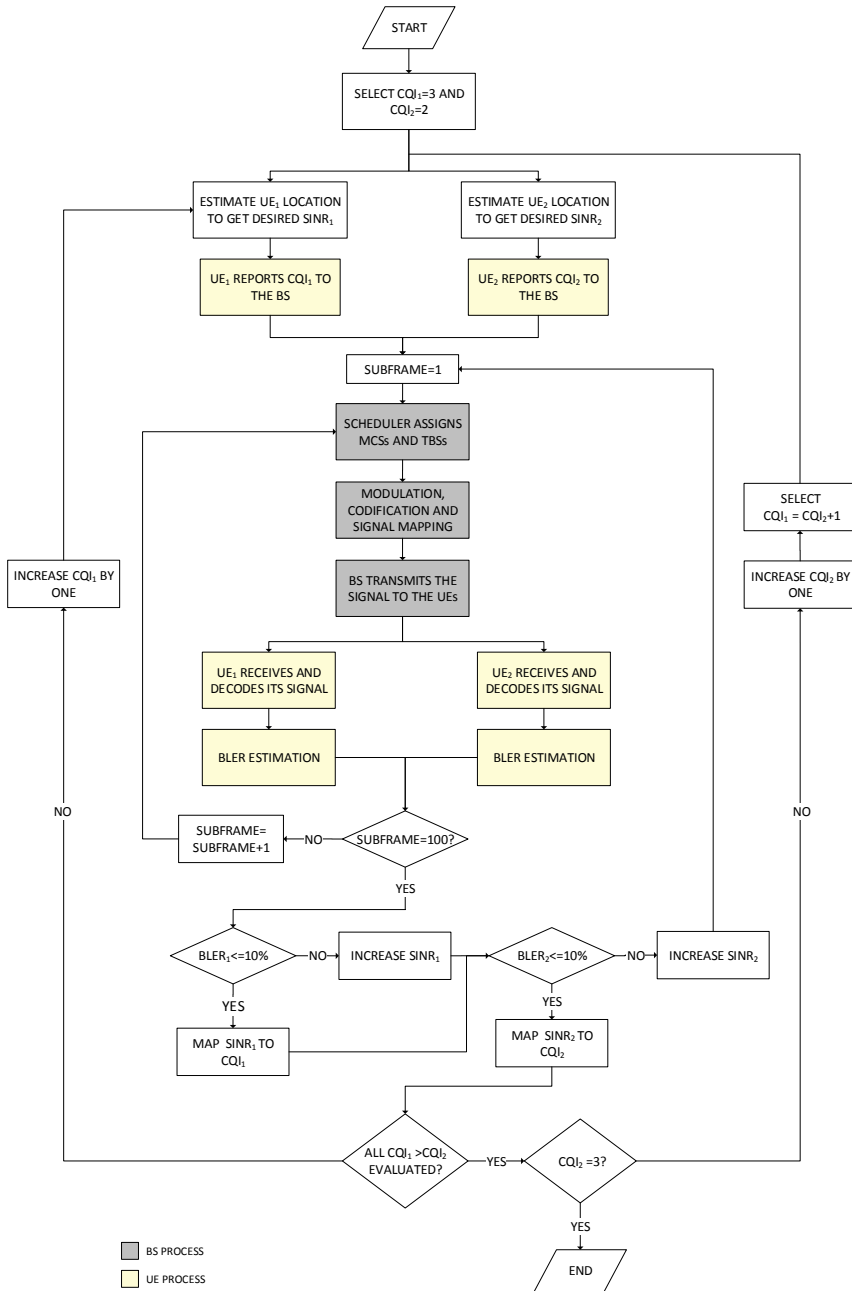
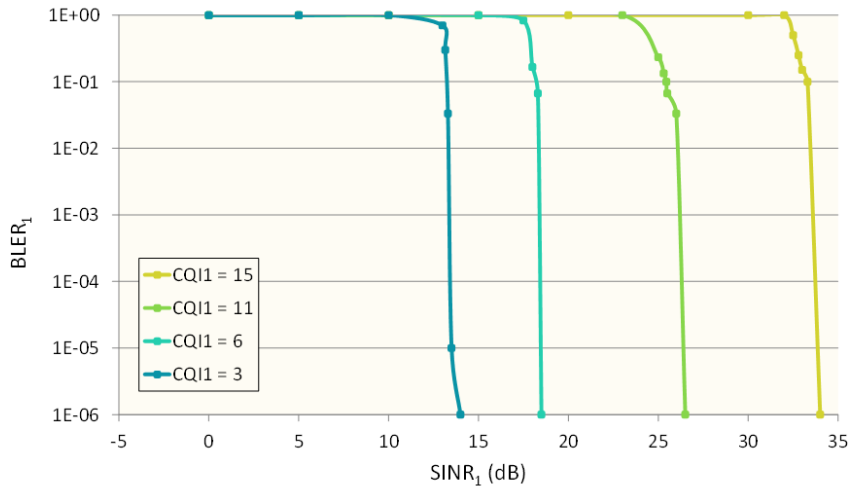
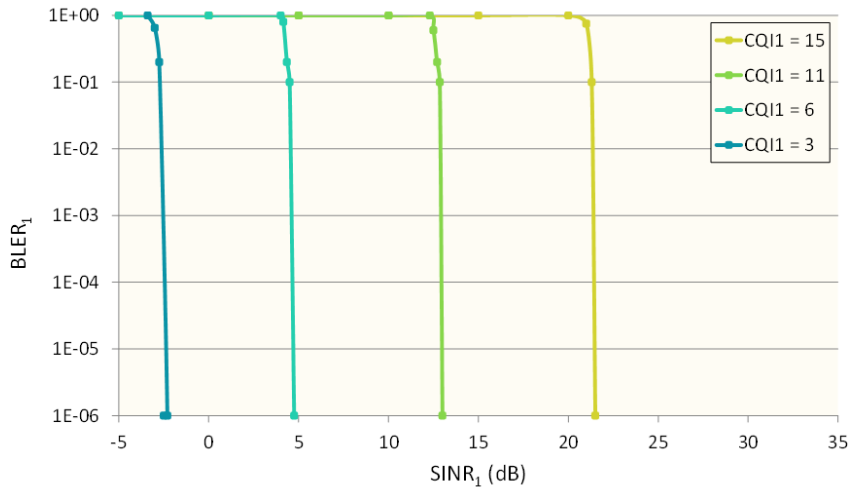


Figure 5.12: Logical process used for estimating the SINR for NOMA as a function of the BLER.



**NOTE:** FOR ALL CASES, CQI<sub>2</sub>=2 WITH SINR<sub>2</sub> = 5 dB FOR A 10% BLER

(a) NOMA



**NOTE:** FOR ALL CASES, CQI<sub>2</sub>=2 WITH SINR<sub>2</sub> = -5.75 dB FOR A 10% BLER

(b) OMA

Figure 5.13: BLER for UE<sub>I</sub> as a function of the SINR for four values of CQI<sub>I</sub> and CQI<sub>2</sub>=2.

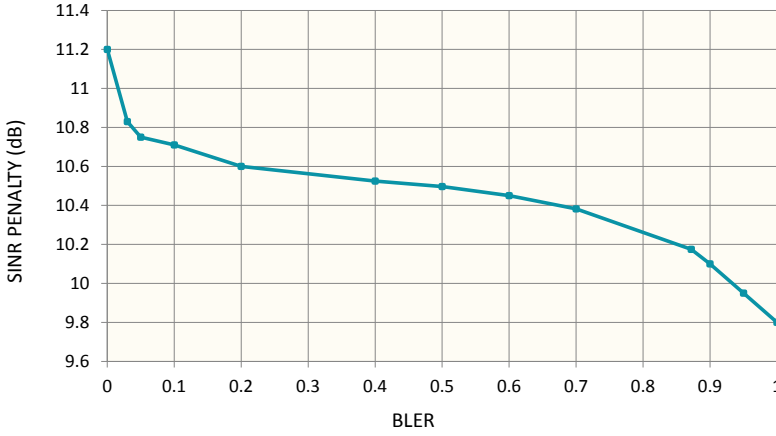


Figure 5.14: SINR penalty for choosing NOMA over OMA, for UE<sub>2</sub> with CQI<sub>2</sub>=2

new constellation diagrams are shown in Figure 5.15, where a significant improvement can be seen after the equalization, which allows for a successful decoding of the signals. Moreover, it can be seen that the composite constellation has less overlapping, this is because as the SINR<sub>2</sub> improves, the power ratio  $P_2$  increases (Equation 5.5); thus there is a higher difference in the quadrature of the superposed constellations. However, although this SINR adjustment can be done very simply in the simulations, the same does not apply for real implementations.

There are basically two ways of increasing the SINR, either by increasing the transmission power or by bringing the UE closer the BS, thus reducing the propagation losses. The first is not always possible due to transmission power regulations; the second cannot be controlled by the system. After analyzing these results, two main issues related to NOMA transmissions arise. First, how to use NOMA to increase the system capacity without degrading the quality of service (QoS) for neither the UEs nor the system. A failed signal decoding implies a retransmission, and a high number of retransmissions seriously compromise the QoS and the network performance because of the retransmissions overhead and the extra delays; this is particularly harmful to real-time and mission-critical services. The second issue is how to select the UEs to multiplexed with NOMA; this must be done in such way that guarantees that the signal will be successfully decoded for both UEs and that there will be a capacity gain over OMA.

One could think that a straightforward solution for the SINR penalty with NOMA is to regenerate the SINR to CQI mapping table, and although this would solve the failed decoding issue, such mapping would only be accurate for the NOMA UEs. As mentioned in Section 6.19a hybrid MA systems are expected for 5G where OMA and NOMA coexist, since it is not realistic to consider that all the UEs can be paired. Therefore, in such a



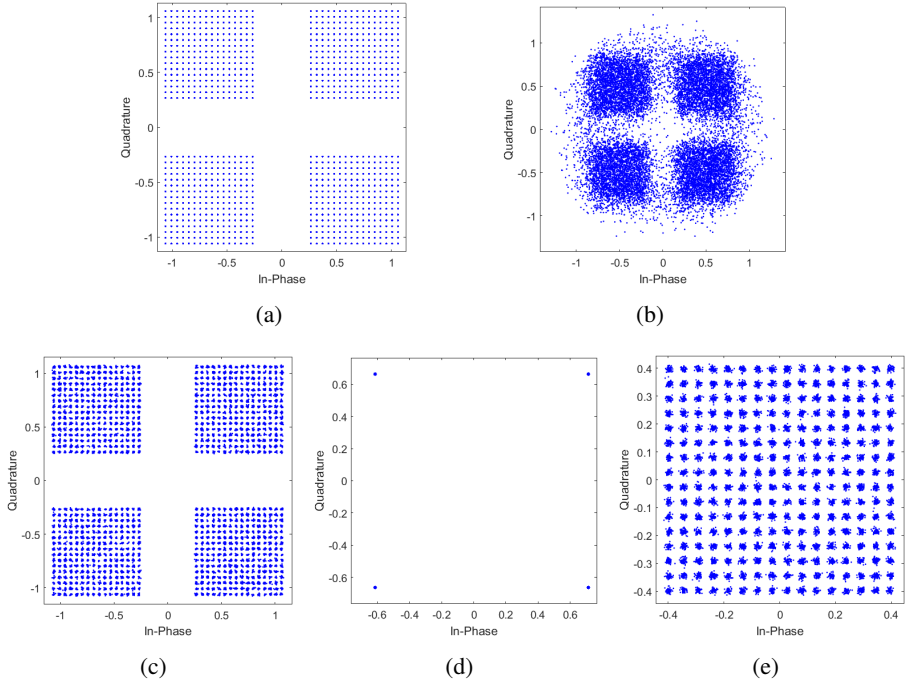


Figure 5.15: Constellation diagrams during the transmission and reception of signals with NOMA for QPSK and 256QAM with improved SINR: (a) Composite constellation for QPSK and 256QAM sent to  $UE_1$  and  $UE_2$ ; (b) Equalized constellation received by  $UE_2$ ; (c) Equalized constellation received by  $UE_1$ ; (d) Regenerated QPSK constellation by  $UE_1$  corresponding to  $UE_2$ 's signal; (e) Recovered 256QAM constellation by  $UE_1$ .

system the UEs would have to know a priori what kind of MA will be used for their transmissions to report the CQI accordingly; since the use of one MA or the other can vary in each subframe transmission this approach is not recommended. Furthermore, if a more accurate SINR to CQI adjustment is desired for NOMA, then the extra intra-cell interference would have to be estimated by each UE depending on the conditions of its pair UE; this would add complexity to the CQI reporting process. Another solution could be to have all the UEs report the CQI according to the SINR to CQI mapping for NOMA. In this case, the failed decoding issue is also solved but then the system capacity could be compromised, since the OMA UEs would have the capacity of supporting higher data rates than those achieved by the reported CQI; it would then have to be evaluated if the implementation of NOMA generates significant capacity gains to the system and the UEs.

Then, the most beneficial solution could be to have all the UEs report the CQI

according to OMA, and have the BSs take care of adjusting the MCS for those UEs whose transmission will be done using NOMA. In this regard, the BSs select the UEs to multiplex and assigns a lower MCS than that corresponding to the reported CQI. For the BS doing this adjustment is easier, since it can be integrated as an extra task during the algorithm to select the UEs to be paired. As a lower MCS implies lower data rates, the BSs can evaluate if this can be compensated with the extra bandwidth that will be allocated to the NOMA UEs because of the pairing. Nevertheless, this approach requires extra processing along with an optimization algorithm to ensure that the overall system performance will increase.

Although many research works have focused on the capacity benefits of NOMA, the work related to methods that help to account for the intra-cell interference in NOMA and select the appropriate transmission parameters in hybrid MA deployments, is rather scarce. Therefore, research work on this topic could be highly beneficial for the successful implementation of NOMA, especially since said interference will depend on the conditions of the multiplexed UEs. In the following section, a proposed method for addressing the issues here highlighted is presented.

### 5.3 Power allocation and MCS adjustment

Up to date, the research works addressing the SINR degradation and the CQI mapping issues that arise in hybrid MA systems with OMA and NOMA, are scarce. In [Ben+14] CQI adjustments for NOMA based on the power allocation ratio is proposed, while in [Ria+17; Zhu+15] the authors point-out the SINR difference between NOMA and OMA and a mismatch between the reported CQI and the actual supported one.

In this respect, a method for improving the performance of a hybrid MA system is proposed, focusing on techniques that help to increase the overall system capacity while maintaining the desirable BLER. To not add complexity to the CQI estimation, all the UEs report the CQI assuming an OMA transmission; then, at the BS we rely on MCS adjustments and extra transmission power ( $P_{TX}$ ) for the NOMA UEs to compensate for the extra intra-cell interference in NOMA. These adjustments are done during the scheduling process, which is divided into two steps.

In the first step, a preliminary PRBs assignment is done using only the CQI reported by each UE for the selection of the MCS. In the second step, the scheduler evaluates which UEs, if any, are candidates for NOMA; this eligibility is determined during the pairing process, where the scheduler tries to find UEs to pair while guaranteeing that the throughput of each UE will not be degraded by switching from OMA to NOMA. This can be done by maintaining or increasing the TBS for the NOMA UEs; nevertheless, when the MCS is adjusted to a more robust value, the only way to maintain the same TBS is to assign more PRBs. With NOMA, it is possible to assign more PRBs to a UE without compromising the PRBs already assigned to the other UEs in the system. When two UEs are paired with NOMA, they both have access to the sum of the PRBs individually

assigned to each one prior to the pairing. However, if we consider that the average extra 12 dB needed in the SINR with NOMA (Section 5.2), the MCS adjustment might not be enough by itself to make up for the extra dBs and not compromise the UEs throughput, which could lead to a high probability that only a few (e.g., 2 or 4) or none of the UEs can be paired. Moreover, as QPSK should be used for the paired UE with the lower channel gain in order keep the complexity of the composite constellation to a minimum, the probabilities of finding suitable UEs to pair reduce even further.

To overcome this limitation, extra  $P_{TX}$  can be considered for the NOMA UEs. For example, let us assume that for OMA the BS is not transmitting at the maximum regulated power, since a lower value is enough to guarantee the desired BLER and to cover the desired area; Then, a  $P_{TX}$  headroom can be considered in a hybrid MA system for the subcarriers that are using NOMA, as long as the total transmission power does not exceed the regulated maximum limit. By allocating extra  $P_{TX}$  when needed in NOMA, it is then possible to provide the extra dBs that cannot be provided with the MCS adjustment to reach the desired SINR. Although allocating extra  $P_{TX}$  to some UEs could be considered going against the idea of having 5G networks that are more energy efficient, if such extra power does not exceed the regulations, it helps to increase the probability of having UEs paired which directly impacts on the system capacity and on the QoS.

With this method, although the MCS adjustment can be done without any other reported information but the CQI, when it comes to the extra  $P_{TX}$ , extra signaling information needs to be shared between the BS and the UEs. Since the BS does not know the exact value of the SINR that the UE experiences, the latter needs to inform the BS if the MCS adjustment is enough or if extra Tx power needs to be sent to cover the extra dBs needed in the SINR. This implies that during the first transmission for a given NOMA pair, the BLER might be above the desired value, while for the following the issue is overcome.

### 5.3.1 Single cell site model

To evaluate the performance of this proposed method, a fully loaded mmWave cell site with UEs randomly located inside its coverage area; the maximum reach of the cell was set to 200 m. The number of UEs is selected randomly following a normal distribution with mean  $\mu=25$  and standard deviation  $\sigma=10$ ; the maximum number of UEs is set to 50. For simplicity reasons, no mobility is considered; the reason for this is that since the channel can considerably vary when using mmWave, an approach as the one proposed at these frequencies would be limited to stationary or semi-stationary environments. The size of the packet to be transmitted to the scheduled UEs is chosen randomly following a uniform distribution, with values ranging from 16 to 97880 bits (corresponding to the minimum and maximum TBSs defined in the LTE standard [3GP17a]). At the scheduler, if the TBS assigned to a UE is not enough to send the whole packet, a number of bits equal to the TBS are sent and the rest of the bits are buffered. The Inter-Arrival Time (IaT) was set to 1 ms, meaning that in every subframe a new packet arrived for the UEs.

For the PRBs allocation, the UEs are divided into four groups according to the modulation (i.e., QPSK, 16QAM, 64QAM, and 256QAM) and the PRBs assigned to each group are proportional to the number of UEs; a minimum of two PRBs per UE is set. The same UE grouping is used for the pairing algorithm, where a semi-exhaustive search approach is used to find the UEs to be paired. Each UE group is assigned a value of  $n=2^m$  with  $m=\{2,4,6,8\}$  representing the modulation order of the group. The search for pairs is done between two groups at the time and all the group combinations are evaluated, which lead to a total of six search iterations; in this way, it is guaranteed that the UEs being considered for pairing will not have the same modulation order, thus avoiding pairing UEs with highly similar channel gains. Figure 5.16 shows the group combination for each search iteration. In the first iteration, the search is done between the two groups with the highest difference in the  $n$  value,  $\Delta n$  (i.e., group 1 and group 4, with  $\Delta n=252$ ). During the search, each unpaired UE in the group with the higher  $m$  is evaluated with each unpaired in the other group. In this evaluation is where the MCS adjustment and extra  $P_{TX}$  take place; two UEs are considered a good pair when the following conditions are met:

- **Condition 1:** the data rates of both UEs must not be degraded after adjusting the MCS to a more robust value.
- **Condition 2:** an MCS corresponding to a QPSK modulation must be assigned to the UE with the lower channel gain after the MCS adjustment.
- **Condition 3:** the extra  $P_{TX}$  needed does not exceed the power headroom available.

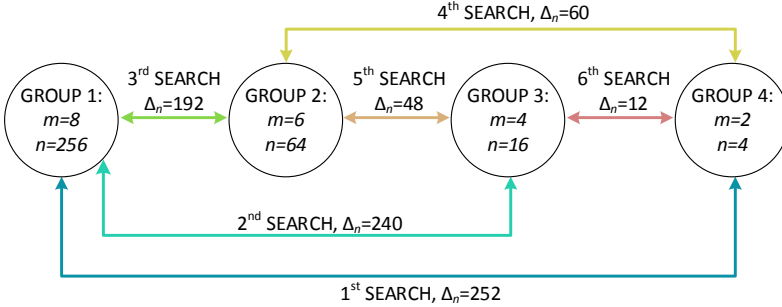


Figure 5.16: UEs group combination for each iteration of the pairing process proposed for NOMA.

After evaluating all possible pairs between two groups, the pair search process continues with the other group combinations in a descending order according to  $\Delta n$ . When all the group combinations are evaluated, the search for NOMA pairs in the cell ends; this pairing search process gives priority to the NOMA pairs with higher channel difference. In Figure 5.17 the proposed pairing algorithm is shown.

For those UEs that are left unpaired after the algorithm is ran in the scheduler, an OMA transmission is used with the same resources assigned during the preliminary allocation. The simulation parameters are summarized in Table 5.4; here it can be seen that the power allocation for the NOMA users is chosen to be fixed instead of dynamic with 0.25 for  $UE_2$  (UE with lower channel gain) and 0.75 for  $UE_I$ .

---

**PROPOSED PAIRING ALGORITHM**


---

**INPUT:**

UEs groups  
Maximum allowed extra transmission power (ePtx\_max)  
Preliminary PRBs, MCS and TBS assignment for all UEs in each group

```

1: for iteration=1 to 6 do
2:   Select the two groups for the search=iteration
3:   Group 1 = group with higher modulation order
4:   Group 2 = group with lower modulation order
5:   M=number UEs in Group 1, N= number UEs in group2, m=1, n=1
6:   while m<=M do
7:     if  $UE_m$  not paired then
8:       while n<=N do
9:         if  $UE_n$  not paired do
10:          TPRBs = Sum preliminary PRBs for  $UE_m$  and  $UE_n$ 
11:          if MCS adjustment is activated do
12:            Find new MCS < preliminary MCS, new TBS>=preliminary TBS for  $UE_m$ 
              and  $UE_n$  with TPRBs and throughput constraints
13:          end if
14:          if new modulation for  $UE_n$  equals QPSK do
15:            Calculate extra transmission power needed, ePtx
16:            if  $ePtx \leq ePtx\_max$  do
17:              Pair  $UE_m$  with  $UE_n$ 
18:              Update MCS, TBS, PRBs, and extra transmission power for  $UE_m$  and  $UE_n$ 
19:            end if
20:          end if
21:        end if
22:        n=n+1
23:      end while
24:    end if
25:    m=m+1
26:  end while
27: end for

```

**OUTPUT:** UEs paired and their new MCS, new TBS, and ePtx for each pair

---

Figure 5.17: Proposed pairing algorithm based on MCS adjustments and extra transmission power allocation for NOMA.

Table 5.4: Simulation parameters for a hybrid MA system with MCS adjustments and extra transmission power allocation for NOMA.

	OMA	NOMA
CARRIER FREQUENCY (GHz)	73	
CARRIER BANDWIDTH (MHz)	20	
CARRIER COMPONENTS	40	
CODING/DECODING	TURBO CODING	
MODULATION SCHEME	QPSK, 16QAM, 64QAM, 256QAM	
TRANSMISSION POWER (dBm)	15	
POWER HEADROOM (dBm)	15	
POWER ALLOCATION PER UE	1	UE <sub>1</sub> =0.25
		UE <sub>2</sub> =0.75
WAVEFORM	OFDM	
TRANSMISSION MODE	SU-MIMO (2X2)	
TRANSMITTER ANTENNA GAIN (dBi)	37	
PATH LOSS MODEL	mmWave OUTDOOR FOR 73 GHz ACCESS (*)	
CHANNEL ESTIMATION	MMSE	
RECEIVER ANTENNA GAIN (dBi)	0	
RECEIVER NOISE FIGURE (dB)	6	
RECEIVER SCHEME	LMMSE	UE <sub>1</sub> = SLIC UE <sub>2</sub> = LMMSE

\* SEE EQUATION 4.8

Choosing fixed values for the power allocation with NOMA can bring significant savings in the signaling overhead while still offering high capacity gains [Ben+14]; therefore, the proposed system was evaluated for different power sets and for a dynamic power allocation following Equation 5.5, concluding that, for the used setup, a set  $(P_1, P_2) = (0.25, 0.75)$  offers a good trade-off between capacity gain and power allocation complexity. For the path losses, the same model for mmWave outdoor at 73 GHz presented in Section 5.2 was used, but in this case, both LOS and NLOS scenarios are considered, thus such losses can be estimated as [Rap+14]:

$$P_L = 69.8\text{dB} + A\log(d) + x_\sigma \quad (5.8)$$

where  $x_\sigma$  represents the shadowing factor and it is a random Gaussian variable with mean zero and standard deviation  $\sigma = 5.2\text{dB}$  and  $\sigma = 7.6\text{dB}$  for LOS and NLOS, respectively. For LOS  $A=20$ , whereas for NLOS  $A=33$ .

### 5.3.2 Performance evaluation

For the system evaluation, six study cases are defined with different percentages of the  $P_{TX}$  allowed as extra power for the NOMA UEs and with or without MCS adjustment. For Case 1 no extra  $P_{TX}$  is allowed, therefore only the MCS adjustment is used to improve the performance of NOMA UEs; in Case 6 no MCS adjustment is used and the system relies only on extra  $P_{TX}$  for the NOMA transmissions. The study cases are summarized in Table 5.5. For all the results an OMA system was used as a benchmark and the values shown were averaged from 65 runs, with 100 subframes being sent in each run.

Table 5.5: Study cases for a hybrid MA system with MCS adjustments and extra transmission power allocation for NOMA.

STUDY CASE	1	2	3	4	5	6
MCS ADJUSTMENT	YES	YES	YES	YES	YES	NO
EXTRA $P_{TX}/P_{TX}$ (%)	N/A	10	25	50	75	100

First, we focus on the UEs data rates which are shown in Figure 5.18. It can be seen that for 90% of the UEs a gain of up to 9.4x is experienced for Case 6, with some UEs even experiencing a gain of 30x. The reason for the high increase, in this case, is that there is no limit in the extra  $P_{TX}$  that can be allocated to the NOMA UEs, therefore two UEs which require a large amount of extra  $P_{TX}$  can still be paired. Hence, their data rates will significantly increase since NOMA is more effective as the difference in the channel gains is larger. In contrast, Case 2 was the one that offered the lowest performance, with no UEs paired; the reason being that the maximum 10% extra  $P_{TX}$  for this case was not enough to make up for the extra dBs needed in the SINR.

Cases 4 and 5 showed a similar performance, with a gain of 6x and 2x for 90% and 50% of the UEs, respectively. For Cases 1 and 3, 80% of the UEs experienced a 2x increase whereas for 90% a gains of up to 2.6x and 3.6x were experienced, respectively for each case. In Table 5.6 the UE data rates are summarized for the 90th, 80th, and 50th percentile. With these results, we can verify the proposed method allows for significant improvements in the UEs data rates.

To now analyze the system capacity, we can refer to Figure 5.19a where it can be seen that the best performance was achieved in Case 5 with a 78% capacity gain over an OMA implementation, corresponding to a channel capacity of approximately 13 Gbps. The pairing probability (Figure 5.19b) for this case was also the highest, with a 0.48 probability; this means that almost 50% of the UEs were paired and NOMA was used for their transmissions, which reflects directly on the high capacity gain of the system.

The high value of extra  $P_{TX}$  allowed in Case 5 (i.e., 75%) is the reason behind reaching such high gain, since it is more likely to reach the extra dBs needed in the SINR for NOMA for more pairs. For Case 4 a highly similar performance from that of Case 5 was achieved, with 72% system capacity gain and 0.42 pairing probability. A high pairing

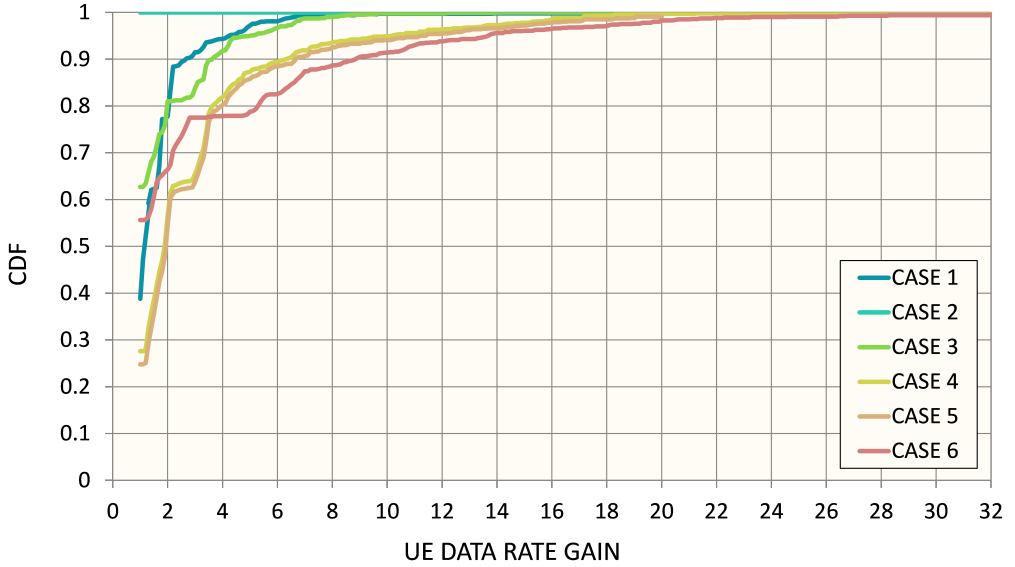


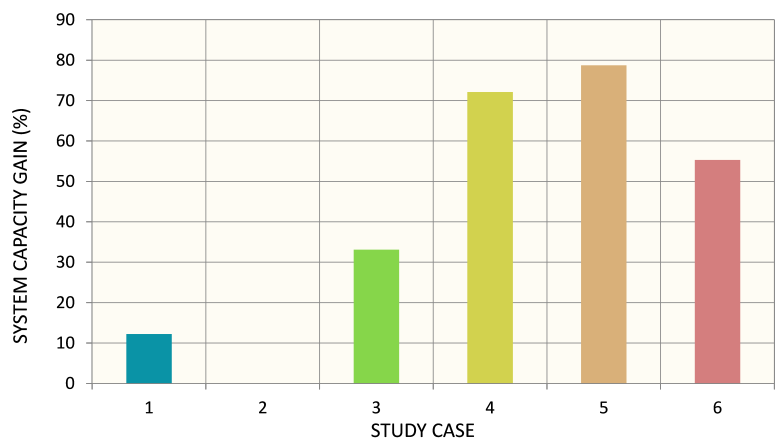
Figure 5.18: Cumulative probability for the UEs data rate gain for a hybrid MA system with MCS adjustments and extra transmission power allocation for NOMA.

Table 5.6: UE data rate gain for the 90<sup>th</sup>, 80<sup>th</sup>, and 50<sup>th</sup> percentile for a hybrid MA system with MCS adjustments and extra transmission power allocation for NOMA.

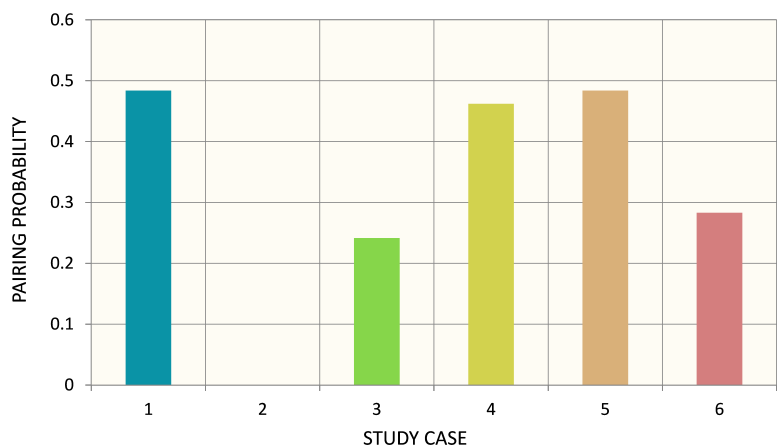
		Case					
		1	2	3	4	5	6
Percentile	90 <sup>th</sup>	2.7	1	3.8	6.3	6.7	9.4
	80 <sup>th</sup>	2.1	1	2	3.7	4	5.3
	50 <sup>th</sup>	1.1	1	1	1.9	1.9	1

probability was also experienced in Case 1 with a 0.48 probability; however, as no extra  $P_{TX}$  is allowed in this case, the gains in the UEs data rates and the system capacity are limited in comparison to the other cases. Only a 12% system capacity gain was achieved for Case 1, which indicates that even when many UEs are paired, their performance is not improved because of the system limitation to accurately compensate for the extra intra-cell interference in NOMA; this confirms what was addressed above that for many cases the MCS adjustment alone is not enough. For Case 2, the system capacity did not change since no suitable pairs were found due to the low allowed extra  $P_{TX}$ ; for Case 3,





(a) Capacity gain



(b) Pairing probability

Figure 5.19: System capacity gain and pairing probability for a hybrid MA system with MCS adjustments and extra transmission power allocation for NOMA.

the system did benefit from the proposed method with a 33% capacity gain and a pairing probability of 0.24. A similar pairing probability of 0.28 was achieved with Case 6 for a system capacity gain of 53%. The reason behind the similar pairing probability of Cases 4 and 6 is that, although 100% extra  $P_{TX}$  is set in Case 6, there is no MCS adjustment. As stated earlier, one of the pairing conditions is that the UE with lower channel gain has to have a QPSK modulation after the MCS adjustment; with such adjustment some UEs with higher order modulation can be assigned a QPSK modulation, which increases the probability of the UEs to be candidates for NOMA.

Furthermore, we can refer to Figure 5.20 to analyze the system BLER; we can see that although for all cases 95% of the UEs experiences a BLER below 10%, for Cases 1 and 6 the BLER was higher. This was expected since in both cases there are limitations related to the improvements of the UEs channel conditions for NOMA, resulting in a higher probability of erroneous decoding. For the remaining cases the BLER for all the UEs was below 10%.

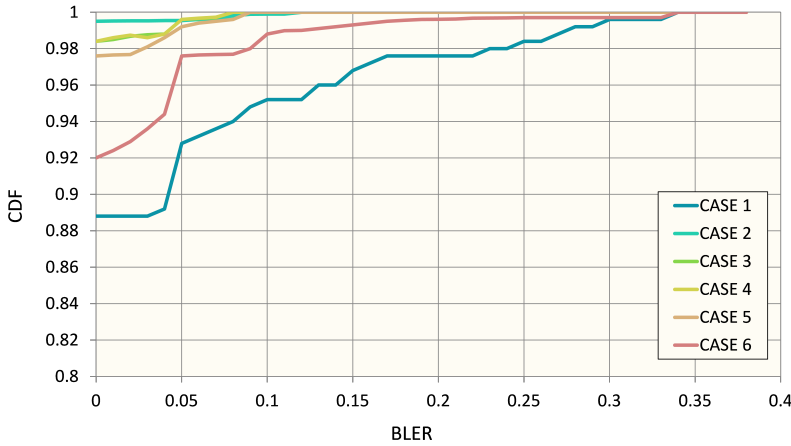


Figure 5.20: BLER for a hybrid MA system with MCS adjustments and extra transmission power allocation for NOMA.

From a power consumption point of view, the results are shown in Figure 5.21, where it can be seen that Case 6 requires an average of 73% extra  $P_{TX}$  even though there is no limit on how much extra power can be used in this case. For Case 2 a higher power allocation should be considered if the main concern relies on the overall system capacity, since when accompanied with the MCS adjustment a higher pairing probability can be obtained and thus a higher system capacity. Case 3 requires on average 14% of extra  $P_{TX}$ , which is 11% less than the maximum allowed; this is an indication that 25% maximum of extra  $P_{TX}$  offers a good trade-off between power consumption and system performance. Cases 4 and 5 also require on average less power than the maximum allowed, with 18%

and 19%, respectively. These results confirm that no more than 14-19% of extra  $P_{TX}$  would be necessary to expect a significant performance improvement.

From these results, it can be concluded that the best trade-offs between the UEs data rates, overall system capacity, BLER, and power consumption are achieved when a hybrid MA system is implemented along with MCS adjustment and extra  $P_{TX}$  allocation for the NOMA transmissions; that is as long as no degradation is allowed in the data rates for the NOMA UEs.

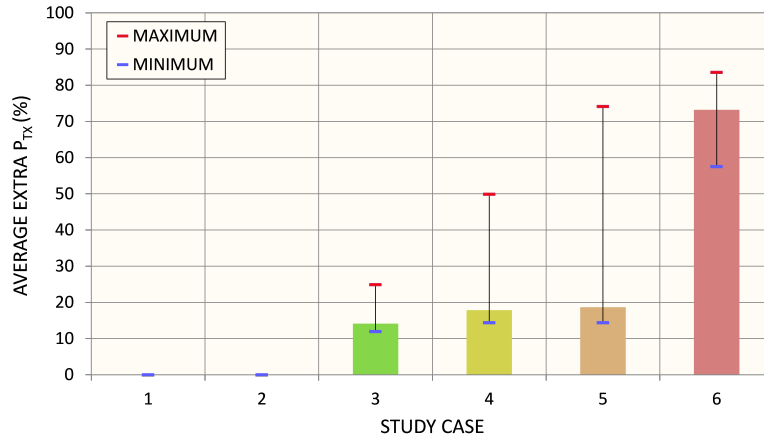


Figure 5.21: Extra transmission power consumption for a hybrid MA system with MCS adjustments and extra transmission power allocation for NOMA.

In the case of applying MCS adjustment but not extra  $P_{TX}$  allocation, the BLER values might be higher than the maximum desired, which eventually affects the QoS since the number of retransmissions will increase. If on the contrary, no MCS adjustment is performed, and no limit is set to the maximum  $P_{TX}$ , there will be an improvement in the system performance, but not a significant one if compared with other cases, at the expenses of using on average 73% extra  $P_{TX}$ .

Through this results analysis, the implementation of a hybrid MA system with a pairing algorithm based on MCS adjustment and an allowed extra  $P_{TX}$  power allocation between 14%-19% (e.g., cases 3, 4 and 5) is recommended. With this configuration an average gain between 1.75x-3.31x can be expected for the UE data rates, the overall system capacity could also improve with gains between 33-78%, corresponding to approximately 10-13 Gbps, and the BLER will remain below 10% for all UEs.

With this proposed method a significant gain can be achieved in the overall system capacity as long as a power headroom is available. Moreover, the only additional signaling needed, in comparison with any NOMA implementation, is to verify whether the MCS adjustment was enough for the UE to achieve the desired SINR with NOMA or if extra  $P_{TX}$

is needed. Hence, the impact of this method in the signaling overhead is expected to be low. If no power headroom is available then the MCS adjustment can still be implemented at the scheduler but alternative scheduling/pairing processes should be considered in order to guarantee a performance improvement.

The implementation of NOMA is not a trivial task; therefore, further work related to solutions of how to mitigate the effects of the intra-cell interference in NOMA, especially in hybrid MA cases, is very much needed. What it is for sure, is that some type of adjustment needs to be considered in order reach an optimal point of capacity gain and QoS delivered. In Section 6.4 an alternative approach for a NOMA implementation is analyzed.

## 5.4 User pairing methods evaluation

The appeal of NOMA for 5G networks relies on its more effective utilization of scarce resources (e.g., spectrum) than 4G; therefore, to really exploit the capacity benefits offered by NOMA, resource management must be done in the most effective possible way. In NOMA, there are three resources that must be carefully allocated: power, frequency and time. Since a group of UEs will be assigned to the same frequency channel during the same time, such UEs must be chosen to guarantee that there will be a capacity gain and that resources will not be wasted. Moreover, the power allocation for each multiplexed UE in NOMA must also be carefully chosen to allow the correct decoding of the signals on the receiver side. Both user-pairing and power allocation, are complex processes that require optimization algorithms to allow for the best results with the minimum resources. Some research works have been focused on these two processes, as outlined in the following.

The work in [DFP16] deals with user pairing for two NOMA systems: NOMA with fixed power allocation and cognitive-radio-inspired NOMA. Results show that each of these systems exhibits a different behavior when selecting the UEs to be paired, and that the gains of fixed power NOMA over OMA can be further increased by selecting UEs whose channel conditions are more distinctive. In [LOP15], a user pairing and power allocation approach based on a proportional fair (PF) metric is used to achieve a balance between transmission efficiency and user fairness. The proposed scheme offers low computational complexity by deriving the prerequisites for user pairing and avoiding comparison of candidate user pairs.

The authors in [MS17] propose three user pairing methods based on the CQI; results are presented for cases with perfect and imperfect SIC and compared with OMA. Matching theory is proposed in [Lia+17] as an approach to optimize user pairing and power allocation in the downlink in a cognitive radio NOMA; the results show that the low complexity proposed algorithm results in a stable matching and outperforms an OMA system. In [Sha+16] two user pairing strategies are proposed, where all the users, including those in the middle of the cell who are typically left unpaired, are considered; results show that the proposed algorithm can outperform the near-far pairing, especially in scenarios

with imperfect SIC. In [Son+17] a comprehensive review of resource management in NOMA is presented; here the authors propose a resource management framework based on game-theoretic models for power-domain and code-domain NOMA. Moreover, our work from Paper D proposes a pairing method based on the UEs CQI and on MCS adjustments and extra transmission power allocation to help mitigate the extra intra-cell interference added in NOMA; the details of our work are explained in Section 5.3.

Nevertheless, the user pairing in NOMA is still at its early stages and since in HetNets with hybrid MA the decision of using OMA or NOMA relies on the pairing algorithm, research work on the topic is highly anticipated. Therefore, and as part of this thesis contribution, four generic pairing algorithms are compared and the use of a cost matrix to help in the pairing selection is proposed. The main idea of such comparison is to determine how they will affect the complexity of the pairs search and how much gain they can provide; the advantage of using generic algorithms is that they can allow for a rapid rollout of NOMA. The analysis presented in this section is part of our research work included in Paper F.

#### 5.4.1 Single cell site model

For the pairing evaluation, UEs are divided into two groups in the scheduler. Group A corresponds to those UEs that have been already selected by the scheduler to transmit in the following subframe; we refer to this as pre-scheduling. Group B corresponds to those UEs that are in need of resources but were not selected to transmit during the pre-scheduling because of lack of resources. No particular order is used to sort the UEs in their corresponding groups.

A proportional fair scheduling algorithm is used and the priority of each UE is assigned according to the following metric,  $PF_i[t]$  [Mia+16]:

$$PF_i[t] = \frac{R_i[t]}{S_i[t-1]} \quad (5.9)$$

where  $t$  is the subframe number,  $R_i[t]$  is the target data rate and it depends on traffic model for the application in use by  $UE_i$ , and  $S_i[t-1]$  is the average experienced data rate and can be estimated as:

$$S_i[t] = \frac{t-1}{t} S_i[t-1] + \frac{1}{t} R_i[t] \quad (5.10)$$

The UEs are selected by the scheduling based on their priority, with those with the highest  $PF_i[t]$  being scheduled first.

For the first pairing approach, the pairing can be treated as an assignment problem, using the Hungarian method [KB55] to find an optimal solution by which the systems gets the maximum capacity. For the second approach, the Gale-Shapley algorithm [GS62] is used to find a stable pairing. Unlike the Hungarian method that finds the optimal solution by minimizing (or maximizing) a cost associated with a set of pairs, the Gale-Shapley

algorithm finds an optimal solution based on the stable marriage criterion. Up to date, no work on user pairing for NOMA in HetNets has evaluated these algorithms. The third algorithm is a simple random pairing, in which UEs from Group A choose the best available pair from Group B. The fourth algorithm is an exhaustive search over all possible pair combinations.

For all four algorithms, a cost matrix that reflects the cost of each possible pair is first generated. Table 5.7 shows an example of such matrix. Here  $C_{i,j}$  is the cost function associated with the expected sum rate of the NOMA UEs paired, and can be represented as:

$$C_{i,j} = \frac{1}{1 + R_T \Delta_{\text{SINR}}}; \quad \text{with } i \leq n; \quad n < j \leq m \quad (5.11)$$

where  $n$  is the number of UEs in Group A and  $m-n$  is the number of UEs in Group B;  $R_T$  is the sum of the  $\text{UE}_i$  and  $\text{UE}_j$  data rates (see Equations 3.5 and 3.6) and  $\Delta_{\text{SINR}}$  is their difference in channel gain. Both values can be calculated respectively as:

$$R_T = R_i + R_j \quad (5.12)$$

$$\Delta_{\text{SINR}} = |\text{SINR}_i - \text{SINR}_j| \quad (5.13)$$

Table 5.7: Cost matrix proposed for pairing methods evaluation in NOMA.

		GROUP B									
		$\text{UE}_{n+1}$	$\text{UE}_{n+2}$	$\text{UE}_{n+3}$	$\cdot$	$\cdot$	$\cdot$	$\text{UE}_j$	$\cdot$	$\cdot$	$\text{UE}_m$
GROUP A	$\text{UE}_1$	$C_{1,n+1}$	$C_{1,n+2}$	$C_{1,n+3}$	$\cdot$	$\cdot$	$\cdot$	$C_{1,j}$	$\cdot$	$\cdot$	$C_{1,m}$
	$\text{UE}_2$	$C_{2,n+1}$	$C_{2,n+2}$	$C_{2,n+3}$	$\cdot$	$\cdot$	$\cdot$	$C_{2,j}$	$\cdot$	$\cdot$	$C_{2,m}$
	$\text{UE}_3$	$C_{3,n+1}$	$C_{3,n+2}$	$C_{3,n+3}$	$\cdot$	$\cdot$	$\cdot$	$C_{3,j}$	$\cdot$	$\cdot$	$C_{3,m}$
	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
	$\text{UE}_i$	$C_{i,n+1}$	$C_{i,n+2}$	$C_{i,n+3}$	$\cdot$	$\cdot$	$\cdot$	$C_{i,j}$	$\cdot$	$\cdot$	$C_{i,m}$
	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
GROUP A	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$
	$\text{UE}_n$	$C_{n,n+1}$	$C_{n,n+2}$	$C_{n,n+3}$	$\cdot$	$\cdot$	$\cdot$	$C_{n,j}$	$\cdot$	$\cdot$	$C_{n,m}$

The selection of this cost function aims at facilitating the pairing mainly for the Hungarian and Gale-Shapley algorithms, although it also applies to the other methods.

The two former methods are defined to minimize the cost associated with certain pair selection; for this, they give preference to the pairs whose cost is lower (as explained ahead). Because the rate gain in NOMA is higher as the channel gain difference between the paired UEs increases, as  $\Delta_{\text{SINR}}$  increases so does  $R_T$  and thus the cost function proposed approximates to zero. This means that as the paired UEs offer higher sum rates, a lower cost will be associated to them, giving such pair a higher probability of being chosen during the pairing process due to their lost cost.

However, when calculating  $C_{i,j}$  the following restrictions apply:

(i) Guarantee that the pairing in NOMA will result in a capacity gain or that, at least, the capacity will be the same as with OMA. This can be achieved by having:

$$R_T \geq R_i$$

with  $R$  being the rate of the pre-scheduled  $\text{UE}_i$ , with  $i \leq n$ .

(ii) Guarantee that the unscheduled  $\text{UE} \in \text{Group B}$  has higher channel gain than the pre-scheduled  $\text{UE} \in \text{Group A}$ ; this will prevent damaging the QoS of pre-scheduled UEs with high channel gains that benefit the performance of the system at the expenses of scheduling UEs with low channel gain. To guarantee this, the MCS index is considered:

$$I'_{\text{MCS}_i} \leq I'_{\text{MCS}_j}$$

where  $I'_{\text{MCS}_i}$  and  $I'_{\text{MCS}_j}$  are the MCS index of  $\text{UE}_i$  and  $\text{UE}_j$  after the MCS adjustments, respectively.

For the MCS adjustments, a different approach than that presented in Section 5.3 is used. In this case, the methodology proposed in [Ben+14] is used, in which the BS adjusts the CQI reported based on OMA (and thus the MCS) so it can approximate to an adequate value for NOMA as follows:

$$CQI'_i = \frac{P_2 CQI_i}{P_1 CQI_i + 1} \quad (5.14)$$

$$CQI'_j = P_1 CQI_j \quad (5.15)$$

where  $CQI'_i$  and  $CQI'_j$  are the estimated CQIs for NOMA, and  $CQI_i$  and  $CQI_j$  are the reported CQIs for  $\text{UE}_i$  and  $\text{UE}_j$ , respectively. The values of  $P_1$  and  $P_2$  are calculated according to Equation 5.5. The reason for choosing this alternative MCS adjustment method is because it can offer higher accuracy when dynamic power allocation is used, since it uses the power ratio for the adjustment.

The UE pairs that do not fulfill the restrictions (i) and (ii) are considered as non-suitable pairs; thus, in the cost matrix, a cost much higher than the maximum  $C(i, j)$  for the suitable pairs is assigned, so that such pairs have the lowest probability of being considered during the pair selection. Once the cost matrix is generated, we proceed

with running the algorithms to find the best pairs that minimize the cost and therefore maximize the system capacity. The number of pairs should be equal to the number of UEs in the group with fewer members, unless there are two or more UEs that can only be paired with the same UE from the opposite group, because of the pairing restrictions set. In case of the latter, there will be fewer pairs than expected. For the Hungarian and Gale-Shapley algorithms, if the number of UEs in Group A and Group B are not equal, dummy rows/columns should be used to generate a square matrix, since both algorithms require square matrixes to find the best solution.

### 5.4.2 Implementation example

To exemplify the use of the considered pairing methods in a single cell, let us assume that we have a total of six UEs; let us also assume that all six UEs need resources in the following subframe and that after the pre-scheduling and the grouping, UEs are divided as shown in Figure 5.22. After calculating the cost of each possible pair according to Equation 5.11 and assuming that conditions (i) and (ii) are fulfilled, the proposed cost matrix would look as shown in Table 5.8. The following subsections explain the pairing process and their complexity (i.e., number of iterations) for all four methods considering the example in Figure 5.22.

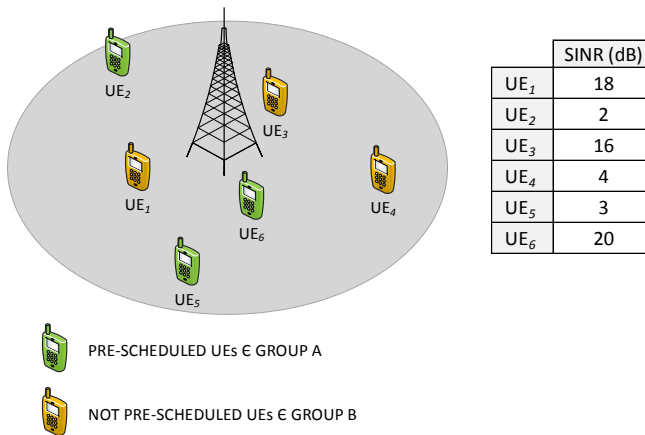


Figure 5.22: Single cell scenario for comparing pairing methods with NOMA.

### Hungarian pairing

The Hungarian method is a combinatorial optimization algorithm used for solving a two-sided one-to-one matching problem. For this method, the problem can be mathematically expressed as:



Table 5.8: Cost matrix for single cell scenario for comparing pairing methods with NOMA.

		B		
		UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>
A	UE <sub>2</sub>	0.03	0.04	0.25
	UE <sub>5</sub>	0.03	0.03	0.38
	UE <sub>6</sub>	0.1	0.05	0.01

$$\text{Minimize} \quad \sum_{i=1}^n \sum_{j=1}^m C_{i,j} x_{i,j}$$

where

$$x_{i,j} = \begin{cases} 1; & \text{if the UE}_i \text{ is already paired with UE}_j \\ 0; & \text{if the UE}_i \text{ is not paired with UE}_j \end{cases}$$

with the restrictions:

- (iii)  $\sum_{i=1}^n x_{i,j} = 1$ ; with  $j = \{1, 2, \dots, n\}$ , to guarantee that the UE<sub>i</sub> only has one pair.
- (iv)  $\sum_{j=1}^m x_{i,j} = 1$ ; with  $i = \{1, 2, \dots, n\}$ , to guarantee that the UE<sub>j</sub> only has one pair.

In Figure 5.23 the basic steps of the Hungarian algorithm are applied to the cost matrix in Table 5.8, until obtaining the final matrix from which pairs are selected by choosing those with  $C_{i,j} = 0$ . From here it can be seen that a total of three pairs were found as required by the algorithm (i.e., the number of pairs has to be the same as the cost matrix dimension). If after the row and column reduction the number of pairs is not optimal, further steps are taken to optimize the solution. Such steps can be found in detail in [KB55].

For the analysis, the complexity of this algorithm is calculated as the number of iterations needed to find the optimal pairing. The execution of each step defined by the algorithm is considered an iteration (e.g., each row/column reduction, each zero assignment). If after the algorithm has been run some of the found pairs are those previously defined in the cost matrix as non-suitable, such pairs are omitted during the scheduling and the UEs involved from Group A continue with OMA, whereas the ones from Group B are not scheduled.

### Gale-Shapley pairing

The Gale-Shapley algorithm uses the stable marriage criterion to find stable assignments (i.e., pairs). Once the cost matrix is generated, each UE in Group A sorts the UEs

**STEP 1**

Row reduction: for each row, subtract the minimum value of the row from all the elements in that row. Iterations = 3

		B		
		UE1	UE3	UE4
A	UE2	0	0.01	0.22
	UE5	0	0	0.35
	UE6	0.09	0.04	0

**STEP 2**

Column reduction: for each column, subtract the minimum value of the column from all the elements in that column. Iterations = 3.


		B		
		UE1	UE3	UE4
A	UE2	0	0.01	0.22
	UE5	0	0	0.35
	UE6	0.09	0.04	0

**STEP 3**

Zero assignment: starting from the first row, find the rows with only one non-selected zero and select the corresponding pair. Cross out all other zeros in the row and column where the pair was selected. Iterations = 3.

		B		
		UE1	UE3	UE4
A	UE2	0	0.01	0.22
	UE5	0	0	0.35
	UE6	0.09	0.04	0

 Selected pair

 Crossed-out pair

**SOLUTION**

UE2 → UE1; UE5 → UE3; UE6 → UE4

**COMPLEXITY**

TOTAL ITERATIONS = 9

Figure 5.23: Example of the Hungarian method application for user pairing in NOMA.

in Group B in order of preference, based on the cost functions defined in Equation 5.11. The lower the cost the higher the preference. The iterative algorithm is then applied to the sorted preference, during which the UEs from Group A “propose” as a pair to the UEs in Group B. The UEs in the latter either accept (if they are unpaired) or reject (if they are paired and prefer their current pair to the one proposing) the proposition. The solution is said to be stable if, and only if, there exists no  $UE_i$  and  $UE_j$  who are not paired with each other but who would both prefer each other over their present partners.

Assuming the same cell scenario shown in Figure 5.22, we apply the Gale-Shapley algorithm to the matrix in Table 5.8, until obtaining the final pairing for a total of three pairs. The steps applied are shown in Figure 5.24.

The complexity of this algorithm is calculated as the iterations for the UEs ranking plus the number of proposals done (accepted and rejected) until the stable solution is found. Even though the solution provided by the Gale-Shapley algorithm is stable, it is not necessarily an optimal solution. In general, there are several solutions to the pairing when applying this algorithm [MW71]; the solutions depend on the group that does the proposal. Then, and if we follow the dynamic explained above, the stable solution is optimal for the UEs in Group A, but not necessarily for the UEs in Group B. Similarly, if the proposal is done by the UEs in Group B, the solution would be optimal for those UEs. It could, however, be the case that the stable optimal solution is the same regardless of which group proposes. Same as with the Hungarian method, if the stable solution considers pairs that have been marked as non-suitable, such pairs are ignored during the

<b>STEP 1</b> UEs rank the UEs in the opposite group in order of preference. Iterations = 6.	Group A preferences by rank				Group B preferences by rank			
	UE <sub>2</sub>	UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>	UE <sub>1</sub>	UE <sub>2</sub>	UE <sub>5</sub>	UE <sub>6</sub>
	UE <sub>5</sub>	UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>	UE <sub>3</sub>	UE <sub>5</sub>	UE <sub>2</sub>	UE <sub>6</sub>
	UE <sub>6</sub>	UE <sub>4</sub>	UE <sub>3</sub>	UE <sub>1</sub>	UE <sub>4</sub>	UE <sub>6</sub>	UE <sub>2</sub>	UE <sub>5</sub>
<b>STEP 2</b> UE <sub>2</sub> proposes to UE <sub>1</sub> ; UE <sub>1</sub> is unpaired and accepts. Iterations = 1.	Group A preferences by rank				Group B preferences by rank			
	UE <sub>2</sub>	UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>	UE <sub>1</sub>	UE <sub>2</sub>	UE <sub>5</sub>	UE <sub>6</sub>
	UE <sub>5</sub>	UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>	UE <sub>3</sub>	UE <sub>5</sub>	UE <sub>2</sub>	UE <sub>6</sub>
	UE <sub>6</sub>	UE <sub>4</sub>	UE <sub>3</sub>	UE <sub>1</sub>	UE <sub>4</sub>	UE <sub>6</sub>	UE <sub>2</sub>	UE <sub>5</sub>
<b>STEP 3</b> UE <sub>5</sub> proposes to UE <sub>1</sub> ; UE <sub>1</sub> is paired with UE <sub>2</sub> and prefers UE <sub>2</sub> over UE <sub>5</sub> , so it rejects the pairing. UE <sub>5</sub> then proposes to UE <sub>3</sub> ; UE <sub>3</sub> is unpaired and accepts. Iterations = 2.	Group A preferences by rank				Group B preferences by rank			
	UE <sub>2</sub>	UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>	UE <sub>1</sub>	UE <sub>2</sub>	UE <sub>5</sub>	UE <sub>6</sub>
	UE <sub>5</sub>	UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>	UE <sub>3</sub>	UE <sub>5</sub>	UE <sub>2</sub>	UE <sub>6</sub>
	UE <sub>6</sub>	UE <sub>4</sub>	UE <sub>3</sub>	UE <sub>1</sub>	UE <sub>4</sub>	UE <sub>6</sub>	UE <sub>2</sub>	UE <sub>5</sub>
<b>STEP 4</b> UE <sub>6</sub> proposes to UE <sub>4</sub> ; UE <sub>4</sub> is unpaired and accepts. Iterations = 1.	Group A preferences by rank				Group B preferences by rank			
	UE <sub>2</sub>	UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>	UE <sub>1</sub>	UE <sub>2</sub>	UE <sub>5</sub>	UE <sub>6</sub>
	UE <sub>5</sub>	UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>	UE <sub>3</sub>	UE <sub>5</sub>	UE <sub>2</sub>	UE <sub>6</sub>
	UE <sub>6</sub>	UE <sub>4</sub>	UE <sub>3</sub>	UE <sub>1</sub>	UE <sub>4</sub>	UE <sub>6</sub>	UE <sub>2</sub>	UE <sub>5</sub>
<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="background-color: #90EE90; width: 20px; height: 10px; display: inline-block;"></div> Selected pair <div style="background-color: #FFD700; width: 20px; height: 10px; display: inline-block;"></div> Crossed-out pair </div>								
<b>SOLUTION</b> UE <sub>2</sub> → UE <sub>1</sub> ; UE <sub>5</sub> → UE <sub>3</sub> ; UE <sub>6</sub> → UE <sub>4</sub>		<b>COMPLEXITY</b> TOTAL ITERATIONS = 10						

Figure 5.24: Example of the Gale-Shapley method application for user pairing in NOMA.

scheduling process.

### Random and exhaustive pairing

In the random method, UEs in Group A simply choose the best unpaired UE in Group B, according to the cost of each pair. For the cost matrix in Table 5.8, the pair choosing process can be seen in Figure 5.25. The complexity of this method is calculated as the number of iterations needed until all the pairs have been found.

For the exhaustive pairing, all possible pairs are evaluated to find the combination of pairs that yields the minimum cost. Although the solution from this method is the optimal solution, its complexity makes it computationally expensive. For a cost matrix of size  $C(n, m)$  with  $n \geq m$ , a total of  $n!/(n - m)!$  iterations are needed to evaluate all the pairs. Each iteration corresponds to the evaluation of one of the possible permutations. When applied to the cost matrix from Table 5.8, the pairing from the exhaustive search is shown in Figure 5.26; six iterations are needed with this method. Although the number of iterations for this example is lower with the exhaustive method, this would not be the case as the dimensions of the cost matrix increase.

Starting from the first row, each UE in Group A chooses the best free pair from Group B.

		B		
		UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>
A	UE <sub>2</sub>	0.03	0.04	0.25
	UE <sub>5</sub>	0.03	0.03	0.38
	UE <sub>6</sub>	0.1	0.05	0.01

 Selected pair

SOLUTION	COMPLEXITY
UE <sub>2</sub> → UE <sub>1</sub> ; UE <sub>5</sub> → UE <sub>3</sub> ; UE <sub>6</sub> → UE <sub>4</sub>	TOTAL ITERATIONS: 3

Figure 5.25: Example of the random method application for user pairing in NOMA.

OPTION 1

COST = 0.07

ITERATIONS = 1

		B		
		UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>
UE <sub>2</sub>		0.03	0.04	0.25
UE <sub>5</sub>		0.03	0.03	0.38
UE <sub>6</sub>		0.1	0.05	0.01

OPTION 2

COST = 0.46

ITERATIONS = 1

		B		
		UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>
UE <sub>2</sub>		0.03	0.04	0.25
UE <sub>5</sub>		0.03	0.03	0.38
UE <sub>6</sub>		0.1	0.05	0.01

OPTION 3

COST = 0.08

ITERATIONS = 1

		B		
		UE <sub>1</sub>	UE <sub>3</sub>	UE <sub>4</sub>
UE <sub>2</sub>		0.03	0.04	0.25
UE <sub>5</sub>		0.03	0.03	0.38
UE <sub>6</sub>		0.1	0.05	0.01

⋮

Continue until evaluating all six combinations. Select the one with lowest cost.

SOLUTION	COMPLEXITY
UE <sub>2</sub> → UE <sub>1</sub> ; UE <sub>5</sub> → UE <sub>3</sub> ; UE <sub>6</sub> → UE <sub>4</sub>	TOTAL ITERATIONS = 6

Figure 5.26: Example of the exhaustive method application for user pairing in NOMA.

From Figures 5.23 to 5.26 we can see that the same pairing was obtained from all four algorithms. Nevertheless, the process of finding the pairs becomes more complex as more UEs are considered and as some of the pairs do not fulfill the criteria explained above.

### 5.4.3 Single cell site model

Aiming at analyzing the impact of the user pairing method on the system performance, a model consisting of a highly loaded macro cell (i.e., 100% load) with hybrid MA, operating at 2.6 GHz and with a maximum reach of 600 m, is considered; 20 UEs are randomly deployed inside its coverage area following a uniform distribution. The four pairing algorithms introduced in this section are compared for this cell, using the proposed cost matrix. The results are shown for the DL using OMA as a benchmark and were obtained from 100 runs of the model, with 50 subframe transmissions per run.

### 5.4.4 Performance evaluation

In Figure 5.27 the capacity gain is shown for all four pairing methods. It can be seen that the Hungarian method offers a gain highly similar to the exhaustive method, with an average of 23.3% and 24%, respectively. The variation for both methods is approximately 12-39%. The results from the Gale-Shapley and random algorithms are also highly similar to each other, with an average gain of 18.3% and 18.5%, respectively; the variation for these methods is wider, which implies a higher uncertainty in the gain that can be obtained with their implementation. For the Gale-Shapley algorithm, such variation is between 4.6-42.4%, whereas for the random algorithm is 3.6-42%.

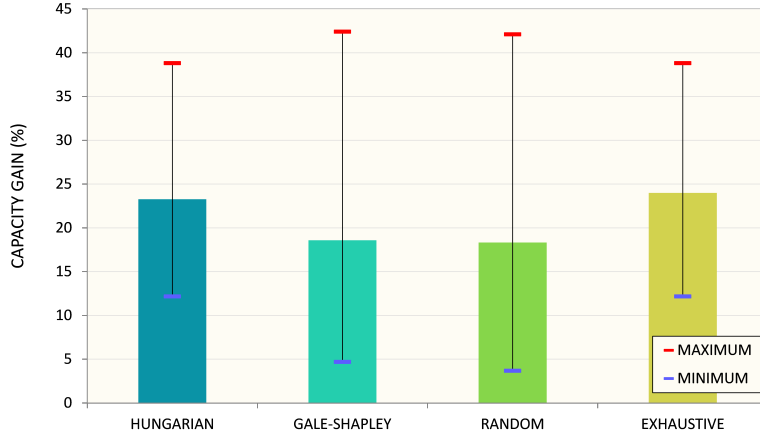
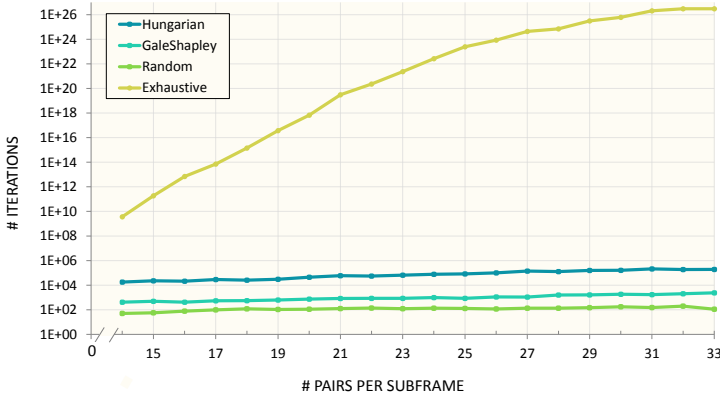
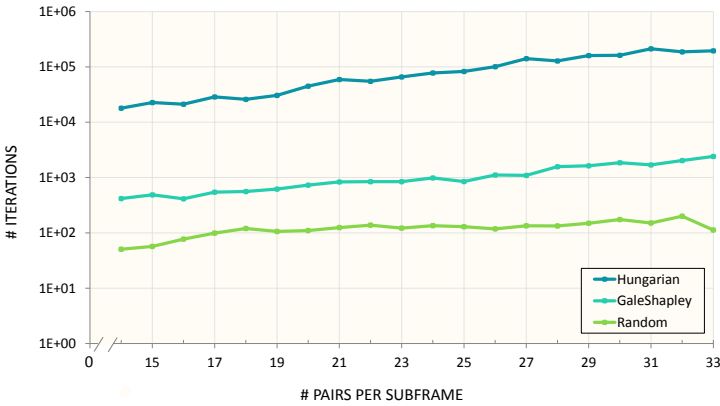


Figure 5.27: Capacity gain for a single cell for four pairing algorithms for NOMA; OMA is used as the benchmark.

From Figure 5.27 is it clear that the Exhaustive algorithm is the best option for increasing the system capacity, followed by the Hungarian method. Nevertheless, the speed/complexity of the methods should also be considered. Figure 5.28a shows the computational complexity in terms of the number of iterations required for each pairing



(a)



(b)

Figure 5.28: Complexity of four pairing algorithms for NOMA, represented as the number of iterations versus the number of pairs per subframe: (a) for the four pairing methods considered; (b) for the three methods with lower complexity.

method. Here it can be seen that the implementation of the exhaustive pairing results in the highest complexity (i.e., number of iterations), with values up to  $1.5E+22$  times higher than the highest complexity for the other three methods. With such complexity, the use of the exhaustive method would likely be time prohibited in a real implementation where every 1 ms a new subframe must be sent. In the used implementation, the performance of the exhaustive pairing could only be obtained from simulations results for up to 10 pairs, due to software limitations. For higher number of pairs, a combination of simulations

results and numerical estimations is used.

In Figure 5.28b the complexity of the pairing methods is shown excluding the exhaustive pairing for a better perspective of the performance of the remaining methods. For the Hungarian method, although its complexity is lower than that for the exhaustive pairing, it is on average 85x higher than for the Gale-Shapley algorithm. The complexity difference between the Hungarian and Gale-Shapley algorithms is due to the fact that with the former an optimal solution must be obtained, whereas with the latter only a stable solution is needed, which is not necessarily optimal. The selection of one or the other depends on the computational speed and the time constraints. The random algorithm is the one with the lowest complexity, with 9x fewer iterations needed than with the Gale-Shapley algorithm. Since the capacity gain of the Gale-Shapley over the random algorithm is negligible, the implementation of the random pairing (following the cost function and the pair restrictions proposed in this section) might offer a better trade-off between complexity and capacity gain. A deeper analysis of the influence of the pairing methods is presented in Section 6.4 for a HetNet implementation.

For the pairing approaches here presented, it could result interesting to propose alternative cost functions and/or UEs grouping, aiming at having a broader perspective of the pairing can be adjusted to increase the capacity.

Through these results, the relationship between the effectiveness of the user pairing method in NOMA and the capacity achieved can be confirmed. The selection of such method should be done based on its complexity and the capacity gains it can offer. But since the pairing selection depends on the UEs location and propagation conditions, choosing the method that offers the best performance is not a simple task. Furthermore, as mobility is added, the pair choosing becomes more complex since the channel gains of the UEs involved can rapidly change. Future work related to this topic is of great value for the implementation of 5G networks.

## 5.5 Chapter summary

The capacity gains that can be achieved with NOMA thanks to its high spectral efficiency, make it a promising MA scheme for 5G networks. Although implementing NOMA comes with many challenges for the network, these can be overcome by selecting appropriate scheduling/resource allocation algorithms, thus getting the benefits of NOMA without compromising the QoS (e.g., the BLER). The channel conditions of the NOMA UEs will be affected by the extra intra-cell interference added with this MA scheme, in comparison to OMA; this poses challenges in hybrid MA access where the CQI is reported by the UEs to the BS assuming an OMA transmission. An average difference of 12 dB can be expected between the NOMA SINR and the OMA SINR; therefore, CQI and MCS adjustments must be applied to meet the UEs QoS requirements. In this regard, an approach based on MCS adjustments and extra transmission power for NOMA has been proposed, aiming at targeting the desired BLER and improving the data rates for the UEs

when switching from OMA to NOMA. With this approach system capacity gains of up to 78% can be achieved, at the expenses of extra transmission power between 14-19% for the NOMA UEs to compensate for the additional intra-cell interference.

The actual gain that can be achieved with NOMA is, however, tightly related to the user-pairing and power allocation algorithm. To analyze how the former affects the system performance, four generic pairing algorithms with dynamic power allocation are evaluated, showing that depending on the efficiency/complexity of the method such gain can vary from 18.3% to 24% in a single cell scenario.

The options for implementing NOMA are still very diverse and further research work is very much needed on this topic. The pairing selection and number of multiplexed UEs per pair, resource allocation, interference management, integration with mmWave and MIMO, are some of the issues that are in their early stages and therefore their solutions have not yet been concretely defined for NOMA and 5G.



## CHAPTER 6

# 5G HetNet with NOMA and mmWave

### 6.1 Introduction

So far the performance of NOMA and mmWave have been evaluated for single cell scenarios, and the benefits that they bring from a capacity perspective have been highlighted. Once these benefits are characterized on a small scale, doing so in a bigger scale with network level approach results easier since a certain behavior is already expected.

This chapter focuses on the evaluation of techniques and settings that help to improve the capacity of a HetNet with mmWave and NOMA; such as the one shown in Figure 6.1. Most of the research work based on these technologies are focused on single cell scenarios. Therefore, analyses with a system-level perspective that offer a notion of the performance that can be expected in a network implementation, are highly anticipated.

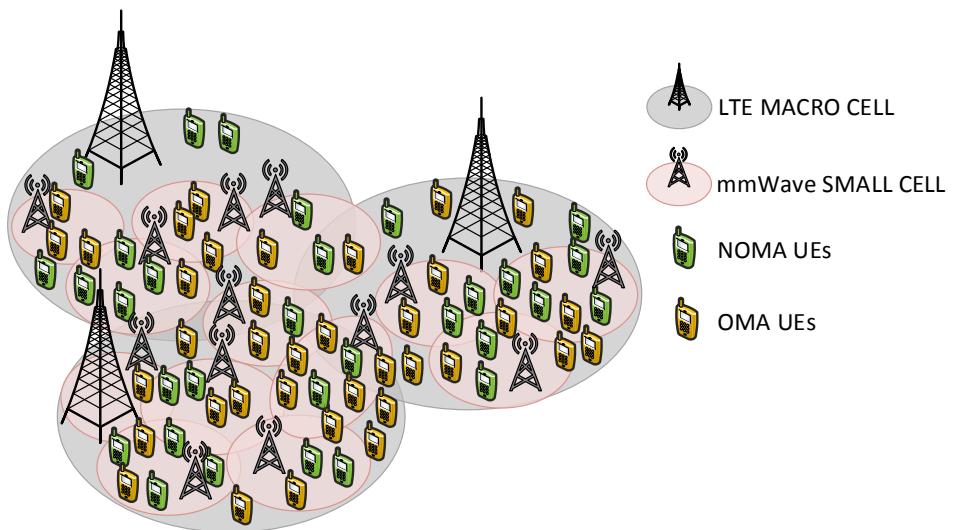


Figure 6.1: Example of a HetNet with mmWave and NOMA.

The results presented in this chapter are based on simulations of anticipated network scenarios for 5G. In Section 6.2 a proposed technique based on load balancing for NOMA implementations is presented. In Section 6.3 mmWave with narrow beamforming is used to help improve the performance of NOMA. Finally, in Section 6.4 a capacity dimensioning of a HetNet with mmWave and NOMA is presented.

As a side note, all the gains in the results analyses are calculated as the ratio between the parameter being evaluated and the benchmark; such gains are therefore dimensionless. The simulation model used for all the analyses is a 5G model developed in MATLAB using the LTE Toolbox; a subcarrier spacing  $\Delta_f = 2^\mu 15[\text{kHz}]$  with  $\mu = 0$  [3GP17c] is used, offering backward compatibility with 4G LTE.

## 6.2 Proposed load balancing technique: LB-NOMA

The integration of NOMA in radio access is not a simple implementation. As previously explained, with the extra intra-cell interference added at the transmitter in NOMA, a decrease in the SINR experienced by the NOMA UEs is expected; therefore, techniques to compensate for such deterioration should be considered. In this regard, the scenarios where NOMA results in an interesting implementation should be identified.

A fully loaded cell, for example, is a scenario in which NOMA helps increasing the system performance, whether this is measured as the number of UEs served within certain time and with certain resources, or as the increase in the UEs and/or system throughput. On the contrary, in a scenario where there is no congestion in the cell and therefore each UE can have access to the needed resources, the implementation of NOMA would only make the signal transmission and reception more complex. Let us assume that we have two UEs with a high difference in their channel gain and that each one only requires half of the available resources for the on-going transmission; both UEs could qualify as NOMA candidates to be paired, but such pairing would not be necessary since each UE already has access to the required resources. Nevertheless, scenarios where NOMA is required would be highly probable in 5G networks.

Section 5.3 shows how the implementation of NOMA can significantly improve the capacity in a hybrid MA systems and how this improvement is directly related to the number of NOMA UEs in the systems. In this respect, the capacity of a hybrid MA system could be increased even further if more UEs could be paired in fully loaded cells; when resources are scarce, the higher the number of NOMA UEs, the higher the system capacity because of the resource reuse. Therefore, the scheduling process in a fully loaded cell should be optimized aiming at pairing as many UEs as possible and keeping the OMA UEs to a minimum.

As an approach for such optimization, LB-NOMA is proposed, a cooperative scheduling based on load balancing techniques among a group of cells with unequal load; this work is part of our research presented in Paper E. With the randomness associated with how, when and where the mobile UEs use and demand data from the network, unequal

load among neighboring cells is highly common. Thus, a cell can be overloaded, while its neighbors have available resources that are not being utilized. This load imbalance affects the performance of the network, and load balancing techniques can be used to use the resources more efficiently.

The main idea behind LB-NOMA is to take advantage of the inter-cell interference found on the edges of the cells. If a UE<sub>A</sub> that is located between the overlapping edges of two cells cannot be paired in its fully loaded serving cell (home cell), perhaps a neighbor interfering cell (target cell) has either enough free resources or/and a good pair for UE<sub>A</sub>; if so, a forced load-based handover is performed for UE<sub>A</sub> from the home cell to the target cell.

To exemplify the LB-NOMA scheme, let us assume that we have two cells, Cell<sub>1</sub> with three UEs and Cell<sub>2</sub> with two UEs, as shown in Figure 6.2, and that the scheduling for subframe  $N$  is currently ongoing; Cell<sub>1</sub> is fully loaded while Cell<sub>2</sub> has available resources. NOMA, with a maximum of two multiplexed UEs per pair, is then implemented in Cell<sub>1</sub> and UE<sub>A</sub> and UE<sub>B</sub> are paired, leaving UE<sub>C</sub> with no pair; in Cell<sub>2</sub> there is no need to use NOMA. As UE<sub>C</sub> is located at the edges of both cells, Cell<sub>1</sub> exchanges UE<sub>C</sub> information with Cell<sub>2</sub>, aiming at reducing the OMA UEs and have more resources for the NOMA UEs (i.e., UE<sub>A</sub> and UE<sub>B</sub>). Cell<sub>2</sub> then determines if it can allocate UE<sub>C</sub>; at this point, two options arise (Figure 6.2):

- **Option 1** *UE<sub>C</sub> with OMA in Cell<sub>2</sub>*: there are enough free resources for UE<sub>C</sub> in Cell<sub>2</sub> so there is no need to implement NOMA, or there is not a suitable pair for UE<sub>C</sub> in Cell 2. In either case, the free resources can be allocated to UE<sub>C</sub>.
- **Option 2** *UE<sub>C</sub> with NOMA in Cell<sub>2</sub>*: UE<sub>C</sub> needs more resources than those available in Cell<sub>2</sub> and there is a suitable pair for UE<sub>C</sub>.

Once the cooperative scheduling is done, the definitive resource allocation is performed. Cell<sub>2</sub> indicates to Cell<sub>1</sub> that it can schedule UE<sub>C</sub> and then a forced handover is performed – Cell<sub>1</sub> sends handover signaling to UE<sub>C</sub> without such action being triggered by handover events reported by UE<sub>C</sub>. Since the forced handover needs to be indicated to UE<sub>C</sub> in the current subframe  $N$ , UE<sub>C</sub> would not be allocated any data resources until subframe  $M$ , with  $M > N$ , in its new serving cell, Cell<sub>2</sub>.

The implementation of LB-NOMA is most beneficial when the target cell is not fully loaded. Otherwise, the throughput of the UEs in the target cell would be affected, which at the ends could generate insignificant or no capacity gain to the system, as shown ahead.

LB-NOMA is an attractive implementation to enhance the system capacity in, for example, a hotspot area, where some cells are highly loaded while the neighbor cells are not. Load balancing techniques have been studied before for the current networks [SBT09] and are also implemented nowadays in hotspot areas. However, with LB-NOMA not only can the load be balanced in the system, but also the system capacity can be significantly increased by minimizing the OMA UEs and increasing the resources for NOMA UEs in a cell where resources are scarce and highly efficient reuse of the resource is needed.

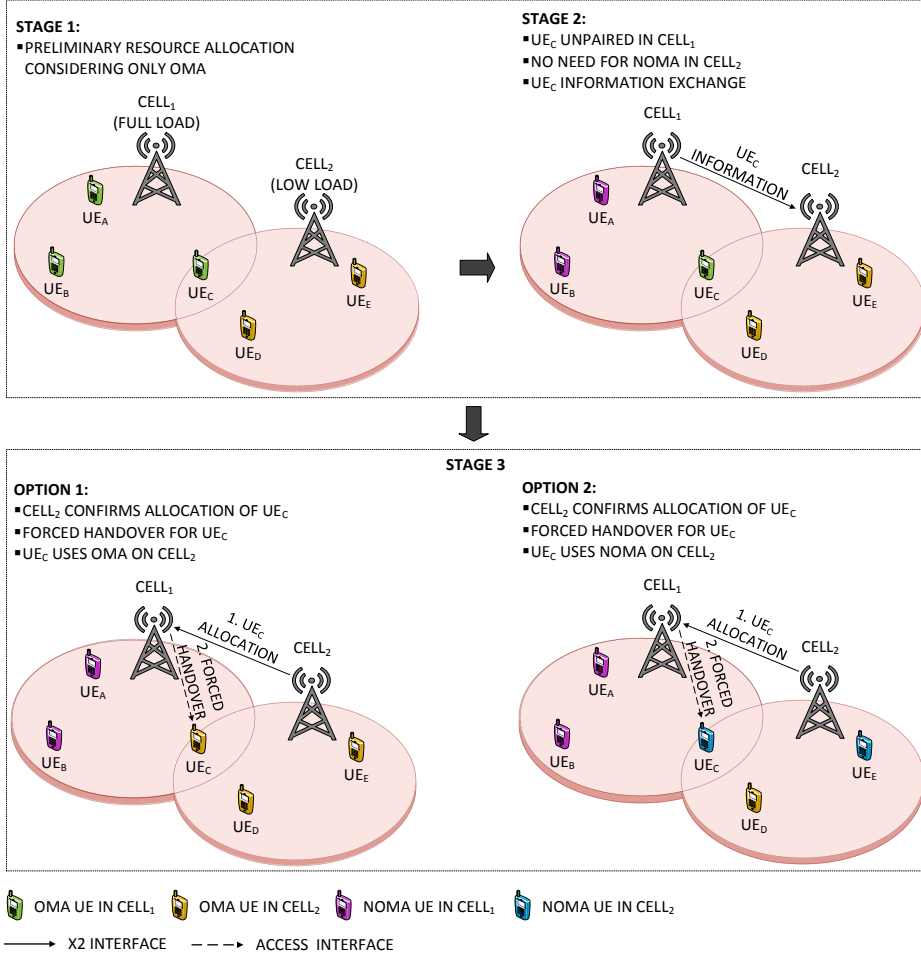


Figure 6.2: LB-NOMA scheme proposed for hybrid MA systems, aiming at reducing the OMA UEs in fully loaded cells.

### 6.2.1 Network model

An improvement in the overall system performance is expected from LB-NOMA in comparison to NOMA for a group of cells with unequal load, since free resources can be utilized more efficiently; to verify this, we focus on the capacity and UEs data rate gains of LB-NOMA over NOMA. A wrap-around model of seven clusters with seven small cell sites each (with three cells per site) is used for this purpose. The path loss model presented in Equation 5.8 is used as well as the propagation parameters from Table 5.4. The pairing method explained in Figure 5.17 is used for the simulations. The context assumed for the simulations is a hotspot area being covered by small cells, aiming at increasing the system capacity in such area.

Two scenarios were considered for the simulations; in the first one, the cells are unequally loaded. In the second scenario, all cells are fully loaded, therefore the target cell does not have available resources for more UEs; nevertheless, it can still try to find pairs for the incoming UEs. The second scenario is considered only for comparison purposes; as the cells are working at its maximum capacity, no gain is expected from the use of LB-NOMA. The cells load was determined based on the number of available PRBs, with zero free PRBs corresponding to a load of 100%. For simplicity reasons it is assumed that all the cells in a cell site have the same load. Table 6.1 summarizes the characteristics of each scenario for one cluster, with cells sites 1 and 6 fully loaded, 5 and 7 medium loaded, and 2-4 lightly loaded in Scenario 1. Figure 6.3 shows the grid configuration of one cluster and its direct neighbors for each scenario.

Table 6.1: Load for each cell site in a cluster for the two scenarios defined for LB-NOMA evaluation.

	CELL LOAD (%)	
CELL SITE	SCENARIO 1	SCENARIO 2
1	100	100
2	30	
3		
4		
5	60	
6	100	
7	60	

The UEs were placed randomly inside the coverage area of the cell sites following a uniform distribution, with 20, 12, and 6 UEs for the fully, medium and lightly loaded cell sites, respectively. The packet size for the UEs was selected randomly, while guarantying that the load of the serving cell would not exceed the set value (Table 6.1). In the scheduler, the pairing process proposed for NOMA based on MCS adjustments and extra  $P_{TX}$  allocation (Section 5.3, Figure 5.17) is used, with a 20% extra  $P_{TX}$  allowed. Once the UEs to be paired with NOMA have been selected, the implementation of LB-NOMA starts for the OMA UEs located in the overlapping areas of neighboring cells. If one UE

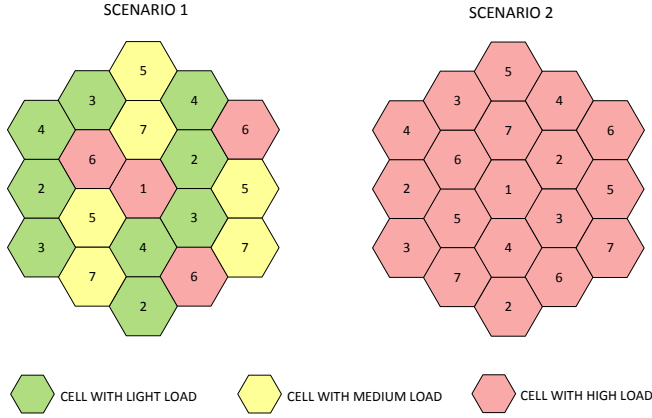


Figure 6.3: Grid configuration for one cluster and its direct neighbors for the two scenarios defined for LB-NOMA evaluation.

is located in the overlapping area of more than one cell, all the interfering cells participate in the scheduling of such UE. The best target cell will be that one that offers the highest throughput for the UE in question, as shown in Figure 6.4.

### 6.2.2 Network performance

The performance of the system is evaluated in terms of the capacity and the UEs data rate improvements, using an OMA system as a benchmark; the results are shown only for one cluster of seven cell sites. First, Scenario 1 was evaluated and the results for the capacity gain are shown in Figure 6.5, where it can be seen that the implementation of LB-NOMA brings significant improvements to the capacity for all seven cells. With LB-NOMA the maximum gain was experienced by the fully loaded cells with an average of 88%; this high gain is due to the use of NOMA when resources are scarce while minimizing the OMA UEs. The less loaded cells (i.e., cells 2-4) experienced an average gain of 16% whereas the medium loaded cells (i.e., cells 5 and 7) experienced an average gain of 12%. For NOMA the fully loaded cells experienced an average gain of 75%, while the cells that are not fully loaded did not experience any gain; this was expected since if there are enough resources for all UEs there is no need to apply NOMA.

These results show that when LB-NOMA is implemented in a group of cells with unequal load, a capacity gain can always be expected. With LB-NOMA the system resources are used more effectively by minimizing the OMA UEs and increasing the resources for NOMA UEs in loaded cells, while having the less loaded cells help in balancing the load. In overall, LB-NOMA offered between 12-18% capacity gain for the cells in comparison to NOMA. It is important to highlight that these results are highly

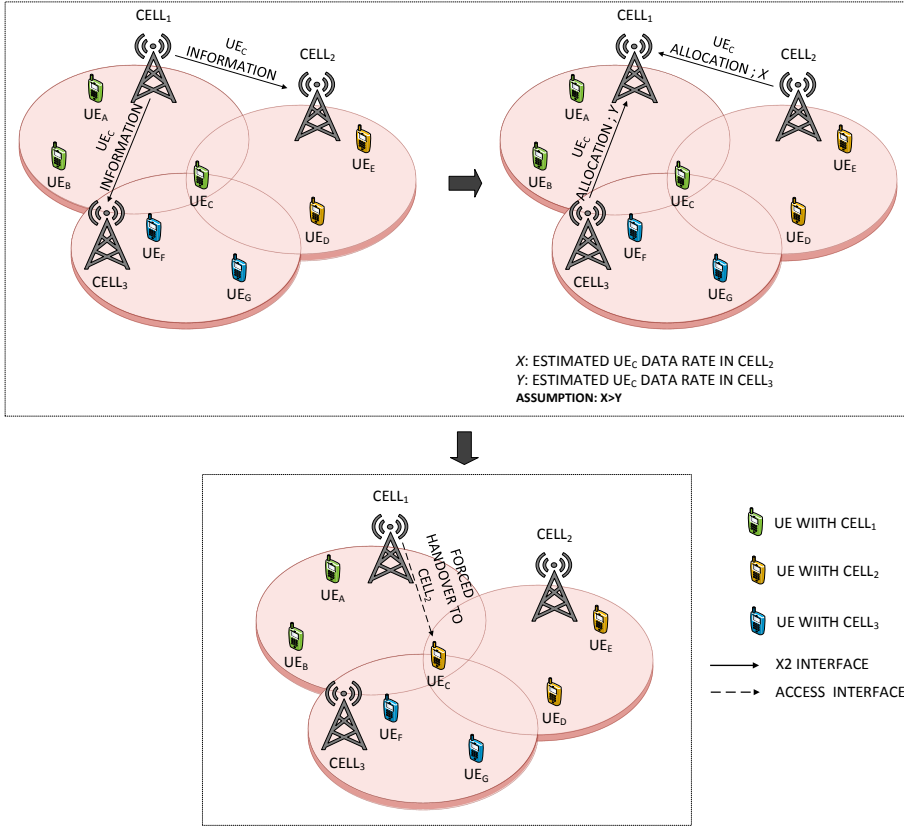


Figure 6.4: LB-NOMA for OMA UEs receiving interference from more than one neighboring cell.

dependable on the pairing method and power allocation selected; hence, these number may vary depending on the network configuration. Nevertheless, the same trend is expected.

To analyze the impact of LB-NOMA on the UEs data rates, Figure 6.6 shows the probabilities of changing or maintaining such rates, using NOMA as a benchmark. Here it can be seen that a UE has approximately 0.41 probability of increasing its data rate when LB-NOMA is implemented, 0.46 of maintaining it and only 0.13 of decreasing it. These results confirm that LB-NOMA is also an attractive implementation from the perspective of most of the UEs in the network.

Now let us refer to Figure 6.7 to analyze the performance of LB-NOMA in Scenario 2; it can be seen that the cell sites did not experience a capacity improvement. Although LB-NOMA still offers a considerably higher system capacity in comparison to OMA,

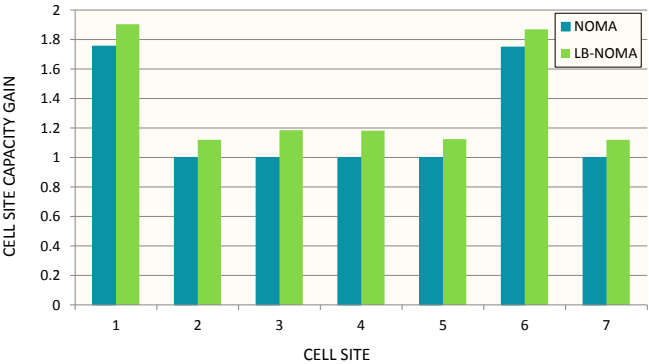
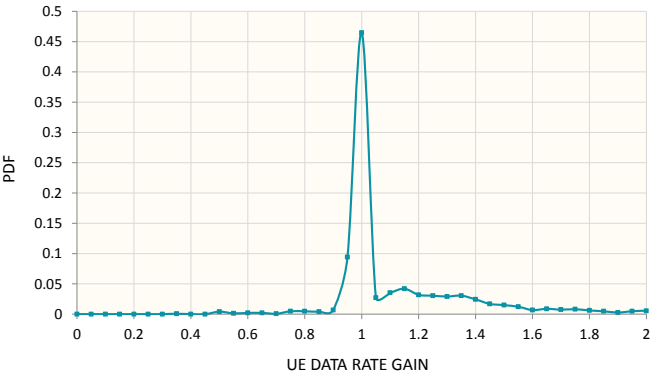
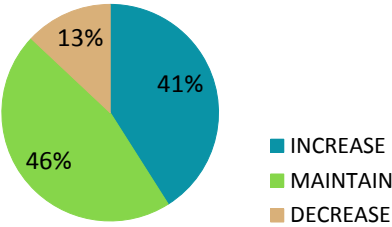


Figure 6.5: Network capacity gain comparison for Scenario 1 with NOMA and LB-NOMA, using OMA as a benchmark.



(a)



(b)

Figure 6.6: UEs data rates gain with LB-NOMA in Scenario 1, using NOMA as a benchmark: (a) PDF; (b) Probability of changing or maintaining the data rates.



with up to 70% gain, NOMA offers higher capacity and less complexity in groups of equally loaded cells. Under such load conditions, the UEs have a 0.27 probability of experiencing a gain of 0.95 in their data rates as shown in Figure 6.7. Moreover, a 0.57 probability of experiencing a decrease in the UEs data rates was experienced, along with a 0.24 and 0.19 probability of maintaining and increasing them, respectively. It can also be seen from Figure 6.8 that a maximum increase of 1.35x was experienced in the data rates.

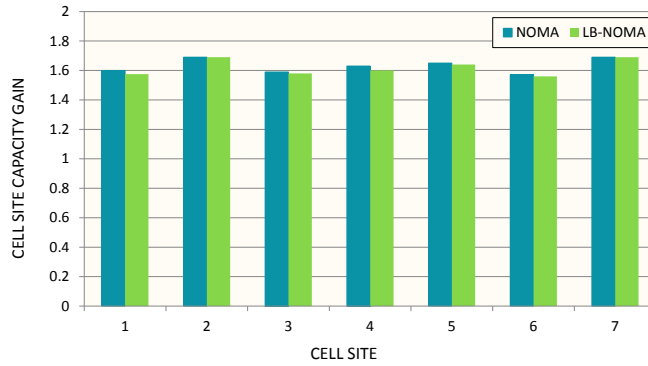


Figure 6.7: System capacity gain comparison for Scenario 2 with NOMA and LB-NOMA, using OMA as a benchmark.

These results show that LB-NOMA in a scenario with all the cells in under the same load conditions has a higher probability of deteriorating the UEs data rates rather than increasing them. The reason for this behavior is that since the cells have no available resources, taking UEs from other cells, even if it is to pair them, requires resources that need to be taken away from other UEs. Eventually, what is a gain in one cell turn into a loss in another. Therefore for scenarios were all the cells have an equal load, the implementation of LB-NOMA would not offer a good trade-off between system capacity and implementation complexity. Thus, under these conditions, it is preferable to implement NOMA.

Because of the typical inequality in the load among neighboring cells, considering load balancing techniques optimized for NOMA can help to significantly increase the capacity in the network. Unlike conventional load balancing where all the UEs at the edge of the cell are affected by the handover threshold adjustments, the load balancing in NOMA should aim at offloading the cell only from the unpaired OMA UEs; the latter restrict the capacity gains that can be achieved by reusing the limited available resources.

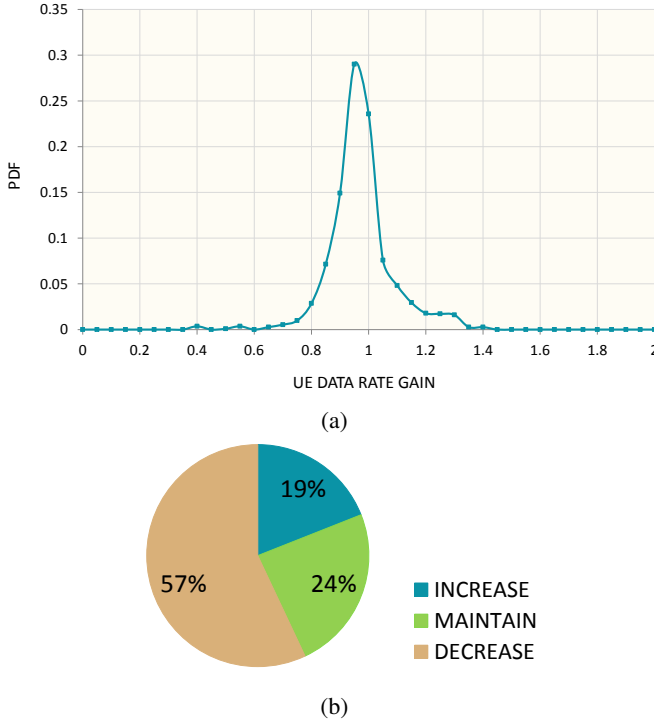


Figure 6.8: UEs data rates gain with LB-NOMA in Scenario 2, using NOMA as a benchmark: (a) PDF; (b) Probability of changing or maintaining the data rates.

### 6.3 Influence of mmWave on user pairing in NOMA

As previously highlighted, the integration of NOMA and small cells is a natural one since both technologies are key enablers for fulfilling the 5G requirements. In this regard, the integration of NOMA and mmWave is also natural. Such combination results particularly interesting for UDNs to offer multi-Gbps speeds to a large number of UEs. The use of NOMA in mmWave small cells can lead to a capacity increase of up to 78% in comparison to OMA, as shown in Section 5.3. It was also shown that such gain is directly related to the user pairing probability and to the extra  $P_{TX}$  allocated for NOMA. The latter is needed in order to compensate for the extra intra-cell interference added during a NOMA transmission. Nevertheless, when a power headroom is not available, relying on extra  $P_{TX}$  for enjoying the capacity benefits of NOMA is no longer feasible. In such cases, the use of mmWave and beamforming can help overcome this limitation.

When, for example, narrow beams are used for signal transmissions the reach and strength of the transmission is improved, since the same energy is concentrated in a

smaller area. If we consider a scenario where a UE is reached with, let us assume, a  $120^\circ$  beamwidth mmWave transmission, that same UE can be reached with higher power at the same position if a narrower beam of, for example,  $20^\circ$  is used as illustrated in Figure 6.9. This can be achieved without modifying the transmission power. Therefore, for NOMA, the use of mmWave with narrow beams can allow the UEs to experience high enough SINR values to correctly decode the superposed signal; thus, achieving the high capacity gains without requiring extra  $P_{TX}$  while guaranteeing the desired BLER. However, this gain in the received power comes at the expenses of a lower pairing probability.

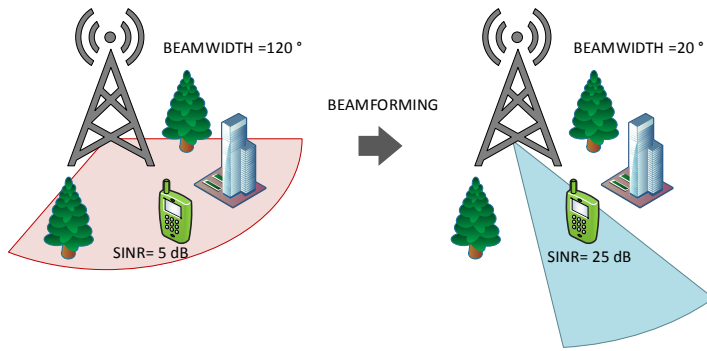


Figure 6.9: SINR increase due to the beamforming gain.

To better understand this, let us assume that we have a cell with two UEs as shown in Figure 6.10; let us also assume that both UEs are good candidates to be paired to each other with NOMA. If a wide beam with a beamwidth  $\theta_1$  is used, both UEs can be covered by the same beam in the same subframe. If, on the contrary, a narrower beam with beamwidth  $\theta_2$  is used, although both UEs will receive a stronger signal, because of their location they cannot be served by the same beam during the same subframe; therefore, even when they *could* be paired with NOMA, the use of a narrow beam limits their pairing. This limitation directly impacts the capacity gain of the system.

Therefore, the influence of mmWave with narrow beams in the performance of a network with hybrid MA is presented in this section. The analysis is done considering three key performance indicators: the extra  $P_{TX}$ , the pairing probability, and the overall network capacity gain. This work is part of our research presented in Paper G.

### 6.3.1 Network model

The simulated network is a wrap-around model of a cluster, consisting of seven outdoor mmWave small cells sites with three cells per site covering  $120^\circ$  each. A carrier frequency of 28 GHz is used and eight carriers of 100 MHz bandwidth each are considered.

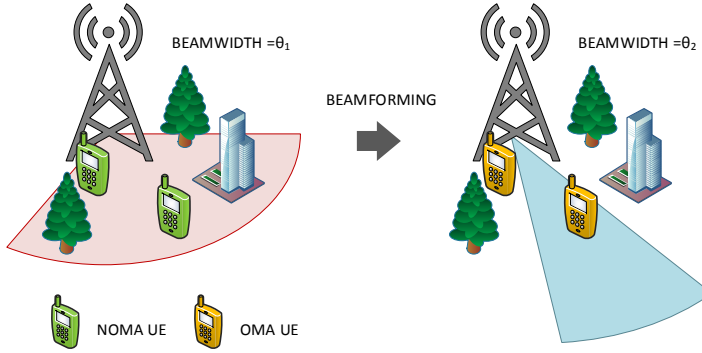


Figure 6.10: Influence of narrow beams on the user pairing for NOMA.

The network context for the simulations is HetNet with ultra-dense deployment of small cells. In Figure 6.11 the system model is shown.

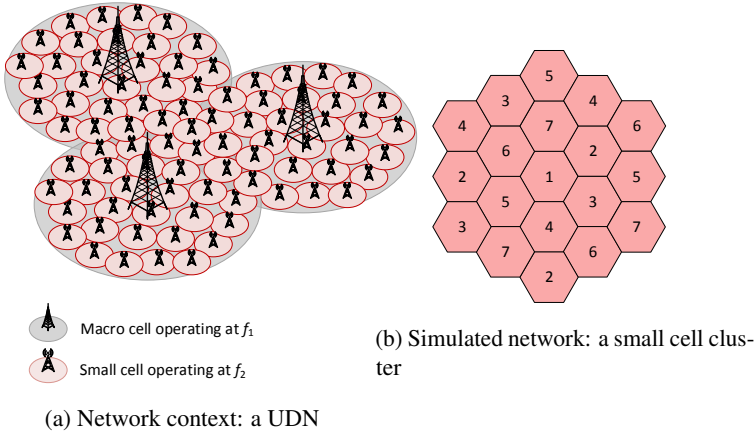


Figure 6.11: Network model used for evaluating the performance of mmWave with narrow beams and NOMA.

The users are located randomly inside the coverage area of the cell, following a uniform distribution. The number of UEs per cell was determined through a normal distribution with mean  $\mu=56$  and standard deviation  $\sigma=2$ . As baseline, broadcasts and pilot signals are assumed to be transmitted with a  $120^\circ$  beamwidth. This value, therefore, determines the coverage area of the cell. The maximum reach for each cell was set to 100 m. For the path loss calculation, the 3GPP recommendation for frequencies above 6

GHz is used [3GP17b]. Specifically, the urban micro (UMi) path loss model for outdoor scenarios is chosen, since it is the most suitable for outdoor mmWave small cells in UDNs. Depending on the LOS or NLOS conditions the path losses are calculated following Equations 6.1 and 6.2 for distances, respectively. These equations are valid for a distance between the UE and BS,  $d \leq 100$  m; a UE height  $h_{UE} = 1.5$  m; and a BS height,  $h_{BS} = 10$  m.

$$PL_{LOS} = 32.4 + 21\log_{10}(d_{3D}) + 20\log_{10}(f_c) + x_\sigma \quad (6.1)$$

$$PL_{NLOS} = 32.4 + 31.9\log_{10}(d_{3D}) + 20\log_{10}(f_c) + x_\sigma \quad (6.2)$$

where  $f_c$  is the carrier frequency in GHz,  $d_{3D}$  is the distance from the top of the UE to the top of the BS in meters, as shown in Figure 6.12. The distance between the BS and the UE,  $d$  is determined in the simulation by the location of the UE. The shadowing factor is represented by  $x_\sigma$  and it is Gaussian variable with mean zero and standard deviation  $\sigma = 4$  and  $\sigma = 8.2$  for LOS and NLOS, respectively. The probability of LOS is estimated also following the recommendation in [3GP17b] as:

$$P_{LOS} = \begin{cases} 1 & ; \quad d \leq 18m \\ \frac{18}{d} + \exp\left(-\frac{d}{36}\right) \left(1 - \frac{18}{d}\right) & ; \quad 18 < d \end{cases}$$

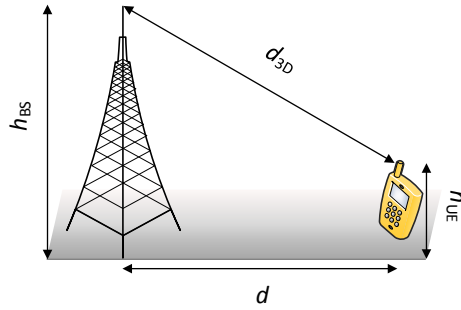


Figure 6.12: Reference for the distances for the UMi propagation model.

Once the path losses are calculated, the received power and SINR can be estimated using Equations 5.3 and 5.1, respectively. As ideal directional antennas are considered for the transmissions, the antenna gain  $G$  can be calculated as:

$$G = \frac{2\pi}{\theta} \quad (6.3)$$

where  $\theta$  represents the beamwidth. For simplicity reasons, analog beamforming is considered; therefore, a single beam per cell is transmitted at the same time. This means that when a beamwidth narrower than  $120^\circ$  is chosen, each cell area will be divided into  $N = 120^\circ/\theta$  narrow sectors. The transmissions to each narrow sector are done following a round robin scheduling.

For the resources allocation, the methodology explained in 5.3 is used. Therefore, the UEs are divided into four groups depending on their modulation order and PRBs are proportionally assigned to each group according to their number of UEs; a minimum of two PRBs per UE is set. For NOMA, the user pairing and the proposed MCS and extra  $P_{TX}$ , also presented in Section 5.3, is used; only in this case, for the latter, the extra  $P_{TX}$  is tried to be covered by the beamforming gain, thus reducing the power headroom needed. For the power allocation of the multiplexed UEs in NOMA, a factor of 0.25 is used for  $UE_1$  and 0.75 is used for  $UE_2$ . Table 6.2 summarizes the model parameters.

Table 6.2: Simulation propagation parameters.

	OMA	NOMA
CARRIER FREQUENCY (GHz)	28	
CARRIER BANDWIDTH (MHz)	100	
COMPONENT CARRIERS	8	
CODING/DECODING	TURBO CODING	
MODULATION SCHEME	QPSK, 16QAM, 64QAM, 256QAM	
TRANSMISSION POWER (dBm)	13	
POWER ALLOCATION PER UE	1	ADAPTIVE (*)
WAVEFORM	OFDM	
TRANSMISSION MODE	SU-MIMO (2x2)	
TRANSMITTER ANTENNA GAIN (dBi)	BEAMWIDTH DEPENDENT (**)	
PATH LOSS MODEL	Umi (†)	
CHANNEL ESTIMATION	MMSE	
RECEIVER ANTENNA GAIN (dBi)	0	
RECEIVER NOISE FIGURE (dB)	6	
RECEIVER SCHEME	LMMSE	$UE_1$ = SLIC $UE_2$ = LMMSE

\* SEE EQUATION 4.6

\*\* SEE EQUATION 5.3

† SEE EQUATIONS 5.1 AND 5.2

### 6.3.2 Network Performance

For the performance evaluation, two scenarios are defined. The first one, a benchmark scenario with OMA as the only multiple access scheme and  $120^\circ$  beamwidth. The second scenario uses hybrid MA and the beamwidth varies from  $5^\circ$  to  $120^\circ$ . Figure 6.13

illustrates both scenarios. Five Monte-Carlo simulations were performed with 100 runs per simulation; each run equals to subframe transmission.

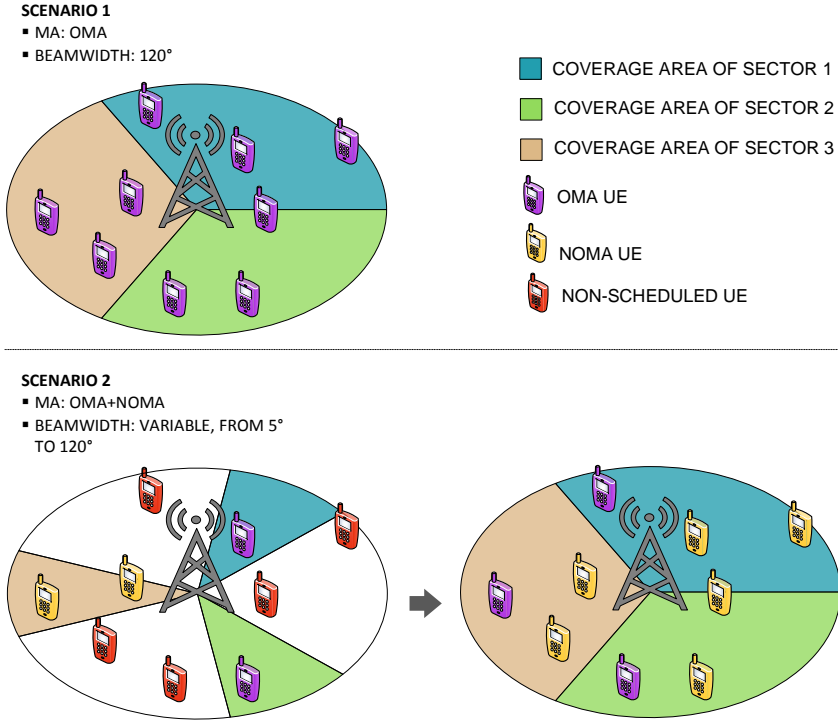
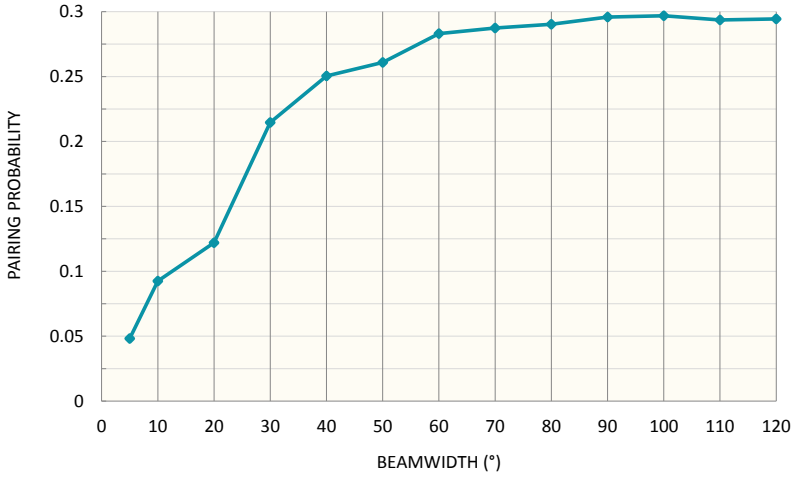


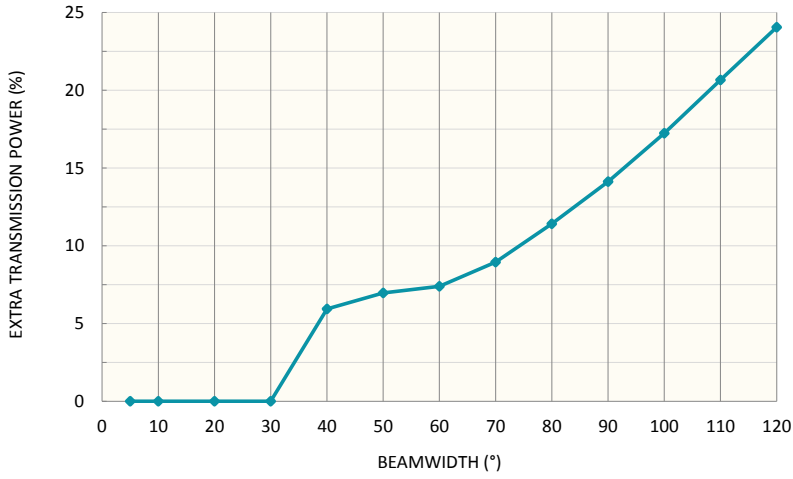
Figure 6.13: Simulation scenarios.

For the first analysis, we refer to Figure 6.14, where the  $P_{TX}$  and pairing probability are shown as a function of the beamwidth for Scenario 2. Here it can be seen that from a power consumption perspective, beamwidths below 30° offer the best performance, with no power headroom requirement; however the pairing probability is rather low with values between 0.05-0.21. This behavior was expected since by reducing the coverage area of each sector, the pairing probability is also reduced, as explained earlier. For beamwidth values between 40°-70°, an extra  $P_{TX}$  between 6-10% is required, resulting in pairing probabilities between 0.25-0.28. For higher beamwidths, the extra  $P_{TX}$  increases almost linearly reaching values of up to 24% and pairing probabilities between 0.28-0.3. These latter beamwidth values do not offer an attractive solution since the pairing probability does not increase significantly and they require considerably higher extra  $P_{TX}$ . Therefore, beamwidths below 70° offer a better trade-off.

As the pairing probability is directly related to the capacity gain, in Figure 6.15 the



(a) Pairing probability



(b) Extra  $P_{TX}$

Figure 6.14: Pairing probability and extra  $P_{TX}$  for mmWave with narrow beams and NOMA.



system capacity gain is shown for Scenario 2, using Scenario 1 as a benchmark. It can be seen that for the beamwidths values below  $22^\circ$  no gain is achieved by including NOMA in the system; in fact, a capacity loss was experienced for those values. The reason for this is that besides the low pairing probability, UEs are being served less often; as the beams are narrower it takes more time slots for the BS to cover each  $120^\circ$  area in Scenario 2, whereas in Scenario 1 it only takes one time slot.

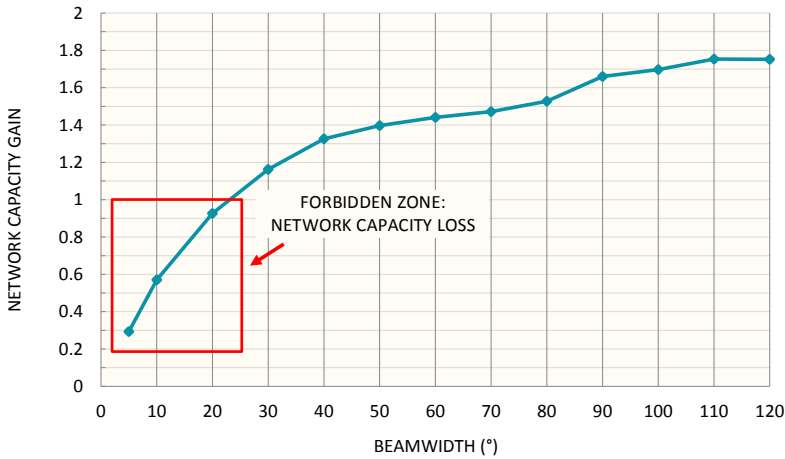


Figure 6.15: Network capacity gain for NOMA, using OMA with  $120^\circ$  beamwidth as a benchmark.

For beamwidths above  $30^\circ$ , a capacity gain between 20% and 80% start to be experienced, reflecting that even when it still takes more time for the BS to cover each  $120^\circ$  cell area, the resources reuse in NOMA and a higher pairing probability compensate this delay. If no power headroom is available, then a beamwidth between  $25$ - $30^\circ$  offer the best solution with a capacity gains up to 20% and no extra  $P_{TX}$  required. On the contrary, if a power headroom is available, using beamwidth values between  $40$ - $70^\circ$  offer a good trade-off, with no more than 10% extra  $P_{TX}$  needed for system capacity gains between 32% and 47%.

These results confirm the benefits of combining directional mmWave transmissions and NOMA when UEs that can be paired are located inside the coverage area of the same sector. The mmWave directional transmission allows reaching the higher channel gains needed for the NOMA UEs (in comparison to OMA) while requiring little to no extra  $P_{TX}$ , thus improving the overall system performance from both a capacity and a power-efficiency perspective. Not needing extra  $P_{TX}$  is interesting for UDNs, since the massive number of small cells deployed will most likely use the same frequency band, which makes them sensible to inter-cell interference. Therefore, careful cell planning is

needed in UDNs to define as precisely as possible the coverage area of each cell. Using extra  $P_{TX}$  for the NOMA users equals to having the user farther located from the cell edge, which means that the range of the cell is increased. Since this increase would depend on the highly variable network conditions, the coverage area of the cells using NOMA with extra  $P_{TX}$  will be difficult to define. In this regard, the benefits of combining mmWave directional transmissions with NOMA proves to be an attractive solution for high-capacity UDNs in 5G deployments.

Nevertheless, the results also showed that the gains of mmWave NOMA are highly dependent on the UEs location. Although this might not pose an issue for highly loaded cells with a high UEs density, it could still compromise the performance of the system, especially when single-beam analog beamforming is used. Alternative techniques can be considered to overcome this limitation such as multi-beam analog beamforming [Xia+17] and hybrid beamforming [SY17], which allow to form multiple narrow beams and serve multiple UEs simultaneously. Although digital beamforming also allows forming multiple beams at the same time, this approach is not recommended for mmWave implementations due to the higher complexity, especially when large-antenna arrays as the ones expected for mmWave are used [SY17]. Furthermore, multi-beam techniques allow tuning the beamforming vectors individually, which would affect the power received by each UE, allowing to more accurately control the signal propagation and achieve better performance.

It can be concluded from the performance analysis that the capacity gain expected from the mmWave NOMA system will depend on the right selection of parameters such as beamwidth, user pairing and power allocation, and beamforming technique. The beam characteristics chosen for mmWave will help overcome the limitations introduced by the extra intra-cell interference for the NOMA UEs, at the expenses of a more restricted user pairing selection for NOMA. Depending on the network environment, said parameters should be adjusted to offer the best performance for the network and for the UEs. For example, narrower beams could be chosen when the cell becomes highly loaded, since the probability of finding suitable pairs even in narrow sectors is higher; a typical scenario for this could be a stadium or a concert venue. For less crowded scenarios, then it comes to finding a balance between the QoS delivered and the capacity gain.

Further work regarding the relationship between the power allocation, user pairing, and beamforming is needed for mmWave NOMA implementations, since their combination is a fundamental step for increasing the spectral efficiency and boost the capacity in small cells.

## 6.4 HetNet capacity dimensioning

So far different techniques and approaches have been evaluated aiming at increasing the capacity for 5G networks. Two main RATs have been considered: NOMA and mmWave; along with small cell deployments. Furthermore, MCS and power adjustments have also been analyzed, as well as user pairing methods for NOMA.

Now, a capacity analysis of a typical expected 5G network is done, combining the previous work here shown and with the goal of presenting a system-level performance analysis of a more realistic deployment. This work is part of our research presented in Paper F.

### 6.4.1 Network model

For evaluating how NOMA and mmWave affect the DL capacity dimensioning of a HetNet, a two-tier out-of-band deployment is considered. The first tier corresponds to the macro cells operating at 2.6 GHz, whereas the second tier corresponds to the small cells operating at 28 GHz. The 5G model consists of a wrap-around cluster model of seven macro cell sites, with small cell sites deployed inside their coverage area. The density of small cell sites on each macro cell site depends on the UEs density. Sparse deployment of small cells sites is used for areas that have identified hotspots, whereas a dense deployment is used for macro cell sites that are constantly fully loaded. The small cell sites are modeled as clusters of 7, 3 or 2 sites, or as a single site; all cells are considered to be located outdoors. For the small cells tier, inter-cell interference is not considered; because of the propagation characteristics of mmWave, the signal power will be concentrated in the main link and any secondary link will dissipate rather fast, not affecting neighboring cells. Figure 6.16 shows the network model, whereas the macro and small cell sites configuration are summarized in Tables 6.3 and 6.4, respectively.

For the coverage calculation of the macro cells, an ISD of 600 m is used, along with the path loss model for urban macro deployment, 3D-UMa, from the 3GPP recommendation in [3GPP17a]. The Equations 6.4 and 6.5 represent the propagation model applicable to the network model in this section for LOS and NLOS, respectively.

$$PL_{\text{LOS}} = \begin{cases} 22\log_{10}(d_{3D}) + 28 + 20\log_{10}(f_c) + x_\sigma; & d \leq 416\text{m} \\ 40\log_{10}(d_{3D}) + 28 + 20\log_{10}(f_c) \\ \quad - 9\log_{10}(812.25) x_\sigma; & 416 < d < 5000\text{m} \end{cases} \quad (6.4)$$

$$PL_{\text{NLOS}} = \max(PL'_{\text{NLOS}}, PL_{\text{LOS}}) \quad (6.5)$$

$$\begin{aligned} PL'_{\text{NLOS}} = & 161.04 - 7.1\log_{10}(20) + 7.5\log_{10}(20) - 22\log_{10}(25) \\ & + (43.42 - 3.1\log_{10}(25)) (\log_{10}(d_{3D}) - 3) + 20\log_{10}(f_c) \\ & - \left( 3.2 (\log_{10}(17.625))^2 - 4.97 \right) + x_\sigma \end{aligned} \quad (6.6)$$

where  $x_\sigma$  is the shadowing factor represented by a Gaussian variable with mean zero and standard deviation  $\sigma=4$  and  $\sigma=6$  for LOS and NLOS, respectively. The distance from the top of the UE to the top of the BS is represented by  $d_{3D}$  (Figure 6.12), whereas  $d$

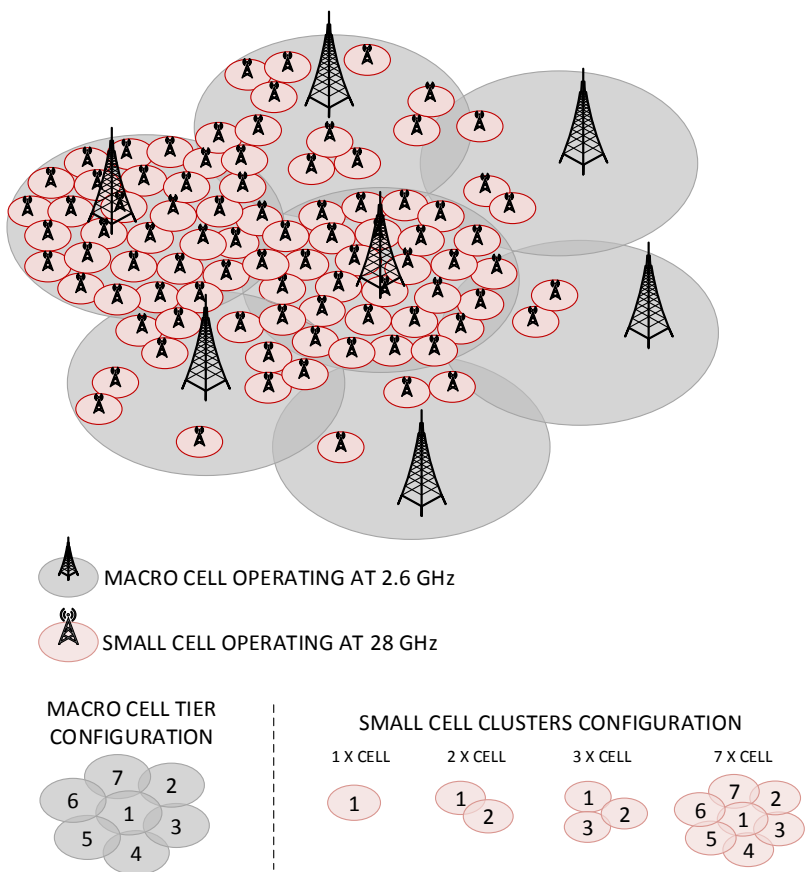


Figure 6.16: Two-tier HetNet model used for capacity dimensioning.

is the distance between the UE and the BS at ground level; both variables are in meters. Heights of 30 m and 1.5 m are considered for the BS and the UEs, respectively. The carrier frequency in Hz is represented by  $f_c$ . For the small cells, the ISD is 100 m and the path loss model UMi for outdoor deployments is used; Equations 6.1 and 6.2 are applied for these calculations.

The UEs in the network are randomly located inside the coverage area of each cell using a uniform distribution; the small cells provide the data connections for the UEs inside their coverage area, while the macro cells provide such connections for the rest of the UEs. A total of 2376 UEs are served by the network (Table 6.3).

Four UE profiles are considered, which determine the size and inter-arrival time of

Table 6.3: Macro cells configuration.

MACRO CELL SITE	UE DENSITY PER Km <sup>2</sup>	# UEs IN MACRO CELL SITE AREA	# SMALL CELL SITES DEPLOYED INSIDE THE MACRO CELL SITE	TYPE OF SMALL CELL SITE DEPLOYMENT	% MACRO CELL SITE COVERED BY THE SMALL CELL SITES	AVERAGE LOAD (%)
1	1200	772	35 (7 x cluster of 7)	Dense: covering most of the macro cell site area	0.97	100
2	93	60	3 (1 x cluster2; 1 single)	Sparse: clustered small cell sites located at the edges	0.08	30
3	62	40	2 (1 x cluster of 2)		0.06	
4	93	60	3 (1 x cluster2; 1 single)		0.08	
5	420	270	10 (2 x cluster of 3; 1 x cluster of 2; 2 single)	Small-area sparse: small clusters and/or isolated cell sites covering small areas	0.28	60
6	1200	772	35 (7 x cluster of 7)	Dense: covering most of the macro cell site area	0.97	100
7	625	402	12 (3 x cluster of 3; 1 x cluster of 2; 1 single)	Small-area sparse: small clusters and/or isolated cell sites covering small areas	0.33	60

Table 6.4: Small cells configuration.

SMALL CELL SITE	UE DENSITY PER Km <sup>2</sup>	# UEs SERVED BY THE CELL SITE	AVERAGE LOAD (%)
1	1118	20	100
2	335	6	30
3	335	6	
4	335	6	
5	670	12	60
6	1118	20	100
7	670	12	60

the UEs' packets: video streaming, File Transfer Protocol (FTP), web browsing and IoT sensors. The work in [Por+16] is used for the characteristics of the first three profiles, whereas the work in [MMB16] is used for the IoT sensors. The profile for each UE is randomly selected while guaranteeing that the average load of the respective serving cell is maintained. For the MA scheme selection, the BS decides between OMA and NOMA depending on the results of the user pairing process. Same as in Section 5.4, the cost matrix from Table 5.7 is used along with its corresponding methodology. Similarly, the same four generic pairing methods are considered: Hungarian, Gale-Shapley, random, and exhaustive; in order to evaluate their influence on the system capacity. The proportional fair scheduler from Equation 5.9 is used for the PRBs allocation. NOMA is applied independently in each network tier; this is possible because of the out-of band implementation. In Table 6.5 the parameters used for the link budget calculations and the signal generation are summarized.

Table 6.5: Simulations parameters.

	MACRO CELL	SMALL CELL
CARRIER FREQUENCY (GHz)	2.6	28
CARRIER BANDWIDTH (MHz)	20	
COMPONENT CARRIERS	5	40
TRANSMISSION POWER (dBm)	43	13
TRANSMITTER ANTENNA GAIN (dBi)	16	10
RECEIVER ANTENNA GAIN(dBi)	0	0
NOISE FIGURE (dB)	7	6
RECEIVER SENSITIVITY (dBm)	-120	-85
PROPAGATION MODEL	3D-Uma	Umi
ISD (m)	600	100
COVERAGE AREA (km <sup>2</sup> )	0.643	0.0179
MULTIPLE ACCESS METHOD	HYBRID: NOMA/LB-NOMA + OMA	
MODULATION SCHEMES	QPSK, 16QAM, 64QAM, 256QAM	
NOMA POWER ALLOCATION	SEE EQUATION 9	

\* SEE EQUATIONS 5.4 AND 5.5

\*\* SEE EQUATIONS 5.1 AND 5.2

† SEE EQUATION 4.6

For simplicity reasons, a beamwidth of  $120^\circ$  per cell (each cell site has three cells) is assumed as well as single-beam analog beamforming. For this network model, using a wide beam does not require the use of extra  $P_{TX}$  because during the scheduling process,

as long as there is a capacity gain from the combined rates of the multiplexed UEs in NOMA, no requirements regarding the individual rates are set. As pre-scheduled and not pre-scheduled UEs are considered during the pairing process, a rate gain is guaranteed for the not pre-scheduled UEs that can be paired. Furthermore, the implementation of LB-NOMA is also considered in the model.

### 6.4.2 Network performance

For the first performance analysis, the system capacity is shown in Figure 6.17a, whereas the capacity gain is shown in Figure 6.17b using OMA as a benchmark. These results reflect the improvement in the network performance that can be achieved by incorporating NOMA in a HetNet. Furthermore, the gain of using LB-NOMA can also be noted, especially for the random pairing. For the NOMA/OMA cases the same trend as in the single cell analysis presented in Section 5.4 can be seen; the Hungarian algorithm offers the second highest capacity and capacity gain after the exhaustive algorithm, achieving an average of 398 Gbps for a gain of 22% and 408 Gbps for a gain of 24%, respectively. In contrast, with the Gale-Shapley algorithm, a capacity of 392 Gbps is achieved for 21% gain, whereas with a random pairing the capacity is approximately 385 Gbps for 18% gain. The combination of NOMA with mmWave in the small cells allows having such high rates because of the wide spectrum available and its more effective utilization.

Interestingly, it can be seen that when LB-NOMA is used, the random method offers the highest capacity gain, with a 29% gain corresponding to a sum rate of 420 Gbps; this means that an extra 11% was gained in the capacity because of the use of load balancing in NOMA. To understand this behavior, let us remember that NOMA is more effective as the difference in the channel gain of the paired UEs becomes larger. This typically occurs between UEs located close to the BS and near the edge of the cell. Therefore, when an optimal method is used to find the pairs (e.g., exhaustive or Hungarian pairing), the UEs located at the edge of the cell will be chosen with a higher probability, since pairing them with UEs close to the BSs yields the highest system gain. With pairing methods that are not optimal (e.g., Gale-Shapley or random pairing), the probability of having active UEs at the edge of the cell that are not paired is higher. Hence, when LB-NOMA is implemented, those unpaired UEs at the cell edge are likely to be moved to neighboring cells that are not fully loaded. This makes the LB-NOMA more efficient when the pairing method fails to choose the best possible pairs. Then, in scenarios where the cells are not equally loaded, the implementation of a simple pairing method such as the random can be chosen along with LB-NOMA. On the contrary, if all cells tend to be fully loaded, thus limiting the need for LB-NOMA, optimal pairing methods should be considered.

### 6.4.3 Capacity dimensioning

To now illustrate how the implementation of NOMA affects the capacity dimensioning of a HetNet, let us consider the network model from Figure 6.16. For the estimations traffic volume based dimensioning is used; assuming that during the busy hour the average

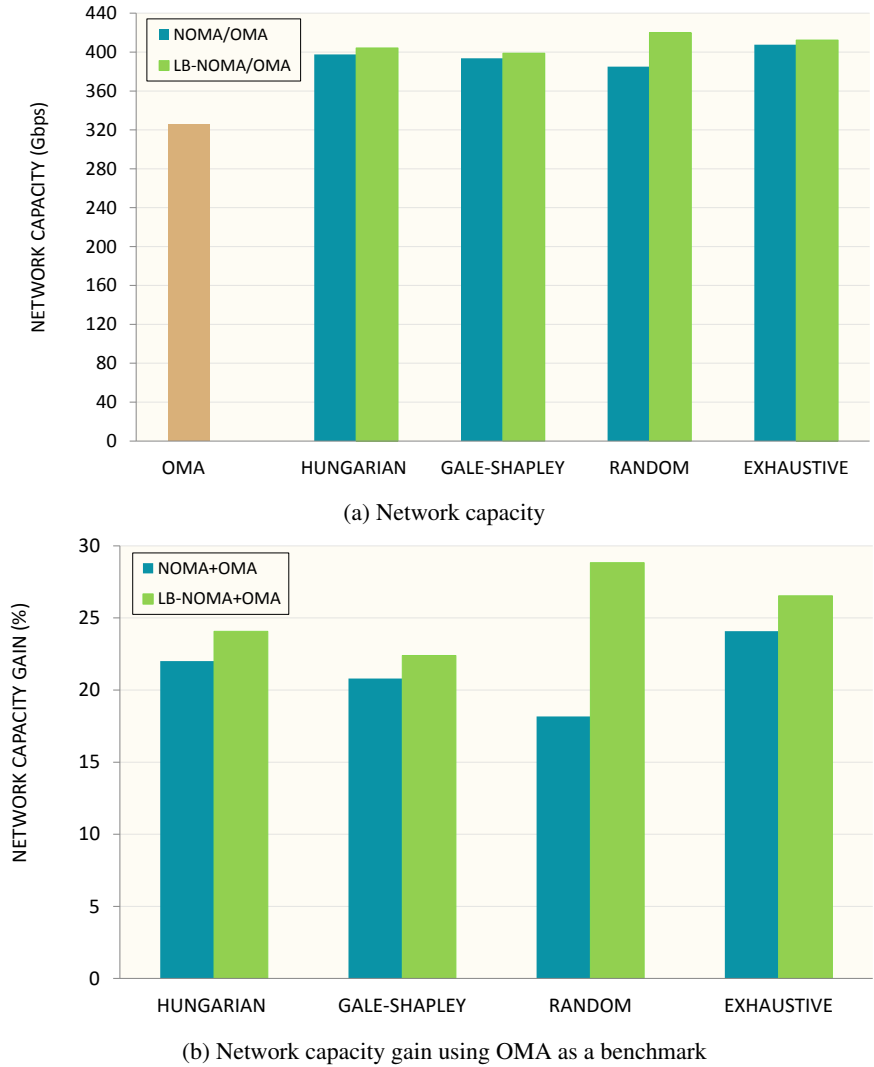


Figure 6.17: Capacity for the modeled HetNet for four pairing algorithms.



load of the cells is 50% and that the busy hour carries 15% of the daily traffic, the traffic volume  $T$  in GB/month/km<sup>2</sup> can be estimated following Equation 4.2.

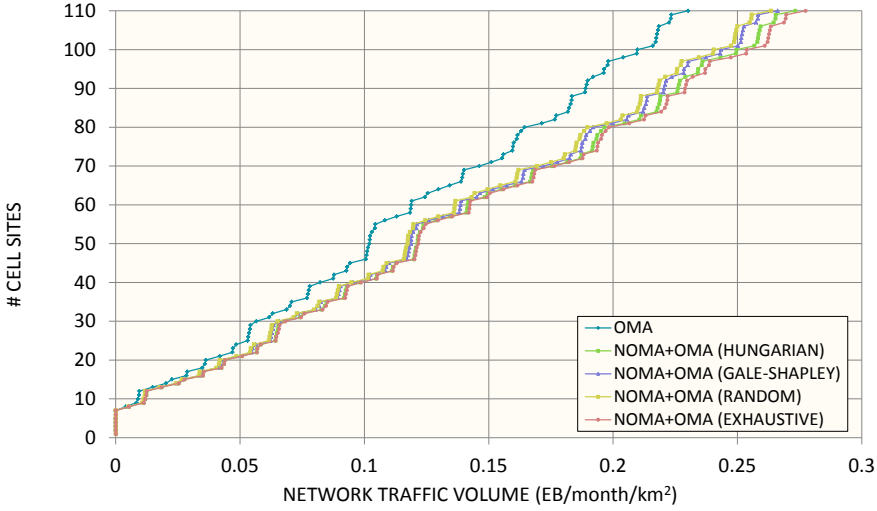
The results are shown in Figure 6.18, and its corresponding zoomed-in version is shown in Figure 6.19. The seven macro cells are fixed for every case, and the capacity expansion is done by adding small cells. From these results, we can see the advantages of NOMA from a dimensioning perspective. In a first glance, the most noticeable gain, in terms of the cells needed to support certain traffic volume, is that of including NOMA (independently of the user pairing method) as an MA scheme. This gain is clearer as the traffic volume increases. Because of the massive amount of data expected for 5G networks, the implementation of NOMA is an attractive feature to meet the capacity requirements while minimizing the deployment costs associated.

For a better deployment requirements analysis, let us assume that the modeled network needs to handle a volume of 0.2 EB/month/km<sup>2</sup>. With OMA, 98 cells are needed (seven macro cells plus 91 small cells). On the contrary, with a hybrid MA system that combines OMA/NOMA (Figure 6.19a) and uses the Hungarian algorithm for the user pairing, 81 cells are needed; same as with the exhaustive pairing. The highest cell requirement from the hybrid MA cases comes with the use of the Gale-Shapley or random algorithms, with 82 cells needed to support such traffic volume; nevertheless, both methods offer a saving of 16 cells over OMA. Even when the saving of only one cell depending on the pairing method (81 vs 82 cells) is not highly attractive, it is important to highlight that, as shown in Figure 6.18, as the traffic volume increases the gain in the number of cells needed becomes higher for the more effective pairing methods.

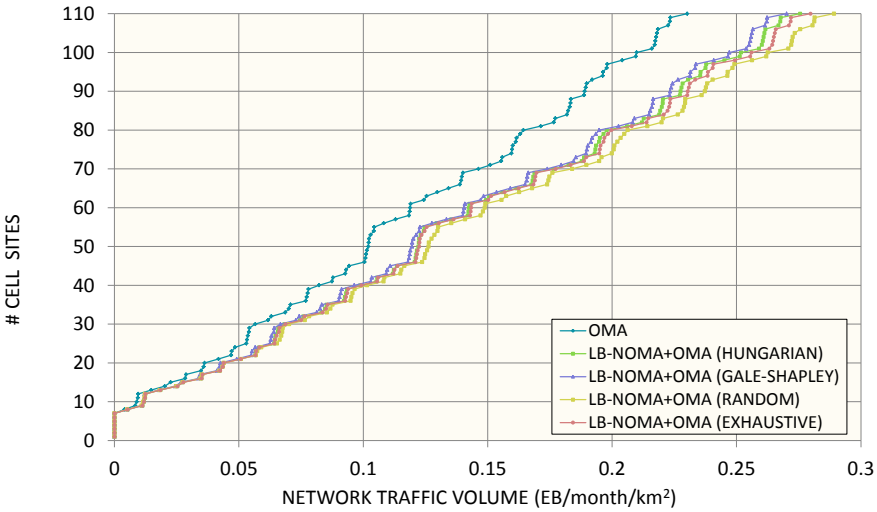
For the hybrid MA system with LB-NOMA/OMA (Figure 6.19b) the best performance is offered when the random pairing is used, as expected from the results shown in Figure 6.17. For the same traffic volume of 0.2 EB/month/km<sup>2</sup>, 75 cells are needed with the random pairing when LB-NOMA is used. The remaining three methods, each needs 81 cells; this represents little to no improvement compared to their equivalents in the NOMA/OMA cases, especially for scenarios with lower traffic volume. Thus, the benefits of LB-NOMA highly depend on the chosen pairing method. When close-to-optimal methods are used, the space left for improvements with LB-NOMA is limited. On the contrary, simpler algorithms can be used if their weaknesses are balanced with other optimization techniques, such as load balancing.

In Figure 6.20 the traffic volume gain and the gain in the number of deployed cells are shown, using OMA as the benchmark. For a NOMA/OMA implementation, the use of the Hungarian method is preferred, since it offers an average gain of 20% in the traffic volume that can be supported (Figure 6.20a) and its complexity is lower than the exhaustive method, which offers a 21% traffic volume gain. The Gale-Shapley algorithm offers a 17% gain, whereas the gain for the random method is 16%. On the contrary, for an LB-NOMA/OMA implementation, the random method offers the highest gain in the traffic volume supported, with 24.6%.

For the number of cells needed, it can be seen from Figure 6.20b that for NOMA/OMA the highest saving in the number of cells deployed is achieved with the exhaustive method,

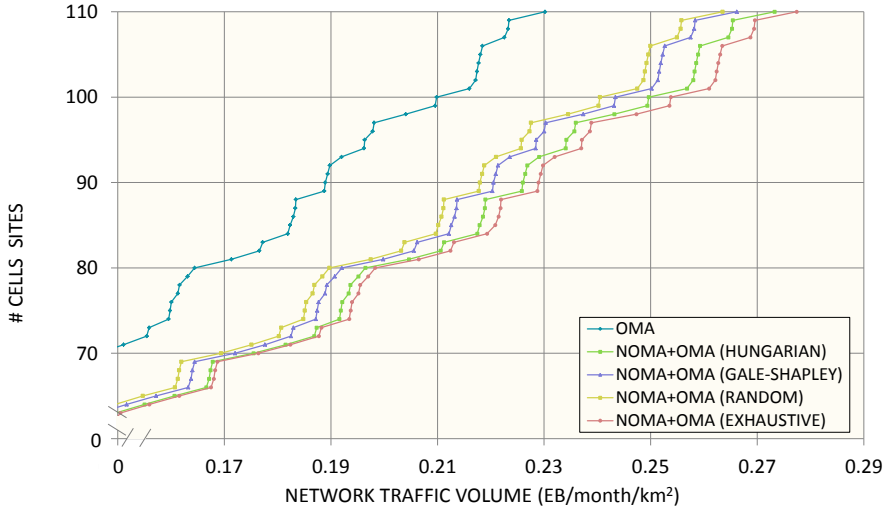


(a) NOMA/OMA

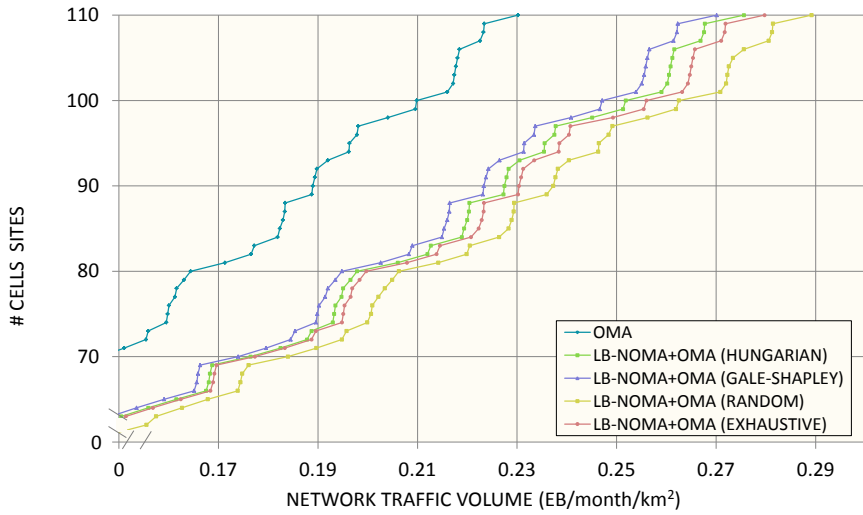


(b) LB-NOMA/OMA

Figure 6.18: Number of cells needed as a function of the traffic volume for OMA and NOMA for four pairing algorithms.

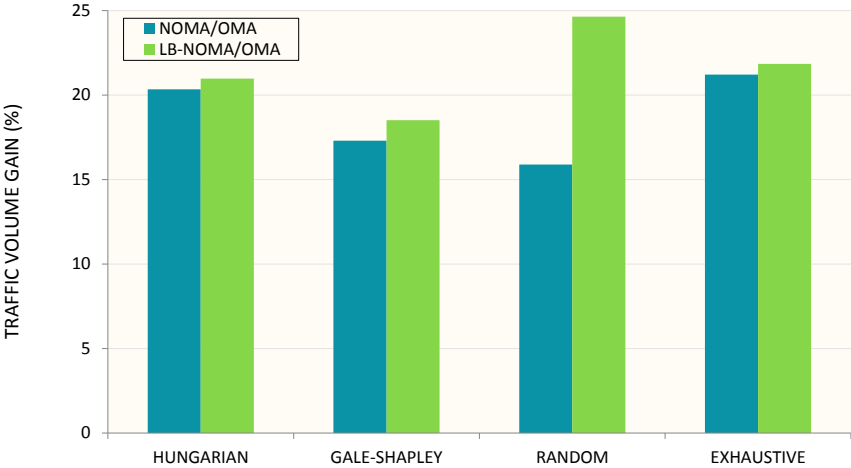


(a) NOMA/OMA

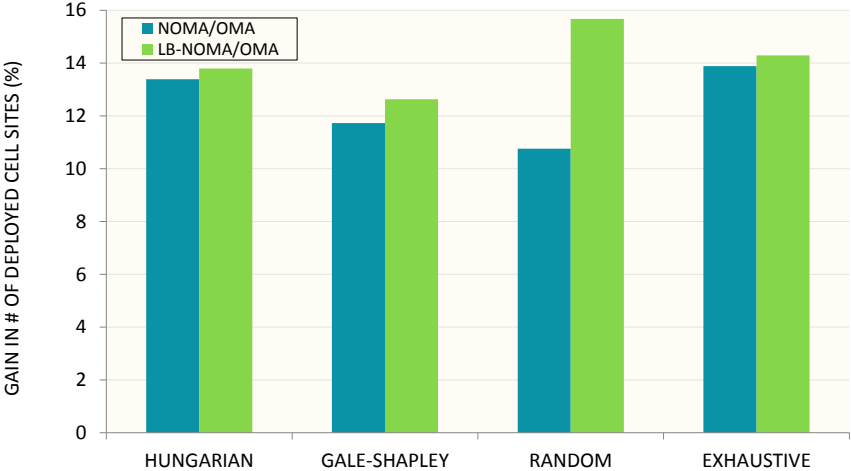


(b) LB-NOMA/OMA

Figure 6.19: Zoomed-in version for the number of cells needed as a function of the traffic volume for OMA and NOMA for four pairing algorithms.



(a) Traffic volume gain



(b) Cell deployment gain

Figure 6.20: Gain for the modeled HetNet for four pairing algorithms for NOMA; OMA is used as a benchmark.

with 14% fewer cells. However, considering the complexity of the exhaustive method, the Hungarian method could offer a better trade-off between complexity and gain, with 13% saving in cells needed. For LB-NOMA/OMA the highest gain is achieved with the random method, with 15.6%. The gain in the number of deployed cells can be translated into a gain in the deployment cost of the network; in this respect, if 15.6% fewer cells are needed to cope with the capacity demand, roughly 15.6% can be saved from the deployment cost, while maintaining the network revenues.

The selection of the pairing method in a hybrid MA HetNet could then be flexible and subject to the load conditions of the cell and its neighbors. The BSs could choose the best user pairing method for NOMA according to the network conditions. This will allow using simpler and faster, but less efficient, pairing algorithms when the load in the cells is unequal, and compensate the inefficiency of the pairing by using LB-NOMA. The optimal or more efficient algorithms could then be reserved for cases where LB-NOMA is not applied, either because all cells have a similar load or because such feature is not available. By having this flexibility in the implementation of NOMA, the network capacity can be improved while lowering the deployment costs.

For example, in the modeled network the Hungarian method could be selected for NOMA, whereas the random method could be used for cases when LB-NOMA is beneficial (unequally loaded cells). With such implementation, and assuming that 100 cells can be deployed (7 macro cells plus 93 small cells), an average network traffic volume of 0.26 EB/month/km<sup>2</sup> can be supported, as shown in Figure 6.21.

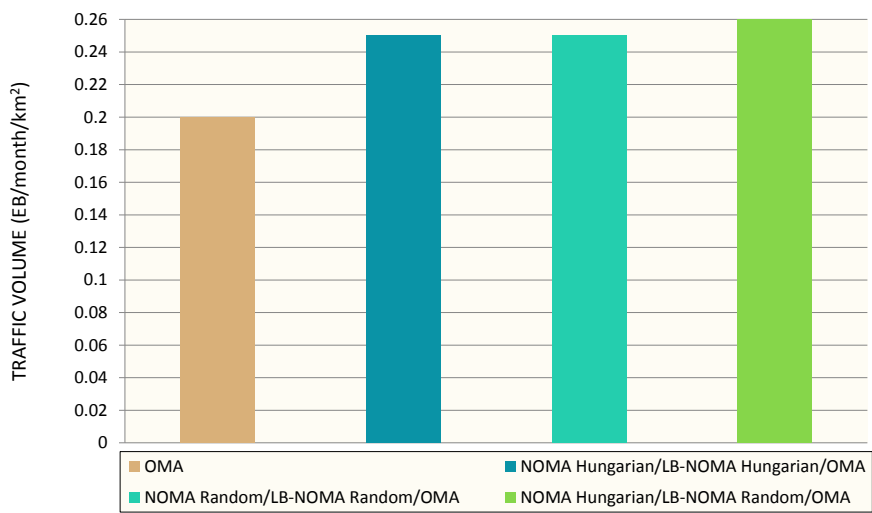


Figure 6.21: Traffic volume that can be supported by the modeled HetNet, with 100 cells deployed for OMA and for NOMA.

In contrast, if only the Hungarian or the random method is used, the same 100 cells can handle 0.25 EB/month/km<sup>2</sup>. This difference of 4% is only due to the flexible choice of the pairing method and directly translates into a 4% gain in the network revenues, since either more UEs or higher data plans can be supported. Moreover, either combination of hybrid MA offers significantly higher capacity than the use of only OMA, with which 0.20 EB/month/km<sup>2</sup> can be supported with the same number of cell cells; this corresponds to capacity gains between 25-30% in the hybrid MA implementation.

For a simple revenue estimation of the network, let us refer to Table 6.6, where 100 cells are considered, as well as the pairing methods with the best trade off between complexity/capacity gain for the modeled network (i.e., Hungarian for NOMA and random for LB-NOMA). Only end users revenues are considered and it is assumed that the price of each GB/month is \$2 (example value estimated from data plans commercially offered nowadays). The highest revenues correspond to the hybrid MA with flexible pairing method, with 2.34 million dollars per month, offering a 4% gain over the hybrid MA with other pairing combinations considered, and a 30% gain over OMA.

Table 6.6: Simple network revenue estimation based only on the end users revenues, for three combinations of multiple access.

Multiple Access	OMA	NOMA + OMA	NOMA/LB-NOMA + OMA		
Pairing Method	N/A	Hungarian	NOMA: Hungarian LB-NOMA: Random	Only Hungarian	Only Random
# Cells	100				
Traffic volume (EB/month/km <sup>2</sup> )	0.2	0.25	0.26	0.25	0.25
Network area (km <sup>2</sup> )	4.501				
Price per GB (\$)	2				
Revenue (MM\$/month)	1.80	2.25	2.34	2.25	2.25
Revenue gain (Benchmark: OMA)	N/A	25	30	25	25

Furthermore, when considering the 2376 UEs simulated, an average of 493 GB/month/UE can be offered with the hybrid MA and flexible pairing as shown in Table 6.7, being this the highest monthly data allowance. Since such a huge amount of data will be likely too high for the average monthly UE consumption, a data plan can be set to make a reverse estimation of how many UEs could be served monthly. For a plan of 30 GB/UE/month, with the hybrid MA and flexible pairing up to 58,513 UEs could be served by the network, versus 56,263 and 45,010 for the other hybrid MA options and OMA, respectively. These numbers result interesting considering the high data demands and

increasing number of connected devices connected that are expected for 5G networks. It is important to remember though, that the high data rate values and number of UEs served are not only due to the implementation of NOMA or LB-NOMA; the 800 MHz channel bandwidth possible at mmWave frequencies plays a key role in achieving this performance.

Table 6.7: Simple network user capacity estimation, for three combinations of multiple access.

Multiple Access	OMA	NOMA + OMA	NOMA/LB-NOMA + OMA		
Pairing Method	N/A	Hungarian	NOMA: Hungarian LB-NOMA: Random	Only Hungarian	Only Random
# Cells	100				
Traffic volume (EB/month/km <sup>2</sup> )	0.2	0.25	0.26	0.25	0.25
Network area (km <sup>2</sup> )*	4.501				
GB/month/UE (for 2299 simulated UEs)	379	474	493	474	474
UEs/month (for 30 GB/UE/month)	45,010	56,263	58,513	56,263	56,263

To also analyze how the load of the neighboring cells affect the traffic that can be handled by the network, let us consider 100 cells and the scenarios shown in Table 6.8. Different loads have been set for the small cells in the model, and an MA configuration of LB-NOMA Random + NOMA Hungarian + OMA is used. The results are shown in Figure 6.22, where it can be seen that when only two of the cells (i.e., small cells 1 and 6) are fully loaded, the lower the load in the remaining cells the higher the traffic volume. This is because the cells with load low can offload those with higher load thanks to the use of LB-NOMA, thus improving the network performance. However, as more cells start to reach their limit load (e.g., a load higher than 80%) as is the case for Scenarios 4 and 5, the growth in the traffic supported starts to stabilize. The reason for this is that the cells that previously helped to balance the load, now have fewer available PRBs to help offload their neighboring cells. For such scenarios, LB-NOMA is no longer used and the system relies only on the Hungarian pairing for NOMA and on OMA. From these results it can also be seen, that when using such MA configuration the variation of the traffic supported is not too large, with a minimum of 0.24 EB/month/km<sup>2</sup> for Scenario 5 and a maximum of 0.28 EB/month/km<sup>2</sup>. This is beneficial when it comes to dimensioning the network, since a highly accurate estimation of the capacity and number of cells/served UEs can be done regardless of how variations in the cells traffic.

The results obtained in this section offer an overview of the performance gains and

Table 6.8: Scenarios with variable load for the small cells in the modeled HetNet.

SCENARIO	SMALL CELL #						
	1	2	3	4	5	6	7
1	100	20			40	100	40
2		30			60		60
3		60			80		80
4		80			100		100
5	100						

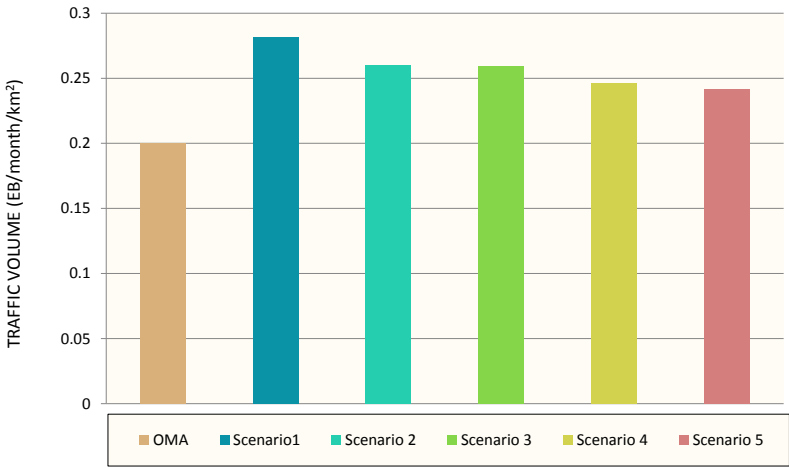


Figure 6.22: Traffic volume that can be supported by the modeled HetNet with variable load of the small cells, for 100 cells deployed for OMA and for NOMA.

challenges that can be expected in 5G HetNets with mmWave and NOMA. In particular, the selection of the user pairing method is a topic that requires detailed planning, since it has a notorious influence on the network capacity. Furthermore, not only one pairing method should be considered; rather a pool of available approaches for user pairing should be considered and have the network select the one that best fits for each cell according to its load conditions and that of its direct neighbors. The effectiveness of the pairing process can be further enhanced when load balancing techniques, such as the proposed LB-NOMA, are used; hence, continuing benefiting from the high spectral efficiency of NOMA and the large bandwidth of mmWave. Moreover, through the appealing results obtained from the simple capacity dimensioning shown, it can be concluded that with a multi-RAT air interface and dense HetNets deployments, 5G networks can cope with the demanding capacity requirements that are set by the IMT-2020.



The analysis of HetNets with hybrid MA and mmWave is still at early stages, and more research work on how the resource management on such networks influences the dimensioning and deployment of 5G networks is highly anticipated. Considering, for example, mMIMO and 3D beamforming can give a more accurate overview of the capacity that can be expected for the 5G RAN. Using a mix of mmWave and sub 6 GHz for the small cells can also provide an interesting inside to the network dimensioning.

## 6.5 Chapter summary

One of the main drivers of 5G networks is the delivery of much higher capacity than the one offered in nowadays networks. As 5G will be characterized by dense HetNets with a multi-RAT architecture and a combination of key technologies, the capacity dimensioning of a network with such characteristics is highly attractive.

When considering the MA schemes, the implementation of NOMA and its coexistence with OMA brings both benefits and challenges to the network. To maximize the capacity benefits of NOMA, the network should be designed to take advantage of the inequality in the load of neighboring cells and the inter-cell interference, and use it to enhance the network performance. This is the basis of the proposed technique LB-NOMA that helps to balance the load and to maximize the NOMA UEs in highly loaded cells while minimizing the OMA UEs. With LB-NOMA, a system capacity gain between 12-18% can be achieved in comparison to NOMA.

The success of LB-NOMA is tightly related to the user pairing process, which is one of the major challenges with NOMA. This is also one of the obstacles to overcome when combining NOMA with mmWave transmissions. Although the use of narrow beamwidths with mmWave brings benefits related to the much-needed improvements in the SINR perceived by the NOMA UEs, it also reduces the pairing probability which is directly related to the capacity gain in the system. Therefore, when combining these RATs the operators must carefully design the network to adjust the beamforming parameters according to, for example, the UEs density, UE location, and QoS desired.

Nevertheless, with mmWave, NOMA and LB-NOMA, huge improvements are expected for the capacity performance of the 5G HetNets. For the modeled network with 100 cells deployed (7 macro cells and 93 small cells), monthly traffic volume of up to 0.26 EB/month/km<sup>2</sup> can be handled, serving around 55,000 mobile devices a month with data plans of up to 30 GB/months. These huge numbers give a clear idea of what can be expected for 5G networks and their capacity to cope with the exponential traffic growth and the IoT.

In overall, although the implementation of mmWave NOMA in HetNets comes with many challenges related mainly to the beamforming parameters and the user pairing, significant capacity improvements can be achieved with such a network architecture. Leading this to the need to propose or optimize solutions that can allow for a smooth integration of these technologies into 5G.



## Conclusion and outlook

---

Capacity enhancements and providing connectivity to a massive amount of devices are two of the main drivers of the 5G mobile networks. To achieve this, 5G will rely on multiple RATs tightly interworking on a flexible network architecture. As more and more capacity is needed, network densification by a massive deployment of small cells will characterize the 5G RAN, along with more spectrum and a better utilization of it. Therefore, 5G will exploit the benefits of the large amount of free spectrum available at mmWave frequencies; these frequencies have not been used before for mobile communications. Thus, their use represents a new paradigm for 5G. Furthermore, by effectively reusing the available resources, NOMA can help increase the spectral efficiency and contribute to the capacity enhancement for 5G.

The performance analysis of mmWave and NOMA on HetNets from a capacity perspective has been presented in this thesis. Single site scenarios as well as network scenarios have been considered, and through computational simulations and results analysis, some of the challenges that the use of mmWave and NOMA impose on the network are addressed and solutions are proposed to overcome the limitations.

### **In relation to mmWave**

Because of the propagation characteristics at these frequencies, beamforming techniques have to be used to benefit from the large spectrum that they offer. We have addressed that this aspect affects the cell discovery process for small cells operating at mmWave, making it differs from that used in microwave small cells. To tackle this issue, a proposed search method for the cell discovery, named JSM and based on cooperation between the macro cell and the small cell, was considered and its performance compared to generic search methods. The results show that the JSM can offer significant improvements with a reduction of up to 96% in the number of pilot signal transmissions that the small cell need to send to find the UE. Furthermore, this improvement comes along with the use of narrow beamwidths, allowing to have a network that offers a high probability of connection to small cells with few cells deployed, low discovery delay, and high data rates. Therefore, this kind of approach is recommended for the future 5G deployments since it would speed up the search process and reduce the initial access delay, while maintaining the directionality that characterizes mmWave transmissions.

The benefits of using mmWaves are further evaluated from a dimensioning perspective by comparing two HetNet deployments; in one, microwave frequencies are used for all cells in the network, while in the other, mmWaves are used for the small cells and

microwave for the macro cells. The results showed that for the proposed scenario up to 12% improvement can be achieved in the number of cells needed to handle the same network traffic volume when mmWave is used. This confirms the advantages that mmWave brings to dimensioning and deployment of 5G HetNets.

### **In relation to NOMA**

Although implementing NOMA comes with many challenges for the network, the results and analysis provided confirm that these can be overcome by selecting appropriate scheduling/resource allocation algorithms, thus getting the benefits of NOMA without compromising the QoS (e.g., the BLER). The channel conditions of the NOMA UEs will be affected by the extra intra-cell interference added with this MA scheme, in comparison to OMA; this poses challenges in hybrid MA access where the CQI is reported by the UEs to the BS assuming an OMA transmission.

Through simulations it was determined that an average difference of 12 dB can be expected between the NOMA SINR and the OMA SINR; therefore CQI and MCS adjustments must be applied to meet the UEs QoS requirements. As the research work related to this topic is rather scarce, future work aiming at characterizing such difference in a more accurate and flexible way, depending on the channel conditions of the multiplexed UEs, is very much needed.

A solution to mitigate the intra-cell interference in NOMA, based on MCS adjustments and extra transmission power for NOMA done at the BS has been proposed, aiming at targeting the desired BLER and improving the data rates for the UEs when switching from OMA to NOMA. With this solution, high capacity gains can be achieved, at the expenses requiring a power headroom. However, when no power headroom is available then the MCS adjustment can still be implemented at the scheduler but alternative scheduling/pairing processes should be considered in order to guarantee a performance improvement. Further work related to solutions of how to mitigate the effects of the intra-cell interference in NOMA, especially in hybrid MA cases, is very much needed. So far it is known though that some kind of adjustment must be done at the BS side in order reach an optimal point of capacity gain and QoS delivered for NOMA.

The actual gain that can be achieved with NOMA is, however, tightly related to the user-pairing and power allocation algorithm. Five pairing methods have been considered throughout the work presented, showing how dynamic and fundamental is the selection of the UEs to be multiplexed. The selection of the pairing method should be done based on its complexity and the capacity gains it can offer. But since the pairing selection depends on the UEs location and propagation conditions, choosing the method that offers the best performance is not a simple task.

The complexity of the pairing method becomes even more complex when mobility is considered, since the channel conditions of the involve UEs can rapidly change. In this regard, future work focusing on dynamic pairing solutions for NOMA will result beneficial for the deployment of 5G networks.

Special emphasis on the number of multiplexed UEs per pair, resource allocation (i.e., spectrum and power), interference management, integration with mmWave and MIMO, are some of the issues that are at their early stages and therefore their solutions have not yet been concretely defined for NOMA and 5G.

### **In relation to HetNets capacity dimensioning with mmWave and NOMA**

As 5G will be characterized by dense HetNets with a multi-RAT architecture and multiple new key technologies coexisting, the capacity dimensioning of a network with such characteristics is highly attractive.

When considering the MA schemes, the implementation of NOMA and its coexistence with OMA brings both benefits and challenges to the network. To maximize the capacity benefits of NOMA, the network should be designed to take advantage of the inequality in the load of neighboring cells and the inter-cell interference, and use it to enhance the network performance. This is the basis of the proposed scheme LB-NOMA based on cooperative scheduling and load balancing among a group of unequally loaded neighboring cells. With LB-NOMA force handovers are performed in highly loaded cells on the OMA UEs located in the overlapping area of neighboring cells. Significant capacity gain can be achieved with LB-NOMA in comparison to NOMA, making it an attractive solution for scenarios like UDNs and hot spots. The success of LB-NOMA is tightly related to the user pairing process.

When combining NOMA with mmWave transmissions the performance is also highly dependent on the pairing method. The use of narrow beamwidths with mmWave allows improving the channel conditions for the NOMA UEs, so the extra transmission power considered for NOMA in hybrid MA system can come from the beamforming gain rather than a power headroom. Nevertheless, the drawback is that such directional transmissions reduce the pairing probability, which is directly related to the capacity gain in the system. Therefore, when combining these mmWave and NOMA the operators must carefully design the network to adjust the beamforming parameters according to, for example, the UEs density, UE location, and QoS desired.

A capacity dimensioning based on the number of UEs served and the network revenues was presented for a 5G HetNet with mmWave, NOMA, and LB-NOMA. Huge capacity improvements were obtained for this scenario, demonstrating that such kind of deployment is an important component of the solution to the exponential traffic growth and the IoT.

The performance of HetNets with hybrid MA and mmWave is still at early stages, and more research work on how the resource management on such networks influences the dimensioning and deployment of the network is highly anticipated. Considering, for example, mMIMO and 3D beamforming can give a more accurate overview of the capacity that can be expected for the 5G RAN. Using a mix of mmWave and sub 6 GHz for the small cells can also provide an interesting inside to the network dimensioning.



# Combined references

---

## References from Chapter 1

- [Eri17a] Ericsson. *Ericsson White Paper: 5G Systems – Enabling the transformation of industry and society*. Tech. rep. January. 2017. URL: <https://www.ericsson.com/res/docs/whitepapers/wp-5g-systems.pdf> (cit. on p. 3).
- [Eri17b] Ericsson. *Mobility Report*. Tech. rep. November. 2017. DOI: 10.3103/S0005105510050031. URL: <https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf> (cit. on pp. 1, 2).
- [Qua16] Qualcomm Technologies. *Making 5G NR a reality - Leading the technology inventions for a unified, more capable 5G air interface*. Tech. rep. December. 2016 (cit. on p. 3).

## References from Chapter 2

- [3GP17a] 3GPP. *3GPP TR 36.932 V14.0.0: Scenarios and Requirements for Small Cells Enhancements for E-UTRA and E-UTRAN (Release 14)*. 2017 (cit. on p. 10).
- [3GP17b] 3GPP. *3GPP TS 36.213 V13.7.0: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 13)*. 2017 (cit. on p. 21).
- [3GP17c] 3GPP. *3GPP TS 36.213 V14.4.0: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 14)*. 2017 (cit. on pp. 7, 22).
- [Ben+14] A. Benjebbour *et al.* “System-Level Performance of Downlink NOMA Combined with SU-MIMO for Future LTE Enhancements”. In: *Globecom Workshops*. 2014 (cit. on p. 24).
- [Boc+14] F. Boccardi *et al.* “Five Disruptive Technology Directions for 5G.pdf”. In: *IEEE Communications Magazine* 52.February (2014), pp. 74–80. URL: <http://dblp.uni-trier.de/db/journals/cm/cm52.html%7B%5C%7DBoccardiHLMP14> (cit. on p. 22).
- [BL16] T. E. Bogale and L. B. Le. “Massive MIMO and mmWave for 5G Wireless HetNet: Potential Benefits and Challenges”. In: *IEEE Vehicular Technology Magazine* 11.1 (2016), pp. 64–75 (cit. on pp. 8, 22).

- [Deh+14] C. Dehos *et al.* “Millimeter-Wave Access and Backhauling: The Solution to the Exponential Data Traffic Increase in 5G Mobile Communications Systems?” In: *IEEE Communications Magazine* 52.9 (2014), pp. 88–95 (cit. on pp. 11, 18).
- [Din+17] Z. Ding *et al.* “Application of Non-orthogonal Multiple Access in LTE and 5G Networks”. In: *IEEE Communications Magazine* 55.2 (2017), pp. 185–191. arXiv: 1511.08610. URL: <http://arxiv.org/abs/1511.08610> (cit. on pp. 23, 24).
- [Eri17a] Ericsson. *Ericsson and Telstra conduct world’s first 5G data call on 26 GHz*. Nov. 2017. URL: <https://www.ericsson.com/en/news/2017/11/ericsson-and-telstra-conduct-worlds-first-5g-data-call-on-26ghz> (cit. on p. 16).
- [Eri17b] Ericsson. *Ericsson White Paper: 5G Systems – Enabling the transformation of industry and society*. Tech. rep. January. 2017. URL: <https://www.ericsson.com/res/docs/whitepapers/wp-5g-systems.pdf> (cit. on p. 12).
- [GF15] M. Giovanni and L. Frecassetti. *ETSI White Paper No. 9: E-Band and V-Band - Survey on Status of Worldwide Regulation*. Tech. rep. 2015 (cit. on p. 16).
- [HTR16] H. Holma, A. Toskala, and J. Reunanen. *LTE Small Cell Optimization: 3GPP Evolution to Release 13*. John Wiley & Sons, 2016 (cit. on pp. 8, 10).
- [ITU] ITU. *ITU Towards “IMT for 2020 and Beyond”*. URL: <http://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Pages/default.aspx> (cit. on p. 7).
- [Ji+17] H. Ji *et al.* “Overview of Full-Dimension MIMO in LTE-Advanced Pro”. In: *IEEE Communications Magazine* 55.2 (2017), pp. 176–184. ISSN: 01636804. DOI: 10.1109/MCOM.2016.1500743RP. arXiv: 1601.0019 (cit. on pp. 21, 22).
- [MSR15] G. R. Maccartney, M. K. Samimi, and T. S. Rappaport. “Exploiting directionality for millimeter-wave wireless system improvement”. In: *IEEE International Conference on Communications*. London, UK, 2015, pp. 2416–2422. ISBN: 9781467364324. DOI: 10.1109/ICC.2015.7248687. arXiv: arXiv:1503.05265v1 (cit. on p. 18).
- [Moh+16] A. Mohamed *et al.* “Control-Data Separation Architecture for Cellular Radio Access Networks: A Survey and Outlook”. In: *IEEE Communications Surveys & Tutorials* 18.1 (2016), pp. 446–465 (cit. on p. 11).
- [Mus+16] H. A. U. Mustafa *et al.* “Separation Framework: An Enabler for Cooperative and D2D Communication for Future 5G Networks”. In: *IEEE Communications Surveys & Tutorials* 18.1 (2016), pp. 419–445 (cit. on p. 11).



- [NTT14] NTT DOCOMO. *DOCOMO 5G White Paper - 5G Radio Access: Requirements, Concept and Technologies*. Tech. rep. 2014. URL: [https://www.nttdocomo.co.jp/english/corporate/technology/whitepaper%7B%5C\\_%7D5g/](https://www.nttdocomo.co.jp/english/corporate/technology/whitepaper%7B%5C_%7D5g/) (cit. on p. 11).
- [OMM16] A. Osseiran, J. F. Monserrat, and P. Marsch. *5G Mobile and Wireless Communications Technology*. Cambridge University Press, 2016 (cit. on p. 12).
- [Qual16] Qualcomm Technologies. *Making 5G NR a reality - Leading the technology inventions for a unified, more capable 5G air interface*. Tech. rep. December. 2016 (cit. on p. 12).
- [Qual17a] Qualcomm Technologies. *Mobilizing 5G NR Millimeter Wave: Network Coverage Simulation Studies for Global Cities October 2017*. Tech. rep. October. 2017. URL: <https://www.qualcomm.com/documents/white-paper-5g-nr-millimeter-wave-network-coverage-simulation> (cit. on pp. 18, 20).
- [Qual17b] Qualcomm Technologies. *Spectrum for 4G and 5G*. Tech. rep. 2017. URL: <https://www.qualcomm.com/media/documents/spectrum-4g-and-5g> (cit. on p. 16).
- [RG06] A. Rahman and P. Gburzynski. “Hidden problems with the hidden node problem”. In: *23rd Biennial Symposium on Communications*. Kigston, Ont., Canada, 2006, pp. 271–273. ISBN: 078039528X. DOI: 10.1109/BSC.2006.1644620 (cit. on p. 20).
- [Rap+14] T. S. Rappaport *et al.* *Millimeter Wave - Wireless Communications*. Prentice Hall, 2014 (cit. on pp. 16, 18, 19).
- [Rap+13a] T. S. Rappaport *et al.* “Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor Urban cellular communications”. In: *IEEE Transactions on Antennas and Propagation* 61.4 (2013), pp. 1850–1859. ISSN: 0018926X. DOI: 10.1109/TAP.2012.2235056 (cit. on p. 18).
- [Rap+13b] T. S. Rappaport *et al.* “Millimeter wave mobile communications for 5G cellular: It will work!” In: *IEEE Access* 1 (2013), pp. 335–349. ISSN: 21693536. DOI: 10.1109/ACCESS.2013.2260813. arXiv: 1507.04592 (cit. on pp. 8, 15).
- [Ria+17] S. M. Riazul Islam *et al.* “Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges”. In: *IEEE Communications Surveys and Tutorials* 19.2 (2017) (cit. on p. 24).
- [Roh+14] W. Roh *et al.* “Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results”. In: *IEEE Communications Magazine* 52.2 (2014), pp. 106–113. ISSN: 01636804. DOI: 10.1109/MCOM.2014.6736750 (cit. on p. 18).

- [Sai+13] Y. Saito *et al.* “Non-orthogonal multiple access (NOMA) for cellular future radio access”. In: *IEEE Vehicular Technology Conference* (2013), pp. 0–4. ISSN: 15502252. DOI: 10.1109/VTCSpring.2013.6692652 (cit. on p. 23).
- [Sam13] Samsung. *Samsung Announces World’s First 5G mmWave Mobile Technology*. May 2013. URL: <https://news.samsung.com/global/samsung-announces-worlds-first-5g-mmwave-mobile-technology> (cit. on p. 18).
- [San+16] Y. Sano *et al.* “LTE-Advanced Release 13 Multiple Antenna Technologies and Improved Reception Technologies”. In: *NTT DOCOMO Technical Journal* 18.2 (2016), pp. 39–51 (cit. on p. 21).
- [Tey+17] O. Teyeb *et al.* “Evolving LTE to Fit the 5G Future”. In: *Ericsson Technology Review* (Jan. 2017), pp. 1–17 (cit. on p. 7).
- [Tit+12] D. Titz *et al.* “Design and characterization of cmos on-chip antennas for 60 ghz Communications”. In: *Radioengineering* 21.1 (2012), pp. 324–332. ISSN: 12102512. URL: <https://dspace.vutbr.cz/handle/11012/37048> (cit. on p. 19).
- [WXP06] P. Wang, J. Xiao, and L. P. “Comparison of Orthogonal and Non-Orthogonal Approaches to Future Wireless Cellular Systems”. In: *IEEE Vehicular Technology Conference*. Vol. 1. 3. 2006, pp. 4–11 (cit. on p. 22).
- [Yua+16] Y. Yuan *et al.* “Non-Orthogonal Transmission Technology in LTE Evolution”. In: *IEEE Communications Magazine* 54.7 (2016), pp. 68–74 (cit. on p. 24).
- [Zha+17] J. Zhang *et al.* “5G Millimeter-Wave Antenna Array: Design and Challenges”. In: *IEEE Wireless Communications* 24.2 (2017), pp. 106–112. ISSN: 15361284. DOI: 10.1109/MWC.2016.1400374RP (cit. on p. 19).
- [ZL09] Y. P. Zhang and D. Liu. “Antenna-on-chip and antenna-in-package solutions to highly integrated millimeter-wave devices for wireless communications”. In: *IEEE Transactions on Antennas and Propagation* 57.10 PART 1 (2009), pp. 2830–2841. ISSN: 0018926X. DOI: 10.1109/TAP.2009.2029295 (cit. on p. 19).

## References from Chapter 3

- [3GP16] 3GPP. *3GPP TR 36.859 V13.0.0: Study on Downlink Multiuser Superposition Transmission (MUST) for LTE (Release 13)*. 2016 (cit. on pp. 27, 33).
- [Ben+15] A. Benjebbour *et al.* “Non-orthogonal Multiple Access (NOMA): Concept, Performance Evaluation and Experimental Trials”. In: *International Conference on Wireless Networks and Mobile Communications (WINCOM)*. 2015 (cit. on pp. 29, 33).

- [Cov72] T. M. Cover. "Broadcast Channels". In: *IEEE Transactions on Information Theory* 18.1 (1972) (cit. on pp. 27, 28).
- [Dai+15] L. Dai *et al.* "Non-Orthogonal Multiple Access for 5G: Solutions, Challenges, Opportunities, and Future Research Trends". In: *IEEE Communications Magazine* 53.9 (2015), pp. 74–81 (cit. on pp. 32, 33).
- [DFP16] Z. Ding, P. Fan, and H. V. Poor. "Impact of User Pairing on 5G Nonorthogonal Multiple-Access Downlink Transmissions". In: *IEEE Transactions on Vehicular Technology* 65.8 (2016) (cit. on p. 32).
- [Din+17a] Z. Ding *et al.* "A Survey on Non-Orthogonal Multiple Access for 5G Networks: Research Challenges and Future Trends". In: *IEEE Journal on Selected Areas in Communications* (2017) (cit. on pp. 32, 33).
- [Din+17b] Z. Ding *et al.* "Application of Non-orthogonal Multiple Access in LTE and 5G Networks". In: *IEEE Communications Magazine* 55.2 (2017), pp. 185–191. arXiv: 1511.08610. URL: <http://arxiv.org/abs/1511.08610> (cit. on pp. 31–33).
- [Ria+17] S. M. Riazul Islam *et al.* "Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges". In: *IEEE Communications Surveys and Tutorials* 19.2 (2017) (cit. on pp. 29, 33).
- [Sai+13] Y. Saito *et al.* "Non-orthogonal multiple access (NOMA) for cellular future radio access". In: *IEEE Vehicular Technology Conference* (2013), pp. 0–4. ISSN: 15502252. DOI: 10.1109/VTCSpring.2013.6692652 (cit. on p. 29).
- [Shi+17] W. Shin *et al.* "Non-Orthogonal Multiple Access in Multi-Cell Networks: Theory, Performance, and Practical Challenges". In: *IEEE Communications Magazine* October (2017), pp. 176–183. ISSN: 01636804. DOI: 10.1109/MCOM.2017.1601065. arXiv: 1611.01607 (cit. on p. 33).
- [Van+12] S. Vanka *et al.* "Superposition Coding Strategies: Design and Experimental Evaluation". In: *IEEE Transactions on Wireless Communications* 11.7 (2012), pp. 2628–2639 (cit. on p. 29).
- [Yua+16] Y. Yuan *et al.* "Non-Orthogonal Transmission Technology in LTE Evolution". In: *IEEE Communications Magazine* 54.7 (2016), pp. 68–74 (cit. on p. 33).

## References from Chapter 4

- [Bar+15] C. N. Barati *et al.* "Directional Cell Discovery in Millimeter Wave Cellular Networks". In: *IEEE Transactions on Wireless Communications* 14.12 (2015), pp. 6664–6678. ISSN: 15361276. DOI: 10.1109/TWC.2015.2457921. arXiv: 1404.5068 (cit. on p. 37).

- [CFS15] A. Capone, I. Filippini, and V. Sciancalepore. “Context Information for Fast Cell Discovery in mm-wave 5G Networks”. In: *Proceedings of European Wireless 2015; 21th European Wireless Conference*. Budapest, Hungary, 2015, pp. 1–6. ISBN: 3800739763. arXiv: 1501.02223. URL: <http://arxiv.org/abs/1501.02223> (cit. on p. 38).
- [Deh+14] C. Dehos *et al.* “Millimeter-Wave Access and Backhauling: The Solution to the Exponential Data Traffic Increase in 5G Mobile Communications Systems?” In: *IEEE Communications Magazine* 52.9 (2014), pp. 88–95 (cit. on p. 38).
- [Des+14] V. Desai *et al.* “Initial beamforming for mmWave communications”. In: *48th Asilomar Conference on Signals, Systems and Computers*. Pacific Grove, CA, USA, 2014, pp. 1926–1930 (cit. on p. 37).
- [GMZ16] M. Giordani, M. Mezzavilla, and M. Zorzi. “Initial Access in 5G mmWave Cellular Networks”. In: *IEEE Communications Magazine* 54.11 (2016), pp. 40–47. ISSN: 01636804. DOI: 10.1109/MCOM.2016.1600193CM. arXiv: 1605.00101 (cit. on p. 38).
- [Gio+16] M. Giordani *et al.* “Comparative Analysis of Initial Access Techniques in 5G mmWave Cellular Networks”. In: *Annual Conference on Information Science and Systems (CISS)*. Princeton, NJ, USA, 2016. ISBN: 9781467394574 (cit. on p. 38).
- [Hab+17] S. Habib *et al.* “Millimeter wave cell search for initial access: Analysis, design, and implementation”. In: *13th International Wireless Communications and Mobile Computing Conference, IWCMC 2017*. Ci. Valencia, Spain, 2017, pp. 922–927. ISBN: 9781509043729. DOI: 10.1109/IWCMC.2017.7986409 (cit. on p. 38).
- [JPY15] C. Jeong, J. Park, and H. Yu. “Random access in millimeter-wave beamforming cellular networks: Issues and approaches”. In: *IEEE Communications Magazine* 53.1 (2015), pp. 180–185. ISSN: 01636804. DOI: 10.1109/MCOM.2015.7010532 (cit. on p. 38).
- [Li+13] Q. C. Li *et al.* “Anchor-booster based heterogeneous networks with mmWave capable booster cells”. In: *2013 IEEE Globecom Workshops*. Atlanta, GA, USA, 2013, pp. 93–98. ISBN: 9781479928514. DOI: 10.1109/GLOCOMW.2013.6824968 (cit. on p. 38).
- [Med+09] J. Medbo *et al.* “Propagation Channel Impact on LTE Positioning Accuracy - A Study Based on Real Measurements of Observed Time Difference of Arrival”. In: *IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC*. Tokyo, Japan, 2009, pp. 2213–2217. ISBN: 9781424451234. DOI: 10.1109/PIMRC.2009.5450144 (cit. on p. 47).

- [PZ17] R. Parada and M. Zorzi. “Cell discovery based on historical users location in mmWave 5G”. In: *Proceedings of European Wireless 2017; 23th European Wireless Conference*; Dresden, Germany, 2017, pp. 1–6. ISBN: 9783800744268. URL: <http://ieeexplore.ieee.org/document/8011308/> (cit. on p. 38).
- [Pra+16] A. Prasad *et al.* “Network assisted small cell discovery in multi-layer and mmWave networks”. In: *2015 IEEE Conference on Standards for Communications and Networking, CSCN 2015*. Tokyo, Japan, 2016, pp. 118–123. ISBN: 9781479989287. DOI: 10.1109/CSCN.2015.7390431 (cit. on p. 38).
- [Rap+14] T. S. Rappaport *et al.* *Millimeter Wave - Wireless Communications*. Prentice Hall, 2014 (cit. on p. 35).
- [Rap+13] T. S. Rappaport *et al.* “Millimeter wave mobile communications for 5G cellular: It will work!” In: *IEEE Access* 1 (2013), pp. 335–349. ISSN: 21693536. DOI: 10.1109/ACCESS.2013.2260813. arXiv: 1507.04592 (cit. on p. 56).
- [Riv] Riverbed. *Riverbed Technology*. URL: <https://www.riverbed.com/dk/index.html> (cit. on p. 44).
- [Sho+15] H. Shokri-Ghadikolaei *et al.* “Millimeter wave cellular networks: A MAC layer perspective”. In: *IEEE Transactions on Communications* 63.10 (2015), pp. 3437–3458. ISSN: 00906778. DOI: 10.1109/TCOMM.2015.2456093. arXiv: 1503.00697 (cit. on pp. 37, 44).
- [Wei+14] R. J. Weiler *et al.* “Enabling 5G Backhaul and Access with millimeter-waves”. In: *European Conference on Networks and Communications, EuCNC*. Bologna, Italy, 2014. ISBN: 9781479952809. DOI: 10.1109/EuCNC.2014.6882644 (cit. on p. 54).

## References from Chapter 5

- [3GP14] 3GPP. *3GPP R1-142212: On configuration and design of 256QAM tables in downlink*. 2014 (cit. on p. 60).
- [3GP17a] 3GPP. *3GPP TS 36.213 V14.4.0: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 14)*. 2017 (cit. on pp. 61, 67, 79).
- [3GP17b] 3GPP. *3GPP TS 38.211 V1.2.0: NR; Physical channels and modulation (Release 15)*. 2017 (cit. on p. 59).
- [Ben+14] A. Benjebbour *et al.* “System-Level Performance of Downlink NOMA Combined with SU-MIMO for Future LTE Enhancements”. In: *Globecom Workshops (GC Wkshps)*. 2014 (cit. on pp. 78, 82, 91).

- [Ben+15] A. Benjebbour *et al.* “Non-orthogonal Multiple Access (NOMA): Concept, Performance Evaluation and Experimental Trials”. In: *International Conference on Wireless Networks and Mobile Communications (WINCOM)*. 2015 (cit. on p. 59).
- [DFP16] Z. Ding, P. Fan, and H. V. Poor. “Impact of User Pairing on 5G Nonorthogonal Multiple-Access Downlink Transmissions”. In: *IEEE Transactions on Vehicular Technology* 65.8 (2016), pp. 6010–6023 (cit. on p. 88).
- [Din+17] Z. Ding *et al.* “Application of Non-orthogonal Multiple Access in LTE and 5G Networks”. In: *IEEE Communications Magazine* 55.2 (2017), pp. 185–191. arXiv: 1511.08610. URL: <http://arxiv.org/abs/1511.08610> (cit. on p. 59).
- [GS62] D. Gale and L. S. Shapley. “College Admissions and the Stability of Marriage”. In: *The American Mathematical Monthly* 69.1 (1962) (cit. on p. 89).
- [KB55] H. W. Kuhn and Bryn Yaw. “The Hungarian Method For The Assignment Problem”. In: *Naval Research Logistics* (1955) (cit. on pp. 89, 93).
- [Lia+17] W. Liang *et al.* “User Pairing for Downlink Non-Orthogonal Multiple Access Networks using Matching Algorithm”. In: *IEEE Transactions on Communications* (2017) (cit. on p. 88).
- [LOP15] F. Liu, P. M. Onen, and M. Petrova. “Proportional Fairness-Based User Pairing and Power Allocation for Non-Orthogonal Multiple Access”. In: *IEEE 26th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. 2015 (cit. on p. 88).
- [MW71] D. McVitie and L. Wilson. “The Stable Marriage Problem”. In: *Communications of the ACM* 14.7 (1971) (cit. on p. 94).
- [Mia+16] G. Miao *et al.* *Fundamentals of Mobile Data Networks*. Cambridge: Cambridge University Press, 2016. ISBN: 9781316534298. DOI: 10.1017/CBO9781316534298. URL: <http://goo.gl/eFpcwy> (cit. on p. 89).
- [MS17] F. W. Murti and S. Y. Shin. “User Pairing Schemes Based on Channel Quality Indicator for Uplink Non-Orthogonal Multiple Access”. In: *Ninth International Conference on Ubiquitous and Future Networks (ICUFN)*. 2017 (cit. on p. 88).
- [Rap+14] T. S. Rappaport *et al.* *Millimeter Wave - Wireless Communications*. Prentice Hall, 2014 (cit. on pp. 65, 82).
- [Ria+17] S. M. Riazul Islam *et al.* “Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges”. In: *IEEE Communications Surveys and Tutorials* 19.2 (2017) (cit. on p. 78).
- [Sai+13] Y. Saito *et al.* “Non-orthogonal multiple access (NOMA) for cellular future radio access”. In: *IEEE Vehicular Technology Conference* (2013), pp. 0–4. ISSN: 15502252. DOI: 10.1109/VTCSpring.2013.6692652 (cit. on p. 59).

- [Sha+16] M. B. Shahab *et al.* “User Pairing Schemes for Capacity Maximization in Non-Orthogonal Multiple Access Systems”. In: *Wireless Communications and Mobile Computing* (2016). DOI: 10.1002/wcm.2736 (cit. on p. 88).
- [Son+17] L. Song *et al.* “Resource Management in Non-orthogonal Multiple Access Networks for 5G and Beyond”. In: *IEEE Network* (2017) (cit. on p. 89).
- [Zhu+15] Y. Zhu *et al.* *Non-Orthogonal Multiple Access (NOMA) Wireless Systems and Methods*. 2015 (cit. on pp. 65, 78).

## References from Chapter 6

- [3GP17a] 3GPP. *3GPP TR 36.873 V12.6.0: Study on 3D channel model for LTE (Release 12)*. 2017 (cit. on p. 119).
- [3GP17b] 3GPP. *3GPP TR 38.900 V14.3.1: Study on channel model for frequency spectrum above 6 GHz (Release 14)*. 2017 (cit. on p. 113).
- [3GP17c] 3GPP. *3GPP TS 38.211 V1.2.0: NR; Physical channels and modulation (Release 15)*. 2017 (cit. on p. 102).
- [MMB16] C. Mavromoustakis, G. Mastorakis, and J. M. Batalla. *Internet of Things (IoT) in 5G Mobile Technologies*. 1st ed. Springer International Publishing, 2016. DOI: 10.1007/978-3-319-30913-2 (cit. on p. 122).
- [Por+16] R. Porat *et al.* *IEEE P802.11 Wireless LANs: 11ax Evaluation Methodology (IEEE 802.11-14/0571r12)*. Tech. rep. 2016. URL: [http://grouper.ieee.org/groups/802/11/Reports/tgax%7B%5C\\_%7Dupdate.htm](http://grouper.ieee.org/groups/802/11/Reports/tgax%7B%5C_%7Dupdate.htm) (cit. on p. 122).
- [SBT09] S. Sesia, M. Baker, and I. Toufik. *The UMTS Long Term Evolution: From Theory to Practice*. Wiley, 2009 (cit. on p. 103).
- [SY17] F. Sotiraki and W. Yu. “Hybrid Analog and Digital Beamforming for mmWave OFDM Large-Scale Antenna Arrays”. In: *IEEE Journal on Selected Areas in Communications* 35.7 (2017), pp. 1432–1443. ISSN: 07338716. DOI: 10.1109/JSAC.2017.2698958. arXiv: 1711.08408 (cit. on p. 118).
- [Xia+17] Z. Xiao *et al.* “Millimeter-Wave Communication with Non-Orthogonal Multiple Access for 5G”. In: (2017), pp. 1–16. arXiv: 1709.07980. URL: <http://arxiv.org/abs/1709.07980> (cit. on p. 118).