PhD dissertation

Cellular and Wi-Fi technologies evolution: From complementarity to competition

Author

Biljana Bojović
Centre Tecnològic de Telecomunicacions de Catalunya

PhD Advisor

Dr. Lorenza Giupponi
Centre Tecnològic de Telecomunicacions de Catalunya

PhD Tutor

Prof. Miguel Soriano Ibáñez
Universitat Politècnica de Catalunya

Universitat Politècnica de Catalunya

I dedicate this PhD thesis to my mother Vinka
who has been always inspiring me with her continuous will to learn and grow

* * *

Le dedico esta tesis doctoral a mi madre Vinka
quien siempre me ha estado inspirando con su continua voluntad de aprender y crecer

* * *

Posvećujem ovu doktorsku tezu mojoj majci Vinki
koja me je oduvek inspirisala svojom neprestanom voljom za učenjem i napredovanjem

Abstract

This PhD thesis has the characteristic to span over a long time because while working on it, I was working as a research engineer in CTTC with highly demanding development duties. This has delayed the deposit more than I would have liked. On the other hand, this has given me the privilege of witnessing and studying how wireless technologies have evolved during a decade from 4G to 5G and beyond.

When I started my PhD thesis, IEEE and 3GPP were defining the two main wireless technologies at the time, Wi-Fi and LTE, covering two substantially complementary market targets. Wi-Fi was designed to operate mostly indoor, in unlicensed spectrum, and was aimed to be a simple and cheap technology. Its primary technology for coexistence was based on the assumption that the spectrum on which it was operating was for free, and so it was simply designed with interference avoidance through the famous CSMA/CA protocol. On the other hand, 3GPP was designing technologies for licensed spectrum, a costly kind of spectrum. As a result, LTE is designed to take the best advantage of it while providing the best QoE in mainly outdoor scenarios.

The PhD thesis starts in this context and evolves with these two technologies. In the first chapters, we study radio resource management solutions for standalone operation of Wi-Fi in unlicensed and LTE in licensed spectrum. We anticipated the now fundamental machine learning trend by working on machine learning-based radio resource management solutions to improve LTE and Wi-Fi operation in their respective spectrum. We pay particular attention to small cell deployments aimed at improving the spectrum efficiency in licensed spectrum, reproducing small range scenarios typical of Wi-Fi settings.

IEEE and 3GPP followed evolving the technologies over the years: Wi-Fi has grown into a much more complex and sophisticated technology, incorporating the key features of cellular technologies, like HARQ, OFDMA, MU-MIMO, MAC scheduling and spatial reuse. On the other hand, since Release 13, cellular networks have also been designed for unlicensed spectrum. As a result, the two last chapters of this thesis focus on coexistence scenarios, in which LTE needs to be designed to coexist with Wi-Fi fairly, and NR, the radio access for 5G, with Wi-Fi in 5 GHz and WiGig in 60 GHz. Unlike LTE, which was adapted to operate in unlicensed spectrum, NR-U is natively designed with this feature, including its capability to operate in unlicensed in a complete standalone fashion, a fundamental new milestone for cellular. In this context, our focus of analysis changes. We consider that these two technological families are no longer targeting complementarity but are now competing, and we claim that this will be the trend for the years to come.

To enable the research in these multi-RAT scenarios, another fundamental result of this PhD thesis, besides the scientific contributions, is the release of high fidelity models for LTE and

NR and their coexistence with Wi-Fi and WiGig to the ns-3 open-source community. ns-3 is a popular open-source network simulator, with the characteristic to be multi-RAT and so naturally allows the evaluation of coexistence scenarios between different technologies. These models, for which I led the development, are by academic citations, the most used open-source simulation models for LTE and NR and have received fundings from industry (Ubiquisys, WFA, SpiderCloud, Interdigital, Facebook) and federal agencies (NIST, LLNL) over the years.

Resumen

Esta tesis doctoral tiene la característica de extenderse durante mucho tiempo porque mientras trabajaba en ella, trabajaba como ingeniera de investigación en CTTC con tareas de desarrollo muy exigentes. Esto ha retrasado el depósito más de lo que me hubiera gustado. Por otro lado, gracias a ello, he tenido el privilegio de presenciar y estudiar como las tecnologías inalámbricas han evolucionado durante una década, de 4G a 5G y más allá.

Cuando comencé mi tesis doctoral, IEEE y 3GPP estaban definiendo las dos principales tecnologías inalámbricas en ese momento, Wi-Fi y LTE, cumpliendo dos objetivos de mercado sustancialmente complementarios. Wi-Fi fue diseñado para funcionar principalmente en interiores, en un espectro sin licencia, y estaba destinado a ser una tecnología simple y barata. Su tecnología primaria para la convivencia se basaba en el supuesto en que el espectro en el que estaba operando era gratis, y así fue diseñado simplemente evitando interferencias a través del famoso protocolo CSMA/CA. Por otro lado, 3GPP estaba diseñando tecnologías para espectro con licencia, un tipo de espectro costoso. Como resultado, LTE está diseñado para aprovechar el espectro al máximo proporcionando al mismo tiempo el mejor QoE en escenarios principalmente al aire libre.

La tesis doctoral parte de este contexto y evoluciona con estas dos tecnologías. En los primeros capítulos, estudiamos las soluciones de gestión de recursos de radio para operación en espectro Wi-Fi sin licencia y LTE con licencia. Anticipamos la tendencia ahora fundamental de aprendizaje automático trabajando en soluciones de gestión de recursos de radio para mejorar LTE y funcionamiento de Wi-Fi en su respectivo espectro. Prestamos especial atención a las implementaciones de células pequeñas destinadas a mejorar la eficiencia de espectro licenciado, reproduciendo los típicos escenarios de rango pequeño de la configuración Wi-Fi.

IEEE y 3GPP siguieron evolucionando las tecnologías a lo largo de los años: Wi-Fi se ha convertido en una tecnología mucho más compleja y sofisticada, incorporando las características clave de las tecnologías celulares, como HARQ, OFDMA, MU-MIMO, MAC scheduling y la reutilización espacial. Por otro lado, desde la Release 13, también se han diseñado redes celulares para espectro sin licencia. Como resultado, los dos últimos capítulos de esta tesis se centran en estos escenarios de convivencia, donde LTE debe diseñarse para coexistir con Wi-Fi de manera justa, y NR, el acceso por radio para 5G con Wi-Fi en 5 GHz y WiGig en 60 GHz. A diferencia de LTE, que se adaptó para operar en espectro sin licencia, NR-U está diseñado de forma nativa con esta función, incluyendo su capacidad para operar sin licencia de forma completamente independiente, un nuevo hito fundamental para los celulares. En este contexto, cambia nuestro enfoque de análisis. Consideramos que estas dos familias tecnológicas ya no tienen como objetivo la complementariedad, sino que ahora están compitiendo, y afirmamos que esta será la tendencia para los próximos años.

Para permitir la investigación en estos escenarios de múltiples RAT, otro resultado fundamental de esta tesis doctoral, además de los aportes científicos, es el lanzamiento de modelos de alta fidelidad para LTE y NR y su coexistencia con Wi-Fi y WiGig a la comunidad de código abierto de ns-3. ns-3 es un simulador popular de red de código abierto, con la característica de ser multi-RAT y así, naturalmente, permite la evaluación de escenarios de convivencia entre diferentes tecnologías. Estos modelos, para los cuales lideré el desarrollo, son por citas académicas, los modelos de simulación de código abierto más utilizados para LTE y NR y han recibido fondos de la industria (Ubiquisys, WFA, SpiderCloud, Interdigital, Facebook) y agencias federales (NIST, LLNL) a lo largo de los años.

Resum

Aquesta tesi doctoral té la característica d'allargar-se durant un llarg període de temps ja que mentre treballava en ella, treballava com a enginyera investigadora a CTTC amb tasques de desenvolupament molt exigents. Això ha endarrerit el dipositar-la més del que m'hagués agradat. D'altra banda, això m'ha donat el privilegi de ser testimoni i estudiar com han evolucionat les tecnologies sense fils durant més d'una dècada des del 4G fins al 5G i més enllà.

Quan vaig començar la tesi doctoral, IEEE i 3GPP estaven definint les dues tecnologies sense fils principals en aquell moment, Wi-Fi i LTE, que cobreixen dos objectius de mercat substancialment complementaris. Wi-Fi va ser dissenyat per funcionar principalment en interiors, en espectre sense llicència, i pretenia ser una tecnologia senzilla i barata. La seva tecnologia primària per a la convivència es basava en el supòsit que l'espectre en el que estava operant era de franc, i, per tant, es va dissenyar simplement evitant interferències a través del famós protocol CSMA/CA. D'altra banda, 3GPP estava dissenyant tecnologies per a espectres amb llicència, un tipus d'espectre costós. Com a resultat, LTE està dissenyat per treure'n el màxim profit alhora que proporciona el millor QoE en escenaris principalment a l'aire lliure.

La tesi doctoral comença amb aquest context i evoluciona amb aquestes dues tecnologies. En els primers capítols, estudiem solucions de gestió de recursos de radio per a operacions en espectre de Wi-Fi sense llicència i LTE amb llicència. Hem anticipat l'actual tendència fonamental d'aprenentatge automàtic treballant solucions de gestió de recursos de radio basades en l'aprenentatge automàtic per millorar l'LTE i Wi-Fi en el seu espectre respectiu. Prestem especial atenció als desplegaments de cèl·lules petites destinades a millorar la eficiència d'espectre llicenciat, reproduint escenaris de petit abast típics de la configuració Wi-Fi.

IEEE i 3GPP van seguir evolucionant les tecnologies al llarg dels anys: El Wi-Fi s'ha convertit en una tecnologia molt més complexa i sofisticada, incorporant les característiques clau de les tecnologies cel·lulars, com ara HARQ i la reutilització espacial. D'altra banda, des de la versió 13, també s'han dissenyat xarxes cel·lulars per a espectre sense llicència. Com a resultat, els dos darrers capítols d'aquesta tesi es centren en aquests escenaris de convivència, on s'ha de dissenyar LTE per conviure amb la Wi-Fi de manera justa, i NR, l'accés a la radio per a 5G amb Wi-Fi a 5 GHz i WiGig a 60 GHz. A diferència de LTE, que es va adaptar per funcionar en espectre sense llicència, NR-U està dissenyat de forma nativa amb aquesta característica, inclosa la seva capacitat per operar sense llicéncia de forma autònoma completa, una nova fita fonamental per al mòbil. En aquest context, el nostre focus d'anàlisi canvia. Considerem que aquestes dues famílies de tecnologia ja no estan orientades cap a la complementarietat, sino que ara competeixen, i afirmem que aquesta serà el tendència per als propers anys.

Per permetre la investigació en aquests escenaris multi-RAT, un altre resultat fonamental d'aquesta tesi doctoral, a més de les aportacions científiques, és l'alliberament de models d'alta fidelitat per a LTE i NR i la seva coexistència amb Wi-Fi a la comunitat de codi obert ns-3. ns-3 és un popular simulador de xarxa de codi obert, amb la característica de ser multi-RAT i, per tant, permet l'avaluació de manera natural d'escenaris de convivència entre diferents tecnologies. Aquests models, pels quals he liderat el desenvolupament, són per cites acadèmiques, els models de simulació de codi obert més utilitzats per a LTE i NR i que han rebut finançament de la indústria (Ubiquisys, WFA, SpiderCloud, Interdigital, Facebook) i agències federals (NIST, LLNL) al llarg dels anys.

Acknowledgements

This doctoral thesis was a long journey that made me grow not only professionally, but also personally through meeting many beautiful people.

I would especially like to thank:

Lazar Berbakov, for helping me start this PhD adventure, and for being a great friend along the way. So Ivana, his wife, for being my honest friend.

Nicola Baldo, for transmitting to me for many years his curiosity and his love for telecommunications, and for teaching me how to have fun while doing the research that matters.

Jaume Nin, for helping me to learn and start speaking Spanish that greatly improved my everyday life, work, and helped me to better understand local people and culture.

Lorenza Giupponi, for helping me when it was the most difficult time of this PhD journey, for believing in me, for giving me the honor of working with her, for transmitting to me her incredible passion and love for telecommunications, research and work in general; for showing me how to be a better researcher and better me, for being not only a great supervisor and an inspiring example to follow, but also a great friend.

Zoraze Ali, for being a great person and a great friend.

All 5G-LENA team: Lorenza, Sandra, Natale, Katerina, and Zoraze, for giving me that amazing feeling of having a great team, and for being great friends.

Silvia Garcés, Marc Majoral and Laia Nadal, for being so great colleagues and friends.

Dr. Giorgio Quer, for giving me the opportunity and privilege to be a visiting researcher and carry out a part of this PhD thesis at the Qualcomm Institute at the University of San Diego, California, and for being a great supervisor and friend.

Professor Thomas R. Henderson for great collaborations and for taking care of the ns-3 network simulator without which this PhD thesis would not be possible.

My tutor professor Miguel Soriano Ibáñez, Networking Engineering PhD Program Coordinator Dr. Xavier Hesselbach, PhD program administrative Aurora Paula Rubio Rodriguez, and Research Director of CTTC Mobile Networks Department Dr. Josep Mangues Bafalluy for their support.

Dr. Ana Maria Galindo Serrano and Dr. Christian Ibars for reviewing this PhD thesis and for their support.

* * *

Посебно бих волела да се захвалим:

Мојој МАМИ, и мојој браћи ГОРАНУ и СРЂАНУ што су увек били највећа могућа подршка и ветар у леђа, што су ме увек безусловно волели, подржавали, веровали у мене и бодрили ме.

Јелени, Ксенији и Драгославу, што сте на најлепши могући начин подржавали вашу тетку.

Мојој дивној тетки Коли, Бошку, тетки Дули, Жики, Шубари и Ружи. Хвала Срђиној Драгани, сестри Драгани и Зорану, брату Слободану и Катарини, брату Зорану доктору, Зорану новинару и Боби, Бошку Бојовићу, и драгим кумовима Ани и Вуку, и Николи и Глигору.

Мојим дивним пријатељима Соњи и Јовану, Тамари и Дарку, Снежи и Бошку, Ивани, Мири, Анамарији и Ивану, Вилми и Стевану, Јелени, Велибору, и кумовима Јасмини и Стеву, што су све ове године били уз мене, веровали у мене и бодрили ме.

Волела бих да споменем и оне који више нису са нама а који су ме испратили на овај дуг и далек пут и подржавали, посебно мог драгог ТАТУ Драгослава Бојовића, а затим и моју милу тетку Бују и драгог брата Нешка.

* * *

Me gustaría agradecer especialmente a:

Mi cariño Joan, por estar siempre allí durante este camino tan repleto de retos y obstáculos. Por apoyarme, creer en mi, darme fuerzas y ánimo cuando más los necesitaba. Sin ti no habría esta tesis.

Mis hijos Joanet e Irina por darme siempre ánimo y ser tan preciosos como sois.

Mi preciosa familia española: Joan Parera Saigi, Mireia, Rafa, Jordi, Lara, el abuelo Pepe, Maria Antonia y Antonio, Jordi, Roger, Sergi, Esmeralda y Pepe, y estupendos amigos de la familia: Prado e Iñaki, Pilar Polo, Angelines, Chema, Jesús, Benaiges y Mercé por darme siempre animo y apoyo.

Montserrat Gallet, Elena y Hector, Vanja y Sergi, Dragana e Ilija, Camelia y Kay, por siempre estar allí y apoyarme.

Me gustaría mencionar a los que ya no están con nosotros, especialmente a Aurora y la abuela Misi.

* * *

Biljana Bojović

Contents

Acronyms

3GPP	3rd Generation Partnership Project
5G	5th Generation
A-MPDU	Aggregate MPDU
ACK	Acknowledgment
AP	Wi-Fi Access Point
B5G	Beyond 5G
BA	Block Acknowledgment
BAR	Block Acknowledgment Request
BN	Bayesian Networks
BS	Base Station
CA	Carrier Aggregation
CAC	Call Admission Control
CAM	Channel Access Manager
CBR	Constant Bit Rate
CBRS	Citizens Broadband Radio Service
CCA	Clear Channel Assessment
CCM	Component Carrier Manager
COT	Channel Occupancy Time
CQA	Channel and QoS Aware
CQI	Channel Quality Indicator
CRS	Cell-Specific Reference Signal
CSAT	Carrier Sense Adaptive Transmission
CSI-RS	Channel State Information - Reference Signal
CSMA/CA	Carrier Sense Multiple Access Collision Avoidance
CW	Contention Window
CWS	Contention Window Size
DCF	Distributed Coordination Function
DCI	Downlink Control Information
DFBA	Dynamic Frequency and Bandwidth Allocation
DL	Downlink
DMTC	Discovery signals Measurement Timing Configuration
DRS	Discovery Reference Signal
eCCA	Extended CCA
ED	Energy Detection

eLAA	Enhanced LAA
eNB	Evolved Node B
EPS	Evolved Packet System
feLAA	Further Enhanced LAA
FTP	File Transfer Protocol
gNB	next-Generation Node B
HARQ	Hybrid Automatic Repeat Request
HOL	Head-of-line
HSPA	High-Speed Packet Access
IP	Internet Protocol
ITU	International Telecommunication Union
KPI	Key Performance Indicator
L2SM	Link-To-System Mapping
LAA	Licensed-Assisted Access
LBT	Listen-Before-Talk
LDS	LTE-U Discovery Signal
LTE	Long Term Evolution
LTE-A	The Long Term Evolution Advanced
LTE-U	LTE Unlicensed
LWA	LTE-WLAN Aggregation
LWIP	LTE-WLAN Radio Level Integration with IPsec Tunnel
MAC	Medium Access Control
MCOT	Maximum Channel Occupancy Time
MCS	Modulation and Coding Scheme
MIB	Master Information Block
MIMO	Multiple Input Multiple Output
mmWave	millimeter-wave
MPDU	MAC Protocol Data Unit
NACK	Negative Acknowledgement
NN	Neural Networks
NR	New Radio
NR-U	NR-based access to unlicensed spectrum
OCB	Occupied Channel Bandwidth
OFDMA	Orthogonal Frequency-Division Multiple Access
PBCH	Physical Broadcast Channel
PCell	Primary Cell
PDCCH	Physical Downlink Control Channel
PDSCH	Physical Downlink Shared Channel
PHY	Physical Layer
PLMN	Public Land Mobile Network
PRACH	Physical Random Access Channel
PRB	Physical Resource Block
PSS	Primary Synchronization Signal
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel

QCI	QoS Class Identifier
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RCLWI	RAN controlled LTE-WLAN Interworking
RLC	LTE Radio Link Control
RLC-AM	RLC Acknowledged Mode
RLC-UM	RLC Unacknowledged Mode
ROI	Return On Investment
RRM	Radio Resource Management
RTT	Round Trip Time
SCell	Secondary Cell
SDL	Supplemental DownLink
SIB1	System Information Block Type 1
SNR	Signal-to-Noise Ratio
SON	Self-Organizing Networks
SRS	Sounding Reference Signal
SS	Synchronization Signal
SSS	Secondary Synchronization Signal
STA	Wi-Fi Station
TBFQ	Token Bank Fair Queue
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TTI	Transmission Time Interval
TxOP	Transmission Opportunity
UDP	User Datagram Protocol
UE	User Equipment
UL	Uplink
ULTE	Unlicensed LTE
VoLTE	Voice over LTE
WLAN	Wireless Local Area Networks

Chapter 1

Introduction

1.1 Motivation

It has passed almost half a century since Motorola made the first handheld mobile cell phone in 1973 (the inventor was Martin Cooper), and it has passed just a little more than two decades since the IEEE released the first 802.11 Wi-Fi standard in 1977. Since their beginnings, cellular and Wireless Local Area Networks (WLAN) technologies follow a trend of an exponential increase in the data rates (approximately predicted by the empirical Cooper's law[1]). Additionally, many digital technological advancements have resulted in an even faster exponential growth in wireless capacity requirements, which further push and drive cellular and WLAN evolution.

To keep up with such rapidly increasing wireless traffic demands, operators need to find the right portfolio of strategies to meet the requirements while ensuring a desirable return on investment (ROI)[2]. The three dimensions through which operators move in search for the best solutions for wireless capacity improvements in terms of ROI are: 1) improving spectral efficiency (e.g., through a better radio resource management and Self-Organizing Networks (SON)s), 2) spectrum aggregation (e.g., through carrier aggregation in both licensed and unlicensed bands, such as LAA, LTE-U, and NR-U technologies), and 3) densification (e.g., dense and ultra-dense networks in 4G and 5G, respectively).

[1]Cooper's law is the prediction made by Martin Cooper that the number of voice calls or equivalent data transactions that can be carried over all the useful radio spectrum doubles every 30 months [1].

[2]Return on investment (ROI) is a metric used to denote how much profit has been generated from an investment that's been made (definition by "Harward Business School Online"). The metric can be either expected (calculated before project kicks off) or actual (calculated after the project is concluded).

All these three dimensions and strategies have found their main representation in the so called small cell networks vision, which should be incorporated in any technology needed to reach extremely high capacity improvements. Small cell technologies should support: i) an extremely high level of spatial reuse, ii) an optimal spectral efficiency and radio resource management of highly and extremely dense networks, and iii) aggregation of spectrum through various spectrum sharing paradigms (i.e., licensed, unlicensed, and shared). Additionally, they should support a wide range of spectrum frequencies, i.e., 5G supports 0.5-100 GHz frequencies which include ultra[3], super[4], and some of the extremely high[5] frequency bands, all of which are characterized by very different propagation conditions.

The resulting small cell networks are highly complex multi-tier, multi-Radio Access Technology (RAT), multi-frequency, and multi-spectrum sharing paradigm systems. To minimize the costs of such complex systems, it is necessary to automate different operations and reduce human interventions. For this reason 3GPP introduced in Release 8 the SON concept, which offers autonomic functionalities such as self-configuration, self-optimization, and self-healing. While in current networks, most of the automation is achieved through expert systems, it is expected that the next paradigm shift will happen when the automation in wireless networks is mostly based on machine learning. Consequently, many standardization bodies are already working on unifying machine learning solutions for wireless networking systems to allow their large-scale worldwide adoption in the near future. In addition, in recent years, there has been a growing interest in unlicensed carrier aggregation for capacity improvements in cellular systems (i.e., LAA, LTE-U, MulteFire technologies, and recent NR-U specification). Using an unlicensed spectrum means to adhere to different regulatory requirements which can significantly impact the spectral efficiency and radio resource management.

This PhD dissertation focuses on designing radio resource management mechanisms that allow the capacity improvements in unlicensed and licensed spectrum, by considering also other aspects of small cell networks capacity improvements, such as densification and spectrum aggregation. We focus on the main technologies that have exploded during the last decade, IEEE Wi-Fi and 3GPP LTE and NR, underlining how, over the years, the two main families of technology standards, IEEE and 3GPP, have passed from being complementary to become increasingly more similar. In addition, we focus on the coexistence characteristics of the two technology families, as they need to be designed for small cells to coexist in the same unlicensed band. To achieve this objective, the PhD thesis investigates the impact of the coexistence parameters (i.e., focusing on cellular system parameters) on the efficient usage of radio resources in multi-RAT coexistence scenarios (LTE/Wi-Fi and NR/WiGig) while coexisting fairly with the other unlicensed technology (e.g., Wi-Fi or WiGig).

Keeping also in mind the sustainability constraint, and the need to introduce automation in the management of future wireless technologies, this PhD thesis also introduces as a main ingredient in the definition of the proposed radio resource management solutions of the different considered technologies, machine learning, as the theoretical tool that allows to introduce self-organization. In this area, this thesis seeks to answer to the following questions:

[3]Ultra high frequency (UHF) is the International Telecommunication Union (ITU) designation for radio frequencies in the range between 300 MHz and 3 GHz, also known as the decimetre band.

[4]Super high frequency (SHF), aka centimetre wave band, is the ITU designation for radio frequencies in the range between 3 and 30 GHz.

[5]Extremely high frequency (EHF) according to ITU, aka millimeter wave frequency, the radio frequency in the range from 30 GHz to 300 GHz.

How to employ machine learning into radio resource management to achieve better spectral efficiency? Can a machine learning scheme be equally good performing in an unlicensed asynchronous system such as a Wi-Fi network and in a licensed system such as an LTE small cell network? How to extend the machine learning concept to a broader context where the optimization would be done at a network level instead of the user-centric approach? What are the machine learning schemes that are suitable for the real-world application and to be used in SONs? What are the main coexistence parameters that impact radio resource management in LTE and NR when coexisting with IEEE Wi-Fi and WiGig technologies? What considerations should be taken into account when modeling and designing a coexistence mechanism for NR-U and WiGig coexistence? Moreover, we seek to answer whether the road ahead leads us toward complementarity or competition of cellular and IEEE technologies in highly densified small cell deployments.

The common denominator of all the SON oriented radio resource management solutions proposed in the tesis for IEEE and 3GPP technologies is that they have been evaluated over the same simulation platform. For this our bet has been on focusing on realistic, high-fidelity, full stack, end-to-end similation models based on the open source ns-3 simulator. Due to my need for availability of the timely models, I could not wait for the open source community to provide them, and so I became over the years, with my team, an active and fundamental member of the open source community, contributing to its main 3GPP models. Hence, one of the main achievements of the present thesis has been the design and development of pioneer, standard-compliant, and open-source ns-3 network simulator modules for simulating the main 3GPP technologies, LTE and NR. The author of this thesis is the principal designer and developer of LTE and NR modules and of their coexistence features (i.e., LAA, LTE-U and NR-U) with IEEE technologies (i.e., Wi-Fi and WiGig).

Specifically, the ns-3 LTE module was developed at CTTC in the context of a long term collaboration with Ubiquisys. Successively, the LAA and LTE-U coexistence modules are developed and evaluated in close collaboration with the industry, i.e., Wi-Fi Alliance and SpiderCloud Wireless, and academy, i.e., the University of Washington (professor Thomas R. Henderson, who was responsible for developing Wi-Fi coexistence features). Additionally, for the NR-U coexistence study, and throughout collaboration first with the industry, i.e., Interdigital, and afterwards with the USA Defense Spectrum Organization (DSO) under the Spectrum Sharing Test and Demonstration (SSTD) Program, it has been developed in CTTC a standard-compliant and open-source NR and NR-U simulator, called 5G-LENA. The author of this thesis is one of the principal designers and developers of this work [2–11], which ends up being the main evaluation platform of this thesis. These simulators, which are also one of the main outputs of this thesis, are not only representing a significant contribution to academy and education in general, but also represent a vital tool that can be used by industry and standardization bodies to develop and evaluate prototypes, patents, and investment strategies through estimating ROI.

The following section outlines the contents of this PhD dissertation. Afterward, we provide a full list of publications that are published as a result of this work. Finally, we provide the list of scientific contributions that are directly related to this thesis but are outside its scope, contributions through a patent, and through ns-3 network simulator modules for LTE and NR, and through their extensions and features (i.e., LAA, LTE-U, NR-U, LTE QoS MAC scheduler and LTE Carrier Aggregation, NR frequency division multiplexing, NR realistic beamforming, 3GPP channel model, etc.).

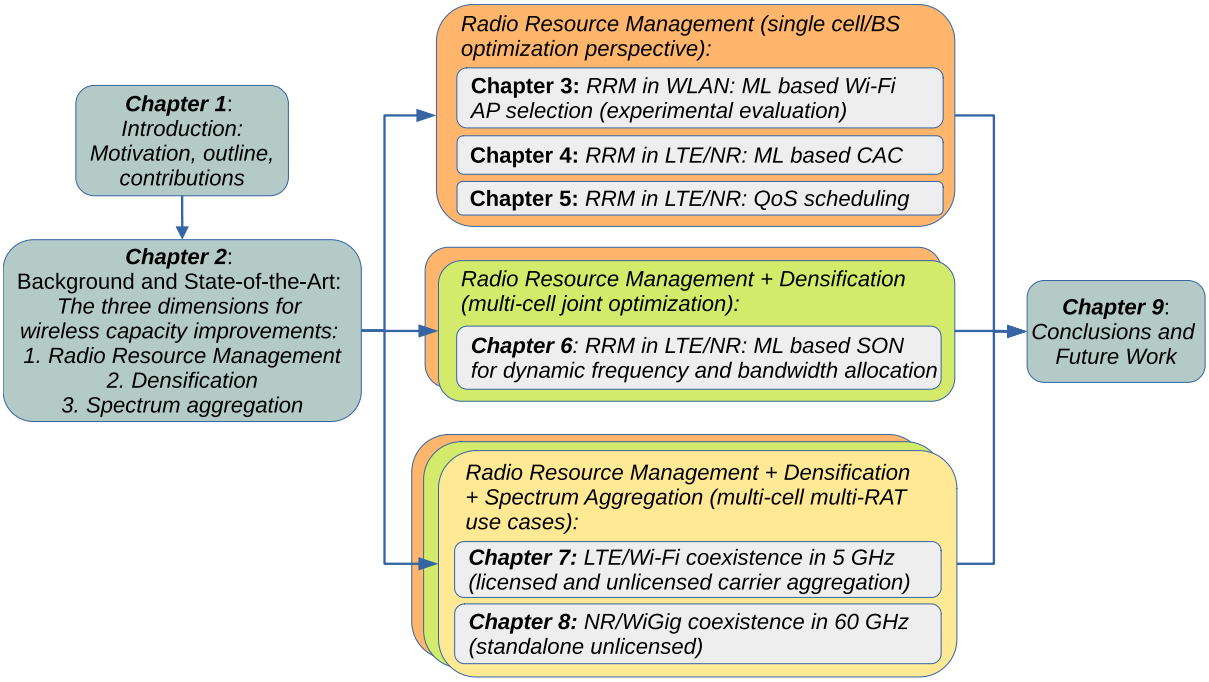


Figure 1.1 Organization of the thesis

1.2 Outline of the dissertation

This PhD thesis focuses on the design and analysis of radio resource management mechanisms for wireless networks in licensed and unlicensed spectrum considering different IEEE and 3GPP technologies in standalone or coexistence settings. It investigates different machine learning and statistical learning-based approaches that can be leveraged for radio resource management and SON paradigm for current and future mobile networks. The work touches different technologies, like IEEE Wi-Fi, 3GPP LTE, 3GPP NR and also the coexistence of these technologies, like LTE and Wi-Fi, through LAA and LTE-U, and NR and Wi-Fi, through NR-U. Additionally, it analyses different NR-U technology implementations. Figure 1.1 illustrates the organization of the PhD dissertation.

Chapter 2 provides the context and background of the present thesis. We first give a brief historical overview of wireless technologies' development to remind the reader what were the principal drivers of the cellular and Wi-Fi technological evolution. Then we analyze the problem of the continuous and exponential increase in capacity demands and explore the three dimensions through which the next capacity targets could be reached for 5th Generation (5G) and Beyond 5G (B5G) cellular networks. We explain how each of these dimensions is related to each part of the thesis. We discuss why these three dimensions lead to more complex and heterogeneous networks that require increasing investments. We explain what the main reasons to introduce SON are and why we believe that it will be a key player also in future mobile networks. We introduce machine learning, describe the main approaches, examples of machine learning for wireless networking, and finally describe its standardization for wireless networks. Finally, we discuss selected performance evaluation methodology and provide the arguments that lead us to choose a network simulation approach and, in particular, the ns-3 network simulator.

In Chapter 3, we study the problem of the access point selection in Wi-Fi networks. A radio resource management scheme for access point selection based on learning is proposed. To add the learning capability to the system, we use Artificial Neural Network (ANN) as the basis of the supervised learning engine. The objective is to investigate whether a machine learning-based scheme can yield better throughput performance than other methods proposed in the literature. We want to study whether the machine learning-based scheme can work well in a real-world environment and learn from real-world measurements, and how will this affect learning performance and the overall access point selection performance.

In Chapter 4 and Chapter 5, we study radio resource management in cellular networks, particularly in LTE. We deal with two radio resource management issues: call admission control in Chapter 4 and MAC scheduling in Chapter 5. In both, the goal is to propose solutions that can guarantee QoS requirements of LTE flows, and in both KPIs are either per user or per cell.

In Chapter 4, the goal is to investigate whether a machine learning-based scheme similar to that presented in Chapter 3 can also be used for the environments in which there are very tight QoS requirements defined by QoS Class Identifier (QCI). Reaching the limitation of the black-box neural network, we study and evaluate whether some other model is able to achieve similar performance, but with more controllable parameters. To this end, we investigate a probabilistic model based on Bayesian Networks and compare its performance and controllability with the Neural Network based model.

In Chapter 5, we investigate, what happens once the call is accepted and what MAC scheduling algorithm is needed to guarantee the QoS requirements specified in QCI. We focus on MAC scheduling, and specifically on a channel and QoS aware scheduling algorithm at the same time. Then we investigate whether the proposed MAC scheduling can enhance the capacity of ongoing Voice over LTE (VoLTE) call, comparing to other state-of-the-art MAC scheduling algorithms available in the literature.

In Chapter 6, we go significantly beyond our previous work. Our goal is to investigate how machine learning lessons learned in the previous chapters can be applied for radio resource management in a more complex scenario in the context of SON and to achieve spectral efficiency optimization on a system level. Hence, instead of considering only the user-centric Key Performance Indicator (KPI)s, we are interested in solutions that would provide an improvement in overall network performance. To achieve this goal, differently from previous works, we consider a centralized machine learning approach which forms an intelligent part of the SON framework. We investigate the problem of dynamic frequency and bandwidth assignment in a dense small cell deployment. We explore various supervised learning approaches and analyze their applicability to the proposed scenario. In this chapter, we would like to determine whether the proposed learning-based performance prediction approach can yield better performance than the baseline approach, whether it is suitable for the application in real-world cellular systems, and under which conditions.

Chapter 7 and Chapter 8 target coexistence of IEEE and 3GPP in two main unlicensed spectrum bands, 5 GHz and 60 GHz, respectively. In Chapter 7 spectrum aggregation is achieved by aggregating a supplemental downlink carrier in the unlicensed spectrum, while in Chapter 8 more spectrum is achieved by using a wide system bandwidth in mmWave band. Differently for LTE-LAA which only can operate as a supplemental carrier anchored to a licensed carrier,

NR-U can fully operate in standalone fashion, and so also this complete operation in unlicensed in considered in Chapter 8.

Chapter 7 considers LTE and Wi-Fi coexistence in the 5 GHz spectrum. It elaborates on parameters impacting radio resource management, and that could be tuned to achieve maximum performance while respecting the regulatory requirements (i.e., the maximum transmission power, the energy detection threshold, listen before talk mechanism, and its parameters). In this study, we also want to carry out an extensive comparative evaluation about the available channel access mechanisms offered by industry, namely LBT and CSAT. It is generally accepted that LBT is a superior access mechanism in terms of interference generation and coexistence with other technologies, but we prove through extensive simulation campaigns, that this claim is not true in all scenarios and all the settings. To achieve these goals, we built a full-stack, open source, specs-compliant LAA, and LTE-U modules for the ns-3 simulator. We elaborate on the research question whether the common conclusions in industry and academia concerning these two technologies are true or should be investigated more so to revisit the selection of the scheme for 5G and B5G networks.

On the other hand, Chapter 8 deals with NR and WiGig coexistence in 60 GHz band, and elaborates on how channel access mechanisms should deal with the beam based access. To carry out this NR and WiGig coexistence study, we built a full-stack, open source, specs-compliant NR and NR-U modules for the ns-3 simulator. We investigate on what lessons learned from LTE and Wi-Fi coexistence hold also for NR-U and WiGig, and on the advantages and new challenges introduced by the directionality of transmission and reception typical of communication in mmWave bands. Finally, we identify open research areas for the years to come.

Chapter 9 concludes the thesis and highights the main contributions of this PhD thesis.

1.3 The list of contributions presented in the thesis

In this section we list the main contributions of the present thesis.

Contributions of **Chapter 3** have been published in:

- **Biljana Bojović**, Nicola Baldo, J. Nin, Paolo Dini, "A Supervised Learning Approach to Cognitive Access Point Selection", IEEE Global Communications Conference (GLOBECOM 2011) Workshop, Houston, Texas, United States, 5-9 December 2011
- **Biljana Bojović**, Nicola Baldo, Paolo Dini, "A Neural Network Based Cognitive Engine for IEEE 802.11 WLAN Access Point Selection", IEEE Consumer Communications & Networking Conference (CCNC), Las Vegas, Nevada, United States, 14-17 January 2012

Contributions of **Chapter 4** have been published in:

- **Biljana Bojović**, Nicola Baldo, Paolo Dini, "A Cognitive Scheme for Radio Admission Control in LTE systems", 3rd International Workshop on Cognitive Information Processing (CIP), Invited Paper, Baiona, Spain, 28-30 May 2012

- **Biljana Bojović**, Nicola Baldo, "A Neural Network based Radio Admission Control Scheme for LTE systems", ICT-ACROPOLIS Network of Excellence Workshop, Brussels (Belgium), 27-28 June 2012
- **Biljana Bojović**, Nicola Baldo, Giorgio Quer, Ramesh R. Rao, "Bayesian and Neural Network Schemes for Call Admission Control in LTE systems", IEEE Global Communications Conference (IEEE Globecom 2013), Atlanta (USA), 9-3 December 2013

Contributions of **Chapter 5** have been published in:

- **Biljana Bojović**, Nicola Baldo, "A new Channel and QoS Aware Scheduler to enhance the capacity of Voice over LTE systems", 11th International Multi-Conference on Systems, Signals & Devices (SSD'14), Castelldefels, Spain, February 2014

Contributions of **Chapter 6** have been published in:

- [**Q2**] **Biljana Bojović**, Elena Meshkova, Nicola Baldo, Janne Riihijärvi, Maria Petrova, "Machine learning-based dynamic frequency and bandwidth allocation in self-organized LTE dense small cell deployments", EURASIP Journal on Wireless Communications and Networking, Vol. 2016, No. 183, August 2016.

Contributions of **Chapter 7** have been published in:

- Lorenza Giupponi, Thomas Henderson, **Biljana Bojović**, Marco Miozzo, "Simulating LTE and Wi-Fi Coexistence in Unlicensed Spectrum with ns-3", *Arxiv* (Submitted on 22 Apr 2016 (v1); last revised 1 Jul 2016 (v2)) https://arxiv.org/abs/1604.06826
- [**Q1**] **Biljana Bojović**, Lorenza Giupponi, Zoraze Ali, Marco Miozzo, "Evaluating Unlicensed LTE technologies: LAA vs LTE-U", *IEEE Access*, Vol. 7, pp. 89714-89751, July 2019.
- **Biljana Bojović**, Melchiorre Danilo Abrignani, Marco Miozzo, Lorenza Giupponi, Nicola Baldo, "Towards LTE-Advanced and LTE-A Pro Network Simulations: Implementing Carrier Aggregation in LTE Module of ns-3", AMC Workshop on ns-3 (*WNS3*), Porto, Portugal, 13-14 June 2017.

Contributions of **Chapter 8** have been published in:

- **Biljana Bojović**, Sandra Lagen, Lorenza Giupponi, "Implementation and Evaluation of Frequency Division Multiplexing of Numerologies for 5G New Radio in ns-3", in Proceedings of the *WNS3* 2018, Surathkal (India) June 13-14, 2018.
- [**Q1**] Natale Patriciello, Sandra Lagen, **Biljana Bojović**, Lorenza Giupponi, "An E2E simulator for 5G NR networks", *Elsevier Simulation Modelling Practice and Theory*, Vol. 96, 101933, Nov. 2019.
- [**Q1**] Natale Patriciello, Sandra Lagen, **Biljana Bojović**, Lorenza Giupponi, "NR-U and IEEE 802.11 Technologies Coexistence in Unlicensed mmWave Spectrum: Models and Evaluation", *IEEE Access*, Vol. 8, pp. 71254-71271, April 2020.

- [**Q1**] Sandra Lagen, Lorenza Giupponi, Sanjay Goyal, Natale Patriciello, **Biljana Bojović**, Alpaslan Demir, Michaela Beluri, "New Radio Beam-based Access to Unlicensed Spectrum: Design Challenges and Solutions", *IEEE Communications Surveys & Tutorials*, Vol. 22, No. 1, pp. 8-37, March 2020.
- **Biljana Bojović**, Sandra Lagen, Lorenza Giupponi, "Realistic Beamforming Design using SRS-based Channel Estimate for ns-3 5G-LENA Module", in Proceedings of the *WNS3* 2021, Virtual Event June, 2021.
- **Biljana Bojović**, Sandra Lagen, Lorenza Giupponi, "The Impact of Traffic Models on NR-U and WiGig Coexistence" (work in progress)

1.4 Other research contributions closely related to the thesis content

Patents

Part of the work that is closely related to the content of **Chapter 8** has been published in the patent application [12]:

- Sanjay Goyal, Arnab Rob, Alpaslan Demir, J.Patrick Tooher, Janet A. Stern-Berkowitz, Moon-il Lee, Lorenza Giupponi, Sandra Lagén Morancho, **Biljana Bojović**, Michaela C. Beluri: "Channel access procedures for directional systems in unlicensed bands"; Filing date: 17.10.2018., WO2019079500 (April 2019), EP3698599 (August 2020).

International conferences and workshops

Parts of the work closely related to the content of **Chapter 7** have been published in:

- Zoraze Ali, Lorenza Giupponi, Josep Mangues, **Biljana Bojović**, "Machine Learning Based Scheme for Contention Window Size Adaptation in LTE-LAA", in Proceedings of 28th Annual *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications* (*IEEE PIMRC* 2017), October 2017, 8-13 October 2017, Montreal
- Zoraze Ali, **Biljana Bojović**, Lorenza Giupponi, Josep Mangues, "On fairness evaluation: LTE-U vs. LAA", in Proceedings of the 14th ACM *International Symposium on Mobility Management and Wireless Access* (*MOBIWAC* 2016), 13-17 November 2016 (Malta).

Parts of the work closely related to the content of **Chapter 8** have been published in:

- Katerina Koutlia, **Biljana Bojović**, Sandra Lagen, Lorenza Giupponi, "Novel Radio Environment Map for the ns-3 NR Simulator", in Proceedings of the *WNS3* 2021, Virtual Event June, 2021.

- Tommaso Zugno, Michele Polese, Natale Patriciello, **Biljana Bojović**, Sandra Lagen, Michele Zorzi, "Implementation of A Spatial Channel Model for ns-3", in Proceedings of *WNS3* 2020, 15 June 2020, Gaithersburg, Maryland (US).

- Natale Patriciello, Sandra Lagen, Lorenza Giupponi, **Biljana Bojović**, "The impact of NR Scheduling Timings on End-to-End Delay for Uplink Traffic", in Proceedings of *IEEE Global Communications Conference* (*IEEE GLOBECOM*), 9-13 December 2019, Waikoloa, HI (USA).

- Natale Patriciello, Sandra Lagen, Lorenza Giupponi, **Biljana Bojović**, "An Improved MAC Layer for the 5G NR ns-3 module", in Proceedings of the *WNS3* 2019, 19-20 June 2019, Firenze (Italy).

- Sandra Lagen, **Biljana Bojović**, Sanjay Goyal, Lorenza Giupponi, Josep Mangues, "Subband Configuration Optimization for Multiplexing of Numerologies in 5G TDD New Radio", In Proceedings of *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.* (*PIMRC*'18), 9-12 September 2018, Bologna (Italy).

- Natale Patriciello, Sandra Lagen, Lorenza Giupponi, **Biljana Bojović**, "5G New Radio Numerologies and their Impact on the End-To-End Latency", in Proceedings of *IEEE International Workshop on Computer-Aided Modeling Analysis and Design of Communication Links and Networks* (*IEEE CAMAD*), 17-19 September 2018, Barcelona (Spain)."

- Sandra Lagen, Lorenza Giupponi, **Biljana Bojović**, Alpaslan Demir, Michaela Beluri, "Paired Listen Before Talk for multi-RAT Coexistence in Unlicensed mmWave Bands", In Proceedings of *IEEE Int. Conf. on Commun.* (*ICC*) - WDN-5G WS, 20-24 May 2018, Kansas City (USA).

Open-source network simulator related contributions

- Developer and maintainer of LTE module of ns-3 network simulator since 2013 and 2015, respectively [13].

- *Principal developer* of LAA and LTE-U modules for ns-3 network simulator. The code is available in [14].

- *One of the principal developers* of NR and NR-U modules of ns-3 network simulator since 2017 [15].

- Mentor and co-mentor of ns-3 Google Summer of Code in 2017 and 2019, respectively.

As being an **LTE ns-3 maintainer and developer**, my contributions to computer networking were recognized by SIGCOMM in 2020 through **SIGCOMM Networking Systems Award** [16] which was awarded to the most active authors of the *ns family of network simulators* (ns-1, ns-2, and ns-3) in order to recognize that our work on a development of a networking system software has a significant impact on the world of computer networking. Additionally, it is recognized **CTTC**, whose Mobile Networks group aka **MONET**, of which I am a member, is one of the most active research groups in ns-3 development worldwide for over a decade.

Chapter 2

Background and State-of-the-art

2.1 Introduction

As introduced in Chapter 1, this PhD thesis focuses on radio resource management of wireless technologies of both 3GPP (LTE and NR) and IEEE (Wi-Fi) standards, and their coexistence in 5 GHz (LTE/Wi-Fi) and 60 GHz (NR/WiGig) bands. Also, since the thesis spans over evolutionary generational jumps from 4G to 5G, and from WiFi to WiGig, it provides an important insight of how, over time, in order to reach ever-increasing capacity requirements, these two different families of technologies have evolved from being complementary to become increasingly more similar. Hence, to provide a broader context, **Section 2.2 provides a historical overview of 3GPP and IEEE technologies**, and concludes by identifying the key technological advancements that made them increasingly similar. **Section 2.3 introduces the three dimensions for improving the wireless system capacity (in terms of bits/s/km^2)**: densification, spectral efficiency, and spectrum aggregation, and explains how each of them is related to different contributions of this PhD thesis. **Section 2.4 introduces SON**, and explains why SON plays a fundamental role to design future mobile networks that are sustainable from the cost and management perspective. In Section 2.4.1 we describe the role of machine learning for SON, and we provide its formal definition. Section 2.4.2 goes through the three main machine learning approaches in the current literature. Section 2.4.3 provides some of the typical applications of machine learning for wireless networking. Section 2.4.4 describes current machine learning standardization efforts for 5G and beyond. **Section 2.5 explains why network simulation is selected as the tool to achieve research objectives of this thesis, and why in particular, ns-3 network simulator**. Finally, **Section 2.6 concludes this chapter**.

2.2 Wireless systems: A brief history review and the road ahead

Wireless systems nowadays represent an integral part of our everyday lives. Not only they define the way we communicate with each other, have an impact on our personal and working habits, but also have redefined entire segments of the economy [17], and are currently paving the way to transform the industry. Wireless technologies were already part of the digital revolution that started in the second half of the 20th century with the inventions in electronics and telecommunications that transformed the industry by initiating Industry 3.0 or the Third Industrial Revolution. There were many advancements in cellular, Wi-Fi, satellite, and sensor networks communications [18, 19]. However, the impact of wireless systems on the industry was limited up to now since they were not fulfilling some of the critical industrial requirements [20–23]. Nowadays, we are witnessing the dawn of the Industry 4.0, or the Fourth Industrial Revolution [24], in which the fifth-generation wireless technologies (5G) play a leading role as the key catalysts for large-scale industrial adoption of technologies such as: the internet of things (IOT), the industrial IOT (IIOT) [25], tactile internet [26], internet of skills (IOS) [23], digital twins [21], fully autonomous vehicles, robotics, augmented reality, artificial intelligence, blockchain [27], etc. Such a disruptive impact is possible since 5G promises a dramatically better quality of service in terms of bandwidth, latency, reliability, and connection density than its predecessor, the fourth-generation wireless technologies (4G). The 5G connectivity is an essential enabler of many novel industrial applications where mobility plays a critical role or in which is not feasible having the wired connectivity, like, for example, in massive sensing. Full digitalization of industry-specific business processes by leveraging 5G technologies creates a vast opportunity for the telecommunications industry to offer not only information and communications technology (ICT) services, but also a new strategic direction through data representation and interpretation services [28], "*anything as a service*" paradigm [29], and through data-driven services in smart manufacturing paradigm [30] such as production data analysis, patterns discovery, factory-wide schedule adjustments, and per link control in the production line [20]. Moreover, it is expected that multi-service 5G transport networks will evolve in terms of transaction models, delivery content, and delivery methods [31]. It sounds that we are really in the middle of a historic breakthrough for the wireless technologies and the industry as a whole. But how did we get here? Who are the main players, and where the road ahead is bringing us? In order to understand the vision of the future and the journey on which wireless communications are bringing us, we should first take a look back and understand the history.

The history of wireless communications starts back in the latter half of the nineteen century with Maxwell's electromagnetic wave theory, Hertz's demonstration of the existence of electromagnetic waves; and with the first demonstrations of transmission of the information over radio waves, first by Tesla and later Marconi [18]. During the Second World War, military applications drove the research on bidirectional mobile communications, which led to new achievements in the theoretical foundations of wireless communications, such as Claude's Shannon groundbreaking work "*A Mathematical Theory of Communications*" in 1948 [32]. We continue wireless technological history review focusing on cellular and Wi-Fi-based systems.

Advances in wireless cellular technologies have come in several generations: 0G, 1G, 2G, 3G, 4G, and 5G. The general concept behind different technology "generations" is that each new

Figure 2.1 Pre-cellular: Jerry Lewis in the 1961 movie, "The Errand Boy" (A copyrighted photo purchased from Alamy Ltd for a limited non-commercial use)

generation offers a significant improvement in performance and capabilities compared to its predecessor. Often, a new "overlay" network belonging to a new generation required a new frequency band.

Zero generation or pre-cellular systems were telephone systems of a wireless type which were based on technologies such as Push to Talk (PTT), Mobile Telephone Service (MTS), Improved Mobile Telephone Service (IMTS), and Advanced Mobile Telephone Service (AMTS) systems. These systems were part of a public switched telephone network. Pre-cellular mobile telephones were usually mounted in cars or trucks (aka car phones). In Image 2.1 is shown a scene from a 1961 movie in which is shown a car phone which was offering low quality and no privacy, but at the time was a popular gadget that was representing a status symbol.

The first hand-held mobile cell phone was produced by Motorola in 1973 [33], while NTT launched the first commercial cellular network in Japan in 1979. This was the beginning of the 1G, which used analog radio technologies and supported voice calls only.

The transition from the analog to digital radio networks in the 1990s marked the beginning of a new second generation of mobile cellular networks. The 2G was based on Global System for Mobile Communications (GSM) standard developed by the European Telecommunications Standards Institute (ETSI). GSM allowed the first roaming services, and it was based on a type of time division multiple access (TDMA) technology. 2G systems allow voice and data services.

The transition from the 2G to 3G did not have an exact turning point because 2G digital networks have evolved significantly throughout their long service life, incorporating a series of new technologies such as General Packet Radio Service (GPRS or aka 2.5G) and Enhanced Data Rates for GSM Evolution (EDGE or aka 2.75G). Because of this, the International Telecommunication Union (ITU) set key performance indicators (KPI) significantly higher for 3G comparing to those of original 2G. According to these, the 3G technologies that are considered truly "revolutionary" are: Wideband Code Division Multiple Access (WCDMA), Time Division-Synchronous Code Division Multiple Access (TD-SCDMA) and Digital Enhanced Cordless Telecommunication (DECT). These technologies were defined in the scope of IMT-

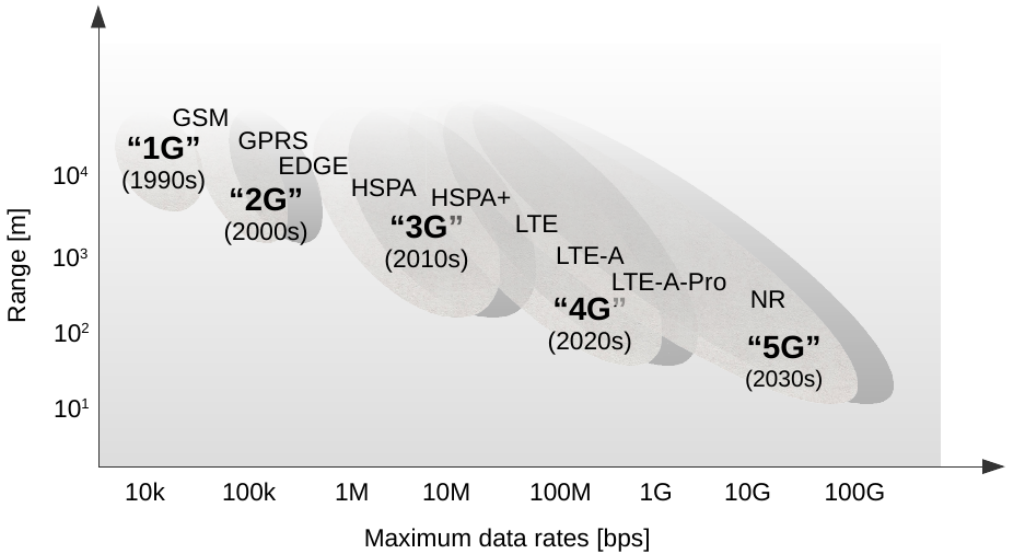


Figure 2.2 Illustration of data rate versus range of cellular wireless generations from 1980 to 2030

2000 standard, which resulted from the collaboration of many entities, inside and outside the ITU, including the 3rd Generation Partnership Project (3GPP). Some of the goals of IMT-2000 were to provide high transmission data rates, improved spectrum efficiency, and to harmonize worldwide 3G systems to provide global roaming. Further evolutions of WCDMA include the High-Speed Packet Access (HSPA) and Evolved High-Speed Packet Access (HSPA+) released in 2008.

In the mid-2010s, when it became clear that 3G networks will be overwhelmed by the need for higher data rates, the requirements for 4G were set, and three competing standard bodies started to work on potential solutions for 4G. The 3GPP standards organization worked on a system called Long Term Evolution (LTE), 3GPP2 on its own standard called the Ultra Mobile Broadband (UMB), and IEEE on a system called WiMAX. Finally, the LTE system developed by 3GPP became the global standard for 4G. The LTE standards are based on OFDMA (Orthogonal Frequency Division Multiple Access) and offer higher throughput, low latency, and improved quality of service. Further enhancements of 4G include LTE-advanced (LTE-A) [34] and LTE-A Pro [35, 36].

In 2016, 3GPP began to work on 5G global standards for a New Radio (NR) access technology [37]. Unlike previous generations, no competing standard bodies are designing potential solutions for 5G. The 5G is envisioned to support a broad range of use cases, including enhanced mobile broadband (eMBB), massive IoT, and mission-critical control. These use cases have diverse requirements, ranging from high capacity and data rate for eMBB, to low latency with high reliability for mission-critical control. To ensure spectrum availability, 5G must operate on diverse spectrum types and bands, including licensed, shared, and unlicensed, and from sub-7GHz to mmWave bands [38].

In Figure 2.2, we illustrate the advancements of cellular technologies and how in order to achieve a higher data rates each generation was adding more capacity by increasing the level of spectrum spatial reuse. We illustrate with the shadow how each generation of cellular

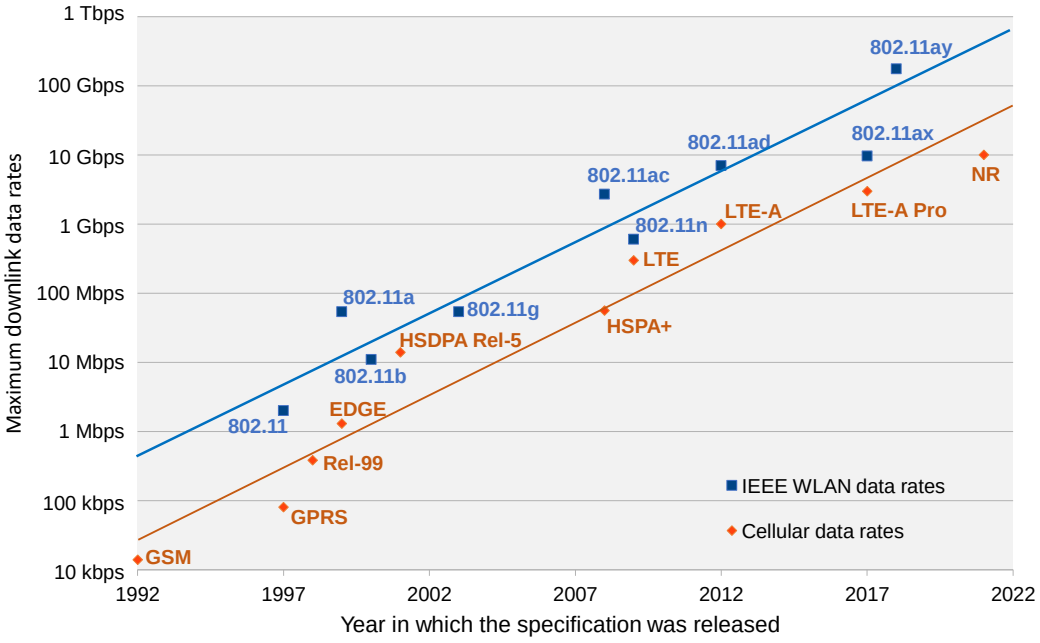


Figure 2.3 Illustration of approximate data rates of cellular (mostly 3GPP) versus IEEE systems from 1992 to 2022

technology has been undergoing a significant evolution during its service life. As noted back in 2007 by authors of [1], while moving to higher frequency bands (assume larger path losses) each new generation has accepted a shorter range in return for a higher data rate. This trend was followed by 4G and has become even more prominent with the 5G NR in mmWave bands.

In parallel with wireless cellular technology evolution, Wi-Fi[1] standards were evolving, even though they started later (in the late 2000s). The first release of the IEEE 802.11 standard was in 1997. This initial standard provided a data rate of up to 2 Mbps in the 2.4 GHz range, which was already significantly higher than the cellular standard at that time, which was still measured in hundreds of kbps. In 1999, the 802.11b standard was released with a top data rate of 11 Mbps in the 2.4 GHz band. This was followed by the release of the 802.11a protocol, which achieves data rates of up to 54 Mbps in the 5 GHz band. The 802.11b protocol was improved with the 802.11g protocol in 2003, which provided a 54 Mbps data rate in the 2.4 GHz band, matching the 802.11a data rate. The 802.11n protocol (aka Wi-Fi 4 [2]), released in 2009, operated in both the 2.4 and 5 GHz bands, achieves data rates up to 600 Mbps, and includes MIMO and channel bonding to achieve these rates. In 2014 802.11ac (aka Wi-Fi 5) was released for the 5 GHz band, with data rates going up to almost 7 Gbps. In 2019 802.11ax (aka Wi-Fi 6) was released for different sub-6 GHz bands reaching maximum link data rates of about 9.6 Gbps. Development of the 802.11be amendment is currently ongoing and its final version is expected in 2024. Some of the main features described in project request are:

[1]Wi-Fi is a term trademarked by the Wi-Fi Alliance for a collection of wireless technologies that meet the IEEE 802.11 wireless standards. The Wi-Fi Alliance is a consortium of companies that certifies 802.11 products.

[2]In 2018, the Wi-Fi alliance standardized consumer-friendly generation numbering so that equipment can indicate which Wi-Fi generation it supports, i.e., Wi-Fi 4 (if the equipment supports 802.11n), Wi-Fi 5 (802.11ac) and Wi-Fi 6 (802.11ax). Also, new 802.11be standard is a candidate for Wi-Fi 7. However, there is no such numbering yet for Wi-Fi in 60 GHz, aka WiGig. For example, WiGig includes 802.11ad and new 802.11ay standards.

Spec.	Aggr. band.	Max. MIMO streams	Max. MU-MIMO streams	Max. Mod.	HARQ	Unlicensed bands[3]	Unlicensed techology
LTE Rel-12	60 MHz	8	4	256-QAM	yes	-	IEEE RAT: RCLWI
LTE Rel-13	80 MHz	8	8	256-QAM	yes	5 GHz	3GPP RAT: LAA IEEE RAT: LWA and LWIP
LTE Rel-14	80 MHz	8	8	256-QAM	yes	5 GHz	3GPP RAT: eLAA IEEE RAT: eLWA, eLWIP
LTE Rel-15	100 MHz	8	8	256-QAM	yes	5 GHz	3GPP RAT: FeLAA
NR-U Rel-17	800 MHz	8	12	1024-QAM	yes	2.4, 3.5, 5, 6, 37, 60 GHz	3GPP RAT: NR-U

Table 2.1 3GPP technologies using unlicensed spectrum.

320 MHz bandwidth, 16 MU-MIMO spatial streams, 4096-QAM, HARQ, OFDMA, multi-band operation (simultaneous use of 2.4 GHz for control, 5 GHz for uplink/control and 6 GHz for downlink/data). It is expected that the maximum MAC throughput will be around 30 Gbps.

In Figure 2.3, we illustrate how cellular and Wi-Fi/WiGig technologies have been evolving over time. Represented data rates are the maximum theoretical achievable data rates as advertised, e.g., in LTE-advanced using the maximum number of carriers and modulation and coding schemes and MIMO. The year represents the year in which the wireless technology standard was released. For example, for IEEE technologies are used dates of the first version of the corresponding standard (denoted with "D1." by the IEEE in [39]). Also for the cellular technologies are used years in which the corresponding specification was released. The data rates used to create this figure were gathered from various references in [40], and always represent the maximum theoretical downlink data rate and not the actual achievable rates.

Spec.	Aggr. band.	Max. MIMO streams	Max. MU-MIMO streams	Max. Mod.	HARQ	Unlicensed bands	Unlicensed techology
Wi-Fi 4	40 MHz	4	no	64-QAM	no	2.4GHz, 5 GHz	802.11n
WiGig (1st)	2.16 GHz	8	no	64-QAM	no	60 GHz	802.11ad
Wi-Fi 5	160 MHz	8	4	256-QAM	no	5 GHz	802.11ac
Wi-Fi 6	160 MHz	8	8	1024-QAM	yes	2.4, 5, 6 GHz	802.11ax (uses OFDMA)
WiGig (2nd)	8.64 MHz	8	8	64-QAM	no	60 GHz	802.11ay
Wi-Fi 7	320 MHz	16	16	4096-QAM	yes	2.4, 5, 6 GHz	802.11be (uses OFDMA)

Table 2.2 IEEE technologies starting from Wi-Fi 4.

Over time Wi-Fi experienced many updates and improvements, which make it today a much more complex and sophisticated family of technologies, which with the new standards currently under discussion, like 11be, includes features which in past were key for cellular technologies, e.g., HARQ, OFDMA, MU-MIMO, scheduling, spatial reuse, etc. For example, while cellular relies on HARQ and OFDMA since the introduction of LTE in mid-late 2010s, Wi-Fi has adopted both of these technologies only recently with 802.11 ax in 2018, hence more than a decade later. On the other hand, Wi-Fi started to use 256 QAM with 802.11 ac around 2012, while LTE adopted it three years later, in Release 12 in 2015. Additionally, cellular technologies, were originally natively designed to operate in scarce licensed spectrum, and so have been always designed to deal with intense spatial reuse and interference conditions.

[3]Unlicensed band column lists only bands used by 3GPP RAT.

Since Release 13, cellular started to operate also in unlicesed spectrum in 5 GHz with LAA technology, and since Release 15, also NR started being designed to natively operate in unlicensed 5 GHz and 60 GHz. This calls for the definition of smart solutions to coexist in unlicensed spectrum, and complex simulators to evaluate multi-RAT scenarios. With the latest amendments and versions (e.g., IEEE 802.11ax and NR-U), both technologies are designed to use large bandwidths in a very efficient manner. Finally, the evolution of 3GPP and IEEE standards is evolving into a direct competition among technologies which were originally designed to occupy complementary spots: one outdoor, the other indoor, one in licensed, the other in unlicesed spectrum.

Table 2.1 and Table 2.2, illustrate, respectively, evolution of 3GPP cellular and IEEE Wi-Fi/WiGig features starting from the point in time when they began to compete in unlicensed spectrum, which makes almost a decade.

2.3 Wireless evolution: Reaching n-fold capacity requirements

Wireless systems requirements are continuously increasing. We could see throughout the wireless history, that shortly after a new milestone is reached, new services become available, which in turn, results in even higher requirements.

They are usually expressed through the target data rates, latency, bandwidth per unit area, number of connected devices, reliability, energy efficiency, mobility. As an example, some of the requirements for 5G [41], compared to those of 4th Generation (4G), include:

- 1000 times more mobile data volume,
- 1000 times more connected devices per unit area,
- 100 times higher user data rate,
- 10 times lower energy consumption,
- 5 times lower end-to-end latency,
- 5 times lower Operating Expenditures (OPEX), and
- 1000 times lower service deployment time.

Additionally are set:

- 99.999% reliability for specific mission-critical services,
- mobility support at speeds higher than 500 km/s, etc.

To meet such ever-increasing requirements, standard bodies, vendors, and operators are continuously working on different solutions for capacity improvements. Figure 2.4 illustrates a three dimensional figure which represents the wireless system capacity in terms of bits/s/km^2: i) *Density* (cells/km^2) ii) *Bandwidth* (Hz), and iii) *Spectral efficiency* (bits/s/Hz/cell) [42]. These three dimensions not only determine directions of future wireless networks technological

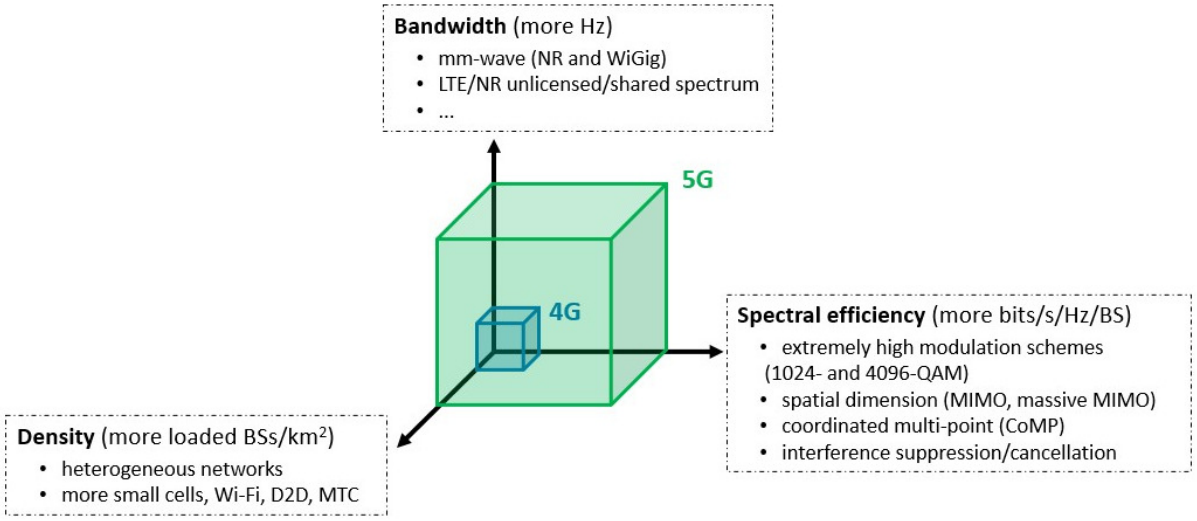


Figure 2.4 Illustration of the three dimensions of the wireless capacity expansion. The illustration is based on the figure from [42], and it is used and slightly updated with the permissions of the author of [42].

advancements, but also define three different strategies that vendors and operators can pursue in order to achieve desired Return On Investment (ROI) while improving system capacity. These strategies include:

- *Network densification to enhance spatial reuse*, e.g., through the planning and deployment of heterogeneous networks and small cells. Most of the wireless capacity gains are achieved through densification [1, 43]. Deployments of small cells result in complex multi-tier, multi-RAT and multi-frequency systems. Network densification adds more complexity to network deployment (e.g., location planning, fronthauling and backhauling) and network management (e.g., radio resource management, mobility management, operation in higher frequency bands, smart idle mode capabilities for energy savings, etc.) which increases both Capital Expenditures (CAPEX) and OPEX. It is expected that hyper-densification [4] will most likely occur in 5G and B5G through the use of mmWave spectrum [45]. Automation will become not just desirable but essential to enable high levels of densification in the 5G era and beyond, to support effective orchestration of many physical and virtual elements.

- *Spectrum aggregation in licensed, unlicensed and shared spectrum*, which typically consists of exploiting higher frequency bands in licensed, shared (e.g., 3.5 GHz) and unlicensed spectrum (5 GHz, recently opened 6 GHz band, and the mmWave bands at 60 GHz band). Some of the cellular 3GPP technologies that are designed to enable spectrum aggregation are: High-Speed Packet Access (HSPA) [46] (only contiguous intra-band), The Long Term Evolution Advanced (LTE-A) [34] (which allows aggregation of contiguous and non-contiguous carriers in both intra- and inter-spectrum licensed bands), LTE Unlicensed (LTE-U) [47] (Based on LTE Releases 10-12, which allow aggregation with unlicensed spectrum bands also in regions where Listen-Before-Talk (LBT) is not a mandatory requirement), LAA [48] (LTE Release 13-15,

[4]The term hyper-dense networks, also called 'ultra-dense networks' in literature, is used to denote the new level of densification that is expected to take place in 5G and beyond [44].

incorporating LBT as a feature to operate in unlicensed spectrum also in regions where this is a mandatory requirement), New Radio (NR) (which is designed to natively support operation in a wide range of frequencies, combining licensed, shared and unlicensed spectrum [49], etc. In the area of IEEE technologies, also Wi-Fi introduces new features for spectrum aggregation, like channel bonding [50] in IEEE 802.11n (Wi-Fi 4), which allows aggregation of channels withing the same band.

- *Adopting technologies that improve spectral efficiency.* In the past, significant improvements in terms of spectral efficiency have been achieved with fundamental techniques such as Turbo codes, low-density parity check codes, Orthogonal Frequency-Division Multiplexing (OFDM) and Multiple Input Multiple Output (MIMO) [43]. Further improvements include enhancements of MIMO and massive MIMO techniques, coordinated multi-point (CoMP) strategies and interference suppression and cancellation techniques, new modulation and coding scheme, etc [42]. Spectral efficiency can be further increased by applying Radio Resource Management (RRM) whose objective is to utilize the limited radio-spectrum resources and radio network infrastructure as efficiently as possible, i.e., maximize the system spectral efficiency while providing a certain level of QoS (e.g., admission control, channel-dependent scheduling, QoS-aware RRM and many others).

In the context of capacity improvements and network densification, it is worth mentioning Cooper's law (aka Law of Spectral Efficiency). According to Cooper's law, the number of voice calls or equivalent data transactions that can be carried over all the useful radio spectrum doubles every 30 months [51]. He predicted that this would continue for the foreseeable future. Focusing on the most recent period, Cooper suggests that the number of voice calls has improved a million times since 1950. In the opinion of Cooper, of that million-times improvement, 15 times was due to using more spectrum, 5 times was thanks to using frequency division (narrower channels), 5 times due to various modulation techniques, and a remarkable, 2700 times improvement was the result of spatial division or spectrum reuse [1] through network densification. While Cooper's prediction is hard to verify [1], an important takeaway is that despite being close to the Shannon limit for a single channel, there is a wast space for wireless capacity increase by investing in more dense networks.

All of these three dimensions are highly relevant for this PhD thesis. For example, in Chapter 3, Chapter 4, Chapter 5 and Chapter 6, the thesis investigates different solutions that could improve the *spectrum efficiency* through various radio resource management functionalities: access point selection, call admission control, MAC scheduling, and inter-cell interference coordination through dynamic bandwidth and frequency configuration. The thesis considers densification through dense scenario use cases in Chapter 6, Chapter 7 and Chapter 8. In particular, in Chapter 7 and Chapter 8, we focus on capacity improvement through spectrum aggregation in the unlicensed spectrum. Here, the purpose is to analyze, respectively, the Wi-Fi/LTE and WiGig/NR-U coexistence problem, which in the end transaltes into a RRM problem in a multi-RAT environment. We investigate which coexistence parameters affect the most the coexistence performance and search for the answer on how to achieve the maximum spectral efficiency while fairly sharing the spectrum and respecting regulations. Hence, in Chapter 7 and Chapter 8, we consider all of these three dimensions together.

In the following sections, we provide an overview of each of these three dimensions for capacity improvements, focusing on and highlighting the sub-set of technologies that are in the scope of

the present thesis.

2.3.1 Network densification: Toward ultra-dense small cell deployments for 5G and beyond

Cellular networks have evolved from the traditional macrocell scenario with high-power macro cell towers providing service to a wide geographic area, to the current ones in which the power is reduced, but network density has significantly increased, as seen in Figure 2.2. Many operators have already been deploying the small cell paradigms for a long time, not only in the form of home small cells, but also in the outdoor deployments to complement macro cell coverage [52].

When Bell Labs introduced the concept of small cells in 1947, they had a radius of about 8 km. In the 1990s, even smaller cell sizes appeared, called picocells (tens of meters to hundreds of meters in diameter). These small cells were just a smaller size of macrocells; hence they needed to be planned (deployment) and used (operation) in the same way as macrocells. These kinds of cells were expensive, and the cost was not always justifying the advantage of its usage. The next shift in the small cells paradigm happened with the introduction of even smaller and low-cost femto cells [53]. The two main advantages of femtocells are their low cost and that they leverage a broadband internet connection as backhaul, which allows better scaling and continuous increment in the density of femtocell deployments per geographic area. In current deployments a small cell radius is typically 500 m or less; a range reduction is expected to increase even more in 5G (tens of meters) [52,54–56]. Most visions agree on that 5G networks will be ultra-dense, populated by a hybrid combination of heterogeneous cells, including different generations 3G, 4G, and 5G, and different types of cells such as macro, pico, and small cells [57,58]. Multi-antenna systems such as massive MIMO and Distributed Antenna Systems (DAS) can also be considered as another densification method [57]. DAS is similar to picocells from a capacity and coverage standpoint, but it refers to another architectural vision, aigned with centralized RAN paradigm, where the base stations centralize the base band processing in a centralized node, or cloud, while the edge of the network is covered by the so called remote radio heads (RRHs) which share cell ID.

Deployments of small cells result in complex multi-tier, multi-RAT and multi-frequency systems. Depending on the type of deployment (planned vs. unplanned), these systems may suffer from inter-cell interference, especially when transmissions of neighboring cells are at the same time/frequency resource. In planned deployments, it was suggested very early [53] to use different bands for adjacent cells. Frequency planning has been a very active area of research. However, the drawback of this method is that it does not only reduce the interference but also the spectrum per cell. To overcome this issue, LTE systems are designed to use full frequency reuse (reuse of 1). Operating in this way causes interference between neighboring cells, especially at cell edges where users will experience a poor Signal to Interference-plus-Noise Ratio (SINR) which, in turn, reduces spectral efficiency. This drove the need for mechanisms to mitigate interference between cells and network layers, such as Inter-Cell Interference Coordination (ICIC) and enhanced Inter-Cell Interference Coordination (eICIC). Additionally, load balancing must be carefully optimized. As authors in [58] discuss, in HetNets, the association based on the signal strength or the signal quality is far from the optimal operation of the network. With the large gap in the transmission power (macrocells: 46 dBm, picocells:

33 dBm, and femtocells: 20 dBm), offload to smaller cell might not happen since the user can find the signal received from a macrocell stronger than that of a small cell and thus associate to the macrocell [59]. Since higher frequencies, in general, are characterized by a higher pathloss, and hence smaller coverage area, small cells will be a candidate to operate in higher frequency bands (e.g., in the 3.5 GHz band, 5 GHz, 6 GHz, and above 6 GHz, such as 60 GHz band in which large amounts of spectrum are available). An extreme network densification as a combination of spatial densification and spectral aggregation of higher frequency bands, as mmWave, is one of the major paradigm shifts in the 5G technology ecosystem [55]. We will discuss the operation in unlicensed bands in more detail in Section 2.3.2.3.

Research on small cells started to increase significantly from 2007 [60], focusing on the femtocells. In the same year, the Femto Forum was formed for promoting the wide-scale adoption of small cells, developing common approaches, standards, and practices. Femto Forum changed its name in 2012 to Small Cell Forum (SCF)[5] to reflect that its work embraces residential, enterprise, metro, and rural small cells, and not only femto cells. SCF is working on solutions that include small cell/Wi-Fi integration, SON evolution and automation, virtualization of the small cell layer, driving mass adoption via a multi-operator neutral host, ensuring a common approach to service Application Programming Interface (API)s to drive commercialization and the integration of small cells into 5G standards evolution [62]. SCF and Next Generation Mobile Networks (NGMN) study several implementation challenges related to the deployment of small cells [56]: (i) reliability and interoperability in a multi-vendor, multi-operator environment, (ii) complexity in parametrization, optimization, and algorithm development, (iii) network performance variations based use case scenarios, and (iv) the added complexity of network topologies.

2.3.2 Spectrum aggregation: Opportunities in licensed, unlicensed and shared spectrum

Concerning spectrum aggregation, operators have currently three options: 1) aggregate more licensed spectrum (e.g., HSPA, LTE-A technologies) 2) offload traffic to unlicensed spectrum (via Wi-Fi, Wi-Fi/LTE aggregation LTE in unlicensed, NR in unlicensed, etc.), or 3) aggregate shared licensed spectrum such as, e.g., Citizens Broadband Radio Service (CBRS) in 3.5 GHz band.

2.3.2.1 Licensed spectrum

Licensed spectrum is not an easy option for spectrum aggregation for the two reasons: i) the low-band (400-1GHz) and the medium-low (1GHz-3GHz) band frequencies that are of the most interest because of good propagation conditions are very scarce, and even when some additional portion of the spectrum becomes available for cellular use, its bandwidth is relatively

[5]SCF provides a definition of a small cell [61]: "A small cell is a radio access point with low radio frequency power output, footprint and range. It is operator-controlled, and can be deployed indoors or outdoors, and in licensed, shared or unlicensed spectrum. Small cells complement the macro network to improve coverage, add targeted capacity, and support new services and user experiences. There are various types of small cell, with varying range, power level and form factor, according to use case. The smallest units are for indoor residential use; the largest are urban or rural outdoor picocells."

small for nowadays data rate demands, and ii) it represents substantial CAPEX for operators. Cellular technologies that allowed aggregation of multiple carriers in the licensed spectrum are HSPA [46] and LTE-A [34]. LTE-A carrier aggregation is one of the key technologies that enabled 1 Gbps dowlink cellular internet, aka Gigabit LTE. An example is Qualcomm's *Snapdragon X16 LTE Modem*, which is the first in the serie of Snapdragon chips to reach 1 Gbps in downlink by aggregating 4x20 MHz while using 4x4 MIMO and 256 QAM. The latest *Snapdragon X24 LTE Modem* uses 7x20 MHz and reaches 2 Gbps in the downlink.

However, due to an extremely high cost and scarcity of dedicated licensed spectrum bands, there is more and more interest in the unlicensed and shared spectrum in recent years.

2.3.2.2 Unlicensed spectrum

Unlicensed spectrum is traditionally used by Wi-Fi technologies, at 2.4 GHz and 5 GHz bands, and since recently, 60 GHz bands by WiGig. In current networks, a significant part of cellular traffic is on-the-spot offloaded through Wi-Fi and WiGig whenever possible as a typical feature already offered through smartphones [63, 64]. According to a recent Cisco study, currently, around 74% of mobile phone traffic runs over Wi-Fi, and by 2023 it is estimated that it will increase to 79% [65]. Wi-Fi standard that enables seamless roaming among Wi-Fi networks and between Wi-Fi and cellular is called HotSpot 2.0 [66], which was developed by Wi-Fi Alliance (WFA) to improve access of mobile devices to Wi-Fi with enhanced authentication and network selection. Hotspot 2.0 uses real-time load metrics to provide QoS and allows operators to improve traffic management in Wi-Fi networks. On-the-spot offloading over Wi-Fi suffers security and management issues since there is no integration between cellular and Wi-Fi RANs.

On the other hand, vendors and operators searching for strategies to increase the cellular system capacity became very interested in recent years in the usage of unlicensed spectrum, especially in the 5 GHz band, recently opened 6 GHz, and mmWave bands, such as the 60 GHz band. This has resulted in various technologies, most of which are specified by 3GPP, LTE-U Forum and MulteFire Alliance. There are two types of these technologies: i) technologies that rely on Wi-Fi RAT to access unlicensed spectrum, e.g., RAN controlled LTE-WLAN Interworking (RCLWI), LTE-WLAN Aggregation (LWA), and LTE-WLAN Radio Level Integration with IPsec Tunnel (LWIP), ii) technologies that use some unlicensed variant of cellular RAT, e.g., LAA or LTE-U, or NR-U. A recent and innovative beyond 5G use case considers unlicensed spectrum for cellular in private mobile networks for industry. An example of such technology is LTE-based technology called MulteFire[6] [67]. Since this use case is outside the scope of the present thesis we will not go into more details about this technology. In the following, we discuss different technologies that allow cellular networks based on LTE and NR standards to make use of unlicensed spectrum to increase capacity, including both Wi-Fi RAT-based and cellular RAT-based technologies.

[6]MulteFire is an LTE-based technology, specified by MulteFire Alliance, that operates standalone in unlicensed and shared spectrum including 5 GHz band. It is based on 3GPP Release 13 and 14. MulteFire technology supports LBT for fair coexistence with other technologies operating in the same spectrum, e.g., Wi-Fi in 5 GHz band.

2.3.2.3 Technologies for cellular usage of unlicensed spectrum

Technologies that allow cellular network to make use of unlicensed spectrum in order to increase capacity include:

- I-WLAN (Release 6 and 8 [68, 69])
- RAN controlled LTE-WLAN Interworking (RCLWI) (Releases 12 and 13 [34, 35])
- LTE Wi-Fi aggregation (LWA) (downlink and uplink, in Releases 13 and 14, respectively, [35, 36])
- LTE-WLAN radio level integration with IPsec tunnel (LWIP) (Release 13 [35])
- License assisted access (LAA) (Release 13 and 14 [35, 36]) and LTE unlicensed (LTE-U) [70]
- New Radio Unlicensed (NR-U) (NR Release 16 [49])
- Multipath TCP (MPTCP) (Release 16 [71])

2.3.2.3.1 I-WLAN An initial 3GPP-WLAN interworking standardization by 3GPP started with the definition of I-WLAN in Release 6 [68]. According to 3GPP specs, a simultaneous connection to an I-WLAN and the 3GPP systems should be possible. Hence, Wi-Fi has been used for mobile data offloading even before introducing LTE in 3GPP Release 8 [69]. In Release 8, 3GPP defined the Access Network Discovery and Selection Function (ANDSF) [72] framework to provide support for network selection (e.g., between Wi-Fi and 3GPP networks), traffic routing, different WLAN deployments (trusted and non-trusted WLAN), etc. The next major progress in LTE and Wi-Fi inter-working was made in Release 12, with the specification of Radio Access Network (RAN)-level inter-working, according to which Wi-Fi offload performance is improved by exploiting real-time channel and traffic information available at the RAN level [34].

2.3.2.3.2 RCLWI RCLWI started to be specified in Release 12 as *RAN assisted WLAN interworking* [34], to be later on enhanced to RCLWI in Release 13 [35]. In RCLWI, a traffic steering decision is added to the 3GPP handover concept [73]. RCLWI could be considered as a bearer handover. User plane bearers are routed through a core network through a Wi-Fi legacy link, instead of through the LTE eNB. The point of aggregation is at the core network [73].

2.3.2.3.3 LWA LWA is an evolution of dual connectivity[7], where the secondary link is provided by the Wi-Fi Access Point (AP). In LWA, the integration of LTE and Wi-Fi occurs at the PDCP layer [8]. The eNB schedules packets to be served on each link and the UE provides feedback about data delivery over WLAN. Optionally, the Wi-Fi access point can provide

[7] Dual connectivity allows terminals to connect to two cells simultaneously and was introduced by 3GPP in Release 12 as one of the key features to achieve desired capacity gains.

[8] PDCP aggregation is being considered as a baseline for aggregation with LTE in 5G.

feedback about the successful delivery of user data over Wi-Fi. The benefit of such tight integration of the two RATs is that resource allocation can be dynamically balanced, based on the latest channel and loading conditions on each RAT. LWA uses the Dual Connectivity (DC) feature [34], which handles the reordering of packets if they arrive out of sequence at the UE through the two links.

2.3.2.3.4 LWIP LWIP is similar to LWA in terms of the high-level concept, which is to aggregate resources from LTE and WLAN for a UE in a connected mode. While LWA requires an upgrade of the WLAN network, LWIP can be applied to legacy WLAN deployments. The LTE eNB manages Wi-Fi link, but instead of LWA-like flow, an IPsec tunnel is established between UE and eNB. The splitting of the bearer is not possible as the aggregation is done at the IP level. As LWA, LWIP integrates LTE and Wi-Fi at RAN.

2.3.2.3.5 LAA/eLAA/feLAA and LTE-U Licensed-Assisted Access (LAA) [35,36,74,75] and LTE-U [70] technologies leverage the Carrier Aggregation (CA) feature (designed originally to be used in licensed bands) to aggregate license-exempt bands. LAA was defined in Release 13, and further enhanced in Release 14, when the uplink operation was also defined, in the context of Enhanced LAA (eLAA) [74]. Additional features were defined in the scope of Further Enhanced LAA (feLAA) Study Item [75]. LAA technologies (LAA/eLAA/FeLAA) use LBT channel access mechanism. An LBT mechanism for Clear Channel Assessment (CCA) before accessing the 5 GHz unlicensed channel is required in some markets such as Europe and Japan. In others, such as the USA, China, India and Korea, there is no such requirement. For markets that do not require LBT, the industrial consortium LTE-U Forum specified a proprietary solution for unlicensed LTE based on Release 12, referred to as LTE-U [70]. LTE-U channel access is based on duty cycling. Thanks to the LBT mechanism, LAA technology is viable for the worldwide market, while LTE-U is only viable in zones that do not require LBT. Both technologies LAA and LTE-U use the LTE air interface to access to unlicensed spectrum. The main control channel and basic voice or data calls remain on a lower licensed LTE spectrum, while during peak traffic periods, the unlicensed spectrum is used as a supplemental data channel. In this thesis, among LTE-based unlicensed technologies, our main focus is on LAA and LTE-U since they represent the most promising technologies for LTE access to unlicensed spectrum. This is because they allow using the same RAN in both licensed and unlicensed spectrum, which allows unified mobility, authentication, security, and management. Additionally, since they leverage CA with the licensed carrier, they also guarantee wide-area coverage and the Quality of Service (QoS) typical of the licensed carrier.

2.3.2.3.6 NR-U Differently from previous experiences with LTE, which was first designed and then adapted to work in unlicensed, NR is designed to include native support for unlicensed spectrum and it includes since Release 16 support for operation in 5 and 6 GHz through the so called NR-U feature. Operation in mmWave 60 GHz is currently under study in Release 17. The design of NR-U started in a study item of NR Release 16 in 2018 focusing on sub 7 GHz bands [49], and it is currently being developed as one of the NR Release 16 work items. This inclusion in standards of NR-U standalone operation in unlicensed bands, is a new major milestone for cellular technologies. Differently from LAA and LTE-U that are based on carrier aggregation using the unlicensed 5 GHz band, and from MulteFire that uses standalone

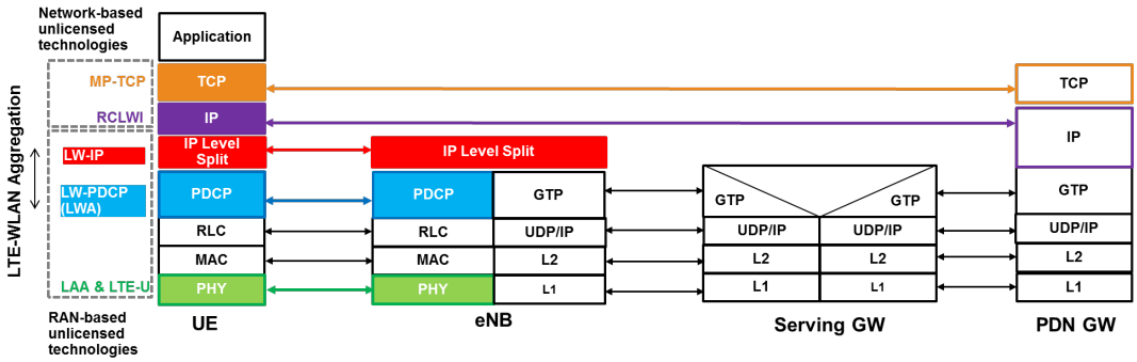


Figure 2.5 Illustration of the aggregation point of different unlicensed cellular technologies (from [76])

operation in the 5 GHz band so far, NR-U is being designed to work in combination with dual connectivity or carrier aggregation, and also in standalone mode. Different bands have been discussed for NR-U operation, including 2.4 GHz, 5 GHz, 6 GHz, and 60 GHz unlicensed bands as well as 3.5 GHz and 37 GHz bands, which are devoted to shared access in the USA.

The author of the thesis has been working on NR-U, LAA, LTE-U during many years, in collaboration with Wi-Fi Alliance, SpiderCloud Wireless and Interdigital.

2.3.2.3.7 MPTCP Multi-path TCP (MPTCP) is an extension of TCP that enables multiple paths for data transmission. Thanks to this feature, MPTCP can automatically use multiple wireless links of a device (e.g., Wi-Fi and LTE, LTE and 5G-NR, etc). At first, a single network connection is established to a destination host. Afterwards, a source host opens additional sub-flows to the destination host. Each sub-flow is treated as a standard TCP connection when transported through the network. MPTCP is now an integral part of 5G mobile networks as a standard feature of 3GPP Release 16 [71] and represents a foundational capability for some of the 3GPP 5G mobile core Access Traffic Steering, Switching and Splitting (ATSSS) features.

Finally, in Figure 2.5, we illustrate where aggregation is done at different layers of protocol stack for different LTE based technologies discussed in this section [9].

2.3.2.4 Shared spectrum

As wireless technologies evolve, the dichotomy licensed vs. unlicensed is not sufficient anymore, and the spectrum policy needs to evolve to converge towards a more efficient use of resources. It is expected that efficient use of spectrum in 5G and beyond networks will rely on spectrum sharing rather than exclusive licenses to reduce congestion in licensed bands and increase capacity [77, 78]. In this line, 3GPP Release 16 [37], includes NR operation in the shared spectrum.

CBRS in the United States in 3.5 GHz band aims to support three tiers (incumbents,

[9] Note that due to the limited transmit power in unlicensed spectrum, these technologies are expected to be implemented in small cells.

prioritized access license holders, and general authorized access) using dynamic sharing, where incumbents (e.g., radars, satellite, wireless ISP) are the top tier and have the highest priority.

In March 2014, the FCC [10], in collaboration with National Telecommunications and Information Administration (NTIA), developed rules for Advanced Wireless Service (AWS-3) auction and reallocation of the 1695-1710 MHz and 1755-1780 MHz radio frequency bands, initially dedicated only to federal use. Most Federal systems using the 1755-1780 MHz band, including those of the United States Department of Defense, will relocate out of the band; however, the FCC's rules provide for indefinite spectrum sharing with selected Federal systems. Another example of usage of the 3.5 GHz are the LTE systems commercially deployed by UK broadband in central London, densely populated urban areas [77].

In 2015, the FCC established the CBRS for shared wireless broadband use of the 3550-3700 MHz (3.5 GHz) band. Then FCC proposed Spectrum Access Systems (SAS) to dynamically manage the CBRS. The SAS are advanced, highly automated radio spectrum coordinators tasked with protecting priority users in the band while optimizing efficient use of available spectrum for everyone else. The SAS achieve this by keeping centralized lists of all the devices that want to use the band, including their type and geographic location. The SAS achieve this information to assign specific frequencies, manage power levels, and prevent interference between devices and services. The result is on-demand sharing in the 3.5 GHz band. This proposal was a major milestone in spectrum sharing and management.

The author of thesis is currently working on spectrum sharing approaches for AWS-3 bands, with the Lawrence Livermore National Lab, in the context of a project funded by the United States Department of Defense.

Other examples of spectrum sharing are Licensed Shared Access (LSA) and Concurrent Shared Access (CSA) [78]. LSA architecture is composed of two tiers: incumbents and secondary users (e.g., mobile operators). On the other hand, CSA allows one class of user to share the spectrum in a coordinated way, e.g., between operators to improve data rates. CSA is currently used in 10 countries. Policies for spectrum sharing are still under development by regulators in many countries.

2.3.3 Spectral efficiency: Toward self-organized radio resource management

Radio resource management is responsible for efficient utilization of the air interface resources in the Radio Access Network (RAN) [36]. Radio resource management concerns multi-user and multi-cell network capacity issues, rather than single-user single-cell channel capacity. Radio resource management is particularly important in systems limited by co-channel interference, such as cellular or WLAN systems, in which many adjacent access points can reuse the same channel frequencies. While classical radio resource management primarily considers the time and frequency domain, recent advances in multi-user Radio resource management strategies should also guarantee a certain a priory agreed QoS level to maintain the planned coverage area while offering high capacity. According to the 3GPP definition

[10] The Federal Communications Commission (FCC) is an independent agency of the United States government that regulates communications by radio, television, wire, satellite, and cable across the United States.

[36], RRM functions include radio bearer control, radio admission control, dynamic resource allocation (aka packet scheduling or MAC scheduling), connection mobility control, and inter-cell interference coordination (ICIC), load balancing between cells, inter-RAT radio resource management (e.g., inter-RAT handover), subscriber profile ID for RAT, inter-eNB CoMP (i.e., coordinate multiple eNBs to improve the coverage of both, high data rate and cell-edge users), and cell on/off and cell discovery. With recent advances in unlicensed cellular technologies, there is an increased interest in academy and industry for radio resource mechanisms in the unlicensed spectrum, e.g., some of the radio resource management topics to be revisited for unlicensed spectrum are listed in [79]. Many future cellular systems are expected to use unlicensed spectrum, and thus should dispose of appropriate radio resource management mechanisms. Such cellular systems include mobile private networks, device-to-device in unlicensed (D2D-U), vehicle-to-everything (V2X) communications (could also be extended to use unlicensed spectrum), unmanned aerial vehicle (UAV) (could also use unlicensed spectrum in public safety scenarios).

The present PhD thesis focuses on various radio resource management studies for Wi-Fi, LTE, NR, and LTE/NR unlicensed. In particular, we focus on: i) radio admission control in unlicensed spectrum (Wi-Fi use case in Chapter 3), and licensed spectrum (LTE use case in Chapter 4), ii) dynamic resource allocation for cellular systems (Chapter 5), iii) inter-cell interference coordination and mobility robustness optimization for cellular systems (Chapter 6). Finally, in Chapter 7 and Chapter 8, we focus on: iv) the analysis of different channel access mechanisms for unlicensed spectrum, and their parameters, and analyze the impact of these parameters on overall system performance and radio resource management for unlicensed spectrum.

High densification and new frequency bands lead to more flexible RAN architectures [80] that require more sofisticated solutions for RRM. In this context, SON paradigm is expected to be one of the key pilars of 5G and B5G networks that will be capable of end-to-end RRM optimization. To achieve their goals, SONs technologies often leverage machine learning. In the following sections we provide an introduction to SON along with a brief overview of the machine learning for SON, and we highlight different SON RRM solutions proposed in the present thesis.

2.4 Self-organizing networks: Minimizing OPEX/CAPEX of 4G, 5G and beyond

In Section 2.3, we described how densification leads to highly complex network deployments, which increase both CAPEX and OPEX. Additionally, operators face the problem that the data volumes have grown exponentially during the last decade, while the average revenue per user has remained almost flat [81]. Also, some countries are much slower in adopting new infrastructures [82]. Consequently, the vast majority of cellular networks' coverage and capacity are still supplied by traditional macrocells [44]. The operators seek solutions to get the most out of the existing and new network assets while minimizing costs (planning, installation, power, backhaul) and human intervention.

For these reasons, SONs are introduced. SONs can automate processes that have been done

to a great extent manually in the past. Hence, SONs can significantly reduce both CAPEX and OPEX. As seen by many, SONs will be a pivotal element to deal with the tremendous complexity and the stringent QoS requirements associated with future 5G networks and beyond [60, 83–87].

SON use cases were introduced by NGMN alliance in 2007 [88], and shortly after 3GPP started standardization activities of SON for LTE in Release 8 [89]. SON use cases (initially defined for LTE) include the following: i) BS configuration, ii) automatic neighbor relation, iii) physical cell identity planning, iv) mobility robustness optimization, v) mobility load balancing, vi) inter-cell interference coordination (ICIC), v) coverage and capacity optimization, vi) random access channel optimization (RACH), vii) energy savings, viii) cell outage detection and compensation, and ix) Minimization of Drive-Tests (MDT). SON for NR is defined in 3GPP TS 38.300 Release 15 [90] and it includes common interfaces, signaling, and measurements. SON has evolved to support new functions specific to NR architecture [91], i.e., 3GPP TS 38.300 includes definitions for the UE support for SON, self-configuration by the dynamic configuration of the NG-C interface, dynamic configuration of the Xn interface, automatic neighbor cell relation (NCR) function, etc.

While the implementations of SON functionalities are vendor-specific, many network elements need to communicate and exchange information [60]. To facilitate multi-vendor deployments of SON-capable small cells, SCF has published in 2015 SON API [92], a standardized interface between a distributed SON (dSON) and a centralized SON (cSON) function, which allows hybrid SON solutions. This multi-vendor compatibility is important, because SON is evolving to become one of the main parts of a broader management and orchestration (MANO) platform [44] and Cloud RAN[11] architecture [93].

SON offers autonomic functionalities such as **self-configuration**, **self-optimization**, and **self-healing**. **Self-configuration** is essential during deployment, extension, upgrade, change, and failure of any network part. Self-configuration replaces the conventional manual configuration process [84]. For example, the number of configurable parameters in a typical 2G, 3G, and 4G node is respectively 500, 1000, and 1500. If this trend continues, a typical 5G node is expected to have 2000, or more parameters [94]. For this reason self-configuration in 5G and B5G is even more needed than in previous generations of mobile networks. During operation, networks need to continuously self-optimize. **Self-optimization** is usually done by leveraging periodic drive tests and log reports. Drive tests are conducted to obtain information about the coverage. They can be also exploited to support SON functions. For example, it was proposed by 3GPP [95] to use this information to improve QoE for UEs. On the other hand, SON solutions can be used for MDT to reduce costs. This can be achieved through SON solutions based on a big data-empowered set of functionalities and machine learning. Finally, **self-healing** assumes remote detection, diagnosis, and recovery actions to ”heal” or fix any network fault.

The present thesis seeks to provide radio resource management algorithms and solutions for self-configuration and self-optimization SON functionalities. Chapters 3, 4 deal with the mobility load balancing optimization use cases. Chapter 5 contributes with RRM solutions to support the mobility load balancing optimization. Chapter 6 deals with inter-cell interfecence coordination, BS configuration and minimization of drive tests SON use cases. On the other

[11]Cloud RAN architecture exploits a combination of virtualization (of software), centralization (of hardware) and coordination techniques (between cells and bands) in order to reduce CAPEX and OPEX [57].

hand, Chapter 7 and Chapter 8 deal with configuration parameters in unlicensed spectrum. This use case is not yet included to NR SON, however, according to [96], it should be included in order to support complex emerging multi-frequncy multi-RAT network scenarios (LTE/Wi-Fi, NR/Wi-Fi, NR/WiGig).

2.4.1 Machine learning for wireless networking

Many SON functionalities are based on intelligent algorithms. Artificial intelligence helps to automate regular network management engineering tasks: planning (e.g., site location planning), configuration (e.g., automating static configuration), operation (e.g., optimizing spectrum usage during operation through the dynamic spectrum and RRM techniques), and maintenance (e.g., identifying and solving issues in networks during their operation). Intelligent algorithms leverage various fields such as machine learning, optimization theory, game theory, control theory, and meta-heuristics. Among them, machine learning is one of the most important sub-fields [94].

Machine learning-based systems can be trained to operate without having a-priori encoded expert knowledge, since they can learn from past experience and adapt to the changing environment without the need for human intervention. Additionally, thanks to very advanced machine learning algorithms, their performance has potential in many use cases to exceed the performance of rule-based expert systems that are based on hard-coded static rules.

Before entering into more details about machine learning for wireless networking, it is noteworthy to provide a more general description of "machine learning" and a more precise definition of "learning". In London in 1947, Alan Turing gave, what was most probably the first known so far, public lecture to mention computer intelligence in which he mentions machine learning by saying: "*What we want is a machine that can **learn from experience** ... possibility of letting the machine alter its own instructions provides the mechanism for this*" [97]. A widely cited formal definition of "learning" is formulated by Tom M. Mitchell [98] in 1997: "*A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.*"

Hence, in contrast to hard-programmed algorithms whose ability to dynamically adapt to continually fluctuating conditions, demands, and system state is limited (if any), **machine learning algorithm** should have the capability to learn continuously, self-optimize, and adapt to changing environment, demands, and states. Machine learning algorithms use computational methods to "learn" information directly from data without relying on a predetermined equation as a model. Machine learning is closely related to statistical learning [99] since machine learning in its simpler form can make use of different algorithms that were already used in statistics to learn from data (e.g., regression algorithms used in statistics), while for modeling more complex non-linear problems algorithms such as neural networks, or reinforcement learning methods are often used. Machine learning is also closely related to data mining[12] since data mining often leverages machine learning models.

In the following, we describe different machine learning approaches. The provided description

[12]Data mining is an interdisciplinary sub-field of computer science whose objective is to extract information using intelligent methods from a data set leveraging on statistics and machine learning.

is far from exhaustive because its purpose is to provide a context of the machine learning background relevant to the present thesis's scope. For a more comprehensive insight on machine learning and related concepts, the interested reader is referred to literature, such as [98–104].

2.4.2 Types of machine learning approaches

Machine learning can be classified based on the inputs that drive learning, the model representing the knowledge, and how the learned knowledge is utilized. Machine learning approaches are commonly classified into three groups: i) **supervised**, ii) **unsupervised**, and iii) **reinforcement learning** [103].

2.4.2.1 Supervised learning

Supervised learning is based on the concept that the system disposes of a set of labeled training data, and the objective of learning is to generate a function that can generalize and predict values on some unseen set of data. In supervised learning, the training data is organized into an input vector ($\mathbf{x}$) and a desired output value (y). The goal is to develop a predictive model by inferring a function $f(\mathbf{x})$, that returns the predicted output $\hat{y}$ [96]. We use the hat symbol to denote an estimate. Training data contains training samples for a specific set of features and is usually divided into two sets. The training set, which is used to train the model, and the test set, which is used to validate the prediction model and evaluate how well the learning model performs the training set samples. The goal of the training is to minimize the error between the predictions and the actual values. The input space is represented by a D-dimensional input vector $\mathbf{x} = (x^{(1)}, \ldots, x^{(n)})^D \in \mathbf{R}^D$. Each dimension is an input variable. In addition, a training set involves m training samples $((\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m))$. Each sample consists of an input vector $\mathbf{x}_i$, and a corresponding output y_i. Hence $x_i^{(j)}$ is the value of the input variable $x^{(j)}$ in training sample i, and the error is usually computed via $|\hat{y}_i - y_i|$ [96]. Each training input $\mathbf{x}_i$ is D-dimensional vector of values representing **features**, **attributes** or **covariates** [103]. In general, $\mathbf{x}_i$ could be a complex structured object, such as an image.

The output $\hat{y}$ can be a *qualitative*[13] or *quantitative*[14] variable. If $\hat{y}$ is qualitative or discrete quantitative, i.e., there is a finite set of output values, the problem is known as **classification**, while when $\hat{y}$ is continuous quantitative, i.e., real-valued variable, the problem is known as **regression**.

Classification is probably the most widely used form of machine learning. In wireless networking, an example of a classification task could be to discover a cause of the low throughput of wireless LAN users: a load of the access point, the presence of hidden node, etc. In this example, different cases belong to different classes, and the classification needs to learn to label new data. An example of regression would be to estimate the throughput that the mobile user would have if it would connect to the specific access point. Regression

[13] Qualitative variables, also called *categorical* or *nominal* variables, have states, levels or categories that are defined by a set of mutually exclusive and exhaustive sub-classes [105], e.g., gender or color.

[14] Quantitative variables are those whose levels are expressed numerically. There are two types of quantitative variables: discrete and continuous [105].

is widely used for channel estimation in cellular networks [94]. For both classification and regression, performance heavily depends on the features with which is modeled the system, size, and accuracy of the training data set.

Throughout the present thesis, various supervised learning methods for regression were investigated and used to develop machine learning models for the wireless networking context. In Chapter 3 and Chapter 4, *artificial neural networks* have been applied for radio resource management, respectively, in Wi-Fi and cellular network. In Chapter 4, a probabilistic model based on *Bayesian networks* is also investigated for radio resource management in cellular networks. In Chapter 6, various regression methods are used for self-configuration and inter-cell interference management in dense small cells cellular networks, such as linear regression method, *decision trees* (*bagging* and *boosted*), *self-organized maps* (aka *Kohonen network*), *support vector machine*, *K-nearest neighbor*, *projection pursuit regression*. Corresponding chapters include descriptions of each of these methods and how they are used to build a machine learning model.

2.4.2.2 Unsupervised learning

Unsupervised learning aims to discover hidden patterns and find a suitable representation in the input data. In unsupervised learning, the input data set contains the data points without any additional information, i.e., the learning model is not told what the desired output is for each input [103]. While the supervised learning approach heavily relies on the quality of data samples labeling, the unsupervised learning approach depends on the accuracy or suitability of parameter settings, e.g., thresholds [94]. In recently published work [106], authors propose an unsupervised cellular planning technique (including the number of required BSs, their positions, and their radiation antenna patterns) based on statistical machine learning (SML), namely, Bayesian learning. In this work, the authors provide mathematical derivations that link probabilistic model parameters to the planning parameters. The advantage of unsupervised learning over supervised is that it does not require a human expert to label the data manually. Labeled data used in supervised learning is not only expensive to acquire, but it also contains relatively little information to estimate the complex models' parameters reliably.

2.4.2.3 Reinforcement learning

The reinforcement learning approach is a type of learning in which a learning agent should take actions in an environment to maximize its long-term reward [102]. The learner is not explicitly commanded (i.e., which actions to take) but instead can discover on its own which actions yield the most reward by trying them. This ability distinguishes reinforcement learning from supervised and unsupervised learning approaches whose models lack the ability to model and represent cognitive choices for future actions.

Elements of reinforcement learning are a *policy*, a *reward signal*, a *value function*, and optionally, a *model* of the environment [104]. A *policy* is a mapping from the perceived states of the environment to the actions taken in these states. For example, the policy may be a lookup table. An environment sends a *reward signal* to the reinforcement learning agent. The reward signal is the primary basis for altering the policy. The agent's objective is to maximize the

total reward it receives over the long run, which is a *value function*. How does learning work? For example, if a low reward follows an action selected according to the current policy, then the policy may be changed to select a different action when the agent enters the same state again. A *model* can optionally be used to predict the resulting reward for a given state and action. Hence, the model can be used for planning. If there is no model, the learning is based on trial-and-error.

Reinforcement learning methods are closely related to optimal control problems, particularly stochastic optimal control problems formulated as *Markov Decision Processes*[15]. Therefore, all the methods that are used for solving optimal control can be used for reinforcement learning, such as, e.g., *dynamic programming* algorithms. A reinforcement learning task that satisfies Markov property[16] is called a *Markov decision process*, and if the state and action spaces are finite, then it is a finite Markov decision process. Given any state and action s and a, and the probability distribution Pr, the probability of each possible pair of next state and reward, s', r, is denoted by:

$$p(s', r|\, s, a) = Pr\{S_{t+1} = s',\ R_{t+1} = r \mid S_t = s, A_t = a\} \tag{2.1}$$

where S_{t+1} and R_{t+1} are, respectively, the next state and the next reward due to action A_t at time t. These probabilities define the dynamics of a finite MDP entirely. Solving a reinforcement learning task means to find an optimal policy. An optimal policy is the one that is better than or equal to all other policies (i.e., there can be more than one). Policy π is said to be better than or equal to policy π' if and only if $v_\pi(s) \geq v_{\pi'}(s)$ for all $s \in \mathcal{S}$, where $\mathcal{S}$ is the set of possible states.

The most common algorithms used in reinforcement learning are: multi-arm bandits (the most simple form of reinforcement learning), dynamic programming[17] (computationally expensive and it requires a perfect model of environment), Monte Carlo methods, and temporal difference (TD) learning. The advantage of TD methods is that they do not require a model of the environment, and work well in the online learning regime. Examples of TD techniques are Q-learning and SARSA ("state-action-reward-state-action"). Some of the compelling applications of TD methods in wireless networks include dynamic channel allocation, resource and interference management, etc [107].

We just skimmed over the reinforcement learning theory to provide a high-level context for the reader of this thesis. A more thorough insight can be found in [104].

2.4.2.4 Deep learning

Lately, **deep learning** has emerged as a popular machine learning approach. Inspired by how human brain works (in particular, visual cortex), machine learning scientists have proposed

[15]Relies on Markov Processes named after A. A. Markov, who defined them and analyzed in 1907. Markov processes are those in which future states of a process are independent of the past and depend only on the present. A discrete-state Markov Process is called *Markov chain* [105].

[16]A *state* or *environment* signal that succeeds in retaining all relevant information is said to be Markov, or to have *the Markov Property* [104], (i.e., it is not necessary to remember the complete past trajectory). If the state signal has the Markov property then the environment's response at time $t + 1$ depends only on the state and action representation at t. If the state signal is non-Markov, then an approximation could be used. This property makes the process easier to analyze.

[17]The dynamic programming includes a collection of algorithms that can be used to compute optimal policies given a perfect model of the environment as a Markov decision process (MDP).

to construct machine learning models that will consist of many layers of processing where each layer is responsible for learning features at a different level of abstraction [103]. This is the main idea of deep learning, where the term "deep" refers to having multiple layers in the model. This idea is not entirely new, just that previously, computing hardware did not have enough computational capabilities for such complex models. While many authors in wireless networking literature mention it as yet another learning category, we consider that it is not, but it rather represents a variant of supervised and unsupervised approaches, which can be further used in combination with reinforcement learning methods to achieve even better performance (these ensemble methods are aka *deep reinforcement learning*).

In Figure 2.6, we show a visual example of deep supervised learning, and in particular how the convolutional neural network builds up its understanding of images layer by layer, from the most simple features, i.e., edges, to the most complex level of abstraction, i.e., objects (generated by *GoogLeNet* [108] which is trained on the *ImageNet* [109] dataset).

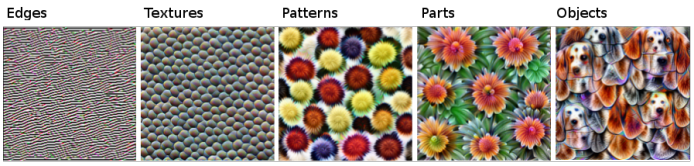


Figure 2.6 Visualization of how a convolutional neural network builds up its understanding of images over many layers, from more simple features on the left to more abstract features on the right (Individual images origin from Google's research blog [110]).

In [111], the authors provide a comprehensive survey of deep learning applications in wireless networking. The authors see the opportunities for the application of deep learning for mobility analysis, user localization, network control and security, signal processing, etc. As an important prerequisite, they highlight the availability of massive high-quality mobile big data. On the other hand, they emphasize the importance of tailoring deep learning models to mobile devices, systems, and networks, i.e., adapting models to their computational capability, energy constraints, privacy concerns, and other requirements.

2.4.3 Applications of machine learning in wireless networking

Machine learning has been extensively studied and applied in the context of wireless networks over the past two decades, and related work has been reviewed in the literature in [94, 96, 107, 111–114]. In the following we discuss a non exhaustive list of some of the most typical applications of machine learning algorithms in wireless networking:

- *Network planning*: Operators face different issues while performing network planning, in complex multi-RAT multi-frequency scenarios, which directly affect ROI. Recently, the authors of [115] have shown results of the application of machine learning for network planning by different companies (AT&T, iBwave, Nokia, and Keima) and some of the compelling results we illustrate in Figure 2.7. The overall conclusion is that machine

learning-based approaches can provide coverage[18] and dominance[19] while reducing the number of sites, which results in direct savings in site investments of about 40% and better coverage (optimized in terms of dominance: 83.8% for manual versus 94.7% for machine learning-based design) in this example of Manhattan area.

- *Variable estimation by classification or regression*: Machine learning algorithm is used, for example, for estimating the QoS or the QoE of the network, predicting different KPIs, or identifying different states of the network, by learning from the analysis of data obtained from past network measurements. SON functions where these tasks are useful are QoS estimation and different MDT use cases that can make use of a predicted values instead of having to wait for new set of measurement to be gathered, etc.
- *Diagnosis of network faults or misbehaviors*: The tasks belonging to this class of problems aim at detecting issues ongoing in the network, which may be associated to faults and anomalous setting of network parameters [96]. This kind of problem could be represented as a classification problem in which the output can take a discrete value from a finite set of different possible network fault causes or through an unsurpevised anomaly detection problem [96]. In SON, this kind of machine learning application is used for self-healing functions.
- *Pattern recognition*: Machine learning algorithms can be used to identify patterns, clusters with similar characteristics. This application of machine learning can be used for self-configuration functions.
- *Sequential decision problems for online parameter adjustment*: Application of machine learning algorithms in this domain is extremely common in the area of autonomous management, where control decisions must be taken online, during the operation, to adjust network parameters during its operation, to meet certain performance metric targets [96]. This kind of machine learning methods is typically used in self-optimization functions.

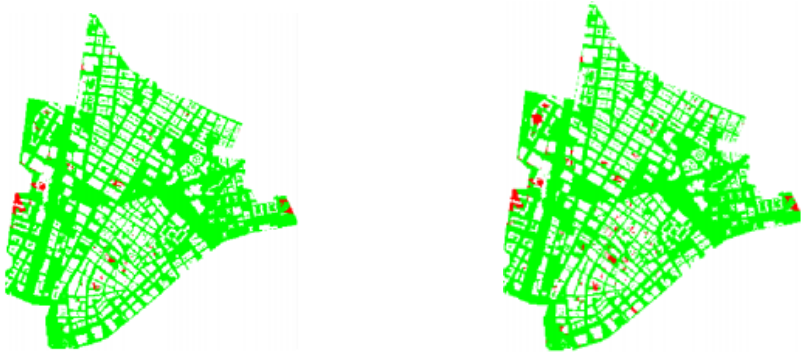
2.4.4 Machine learning standardization for 5G and beyond networks

In recent years, network operators along with service providers started to put a significant effort in the standardization of machine learning and artificial intelligence for 5G and B5G with the objective to define universal architectures, interfaces, and protocols, and to perform feasibility studies of machine learning and artificial intelligence applicability, and also to develop a unified way of testing and creating datasets, as noted in [116]. In this context, there are many ongoing standardization activities led by some of the leading standardization bodies and industrial specification groups, such as the following:

1. *European Telecommunications Standards Institute* (ETSI) has established two industrial specification groups:

[18] An outdoor location is considered covered if the strongest signal at that location has an RSRP greater than -112 dBm

[19] A location has satisfactory dominance if there is at least 5 dB of separation between the strongest and third strongest signal at that location.



(a) Manual planning using 185 sites and achieving 98.5% coverage

(b) Machine learning-based planning using 111 sites and achieving 97.5% coverage

Figure 2.7 Coverage (RSRP) results when using manual vs. machine learning-based site planning in Manhattan studied in [115].

(a) "*Experimental Networked Intelligence*" (ENI) [117] was officially launched by ETSI in February 2017 as the first networked artificial intelligence work group, with the objective to define a cognitive network management architecture [118]. ENI already has over 30 members, including many operators such as Vodafone, Telefonica, Verizon, China telecom, etc., and some vendors such as Huawei, ZTE, Samsung, Intel, etc.

(b) "*Zero-touch network and Service Management*" (ZSM) [119] was formed by ETSI in December 2017. The goal of ZSM[20] is to have all operational processes and tasks such as delivery, deployment, configuration, assurance, and optimization, executed automatically, ideally with 100% automation [116]. ZSM is initially focused on the 5G end-to-end network and service management (e.g., network slicing management) and will extend to the management for future generation networks. The major milestone was the publication of a document defining the reference architecture for the end-to-end ZSM framework [120].

2. *Third Generation Partnership Project* (3GPP) has defined Network Data Analytics Function (NWDAF) in the 5G System Architecture specification in [121] in May 2017, with a purpose to provide a centralized data collection and analytics.

3. *International Telecommunication Union* (ITU) established "*Focus Group on Machine Learning for Future Networks including 5G*" (FG-ML5G) in November 2017 whose objective was to draft technical reports and specifications for machine learning for future networks, including interfaces, network architectures, protocols, algorithms and data formats. An important milestone for FG-ML5G was the publication of the unified architecture for machine learning in future networks in July 2019 [122]. In FG-ML5G various operators such as Vodafone, China Mobile, Deutche Telekom, etc., various network vendors such as Huawei, ZTE, etc., and some research institutes and

[20]Zero-touch is a term used to describe the transformation of a manual IT process into a fully automated workflow.

universities, e.g., China Academy of Information and Communications Technology are actively involved. FG-ML5G work is divided into i) definition of use cases, services, and requirements for machine learning, ii) data format and machine learning algorithms, and iii) network architecture development for machine learning [116].

4. *China Communications Standards Association* (CCSA) has a network management technical committee (TC7), which is investigating the applicability of artificial intelligence techniques in 5G/B5G networks [116].

2.5 Network simulation framework

The present thesis aims to provide radio resource management solutions for LTE/NR and WiFi/WiGig systems in standalone and coexistence scenarios. Our goal is not only to make an analytic study, but also to perform evaluations based on high fidelity models. Hence, it is necessary to consider and represent model with accuracy of as many features as possible of real deployed wireless standards, including: i) technology-specific and standard-compliant protocol stack implementations, ii) end-to-end network system architecture, iii) coexistence of multi-RAT systems, iv) realistic transport protocol implementations, such as different flavors of Transmission Control Protocol (TCP) and Internet Protocol (IP), v) realistic traffic patterns, vi) obstacles such as buildings, vii) realistic channel and propagation models, high fidelity Link-To-System Mapping (L2SM) to abstract the PHY and reduce computational effort, in order to favor evaluation of scalable networks, realistic mobility models, antenna radiation, etc.

To achieve this objective, in the present thesis, we use *network simulation* as the principal tool to model the wireless networking system, to carry out the performance evaluation and validation of proposed wireless networking algorithms, schemes and models; and also to gather the network measurements that are used to build a knowledge base for supervised machine learning and statistical learning models for RRM and SON. Whenever possible, we also have carried out validation of the proposed algorithms, scheme and models through testbeds (e.g., such as in Chapter 3 and Chapter 7).

The network simulation approach was a natural choice over other alternatives such as experimental platforms (or testbeds) and analytical models. While experimental platforms have real implementation and thus do not introduce almost any modeling assumptions and simplifications, they also have some disadvantages: 1) not always available if the technology is new, as is often case in research, e.g., this was the case in my work with LAA, LTE-U, WiGig, NR, NR-U, etc., 2) not scalable, i.e., not suitable for complex scenarios such as dense small-cell deployments (e.g., such as deployments in Chapter 6, Chapter 7 and Chapter 8), 3) not practical for highly-mobile scenarios (such as those in Chapter 4 and Chapter 5), etc. We also consider that analytical models are interesting to build intuition on the general trend of the expected results, but are often insufficient to understand the deep inter-relations of all the features of the technology.

However, not all simulation tools are suitable for system-level simulations, such as we described previously. In general, there are two types of simulation tools: i) *link-level simulators*, and ii) *system-level simulator*. *Link-level simulators* focus on the physical layer, modeling it meticulously with the goal to replicate the physical layer functionalities of the real

wireless system. To achieve that goal, link-level simulators use detailed and computationally intensive PHY layer models of the air interface, such as, channel coding/decoding, multi-antenna gains, OFDM modulation, etc. Due to such detailed modeling, only a limited number of links can be simulated assuming an acceptable computational effort. As a result, this kind of simulators are not suitable for more complex multi-cell and multi-RAT scenarios, such as the ones that are highly relevant for the present thesis. On the other hand, *System-level simulators* use abstractions and simplifications at PHY based on L2SM, and thus can simulate many more links and model the interference in multi-cell and multi-RAT scenarios. Such system-level simulators normally provide the capability to evaluate end-to-end system performance, which is one of the performance evaluation requirements of the present thesis. Differently from link-level simulators, system-level network simulators may include modeling of the full protocol stack, the core network, the network topology, end-to-end connections, etc. Hence, for the present thesis we need to rely on a system-level simulator.

An important criteria for selecting a network simulator is the *open-source* availability. Commercial tools are not adequate for research because they are normally composed of black-box components whose underlying code and parameters remain unknown and are not accessible. In contrast, open source network simulators allow access to all implementation details and parameters. Additionally, the open source availability is highly appreciated in academy and research since it favors the results **reproducibility** [123] and allows collaborative developments. Because of this, we narrowed our search only to *open-source* network simulators.

There are various *open-source* network simulators available: ns-2, ns-3, OMNeT++, JiST and SimPy [124]. Among them, ns-3 demonstrated the best overall **scaling performance** in terms of simulation runtime and memory usage, especially for large-scale simulations [124].

In addition, by one measure (academic citations), ns-3 is the leading packet simulation tool. ns-3 is the most frequently cited tool used in computer network research (in the IEEE and ACM Digital Libraries). Additionally, as an example, Google scholar in September 2020 counts 7320 publication when searched for “ns-3 LTE”; while for the next most popular LTE simulator, *SimuLTE* [125] (based on OMNET), searching for “SimuLTE”, counts significantly fewer, 1320. If we search “ns-3 wireless” it counts 19,000 publications, while “OMNET wireless” counts again significantly fewer, 3,950. ns-3 is supported by a **large and active community** of users from both academia and networking industries, which continously extend, test, validate and improve ns-3 models.

Unlike other simulators openly available to the community, such as OMNET’s SimuLTE, ns-3 offers **multi-RAT models** for coexistence. While SimuLTE supports multi-RAT scenarios in which the same node may have various network interface cards (NICs), there does not seem to exist much support for inter-technology coexistence studies. SimuLTE is mentioned for the first time in literature in 2019 [126] in the context of coexistence among different IEEE standards, namely 802.11 and 802.15.4. On the other hand, since 2009 ns-3 simulator has *spectrum module* [127] that implements frequency-dependent aspects of communications allowing a more accurate spectrum modeling, which set a foundation for simulating **inter-technology coexistence** (ns-3 supports different IEEE standards coexisting since 2009, and since 2015 ns-3 has the external module for LTE and Wi-Fi coexistence [3], and its extension [5] developed by myself).

Another highly appreciated aspect for network simulator is whether it was being **validated** and **calibrated** against the testbed or external reference. In that sense, the ns-3 simulator fulfills since its wireless modules have been validated and calibrated in various independent studies. For example, the ns-3 Wi-Fi module was validated against a real testbed in [128]. Also, ns-3 LTE module is validated with an extensive set of tests that cover all the main models/functionalities using the official ns-3 test framework. Additionally, ns-3 LTE has been validated against a real world testbed, demonstrating that it can deliver voice quality and latency as good as an experimental testbed using actual LTE equipment over a range of signal-to-noise ratios [129]. Finally, calibration campaigns have also been performed in 3GPP reference scenarios [130]. Results show that the ns-3 LTE module achieves similar performance to those obtained by the 3GPP industrial simulators in the evaluated cases, both in terms of SINR distributions and users' throughput.

Last but not least, I pick for the evaluations of this thesis the ns-3 simulator, because my research group has a long tradition and involvement in this open source community. In particular, the MONET research group of CTTC initiated the design and development of the LTE module in 2010 (at the time of Release 10), and in 2019 launched the first release of the new NR module. Personally, I have been over all my years at CTTC intensively involved in the design and development of these two models, and I have been participating during my PhD thesis in many projects to improve, and validate our simulators, with funding of different US agencies and companies, like WFA, Spidercloud Wireless, Interdigital, and the Lawrence Livermore National Lab. I also am an official developer and maintainer of the LTE and NR modules, and I regularly provide courses and tutorials addressed to the ns-3 community and to industries. As such, I consider that I am very expert in the simulator and consequently this is the natural choice for the evaluations in my PhD thesis work.

To summarize, ns-3 simulator fulfills the following requirements: i) supports system-level studies, ii) it is open-source, iii) has low computational complexity, iv) has a broad academic and industrial community, v) has been validated and calibrated, vi) it is a multi-RAT simulator, with LTE/NR and WiFi/WiGig being its most popular modules, and as such it contains models that are necessary to support scenarios that are considered in this thesis, e.g., inter-technology coexistence, vii) it is highly modular, which allowed us to extend the simulator when it was necessary by building upon the existing blocks and modules of the ns-3 code-base, and, viii) the author of the thesis has an expert knowledge of the ns-3 simulator, which was needed to create new models, algorithms, test, validate, calibrate, and to carry out exhaustive and complex performance evaluation studies to achieve research objectives of this PhD thesis. Because of all these reasons, we have identified ns-3 network simulator as the best simulation framework to reach the objectives of the present thesis.

2.5.1 ns-3: A brief overview

ns-3 is a discrete-event[21] network simulator, entirely written in *C++* with optional *python* bindings. It has been under continuous development since 2005, courtesy of funding from the *US NSF*, and several other public and private organizations. It is licensed under the GNU General Purpose License, version 2.

[21]The key property of discrete-event simulations is that the state of the simulation model can only change at discrete points in time which are referred to as events [123].

ns-3 offers very mature, well tested and validated implementations of Wi-Fi (802.11a/ b/ g/ p/ e/ n/ ac standards) and LTE, and evolutions of these technologies inside IEEE and 3GPP. Lately, a new implementation for .11ax has been made available to the community, even if still not included under the mainline. The models have been funded by *Cisco* and *Intel*, and are currently still under development. Parallel branches also include support for .11ad.

The ns-3 LTE module, commonly referred to as LENA, combines a simulated channel and physical layer model with an implementation of the LTE and EPC (Evolved Packet Core) protocol stack that closely follows 3GPP specifications. Most of the current version of LENA was developed between 2011 and 2013 as part of an industrial project funded by *Ubiquisys Ltd.* (now part of *Cisco*) and carried out by Mobile Networks group at CTTC. The LTE model also supports different features in separate branches, like the D2D model that was the result of a collaboration of CTTC and the *National Institute of Standard and Technologies in the US (NIST)* during the last two years [22][21]. LTE module also has the LAA and the LTE-U branches, which result from almost 3 years long collaborations of CTTC and the *University of Washington* between 2015 and 2017. The LAA was developed with *Wi-Fi Alliance (WFA)*, and LTE-U was deveveloped with *SpiderCloud Wireless*. The author of the thesis is the principal developer and maintainer of both ns-3 LTE unlicensed models, LAA and LTE-U [5].

In February 2019, my group released the first version of the first open source NR simulator [8], based on ns-3, see [15] for more information or to download. The model is a fork of LTE and ns-3 mmWave models, and it mainly focuses on refactoring the PHY and MAC layers of the LTE module in order to provide a standard-compliant implementation of NR Release 15 [90]. The RRC (Radio Resource Control) and upper layers, still rely, as of today, on the LTE implementation, as much as the EPC (Evolved Packet Core), which makes the proposed NR model an NSA (non-standalone) implementation. *Interdigital* funded this work. As a result new models were released recently in 2020, including support for NR-U and coexistence with WiGig in 60 GHz. The NR model is currently being extended in the context of the *S3 (Spectrum Sharing Simulator)* project, led by the *Lawrence Livermore National Lab*, under funding of the *US Department of Defense (DoD)*. The extensions will enable the simulation of mixed 3GPP multi-RAT scenarios, and evaluate spectrum sharing capabilities of 3GPP technologies in realistic deployments in multiple bands, subject to shared access rules. My team is also extending the NR module to support C-V2X and NR V2X in collaboration with *National Institute of Standard and Technology in the US (NIST)*. I am a key player in all these achievements, and this is the main platform used for all the studies carried out in my PhD thesis.

2.6 Conclusions

In this chapter, we have provided the overall context and the background of the present thesis. To achieve this goal, in Section 2.2, we first have provided a brief historical overview of the development of wireless technologies with a focus on 3GPP and IEEE technologies. The purpose of this historical overview is to remind the reader about the main drivers of the cellular and WLAN technological development up to now, and besides the main technological advancements that allowed their capacity to keep growing over time. In Section 2.3, we have analyzed the problem of an exponential increase in capacity demands, and explored the three

dimensions through which the next capacity targets could be reached, for 5G and B5G. The three dimensions are: densification, spectrum efficiency and spectrum aggregation. We have then explained how each of these dimensions is related to the different parts of the thesis. Successivelly, in Section 2.4, we have introduced the SON paradigm, and explained why it is expected that SON will be a key player in future mobile networks, through the introduction of machine learning. In this context, we have provided different examples of how machine learning is used in networking, after briefly introducing high level machine learning concepts. Finally, in Section 2.5, we have discussed the main reasons why we choose ns-3 as our basic system simulation strategy. We have also listed the contributions made in this community by my group, and highlighted my personal contributions.

Chapter 3

RRM in WLAN: A ML based approach to IEEE 802.11 WLAN access point selection

In this chapter, we deal with the problem of Access Point (AP) selection in IEEE 802.11 Wireless LANs (WLANs). We propose a machine learning based radio resource management scheme for access point selection. To add the learning capability to the system, we use Artificial Neural Network (ANN) as the basis of a supervised learning engine. The objective is to investigate whether a machine learning scheme can yield better throughput performance compared to other methods already proposed in the literature. Then we validate the proposed machine learning scheme in a real-world environment by using the experimental tetbed platform. The objective is to understand whether the machine learning based scheme can learn equally well from real-world measurements as from the simulation environment. Furthermore, we seek to answer questions such as how realistic measurements impact the learning performance, and the overall access point selection performance.

As mentioned in the previous chapter, almost 80% of mobile data traffic is offloaded to Wi-Fi networks. Due to wireless networks densification and different technologies that allow to offload traffic over Wi-Fi, the access point selection algorithm is one of the key algorithms to ensure reliable and high throughput connectivity through Wi-Fi. This is especially the case in situations when many APs are available; typical scenarios are residential and enterprise buildings, university campuses, and airports. This scenario is similar to small cell deployments with the open subscriber group paradigm, in which the mobile user needs to select the small cell to attach to among several alternatives with partially overlapping coverage areas. Depending on the propagation environment and the traffic load, the performance that the mobile user can perceive from different APs may vary significantly; as a consequence, it is interesting for the

mobile user to identify and select the AP that will provide the best performance.

Since the IEEE 802.11 standard [131] does not specify how to select the most appropriate AP, manufacturers implement different solutions. In fact, in order to be competitive in the market, it is interesting for a manufacturer to implement an AP selection scheme which will provide good throughput performance to the mobile user. The most commonly implemented scheme in mobile devices is the one based on the Received Signal Strength Indicator (RSSI) measurement and consists of selecting the AP from which the device receives beacon frames with the strongest signal strength. This type of scheme is easy to implement and neither requires changes on the APs nor additional exchange of control information. However, its major drawback is that the APs which are close to the majority of the users can become overloaded, while other APs which are at a longer distance remain underutilized. This is due to the fact that the RSSI scheme is neglecting different traffic loads of the available APs.

3.1 Related work

To address this problem, several schemes that consider the traffic load for the AP selection process have been proposed in literature. These schemes can be divided into two main groups: centralized and decentralized. In centralized schemes, a separate management system connected to the WLAN provides mobile users with information regarding the load of the APs; such a scheme is proposed in [132]. The benefit of this scheme is that it has all the information which is necessary to perform load balancing among the APs. A major issue of this approach is that it can be used only at locations where this management system is deployed, and therefore it does not allow an efficient AP selection in legacy WLAN deployments. For this reason, we do not consider centralized schemes.

On the other hand, decentralized load-based AP selection schemes are entirely implemented on the mobile stations and neither require changes to APs nor to the specialized WLAN management system. These schemes typically use metrics that are representatives of the load of each AP. In [133], the authors use the probe delay metric, which is defined as the difference between the probe request time and the probe response time for a specific AP. The problem with this algorithm is that it only considers those APs whose Signal-to-Noise Ratio (SNR) is greater than a predefined threshold and excludes others, which could offer better performance in situations where the APs with the higher SNRs are overloaded. Moreover, since the probe delay mostly depends on the uplink traffic, the schemes based on this metric will give bad results when most traffic is downlink (e.g., web browsing). In [134] and [135], the authors propose schemes based on the average transmission time metric to estimate of the throughput. Since the average transmission time is calculated using measurements gathered by monitoring the wireless medium, and since the mobile station's coverage region is different from that of the AP, the mobile station cannot always decode all the frames exchanged between the AP and other stations. Hence, in the presence of hidden nodes, the estimated value of the throughput becomes less accurate, potentially yielding incorrect AP selection decisions. For this reason, some authors, e.g., in [134] and [136], suggest modifications in an AP to provide the stations with additional information about the AP load.

Regardless of the used technology (Wi-Fi, LTE, etc.), the fundamental issue of the AP selection problem is that the performance achievable from a particular AP depends on many

environmental factors and complicated relationships that is not feasible to model it analytically without making many assumptions and significant simplifications. For example, in the case of the 802.11 technology, AP schemes based solely on theoretical considerations [133, 134, 136] work adequate only in specific situations, but fail to work properly over a large variety of conditions that are encountered in realistic scenarios, as we will show in the performance evaluation Section 3.5.3 and Section 3.6.3. In order to design an AP selection scheme which can provide better performance than state-of-the-art schemes, we propose a decentralized scheme based on machine learning and which leverages both the RSSI and load-based metrics. Because of the complexity of designing an analytical model based on these metrics that can perform well also in realistic conditions (e.g., in the presence of hidden nodes), we choose a machine learning approach [137]: the mobile station learns from its past experience how the environmental conditions influence the throughput performance; the cognitive engine then uses this knowledge to select the AP that is expected to provide the best performance. The learning is said to be *supervised* since it is based on known labeled training data, which in our case consists of the measurement data gathered by the mobile station. After learning has been accomplished, the mobile station uses the acquired knowledge to estimate the throughput of all available APs, and then selects the AP with the highest value of the estimated throughput. As we will show in the performance evaluation section, the proposed AP selection scheme achieves a significant performance enhancement with respect to the RSSI and load-based decentralized AP selection schemes.

3.2 Machine learning based AP selection scheme

Our primary objective is to design an AP selection scheme to enable the mobile station to select the AP that offers the best performance. To achieve this goal, we use a supervised machine learning based scheme to estimateD the performance for all available APs. In Figure 3.1, we illustrate the flow chart of the proposed machine learning based scheme. First, the station performs a scan in order to discover the available access points. For each discovered AP, the station gathers the measurements regarding the environmental conditions. Then, the cognitive engine uses these measurements as inputs and provides as output the estimated performance of each discovered AP. This procedure is repeated for all the available APs, and depending on the estimated performance of the APs, the best AP candidate is updated. Once that all the APs are evaluated, the station connects to the best AP (according to cognitive engine prediction) and measures the obtained communication performance to prepare a new sample for the training set. The data set is built using the environmental measurements gathered before and after the connection is established, and this data set represents the past experience from which the cognitive engine learns during the training phase the correlation between inputs and the output. Each time the knowledge base is updated, the cognitive engine learns more and improves the prediction accuracy for future predictions. As the authors of [138] discuss, this is the key aspect of the cognitive process. Once the station is connected to the selected access point, it continues to observe the environment to check whether it is convenient to switch to a different AP. This happens if the difference between the estimated throughput for the new best access point and the currently selected access point is higher than a given threshold. The threshold is needed to avoid frequent reconnections and the ping-pong effect in response to minor environmental conditions variations.

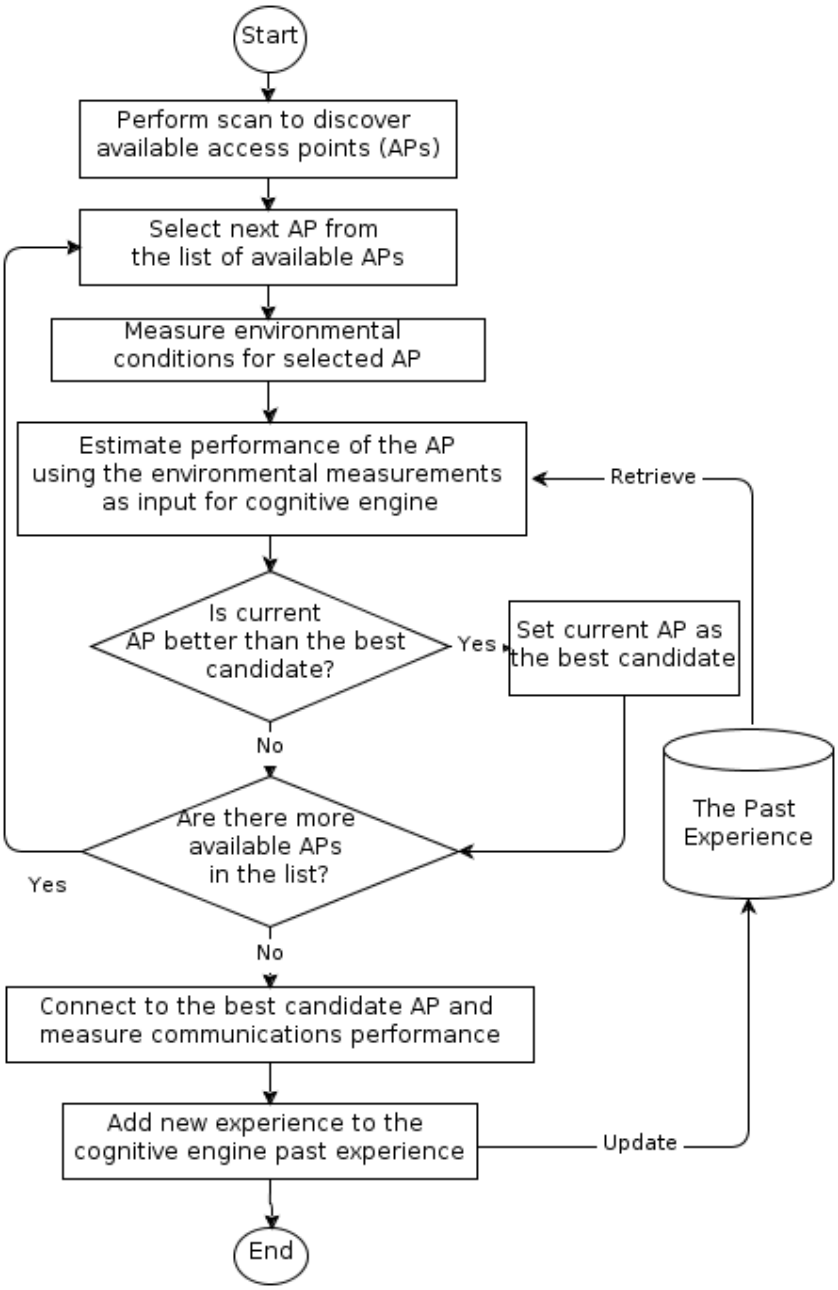


Figure 3.1 The flow chart of proposed machine learning based AP scheme

3.3 Feed Forward Neural Network

For the implementation of the machine learning based scheme, we use the Feed Forward Neural Network (FFNN), a machine learning technique that can model non-linear functions between inputs and outputs. We choose a MFNN because it gives a more compact model than other prediction techniques with the same generalization performance, such as support vector machines [101]. Furthermore, this choice is also supported by other works that considered different prediction techniques for wireless networking applications. For example, the authors of [139] did a study on the network traffic prediction using MFNN and Auto Regressive Integrated Moving Average, and as a result they recommend the use of a MFNN

as a prediction technique with less complexity and better results. MFNN are widely used for pattern recognition purposes [101].

A MFNN is an adaptive system that can be used to infer the future performance of the system as a function of the available network measurements. We can write the MFNN function [101] as:

$$y(\mathbf{x}, \mathbf{w}^{(1)}, \mathbf{w}^{(2)}) = f\left(\sum_{h=1}^{H} w_h^{(2)} f\left(\sum_{m=0}^{M} w_{hm}^{(1)} x_m\right) + w_0^{(2)}\right) , \tag{3.1}$$

where $\mathbf{x}$ is the vector of the inputs x_m, with $m = 1, \ldots, M$, and $x_0 = 1$; H is the number of nodes in the hidden layer; $\mathbf{w}^{(1)}$ is the vector of adaptive weights $w_{hm}^{(1)}$ for the hidden layer, with $h = 1, \ldots, H$; $\mathbf{w}^{(2)}$ is the vector of adaptive weights $w_h^{(2)}$ for the output layer. The sigmoidal activation function is defined as:

$$f(a) = \frac{1}{1 + e^{-a}} . \tag{3.2}$$

To reduce the complexity of the MFNN architecture, we seek the topology with the minimum number of hidden layers sufficient to approximate any continuous function on an input domain to arbitrary accuracy, provided the network has a sufficiently large number of hidden units. According to [101], this is a two-layer topology. Hence we use a two-layer[1] FFNN.

The MFNN output value is obtained as a function of all the input parameters in $\mathbf{x}$. For the output variable of MFNN, we select the throughput that is the key performance metric used in our scheme to evaluate available APs and make the AP selection decision. We select as inputs of the MFNN the set of metrics observable by the mobile device and are highly relevant for the selected MFNN output variable. We will describe more in detail, all considered input and output metrics in Section 3.4.

3.4 IEEE WLAN 802.11 feature extraction

For the inputs and the Neural Network output, we use metrics representing environmental conditions and the throughput that the mobile station can achieve by connecting to specific AP. Since the coverage and interference regions of the mobile station and the AP are, in general, different, it is obvious that the station cannot gather perfect environmental information regarding the AP. Therefore, the environmental measurements will represent how the mobile station "sees" the communication environment. Since we design an AP selection scheme in a decentralized manner, we need to select the features relevant for the AP selection algorithm in IEEE WLAN 802.11, and that can be calculated by leveraging measurements gathered solely by a mobile device. For this purpose, we define the following metrics that describe the environmental conditions:

- the signal to noise ratio $\gamma \in \mathcal{R}$
- the probability of failure $p_f \in [0, 1] \subset \mathcal{R}$

[1] We use the same terminology as in [101] where the number of layers refers to the number of layers of adaptive weights.

- the business ratio $b_r \in [0, 1] \subset \mathcal{R}$
- the average beacon delay $t_b \in \mathbb{R}_{\geq 0}$
- the number of detected stations $n_s \in \mathcal{Z}$

and the metric that describes the perceived performance:

- the throughput R_d

γ is defined as the ratio of the received signal power (RSSI) to the noise power corrupting the signal; both measurements are available on most commercial WLAN devices.

p_f represents the ratio of the number of retransmitted frames and the total number of frames. We calculate p_f using the $retry$ frame flag from the MAC header of IEEE 802.11. The $retry$ flag takes 0 as a value when a frame is transmitted for the first time and a value of 1 when it is retransmitted. Let s and r be the numbers of DATA frames that are successfully decoded by the mobile station and have the value of the retry flag 0 and 1, respectively. We define p_f as:

$$p_f = r/(r + s) \tag{3.3}$$

b_r represents the ratio of time in which the channel is occupied by the frame transmissions that the mobile node can decode successfully. The calculation of b_r is done in the following way:

$$b_r = \frac{\sum_{i=1}^{n} T_i}{T} \tag{3.4}$$

where T is the duration of the time interval in which the mobile node gathers measurements, n is the total number of frame exchange sequences which are transmitted during the interval T and T_i is the duration of the $i - th$ frame exchange sequence. T_i is calculated as the sum of the duration of the DATA frame, the ACK frame and the DIFS (or AIFS in the case of QoS support) and SIFS, as is defined by the standard [131].

t_b is the average beacon delay in the beacon transmissions from an access point during the time T in which the mobile node gathers the measurements. The beacon delay is equal to the difference between the timestamp when the beacon is transmitted and target beacon transmission time (TBTT), as is illustrated in Figure 3.2. The value of the beacon timestamp is obtained from the management frame field and TBTT is calculated and updated using the beacon interval time and beacon sequence numbers, which are obtained from the beacon frame.

n_s is the number of stations detected by the considered station to be exchanging frames with the AP. This metric is calculated by counting the distinct values of the source and the destination address fields in the decoded frames. Since we consider scenarios in which one dedicated orthogonal channel is used for each AP, we can calculate n_s by counting all the different station addresses which are detected by the monitoring station.

R_d is the downlink throughput of the TCP flow. We choose this metric to describe the perceived performance since it is a relevant metric for most of the operations commonly performed by the mobile users in a WLAN, such as a web browsing and file transferring.

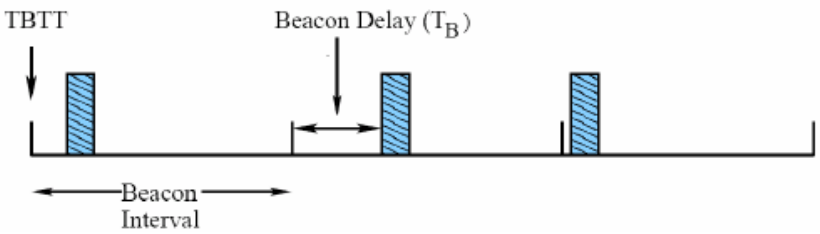


Figure 3.2 Beacon transmission delays and TBTT on an AP

3.5 Simulation based modeling and evaluation of machine learning based AP selection scheme

3.5.1 Simulation scenario setup in ns-3 network simulator

For the performance evaluation of the proposed scheme, we used the ns-3 network simulator [13]. We choose ns-3 for this study because it is an open source, it has good TCP/IP and Wi-Fi models (see, for example, the validation study in [140]), and it has excellent run-time performance and memory usage [141]. Moreover, ns-3 supports saving packet traces in the PCAP format with RADIOTAP header, which is the same format commonly used for real Wi-Fi devices. We consider this feature important to make it easy to implement the proposed scheme in a real testbed in the future. For the processing of the PCAP traces and the calculation of the metrics described in Section 3.4, we use the PCAP Trace Parser [142], which we developed at the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC). The data obtained from the parser is stored in a MYSQL database, from where it is fetched for training and testing the Neural Network.

The performance evaluation of our scheme is done in two parts. In the first part, we build the scenario to obtain the measurements needed to train the Neural Network and identify the values of the training parameters that provide the best prediction accuracy. Then, in the second part, we evaluate the AP selection scheme on the test set. The training of the Neural Network is done with measurements obtained in a scenario in which one mobile station, which in the remainder of the Chapter is referred to as the tagged station, is in coverage of a single AP. A number N of other stations, called background stations, are connected to the AP. An example of this scenario is represented in Figure 3.3, where we have one AP with 7 background stations (STA1-STA7) plus the tagged station (STA), which is at a distance d from the AP. We fix the number and position of the stations for each simulation. We run a TCP file download on all the stations. The background stations are uniformly randomly distributed within a disc centered at the AP and having a radius equal to 150 meters. We choose a radius equal to 150 meters because this is approximately the radius of the coverage area of an AP using the default ns-3 parameters. For rate adaptation, we choose the algorithm, which in ns-3 is called "IdealWifiManager". This algorithm is similar to Receiver-Based AutoRate (RBAR) [143], according to which each transmitter keeps track of the last SNR sent back by a receiver and uses it to pick a transmission mode based on a set of SNR thresholds built from a target Bit Error Rate (BER) and transmission mode-specific SNR/BER curves. During each simulation, the behavior of the tagged station is a bit different from that of background stations. In the first part of the simulation, the tagged station monitors the wireless medium and gathers the measurements

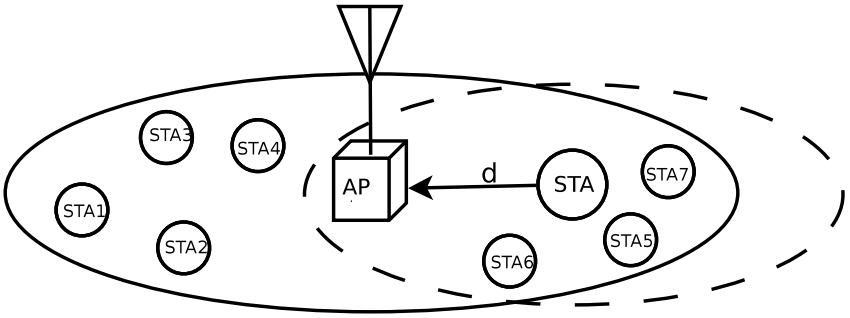


Figure 3.3 Neural Network training scenario

necessary for calculating the metrics described in Section 3.4, and it does not perform any data communication. In the second part the tagged station connects to the AP, performs a TCP file download and measures the throughput of this flow. The gathered environmental and throughput measurements are used as inputs and output for Neural Network training. In this way, the Neural Network will learn how the throughput depends on the environmental conditions measured before the connection with the AP is established.

We run 192 different simulation campaigns where each has a different combination of N and d values, and in each simulation campaign, we run 30 independent repetitions, which gives 5760 samples. One sample represents a set of inputs and related outputs.

3.5.2 Neural Network configuration, training and testing on simulation data

In this section, we focus on the Neural Network configuration, training, and testing. Neural Network configuration consists of input variable and key neural network training parameters selection. The testing consists of calculating the error between the throughput value predicted by Neural Network and the actual value measured by the device. For the Neural Network implementation we used FANN, a publicly available software library for building neural network models available in [144]. The 5760 samples are divided into two sets of equal size, where the first is used for the Neural Network training and the second for the testing. The training is done by using cross-validation based on a random subsampling. Following the recommendations from [145], we use the iRPROP- batch training algorithm [146]. The input variable selection is a fundamental part of the Neural Network configuration as it directly affects the computational complexity and the prediction performance. There are many different algorithms proposed in the literature that can deal with high dimensionality in models, complex correlations between inputs, complex noise function, and satisfy the optimality criterion. The advantage of our model in this sense is that it has relatively low dimensionality, so it is feasible to apply plain exhaustive search algorithm for the Neural Network inputs selection, i.e., to select the subset of inputs for which the neural network achieves the best prediction performance. In Figure 3.10, we show the resulting Neural Network configuration. In general, the accuracy of a Neural Networks model depends on the parameters which are used for its training [147]. To discover the values of the training parameters for which our Neural Network gives the best performance we carried out a neural network training campaign varying the following parameters:

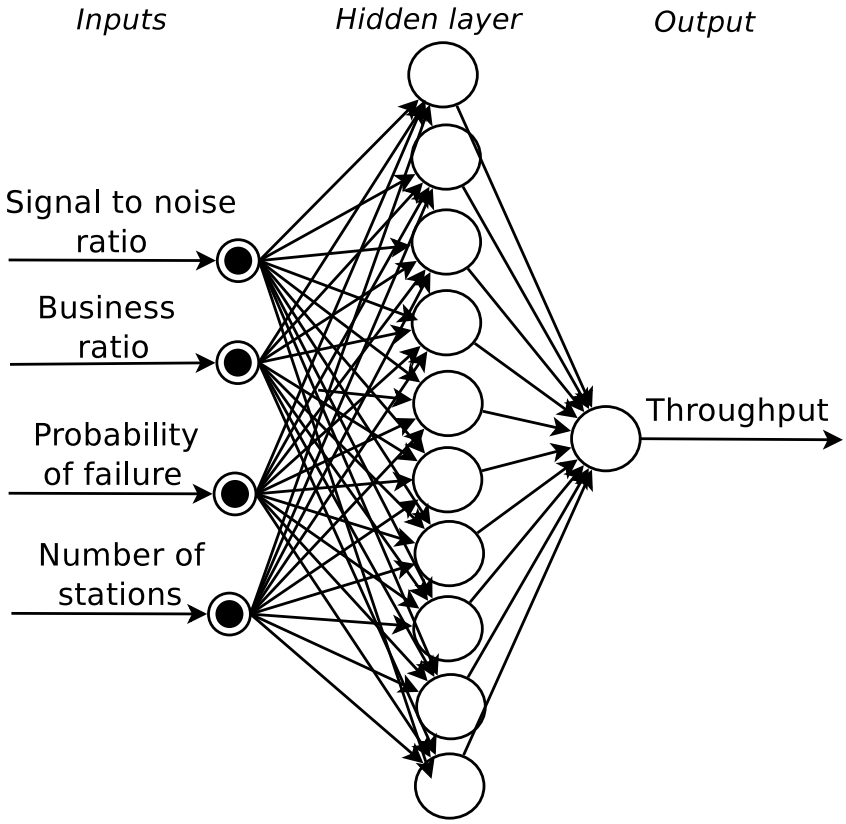


Figure 3.4 Architecture of neural network based machine learning model with four inputs, a hidden layer and one output

- the number of nodes in the hidden layer $H \in \mathcal{Z}$
- the maximum number of epochs $E \in \mathcal{Z}$
- the learning rate $L \in [0, 1]$.

In Figure 3.5, we report the Normalized Root Mean Squared Error (NRMSE) of the estimated throughput for different learning parameters configurations. The RMSE is calculated in the following way:

$$NRMSE = \frac{RMSE}{X_{max} - X_{min}}, \tag{3.5}$$

where RMSE is the Root Mean Squared Error calculated on the testing set and X_{max} and X_{min} are, respectively, the maximum and the minimum values of the measured throughput.

As evident from the figure, the most important parameter for the training of the Neural Network is E, while H and L do not appear to play an important role. The best performance of the Neural Network is achieved with E=20, for which the NRMSE has its minimum value. For values of E lower than 20, the error is bigger because the Neural Network does not have enough knowledge yet; conversely, for values higher than 20 the Neural Network becomes over-trained and too specialized on the training data set, so it looses generalization, which is needed for the good interpretation of the testing data set. Since the Neural Network provides the best estimation for E=20, we will use this value for the training of the Neural Network used in performance evaluation of the AP selection scheme, which we will describe in the next subsection.

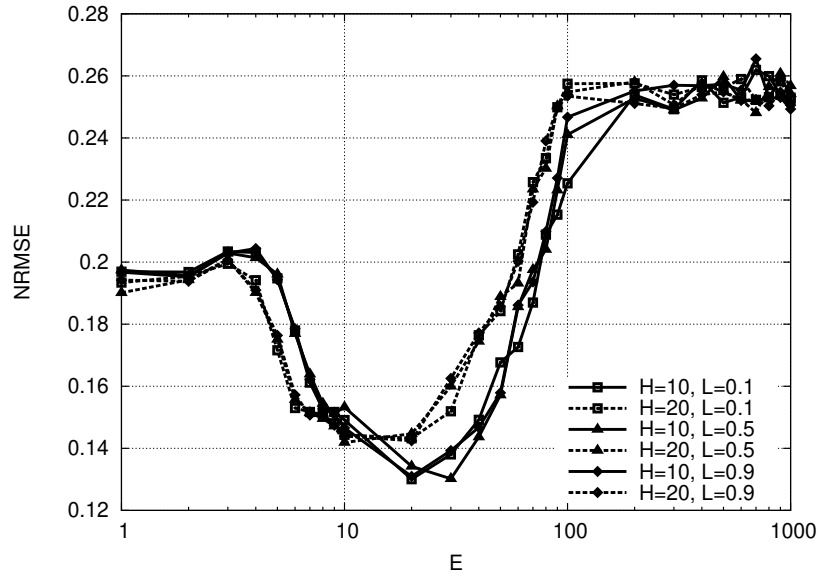


Figure 3.5 Performance of Neural Network based estimation of AP throughput

3.5.3 Simulation evaluation of machine learning based AP selection scheme

For the performance evaluation of our AP selection scheme, we consider a scenario in which the tagged station is in coverage of two APs, respectively called AP1 and AP2, operating on orthogonal channels. Even if we do performance evaluation in a scenario with two APs only, our scheme can be applied to any number of APs. This is illustrated in Figure 3.6. In this scenario, the APs have, in general, different numbers of background stations, respectively N_1 and N_2, and are at different distances from the tagged station, respectively d_1 and d_2. For each simulation run we fix the number of background stations and the positions of all stations and APs. The background stations for each AP are uniformly randomly distributed within a disc of radius 150 meters centered on the AP. On each background station, a TCP file download is performed. The tagged station is passively monitoring the wireless channel to calculate the metrics that are described in Section 3.4. We used the Neural Network trained with the training parameters and data described previously in Section 3.5.2. The trained Neural Network then provides the value of the estimated throughput for both access points, AP1 and AP2. After obtaining these values, the tagged station selects the AP with the higher value of estimated throughput, connects to it, and starts a TCP file download. The simulation campaign is carried out by varying N_1, N_2, d_1 and d_2.

3.5.3.1 Simulation scenario with varying load

To present how the performance of our scheme depends on the difference in the load of the APs, we first consider the subset of experiments with a fixed value of d_1, d_2 and N_1, while we vary N_2. The values of the parameters are summarized in Table 3.1. We compare our machine learning based AP selection strategy with the RSSI and the load-based schemes. In Figure 3.7, we compare the performance of all the AP selection schemes for the scenario described above. The X axis represents N_2, and the Y axis represents the average throughput which is obtained

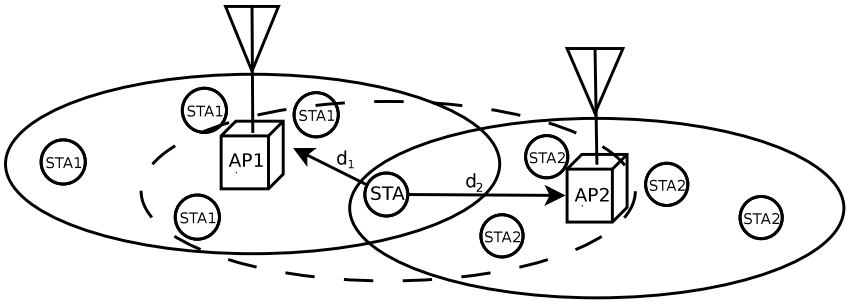


Figure 3.6 Tested scenario for AP selection

Table 3.1 Configuration parameters for different network scenario simulations

Scenario	N_1	N_2	$d_1[m]$	$d_2[m]$
Varying load	5	1-11	120	30
Varying distance	2	11	80	0-150

by the tagged station using each AP selection scheme. Since AP2 is closer to the tagged station, AP2 is a better choice when $N_2 < N_1$, because it is less loaded and closer. Conversely, when $N_2 > N_1$, AP1 is a better choice because it is less loaded. Thus, the expected behavior for all schemes is that the throughput obtained by the tagged station decreases with an increase of the load on AP2. In this scenario, the RSSI scheme has good performance while AP2 is less loaded, but when it becomes more loaded ($N_2 > N_1$) the RSSI scheme performs worse, because it selects the wrong AP. On the other hand, the load-based AP selection scheme has inferior performance because it has only partial information about the load of AP1 due to its distance. The load-based scheme in many cases chooses AP1 because it seems to be less busy, which is wrong when $N_2 < N_1$. When $N_2 > N_1$, AP1 becomes the best choice, and the load-based scheme achieves good performance. While the RSSI scheme gives good results for $N_2 < N_1$ and the load-based scheme for $N_2 > N_1$, our scheme gives good results for all values of N_2. To summarize, compared to the RSSI and the load-based schemes, our scheme performs better in all situations that arise in this scenario, thanks to its ability to learn all the environmental aspects that affect performance.

3.5.3.2 Simulation scenario with varying distance

The second group of simulations is characterized by fixed values N_1, N_2, d_1 and variable d_2. The settings for this scenario are summarized in Table 3.3. Figure 3.8 shows the performance of our scheme, compared with the RSSI and the load-based schemes. For $d_2 < d_1$ we see that the RSSI based scheme results in a wrong decision because it always connects to AP2, which is closer but much more loaded than AP1. Conversely, for $d_2 > d_1$ the RSSI scheme will perform better because it will always choose AP1, which is in this case closer to the tagged station. The load-based scheme provides better performance than RSSI, but as the distance from AP2 increases, it starts to choose more often AP2, since it considers it as less loaded. This is due to the fact that environmental information becomes more partial as the distance increases, as we explained in the introductory part of this chapter. It can be noticed that the Neural Network based scheme has excellent performance for all values of d_2 and performs better than the RSSI and the load-based schemes.

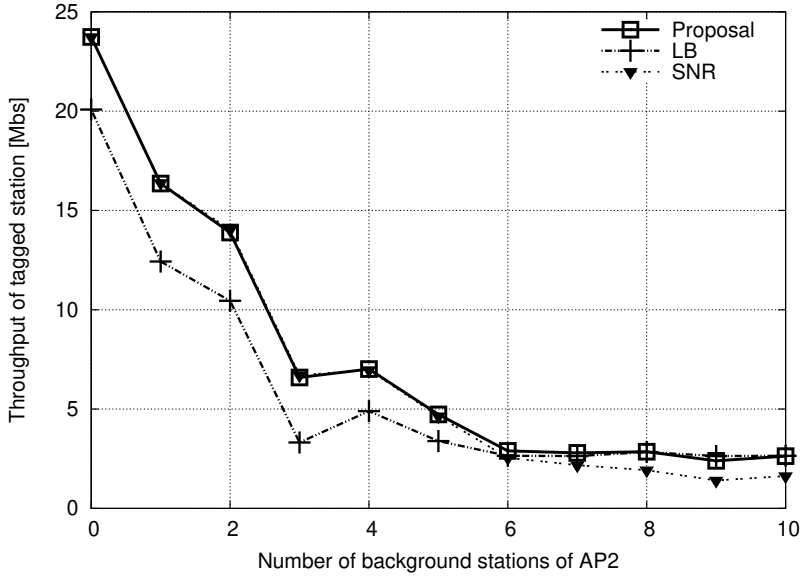


Figure 3.7 Performance of different AP selection schemes in a scenario with two AP, for different values of load of AP2, while load on AP1 is fixed as well as distance from tagged station to both APs

3.6 Experimental validation of machine learning based AP selection scheme

3.6.1 Experiment setup in EXTREME Testbed

We carried out an experimental performance evaluation of the proposed scheme using the EXTREME Testbed® [148]. The configuration that we set up for experiments in the testbed is shown in Figure 3.9. All stations are regular PCs running the Linux operating system, using the *Madwifi* driver for the wireless interface, and they are connected via Ethernet to another PC which is used as a central experiment controller. We use the central controller PC to run the experiments, to monitor their execution and to gather the measurements from all the stations. The wireless interfaces of all the stations are interconnected using RF cables to minimize noise-related channel errors. We used a commercial *CISCO Aironet 1200 Series* access point configured in 802.11g mode. The settings of the access point are shown in Table 3.2. On each station we installed the *Iperf* tool (version 2.0.4) [149] to create TCP data streams in the downlink and to measure the obtained throughput; more information on the Iperf tool can be found in [150]. One station, referred to as the *tagged* station, is the node running the AP selection scheme, which is connected to the AP via attenuator. The other stations are referred to as *background* stations; the signal toward them is not attenuated, and hence they ideally can use the highest transmission rate of 802.11g, which is 54 Mbps. For all the stations, the actual rate being used is dynamically selected by *Madwifi* driver's default rate adaptation algorithm.

We note that 0 dB corresponds to the station being able to use the maximum physical rate, and 25 dB corresponds to the station starting to loose connection with the AP. In each experiment we fix the number of active background stations N and the attenuator value t. During each experiment the active background stations perform a TCP downlink transfer using the *Iperf*

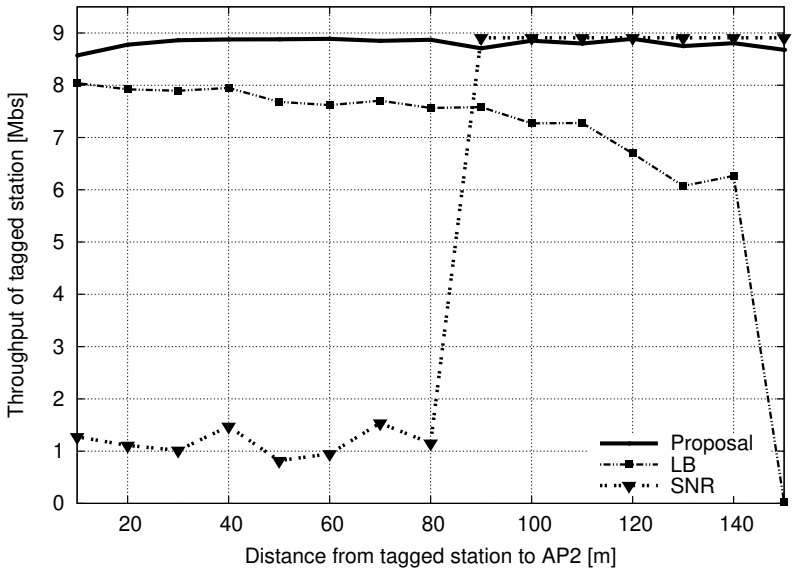


Figure 3.8 Performance of different AP selection schemes in a scenario with two AP, for different values of distance from tagged station to AP2, while distance to AP1 is fixed as well as load on both APs

tool. In the first part of the experiment, the tagged station monitors the wireless channel and gathers the measurements in PCAP traces using its wireless interface in the monitor mode. In the second part of the experiment, the tagged station connects to the AP, performs a TCP file download and measures the obtained throughput using the Iperf tool.

We run 48 different experiment configuration changing N and t; for each configuration we run experiments using 8 different permutations of the available machines from the EXTREME Testbed®, in order to minimize the eventual bias of the measurements on the specific hardware being used. For each setup we run 4 independent repetitions, which results in a total of 1536 samples. One sample represents a set of inputs and its related outputs.

Once the experiments are done, we use the PCAP Trace Parser tool [142] to process the PCAP traces and to calculate the metrics described in Section 3.4. Since the PCAP parser is not providing the beacon delay as output, we developed another software which uses the PCAP trace to calculate the average beacon delay for the whole measurement interval. The Iperf tool provides as output value the measured throughput, so there was no need for additional processing of the throughput measurements. Once all the experimental results are processed, all the obtained data is stored in a MYSQL database, from where it will be fetched for the training and the testing the Neural Network.

The performance evaluation is done in two phases. In the first phase we identify the values of the training parameters that provide the best MFNN performance; in the second phase, we evaluate the actual AP selection scheme using the experimental results just described. These phases are explained in the next subsections.

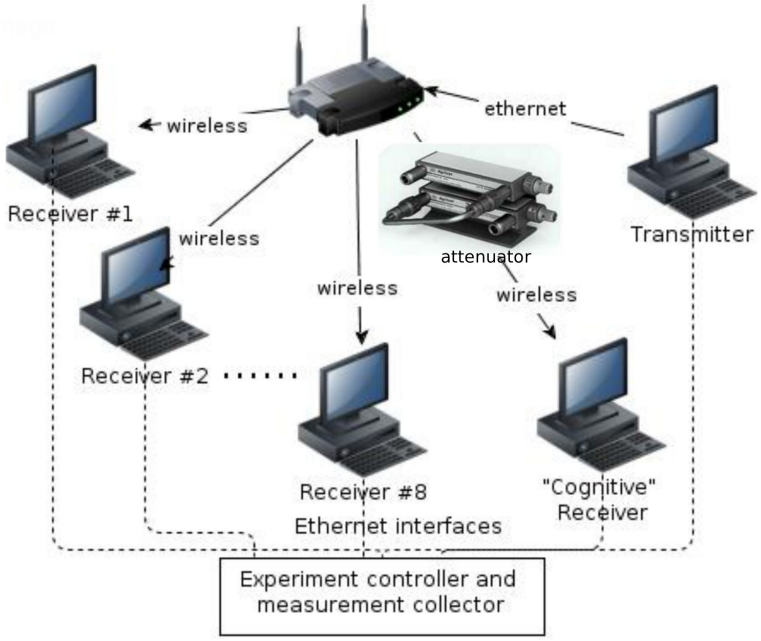


Figure 3.9 Setup of the EXTREME Testbed

Table 3.2 Access point setup

Network interface	802.11g
Operational Rates	1.0, 2.0, 5.5, 6.0, 9.0, 11.0, 12.0, 18.0, 24.0, 36.0, 48.0, 54.0 Mb/sec
Radio Channel	2412 MHz Channel 1
Short Slot-Time	Enabled
Beacon period	51200 μ

3.6.2 Neural Network configuration, training and testing on experimental data

The training of the Neural Network is done based on the experimental results described in Section 3.6.1. The samples are divided in two sets of equal size, where the first is used for the training of the MFNN and the second is used for the testing. For the training, the environmental and throughput measurements are used, respectively, as inputs and output for the Neural Network. In this way, the MFNN learns how the throughput depends on the environmental conditions measured before the connection with the AP is established. For the implementation of the MFNN we used FANN, which is a publicly available software library [144]. Following the recommendations from [145], we set up FANN to use the iRPROP- batch training algorithm [146]. By testing the different combinations of the metrics defined in Section 3.4 we get the architecture of the neural network based machine learning engine which is shown in Figure 3.10.

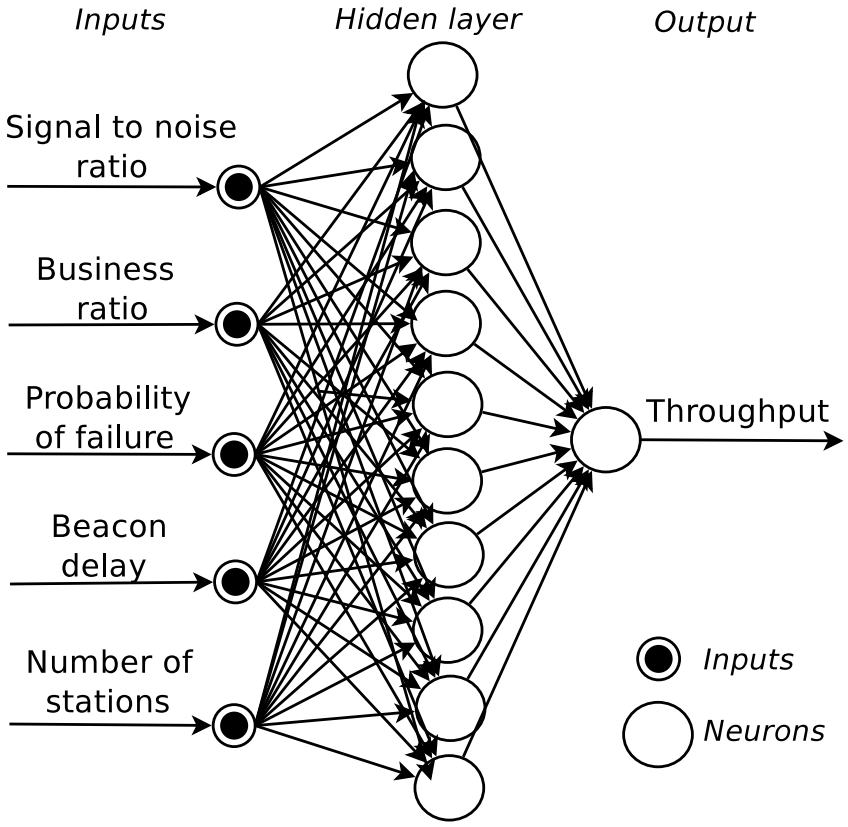


Figure 3.10 Architecture of the MFNN with five inputs, a hidden layer and one output

Similarly to the procedure described in Section 3.5.2, the goal is first to determine the MFNN configuration parameters, $H \in \mathcal{Z}$, $E \in \mathcal{Z}$, and $L \in [0, 1]$ for which it has the best learning performance. In Figure 3.11, we report the Normalized Root Mean Squared Error (NRMSE) of the estimated throughput. RMSE is the Root Mean Squared Error calculated on the whole testing set. As evident from the figure, and similar to what we have seen already in Section 3.5.2, the most important parameter for the training of the MFNN appears to be E, while H and L do not impact so much the learning performance of our Neural Network model. The best performance of the MFNN is achieved for $E = 70$, for which the NRMSE reaches its minimum value equal to 0.0835. For values of E lower than 70, the error is bigger because the MFNN does not have enough knowledge yet. Conversely, for values higher than 70, the MFNN becomes over-trained and too specialized on the training data set; in other words, it looses generalization, which is needed for the good interpretation of the testing data set. Since the MFNN provides the best estimation for $E = 70$, we will use this configuration for the training of the MFNN for the AP selection scheme. Comparing to the MFNN model that we have created in the simulator, the MFNN based on the experimental data achieves a much higher predictive accuracy. We attribute that to having more information about the environment thanks to an additional input metric based on the beacon delay, which in the case of the experimental environment appeared to have more useful information than in the simulation environment.

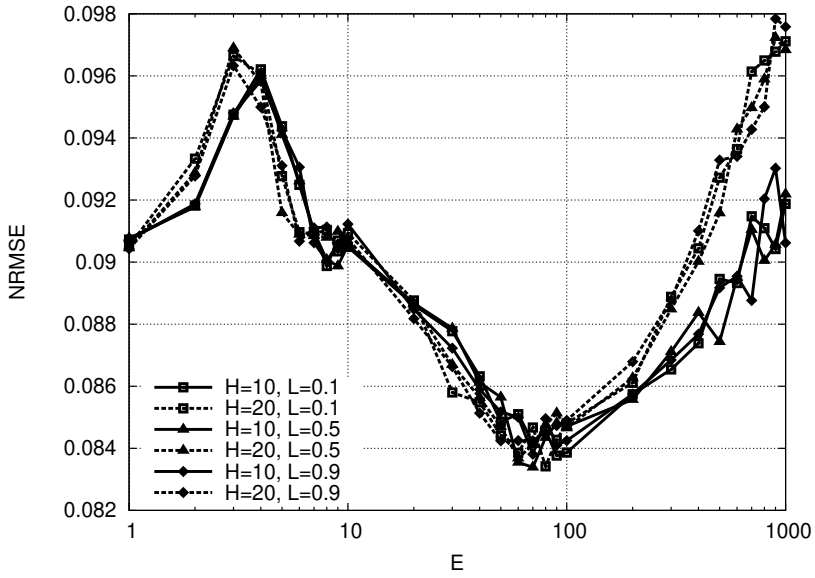


Figure 3.11 Performance of the MFNN based estimation of the AP throughput

3.6.3 Performance evaluation of machine learning based AP selection scheme in experimental environment

For the performance evaluation of the machine learning based AP selection scheme we consider a scenario in which the mobile station is in coverage of two APs, respectively AP1 and AP2. In general, the APs have different numbers of background stations, respectively N_1 and N_2. The values of the signal to noise ratio that the tagged station receives from the APs are, in general, different and respectively denoted as γ_1 and γ_2. Since running experiments with two APs is cumbersome due to the many combinations of N_1, N_2, γ_1 and γ_2, we recreate this scenario artificially by using the Cartesian product of two subsets of the results obtained from the experiments with a single AP. Let S be the result set of the single AP experiments, and let $s = (N, \gamma, r)$ be the generic element of S, where r is the repetition identifier. Then, we define two subsets of S in the following way:

$$S_1 = \{s : N_{1min} \leq N_1 \leq N_{1max}, \gamma_{1min} \leq \gamma_1 \leq \gamma_{1max}, r < r_{thr}\} \tag{3.6}$$

$$S_2 = \{s : N_{2min} \leq N_2 \leq N_{2max}, \gamma_{2min} \leq \gamma_2 \leq \gamma_{2max}, r \geq r_{thr}\} \tag{3.7}$$

The sample from the Cartesian product $S_{(1\times2)} = S_1 \times S_2$ is the ordered pair (s_1, s_2), where s_1 and s_2 are respectively, elements from S_1 and S_2, and represent the performance of AP1 and AP2 respectively. In Table 3.3 the subsets constraint values are shown for the different scenarios with 2 APs that we consider.

3.6.3.1 Experimental evaluation scenario 1 (Varying N_2)

In order to analyze the performance of the machine learning based scheme for the different load of the APs, we consider the experiments with a fixed N_1, and with γ_1 and γ_2 varying in narrow

Table 3.3 Experiments configuration settings

	Scenario name		
	Varying N	Varying γ	General
$N_{1min} - N_{1max}$	5-5	1-1	1-8
$N_{2min} - N_{2max}$	1-8 (N)	4-8	1-8
$\gamma_{1min} - \gamma_{1max}[dB]$	50-60	50-55	35-60
$\gamma_{2min} - \gamma_{2max}[dB]$	35-50	35-60 (γ)	35-60

intervals, while we vary N_2 for all the values of N. The settings for all the experiments which are run for this scenario are in Table 3.3. We compare our solution with the scheme based on the SNR, the load-based scheme and the beacon delay scheme. The load-based scheme uses the business ratio as selection criteria, while the beacon delay scheme uses average beacon delay. Figure 3.12 shows a comparison of the performance of these AP selection schemes for the scenario described above. The X axis represents N_2, and the Y axis represents average throughput, which is obtained by the tagged station using each AP selection scheme. Since AP2 offers a stronger γ, AP2 is expected to be a better choice when $N_2 < N_1$, because it is less loaded and closer. When $N_2 > N_1$, AP1 is expected to be a better choice because it is less loaded. Thus, the expected behavior for all schemes is that the throughput obtained by the tagged station decreases with an increase of the load on AP2. In this scenario, since the SNR scheme neglects the load of APs, the fraction of correct decisions remains constant for any value of N_2. The load-based AP selection, which performs selection using the business ratio, in this scenario has good performance, but still, in many cases, it makes wrong decisions because it does not consider γ. The AP selection scheme based on the beacon delay has similar performance to the load-based scheme, but for $N_2 > N_1$ in many cases it selects the wrong AP, because the average beacon delays for the AP1 and the AP2 start to have very similar values. Our scheme gives a very good performance for all values of N_2. Due to possible errors in predicted values described in Section 3.6.2, the machine learning scheme may sometimes select a wrong AP, especially when both APs have very similar load and when a similar γ values for different APs are detected by the station. However, the machine learning based scheme shows an overall robust behavior and it chooses a correct AP with much higher probability than the other schemes as is shown in Figure 3.14, in which this scenario is denoted as "Scenario1".

3.6.3.2 Experimental evaluation scenario 2 (Varying γ_2)

The second scenario is characterized by a fixed N_1, N_2 and γ_1 varying in a narrow interval, while γ_2 is varying in the whole interval of γ. We choose this scenario to evaluate the performance of the machine learning based AP selection scheme when the detected values of γ from AP1 and AP2 are very different. Also, the settings for this scenario are summarized in Table 3.3. Figure 3.13 shows the performance of our scheme, compared with the SNR scheme, the load-based scheme and the beacon delay scheme. As expected the SNR scheme has good performance when $\gamma_1 > \gamma_2$, since it selects AP1 which is less loaded in all experiments. Conversely, for $\gamma_1 < \gamma_2$ the SNR scheme performs worse because it will always choose AP2 which is more loaded. The load-based scheme and the beacon delay scheme provide more stable performance than the SNR; also, those schemes cannot detect high differences between γ_1 and γ_2 which are present in this scenario, and react appropriately to this condition. We

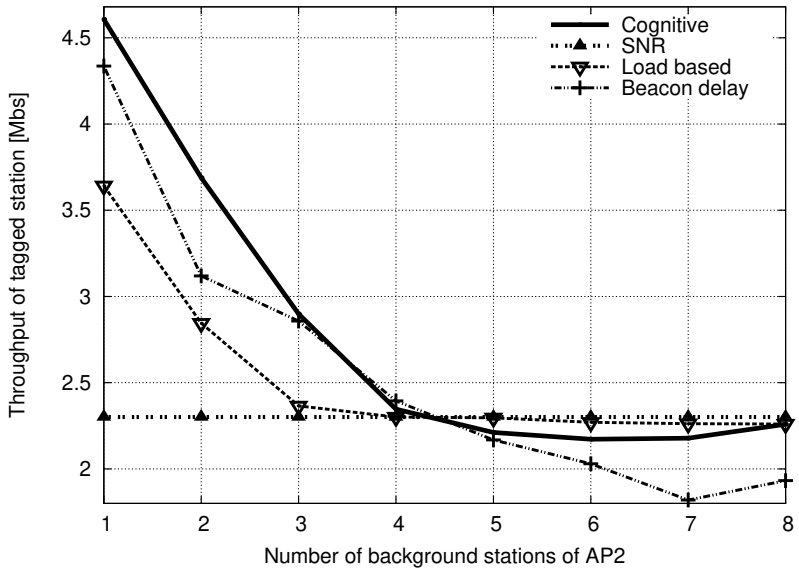


Figure 3.12 Performance of different AP selection schemes in a scenario with two AP only, for different values of N_2

notice that the machine learning based scheme has very good performance and stable behavior in AP selection for all values of γ_2, as we can see from Figure 3.13 and 3.14, in which this scenario is denoted as "Scenario2".

3.6.4 Experimental evaluation on a more general use case

We already discussed that, in cases with a significant difference in N and γ, the machine learning scheme has very good performance. To evaluate the performance in a more generic way we choose the subsets for the Cartesian product so that we involve all possible scenarios. To achieve this we do not enforce any constraint on N_1, N_2, γ_1 and γ_2, as we show in Table 3.3. We use r to divide results of the single-AP experiments in two exclusive subsets, from which we randomly select a limited number of experiments (1000) which will be used in the Cartesian product. As a result, we get 100000 two-AP experiments for this scenario. In Figure 3.14 we can see that in this general scenario, the machine learning scheme performs better than the other schemes, with an average of +15% correct decisions.

3.7 Conclusions

In this chapter, we have treated AP selection in Wi-Fi networks. Our scheme aims at selecting the AP that is expected to yield the best throughput according to the past experience. We have focused on NN based solutions, which allow to adapt the decision based on the status of the environment and the evolution of performances. After designing the NN, we build a realistic scenario in a high fidelity network simulator, on ns-3, and then we validate the results through experimentation in our group's testbed. The simulation results show that the proposed machine learning based AP scheme can achieve significant improvements in

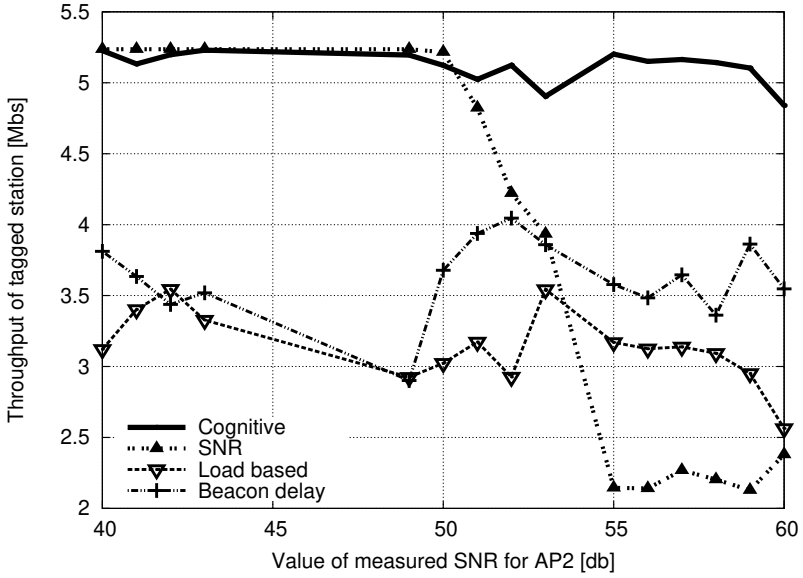


Figure 3.13 Performance of different AP selection schemes in a scenario with two AP, for different values γ_2

throughput performance compared with state-of-the-art decentralized AP selection schemes in a variety of scenarios. The validation of the proposed scheme through an experimental platform confirms that the proposed NN based AP selection scheme can achieve significant performance improvements with respect to state-of-the-art schemes, especially in scenarios with high variations in environmental data. The main advantage of the proposed scheme with respect to state-of-the-art schemes is that it is more robust as it copes much better with the complexity of today's communication systems and the variability and unpredictability of real-world scenarios.

In real devices, learning from the past experience is expected to be implemented either by using experience gathered in the laboratory (e.g., training the device before it is sold) or gathering experience during the ordinary usage in real-life scenarios. As we could see in Sections 3.5.2 and 3.6.2, it is essential to provide a priory training, but there is a risk of entering into the overfitting learning region. In this sense, the first approach is better since the AP selection scheme relies on prior knowledge since its first usage. However, it is important not to over-train since it would lead to a very bad performance in the realistic environment, which includes many non-ideal phenomenona that are not present in the laboratory conditions. This performance could be even worse than if the scheme would not have any prior knowledge, i.e., sometimes it is worse to have over-trained than under-trained Neural Network, e.g., see Figures 3.5 and 3.11. On the other hand, in the second approach, knowledge is gathered from real-life scenarios, but since the mobile station starts using the AP selection scheme without having prior knowledge, i.e., it is under-trained at the beginning, it will likely make many wrong AP selection decisions until it gathers enough experience. Hence, we suggest that the best approach would be to combine the two approaches, i.e. to install the AP selection scheme on the mobile station with a machine learning engine that is pre-trained in the laboratory, and then to enhance the knowledge and prediction accuracy during real-life usage.

Regardless of the AP selection scheme being used, the duration of the measurements window influences the metrics accuracy and hence the performance of the AP selection scheme. To

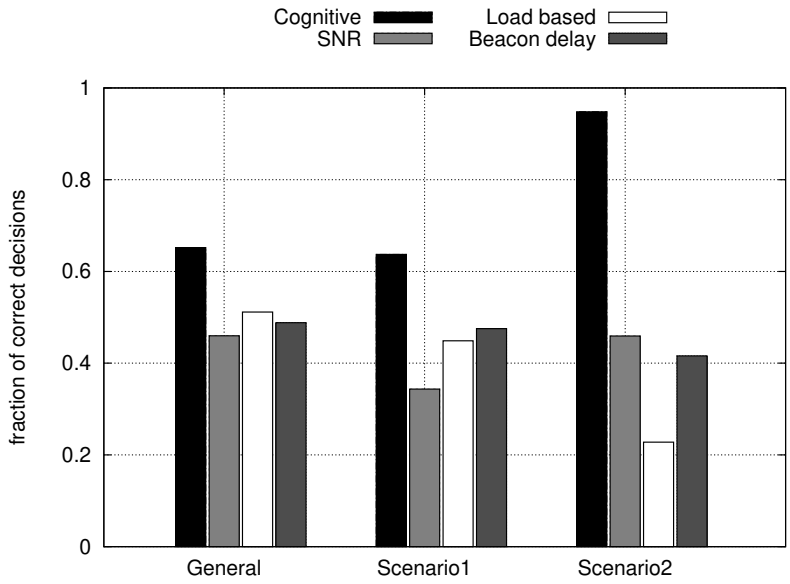


Figure 3.14 Histogram for different algorithms

achieve accurate metrics on which to base the AP decision, especially for the load-based schemes [133, 134], the measurement time is typically set to a few seconds. This duration is at least 3 orders of magnitude higher than the time necessary for the MFNN to provide its prediction. Similarly, the time needed to accumulate the past experience is orders of magnitude greater than the time needed to run the iRPROP batch training algorithm. For this reason, we consider that the computational time of using an MFNN for AP selection is negligible and thus this solution is viable for the real-world devices.

However, a drawback of MFNNs is that their output is a predicted value whose confidence is not known. It also stays quite uncertain and unknown what is the impact of each of the input variables on the prediction accuracy. For the scenarios in which we would like to provide some QoS guarantees, e.g., in LTE call admission control use case, and in which we would like to measure the impact of each variable with more insight than just an average RMSE; it would be interesting to find a technique that would provide not only a predicted value, but rather a probability distribution of the output value, and that would allow providing some guarantees related to QoS requirements. In the next chapter, we propose to use a graphical model, a Bayesian Network model that represents the probabilistic relationships among a set of input and output variables, and we apply it to the LTE call admission control use case, in which different users can have, in general, different QoS requirements. We then provide a comparative study of neural network and a probabilistic graphical model.

Chapter 4

RRM in LTE/NR: ML based CAC

In this and the following chapter, we focus on the radio resource management from the single-cell optimization perspective, i.e., the goal is to optimally utilize the available radio resources focusing on a capacity per cell. In this context, we consider two closely related radio resource management tasks, Call Admission Control (CAC) and Medium Access Control (MAC) scheduling in LTE/NR systems. These two functionalities are responsible for the QoS of the calls from their establishment until their termination. In this chapter, we focus on the CAC, while in the following, Chapter 5, we focus on the MAC scheduling functionalities.

The task of CAC is to admit or reject the establishment requests for new radio bearers (voice/video call, file transfer, gaming traffic, etc.). To achieve this, CAC should take into account many variables: the overall resource situation (estimated cell capacity and current cell load), the QoS requirements (defined by QCI) for the already admitted calls, and for a new call, the actual (measured) QoS of ongoing calls, the channel quality for ongoing calls (i.e., in terms of interference), etc. All these variables together determine the real instantaneous cell capacity (i.e., in terms of the number of calls that can be served while providing the specific QoS requirements), and also determine whether a new call can be admitted or not. The goal of CAC is to maximize the number of accepted radio bearer requests and at the same time to ensure proper QoS for all ongoing calls (i.e., by rejecting radio bearer requests when they cannot be accommodated).

CAC can be described as a radio resource management optimization problem, in which there is a trade-off between the QoS level perceived by the user (e.g., in terms of the call dropping probability, or transmission rate, delay, packet error loss rate and jitter), and the optimal utilization of wireless radio resources. The performance of CAC is typically evaluated through call blocking and call dropping probabilities, transmission rate, packet delay parameters, and

loss rate. The distinction between call dropping and blocking is important because dropping a call in progress is much more annoying than blocking a new call request, and this is especially the case for the voice calls [151]. Because of this, ongoing voice calls, for example during the handoff process, are typically given higher priorities than new incoming calls.

In this chapter, we propose a CAC scheme that can take into account the dynamics of a variety of parameters and variables (traffic, QoS requirements, mobility, etc.) in order to maximize the user's perceived QoS. Furthermore, this chapter provides a comparative study of the application of machine learning and statistical learning approaches, respectively, Neural Networks (NN) and Bayesian Networks (BN), for the CAC task. This chapter also provides the peformance comparison of these two approaches. We study a relevant research topic in the field of mobile communications: application of machine learning and statistical learning approaches for the problem of CAC in LTE systems, that can be easily generalized and applied to the future 5G NR systems. CAC in general in cellular networks deals with an automatic decision whether a call will be served or blocked while trying to avoid the droppings. As such, CAC represents one of the most essential methods for optimal resource management [152]. The CAC decision is determined by studying several dynamic parameters of the network, pertaining to existing network resources and their utilization. The CAC problem is challenging due to the high complexity of the cellular systems, such as LTE and NR, which can be traced back to their advanced features such as adaptive modulation and coding, HARQ, and dynamic packet scheduling with QoS support. All these factors together make the performance of a cellular system very hard to predict due to the users' mobility and to the variations in propagation conditions, as well as in the type and amount of traffic. Because of this, machine learning and statistical learning approaches have a potential to achieve better performance than traditional approaches.

The main contributions presented in this chapter are:

- the design of machine learning and statistical learning-based LTE CAC schemes based on a NN and a BN, respectively, including the parameter learning phase for both models;
- the definition of a new feature extraction model for measurements available in an LTE system used for a learning-based CAC scheme;
- the design and implementation of a realistic simulation scenario representative of an LTE system with heterogeneous types of traffic, realistic propagation conditions, and user mobility;
- and a comparative analysis of the performance, in terms of CAC decision accuracy, of the NN, the BN and a state-of-the-art CAC scheme for LTE;

Section 4.1 gives an overview of the related work. Section 4.2 discusses various CAC design aspects. Section 4.3 details the system model. Section 4.4 describes the feature extraction model for the CAC schemes. Section 4.5 and Section 4.6 describe the NN and the BN models, respectively. Section 4.7 presents the CAC decision schemes for both the NN and the BN models. Section 4.8 describes the experimental setup and the main results. Finally, Section 4.9 summarizes the main contributions.

4.1 Related work

As discussed in Chapter 2 in Section 2.3, wireless communications networks are becoming more and more complex, to the point that their operation and maintenance cannot be sustained without resorting to automation. In this respect, the use of machine learning techniques that can extract useful information from massive amounts of measurement data is a very promising line of investigation. The idea is to use these techniques to realize a cognitive network [138], i.e., a cognition process that spans the whole network and learns from the observation of the environment in order to reconfigure the network parameters to optimize its performance.

In recent literature, different machine learning techniques have been proposed for various applications in the wireless networking context. In particular, the application of supervised learning via NNs was proposed to learn how different system configurations affect communication performance in the presence of different environmental conditions [153]. This approach was shown to be effective in a number of cases, such as access point selection [154] in WLANs, radio admission control in Long Term Evolution (LTE) systems [155], and optimization of cognitive radio systems [156].

In spite of their success, as discussed in Chapter 3, a potential drawback of NN is that their output is a predicted value whose confidence is not known. In this respect, a machine learning technique that could provide not just a crisp value as output but rather a probability distribution of the value of interest seems more promising. A popular mathematical tool of this kind is BN. BN is a graphical model that is used to represent the probabilistic relationships among a set of variables. In a machine learning context, this tool is used to carry out statistical inference in a computationally efficient way, e.g., to predict the probability distribution of a certain variable conditioned to the known value of some other variables. An example of the application of this approach in a wireless networking context was presented in [157], where the particular application to CAC in WLANs was considered.

Several CAC schemes have already been proposed for LTE systems. In [158] a scheme based on modeling the call arrival process with the queueing theory is proposed, and the concept of resource reservation is applied to this scheme. In the presence of an incoming call, extra resources are reserved to avoid QoS degradation. The amount of these extra resources is determined a priori based on the knowledge of the user mobility patterns. The major problem is that LTE is expected to be used with a mixture of heterogeneous cells of different sizes (macro/micro/pico/femto cells), deployed in a loosely coordinated fashion, with minimum to no planning. In such conditions, the statistics of the mobility pattern of the users are expected to vary significantly among different cells and cannot be known in advance; furthermore, overcoming these variations by a conservative estimation of the extra resources to be reserved would lead to poor resource utilization. Other CAC schemes for LTE follow a ring-based modeling approach [159, 160]. This modeling approach assumes that users belonging to a cell can be grouped into rings according to their distance from the base station and that the users located in the same ring consume the same amount of radio resources. Even though these schemes show significant performance improvements compared to the previous CAC models for LTE, they still rely on a priori modeling of the radio environment, which is not realistic. In particular, in the presence of frequency selective fading, as well as path loss variations due to obstacles such as walls and buildings, the prediction of the needed amount of resources based solely on the distance between users and the base station may be highly inaccurate.

In other words, the vast majority of previously proposed CAC algorithms for LTE are based on some a priori known analytical models to predict the variations in resource utilization. Thus, they may fail as soon as these models do not match the actual deployment conditions, which is, in fact, expected to occur in the vast majority of LTE deployment scenarios. For this reason, an LTE CAC scheme relying on a learning-based approach, which can react to the actual conditions faced by each cell, looks more promising. In [152], the authors formulate the CAC as a constrained optimization problem and solve it by leveraging a Genetic Algorithm. The advantage of this algorithm is that it gives a complete view of the network and considers various dynamic network parameters. However, there is no differentiation among calls belonging to different QoS classes; hence no QoS level of management is supported. Additionally, the method requires having a central unit that makes decisions for the whole network. In preliminary work [155], we considered machine learning scheme based on NNs. In that initial work, we considered a scenario without mobility and with simplified propagation conditions, thus not realistic. In this chapter, we go significantly beyond that previous work by considering a much more realistic scenario with user mobility and frequency selective fading, by improving the performance of the NN-based LTE CAC scheme, and by additionally designing and evaluating a new LTE CAC scheme based on BNs.

4.2 Design of call admission control scheme

The CAC [1] is located at the radio resource control entity in Layer 3 (RRC) of the LTE/LTE-A protocol stack (i.e., alongside handover management). The purpose of the CAC scheme is to decide in an intelligent way whether a new radio bearer admission request should be accepted. As such, CAC represents an essential part of the handover process.

CAC scheme design involves various design options and each of them brings certain advantages and disadvantages. In the following, we discuss some of them that are highly related to the design choices that were made in the design of the proposed CAC schemes in this thesis.

CAC schemes can be either static or dynamic [152]:

- *Static* or *deterministic* CAC: QoS parameters are guaranteed with 100% confidence. Typically, these schemes require extensive knowledge of system parameters (i.e., such as some a priori knowledge about traffic and users' mobility) and they tipically dedicate the scarce radio resources to guarantee the deterministic QoS requirements.
- *Dynamic* or *stochastic* CAC: QoS parameters are guaranteed with some probabilistic confidence, by relaxing QoS guarantees. These schemes are more flexible and can achieve on average much higher utilization than deterministic approaches. Obviously, the research work is in the domain of dynamic CAC schemes.

From the optimization point of view, CAC schemes can be *optimal* or *subobtimal* [151]. The optimal CAC schemes are always preferred, but they are hardly feasible to achieve in practice in nowadays highly dynamic, complex and heterogeneous wireless systems. On the other

[1] CAC is sometimes refered to as radio admission control (RAC). However, we prefer to use in this thesis CAC since it is a more general term in the literature.

hand, suboptimal solutions are more realistic to achieve in most cases. As such, heuristics and intelligent techniques are widely used to implement a CAC scheme [151].

Yet another design choice to make is the number of services/classes for which they are designed to, and depending on it a CAC scheme can be either *single-service* or *multi-service* [151]:

- *Single-service* CAC: A lot of research work focuses on single-class CACs, which has been dominant in the first and second generation (1G and 2G) of cellular wireless networks when voice service was the main (and the only) offered service.
- *Multi-service* CAC: With the arrival of data and multimedia services in 3G, 4G, and 5G, single-class CAC schemes were no longer sufficient, and as a result, multiple-service/class CAC schemes were needed. The design of multiple-service/class CAC schemes adds more complexity to the CAC, because it has to consider service prioritization based on QCI, while still taking care of other aspects such as perceived QoS, radio resource utilization, fairness, etc. The CAC schemes that are investigated in the scope of the present thesis fall into the multi-service category.

When designing RRM schemes, one of the main design choices to make is whether the scheme will be centralized, distributed, or hybrid:

- *Centralized* CAC: In centralized schemes, a single entity controls the admission in the whole network. Central CAC schemes tend to perform better (thanks to the global information availability) in terms of block and drop rates, fairness, transmission rate, but their implementation tends to be complex because the entity and procedures must be implemented, which can make them not work well in a multi-RAT environment and could have a lot of control overhead signaling. Moreover, too much signaling may add up additional delays, which may not be appropriate for some QoS types of traffic that are bound by very tight latency requirements, e.g., ultra-reliable low-latency traffic in NR 5G should be below 1 ms. Due to their complexity, centralized schemes are not often used in real wireless networks.
- *Distributed* CAC: Distributed CAC is performed in each base station. Future cellular networks will be overlaid with densely deployed small cells, thus the frequency of UEs traveling across many small cells will be very high, especially for high-mobility UEs. Such high-density environments require a more efficient CAC scheme, in which the delays caused by control signals are minimum. To achieve this, the base station should be able to decide on its own.
- *Hybrid* or *semi-centralized* CAC: Different hybrid mechanism can be made to join the best of the distributed and centralized CAC schemes. E.g., decisions for URLLC could be made in a distributed fashion, while some parts of functionality could be centralized to leverage global knowledge and based on it to try to avoid some well-known issues, such as ping pong effects.

All in all, a dynamic distributed multi-service admission control scheme with low complexity is required for complex systems such as LTE/LTE-A and NR [161], and this thesis focuses on that type of CAC schemes.

4.3 System Model

We consider an LTE/LTE-A evolved packet system (EPS) with one base station (eNB) and N users (UEs). According to the LTE/EPS specifications, an EPS bearer defines the QoS requirements of a particular class of traffic for a UE, and all the packets transmitted to and from the UE should meet the QoS requirements of the EPS bearer they are mapped to. EPS bearers are classified into two main categories: guaranteed bit rate (GBR) and non-guaranteed bit rate (non-GBR). For each GBR bearers, a certain value of bitrate is specified, which should be guaranteed by the LTE/EPS system. On the other hand, non-GBR bearers are served on a best effort basis. Requirements other than the bitrate are normally defined by standardized QoS class identifier (QCI) values, which correspond to a pre-defined set of requirements.

For the LTE radio interface, we consider a frequency division duplex with a system bandwidth of S resource blocks (RBs). At each transmission time interval (TTI), the MAC scheduler located at the eNB decides which RBs are dedicated to the transmission of data belonging to each bearer, with the aim to satisfy the QoS requirements of the bearer. Clearly, meeting the QoS requirements for all bearers is only possible if the amount of radio resources available is sufficient in the current conditions. These conditions include the amount of data to be transmitted, and the number of bytes which can be transmitted within an RB to/from each UE, as determined by the adaptive modulation and coding (AMC) functionality. The AMC functionality depends on the propagation conditions, as well as on the distance between the UE and the eNB. The role of the CAC functionality is to determine if a new EPS bearer can be activated or not, i.e., if the available resources are sufficient to satisfy the QoS requirements of all active bearers plus the new one, or alternatively if it is better to reject the activation of the new bearer in order to preserve the QoS of the previously activated ones [162].

4.4 Feature extraction for CAC

In our model, without loss of generality, we consider that there is one EPS bearer per UE. We have N_v GBR UEs and N_b non-GBR UEs. Each GBR UE is performing a VoIP call, while each non-GBR UE is transmitting a bulk file via a TCP connection. We measure the quality of service of the VoIP calls by using the R-factor defined by the E-model [163].

The R-factor after the acceptance (aa) of the incoming call, for every GBR UEs $n = 1, ..., N_v$, is named $R_n^{(aa)}$. This quantity depends on the future performance of the network, so it can not be measured at the moment of the CAC decision, but it can be estimated with our NN and BN models. According to the QoS constraints, all the GBR UEs should have at least a minimum R-factor, namely τ_R. The eNB should accept the incoming call if $R_n^{(aa)} \geq \tau_R$ for all the GBR UEs, and drop the incoming call otherwise. Thus, we define the R-factor for the system as:

$$R^{(aa)} = \min_n (R_n^{(aa)}) \ . \tag{4.1}$$

If $R^{(aa)} \geq \tau_R$, the QoS of all GBR UEs is satisfied even after the acceptance of the incoming call. On the contrary, if $R^{(aa)} < \tau_R$, then at least one UE does not receive a sufficient QoS. If it is possible to predict this event, the CAC scheme should drop the incoming call, thus preserving the QoS for the other GBR UEs.

In the following, we select a set of metrics that are representative of the network conditions, and that will be used as the inputs to the cognitive models to predict $R^{(aa)}$ at the moment in which the CAC decision should be made.

One of the most important factors that affect the capacity in the LTE system is the radio resource allocation scheme. Radio resource allocation in an LTE system is not specified by the 3GPP standard, and it is implemented by the operators. Thus, the ratio of resources that are assigned to GBR and non-GBR UEs depends on the implementation of the scheduler at the MAC layer. In our system, we model this radio resource allocation scheme as a black box. The performance of this scheme can be observed through the variables that describe the radio resource allocation over time, such as the ratio of resources that are assigned to GBR and non-GBR UEs, or the ratio of radio resources that are not assigned to any UE. Thus, we define the metric $\phi_v(bd)$ which represents the ratio of resource blocks that are dedicated to the GBR UEs during the last J time intervals before the CAC decision (bd):

$$\phi_v^{(bd)} = \frac{1}{J}\sum_{j=1}^{J} K_v^{(j)}/K \ , \tag{4.2}$$

where $K_v^{(j)}$ is the number of resource blocks assigned to GBR flows in the interval j, while K is the total amount of available resource blocks. We also define the metric that describes the amount of resources that are allocated to non-GBR UEs in the following way:

$$\phi_b^{(bd)} = \frac{1}{J}\sum_{j=1}^{J} K_b^{(j)}/K \ . \tag{4.3}$$

Both $\phi_v^{(bd)}$ and $\phi_b^{(bd)}$ are providing information on the resource allocation, not on the actual MAC layer throughput. Furthermore, in case the resources for GBR UEs are not sufficient, a fraction of the $\phi_b^{(bd)}$ can be allocated to GBR UEs.

Another important variable to describe the QoS of the system before the CAC decision is the R-factor calculated at the moment of the decision for every GBR UE $n = 1, \ldots, N_v$, i.e., $R_n^{(bd)}$. As a feature for the CAC scheme we consider:

$$R^{(bd)} = \min_n(R_n^{(bd)}) \ . \tag{4.4}$$

Since the performance of non GBR UEs may also be relevant to predict the value of $R^{(aa)}$, we consider $P_n^{(bd)}$ (with $n = 1, \ldots, N_b$), the successful packet rate for non GBR UEs, and we define another feature for the CAC scheme as:

$$P^{(bd)} = \min_n(P_n^{(bd)}) \ . \tag{4.5}$$

Another variable that has a high impact on overall system capacity is the channel quality indicator (CQI). The CQI is an indication of the data rate that can be supported by the channel for a certain UE. It depends on the signal to interference plus noise ratio (SINR) as well as on the characteristics of the UE receiver [164]. The CQI value is reported by each UE to the

eNB and is used by the eNB to determine the modulation scheme and the code rate that can be supported by the channel for the downlink data transmissions. Since the network environment is changing quite fast in a realistic scenario, due to the mobility of the UEs, the CQI values may significantly vary over time. E.g., in case of wrong estimation of the code rate, packet losses, and high delays for the packets received may occur. This would affect the perceived QoS. For these reasons, it is important to include some features which represent the variations in channel quality. Thus, we define a variable that represents the level of variations of the perceived channel quality by all UEs along all their subbands. For each UE $n = 1, \ldots, N$, the eNB node collects the past U values of the CQI received from that node, $CQI_n^{(u)}$ with $u = 1, \ldots, U$, which are referred to the past J time intervals and to S different subbands, with $J \cdot S = U$. With these values, the standard deviation of the CQI for each UE n can be calculated as:

$$\sigma_{CQI_n}^{(bd)} = \left(\frac{1}{U} \sum_{u=1}^{U} \left(CQI_n^{(u)} - \frac{1}{U} \sum_{u_1=1}^{U} CQI_n^{(u_1)} \right)^2 \right)^{1/2} . \tag{4.6}$$

By using the previous expression, we define the feature representing the CQI variation over all UEs as:

$$\sigma_{CQI}^{(bd)} = \frac{1}{N} \sum_{n=1}^{N} \sigma_{CQI_n}^{(bd)} . \tag{4.7}$$

Another important feature related to the channel quality is the average CQI among all UEs, defined as:

$$m_{CQI}^{(bd)} = \frac{1}{N} \sum_{n=1}^{N} \left(\frac{1}{U} \sum_{u=1}^{U} CQI_n^{(u)} \right) . \tag{4.8}$$

4.5 Neural Network model

A NN is an adaptive system that changes its structure based on external or internal information that flows through the network. This adaptive system can be applied to different optimization problems in telecommunications, like pattern recognition, function approximation or classification. A NN model is adopted to infer the future performance of the system as a function of the available network measurements. We use a Feed Forward NN (FFNN) with two layers of adaptive weights. We can write the NN function [101] as:

$$y(\mathbf{x}, \mathbf{w}^{(1)}, \mathbf{w}^{(2)}) = f \left(\sum_{h=1}^{H} w_h^{(2)} f \left(\sum_{m=0}^{M} w_{hm}^{(1)} x_m \right) + w_0^{(2)} \right) , \tag{4.9}$$

where $\mathbf{x}$ is the vector of the inputs x_m, with $m = 1, \ldots, M$, and $x_0 = 1$; H is the number of nodes in the hidden layer; $\mathbf{w}^{(1)}$ is the vector of adaptive weights $w_{hm}^{(1)}$ for the hidden layer, with $h = 1, \ldots, H$; $\mathbf{w}^{(2)}$ is the vector of adaptive weights $w_h^{(2)}$ for the output layer. The sigmoidal activation function is defined as:

$$f(a) = \frac{1}{1 + e^{-a}} . \tag{4.10}$$

We select as inputs of the NN the set of metrics that are observable and available at the eNB, i.e.,

$$\mathbf{x} = \{N_v, N_b, m_{\text{CQI}}^{(bd)}, \sigma_{\text{CQI}}^{(bd)}, \phi_v^{(bd)}, \phi_b^{(bd)}, P^{(bd)}, R^{(bd)}\} \ . \tag{4.11}$$

The NN output is the value of $R^{(aa)}$ obtained as a function of all the parameters in $\mathbf{x}$, which will be used to make the CAC decision, i.e., to accept or not the incoming call.

4.5.1 Neural Network training

NN training is performed in a supervised manner. During the training phase, we collect a training dataset $\mathcal{D}$, i.e., for every tested scenario $t = 1, \ldots, T$, all the metrics in the set $\mathbf{x}(t)$, as well as the value $R^{(aa)}(t)$ of the R-factor after the new call has been accepted. The weights $\mathbf{w}^{(1)}$ and $\mathbf{w}^{(2)}$ of the NN are obtained by minimizing the error function, i.e.,

$$E(\mathbf{w}^{(1)}, \mathbf{w}^{(2)}) = \frac{1}{2} \sum_{t=1}^{T} \| y(\mathbf{x}(t), \mathbf{w}^{(1)}, \mathbf{w}^{(2)}) - R^{(aa)}(t) \|^2 \ . \tag{4.12}$$

To optimize the NN training, we use the improved Resilient back-PROPagation (RPROP) without back-tracking, aka iRPROP-. iRPROP- performs better than traditional RPROP algorithm (the first-order learning algorithm for neural networks) , and it has lower computational complexity than improved RPROP with back-tracking, aka iRPROP+, while having similar accuracy performance [165]. During the training phase, we vary three training parameters to evaluate how they affect the accuracy performance and to empirically select their optimal value: the learning rate, L, the maximum number of epochs, E, and the number of nodes in the hidden layer, H.

In Fig. 4.1, we show the performance of the NN training phase as a function of the three learning parameters. The accuracy performance is represented by the normalized root mean square error, which is defined as:

$$\varepsilon(E, H, L) = \frac{\sqrt{\frac{1}{T} \left(\sum_{t=1}^{T} y(\mathbf{x}(t), \mathbf{w}^{(1)}, \mathbf{w}^{(2)}) - R^{(aa)}(t) \right)^2}}{max(R^{(aa)}) - min(R^{(aa)})} \ . \tag{4.13}$$

4.6 Bayesian Network model

A BN is a probabilistic graphical model [166] that can describe in a compact way the conditional independence relationships among a set of random variables. In our case, such variables are represented by the metrics available at the moment of a decision whether to accept the incoming call, i.e., the same variables selected as inputs of the NN, in the set $\mathbf{x}$, as well as the R-factor observed after the incoming call has been accepted, $R^{(aa)}$. Indeed, during the learning phase of the BN model, we do not discriminate between observed variables and variables to infer, so we consider in the same way all the variables in the set $\mathbf{v} = \{\mathbf{x}, R^{(aa)}\}$.

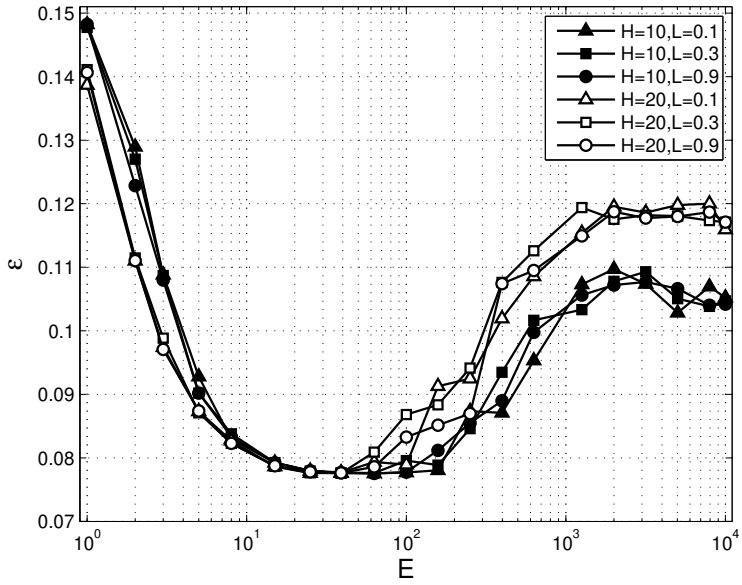


Figure 4.1 Neural network training performance in terms of normalized root mean square error (ε) for different values of training parameters E, H, L.

The probabilistic relationship structure among these variables is represented by a directed acyclic graph (DAG). A DAG is a graphical representation of the conditional dependencies among the variables. It defines the structure of the joint probability among these variables. We use a structure learning algorithm [166] to select the DAG that best represents the probabilistic relationships among the variables, using the samples in the training dataset $\mathcal{D}$, described in Section 4.5. A significant problem with this approach is that the set of all possible DAGs grows exponentially with the number of variables (or nodes) in the DAG. For this reason, to find the DAG that best fits the observed data, we use a local search algorithm, the hill climbing (HC) random search [167], that performs a walk on the space of all the possible DAGs. This algorithm selects a local minimum at each repetition of the algorithm, i.e., a local best fitting DAG. With a sufficient number of repetitions, it is possible, in our case, to reach a solution that is good enough for our inference purposes, i.e., that provides good performance results. In any case, it is not guaranteed that the optimal solution can be reached with a finite number of repetitions of the algorithm.

For every iteration of the structure learning algorithm, the best DAG is chosen among a set of DAGs to be 1) the best representation of the probabilistic relationships among the variables and 2) as simple as possible to allow the BN algorithm to learn the quantitative relationships among the variables from a limited training set. To meet these two goals, we score the DAGs with the Bayesian information criterion (BIC) scoring function, see, e.g., [168]. This criterion assigns a score to each DAG as a function of how well the data in the training dataset $\mathcal{D}$ is represented by the DAG chosen (in terms of maximum likelihood), and it also penalizes based on the number of edges of the DAG, thus favoring more superficial DAG structures.

In Fig. 4.2, we show the BN obtained with the HC random search and the BIC function to score each selected DAG. Observing this DAG, we notice that there exists a set of nodes $\mathbf{z}$ that separates the $R^{(aa)}$ from the rest of the network according to the d-separation rule [166]. This means that if all the values of the nodes in $\mathbf{z}$ are observed, $R^{(aa)}$ becomes independent from

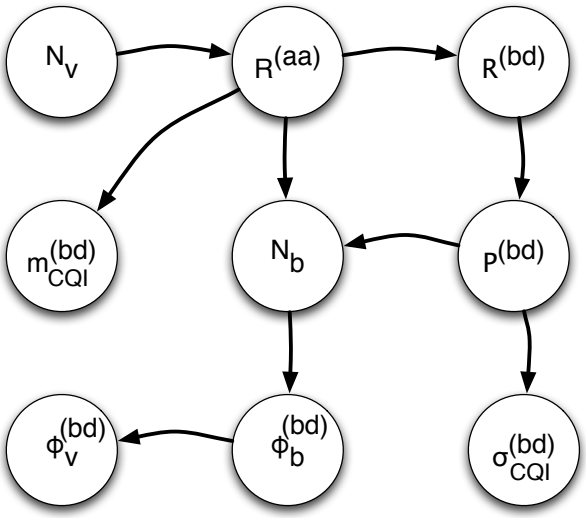


Figure 4.2 Bayesian network showing the probabilistic relationships among all the observable variables in **x** and the variable we want to infer, $R^{(aa)}$.

any other node in the network. This set of nodes is composed of:

$$\mathbf{z} = \{N_v, N_b, m_{\text{CQI}}^{(bd)}, P^{(bd)}, R^{(bd)}\} \ . \tag{4.14}$$

Moreover, these are the nodes (features) observed to make a prediction with the BN.

Regarding the parameter learning phase, we adopt an ML model [101], coherently with the choice of the BIC as a scoring function, to infer quantitatively the probabilistic relationships between the variables of the nodes in **z** and the variable we want to infer, $R^{(aa)}$.

The problem is that it is impossible to learn the probabilistic relationships among the variables based on a finite set of data with such a complicated structure. Usually, some arbitrary assumptions allow the study of the joint probability. In our case, we exploit a structured learning technique to identify conditional independence conditions among the variables. If such conditions exist, it is possible to simplify the DAG by removing some edges (links) and represent the joint probability with a simplified structure. The approach to select the BN structure includes a score based method [166] and a Hill Climbing random search [167], and can be summarized in the following steps:

- we select a valid (and random) BN structure (DAG_i);
- we score this structure based on how well it represents the conditional distribution of the data in the training set, using the Bayesian Information Criterion (BIC) [169];
- we select all the DAGs obtained from DAG_i by adding, removing, or changing the direction of a single edge (link);
- we score each selected DAG and pick the one with the highest score, DAG^*(best fitting with the data);
- if $\text{score}(\text{DAG}^*) > \text{score}(\text{DAG}_i)$, then $\text{DAG}_i \leftarrow \text{DAG}^*$, and the process is repeated until convergence.

This technique can only find a local maximum of the best fitting DAG, so to improve the accuracy, it should be repeated multiple times with different initial DAGs. The optimality can not be guaranteed, but this is a good technique to find local optima, and in our case, we observed that they are close enough to the global optimum.

With the best fitting structure obtained as described above, it is possible to quantitatively learn the probabilistic relationships among the data, using a maximum likelihood parameter learning method. We refer to [170] for further details on the specific choices made to learn a BN from the data and about the learning techniques involved.

4.7 Call Admission Control (CAC) decision

A CAC control scheme should provide a binary output every time a user (UE) starts a new call, i.e., to accept or drop the incoming call. The policy is that a call should always be accepted, except for the case in which the acceptance of a new call may affect the quality of the calls from other UEs. In other words, the eNB should accept a new call if the $R^{(aa)}$ will be above a certain threshold τ_R, and drop the call otherwise. The value of $R^{(aa)}$ depends on the network's performance in the future, so it can only be estimated.

The CAC scheme exploiting the NN approach works as follows. In the presence of an incoming call, the NN estimates the value of $R^{(aa)}$ as a function of all the variables in $\mathbf{x}$, defined in Eq. (4.11). If the estimated value $\widehat{R}^{(aa)} \geq \tau_R$ the incoming call is accepted, otherwise if $\widehat{R}^{(aa)} < \tau_R$ the incoming call is dropped.

The CAC scheme exploiting the BN is slightly different. In the presence of an incoming call, the BN estimates the probability distribution of $R^{(aa)}$ as a function of all the variables in $\mathbf{z}$, defined in Eq. (4.14). Then, if the probability

$$p(R^{(aa)} \geq \tau_R) > p_T \ , \tag{4.15}$$

the call is accepted; otherwise, the call is dropped. The value of p_T is another parameter of the CAC scheme that will be discussed in Section 4.8.

For every incoming call, we should discriminate among four cases:

1. the CAC scheme suggests to accept the incoming call, and the value of the R-factor, calculated a posteriori, is $R^{(aa)} \geq \tau_R$; this corresponds to a correct decision;
2. the CAC scheme suggests to accept the incoming call, but $R^{(aa)} < \tau_R$; this corresponds to an incorrect decision;
3. the CAC scheme suggests not to accept the call; for testing purposes, the system accepts the incoming call, and the R-factor calculated in the case the call is accepted is $R^{(aa)} \geq \tau_R$; this corresponds to an incorrect decision;
4. the CAC scheme suggests not to accept the call; as in the previous case, the call is accepted for testing purposes and $R^{(aa)} < \tau_R$; this corresponds to a correct decision.

In the following, we refer to case 2) as a False Negative (FN). In this case, the CAC scheme erroneously decides to accept the call, but the R-factor after the call has been accepted is $R^{(aa)} < \tau_R$. This means that the scheme can not meet the QoS requirements.

On the other hand, we refer to case 3) as a False Positive (FP). In this case, the CAC scheme erroneously decides not to accept the call, but the R-factor if the call would be accepted is $R^{(aa)} < \tau_R$. Thus, the radio resources are not fully exploited since the CAC scheme drops a call that could be supported by the system.

4.8 Performance evaluation

4.8.1 Scenario and experiment setup

We model a macro cell LTE CAC scenario with the ns-3 simulator [13]. In this scenario, several users (UEs) are connected to a single base station (eNB). To simulate the UEs' mobility, we adopt the steady-state random waypoint mobility model [171], which is implemented in ns-3. With this model, we can simulate random mobility of the UEs in a rectangular area, and the initial distribution of the UEs' positions is the steady-state distribution. For this reason, each simulation run reaches its steady-state in a very short simulation time. The simulation area is square and the eNB is positioned in the center of the square area. As a radio propagation model, we adopt the COST-231 path loss model [172], which is a common choice to simulate macro cell outdoor scenarios. The size of the square area is chosen such that, given the path loss model and the other network parameters, we have a wide range of SINR values. Furthermore, we use a Rayleigh multipath fading model, whose parameters are specified by 3GPP specification [173] for extended pedestrian A model (EPA). Regarding the configuration of the higher layers, we divide the UEs into two categories. The UEs in the first category are performing VoIP calls over a GBR bearer, while the other UEs are performing a TCP file download over a non-GBR bearer. The scenario setup described above is very demanding for the LTE MAC scheduler at the eNB, which has to satisfy the QoS requirements while the network conditions vary significantly as a function of time and frequency due to the mobility and the fast fading. For this reason, we choose the Priority Set (PS) scheduler which is able to successfully adapt to the changes in the channel conditions while providing the guaranteed bit rate to the GBR UEs [174]. Other relevant experimental setup parameters are provided in Tab. 4.1.

In the simulations, we vary the values of the number of GBR UEs, $N_v = 12, 13, \ldots, 17$; the number of non-GBR UEs, $N_b = 0, 2, 4, 6$; and the average speed of the UEs, $v = 0.3, 3, 10, 20$ km/h. These parameters are chosen in order to create scenarios in which the network is close to the limit of its capacity, and an incoming call may effectively deteriorate the overall QoS. For every combination of these parameters we run several independent simulations.

In each simulation, the number of UEs does not change for the first 20 seconds. Then, a new call request event occurs. If the incoming call is accepted, the new UE connects to the eNB and performs a VoIP call during the following 20 seconds of the simulation, competing for spectrum resources with the other active UEs. On the contrary, if the incoming call is not accepted, only the previously active UEs continue competing for the spectrum resources for the second part of the simulation.

Table 4.1 System parameters

Transmission power (dBm)	40
Downlink carrier frequency (MHz)	2120
Uplink carrier frequency (MHz)	1930
Bandwidth (MHz)	1.4
Bandwidth (number of RBs)	6
eNB antenna height (m)	50
UE height (m)	2
Simulation area (m^2)	1560x1560
Simulation time ($seconds$)	40
VoIP codec	G711

4.8.2 Results

The simulation performance of the two CAC schemes exploiting the NN and the BN learning modules are detailed in this section, and it is compared to the performance of a state-of-the-art CAC scheme, i.e., the resource reservation (RR) CAC scheme [158]. The RR scheme works as follows. It measures the fraction f_G of the system resources used by the GBR UEs, and it reserves a fraction $(1 + \alpha_G) f_G$ of the resources for these UEs. The incoming call is accepted only if there are enough available resources. We set $\alpha_G = 0.1$.

In order to calculate the performance of the CAC schemes, we consider two types of CAC decision errors, the FP and the FN, described in Section 4.7. We run $T = 6000$ testing simulations, and we can divide the results of the simulations in T_a cases in which the CAC scheme is accepting the incoming call, and T_d cases in which the CAC scheme is dropping the incoming call, with $T_a + T_d = T$. We define the ratio of FN as:

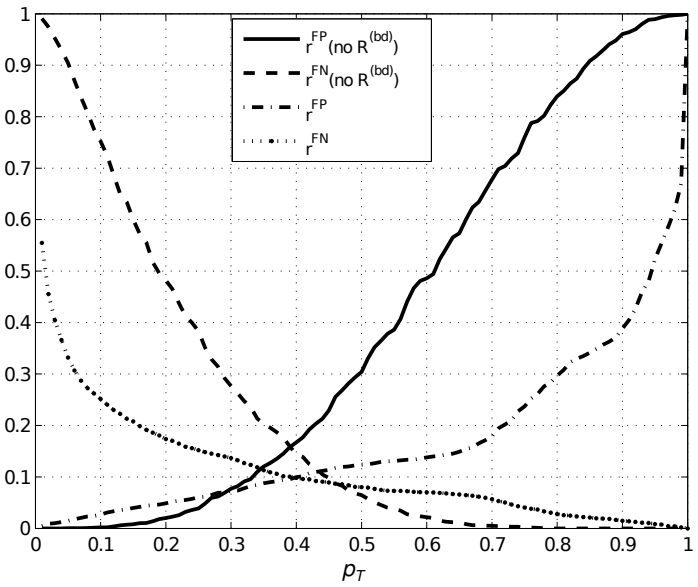
$$r^{FN} = \frac{1}{T_a} \sum_{t_a=1}^{T_a} \infty \left(R^{(aa)}(t_a) < \tau_R \right) , \tag{4.16}$$

where $R^{(aa)}(t_a)$ is the $R^{(aa)}$ relative to the t_a-th simulation with incoming call accepted, and $\infty(\cdot)$ is the indicator function. In a similar way, we define the ratio of FP as:

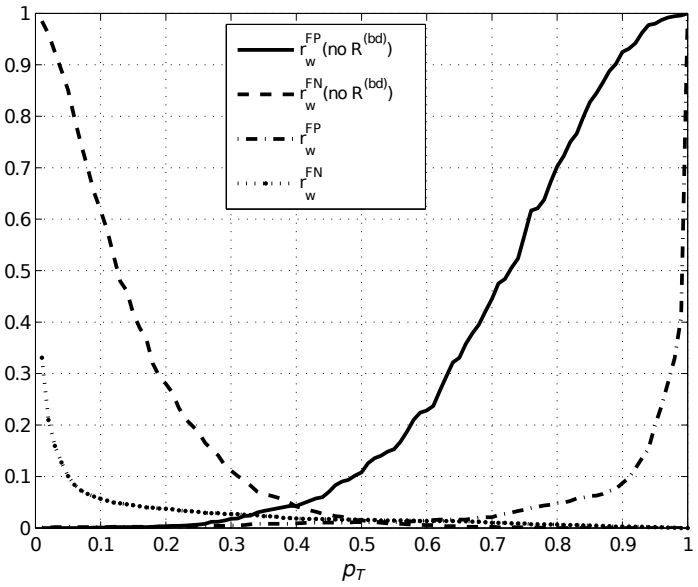
$$r^{FP} = \frac{1}{T_d} \sum_{t_d=1}^{T_D} \infty \left(R^{(aa)}(t_d) \geq \tau_R \right) , \tag{4.17}$$

where for each of the T_d simulations in which the CAC is choosing to drop the incoming call, we need to simulate also the case in which the call is accepted, in the same exact conditions, in order to be able to calculate $R^{(aa)}(t_d)$.

Before comparing the performance of the two CAC schemes, it is important to make another consideration. While the NN is directly predicting the value of the $R^{(aa)}$, the BN is providing a distribution over the possible values of the $R^{(aa)}$. This gives the CAC the additional flexibility to choose the probability threshold introduced in Eq. (4.15). To study how p_T is affecting the performance, in Fig. 4.3-(a) we plot r^{FN} and r^{FP} as a function of p_T, in the case in which the value of $R^{(bd)}$ is available and in the case in which such value is not available to the BN



(a) r^{FN} and r^{FP}.



(b) r_w^{FN} and r_w^{FP}.

Figure 4.3 Ratio of FP and FN occurrences for a BN based CAC scheme, that is receiving or not receiving as input the $R^{(bd)}$, as a function of the probability threshold p_T.

based CAC scheme. We observe that in both cases, there exists an optimal value of p_T to jointly minimize FN and FP. We can observe from the figure that the availability of $R^{(bd)}$ is very important, in particular, because it allows the scheme to jointly keep the r^{FN} and the r^{FP} at a value below 0.15 even for a non-optimal choice of p_T.

In Fig. 4.3-(b) we use another definition of the ratio of FN:

$$r_w^{FN} = \frac{1}{T_a} \sum_{t_a=1}^{T_a} \infty \left(R^{(aa)}(t_a) < \tau_R - \Delta_\tau \right) . \tag{4.18}$$

In other words, we consider as a FN (wrong CAC decision) only the case in which the incoming call is accepted and the $R^{(aa)}$ falls significantly under the threshold τ_R. We have set $\Delta_\tau = 5$. Similarly, we define the FP as:

$$r_w^{FP} = \frac{1}{T_d} \sum_{t_d=1}^{T_d} \infty \left(R^{(aa)}(t_d) \geq \tau_R + \Delta_\tau \right) . \tag{4.19}$$

We notice that in this case, especially when $R^{(bd)}$ is available to the CAC, the error is close to 0 for both FN and FP, and for a wide range of p_T.

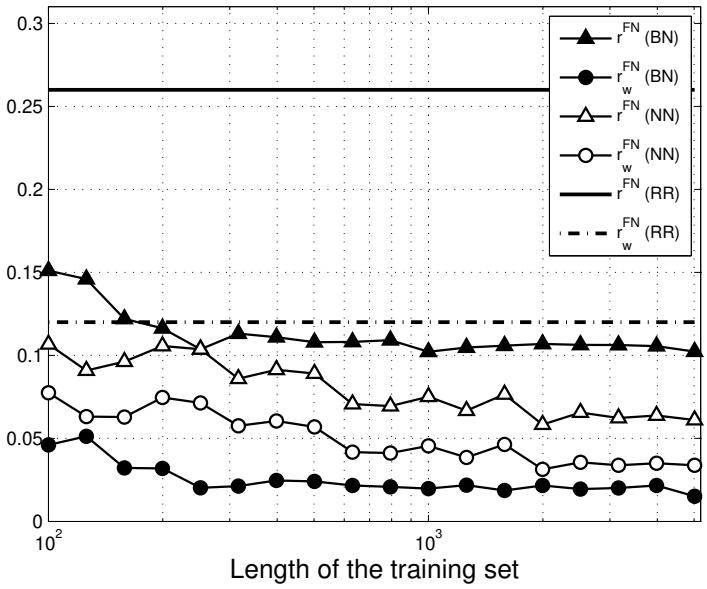
Finally, in Fig. 4.4, we show a performance comparison among the NN, the BN, and the RR CAC schemes. The performance of the NN and the BN changes as a function of the length of the training set available to learn the inference engine, while the performance of the RR scheme does not depend on the training set. In order to evaluate the true performance of each scheme, it is necessary to jointly observe the r^{FN} and the r^{FP}, or r_w^{FN} and r_w^{FP}. In Fig. 4.4-(a) we plot the r^{FN} and the r_w^{FN}, while in Fig. 4.4-(b) we plot the r^{FP} and the r_w^{FP}. We notice that, as expected, the RR scheme performs well in terms of r^{FP}, while it performs poorly in terms of r^{FN}. In other words, it can not balance well between the need to meet QoS requirements, which requires a low value of r^{FN}, and full exploitation of the system resources, which requires a low value of r^{FP}. On the other hand, NN and BN based CAC schemes perform well in both cases. In particular, for the r^{FN} (r^{FP}), the NN CAC scheme outperforms in terms of prediction accuracy the BN CAC scheme. On the contrary, when we evaluate r_w^{FN} (r_w^{FP}), i.e., when we do not consider as an error the case in which $R^{(aa)}$ is close to the threshold τ_R, then the BN CAC scheme has an error close to 2%, and it outperforms the NN CAC scheme.

4.9 Conclusions

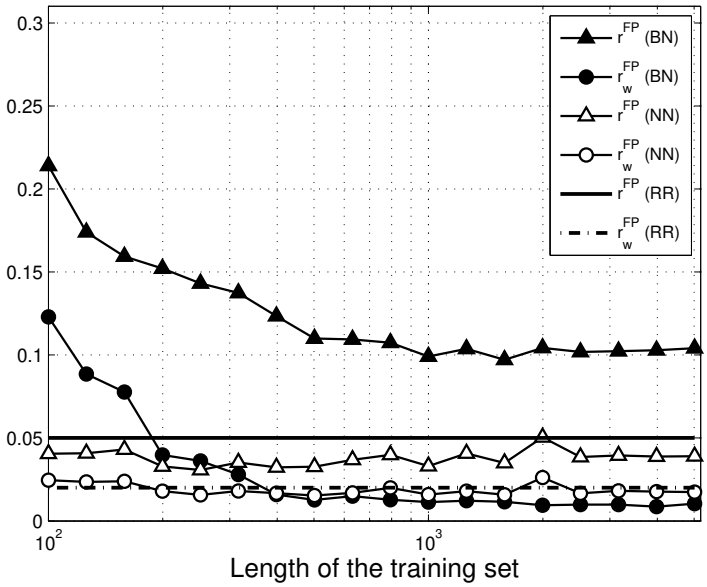
In this chapter, we consider the CAC problem in LTE systems. The CAC scheme needs to decide whether to admit or reject the incoming radio bearer establishment request while trying to maximize the number of accepted radio bearer requests and ensuring proper QoS for new and ongoing sessions.

First, we propose NN and BN based models for CAC for LTE systems and study how best to select the most relevant features that can be observed by the base station. The NN and the BN models learn the network behavior by observing selected features during a training phase. While the NN is directly predicting the KPI value (i.e., $R^{(aa)}$) representing the overall QoS of the new and ongoing calls, the BN is providing a distribution over the possible values of the KPI value ($R^{(aa)}$).

To be able to evaluate the performance of the BN technique empirically, we discriminate among different cases when the CAC scheme suggests to accept or reject the call, and what the correct decision in each of these cases is. We do this through the simulations campaign where for each



(a) r^{FN} and r_w^{FN}.



(b) r^{FP} and r_w^{FP}.

Figure 4.4 Ratio of False Positive (FP) and False Negative (FN) occurrences for the BN, the NN and the RR CAC scheme.

scenario, we obtain measurements that cover both cases: the call is accepted, and the call is rejected. Then, to evaluate the performance of the proposed BN based approach, we calculate the occurrences of the false positive and false negative decisions. In this way, we are able not only to evaluate a single aspect, i.e., the achieved QoS of the VoIP calls but also to inspect the trade-off (among maximizing the resource utilization and meeting the QoS KPIs of calls) and have a more detailed insight into the behavior of the BN scheme for different probability thresholds, i.e., we can see whether the scheme overprotects the system by rejecting the new

calls and thus wastes the radio resources, or instead allows some calls to be accepted that should not be, and thus degrades the performance of all calls. The advantage of the BN based CAC mechanism is its flexibility to choose the probability threshold introduced in Eq. (4.15) based on the desired behavior of the scheme. One of the key conclusions regarding BN is that having all the essential measurements available, such as $R^{(bd)}$, is very important to allow the scheme to jointly keep the r^{FN} and the r^{FP} low even for a non-optimal choice of p_T. In any case, regardless of the BN inputs, we could see that there is an optimal value of p_T to jointly minimize FN and FP.

The results show that the NN CAC scheme outperforms in terms of prediction accuracy of the BN CAC scheme. On the contrary, when we evaluate FN or FP, i.e., when we do not consider as an error the cases in which $R^{(aa)}$ is very close to the threshold τ_R, then the BN CAC scheme has a low error, and it outperforms the NN CAC scheme. Even the overall performance of the NN approach is better than that of BN. The disadvantage is its shallow level of interpretability (i.e., black-box function approximator) and impossibility to tune parameters in order to adjust the performance (i.e., the trade-off between maximizing the radio resource utilization vs. meeting the QoS requirements with a certain level of guarantees).

Furthermore, the performance of the proposed schemes (NN and BN based), in terms of CAC accuracy, is compared with the state-of-the-art CAC scheme for the LTE system. Results show that both proposed schemes outperform the state-of-the-art scheme. Moreover, the BN scheme can also meet different system requirements by opportunely tunning its parameters.

Chapter 5

RRM in LTE/NR systems: Channel and QoS aware MAC scheduling for VoLTE

Back in 2012, VoLTE was foreseen to become the dominant solution for the provisioning of voice services over 4G systems [175]. And indeed, in March of 2020, according to the latest official report by Global mobile Suppliers Association (GSA) [176], there were more than 272 operators worldwide investing in VoLTE in 119 countries globally, including 211 operators with commercially launched VoLTE-HD voice service in 100 countries. According to the same report, there is an annual rise of 19% in the total count of devices supporting VoLTE. The main advantage of VoLTE is its superior quality of voice call, offering to the end-user much reacher quality of experience. While in previous mobile network technologies, such as 3G, voice traffic is conveyed over dedicated circuit-switched networks, in VoLTE system, it is transferred over packet-switched networks along with other data traffic, such as best-effort HTTP traffic. Voice traffic has a very tight QoS requirements, such as bit rate and delay. To support the QoS requirements for different types of service, LTE already provides the possibility of setting up different bearers within the Evolved Packet System (EPS), each being associated with a different traffic flow mapped to its specific QoS requirements [164]. Vendor-specific QoS-aware solutions are then expected to be deployed in order to fulfill these requirements.

We focus on such QoS-aware solutions related to the downlink part of the LTE radio interface, which is based on the Orthogonal Frequency-Division Multiple Access (OFDMA). OFDMA allows a fine-grained dynamic allocation of radio resources both in the Time Domain (TD) and Frequency Domain (FD). In 3GPP specification this functionality is often referred to as Dynamic Packet Allocation or Packet Scheduling [177], while we normally throughout this thesis call it MAC scheduling. MAC scheduling resides in the base station, Evolved

Node B (eNB). The design of efficient MAC scheduling algorithms is left open for LTE eNB manufacturers to come up with advanced solutions that are envisioned to become key product differentiation factors. Considering VoLTE services, in particular, a good MAC scheduling solution is required to include a radio resource allocation mechanism that is aware of the QoS requirements as well as of the channel conditions, in order to maximize the voice capacity, i.e., the number of voice flows that can be served by the eNB with guaranteed QoS requirements.

5.1 Related work

Due to the increasing popularity of the LTE technology systems worldwide, there has been a growing interest in the design of efficient LTE/NR MAC scheduling algorithms, and several downlink scheduling algorithms have already been proposed in the scientific literature. An abundant survey on the topic is provided in [178]. However, most of the scheduling algorithms mentioned in this survey, such as Round Robin, Proportional Fair, Maximum Throughput, Throughput to Average, and Blind Equal Throughput, actually are not QoS-aware, and hence are not suitable for VoLTE systems. For this reason, we do not consider such schedulers in this study. Instead, we focus on the most relevant QoS-aware LTE downlink scheduling algorithms.

The first category of such algorithms includes those aiming to satisfy the delay requirement of real-time traffic, such as the scheme proposed in [179], which prioritizes the data flows to be scheduled based on the Head-of-line (HOL) delay parameter. A downside of this approach is that it does not take into consideration the variable channel conditions; in particular, in realistic scenarios in which the presence of fast and frequency-selective fading is expected, assigning radio resources based only on the HOL metric often results in the selection of lower modulation and coding schemes, which is spectrally inefficient and thus does not allow to achieve a high voice capacity.

Among the channel-aware approaches, we consider the Token Bank Fair Queue (TBFQ) scheduler [180], which is a queue- and channel-aware scheduling algorithm, which attempts to maintain fairness among users. TBFQ is based on the leaky-bucket principle, and it is mainly designed to support bursty traffic by assigning a higher amount of resources to the users with more data in the queues. This feature of the TBFQ approach is not adequate for voice traffic since it is characterized by small packet sizes and low expected queue fill levels. Furthermore, TBFQ scheduler does not explicitly take into account the delay requirements.

A better candidate for voice traffic is the PS scheduler [174], which is a channel-aware scheduler that aims at guaranteeing a predefined bit rate to each user. This algorithm has an outstanding performance because it successfully combines TD and FD scheduling to achieve higher spectral efficiency and increase the overall system capacity. Regarding the QoS support, the main drawback of this scheduler is that it only considers the Guaranteed Bit Rate (GBR) parameter specified within the EPS bearer. This means that delay-sensitive classes of traffic, such as voice, video, and gaming, may suffer poor quality even if their GBR requirement is satisfied. This limits the application of this scheduler to delay tolerant types of traffic.

As a step forward in this research line, we propose a new LTE downlink scheduling algorithm called Channel and QoS Aware (CQA) scheduler. The QoS parameters that it considers are the HOL and the GBR parameters. The CQA scheduler performs the scheduling according to

different criteria in the TD and FD, to achieve high spectral efficiency while at the same time taking care of satisfying the delay requirements of the traffic.

5.2 User-perceived QoS of voice calls

According to the ITU-T E-model [163] the quality of a voice call can be estimated by calculating the R factor, which we denote as R_f:

$$R_f = Ro - Is - Id - Ie_{eff} + A \ , \tag{5.1}$$

where Ro is the basic signal-to-noise ratio, including noise sources such as circuit noise and room noise; Is are impairments simultaneous to voice signal transmission, such as too loud speech level, non-optimum sidetone, quantization noise; Id are impairments caused by delay and echo effects; Ie_{eff} represents impairments caused by low bit-rate codecs and impairment due to randomly distributed packet losses; finally, the advantage factor A allows for compensation of impairment factors when the user benefits from other types of access to the user. A user is satisfied with the QoS [1] of voice call when the R_f is greater than a threshold [163]. In order to provide satisfactory R_f, the scheduling mechanism should optimize all the metrics at the MAC layer that affect QoS of voice calls. Those metrics are: MAC layer throughput, MAC layer queuing delay and packet losses caused by buffer overflows. Taking this into account, we propose an algorithm that aims at simultaneously:

- minimizing delay by giving priority to the user with greater HOL delay
- maximizing MAC layer throughput by improving radio resource utilization
- allocating to each user the amount of radio resources that is necessary to achieve the guaranteed bit-rate specified by the GBR parameter in the EPS bearer

5.3 Proposed LTE/NR MAC scheduling algorithm

The CQA scheduler is based on joint TD and FD scheduling, which has been shown in many studies to be a more efficient approach than only TD or FD scheduling; an example of such performance comparison for the LTE system can be found in [182]. The proposed algorithm runs every transmission time interval (TTI), which is equal to 1 ms. The TTI is the smallest resource unit in the time domain. In the FD, the smallest resource unit is resource block (RB), which forms RB groups (RBGs). Depending on the system settings, such as bandwidth and

[1] Note that according to [181] for communication services it is possible to distinguish three approaches to quality: QoS, user-perceived QoS, and QoE. QoS is mainly used to define technical parameters of telecommunication applications such as network delay, data rate, and packet loss. User-perceived QoS is typically measured with a subjective rating scale analyzed as a "Mean Opinion Score" (MOS). QoE is a measure of user performance based on both objective and subjective (i.e., psychological) measures of using a communication service (i.e., voice or audio call). QoE takes into account technical parameters (e.g., QoS) and subjectively perceived quality of service. For example, it may happen that certain offered QoS, in reality, offers very poor QoE, even if its QoS parameters meet the QoS requirements.

type of allocation, one RBG can contain different numbers of RBs. The smallest resource unit that our scheduling algorithm assigns is one RBG.

In the TD, at each TTI, the CQA scheduler selects from all the users $j = 1, ..., N$ those that did not yet reached the maximum bit rate (MBR) and groups them by HOL delay calculating the metric m_{td} in the following way:

$$m_{td}^{j}(t) = \lceil \frac{d_{HOL}^{j}(t)}{g} \rceil \, , \tag{5.2}$$

where $d_{HOL}^{j}(t)$ is the current value of HOL delay of flow j, and g is a grouping parameter that determines the granularity of the groups, i.e., the number of flows that will be considered in the FD scheduling iteration. The grouping is used to select the most urgent flows, i.e., with the highest value of HOL delay, and to enforce the scheduling mechanism to consider those flows in the following FD scheduling iteration. A low value for g reduces the users' diversity, thus decreases the scheduler's gains in the FD; on the other hand, it gives more importance to the d_{HOL} metric in the scheduling. This can be useful in network scenarios in which all users have relatively good channel conditions, and the fast fading is negligible. On the contrary, a high value for g increases the users' diversity, thus increases FD gains, but d_{HOL} has less impact on scheduling decisions. However, this parameter should be set up according to the network capacity and the expected average number of users in the system. The groups of flows selected by TD iteration are forwarded to FD scheduling starting from the flows with the highest value of the m_{td} metric until all RBGs are assigned in the corresponding TTI.

In the FD, for each RBG $k = 1, ..., K$, the CQA scheduler assigns the current RBG to the user j that has the maximum value of the FD metric which we define in the following way:

$$m_{fd}^{(k,j)}(t) = d_{HOL}^{j}(t) \cdot m_{GBR}^{j}(t) \cdot m_{ca}^{k,j}(t) \, , \tag{5.3}$$

where $m_{GBR}^{j}(t)$ is calculated as follows:

$$m_{GBR}^{j}(t) = \frac{GBR^{j}}{\overline{R^{j}}(t)} = \frac{GBR^{j}}{(1-\alpha) \cdot \overline{R^{j}}(t-1) + \alpha \cdot r^{j}(t)} \, , \tag{5.4}$$

where GBR^{j} is the bit rate specified in EPS bearer of the flow j, $\overline{R^{j}}(t)$ is the past averaged throughput that is calculated with a moving average, $r^{j}(t)$ is the throughput achieved at the time t, and α is a coefficient such that $0 \leq \alpha \leq 1$. In (5.3) the purpose of the d_{HOL} and m_{GBR} metrics is to provide to all flows the same level of QoS regarding delay and GBR by prioritizing the flows that have higher HOL delay and the flows which ratio of GBR^{j} to $\overline{R^{j}}(t)$ is larger. For example, if the GBR is achieved, but not also the MBR, $GBR^{j} < \overline{R^{j}}(t) < MBR^{j}$, to the flow j will be assigned lower priority in scheduling since $m_{GBR} \leq 1$. The purpose of $m_{ca}^{(k,j)}(t)$ is to add channel awareness to the system in order to maximize the capacity by assigning the resources to the flows that can use them more efficiently. For $m_{ca}^{(k,j)}(t)$ we consider two different metrics: $m_{pf}^{(k,j)}(t)$ and $m_{ff}^{(k,j)}(t)$. The m_{pf} is the Proportional Fair metrics which is defined as follows:

$$m_{pf}^{(k,j)}(t) = \frac{R_e^{(k,j)}}{\overline{R^j}(t)}\ , \tag{5.5}$$

where $R_e^{(k,j)}(t)$ is the estimated achievable throughput of user j over RBG k calculated by the Adaptive Modulation and Coding (AMC) scheme that maps the channel quality indicator (CQI) value to the transport block size in bits. The CQI value is the indication of the data rate which can be supported by the channel, taking into account the signal to interference plus noise ratio (SINR) and the characteristics of the UE's receiver [164]; this value is reported by UE to eNB for each RB as part of channel state information (CSI) reporting procedures that are defined in [183]. We consider the m_{pf} metric as a good channel awareness metric since it aims at simultaneously achieving fairness among flows and maximizing system capacity by prioritizing the users that have suffered lower channel quality and the users that have extremely good instantaneous channel quality; we denote the CQA scheduler that uses this channel awareness metric CQA_{PF}. The other channel awareness metric that we consider is m_{ff} which is proposed in [174] and it represents the frequency selective fading gains over RBG k for user j and is calculated in the following way:

$$m_{ff}^{(k,j)}(t) = \frac{CQI^{(k,j)}(t)}{\sum_{k=1}^{K} CQI(t)^{(k,j)}}\ , \tag{5.6}$$

where $CQI^{(k,j)}(t)$ is the last reported CQI value from user j for the k-th RBG. We consider this metric as a good channel awareness metric since it aims at increasing the overall system capacity by prioritizing users that can use available resources more efficiently. We name the CQA scheduler that uses this channel awareness metric CQA_{FF}.

5.4 Description of the scenarios: a simplified and EPA channel model

To evaluate the proposed scheduler we simulate a typical outdoor scenario in which N UEs are attached to a single eNB. All users perform voice over IP (VoIP) calls and have corresponding GBR EPS bearers set up in the EPS. We consider a single cell scenario, thus inter-cell interference is not considered in this work. The users are randomly distributed in a squared area around the macro cell. We consider two channel scenarios that are based on different channel models:

- Simplified channel model: the UEs are static ($v = 0\ km/h$) and no model for time and frequency selective fading is used. Thus, in this scenario, the SINR perceived by the UEs remains unchanged during the simulation. Even if the simulation scenario based on this model does not represent a realistic LTE system, we consider that using this model can help correlation understanding the performance of different schedulers for different channel conditions.
- EPA model: the UEs are mobile ($v = 3\ km/h$) and Extended Pedestrian A model (EPA) is used to simulate fading with values of model parameters defined in [173]. Due to the

Table 5.1 Simulation parameters

Macro cell transmission power	40 dBm
Bandwidth	6 RB (1.4 MHz)
Downlink/Uplink carrier frequency	2120/1930 MHz
RBs per RB group (RBG)	1
Pathloss model	Cost 231 (Hata Model PCS Extension)
Fading loss model	Pedestrian EPA model 3 km/h
Adaptive modulation and coding scheme	PiroEW2010
Users mobility model	Steady state random waypoint
Simulation area	1560x1560 m^2
Voice EPS bearer settings	GBR, conversational voice, 80 kb/s
Voice codec	G711
TTI	1 ms
Number of UEs in system	1-20
Replications of each simulation setup	10
Simulation time	8 s

UE's mobility along with the fading model, the quality of channel varies over time and frequency, thus the scenario that is based on EPA model can be considered as a realistic LTE scenario.

5.5 Simulation setup

We use the LTE-EPC network simulator (LENA) [184] to carry out the performance evaluation. To simulate the performance of the state of the art algorithm we used the implementation of the PS scheduler provided by [185], considering both versions of PS scheduler: PSS_{pf} and PSS_{coita}. We implemented in the LENA simulator the HOL scheduler according to [179] and both versions of the proposed CQA scheduler that we described in Sec. 5.3. The simulation parameters are shown in Table 5.1 and the system configuration is as follows.

The macro cell is connected via the PDN Gateway (PGW) to the internet and for each UE a separate remote peer was placed in the internet and connected to the gateway of the LTE network (the PGW) via a separate point to point link with overprovisioned bandwidth in order to simulate end-to-end performance of voice calls. For the path loss model we adopt the COST-231 path loss model [172], which is a common choice to simulate macro cell outdoor scenarios. The size of the square area is chosen such that, given the path loss model and the other network parameters, we have a wide range of SINR values, which we verified by observing the CQI values reported by the UEs to be in range $[1, 15]$. Note that CQI 1 correspond to a UE using the lowest MCS but still connected to the eNB, whereas CQI 15 correspond to a UE using the highest MCS. The tunable parameter of CQA scheduler g was determined empirically to give the best throughput performance for CQA scheduler for this network scenario. Its value is constant for both scenarios and equal to 300. While in simplified scenario we use a constant position mobility model, in the EPA scenario we simulate the mobility of UEs by using steady

state random waypoint mobility model [171]. The initial distribution of UEs is the same in both scenarios. We simulate the fast fading model described by EPA model using a Rayleigh multi-path fading model and we set the model parameters by using values defined in [173] for the EPA model. For each experiment setup we run 10 independent replications of each experiment which gives different position topologies.

5.6 Performance evaluation

The performance of the proposed CQA scheduler and the state of the art schedulers is evaluated in terms of QoS satisfaction and system performance. To measure QoS satisfaction we use the R-factor R_f that we described in Section 5.2. According to the E-model [163] the users are satisfied with QoS of voice call when $R_f \geq 70$. We use this threshold to evaluate if the QoS requirements of calls are satisfied. We consider the minimum R_f value $R_{f_{min}}$ over all the users in the system as the strictest metric to evaluate the performance of schedulers. We aim to evaluate the average number of users that each scheduling algorithm can support while providing to all the users satisfactory QoS. For this purpose, we define the metric $\overline{R_{f_{min}}}$, which represents the average $R_{f_{min}}$ over the set of independent replications of simulations. We denote the number of users for which $\overline{R_{f_{min}}} = 70$ as $\overline{N}_{QoS}$, which represents the voice capacity of the system.

5.6.1 Simplified channel model

In Fig. 5.1 we show the $\overline{R_{f_{min}}}$ for all the schedulers. The HOL scheduler achieves on average the highest number of satisfied users $N_{QoS} = 16$. In this scenario, the schedulers that are channel aware cannot benefit from the UE diversity; thus, their performance degrades. The schedulers that use the proportional fair metric, such as CQA_{PF} and PSS_{PF}, perform worst, while the schedulers based on the m_{ff} metric, such as CQA_{FF} and PSS_{coita}, perform slightly better.

In Fig. 5.2, we show the system throughput, which is calculated as the average total VoIP throughput at the application layer over all simulations having the same number of users. In a scenario with VoIP calls, the application layer throughput is significantly lower than the MAC throughput because of $RLC + PDCP + IP + RTP$ overhead; moreover, the transport blocks assigned to the user is is often greater than the VoIP packet. Because of this, a portion of resources that are assigned to the user is wasted. Also, it is important to notice that channel quality varies among UEs, and the UEs' achievable rate depends on AMC, so the system throughput is significantly lower than the peak LTE throughput that is often advertised for the given bandwidth. From Fig. 5.1 and Fig. 5.2, we notice that the HOL scheduler achieves good performance in terms of the QoS while has poor system throughput performance. In fact, low throughput performance is expected for the HOL scheduler since it has low radio resource utilization efficiency. On the other hand, CQA_{PF} and PSS_{PF} have the highest throughput performance but the worst QoS.

In Fig. 5.3 we show the cumulative distribution function of R_f. The HOL scheduler has the lowest probability to have unsatisfied VoLTE users and this probability is 5% lower comparing

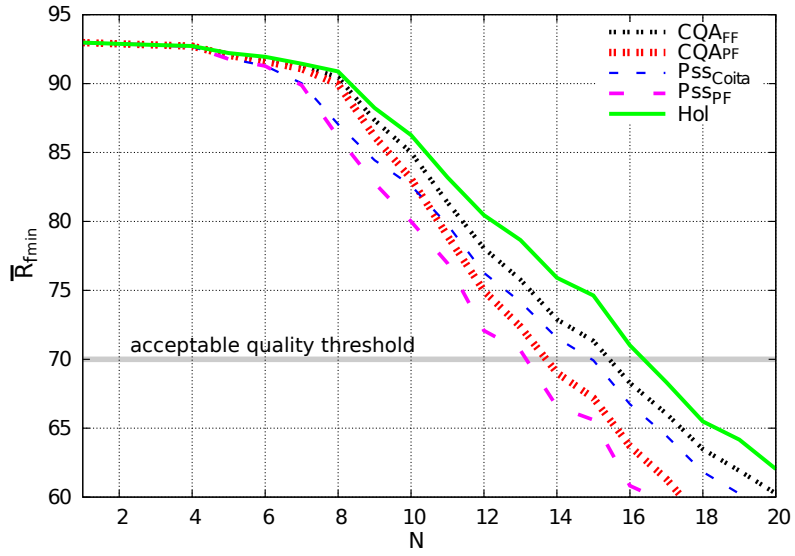


Figure 5.1 QoS performance comparison by using $R_{f\,min}$ metric in the static scenario

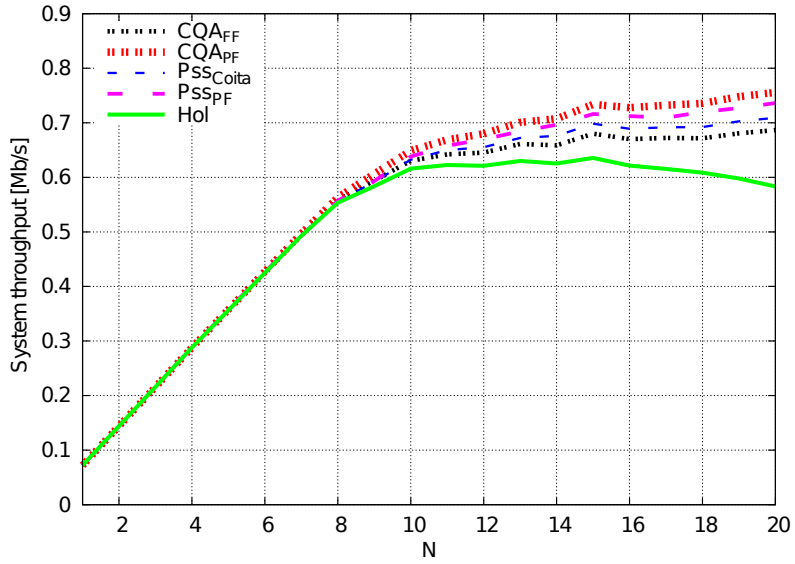


Figure 5.2 System throughput performance comparison in the static scenario

to other schedulers. We notice from the same figure that the second scheduler with the lowest CDF is CQA_{PF} scheduler, and we argue that this is because it achieves high throughput gains (See. Fig. 5.2), which reduces delays and packet losses due to buffer overflow.

5.6.2 EPA model

We evaluate the performance of all schedulers in the scenario in which UEs are moving, and fast frequency selective fading is present. From Fig. 5.4 we notice that the proposed CQA_{PF} scheduler achieves significantly better performance regarding the provided QoS than the other schedulers. The CQA_{PF} scheduler achieves increase in VoIP capacity up to 27% compared to CQA_{FF} and Pss_{PF} gains, and up to 100% compared to the Pss_{coita} and the HOL schedulers. We explain these performance gains by the use of m_{fd} metric from (5.3), which balances

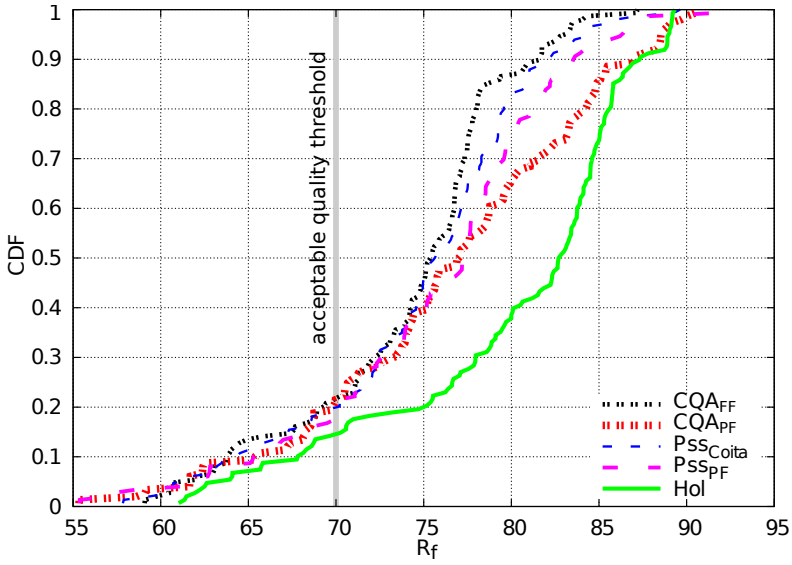


Figure 5.3 Cumulative distribution function of R_f in the static scenario when $N = 15$

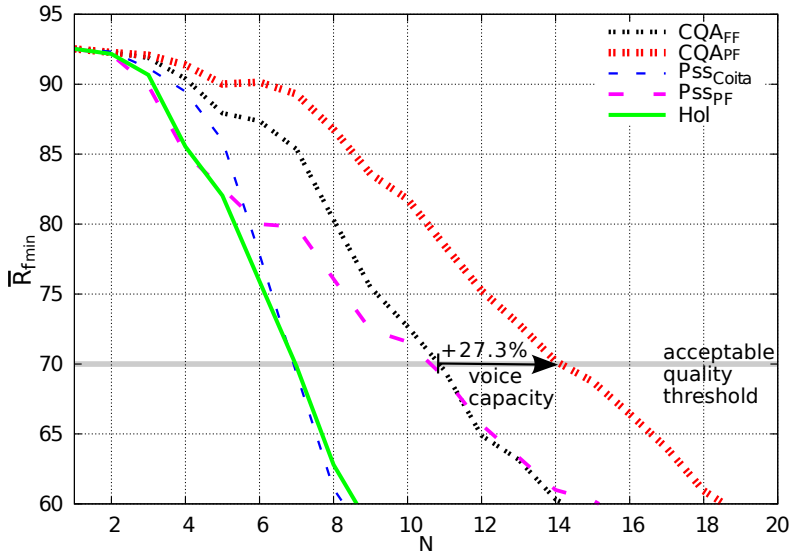


Figure 5.4 QoS performance comparison by using $R_{f_{min}}$ metric in the EPA scenario

the delay and GBR requirements with the capacity maximization objective. On the other hand, we explain the low performance of the HOL scheduler because it is not leveraging fast fading and users' diversity. The Pss_{PF} and the CQA_{FF} schedulers perform similarly, which can be explained by the fact that CQA_{FF} is gaining more performance by being HOL delay aware while Pss_{PF} is gaining higher performance by being channel aware. We notice that the schedulers that are using the Proportional Fair metric, i.e., CQA_{PF} and Pss_{PF}, achieve much higher performance gains than the schedulers that are using m_{ff} metric.

In Fig. 5.5 we show the application layer system throughput performance for the EPA scenario for the case $N = 15$. From the figure, we notice that the proposed CQA_{PF} scheduler achieves significant gains in terms of system throughput comparing to all the other schedulers. We again notice that the schedulers that use the m_{pf} metric achieve higher performance.

In Fig. 5.6 we show the cumulative distribution function of R_f for EPA scenario for the case

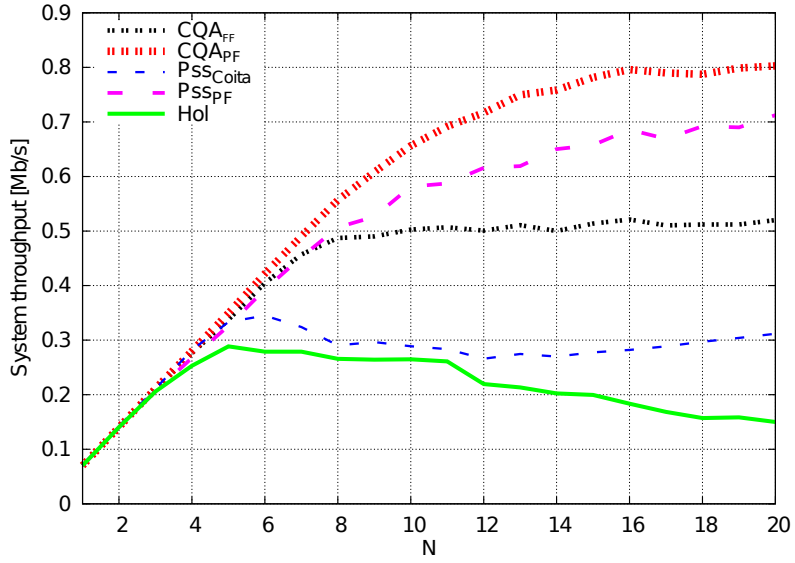


Figure 5.5 System throughput performance comparison in the EPA scenario

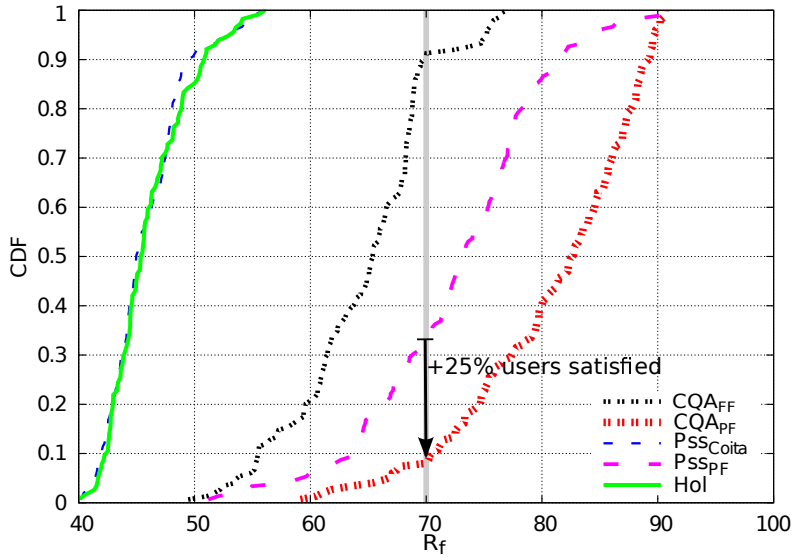


Figure 5.6 Cumulative distribution function of R_f in the EPA scenario when $N = 15$

$N = 15$. We notice that the CQA_{PF} scheduler has 25% higher probability to have satisfied users than the Pss_{PF} scheduler, around 80% higher than the CQA_{FF} scheduler, and approximately 90% higher than the Pss_{coita} and HOL schedulers. Finally, we conclude that in a realistic scenario with fast and frequency selective fading, schedulers which use Proportional Fair metric, such as CQA_{PF} and Pss_{PF}, significantly outperform schedulers based on m_{ff} metric, such as CQA_{FF} and Pss_{coita}.

5.7 Conclusions

In this chapter, we consider the MAC scheduling in LTE systems. We proposed a scheduling algorithm that is both channel-aware and QoS-aware. The proposed CQA scheduling algorithm

aims to enhance VoLTE capacity. In order to provide satisfactory VoLTE QoS (in terms of R_f), the scheduling mechanism should optimize all the metrics at the MAC layer that affect the QoS of voice calls. We identify the following metrics: the MAC layer throughput, the MAC layer queuing delay, and the packet losses caused by buffer overflows. Taking into account previous observation, we propose an algorithm that aims to simultaneously:

- minimize delay by giving priority to the user with more significant HOL delay
- maximize MAC layer throughput by improving radio resource utilization
- allocates to each user the amount of radio resources that are sufficient to achieve the guaranteed bit-rate specified by the GBR parameter in the EPS bearer

In order to meet the HOL delay requirement, the scheduler performs the grouping of the calls based on the HOL delay, i.e., to enforce the scheduling mechanism to consider the most urgent flows first. We elaborate on the size of the group as an essential configuration parameter. A small group size reduces the users' diversity, thus decreases the scheduler's gains in the frequency domain. Hence it is harder to maximize the radio resource utilization; on the other hand, it gives more weight to the HOL metric in the scheduling process, thus adds more importance to the QoS requirements related to delay and throughput. We observe that this can be useful in network scenarios in which all users have relatively good channel conditions, and the fast fading is negligible. On the contrary, a high value of group size increases the users' diversity, thus increases the gains in the frequency domain by maximizing the system capacity, but HOL delay has less impact in scheduling decisions. This parameter should be set up considering the desired overall behavior of the scheduler, and it depends on the network capacity and the expected average number of users in the system.

We propose two different metrics to use to prioritize flows, and based on these metrics; we propose two versions of the CQA algorithm, namely, CQA_{PF} and CQA_{FF}. We carried out performance evaluation by simulation and compared the proposed solution with the state-of-the-art scheduling algorithms: the PSS and the HOL delay schedulers.

To understand the impact of different metrics and parameter configurations, we consider two scenarios based on different channel models: a simplified channel model and EPA model. The first model serves us to model a static scenario where time and frequency selective fading is negligible. Even if this kind of channel model does not represent a realistic LTE system, we justify using this model to help understand the performance of different schedulers for different channel conditions. On the other hand, the second channel model (EPA) is used to model a typical real-world pedestrian scenario in which the quality of the channel varies over time and frequency.

We evaluate the QoS of VoLTE calls in terms of the R factor (i.e., according to the ITU-T E-model) and the application throughput (calculated as the average total VoIP throughput at the application layer over all simulations having the same number of users). Results show, as expected, when a simplified channel model is used that the schedulers that are channel aware cannot benefit from the UE diversity. While in a realistic pedestrian scenario in which fast fading is present, the proposed CQA scheduler gains approximately to 27% of VoLTE capacity compared to the PSS scheduler and almost 100% compared to the HOL scheduler.

In summary, in this and the previous chapter, Chapter 4, we could observe through the performance analysis that in order to meet the QoS requirements and achieve good QoS while maximizing the radio resource utilization, it is necessary that RRM entity employs CAC and MAC scheduling algorithms that take into account the QoS requirements, and can efficiently and dynamically determine how to use the radio resources. As shown in the performance evaluations, both of these RRM functionalities, CAC and scheduling, are important to achieve efficient RRM, and the performance of each of them depends on the other. For example, even the best MAC scheduler cannot perform well if the CAC scheme accepts too many calls (i.e., many false-positive decisions), and also even the best CAC mechanism cannot provide adequate QoS guarantees if the MAC scheduler is not aware of the QoS requirements and is not able to prioritize flows while the scheduling. While these RRM functionalities focus on a single-cell optimization and an operation in the licensed spectrum, in the following chapter, Chapter 6, we consider a broader scenario and elaborate on how the radio resources could be efficiently used in a multi-cell environment in licensed spectrum. Further on, in Chapter 7 and Chapter 8, the multi-RAT coexistence in unlicensed spectrum is analyzed also in a multi-cell scenario context.

Chapter 6

RRM in LTE/NR systems: Dynamic frequency and bandwidth allocation

6.1 Introduction

In recent years, mobile networks have been rapidly growing in size and complexity. As discussed in Chapter 2, operators are continuously seeking to improve the network capacity and the QoS by adding more cells of different types to the current deployments consisting of macro-, micro-, pico-, and femto-cells. These heterogeneous deployments are loosely coupled, increasing the complexity of cellular networks (4G, 5G, and B5G). This increase in complexity brings a significant growth in the operating and capital expenditures (OPEX/CAPEX) of the mobile network providers. To reduce these costs on a long-term scale, operators are seeking network solutions that will provide automatic network configuration, management, and optimization, and that will minimize the necessity for human interventions. Initially, the NGMN alliance recommended SONs as a key concept for next-generation wireless networks, and defined operator use cases in [88]. Shortly after, the SON concept was recognized by the Third-Generation Partnership Project (3GPP) as an essential functionality to be included in 3GPP technologies, and consequently, it was introduced into the 3GPP standard in [186]. SONs are expected to reduce the OPEX/CAPEX and to increase the capacity and the QoS in future cellular networks (5G and B5G). As described in Section 2.4, all self-organizing tasks in SONs are described at a high-level by the following features: self-configuration, self-healing, and self-optimization.

Various studies show that roughly 80% of mobile data traffic is indoor [187]. Still, operators

are failing at providing good QoS (coverage, throughput) to the indoor users. In order to solve these issues while saving OPEX/CAPEX, operators are deploying small cells. These are low cost cells that can be densely deployed in residential areas which are connected to the core network via broadband. In current small cell deployments, various technical issues have been detected. Small cells are increasingly being deployed according to traffic demands rather than by traditional cell planning for coverage. Such small cell networks are characterized by unpredictable interference patterns, which are caused by the random and dense small cell placements, the specific physical characteristics of the buildings (walls, building material, etc.), and the distance to outdoor cells, e.g., macro or micro base stations. Thus, such deployment scenarios are characterized by complex dynamics that are hard to model analytically. However, in the research literature, solutions are often proposed based on simplified models, e.g., assuming interference models with uniform distribution of small cells over the macrocell coverage, which differs significantly from realistic urban deployments [188]. Therefore in this kind of deployments, the classical network planning and design tools become unusable, and there is an increasing demand for small cells solutions that are able to self-configure and self-optimize [189].

Recently, several network infrastructure providers have been developing SON solutions based on machine learning and big data analytics. For example, Reverb, one of the pioneers in self-optimizing network software solutions, has created a product called InteliSON [190], which is based on machine learning techniques, and its application to real networks results in lower drops, higher data rates, and lower costs for the operator. Similarly, Zhilabs and Stoke [191] are developing solutions based on big data analytics. Samsung developed a product called Smart LTE [192] that is leveraging the SON solution that gathers radio performance data from each cell and adjusts a wide array of parameters at each small cell directly.

Similarly to the previously described industrial approaches, in this work, we focus on the application of machine learning to improve SON functionalities by providing more accurate estimates of the key performance indicators (KPIs) as a function of the network configuration. The KPIs are mainly important for operators to detect changes in the provided QoS and QoE, for example, in order to reconfigure the network in response to a detected degradation in QoS. The estimation of the KPIs based on limited network measurements is one of the main requirements of the Minimization of Drive Tests (MDT) functionality and represents a key element for the realization of the Big Data Empowered SON approach introduced in [193]. In this work, we apply learning based KPI estimation approach to the specific use case of LTE small cell frequency and bandwidth assignment. We investigate the potential of LTE's frequency assignment flexibility [164] in small cell deployments, i.e., exploiting the possibility of assigning different combinations of carrier frequency and system bandwidth to each small cell in the network in order to achieve performance improvements. Currently, most small cell deployments rely on same-frequency operation with the reuse factor of one, whose main objective is to maximize the spectral efficiency. However, the spectrum reuse factor is subject to a trade-off between spectral efficiency and interference mitigation. Since interference may become a critical issue in unplanned dense small cell deployments reconsidering spectrum reuse factors in this kind of deployments may be necessary. Moreover, the same-frequency operation is not expected to be the standard practice in the future since additional spectrum will be available at higher frequencies, e.g., 3.5 GHz [194]. Thus, for future network deployments, it will be more relevant to consider band-separated local-area access operating on higher-frequency bands, with the overlaid macro layer operating on lower cellular bands.

In this chapter, we investigate how to exploit this flexibility in order to maximize the performance of small cell network deployments. We show that the proposed learning based KPI estimation can be successfully employed to effectively optimize such multi-frequency multi-bandwidth small cells deployment strategy.

6.2 Related Work and Proposed Contributions

Frequency assignment is one of the key problems for the efficient deployment, operation, and management of wireless networks. For earlier technologies, such as 2G and 3G networks, as well as Wi-Fi access point deployments, relatively simple approaches based on generalized graph coloring [195] were sufficient to obtain a good performance. This is because the frequency assignment for these networks was often orthogonal and with a low degree of frequency reuse, and the runtime scheduling of radio resources had a highly predictable behavior due to the simplicity of the used methods. Additionally, due to the predictable system load, the frequency assignment was often based on static planning, which could be done easily offline.

However, the new 4G and 5G technologies, such as LTE and NR, adopt more flexible spectrum access approaches based on dynamic frequency assignment (DFA) and inter-cell interference coordination in order to allow high frequency reuse. In particular, DFA is recognized as one of the key aspects for high performance small cell deployments [196]. According to DFA, the available spectrum is allocated to base stations dynamically as a function of the channel conditions to meet given performance goals. Furthermore, the LTE technology is highly complex due to the inclusion of advanced features such as OFDMA and SC-FDMA, adaptive modulation and coding (AMC), dynamic MAC scheduling, and HARQ [164]; hence, it is much more difficult to predict the actual system capacity in a given scenario than it was for previous mobile technologies. Because of this, it is very challenging to design a DFA solution that can work well not only on paper, but also in a realistic small cell deployment.

On this matter, while several publications appeared in the literature to deal with the general problem of cellular radio resource management, considering aspects ranging from power control [197] to frequency reuse between macro and small cells [198], only few works focus on DFA for small cell networks. Among these, we highlight [199] and [200] whose authors propose DFA solutions based on graph coloring algorithms. The key aspect of these papers, and of many other similar works, is that they assume that the achieved rate on a specific channel is given by simple variants of Shannon's capacity formula, thus neglecting some important aspects that affect the performance of cellular systems, such as MAC Scheduling, HARQ, and L3/L4 issues. Following this approach may result in significant capacity estimation errors, which may end up in sub-optimal of low performance frequency assignments. Because of this, we argue that solutions like [199] and [200] offer an idea on the general trend of the expected results, but need to be extended with system level studies, in order to capture the complexity of real deployments.

Additionally, as argued in [161], the existing techniques for small cell aware spectrum allocation need further investigation, i.e., co-tier interference and global fairness require more research, and there are still open issues. The main issue is to strike a good balance between spectrum efficiency and interference, i.e., to mitigate the trade-off between orthogonal

spectrum allocation and co-channel spectrum allocation. Still, the existing approaches are highly complex, difficult to be implemented by the operator, and they mainly aim to address the cross-tier spectrum-sharing issues.

We believe that a learning based approach can successfully address these issues while keeping the overall implementation and computational complexity very low. The main advantage of the machine learning based approach over other techniques is its ability to learn the wireless environment and to adapt to it. An example is the work in [201] where the authors propose a machine learning approach based on reinforcement learning in a multi-agent system according to which the frequency assignment actions are taken in a decentralized fashion without having complete knowledge on actions taken by other small cells. Such a decentralized approach may lead to frequent changes in frequency assignments, which may cause unpredictable levels of interference among small cells and degradation of performance.

In this chapter, we propose different machine learning and advanced regression based techniques for the performance prediction (e.g., that a user would experience in a small cell network) by leveraging a small sample of performance measurements. These techniques take as inputs different frequency configurations and measured pathloss data, and hence allow to estimate the impact of configuration changes on various KPIs. Differently to the previously described work [201], in our approach frequency assignments of the small cells are determined in a centralized fashion, by selecting the parameters which will lead to the best network performance.

6.3 Learning Based Dynamic Frequency and Bandwidth Assignment

6.3.1 Optimization Problem and Real System Constraints

Our specific optimization problem consists of selecting, for each deployed eNB $i = 1, \ldots, N$, the frequency f_c^i and the system bandwidth B^i that achieves the best network performance in terms of selected KPI. The number of possible configurations, C, is exponential with N; the base of the exponent depends of the number of allowed combinations of f_c and B for each eNB, which depends on the total bandwidth available for the deployment by the operator and is constrained by the operator's deployment policy. Let $x_{conf}^{(i,j)} = (f_c^{(i,j)}, B^{(i,j)})$ be the configuration of the i-th eNB in the configuration j; then the j-th network configuration may be represented as a vector $\vec{x}^j = [x_{conf}^{(1,j)}, \ldots, x_{conf}^{(N,j)}]$, where $j = 1, \ldots, C$. Let γ_{kpi}^j be the network performance for the selected KPI. The network configuration that maximizes the network performance is formally given by

$$x^{\vec{(opt)}} = \arg\max_{\vec{x}^j} \gamma_{kpi}^j. \tag{6.1}$$

If the values γ_{kpi}^j are known for all frequency and bandwidth configurations, then the $x^{\vec{(opt)}}$ can be found by performing an exhaustive search on the set of samples $(\vec{x}^j, \gamma_{kpi}^j)$. However, the application of the exhaustive search is not feasible in a real system. The practical constraints of this solution are the cost and the time of performing the network measurements for all

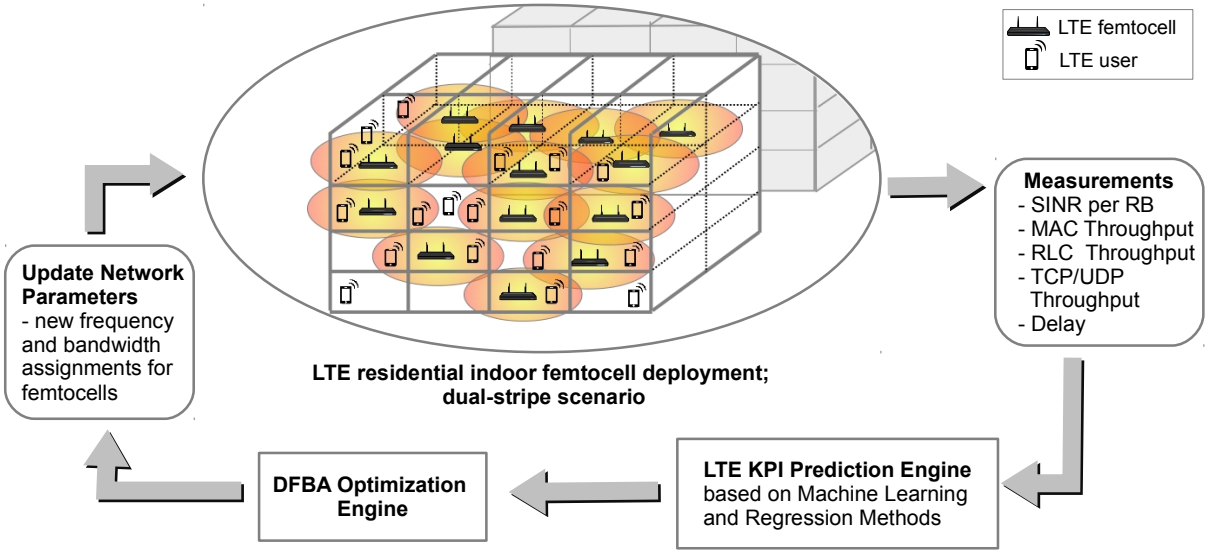
possible configurations. The measurements may be obtained by performing drive tests, but as these tests are expensive for the operator, the number of tests would need to be very limited. To reduce costs, the MDTs measurements may be used. Even so, time would be a significant constraint since the time to obtain all measurements linearly grows with the number of possible configurations.

As an example, in a four small cell network deployment with a total available bandwidth of 5 MHz, considering f_c values multiple of 300 kHz (three times the LTE channel raster, i.e., one third of the possible frequencies), and limiting the choice of B to $\mathcal{B} = \{6, 15, 25\}$ for simplicity, there are already 4625 physically distinct configurations. For a five small cell network this number grows to 34340. If the measurement time per configuration is only 1 hour, then the time necessary to gather measurements for a four small cell network is 193 days, and for a five small cell network is close to 4 years. In order to overcome this constraint, we aim at designing a solution that is capable of performing nearly optimally, while leveraging only a limited number of KPI measurements.

Finally, we consider two constraints of the real small cell deployments: the number of possible configurations, C, and the frequency of the configuration changes in the network. Even if an LTE carrier could be positioned anywhere within the spectrum, while respecting the channel raster constraint, and the basic LTE physical-layer specification does not specify about the exact frequency location of an LTE carrier, including the frequency band, the number of allowed combinations needs to be limited for practical reasons [194], e.g., to reduce search time when an LTE terminal is activated. As we will show in this work, even with a limited number of combinations of parameters, significant performance gains can be achieved.

6.3.2 Proposed Approach

In a nutshell, our goal is to design a general framework for LTE network performance prediction and optimization that is easy to deploy in a real LTE system and able to adapt to the actual network conditions during normal operation. Figure 6.1 illustrate the proposed approach. As shown in the figure, we mainly focus on LTE indoor small cell network deployments and, in terms of evaluation, on the typical LTE residential dual-stripe scenario described in [202]. This scenario characterizes not only interactions among neighboring small cells within the same building, but also among small cells belonging to adjacent buildings. According to our approach, the measurements are gathered from both, LTE users and LTE small cells. On the user side, we gather measurements related to the performance achieved by the user, i.e., throughput, and delay, and the corresponding measurements related to channel conditions, i.e., SINR per RB. On the small cell side, we gather RLC and MAC layer statistics and various throughput performance measurements. These measurements are then used to calculate different metrics, which are then used for network performance predictions by the *LTE KPI Prediction Engine*. This engine is leveraging different machine learning and regression methods to realize the LTE KPI prediction functionality. The predicted LTE KPI values are then forwarded to the *Dynamic Frequency and Bandwidth Allocation (DFBA) Optimization Engine*, which is using these values together with network measurements to inspect how near the current network performance is to the estimated optimal one, for the currently measured network conditions. If the DFBA Optimization Engine estimates that the change in network configuration will compensate possible trade-offs (e.g., interruption in service), it schedules the



(a)

Figure 6.1 Proposed learning based approach.

reconfiguration of the frequency and bandwidth assignment.

Our approach follows the centralized SON (CSON) architecture, according to which there is a centralized node that oversees the operation of all small cells and controls their behavior. In CSON architecture, the centralized node receives inputs from small cells and determines their configuration. Thus, the *LTE KPI Prediction Engine* and the *DFBA Optimization Engine* are placed in the centralized node. Since the configuration parameters are not going to be changed frequently, the proposed solution should not be affected by the latency due to the communication exchange between small cells and the centralized node. Also, the network overhead is low since the measurement information from the small cells to the central node can be scheduled per best-effort basis. Note that this architecture is compliant with the control plane solution for MDT, which is discussed in 3GPP TR32.827 [203]. Thus, the main message exchanges in our approach are between UEs and small cells, and between small cells and the centralized management node, and all the interfaces needed for implementing our solution are already present in the standard.

The main contribution of the proposed approach is the learning based LTE KPI performance estimation. Even if in this work we apply this approach to the frequency and bandwidth assignment use case, without loss of generality, we argue that it is much more general and may be used for a larger set of configuration parameters and for different utility-based network planning and optimization tasks [204], where the accurate prediction of KPIs are necessary for an effective optimization.

6.3.3 LTE KPI Prediction Engine

To realize the LTE KPI prediction engine, we propose a learning based approach according to which different KPIs are accurately predicted by using regression analysis and machine learning techniques based on basic pathloss and configuration information combined with a limited number of feedback measurements that provide the throughput and the delay metrics

Table 6.1 Considered combinations of covariates.

Covariates	SINR			SINR/MAC THR Mapping		
Usage	per RB	Aggr.	Min.	per RB	Aggr.	Min.
AGGR		×	×		×	
1RB+			×	×		
2RB+	×			×		
AGGR-1RB+		×	×	×		

for a particular frequency and bandwidth setting. As discussed in Section 6.3.1, we aim at designing a solution that requires a minimal amount of training for active exploration. Moreover, the prediction engine should be able to predict different KPIs, e.g., the network-wide and per-user LTE KPIs. To achieve all these requirements and to select the best candidate for the prediction engine, we study and compare the performance of various classical and modern prediction techniques. We list and explain these techniques in Section 6.3.4.

These prediction techniques leverage various parameters, metrics, and derived inputs. The latter are usually called *covariates* or *regressors* in the statistical and machine learning literature. Among the covariates used in this chapter, the majority is calculated by means of the SINR/MAC throughput mapping. This mapping represents the network MAC layer throughput calculation based on the actual network measurements. We calculate this mapping in the following way: According to the LTE standard, UEs are periodically reporting to the base station a channel quality indicator (CQI) both per each subband and wideband. We use this value at the MAC layer of the base station for AMC mapping, i.e., to determine the size of the transport block (TB) to be transmitted to the UE. A typical AMC behavior is to select a TB size that yields a BLER between 0 and 10% [205]; the TB size for each given modulation and coding scheme and number of RBs are given by the LTE specification in [206].

Moreover, we investigate the performance for different combinations of covariates. Since the covariates can be combined on a per-RB basis, or aggregated together in various ways (such as considering the minimum or the sum of SINRs over the band), the number of different combinations of covariates is very large. Here we limit our attention to a small number of representative combinations summarized in Table 6.1.

Additionally, we consider the effect on the prediction performance of different sampling methods, i.e., *random* and *stratified sampling*. In statistics, *stratified sampling* is obtained by taking samples from each stratum or sub-group of a population, so that a mini-reproduction of the population is achieved; conversely, according to the *random sampling* method, each sample is chosen entirely by chance in order to reduce the likelihood of bias. While *stratified sampling* requires more effort for data preparation, it is appealing for its higher prediction accuracy in scenarios where the performance of different sub-groups of population or sampling regions varies. For the *stratified sampling* method, we define the sampling regions by calculating the aggregated network throughput based on the SINR/MAC throughput mapping previously described.

Finally, we analyze performance prediction by means of the goodness of fit metrics, such as the prediction error in network-wide and per-user throughput estimation, evaluating how they

depend on the size of the training set. This allows us to determine the ability of the proposed solution to learn during real-world operation.

6.3.4 Statistical and Machine Learning Methods for LTE KPI Prediction Engine

In this section, we provide an overview of the different statistical and machine learning methods studied for the realization of the LTE KPI prediction engine. We begin with a basic overview of the principles and terminology involved, and then give a concise summary on the principles of the used methods. For further information on the considered prediction techniques, the interested reader is referred to [207] and [208].

The objective of all of the considered methods, regardless of whether statistical or machine learning based, is to find a function that *predicts* the value of a dependent variable $y = f(x_1, \ldots, x_n)$ as a function of various *predictors* or *covariates* x_i. Usually, this is done by conducting a limited number of experiments that yield the value of y for known values of the covariates that are then used to *fit* or *train* the model. The functional form of the model, as well as the used training procedure, are the main differences between the different methods. In our case, the y corresponds to a performance metric of interest, and the different x_i are the measurements of network conditions (SINR values for different nodes), as well as available prior data (such as theoretical MAC layer throughput at given SINR).

- **Linear regression method (LM)**: The simplest method used for establishing a baseline prediction performance is *linear regression* method, that simply models y as a linear function of the covariates, as in

$$y = a_0 + \sum_i a_i x_i. \tag{6.2}$$

  The coefficients a_i are determined based on the training data, for example, by minimizing the root mean squared error (RMSE) of the predictor. Linear regression also has in our context a simple communication-theoretic interpretation: in the high SINR regime, linear functions approximate well the Shannon capacity formula, and y becomes simply the best approximation of the network throughput as an optimal weighted sum of the individual Shannon capacity estimates. Thus linear regression can be used as an improved proxy for simple Shannonian SINR-based network capacity models. The simple generalization of this basic scheme is to apply a transformation function to each of the terms $a_i x_i$. The generalized regression techniques thus obtained are usually called *projection pursuit regression* (PPR) methods.

- **K Nearest Neighbor (KNN)**: KNN is the simplest non-traditional prediction method that we consider. For KNN, we consider the covariates x_i as defining a point in an Euclidean space, with the value of y obtained from the corresponding experiment assigned to that point. When predicting y for x_i' for which experimental data is not available, we find the K nearest neighbor of the point x_i' from the training data set in terms of the Euclidean distance. Our prediction is then the distance-weighted average of the corresponding values of y. The KNN algorithm is an example of a non-parametric method that requires no estimation procedure. This makes it easy to apply but limits both its ability to

generalize beyond the training data and the amount of smoothing it can perform to counter the effects of noise and other sources of randomness on the predictions.

- **Tree graph models**: A much more general and powerful family of regression techniques is obtained by considering *trees* of individual regression models. The model corresponds to a tree graph, with each non-leaf vertex corresponding to choosing a subspace by imposing an inequality of some of the x_i. The leaves of the tree finally yield the predictions y as the function of the ancestor vertices partitioning the space of x_i into subsequently finer subspaces. The various regression tree algorithms proposed in the literature differ mainly in the method used to choose the partitioning in terms of the covariates x_i, as well as in the way training data is used to find the optimum selection of decision variables in terms of the chosen partitioning scheme. We consider both **boosting tree (BTR)** and **bagging tree (TBAG)** in the process of finding optimal regression tree. Of these, bagging uses *bootstrap* (sampling with replacement to obtain a large number of training data sets from a single one) with different sample sizes to improve the accuracy of the involved parameter estimates. Boosting, on the other hand, performs retraining of the model several times, with each iteration giving increased weight to samples for which the previous iterations yielded poor performance results. The final prediction from a boosted tree is a weighted average of the predictions from the individual iterations. In general, regression trees are a very powerful and general family of prediction methods that should be considered as a potential solution to any non-trivial prediction or learning problem.

- **Support Vector Machines (SVMs)**: SVMs are a type of the machines often used for classification and pattern recognition but can also be used for regression problems. These methods have efficient training algorithms and can represent complex non-linear functions. The core of this method is the transformation of the studied data into a new, often higher dimensional space so that this data is linearly separable in this new space and, thus, the classification or regression is possible. The representation of data using a high-dimensional space carries the risk of overfitting. SVMs avoid this by finding the optimal linear separator, a hyper-plane that is characterized by the largest margins between itself and the data samples from both sides of the separator. A separator is obtained by the solution of a quadratic programming optimization problem, which is characterized by having a global maximum, and is formulated using dot products between the training data and the support vectors defining the hyperplanes. While it is rare that a linear separator can be found in the original input space defined by the x_i, this is often possible in the high dimensional feature space obtained by mapping the covariates with the *kernel function*. In the following, we use SVMs with basic radial basis functions for regression.

- **Kohonen networks (KOH)**: The last family of machine learning techniques that we consider in our study is that of *self-organizing maps*, also known as *Kohonen networks*. These form a family of artificial neural networks, for which each *neuron* (a vertex on a lattice graph) carries a vector of covariates initialized to random values. The training phase iterates over the training data set, finds the nearest neighbor to each vector of covariates from this set and the neural network, and updates the corresponding neuron and its neighbors to have a higher degree of similarity with the training vector. Over time, different areas of the neural network converge to correspond to different common types occurring often in the training data set. While originally developed for classification

Table 6.2 Considered KPIs, prediction methods and metrics, regressors and sampling methods.

KPIs	Network throughput; User throughput
Prediction methods	Bagging tree (TBAG, Treebag) Boosted tree (BTR) Kohonen network (KOH) SVM radial (SVM) K-nearest neighbor (KNN) Projection pursuit regression (PPR) Linear (LM)
Prediction metrics	RMSE user fit; 95th %ile RMSE
Regressors	SINR; SINR/MAC throughput mapping
Sampling	Random; Stratified

problems, the Kohonen network can be used for regression by assigning a prediction function (such as the simple linear regression) to each class discovered by the neural network.

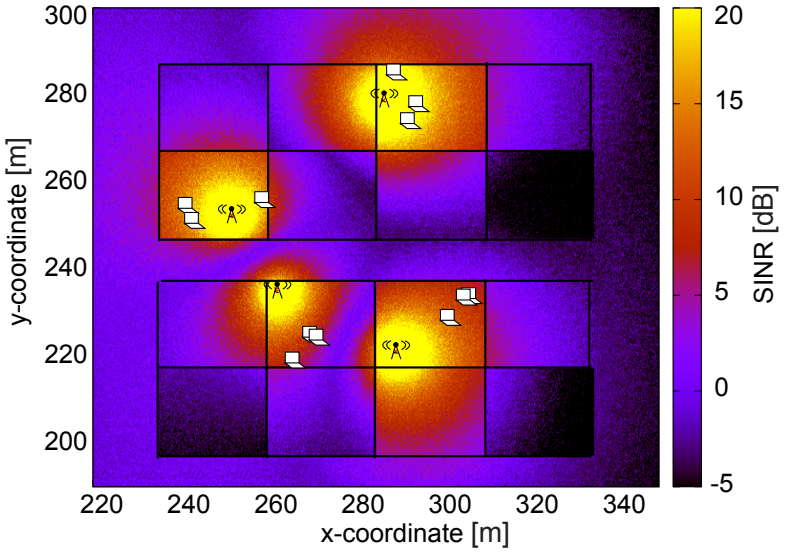
The considered KPIs, prediction methods and metrics, regressors, and sampling methods are summarized in Table 6.2. We use the R environment, and, in particular, the caret package, as the basis of our computations [208].

6.4 Performance Evaluation

6.4.1 Evaluation Setup

We consider a typical LTE urban dual stripe building scenario defined in [202] and the corresponding simulation assumptions and parameters defined in [209]. In Figure 6.2, we show a radio environmental map of one instance of the simulated scenario. Each building has 1 floor, which has 8 apartments. The small cells (home eNBs) and users are randomly distributed in the buildings. Each home eNB has an equal number of associated UEs and is placed in a separate apartment along with its associated UEs. By using the random distribution, we aim at simulating the scenario that corresponds to the greatest extent to a realistic residential small cell deployment. The random placement of small cells in each independent simulation, along with the random placement of the users, adds to the simulation an additional degree of randomness, which is consequently increasing the credibility of obtained simulation results. We concentrate on studying the following network configurations:

- 4 home eNBs, 12 users and a total system bandwidth of 2 MHz,
- 4 home eNBs, 8 users and a total system bandwidth of 5 MHz, and
- 2 home eNBs, 20 users and a total system bandwidth of 2 MHz.



(a)

Figure 6.2 Radio environmental map of dual-stripe scenario with 1 block of 2 buildings. Each home eNB has connected three UEs that are located in the same apartment.

For propagation modeling, we use the ITU-R P.1238 model with additional loss factors for internal and external wall penetration. We consider both TCP and UDP as transport layer protocols to investigate the performance of our approach for different types of a transport protocol. In both cases, we configure traffic parameters to send packets with a constant rate that can saturate the system. Additionally, we consider the effect of the MAC scheduler on the LTE KPI prediction performance. The purpose of the MAC scheduler is to decide which RB will be assigned to which UE; different policies are used for this purpose, resulting in different performance trade-offs. We select two schedulers that are widely used as a reference in literature: Round Robin (RR) and Proportional Fair (PF). For more information on MAC schedulers, the interested reader is referred to [178]. To avoid the effects of the network initializations and starting up of the user applications, we neglect the first interval of 5 s of each simulation execution. We configure simulations by using different combinations of B and f_c, and we configure other network parameters according to Table 6.3. Different random placements of small cells and users are achieved by running each simulation configuration with different values of the seed of the random number generator.

To simulate the scenarios, we use the ns-3 based LTE-EPC network simulator (LENA) [184], which features an almost complete implementation of the LTE protocol stack, from layer two and above, together with an accurate simulation model for the LTE physical layer [205]. The use of such a detailed simulator provides a performance evaluation that is reasonably close to that of an experimental LTE platform.

6.4.2 Results on the correlation between Covariates and KPIs

We begin by illustrating the challenges of MAC layer throughput prediction based on an SINR metric. For this analysis we select the *Sum SINR*, *Sum/THR Mapping* and *Min SINR per RB*

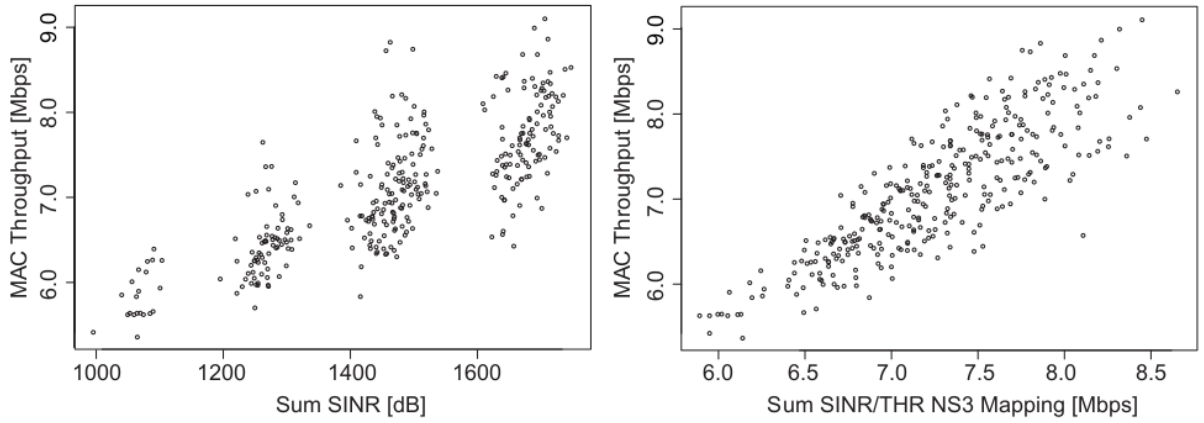
Table 6.3 Evaluation configuration parameters

Parameter name and units	Values
Scenario type	Dual-stripe [202] and [209]
No. of small cells per scenario setup	$\{2, 4\}$
No. of users per small cells	$\{2, 3, 10\}$
Small cell placement inside of apartment	Random
User placement inside of apartment	Random
Lower bound of downlink carrier frequency values	2110 MHz
Upper bounds of downlink carrier frequency values	$\{2112, 2115\}$ MHz
Lower bound of uplink carrier frequency values	1920 MHz
Upper bounds of uplink carrier frequency values	$\{1922, 1925\}$ MHz
Total system bandwidth	$\{2, 5\}$ MHz
Carrier frequency spacing	300 kHz
Small cell bandwidth	$\{1.4, 3, 5\}$ MHz
Small cell bandwidth in RBs	$\{6, 15, 25\}$
Home eNB transmission power	20 dBm
Transport protocols	$\{TCP, UDP\}$
MAC Schedulers	*Proportional Fair*, *Round Robin*
Simulation time in seconds	15
Measurements start time in seconds	5
Measurements update interval in milliseconds	100
No. of independent simulations	50

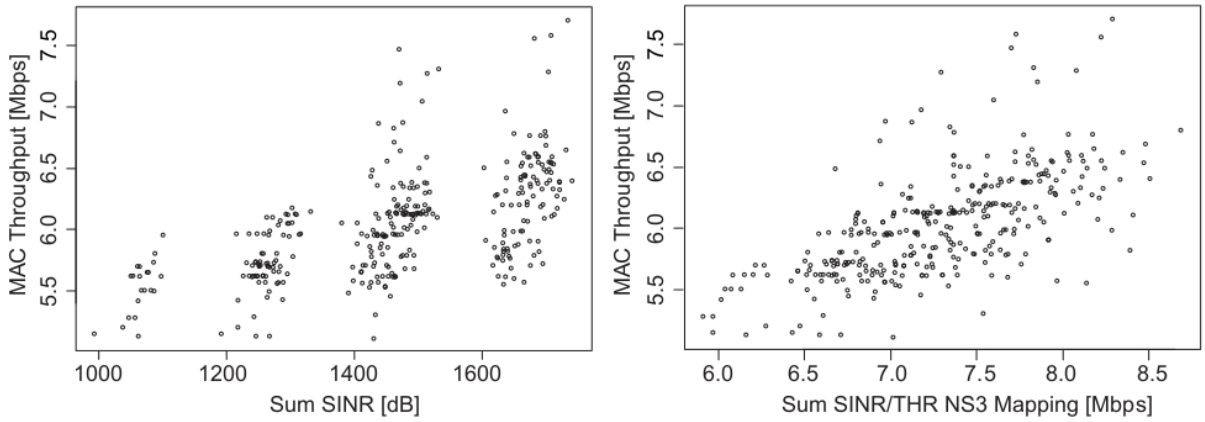
covariates, that were introduced in Section 6.3.3. The *Sum SINR* covariate is calculated as the raw sum of SINRs per RB. The *SUM/THR Mapping* represents the MAC layer throughput calculated as a function of the raw sum of SINRs per RB according to the throughput calculation based on the AMC scheme, which is explained in Section 6.3.3. The *Min SINR per RB* covariate is the minimum SINR perceived per RB. The SINR metric is calculated by leveraging on the pathloss measurements gathered at each UE. In Figure 6.3 we show the actual measured system-level MAC layer throughput as a function of either *Sum SINR* or *Sum/THR Mapping* based on 337 simulation results. Points in the figures correspond to measurements obtained from different simulation executions.

From Figure 6.3(a) and (b), we note that: 1) a low positive correlation is present between actual measured throughput and covariates, which confirms the need for advanced prediction techniques for KPI predictions; and 2) the correlation between the system-level MAC throughput and the covariates, *Sum SINR* and *Sum/THR Mapping*, is very similar for Proportional Fair and Round Robin schedulers, i.e. it is expected that the KPI prediction engine that is predicting system-level KPIs by using these covariates will perform equally well regardless of the MAC scheduler used at eNBs. This is not the case for the user-level KPI estimation. Namely, from Figure 6.3(c), we note that when the Proportional Fair scheduler is used, there is no linear correlation between the MAC throughput and *Sum SINR/THR Mapping*, while when the Round Robin scheduler is used, there is a positive correlation. This indicates that the choice of MAC scheduler significantly affects the correlation function between the actual measured MAC throughput and the selected covariate, when the user-level KPIs are predicted. This can be explained by the fact that the Round Robin scheduler allocates an approximately equal amount of resources to each UE, while the resources allocated by the Proportional Fair scheduler strongly depend on the actual environment, e.g., distributions of small cells and users, and on the channel conditions of all users; thus, the KPIs obtained when using Round Robin scheduler should be easier to predict. Still, the results obtained for the Round Robin scheduler show a large dispersion of correlation. This behavior may be the consequence of assigning the resources to UEs always in the same order during the simulation; thus, if there is a significant difference between SINRs among RB, this will affect the performance of the user. For example, if some user always gets assigned an RB with low SINR, the performance will be poor, even if the average SINR value over all RBs allows for much better performance. Another reason could be that the transport block size assigned to the user is affected by the presence of RB with very low SINR; because of this, in the following, we consider the correlation of *Min SINR per RB* and *MAC throughput*.

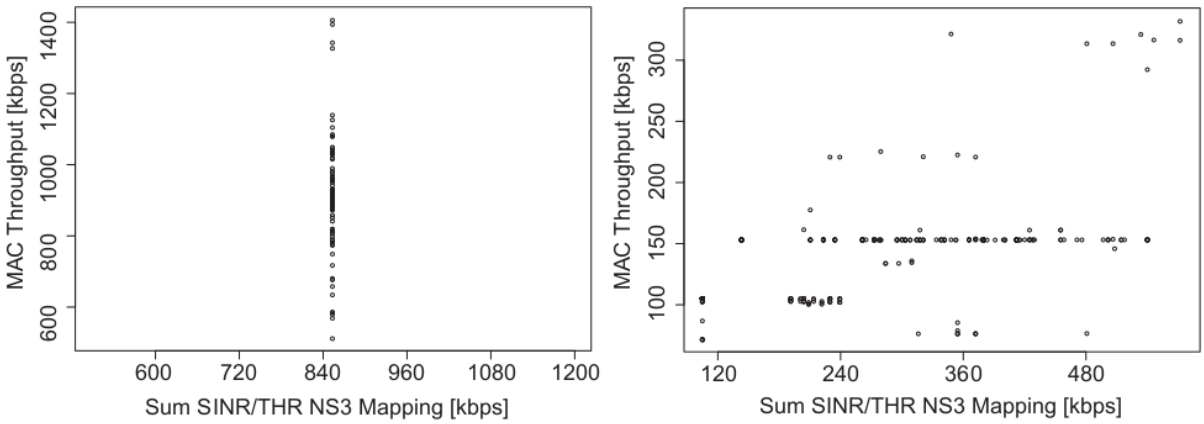
In Figure 6.4, we illustrate the correlations between the KPI and the selected covariates on a much larger data set, which contains 4625 samples. These samples are achieved by configuring a larger system bandwidth, 5 MHz, which allows for a much larger number of frequency and bandwidth assignment combinations, as are explained in Section 6.3.1. As we show in the following discussion, the analysis on a larger data set confirms the trends that were observed for smaller data sets in Figure 6.3. In Figure 6.4(a) and Figure 6.4(b), we note the strong correlation between the transport protocol type and the measured MAC layer throughput. When the transport protocol is UDP, there is a strong correlation between the *MAC throughput* and the *Sum SINR/MAC THR Mapping* covariate. On the other hand, when TCP is being used, there is a weak correlation, i.e., it is harder to predict the KPIs. This is expected behavior because of the complex interplay between the TCP congestion control and the LTE PHY, MAC, and RLC layers. We also note from these two figures that there is no strong correlation



(a) Measured MAC throughput vs. min SINR (left) and SINR/MAC thoughput mapping (right) for a scenario with 4 eNBs with 3 users each, Proportional Fair scheduler, and TCP traffic.



(b) Measured MAC throughput vs. min SINR (left) and SINR/MAC thoughput mapping (right) for a scenario with 4 eNBs with 3 users each, Round Robin scheduler, and TCP traffic.



(c) Measured MAC thoughput for two selected users vs. SINR/MAC THR mapping for a scenario with 4 eNBs 3 users each, Proportional Fair (left) and Round Robin schedulers (right) for TCP traffic.

Figure 6.3 Actual measured performance vs. pathloss-based SINR, and 3GPP based mapping of these values to MAC throughput. System bandwidth of 2 MHz, 4 small cells with 3 associated users each.

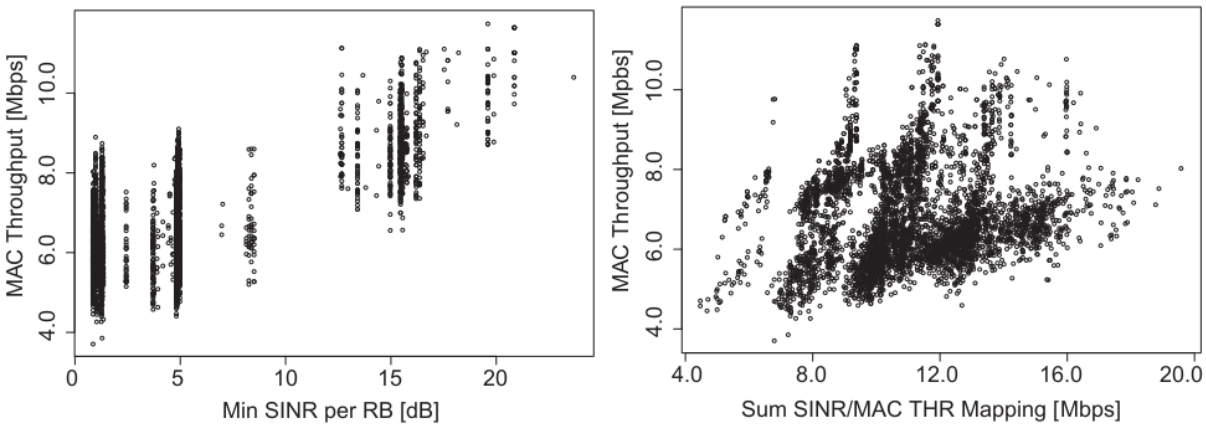
between *MAC throughput* and *Min SINR per RB*, so that the dispersion of results for the Round Robin scheduler shown in Figure 6.3(c) is not caused by assigning the min SINR per RB to UEs. Figure 6.4(c) shows that the correlation remains strong when the eNBs are configured to use the Round Robin scheduler instead of Proportional Fair, and that the *Sum SINR* and the *Sum SINR/MAC THR Mapping* covariates can be used almost interchangeably for predictions. We also note that a smaller number of users increases the dispersion in the SINR vs. MAC throughput dependency even further.

From these results, we note that the correct selection of covariates is fundamental for the robust and effective prediction engine. Moreover, we expect that designing a solution that can perform good in a variety of network configurations, and that can perform equally good while predicting both system-level and user-level KPIs, is a challenging problem.

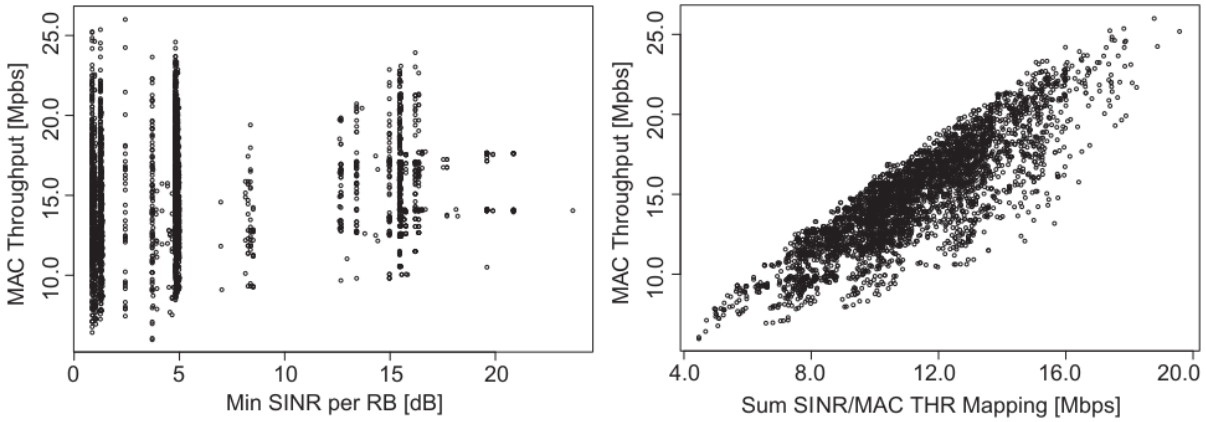
6.4.3 Performance of Prediction Methods

Following the conclusions derived in Section 6.4.2, we select the scenario setup and regressors for the performance comparison of the LTE KPI prediction methods. Namely, we select the configuration that appears the most complex for prediction, i.e., the configuration that offered low or lack of linear correlation between the predicted KPI and covariates, that is the network configuration in which small cells operate with the Proportional Fair MAC scheduler and UEs traffic goes over TCP. Additionally, based on a study from Section 6.4.2, we select the aggregate regressors since they appear to have a higher correlation with KPI than *Min SINR per RB*. A total of 4625 samples are obtained by running the small cell network scenario that consists of four small cells with two users associated with each of them, while the total system bandwidth is 5 MHz. The training data for each prediction method is obtained by selecting 10% of samples by random sampling method. The testing data samples are generated based on measurements for each user in the scenario, with a total of 50 independent samplings and regression fittings samples. We consider the following prediction techniques: Bagging tree (TBAG), Boosted tree (BTR), Kohonen network (KOH), SVM radial (SVM), K-nearest neighbor (KNN), Projection pursuit regression (PPR), and Linear regression method (LM), as introduced in detail in Section 6.3.4. Finally, Figure 6.5 shows the results of the prediction performance of different prediction methods. For boxplots, the three lines of the box denote the median together with the 25th and 75th percentile, while the whiskers extend to the data point at most 1.5 interquartile ranges from the edge of the box.

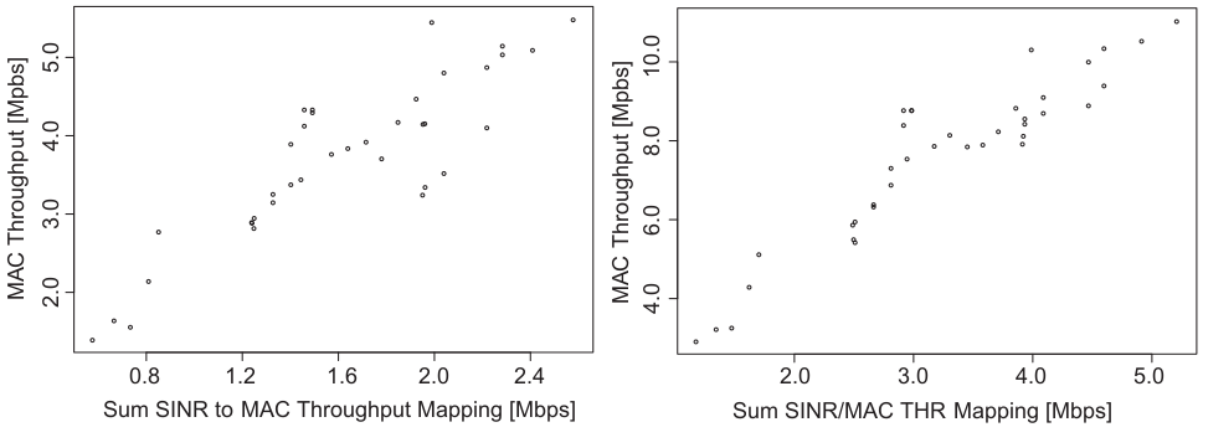
As expected for the selected network scenario with complex non-linear nature of the additional information, the simplest prediction method, LM, has the highest Root Mean Square Error (RMSE) and consequently, the poorest prediction performance ratio. The poor performance of the LM method indicates that analytical models based on Shannonian capacity estimates are also expected to perform poorly. Note also that the gain of more advanced methods over LM lower bounds the gain compared to even simpler schemes, such as full frequency reuse or orthogonalized channelization. More advanced prediction techniques based on regression, PPR and KNN, are computationally extremely fast ($\ll$ 1 ms for the tested sample set), which can thus be useful to offer an intermediate solution in situations in which more computationally expensive methods are not feasible. Among advanced machine learning techniques, SVMs and KOH networks perform the poorest, and the latter technique shows additionally a large variability in the performance prediction accuracy. Both tree-based methods (TBAG and



(a) Measured MAC throughput vs. min SINR (left) and SINR/MAC thoughput mapping (right) for a scenario with 4 `eNBs` with 2 users each, Proportional Fair scheduler, and TCP traffic.



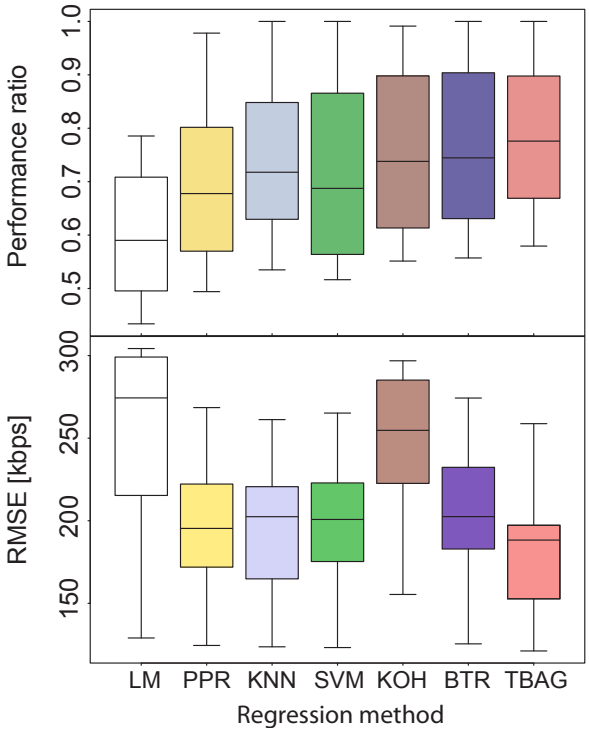
(b) Measured MAC throughput vs. min SINR (left) and SINR/MAC thoughput mapping (right) for a scenario with 4 `eNBs` with 2 users each, Proportional Fair scheduler, and UDP traffic.



(c) Measured MAC throughput vs. SINR/MAC thoughput mapping for a selected `eNodeB` (left) and a whole network (right) for a scenario with 2 `eNBs` with 10 users each, Round Robin scheduler, and UDP traffic.

Figure 6.4 Actual measured performance vs. pathloss-based SINR, and 3GPP based mapping of these values to MAC throughput. System bandwidth of 5 MHz. Setup with 4 small cells each having associated 2 users in (a) and (b); 2 small cells each having associated 10 users in (c)

.



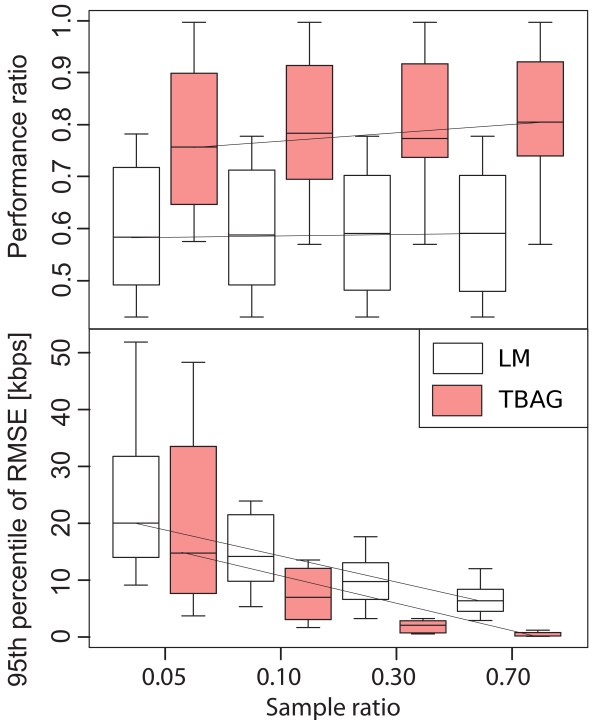
(a)

Figure 6.5 Comparison of prediction methods over random sampling of 10%.

BTR) perform consistently better than all previous methods in terms of raw performance and variability of results; finally, the TBAG method achieves the best prediction performance. This superior performance is expected due to the nature of TBAG and BTR. The use of bootstrap samples results in both of these methods being essentially not an individual machine learning optimizer, but an *ensemble learner* conducting voting between a large number of individual models. Such combinations of models usually outperform individual ones by a wide margin at the cost of larger storage and training overhead [207]. Based on the latter discussion, we conclude that TBAG is the most promising method for the prediction engine.

6.4.3.1 Prediction Performance Validation for Different Sizes of The Training Set

In the following, we evaluate the prediction performance of TBAG as a function of the size of the training set, i.e., in order to assess how fast it can learn when deployed in an actual scenario. We carry out a performance evaluation study using the same small cell network scenario setup that we used for the comparison of the prediction techniques. We compare TBAG with the LM method in order to analyze the advantage of the application of advanced prediction techniques instead of simple prediction techniques for different sizes of the training set. For this performance evaluation, we define the performance ratio metric as the ratio of the network throughput of the frequency and bandwidth allocation chosen by solving the optimization problem with the considered model to the network throughput of the best possible frequency and bandwidth configuration, i.e., the one that would be allocated by an exhaustive search algorithm. The purpose of this metric is to give a measure of how close a given solution is to the optimal frequency and bandwidth assignment. In Figure 6.6, we show the results



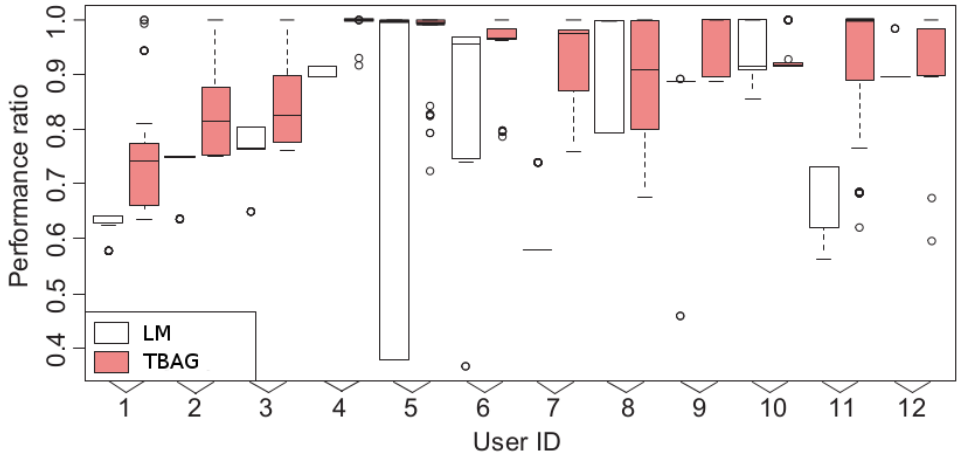
(a)

Figure 6.6 Linear and bagging tree methods for different sizes of the training set (random sampling with 5% to 70%).

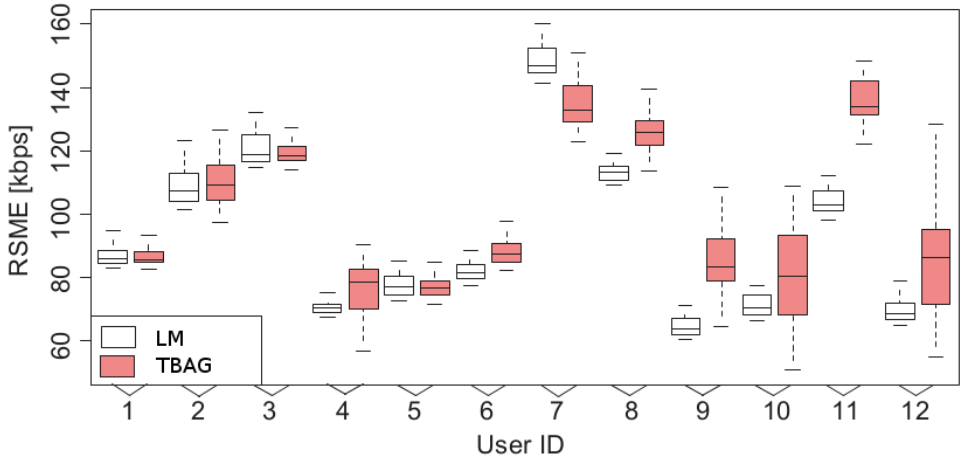
of the prediction performance for different sizes of the training set. The black lines in the figures show the tendencies in the plot, while the boxplots are generated in the same way as for the results shown in Figure 6.5. By observing the RMSE from Figure 6.6, we note that for more accurate performance more samples need to be taken, though this does not necessarily translate into a better network optimization performance, which is the case for the LM method. Additionally, we conclude that the benefit of advanced prediction techniques over simpler prediction techniques is not only the ability to learn on a very small sample set, but is also the ability to improve its performance over time. Both characteristics are crucial for real network deployment, as we seek a solution that can work well with minimum a priori knowledge, and that is able to improve performance by exploiting real- time network measurements.

6.4.3.2 Prediction Performance Validation for Different LTE KPIs and Network Configurations

We continue the TBAG performance analysis by testing the prediction of different KPI metrics under various network configurations. Specifically, whereas before we evaluated TBAG in the context of optimizing system-level KPI, we now focus on the performance prediction of TBAG in terms of user-level KPIs. We evaluate the performance obtained with differently configured small cell network setups. The fixed scenario parameters are: the system bandwidth is 2 MHz, the network has 4 small cells, and a total of 12 users. We run independent batch simulations that have in common the small cell network topology but differently configured transport protocols used by UEs' applications (TCP or UDP) and different MAC scheduler (Proportional Fair or



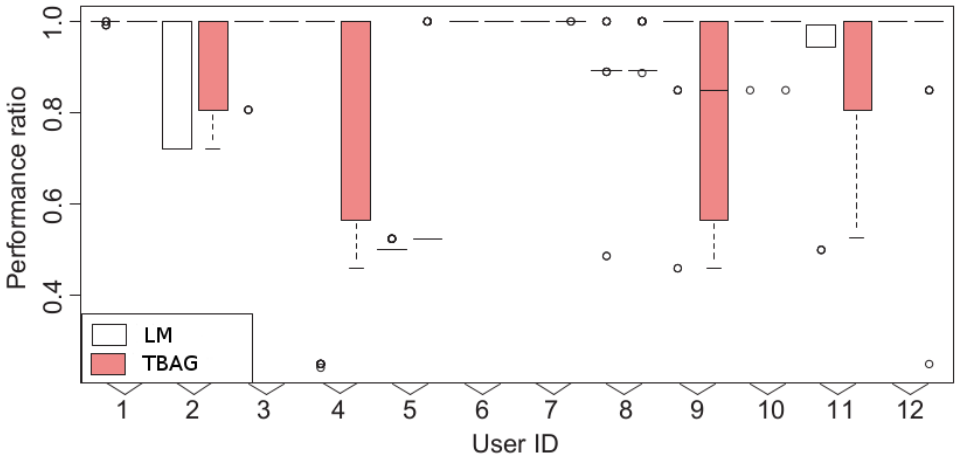
(a) Normalized performance considering per user optimization.



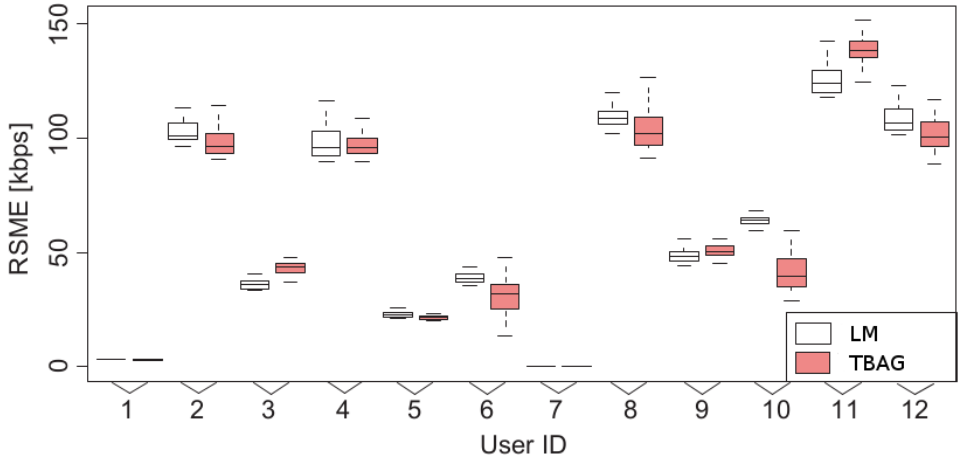
(b) RMSE over user performance regression fitting with actual measured performance ranging from 538 kbps to 1545 kbps.

Figure 6.7 Random sampling with 10% of 337 permutations being explored with the linear and the bagging tree regression methods and the aggregate regressor for the TCP, Proportional Fair scheduler.

Round Robin). Out of the four combinations (two different schedulers, two different transport protocol types) that we evaluated, we illustrate the performance of TBAG vs. LM in Figure 6.7 and Figure 6.8 for the two most interesting cases: 1) eNBs employing the Proportional Fair scheduler and UEs traffic going over TCP, and 2) eNBs employing the Round Robin scheduler and UEs traffic over UDP. Our results confirm that the TBAG method performs well for different scenario setups. Here the TBAG method outperforms the LM method, especially in case of TCP and the Proportional Fair scheduler (panels (a)–(b)). We note that the results shown in Figure 6.5 also hold on a per-user basis, as well as in a more complex and dynamic network scenario (TCP and Proportional Fair scheduler being used). Figure 6.7(a) shows a similar collection of results for the case of UDP with the Round Robin scheduler. Here even the simple LM method performs nearly optimal due to the simplified higher-layer interactions explained in Section 6.4.2. These figures confirm the previously formulated hypothesis that the network configuration with the simpler setup (UDP and more simple scheduler, such as Round Robin) results in higher predictability.



(a) Normalized performance considering per user optimization.



(b) RMSE over user performance regression fitting with actual measured performance ranging from 1521 kbps to 2871 kbps.

Figure 6.8 Random sampling with 10% of 337 permutations being explored with the linear and the bagging tree regression methods and the aggregate regressor the UDP, Round Robin scheduler.

6.4.3.3 Prediction Performance Validation for Different Covariates and Sampling Methods

We evaluate the performance of the TBAG prediction method for different covariates and different sampling methods: the random and the stratified sampling, introduced in Section 6.3.3. Figure 6.9 summarizes the performance of the TBAG method and the LM with different used covariates, together with two sampling methods, taking over 5% and 30% samples being taken. The stratified sampling results in better performance than the simple random selection of the configurations used to train the predictor. The basic AGGR covariate is outperformed by the 1RB+ regressor if complex machine learning based methods are applied, as those can make use of the additional information available through them (see Table 6.1 for the covariate abbreviations). For LM in Figure 6.10, due to the non-linear nature of this additional information, the performance impact is actually negative. In general, only advanced machine learning and regression techniques are able to benefit from more complex covariates, such as per-RB mea-

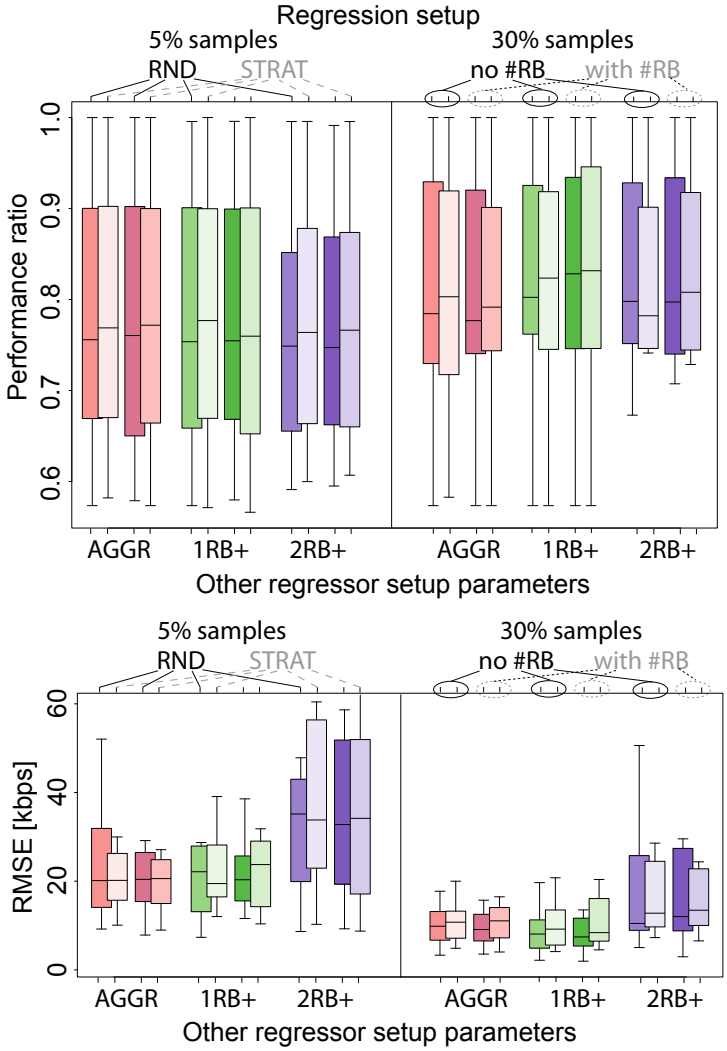


Figure 6.9 Bagging tree regression method. Four small cells scenario with two users per small cells and bandwidth of 5 MHz. Random sampling with 5% and 30% of 4625 permutations. Per user network optimization is considered with actual measured performance ranging from 1521 kbps to 2871 kbps.

surements, provided that a large enough sampling base is available (which was not the case for the 2RB+ regressor).

6.4.4 Performance Evaluation of Proposed Learning Based DFBA Approach

Finally, in this section we present the major results of this work by evaluating the network performance achieved for DFBA when the proposed learning based approach is used, and comparing it with the case where prediction methods based on pathloss-based mathematical models that use SINR and MAC throughput mapping estimates (sum or minimum of those over the RBs) are used. The performance gain is expressed as the percentage of the maximum achievable network performance obtained by applying an exhaustive search method to solve the DFBA problem. The learning based DFBA approach is using the TBAG method for LTE KPI

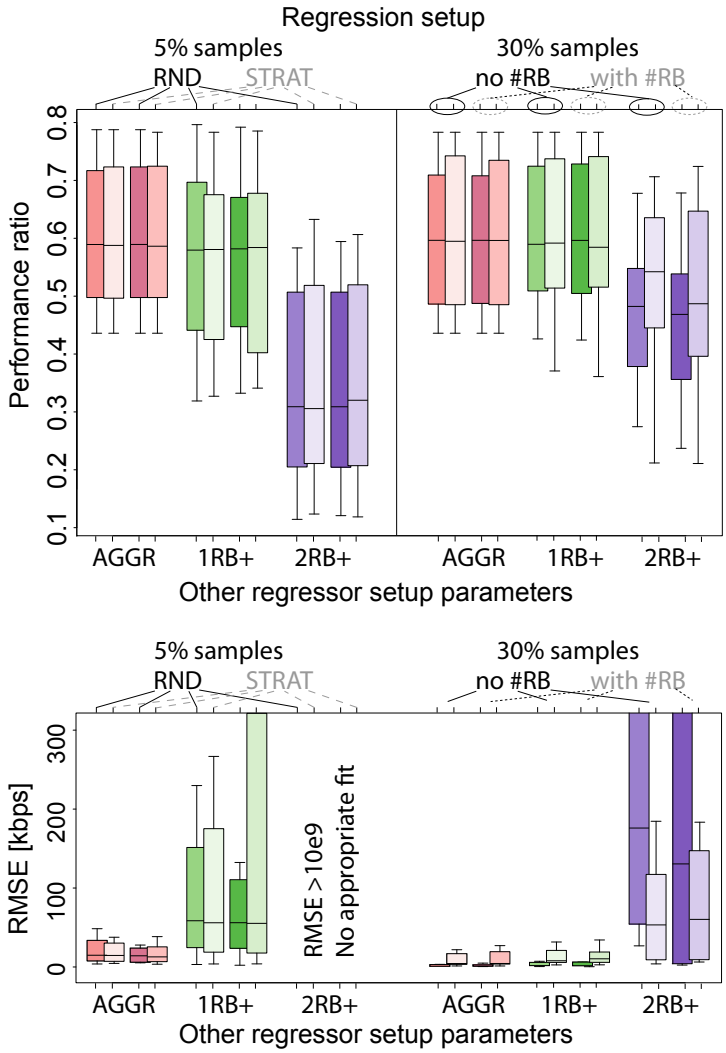


Figure 6.10 Linear regression method. Four small cells scenario with two users per small cells and bandwidth of 5 MHz. Random sampling with 5% and 30% of 4625 permutations. Per user network optimization is considered with actual measured performance ranging from 1521 kbps to 2871 kbps.

predictions which is trained by using the stratified sampling method and is employing active probing in addition to pathloss values. Table 6.4 shows the performance obtained when using different prediction methods for solving the frequency and bandwidth optimization problem explained in Section 6.3.1 with the goal of total network throughput maximization.

The scenario label identifies the number of small cells/number of users, the percentage of taken samples, and the employed transport layer and schedulers. The gains obtained by using the learning based DFBA range between 6% and 43%. We note that the gain is largest for the more complex scenarios, which means that even larger gains are expected for more complicated performance optimization goals, e.g., ones that include a fairness metric. Overall, the results provided in Table 6.4 show that the learning based DFBA approach results in the selection of a network configuration that performs better compared to the SINR-based models, and is close-to-optimal.

Table 6.4 Comparison of DFBA performance when different prediction approaches are used.

Scenario conf. and sample set size	4/12, 2 MHz, 10%				4/8, 5 MHz, 5%			
Transport protocol	TCP		UDP		TCP		UDP	
MAC scheduler	PF	RR	PF	RR	PF	RR	PF	RR
SINR	85%	83%	83%	86%	53%	54%	53%	42%
Min SINR	91%	82%	89%	89%	71%	77%	58%	49%
Sum SINR/MAC THR Mapping	72%	72%	70%	81%	61%	61%	35%	43%
Min SINR/MAC THR Mapping	89%	81%	88%	89%	55%	64%	55%	50%
Learning based (TBAG)	100%	85%	100%	95%	96%	95%	97%	92%
Exhaustive search	100%	100%	100%	100%	100%	100%	100%	100%
Exhaustive search [Mbps]	9	8	9	7	12	10	26	25

6.5 Conclusions

We investigated the problem of performance prediction in LTE small cells and we studied its application to dynamic frequency and bandwidth assignment in an LTE small cells network scenario. We proposed a learning based approach for LTE KPI performance prediction and we evaluated it by using data obtained from realistic urban small cell network simulations.

Summarizing, the key contributions of the work presented in this chapter are the following:

- We propose a learning based KPI estimation and we study its application to the use case of Dynamic Frequency and Bandwidth Assignment (DFBA) for self-organizing LTE small cell networks.
- We select and investigate various machine learning and statistical regression techniques for predicting network and user level KPIs accounting for the impact on the performance of the whole LTE stack, based on small number of measurements. The focus of this study is specifically on well-established machine learning and regression techniques rather than on developing some new ad hoc solutions. Furthermore, at the time of publishing this work, it was the first one to include both machine learning and regression techniques in a comparative integrated study applied to small cell SONs.
- We study the impact of the choice of covariates (measurement or configuration information made available to the performance prediction algorithm) and different sampling strategies (effectively deciding which measurements of network performance to carry out in a given deployment) on the efficiency of the KPI prediction. Additionally, the prediction performance is tested for different network configurations, different sizes of training sets, and different KPIs.
- We evaluate the performance of a DFBA solution based on the proposed learning based KPI estimation, comparing with a legacy approach as well as with an optimal exhaustive search approach.

The results firmly show that the learning based performance prediction approach can yield very high performance gains. The outstanding aspect of the learning based DFBA approach is that the high performance gains are obtained for a reasonably small number of measurements, which allows for its implementation in a real LTE system. Among the studied prediction methods, the Bagging tree prediction method results to be the most promising approach for LTE KPI predictions compared to other techniques, such as Boosted trees, Kohonen networks, SVMs, K-nearest neighbor, Projection pursuit regression, and Linear regression methods. Another conclusion of the comparative study on the prediction methods for the LTE network performance prediction is that the used performance metric and RMSE should be considered together when evaluating the different performance prediction methods. In particular, a high RMSE does not always lead to poor optimization results, and, if maximum performance grows, RMSE may also increase due to higher variance, but the main tendency of prediction might not change. Finally, we show that the DFBA based on LTE KPI prediction achieves in average performance improvements of 33% over approaches involving simpler SINR-based models. Moreover, the learning based DFBA performs very close to optimal configuration, achieving on average 95% of the optimal network performance.

Chapter 7

Coexistence: LTE/Wi-Fi in 5 GHz band

In recent years, unlicensed bands have gained a lot of attention from the cellular wireless industry as a way to increase the system capacity. Unlicensed bands that are currently available worldwide are: 2.4 GHz, 5 GHz, 60 GHz, and recently opened in 2020, the 6 GHz band [210]. Unlicensed spectrum has been used traditionally by Wi-Fi technologies, e.g., 2.4 GHz and 5 GHz bands, and in recent years, 60 GHz bands by WiGig. In current networks, a significant part of cellular traffic is offloaded through Wi-Fi and WiGig. According to a recent Cisco study, currently, around 74% of mobile phone traffic runs over Wi-Fi, and by 2023 it is estimated that it will increase to 79% [65]. In recent years, cellular vendors and operators became interested in the usage of the 5 GHz band, 6 GHz, and mmWave bands (such as the 60 GHz band). This has resulted in various technologies, which have been presented in detail in Chapter 2 in Section 2.3, and which are standardized by 3GPP, LTE-U Forum, or MulteFire Alliance. In this and the following chapter, we focus on cellular unlicensed technologies that use cellular RAT to access to unlicensed spectrum, e.g., LAA, LTE-U, and NR-U.

With recent advances in unlicensed cellular technologies, there is an increased interest in academy and industry for radio resource management mechanisms in the unlicensed spectrum [79]. The scope of this chapter is to provide an exhaustive analysis of the impact of different variables and mechanisms on on both cellular and WiFi system performance, while operating in unlicensed spectrum and coexisting one with each other.

In this chapter, we focus on LAA and LTE-U technologies, and provide a detailed study of these technologies, main features, comparison with Wi-Fi technology, description of the model for the simulator and a detailed evaluation study of the main variables that impact the system performance.

7.1 Introduction

According to a recent study commissioned by the WFA [211], between 2020 and 2025 users worldwide are likely to experience a spectrum shortfall. As a result of the increasing demand for traffic and bandwidth, mobile operators are increasingly interested in deploying a complementary access utilizing unlicensed spectrum. There has been a special interest lately for accessing the 5 GHz band, traditionally mainly used by Wi-Fi technologies, with LTE. This has generated the definition of novel LTE-based access technologies capable of operating in unlicensed spectrum, while coexisting with other technologies [48], [47], [212]. Moreover, recently, the use of unlicensed spectrum above 6 GHz has attracted a lot of attention by industry, regulation and standardization bodies.

With emphasis on the 5 GHz band, depending on the radio access technology used to access the unlicensed spectrum, these technologies can be divided into two main groups: 1) technologies based on integration of LTE and Wi-Fi radio links and using Wi-Fi to access the unlicensed spectrum, and 2) technologies using LTE RAN in unlicensed spectrum. The integration of LTE and Wi-Fi radio links has been proposed since 3GPP Release 13 [213]. Examples of these technologies are: LWA and LWIP. As for unlicensed LTE technologies, using LTE RAN to access the unlicensed spectrum, their main challenge is the fair coexistence with other wireless technologies operating in the same band. While LTE is designed to have an exclusive access to channel and perform in uninterrupted and synchronous fashion, the existing unlicensed technologies operate in a decentralized, asynchronous manner employing protocols typically based on carrier sensing in order to achieve a fair usage of the spectrum. Some of the challenges of such coexistence scenario are explained in [214], [3].

Therefore, a critical requirement for the design of unlicensed LTE is that it has to coexist with other technologies, like Wi-Fi, on a "fair" and "friendly" basis [48, 215], by extending its synchronous design. In some markets, like Europe and Japan, a LBT feature for CCA before accessing the 5 GHz unlicensed channel is required, while in others, such as the USA, China, India and Korea, there is no such requirement. For markets that do not require LBT, the industrial consortium LTE-U Forum specified a proprietary solution for unlicensed LTE based on Release 12, which is referred to as LTE-U [70]. On the other hand, to meet LBT requirement, 3GPP has produced in Release 13 [48] LAA specification, for Supplemental DownLink (SDL) in unlicensed band. In Release 14, the uplink operation was also defined, in the context of eLAA [74] and new features are under definition in the feLAA Study Item [75]. Several products have already been presented at Mobile World Congress 2017 and 2018 by companies such as Qualcomm, to reach the 1 and 2 Gigabit LTE, respectively, with and without aggregation of unlicensed bands. On the other hand, recently, the Multefire alliance [212] proposed a solution for unlicensed LTE that operates in a completely stand-alone manner in unlicensed spectrum by capitalizing on 3GPP Release 13 and 14 LAA.

In this work, we focus on LAA and LTE-U since they represent the most promising and widespread LTE-based unlicensed technologies. This is because these two technologies use the same RAN in both licensed and unlicensed spectrum, which allows a unified mobility, authentication, security and management. Additionally, since they leverage CA with the licensed carrier, they also guarantee wide-area coverage and the QoS typical of the licensed carrier. We carry out a detailed evaluation study of these technologies, and compare their performance towards their users and in terms of coexistence. In particular, in many commercial

deployments around the world, currently LAA is used as supplemental downlink to deliver Gigabit LTE.

There is a common belief that LAA, thanks to the LBT feature, is in general superior to LTE-U when it comes to coexistence performance with Wi-Fi. We show in this work that the reality is not so straightforward and we try to deeply understand the limits and strengths of each of the two technologies. A key challenge to evaluate and compare these technologies is that despite the large body of simulation results by industry [47, 48] and in the literature [216], [217], [218], the simulators are not publicly available, the two technologies have been evaluated in a standalone fashion, and they have not been compared yet over the same scenarios and the same simulation or test platforms. As a consequence, the obtained results are nor reproducible, neither comparable, and system performance metrics are presented without much details revealed about the underlying models and assumptions. Few analytical models have been proposed in the literature to study channel access of both technologies [219], [220], and [221]. However, these models employ different assumptions, which limit their capability to evaluate the impact of some of the key coexistence parameters.

In order to perform a coexistence study and comparison of LAA and LTE-U technologies we have built a detailed simulation platform, strictly complying with LTE-U Forum and LAA 3GPP specifications, extending the popular open source network simulator ns-3 [13]. This allows access to the full configuration of the system (i.e., from the application to the network interface) and the reproducibility of results. The simulator that we have built based on ns-3, allows to reproduce both 3GPP and evaluation approaches, and differently from any other simulator used in literature or in 3GPP studies, allows a full protocol stack simulation and an end-to-end performance evaluation. Based on this simulation platform, we have carried out an extremely detailed simulation campaign, analyzing many aspects affecting coexistence, and comparing LAA and LTE-U coexistence performance. This has allowed us to reach meaningful conclusions, which are discussed throughout this Chapter. The work on simulator has been supported by the WFA and by a small cell vendor, SpiderCloud Wireless, intensively working in unlicensed spectrum, and so it has been designed in close consultation with industry. The implementations of LAA and LTE-U models along with the documentation, and the simulation scenarios presented in this Chapter are publicly available at [14] to facilitate results reproducibility and further collaborative developments.

In the following, we first provide a detailed description of state-of-the-art contributions in the area of performance evaluation of LAA and LTE-U technologies in Section 7.2. In this chapter we focus on unlicensed LTE implemented as a technology anchored to licensed spectrum, and in particular on LAA and LTE-U technologies. Hence, for the reader's convenience, we provide an overview of these technologies in Section 7.3 and Section 7.4, which will be useful to better understand the simulation models presented in the rest of this chapter, the nomenclature and the evaluation results. In Section 7.5, we discuss the main differences with respect to the Wi-Fi access technology. Section 7.6 provides descriptions of the LAA and LTE-U models that we implemented in the simulator. Section 7.7 provides description of the evaluation methodology including the definition of the fairness definition, deployments, and performance metrics of interest. Section 7.8 provides a description of the validation of the simulator against the results obtained from the testbed. In Section 7.9 and Section 7.10, we describe the performance evaluation of LAA and LTE-U, respectively. Section 7.11 provides comparison of LAA and LTE-U peformance. Section 7.12 provides a summary of our findings and suggestions for future work. Section 7.13 concludes this chapter.

7.2 Related work on LAA and LTE-U evaluation studies

Due to the increasing interest in unlicensed LTE, academia, industry, and standardization bodies, have dedicated lately a lot of work to investigate different spectrum sharing mechanisms for unlicensed LTE and 5G technologies [222], [223] and [224]. However, not much work is devoted to compare the two predominant unlicensed LTE technologies, LAA and LTE-U. As we mentioned earlier, it is generally considered that LAA is fairer to Wi-Fi since it uses an LBT mechanism, but nothing can be found in literature to confirm or deny this common assumption. In this section, we describe available work in the literature, specifically focusing on the performance evaluation of the LAA and LTE-U.

Some of the very first performance evaluations in the literature were mainly focused on evaluating the effects of LTE on Wi-Fi without using any coexistence mechanisms by LTE, e.g., [225]. This study is mostly focused on the interference effects, and the impact of Physical Layer (PHY) parameters such as LTE bandwidth, carrier frequency, the impact of Wi-Fi CCA threshold and MIMO onto coexistence. In [226] the authors propose an analytical model and perform an experimental evaluation, where the LTE device does not implement any coexistence mechanism, but it instead implements a centralized mechanism for coordination of Wi-Fi and LTE. Some other studies evaluate only a single spectrum sharing mechanism, i.e., either the LBT [216], [227] or the LTE-U duty cycle [228].

The performance evaluations carried out by the industry also consider only a single technology, since they were carried out as a part of either LTE-U or LAA study. The results for LTE-U performance evaluation are shown in [47], and for LAA in [48]. In both cases, the performance evaluations were carried out with either proprietary simulators or testbeds that are not publicly available. Also, each of these studies established different methodologies, scenarios, traffic models and performance evaluation metrics. Because of this, it is very hard to draw conclusions on which technology performs better. As a result, we consider that, in order to compare the two technologies in an unbiased manner, it is crucial to establish a unified framework under which they can be evaluated with the same conditions.

To the best of our knowledge, only the authors in [217,219–221] consider both unlicensed LTE technologies. However, in [217,220,221], [229] the authors evaluate LTE-U leveraging a fixed duty cycle scheme, which we do not consider as an appropriate benchmark, or state-of-the-art Carrier Sense Adaptive Transmission (CSAT) scheme for LTE-U, since the duty cycle has to be adaptive based on the activity observed on the channel. An example of highly performing CSAT, has been proposed by Qualcomm [230, 231]. In addition, contributions in [217], [220] and [221] do not consider an exponential backoff in LAA LBT, which though is what it is agreed in 3GPP [213]. Only the work in [219] models both spectrum sharing mechanisms and captures the main characteristics of each mechanism. In particular, the authors propose a novel throughput and interference model for heterogeneous technology coexistence in the same spectrum. While the latter proposal represents an effective analytical tool for the evaluation of different channel access mechanisms, its focus is limited to the MAC and to some coexistence parameters. As a result, the insights cannot be used to evaluate the impact of higher layers and end-to-end effects of coexistence.

Taking into account the limitations of the available literature, in this work we carry out a performance evaluation of both technologies in a unified, full-protocol stack and open source simulation framework. In the proposed simulation model, the spectrum sharing mechanisms

Table 7.1 Related work on performance evaluation studies of LAA and LTE-U

Reference work	Spectrum sharing technology	Evaluation methodology	Modeling assumptions
[48]	LAA only	- 3GPP contributions by a large number of small cell vendors. - Follows 3GPP evaluation methodology. - Coexistence scenarios: a) DL-only LAA and DL-only Wi-Fi (Annex B1) b) DL-only LAA and DL+UL Wi-Fi (Annex B2) c) DL+UL LAA and DL+UL Wi-Fi (Annex B3)	Proprietary simulators and testbeds.
[47]	LTE-U only	- Contributions by different vendors. - Various real-world scenarios: indoor/outdoor, high/low density configurations.	Proprietary simulators and testbeds.
[227]	LAA only (LBT with a fixed CW size)	- Stochastic model based on Markov chain. - Scenario with 1 LAA and 1 AP, and LAA coexisting with another LAA. - Simulations for different traffic rates.	- No exponential backoff in LAA; fixed CW size. - LAA and AP can always sense each other. channel knowledge. - Wi-Fi like model for LAA collision detection and error recovery.
[216]	LAA only	Simulations based on a 3GPP evaluation methodology.	Proprietary simulator.
[228]	LTE-U only (ALOHA-like CSAT)	- A stochastic geometry based model. - Only persistent downlink traffic.	- LTE-U employs ALOHA-like random access scheme. - Interference limited regime. - Perfect energy-detection. - Error model assumption: The transmission is successful if SINR is above a predefined threshold.
[219]	LAA and LTE-U	- A throughput and interference model for inter-technology coexistence. - Indoor/outdoor combinations of scenarios with/without internal walls. - Only downlink saturated traffic.	- Adaptive duty cycle modeled as ideal TDMA MAC. - Simulation limitation: 1 UE per AP.
[220]	LAA (no exponential backoff) and LTE-U (fixed duty cycle)	A throughput model for LTE and Wi-Fi coexistence.	- LTE-U duty cycle fixed. - No exponential backoff in LAA and Wi-Fi access. - Wi-Fi senses the channel busy when LTE transmits. - Fixed coding, 64 QAM for both, LTE and Wi-Fi. - Fixed duration of LTE transmissions.
[221]	LAA (LBT with a fixed CW size) and LTE-U (fixed duty cycle)	- A stochastic geometry based model. - Downlink saturated traffic only.	- Free space path loss model. - LBT fixed CW size. - LTE-U with a static muting pattern, synchronous and asynchronous eNBs patterns. - Channel access priority model options: 1) Wi-Fi and LAA equal priority; 2) LAA has lower priority than Wi-Fi by by setting a larger backoff timer for LAA.
[217]	LAA (LBT with a fixed CW size) and LTE-U (fixed duty cycle)	- Various indoor and outdoor setups. - CSAT muting subframes simulated with almost blank subframes.	- Proprietary simulator. - LBT fixed CW size. - Fixed duty cycle.
[229]	LAA and LTE-U	- 3GPP indoor scenario. - Standard LTE-U implementation without CSAT.	- ns-3 simulator. - LBT dynamic CW size. - LTE-U fixed duty cycle.

are defined according to LTE-U Forum and 3GPP specifications, respectively. As we will explain further in this Chapter, we consider a variety of scenarios for evaluation, being directly inspired by both, 3GPP LAA evaluation study and WFA test plan methodologies, in order to be as complete as possible. We show a summary of related work in Table 7.1.

7.3 LAA technology background

In the following, we discuss the main features of the LAA technology, and in particular:

- LBT procedure
- Energy detection method
- Contention Window (CW) adjustment procedure
- Discovery Reference Signal (DRS)
- LAA reservation signal
- LAA partial subframe

Table 7.2 LAA channel access priority classes

Channel access priority class	m_p	CW_{min}	CW_{max}	MCOT	Allowed CW sizes
1	1	3	7	2 ms	3, 7
2	1	7	15	3 ms	7, 15
3	3	15	63	8 or 10 ms	15, 31, 63
4	7	15	1023	8 or 10 ms	15, 31, 63, 127, 255, 511, 1023

7.3.1 LBT procedure

The main feature of LAA, which distinguishes it from LTE-U, is the support of LBT. LBT is a mechanism with which an equipment applies a CCA to check the availability of the channel before transmitting on it. During the study item on LAA [48], different channel access mechanisms were evaluated, with and without random backoff and with and without variable CW. Category 4 mechanism, considering both random backoff and variable CW was selected for Physical Downlink Shared Channel (PDSCH) transmission. The eNB is allowed to transmit after sensing the channel to be IDLE during the initial CCA defer period T_{d_cca} or after performing the Extended CCA (eCCA). The defer duration T_{d_cca} is composed of duration T_f = 16 μs and m_p CCA slots, where each CCA slot duration is T_{sl_cca} = 9 μs, as explained in Section 15.1.1 in [74]. The value of m_p depends on the channel access priority class, which is used to categorize the type of traffic scheduled in the unlicensed band, as specified in Table 7.2. For example, for the priority class 3, T_{d_cca} is 43 μs. If the channel is idle during T_{d_cca}, the eNB transmits the packet burst. Otherwise, the eNB shall perform the eCCA, during which it first draws a random value of a backoff counter N in the range of [0, CW_p] and starts sensing the channel for an eCCA defer duration, T_{d_ecca}. The CW_p is the current CW size, which ranges between CW_{min} and CW_{max}, also specified by the priority class, as shown in Table 7.2. Once the channel is free for T_{d_ecca}, the eNB senses the channel to be idle for the duration of additional N eCCA slots, and each time that the channel is detected to be idle for a period of one eCCA slot, T_{sl_ecca} = 9 μs (equal to T_{sl_cca}), the backoff counter is decreased by one. The channel is considered to be idle during T_{sl_ecca} if the detected power is lower than the energy detection threshold ED_{th} for at least 4 µs [74]. Otherwise, the slot is considered to be busy. If during the backoff process, the eNB detects that the channel is occupied, the backoff counter is frozen, and the eNB continues to sense the channel until it finds it to be idle for T_{d_ecca}. Once the backoff counter reaches zero, the eNB occupies the channel for a Transmission Opportunity (TxOP), no longer than the Maximum Channel Occupancy Time (MCOT), which depends on the priority class. An LAA eNB can occupy the channel up to 10 ms in case of Best Effort (BE) and Background (BK) traffic, i.e., priority classes 3 and 4, respectively. For other types of traffic, which require a higher quality of service, the length of the MCOT is shorter [74].

7.3.2 ED method

The LBT procedure is based on the Energy Detection (ED) method to determine whether the channel is free to use. The ED method is a CCA mechanism which attempts to determine if the medium is busy by measuring the total energy a device receives. If the received energy is above a certain *ED threshold* the medium is considered busy. Because of this, the ED threshold is one of the most critical parameters to be set for fair coexistence. The maximum configurable ED threshold is defined by the regulations for each region. According to 3GPP agreements the

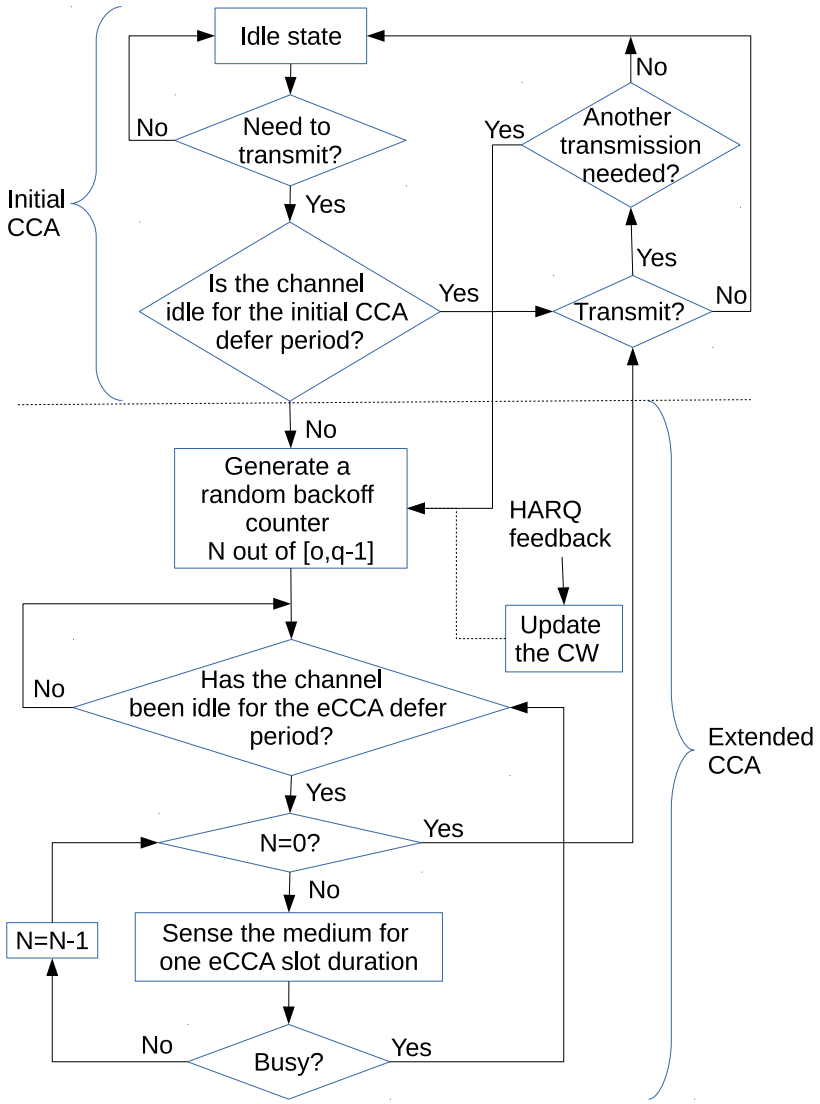


Figure 7.1 LBT 3GPP Category 4 algorithm

established ED threshold for LAA should be -72 dBm for 20 MHz channel. The ED threshold should be adaptive, based on the mechanism defined in Section 15.1.4. in [74].

7.3.3 CW adjustment procedure

3GPP provides a description of CW adjustment procedure in [74]. The initial value for CW window is defined by the priority class, as shown in Table 7.2. The CW size is increased upon collisions, which in LAA are detected by means of Hybrid Automatic Repeat Request (HARQ) feedbacks from a receiving node. In particular, the CW size at the eNB is increased if more than Z percentage of HARQ feedbacks corresponding to the PDSCH transmission in reference subframe k are determined as Negative Acknowledgement (NACK)s [74]. The default value of the Z parameter is 80%. Otherwise, the CW size is reset to the minimum value (CW_{min}). The reference subframe k is typically the first subframe of the most recent transmission burst

for which some HARQ feedback is available.

7.3.4 DRS

The DRS is introduced in Release 12 to allow functionalities such as discovery of the LTE cell, synchronization to the LTE cell, and to perform Radio Resource Management (RRM) measurements. The DRS includes several signals like Primary Synchronization Signal (PSS), Secondary Synchronization Signal (SSS) and Cell-Specific Reference Signal (CRS), and if configured may have Channel State Information - Reference Signal (CSI-RS). The DRS is transmitted for both cells in ON-state and cells in OFF-state. A cell in OFF-state transmits a DRS so that the UE can detect, measure and report it to the network for efficient RRM functionalities. The Release 12 DRS is transmitted with a fixed periodicity of 40, 80 or 160 ms within the Discovery signals Measurement Timing Configuration (DMTC) occasion. The fixed periodicity is relaxed for LAA, to make it compatible with the LBT requirement.

7.3.5 LAA reservation signal

Legacy LTE transmissions are synchronous and start at a subframe boundary. However, the LBT procedure can be completed at any time. Because of this, other neighboring systems with no such restriction, like Wi-Fi, can take a transmission opportunity while the LAA eNB is waiting until the next subframe boundary to start transmitting. To solve this, LAA reserves the channel by transmitting reservation signals after LBT is successful and the access is granted, and until the subframe boundary (or slot boundary, see section 7.3.6). Note that the reservation signal spares channel occupancy time since its duration counts against the MCOT, and so decreases the spectral efficiency and increases unfairness towards Wi-Fi. We illustrate this in Fig. 7.2. To reduce inefficiencies caused by the LAA channel reservation, the partial subframe is proposed for LAA SCell.

7.3.6 LAA partial subframe

The partial subframe is a subframe with a shorter duration than 1 ms. In particular, in case the LBT is successful in the middle of the subframe, the data transmission can start at the slot boundary (1 subframe has 2 slots of 0.5 ms each). Also, depending on the starting time of the transmission and due to the MCOT limitation, the grant may not end at the subframe boundary. To support such partial subframe transmissions for LAA, a new frame structure, Type 3, was introduced in Release 13 [232]. According to this frame structure, the transmissions can start at the slot boundary (symbol 0 or 7), and can end either with the full subframe or with a partial subframe following one of Downlink Pilot Time Slot (DwPTS) durations. The DwPTS is the downlink portion of the subframe of Type 2 for TDD. By using existing DwPTS, the duration of the transmission burst can be 3, 6, 9, 10, 11, or 12 OFDM symbols. The ending partial subframe can be utilized by reusing the DwPTS structure with almost no changes to the existing LTE. The partial subframe increases the complexity of the scheduling. To reduce the complexity, it is possible to limit the symbols at which the transmission can start/end. However,

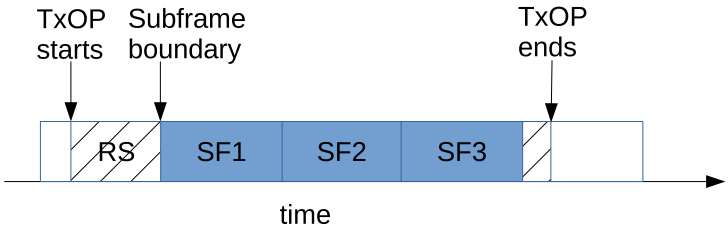


Figure 7.2 An example of MCOT of 4 ms and a reservation signal (RS) being sent upon a successful LBT until a subframe boundary. Subframe (SF) duration is 1 ms. The part of the MCOT since the end of the last full subframe in MCOT (SF3) until the end of the MCOT is not used for transmission.

limiting the number of possible options reduces the efficiency. More information can be found in [233, 234].

7.4 LTE-U technology background

LTE-U refers to the unlicensed LTE technology that was proposed and specified by the LTE-U Forum in [231, 235]. The LTE-U Forum was formed in 2014 by Verizon in cooperation with Alcatel-Lucent, Ericsson, Qualcomm Technologies, Inc., and Samsung, with the objective of collaborating together and generating technical specifications for LTE-U technology. These specifications support LTE operation in the 5 GHz UNII-1 and UNII-3 bands as SDL carriers, in conjunction with an LTE deployment in licensed bands, based on already published 3GPP Release 10 and later specifications. What characterizes LTE-U is the duty-cycling nature of the channel access procedure and the fact that it lacks LBT mechanism. When it comes to transmission according to the duty cycle pattern (i.e., it has an ON-OFF time pattern), LTE-U accesses the channel, regardless of the presence of Wi-Fi transmissions. This kind of the channel access may generate collisions with co-deployed Wi-Fi networks when the duty cycle pattern switches from OFF to ON, and this is the primary concern raised against the LTE-U technology. When the SCell is in ON-state, it transmits as a legacy LTE based on Release 10, 11 or 12. LTE-U SCell differs from the legacy LTE, active in the Primary Cell (PCell), in supporting the following aspects:

- CSAT to access the channel,
- Transmission of LTE-U Discovery Signal (LDS),
- Ability to skip the transmission of Master Information Block (MIB)/System Information Block Type 1 (SIB1), and
- Opportunistic SCell OFF.

7.4.1 CSAT

CSAT is the medium access procedure defined by the LTE-U forum in [231]. Based on CSAT, the Secondary Cell (SCell) first senses the medium to evaluate its occupancy, and then,

depending on this measure, it establishes a duty cycle pattern, i.e., the percentage of time the SCell is to be ON. The duty cycle pattern is periodically adjusted so that the SCell only occupies the channel for a fair share. The LTE-U Forum specification does not define the algorithm according to which the duty cycle is adjusted, and it is left to the vendor to establish its adaptation algorithm. However, the LTE-U Forum defines the following constraints for the CSAT pattern:

- $4ms \leq T_{ON} \leq 20\,ms$
- $1ms \leq T_{OFF} \leq T_{LDS}$,

where T_{ON} is the ON-state duration and T_{OFF} is the OFF-state duration. T_{LDS} is the periodicity of LDS, which will be described in more details in the following subsection. The minimum of 4 ms for T_{ON} is required as long as there is data in buffers and is chosen to avoid frequent OFF/ON state transitions, which are likely to generate collisions. The maximum duration of 20 ms is allowed for uninterrupted transmission of the LTE-U SCell and is chosen to favor coexistence with other technologies. After reaching the maximum uninterrupted transmission duration, the LTE-U has to switch OFF during a period of 1 or 2 ms, which is normally referred to as *puncturing* period. The puncturing subframes are subframes during which the SCell is not permitted to transmit, in order to allow low-latency Wi-Fi traffic to go through. The duty cycle pattern, consisting of alternating T_{ON} and T_{OFF} periods, is adaptive based on the activity sensed during the T_{OFF} times. Additionally, some CSAT implementations, like, e.g., Qualcomm's CSAT [230], define periods during which the sensing is performed specifically to detect the number of active APs that are present on the same channel. This part of the Qualcomm's CSAT method is called *AP scan* [230]. In this work, in the implementation and in the results that are discussed in the following sections, we refer to the CSAT algorithm proposed by Qualcomm. We focus on this algorithm because it is very sophisticated, and it includes several features that favor the coexistence with Wi-Fi.

According to this algorithm, the adjustment of the CSAT duty cycle is performed in the following way. The averaged medium utilization (MU), $\overline{MU}$, is a weighted moving average of Wi-Fi activity over a monitoring window. The MU is obtained by summing up the durations of all Wi-Fi transmissions that were detected by the monitoring Wi-Fi device at the LTE-U node during the corresponding monitoring window. The $\overline{MU}$ is calculated as the weighted moving average in the following way:

$$\overline{MU}(t) = \alpha_{MU} \times MU(t) + (1 - \alpha_{MU}) \times \overline{MU}(t-1), \tag{7.1}$$

where α_{MU} is the weight. According to [230] there are two $\overline{MU}$ thresholds, a lower and a higher, MU_{low} and MU_{high}. If $\overline{MU}$ is greater than MU_{high}, then the T_{ON} is decreased by δ_{down}. On the another hand, if $\overline{MU}$ is lower than MU_{low} then the T_{ON} is increased by δ_{up}. δ_{down} and δ_{up} are CSAT parameters used to adjust the duty cycle, and they directly determine the speed of the convergence of the duty cycle. If $\overline{MU}$ is in between the two thresholds, the CSAT duty cycle remains unchanged. T_{ON} is bounded by the minimum and the maximum value. The minimum T_{ON} value, $T_{ON,min}$, is introduced to guarantee that LTE-U grants a fair share of the spectrum, and the maximum, $T_{ON,max}$ to guarantee that the OFF period is long enough to achieve a proper sensing of Wi-Fi activity. $T_{ON,max}$ is directly determined by the time during each cycle that LTE-U has to be in OFF state to monitor the Wi-Fi activity, $T_{OFF,min}$. Thus,

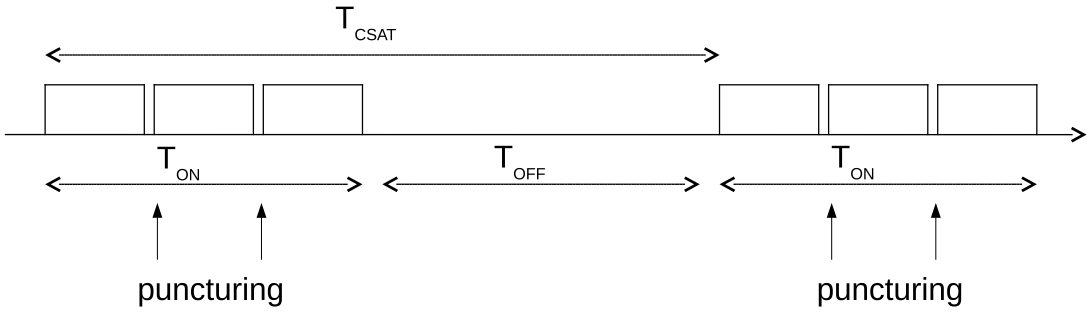


Figure 7.3 An example of CSAT pattern

$T_{ON,max}$ is calculated in the following way:

$$T_{ON,max} = T_{CSAT} - T_{OFF,min} \tag{7.2}$$

On the other hand, $T_{ON,min}$ threshold is adaptable and is determined in the following way:

$$T_{ON,min} = min\left\{C_{min}, \frac{(N_{LTE}+1) \times T_{CSAT}}{M_{LTE} + N_{Wi-Fi} + 1}\right\}, \tag{7.3}$$

where C_{min} is a parameter that is controlling the minimum duty cycle when no Wi-Fi activity is being sensed by the monitoring Wi-Fi device at LTE-U node. N_{LTE} is the number of LTE nodes with the same Public Land Mobile Network (PLMN) ID, and M_{LTE} is the number of all LTE nodes, regardless their PLMN ID. Since in this work we do not consider different PLMN IDs, N_{LTE} is equal to M_{LTE}. N_{Wi-Fi} represents the number of Wi-Fi nodes. N_{Wi-Fi} is obtained while performing the AP scan monitoring, by listening to the beacon signals. AP scan monitoring is independent of MU monitoring. While MU monitoring can last around 20 ms, the AP scan monitoring needs to last for more than 102.4 ms to allow a beacon detection belonging to all neighboring APs. This is because beacons are typically sent with 102.4 ms interval. Fig. 7.3 illustrates an example of CSAT pattern.

7.4.2 Transmission of LDS

The LDS allows User Equipment (UE)s to obtain and keep SCell time and frequency alignment as well as performing SCell tracking and measurements [231]. LDS is a similar message to the discovery reference signal (DRS), introduced in Release 12 for a discovery of LTE cells, in case of sleeping cells, or cells switching OFF due to, e.g., energy saving policies. LDS is broadcasted by the LTE-U eNB, only at subframe 5 occasions, with a fixed time periodicity and with a fixed offset according to the DMTC. The LDS does not contain CSI-RSs. It comprises CRS/PSS/SSS and Physical Downlink Control Channel (PDCCH)/PDSCH for SIB1 transmission. LDS configuration allows periodicities of 40, 80 or 160 ms.

7.4.3 Ability to skip the transmission of MIB/SIB1

The MIB and SIB1 signaling on the SCell allow the detection of eNBs belonging to other PLMNs. MIB and SIB1 shall be transmitted on SCell only when the MIB/SIB1 transmission

period overlaps with a SCell transmission (SCell ON-state). However, the MIB has a minimum allowed periodicity, which is typically 160 ms. So, if it is not transmitted during this period, the LTE-U cell has to access the channel and send the MIB, without waiting for the ON-state. The duration of the MIB transmission (usually 1 ms) is counted towards the ON time.

7.4.4 Opportunistic SCell OFF

SCells in an unlicensed spectrum have to be active only when traffic and coverage conditions really require it, and there is a clear benefit in offloading traffic in these bands. As a result, the SCell has to be turned OFF when there are no UEs in the SCell coverage area, or when the eNB does not have any data in its buffers for the UEs in the SCell coverage. The SCell configuration and activation procedures follow 3GPP Release 10 specifications. When the SCell is activated, the UE monitors the Downlink (DL) subframes at least from the 8th subframe after the subframe including the activation MAC control element command, as in case of the legacy LTE SCell. On the other hand, when the SCell is deactivated, the UE is only required to monitor LDS signals. UEs are not required to be activated for an LTE-U SCell in order to demodulate and perform measurements based on the LDS.

7.5 Differences between Wi-Fi and unlicensed LTE technologies

Even if LAA adds to LTE the fundamental LBT functionality, perfectly imitating the WiFi behaviour, some intrinsic differences between LTE and Wi-Fi, will make that the technologies occupy the channel differently. In this section, we discuss the main technological differences between Wi-Fi and unlicensed LTE technologies, which may impact coexistence performance. This section will facilitate the reader to understand the different coexistence behaviour that will be discussed in the evaluation sections. Differently from other works [223], we focus here on both LAA and LTE-U.

7.5.1 Channel access

Even if Wi-Fi and LAA have similar LBT/CSMA approaches, the way they access the channel is different, due to their technological intrinsic differences. First of all, Wi-Fi can start its transmission at any time, whereas LTE is limited by its frame structure, requiring the use of reservation signals (as introduced in Section 7.3.5). The inefficiency in LAA channel occupancy caused by the reservation signals can be reduced by leveraging the partial subframe feature. However, this further complicates the logic of the MAC and PHY protocols. In case of LTE-U, since there is no LBT, there is no need to transmit any reservation signal or to use partial subframes. The LTE-U transmission always starts at the subframe boundary, so it does not introduce inefficiencies. However, if a Wi-Fi node is transmitting, the probability of collision between LTE-U and Wi-Fi is higher than between LAA and Wi-Fi.

7.5.2 Detection procedures

As we mentioned earlier, LAA uses ED to determine whether the channel is free to use. On the other hand, Wi-Fi uses both ED and preamble detection (PD). In particular, LAA energy detects Wi-Fi at -72 dBm, while Wi-Fi energy detects LAA at -62 dBm and preamble detects other Wi-Fi devices at -82 dBm. As a result, in situations where LAA and Wi-Fi sense each other in the range of -62 dBm and -72 dBm, LAA will defer its access, while Wi-Fi will access the channel considering it to be idle. These setups are not only problematic for LAA, but also for Wi-Fi, since they can result in a variety of hidden node scenarios. To solve this issue, one of the proposals from 3GPP RAN1 was to consider adopting a lower ED threshold (-72 dBm) for new Wi-Fi systems, like 802.11ax. IEEE 802 declined this proposal considering that this would put in disadvantage 802.11a/n/ac systems which are using an ED threshold of -62 dBm, and also taking into account that LAA is not detecting 802.11 preambles at -82 dBm. Additionally, it is under consideration whether LAA shall transmit CTS-to-Self control frames to reserve the medium. This would allow Wi-Fi systems to detect and defer to LAA systems at -82 dBm. However, if Wi-Fi can preamble detect LAA, then it would be fair that also LAA can detect Wi-Fi preambles. The use of CTS-to-Self has been evaluated in Release 14, but its introduction was rejected since it may require installing a Wi-Fi card inside the LAA device.

7.5.3 Coexistence among unlicensed LTE technologies

The addition of LBT procedure to LAA has implications on the performance of LAA, when coexisting with another LAA. It can easily happen that one LAA defers its access to the channel when it detects that another LAA is transmitting. However, this may be undesirable since LTE successfully manages the frequency-reuse 1. Thus, if LBT prevents multiple LAA nodes from transmitting simultaneously, this may degrade the spectral efficiency of the LAA network. On the other hand, in case of LTE-U, there are two options. LTE-U nodes of the same operator can be synchronous, which means that they transmit and are in OFF-state simultaneously, or they can be asynchronous. The first option requires a central network entity that synchronizes the duty cycles. So, it requires more implementation effort, but it may increase the spectral efficiency. On the other hand, if the nodes are asynchronous, it is to expect that the spectral efficiency will drop, but no implementation effort is required.

7.5.4 Contention window update

The CW size in both LAA and Wi-Fi is increased, if a receiving node is unable to decode the packet due to errors caused by poor channel conditions or collisions. The main difference between the CW update procedures of these two technologies lies in how a transmitter identifies a collision, and how a receiving node recovers the erroneous packet. As we mentioned earlier, in LAA, the CW is increased if the percentage of NACKs for a single reference subframe (usually the first subframe) of a transmission burst, is bigger than a predefined threshold. The standard suggests to set the threshold to 80% [74]. Setting such a high threshold value may cause LAA to be unable to detect collisions, especially in dense scenarios. For example, if collisions occur in a dense scenario, it may happen that only a small portion of users experience them. Since a high number of UEs can be scheduled in a single subframe, if only a relatively

small portion of UEs has reported NACKs, it may happen that the 80% threshold is not reached, even if collisions have happened. In this case, the LAA eNB would ignore these collisions, and would not increase its CW, so that potentially needed corrective actions would not be taken. On the other hand, Wi-Fi increases its CW size when an Acknowledgment (ACK) or a Block Acknowledgment (BA) for the packet/s transmitted to a single user is not received. The Wi-Fi CW update scheme is then considering equally the feedback coming from different Wi-Fi Station (STA)s. This difference in channel access mechanism can cause that under specific conditions LAA updates the CW significantly less often than Wi-Fi [236].

7.5.5 Collision detection mechanism

In LAA, HARQ feedback is chosen to detect and declare collisions. However, HARQ does not necessarily reflect collisions. Additionally, LTE HARQ procedure is based on soft combining techniques, i.e., Incremental Redundancy (IR) and Chase Combining (CC), according to which the failed transmissions are not wasted, but combined with the retransmissions. As a result, it may happen that an unsuccessful retransmission, due to a collision, does not result in a NACK, because the combined information with a previous transmission is enough for the UE to decode the data successfully. In contrast, the Automatic Repeat reQuest (ARQ) mechanism that is used in Wi-Fi leverages acknowledgments and timeouts to detect the error. Thus, it always discards the data with errors and asks for a new transmission. For example, if the sender does not receive an ACK before the timeout expires, it will usually retransmit the frame/packet until the sender receives an ACK or exceeds a predefined number of retransmissions. Therefore, due to the efficiency of a soft combining technique used in the LTE HARQ procedure, it may happen that, Wi-Fi detects more collisions than LAA, and consequently it increases the CW more often [236].

7.5.6 Collision detection delay

LAA uses the HARQ feedback corresponding to the starting subframe of the most recent transmission burst. One of the reasons to select the starting subframe as a reference subframe was to minimize the delay between the collision detection and the update of the CW size. Still, the difference in collision detection delay between Wi-Fi and LAA is significant. The time between the transmission and the corresponding feedback in Wi-Fi is in the order of μs, while in LAA is in the order of ms. Wi-Fi detects a failed transmission SIFS time after the transmission is finished, which is typically 16 μs. In Fig. 7.4, we illustrate the delay between the data transmission by the eNB in subframe n and the acknowledgment by the UE in subframe $n+4$, i.e., the delay is 4 ms [164]. Therefore, an update of CW in LAA is significantly delayed compared to Wi-Fi. The difference in this delay may generate very different reactions of these two technologies to the same channel conditions. For example, in the cases when Wi-Fi packets are small, Wi-Fi may detect various collisions with the same LAA node and update the CW accordingly, while the LAA node would be still waiting to receive the HARQ feedbacks of the starting subframe of the corresponding transmission.

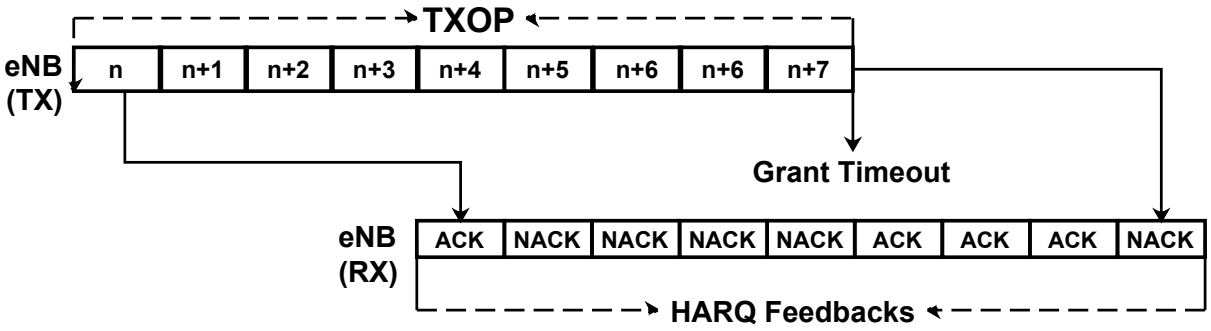


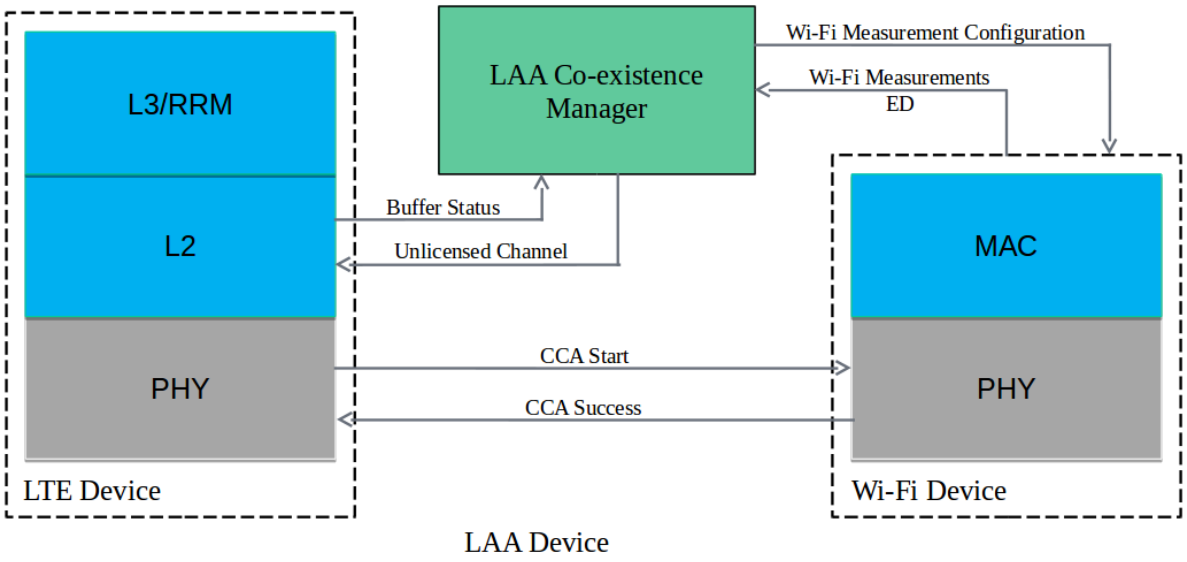
Figure 7.4 LTE eNB transmission and HARQ feedback delay

7.6 LAA and LTE-U ns-3 Models

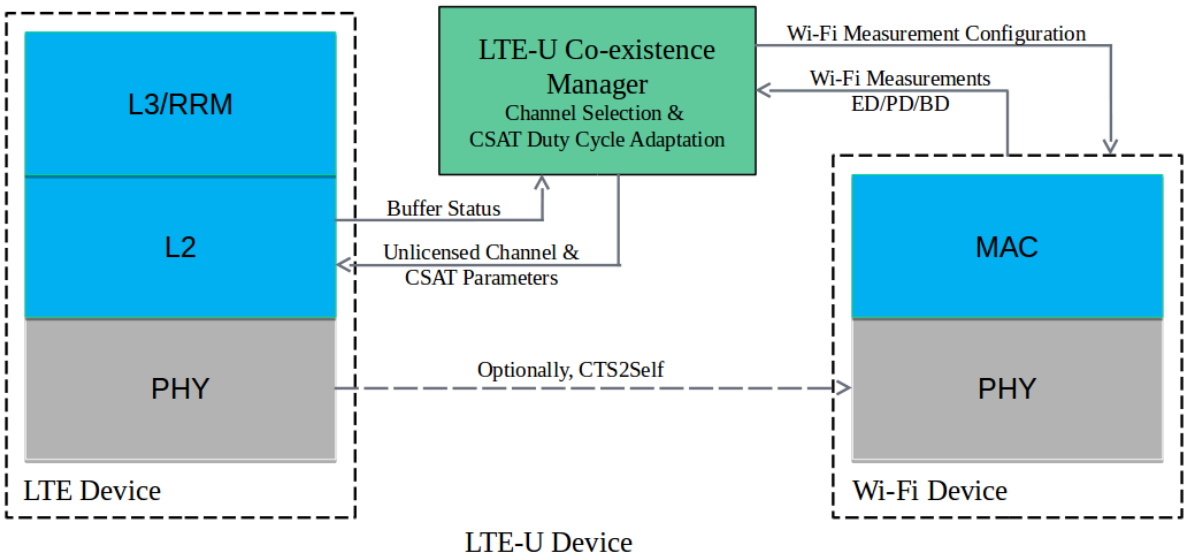
To support LAA and LTE-U evaluations we have developed extensions to ns-3 [13], an open-source system simulator popular in research and academia. As mentioned earlier, the implementation of LAA and LTE-U extensions to ns-3 is publicly available at [14]. Most of the existing coexistence simulators can be classified as link layer simulators with high fidelity models of the channel, PHY and MAC layers, but high levels of abstraction at higher layers. In contrast, ns-3 is a full stack system simulator, with more abstracted PHY layer models with respect to link layer simulators, but with higher fidelity models at higher layers. ns-3 LTE models [184] have been developed according to Femto Application Platform Interfaces (FAPI) by Small Cell Forum and following small cell vendor recommendations. FAPI is an initiative by Small Cell Forum to encourage competition and innovation between suppliers of platform hardware, platform software and application software by providing a common Application Plaftorm Interface (API) around which suppliers of each component can compete. Both LTE unlicensed models, LAA and LTE-U, are leveraging carrier aggregation [4] to allow a simultaneous access through licensed and unlicensed carriers. The ns-3 Wi-Fi models have been developed following the IEEE standards, starting with initial 802.11a models and later extending to many aspects of the 802.11b/g/p/e/n/ac standards. In the following, we discuss the LAA and LTE-U models that we have developed, based on the widely used ns-3 LTE model, and that we will use in the evaluation. Then we go through the evaluation framework, the considered simulation scenarios and validation of coexistence model.

7.6.1 ns-3 LAA model

Fig. 7.5a highlights the architecture of our LAA implementation design. The logic of the channel access operation is placed in a newly introduced entity, referred to as LAA Coexistence manager, which is added to the LTE device, and implements the LBT algorithm following 3GPP specifications as explained in Section 7.3.1. As many actual products in the market, the LAA device in our model contains a Wi-Fi device which is used to perform the CCA based on the ED method. As shown in Fig. 7.5a, the LAA device is composed of a modified LTE device, and of a Wi-Fi device operating in a monitoring mode. LAA coexistence manager is hooked to the MAC and PHY layers of the LTE eNB device to allow the coordination between the MAC scheduler, and the PHY. Every time the LAA entity needs to transmit, it first requests the grant to access the channel from the LBT coexistence manager. The LBT coexistence manager



(a) LAA model



(b) LTE-U model

Figure 7.5 Block diagrams of LAA and LTE-U models for ns-3

performs the LBT procedure, relying on measurements from the Wi-Fi device. Once the LBT has determined that the channel is free to use, it provides a transmission grant to the LAA PHY.

Table 7.3 summarizes the default LAA parameters. The model supports exponential backoff according to the Category 4 design. Initial and extended CCA defer at 43 μs, and the LAA CCA slot time is 9 μs [237]. LAA ED threshold is tunable, separately from the Wi-Fi's threshold, and its value is set to -72 dBm. The maximum length of the TxOP is configurable, and it defaults to 8 ms. The update of the CW is implemented following a HARQ feedback based approach, as agreed in [237, 238].

Data transfer starts at the subframe boundary. We implement reservation signals to occupy the channel and force other nodes to defer, while we are not occupying the channel with data.

Table 7.3 LAA default simulation parameters

LAA parameter	Default value
CCA-ED threshold	-72 dBm
DRS interval	160 ms
MCOT	8 ms
Z threshold	80%
Min CW size	15 slots
Max CW size	63 slots
CCA Slot duration	9 μs
Initial and extended CCA	43 μs

The reservation time is discounted from the transmission opportunity grant time. The DRS is modeled according to 3GPP specifications as discussed in Section 7.3.4. DRS may have a variable position inside the DTMC window, and it should be subject to a priority LBT, with a fixed defer period of only 25 μs, but we consider a normal LBT, as for data. If data is scheduled during the DTMC window, DRS is embedded with data, otherwise, it is sent alone without data, and modeled as a transmission occupying 14 symbols (1 ms). DRS periodicity is configurable. Possible values are 40, 80 and 160 ms, and the default in our tests is 80 ms. Notice that DRS is only modelled to account for its channel occupancy. We assume that DRS impact on synchronization is not evaluated in the model, and UEs do synchronize through the licensed carrier. The system information (MIB/SIB1) is channeled through the PCell. Between the time when the MAC schedules a PDU, and when it actually reaches the air, there are 2 subframes (2 ms) of delay, due to PHY/MAC processing.

7.6.2 ns-3 LTE-U model

Fig. 7.5b shows the architecture of our implementation design of the LTE-U device. The LTE-U device is composed of the modified LTE device where the algorithm handling the access is in the newly introduced entity referred to as LTE-U coexistence manager. Similarly to the LAA model, we introduce a Wi-Fi device in a monitoring mode to LTE-U device in order to support basic Wi-Fi PHY and MAC functionalities. The LTE-U coexistence manager has a central role in the LTE-U model since it performs the CSAT algorithm and it decides the LTE-U ON/OFF patterns. The LTE-U coexistence manager is attached to the Wi-Fi device, which is used for the channel monitoring, i.e., for ED, PD and beacon detection (BD). These measurements are used to adjust the duty cycle parameters. Differently from the LAA model, here the full control over the transmission is in the LTE-U coexistence manager, i.e., the LTE device never requests the grant, but waits for the grants that are generated and notified by the LTE-U coexistence manager. Once the grant is received, the LTE-U device starts transmitting.

The CSAT algorithm that we implemented is inspired by Qualcomm's CSAT, as explained in Section 7.4. Table 7.4 summarizes the CSAT configuration parameters and provides the default values selected in our implementation. We set the default value of duration of AP scan monitoring, T_{apsc}, to 160 ms. During this period the transmission of LTE-U cell is OFF, except for the minimum MIB and LDS required signaling. I_{apsc} represents the frequency of performing the AP scan. This interval is expressed in number of CSAT cycles. E.g., every 8 or

Table 7.4 LTE-U CSAT parameters

CSAT parameter	Description	Default value
d	Duty cycle. Adaptable, but an initial value shall be set.	0.5
T_{csat}	ON + OFF interval. It takes values between 40 and 1280 ms	160 ms
$T_{OFF,min}$	Minimum time to monitor Wi-Fi activity during single T_{CSAT} period	20 ms
$T_{ONminPerTxOP}$	Based on LTE-U Forum specification is the minimum time the LTE-U cell is required to be in ON-state as long as there is data in users' buffer	$4ms$
I_{punct}	The puncturing interval	20 ms
T_{punct}	The puncturing length	1 ms
I_{apsc}	The AP scan periodicity	$16 \times T_{CSAT}$
T_{apsc}	The length of the AP scan period	160 ms
C_{min}	Controls the minimum duty cycle bellow ED threshold	120 ms
MU_{high}	Upper MU threshold	0.6
MU_{low}	Lower MU threshold	0.4
δ_{down}	The delta value used when MU is to be decreased	0.05
δ_{up}	The delta value used when MU is to be increased	0.05
α_{MU}	The weight used for $\overline{MU}$ calculation	0.8

16 CSAT cycles LTE-U cell performs the AP scan. A single CSAT period, T_{csat}, is composed of ON and OFF periods. Standard values of T_{csat} in the industry are 80 ms, and 160 ms, but higher values are also considered. As explained previously, the puncturing is necessary to protect latency sensitive Wi-Fi applications. There are two puncturing parameters, T_{punct} and I_{punct}, the puncturing duration and the puncturing interval, respectively. The duration of puncturing is typically 1-2 ms, and the interval is usually set to 20 ms. The CSAT adaptation is performed per CSAT cycle and MAC is notified in advance to account for PHY processing delay. The adaptation of CSAT cycle is performed at the end of the current CSAT cycle and is applied to the following one.

We model in ns-3 three control signals in LTE-U transmission patterns: MIB, SIB1 and LDS. We model them for channel occupancy purposes. The modeling of the control signals is based on the latest LTE-U requirements defined in [231]:

- MIB/SIB1: As explained in 7.4.3, according to LTE-U specs (section 5.2 in [231]) MIB/SIB1 is transmitted on LTE-U SCell when its transmission period overlaps with a SCell ON-state and at least once every 160 ms. This implies a regular 10 ms transmission for MIB and 20 ms transmission for SIB1 during the ON-state. As a result, whenever the SCell ON period covers the subframe (SF) 0, the MIB is transmitted; whenever the SCell ON period covers SF 5, the SIB1 is transmitted, if it is scheduled in that frame. If MIB transmission is not scheduled during approximately 150 ms, its transmission is scheduled during the following frame, in SF 0, regardless of the SCell to be in ON period. This assures that MIB is transmitted at least every 160 ms.
- LDS: The LDS is defined as an instance of SF5 with CRS/PSS/SSS and PDCCH/PDSCH. As explained in 7.4.2, the UE may assume that the LDS is transmitted at a fixed time periodicity with a fixed offset signaled in the configured DMTC as per 3GPP Release 12 DRS RRC configuration (section 5.3 in [231]). LDS uses the RRC

signaling defined for 3GPP Release 12 DRS. Release 12 DRS RRC configuration allows periodicities, T_{DRS}, of 40, 80 and 160 ms. This means in practice that the LDS is transmitted every $T_{DRS} = T_{LDS}$, and the duration is 1 ms. The transmission time is counted against the overall SCell ON time.

7.6.3 ns-3 LTE Carrier Aggregation

Both, the LAA and LTE-U, models rely on a CA implementation [4] because both technologies are anchored to a licensed PCell. The focus of our design is on the downlink traffic, while the uplink traffic (Physical Uplink Shared Channel (PUSCH), Physical Uplink Control Channel (PUCCH), Sounding Reference Signal (SRS)) is channeled through the primary carrier. The criteria for the CA downlink traffic splitting between PCell and SCell are the following:

- Signaling bearers (SRB0, SRB1) are transmitted over the primary carrier
- Guaranteed bit rate traffic is transmitted only over the primary carrier
- LTE Radio Link Control (RLC) retransmissions are transmitted through the primary carrier
- Secondary carrier is used opportunistically, i.e., only when the PCell is fully occupied

Additionally, the secondary carrier of a UE is activated only when the signal quality of the PCell is better than the activation threshold and after data is available to be transmitted at the secondary carrier. On the other hand, if the signal quality becomes lower than a predefined threshold the secondary carriers are deactivated. If the secondary carriers of all UEs are deactivated, then the transmission over the secondary carrier is disabled, except for the minimum obligatory signaling messages that are configured, e.g., for LTE-U these are MIB and LDS, for LAA the DRS.

In order to receive the data, the UE first needs to decode the Downlink Control Information (DCI), which is sent over PDCCH of the corresponding carrier. We have confirmed that this design is aligned with implementation in real-devices by many vendors. The PDCCH is modeled in the first symbols of the subframe. The number of symbols occupied by the PDCCH is configurable and can take values 1, 2 or 3 symbols. In our implementation it defaults to 1 ms.

In Table 7.5, we show the details of the ns-3 simulator implementation, in comparison to 3GPP [48] and WFA guidelines [215].

7.7 Evaluation methodology and simulation scenarios

7.7.1 Performance evaluation criteria: Fairness definition

The central aspect of the performance evaluation methodology is the definition of the performance criteria, i.e., the measure of fairness of unlicensed LTE towards Wi-Fi. The WFA

Table 7.5 Coexistence scenario configuration according to 3GPP, WFA and ns-3 model

Unlicensed model	**3GPP TR 36.899**	**Wi-Fi alliance guideline**	**ns-3 implementation**
Network layout	Indoor/outdoor scenario	Various simple scenarios	Indoor, outdoor, and simple scenarios
System bandwidth	20 MHz	20/40 MHz	20/40 MHz
Number of carriers	1, 4 (to be shared between two operators), 1 unlicensed for evaluation of DL+UL Wi-Fi coexisting with DL-only LAA	-	1 licensed carrier for LAA DL/UL operation, 1-4 unlicensed for evaluation of DL+UL Wi-Fi coexisting with DL-only LAA
Total Base Station transmission power	18/24 dBm	20 dBm	18 dBm by default
STA transmission power	-	18 dBm	18 dBm by default
Data preamble type	11 n/ac	11 ac	11 n
Distance based path loss shadowing and fading	ITU InH	Computed on the basis of 3-D distance	802.11 ax indoor model
Antenna pattern	2D Omni-directional	2D Omni-directional	2D Omni-directional
BS antenna gain	5 dBi	0 dBi	5 dBi
UE antenna gain	0 dBi	-2 dBi	0 dBi
Number of UEs	Various scenarios	Various scenarios	Supports all 3GPP 36.889 and all WFA guidelines configurations
UE dropping	Randomly dropped within the coverage	Scenario specific	Supports both: random and scenario specific
Traffic model	a) FTP Model 1 and 3 based on TR 36.814 with size of 0.512 MBytes b) optionally: VoIP traffic model	a) FTP traffic model with an exponential reading time and with a model according to which some files are transfered with MTU of 1500 and some of 512 bytes. b) Full buffer traffic with MSDU size of 1500 bytes	a) FTP: FTP Model 1 based on TR 36.814, over UDP and TCP, with file size of 0.512 MBytes b) UDP full buffer traffic of MSDU of 1000 bytes c) VoIP traffic model
UE noise figure	7 dB	9 dB	Supported, 9 dB by default.
LAA cell selection	Based on RSRP (Reference signal received power)	Based on RSRP (Reference signal received power)	Based on RSRP (Reference signal received power)
Wi-Fi AP selection	Based on RSS (Received signal power strength)	Based on RSS (Received signal power strength)	Based on RSS (Received signal power strength)

and 3GPP have in parallel proposed two different evaluation methodologies, which we discuss in this section. The definition of fairness proposed by the WFA in [215] is aligned with the one defined by 3GPP in [48]. In particular, both definitions consider that a deployed system transmitting in an unlicensed channel operates fairly to Wi-Fi if its impact on the Wi-Fi users performance is no worse than the impact that would result from an additional Wi-Fi network introduced into the channel supporting the same traffic load as the deployed system. This definition also determines the evaluation methodology, which typically consists of two steps:

1. The evaluation of the performance of the Wi-Fi network in the scenario in which Wi-Fi coexists with another Wi-Fi network. This step represents the baseline of the evaluation, and in the following, we will refer to this step as "Wi-Fi over Wi-Fi" scenario;

2. The evaluation of the performance of the Wi-Fi network in the scenario in which one of the Wi-Fi networks is replaced by an unlicensed LTE network. We will refer to this step as "Wi-Fi over LAA/LTE-U".

7.7.2 Evaluation topologies

In our simulation study, we consider that Unlicensed LTE (ULTE) is deployed as an SDL system, anchored to a licensed PCell, employing the CA feature. The SCell is deployed in the 5 GHz band, and in particular it is using the Wi-Fi channel 36, which has 20 MHz of bandwidth and is corresponding to 5180 MHz central carrier frequency. The simulator allows for aggregation of up to 4 secondary carriers, either licensed or unlicensed. As for the unlicensed band, we extended the simulator to support additional 5 GHz channels: 32, 40, 44, 48, 149, 153, 157, 161 and 165.

In this study, we focus on two main simulation topologies, Simple and 3GPP indoor. We also simulate two more scenarios, HN simple and BS corners as a variation of the main simulation topologies.

- *Simple*: the simple scenario is composed of two BSs that can be either eNBs or APs, and two users which can be either UEs or STAs. We use this scenario to model WFA's simple model with single clients in which the networks are co-located. To model this scenario we have implemented the simple topology shown in Fig. 7.8a. The distances between the BS and the user, $d1$, and between the two BSs, $d2$, can be configured to model different interference configuration.
- *HN simple*: The HN (hidden nodes) simple scenario is a variant of the simple scenario and we use it to create more specific cases with hidden nodes. To model this scenario we have implemented the modified simple scenario topology shown in Fig. 7.8b.
- *3GPP indoor*: 3GPP TR36.889 proposes two simulation scenarios, indoor and outdoor. Although the simulator supports both, in this simulation study we focus only on the indoor scenario. In fact, the indoor scenario is more challenging than the outdoor for achieving fair coexistence due to close proximity between LTE eNB and Wi-Fi APs and STAs. The indoor scenario consists of two operators deploying 4 small cells in a single floor building. The small cells of each operator are centered along the shorter dimension of the building and they are equally spaced. Fig. 7.6 shows the layout and the dimensions of this scenario. Operators are using the licensed cell PCell, which can be at 3.5 GHz with a bandwidth of 10 MHz and total transmit power per carrier 24 dBm. The unlicensed SCell operates in the 5 GHz carrier with a bandwidth of 20 MHz and a total transmit power of 18 dBm. There are 10 or 20 UEs per unlicensed band per operator, and they are randomly distributed in the rectangular region. All UEs are within the coverage of the small cell in the unlicensed band. In LAA network, the cell selection decision is based on RSRP in the unlicensed carrier, whereas in Wi-Fi, on Received Signal Strength Indicator (RSSI) of Wi-Fi APs.
- *BS corners indoor*: We define a variant of the indoor scenario in which the Base Station (BS)s are placed in the corners. The scenario is illustrated in Fig. 7.7. The objective here is to increase the number of hidden nodes and to study the behavior of ULTE technologies under more challenging interference conditions.

The recommended traffic models by 3GPP are File Transfer Protocol (FTP) Model 1 or FTP Model 3. We carry out the performance evaluation with FTP Model 1. FTP Model 1 is a non-full buffer traffic model, according to which users arrive with a Poisson process with arrival

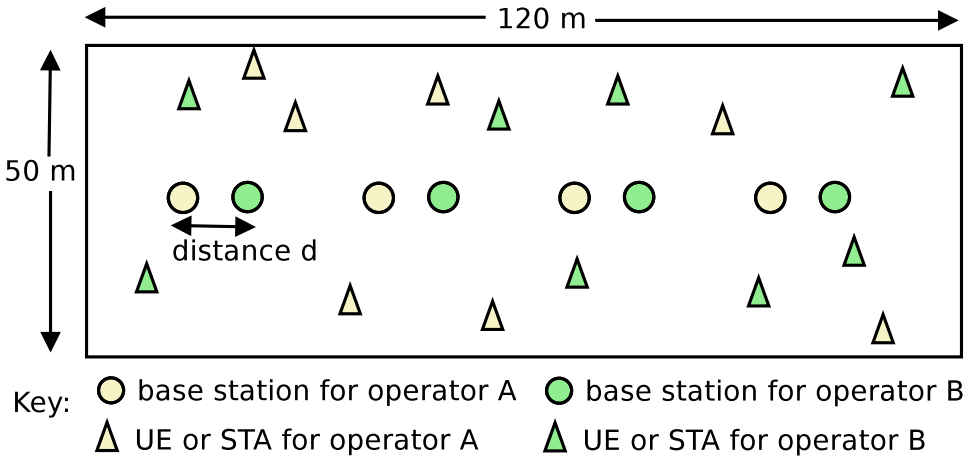


Figure 7.6 3GPP Indoor scenario layout

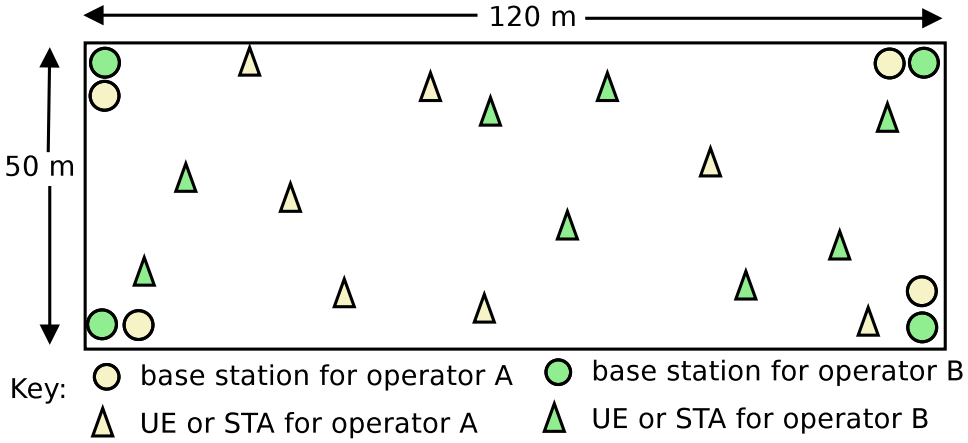


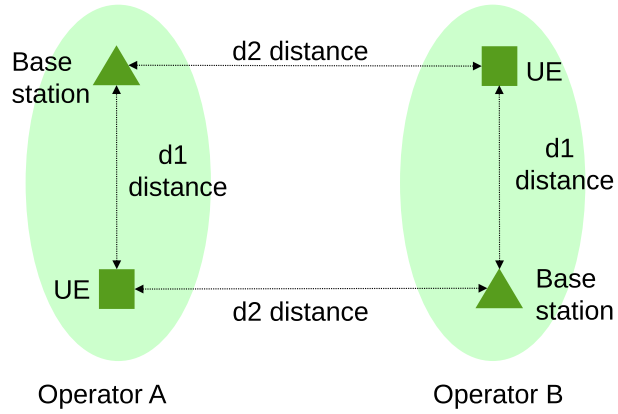
Figure 7.7 A modified version of 3GPP indoor scenario layout in which eNBs are placed in the corners

rate λ. One user downloads a single file which can be 2 Mbytes or 0.5 Mbytes. Parameter λ recommended values are 0.5, 1, 2, 2.5 for file size of 0.5 Mbytes. With this traffic model, the default value of base simulation duration in our simulation campaigns is 480 seconds, which translates into approximately 480 FTP flows per operator for any λ. In simulations with User Datagram Protocol (UDP) Constant Bit Rate (CBR) traffic, since there is a single UDP flow per user, the simulation duration is shorter and by default is set to 20 seconds.

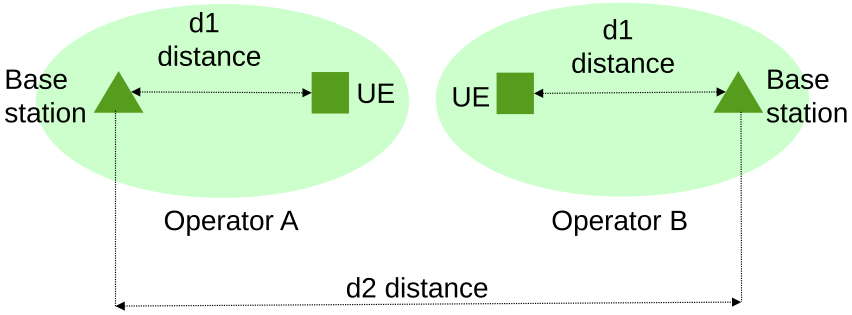
Table 7.6 summarizes common simulation parameters that we use for both LAA and LTE-U evaluations.

7.7.3 Performance evaluation metrics

The key considered performance metrics in this study are "user-perceived throughput" and "latency". In ns-3, we calculate them by using the built-in `FlowMonitor` tool that tracks statistics per flow including throughput and latency, and we then post-process these results to obtain CDFs. To correctly interpret the performance of each UE and flow in various scenario conditions, we also monitor additional metrics and measurements, such as channel occupancy



(a) Scenario Simple



(b) Modified simple scenario with a hidden node effect

Figure 7.8 Simple scenario variants

time by each operator, total number of collisions among different technologies, retransmission rate per Wi-Fi device, percentage of LAA NACK HARQ feedback (i.e., $Z\%$), packet lost, backoff values per each LAA and Wi-Fi device, CW for LAA and Wi-Fi devices, LTE-U duty cycle, medium utilization as detected per every LTE-U device, number of AP detected by each LTE-U device and interference levels among all devices in unlicensed carrier, beacons lost and inter-beacon interval time.

We evaluate performance of different traffic models over both UDP and TCP transport protocols. TCP model in ns-3 integrates advancements made in the real world implementations, having as the main reference the Linux implementation [239]. TCP protocol defaults to TCP NewReno. As for the LAA/LTE-U link layer, we consider UDP over RLC Unacknowledged Mode (RLC-UM) and TCP over both RLC-UM and RLC Acknowledged Mode (RLC-AM).

7.8 Validation of ns-3 coexistence models: A comparison against National Instruments (NI) experimental testbed

In this section, we discuss the validation of the proposed models. In particular, the coexistence simulator was built based on the ns-3 LTE and Wi-Fi modules. These two modules have been previously validated through extensive calibration campaigns and against experimental testbed

Table 7.6 Default common simulation parameters

Parameter	Value
Base simulation duration (FTP Model 1)	480 seconds
Simulation duration (UDP CBR)	20 seconds
Default eNB configuration	1 PCC + 1 SCC
FTP Model 1 file size	0.512 Mbytes
FTP Model 1 lambda for LAA/LTE-U operator	5
FTP Model 1 lambda for Wi-Fi operator	2.5
UDP CBR rate for UE/eNb in scenario simple	150 or 300 Mbps
UDP CBR rate per STA/AP in scenario simple	75 or 150 Mbps
UDP CBR rate per UE in scenario indoor	2 or 4 Mbps
UDP CBR rate per STA in scenario indoor	1 or 2 Mbps
UDP packet size	1000 bytes
TCP initial congestion window	10
TCP segment size	536
TCP RLC mode	RLC-AM
Wi-Fi default CCA-ED threshold	-62 dBm
PDCCH duration	1 symbol (71 μs)

cross-validation. The LTE module was built at CTTC[1], in close collaboration with Ubiquisys (now Cisco) [184]. The LTE module is validated with an extensive set of tests that covers all the main modules/functionalities using the official ns-3 test framework. Calibration campaigns have been performed also in 3GPP reference scenarios [130]. Results show that the ns-3 LTE module achieves similar performance to those obtained by the 3GPP industrial simulators in the evaluated cases, both in terms of SINR distributions and users' throughput. LTE has been also validated against a real world testbed demonstrating that it can deliver voice quality and latency as good as an experimental testbed using actual LTE equipment over a range of signal-to-noise ratios [129]. As for the Wi-Fi module, it was validated against a real testbed in [128].

To further validate the models in the coexistence scenarios, we have also referred to the work in [240]. In this work, the authors propose a new analytical model for throughput of LTE-LAA and Wi-Fi systems in coexistence scenarios, and they validate it against an experimental results using the National Instruments (NI) platform. To perform validation in ns-3, we have implemented a coexistence validation script `lte-wifi-coexistence.cc` that is strictly following the system configuration considered in [240], along with different validation cases and topologies. The mentioned script is publicly available with the simulator that we built, in the repository available in [14]. Following validation strategy from [240], we consider two types of scenarios: Wi-Fi only, which is the baseline for comparison, and Wi-Fi coexisting with LAA. We carry out validation in 3 different scenario topologies where each differs in the number of the interfering AP/LAA nodes: 2, 4, and 6 nodes scenarios. Note that in 4 and 6 nodes scenarios, when Wi-Fi is coexisting with LAA, the number of LAA nodes is 2. As in [240], we further evaluate throughput of Wi-Fi and LAA by configuring 3 different data rates for both, Wi-Fi and LAA, respectively, R_W and R_L. Wi-Fi uses the 802.11a standard with 20 MHz bandwidth, and LAA is configured to use the same 20 MHz channel. LAA PDCCH

[1]The author of this thesis, Biljana Bojović is one of the authors of the LTE module and its maintainer since 2015.

Table 7.7 Validation results

System	NI experiment	Theoretical modeling	Coexistence simulation
2 Wi-Fi	8.0	7.77	8.03
1 Wi-Fi/ 1 LAA	1.80 and 5.22	1.49 and 5.26	1.58 and 5.20
4 Wi-Fi	7.89	7.24	7.69
2 Wi-Fi/ 2 LAA	1.31 and 3.82	1.34 and 4.72	1.48 and 5.52
6 Wi-Fi	7.16	6.90	7.34
4 Wi-Fi/ 2 LAA	1.86 and 3.10	2.01 and 3.56	1.98 and 5.26
2 Wi-Fi	15.0	14.62	14.87
1 Wi-Fi/ 1 LAA	1.69 and 11.09	1.63 and 11.51	1.74 and 11.15
4 Wi-Fi	13.90	13.73	14.27
2 Wi-Fi/ 2 LAA	1.62 and 9.98	1.46 and 10.24	1.54 and 11.54
6 Wi-Fi	13.17	13.12	13.57
4 Wi-Fi/ 2 LAA	2.66 and 7.76	2.31 and 8.19	2.26 and 11.48
2 Wi-Fi	35.40	34.38	35.28
1 Wi-Fi/ 1 LAA	1.40 and 57.80	1.73 and 55.18	1.91 and 54.83
4 Wi-Fi	34.78	34.07	34.53
2 Wi-Fi/ 2 LAA	1.98 and 51.20	1.54 and 48.98	1.94 and 52.94
6 Wi-Fi	33.19	32.85	32.58
4 Wi-Fi/ 2 LAA	2.98 and 43.11	2.57 and 40.99	3.31 and 39.66

occupies 1 symbol. Data traffic is a full buffer and packet length is 2048 bytes. A maximum transmission unit of devices in ns-3 is configured to 2500 bytes to allow the transmission of 2048 bytes without fragmentation that would lead to degradation in throughput performance that is not considered in the analytical model and NI experiment. LAA is configured with channel access priority 3 class (MCOT=8 ms, Min CW = 15 and Max CW = 63), which is also the priority class that is configured in the LAA simulations. In Wi-Fi only scenario Wi-Fi is configured to have Min CW = 15 and Max CW = 1023, while when coexisting with LAA the Max CW of Wi-Fi is set to 63 as in LAA. In Table 7.7, we show the comparison of the results obtained by the simulator versus the theoretical model and experimental platform. We can observe that there is the same trend of Wi-Fi throughput in simulation and theoretical model when changing the date rate and the number of nodes. There is a difference in the trend of LAA performance for lower data rates. Note that according to the analytical model in [240], any partial overlap of the transmitted frames from different nodes results in collisions; however, in experimental and ns-3 platforms the receiver may be able to decode packets which are partially overlapped. Because of this, for lower rates we can notice difference between theoretical results on one side, and experimental and simulation results on the other side. However, this difference reduces when systems use higher data rates, since devices become more sensitive to the interference. Additionally, differently from the theoretical model and experimental platform, in simulator, LAA nodes are synchronized per subframe boundary, so there is a greater probability for them to collide among each other, then with Wi-Fi system. Hence, with lower rates, we do not notice a trend of degradation in performance of LAA when increasing the number of interfering nodes that we can notice in NI experiment and theoretical modeling results. However, for the higher data rates this trend is very clear also in the simulation, and it is the same as in theoretical modeling and experimental results.

7.9 Performance evaluation of LAA/Wi-Fi coexistence

7.9.1 Impact of LAA ED threshold

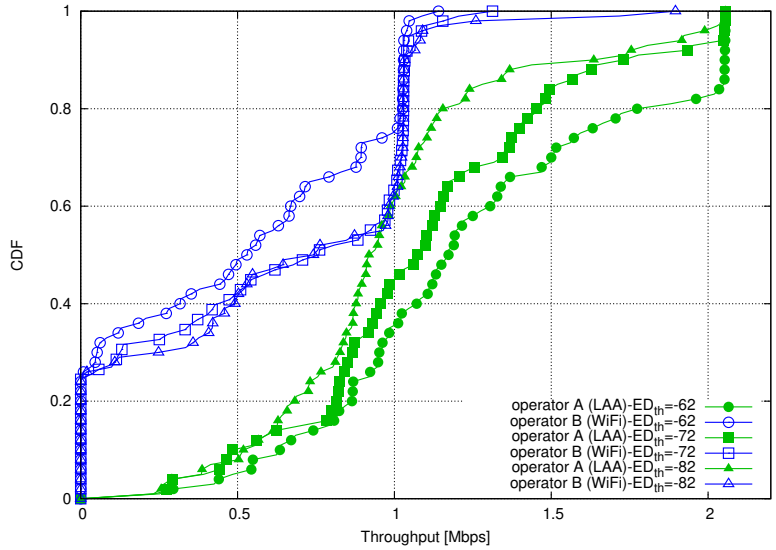
The ED threshold is an essential parameter of the LBT mechanism since it directly influences when the LAA node detects the channel to be busy or idle, and consequently when it transmits. We analyze the impact in different scenarios, the simple and the indoor one, and with two different kinds of traffic, a CBR and FTP traffic. We show that the expected behavior of protecting Wi-Fi simply by lowering the LAA ED threshold is not so direct, and many other Wi-Fi and LAA implementation aspects intervene in determining the coexistence performance of both technologies.

General behavior: We start by analyzing the 3GPP indoor scenario in which the traffic performed is CBR over UDP. CBR rate per UE flow for LAA operator is 2 Mbps, and for Wi-Fi operator is 1 Mbps. First, we consider the ED threshold of LAA at -62 dBm, as Wi-Fi's, and then we lower it to -72 dBm. In Fig. 7.9a and Fig. 7.9b, we show the impact of LAA ED threshold on throughput and latency, and we compare the case of LAA over Wi-Fi, to the baseline Wi-Fi over Wi-Fi. **In this simulation configuration, we observe, as expected, a positive impact of lowering the LAA ED threshold on the Wi-Fi operator, while the LAA operator is negatively impacted. With this traffic type, both operators need to access the channel constantly, and due to the asymmetry in energy detection levels, Wi-Fi has more priority when accessing the channel compared to LAA. Based on the simulation traces, we observe that when lowering the ED threshold, LAA nodes are on average spending more time in backoff, and this makes that the average Wi-Fi throughput increases.**

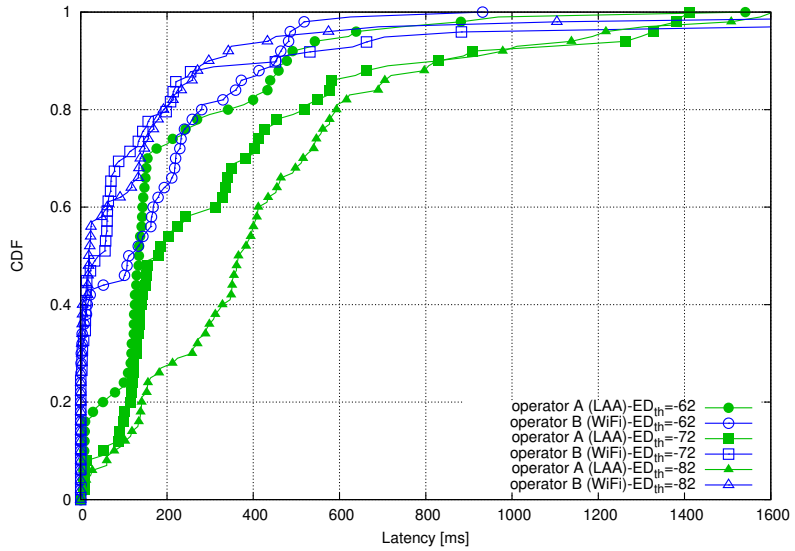
Dependency on the traffic model: We analyze now results with the 3GPP FTP Model 1. We first study the simple scenario case to well understand the multiple aspects that can affect performance. We compare the case when the ED threshold of LAA is set to -62 dBm, equal to that of the Wi-Fi AP, with the case when LAA ED threshold is set to -72 dBm. The distances, $d1$ and $d2$, are set in such a way that the BSs cannot detect each other when the LAA ED threshold is -62 dBm. In particular, $d1$ is 10 , and $d2$ is 30. We show the received power levels in this scenario in Table 7.12.

Note that when the LAA ED threshold is set to -72 dBm, the LAA eNB can energy detect the Wi-Fi AP and STA, but the AP and STA cannot detect the LAA node. We illustrate these two scenarios in Figure 7.10.

In Fig. 7.11, we show the impact of LAA ED threshold on throughput in simple scenario, and we compare the case of LAA over Wi-Fi (Fig. 7.11a) to the baseline Wi-Fi over Wi-Fi (Fig. 7.11b). When the energy detection thresholds of both technologies are set to -62 dBm, we notice a slight degradation of Wi-Fi performance wrt to the baseline case. On the other hand, we observe that when the LAA ED threshold is lowered to -72 dBm, and so LAA properly detects the Wi-Fi transmissions and backs off to them, **there is an unexpected negative impact on Wi-Fi performance.** Getting into the detail of the PHY layer traces we observe that when the LAA ED threshold is decreased from -62 dBm to -72 dBm, the amount of collisions decreases from 16.95% to 11.38%, and the LAA channel occupancy decreases, while the Wi-Fi channel occupancy increases and so its CW size. This means that, even if the overall amount of collisions reduces in the -72 dBm case, the Wi-Fi experiences more failures, and its



(a) Throughput of LAA and Wi-Fi flows

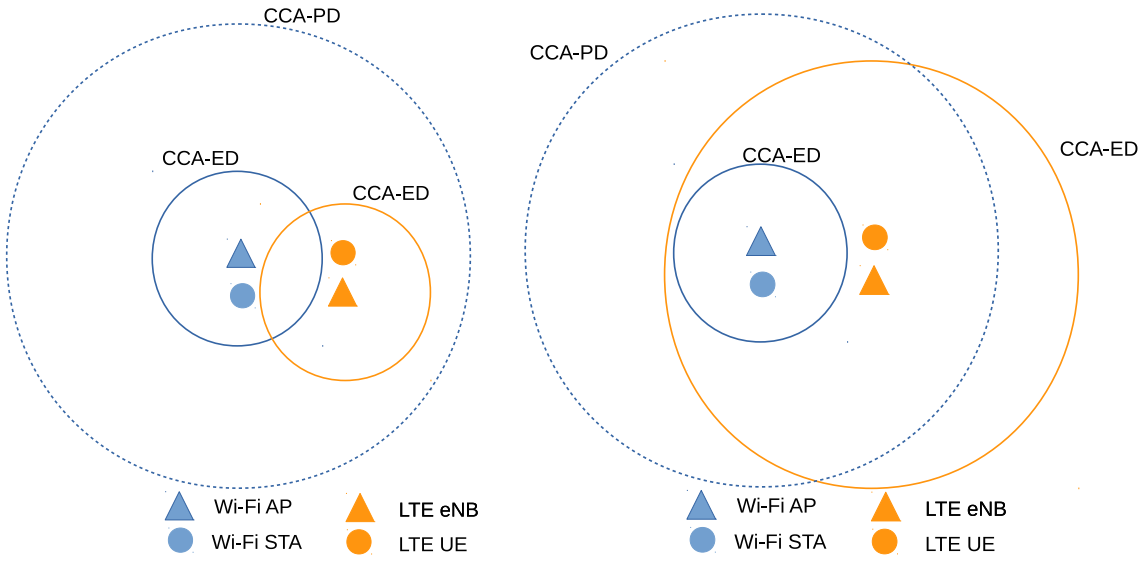


(b) Latency of LAA and Wi-Fi flows

Figure 7.9 Impact of LAA ED threshold, $ED_{th} \in \{-62, -72, -82\}$ dBm, evaluated in 3GPP indoor scenario, when the traffic is the UDP CBR, and with the LAA rate of 2 Mbps and the Wi-Fi rate of 1 Mbps.

performance is more affected compared to the -62 dBm case.

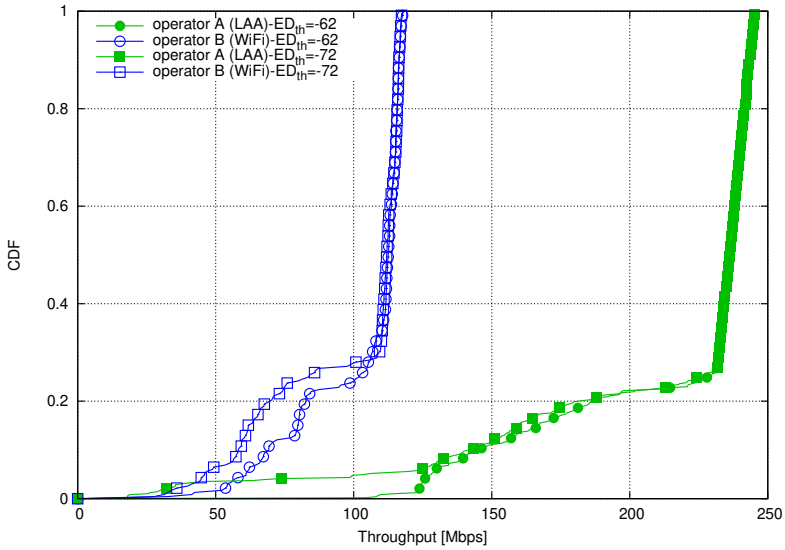
The reason for this behavior is not evident, and it depends on aspects that may be implementation specific and on the specific traffic model we are using. When the LAA ED threshold is -72 dBm, LAA detects the activity of Wi-Fi in the channel, which forces a backoff until the channel is sensed to be idle. Immediately after the AP's transmission, the channel is idle during SIFS (16 μs), and after that the STA transmits the BA. LAA eNB backs off during the AP transmission, and also during the SIFS, since it is shorter than the initial CCA duration of 43 μs. As a result of that, LAA eNB initiates its transmission once the AP has transmitted the entire Aggregate MPDU (A-MPDU) and has received the corresponding BA. This would lead to an expected favorable behavior towards Wi-Fi, in case the Wi-Fi transmission is over.



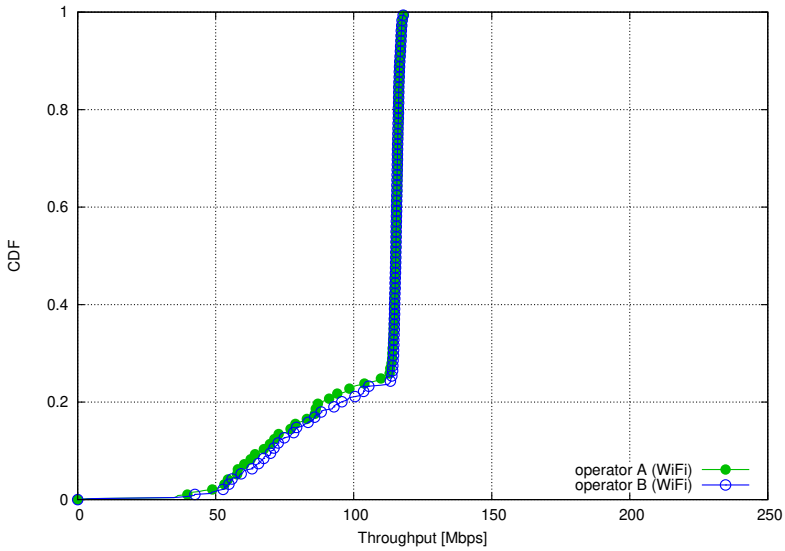
(a) LAA ED_{th} = -62 **(b)** LAA ED_{th} = -72 dBm

Figure 7.10 Illustration of the energy detection coverage when varying the LAA ED threshold in the scenario simple. Wi-Fi CCA-ED threshold is fixed to -62 dBm.

However, with the FTP model 1 we are using, the transmission of a single A-MPDU is not enough to transmit the FTP file of 0.512 MBytes. This means that once LAA has started its transmission, it is likely that Wi-Fi continues to transmit the following A-MPDU, since it does not energy detect the LAA node transmission. For the Wi-Fi AP, the previous transmission concluded successfully, so that in the new transmission it applies the previous transmission rate, when there was no interference by LAA. However, the new A-MPDU rate is most likely too high for the new interference conditions. Thus the STA is not able to decode the A-MPDU and does not send back the corresponding BA. From this point, the following actions taken by the Wi-Fi AP depend on the specific proprietary rate adaptation mechanism. According to the Wi-Fi implementation in ns-3, the Wi-Fi AP waits until the timer expires and then sends the Block Acknowledgment Request (BAR), but these messages are likely not to be received by the STA since the LAA transmission is ongoing. As soon as the Wi-Fi adaptation mechanism retransmits the next A-MPDU with a lower Modulation and Coding Scheme (MCS) that can be decoded by the STA, the Wi-Fi transmission starts to recover. However, some delay is inevitable, regardless of the proprietary rate adaptation algorithm, due to the lost A-MPDU and the time needed for the AP to resume the flow. Because of this delay the Wi-Fi performance drops, while the performance of LAA increases due to reduced collisions with Wi-Fi flows that are waiting for BA timeout. On the other hand, when LAA ED threshold is -62 dBm, it is likely that the LAA transmission starts in the middle of the AP transmission of the A-MPDU generating a collision. Even if counter-intuitive, this situation is better for Wi-Fi since LAA will most likely interfere with only some MAC Protocol Data Unit (MPDU)s of the A-MPDUs, so that the STA will be aware of the A-MPDU and will send back the corresponding BA. Since the BA is typically sent with the lower transmission rate, there is a high probability that the AP can decode it successfully even if there is some interference by LAA. This BA provides an important information to the AP, about missed MPDUs and new interference, which leads to decrease the MCS of the following A-MPDUs. Thanks to this, and differently to the -72 dBm case, the Wi-Fi transmission continues, even though with a lower rate. The



(a) LAA over Wi-Fi scenario

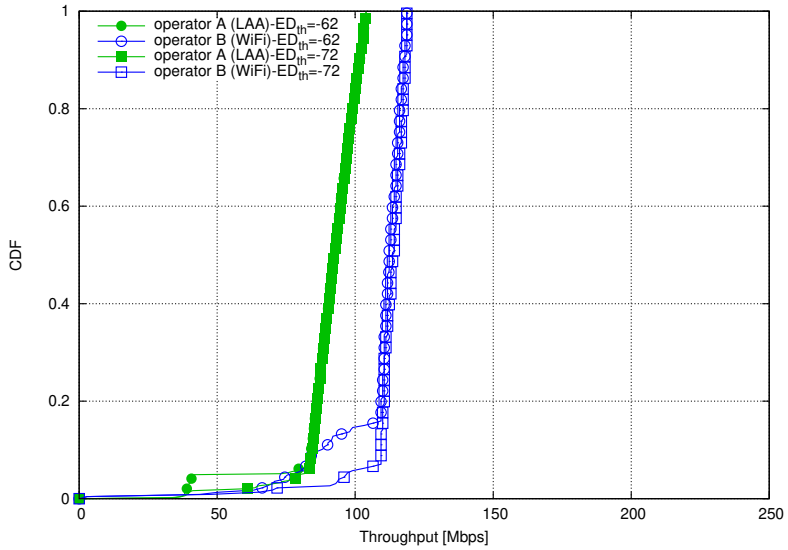


(b) Baseline Wi-Fi over Wi-Fi scenario

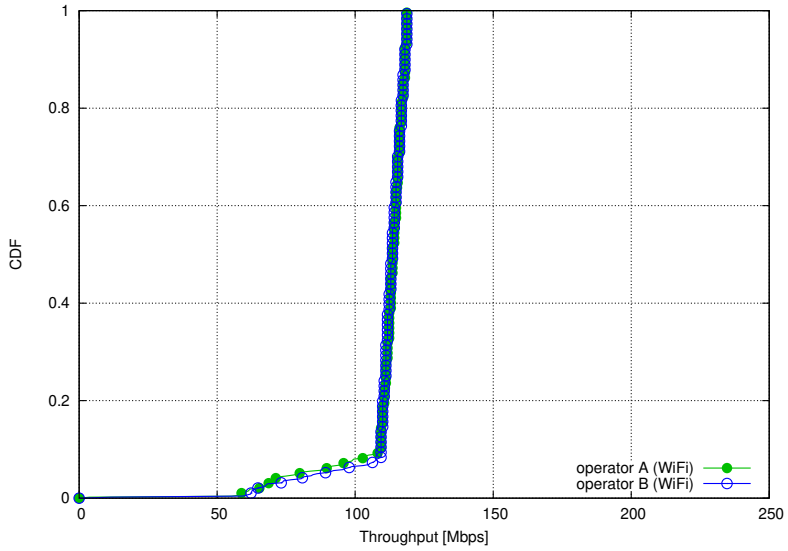
Figure 7.11 Impact of LAA ED threshold, $ED_{th} \in \{-62, -72\}$ dBm, evaluated in simple scenario using SNR-triggered Wi-Fi rate manager with the default FTP settings ($\lambda_{LAA} = 5$, $\lambda_{Wi-Fi} = 2.5$, file size=0.512 MB)

packet capture (PCAP) traces confirm these findings. Specifically, in the -62 dBm case, the maximum rate is used in 92.4% of transmissions, while in the -72 dBm case, it is in 99%. The BAR is sent 123 and 401 times when the ED threshold is, respectively, -62 dBm and -72 dBm case. **We conclude that the coexistence performance are not only affected by the access mechanism and ED procedure, but the impact of traffic model, Wi-Fi Distributed Coordination Function (DCF) and rate adaptation mechanism plays an important role.**

Dependency on packet size: To further confirm the conclusion discussed above, it is reasonable to expect that when flows can be sent within a single A-MPDU, the interplay between the rate adaptation mechanism and the LAA backoff will disappear. In Figures 7.12b and 7.12a, we show the results for a smaller file size of 0.0512 MB. We choose this file size



(a) LAA over Wi-Fi scenario (SNR-triggered Wi-Fi rate manager)



(b) Baseline Wi-Fi over Wi-Fi scenario (SNR-triggered Wi-Fi rate manager)

Figure 7.12 Impact of LAA ED threshold, $ED_{th} \in \{-62, -72\}$ dBm, evaluated in simple scenario using SNR-triggered Wi-Fi rate manager, with the modified FTP settings ($\lambda_{LAA} = 20$, $\lambda_{Wi-Fi} = 10$, file size=0.0512 MB)

because the maximum size of A-MPDU in 802.11n is 65535 bytes, thus in this way the file can be transferred during a single A-MPDU. We increase FTP lambdas to compensate for the lower file size. The results show the expected behavior that we were observing with the CBR traffic, i.e., Wi-Fi benefits from lower LAA ED threshold.

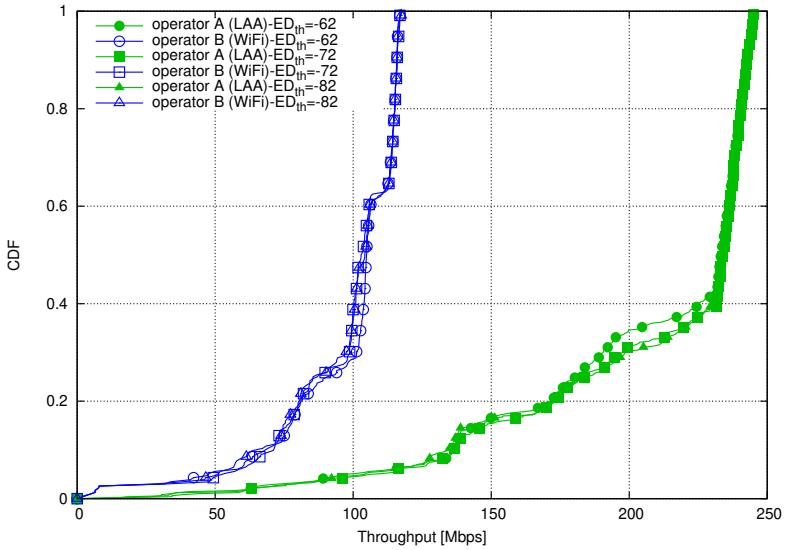
Dependency on LAA node coordination: Finally, we focus our attention on the more general 3GPP indoor scenario, when the traffic type is the 3GPP Model 1 FTP over UDP. In Figures 7.13a and 7.13b, we show the impact of LAA ED threshold on the throughput and latency. In this scenario, lowering the LAA ED threshold impacts negatively Wi-Fi, while it is impacting

positively LAA. The channel occupancy of the LAA operator is 9.8% when ED_{th}=-62 dBm, 10% when ED_{th}=-72 dBm, and 10.3% when ED_{th}=-82 dBm. Besides the already described aspects, which depend on the traffic model and the rate adaptation algorithm, in this scenario we observe an additional effect, which depends on LAA network design. When lowering the ED threshold it becomes less likely that various LAA nodes transmit simultaneously, since they back off to each other, so that the channel occupancy of the LAA network increases. By lowering the LAA ED threshold, the number of collisions reduces (6.34%, 5.35% and 5.18% for ED_{th} of, respectively, -62, -72 and -82 dBm), but we get the side effect of an increased channel occupancy by LAA, which makes that the Wi-Fi nodes spend more time in backoff. The impact of lowering ED_{th} on the LAA network is positive, since the collisions with Wi-Fi are significantly reduced, and the reduction in collisions compensates the increase of time spent in the backoff. This effect should be taken into account when designing an LAA network. LTE is in fact capable of exploiting frequency reuse-1, thanks to interference management procedures inherent in its design (e.g., adaptive rate control, power control, Coordinated Multi- Point (CoMP), enhanced Inter-Cell Interference Coordination (eICIC)). Solutions of self-deferral have been proposed in 3GPP in order to take into account these aspects.

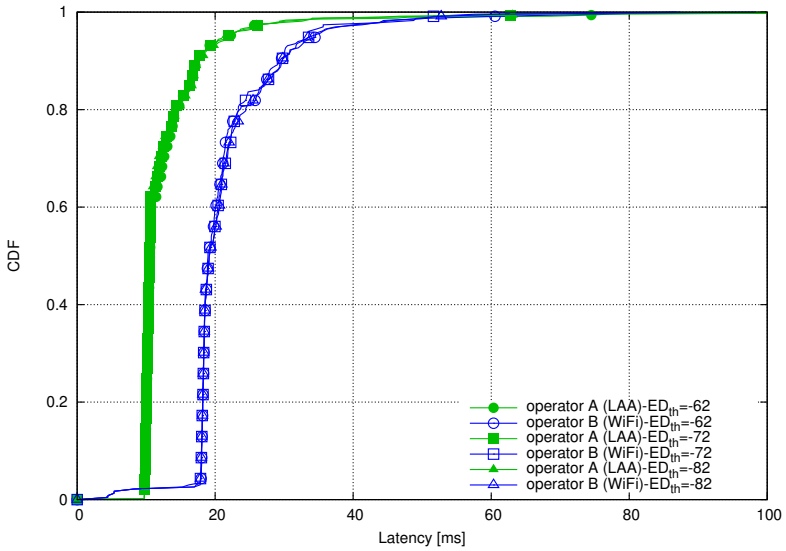
We reach then the following conclusion on the impact of the ED threshold. **In general, it is reasonable to expect that lowering the ED threshold will benefit the coexistence performance. However, this effect is not at all linear, and can be heavily impacted by aspects which have nothing to do with LAA access mechanism, but are related to scenario configurations, implementation issue and the traffic patterns. In addition, while LTE was designed to be able to reuse in frequency the channel, the LAA nodes of the same operator should be coordinated in order to control the LAA to LAA backoff mechanism, which also may affect coexistence with Wi-Fi. For this, solutions like the *self-defer* have been proposed in standard.**

7.9.2 Impact of Wi-Fi Rate Adaptation Algorithm

Since the Wi-Fi standard does not specify any particular strategy for selecting the MCS, rate adaptation mechanisms are implementation specific, and can obey to different cost-functions, strongly impacting final coexistence performance. In our study, we use an SNR-triggered adaptation rate algorithm, which selects the highest transmission rate that allows to obtain a target frame delivery to achieve the target bit error rate (BER), by leveraging transmission mode based-specific SNR/BER curves. To understand better the impact of the rate adaptation mechanisms, we further analyze another representative Wi-Fi rate manager, Minstrel-HT, and compare its behavior in the coexistence scenario to that of SNR-triggered Wi-Fi rate manager. Minstrel-HT is a well-known rate adaptation mechanism that is used by many drivers in Linux kernel. Minstrel-HT keeps evaluating the delivery probability performance of every MCS by using look-around frames during normal operations, and selects on average the MCS with the best performance. In Fig. 7.14a and Fig. 7.14b, we show the coexistence results when Minstrel-HT is employed, instead of the SNR-triggered rate manager, in the simple scenario and when lowering the LAA ED threshold. We observe that with Minstrel, the throughput of a much larger number of flows that are transmitted with the highest rates is reduced when the LAA ED threshold is lowered, compared to the SNR-triggered rate manager, whose performance were already shown in Fig. 7.11. The reduction of the performance is due to the decrease in the MCS. On the other hand, we do not notice the impact of changing the LAA ED threshold on the



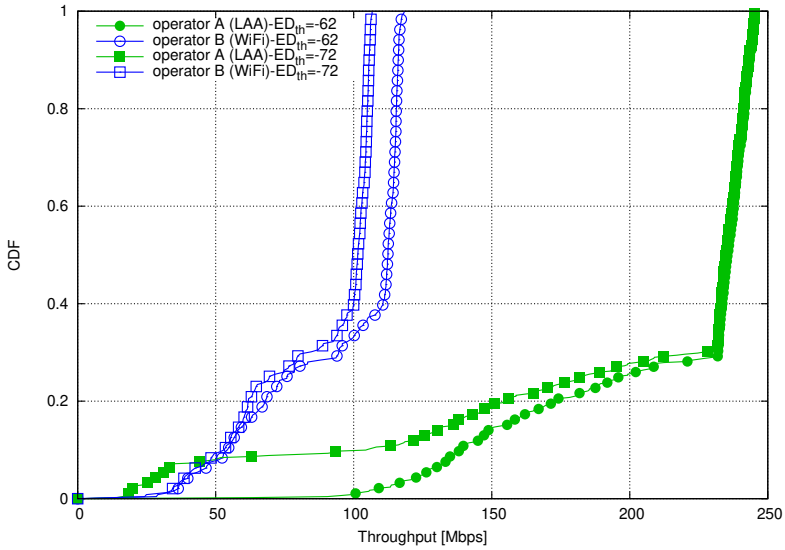
(a) Throughput of LAA and Wi-Fi FTP flows



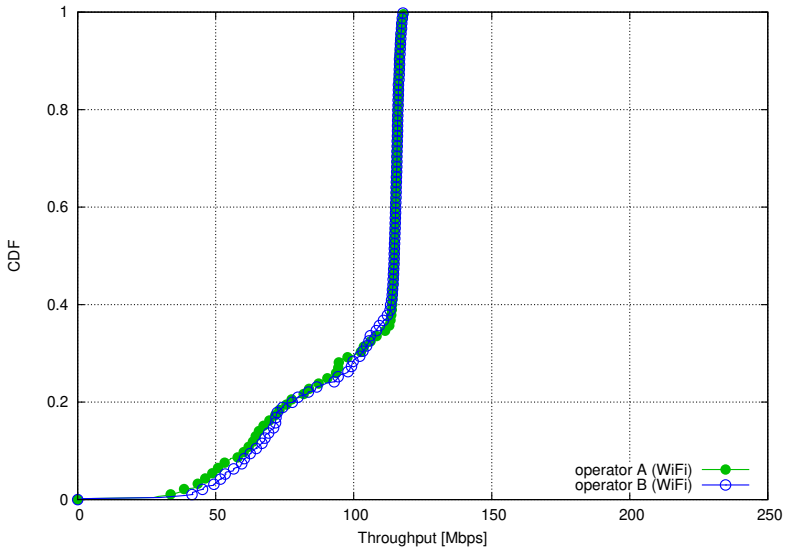
(b) Latency of LAA and Wi-Fi FTP flows

Figure 7.13 Impact of LAA ED threshold, $ED_{th} \in \{-62, -72, -82\}$ dBm, evaluated in 3GPP indoor scenario when the traffic is FTP and Wi-Fi is using the SNR-triggered rate manager.

flows that are transmitted with the lower rates. This is because, when the interference occurs, the Minstrel-HT adjusts the rate, which remains then similar during the whole simulation, so that the impact of interference is more time lasting than in case of the SNR-triggered Wi-Fi rate manager. Similar results, showing an impact of the rate adaptation, were presented by Google's initial investigation on Wi-Fi and LTE-U coexistence in [241], in case of coexistence between LTE-U and Wi-Fi. However, it was generally considered in literature that the existence of an LBT would eliminate the interplay between the Wi-Fi rate adaptation algorithm and channel access mechanism. In this work, however **we show that the Wi-Fi rate adaptation may have a strong impact on coexistence, depending on the traffic model, not only in case of LTE-U, but also for LAA.**



(a) LAA over Wi-Fi (Minstrel-HT)

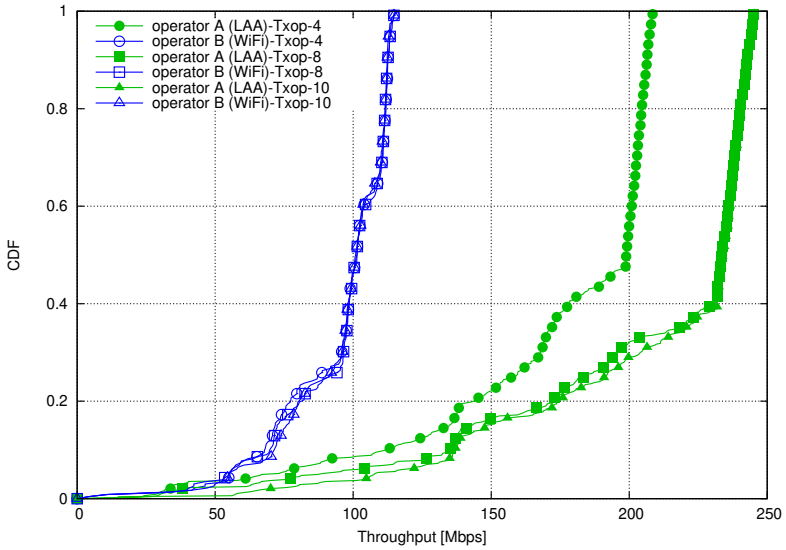


(b) Baseline Wi-Fi over Wi-Fi (Minstrel-HT)

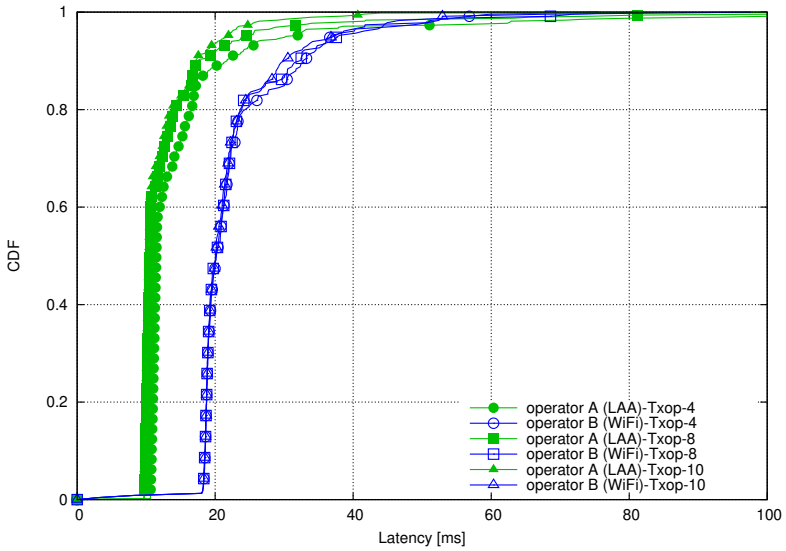
Figure 7.14 Impact of LAA ED threshold, $ED_{th} \in \{-62, -72\}$ dBm, evaluated in simple scenario using Minstrel-HT Wi-Fi rate manager.

7.9.3 Impact of LAA maximum TxOP length

The maximum TxOP length or MCOT is another LAA parameter that needs to be carefully selected. 3GPP selects the TxOP length depending on the addressed application, with values ranging from 2 to 10 ms. In Figure 7.15, we show the impact of MCOT in the 3GPP indoor scenario. The results show that the impact on LAA is significant while on Wi-Fi it is not much pronounced. As expected, **LAA achieves better performance for higher values of MCOT.** This is due to the reduced waste of capacity due to reservation signals, and the decreased number of required accesses to the channel, which results in shortened accumulated backoff time. For example, in this simulation, the LAA channel occupancy is 10.1%, 9%, 8.8%, for MCOT of, respectively, 4, 8 and 10 ms. **Besides the channel occupancy, the MCOT also**



(a) Throughput of LAA and Wi-Fi flows



(b) Latency of LAA and Wi-Fi flows

Figure 7.15 Impact of LAA TxOP evaluated in 3GPP indoor scenario.

affects the probability of collisions. In the previous example, when the MCOT is 4 ms, the LAA node starts to transmit and then releases the channel 6 times during the transmission of a single file. However, in case of 8 and 10 ms MCOT, this will happen only 3 times. The simulation traces confirm that in case of $MCOT = 4$ ms, the percentage of collisions is 4.78% and is higher than in case of $MCOT = 8$ ms and $MCOT = 10$ ms, when it is, respectively, 4.75 and 4.68%. When Wi-Fi is competing for the channel the impact is not so linear as for LAA. **In general, it is better for Wi-Fi when the MCOT is shorter. However, a shorter LAA MCOT also increases the overall channel occupancy by LAA and the probability of collisions, and consequently also the Wi-Fi's CW size. On the other hand, a longer MCOT increases the time Wi-Fi waits for LAA to complete its transmissions, but reduces the channel occupancy by LAA and so the number of collisions.** Among the tested values for MCOT, results show that Wi-Fi attains the best performance when the $MCOT = 10$ ms.

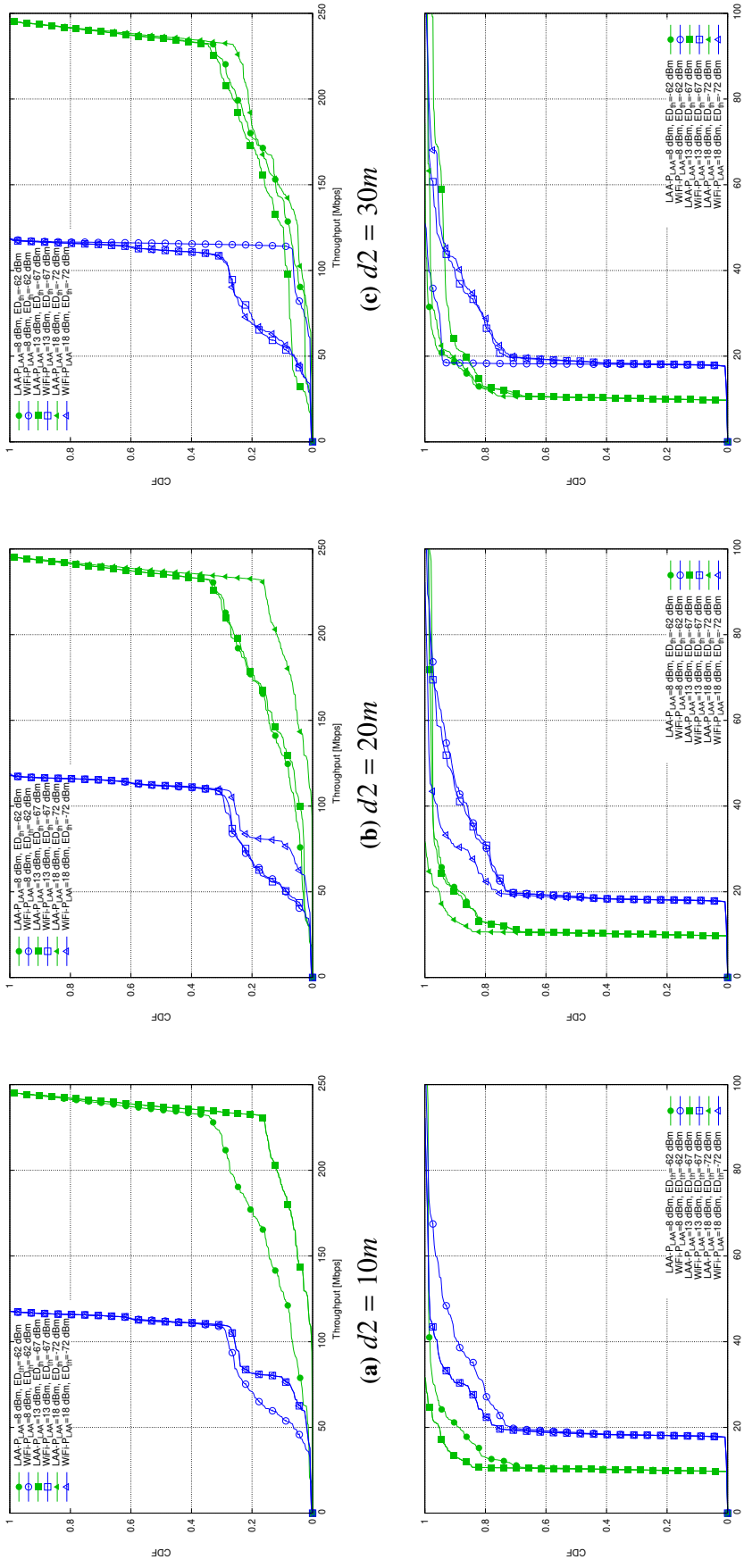


Figure 7.16 Impact of varying the LAA transmission power along with LAA ED threshold evaluated in simple scenario with FTP traffic using SNR-triggered Wi-Fi rate manager with the default FTP settings ($\lambda_{LAA} = 5$, $\lambda_{Wi-Fi} = 2.5$, file size=0.512 MB), $d_1 = 10m$, $d_2 = 10, 20, 30m$.

Table 7.8 How eNB listens to AP depending on $d2$ and P_{LAA}. Wi-Fi $ED_{th} = -62dBm$.

d2	$P_{LAA} = 18\ dBm$ $ED_{th} = -72\ dBm$	$P_{LAA} = 13\ dBm$ $ED_{th} = -67\ dBm$	$P_{LAA} = 8\ dBm$ $ED_{th} = -62\ dBm$
10 m	-54 dBm (yes)	-54 dBm (yes)	-54 dBm (yes)
20 m	-61 dBm (yes)	-61 dBm (yes)	-61 dBm (yes)
30 m	-66 dBm (yes)	-66 dBm (yes)	-66 dBm (no)

7.9.4 Impact of LAA transmission power

In this section, we study the impact of the LAA transmission power on the coexistence performance. It is worth mentioning that the modification of the transmission power has an impact on the ED threshold that can be adopted. In particular, the ED threshold can be adapted depending on the available bandwidth and the maximum transmit power, as discussed in section 15.2.3.1 in [74]. We consider the simple scenario, with nodes at distance 10, 20 and 30 meters. We study three different power levels for LAA eNB, P_{LAA}, (8, 13, 18 dBm) considering the gain of 5 dbi, to which correspond three different LAA ED_{th} values (-62, -67 and -72 dBm), according to the ED adaptation rule discussed in [74]. Depending on the transmission power of the eNB, P_{LAA} and of the related energy detection threshold, the LAA eNB and Wi-Fi AP can either detect or not detect each other, in a symmetric or asymmetric way. In particular, the detection capabilities of each node, when considering only path loss, and eliminating the random aspects of the propagation, are summarized in Tables 7.9 and 7.8: When the distance is 10 m, the eNB receives the emission of the AP, which is transmitted always at 18 dBm, at -54 dBm. So LAA sees and backoffs to Wi-Fi. On the other hand, Wi-Fi detects LAA transmission when they are emitted at 18 dBm and at 13 dBm, but not at 8 dBm, since they are received below threshold. When the distance is 20 m, again LAA detects Wi-Fi, but Wi-Fi only detects LAA when it transmits at the maximum power. Finally, when the distance is 30 m, LAA cannot detect Wi-Fi at -62 dBm, and Wi-Fi can never detect LAA, because it is energy detecting at -62 dBm. We discuss simulation results in Figure 7.16. When the distance between nodes is 10 m, coexistence performance is degraded when Wi-Fi cannot detect LAA, ans its access generates collisions and interference, so with P_{LAA}=8 dBm. When P_{LAA} is higher, instead the performance is better and equivalent for both 13 and 18 dBm. A similar trend is observed with distance of 20 m. Here the coexistence performance is jeopardized, again, when Wi-Fi accesses the channel without detecting LAA. The interference is still quite high and so we see worse performance than for the case of 18 dBm, when Wi-Fi and LAA see each other. Finally, when the distance is 30 m, collision are generated by Wi-Fi in all the cases. However, the interference is lower due to the increased distance. As a result, of that the best performance is achieved when P_{LAA} is 8 dBm, while the worse one when P_{LAA} is 18 dBm. Those results are interesting in that show that any slight modification in scenario parameters can provide different effects in performance evaluation.

7.9.5 Impact of DRS

As mentioned earlier, the periodicity of the DRS signaling, T_{DRS}, can be 40, 80, 160 ms. For RRM measurements, it would be beneficial to have a high periodicity of DRS signals. However, the DRS model does not capture these aspects in the simulator, but only its channel occupancy.

Table 7.9 How AP listens to eNB depending on $d2$ and P_{LAA}. Wi-Fi $ED_{th} = -62dBm$.

d2	$P_{LAA} = 18\ dBm$ $ED_{th} = -72\ dBm$	$P_{LAA} = 13\ dBm$ $ED_{th} = -67\ dBm$	$P_{LAA} = 8\ dBm$ $ED_{th} = -62\ dBm$
10 m	-54 dBm (yes)	-59 dBm (yes)	-64 dBm (no)
20 m	-61 dBm (yes)	-66 dBm (no)	-71 dBm (no)
30 m	-66 dBm (no)	-71 dBm (no)	-76 dBm (no)

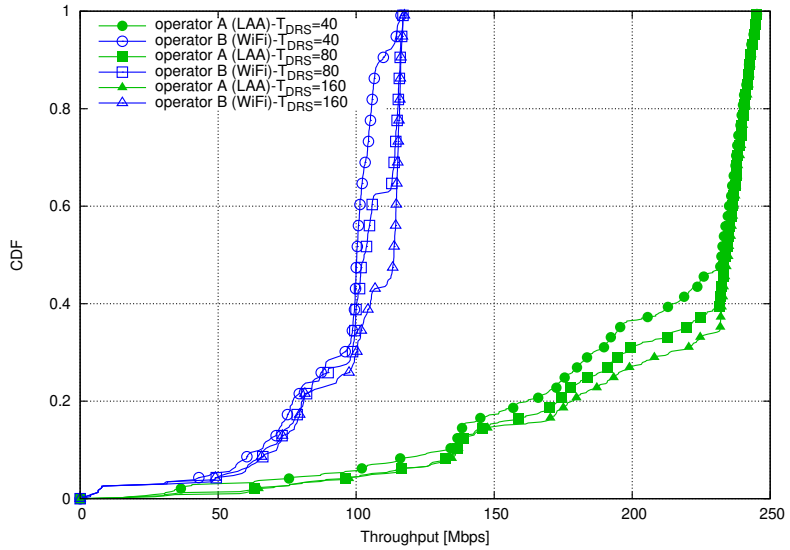
As a result, Fig. 7.17 shows that **sending the DRS more frequently may have a negative impact, not only on a neighboring Wi-Fi but also on LAA network. The reason is that DRS is a quite expensive signaling, which may spare 1 ms transmission, in case it cannot be transmitted with data.** In such cases, it is much harder for Wi-Fi to achieve the highest data rates. In particular, while for $T_{DRS} = 160$ ms more than 55% of Wi-Fi flows obtain the maximum data rate of around 117 Mbps, for $T_{DRS} = 80$ ms the percentage of Wi-Fi flows to attain this rate drops to 35%, and for $T_{DRS} = 40$ ms this value is around 10%. This is due to the increased LAA occupancy time, and increased number of collisions, when T_{DRS} is low (i.e., high periodicity). LAA channel occupancy is 11.22%, 9.3% and 8.3% for the T_{DRS} of, respectively, 40, 80 and 160 ms. There are 4.26%, 4.07% and 3.65% collisions for, respectively, T_{DRS} equal to 40, 80 and 160 ms. The collisions are affecting the CW sizes of both, LAA and Wi-Fi operator, which results in a higher backoff time for both. Additionally, the LAA node enters into the backoff more often for lower T_{DRS}, since it accesses the channel more frequently, hence its delay and throughput performance degrade.

7.9.6 Impact of hidden nodes on LAA and Wi-Fi coexistence

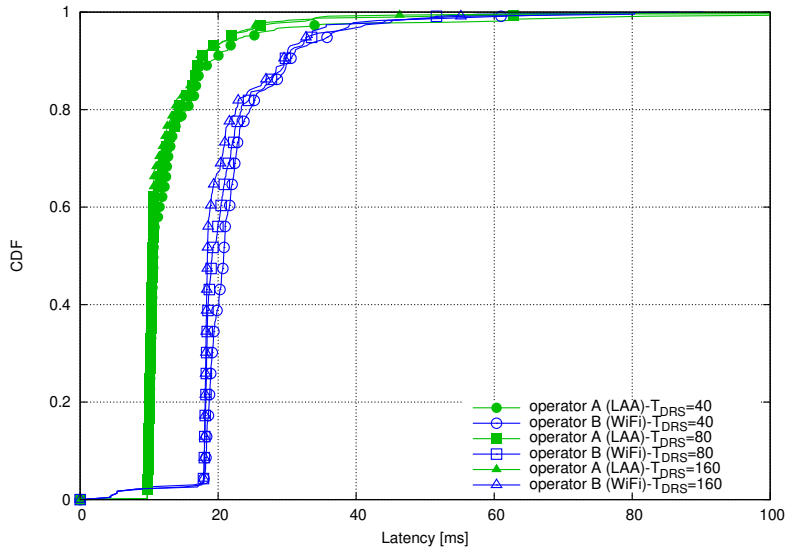
To study the impact of hidden nodes, we compare the performance obtained in BS corners scenario to that of the 3GPP indoor scenario in Fig. 7.18b. We recall that a Wi-Fi device is capable of detecting other Wi-Fi devices at -82 dBm and LAA devices at -62 dBm. On the other hand, an LAA device energy detects all types of devices, Wi-Fi or LAA, at the same energy level of -72 dBm. Both networks are negatively impacted by the hidden nodes, in particular, the percentage of collisions increases from 4.06% in the indoor scenario, to 5.47% in the BS corners scenario. The collisions cause an increase in the CW size and the backoff, and consequently reduce the throughput of both Wi-Fi and LAA. The CW size is especially increased for Wi-Fi. The average backoff counter goes up to 244 slots in the indoor scenario, whereas in the BS corners scenario it reaches 735 slots. Due to the asymmetric detection levels, the number of LAA simultaneous transmissions increases, so that the average MCS for affected UEs reduces. As a result, the LAA network is more affected by the hidden node scenario than the Wi-Fi network.

7.9.7 Impact of LAA CTS-to-Self

As we mentioned in section 7.5.2, and we analyzed in Sections 7.9.1 and 7.9.6 the performance degradation caused by the hidden nodes could be reduced by using the CTS-to-Self mechanism. CTS-to-Self would allow Wi-Fi to preamble detect LAA at -82 dBm or below. We evaluate the impact of CTS-to-Self in 3GPP indoor and BS corners scenario. Fig. 7.20 shows that the **CTS-**



(a) Throughput of LAA and Wi-Fi flows

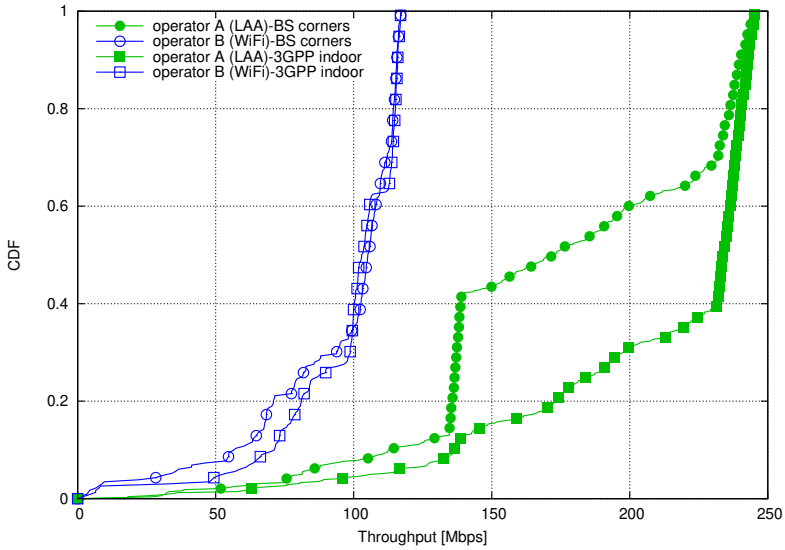


(b) Latency of LAA and Wi-Fi flows

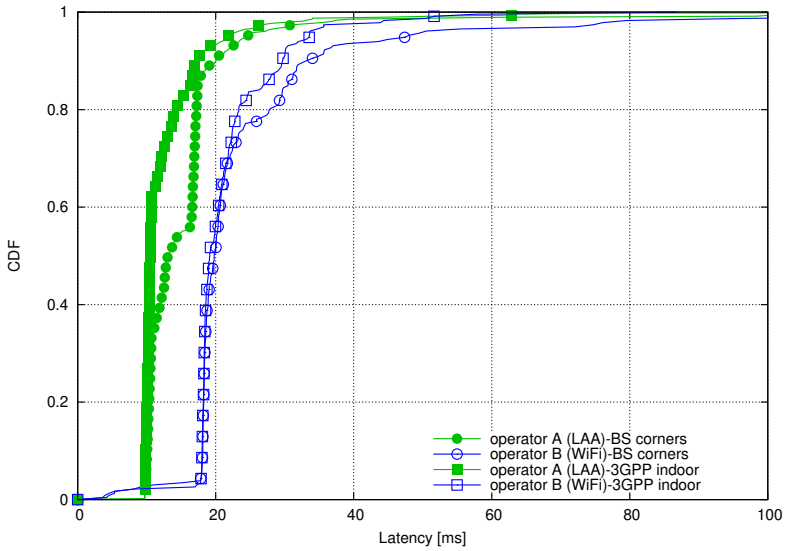
Figure 7.17 Impact of DRS evaluated in 3GPP indoor scenario.

to-Self improves the performance of the LAA network, while the impact on Wi-Fi is, as expected, negative, and depends on the scenario and the number of hidden nodes.

In 3GPP indoor scenario, collisions decrease from 2.68% to 0.43% when CTS-to-Self is active. This benefit is also reflected in terms of lower CW size for both LAA and Wi-Fi, e.g., when CTS-to-Self is not used, the maximum BO counter value of Wi-Fi reaches on average 512 slots, while when CTS-to-Self is used, its maximum value is on average around 32 slots. However, even if the CW size and collisions are reduced, the Wi-Fi performance is impacted negatively. This is mostly due to unnecessary backoffs to far away LAA nodes, which are detected in the range of [-82, -62] dBms, and which would not significantly affect the Wi-Fi communication even in case of collisions. Note that when performing the backoff, the actual time spent in backoff is composed of: the idle slots during which the backoff counter is decreased; and the channel busy slots, when the backoff counter is frozen until the channel is found to be idle for



(a) Throughput of LAA and Wi-Fi flows



(b) Latency of LAA and Wi-Fi flows

Figure 7.18 Impact of hidden nodes evaluated by comparing performance in 3GPP indoor and BS corners scenario.

the duration of eCCA defer period for LAA or DIFS for Wi-Fi. We recall how the backoff counter is frozen and resumed in Fig. 7.19. As a result, even if the average number of idle slots decreases, the number of busy slots is increased since Wi-Fi sees the channel more occupied.

In the BS corners scenario, the trend is similar, but more accentuated since the number of hidden nodes increases. The CTS-to-Self mechanism is almost not impacting the Wi-Fi network, while it is significantly improving and protecting the LAA performance. Collisions drop from 2.75% to 8.09% collisions, when CTS-to-Self is used. In a scenario more affected by hidden nodes, the reduced number of collisions allowed by the introduction of CTS-to-Self is compensated by the extra time that Wi-Fi has to spend in backoff, due to the better detection capabilities. This makes that Wi-Fi performance results not affected by CTS-to-Self in this scenario, differently from the indoor.

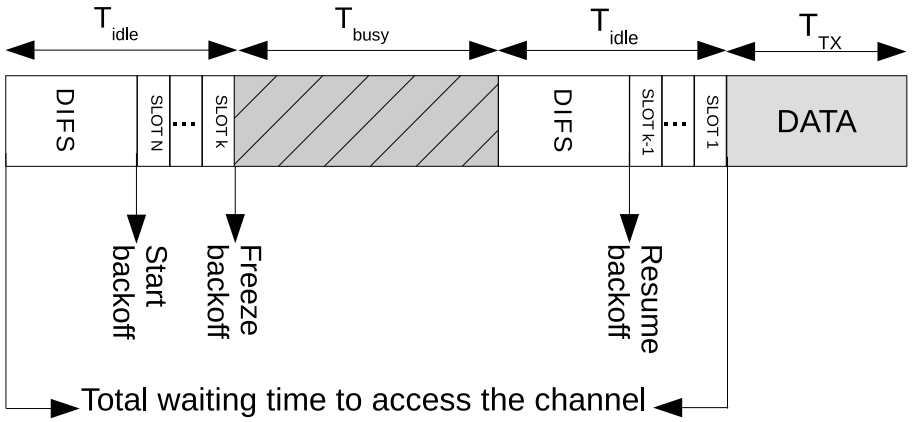
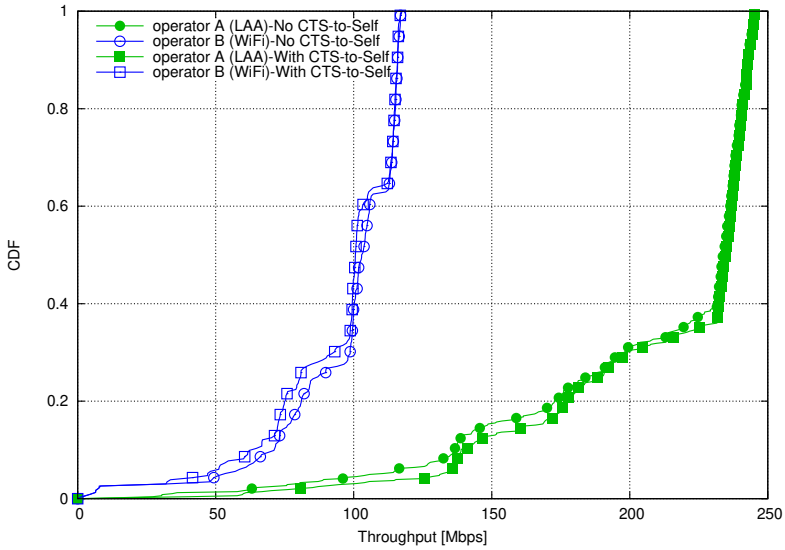


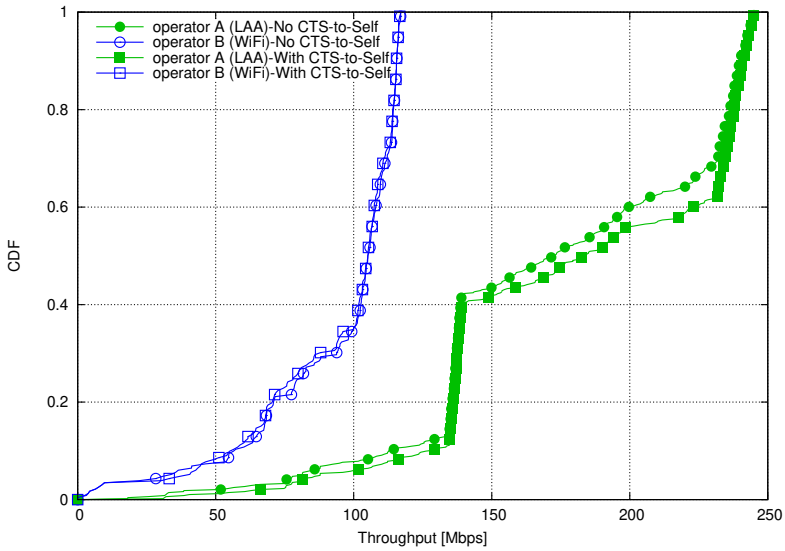
Figure 7.19 Illustration of time that Wi-Fi spends in backoff when the channel is busy during the backoff process. An initial value of backoff counter is N slots. After N-k slots the backoff is frozen due to detection of channel being busy.

7.9.8 Impact of LAA CW Update Rule (Z parameter)

In Fig. 7.21, we show the impact of Z parameter on coexistence performance for different λ. Two CW update rules are compared, CW-nacks80 and CW-any. The CW-nacks80 is the default LAA LBT CW update rule which uses 80% threshold in order to increase the CW. The CW-any is the LAA LBT CW update rule according to which the CW is increased if any NACK belonging to a reference subframe is reported. The comparison results show that there is hardly any impact of the Z parameter on the coexistence performance, when considering λ=5. This is mainly for the reasons explained in the Section 7.5.4. In particular, **the impact of Z parameter depends on the number of flows being scheduled in a single subframe, it also depends on the type of traffic being considered, and on the MAC scheduler algorithm.** For this reason, we evaluate the FTP Model 1 traffic also for a higher λ value in order to achieve more UEs to be scheduled in the same subframe. In Fig. 7.21b we show the performance when λ=10. As expected, when $Z = 80\%$ LAA backs off less and enters in backoff less often, than the case $Z > 0$, while Wi-Fi finds the channel more frequently occupied, and enters more often in backoff. On the other hand, when Z is reduced, we note an increase in LAA backoff time, which results in a reduced number of collisions. There are 8.54% of collisions when $Z = 80\%$, and 8.23% when CW-any rule is used. **We conclude that reducing the Z parameter positively impacts Wi-Fi, while it negatively impacts LAA.** Notice that the impact of the Z policy as observed in Fig. 7.21 is marginal, also due to the maximum contention window value proposed by 3GPP for the class type 3 that we are considering in the simulations, which is 63. With a slot time of 9 μs, even when reaching the maximum contention window, due to collisions, the LAA eNB is normally in the condition to finalize its backoff count before the beginning of the following subframe, whatever the Z rule is. Consequently the impact is marginal.



(a) 3GPP indoor scenario



(b) BS corners scenario

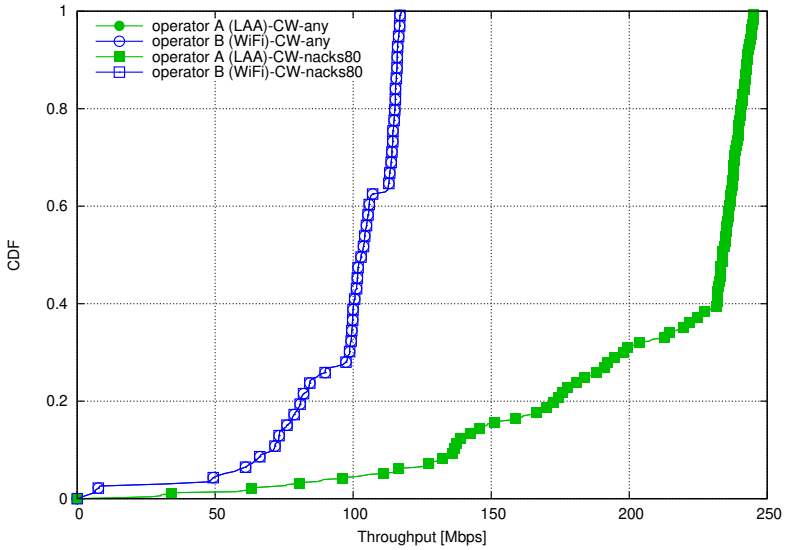
Figure 7.20 Impact of LAA CTS-to-Self

7.10 Performance evaluation of LTE-U/Wi-Fi coexistence

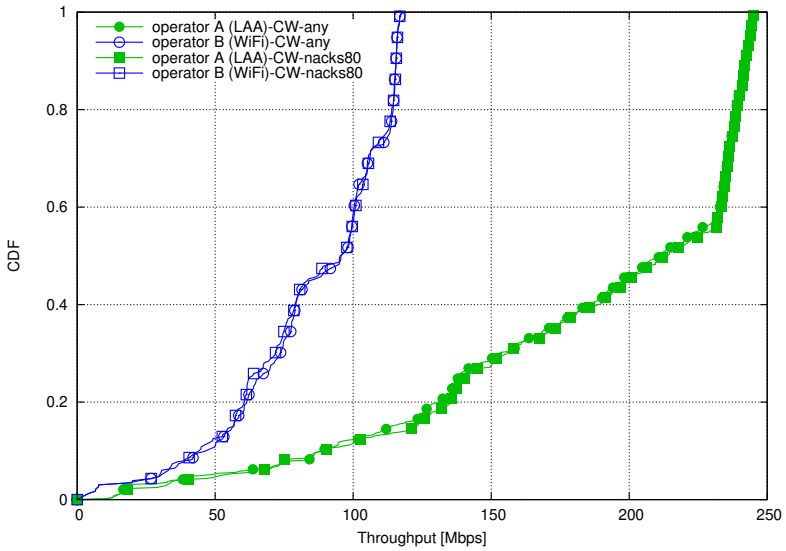
7.10.1 Impact of CSAT dyty cycle duration

In this section, we analyze the impact of T_{CSAT} parameter. As we mentioned earlier, T_{CSAT} represents the duration of the duty cycle, which includes ON and OFF time. This is one of the parameters that could be adaptively adjusted based on the MU measurements. Some of the values considered in industry are {40, 80, 160, 320, 640, 1280} ms. We will show in the following that the impact of T_{CSAT} depends on a scenario and traffic type. We consider the full buffer case, and a low traffic profile.

In Table 7.10, we show the impact of T_{CSAT} in the simple scenario when the traffic is full buffer



(a) $\lambda = 5$



(b) $\lambda = 10$

Figure 7.21 Impact of LAA Z parameter evaluated in 3GPP indoor scenario.

CBR. The distances between nodes, d1 and d2, are selected in such a way that the nodes can energy detect each other (10 mt). Since the traffic is a full buffer traffic over UDP, then during the ON time, LTE-U occupies the channel during almost 100% of time. Thus, it is to expect that the share of the channel remains the same for any value of the T_{CSAT}. However, what changes depending on T_{CSAT} is the time that the LTE-U node needs to adjust the duty cycle and converge to 50% share, based on the medium utilization measurements.

When the intensity of the traffic is lower, though, we observe different behaviours. In Fig. 7.22 we show results obtained for the 3GPP indoor scenario, with FTP traffic. We observe a negative impact on Wi-Fi performance when increasing T_{CSAT}. This is due to the traffic and duty cycle pattern, as we explain in the following. According to the exponential arrival process, for LTE-U operator, $\lambda = 5$, there is on average 1 file transfer every 0.2 second. Similarly, for Wi-Fi operator, $\lambda = 2.5$, the average file transfer will start every 0.4 seconds. In Table 7.11 we show

Table 7.10 Impact of T_{CSAT} in scenario simple, where $d1 = d2 = 10$ meters.

Measurement/T_{CSAT}	40	80	160
LTE-U throughput (Mbps)	198.29	214.5	218.97
Wi-Fi throughput (Mbps)	68.16	55.68	50.65
Average LTE-U duty cycle	0.5	0.58	0.58

Table 7.11 Impact of T_{CSAT} in scenario indoor, where $d1 = d2 = 10$ meters.

Measurements/T_{CSAT}	40	80	160
MU	0.133	0.125	0.107
Average LTE-U duty cycle	0.5	0.7465	0.873

the medium utilization and duty cycle values for different values of T_{CSAT}. We observe that for any value of T_{CSAT} the average MU is around 0.1, which is below the lower MU threshold, MU_{low}, which defaults to 0.4, as shown in Table 7.4. Because of this, the duty cycle during each simulation converges to its maximum value. Note that the effective LTE-U ON time increases with the increase of T_{CSAT}. Consequently, the number of transmissions by LTE-U increases. This is mainly because LTE-U can send more often MIB and LDS signaling, and the signaling is transmitted independently of data. For example, there are 3437, 5853 and 7362 MIB transmissions for T_{CSAT} of, respectively, 40, 80 and 160 ms. Note that the default periodicity for MIB is 10 ms for ON periods, while during OFF periods it drops to the minimum of 160 ms. We observe similar increment of LDS signaling, transmitted 999, 1167 and 1202 times for T_{CSAT} equal to, respectively, 40, 80 and 160. More signaling by LTE-U nodes introduces more collisions in the network. **We conclude that an increase of the T_{CSAT}, benefits the LTE-U network and not Wi-Fi, when the traffic in the network is low and the duty cycle is high.**

7.10.2 Impact of LDS periodicity

As we mentioned earlier, LDS periodicity follows the specification of DRS, thus its periodicity can be of 40, 80 or 160 ms. Its synchronization functionality in the network is not modeled and we only model the channel occupancy of the signal and the collisions it generates. In Fig. 7.23, we show the impact of T_{LDS} on both, Wi-Fi and LTE-U. As expected, the more frequent the LDS, the higher the collision probability. We observe that there is 6.56%, 6.23% and 5.79% of collisions for T_{LDS} of respectively, 40, 80 and 160 ms. Also, LTE-U channel occupancy is higher for lower values of T_{LDS}, and is 9.4%, 8.2% and 7.6% for T_{LDS} of, respectively, 40, 80 and 160 ms. **Finally, similarly to LAA DRS, we conclude that T_{LDS} should be kept as low as possible to ensure a good coexistence performance.** Impact of LDS on Wi-Fi seems to be much weaker than that of DRS, in spite of the fact that DRS is sent with LBT. This is due to a higher channel occupancy in case of DRS, due to the potential need of reservation signal.

7.10.3 Impact of LTE-U puncturing

The concept of puncturing has been introduced to let low latency Wi-Fi traffic go through during the LTE-U ON periods. In Fig. 7.24, we show the impact on puncturing when both operators perform the FTP Model 1 traffic. We observe that for this type of traffic the puncturing is not playing an important role, and it is better not to puncture. **The reason is that the puncturing increases the probability of collisions, due to the impact of the OFF to ON transitions, when 1 ms puncturing is not enough to transfer the A-MPDU, which usually takes about 4 ms.** Collision traces show that there are 3.89% of collisions when there is no puncturing, and 4.16% when there is puncturing. When the traffic type is FTP, it is better for coexistence with Wi-Fi to let the LTE-U node finish its transmission and then have the Wi-Fi node transmit ideally without interruptions by LTE-U.

In Fig. 7.25, we show the impact on latency sensitive traffic performed by Wi-Fi nodes. In this scenario, Wi-Fi is transmitting a VoIP-like type of traffic, with a packet size equal to 160 bytes, sent every 20 ms, while LTE-U transmits a CBR traffic at the rate of 7 Mbps. Fig. 7.25b shows the latency per packet. When puncturing is not used there is 0.65 probability that the latency is lower than 20 ms, while in the case of puncturing, this probability reaches 0.95. The same effect can be observed for the throughput in Fig. 7.25a. These results show that, **for VoIP like traffic, puncturing facilitates that Wi-Fi finds on average the channel free more often, and reduces the need for backoff.**

7.10.4 Impact of LTE-U AP scan

In this section, we analyze the impact of AP scan parameters, i.e., the AP scan duration, T_{apsc} and the AP scan interval, I_{apsc}. We show the results of the impact of these parameters in Figures 7.26a and 7.26b. We notice a negative impact on LTE-U performance when increasing the AP scan duration, due to an increased delay for the flows whose arrival time overlaps with the AP scan interval. Channel occupancy by LTE-U decreases when AP scan duration is increased, and is 8.3%, 8.1% and 7.7% when the AP scan period is, respectively, 120, 240 and 360. The impact on Wi-Fi is low since Wi-Fi observes on average the same activity of LTE-U during the regular CSAT transmissions. As LTE-U nodes perceive low channel utilization, around 0.12, the duty cycle converges to the maximum 0.875 value. The only difference is experienced by Wi-Fi flows which fall into AP scan periods, and which may experience the benefit of extended AP scan periods, as can be observed for few flows in Fig. 7.26a. While Wi-Fi sees less channel occupancy for higher T_{apsc}, it also experiences more collisions since it is more likely for it to collide with mandatory MIB and LDS that are sent during the AP scan periods. In Fig. 7.26b we observe, as the intuition suggests, that the less frequent the AP scan, the higher performance of LTE-U.

7.10.5 Impact of hidden nodes on LTE-U and Wi-Fi coexistence

Fig. 7.27 compares performance in the BS corner scenario, compared to the indoor one. The objective is to show the impact of an increased number of hidden nodes on LTE-U and Wi-Fi coexistence. Unlike LAA, LTE-U does not backoff to Wi-Fi, so that the Wi-Fi transmissions

are more affected by LTE-U transmissions than they are when coexisting with LAA, in case that Wi-Fi is not able to properly sense the activity of LTE-U. In the BS corners scenario the average distance among nodes is higher than in the indoor scenario. In this context, it increases the number of Wi-Fi nodes that cannot detect ongoing LTE-U transmissions when their energy is lower than -62 dBm. Since LTE-U does not backoff, Wi-Fi nodes that energy detect LTE-U nodes in the range [-72, -62] dBm do not have channel access priority that they have when coexisting with LAA. As a result, Wi-Fi will be affected not only by the collisions that it causes, but also by the collisions generated by LTE-U nodes. The maximum size of the Wi-Fi backoff counter increases significantly in the BS corners scenario and is 530 slots, compared to the 42.25 slots in the indoor case. This confirms that the Wi-Fi transmissions are more affected by collisions in the BS corners scenario. **As an additional metric, we observe that the number of collisions in the BS corners scenario is higher when Wi-Fi coexists with LTE-U, than with LAA, which confirms that in a challenging scenario in terms of hidden nodes, LAA provides a more robust technology option than LTE-U.**

7.10.6 Impact of LTE-U CTS-to-Self

We evaluate the impact of the addition of the CTS-to-Self mechanism to LTE-U node on both, 3GPP indoor and BS corners scenario. Results are shown in Fig. 7.28. We observe a slightly negative impact of CTS-to-Self mechanism on Wi-Fi performance, since Wi-Fi backs off more, so the latency increases. We note that Wi-Fi is not impacted significantly by CTS-to-Self, because it is getting enough channel time when LTE-U is OFF. The percentage of collisions reduces in both scenarios, indoor and BS corners, when CTS-to-Self mechanism is employed. In the indoor scenario there are 6.23% collisions when CTS-to-Self is not used, compared to 4.44% when it is in use. Similarly, in the BS corners scenario there is 6.18% of collisions when CTS-to-Self is not employed, while there is 4.23% when CTS-to-Self is enabled. Among other metrics, we observe a significant reduction in CW size when CTS-to-Self is used in BS corners scenario. The backoff counter reaches values of around 530 slots in the case without CTS-to-Self, while when this mechanism is used it drops to 47 slots. In the indoor scenario, we do not observe a similar effect. However, as we discussed earlier, the rate control implementation may have an important influence. In some vendor's rate control implementation, the packet drop may result in severe coding rate backoff and even connection failures, and this is why this feature is in general recommended by WFA to be included in LTE-U products.

We observe that in both cases, LAA and LTE-U coexistence scenarios, the CTS-to-Self mechanism can protect LAA and LTE-U while having a minor negative impact on the Wi-Fi performance. The advantage for Wi-Fi is evident in scenarios with challenging hidden nodes conditions, since the increased latency produced by the augmented backoffs, is compensated by a significant collision reduction when the nodes are placed farther from APs.

Table 7.12 Simple scenario received power as a function of $d2$

Distance (m)	BS<->BS (dBm)	BS<->own UE (dBm)	BS<->neighbor UE (dBm)
10	-54	-48	-48
20	-61	-48	-59
30	-66	-48	-65
50	-73.5	-48	-73.2
1000	-	-48	-

7.11 Comparison: LAA vs. LTE-U coexisting with Wi-Fi or LTE

In this section, we compare the coexistence performance of LAA and LTE-U when with Wi-Fi or another LAA/LTE-U network. We evaluate performance when coexisting with Wi-Fi in both, the simple and the 3GPP indoor, scenarios. We create variants of the simple scenario by changing the parameter $d2$ (distance between the AP and the eNB) in order to examine different interference situations. In Table 7.12, we show the received power between BS/AP and their UE/STA, and between BS/AP and neighboring STA/UE as a function of $d2$. When $d2 = 10$ m, everyone can energy detect everyone. When $d2 = 30$ m, the scenario with LAA and Wi-Fi becomes asymmetric since the LAA node can energy detect the AP/STA, while AP/STA cannot energy detect the LAA node. Also, in case of LTE-U and Wi-Fi coexistence, LTE-U can carrier sense Wi-Fi since it senses it up to CCA-PD threshold, while Wi-Fi cannot energy detect LTE-U. When $d2 = 50$ m, the scenario with LAA and Wi-Fi becomes symmetric, since at this distance they cannot anymore energy detect each other. On the other hand, LTE-U and Wi-Fi coexistence scenario is still asymmetric, since the AP/STA cannot energy detect LTE-U, while LTE-U is able to preamble detect the AP/STA.

7.11.1 LAA vs. LTE-U: A simple scenario with full buffer traffic

In Tables 7.13 and 7.14, we show the coexistence performance comparison of LAA and LTE-U in a simple scenario (shown in Fig. 7.8), with full buffer CBR traffic over UDP. In particular, Table 7.13 shows the throughput achieved by: 1) Wi-Fi when coexisting with Wi-Fi, 2) Wi-Fi when coexisting with LAA, 3) Wi-Fi when coexisting with LTE-U and Table 7.14 shows the throughput perforamnce of: 1) LAA when coexisting with Wi-Fi, 2) LTE-U when coexisting with Wi-Fi. In a simple scenario, there is 1 UE per operator and 1 CBR flow per UE, and therefore, there is a single throughput value per operator. In general, we observe that when the nodes get closer (d2 decreases) the throughput decreases. This is due to the increased interference that the two networks generate to each other and due to the latency introduced by contention.

When $d2 = 10$ m, the transmitting nodes can see each other. In case of LTE-U over Wi-Fi scenario, Wi-Fi backs off properly to LTE-U, and LTE-U observes a high medium occupancy, due to Wi-Fi activity, and so it correctly reduces its duty cycle to 0.5. LTE-U channel occupancy is 46.57%. However, the collision probability is high, due to the LTE-U OFF to ON transitions.

Table 7.13 LTE-U vs LAA throughput coexistence comparison in simple scenario with a full buffer traffic as a function of $d2$.

Distance $d2$ (m)	Wi-Fi in Wi-Fi/Wi-Fi (Mbps)	Wi-Fi in Wi-Fi/LAA (Mbps)	Wi-Fi in Wi-Fi/LTE-U (Mbps)
10	54.70	13.99	57.58
30	55.14	2.09	77.87
50	60.81	113.48	114.35
1000	115.26	115.28	115.31

Table 7.14 LTE-U vs LAA throughput coexistence comparison in simple scenario with a full buffer traffic as a function of $d2$.

Distance $d2$ (m)	LAA in Wi-Fi/LAA (Mbps)	LTE-U in Wi-Fi/LTE-U (Mbps)
10	188.37	209.44
30	208.76	201.79
50	279.85	226.68
1000	279.85	273.34

Additionally, the interferers are at sufficient distance to cause harmful interference to each other. There are 21% of collisions when LTE-U coexists with Wi-Fi. Still, we observe a slightly better performance of Wi-Fi when coexisting with LTE-U, compared to the baseline Wi-Fi over Wi-Fi scenario, where we do not observe any collision. This means that the interference generated by the LTE-U OFF to ON transitions is well compensated by the LTE-U OFF time, during which Wi-Fi can transmit without sharing the medium with any other network. When coexisting with LTE-U, Wi-Fi network occupies the channel 45.78%, while when coexisting with another Wi-Fi this value drops to only 24.4%, due to the time spent in backoff backoff. Because of this, coexistence with Wi-Fi results in poorer performance than coexistence with LTE-U. As a result, **this is one interesting scenario configuration in which it is true the claim that LTE-U can be a better neighbor to Wi-Fi than Wi-Fi itself**. When $d2 = 10$ m and LAA coexists with Wi-Fi, we note a negative impact on the Wi-Fi throughput. The reason is that the full buffer model makes LAA constantly compete for the channel. As soon as Wi-Fi releases the channel, LAA accesses and transmits for the MCOT duration, since it has always data available in the RLC buffer. Since the average Wi-Fi A-MPDU takes around 4 ms, and LAA MCOT is 8 ms, for channel access category 3, the channel share is unfair to Wi-Fi. In particular, we observe that the channel is occupied by LAA around 84% of the time, while Wi-Fi uses only 11% of the channel share. Note that, the channel occupancy of Wi-Fi is much lower than when it coexists with another Wi-Fi or LTE-U. On the other hand, we do not observe any collision when Wi-Fi coexists with LAA. **To sum up, in a simple scenario when the interferers can hear each other,** $d2 = 10$ **m, and a full buffer traffic model is performed, LTE-U is a better neighbor to Wi-Fi than LAA and than Wi-Fi itself. The LAA configuration parameter that makes the difference to jeopardize Wi-Fi, is the MCOT, considered for the channel access category we used in simulations, which is higher than the maximum Wi-Fi transmission opportunity.** Lower values of MCOT would benefit the coexistence with Wi-Fi, as already studied in previous sections.

When $d2 = 30$ m, the Wi-Fi transmitter does not detect the LAA/LTE-U signal since it is below

Wi-Fi's CCA-ED threshold. Consequently, it does not backoff, which produces an increment in collisions, with respect to the 10 mt case, from 21% to 64.6%. On the other hand, both LTE-U and LAA detect Wi-Fi. Despite the increment of collisions, we observe an improvement in Wi-Fi performance of about 35%, due to the fact that its transmissions are not blocked during LTE-U ON periods. On the other hand, these collisions negatively impact LTE-U. In case of LAA/Wi-Fi, we notice worse coexistence performance than that of $d2 = 10$ m case. The reason for a significant degradation in Wi-Fi performance, and at the same time an increase of LAA performance, is twofold. First, there are a lot of collisions since Wi-Fi does not detect LAA. Second, while Wi-Fi is increasing its backoff and deferring its access to the channel due to failed transmissions, collisions and missed BAs, the LAA node occupies the channel even more often. There are 21% of collisions in this scenario, while at 10 m distance there were no collisions. LAA occupies the channel almost 93% of the time, while Wi-Fi occupies the channel only 2% of the time.

When we further increase the distance to $d2 = 50$ m, the coexistence performance of both LAA and LTE-U improve, and this is because the interference generated to each other at the colliding packets is not high enough to cause a significant disruption. The performance of LTE-U cell increases even if there are more collisions, 73%. However, the LTE-U performance is upper bounded by its duty cycle since it preamble detects the Wi-Fi operator, and consequently, its duty cycle drops to 0.5. Therefore, LTE-U achieves over the unlicensed carrier only around 50% of the maximum achievable throughput per carrier, which is around 150 Mbps, so that, its total throughput is approximately 225 Mbps. In case of LAA/Wi-Fi, both operators reach almost the maximum throughput performance. LAA does not energy detect Wi-Fi, there are 56% collisions, but they do not harm Wi-Fi due to the distance. When Wi-Fi coexists with Wi-Fi, there are only 7% of collisions, but since Wi-Fi preamble detects the Wi-Fi their performance remains limited by backoff. As a result, when $d2 = 50$ m, both ULTE technologies, LAA and LTE-U are better neighbors to Wi-Fi than Wi-Fi itself. LTE-U is a slightly better neighbor than LAA, but at the cost of significantly lower throughput than LAA.

Finally, when $d2 = 1000$ m, the transmitters cannot detect each other, and they are at a safe distance so they do not affect each other's transmissions. LTE-U observes no medium utilization, and so its duty cycle converges to its maximum value of 0.875 (20 ms out of 160 ms have to be OFF for MU monitoring), reaching the maximum throughput of 273 Mbps, which corresponds to the sum of approximately 150 Mbps achieved in the PCell plus about 123 Mbps achieved in SCell. In case of LAA coexisting with Wi-Fi, LAA does not energy detect Wi-Fi, so it achieves the maximum performance. The same happens when Wi-Fi coexists with another Wi-Fi.

As a summary, **in this simple scenario setup where we consider a full buffer traffic over UDP, and we only vary the distance of interferers, results show that we can easily set-up scenario configurations where, unexpectedly, LTE-U is a better neighbor to Wi-Fi than LAA and Wi-Fi itself. This is mainly due to the MCOT value chose by 3GPP for the channel access category 3 that we are considering. More favourable coexistence values would be achieved with lower MCOT.** Regarding ULTE performance, however, LAA is on average achieving better performance than LTE-U.

Table 7.15 CSAT measurements

Distance	MU	Avg. duty cycle
10	0.2	0.86
30	0.21	0.83
50	0.18	0.86

7.11.2 LAA vs. LTE-U: Simple scenario and FTP UDP traffic

In Figures 7.29, 7.30 and 7.31, we show results for the simple scenario when traffic is FTP over UDP and distance d2 is equal to, respectively, 10, 30 and 50 m. In general, both ULTE technologies impact Wi-Fi throughput negatively for the lower values of d2. LAA is a slightly better neighbor to Wi-Fi at 10 m case, while LTE-U is better at coexisting than LAA at 30 m case. However, when $d2$ = 50 m both technologies are better neighbors to Wi-Fi than another Wi-Fi.

When $d2$ = 10 m, the LTE-U cell observes the average medium utilization of around 0.2. In Table 7.15, we show LTE-U CSAT measurements. We observe that with FTP traffic the medium utilization is low to appreciate a balanced share of the channel. We observe 6% of collisions when LTE-U coexists with Wi-Fi, while there are no collisions when Wi-Fi coexists with LAA or another Wi-Fi. **Because of collisions, the performance of the Wi-Fi cell is lower when coexisting with LTE-U, than when coexisting with LAA, even considering MCOT of 8 ms.** On the other hand, LAA impacts the Wi-Fi performance negatively because it occupies the channel more than it would another Wi-Fi network, and this causes an additional delay for some Wi-Fi flows. In particular, the channel occupancy of LAA network when coexisting with Wi-Fi is 8.91%, while the channel occupancy of a replaced Wi-Fi network is only 4.1%. **To sum up, when** $d2$ = 10 **m, and when, differently from before, we consider a bursty, non full buffer traffic model, Wi-Fi suffers degradation in performance when coexisting with both LTE-U and LAA. LTE-U performs worse than LAA since its duty cycle is upper bounded by 0.875 value, and LTE-U occupies the channel less than LAA, i.e. 8.1% of the channel time, and introduces more collisions**.

When $d2$ = 30 m, the negative impact of LTE-U over Wi-Fi decreases, since the power of the interfering signal is lower than the $d2$ = 10 m case. However, LAA impact over Wi-Fi becomes worse, due to an increased number of collisions, because Wi-Fi does not energy detect LAA. At this distance, there are 14.8% of collisions when LAA coexists with Wi-Fi, and 20.06% when LTE-U coexists with Wi-Fi. However, LTE-U occupies the channel less than LAA (8.6% of channel occupancy for LTE-U, against 9.6% for LAA). As a result, when $d2$ = 30 m, LTE-U is a slightly better neighbor to Wi-Fi, than LAA.

When $d2$ = 50, we observe that LAA and LTE-U are better neighbors to Wi-Fi than Wi-Fi itself. This is because, at this distance, Wi-Fi sees the other Wi-Fi and it backs off, while the interference is low compared to the useful signal. **There are then setups where the time lost during the backoff process affects more the performances of Wi-Fi than the interference itself.** LTE-U observes on average a low medium utilization of 0.18, due to the traffic model, so that the duty cycle again converges to its highest value, which is 0.875. On the other hand, LAA does not detect Wi-Fi at -72 dBm, and therefore it does not back off, neither reserves the channel. Thus, LAA attains a better performance.

As a result, we conclude from these results that **the coexistence performance does not only depend on the access mechanism itself. The apparent conclusion that LAA is a better neighbor to Wi-Fi than LTE-U is not always true, the interference level received, the detection capabilities of the eNB, or the traffic pattern, may make one technology perform better than the other, in such a way that it is not possible to claim that one technology is superior to the other in terms of coexistence.**

7.11.3 LAA vs. LTE-U: 3GPP Indoor scenario with UDP CBR traffic

In Fig. 7.32 we show the performance of Wi-Fi, LAA and LTE-U networks evaluated in the indoor scenario when considering UDP CBR traffic of 4 Mbps of each flow of LAA or LTE-U operator, and 2 Mbps of Wi-Fi. Fig. 7.32 shows results of the following topologies: 1) the baseline Wi-Fi over Wi-Fi, 2) LAA over Wi-Fi and 3) LTE-U over Wi-Fi. In this scenario and with the proposed traffic model, we observe that LTE-U is definitely a better neighbor to Wi-Fi than Wi-Fi to itself, in terms of throughput, which can be observed in Figures 7.32a and 7.32c. In terms of latency, the performance of Wi-Fi in Wi-Fi over Wi-Fi scenario, and LTE-U over Wi-Fi is very similar, as shown in Figures 7.32d and 7.32f. Contrarily, LAA has extremely negative coexistence performance. Getting more into the detail of the statistics, the channel occupancy of Wi-Fi when it coexists with another Wi-Fi is 32.9%, with LAA is 19.1%, and with LTE-U is 23.3%. The channel occupancy of LAA is 80.3% and of LTE-U is 52.1%. LTE-U observes a high medium utilization of around 0.61 and its duty cycle drops to 0.5. There are 20% of collisions when Wi-Fi coexists with LAA, 11.29% when coexisting with another Wi-Fi and 10.71% when coexisting with LTE-U. Based on these stats, we observe that LAA offers extremely poor coexistence performance, because it occupies the channel much more than the other technologies. The reason is that for each LAA or LTE-U flow, one UDP packet arrives every 2 ms to RLC queues. Every time a new packet is available, LAA attempts to access the channel, and when it gets it, it cannot fill all the subframe capacity, because only one MAC PDU is available to be sent. As a result, the LAA transmission opportunity is not used at its best, and the channel is not efficiently occupied during 1 ms. To this we need to add the reservation inefficiency to access the channel with LAA. The spectral efficiency of this access is extremely low. The channel occupancy increases, then, dramatically, compared to Wi-Fi, and this jeopardizes the coexistence performance. This effect, though, is not present in LTE-U, where the channel can be used continuously during the ON period. As a result, **we conclude from these simulation results, that the coexistence performances are highly affected by the traffic pattern, besides that by the channel access approach. In particular, when considering an unbursty traffic model that does not allow to take full advantage of the capacity of the subframe, and of the transmission opportunity, LAA occupies the channel for 1 ms, to transmit little data, while Wi-Fi to transmit the same data would occupy the channel in the order of us. In this case, the coexistence performance is significantly poorer than LTE-U.** This problem in LAA could be reduced at MAC scheduler level, by not allowing access to the unlicensed channel with small packets, or without having enough data to transmit to fill the transmission opportunity. Also, we have to consider that this kind of LAA scheduling inefficiencies in unlicensed access will be less evident with New Radio, since this novel technology will allow enough flexibility through the concept of the flexible Transmission Time Interval (TTI)s, to use more appropriately the scheduling opportunities.

7.11.4 LAA vs. LTE-U: 3GPP Indoor scenario with FTP UDP traffic

Fig. 7.33 shows LTE-U vs. LAA coexistence performance, in the 3GPP indoor scenario, when the traffic is FTP over UDP. **In this set-up, we finally observe some of the results that one would expect when comparing LTE-U and LAA. In fact, this setup was the one selected for evaluation in 3GPP TR36.889.** Let us first go through some statistics. The Wi-Fi channel occupancy is 5.5% when coexisting with LAA, 5.9% when coexisting with LTE-U, and 7% when coexisting with another Wi-Fi. We also observe 0.11% of collisions when coexisting with Wi-Fi, 1.9% when coexisting with LAA, and 4.01% when coexisting with LTE-U. From the figure, we observe that for the flows with higher data rates, the throughput significantly reduces when coexisting with LAA (see section 7.11.2). For this kind of flows, there is a high chance of losing a complete A-MPDU and the corresponding BAs. The impact on these higher data rates flows, mainly depends on Wi-Fi's adaptation rate mechanism. Wi-Fi flows with lower and medium rate are positively impacted by LAA LBT mechanism, since in the coexistence there are not unnecessary backoffs to far away nodes. The negative impact of LTE-U on Wi-Fi is mainly due to the increased number of collisions, and similarly to the LAA case, there is a negative impact on the flows with the higher data rates since there is more probability to lose A-MPDUs due to too high MCS in interference situations. However, we see that both LAA and LTE-U positively impact the Wi-Fi performance for the flows transmitted with the lower MCS.

Differently to what happens with the CBR UDP traffic model, we observe here that LAA achieves better performance than LTE-U concerning the throughput and the latency. LAA occupies the channel 9.3% of the time, while LTE-U only 8.3%. It is not only the channel occupancy the reason why LTE-U achieves worse performance, it is also due to OFF times during which the transmission stops and the delay of the flow increases. We note this effect even if the LTE-U duty cycle converges quickly to its maximum value of 0.875. (min OFF time is 20 ms).

These results obtained in the indoor scenario are aligned with those obtained in the simple scenario discussed in Section 7.11.2. What is generating a great difference in coexistence behaviour, wrt. what we observed in previous section, is the pattern of packet arrivals at application level. **The arrival of data per burst that we get with the FTP UDP model, allows LAA to take advantage of the full capacity of the transmission opportunity, and, consequently avoid to use inefficiently the TxOp as it happens with less bursty traffic models like the CBR shown in the previous section. When LAA uses properly its transmission opportunity capacity, its channel occupancy becomes more similar to that of Wi-Fi, and consequently, the coexistence improves. In general, we observe that LAA coexistence performance is more sensitive than LTE-U's to the characteristic of the traffic model.**

The results that we have obtained in this section are similar to those obtained by 3GPP in TR36.889, since the traffic model, the scenario set-up and the LTE configuration is the same. In particular, here we are simulating an FTP application over UDP and RLC-UM, which is similar to what is simulated in 3GPP groups, where normally only PHY and MAC are modeled. However, FTP applications in reality never run over UDP, but normally they do over TCP. This is why in next subsection we further study the impact of TCP and different RLC implementations.

7.11.5 LAA vs. LTE-U: 3GPP Indoor scenario with FTP TCP traffic

In the current Internet, file transfers are typically run over TCP. In simulations where we simulated the FTP application operating over UDP, the RLC protocol at LTE was configured to be RLC-UM, to avoid retransmissions at both link and transport layers. However, TCP commonly runs over a reliable link, which is supported by RLC-AM in LTE networks. We, therefore, test the FTP over TCP and RLC-AM. In this case, **the coexistence of two windows, TCP and RLC, generates a flow control effect that can alter the data arrival pattern as compared with the bursty arrival rate observed when FTP is run over UDP.** Because of this, to better understand this effect, we further simulate as well TCP over RLC-UM. Note that, since our simulation study is focused on the coexistence in an unlicensed band for DL only, the TCP ACKs in ULTE are transmitted in the licensed band without contention. In Fig. 7.34, we show the Wi-Fi performance for TCP over RLC-AM and RLC-UM, and in Fig. 7.35 we show the performance of LAA and LTE-U.

In Fig. 7.34, we observe that the throughput of Wi-Fi is significantly lower than for the FTP over UDP traffic model. When RLC-AM is used, both ULTE technologies impact Wi-Fi throughput negatively. This can be observed in Fig. 7.34a. This performance degradation is due to the additional contention that LAA and LTE-U cause because they occupy the channel much more than in case of FTP over UDP. In particular, LAA occupies the channel 16.4% of the time, LTE-U 12.3%, while another Wi-Fi occupies it 8.3%. **An increase in the channel occupancy of ULTE is because TCP is affected by longer Round Trip Time (RTT) due to the inherent latencies of the LTE protocol stack**, which we will explain in more details later in this section. What concerns the Wi-Fi coexistence is that the **longer RTT in LTE could cause transmission timeouts, which results in the shrink or reset of the TCP congestion window, which may drop down to few or even only one segment. Once this happens, LTE resources are underutilized, since the subframe (i.e., 1 ms long) is the minimum granularity in LTE for resource allocation.** Thus, resources are in general less efficiently used in ULTE than in Wi-Fi, which holds the channel only the amount of airtime that corresponds to the size of the data to transmit. Additionally, when coexisting with LAA, there are 3.5% of collisions, when coexisting with LTE-U, there are 5.1% and when coexisting with another Wi-Fi, there are 0.88% of collisions.

When RLC-AM is used, we observe that the Wi-Fi throughput is better when coexisting with LTE-U than with LAA, even if there are more collisions when coexisting with LTE-U. The reason for this is that LAA holds the channel much more than LTE-U, which increases the time Wi-Fi spends in backoff, and thus its delay. We notice the same effect for all the flows, independently from the rate at which they are being transmitted. When RLC-UM is used, the Wi-Fi throughput curves intersect, and the impact of both ULTE technologies on Wi-Fi is similar. This is shown in Fig. 7.34c. The channel occupancy of LAA and LTE-U is much lower from that of RLC-AM case, i.e., LAA occupies the channel 8% of the time, and LTE-U 8.1%. There are also less collisions when coexisting with LAA (2.7%), than when coexisting with LTE-U (4.37%). However, these coexistence statistics are still poor when compared to the Wi-Fi over Wi-Fi case, when we observe 4.1% of channel occupancy time, and only 0.88% of collisions.

Concerning the LAA and LTE-U results, in Fig. 7.35 we observe that LAA achieves significantly better performance than LTE-U regarding throughput and latency, for both RLC

implementations. The reason is that both, TCP and RLC-AM, are sensitive to delays, which are much more pronounced in LTE-U due to OFF periods. As a final observation, from Fig. 7.35, we notice that the LAA/LTE-U FTP over TCP throughput is significantly lower than that of FTP over UDP. In particular, depending on the segment size and file size, TCP requires multiple RTT to transmit the data. Due to LTE protocol stack delays, RTT is higher in LTE than in Wi-Fi. In particular, the RTT when transmitting one FTP file of 0.512 MBytes in LTE is around 15 ms. 3-4 ms are required for the TCP segment to be transmitted from the eNB and received at the UE. Then a TCP ACK is generated and approx. 12 ms are required to transmit it, without contention, because the uplink traffic is transmitted in a licensed band. When the UE detects the ACK in the RLC queue, or has something to transmit, the UE RLC sends a Buffer Status Report (BSR) to the eNB MAC. From this moment, it takes 4 ms for the BSR to reach the eNB MAC. Then, the eNB receives the BSR and has to schedule and generate an Uplink (UL) DCI. It takes additional 4 ms before the UL DCI is received at the UE, with the information for it to transmit. When the UE receives the UL DCI, it takes additional 4 ms before the TCP ACK reaches the end destination. These delays are standard LTE delays. To transmit files of 0.512 Mbytes with TCP segment size of 1440 bytes, and using a somewhat aggressive initial congestion window of 10 segments, we found by examining traces that we need about 10 RTTs of at least 12 ms, sometimes more or less, depending on buffer occupancy of each component carrier and timings involved. As we are focusing on a DL only, this RTT length is accentuated by the fact that traffic is not sent in the UL. This means that every time that UE has to send a TCP ACK in UL, it needs first to send a BSR from the UE RLC to the eNB MAC and then wait for the UL DCI. This handshake is expensive from a latency standpoint. If there was data in UL, the TCP ACK could be piggybacked to the data, and this would shorten the RTT since the UE would not have to request and wait for the UL DCI grant, which takes around 8 ms. These TCP performance trends not only are confirmed with LAA and LTE-U scenarios, but also with LTE in general. The issue with TCP in LTE networks was discussed in [242]. However, in case of LTE-U and LAA these delays are further increased due to discontinuous LBT or LTE-U transmissions, and subframes being wasted due to reservation, backoff, puncturing, and OFF periods.

7.11.6 LAA vs. LTE-U: The impact on Wi-Fi beacons

In this section, we study the impact that different LTE technologies may have on the Wi-Fi beacons. Several critics have been raised by industry that coexistence with ULTE technologies may increase the percentage of beacons that are lost, or not correctly received by Wi-Fi, or received in a non-timely fashion. These effects may jeopardize the Wi-Fi association procedures. We have instrumented the simulator with traces to track events of collisions with Wi-Fi beacons. Wi-Fi network sends one beacon every 102,4 msec. We have evaluated the impact on percentage of lost beacons, and standard deviation of the beacon interval, for all the scenarios and traffic models considered in the paper (indoor, simple, BS corner and hidden node scenario, with FTP over UDP and FTP over TCP, DP CBR; in the simple scenario we have considered different distance situation between the nodes). Due to space constraints, and without loss of generality, based on the observation of the gathered results, we only represent a subset of these results, while the rest can be reproduced by the interested reader using the scripts available in the public repository of this work. We consider the indoor scenario with FTP traffic over UDP. In this scenario we observe that when Wi-Fi coexists with Wi-Fi or

with LAA, thanks to the implementation of LBT assessment, no beacons are lost. However, the contention of the channel results in a standard deviation of approximately 13 ms for the coexistence with Wi-Fi, 8 ms with LAA. Coexistence with LTE-U in turn results in 1,49% of lost beacons and approximately 12 ms of standard deviation of the beacon interval. These results show that LAA does not affect Wi-Fi association performance more than Wi-Fi already does, while LTE-U increases of less than 2 points the probability of not receiving a beacon, and introduces the same delay to receive it. Similar results are obtained in the other scenarios.

7.11.7 LAA vs. LTE-U: LAA/LTE-U coexistence performance with another LAA/LTE-U network

In this section, we evaluate LAA and LTE-U performance when coexisting with another LAA/LTE-U network. In Fig. 7.36, we illustrate LAA and LTE-U performance in terms of throughput and latency, when coexisting with another unlicensed technologies of the same type and compare this case with the cases when they coexist with the other unlicensed LTE technology, and with Wi-Fi. Similarly to the procedure followed in previous sections, to evaluate coexistence of LAA/LTE-U and Wi-Fi, we have evaluated coexistence in both simple and indoor scenarios, and with both FTP and constant bit rate traffic models. However, due to space constraints, and without loss of generality, we only represent in this section results for the most general indoor scenario, with FTP traffic model. Scripts to reproduce also the other scenarios are available for the interested reader in the public repository of the work.

LTE-U, when coexisting with Wi-Fi, works by preamble detecting the Wi-Fi headers received from the surrounding APs. In case of coexistence with LAA, we consider that LTE-U has knowledge on the number of surrounding eNBs, through the Automatic Neighbour Relation (ANR) SON function. In this way, it is able to accordingly adjust the duty cycle. Results show that Wi-Fi is the best neighbor to both LAA and LTE, meaning that throughout curves of both LTE-U and LAA are better, when they coexist with Wi-Fi, than when they coexist with themselves. On the other hand, the performance of LAA and LTE-U when coexisting with each other are similar. In particular, LTE-U has a slightly better performance when coexisting with Wi-Fi and LAA than with LTE-U. On the other hand, it is hard to claim if LAA coexists better with itself or with LTE-U, since the CDF curves cross in many points. It is interesting to note that LTE-U performance is more stable since it depends mainly on the duty cycle which in all three cases converges quickly to 0.5; while LAA depends on the availability of the channel, and since LAA and LTE-U are occupying channel more in average than Wi-Fi, 8.62%(LAA) and 8.2%(LTE-U) comparing to 5.26% (Wi-Fi), its performance quickly degrades when coexisting with these two technologies.

7.12 Summary of Findings and Suggestions for future work

In the paper, we have presented a detailed analysis of the performance of ULTE technologies through an accurate simulation campaign, run over a high fidelity, standard compliant network simulator. While discussing the results, we have tried to highlight the main aspects that the simulation insights were reveling. In what follows, we summarize the main findings here.

- The main message that we want to convey in the paper, after this very extensive collection of results, is that there is a general belief that LAA is superior to LTE-U in terms of coexistence. This is true in some setups, as those described in section VI.C.3. However, studying also the impact of other aspects like the interference level (distance among nodes), the detection capability of the nodes, the details of LTE-U CSAT algorithm, the selection of the LAA parameters (e.g. MCOT), the traffic pattern injected in the network (if CBR, bursty or full buffer), it is possible to see that the conclusion is not at all straightforward and the absolute superiority of one access mechanism over the other, cannot be claimed. The common thought in industry and literature that LTE-U is necessarily a worse neighbour to Wi-Fi than LAA is, is also not always true. We have shown and discussed setups, scenario configurations, and traffic patterns, where LTE-U coexists much more fairly with Wi-Fi than LAA.

- Coexistence performance is highly sensitive to factors that affect the channel occupancy (e.g. control signals), and in some cases even more than to the parameter choices of the LBT CCA and backoff algorithms.

- Channel occupancy, and consequently coexistence, is not only affected by the behavior of the PHY-MAC layers, and of the LAA access in particular, which have been evaluated in 3GPP RAN1 and literature, but also by other aspects, related with upper layer protocols, such as TCP and RLC. However, no other previous study has included evaluations of TCP performance, to our knowledge.

- The characteristic of the traffic plays a very important role in coexistence performance, much more than it has been highlighted up to date in 3GPP studies or in other references. For instance, a bursty traffic pattern, such as the FTP run over a UDP or raw transport, may be a best-case scenario for coexistence in LAA scenarios, because inefficiencies of LAA in accessing the channel, due to the resource allocation granularity of 1 ms, can be amortized when transmissions are bursty. Alternatively, less bursty traffic models, or other transport protocols, e.g., TCP, may cause LAA to occupy the channel more frequently and inefficiently and impact the coexistence with other technologies.

- The proprietary rate control in Wi-Fi, can have a very important role in the final coexistence performance.

- There has been much attention on the importance of the ED threshold definition for LAA. However, the expected result that it is always beneficial that the LAA ED threshold is lowered, is untrue in many cases. We have shown unexpected effects, which could not be revealed without the support of a high fidelity network simulator.

- Some design parameters of both LAA and LTE-U may impact on coexistence more than expected: the expensive control messages which need to be sent in unlicensed, like DRS and LDS, the long delays required in LAA to update the contention window, the MCOT definition, the puncturing. In particular, the MCOT appears to be a very important parameter, which for two out of four channel access procedures, has been set to a value quite higher (8 or 10 ms) than the average Wi-Fi transmission opportunity. This leads to a clearly unfair behaviour to Wi-Fi, in case of full buffer traffic.

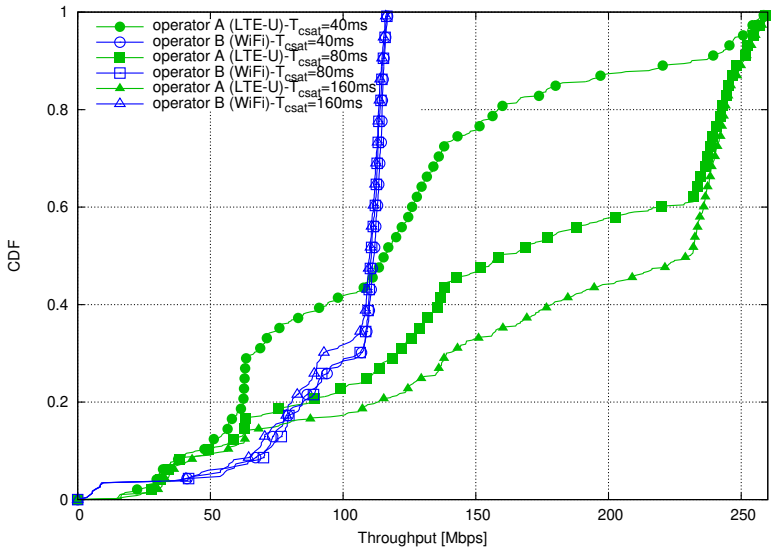
- We have presented in general encouraging results for LTE-U. However, the reader should take into account that we have implemented and studied the Qualcomm's CSAT, which is

only one possible implementation of CSAT, and definitively an excellent one, including many interesting features to achieve fair coexistence, which are not mandatory according to LTE-U specifications.

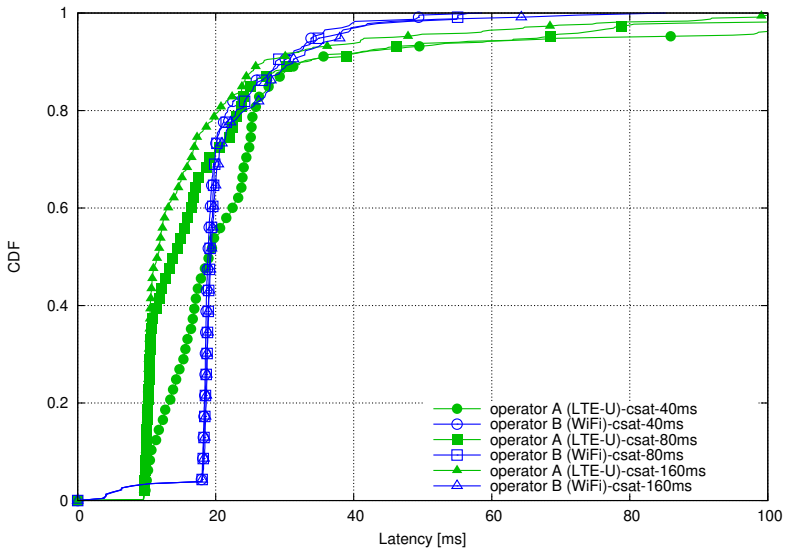
For future research, we recommend the design of smart MAC scheduling approaches capable of solving the inefficiency and granularity issues in LAA resource allocation, which are highlighted in cases of applications run over TCP, or in the case of constant bit rate applications, such as voice. In addition, we recommend investigating the effectiveness of the HARQ based collision detection approach. Moreover, the development roadmap of the simulator may include future evolutions of LAA related with uplink transmissions, i.e., Release 14 eLAA, as well as the MulteFire technology to develop LTE entirely in unlicensed band, without an anchor in the licensed band [212].

7.13 Conclusions

While 3GPP prioritizes LAA and its successors for both, downlink and uplink access in unlicensed spectrum, vendors are considering different unlicensed technologies based on the regulatory requirements, and the operating and capital expenditure of each of them. As ULTE deployments continue to roll out, validating these technologies and an early detection of potential problems, becomes essential for both, vendors and network operators. In this work, we provide a detailed comparison study of the two predominant technologies for LTE in unlicensed, LAA and LTE-U. We have done so by building a full stack, specs-compliant network simulator in the popular ns-3 simulator, and we have open its access to favour results reproducibility. A considerable portion of this work elaborated on examining each of these technologies, its advantages and limitations. Some meaningful and unexpected conclusions have been generated by the analysis of the intense simulation campaign that we have carried out, showing that some of the common thoughts in industry and academia, with respect to these two technologies, are not necessarily true. Many aspects like the considered traffic patterns, the simulation setups, the implementation choices for both LTE and Wi-Fi, play a significant role in the coexistence performance, besides the specific details of the channel access procedures that have long been under discussion.

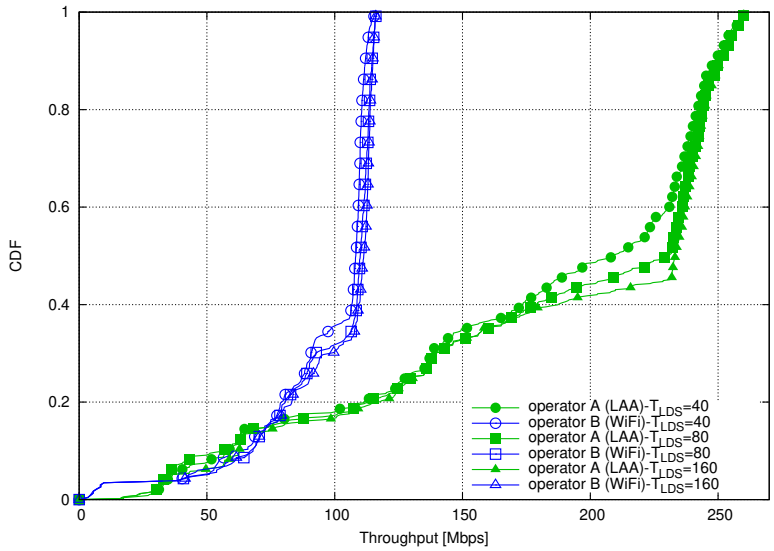


(a) Throughput of LTE-U and Wi-Fi flows

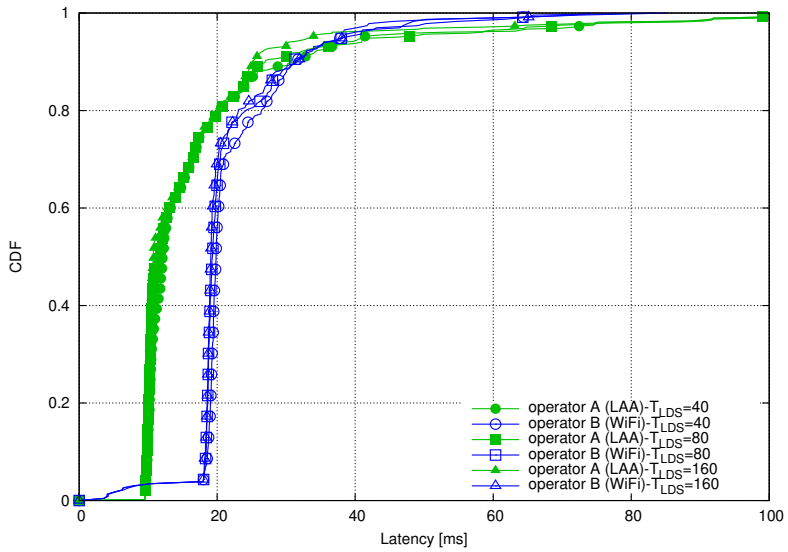


(b) Latency of LTE-U and Wi-Fi flows

Figure 7.22 Impact of T_{CSAT} evaluated in 3GPP indoor scenario.

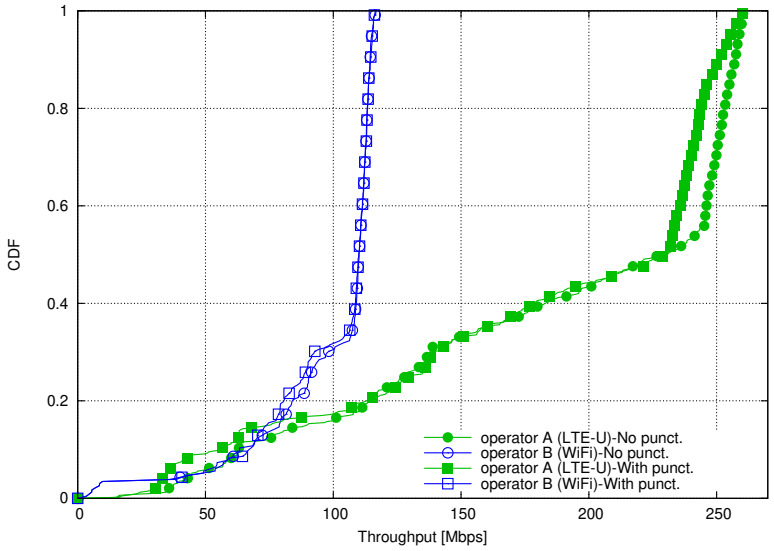


(a) Throughput of LTE-U and Wi-Fi flows

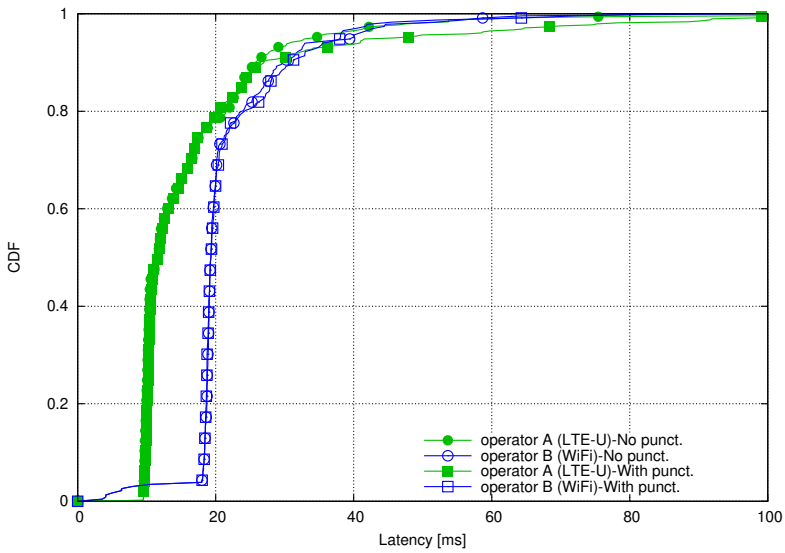


(b) Latency of LTE-U and Wi-Fi flows

Figure 7.23 Impact of T_{LDS} in 3GPP indoor scenario with FTP Model 1 traffic.

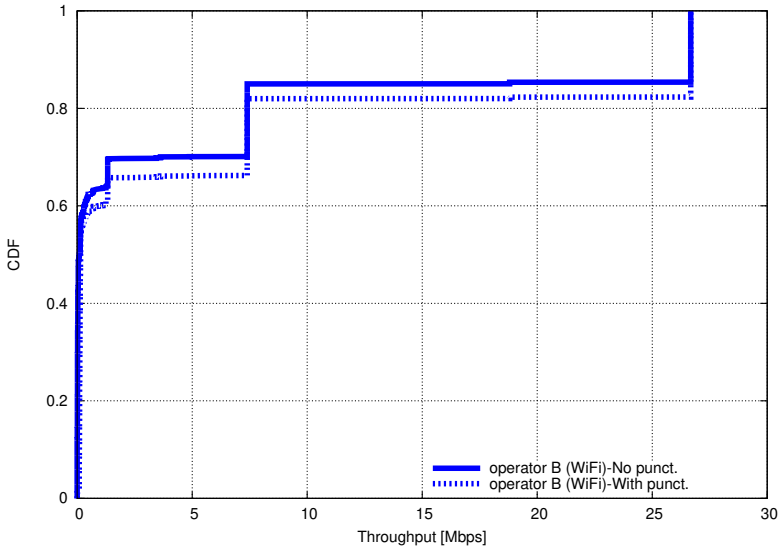


(a) Throughput of LTE-U and Wi-Fi flows

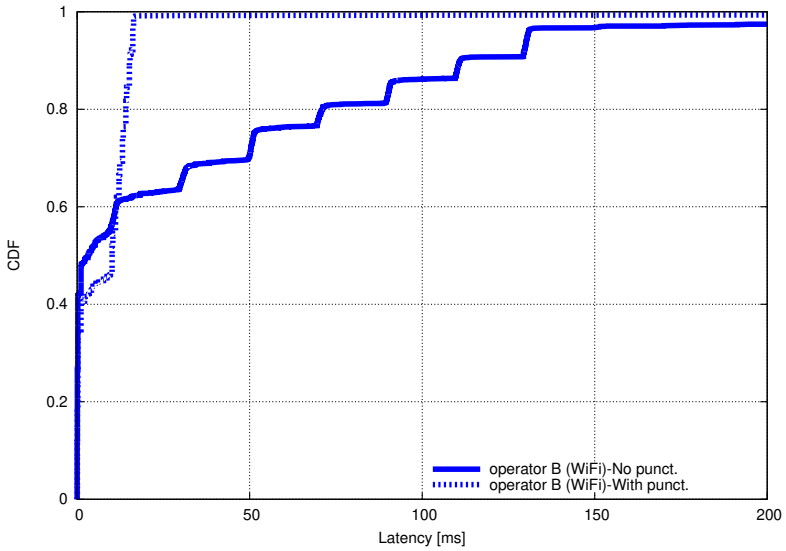


(b) Latency of LTE-U and Wi-Fi flows

Figure 7.24 The impact of puncturing on FTP Model 1 traffic evaluated in 3GPP indoor scenario.

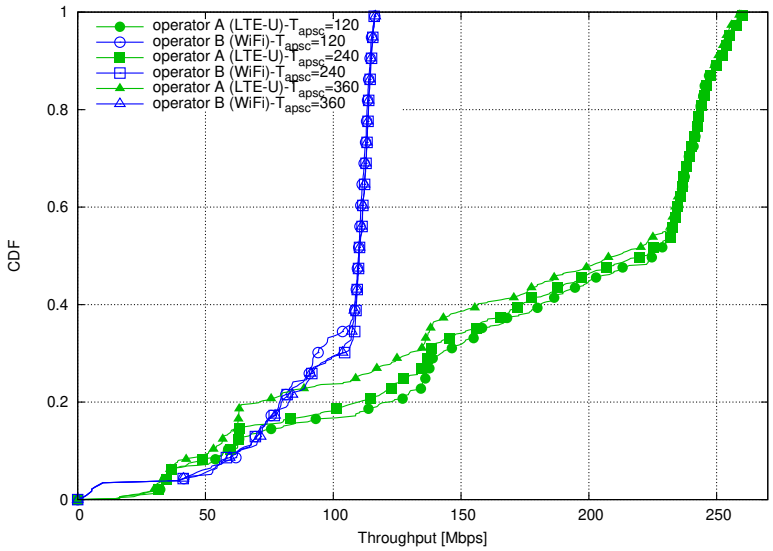


(a) Throughput of LTE-U and Wi-Fi flows

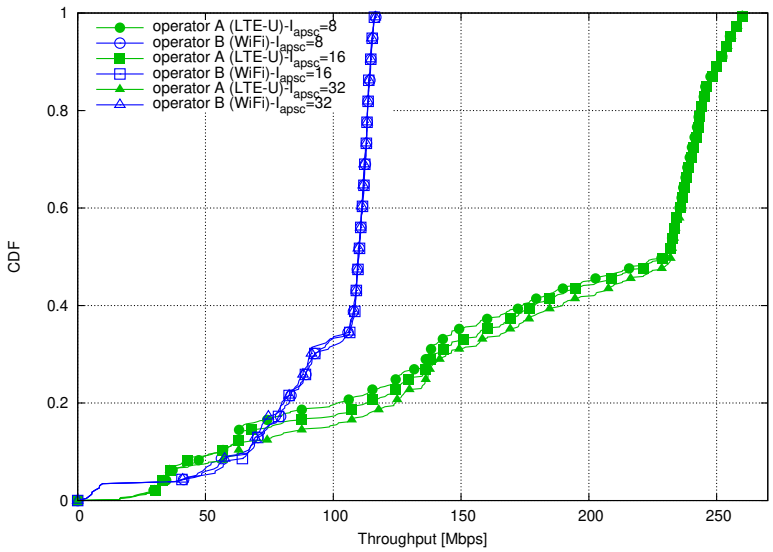


(b) Latency of LTE-U and Wi-Fi flows

Figure 7.25 The impact of puncturing on latency sensitive Wi-Fi traffic evaluated in 3GPP indoor scenario.

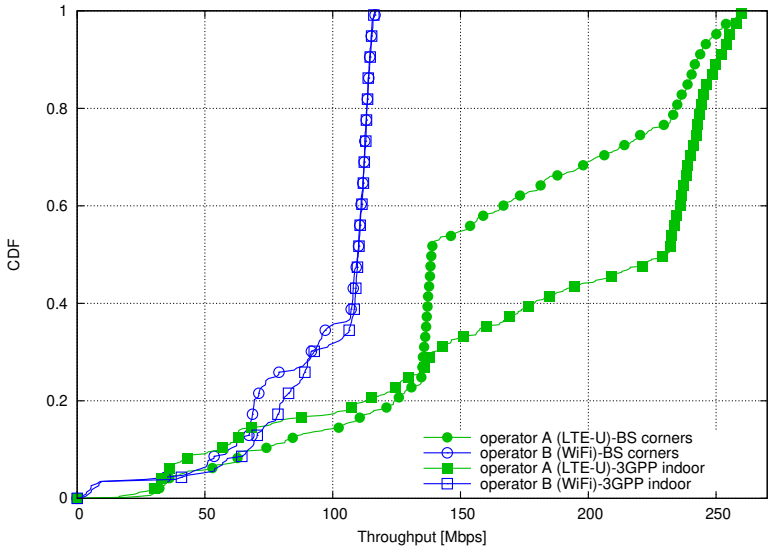


(a) The impact of duration of AP scan, T_{apsc}.

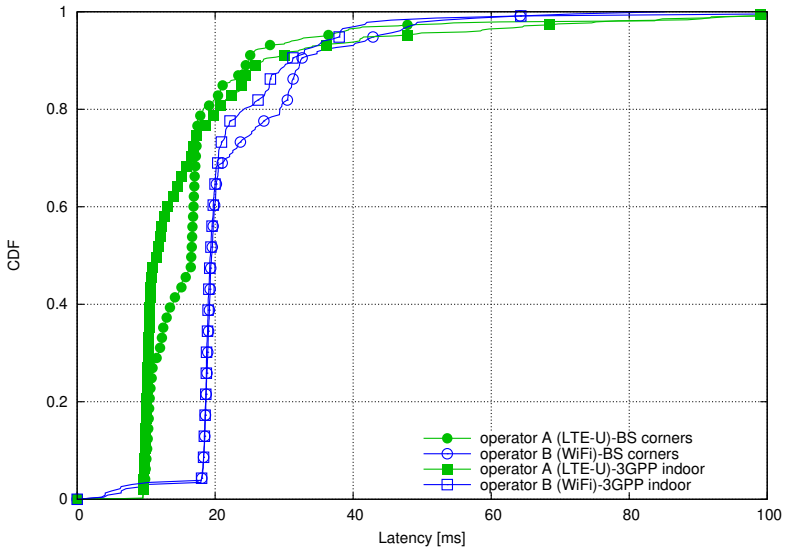


(b) The impact of AP scan interval, I_{apsc}.

Figure 7.26 Impact of AP scan parameters evaluated in 3GPP indoor scenario.

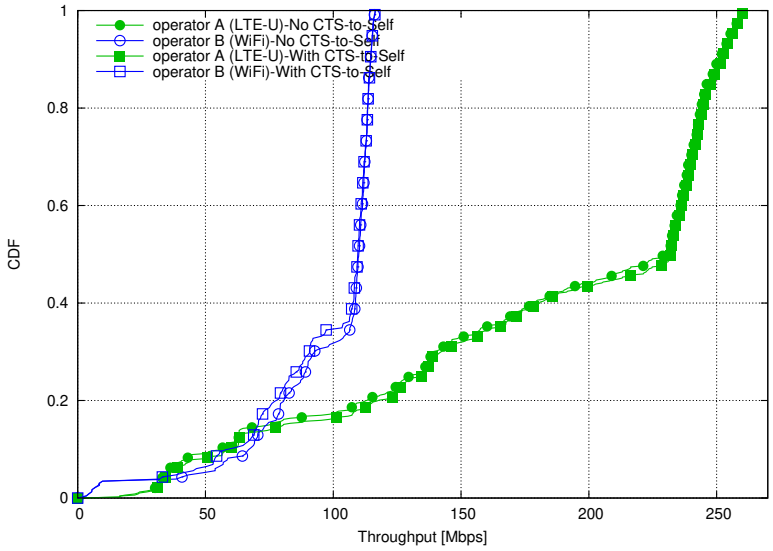


(a) Throughput of LTE-U and Wi-Fi flows

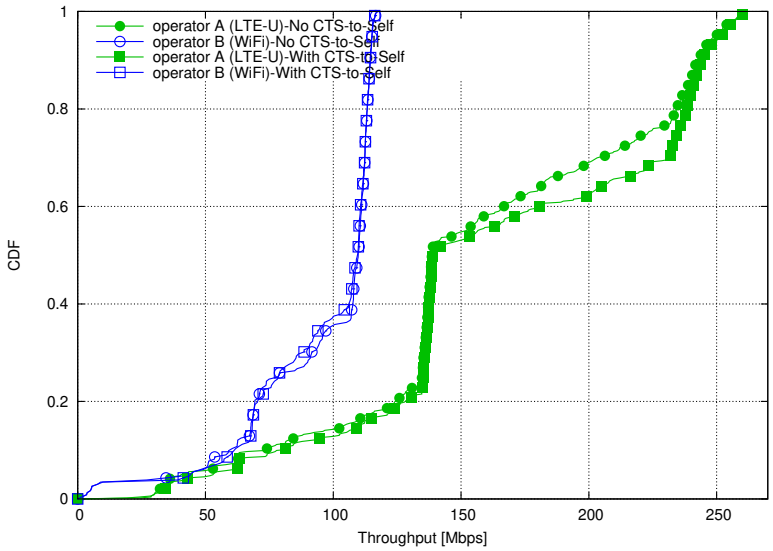


(b) Latency of LTE-U and Wi-Fi flows

Figure 7.27 Impact of hidden nodes evaluated by comparing the performance in 3GPP indoor and the BS corners scenario.

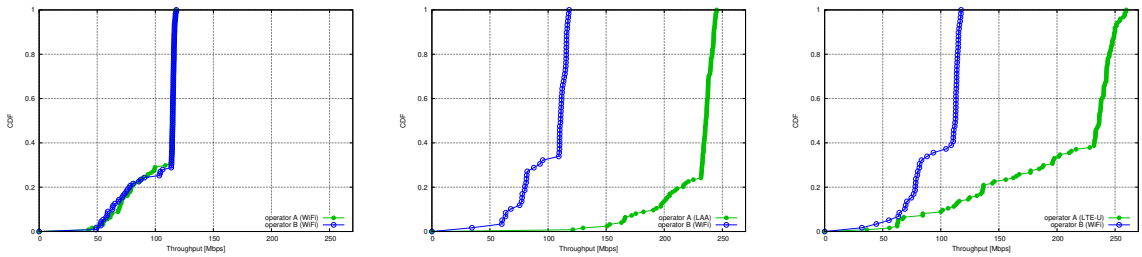


(a) 3GPP Indoor scenario



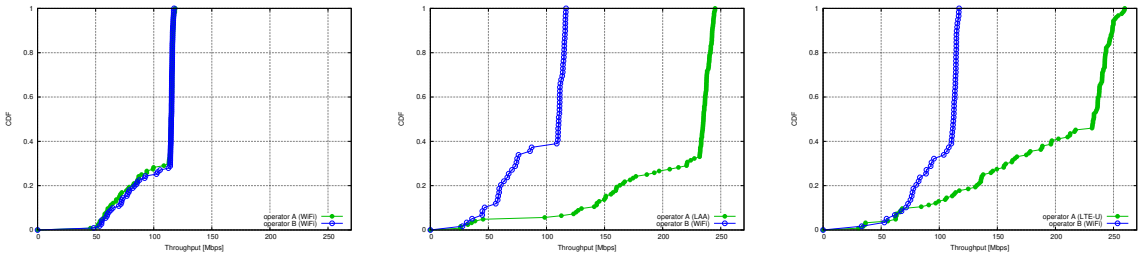
(b) BS corners scenario

Figure 7.28 Impact of LTE-U CTS-to-Self.



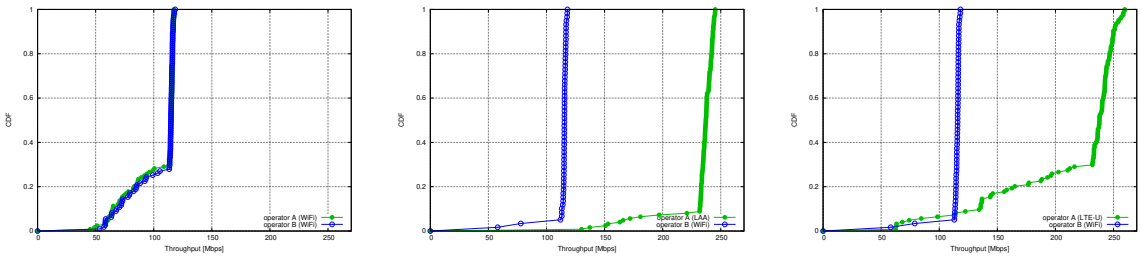
(a) Wi-Fi/Wi-Fi **(b)** LAA/Wi-Fi **(c)** LTE-U/Wi-Fi

Figure 7.29 LTE-U vs LAA coexistence performance comparison in simple scenario when traffic is FTP over UDP. Distance $d2$ = 10 meters.



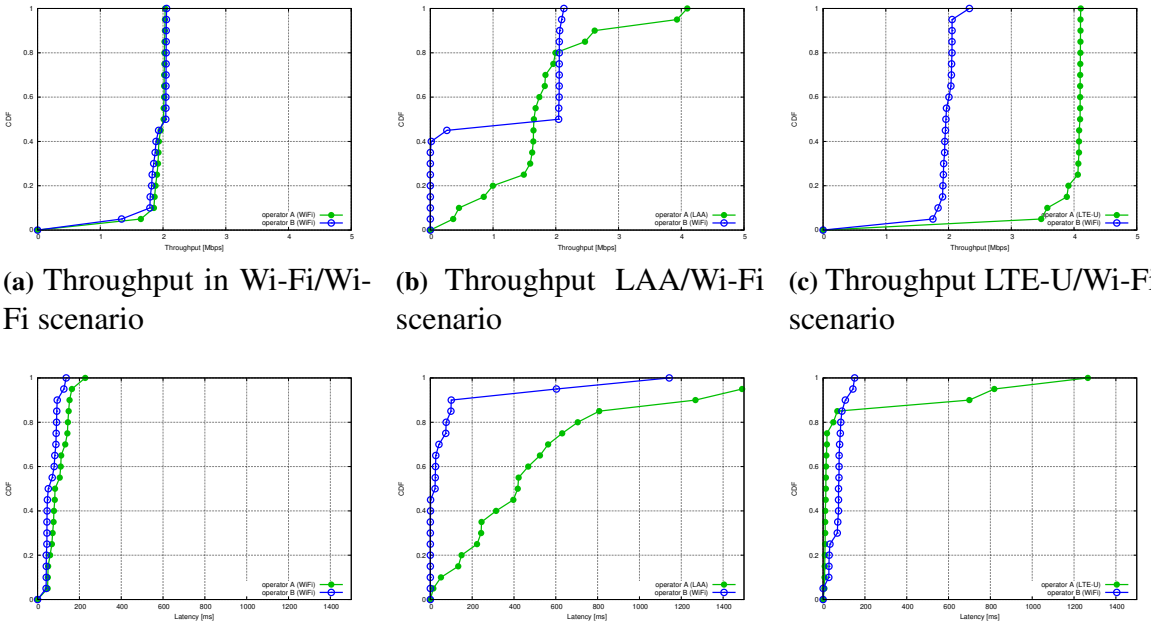
(a) Wi-Fi/Wi-Fi **(b)** LAA/Wi-Fi **(c)** LTE-U/Wi-Fi

Figure 7.30 LTE-U vs LAA coexistence performance comparison in simple scenario when traffic is FTP over UDP. Distance $d2$ = 30 meters.



(a) Wi-Fi/Wi-Fi **(b)** LAA/Wi-Fi **(c)** LTE-U/Wi-Fi

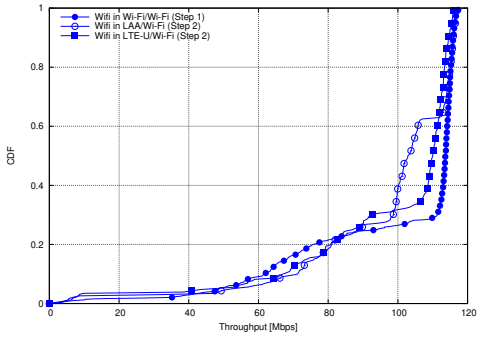
Figure 7.31 LTE-U vs LAA coexistence performance comparison in simple scenario when traffic is FTP over UDP. Distance $d2$ = 50 meters.



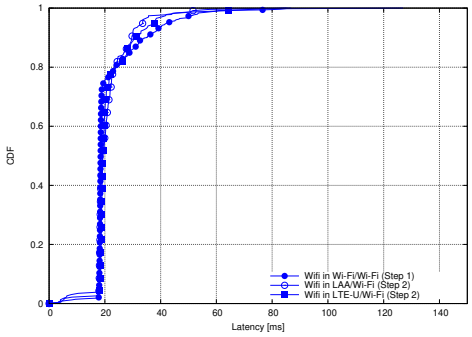
(a) Throughput in Wi-Fi/Wi-Fi scenario **(b)** Throughput LAA/Wi-Fi scenario **(c)** Throughput LTE-U/Wi-Fi scenario

(d) Latency in Wi-Fi/Wi-Fi scenario **(e)** Latency in LAA/Wi-Fi scenario **(f)** Latency in LTE-U/Wi-Fi scenario

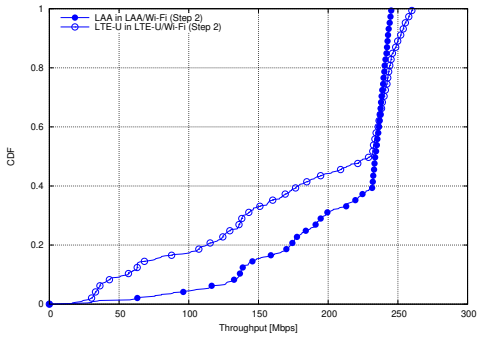
Figure 7.32 LTE-U vs LAA coexistence comparison in 3GPP indoor scenario when traffic is UDP CBR.



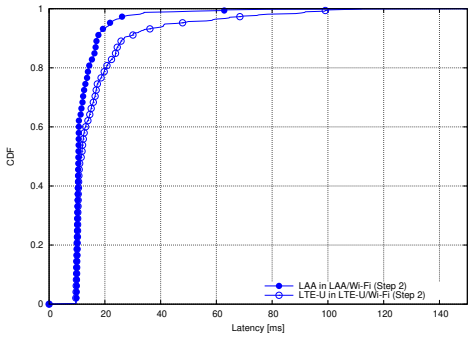
(a) Wi-Fi throughput when coexisting with Wi-Fi, LTE-U and LAA



(b) Wi-Fi latency when coexisting with Wi-Fi, LTE-U and LAA

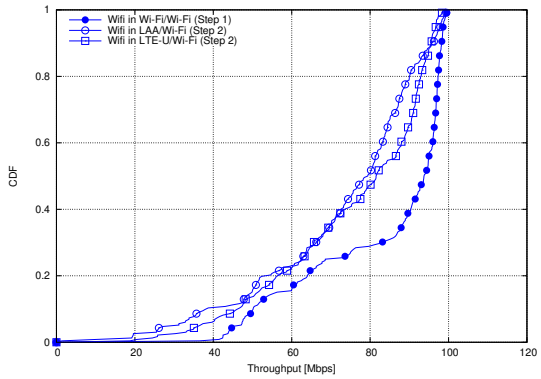


(c) LAA and LTE-U throughput when coexisting with Wi-Fi

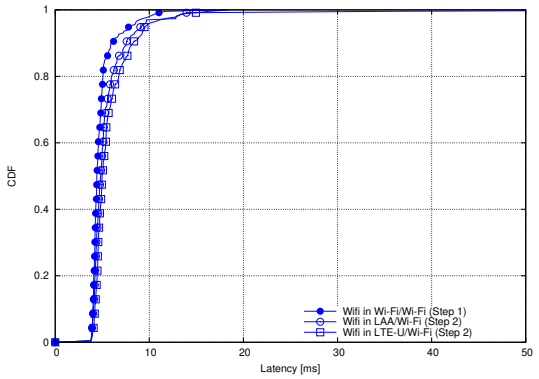


(d) LAA and LTE-U latency when coexisting with Wi-Fi

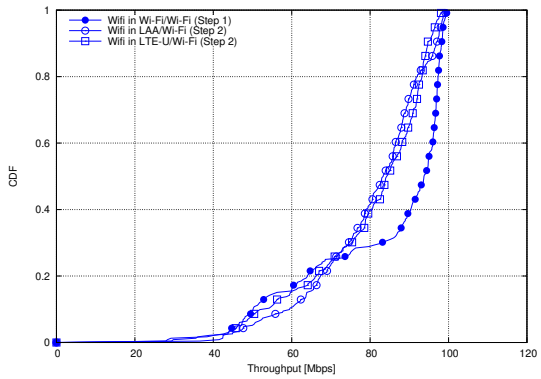
Figure 7.33 LTE-U vs LAA coexistence comparison in 3GPP indoor scenario when traffic is FTP over UDP.



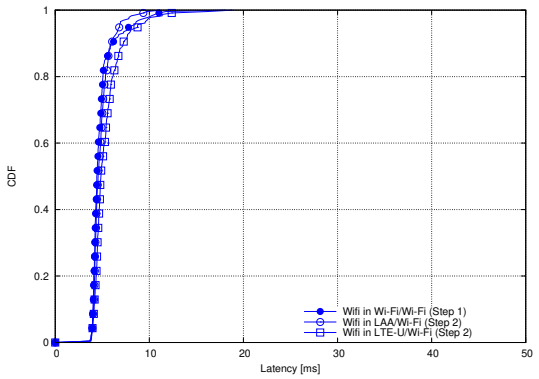
(a) Wi-Fi throughput when coexisting with Wi-Fi, LAA and LTE-U, and LTE RLC mode is RLC-AM



(b) Wi-Fi latency when coexisting with Wi-Fi, LAA and LTE-U, and LTE RLC mode is RLC-AM

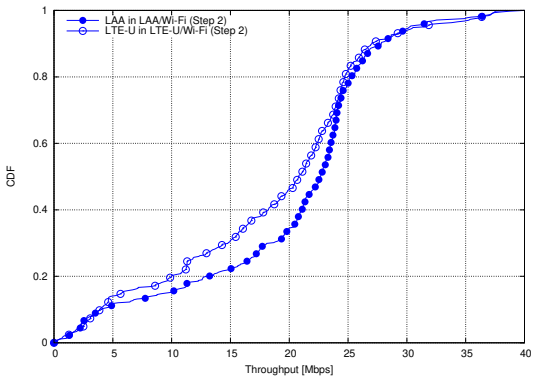


(c) Wi-Fi throughput when coexisting with Wi-Fi, LAA and LTE-U, and LTE RLC mode is RLC-UM

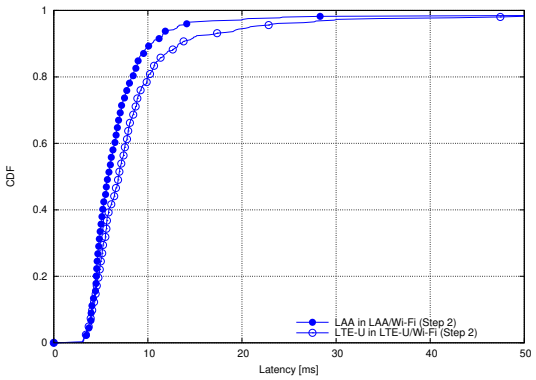


(d) Wi-Fi latency when coexisting with Wi-Fi, LAA and LTE-U, and LTE RLC mode is RLC-UM

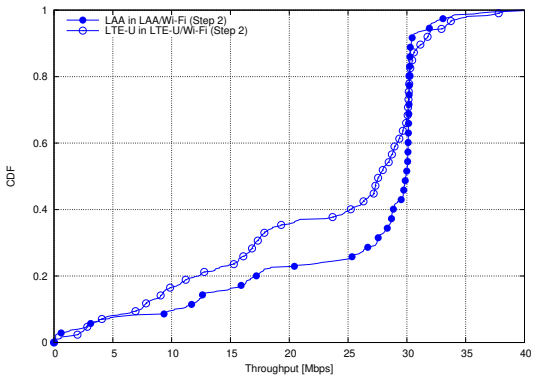
Figure 7.34 Wi-Fi performance when coexisting with Wi-Fi, LAA and LTE-U, the traffic is FTP over TCP, and LTE RLC mode is RLC-AM or RLC-UM.



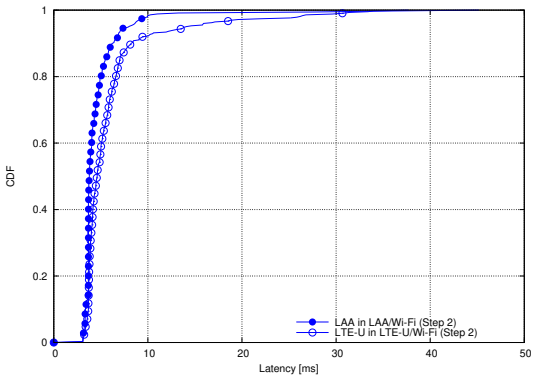
(a) LAA vs LTE-U throughput when coexisting with Wi-Fi and employing RLC-AM



(b) LAA vs LTE-U latency when coexisting with Wi-Fi and employing RLC-AM

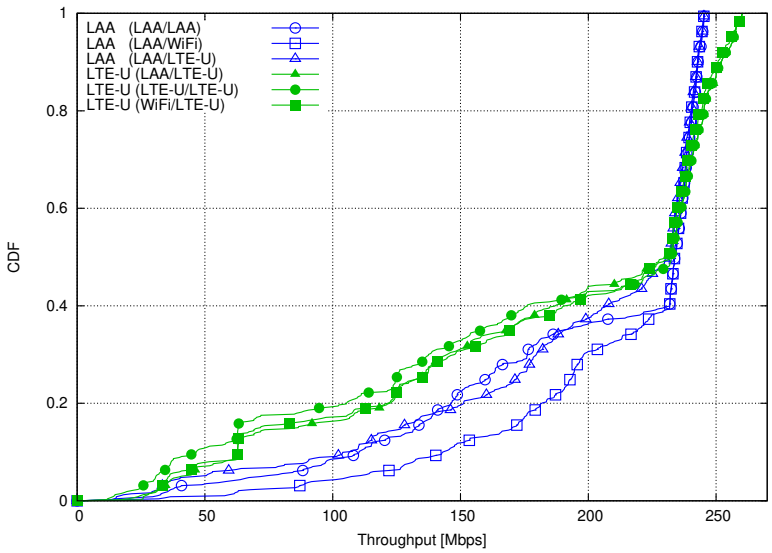


(c) LAA vs LTE-U throughput when coexisting with Wi-Fi and employing RLC-UM

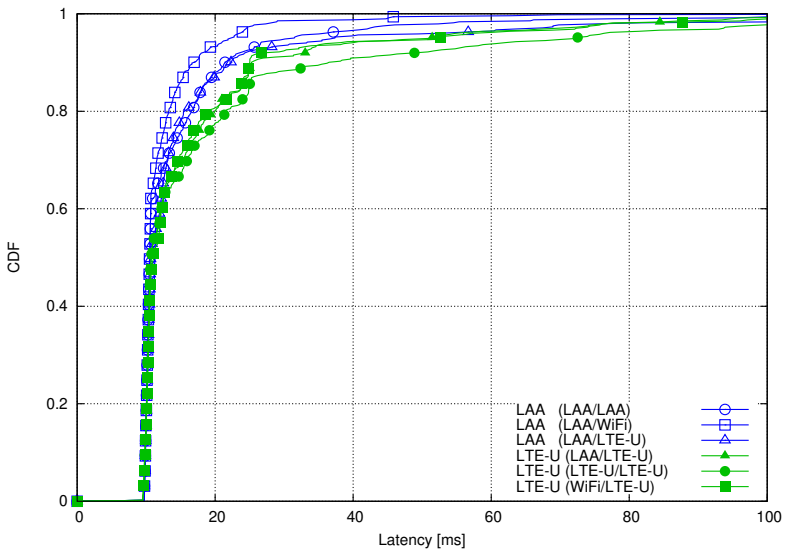


(d) LAA vs LTE-U latency when coexisting with Wi-Fi and employing RLC-UM

Figure 7.35 Performance of LAA and LTE-U when the traffic is FTP over TCP, and LTE RLC mode is RLC-AM or RLC-UM.



(a) LAA and LTE-U throughput when coexisting with LAA, LTE-U, and Wi-Fi



(b) LAA and LTE-U latency when coexisting with LAA, LTE-U and Wi-Fi

Figure 7.36 Performance of LAA and LTE-U when coexisting with LAA, LTE-U and Wi-Fi in indoor scenario, when the traffic is 3GPP FTP M1.

Chapter 8

Coexistence: NR-U/WiGig in 60 GHz band

8.1 Introduction

As discussed in previous chapter, in recent years, unlicensed bands gained a lot of attention from the wireless industry as a way to increase the system capacity. First, the 5 GHz band was used only by military and meteorological radar. Later on, parts of it were opened up for unlicensed use in 1997 and 2003, where some sub-set of bands could be accessed only by devices capable of dynamic frequency selection (DFS) to avoid the interference mainly with radar. In 2009, Wi-Fi devices started to occupy 5 GHz band, and from 2017 the first cellular wireless technologies based on LAA [48, 243] standard began to use it. On the other hand, the 60 GHz spectrum band, in the millimeter-wave (mmWave) range, offers huge opportunities in a much wider spectrum [244], once the challenging propagation conditions have been faced through key technology advancements like the beamforming. Currently, two wireless technologies are occupying the 60 GHz band: IEEE 802.11ad (aka WiGig) [245] and its successor IEEE 802.11ay [246].

The importance of an unlicensed spectrum for the future cellular systems is well recognized by the 3GPP, which, between Release 16 and 17 is standardizing the use of NR for operation in unlicensed bands through an NR-based access to unlicensed spectrum (NR-U) technology [38]. The design of NR-U for sub-7 GHz [1] bands (2.4, 5, and 6 GHz) started in 2018 and a study

[1]Recently in 2020, 6 Ghz band was opened up. Because of this, when refering to lower unlicensed frequency including 6 GHz band we use expression sub-7 GHz.

item on this topic was included in Release 16 [49], while its design for 60 GHz band just began in 2020 when a new study item for unlicensed operation of NR in 60 GHz was added and currently is ongoing in Release 17 [247, 248]. The development of NR-U drives a need in industrial and research communities, for an NR-U simulator that would allow simulating multiple radio access technology (multi-RAT) coexistence in unlicensed bands. The results of simulations of NR-U coexisting with Wi-Fi in 5 GHz band are already obtained by numerous companies and reported in 38.889 [49]; however, the simulators are not publicly available. Additionally, there is a growing research interest in simulating NR-U in unlicensed mmWave bands, especially 60 GHz band, but there is not yet any simulator openly available that supports such simulations.

A specialized NR-U simulator for mmWave frequencies is needed because mmWave bands are characterized by very different propagation characteristics compared to those in sub-7 GHz bands, i.e., due to the short wavelengths, mmWaves signals suffer high penetration losses and are very sensitive to blockage effects. To compensate such propagation losses, it is necessary to use antenna arrays to achieve highly directional transmissions and receptions. Due to the high directionality of transmissions and receptions, and depending on the channel access design, more complex scenarios with hidden and exposed nodes can appear than those in 5 GHz band with omni-directional transmissions and receptions. Consequently, the multi-RAT coexistence model is much different to that studied in previous chapter 7 for the 5 GHz band. Therefore, coexistence models and simulators that are designed for 5 GHz band cannot be used for 60 GHz band without major changes.

To fill this gap, we focused on extending our open-source and ns-3 based NR module [8] to support NR-U coexistence evaluation. As a result, we developed **the first open-source platform for NR-U coexistence evaluations that allows simulations not only in sub-7 GHz but also in mmWave unlicensed bands**. The NR-U simulator extends the NR [8] module by incorporating the 60 GHz regulatory requirements and by enabling multi-RAT coexistence in mmWave bands. Additionally, the WiGig [249, 250] model is extended to support coexistence with NR-U. Due to the directionality of transmissions and receptions in mmWave systems, there are many different options for LBT design and its integration into a system. However, regulatory requirements related to LBT are not very specific, and leave a vast space for possible implementations. At the time of writing, NR-U specification in mmWave unlicensed spectrum is still discussing many key aspects such as whether the sensing should be omni-directional or directional, whether the receiver should also include the LBT procedure and how, how the energy detection threshold should vary, considering also the configuration of the beams, how the access grant should be shared between transmitter and receiver, etc. Considering all these important open aspects, and the timeliness of our work and NR-U simulation platform, we believe that this work is potentialy of high research impact, since it offers a first initial study, and also allows timely simulations and evaluations of different NR-U channel access candidate proposals. The simulation platform can be downloaded from [15].

The final objective of this research work is to study the coexistence between 3GPP and IEEE technologies in unlicensed mmWave spectrum from an end-to-end perspective. To achieve this goal, we carry out the evaluation study of coexistence of NR-U and WiGig technologies in a realistic 3GPP indoor scenario. In this evaluation study, we focus, in particular, on the LBT channel access mechanism, which we consider to be the most critical component of NR-U systems because of two main reasons. First, it is the only channel access mechanism that fulfils regulatory requirements worldwide. And, secondly, its careful design is fundamental to achieve

the fair coexistence and optimal radio resource utilization. This study is unprecedented since there are not yet publicly available simulators or performance evaluation results neither from academia nor from industry and standardization bodies.

In the following, in Section 8.2, we describe some of the most important NR-U features, highlight technological design challenges related to unlicensed channel access and LBT mechanism. Additionally, Section 8.3 provides a brief introduction to WiGig technology. Section 8.4 provides descriptions of the NR-U and WiGig coexistence simulation models for mmWave bands. Section 8.5 provides scenario description, fairness definition and simulation setup. Section 8.6 describes the evaluation study and discusses results. Section 8.7 concludes this chapter.

8.2 3GPP NR-U technology highlights

This section reviews the standardization process of NR-U technology in 3GPP, key design features and functionalities. The work item for NR-U in sub-7 GHz has been part of Release 16 [49, 251], while NR-U for mmWave bands is scheduled for Release 17 and beyond. The objective of these work items is to identify NR enhancements to achieve a single global solution framework for NR based access to unlicensed spectrum.

Release 16 considers the following NR-U deployment options [251]:

- **Carrier aggregation** between NR in licensed spectrum (PCell) and NR in unlicensed spectrum (SCell) with two options: a) SCell has DL only, and b) SCell has both DL + UL. This option is evolution of LTE LAA and eLAA that were proposed in Releases 13 and 14, respectively.
- **Dual connectivity** between LTE/NR in licensed spectrum (PCell) and NR in unlicensed spectrum (SCell); which is based on LTE eLAA introduced in Release 14.
- **Standalone NR-U** in which primary cell uses unlicensed spectrum. This is similar to MulteFire for LTE. The standalone NR-U requires redesign of many procedures to adapt to the fact that all signals go through the unlicensed band, e.g., initial access, scheduling, HARQ should be redesigned.

The key basis for all the NR enhancements for NR-U is to be compliant with the regulatory requirements [38], which in case of 5 GHz and 60 GHz bands include LBT, maximum Channel Occupancy Time (COT), Occupied Channel Bandwidth (OCB), power limits (in terms of maximum equivalent isotropically radiated power and maximum power spectral density) and specific functionalities (such as dynamic frequency selection and frequency reuse). Such requirements impose certain redesign of the standard procedures, channels and signals, as well as challenges at an implementation level. For example, the requirement of LBT creates uncertainty for the channel availability, which is fundamentally different from the license-based access, where all the transmissions occur at prescheduled and fixed times.

Modifications to several NR Release 15 features to support NR-U include:

- Initial access procedures and signals. For example, Synchronization Signal (SS)/Physical Broadcast Channel (PBCH) transmissions and the random access procedure should take into account LBT. Also Physical Random Access Channel (PRACH) preamble transmissions should be changed to meet the OCB requirement.
- DL channels and signals. For example, dynamic PDCCH monitoring and PDSCH transmissions/receptions should support flexible starting time due to LBT.
- UL channels and signals. For example, PUCCH and PUSCH should take into account the OCB requirement, flexible starting points for PUSCH transmissions due to LBT, and sounding reference signal enhancements.
- Paging procedures. This includes flexibility in monitoring paging signal, which may not be transmitted at prescheduled time due to LBT.
- HARQ procedures. For example, additional ACK/NACK transmission opportunities, to avoid declaring a NACK in case of failed ACK transmission due to LBT.
- Configured grant mechanisms. For that, flexibility in time-domain resource allocation can overcome the LBT impact.
- Measurement framework. This envisions changes to the radio link monitoring procedures due to the LBT requirement.

Some of the recent agreements in 3GPP for the NR-U in sub-7 GHz bands include: i) two SS/PBCHs transmissions per slot, ii) PRACH transmission with repetition of sequence in frequency domain, iii) increase of the random access response window from 10 ms to 20 ms, iv) Physical Resource Block (PRB)-based interlace design for PUSCH and PUCCH, and v) additional PDCCH monitoring paging occasions within a single paging occasion.

8.2.1 Channel access procedure

Channel access procedures are also widely discussed in the standard, such as selection of the LBT category and the corresponding parameters for each of the downlink and uplink channels under different conditions. The LBT protocol for NR-U follows the LBT procedure defined for LTE LAA, which was inspired by the Wi-Fi Carrier Sense Multiple Access Collision Avoidance (CSMA/CA) mechanism. A state machine for the LBT CCA process is presented in Section 7.3.1. In particular, for NR-U, all four LBT categories have been considered [49]:

- **Category 1**: Immediate transmission after a short switching gap of 16 μs.
- **Category 2**: LBT is performed without random back-off, in which the CCA period is deterministic (e.g., fixed to 25 μs).
- **Category 3**: LBT with random back-off with a contention window of fixed size, in which the extended CCA period is drawn by a random number within a fixed contention window.

- **Category 4**: LBT with random back-off with a contention window of variable size, in which the extended CCA period is drawn by a random number within a contention window, whose size can vary based on the HARQ feedback.

As in LAA, category 4 is tipically used for data transmissions, while category 2 can be used for specific signaling like discovery reference signals (see details in [49]).

In December 2019 in 3GPP plenary meeting, it was planned to extend NR operation up to 71 GHz, including 60 GHz band, [247, 248]. The key design aspects related to the coexistence of NR-U at 60 GHz unlicensed bands, such as directional LBT and the corresponding beam management impacts [38], are expected to be considered in Release 17 as part of the newly approved study items.

8.2.2 Shared COT

The rules for shared COT have also been defined for NR-U in [49]. When the gap between DL/UL transmissions is less than 16μs the responding device is allowed to transmit without performing a CCA check, i.e., category 1 LBT can be used. When the gap is more than 16μs and less than 25μs, within the same COT, only a short sensing is needed at the responding devices, i.e., category 2 LBT can be used. Finally, if the gap is longer than 25μs, regular LBT has to be done at responding devices, i.e., category 4 LBT. Differently from LAA, which supports a single DL/UL switching point within the COT, NR-U supports multiple DL/UL switching points within the same COT.

8.3 IEEE 802.11ad/ay technology highlights

IEEE 802.11 WLANs standards have started technology development to use the unlicensed spectrum at 60 GHz bands with multi-Gigabit/s data rates, through 802.11ad specification, in 2013 [245], and its recent enhancement through 802.11ay specification in 2019 [246, 252]. As compared to legacy IEEE 802.11 systems, IEEE 802.11ad/ay (or WiGig) includes several fundamental changes and additions. For example, a hybrid MAC layer that has three different mechanisms for transmissions, specific beam training processes for directional transmissions, more pronounced frame aggregation, fast session transfer to 802.11ac/n, etc.

Differently from Wi-Fi in 5 GHz bands, WiGig uses a hybrid MAC approach that has three different mechanisms: contention-based access, scheduled channel time allocation, and dynamic channel time allocation or polling. However, in all the access modes, CSMA/CA is used with a CCA that considers an omnidirectional sensing and a random back-off with a contention window of variable size. In WiGig, differently than NR-U, there is no notion of shared COT, and every device has to apply a CCA before accessing to the channel. However, different duration of the sensing stages are defined for different control/data frames.

The beam steering process has been carefully designed in WiGig in order to precisely align, transmit and receive beams. Particularly, the beamforming training is composed of two phases: sector level sweep and an optional beam refinement phase. Both are based on

beam sweeping processes, thus incurring overheads and additional frames that are defined and devoted exclusively for the beam training.

As previously mentioned, IEEE 802.11ad/ay is particularly designed to operate standalone in the 60 GHz bands. While IEEE 802.11ad considers only operation in a single channel of 2.16 GHz bandwidth, IEEE 802.11ay enables aggregation of up to 4 channels, thus leading to a total aggregated bandwidth of 8.64 GHz. However, both also support fast session transfers to IEEE 802.11ac/n operating in the 5 GHz bands in case of channel blockage, and relay support.

Differently from NR-U, IEEE 802.11ad/ay specifications have been already completed and published. So, for more details about IEEE 802.11ad/ay specifications, we refer interested readers to [245, 246, 252, 253].

8.4 Simulation models: NR-U and WiGiG

We built an NR-U simulator extending the ns-3 NR [8] model. We extended the WiGig models [249, 250] to support the coexistence multi-RAT simulations. In order to simulate 3GPP scenarios, we made these two simulators compatible with the 3GPP channel model that at the time of the work was only part of NR and mmWave ns-3 models, and consequently not compatible with any other model. This extension of the channel model has then been further elaborated in the context of a Google Summer of Code project [9] which I supervised, and included in ns-3 mainline, in order to offer the possibility to use the 3GPP channel model by all the other ns-3 modules. In the following we describe NR-U model and very briefly the WiGig extensions that were needed to enable this coexistence evaluation study.

We started from the NR model described in [8] and we extended it with NR-U specific features. Since NR and NR-U standardization for above 52.6 GHz recently started [247, 248], we have used the latest NR and NR-U Release 16 specification as the basis of our NR-U model, and we adapted it to support 60 GHz specific regulatory requirements in terms of the maximum COT, LBT, OCB, power limits, etc.

Fig. 8.1 shows the main architecture of our NR-U device design. There is a single Component Carrier Manager (CCM) entity per device (next-Generation Node B (gNB) or UE). CCM is an entity that manages the traffic, control, and signals over different carriers. Channel Access Manager (CAM) is an entity that is used by the PHY to access to unlicensed spectrum. CAM implements channel access procedure and determines when the device can access the channel and transmit. Device can have various CAMs, one per each carrier, and each of them can be of different type. CAM communicates with the PHY, i.e., it notifies PHY when it is allowed to transmit (this applies for all types of CAM), receives notifications from PHY regarding CCA status (only in the case of LBT CAM), etc.

The proposed NR-U device model supports non-standalone and standalone deployments, described in Section 8.2. In the case of non-standalone NR-U, the CCM entity is responsible for distributing the traffic between the licensed and unlicensed carriers. For each carrier, we can select the operational mode:

1. **licensed mode**, i.e., no LBT is performed, it uses a CAM that always allows access to channel, or

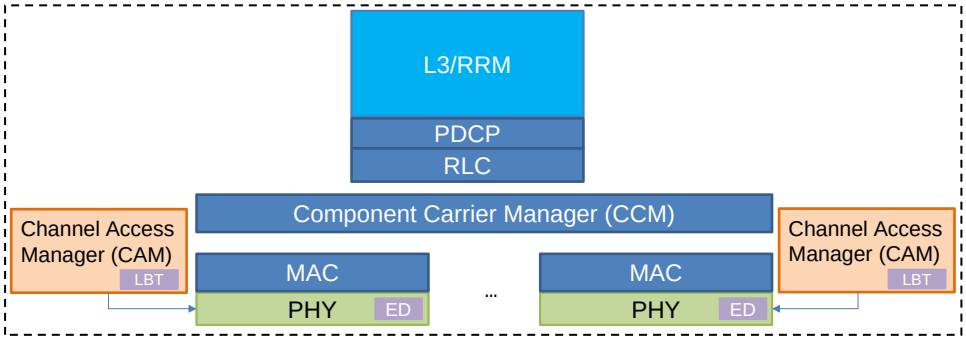


Figure 8.1 NR-U device architecture with multiple component carriers and LBT after MAC processing implementation

2. **unlicensed mode**, i.e., it uses a CAM that restricts the access depending on an algorithm type (LBT or duty-cycle).

For example, NR-U deployment with carrier aggregation can use NR in licensed carrier and NR-U in unlicensed carrier. Additionally, the model allows to configure to use only uplink in licensed and downlink in unlicensed, etc.

As mentioned before, CAM can be configured to use different algorithms. For uninterrupted transmission, such as in licensed band, there is AlwaysOn CAM. For unlicensed spectrum there are two modes: duty-cycle based CAM and LBT CAM. Duty-cycle based CAM further in the text we call OnOff CAM, since it alternates ON and OFF periods. OnOff CAM does not perform any CCA. Finally, LBT CAM implements 3GPP LBT procedure.

8.4.1 NR-U CCA

The sensing capability is implemented in the NR-U PHY model through the ED functionality at the PHY, which performs CCA when indicated by the CAM. For LBT at the gNBs we have considered **omnidirectional sensing**, i.e., omnidirectional LBT, since for the time being it is the only kind of sensing considered for NR-U in 3GPP [254–257]. Release 17 [258] might include more studies on omnidirectional versus directional sensing. Previous studies on omnidirectional versus directional sensing trade-offs, are discussed here [259]. There are two reasons related to LBT design that go in favor of omnidirectional sensing at the gNB. First, it allows to gNB to transmit to multiple UEs using the same COT, without having to perform a time-consuming LBT for each UE direction. Another reason in favor of omnidirectional sensing is that directional sensing at gNB requires complex interactions between the MAC, the scheduler, the PHY and the CAM, and thus directional sensing would require much more complex NR-U design.

At the UE side, instead, we consider **directional sensing**, i.e., directional LBT, since differently from the gNB, the UE only has to communicate with its gNB and because of that we consider it to be a natural design choice. For the UE, for directional sensing is used the same beam that is used for transmission/reception (i.e., assuming beam reciprocity) towards/from the serving gNB.

Figure 8.2 shows the current model at the gNB side, including the omnidirectional LBT sensing

and the Time Division Multiple Access (TDMA) beam-based access, considering three beams (B_1, B_2, B_3), and two UEs scheduled within the third beam (UE_c, UE_d).

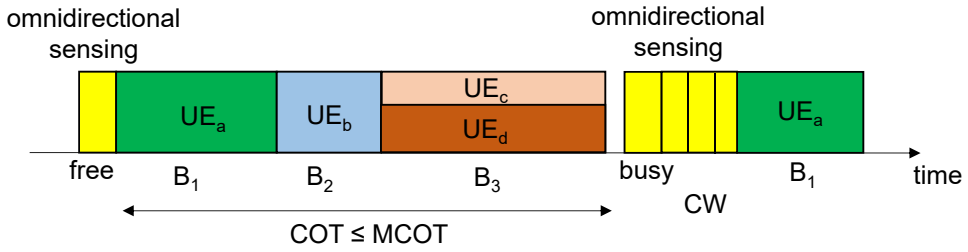


Figure 8.2 Omnidirectional LBT model at gNB side and related scheduling operations to meet the OCB requirement.

8.4.2 NR-U LBT

We have implemented the 3GPP LBT procedure, which is also used for LTE LAA [5]. All four LBT categories are supported in the DL (i.e., at gNB side) (see Section 8.2). In UL, category 1 and 2 are implemented up to now. This means that the model supports DL data and DL control transmissions, as well as UL control transmissions, which are necessary for the performance evaluation of this study.

Every time LBT is successful, the channel is granted for the duration of the MCOT. All the LBT categories have different attributes to configure: the ED threshold, the CCA slot duration, the defer interval during CCA, and the MCOT duration. In addition, the simulator allows configuration of: the minimum and maximum values of CW size for category 4, the CW size for category 3, and the defer period for category 2.

During a single COT, multiple DL-UL switching points are possible and are supported by our model. However, we are focusing on a single switching point. Hence, according to our model the channel is released by gNB whenever there is a DL to UL switch, or when between two DL transmissions there is a gap that is larger than 25 us, as per NR-U specification.

8.4.3 MAC scheduling and LBT

Another very important design aspect is the decision of when LBT is performed at the gNB with respect to the MAC scheduling. In our design, we considered the following two options:

1. **LBT before MAC scheduling**: Start the LBT procedure before the MAC starts the scheduling decisions, i.e., trigger MAC scheduling once the channel is idle/clear;
2. **LBT after MAC scheduling**: Start the LBT procedure after the MAC has scheduled the data, i.e., once that the channel is found idle the transport block is ready to be transmitted (i.e., the frame structure that the PHY will send is already decided).

Note that, in general, in LTE/NR, the MAC works ahead with respect to the slot in which the data actually occupies the channel. For example, LTE eNB MAC scheduling works 2

subframes ahead with respect to the time instant in which the data is transmited over the air in the downlink. Therefore, the two options are not equivalent. Option (1) may generate an inefficiency in spectrum usage because there is a gap between the time instant when the channel is granted and the time instant when it starts to be used for the transmission. On the other hand, in option (2), there is a risk that the channel is not granted when the scheduler has decided to occupy it. Option (2) reduces the inefficiencies in channel occupancy, but it has a more complex model than option (1). Additionally, option (2) guarantees that this model is also adequate for NR-U operation in sub-7 GHz bands, for which the duration of a slot is larger than the one employed for mmWave bands and, consequently, the inefficiency of option (1) would be significant, and would affect the coexistence performance. We have selected the option (2) to add to our model, i.e., LBT after MAC option.

Similarly to LTE, NR transmissions follow a certain frame structure and the beginning of the transmissions need to be synchronized with the slot boundary. LBT introduces randomness in the instant in which the channel is granted. If it happens in the middle of the slot, our LBT algorithm waits for the slot boundary, and it does not reserve the channel, which was a normal practice in LTE LAA. We made this choice because the slot duration in mmWave bands is much lower than it was for LTE, so we consider that the impact is limited. This might generate inter-RAT collisions as WiGig may start transmission in this gap, but the advantage of this approach is that it reduces channel occupancy of NR-U.

8.4.4 NR-U OCB requirement

To meet the OCB requirement, for DL data transmission, we use a TDMA beam-based access, in which OFDM symbols are distributed among beams, and UEs associated to the same beam get allocated during the same OFDM symbols but over different PRBs. For control channels to meet the OCB requirement, we spread such signals through the whole bandwidth. Finally, we constraint the maximum radiated power according to the regulatory requirements, and distribute the available power uniformly among the PRBs to meet the power spectral density constraint.

Note that the developed NR-U model does not include yet the enhancements related to RACH, SS/PBCH, PDCCH, PDSCH, PUSCH, PUCCH transmissions that are being considered in the 3GPP for sub 7 GHz bands (as discussed in Section 8.2).

8.4.5 WiGig simulation model

For WiGig, we have used, as a basis, the models described in [249, 250]. Additionally, to enable the multi-RAT scenario we have modified the 802.11ad model so that (i) interference from other RATs can be represented and taken into account, and (ii) we have unified the beamforming representation through antenna weights, a.k.a., beamforming vectors, rather than using spatial radiation patterns that were used in the WiGig model. This was needed since the ns-3 3GPP channel model (which is the channel model to be used in the NR-U coexistence study) requires beamforming vector representation of beams [9]. Therefore, this was one of the main prerequisites to meet so that NR-U and WiGig can attach to the same channel model and coexist.

In addition, the WiGig model was not considering any link adaptation algorithm and was working only with a fixed and preconfigured MCS during the whole simulation. From the standardization point of view, WiFi/WiGig do not specify any particular strategy to select the MCS. However, adaptive modulation and coding is a very important feature not only for evaluation of WiGig operation itself, but also for coexistence setups, in which it is important to adapt the transmission strategy to the channel observations. In this regard, we have extended the WiGig rate manager, beyond the constant rate manager approach, and we have created a new rate manager that selects and updates the MCS based on the perceived SINR.

Finally, we did other small improvements to the WiGig model, always keeping in mind the coexistence evaluation. These include: random initialization of the beacon intervals (or frame structures) of different APs, and inclusion of a beam reciprocity assumption (as it is assumed in the NR model, and by extension in our NR-U). In case of WiGig, this removes the need to train e.g., the beam of the AP used for transmission and for reception towards a specific STA. Instead, the same beam is used for each AP/STA and so a single training of the transmit/receive beam of such node is required.

In Table 8.1, we summarize the features and functionalities that are used and available in the NR-U and WiGig system-level simulators.

8.5 Performance evaluation setup

This section discusses the simulation scenario, the simulation campaign, and a selection of the end-to-end performance results, which can be obtained through the NR-U simulator that we developed.

8.5.1 Simulation scenario

The 3GPP has defined the deployment scenarios to evaluate the NR-U coexistence performance in sub-7 GHz bands in [49]. Two main layout scenarios are defined based on the deployment and propagation environment conditions: indoor and outdoor sub-7 GHz. We adapt the indoor sub-7 GHz scenario to indoor mmWave [38], to evaluate 60 GHz coexistence scenarios. Since the coverage ranges at the mmWave frequencies are shorter than the 5 GHz band, we define a shorter minimum inter-site distance of 20 m, compared to 40 m in the 5 GHz band. The scenario consists of two operators deploying 3 gNBs/APs each, in a single floor building of 60 m $\times$ 20 m area. Each operator deploys WiGig or NR-U and serves 12 users randomly distributed in the building. We focus on NR-U standalone operation, for being more challenging from the coexistence perspective. The remaining simulation parameters are given in Table 8.2. We illustrate the scenario deployment in Figure 8.3.

8.5.2 Fairness definition

The coexistence evaluation procedure considers two operators, that can have different RATs, e.g., NR-U/WiGig, or the same RAT, e.g., WiGig/WiGig or NR-U/NR-U. The coexistence

Table 8.1 NR-U and WiGig models.

	NR-U	**WiGig**
Frame structure	TDD NR-compliant frame structure with slots and OFDM symbols of numerology-dependent length [8, 90]: - frame: 10 ms, subframe: 1 ms - each subframe has 2^{μ} slots (associated to $15{\times}2^{\mu}$ kHz SCS) - numerologies μ=0,1,2,3,4 are supported - each slot is composed of 14 OFDM symbols - 1st symbol: DL control, 14th symbol: UL control, 2nd to 13th symbols flexibly allocated to DL and UL data	TDD WiGig-compliant [245, 249]: - Beacon interval of 102.4 ms, including BTI, A-BFT, ATI and DTI phases - Beacon Transmission Interval (BTI) of 1.2 ms - Association Beamforming Training (A-BFT), composed of 8 slots and 16 frames per slot - Announcement Transmission Interval (ATI): currently deactivated - Data Transmission Interval (DTI) of 98 ms: currently based on contention, but also supports contention-free and poling [249]
Antenna models	3GPP-compliant [260]: - Antenna arrays: 1 uniform planar array per AP/STA, $M{\times}N$ antenna elements, no polarization - Antenna elements: isotropical radiation and directional radiation are supported	3GPP-compliant [260]: - Antenna arrays: 1 uniform planar array per AP/STA, $M{\times}N$ antenna elements, no polarization - Antenna elements: isotropical radiation and directional radiation are supported
Beamforming	Two methods are available: beam-search method and singular value decomposition (SVD)-based method [261]. Both are ideal in the sense that no resources are used for beam training.	Beam-search method, implemented with a real training through BTI phase (to train AP beam) and A-BFT phase (to train STA beam) [249]
DATA/CTRL channels	- DL/UL data: transmitted and received directionally - DL control: sent quasi-omnidirectionally from gNBs and received directionally at UEs - UL control: sent directionally from UEs and received quasi-omnidirectionally at gNBs	- DL/UL data: transmitted and received directionally - DL control: sent directionally from APs and received quasi-omnidirectionally at STAs - UL control: sent directionally from STAs and received quasi-omnidirectionally at APs
Error models	- NR PHY abstraction for DL and UL data channels [262] including support for MCS Table1 and MCS Table2 [263], LDPC coding and block segmentation [264] - No error model for DL/UL control	- 60 GHz sensitivity error model for DL/UL data and control frames
Modulation	OFDM	Both single carrier and OFDM
Channel Coding	LDPC	LDPC
MCS	QPSK, 16-QAM, 64-QAM, 256-QAM	BPSK, QPSK, 16-QAM, 64-QAM
HARQ	- NR PHY abstraction for HARQ including support for HARQ-IR and HARQ-CC	Not supported by the standard
Retransmissions	Up to 4 with retransmission combining	Up to 7 without retransmission combining
MAC	Scheduled-based access: - In DL, OFDMA and TDMA accesses are supported with round-robin, proportional-fair and maximum rate rules for the MAC scheduler - In UL, TDMA access is supported	Contention-based access for DL and UL
Link adaptation	Two adaptive modulation and coding schemes are supported: Error model and Shannon bound	Link adaptation based on the Shannon bound
Operational modes	Standalone NR-U and Carrier Aggregation NR-U	Standalone WiGig
Channel access	LBT, OnOff, AlwaysOn	CSMA/CA

performance for NR-U is measured in terms of a fairness, which is defined as it was for LAA in Release 13: NR-U is coexisting fairly with WiGig if the peformance of WiGig when coexisting with NR-U is not impacted more than it would be impacted by another WiGig network [38, 265]. Therefore, the standard way to evaluate the fairness is to first consider a baseline WiGig/Wigig deployment (denoted by *WiGig only* further in the text and figures), and then replace one WiGig network by an NR-U. The final step is to compare the baseline scenario peformance with the case when NR-U coexists with the Wi-Fi system.

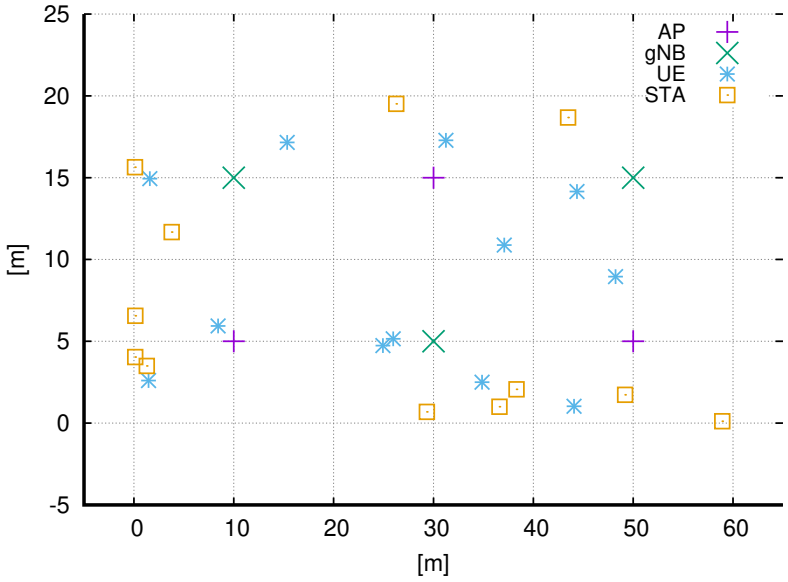


Figure 8.3 Indoor scenario with 3 gNBs, 3 APs, 12 UEs, and 12 STAs.

8.5.3 Simulation campaign

In this performance evaluation we focus on the impact of channel access mechanism on the coexistence performance. We focused on an unbursty traffic model representing a worst case scenario for coexistence. To achieve thhis we configure each UE to perform downlink constant bit rate traffic. The amount of the traffic is adjusted to not reach saturation point but to rather be a representative of a medium load. For the configuration of the system as shown in Table 8.2, we found that 50 Mbps per device represents the medium load. In each figure, we show the results for each of the following NR-U system configurations at gNB and UE:

- *On/On*: NR-U with AlwaysOn CAM at both the gNBs and the UEs;
- *OnOff/OnOff*: NR-U with OnOff CAM, with a 50% duty cycle of 18 ms, i.e., 9 ms ON and 9 ms OFF, at both the gNBs and the UEs, assuming that all gNBs/UEs use time alignment of ON and OFF periods. Duty cycle of 18 ms has been selected because, according to ETSI regulation in 60 GHz bands, a node is allowed to transmit therein in uninterrupted manner during 9 ms;
- *Cat4/On*: NR-U with category 4 LBT at the gNBs and AlwaysOn at the UEs;
- *Cat4/Cat2*: NR-U with category 4 LBT at the gNBs and category 2 LBT at the UEs;
- *Cat3/On*: NR-U with category 3 LBT at the gNBs and AlwaysOn at the UEs;
- *Cat3/Cat2*: NR-U with category 3 LBT at the gNBs and category 2 LBT at the UEs.

For each configuration we run simulation campaign of 20 independent simulations by changing the random seed value.

The performance metrics presented in this performance evaluation are:

Table 8.2 Main simulation campaign parameters

Parameter	Value
Deployment and configuration:	
Channel model	3GPP Indoor Hotspot [266]
Channel bandwidth	2.16 GHz
Central frequency	58 GHz
Mobility model	static positions
Link adaptation	Adaptive MCS
gNB/AP antennas	Uniform Planar Array 8x8
UE/STA antennas	Uniform Planar Array 4x4
gNB/AP/UE/STA transmit power	17 dBm
NR-U subcarrier spacing	120 kHz
Noise power spectral density	-174 dBm/Hz
gNB/AP/UE/STA noise figure	7 dB
NR-U LBT CAM:	
gNB ED threshold	-79 dBm (omniLBT)
UE ED threshold	-69 dBm (dirLBT)
CCA slot duration	5 us
defer interval during CCA	8 us
Maximum COT	9 ms
Cat 4 LBT minimum CW size	15
Cat 4 LBT maximum CW size	1023
Cat 3 LBT CW size	15
Cat 2 LBT defer period	25 us
NR-U OnOff CAM:	
duty cycle	50%: ON and OFF periods of 9 ms
WiGig CSMA/CA:	
AP/STA ED threshold	-79 dBm (omniLBT)
CCA slot duration	5 us
defer interval during CCA	8 us
CSMA/CA minimum CW size	15
CSMA/CA maximum CW size	1023

- the channel occupancy, which represents the percentage of the time that the operator occupies the channel
- latency, which represents end-to-end application layer packet latency, and
- throughput, which represents end-to-end application layer throughput.

8.6 Performance evaluation

Figures 8.4, 8.5 and 8.6 show the occupancy, latency and throughput, respectively. We represent the maximum and minimum value as whiskers, and the values from 5% percentil untill 95% as a box. In each box, a horizontal solid line represents the 50% percentile. The WiGig-only represents the baseline case in which WiGiG coexists with another WiGig. Hence, in WiGig-Only scenario there are in total 24 WiGig users (STAs), distributed between two operators. For the other, non baseline cases, 12 devices belong to a WiGig operator, and the other 12 to NR-U operator.

8.6.1 Impact on channel occupancy

As shown in Figure 8.4, the channel occupancy of NR-U devices is significantly higher than that of the WiGig devices. The reason is that the minimum resource allocation granularity of NR-U in time domain is the entire OFDM symbol, while WiGig's channel occupancy depends on the time needed to transmit the frames, which varies depending on MCS. From the simulation data, we observe that the average length of a WiGig transmission is 3.5μs, while NR-U occupies at least 8.92μs, which is the symbol duration for the subcarrier spacing of 120 kHz. This means that to transmit the same amount of unbursty data, NR-U occupies the channel almost three times more than WiGig. A similar behavior was found in LAA and Wi-Fi coexistence [5], where the difference was even higher, because of the higher LTE resource allocation granularity (1 msec). Hence, in the case of ubursty traffic, and if the data to be transmitted in each symbol is not enough to fully occupy the OFDM symbols in frequency domain, the channel will be inefficiently used by NR-U and this can impact .

On the other hand, as expected, different NR-U channel access techniques result in different channel occupancy, i.e., the channel occupancy of OnOff and LBT strategies (i.e., all 4 categories) are lower than the AlwaysOn strategy. The reason is that during the period when device cannot transmit (OFF period in OnOff mode, or unsucsessful CCA in LBT approach), the data gets accumulated, which allows to more efficiently utilize the symbols in the frequency domain during the transmission opportunities. For example, in OnOff approach, data gets accumulated in the RLC buffers during the 9 ms OFF period. Similarly, any LBT-based implementation results in accumulation of data during the sensing time and while the channel is busy. Among the LBT based techniques, we observe that the more conservative the implementation, due to the higher requirements to access the channel, the more the NR-U channel occupancy is reduced.

In summary, with unbursty traffic models, and due to the fact that the minimum transmission time of NR-U does not depend on the amount of data to be transmitted but on the symbol duration, NR-U can be unfair to WiGig from channel occupancy perspective. To reduce this effect, appropriate scheduling techniques should be implemented to take optimal advantage of the transmission opportunity while taking into account QoS requirements with respect to latency and throughput. In the following, we discuss the impact of coexistence on latency and throughput of each of the RATs.

8.6.2 Impact on latency

Figure 8.5 shows that, in terms of delay, NR-U performs considerably better than WiGig. This is achieved thanks to NR-U features. First, thanks to NR-U scheduling, the intra-technology "blocking" are reduced wrt to WiGig. WiGig has less efficient channel access which causes that on average WiGig devices spend more time than NR-U device in contention, waiting for the channel to transmit. Second, if transmissions collide, HARQ in NR-U still allows to successfully decode the frame through data recombination, while WiGig keeps retransmitting without combining until the maximum number of retransmissions is reached, or a positive feedback is received, which eventually increases the latency. We also observe that for these reasons, WiGig performance presents a higher standard deviation compared to NR-U. In particular, NR-U appears to be more robust to cell edge users, and serves them in a more

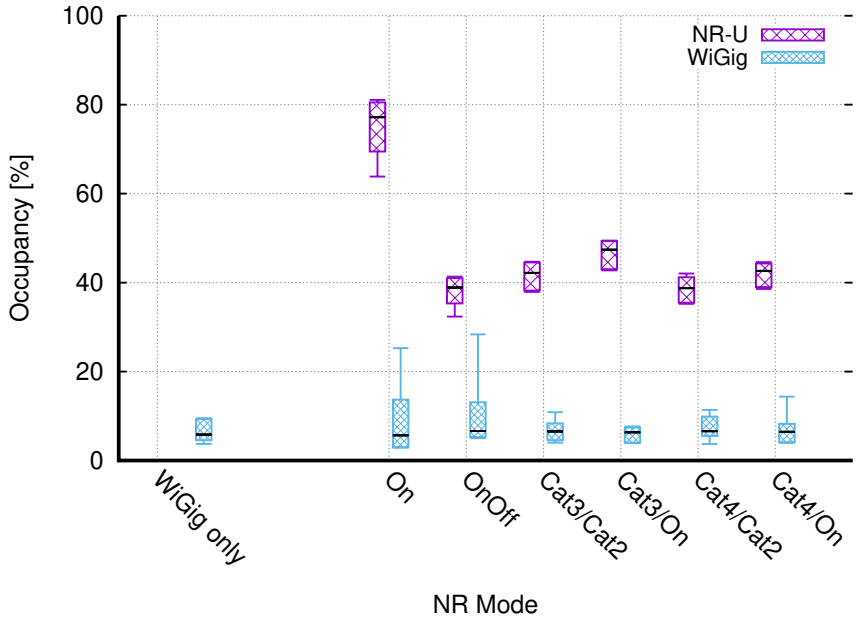


Figure 8.4 Occupancy.

reliable manner, which results in reduced standard deviation of the latency.

If we focus now on median values, WiGig's latency slightly increases when the neighbour operator deploys NR-U, compared to the baseline. The reason lies in the higher channel occupancy of NR-U, compared to WiGig, which results in more contention for WiGig. Comparing the different types of LBT, we can see that more conservative implementations (that consider LBT also at the UE side) offer a slightly lower median latency than the less conservative schemes. The OnOff scheme offers similar performance to LBT, considering that WiGig in this case benefits from a reduced channel occupancy of NR-U during the OFF periods. Finally, the AlwaysOn NR-U is unfriendly to WiGig as expected, as it continuously blocks WiGig's attempts to access. This shows the need of appropriate coexistence schemes to access the channel in unlicensed 60 GHz bands.

8.6.3 Impact on throughput

Figure 8.6 presents throughput results. First, we observe that in all the scenarios, WiGig throughput presents a higher standard deviation than NR-U, independently of whether it coexists with WiGig or NR-U. This is due to the fact that in the proposed deployment, few STAs/UEs are located at the cell edge and WiGig experiences more difficulties to serve this kind of users, compared to NR-U. This makes that the WiGig MCS drops faster than for NR-U's. NR technology appears to be more robust than WiGig to serve cell edge users, and this is due to key features like scheduling and HARQ retransmission combining. If we observe the median values, WiGig throughout is not affected by coexisting with NR-U, independently of the considered channel access scheme, except for the case in which coexists with AlwaysOn NR-U, where WiGig obtains a lower performance. This demonstrates that in terms of throughput, NR-U, with either duty cycle or LBT, fulfills its coexistence design objective.

Impact of CAM. We observe that all NR-U's channel access options, based on LBT

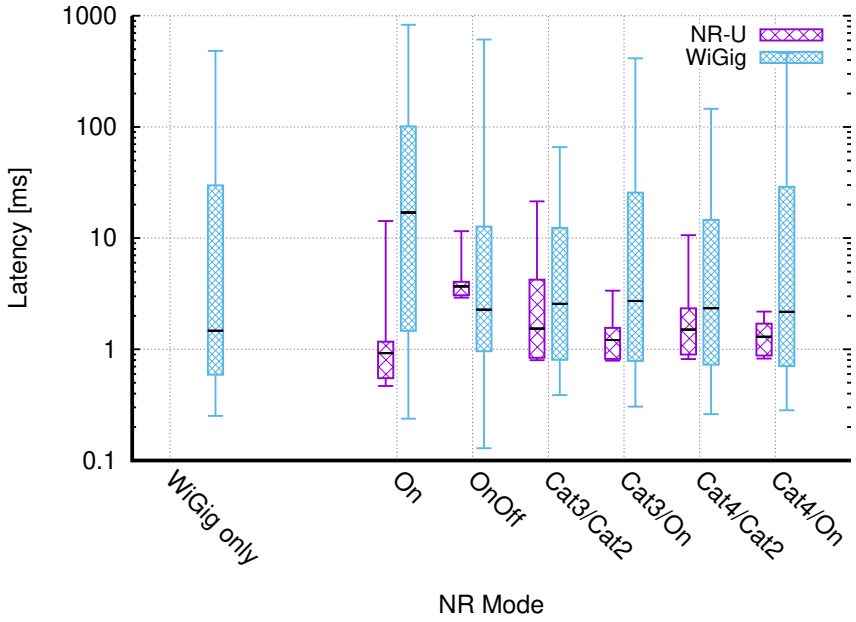


Figure 8.5 End-to-end latency.

or duty cycle (OnOff), are similarly friendly to WiGig. Thanks to the directionality of transmissions/receptions and to the propagation conditions of the mmWave bands, the concrete access schemes do not significantly impact fairness.

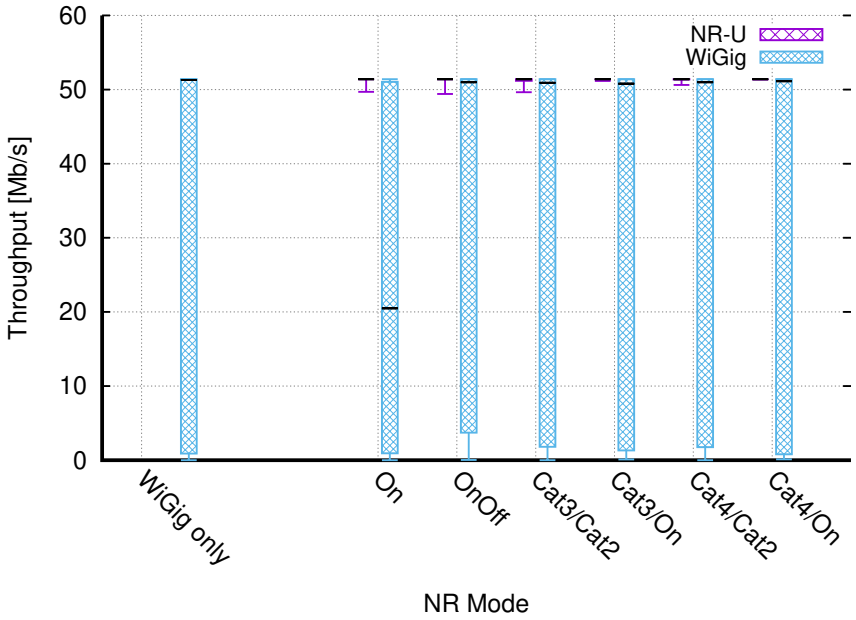


Figure 8.6 End-to-end throughput.

8.7 Conclusions

We have presented an open-source extension to the ns-3 simulator that allows researchers, academia and industry to perform system-level simulation studies of the coexistence between 3GPP NR and IEEE 802.11 technologies in a broad range of unlicensed spectrum bands, from

an end-to-end perspective. The simulator and this study are made available in a timely manner, when the work and study items in 3GPP, targeting NR-U operation in 60 GHz band, have not started yet.

In particular, we have focused on the NR-U and IEEE 802.11ad (WiGig) coexistence in the 60 GHz bands. For that, we have developed an NR-U model, which is based on an extension of the NR Release-15 model to account for the regulatory requirements, such as maximum COT, LBT, OCB and power limits. Then, with such models, we carried out a simulation campaigns. In this study, we specifically focused on the impact of different NR-U channel access schemes on coexistence performance with IEEE WiGig in a 3GPP indoor scenario. We focused on an unbursty traffic model representing a worst case scenario for coexistence.

Under these conditions, we observe that NR-U devices are occupying the channel longer than WiGig to transmit the same data. In spite of this, WiGig latency is marginally impacted by NR-U, and the throughput can be served without degradation, when considering any of the proposed LBT and duty cycle solutions for NR-U. We conclude that the directionality of transmissions/receptions and the particular propagation conditions in the 60 GHz band, favors NR-U and WiGig coexistence. The absence of a specific coexistence oriented access instead, generates unfairness towards WiGig, as it is to expect if NR-U operates in uninterrupted manner.

As future work, we plan to include 3GPP FTP models to NR-U simulator to evaluate the coexistence with bursty type of traffic. Bursty traffic models usually result in different simulation conclusions as compared to constant bit rate traffic models, specially for what regards to the channel access procedure and the channel occupancies. So, we plan to analyze and evaluate the NR-U/WiGig coexistence under different traffic models. Also, it would be interesting to design some more sophisticated sensing strategies that may use information from the receiver, as initially investigated in [267].

Chapter 9

Conclusions

This PhD thesis has been a long journey. During most of the time of my studies I have in fact been working full time as a research engineer at CTTC. In addition, during the process, I have been involved in huge open source developments in the area of LTE and NR. As a result, I had the privilege to see the evolution of both IEEE (Wi-Fi/WiGig) and 3GPP (LTE/NR) technologies across generations and to investigate the research topics before their standardization (i.e., LAA, LTE-U, NR, and NR-U), with some of the main industrial players in IEEE and 3GPP, such as Wi-Fi Alliance, SpiderCloud Wireless, Interdigital, Facebook, and US agencies like LLNL.

The main storyline around which my PhD thesis has been elaborated deals with the radio resource management and strategies to improve the wireless system capacity. We believe that future wireless networks will reach next generations' capacity requirements through a combination of an extreme network densification and spectrum aggregation of licensed, shared, and unlicensed bands (including mmWave frequencies). In complex future wireless systems, radio resource management is expected to be implemented through a self-organized network paradigm in order to achieve an extremely high level of automation (aka *zero-touch, lights-out*) that would allow more sustainable wireless networks, and that would require less investment. In addition to that, another fundamental ingredient of the vision of this thesis has been that future wireless networks automation will be mostly based on AI/ML. As the reader may know well, at this point of the story, this vision is gradually becoming a reality in Beyond 5G networks, with new standardization work already ongoing in the main fora.

In this line, this PhD thesis investigates different approaches for radio resource management based on machine learning and statistical learning for both unlicensed and licensed technologies, and deals with the two main families of wireless technologies, IEEE and 3GPP. Fur-

thermore, the PhD thesis proposes a more general architecture that allows integration of such approaches into SON in a centralized manner, which is suitable for future wireless networks architectures such as Cloud RAN. Finally, the thesis provides a detailed study of the coexistence of the two main families of wireless technologies in unlicensed spectrum, IEEE and 3GPP. First, it analyses different LTE technologies for unlicensed bands, namely LTE-U and LAA, and their coexistence with Wi-Fi in the 5 GHz band, and successively, it focuses on the coexistence of NR-U and WiGig in 60 GHz unlicensed mmWave spectrum.

During my PhD thesis elaboration, I have first focused on Wi-Fi, and so, unlicensed spectrum. Specifically, in Chapter 3 I have presented a ML based AP selection scheme. The performance evaluation showed that the proposed scheme achieves significant improvements in throughput performance compared to other baseline approaches. Furthermore, we validated the proposed scheme through an experimental platform (using Wi-Fi testbed). We confirmed that the proposed scheme could also achieve significant performance improvements in a real setup characterized by much higher variations in the channel conditions.

This work has demonstrated the main advantage of applying machine learning-based models for the radio resource management wrt to the analytical models. This advantage lies in the fact that machine learning-based methods, in their essence, have the native ability to adapt their behavior dynamically through the learning process based on the knowledge base (the past experience represented through different correlated measurements). In contrast, most of analytical methods would need some manual tuning of many of the parameters to adapt to a specific environment to operate efficiently. With this work, we anticipated the now fundamental machine learning trend in wireless networks.

In the following work, shown in Chapter 4, we focused on LTE operation in licensed spectrum. A significant part of the work presented in this chapter has been done during my 6 month visit to Qualcomm Institute at the University of California San Diego (UCSD) under the supervision of Dr Giorgio Quer. In this work, we propose two radio resource management algorithms, one based on machine learning and the other based on statistical learning. In particular, we focus on call admission control taking into account different QoS requirements. The CAC scheme needs to decide whether to admit or reject the incoming radio bearer establishment request while maximizing the number of accepted radio bearer requests and ensuring proper QoS for new and ongoing sessions. While in previous Chapter 3, I have proposed a supervised learning approach based on neural networks, in this work, we have built and evaluated algorithms based on the two different mathematical models: the first one is similar to the approach done in Chapter 3 and is also based on the feed-forward neural network model, while the second is based on the graphical probabilistic model, i.e., Bayesian Networks. The results show that the overall performance of the NN approach is better than that of BN, the general disadvantage of NNs is its shallow level of interpretability and impossibility to tune parameters in order to adjust the performance (i.e., the trade-off between maximizing the radio resource utilization vs. meeting the QoS requirements with a certain level of guarantees).

Once the call is accepted, the QoS management of radio bearers falls into the MAC scheduling domain. For this reason, in Chapter 5 we have focused our attention to the radio resource management at MAC. Namely, the MAC scheduler has a complex optimization task to perform, and that is to maximize the usage of the radio resources while managing the QoS requirements of each of the calls. Consequently, in Chapter 5, we have dealt with the RRM at the MAC and propose, at the time, novel MAC scheduling algorithm that optimizes radio resource utilization

based on the channel feedback, while taking into account QoS requirements. We evaluated the proposed scheme on a scenario in which UEs perform VoLTE traffic, and we compared our scheme with other state-of-the-art QoS-aware LTE downlink schedulers. Evaluation results have shown that the proposed MAC scheduling scheme significantly outperforms other state-of-the-art algorithms available in the literature. In this study, we have observed a huge impact of the MAC scheduling on the cellular system performance, and the importance of employing MAC scheduler that is able to leverage channel quality information in combination with, at least, the QoS parameters and the buffer status reports.

In the following work, we have investigated how machine learning approaches studied in the previous chapters can be extended and applied to a more complex urban small-cell scenario in the context of SON and to achieve a spectral efficiency optimization on a system level, considering also system level KPIs, instead of considering only the user-centric KPIs. To achieve this goal, in Chapter 6, differently from our previous works, we have considered a centralized machine learning approach which forms an intelligent part of the SON framework. We have investigated the problem of a dynamic frequency and bandwidth assignment in a dense small cell deployment. We explored various supervised learning and statistical approaches and analyzed their applicability to a realistic urban dense small cell deployment scenario. We evaluated the proposed learning-based performance approach. We could observe that it yields significantly better performance than the baseline approach and is suitable for real-world cellular systems. With the advent of new network architectures for 5G and B5G, such as Cloud RAN, real systems can relatively easily employ centralized solutions like the one proposed in this chapter.

In 2015, the scientific and industrial community became increasingly interested in using unlicensed spectrum for the cellular systems to increase capacity by gaining access to the spectrum that can be used for free. Firstly, in the 5 GHz band, which was not so saturated at the time as the 2.4 GHz band, and later, with the start of 3GPP work on the NR standardization, in 2017, there was the beginning of industrial interest in NR-U operation in mmWave bands, and in particular the 60 GHz band. Since 2015, the thesis has followed this very engaging research trend of studying the coexistence of these two main families of technologies, IEEE and 3GPP. In Chapter 7, I have proposed a very deep study of the coexistence of LTE and Wi-Fi based on the Release 13, which was at the time the latest 3GPP Release. The research carried out in Chapter 7 has been performed with the key actors in the IEEE and 3GPP. Namely, I have been working on LAA and Wi-Fi coexistence in the 5 GHz band with Wi-Fi Alliance, and afterward, I have been working on LTE-U and Wi-Fi coexistence in the 5 GHz band with SpiderCloud Wireless that has been working closely with Qualcomm. The results were also presented to Verizon. Consequently, this PhD thesis has provided one of the most detailed studies on the matter, an exhaustive and holistic analysis of the impact of different coexistence parameters on the overall system performance of both technologies. Additionally, it has identified the parameters and components of these technologies that play an important role in efficient radio resource management in a coexistence scenario. Finally, it analyses deeply and compares the coexistence performance of LBT and non-LBT channel access schemes (i.e., duty cycle based). It is generally accepted that LBT is a superior access mechanism in terms of interference generation and coexistence with other technologies. However, in this study, we have proven that this claim is not true in all scenarios and settings.

Finally, since 2017, I have been working in the 5G and Beyond area, while studying the coexistence of NR and Wigig in mmWave frequencies, pioneering with Interdigital this area of

research, which is undergoing standardization discussion only now in Release 17. This highly synergical collaboration between my research group and Interdigital has resulted in various contributions: a patent, several high quality scientific publications, and the first open source NR-U module for ns-3 simulator. In Chapter 8, I presented the NR-U and WiGig coexistence models that I have been working on, and the NR-U and WiGig evaluation study. We provided a detailed and holistic system-level performance evaluation of the NR-U/WiGig coexistence that helps better to understand dependencies and trade-offs between channel access mechanism, parameters, traffic, and scenarios.

This PhD thesis contains the description of results and solutions which I have tested, and the work has been accompanied by a huge development effort in ns-3. Each and every line of code that I wrote is released and part of open source ns-3 LTE mainline and ns-3 NR models.

The work presented in this PhD thesis has also been funded by many projects and industrial partners like:

- ARTICO (Chapter 3)
- SYMBIOSIS (Chapters 4)
- FP7 Acropolis (Chapters 3, 4, and 6)
- Wi-Fi Alliance (Chapter 7, the LAA/Wi-Fi coexistence part)
- SpiderCloud Wireless (Chapter 7, the LTE-U/Wi-Fi coexistence part)
- Interdigital (Chapter 8).

My future work will likely continue to be strongly oriented towards multi-RAT coexistence scenarios (NR-U and Wi-Fi/WiGig coexistence), which is definitively the topic that I find very engaging. I was also fortunate recently that my group won an industrial project with Facebook. This project deals with LTE and 5G cellular coexistence system-level simulations for Augmented Reality (AR)/Virtual Reality (VR) use cases. I also plan to continue my engagement in ns-3 and 5G-LENA simulators developments.

Bibliography

[1] William Web (editor), *Wireless Communications: The Future*. John Wiley & Sons, Ltd, 2007.

[2] B. Bojović and N. Baldo, "A new Channel and QoS Aware Scheduler to enhance the capacity of Voice over LTE systems," in *11th International Multi-Conference on Systems, Signals & Devices (SSD'14)*, Castelldefels, Spain, February 2014.

[3] L. Giupponi, T. Henderson, B. Bojović and M. Miozzo, "Simulating LTE and Wi-Fi Coexistence in Unlicensed Spectrum with ns-3," https://arxiv.org/abs/1604.06826, 2016.

[4] B. Bojović, D. Abrignani, M. Miozzo, L. Giupponi and N. Baldo, "Towards LTE-Advanced and LTE-A Pro Network Simulations: Implementing Carrier Aggregation in LTE Module of ns-3," in *WNS3 Proceedings of the Workshop on ns-3*. Association for Computing Machinery (ACM), Jun 2017.

[5] B. Bojović, L. Giupponi, Z. Ali, and M. Miozzo, "Evaluating Unlicensed LTE Technologies: LAA vs LTE-U," *IEEE Access*, vol. 7, pp. 89 714–89 751, 2019.

[6] B. Bojović, S. Lagén, and L. Giupponi, "Implementation and Evaluation of Frequency Division Multiplexing of Numerologies for 5G New Radio in ns-3," in *Proceedings of the 2018 Workshop on ns-3 (WNS3 2018)*. New York, NY, USA: Association for Computing Machinery (ACM), 2018.

[7] N. Patriciello, S. Lagén, B. Bojović, and L. Giupponi, "NR-U and IEEE 802.11 Technologies Coexistence in Unlicensed mmWave Spectrum: Models and Evaluation," *IEEE Access*, vol. 8, pp. 71 254–71 271, 2020.

[8] N. Patriciello, S. Lagén, B. Bojović, and L. Giupponi, "An E2E simulator for 5G NR networks," *Simulation Modelling Practice and Theory*, vol. 96, p. 101933, 2019.

[9] T. Zugno, M. Polese, N. Patriciello, B. Bojović, S. Lagén, and M. Zorzi, "Implementation of a Spatial Channel Model for ns-3," in *Proceedings of the 2020 Workshop on ns-3 (WNS3 2020)*. New York, NY, USA: Association for Computing Machinery (ACM), 2020, pp. 49–56.

[10] B. Bojović, S. Lagén, and L. Giupponi, "Realistic beamforming design using SRS-based channel estimate for ns-3 5G-LENA module," in *Proceedings of the 2021 Workshop on ns-3 (WNS3 2021)*. New York, NY, USA: Association for Computing Machinery (ACM), 2021.

[11] K. Koutlia, B. Bojović, S. Lagén, and L. Giupponi, "Novel Radio Environment Map for the ns-3 NR Simulator," in *Proceedings of the 2021 Workshop on ns-3 (WNS3 2021)*. New York, NY, USA: Association for Computing Machinery (ACM), 2021.

[12] S. Goyal, A. Rob, A. Demir, J. Tooher, J. A. Stern-Berkowitz, M. il Lee, L. Giupponi, S. L. Morancho, B. Bojović, and M. C. Beluri, "Channel access procedures for directional systems in unlicensed bands," U.S. Patent WO/2019/079 500, April, 2019.

[13] "The ns-3 network simulator," http://www.nsnam.org/, (Last accessed September 2020).

[14] LAA and LTE-U extensions for ns-3, http://bitbucket.org/cttc-lena/ns-3-lena-dev-lte-u, (Last accessed September 2020).

[15] "5G LENA Project," https://5g-lena.cttc.es/download/, (Last accessed September 2020).

[16] ACM SIGCOMM, "SIGCOMM Networking System Award," 2020, Last accessed September 2020. [Online]. Available: https://www.sigcomm.org/content/sigcomm-networking-systems-award

[17] S. Yang, "The marketing chain in the mobile internet era," in *2011 International Conference on Machine Learning and Cybernetics*, vol. 3, 2011, pp. 1058–1061.

[18] Andreas F. Molisch, *Wireless Communications*, 2nd ed. Wiley Publishing, 2011.

[19] Andrea Goldsmith, *Wireless Communications*, 1st ed. USA: Cambridge University Press, 2005.

[20] The 5G Infrastructure Public Private Partnership (5G PPP), "White paper: 5G and the Factories of the Future," https://5g-ppp.eu, 2015, (Last accessed September 2020).

[21] G. Szabó, S. Rácz, N. Reider, H. A. Munz, and J. Pető, "Digital twin: Network provisioning of mission critical communication in cyber physical production systems," in *2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, 2019, pp. 37–43.

[22] X. Vilajosana, C. Cano, B. Martinez, P. Tuset-Peiro, J. Melià, and F. Adelantado, "The wireless technology landscape in the manufacturing industry: A reality check," *IEEE COMSOC MMTC Communications - Frontiers: Special Issue On Multiple Wireless Technologies and IoT in Industry: Applications and Challenges*, vol. 12, no. 6, pp. 38–41, Nov 2017.

[23] M. Dohler, T. Mahmoodi, M. A. Lema, M. Condoluci, F. Sardis, K. Antonakoglou, and H. Aghvami, "Internet of skills, where robotics meets AI, 5G and the Tactile Internet," in *2017 European Conference on Networks and Communications (EuCNC)*, 2017, pp. 1–5.

[24] K. Schwab, *The Fourth Industrial Revolution*. USA: Crown Publishing Group, 2017.

[25] J. Yang, B. Ai, I. You, M. Imran, L. Wang, K. Guan, D. He, Z. Zhong, and W. Keusgen, "Ultra-reliable communications for industrial internet of things: Design considerations and channel modeling," *IEEE Network*, vol. 33, no. 4, pp. 104–111, 2019.

[26] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5g-enabled tactile internet," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 460–473, 2016.

[27] K. Valtanen, J. Backman, and S. Yrjölä, "Creating value through blockchain powered resource configurations: Analysis of 5g network slice brokering case," in *2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2018, pp. 185–190.

[28] "ITU-T Technology Watch Report - Analysis of Digital Data Technologies Toward Future Data Eco-Society," https://www.itu.int, 2016, (Last accessed September 2020).

[29] D. Soldani and A. Manzalini, "Horizon 2020 and Beyond: On the 5G Operating System for a True Digital Society," *IEEE Vehicular Technology Magazine*, vol. 10, no. 1, pp. 32–42, 2015.

[30] H. S. Kang, J. Y. Lee, S. Choi, H. Kim, J. H. Park, J. Y. Son, B. H. Kim, and S. D. Noh, "Smart manufacturing: Past research, present findings, and future directions," *International Journal of Precision Engineering and Manufacturing-Green Technology*, vol. 3, no. 1, pp. 111–128, Jan 2016.

[31] Li Guang, "New business models for multi-service 5G transport networks," https://www.huawei.com, January 2019, (Last accessed September 2020).

[32] Claude E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[33] Martin Cooper, et al. Assignee Motorola, "Radio Telephone System," U.S. Patent 3,906,166, September, 1975.

[34] 3GPP TS 36.300 V12.10.0, "Overall description; Stage 2 (Release 12)," June 2016.

[35] 3GPP TS 36.300 V13.14.0, "Overall description; Stage 2 (Release 13)," March 2020.

[36] 3GPP TS 36.300 V14.12.0, "Overall description; Stage 2 (Release 14)," March 2020.

[37] 3GPP TS 38.300 V16.2.0, "NR and NG-RAN Overall Description; Stage 2 (Release 16)," July 2020.

[38] S. Lagén, L. Giupponi, S. Goyal, N. Patriciello, B. Bojović, A. Demir, and M. Beluri, "New Radio Beam-Based Access to Unlicensed Spectrum: Design Challenges and Solutions," *IEEE Communications Surveys Tutorials*, vol. 22, no. 1, pp. 8–37, 2020.

[39] IEEE, "Official IEEE 802.11 Working Group Project Timelines - 2020-09-24, IN PROCESS - Standards, Amendments, and Recommended Practices," Last accessed September 2020. [Online]. Available: https://www.ieee802.org/11/Reports/802.11_Timelines.htm

[40] "Wi-Fi, From Wikipedia, the free enciclopedia," https://en.wikipedia.org/wiki/Wi-Fi, (Last accessed September 2020).

[41] 5G-PPP and ERTICO and EFFRA and EUTC and NEM and CONTINUA and Networld2020 ETP, "5G empowering vertical industries," https://5g-ppp.eu, February 2016.

[42] Sandra Lagén Morancho, *Coordination strategies for interference management in MIMO dense cellular networks (PhD Thesis)*. Universitat Politècnica de Catalunya, 2017.

[43] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G Network Capacity: Key Elements and Technologies," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 71–78, 2014.

[44] Small Cell Forum, "Small Cell SON and Orchestration from 4G to 5G, V1.00," www.scf.io, February 2020, (Last accessed September 2020).

[45] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What Will 5G Be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[46] 3GPP TS 25.308 V16.0.0, "High Speed Downlink Packet Access (HSDPA); Overall description; Stage 2 (Release 16)," July 2020.

[47] LTE-U Forum, "LTE-U Technical Report - Coexistence Study for LTE-U SDL, V1.0," February 2015.

[48] 3GPP TR 36.889 V13.0.0, "Study on Licensed-Assisted Access to Unlicensed Spectrum (Release 13)," June 2015.

[49] 3GPP TR 38.889 V16.0.0, "NR; Study on NR-based access to unlicensed spectrum (Release 16)," December 2018.

[50] Z. Khan, H. Ahmadi, E. Hossain, M. Coupechoux, L. A. Dasilva, and J. J. Lehtomäki, "Carrier aggregation/channel bonding in next generation cellular networks: methods and challenges," *IEEE Network*, vol. 28, no. 6, pp. 34–40, 2014.

[51] Engineering and Technology History Wiki, "Martin Cooper's Biography," https://ethw.org/, (Last accessed September 2020).

[52] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 Gbps/UE in Cellular Systems: Understanding Ultra-Dense Small Cell Deployments," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2078–2101, 2015.

[53] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59–67, 2008.

[54] E. Dahlman et al., "5G radio access," *Ericsson Review (The communications technology journal since 1924)*, vol. 6, pp. 2–7, 2014.

[55] O. Galinina, A. Pyattaev, S. Andreev, M. Dohler, and Y. Koucheryavy, "5G Multi-RAT LTE-WiFi Ultra-Dense Small Cells: Performance Dynamics, Architecture, and Trends," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1224–1240, 2015.

[56] M. Z. S. Muhammad Ali Imran, Syed Ali Raza Zaidi, Ed., *Access, Fronthaul and Backhaul Networks for 5G & Beyond*. Institution of Engineering and Technology (IET), 2017.

[57] F. Marzouk, R. Alheiro, J. Rodriguez, and A. Radwan, "Perspectives for 5G Network Sharing for Mobile Small Cells," in *Broadband Communications, Networks, and Systems*, V. Sucasas, G. Mantas, and S. Althunibat, Eds. Springer International Publishing, 2019, pp. 377–386.

[58] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-Dense Networks: A Survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 4, pp. 2522–2545, 2016.

[59] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. S. Dhillon, "An overview of load balancing in hetnets: old myths and open problems," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 18–25, 2014.

[60] H. Claussen, D. Lopez-Perez, L. Ho, R. Razavi, and S. Kucera, *Small Cell Networks: Deployment, Management, and Optimization*. Wiley-IEEE Press, January 2018.

[61] Small Cell Forum, "Small cell definition," https://www.smallcellforum.org/what-is-a-small-cell/, (Last accessed September 2020).

[62] Small Cell Forum, "5G small cell architecture and product definitions," www.scf.io, July 2020, (Last accessed September 2020).

[63] M. H. Cheung, F. Hou, J. Huang, and R. Southwell, "Congestion-Aware DNS for Integrated Cellular and Wi-Fi Networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1269–1281, 2017.

[64] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: How much can wifi deliver?" *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 536–550, 2013.

[65] C. Szymanski, Broadcom Inc., "20 years of wifi with Broadcom," https://www.wi-fi.org/, August 2019, (Last accessed April 2020).

[66] Wi-Fi Alliance, "Hotspot 2.0 (Version 3.1)," 2019.

[67] C. Rosa, M. Kuusela, F. Frederiksen, and K. I. Pedersen, "Standalone LTE in Unlicensed Spectrum: Radio Challenges, Solutions, and Performance of MulteFire," *IEEE Communications Magazine*, vol. 56, no. 10, pp. 170–177, October 2018.

[68] 3GPP TS 23.234 V6.10.0, "3GPP system to Wireless Local Area Network (WLAN) interworking; System description (Release 6)," September 2006.

[69] 3GPP TS 24.327 V8.6.0, "Mobility between 3GPP Wireless Local Area Network (WLAN) interworking (I-WLAN) and 3GPP systems; General Packet Radio System (GPRS) and 3GPP I-WLAN aspects; Stage 3 (Release 8)," June 2010.

[70] Qualcomm Technologies Inc., "LTE-U coexistence mechanism," May 2015.

[71] 3GPP TS 23.793 V16.0.0, "Study on Access Traffic Steering, Switching and Splitting support in the 5G system architecture (Release 16)," December 2018.

[72] 3GPP TS 24.312 V8.6.0, "Access Network Discovery and Selection Function (ANDSF) Management Object (MO) (Release 8)," June 2012.

[73] M. Rahnema and M. Dryjanski, *From LTE to LTE-Advanced Pro and 5G*. Artech House, 2017.

[74] 3GPP TR 36.213 v14.15.0, "Physical layer procedures (Release 14)," June 2020.

[75] 3GPP TSG RAN 75 Meeting, "RP-170848: New Work Item on enhancements to LTE operation in unlicensed spectrum (Release 15)," March 2017.

[76] Small Cell Forum, "Combining the benefits of licensed and unlicensed technologies (Release 7.0)," www.scf.io, June 2016, (Last accessed September 2020).

[77] N. Neokosmidis et al., "Are Small Cells and Network Intelligence at the Edge the Drivers for 5G Market Adoption? The SESAME Case," in *Engineering Applications of Neural Networks (EANN)*. Springer International Publishing, 2017, pp. 693–703.

[78] GSMA, "Spectrum Sharing: GSMA Public Policy Position," www.gsma.com, July 2019, (Last accessed April 2020).

[79] R. LIU, Q. CHEN, G. YU, G. Y. LI, and Z. DING, "Resource management in lte-u systems: Past, present, and future," *IEEE Open Journal of Vehicular Technology*, vol. 1, pp. 1–17, 2020.

[80] Telefonica and Ericsson, "Cloud RAN Architecture for 5G," http://www.tid.es/, June 2016, (Last accessed September 2020).

[81] G. Miao, J. Zander, K. W. Sung, and S. Ben Slimane, *Fundamentals of Mobile Data Networks*. Cambridge University Press, 2016.

[82] F. Grijpink, E. Kutcher, A. Ménard, S. Ramaswamy, D. Schiavotto, J. Manyika, M. Chui, R. Hamill, and E. Okan, "Connected world: An evolution in connectivity beyond the 5G revolution (Discussion paper)," February 2020.

[83] I. P. Chochliouros et al., "Inclusion of "Self-x" Properties in the SESAME-Based Wireless Backhaul for Support of Higher Performance," in *Engineering Applications of Neural Networks (EANN)*. Springer International Publishing, 2017, pp. 716–727.

[84] M. Agiwal, A. Roy, and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

[85] S. Fortes, R. Barco, and A. Aguilar-Garcia, "Location-based distributed sleeping cell detection and root cause analysis for 5G ultra-dense networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, no. 1, p. 149, June 2016.

[86] "Self-Organizing Networks in the 5G Era: 2019-2030 Opportunities, Challenges, Strategies & Forecasts," http://www.snstelecom.com/son, September 2018, (Last accessed September 2020).

[87] Telus and Huawei, "Next generation SON for 5G (White Paper)," 2016.

[88] NGMN Technical Working Group Self Organising Networks, "Next generation mobile networks use cases related to self organising network, overall description," 2008.

[89] 3GPP TS 32.500 v8.0.0, "Telecommunication management; Self-Organizing Networks (SON); Concepts and requirements," December 2008.

[90] 3GPP TS 38.300 V15.3.1, "NR; Overall description; Stage 2 (Release 15)," October 2018.

[91] T. Jyrki and J. Penttinen, "5G Explained: Security and Deployment of Advanced Mobile Communications," April 2019.

[92] Small Cell Forum, "SON API for small cells (Release 7.0)," www.scf.io, March 2015, (Last accessed September 2020).

[93] A. Tukmanov, M. A. Lema, I. Mings, M. Condoluci, T. Mahmoodi, Z. Al-Daher, and M. Dohler, *Fronthauling for 5G and beyond*. Institution of Engineering and Technology (IET), 2017.

[94] R. Li, Z. Zhao, X. Zhou, G. Ding, Y. Chen, Z. Wang, and H. Zhang, "Intelligent 5G: When Cellular Networks Meet Artificial Intelligence," *IEEE Wireless Communications*, vol. 24, no. 5, pp. 175–183, 2017.

[95] 3GPP TS 36.805 v9.0.0, "Study on minimization of drive tests in next generation networks," May 2010.

[96] J. Moysen and L. Giupponi, "From 4G to 5G: Self-organized network management meets machine learning," *Elsevier Computer Communications*, vol. 129, pp. 248 – 268, 2018.

[97] *Encyclopædia Britannica*. Encyclopædia Britannica, Inc., (Last accessed September 2020).

[98] Tom M. Mitchell, *Machine Learning*. New York: McGraw-Hill Science / Engineering / Math, March 1997.

[99] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*. Springer, 2016.

[100] T. G. Dietterich and P. Langley, *Cognitive Networks: Towards Self-Aware Networks*, Q. H. e. Mahmoud, Ed. John Wiley & Sons, Ltd, Chichester, UK, 2007.

[101] Christopher M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[102] D. Kibler and P. Langley, "Machine learning as an experimental science," Glasgow, UK, pp. 81–88, 1988.

[103] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, Massachusetts, London, England: The MIT Press, 2012.

[104] Richard S. Sutton and Andrew G. Barto, *Reinforcement Learning: An Introduction, Second Edition (in progress)*. Cambridge, Massachusetts, London, England: The MIT Press, 2014, 2015.

[105] Raj Jain, *The art of computer systems performance analysis - techniques for experimental design, measurement, simulation, and modeling*. Wiley, 1991.

[106] M. Chraiti, A. Ghrayeb, C. Assi, N. Bouguila, and R. A. Valenzuela, "A framework for unsupervised planning of cellular networks using statistical machine learning," *IEEE Transactions on Communications*, vol. 68, no. 5, pp. 3213–3228, 2020.

[107] J. Wang, C. Jiang, H. Zhang, Y. Ren, K. Chen, and L. Hanzo, "Thirty Years of Machine Learning: The Road to Pareto-Optimal Wireless Networks," *IEEE Communications Surveys Tutorials*, vol. 22, no. 3, pp. 1472–1514, 2020.

[108] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[109] J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[110] C. Olah, "The building blocks of interpretability," https://ai.googleblog.com, March 2018.

[111] C. Zhang, P. Patras, and H. Haddadi, "Deep Learning in Mobile and Wireless Networking: A Survey," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.

[112] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks," *IEEE Communications Surveys Tutorials*, vol. 19, no. 4, pp. 2392–2431, 2017.

[113] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "A Survey of Anticipatory Mobile Networking: Context-Based Classification, Prediction Methodologies, and Optimization Techniques," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1790–1821, 2017.

[114] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3072–3108, 2019.

[115] Small Cell Forum, "Precision Planning for 5G Era Networks with Small Cells," www.scf.io, Oct 2019, (Last accessed September 2020).

[116] M. Lin and Y. Zhao, "Artificial intelligence-empowered resource management for future wireless communications: A survey," *China Communications*, vol. 17, no. 3, pp. 58–77, 2020.

[117] ETSI, "Experiential Networked Intelligence (ENI)," https://www.etsi.org/technologies/experiential-networked-intelligence, (Last accessed September 2020).

[118] ETSI GR ENI 007 V1.1.1, "Experiential Networked Intelligence (ENI); ENI Definition of Categories for AI Application to Networks," November 2019.

[119] ETSI, "Zero touch network & Service Management (ZSM)," https://www.etsi.org/technologies/zero-touch-network-service-management, (Last accessed September 2020).

[120] ETSI GS ZSM 002 V1.1.1, "Zero-touch network and Service Management (ZSM); Reference Architecture," August 2019.

[121] 3GPP TS 29.520 version 15.3.0, "5G; 5G System; Network Data Analytics Services, Stage 3 (Release 15)," April 2019.

[122] International Telecommunication Union (ITU), Focus group on Machine Learning for Future Networks including 5G (FG-ML5G), "Unified architecture for machine learning in 5G and future networks (Technical Specification)," January 2019.

[123] Riley, George F. and Henderson, Thomas R., *The ns-3 Network Simulator*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 15–34.

[124] E. Weingaertner, H. Lehn, and K. Wehrle, "A performance comparison of recent network simulators," 07 2009, pp. 1 – 5.

[125] A. Virdis, G. Stea, and G. Nardini, "Simulte — a modular system-level simulator for lte/lte-a networks based on omnet++," in *Proceedings of the 4th International Conference on Simulation and Modeling Methodologies, Technologies and Applications*, ser. SIMULTECH 2014. Setubal, PRT: SCITEPRESS - Science and Technology Publications, Lda, 2014, p. 59–70. [Online]. Available: https://doi.org/10.5220/0005040000590070

[126] A. M. Mamadou, J. Toussaint, and G. Chalhoub, "Interference study of coexisting ieee 802.11 and 802.15.4 networks," in *2019 8th International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks (PEMWN)*, 2019, pp. 1–5.

[127] N. Baldo and M. Miozzo, "Spectrum-aware channel and phy layer modeling for ns3," in *Proceedings of the Fourth International ICST Conference on Performance Evaluation Methodologies and Tools*, ser. VALUETOOLS '09. Brussels, BEL: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009. [Online]. Available: https://doi.org/10.4108/ICST.VALUETOOLS2009.7647

[128] N. Baldo, M. Requena-Esteso, J. Núñez Martínez, M. Portolès-Comeras, J. Nin-Guerrero, P. Dini, and J. Mangues-Bafalluy, "Validation of the IEEE 802.11 MAC Model in the Ns3 Simulator Using the EXTREME Testbed," in *Proceedings of the 3rd International ICST Conference on Simulation Tools and Techniques (SIMUTools '10)*, 2010, pp. 64:1–64:9.

[129] S. Sakai, G. Gambugge, R. Takaki, J. Seki, J. Bazzo, and J. a. P. Miranda, "Performance comparison of a custom emulation-based test environment against a real-world lte testbed," in *Proceedings of the 2015 Workshop on Ns-3 (WNS3 '15)*, 2015, pp. 106–111.

[130] A. Marinescu, I. Macaluso, and L. A. DaSilva, "System level evaluation and validation of the ns-3 lte module in 3gpp reference scenarios," in *Proceedings of the 13th ACM Symposium on QoS and Security for Wireless and Mobile Networks (Q2SWinet '17)*, 2017, pp. 59–64.

[131] IEEE Std. 802.11-2007, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications," June 2007.

[132] Li-Hsing Yen and Tse-Tsung Yeh, "Snmp-based approach to load distribution in ieee 802.11 networks," in *2006 IEEE 63rd Vehicular Technology Conference*, vol. 3, May 2006, pp. 1196–1200.

[133] J. Chen, T. Chen, T. Zhang, and E. van den Berg, "Wlc19-4: Effective ap selection and load balancing in ieee 802.11 wireless lans," in *IEEE Global Telecommunications (IEEE Globecom) Conference*, December 2006, pp. 1–6.

[134] Jae-wook Jang and Yeon-sup Lim and Chong-kwon Kim, "Traffic-aware decentralized AP selection for multi-rate in WLANs," in *Proceedings of International Conference on Advanced Communication Technology (ICACT)*, February 2010.

[135] V. Ghini, S. Ferretti, and F. Panzieri, "A strategy for best access point selection," in *Proceedings of Wireless Days, 2010*, October 2010.

[136] Y. Fukuda, T. Abe, and Y. Oie, "Decentralized access point selection architecture for wireless LANs," in *Wireless Telecommunications Symposium (WTS)*, May 2004.

[137] C. Fortuna and M. Mohorcic, "Trends in the development of communication networks: Cognitive networks." *Computer Networks*, 2009.

[138] R. W. Thomas, D. H. Friend, L. A. DaSilva, and A. B. MacKenzie, "Cognitive Networks: Adaptation and Learning to Achieve End-to End Performance Objectives," *IEEE Communications Magazine*, vol. 44, no. 12, December 2006.

[139] H. Feng and Y. Shu, "Study on network traffic prediction techniques," in *Proceedings of Wireless Communications, Networking and Mobile Computing (WiCOM)*, September 2005.

[140] N. Baldo, M. Requena, J. Nunez, M. Portoles, J. Nin, P. Dini, and J. Mangues, "Validation of the ns-3 IEEE 802.11 model using the EXTREME testbed," in *Proceedings of SIMUTools Conference*, March 2010.

[141] E. Weingartner, H. vom Lehn, and K. Wehrle, "A Performance Comparison of Recent Network Simulators," in *Proceedings of IEEE International Conference on Communications*, June 2009.

[142] "Pcap Trace Parser," http://networks.cttc.es/mobile-networks/software-tools/pcap-trace-parser/, (Last accessed September 2020).

[143] G. Holland, N. Vaidya, and P. Bahl, "A rate-adaptive MAC protocol for multi-Hop wireless networks," in *Proceedings of Mobile Computing and Networking conference*, 2001.

[144] "Fast Artificial Neural Network Library (FANN)," http://leenissen.dk/fann/.

[145] S. Nissen, "Neural Networks made simple," http://fann.sf.net/fann_en.pdf, 2005.

[146] C. Igel and M. Hüsken, "Improving the RPROP learning algorithm," in *Proceedings of the ICSC Symposium on Neural Computation*, May 2000.

[147] B. Wilson, "The Machine learning dictionary," http://www.cse.unsw.edu.au/ billw/mldict.html.

[148] M. Portoles-Comeras, M. Requena-Esteso, J. Mangues-Bafalluy, and M. Cardenete-Suriol, "EXTREME: combining the ease of management of multi-user experimental facilities and the flexibility of proof of concept testbeds," in *Proceedings of Testbeds and Research Infrastructures for the Development of Networks and Communities (TridentCom)*, 2006.

[149] "The IPERF network testing tool for TCP and UDP data streams," http://iperf.sourceforge.net/.

[150] E. Yildirim, I. Suslu, and T. Kosar, "Which network measurement tool is right for you? a multidimensional comparison study," in *Proceedings of International Conference on Grid Computing (GRID)*, October 2008.

[151] M. H. Ahmed, "Call admission control in wireless networks: A comprehensive survey," *IEEE Communications Surveys Tutorials*, vol. 7, no. 1, pp. 49–68, 2005.

[152] S. Ghosh and A. Konar, *An Overview of Call Admission Control in Mobile Cellular Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 1–62.

[153] N. Baldo and M. Zorzi, "Learning and adaptation in cognitive radios using neural networks," in *IEEE Consumer Communications and Networking Conference (CCNC)*, 2008.

[154] B. Bojović, N. Baldo, J. Nin Guerrero and P. Dini, "A Supervised Learning Approach to Cognitive Access Point Selection," in *IEEE Global Telecommunications (IEEE Globecom) Conference Workshop*, Houston, Texas (US), December 2011.

[155] B. Bojović, N. Baldo, and P. Dini, "A Cognitive Scheme for Radio Admission Control in LTE systems," in *Cognitive Information Processing (CIP)*, May 2012.

[156] A. Katidiotis, K. Tsagkaris, and P. Demestichas, "Performance evaluation of artificial neural network-based learning schemes for cognitive radio systems," *Computers & Electrical Engineering*, vol. 36, no. 3, pp. 518–535, 2010.

[157] G. Quer, N. Baldo, and M. Zorzi, "Cognitive Call Admission Control for VoIP over IEEE 802.11 Using Bayesian Networks," in *IEEE Global Telecommunications (IEEE Globecom) Conference*, December 2011.

[158] M. Khabazian, O. Kubbar, and H. Hassanein, "Call Admission Control with Resource Reservation for Multi-service OFDM Networks," in *International Conference on Computing, Networking and Communications (ICNC)*, February 2012.

[159] D. K. Kim, D. Griffith, and N. Golmie, "A Novel Ring-Based Performance Analysis for Call Admission Control in Wireless Networks," vol. 14, no. 4, pp. 324 –326, April 2010.

[160] X. Zhu, X. Li, and Q. Zhang, "A Link Adaptation Based Call Admission Control Algorithm for OFDMA systems," in *International Conference on Advanced Communication Technology (ICACT)*, February 2012.

[161] Y. L. Lee, T. C. Chuah, J. Loo, and A. Vinel, "Recent advances in radio resource management for heterogeneous lte/lte-a networks," *IEEE Communications Surveys Tutorials*, vol. 16, no. 4, pp. 2142–2180, Fourthquarter 2014.

[162] T. Ali-Yahiya, *Understanding LTE and its Performance: Quality of Service*. Springer New York, 2011.

[163] ITU-T Recommendations G.107, *E-model: a computational model for use in transmission planning*, 2011.

[164] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution (From Theory to Practice)*. John Wiley & Sons Ltd, 2009.

[165] C. Igel and M. Hüsken, "Empirical evaluation of the improved Rprop learning algorithm," *Neurocomputing*, vol. 50, pp. 105–123, 2003.

[166] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. The MIT Press, 2009.

[167] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2003.

[168] F. V. Jensen and T. D. Nielsen, *Bayesian Networks and Decision Graphs*. Springer, 2007.

[169] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[170] G. Quer, F. Librino, L. Canzian, L. Badia, and M. Zorzi, "Using Game Theory and Bayesian Networks to Optimize Cooperation in Ad Hoc Wireless Networks," in *IEEE ICC*, June 2012.

[171] W. Navidi and T. Camp, "Stationary distributions for the random waypoint mobility model," vol. 3, pp. 99–108, 2003.

[172] *Digital Mobile Radio: COST 231 View on the Evolution Towards 3rd Generation Systems*. Commission of the European Communities, L-2920, Luxemburg, 1989.

[173] 3GPP TS 36.104 V12.13.0, "Base Station (BS) radio transmission and reception (Release 12)," March 2019.

[174] G. Mongha, K. Pedersen, I. Kovács, and P. Mogensen, "QoS Oriented Time and Frequency Domain Packet Schedulers for The UTRAN Long Term Evolution," in *Proceedings of IEEE Vehicular Technology Conference (VTC)*, 2008.

[175] M. Poikselkä, H. Holma, J. Hongisto, J. Kallio, and A. Toskala, *Voice over LTE (VoLTE)*. John Wiley & Sons, 2012.

[176] Global mobile Suppliers Association (GSA), "VoLTE Update March 2020 - Detailed Report," https://gsacom.com/paper/volte-update-march-2020-detailed-report/, 2020.

[177] 3GPP TS 36.300 V11.7.0, "Overall description; Stage 2 (Release 11)," December 2015.

[178] F. Capozzi, G. Piro, G. L.A, G. Boggia, and P. Camarda, "Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey," vol. 15, no. 2, pp. 678–700, 2013.

[179] I. Ahmed, L. Badia, N. Baldo, and M. Marco, "Design of a Unified Multimedia-Aware Framework for Resource Allocation in LTE Femtocells," in *Proceedings of International Symposium on Mobility Management and Wireless Access (MobiWac)*, 2011.

[180] F. Bokhari, W. Wong, and H. Yanikomeroglu, "Adaptive Token Bank Fair Queuing Scheduling in the Downlink of 4G Wireless Multicarrier Networks," in *Proceedings of IEEE Vehicular Technology Conference (VTC)*, 2008.

[181] ETSI TR 102 643 V1.0.2, "Human Factors (HF); Quality of Experience (QoE) requirements for real-time communication services," January 2010.

[182] K. C. Beh, S. Armour, and A. Doufexi, "Joint time-frequency domain proportional fair scheduler with harq for 3gpp lte systems," in *Proceedings of IEEE Vehicular Technology Conference (VTC)*, 2008.

[183] 3GPP TS 36.213 V11.13.0, "Physical layer procedures (Release 11)," September 2017.

[184] N. Baldo, M. Miozzo, M. Requena-Esteso, and J. Nin-Guerrero, "An open source product-oriented lte network simulator based on ns-3," in *Proceedings of the 14th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '11)*. ACM, 2011, pp. 293–298.

[185] D. Zhou, N. Baldo, and M. Miozzo, "Implementation and Validation of LTE Downlink Schedulers for ns-3," in *Proceedings of Workshop on ns-3 (WNS3)*, 2013.

[186] 3GPP TR 36.902, "Self-configuring and self-optimizing network (SON) use cases and solutions (Release 9)," 2011.

[187] C. W.-F. S. Provider, "A platform for business innovation and revenue generation," http://www.cisco.com, 2012, (Last accessed September 2020).

[188] J. Weitzen, L. Mingzhe, E. Anderland, and V. Eyuboglu, "Large-scale deployment of residential small cells," *Proceedings of the IEEE*, vol. 101, no. 11, pp. 2367–2380, November 2013.

[189] T. Zahir, K. Arshad, A. Nakata, and K. Moessner, "Interference management in femtocells," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 293–311, January 2013.

[190] Reverb, "Inteligent SON solutions," http://www.reverbnetworks.com/, 2015, (Last accessed September 2020).

[191] Stoke and Zhilabs, "Analytics in Secured LTE," http://www.zhilabs.com/, 2014, (Last accessed September 2020).

[192] Samsung, "Smart LTE for Future Innovation," http://www.samsung.com/, 2014, (Last accessed September 2020).

[193] N. Baldo, L. Giupponi, and J. Mangues-Bafalluy, "Big data empowered self organized networks," in *European Wireless 2014; 20th European Wireless Conference*, May 2014, pp. 1–8.

[194] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*. Academic Press (Elsevier), 2013.

[195] W. K. Hale, "Frequency assignment: Theory and applications," *Proceedings of the IEEE*, vol. 68, no. 12, pp. 1497–1514, Dec 1980.

[196] H. Zhuang, D. Shmelkin, Z. Luo, M. Pikhletsky, and F. Khafizov, "Dynamic spectrum management for intercell interference coordination in lte networks based on traffic patterns," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 1924–1934, Jun 2013.

[197] Z. Lu, T. Bansal, and P. Sinha, "Achieving user-level fairness in open-access femtocell-based architecture," *IEEE Transactions on Mobile Computing*, vol. 12, no. 10, pp. 1943–1954, Oct 2013.

[198] L. Tan, Z. Feng, W. Li, Z. Jing, and T. A. Gulliver, "Graph coloring based spectrum allocation for femtocell downlink interference mitigation," in *2011 IEEE Wireless Communications and Networking Conference*, March 2011, pp. 1248–1252.

[199] S. Sadr and R. Adve, "Hierarchical resource allocation in femtocell networks using graph algorithms," in *2012 IEEE International Conference on Communications (ICC)*, June 2012, pp. 4416–4420.

[200] S. Uygungelen, G. Auer, and Z. Bharucha, "Graph-based dynamic frequency reuse in femtocell networks," in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, May 2011, pp. 1–6.

[201] F. Bernardo, R. Agustí, J. Pérez-Romero, and O. Sallent, "Intercell interference management in ofdma networks: A decentralized approach based on reinforcement learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 6, pp. 968–976, November 2011.

[202] 3GPP TS 36.814, "Further advancements for E-UTRA physical layer aspects (Release 9)," 2010.

[203] 3GPP TS 32.827, "Integration of device management information with Itf-N (Release 10)," 2010.

[204] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1136–1159, March 2013.

[205] M. Mezzavilla, M. Miozzo, M. Rossi, N. Baldo, and M. Zorzi, "A Lightweight and Accurate Link Abstraction Model for the Simulation of LTE Networks in Ns-3," in *Proceedings of the 15th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM '12)*. New York, NY, USA: ACM, 2012, pp. 55–60.

[206] 3GPP TS 36.213, "Physical layer procedures (Release 12)," 2014.

[207] T. Hastie; R. Tibshirani and J. J. H. Friedman, *The elements of statistical learning*. Springer New York, 2001.

[208] M. Kuhn, "Building predictive models in r using the caret package," *Journal of Statistical Software, Articles*, vol. 28, no. 5, pp. 1–26, 2008.

[209] 3GPP TS 36.828, "Further enhancements to LTE Time Division Duplex (TDD) for Downlink-Uplink (DL-UL) interference management and traffic adaptation (Release 11)," 2012.

[210] R. Wireless, "What's the future hold for unlicensed use of the 6 GHz band?" rcrwireless.com, (Last accessed September 2020).

[211] S.Methley, W.Webb, "Wi-Fi Spectrum Needs Study, Final Report to Wi-Fi Alliance (3rd edition)," http://www.wi-fi.org, February 2017.

[212] MFA TS 36.300 V1.0.1, "Overall description; Stage 2 (Release 1)," May 2017.

[213] 3GPP TS 36.300 V13.2.0, "Overall description; Stage 2 (Release 13)," January 2016.

[214] R. Zhang, M. Wang, L. Cai, Z. Zheng, X. Shen, and L. Xie, "LTE-unlicensed: the future of spectrum aggregation for cellular networks," *IEEE Wireless Communication Magazine*, vol. 22, no. 3, pp. 150–159, Jun 2015.

[215] Wi-Fi Alliance, "Coexistence Guidelines for LTE in Unlicensed Spectrum Studies, V2.0," February 2016.

[216] L. Falconetti *et al.*, "Design and Evaluation of Licensed Assisted Access LTE in Unlicensed Spectrum," *IEEE Wireless Communications*, pp. 24–30, Dec 2016.

[217] J. Jeon, H. Niu, Q. C. Li, A. Papathanassiou, and G. Wu, "LTE in the unlicensed spectrum: Evaluating coexistence mechanisms," in *IEEE Global Telecommunications (IEEE Globecom) Conference Workshop*, Dec 2014.

[218] A. Mukherjee *et al.*, "System Architecture and Coexistence Evaluation of Licensed-Assisted Access LTE with IEEE 802.11," in *IEEE International Conference on Communications (ICC) Workshop on LTE-U*, London, UK, Jun 2015.

[219] A. M. Voicu, L. Simić, and M. Petrova, "Inter-Technology Coexistence in a Spectrum Commons: A Case Study of Wi-Fi and LTE in the 5-GHz Unlicensed Band," vol. 34, no. 11, pp. 3062–3077, Nov 2016.

[220] C. Cano and D. J. Leith, "Unlicensed LTE/WiFi coexistence: Is LBT inherently fairer than CSAT?" in *IEEE International Conference on Communications (ICC)*, Kuala Lumpur, 2016, pp. 1–6.

[221] Y. Li, F. Baccelli, J. G. Andrews, T. D. Novlan, and J. C. Zhang, "Modeling and Analyzing the Coexistence of Wi-Fi and LTE in Unlicensed Spectrum," vol. 15, no. 9, pp. 6310–6326, Sept 2016.

[222] S. Zinno, G. D. Stasi, S. Avallone, and G. Ventre, "On a fair coexistence of LTE and Wi-Fi in the unlicensed spectrum: A Survey," *Computer Communications*, vol. 115, pp. 35 – 50, 2018.

[223] B. Chen, J. Chen, Y. Gao, and J. Zhang, "Coexistence of lte-laa and wi-fi on 5 ghz with corresponding deployment scenarios: A survey," *IEEE Communications Surveys Tutorials*, vol. 19, no. 1, pp. 7–32, 2017.

[224] S. K. Sharma, T. E. Bogale, L. B. Le, S. Chatzinotas, X. Wang, and B. Ottersten, "Dynamic spectrum sharing in 5g wireless networks with full-duplex technology: Recent advances and research challenges," *IEEE Communications Surveys Tutorials*, vol. 20, no. 1, pp. 674–707, 2018.

[225] Y. Jian, C.-F. Shih, B. Krishnaswamy, and R. Sivakumar, "Coexistence of Wi-Fi and LAA-LTE: Experimental evaluation, analysis and insights," in *IEEE International Conference on Communications (ICC) Workshop*, London, UK, Jun 2015.

[226] S. Sagari, S. Baysting, D. Saha, I. Seskar, W. Trappe, and D. Raychaudhuri, "Coordinated dynamic spectrum management of LTE-U and Wi-Fi networks," in *IEEE DySPAN*, Stockholm, SE, Sept 2015, pp. 209–220.

[227] C. Chen, R. Ratasuk and A. Ghosh, "Downlink Performance Analysis of LTE and WiFi Coexistence in Unlicensed Bands with a Simple Listen-Before-Talk Scheme," in *IEEE 81st Vehicular Technology Conference (VTC Spring)*, Glasgow, 2015.

[228] X. Wang, T. Q. S. Quek, M. Sheng, and J. Li, "Throughput and Fairness Analysis of Wi-Fi and LTE-U in Unlicensed Band," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 1, pp. 63–78, Jan 2017.

[229] Z. Ali, B. Bojović, L. Giupponi, and J. Mangues, "On fairness evaluation: LTE-U vs. LAA," in *ACM MobiWac*, Malta, Nov. 2016.

[230] Qualcomm Technologies Inc., "LTE-U Technology and Coexistence," San Diego, USA, May 2015.

[231] LTE-U Forum, "LTE-U CSAT Procedure TS, V1.3," October 2015.

[232] 3GPP TS 36.211, "Physical channels and modulation (Release 13)," January 2016.

[233] MediaTek Inc., *eNB and UE complexities with respect to partial subframes*, 3GPP TSG RAN WG1 82 Meeting , Beijing, China, August 2015.

[234] Samsung, *Partial subframe for LAA*, 3GPP TSG RAN WG1 82 meeting, Beijing, China, August 2015.

[235] LTE-U Forum, "LTE-U SDL Coexistence Specifications, V1.3," October 2018.

[236] Z. Ali, L. Giupponi, J. Mangues, and B. Bojović, "Machine Learning Based Scheme for Contention Window Size Adaptation in LTE-LAA," in *Proceedings of 28th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Montreal, Canada, Oct. 2017.

[237] 3GPP TSG RAN 70 meeting, "RP-151977: Status Report of WI Licensed-Assisted Access using LTE," December 2015.

[238] Intel and others, *WF on Contention Window adaptation based on HARQ ACK/NACK feedback*, 3GPP TSG RAN WG1, October 2015.

[239] M. Casoni and N. Patriciello, "Next-generation TCP for ns-3 simulator," in *Simulation Modelling Practice and Theory*, vol. 66, 2016, pp. 81–93.

[240] M. Mehrnoush, V. Sathya, S. Roy, and M. Ghosh, "Analytical Modeling of Wi-Fi and LTE-LAA Coexistence: Throughput and Impact of Energy Detection Threshold," vol. 26, no. 4, Aug. 2018, pp. 1990–2003.

[241] N. Jindal and G. I. Breslin, Donald, "LTE and Wi-Fi in Unlicensed Spectrum: A Coexistence Study," 2015.

[242] J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z. Mao, S. Sen, and O. Spatscheck, "An In-depth Study of LTE: Effect of Network Protocol and Application Behavior on Performance," in *ACM SIGCOMM*, August, 2013, pp. 363–374.

[243] H. Kwon, J. Jeon, A. Bhorkar, Q. Ye, H. Harada, Y. Jiang, L. Liu, S. Nagata, B. L. Ng, T. Novlan, J. Oh, and W. Yi, "Licensed-assisted access to unlicensed spectrum in lte release 13," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 201–207, February 2017.

[244] Z. Pi and F. Khan, "c," *IEEE Commun. Magazine*, vol. 49, no. 6, pp. 101–107, Jun. 2011.

[245] T. Nitsche, C. Cordeiro, A. B. Flores, E. W. Knightly, E. Perahia, and J. C. Widmer, "IEEE 802.11ad: directional 60 GHz communication for multi-Gigabit-per-second Wi-Fi [Invited Paper]," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 132–141, December 2014.

[246] Y. Ghasempour, C. R. C. M. da Silva, C. Cordeiro, and E. W. Knightly, "IEEE 802.11ay: Next-Generation 60 GHz Communication for 100 Gb/s Wi-Fi," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 186–192, December 2017.

[247] 3GPP TSG-RAN 86 Meeting, "RP-193259: Study on supporting NR from 52.6 GHz to 71 GHz," December 2019.

[248] 3GPP TSG-RAN 86 Meeting, "RP-193229: Extending current NR operation to 71GHz," December 2019.

[249] H. Assasa and J. Widmer, "Implementation and Evaluation of a WLAN IEEE 802.11ad Model in ns-3," *ns-3 Workshop 2016*, 2016.

[250] H. Assasa and J. Widmer, "Extending the IEEE 802.11ad Model: Scheduled Access, Spatial Reuse, Clustering, and Relaying," *ns-3 Workshop 2017*, 2017.

[251] 3GPP TR 21.916 V1.0.0, "Release 16 Description; Summary of Rel-16 Work Items (Release 16)," December 2020.

[252] P. Zhou, K. Cheng, X. Han, X. Fang, Y. Fang, R. He, Y. Long, and Y. Liu, "IEEE 802.11ay-Based mmWave WLANs: Design Challenges and Solutions," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 1654–1681, thirdquarter 2018.

[253] IEEE 802.11ad Specific requirements, "Part 11: Wireless lan medium access control," 2012.

[254] Intel Corporation, *Channel access mechanism for NR-unlicensed*, 3GPP R1-1902471, TSG RAN WG1 96 Meeting, February 2019.

[255] Huawei, HiSilicon, *Coexistence and channel access for NR unlicensed band operations*, 3GPP R1-1901525, TSG RAN WG1 96 Meeting, February 2019.

[256] Nokia, Nokia Shanghai Bell, *Channel access and co-existence for NR-U operation*, 3GPP R1-1902109, TSG RAN WG1 96 Meeting, February 2019.

[257] Samsung, *Evaluation Results for Directional LBT*, 3GPP R1-1902264, TSG RAN WG1 96 Meeting, February 2019.

[258] 3GPP TSG-RAN 84 Meeting, "RP-191551: Preparing for Rel-17," June 2019.

[259] S. Lagén, L. Giupponi, N. Patriciello, "LBT Switching Procedures for New Radio-based Access to Unlicensed Spectrum," in *IEEE Global Communications Conference, ET5GB WS.*, Dec. 2018.

[260] 3GPP TR 38.802 v14.2.0, "Study on New Radio (NR) Access Technology; Physical Layer Aspects (Release 14)," September 2017.

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

[262] S. Lagén, K. Wanuga, H. Elkotby, S. Goyal, N. Patriciello, and L. Giupponi, "New Radio Physical Layer Abstraction for System-Level Simulations of 5G Networks," in *2020 IEEE International Conference on Communications*, June 2020.

[263] 3GPP TS 38.214 v15.3.0, "NR; Physical layer procedures for data (Release 15)," September 2018.

[264] 3GPP TS 38.212 v15.3.0, "NR; Multiplexing and channel coding (Release 15)," September 2018.

[265] 3GPP TSG RAN 75 Meeting, "RP-170828: Study on NR-based Access to Unlicensed Spectrum (Release 15)," March 2017.

[266] 3GPP TR 38.900 V14.3.1, "NR; Study on channel model for frequency spectrum above 6 GHz (Release 14)," July 2017.

[267] S. Lagén and L. Giupponi, "Listen before receive for coexistence in unlicensed mmWave bands," in *2018 IEEE Wireless Communications and Networking Conference*, April 2018, pp. 1–6.