Radio Resource Management Techniques for Ultra-Reliable Low-Latency Communications in Unlicensed Spectrum

Cuevas, Roberto Maldonado

DOI (link to publication from Publisher):
10.5278/vbn.phd.tech.00060

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Cuevas, R. M. (2020). *Radio Resource Management Techniques for Ultra-Reliable Low-Latency Communications in Unlicensed Spectrum*. Aalborg Universitetsforlag.
https://doi.org/10.5278/vbn.phd.tech.00060

RADIO RESOURCE MANAGEMENT TECHNIQUES FOR ULTRA-RELIABLE LOW-LATENCY COMMUNICATIONS IN UNLICENSED SPECTRUM

BY
ROBERTO MALDONADO CUEVAS

DISSERTATION SUBMITTED 2020

AALBORG UNIVERSITY
DENMARK

Radio Resource Management Techniques for Ultra-Reliable Low-Latency Communications in Unlicensed Spectrum

Ph.D. Dissertation

Roberto Maldonado Cuevas

Aalborg University
Department of Electronic Systems
Fredrik Bajers Vej 7
DK - 9220 Aalborg

Dissertation submitted: October 2020

PhD supervisor: Professor Preben Mogensen
Aalborg University

Assistant PhD supervisors: Professor Klaus I. Pedersen
Aalborg University

Senior Research Engineer Claudio Rosa
Nokia Bell Labs

PhD committee: Associate Professor Jimmy Jessen Nielsen (chairman)
Aalborg University

Dr Sorour Falahati
Ericsson

Professor Raquel Barco
University of Malaga

PhD Series: Technical Faculty of IT and Design, Aalborg University

Department: Department of Electronic Systems

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-833-9

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2020

Curriculum Vitae

Roberto Maldonado Cuevas

Roberto Maldonado Cuevas received his B.Sc. and M.Sc. degrees in telecommunication engineering from Granada University (Spain) in 2014 and 2016, respectively. Since 2017, he pursues his PhD degree at the Electronic Systems Department from Aalborg University (Denmark) in collaboration with Nokia Bell Labs. His research interests include unlicensed spectrum, low-latency high-reliable communications and 5G radio resource management features.

Abstract

The next industrial revolution (commonly referred to as Industry 4.0) aims at improving the efficiency, flexibility and versatility of the factories. It foresees a fully digital industrial paradigm in which all the players along the value chain are interconnected. In this context, a wireless communication technology capable of supporting delay-critical transmissions at very high reliability levels plays a crucial role. The fifth generation (5G) of cellular technologies emerges as a candidate as it is designed to support ultra-reliable and low-latency communications (URLLC). Unlicensed spectrum is seen as an attractive option for industrial verticals due to its easy access and global availability, as well as its simplicity and low cost of deployment as compared to licensed spectrum. However, the strict spectrum regulations impose additional challenges in fulfilling the URLLC requirements in unlicensed bands. The PhD analyzes the capabilities of unlicensed radio access technologies (RATs) in supporting stringent latency and reliability requirements. Radio resource management (RRM) techniques and mandatory channel access procedures are investigated and optimized given the observed limitations. The PhD project answers research questions such as: What is the impact of channel access mechanisms on the overall system performance? Which latency-reliability requirements can be fulfilled by cellular RATs operating in the unlicensed spectrum? Which RRM and channel access enhancements are promising for supporting Industry 4.0 use-cases?

First, the impact of the channel access procedures on the latency and reliability is assessed in a realistic multi-user, multi-cell setting. Initially, MulteFire is adopted as the reference model. The study concludes that the channel access mechanisms based on listen-before-talk (LBT) are the main bottleneck for supporting delay-sensitive applications. At the base stations (BSs) side, the channel access account for 18 % to 42 % of the overall packet delay, depending on the network load and reliability level. At user equipments (UEs) side, it is shown that the channel access also plays an important role. With the insights acquired in the study, promising RRM techniques aiming at reducing the channel access impact are proposed and evaluated. Specifically, multiple hybrid automatic repeat request (HARQ) opportunities for uplink control transmissions, grant-free for uplink data transmissions, and the avoidance of the uplink LBT (under specific conditions defined by spectrum regulations).

The second part of the PhD focuses on 5G operation in the unlicensed spectrum, also known as New Radio Unlicensed (NR-U). Key technology components, initially design for 5G licensed operation, are analyzed for unlicensed operation. Selecting higher subcarrier spacing, faster processing times and shorter transmission time interval (TTI) are shown to provide additional latency-reliability benefits as compared to the reported in licensed deployments. For instance, reducing the TTI leads to faster and more reliable LBT as the channel is occupied for shorter durations. Towards further improving the NR-U latency-reliability performance, novel unlicensed-specific enhancements are developed. A time-diversity technique that mitigates the impact of the channel access in uplink data transmissions is proposed. UEs following grant-based uplink are provided with multiple time resources for uplink transmissions. The target is to reduce the need for several grants for a single uplink (UL) transmission. The proposal is shown to be especially useful in scenarios with high LBT failure probability. Additionally, two enhancements for the channel access are presented. Both are designed based on coordination among the nodes. The first optimization introduces the usage of a periodic structure that dictates the channel access and channel occupancy instances. This ensures a bounded delay at the BS channel access in single network deployments. On top of that, BSs are also coordinated in the frame selection. This approach achieves latencies below 10 ms for any of the considered network loads in the study. As an alternative, an approach based on silent gap coordination is introduced. Nodes following the silencing pattern defer their transmissions for a certain duration to favour the LBT outcome of neighbour nodes. Both enhancements fully remove the uncertainty added by LBT on the uplink transmissions. A 16-fold reduction on the packet delay is achieved by switching from MulteFire (77.3 ms) to a delay-optimized NR-U (4.75 ms), for a network load of 2.5 Mbit/s and 99.99 % reliability.

The third part of the study evaluates the NR-U performance in scenarios with additional RATs coexisting in the 5 GHz band. The presence of unplanned inter-RAT interference harms the NR-U performance, reducing the channel availability and increasing by a factor of 3 the experienced latency. To cope with this, multi-channel techniques such as carrier aggregation, wideband operation and packet duplication are proposed. The study shows that mechanisms that allow for a dynamic adjustment of the transmissions to subchannels with low or absence of interference are essential for fulfilling certain URLLC requirements.

Resumé

Den næste industrielle revolution, generelt refereret til som Industri 4.0, har til formål at forbedre fabrikkernes effektivitet, fleksibilitet og alsidighed. Et fuldt digitaliseret industrielt miljø hvor mennesker, maskiner og varer kommunikerer på tværs af værdikæden forventes. I denne kontekst, spiller trådløs kommunikations teknologi, som er i stand til at understøtte transmissioner der er yderst følsomme overfor forsinkelser, med højt niveau af pålidelighed, en afgørende rolle. Den 5. generation af cellulære teknologier er en kandidat til dette, da den er designet til netop at understøtte ultra-pålidelig, lav latens kommunikation (URLLC). Det licensfrie spektrum er en attraktiv mulighed for industriens vertikale aktører grundet nem adgang og global tilgængelighed, samt fraværet af omkostninger ifbm. spektrum licensering. Men, reglerne for brug af licensfrie bånd er forbundet med udfordringer, specielt når det vedrører opfyldelse af URLLC krav. PhD projektet analyserer licensfrie radioadgangsteknologiers (RATs) evne til at supportere URLLC relaterede krav. Teknikker til styring af radioressourcer (RRM) og spektumadgangsprocedurer undersøges og optimeres. Der besvares forskningsspørgsmål såsom: ″Hvordan påvirker spektrum-adgangsprocedurer det samlede systems evne til at understøtte URLLC?″, ″Hvilke latenstid-pålidelighed krav kan opfyldes af cellulære RATs, der opererer i det licensfrie spektrum?″, ″Hvilke RRM forbedringer har størst potentiale?″.

Første del af PhD projektet omhandler spektrum-adgangsprocedurers indflydelse på latenstid og pålidelighed i et realistisk multibruger, multicelle miljø. Først anvendes MulteFire som referencemodel. Studiet konkluderer at mekanismer baseret på listen-before-talk (LBT) er den største flaskehals i understøttelsen af latenstid følsom kommunikation. Afhængig af netværksbelastning og det krævede pålidelighedsniveau, så bruges 18% til 42% af den samlede latens budget på LBT. Lovende RRM teknikker, der mindsker påvirkning fra LBT er derfor blevet udviklet. Dette er specifikke løsninger, bl.a. multiple hybrid automatic repeat request (HARQ), transmissionsmuligheder, grant-free uplink data transmissioner, og undgåelse af uplink LBT under visse betingelser. Anden del af studiet fokuserer på 5G New Radio Unlicensed (NR-U). Dette inkluderer større subcarrier afstand, hurtigere processeringstider og kortere transmission time interval (TTI), som giver yderligere fordele for latenstid og pålidelighed, sammenlignet med de rapporterede ved licenseret anvendelser. Eksempelvis vil en reduktion af TTI føre til hurtigere og mere pålidelig LBT da kanalen er optaget i en kortere tidsperiode. Nye NR-U forbedringer foreslås: En tidsspredning teknik, der afhjælper påvirkningen fra kanaladgang i uplink data transmissioner. Yderligere præsenteres to kanalindgangforbedringer baseret af koordination. Den første optimering introducerer brugen af en periodisk struktur som foreskriver kanaladgang og kanalbrug. Dette sikrer en begrænset forsinkelse ved base stations (BSs) kanaladgang. Denne tilgang opnår latenstider på mindre end 10ms. Alternativt foreslås en tilgang, der koordinerer de tidsrum, hvor der ingen transmission er mellem netværkselementer. Begge forbedringer eliminerer til fulde den usikkerhed som LBT tilføjer til uplink transmissioner. En faktor 16 reduktion af pakkeforsinkelser opnås ved at skifte fra MulteFire (77.3 ms) til en forsinkelses optimeret NR-U (4.75 ms) ved netværksbelastning på 2.5 Mbit/s, indenfor den 99.99%-pålidelighed. Tredje del af studiet evaluerer NR-U under påvirkning af interferensen fra andre RATs. Inter-RAT interferens skader NR-Us ydeevnen, i kraft af mindsket kanalledighed, hvilket resulterer i en faktor 3 øget latenstid. For at overkomme dette, foreslås multikanal teknikker. Det eftervises, at mekanismer, som tillader dynamisk justering af transmissioner til subkanaler med lav eller ikke eksisterende interferens, er afgørende for opfyldelse af visse URLLC krav.

Contents

List of Abbreviations

1G first generation

2G second generation

3D three dimensional

3G third generation

3GPP 3rd Generation Partnership Project

4G fourth generation

5G fifth generation

ACK positive acknowledgement

BS base station

COT channel occupancy time

CQI channel quality indicator

DL downlink

ED energy detection

eLAA enhanced-LAA

eMBB enhanced mobile broadband

ETSI European Telecommunications Standards Institute

FBE frame-based equipment

FDD frequency division duplex

GB grant-based

GF grant-free

gNB next generation Node-B

HARQ hybrid automatic repeat request

IIoT Industrial Internet of Things

IMT-2020 International Mobile Telecommunications for 2020 and beyond

IoT Internet of Things

ISM industrial, scientific and medical

ITU International Telecommunications Union

KPI key performance indicator

LA link adaptation

LAA Licensed-Assisted Access

LBE load-based equipment

LBT listen-before-talk

LOS line-of-sight

LPWA Low Power Wide Area

LSB LBT synchronization boundary

LTE Long Term Evolution

LTE-A LTE-Advanced

LTE-U LTE-Unlicensed

LWA LTE-WLAN Aggregation

MAC medium-access-control layer

MBB mobile broadband

MCOT maximum channel occupancy time

MCS modulation and coding scheme

MFA MulteFire Alliance

mMTC massive machine type communications

mmWave millimeter waves

NACK negative acknowledgement

NR	New Radio
NR-U	New Radio Unlicensed
OFDM	orthogonal frequency-division multiplexing
PC	power control
PDCP	packet data convergence protocol
PHY	physical layer
PUSCH	physical uplink shared channel
QoS	quality-of-service
RAT	radio access technology
RRM	radio resource management
RTT	round-trip time
SCS	sub-carrier spacing
SR	scheduling request
TDD	time division duplex
TTI	transmission time interval
UE	user equipment
UL	uplink
URLLC	ultra-reliable and low-latency communications
URLLC-u	ultra-reliable low-latency communications over the unlicensed spectrum
V2X	vehicular to anything
WLAN	wireless local area network

Thesis Details

Thesis Title:	Radio Resource Management Techniques for Ultra-Reliable Low-Latency Communications in Unlicensed Spectrum
PhD Student:	Roberto Maldonado Cuevas
Supervisors:	Prof. Preben Mogensen. Aalborg University Prof. Klaus I. Pedersen. Aalborg University Claudio Rosa. Nokia Bell Labs, Aalborg

This PhD thesis is the result of three years of research at the Wireless Communication Networks (WCN) section (Department of Electronic Systems, Aalborg University, Denmark) in collaboration with Nokia Bell Labs (Aalborg). The work was carried out in parallel with mandatory courses required to obtain the PhD degree.

The main body of the thesis consists of the following articles:

Paper A: R. Maldonado, C. Rosa, F. Frederiksen and K. I. Pedersen, "On the Impact of Listen-Before-Talk on Ultra-Reliable Low-Latency Communications", *IEEE Global Communications Conference (GLOBECOM)*, December 2018, pp 1-6.

Paper B: R. Maldonado, C. Rosa, F. Frederiksen and K. I. Pedersen, "Uplink Ultra-Reliable Low Latency Communications Assessment in Unlicensed Spectrum", *IEEE Globecom Workshops (GC Wkshps)*, December 2018, pp 1-6.

Paper C: R. Maldonado, C. Rosa and K. I. Pedersen, "Latency and Reliability Analysis of Cellular Networks in Unlicensed Spectrum", *IEEE Access*, 2016, pp 49412-49423.

Paper D: R. Maldonado, C. Rosa and K. I. Pedersen, "Analysis of High-Reliable and Low-Latency Communication Enablers for New Radio Unlicensed", *IEEE Wireless Communications and Networking Conference (WCNC)*, April 2020, pp 1-6.

Paper E: R. Maldonado, C. Rosa and K. I. Pedersen, "A Fully Coordinated New Radio-Unlicensed System for Ultra-Reliable Low-Latency Applications", *IEEE Wireless Communications and Networking Conference (WCNC)*, April 2020, pp 1-6.

Paper F: R. Maldonado, C. Rosa and K. I. Pedersen, "Multi-link Techniques for New Radio-Unlicensed URLLC in Hostile Environments", *IEEE Vehicular Technology Conference (VTC), April 2021*, **Submitted for publication**

Paper G: R. Maldonado, C. Rosa and K. I. Pedersen, "A silent gap coordination for supporting URLLC in unlicensed deployments", **Work not submitted for publication**

This thesis has been submitted for assessment in partial fulfilment of the PhD Degree. The thesis is based on the submitted or published papers that are listed above. Parts of the papers are used directly or indirectly in the extended summary of the thesis. As part of the assessment, co-author statements have been made available to the assessment committee and also available at the Faculty.

Acknowledgements

This work has been a life-changing journey, personally and professionally. None of this would be possible without Klaus Pedersen and Preben Mogensen. Thanks for offering me the opportunity of pursuing a PhD. I am honoured to be part of Aalborg University and to collaborate with Nokia Bell Labs.

My sincere gratitude to my supervisors for their commitment, guidance, and advice. Especially to Claudio Rosa who day-by-day shared his knowledge and contributed to my development as a researcher. Thanks as well to Timo Nihtilä whose technical support was key for the project. Thanks to my colleagues from Aalborg University and Nokia Bell Labs who also collaborated in this project.

To Melisa, María, Dani, Enric, Pilar and Nacho. Thanks for the help and the unforgettable moments lived together, I feel you all as my second family. My gratitude as well to Majken, thanks for your patience and for helping me to always look at the bright side of things. Also thanks for the good times to my international friends. And to all my friends in Spain, thanks for the support from the distance.

From the bottom of my heart, thanks to my parents, Rosa and Juan, and to my brother, David. I love you all and I am blessed to have you as my family. Thank you for your endless support and your unconditional love. I also want to remember my grandmother, Adoración. Thanks for protecting me wherever you are.

Roberto Maldonado Cuevas
Aalborg University, October 2020

Part I

Introduction

Introduction

Nowadays, wireless communications have become an essential part of our lives, playing an important role in the society and the economy. Ubiquitous connectivity is paving our way towards a fully interconnected world where people, machines and objects collaborate and mutually exchange information. Over the last decades, cellular technologies have experienced a rapid evolution to accommodate the needs of the society [1]. It all started back in 1980 with the first generation (1G) of mobile communications. It employed analogue radio transmissions schemes and circuit-switched networks to provide voice services. In the 1990s, a transition from analogue to digital enabled the second generation (2G). As in 1G's design, 2G supported voice-centric services. During the deployment of 2G, networks were further improved to also provide data services at modest rates, in the order of hundreds of kbit/s (200 kbit/s in downlink (DL) and 90 kbit/s in uplink (UL)). Additionally, it enabled new services such as short message service (SMS). An increase in the demand for higher data rates led to a new cellular generation. The third generation (3G) went beyond the voice-centric perspective of previous generations and it is considered the foundation of mobile broadband services. To ensure global applicability of the 3G specifications, a consortium of several telecommunication standard organizations jointly created the 3$^{\text{rd}}$ Generation Partnership Project (3GPP). 3GPP standards are structured in Releases (Rel), with Rel-99 as the first 3G standard available in 1999. Enhancement to push the 3G potential were conducted in future releases, specifically, until Rel-10 in 2011. As compared to 2G, 3G increased the data transfer capabilities: early specifications supports theoretical DL bit rates in the order of 15 Mbit/s whereas evolved versions of 3G reaches values above 150 Mbit/s using a combination of multi-carrier and multi-antenna techniques [2]. It supported circuit-switched networks, for voice services, and packet-switched networks, for data services. Relying on its advanced capabilities, 3G enabled real-time services such as video conferencing or mobile TV. The society and market trends demanded higher capacity for mobile broadband applications. To fulfil the requirements, the fourth generation (4G) was firstly standardized in Rel-8 (2008), under the name of Long Term Evolution (LTE), and it was commercially available in 2009. It came with a transition from the existing circuit + packet-switched networks to a fully packet-switched approach. It provided superior performance in terms of efficiency and achievable data rates, reaching DL peak data rates of 300 Mbit/s and 75 Mbit/s in DL and UL, respectively. The evolution of the standard continued in 3GPP and a major enhancement to LTE led to the standardization of LTE-Advanced (LTE-A), specified in Rel-10 (2011). LTE-A reaches theoretical data rates of 3 Gbit/s in DL and 1.5 Gbit/s in UL. LTE brought mobile Internet access to hundreds of millions of people [3]. Moreover, it successfully enabled new applications that go beyond the mobile broadband services, expanding to use cases with a large number of low-cost devices with low battery consumptions [4]. Driven by the ambitious communications requirements specified by the International Mobile Telecommunications for 2020 and beyond (IMT-2020), 3GPP initiated the design of a new cellular generation. Apart from the natural evolution in the mobile broadband requirements, novel use cases with diverse and unprecedented requirements, which are not possible to meet with LTE, are envisioned. The fifth generation (5G) shall provide connectivity to all kind of services anywhere, anytime, to anyone and anything.

The focus of the thesis is on the support of ultra-reliable and low-latency communications (URLLC) over unlicensed spectrum. This type of communications requires 5G networks to support transmissions where packets are delivered in a very short time interval (in the order of milliseconds) and with a very high probability of success (above 99.999 %). The achievement of these requirements will enable the applicability of 5G to verticals such as industry, health and safety. Meeting the requirements represents a challenge that brings additional difficulties when transmissions are carried over the unlicensed spectrum. Unlicensed bands are shared among multiple wireless technologies, which imply that a steady quality of service can not always be guaranteed. In the thesis, the implications of using unlicensed spectrum and their impact on latency and reliability are analysed. Given the need for improving the baseline performance, novel radio resource management (RRM) techniques are proposed. The thesis provides recommendations for achieving demanding latency and reliability requirements in unlicensed spectrum. At the time the PhD was started, 5G operation over unlicensed spectrum was not supported. MulteFire was the only 3GPP-based radio access technology exclusively operating in unlicensed spectrum, i.e. with no anchor carrier in licensed spectrum, for which radio specifications were available. MulteFire is based on LTE standard and it was adopted as the reference system model for the initial studies. A transition towards 5G unlicensed is carried out during the PhD. In the remaining of the chapter, 5G service classes, URLLC and 3GPP unlicensed-based technologies are described. Moreover, the term Industry 4.0 and the role of unlicensed spectrum is introduced. The research methodology, thesis contributions and thesis outline are also detailed.

1 5G overview and service classes

As envisioned by the International Telecommunications Union (ITU) in the IMT-2020 recommendation [5], 5G shall support a wide range variate of applications which are far beyond the ones supported by any of the previous cellular generations. ITU identifies three main cornerstones as the main service classes to be supported by 5G:

- Enhanced mobile broadband (eMBB): it corresponds to the natural evolution of the previous cellular generation, driven by the increasing demand for traffic in human-centric communications. It targets to support indoor hotspots scenarios with stringent user density requirements and high peak data rates, targetting up to 20 Gbit/s in DL and 10 Gbit/s in UL. Wide-area scenarios with requirements such as seamless coverage and medium to high mobility (up to 500 km/h) while supporting high data rates should be fulfilled.

- Massive machine type communications (mMTC): this service class represents machine-centric communications. It covers use cases with a large number of devices, in the order of 1 000 000 devices/km^2, sparsely transmitting small packets with delay-tolerant requirements. Deployed devices are required to be low-cost, low-complexity devices with long battery life.

- Ultra-reliable and low-latency communications (URLLC): it comprises human- and machine-centric communications defined by sporadic transmissions of small payloads with stringent requirements for latency, in the range of 1 ms, and reliability, in the order of 99.999 %.

3GPP started the standardization process of the new generation using these service classes as a reference. Different technical solutions were identified and evaluated as 5G candidates as part of the Rel-14 Study Item (mid-2016), resulting in the first version of the 5G standard, known as New Radio (NR), in Rel-15 (end of 2017). With Rel-15, the focus was mainly on eMBB use-cases. Key enhancements for the support of mMTC and URLLC were also standardized. The work for Rel-16 started in 2018, aiming to cover new usage scenarios for different industrial verticals. As a result, important functionalities for the support of factory automation and vehicular to anything (V2X) were standardized. Additionally, unlicensed spectrum operation is standardized as part of Rel-16 specifications [6]. The work towards a new Release started in December 2019. One of the goals in Rel-17 is to keep expanding the versatility of 5G to a wide variety of industry verticals. Enhancements in sidelink, positioning and URLLC - including the support in unlicensed spectrum - are part of the work for enabling new industrial use cases. 3GPP has started working on the definition of the key feature for Rel-17, from January 2020. According to the agreed time-plan, the technical work is expected to be completed by the end of 2021.

2 Ultra-Reliable and Low-Latency Communications

Latency and reliability are the key performance indicators (KPIs) determining URLLC. Both metrics are closely related and they are described as [7]:

- User-plane latency: interval of time that takes to successfully deliver a packet measured in layer 2/3 at both ends of the communication. Potential additional delays from the core network are not included in this definition.
- Reliability: success probability of transmitting X bytes within a specific packet delay and a certain channel quality. A common URLLC reliability requirement is $1 - 10^{-5}$ for a 32 bytes packet transmission with a user-plane latency of 1 ms. In other words, 99.999 % of the total packet transmissions must be successfully delivered within a time boundary of 1 ms.

Meeting the stringent latency and reliability requirements suppose a big challenge as it requires modifications in the system criteria design, which is no longer only devoted to serving mobile broadband (MBB) traffic. Novel and advanced technology enablers are designed to make 5G systems compliant with the URLLC requirements and allow the support of mission-critical communications [8]. For latency reduction purposes, flexible numerology allows the possibility of conveying the information in shorter time intervals. This is achieved by increasing the sub-carrier spacing, which implies shorter orthogonal frequency-division multiplexing (OFDM) symbol duration, and employing small scheduling units such as mini-slots. Fast processing times to prepare and decode data transmissions are also optimized for latency-sensitive applications. Schemes such as self-contained subframe and grant-free UL also bring latency benefits. Reliability can be enhanced, among others, by diversity techniques in different domains: time, frequency and space. Supporting URLLC opens the possibility to expand the 5G applications, providing services to new business models and use cases not supported with previous generations. It will enable applications in different industrial verticals such as medical and health care, transport and safety, media and entertainment and industrial automation. The latter is expected to have a big impact on our society as it enables the achievement of the next industrial revolution.

2.1 Industry 4.0

Industry 4.0 refers to the fourth industrial revolution. It aims to evolve the current manufacturing industry by improving the flexibility, versatility, resource and cost efficiency, worker support and quality of the production and logistics. Connectivity is one of the key enablers of Industry 4.0 as it relies on a powerful network infrastructure to connect people, machines and objects. Traditionally, connectivity within the factory is provided via wired technologies as wireless alternatives unsuccessfully support the industrial-specific communication requirements. Wired solutions present several drawbacks: impose limitations in terms of mobility, increase the installation and maintenance cost, and reduce the reliability as they are prone to suffer from wear and tear. These implications, especially the lack of flexibility, hinder the feasibility of smart factories envisioned for Industry 4.0. In such factories, a reconfigurable production line is essential as real-time adjustment leads to more efficient and costumer-specific manufacturing [9]. Therefore, a transition to wireless deployments is required. Industry-specific wireless solutions are currently available, however, they are tailored solutions specifically designed for very limited applications. Moreover, they often are proprietary solutions with lack of interoperability. In this situation, 5G emerges as a wireless alternative for industrial deployments. It provides a single, standardized and with flexible design wireless technology to the industry players. 5G offers the deployment of private networks with full control of the network parameters given certain requirements. The usage of private networks also provides isolation from other networks, ensuring strong security, privacy, safety. Applications within Industry 4.0 require the support of eMBB, in augmented reality applications, mMTC, in massive wireless networks sensors, and URLLC, in motion control or mobile robots [10] [11].

3 Unlicensed spectrum: description, regulations and technologies

5G is designed to operate at any frequency of the radio spectrum between 400 MHz and 100 GHz [7]. Currently, the 5G standard is limited to frequency ranges up to 52 GHz [12] but on-going work in 3GPP is expected to allow deployments on higher frequencies, e.g. between 52 GHz and 71 GHz [13]. The usage of a specific frequency range depends on the target application, e.g. sub 1 GHz bands are used for wide area coverage and deep indoor penetration whereas millimeter waves (mmWave), i.e. above 30 GHz bands, are intended for providing extreme throughput performance at short-range distances and under primarily line-of-sight (LOS) conditions. Across the available frequencies, the spectrum can be classified into licensed, shared and unlicensed bands. Unlicensed bands, also known as licensed-exempt bands, comprise a set of frequencies which are free of access for any radio access technology (RAT) that operates according to specific regulatory requirements. Examples for unlicensed bands for mobile services are the 2.4 GHz and 5 GHz in sub-7 GHz spectrum as well as 60 GHz in the mmWave range. Additionally, spectrum regulators have recently agreed on releasing a large portion of the spectrum for unlicensed operation in the 6 GHz band, in the United States ranging from 5.925 GHz to 7.125 GHz [14] and in Europe from 5.925 GHz to 6.425 GHz [15]. In the sub-7 GHz spectrum, frequency bands are present in the form of unpaired spectrum and therefore, time division duplex (TDD) is adopted for operation in the unlicensed spectrum. As compared to licensed bands, which ensures an exclusive usage of certain frequencies, in unlicensed bands RATs operate collectively and without any cooperation nor required license. In contrast, they must be compliant with certain guidelines which are defined to govern the spectrum usage. Although the rules are regional- and band-specific, the essence of the spectrum regulatory requirements points to the same direction: ensure fair usage of the unlicensed frequency bands among the different RATs. With that objective, spectrum regulations limit the maximum transmit power levels and the occupied bandwidth, avoid co-channel operation with radar systems and provide guidelines to fairly use the channel [16].

3.1 Unlicensed technologies

Various wireless technologies operate over unlicensed bands. Several examples are found among the IEEE technologies. Bluetooth (also known as IEEE 802.15.1) and Zigbee (also known as IEEE 802.15.4) use the industrial, scientific and medical (ISM) 2.4 MHz frequency band for short-range communications. The 802.11 group of standards [17], commonly known as Wi-Fi, are deployed over the 2.4 GHz and 5 GHz unlicensed frequency bands. IEEE 802.11ax, the latest Wi-Fi version also known as Wi-Fi 6, is also capable of operating over the 6 GHz. Cellular technologies also put interest in the unlicensed bands. The usage of unlicensed spectrum is motivated by the scarceness and high cost of the licensed spectrum. Cellular technologies can leverage from the additional frequency resources by offloading part of the traffic carried over the licensed bands into the unlicensed channels. It allows for the improvement of the system throughput, i.e. it is mainly designed for the support of MBB applications. Licensed-Assisted Access (LAA), based on the LTE standard, is the 3GPP LTE unlicensed technology [18]. It operated over the 5 GHz band and ensures global applicability by following the most stringent regulatory requirements among those standardised in the different regions. It uses carrier aggregation to combine a licensed channel with one or more unlicensed channels. In LAA, an anchor-licensed connection must be maintained since it carries all the control and signalling information. LAA was introduced for offloading DL traffic in Rel-13 and further improved to support UL in enhanced-LAA (eLAA) as part of Rel-14. Other examples of cellular technologies are LTE-Unlicensed (LTE-U) [19], whose mode of operation resembles LAA but with limited applicability as it is not compliant with spectrum regulations in some regions, and LTE-WLAN Aggregation (LWA) [20], where tight collaboration between LTE and wireless local area network (WLAN) allows for LTE data offloading to Wi-Fi links.

MulteFire Driven by the need for overcoming certain limitations presented in LAA, a consortium of companies denoted as MulteFire Alliance (MFA) designed a radio access technology capable of solely operating over the unlicensed bands. This mode of operation expands the usage of unlicensed to new use cases and markets since the usage of an anchor carrier in the licensed spectrum is no longer needed. The first specifications were released in January 2017 under the name of MulteFire. It is based on LAA specifications (Rel-13 and Rel-14) but modified to introduce the support of standalone operation in unlicensed spectrum [21]. The first version of the standard designed MulteFire to be deployed over the 5 GHz frequency band. Further MulteFire releases bring optimizations for Internet of Things (IoT) and Low Power Wide Area (LPWA) support. Hence, the supported frequency ranges were extended to 2.4 GHz and sub-1 GHz bands. MulteFire offers industry verticals the possibility of creating, installing and operating private networks while taking benefit from the LTE technology and ecosystem and with the simplicity of Wi-Fi-like deployments. As compared to Wi-Fi 5 (IEEE 802.11ac), MulteFire supports twice as much capacity, increases the coverage area and allows mobility around the deployment [22]. It is the first cellular standard for standalone unlicensed operation and supposes the first step towards a global wireless unlicensed solution for smart industries.

New Radio Unlicensed Following the trend established during the LTE era, 5G also considered unlicensed spectrum as an attractive resource. 3GPP, in Rel-16, started to investigate the support of 5G-NR over the unlicensed bands. The 5G-NR unlicensed variant is known as New Radio Unlicensed (NR-U) [6] and, based on Rel-16 agreements, it is designed to operate over 5 and 6 GHz. Further releases expect to broaden the NR-U application to the mmWave range (60 GHz). Different NR-U deployment designs are identified for NR-U operation. Among them, carrier aggregation between licensed and unlicensed bands, as in LAA, and standalone unlicensed operation, as in MulteFire, are considered. The latter is the special interest for industrial verticals as connectivity can be provided without the need of a licensed carrier, which implies a significant reduction in the network infrastructure costs.

Moreover, simplicity in the deployment and larger available bandwidth are also strong points in favour of standalone unlicensed. On the other hand, spectrum regulations impose limitations to the NR-U performance. Therefore, for fulfilling the most demanding Industry 4.0 use cases, NR-U needs to be further enhanced. With that purpose, 3GPP has recently agreed on the need for supporting ultra-reliable low-latency communications over the unlicensed spectrum (URLLC-u), as stated in the Industrial Internet of Things (IIoT) Study Item for Rel-17 [23].

3.2 Challenges to achieve URLLC-u

The aforementioned spectrum regulations impose additional challenges in the achievement of URLLC-u requirements in standalone mode. Especially critical for latency and reliability are the mandatory channel access mechanisms. They are meant for guaranteeing a fair usage of the spectrum among devices simultaneously operating on the same unlicensed channel. To achieve this, devices from each unlicensed RAT must ensure there is no activity on the channel before transmitting as well as restrict the amount of time that the channel can be continuously occupied. The channel access mechanisms are built on the concept listen-before-talk (LBT). LBT is a contention-based protocol that allows devices to use the same frequency channel without any coordination. It consists of a sensing procedure, in which devices measure the received interference, and a posterior comparison with a predefined energy detection (ED) threshold. If the measured interference is above the threshold, the device declares the channel as busy, meaning that the channel is currently used by another device. Otherwise, the channel is declared as idle. A device can start a transmission only if the channel is observed as idle for a certain amount of time. These mechanisms impact the channel availability, which is crucial for achieving the URLLC requirements.

4 Scope and Objectives of the Thesis

The conducted study focuses on achieving low-latency and ultra-reliable communications for an industrial deployment over the 5 GHz unlicensed band. As compared to the licensed spectrum, unlicensed operation brings additional limitations which makes the achievement of the URLLC requirements more challenging. The goal is to propose RRM solutions that tackle these difficulties, striving to achieve a performance similar to the equivalent 5G licensed spectrum deployments. Although the focus is on the 5 GHz band, the findings of the PhD can be generalized to any unlicensed band that adopts LBT as the coexisting mechanism.

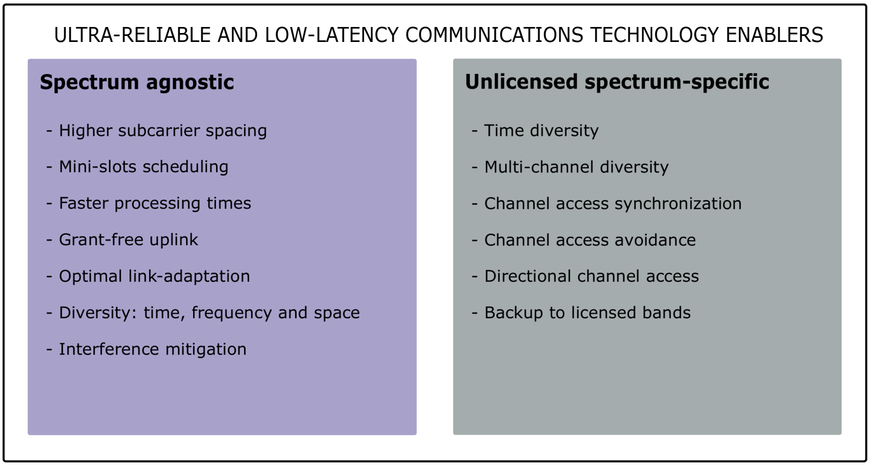


Fig. I.1: 5G technology enablers candidates for achieving URLLC classified into spectrum agnostic and unlicensed band-specific techniques.

In the first stage of the PhD, a study understanding the main bottlenecks limiting the achievement of low-latency and high-reliable communications is conducted. The evaluation is performed in an indoor scenario resembling a factory plant under different network offered load conditions. Our focus is to analyze the impact of the spectrum regulations, specifically the mandatory channel access mechanisms, in DL and UL transmissions. Transmissions are susceptible to be delayed due to these mechanisms, limiting the supported quality-of-service (QoS). In the DL, base stations (BSs) continuously sense the channel until it is declared available. This results in transmissions suffering from additional delays depending on the channel activity. In the UL, channel access procedures impose critical limitations in grant-based transmissions. user equipments (UEs) transmissions are subjected to the interference measured on a single measurement interval. If the channel is declared as occupied during that timespan, the UL transmission is not performed. The relevance of the channel access for UL transmissions is studied from a latency-reliability perspective, highlighting the need for improvement.

The acquired knowledge is then used for design and evaluate novel RRM techniques that improve the system performance from a latency and reliability perspective. As introduced in Fig. I.1, techniques can be classified into two groups: i) spectrum agnostic and ii) unlicensed spectrum-specific. The first group includes well-known techniques defined in the NR standard for latency reduction purposes. The goal is to analyze any additional benefit of these techniques when they are applied to unlicensed operation. Among others, reduced transmission times interval and faster processing times are evaluated. To improve the UL performance, grant-free is also considered as a candidate. The change in the UL paradigm, as compared to grant-based UL, can provide additional benefits in unlicensed operation since fewer channel access procedures are needed before the uplink data transmission. However, it may impose some limitations due to the lack of dedicated resources per UE. Techniques classified as unlicensed spectrum-specific are designed to reduce the influence of the channel access on the system performance. Time and frequency diversity techniques as well as multi-node coordination as part of this group of solutions. Time diversity techniques are evaluated to increase the success probability of accessing the channel at the UEs side. Both control and data transmissions are addressed. Providing more UL resources, in a TDD system, might have an impact in the DL performance. An analysis of the trade-offs is also conducted.

Multi-node coordination is also considered as an enabler for URLLC-u. Previous techniques aim to mitigate the impact of the channel access mechanisms, in this case, we directly tackle the source of the problem. Based on multi-node coordination, the effects on the channel access can be completely removed. Our goal is to propose a coordinated scheme that aligns the channel sensing interval of the nodes in the deployment. Synchronicity at the BSs side is supported by spectrum regulations for the 5 GHz. However, UEs sensing intervals depend on the scheduling decision, which is independently determined at each BS. Consequently, non-zero blocking probability is observed in this case due to interference created by other nodes transmissions. Coordination from a high-level hierarchy node is needed to achieve synchronized channel access at both ends. Thus, a central node is in charge of ensuring the synchronicity. The performance of this type of deployment is addressed. The application of this approach brings certain trade-offs which are also investigated. One of the trade-offs of using a fully synchronous approach is that BSs must use the same frame configuration. This assumption, especially in conditions with very diverse traffic conditions in the scenario, can lead to longer frame alignment delays. To overcome this problem, a coordination scheme based on silencing patterns is proposed and evaluted.

Additionally, frequency diversity techniques are proposed for reducing the channel access delay. To evaluate them, a more challenging scenario is targetted by assuming multiple RATs coexisting over the same frequency band. In our study, an NR-U deployment partially shares the frequency resources with Wi-Fi. In such deployment, the gains of the previously presented mechanisms for achieving URLLC-u are diminished by the presence of Wi-Fi. Wi-Fi is the incumbent technology in the 5 GHz band and it is designed with a more favourable channel access configuration as compared to other RATs. For instance, Wi-Fi uses a higher energy detection threshold (-62 dBm) as compared to NR-U (-72 dBm) which allows for faster channel access. We evaluate the latency-reliability degradation due to Wi-Fi presence and analyze the performance of several frequency diversity techniques. The goal is to provide NR-U nodes with supplementary frequency resources to be employed in case the channel availability is reduced due to Wi-Fi transmissions. Thus, it is always assumed that, at least, a 20 MHz band is free of Wi-Fi interference. The evaluated techniques are solutions currently available for licensed operation and whose target is to boost the available bandwidth, i.e. suitable for eMBB applications, or to increase the packet decoding probability, at the receiver side, by sending copies of the same packet, i.e suitable for URLLC applications. For unlicensed operation, our approach is to adopt them from a novel perspective and exploit the diversity in frequency to increase the channel access probability at the transmitter side. With several channels available for transmission, the impact of Wi-Fi can be mitigated while maintaining the fair coexistence and supporting a stringent latency-reliability QoS.

In the following, the main research questions and their corresponding hypothesis addressed during the PhD are formulated:

Q1 What is the impact of spectrum regulations on the unlicensed cellular latency and reliability performance?

H1 Mision-critical communications are expected to be delayed due to reduced availability of the unlicensed channel. A study for understanding the impact of the spectrum guidelines on the latency and reliability of the system is needed as a starting point for deriving further enhancements.

Q2 Which technology components, initially designed for URLLC in licensed spectrum, are suitable for unlicensed operation?

H2 Mechanisms targetting latency reduction in licensed spectrum can be adopted to unlicensed operation, potentially bringing additional benefits. They are equally valid for unlicensed operation, while their exact benefits may differ from that observed in licensed bands due to e.g. the interaction with the channel access mechanisms. Technology components such as shorter tranmissions interval times, faster processing times or configured-grant are evaluated.

Q3 How to mitigate the performance impact of the channel access by means of RRM techniques?

H3 Grant-based UL transmissions, both control and data, are heavily affected by the channel access mechanisms. These transmissions are subjected to the measured interference during few µs before the allocated resources. An unsuccessful LBT before a data transmission requires a new grant, and therefore, additional channel access attempts at BS and UE sides. In contrast, in for DL data, a BS can simply transmit in the data on the next occasion. For UL control information, e.g. hybrid automatic repeat request (HARQ) feedback, experiencing an LBT failure introduce uncertainty in the BS about the decoding outcome at the UE side. This triggers a dilemma at the BS side to whether retransmit the data or not. In scenarios targetting ultra-reliable QoS, it is assumed that the BS decides to retransmit the data. This implies the performance of additional channel access and postponing any new packet transmission to further occasions. In summary, channel access blocking in UL implies a degradation in the UL and DL performance. To overcome this issue, time-diversity techniques to increase the probability of successful channel access are proposed.

Q4 How to efficiently deploy a NR-U network to cope with the reduced channel availability?

H4 Uncoordinated channel access might lead to cases with large channel access delays. Synchronicity in the channel access can reduce the channel access delay, improving the channel availability, and hence, the latency and reliability performance. Coordination among the nodes is study as an enabler for achieving URLLC targets.

Q5 How to achieve URLLC requirements in a scenario with multiple RATs share part of the available spectrum?

H5 Coexisting with other RATs brings additional challenges. In such deployment, the spectrum guidelines make sure that the resources are shared among the RATs. In order to maintain the URLLC QoS, frequency diversity techniques for a multi-channel deployment are considered as potential solutions.

5 Research Methodology

Based on the research questions and hypothesis, a classical research methodology is applied to acquire knowledge, provide insights and fulfil the study objectives. Several steps are identified as essential during this process and a summary of them is presented as follows:

1. **Problem and objective identification:** A clear problem description and short and long-term targets definition are key to the research process. Solid knowledge about the state of the art is built by extensive literature review. Fruit of the interaction with supervisors and other academy and industry experts, the current status of the unlicensed technologies, their performance and limitations to achieve stringent latency-reliability requirements are identified. A carefully understanding of the spectrum

regulations is required in order to establish realistic objectives which are compliant with the established spectrum rules. As part of this initial phase, it is also important to become familiar with the performance evaluation tool.

2. **Hypotheses formulation and potential solutions:** The acquired knowledge helps us in the formulation of the hypotheses and potential solutions. Additionally, a more detailed literature review searching for state-of-art techniques that might solve the formulated hypotheses is conducted. As part of this stage, preliminary analyses also help in formulating the hypotheses. Discussion with supervisors and colleagues brings new proposals to combat the communications bottlenecks. The proposals are intended to be as feasible as possible to current and future unlicensed deployments while respecting the unlicensed regulations.

3. **Validation of the proposals:** Accounting for all the dynamics involved in a real wireless deployment is a complex task. Evaluation of the proposals following a strictly theoretical approach is proven to be very difficult. Therefore, a system-level simulator is chosen as the evaluation tool. The proposals are implemented in the Nokia Bell Labs proprietary system-level simulator and evaluated following the Monte Carlo method [24]. The system-level simulator is capable of reproducing, with a high degree of realism, the majority of the process involved in a wireless communication. It is built over complex mathematical models that mimic the stochastic and complex nature of mobile networks. It is designed to model multi-cell multi-user deployments, under advanced channel propagation conditions and with most of the 5G radio resource management techniques. It is also inlined with the required guidelines for unlicensed operation. A sufficient number of samples is collected during the Monte Carlo simulations in order to obtain statistically reliable results. The methodology used implies the execution of multiple realizations of the considered scenario. In order to account for all the dynamics in the system, each realization has a different UEs location. The samples of each realization are combined afterwards.

4. **Performance analysis:** The potential solutions are compared against baseline assumptions, highlighting the performance impact of the proposed schemes based on the KPI. In this stage, statistical analysis of the simulation results as well as a sensitivity analysis are performed. It helps in understanding missing behaviours or corner-cases that were not identified when formulating the research questions and potential proposals. Re-adjustment of the proposed radio-resource management techniques might occur during this phase.

5. **Dissemination of the results:** The proposed solutions are described in details and presented to the research community in form of scientific publications, targetting highly-impacting conferences and journals. Dissemination of the acquired knowledge is also achieved by giving presentations in project-related meetings at Wireless Communication Networks (WCN) section and Nokia Bell Labs. As part of ONE5G, a European project driven by the development of new radio technologies for 5G, results have been presented with positive feedback [25]. The conducted work has also influenced Nokia Bell Labs views on its contributions to 3GPP for unlicensed spectrum. Additionally, performance results will be included as part of a Nokia Bell Labs whitepaper in which the company express its view on unlicensed technologies. As an outcome of this research process, novel ideas are generated and protected, acquiring intellectual property rights via patent applications.

6 Contributions

The main findings of the study are summarized bellow:

1. **Identifying the challenges of achieving low latency in unlicensed deployments.** A comprehensive analysis of the bottlenecks limiting the system latency performance is carried out. It is shown that the channel access mechanisms heavily impact the latency performance at both ends of the communication. BSs channel access is shown to increase with the offered load and, already in low-to-moderate loads, it often surpasses the delay budget of delay-sensitive QoS. Nonetheless, due to its continuous sensing approach, the BSs ensure that the channel is accessed in the long run. On the other hand, UEs typically use a single-shot approach, in which the channel is assessed based on a single sensing interval. It leads to frequent blocking in the channel access, increasing the packet delay and limiting the application of unlicensed technologies to delay-tolerant applications. The study highlights the need for improvement to make unlicensed standalone technologies a candidate for URLLC deployments. The analysis is conducted following the MulteFire assumptions but conclusions apply to any unlicensed technology working in standalone mode.

2. **Analyzing the impact of URLLC RRM techniques over unlicensed spectrum, highlighting their additional benefits.** Supporting URLLC technology components, initially designed for licensed operation, has been shown to provide additional benefits in the unlicensed spectrum. Shortening the transmission time interval, apart from providing the inherent reduction in latency, it also reduces the time the channel is

actively used. This has a direct effect on the channel access performance making it faster and with higher success probability. Introducing multiple transitions between downlink and uplink within a single TDD frame configuration also brings benefits as messages exchange between BSs and UEs can be contained in a single channel occupancy time. Grant-free UL is also studied as an alternative to grant-based UL. It is shown to provide latency reductions as less channel access procedures are performed as compared to grant-based UL. Additionally, UL transmissions are more likely to happen as the continuous channel sensing, rather than the single-shot channel access, is followed. However, grant-free performance is bounded by the supported system load as collisions among UEs might appear.

3. **Proposing RRM techniques to cope with the channel access impact.** The channel access performed at the UE side is especially critical for delay-sensitive applications. Assuming grant-based scheduling, UEs transmissions are subjected to two conditions: i) receive a grant from the serving BS with specific time-frequency resources and ii) successfully access the channel right before the allocation starts. In case the channel is declared as busy, the granted resources are not used as transmissions are not allowed to take place. Thus, the BS must issue a new grant, which requires the performance of additional channel access mechanisms. This process can take several iterations until the UE successfully access the channel. To overcome this issue, we propose time-diversity RRM techniques that allow the BS to indicate several time instances for attempting the channel access in UL. Consequently, the channel access probability at the UE side is increased and the number of channel access attempts at the BS side reduced. For the physical control-plane transmissions, specifically HARQ responses to previously received DL transmissions, UEs are provided with K consecutive attempts to perform the channel access. Only after K consecutive channel access failures, the DL transmission is retransmitted due to the uncertainty at the BS of the UE decoding outcome. For the physical user-plane UL transmissions, BSs signal in the grant the M consecutive resources that UEs can potentially use for the UL data transmission. Additionally, the possibility of skipping the performance of LBT as a responding device is studied under different load conditions. Spectrum regulations allow for transmissions without LBT under the condition that the gap interval between the last DL transmission and the next UL transmissions is shorter than 16 µs.

4. **Introducing an unlicensed mode of operation free of channel access impact.** A new concept based on coordination among the nodes is proposed. The technique remains compliant with the spectrum regulations,

as the channel access mechanisms are performed, but the sensing interval timings are common for all the nodes. Synchronicity in the channel access together with coordination in the frame selection leads to full mitigation of the impact of LBT. To do so, the presence of a central node that coordinates the transmissions is required. The set-up is studied under different load points and compared against the asynchronous case, highlighting the benefits and limitations. The application of this proposal ensures that, by having a well-designed deployment, similar latency-reliability performance as NR TDD licensed can be achieved. However, agreeing on the TDD frame among the nodes may introduce unnecessary frame alignment delay, especially in scenarios with large variations in the traffic conditions, due to non-ideal frame selection. To overcome this, an alternative based on silent gap coordination among nodes is proposed. There, nodes are configured with a common silencing pattern that collides with the potential sensing intervals of neighbour nodes. It achieves 0 % channel access blocking probability in the system while allows each node to select the optimal frame that fits best their traffic conditions.

5. **RRM proposals to combat sporadic interference from other RATs and its impact on latency and reliability.** Sharing the channel with multiple RATs unavoidably reduces the channel availability, which in turns, significantly degrades the system delay performance. In order to keep the stringent latency and reliability requirements, we propose to seek additional resources in the frequency domain. Techniques designed for licensed spectrum operation are applied to unlicensed band from a different perspective. Initially, for licensed operation, the frequency diversity techniques aim to either increase the supported throughput, i.e. used for mobile broadband purposes, or increases the decoding probability at the receiver side, i.e. for reliability improvement. In our case, the target is to increase the successful channel access probability at the transmitter side. The evaluation is conducted on an indoor factory scenario with NR-U and IEEE 802.11ax contending for the channel. Both technologies partly coexist over the unlicensed bands, as our assumption is that the NR-U network has more spectrum available than IEEE 802.11ax. Carrier aggregation, wideband operation and packet data convergence protocol (PDCP) duplication are studied from a latency and reliability perspective. Performance comparison against the case in which IEEE 802.11ax is not present is conducted. The study shows that carrier aggregation, under the current standard specifications, do not properly combat the presence of inter-RAT interference. The fact that transmissions are not flexibly re-scheduled to other channel limits its performance. The study concludes that to mitigate the

effect of unplanned interference, schemes with flexible readjustment of the selected channel for transmission are required.

The thesis is composed of a collection of papers. Formulations, models, ideas, and results from these papers are therefore presented throughout the thesis. The main findings and contributions are included in the following publications:

Paper A: R. Maldonado, C. Rosa, F. Frederiksen and K. I. Pedersen, "On the Impact of Listen-Before-Talk on Ultra-Reliable Low-Latency Communications", *IEEE Global Communications Conference (GLOBECOM)*, December 2018, pp 1-6.

Paper B: R. Maldonado, C. Rosa, F. Frederiksen and K. I. Pedersen, "Uplink Ultra-Reliable Low Latency Communications Assessment in Unlicensed Spectrum", *IEEE Globecom Workshops (GC Wkshps)*, December 2018, pp 1-6.

Paper C: R. Maldonado, C. Rosa and K. I. Pedersen, "Latency and Reliability Analysis of Cellular Networks in Unlicensed Spectrum", *IEEE Access*, 2016, pp 49412-49423.

Paper D: R. Maldonado, C. Rosa and K. I. Pedersen, "Analysis of High-Reliable and Low-Latency Communication Enablers for New Radio Unlicensed", *IEEE Wireless Communications and Networking Conference (WCNC)*, April 2020, pp 1-6.

Paper E: R. Maldonado, C. Rosa and K. I. Pedersen, "A Fully Coordinated New Radio-Unlicensed System for Ultra-Reliable Low-Latency Applications", *IEEE Wireless Communications and Networking Conference (WCNC)*, April 2020, pp 1-6.

Paper F: R. Maldonado, C. Rosa and K. I. Pedersen, "Multi-link Techniques for New Radio-Unlicensed URLLC in Hostile Environments", *IEEE Vehicular Technology Conference (VTC), April 2021*, **Submitted for publication**

Paper G: R. Maldonado, C. Rosa and K. I. Pedersen, "A silent gap coordination for supporting URLLC in unlicensed deployments", **Work not submitted for publication**

Additionally, several patent applications have been drafted. Among them, the ones below have been successfully filed:

Patent Application 1: Mechanism to determine the channel access parameters based on the transmission of robustness indication by neighbour nodes

Patent Application 2: Scheduling request enhancement for URLLC/IIoT in NR-U

Patent Application 3: Smart channel access implementation

Patent Application 4: Method for dynamic temporary adaptation of measurement gap configuration in NR-U

Patent Application 5: Mobile device capable to send wakeup signal to nearby devices

Patent Application 6: Relaying paging messages for proximity service remote UEs

The Nokia proprietary system-level simulator is the tool adopted for system performance evaluation in this study. The simulator is built in C++ and it models the majority of the physical layer (PHY) and medium-access-control layer (MAC) mechanisms involved in a cellular communication such as packet scheduling, HARQ and link adaptation, power control. It is capable of three dimensional (3D) channel modelling and accurate interference calculation. It also includes the specific functionalities for unlicensed operation. The simulator has been calibrated against other 3GPP companies simulators and it is extensively used for providing LTE and 5G performance analysis for both academia and industry. Additionally, the simulator supports the modelling of Wi-Fi deployments. It includes the possibility of running co-existence studies with detailed modelling of interference from one RAT to another. During the PhD, the simulator has been developed according to the specific needs of each of the conducted studies. Time is invested in understanding the key functionalities of the simulator as well as in debugging, testing and documenting the new implementations, ensuring their performance consistency under different conditions. The added contributions to the simulator are summarized in the following:

- **Unlicensed spectrum operation:** following the 5 GHz spectrum regulations, a realistic model of the guidelines for unlicensed operation is added to the system-level simulator. Aspects such as the aynchronous and synchronous channel access designs, avoidance of channel access under specific conditions as well as the limitation in the channel occupancy time are taken into consideration.

- **Improvement of RRM functionalities:** a realistic scheduling request model for grant-based UL transmissions and a grant-free UL scheme for UL transmissions are implemented (Papers B and C). Adaptation of the TDD frame based on the instantaneous traffic conditions, i.e. dynamic TDD, is also developed.

- **Uplink control transmissions dependant on channel access:** successful delivery of the signaling information between UEs and next generation Node-Bs (gNBs) is only possible given a successful LBT. Control information such as HARQ responses to previously received DL transmissions, scheduling request (SR) messages prior to any UL data transmissions and channel quality indicator (CQI) reports, are enhanced to be dependant on the LBT outcome. Adding this functionality align the simulator with reality, adding more challenges to unlicensed spectrum operation.

- **Time diversity techniques:** contributions described in Papers C and D are based on providing additional attempts for transmissions in the time domain. Techniques for improving the channel access probability when transmitting UL control and UL data are implemented.

- **Nodes coordination:** as part of the studies conducted in Paper E, a centralized scheme was proposed. For that, a functionality for coordinating the channel access and frame selection among the deployed nodes is added. Additionally, a functionality where nodes are coordinated according to a silent gap pattern is added.

- **Frequency diversity techniques:** the support of different frequency diversity techniques for unlicensed operation was required for the studies in Paper F. Adaptation of carrier aggregation to unlicensed spectrum and implementation of wideband operation were performed. Different algorithms for the channel selection and multi-band channel access are also part of the contributions.

- **Indoor factory channel model:** during the development of the PhD, a specific channel model for indoor factory scenarios was defined. It supposes an alternative to the indoor hotspot scenario used for indoor office deployments. In unlicensed spectrum, apart from the usual gNB-to-UE channel model, it is also important the gNB-to-gNB and the UE-to-UE channel model as they impact the LBT outcome. The possibility of using a specific channel model among the defined for the indoor factory for each of the links is added.

7 Thesis Outline

The thesis is written as a collection of papers and it is structured in 6 parts. In Part I, an introduction describing the conducted work is presented. The main contributions are included in Parts II, III and IV and they are presented in the form of articles. Each of these parts contains a summary highlighting the motivation and the main findings. Conclusions are found in Part V. Part VI includes an appendix with unpublished results. The content of each part is described here:

- **Part I: Introduction -** This part introduces the PhD topic, motivates the research and summarizes the added contributions during the conducted study.
- **Part II: Challenges in supporting URLLC over unlicensed spectrum -** This part includes an analysis of the impact of the channel access mechanisms on the latency-reliability using MulteFire as the reference model. DL and UL directions are analyzed in an indoor deployment. Several mechanisms are also proposed to enhance the baseline performance.
- **Part III: URLLC over unlicensed spectrum in controlled environments -** A transition from MulteFire to NR-U is covered in this part. This part focuses on the evaluation of the gain that key NR technology components can provide when applied to unlicensed spectrum. Additionally, to strive for the achievement of URLLC-u requirements, coordination among the nodes is shown as a promising technique. The analysis is based on controlled environments conditions, i.e. a single NR-U network is deployed in the channel.
- **Part IV: URLLC over unlicensed spectrum in hostile environments -** RRM techniques to overcome the presence of multiple radio access technologies, i.e. in hostile environments, is addressed in this part. We analyze different frequency diversity techniques from a latency and reliability perspective.
- **Part V: Conclusions -** In this part we draw the conclusions of the study, summarize the main findings and provide design recommendations for fulfilling the URLLC-u requirements. Topics that should be addressed in future studies are also mentioned.
- **Part VI: Appendix -** A new coordination scheme is described in this part. Coordination among the nodes is achieved by following a periodic silencing pattern. It represents an alternative to the coordination scheme presented in Part III.

References

[1] E. Dahlman, S. Parkvall, and J. Sköld, "Chapter 1 - Background of LTE," in *4G LTE/LTE-Advanced for Mobile Broadband*. Academic Press, 2011.

[2] H. Holma and A. Toskala, "Chapter 16 - HSPA Multicarrier Evolution," in *WCDMA for UMTS: HSPA Evolution and LTE*. Wiley, 2010.

[3] Ericsson, "Ericsson Mobility Report," Jun. 2020. [Online]. Available: https://www.ericsson.com/49da93/assets/local/mobility-report/documents/2020/june2020-ericsson-mobility-report.pdf

[4] O. Liberg, M. Sundberg, Y.-P. E. Wang, J. Bergman, J. Sachs, and G. Wikström, "Chapter 7 - NB-IoT," in *Cellular Internet of Things (Second Edition)*. Academic Press, 2020.

[5] International Telecommunication Union (ITU), "IMT Vision - Framework and Overall Objectives of the Future development of IMT 2020 and beyond," ITU-R, Tech. Rep., Sep. 2015.

[6] 3GPP, "Physical layer procedures for shared spectrum channel access (Release 16)," Tech. Rep. 37.213 V16.2.0, Jun. 2020.

[7] ——, "Study on Scenarios and Requirements for Next Generation Access Technologies; (Release 16)," Tech. Rep. 38.913 V16.0.0, Jul. 2020.

[8] S. Ahmadi, in *5G NR: Architecture, Technology, Implementation, and Operation of 3GPP New Radio Standards*. Academic Press, 2019.

[9] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, and B. Yin, "Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges," *IEEE Access*, vol. 6, pp. 6505–6519, 2018.

[10] 5G-ACIA, "5G for Connected Industries and Automation - White Paper, 2nd Edition," 2018.

[11] 3GPP, "Service requirements for cyber-physical control applications in vertical domains (Release 17)," Tech. Rep. 22.104 V17.3.0, Jul. 2020.

[12] ——, "Base Station (BS) radio transmission and reception (Release 16)," Tech. Rep. 38.104 V16.4.0, Jun. 2020.

[13] ——, "Study on supporting NR from 52.6 GHz to 71 GHz (Release 17)," Tech. Rep. 38.808 V0.0.1, Sep. 2020.

[14] Federal Communications Comission (FCC), "FCC adopts new rules for the 6 GHz band, unleashing 1200 MHz of spectrum for unlicensed use," 2020. [Online]. Available: https://docs.fcc.gov/public/attachments/DOC-363945A1.pdf

[15] European Telecommunications Standard Institute (ETSI), *System Reference document (SRdoc); Wireless access systems including radio local area networks (WAS/RLANs) in the band 5 925 MHz to 6 725 MHz*, ETSI TR 103 524 V1.1.1 Std., Oct. 2018.

[16] ——, *5 GHz RLAN; Harmonised Standard covering the essential requirements of article 3.2 of Directive 2014/53/EU*, ETSI EN 301 893 V2.1.1 Std., May 2017.

[17] "Ieee draft standard for information technology – telecommunications and information exchange between systems local and metropolitan area networks – specific requirements part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment enhancements for high efficiency wlan," *IEEE P802.11ax/D6.0, November 2019*, pp. 1–780, 2019.

[18] 3GPP, "Study on Licensed-Assisted Access to Unlicensed Spectrum, (Release 13)," Tech. Rep. 36.889 V13.0.0, Jun. 2015.

[19] R. Zhang, M. Wang, L. X. Cai, Z. Zheng, X. Shen, and L. Xie, "Lte-unlicensed: the future of spectrum aggregation for cellular networks," *IEEE Wireless Communications*, vol. 22, no. 3, pp. 150–159, 2015.

[20] D. Laselva, D. Lopez-Perez, M. Rinne, and T. Henttonen, "3GPP LTE-WLAN Aggregation Technologies: Functionalities and Performance Comparison," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 195–203, 2018.

[21] M. Alliance, "MulteFire Release 1.0 Technical Paper: A New Way to Wireless," Tech. Rep. [Online]. Available: https://www.multefire.org/

[22] 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, 2018.

[23] 3GPP, "Enhanced Industrial Internet of Things (IoT) and URLLC support," Tech. Rep. RP-193233, Dec. 2019.

[24] R. F. Coates, G. J. Janacek, and K. V. Lever, "Monte Carlo simulation and random number generation," *IEEE Journal on Selected Areas in Communications*, vol. 6, no. 1, pp. 58–66, 1988.

[25] ONE5G, "Deliverable D3.2 Recommended Multi-Service Performance Optimization Solutions for Improved E2E Performance," May 2019.

Part II

Challenges in supporting URLLC over unlicensed spectrum

Impact of channel access mechanisms on latency and reliability

This part provides insights about the feasibility of supporting high-reliable low-latency applications over unlicensed cellular technologies. It focuses on the analysis of the limiting factors that prevent the support of stringent latency and reliability QoS requirements with an emphasis on the channel access mechanisms. This initial part of the PhD is used as a reference for the rest of the study as it highlights the weaknesses of unlicensed band operation. Under the assumptions of an indoor deployment, the chapter includes an analysis of the impact of the channel access mechanisms at both ends of the communication. Based on the extracted conclusions, novel RRM techniques are introduced and evaluated to minimize the impact of the spectrum regulations on the system performance.

1 Problem Description

Using unlicensed bands for transmissions adds further complexity to the challenge of achieving low-latency and high-reliable communications. Due to the open-access nature of the unlicensed bands, spectrum regulators must ensure a fair coexistence among the multiple radio access technologies sharing the same frequency bands. Regulatory requirements by the European Telecommunications Standards Institute (ETSI) are among the most stringent for operation in the 5 GHz spectrum. LAA, MulteFire and NR-U are designed based on ETSI specifications as baseline for operation in the 5 GHz band. ETSI includes in [1] the technology-agnostic guidelines to govern the 5 GHz spectrum usage. Due to their influence on the channel availability, the channel access mechanisms are thoroughly analyzed in the following. Two different implementations of the channel access are specified by ETSI.

The first channel access design is denoted as frame-based equipment (FBE) and assumes a periodic timing structure that defines the channel sensing and channel occupancy intervals. On the other hand, load-based equipment (LBE) adopts an asynchronous approach in which the channel sensing and the channel occupancy intervals occur at any time instance. The focus on this part is on LBE, the only channel access design commonly supported by LAA, MulteFire and NR-U. NR-U also supports FBE and it is therefore studied at a later stage of the PhD. In the spectrum regulations, ETSI differentiates between initiating and responding devices. In both cases, devices must perform LBT to evaluate the channel activity. However, the sensing protocol is applied differently. Firstly, the channel access is acquired by the initiating devices. To do so, they continuously perform LBT over, at least, N consecutive sensing slots of 9 µs. N is given by a uniform distribution and defines the minimum number of sensing slots where the channel must be assessed as free before LBT can be declared successful. This type of channel acess is denoted as Type 1 or Category 4 LBT [2]. Upon the LBT success, the initiating devices are allowed to transmit and gain access to the channel for a limited amount of time. This time is also known as maximum channel occupancy time (MCOT). During that time interval, responding nodes can leverage from the recently acquired channel and, upon an authorization reception, share it with the initiating devices. In this case, responding devices adopt an LBT with a fixed sensing duration, in the order of tens of µs, to evaluate the channel activity. This type of LBT is denoted as Type 2 or Category 4 LBT. A negative LBT outcome implies the blocking of the responding devices transmissions. Applying these concepts to a 3GPP-based technology leads us to typically define BSs as initiating nodes and user equipments (UEs) as the responding devices. This is in line with the usual BS-centric approach in which base stations are in charge of the scheduling and ultimately provide UEs with the specific time and frequency resources for transmissions. In a few cases, UEs may also act as initiating devices, for instance, during random access and scheduling request procedures.

Channel access mechanisms add uncertainty to the performance of the transmission. Interference generated by other radio access technologies, i.e. inter-system interference, and interference from nodes from the same technology, i.e. intra-system interference, can postpone the access to the medium. In the DL, before any transmission, BSs must acquire the channel by means of the continuous LBT. Due to the lack of control of these source of interference, BSs might experience high channel access delay which, in some cases, can already surpass the latency budget of delay-sensitive applications. Once the channel is acquired and the DL transmission is performed, the intended UEs receive the data. After certain processing time, the UEs need to signal back the positive acknowledgement (ACK) or negative acknowledgement (NACK) based on the decoding outcome. UEs must perform the fixed duration LBT, also known as single-shot LBT, before the transmission. A potential failure in the access of the channel occurs if: a neighbour node accesses the channel during a gap without transmissions or due to the hidden node problem. The hidden node problem is a well-known phenomenon that occurs when, due to differences in the measured interference levels, a node (the BS) declares the channel as idle after a successful LBT and grant access to another node (the UE) that experiences interference higher than the ED. A blockage in the channel access prevents the usage of the reserved UL resources. In such a case, and since certain reliability level must be ensured, the BS retransmits the DL packet independently of the decoding outcome at the UE side. Channel access mechanisms also influence the UL data transmissions. Assuming grant-based (GB) scheduling, UE must go through a scheduling request (SR) procedure before the UL data transmission. During the SR, a handshake between BS and UE is executed to agree on the dedicated time-frequency UL resources to be used. In each step of the process, i.e. scheduling request indication, grant transmission and UL data transmission, a successful channel access procedure is required. Therefore, a total of 3 successful LBTs are required. A failure in any of the LBTs degrades the latency performance. Similar to the ACK/NACK transmission, an LBT blocking before the data transmission implies the impossibility of using the agreed resources and a new allocation needs to be issued by the BS. In summary, transmissions in both directions are susceptible of being delayed due to the mandatory channel mechanisms for operation in the 5 GHz frequency band.

2 Objectives

The goal of this part of the PhD is to provide insights in the following directions:

- Understand the limitations of supporting high-reliable low-latency communications in standalone mode over the unlicensed spectrum by extensive system-level simulations. A decomposition of the DL and UL packet delay into several components, highlighting the unlicensed-specific contributions, is provided.
- Analyze the impact of the channel access procedures in DL transmissions and quantify their contribution to the overall packet delay.
- Study the implications of a channel access blockage at the UE side and propose alternatives to mitigate its negative effects. For UL control transmissions, the possibility to avoid the single-shot LBT, in accordance with the regulations, and a time-diversity technique, are evaluated. For UL data transmissions, grant-free (GF) UL is proposed as an

alternative to GB UL. Performance analysis of the schemes is conducted through highly-detailed system-level simulations.

3 Included Articles

The following articles form the main body of this part of the thesis:

Paper A. On the Impact of Listen-Before-Talk on Ultra-Reliable Low-Latency Communications

The impact of the channel access mechanisms on the DL packets delay is studied in this paper. Firstly, the continuous LBT is analyzed from a pure probability perspective. Different LBT success probabilities are evaluated to model diverse channel congestion conditions, i.e. a high probability implies a less congested scenario in which fewer nodes content for the channel and fewer transmissions are performed. Due to the limited accuracy of this analysis, a similar study is conducted on a system-level simulator which properly represents the dynamics of a real wireless deployment. Using MulteFire as the reference model, the impact of the channel access in the overall packet delay is measured under different load conditions. Furthermore, the channel access performance at the UE side is also considered in the paper. In a DL-only scenario, UEs transmit UL control information containing the HARQ feedback of previously received DL packets. These transmissions are subjected to the outcome of the single-shot LBT. In case of a failure in the channel access, independently of the decoding result, a DL retransmission is triggered. In order to understand the relevance of succeeding in the single-shot LBT, a scenario with real UL LBT is compared against a scenario with ideal UL LBT (with 100 % success probability). The study motivates the need for improvement at both ends of the communication to better cope with the channel access influence.

Paper B. Uplink ultra-reliable low latency communications assessment in unlicensed spectrum

This paper addresses the analysis of UL transmissions in unlicensed bands from latency and reliability perspective. Starting from GB UL, a study of the SR procedure and the LBT associated with each control message exchange is conducted. As in Paper A, the single-shot LBT before the UL data transmission is remarked as critical from a delay perspective. A block in the channel access implies the need for an additional grant with new UL resources. Moreover, the SR procedure in unlicensed operation adds uncertainty and potentially increases the packet delay as multiple LBTs must be performed before the actual UL data transmission. Striving for improving the UL performance and for reducing the impact of the channel access mechanisms, GF UL is evaluated as an alternative radio-resource management mechanism for UL transmissions. With GF UL, UEs skip the SR procedure as the UL allocations are pre-configured. On the other hand, the fact that the reserved resources are shared by multiple UEs increases the probability of collisions and adds complexity at the BS decoding process. Via extensive system-level simulations, both UL procedures are compared in full-UL MulteFire deployment under different load conditions.

Paper C. Latency and Reliability Analysis of Cellular Networks in Unlicensed Spectrum

This contribution combines both previously described papers and analyzes an indoor multi-cell deployment with simultaneous DL and UL URLLC traffic. The paper introduces several enhancements for reducing the observed packet delay. For improving the DL performance, two strategies are studied. The first one is based on the ETSI guidelines for 5 GHz operation and allows the UEs to skip the performance of the single-shot LBT in cases where a gap between the last DL transmission and the next UL transmission is shorter than 16 µs. However, this condition is frequently difficult to achieve. The second proposal tries to cope with those cases with the channel access needs to be inevitably performed. It consists of using a time-diversity technique for increasing the probability of channel access. It is intended for UL control transmissions and UEs are provided with multiple opportunities to transmit the HARQ feedback. For the UL, GF is analyzed in a deployment with bi-directional traffic conditions.

4 Main Findings

Impact of downlink channel access on downlink data transmissions

For the DL performance (Paper A), the pure probability analysis shows that channel access delays of 3 ms to 18 ms at 99.99 % reliability are observed for LBT success probabilities of 0.8, 0.7 and 0.6, respectively. A similar trend is confirmed by evaluations performed using the system-level. It is shown that the continuous LBT accounts for 18 % to 42 % depending on the outage level and the supported load. This leads to the conclusion that a significant part of the latency budget is spent in the performance of the channel access. In both studies, the channel access delay is found to directly depend on the offered load in the scenario. Boosting the offered load increases the number of nodes contending for the channel which substantially increment the channel access delay, and consequently the overall packet delay.

Impact of uplink channel access on downlink data transmissions

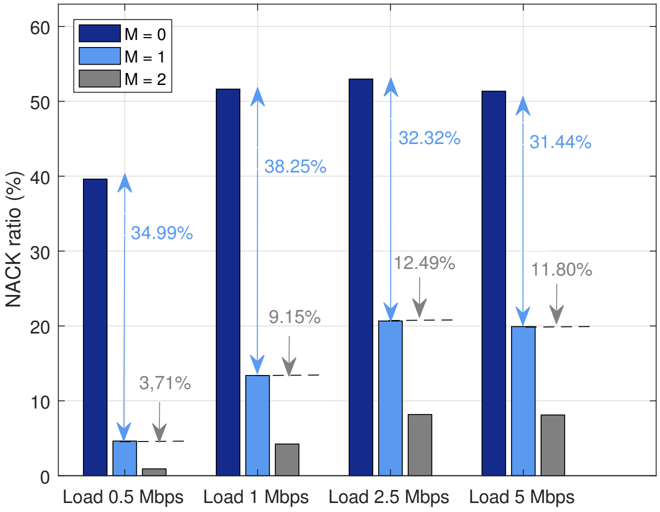


Fig. II.1: Effect of UL LBT failure in the system. The ratio between the number of triggered DL retransmissions due to UL LBT failure and the total received DL packets is shown under different assumptions for the number of additional UL resources (M) and the offered loads. (Source: Paper C)

For the HARQ feedback, it is observed that by assuming ideal LBT conditions, a significant decrease in the DL packet delay is achieved (Paper A). A reduction in delay by a factor of 2.8 is reached in the high-load case. Ensuring the channel access avoids the performance of unnecessary retransmissions (packets decoded correctly at the UE side but retransmitted due to UL LBT failure) which reduces the queuing delay and improves the spectral efficiency of the system. Moreover, it reduces the number of nodes contending for the channel. The ideal LBT conditions studied in Paper A are revisited in Paper C. In this case, the possibility of skipping LBT is only allowed, as described in the spectrum guidelines, given the fulfilment of the gap duration condition. Under these realistic conditions, the LBT blocking rate can only be reduced by 12 % for the high load conditions, and by a modest 2 % in low-loaded cases. At 99.99 % reliability, providing UEs with multiple opportunities for the HARQ transmissions presents an homogeneous gain across the load ranges as shown in Fig. II.1. In Paper C, the gain of this technique is measured according to the NACK ratio. This metric is related to the UL LBT blocking probability and measures the number of DL packets considered as NACK, due to single-shot LBT failure, in relation with the total number of received DL packets. Providing the users with an additional opportunity for the ACK/NACK transmissions gives a NACK ratio reduction ranging from 31 % to 38 % depending on the offered load. At high offered loads, due to a large number of users contending for the channel and transmissions taking multiple subframes, the NACK ratio can be further reduced by an additional 12 % when 2 additional opportunities are provided. However, the benefit at low-load is bellow 4 %. Intuitively, reserving supplementary resources for UL transmissions in a TDD system suppose a degradation in the DL performance. Nonetheless, the DL latency-reliability is improved due to the reduction of the unnecessarily triggered retransmissions which, in turn, directly impact the channel access and packet delays. At 99.99th percentile, a latency reduction of ∼10 % and ∼55 % is achieved at low and high loads, respectively.

Impact of uplink channel access on uplink data transmissions

The fact that fewer DL retransmissions are triggered also benefits the UL performance. Under GB conditions, allowing the skip of the single-shot LBT (Paper C) brings a latency reduction, at 99.99th percentile, of 11 % for low loads and 24 % for high loads. Similar behaviour is observed regarding the provision of multiple HARQ opportunities. It improves the UL system latency performance by 10 % and 18 % for low and high loads, respectively. Changing the UL paradigm to a GF is verified as beneficial in Paper B and Paper C. In Paper B, the analysis is conducted in an UL-only deployment and it is shown that GF is the preferable scheme for UL transmissions at low to moderate loads. However, the performance of GF UL degrades as the load increases due to the non-orthogonality of the UL resources. The trend is also confirmed in Paper C under bi-directional traffic conditions. Nevertheless, at low loads, a latency reduction of 52 % is achieved as compared to GB UL.

5 Key recommendations

After the conducted study, the following recommendations are identified:

- Benefit from the spectrum regulations and adopt the possibility of skipping the UL LBT for improved latency-reliability performance.
- Providing UEs with multiple resources for HARQ transmissions is advised. It reduces the impact of negative UL LBTs in the system.
- An efficient approach to reduce the uncertainty on UL data transmissions is to use GF, as compared to GB. However, its usage is only advised in scenarios with relative modest UL offered loads.

References

[1] ETSI, *5 GHz RLAN; Harmonised Standard covering the essential requirements of article 3.2 of Directive 2014/53/EU*, ETSI EN 301 893 V2.1.1 (2017-05) Std.

[2] 3GPP, "Physical layer procedures for shared spectrum channel access (Release 16)," Tech. Rep. 37.213 V16.2.0, Jun. 2020.

Paper A

On the Impact of Listen-Before-Talk on Ultra-Reliable Low-Latency Communications

Roberto Maldonado, Claudio Rosa, Frank Frederiksen, Klaus I. Pedersen

The paper has been published in the
IEEE Global Communications Conference (GLOBECOM), 2018.

The layout has been revised.

Abstract

In this paper, we study the performance of ultra-reliable and low-latency communications (URLLC) for cellular standalone systems in unlicensed bands. Our focus is on the 5 GHz *band, adopting a system model in coherence with the MulteFire standard, which essentially is a variant of Long Term Evolution (LBT) for unlicensed band operation. We show that especially the mandatory listen-before-talk (LBT) procedure suppose a challenge for achieving low latency communication. Our theoretical analysis indicated that the impact of the LBT procedure prior to downlink transmissions can be significant, taking values of* 3 ms *to* 25 ms *for the considered offered load levels. Furthermore, our advanced dynamic system-level simulations show that the time spent performing LBT often accounts for* 18 % *to* 42 % *of the packet latency budget, depending on the considered outage level and offered traffic. Therefore LBT represents a major challenge for achieving URLLC-alike performance in unlicensed spectrum.*

1 Introduction

One of the key features of fitfth generation (5G) wireless communication systems is the support of mission-critical applications demanding high reliability and low latency, also known as ultra-reliable and low-latency communications (URLLC). URLLC is one of the enablers of the fourth industrial revolution [1], as it facilitates the deployment of more flexible communication infrastructures as compared to wired communication [2]. Requirements and solutions enabling URLLC use cases have been comprehensively addressed by 3$^{\text{rd}}$ Generation Partnership Project (3GPP) during the standardization of the 5G Next Generation Radio Access Network [3]. However, 5G standardization and related research have mainly focused on licensed spectrum. On the other hand, due to its large availability globally and relatively ease of access, unlicensed spectrum is also expected to play a key role in future 5G wireless communication systems. The growing interest in unlicensed spectrum is confirmed by the introduction in 3GPP standard specifications of solutions supporting better interworking and aggregation between Long Term Evolution (LTE) and WiFi [4], as well as solutions that can leverage LTE technology in the 5 GHz unlicensed band, in combination with licensed spectrum and using carrier aggregation. The latter is also known as Licensed-Assisted Access (LAA) [5]. However, these solutions all require an anchor in licensed spectrum, and as such, they are mainly targeted at traditional mobile network operators. More recently, the MulteFire Alliance (MFA) [6] has addressed the need of large-scale technology enterprises and verticals for a wireless radio access technology that can provide reliable access to private networks, globally and without the need for expensive licensed spectrum. MulteFire is based on LTE specifications, but differently from other 3GPPbased radio access technologies, it can provide standalone access to unlicensed spectrum [6]. This trend continues in 5G, with 3GPP recently starting a new study item on New Radio (NR)-based access to unlicensed spectrum, targeting carrier aggregation, but also dual connectivity and standalone deployment scenarios [7]. Nevertheless, so far, research on the topic of unlicensed spectrum access has mainly focused on either providing enhanced mobile broadband services or enabling massive machine type communication in unlicensed spectrum. No studies are available in the open literature that specifically address the problem of providing highly reliable and low latency communication using unlicensed spectrum in the 5 GHz band. Due to the channel access uncertainty introduced by the requirement of clear channel assesstment (CCA) based on listen-before-talk (LBT), fulfilling the URLLC requirements in the 5 GHz unlicensed band presents a big challenge.

In this paper, we study how well URLLC applications can be supported in unlicensed bands, with emphasis on the downlink user-plane performance, using the MulteFire system model. Our hypothesis is that the LBT procedure plays an important role, and influences the achievable latency, so we have a special focus on it throughout our study. We first assess the latency contributed by the LBT procedure by means of analytical results. Secondly, to gain further insight, we present state-of-the-art performance results from dynamic system-level simulations with high degree of realism. Those simulations are based on commonly accepted models, in line with 3GPP simulation guidelines. The rest of the paper is organized as follows: Section 2 presents the system model and explains in detail the LBT algorithm. Section 3 contains a detailed mathematical representation of the overall delay and an analysis of the delay spent by an access point performing LBT. The simulation methodology and system-level simulation results are given in Sections 4 and 5, respectively. Section 7 contains our concluding remarks and a reference to future research.

2 System Model

2.1 Basic assumptions

The system model used throughout the paper follows the MulteFire standard. We assume an indoor scenario deployed in the 5 GHz band. In order to mimic an industrial deployment, we consider a single operator scenario with M evolved Nodes-B (eNBs) and K uniformly distributed user equipments (UEs) as it is depicted in Fig II.1. Only downlink data transmissions are considered. Consequently, the user-plane data is generated at the eNB side and transmitted to the intended UEs. The UEs send control information including hybrid automatic repeat and request (HARQ) feedback in uplink. For

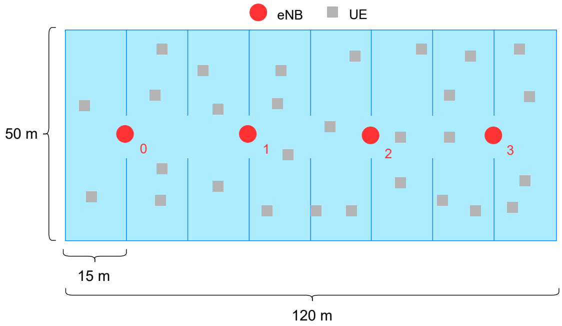


Fig. II.1: Indoor scenario layout.

downlink transmissions, the data intended to the UEs is multiplexed using orthogonal frequency-division multiple access (OFDMA) whereas, for uplink control information, the UE transmissions are multiplexed using block-interleaved frequency division multiple access (B-IFDMA) [8].

2.2 Traffic and frame structure

It is assumed that the packet generation follows a homogeneous Poisson point process with an average packet arrival rate of λ. The packet size is assumed to be fixed to B bytes. This type of traffic is also known as FTP-3 in 3GPP scope [9] [10]. In order to support different offered loads, the packet arrival rate is modified accordingly. Assuming K UEs deployed in the scenario, the offered load expressed in bits per second (bit/s) can be obtained by the following equation:

$$L = K \cdot B \cdot \lambda \cdot 8 \tag{A.1}$$

For transmission in unlicensed spectrum, we assume time division duplexing (TDD) with a very flexible frame structure where each subframe can, in principle, be downlink or uplink. The frame structure is defined based on the dynamic conditions of the buffer's status at the nodes. Additionally, the use of a partial downlink subframe as a starting subframe is allowed. A partial subframe consists of 7 OFDM symbols and allows the eNB to start a frame in the second slot of the subframe. This is adopted due to the different time resolution between the CCA and the LTE time slot. LBT can finish in the middle of an LTE slot and, therefore, a mechanism for achieving frame alignment is needed.

The transition subframe between downlink and uplink subframes, also known as special subframe, is based on the following structure: a partial ending subframe (DwPTS), a guard period and a shortened uplink subframe. This short uplink subframe is known in MulteFire terminology as

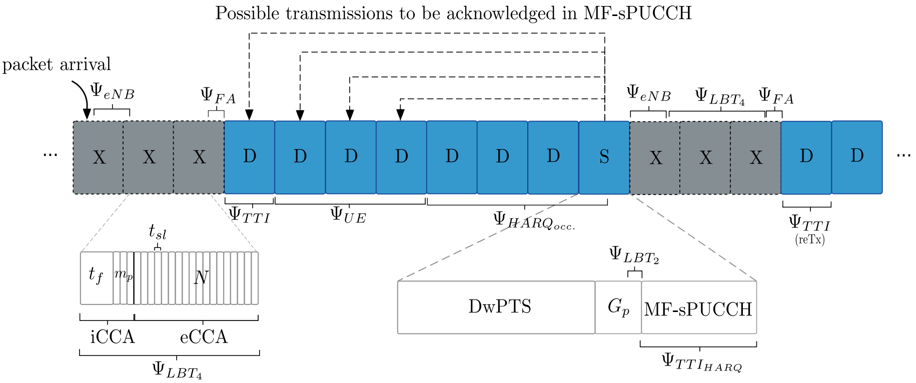


Fig. II.2: Frame structure and LBT procedure schematic.

MF-sPUCCH (MulteFire-short PUCCH). It is used for carrying uplink control information such as scheduling request or HARQ feedback. This feature follows the specifications included in the release 1.0 of MulteFire [6]. The frame structure, the partial subframe and the design of the MF-sPUCCH are depicted in Fig. II.2. Further details included in Fig. II.2 are presented in the following sections.

2.3 Listen Before Talk

In order to be compliant with the regulatory requirements in the 5 GHz unlicensed spectrum, each node must assess the availability of the channel by means of LBT prior to any transmission. LBT is a contention-based protocol that allows devices deployed in the unlicensed spectrum to share the radio channel without pre-coordination. It is based on an energy-threshold detection algorithm and it is performed in a CCA slots basis. Each CCA slot has a duration of 9 µs. According to 3GPP specifications [5], an eNB operating in the unlicensed band that wants to transmit shall first sense the channel to be idle during a defer duration (t_d). The defer duration consists of a fixed portion of time (t_f) with a duration of 16 µs, immediately followed by m_p CCA slots (t_{sl}). During a CCA slot, the channel is considered to be idle if the energy detected by the eNB is lower than a certain threshold for at least 4 µs. This is also known as initial CCA (iCCA).

After this preliminary procedure, the extended CCA (eCCA) starts. During the eCCA, the channel is continuously observed for a duration of at least N CCA slots, where N is a random variable uniformly distributed between zero and the maximum value of the contention window size. While performing the eCCA procedure, if a CCA slot is sensed as idle, the eNB decreases the value of the CCA counter (which is initialized to the value N). If a CCA slot is sensed as busy, the eNB enters the back-off mode, also known as deferral mode. While in back-off mode, the eNB cannot decrease the CCA counter. The eNB shall continuously sense the channel as idle for a period equal to the defer time t_d before it can exit the back-off mode. The LBT is considered successful when the CCA counter equals zero. After a successful LBT, the eNB can transmit and occupy the channel for a duration equal to the maximum channel occupancy time (MCOT). This algorithm is commonly known as category 4 LBT.

3GPP has established 4 different channel access priority classes where several parameters related to LBT are defined [5] (Tab. 15.1.1.1-1). Based on the priority class, the maximum value that N can take are delimited. Among the possible values that the upper bound of N can take, the minimum value is initially chosen. In the following transmission occasions, an adjustment procedure for the contention window is performed. On it, if at least 80 % of the HARQ feedbacks received for a given subframe are negative, then, the upper bound is increased to the next possible value defined in the channel access priority class [5]. The priority class also specifies m_p, which is the number of CCA slots that are observed during the iCCA, and MCOT per channel access.

For uplink transmissions within the eNB acquired MCOT, it is assumed that every UE perform a category 2 LBT. This is also known as Type B LBT in 3GPP terminology [5]. With category 2 LBT, the UE shall sense the channel for a sensing interval of 25 µs prior to any UL transmission. If the channel is sensed as idle, the UE can start the transmission. Otherwise, the UE shall wait until its next transmission opportunity. In accordance with the MulteFire specifications [6], a UE can skip the category 2 LBT and transmit immediately if the MF-sPUCCH transmission starts no later than 16 µs after the end of the downlink transmission. This feature will be employed in our work in order to evaluate the impact of category 2 LBT on the packet delay.

2.4 Hybrid Automatic Repeat and Request

In order to handle possible failures in the decoding process, HARQ is assumed. By means of HARQ, receivers can request a retransmission if needed. Due to the flexible frame structure and the uncertainty in the channel availability, adaptive asynchronous HARQ is considered for downlink transmissions. Unlike synchronous HARQ where retransmissions for certain HARQ process can only occur at specific time occasions, asynchronous HARQ assumes that retransmissions may occur at arbitrary time instants. In downlink, the HARQ feedback is transmitted by the UEs in MF-sPUCCH resources. The delay components of HARQ can be split as follows. The processing time at the UE, that is, the time needed from the UE receives a downlink transmission until it can send the corresponding HARQ feedback, is 4 subframes.

Regarding the eNB, the processing time (from eNB receives the HARQ feedback until it can either consider the transmission acknowledged or it can reschedule a retransmission) is 2 subframes. As it is depicted in Fig. II.2, due to the delay constraints, only the downlink data transmitted up to the subframe $k^{th} - 4$ can be acknowledged in the k^{th} subframe. The HARQ feedback related to the pending downlink subframes that cannot be acknowledged in the current MCOT need be transmitted in the next transmission opportunity.

2.5 Objective

Given the outlined system model (in coherence with MulteFire), our objective is to analysis the URLLC-alike performance for downlink packet transmissions. That is, we study the achievable one-way latency at high-reliability levels on the order of 10^{-4}. The latency, in downlink, is defined as the time from a packet arrives at the eNB until it is correctly decoded and forwarded to higher layers at the UE side. Our hypothesis is that the LBT procedure plays an important role in the latency budget, so we first analyze it by means of simple theoretical assessments. Secondly, we further quantify the performance by using advanced dynamic system-level simulations with a high degree of realism.

3 Latency analysis

Throughout this section, an analysis of the overall delay per packet transmission is presented. Given that the emphasis of our research is on LBT, a probabilistic study of the delay caused by the mandatory LBT procedure is presented. From a packet is generated and stored in the node's buffer at layer 3 until the receiver is able to decode it, the delay can be split into several components. These components are depicted in Fig. II.2. Assuming that the packet is successfully decoded at first transmission, the following equation contains the contributions of the overall packet delay for a downlink transmission:

$$\Psi_{TX} = \Psi_{eNB} + \Psi_{LBT_4} + \Psi_{FA} + \Psi_{TTI} + \Psi_{UE}, \tag{A.2}$$

where Ψ_{LBT_4} corresponds to the delay associated with the category 4 LBT performed by the eNB, Ψ_{FA} is the delay spent due to the frame alignment, Ψ_{TTI} is the time used for the transmission and Ψ_{eNB} and Ψ_{UE} corresponds to the processing time used at the transmitter and receiver to prepare and decode the packet, respectively.

Some of the terms in Eq. (A.2) are fixed. Specifically, the duration of Ψ_{TTI} and the processing time at both ends. However, the delay associated with LBT and frame alignment are random variables. The LBT delay depends on the selection of N and the sensed interference level. The frame alignment

(FA) delay is conditioned to when the eNB successfully finishes the LBT. As it was introduced in Section 2.2, there are two opportunities to start a frame (1st and 7th OFDM symbol of each subframe). Therefore, Ψ_{FA} follows a uniform distribution between 1 and 7 symbols.

Assuming a non-zero probability of having an erroneous reception at first transmission, HARQ mechanism is triggered. The time needed by the UEs to transmit the HARQ feedback and have it processed at eNB side can be expressed as follows:

$$\Psi_{HARQ} = \Psi_{UE} + \Psi_{HARQ_{occ.}} + \Psi_{LBT_2} + \Psi_{TTI} + \Psi_{eNB}, \tag{A.3}$$

where in this case Ψ_{UE} is the time spent by the UE preparing the HARQ feedback and Ψ_{eNB} is the time used at eNB side to decode it. Both processing times and Ψ_{TTI} are considered to be fixed. $\Psi_{HARQ_{occ.}}$ corresponds to the time that a UE needs to wait until there is a HARQ feedback opportunity, that is, the MF-sPUCCH resources. Regarding the delay spent performing category 2 LBT, Ψ_{LBT_2}, it is fixed to 25 µs. The time transmission interval (Ψ_{TTI}) for sending the HARQ feedback is also fixed with a duration equal to the MF-sPUCCH resources (4 OFDM symbols).

Assuming that q retransmissions are needed to successfully decode a given packet, the overall delay can be expressed as:

$$\Psi_{delay} = \Psi_{TX} + q\Big[\Psi_{HARQ} + \Psi_{TX}\Big] \tag{A.4}$$

Regarding the LBT procedure, a Markov chain is used to model the state transitions that an eNB experiences while category 4 LBT is performed. We assumed that with probability p the channel will be sensed as idle within a CCA slot while with $(1-p)$ the channel is considered as busy. This probability is affected by several factors such as the node position, radio channel conditions and the number of nodes competing for the channel. Here, we neglect the contribution of the additional defer time that needs to be performed after sensing the channel as busy in a CCA slot. The average time spent by the eNB for sensing each of the N CCA slots as idle is given by the following expression [11]:

$$\begin{aligned} t_{backoff} &= \lim_{n\to\infty}\Big[p\cdot t_{sl} + p(1-p)\cdot 2t_{sl} + p(1-p)^2\cdot 3t_{sl} + \\ &\quad + p(1-p)^3\cdot 4t_{sl} + ... + p(1-p)^{(n-1)}\cdot nt_{sl}\Big] \\ &= \frac{t_{sl}}{p}, \end{aligned} \tag{A.5}$$

where n is the number of required CCA slots needed for sensing the channel as idle and t_{sl} is the duration of a CCA slot. Since the contention window size

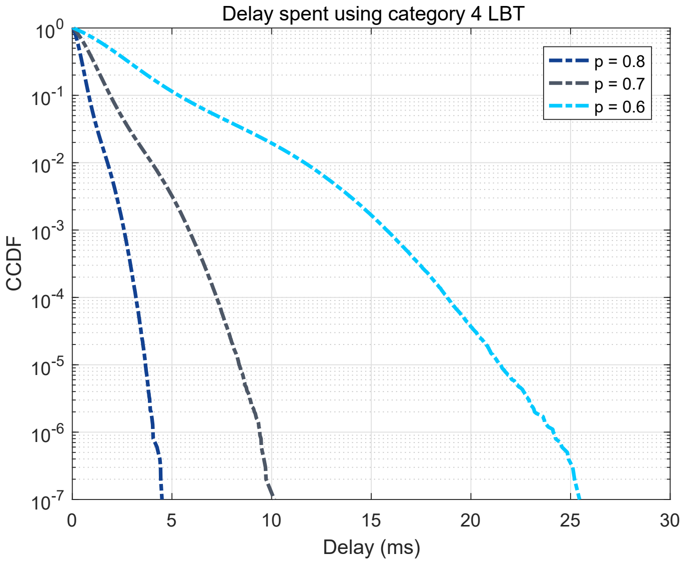


Fig. II.3: CCDF of the delay associated with the cateogry 4 LBT.

follows a uniform distribution limited by 0 and CW_{max}, the average value of N is:

$$\bar{N} = \frac{CW_{max}}{2} \tag{A.6}$$

Taking the equation (A.5) and the definition of iCCA included in Section 2, the average time spent in performing the category 4 LBT can be expressed as:

$$\Psi_{LBT_{Cat.4}} = \underbrace{t_f + m_p \cdot t_{sl}}_{iCCA} + \underbrace{\bar{N} \cdot t_{backoff}}_{eCCA} \tag{A.7}$$

Assuming LBT with channel access priority class 3, we have performed a Matlab simulation where the delay spent in performing category 4 LBT is studied. Based on different probabilities to sense the channel as idle, p, we establish a comparison among the different cases. We assume that the probability of sensing the channel idle and scenario load are closely related. Therefore, we use the terms low, average and high load when referring to high, average and low probability of sensing the channel idle respectively. The values of p employed during the simulation are 0.8, 0.7 and 0.6. As the priority class 3 defines [5], three different values for the upper bound of the contention window are used $\{15, 31 \text{ and } 63\}$. Since the selection of this value is based on the percentage of negative acknowledgment (NACK) transmissions received for a given k^{th} subframe, we model it based on p. We assume an approach where there is a relation between p and the rate of NACK received. For the low load scenario, the probability of choosing the lowest value for CW_{max} is higher and vice-versa for the high load scenario. The set of probabilities used in the numerical model for selecting the 3 possible values for CW_{max} are $\{0.8, 0.15, 0.05\}$, $\{0.75, 0, 2, 0.05\}$ and $\{0.7, 0.2, 0.1\}$ for the low, average and high load respectively. The complementary cumulative distribution function (CCDF) of the time spent performing the category 4 LBT is shown in Fig. II.3.

Meaningful information can be extracted from Fig. II.3. First, it can be concluded that for high load situations, the eNB spends more time performing LBT as it can be deducted from Eq. (A.5). Moreover, different behaviours are observed depending on the probability of sensing the channel idle. For the low load case, the probability of sensing the channel idle in N consecutive CCA slots at the first attempt is higher. Consequently, the CCDF is steeper as compared with the two other cases since it is unlikely that the node enters in back-off mode. For average and high load scenario, the delay associated with category 4 LBT is notably increased. The reasoning for this performance is the fact that every time the channel is sensed as busy, the node enters in a deferral mode where it needs to wait for an additional interval of time of at least 43 µs before it can start decreasing the CCA counter. Moreover, in highly loaded scenarios, the fact that the probability of choosing higher values for CW_{max} is high makes LBT a more time-consuming process.

This analysis of the category 4 LBT delay is based on fixed probabilities and it does not take into consideration the different dynamics present in a real system. In order to better understand the behaviour of LBT in a real scenario, we employ a detailed system-level simulator.

4 Simulation methodology

We consider an indoor scenario which follows the guidelines of 3GPP [10]. It consists of 4 base stations and 50 users randomly placed within the scenario. In order to make the system-level simulations in-line with 3GPP NR assumptions for URLLC traffic [3], we assume a packet with a payload of 50 B. Different traffic loads are supported during the simulations. To achieve it, several values of the generated traffic are obtained by means of modifying the packet arrival rate (λ) of the Poisson process.

Detailed system-level simulations with subcarrier-OFDM symbol resolution inlined with 3GPP and MulteFire assumptions are carried out. The simulator models the majority of the radio resource management techniques such as time-frequency domain scheduling and HARQ. Moreover, channel quality indicator (CQI) transmissions for performing an accurate link adaptation are also implemented. The selection of the modulation and coding scheme (MCS)

Table II.1: Simulation details

Parameter	**Assumption**
Layout	Indoor model (3GPP TR 36.889, Annex A.1.1 [10])
Channel model	International Telecommunications Union (ITU) Indoor Propagation Model
Frame configuration	TDD, MCOT: 8 ms, Bandwidth: 20 MHz
TTI	14 OFDM symbols, 15 kHz sub-carrier spacing
Scheduling metric	Proportional Fair max. scheduled transmissions per TTI: 10
Scheduling type	DL: physical resource block based UL: interlace based
HARQ	Asynchronous HARQ, chase combining 6 retransmissions at maximum
Link adaptation	OLLA enabled, BLER target 1 %
Traffic model	B = 50 bytes, $\lambda = \{20, 50, 100\}$ packets/s
Channel Access	Category 4 LBT with priority class 3 Energy threshold: -72 dBm

is based on it and it is adjusted to fulfil the block error rate (BLER) by means of the outer-loop link adaptation (OLLA) [12]. The rest of the simulation assumptions are summarized in Table II.1.

5 System-level performance

5.1 Impact of category 4 LBT

Fig. II.4 shows the CCDF of the packet delay for different loads for a full downlink scenario. It is observed that when the packet arrival rate is increased the overall delay per packet is also increased. Among other relevant aspects such as retransmissions or queuing delay, this behaviour is due to the likelihood that category 4 LBT senses the medium as busy while trying to access the channel. In that case, the eNB enters in back-off and it needs to sense the channel free for an additional deferral time before it can start decreasing the CCA counter. Consequently, especially in high load cases, the overall delay is strongly impacted by LBT.

In Fig. II.5 we show the CCDF of the time spent by the eNBs performing the category 4 LBT. As it was presented in Section 3, the time used to perform the LBT is inverse proportional to the probability of sensing the channel as

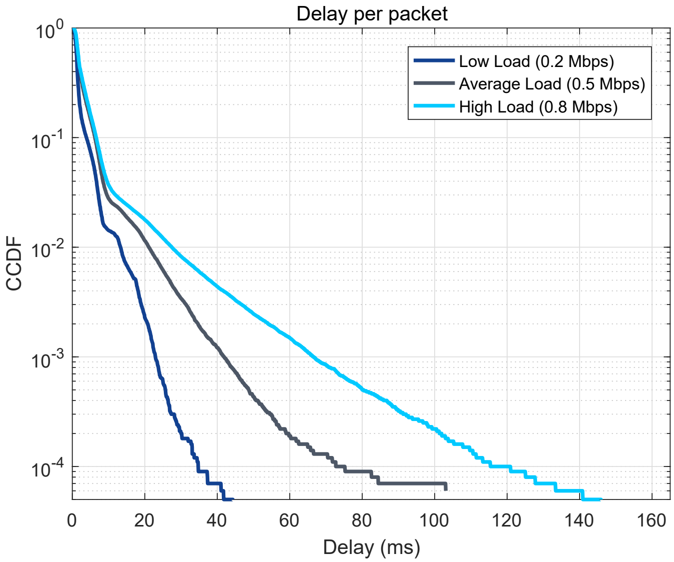


Fig. II.4: CCDF of the overall delay per packet.

idle. In other words, it is proportional to the amount of traffic generated in the scenario. Results from Section 3 are also included in the figure. The probabilities of sensing the channel as idle used in the numerical model are 0.8, 0.7 and 0.65 for low, average and high load, respectively. Comparing Fig. II.4 and Fig. II.5, the impact of category 4 LBT on the overall packet delay can be deducted.

To evaluate the impact of category 4 LBT on the overall packet delay, we introduce a new metric defined as the ratio of the n^{th} percentile of the time spent doing cat 4 LBT to the n^{th} percentile of the overall packet delay. This metric is used to estimate the portion of the packet delay which is spent doing category 4 LBT. Table II.2 summarizes the results obtained for different loads. It can be noticed that LBT has a reasonable impact on the overall delay, e.g., for highly loaded scenario more than 40 % of the packet delay is spent in performing category 4 LBT in 10 % of the cases.

5.2 Impact of category 2 LBT

LBT also impacts the overall delay of the system if the UE transmission is blocked due to category 2 LBT failure. Assuming a full downlink traffic operation, after the downlink transmission, the eNB waits for the HARQ feedback by the UE. However, especially in highly loaded scenarios, the UE may not able to transmit the HARQ feedback due to the failure in the category 2

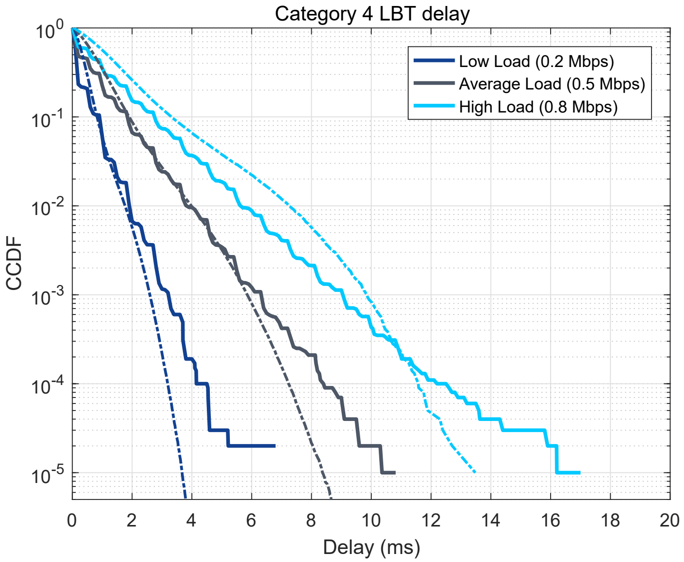


Fig. II.5: CCDF of delay associated with the category 4 LBT. Solid lines correspond to the results obtained from system-level simulator. Dotted lines show the delay obtained using the numerical model described in Section 3.

Table II.2: Impact of category 4 LBT on the overall packet delay

Scenario Load	**LBT delay** [ms] **/ Packet Delay** [ms] (%)	
	90^{th} percentile	99^{th} percentile
Low load (0.2 Mbit/s)	0.7 / 3.9 (18.0 %)	1.9 / 13.3 (14.3 %)
Average load (0.5 Mbit/s)	1.85 / 6.2 (29.4 %)	3.8 / 21.1 (18.0 %)
High load (0.8 Mbit/s)	2.75 / 6.6 (41.7 %)	5.6 / 27.5 (20.4 %)

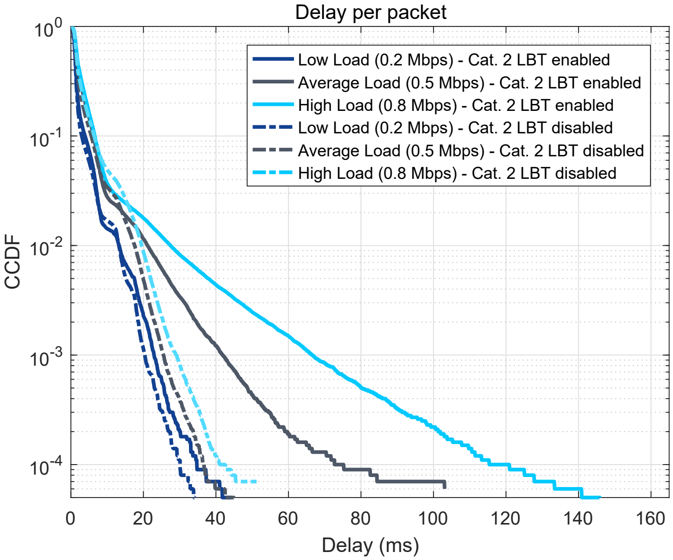


Fig. II.6: CCDF of overall delay per packet. Solid lines correspond to the case when UEs perform category 2 LBT before sending HARQ feedback whereas dotted lines show the packet delay when category 2 LBT is disabled.

LBT. In this case, the eNB is not aware of whether the UE was able to decode the message or not. The eNB will treat the absence of HARQ feedback as a NACK and it will trigger a retransmission, even though, the UE already decoded successfully the packet. This situation will not impact the delay of the current packet transmission since it was already decoded at the UE, but it will have a side effect in the following transmissions. First, the eNB will compete for occupying the channel to transmit data that is not needed. Moreover, once the eNB is within its MCOT, the interference is also increased and, potentially, it does not allow the access to the channel to neighbour eNBs. Potential queuing delay might happen because part of the available resources are used inefficiently for the retransmissions. So that, it prevents the use of the available resources for new data transmissions. This assumption supposes a trade-off between latency and channel efficiency. As asynchronous HARQ is assumed the eNB could wait until, upon the LBT success, the UE is able to send the HARQ feedback. However, this approach will highly impact the packet delay. Therefore, preemptive retransmissions are assumed in case that no feedback is received after the first transmission opportunity.

In order to evaluate this effect, we establish a comparison between a scenario where the UEs need to perform category 2 LBT before sending the HARQ feedback and one where they are allowed to transmit the ACK/NACK avoiding LBT. As it was mentioned in Section 2.3, the possibility of skipping the category 2 LBT is included in MulteFire specifications [6]. The results are shown in the Fig. II.6. It can be observed that there is a latency reduction when the category 2 LBT used for transmitting the HARQ feedback is disabled for any of the loads considered. It is especially noticeable when the load is high since the probability of failing while performing the category 2 LBT is high. Using this approach the effect of the scenario's load in the delay is relaxed since now it only affects the category 4 LBT performed by the eNB.

6 Conclusions

We have analysed the impact of category 4 and 2 LBT in the overall packet delay in a full downlink scenario using system level simulations supported by a numerical analysis. In both cases, the main message is that when the load in the scenario is increased, the impact of LBT is increased. Regarding category 4 LBT, it has been demonstrated that highly impacts the overall packet delay preventing from achieving the URLLC requirements, e.g. LBT contributes with more than 40% in 10% of the cases when the load is high. For the category 2 LBT, we have evaluated the impact of LBT failure. It triggers unnecessary retransmissions that impact on neighbouring eNBs, by avoidable channel accesses, and eNB itself, by increasing its queuing delay. It has been shown that the solution adopted by MulteFire where the category 2 LBT is avoided is a good starting point to reach the URLLC requirements. Using this study as a baseline, the next step will be to add uplink traffic and study the impact on the latency and reliability performance by using different channel access priorities. In order to follow the NR approach, higher sub-carrier spacing and faster processing times will also be considered.

Acknowledgment

Part of this work has been performed in the framework of the Horizon 2020 project ONE5G (ICT-760809) receiving funds from the European Union. Views expressed in this contribution are those of the authors and do not necessarily represent the project.

References

[1] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, and B. Yin, "Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges," *IEEE Access*, vol. 6, pp. 6505–6519, 2018.

[2] D. A. Wassie, I. Rodriguez, G. Berardinelli, F. M. L. Tavares, T. B. Sorensen, and P. Mogensen, “Radio Propagation Analysis of Industrial Scenarios within the Context of Ultra-Reliable Communication,” in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, 2018, pp. 1–6.

[3] *Study on Scenarios and Requirements for Next Generation Access Technologies (Release 14) - 3GPP TR 38.913 V14.3.0*, Std.

[4] D. Laselva, D. Lopez-Perez, M. Rinne, and T. Henttonen, “3GPP LTE-WLAN Aggregation Technologies: Functionalities and Performance Comparison,” *IEEE Communications Magazine*, vol. 56, no. 3, pp. 195–203, 2018.

[5] *Physical layer procedures (Release 15) - 3GPP TS 36.213 V15.1.0*, 3rd Generation Partnership Project Std.

[6] M. Alliance, “MulteFire Release 1.0 Technical Paper: A New Way to Wireless,” Tech. Rep. [Online]. Available: https://www.multefire.org/

[7] *Study on NR-based Access to Unlicensed Spectrum (Release 15) - 3GPP TR 38.889 V0.0.1*, 3rd Generation Partnership Project Std.

[8] T. Svensson, T. Frank, D. Falconer, M. Sternad, E. Costa, and A. Klein, “B-IFDMA - A Power Efficient Multiple Access Scheme for Non-frequency-adaptive Transmission,” in *2007 16th IST Mobile and Wireless Communications Summit*, 2007, pp. 1–5.

[9] *Further advancements for E-UTRA physical layer aspects (Release 9) - 3GPP TR 36.814 V9.0.0*, 3rd Generation Partnership Project Std.

[10] *Study on Licensed-Assisted Access to Unlicensed Spectrum; (Release 13) - 3GPP TR 36.889 V13.0.0*, 3rd Generation Partnership Project Std.

[11] T. Tao, F. Han, and Y. Liu, “Enhanced LBT algorithm for LTE-LAA in unlicensed band,” in *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2015, pp. 1907–1911.

[12] K. I. Pedersen, G. Monghal, I. Z. Kovacs, T. E. Kolding, A. Pokhariyal, F. Frederiksen, and P. Mogensen, “Frequency Domain Scheduling for OFDMA with Limited and Noisy Channel Feedback,” in *2007 IEEE 66th Vehicular Technology Conference*, 2007, pp. 1792–1796.

Paper B

Uplink Ultra-Reliable Low Latency Communications Assessment in Unlicensed Spectrum

Roberto Maldonado, Claudio Rosa, Frank Frederiksen, Klaus I. Pedersen

The paper has been published in the
IEEE Global Communications Conference (GLOBECOM) Workshops, 2018.

Abstract

Through this paper, the performance of uplink transmissions in terms of latency and reliability in unlicensed spectrum is addressed. Our goal is to provide insights into how ultra-reliable and low latency communications behave over the unlicensed spectrum in standalone mode. Our system model is based on the MulteFire radio access technology and operates in the 5 GHz *unlicensed band. Grant-free is an alternative to scheduled uplink scheme aiming at reducing the latency of uplink transmissions. With grant-free uplink, the users can transmit their uplink data without needing a specific grant. In this paper, both grant-free and scheduled uplink are analyzed with a focus on their performance in unlicensed spectrum. Based on extensive system-level simulations, we show that the grant-free approach is also suitable for latency reduction in the unlicensed spectrum under certain load conditions. Nevertheless, scheduled uplink presents some advantages when the number of grant-free users transmitting simultaneously increases.*

1 Introduction

Future wireless communications systems are envisioned to support a wide range of type of communication. Among others, communications with high requirements in terms of latency and reliability are considered as key enablers for new use cases such as vehicular to vehicular communications or industrial control and automation. This type of communications, widely known as ultra-reliable and low-latency communications (URLLC), suppose a big challenge to the research community since the focus is no longer on the throughput performance rather on the reduction of latency under certain reliability constraints. Consequently, innovative techniques must be developed to fulfil the requirements established by the standardization bodies. Due to its potential application to the industry, URLLC is considered as one of the key enablers of the Industry 4.0 [1]. It will bring flexibility into the industry and it will allow mission-critical communication among the machinery which will boost the overall industry productivity. In this scope, unlicensed spectrum plays an important role. This free-of-use frequency band allows deploying private wireless infrastructure with a lower cost and easier deployments as compared to the licensed band. On the other hand, stringent regulations must be fulfilled in order to assure the coexistence among the technologies operating in the unlicensed band. Unlicensed spectrum has been already considered by 3rd Generation Partnership Project (3GPP) as a complementary resource that can boost the system performance. As part of Long Term Evolution (LTE), 3GPP specifies Licensed-Assisted Access (LAA) feature which leverages from the combination of licensed and unlicensed bands, by means of carrier aggregation. It provides the base stations with the possibility to offload part of its traffic to unlicensed spectrum. However, LAA needs an anchor cell in the licensed spectrum that provides quality-of-service (QoS) and reliability for critical communications. Thus, its application is limited to mobile network operators and mobile broadband services. With the goal of proving full operability using only the unlicensed spectrum, the MulteFire Alliance brought a solution called MulteFire [2]. This technology is based on LTE systems and it is capable of operating in the unlicensed spectrum in a standalone manner. Nowadays, the interest of 3GPP in the unlicensed spectrum has been reaffirmed in a recent study related to New Radio (NR) operation in unlicensed spectrum [3]. However, all the solutions studied in 3GPP and MulteFire focused on providing baseline access to unlicensed spectrum and did not consider the support of URLLC. As an alternative to the conventional scheduling request uplink approach, where every user needs to send a scheduling request message and receive a grant before the transmission can be initiated, grant-free uplink (GUL) scheme was proposed. GUL provides a contention based solution for the uplink transmissions where multiple users share the same preconfigured resources. It allows skipping the lengthy scheduling request process adopted in LTE systems, thus, it is considered a mechanism of reducing the latency. GUL has also been considered applicable to unlicensed spectrum. GUL, by means of skipping the transmission of the scheduling request message by the users and the corresponding grant by the eNBs, reduces the number of mandatory clear channel assessments that each node must perform prior any transmission in the unlicensed spectrum. It reduces the impact of the channel availability on the transmission and, thus, it will reduce the latency. The performance of uplink transmission using the unlicensed spectrum, comparing grant-free and scheduling based approach, is addressed in [4] and [5]. However, the analysis is mainly focused on the uplink throughput and Wi-Fi coexistence as the main key parameter indicators (KPIs). Grant-free uplink transmissions and URLLC has been analyzed in several papers in the licensed spectrum. In [6, 7], system level simulations are conducted to evaluate the latency and reliability of grant-free transmissions from different perspectives. To the best of the authors' knowledge, there is no literature available that analyze grant-free scheme in the unlicensed spectrum with a focus on the latency and reliability. In this paper, the uplink performance of a MulteFire system is addressed. Based on extensive system-level simulations, a comparison between scheduling request and grant-less uplink approaches in terms of latency and reliability is provided. The rest of the paper is organized as follows: Section 2 introduces the regulatory requirements of the unlicensed band and an overview of both scheduling-based and grant-less uplink operations where their strengths and drawbacks in terms of latency and reliability are highlighted. Section 3 describes the simulation methodology and the assumptions considered to conduct the system-level simulations. A comparison between both schemes is included in Section 6 whereas, in Section 7, conclusions are drawn.

2 System model

2.1 Scenario set-up

The system model assumed in this paper follows the guidelines provides in MulteFire standard. A single-operator indoor scenario with M evolved nodes-B (eNBs) and K URLLC user equipments (UEs) uniformly distributed along the scenario is assumed. Each node operates in 5 GHz frequency band in a time division duplex (TDD) manner with a bandwidth of 20 MHz. Uplink data transmissions, both information and control and downlink control transmissions are considered in the scenario. Data traffic is generated at UEs side following a homogeneous Poisson point process with an average packet arrival rate of λ. A fixed packet size of B bytes is adopted. The frame configuration is dynamically updated and the downlink-uplink heaviness depends on the buffers' status at the nodes. It is assumed that the M eNBs are synchronized in a subframe level, meaning that all the subframes start and finish at the same point in time for every eNB. However, each eNB selects its own frame configuration based on the buffer status of the nodes and the clear channel assessment outcome.

2.2 Regulatory requirements

In order to operate in the 5 GHz unlicensed band, a node must fulfil certain regulatory requirements and guarantee fair co-existence with other radio access technologies operating in the same band. In this paper, we assume compliance with the requirements that have been embraced by the 3GPP and MulteFire standardization bodies. These include restrictions on the transmission power and power spectral density. It is also required that every device must confine the 99 % of the signal energy within, at least, 80 % of the channel bandwidth. Therefore, 3GPP agreed in the introduction of a new uplink waveform called block-interleave frequency division multiple access (B-IFDMA) [8] as an alternative to single-carrier frequency domain multiplexing access (SC-FDMA) algorithm used in uplink LTE. The solution adopted spans the frequency domain allocation of each UE over the available transmission bandwidth. To do so, each UE is assigned with one interlace as minimum frequency domain allocation. An interlace is a set of Ω physical resource blocks (PRBs) where each of them is constantly equidistant in frequency from the rest. For a 20 MHz bandwidth, it is assumed that each interlace is formed by 10 PRBs ($\Omega = 10$) and, thereby, assuming a sub-carrier spacing of 15 kHz, 10 interlaces are available in each transmission time interval.

2.3 Channel access procedure

Another important requirement in 5 GHz unlicensed spectrum is the mandatory clear channel assessment (CCA) procedure which must be performed by every node prior to any transmission. Listen-before-talk (LBT) is the solution adopted by 3GPP and MulteFire. LBT is a contention-based protocol that allows devices deployed in the unlicensed spectrum to share the radio channel without pre-coordination. It is based on an energy-threshold detection algorithm and it is performed in a CCA slots basis (CCA_{int}) where each slot has a duration of 9 µs. Following the 3GPP and MulteFire specifications for the unlicensed spectrum [9], two different types of LBT are defined: category 4 LBT (Cat.4 LBT) and category 2 LBT (Cat.2 LBT). Cat.4 LBT must be performed by any node that needs to acquire access to the channel. It consists of 1) an initial CCA (iCCA) where the channel is sensed during a defer duration ($16\ \mu s + m_p \cdot CCA_{int}$) and, upon the success of the iCCA, 2) an extended CCA (eCCA) is performed. During the eCCA the channel is sensed during N consecutive CCA slots also known as contention window. N is a random uniformly distributed variable between 0 and CW_p. One regulatory restriction implicitly fulfilled with LBT is the the amount of time that a node can occupy the channel also known as maximum channel occupancy time (MCOT). It is defined as the maximum time that a node can transmit upon the success of a Cat.4 LBT. The MCOT duration, the possible values for the upper bound of N (CW_p), and the number of CCA slots that need to be sensed as part of the iCCA (m_p) vary based on the channel priority classes (p) defined in 3GPP and MulteFire standards. There is a direct relation between CW_p and the MCOT such that, for shorter values of the MCOT, the smaller is CW_p and, in principle, the faster is the LBT procedure. Regarding the Cat.2 LBT, it is limited to uplink transmissions and allows the UEs to share the MCOT previously initiated by the serving eNB by a successful Cat.4 LBT. With this type of LBT, a UE can transmit uplink data immediately after sensing the channel as idle for a sensing interval duration of 25 µs. Detailed information about the different types of LBT, channel access priority classes, procedure for N selection , etc. can be consulted in 3GPP Release 15 ([9], Section 15).

2.4 Scheduling-based approach

With the scheduled uplink (SUL) approach, a UE that intends to transmit uplink data needs to first get a grant from its corresponding serving eNB. Prior to the reception of the grant, the UE needs to inform to the eNB that it has data to transmit. To do so, the UEs transmit a scheduling request on dedicated physical uplink control channel (PUCCH) resources. A description of

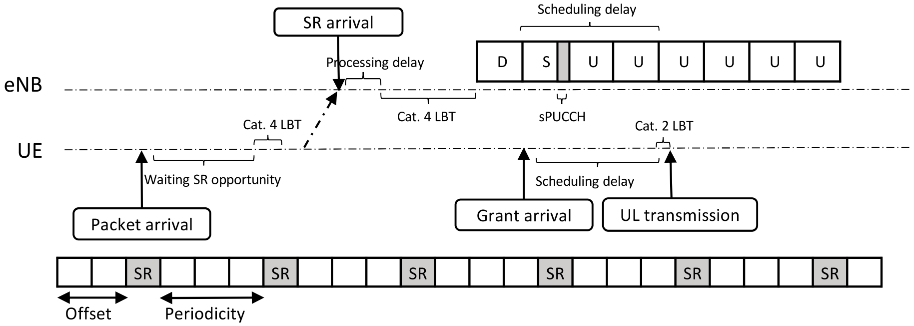


Fig. II.1: Scheduling request procedure. This figure depicts the case where there is no ongoing transmissions by the serving eNB.

the scheduling-based procedure is depicted in the Fig. II.1. The scheduling request (SR) dedicated resources are configured by certain periodicity and frame offset. Different configuration for offset and periodicity are defined in 3GPP and MulteFire specifications [9](Table 10.1.5-1). Thus, SR transmission instances are only possible in those uplink subframes that are defined by both variables. Besides, in the unlicensed spectrum, each UE must perform LBT before accessing the channel. On one hand, SR transmission may occur at pre-configured time occasions provided that there is not an ongoing transmission from/to the serving eNB. In this case, the SR transmission must be preceded by a successful Cat.4 LBT using channel access priority class 1. In case of an ongoing transmission from the serving eNB, the UE can transmit the SR when the preconfigured time occasions are in correspondence with the short PUCCH (sPUCCH). sPUCCH is a MulteFire-specific feature and it is defined as the last 4 OFDM symbols of the special subframe between the downlink and the uplink burst of the MCOT (see Fig. II.1). In this case, the SR transmission must be preceded by a successful Cat.2 LBT procedure.

Once the SR is transmitted and after the SR processing time at the eNB side, the eNB prepares the grant and performs a Cat.4 LBT before accessing the channel to transmit the grant. After the reception of the grant, the UEs need to decode the grant and prepare the data for transmission. This requires some processing time at the UE. In accordance with LTE and MulteFire specifications, we assume a minimum of 4 ms between the transmission of a grant in downlink and the corresponding transmission on pysical uplink shared channel (PUSCH) in uplink. Afterwards and, before the uplink transmission, the scheduled UEs need to perform a successful Cat.2 LBT procedure.

Applying this approach to transmissions with extremely demanding requirements in terms of latency is not considered a suitable solution. One of the drawbacks is the inevitable delay that constrains the UE to transmit the uplink data upon the reception of the grant. Therefore in an optimistic scenario, in which the processing at the eNB can be neglected and every subframe is considered a candidate to send the SR, a packet will experience, at least, 5 ms of delay. Moreover, employing the scheduling-based approach in unlicensed spectrum could be seen as less appropriate due to the multiple LBTs. The uncertainty in the outcome of the LBT will lead to increasing delay in the transmission. Another critical issue is the fact that, in case the UE fails to access the channel due to an unsuccessful Cat.2 LBT after the grant reception, the resources allocated for the PUSCH transmission are wasted. The serving eNB will detect that there was no transmission in the allocated resources and it will re-scheduled the same data by means of sending the grant again. This situation will lead to higher delays for uplink transmission and it also may block the transmission of neighbour cells since the channel must be occupied for a longer time.

2.5 Grant-free uplink approach

Due to the lengthy exchange of control information between UEs and eNBs before the uplink data transmissions and the multiple LBTs, the scheduling-based approach is not considered a good candidate for supporting communications with stringent latency requirements. As an alternative, grant-free uplink (GUL) scheme is considered as a new candidate for providing lower latency and lower signalling overhead in uplink transmissions. With GUL, a UE is allowed to transmit using periodic pre-configured resources at specific time occasions, here referred to as GUL resources, without transmitting SR and receiving any specific grant. Prior to the transmission on GUL resources, the UE needs to perform either Cat.2 or Cat.4 LBT depending on whether or not the transmission is within the MCOT previously acquired by the serving eNB. GUL resources are shared among GUL UEs, meaning that several UEs can transmit simultaneously in the same time-frequency resources. Advanced receiver techniques are considered to alleviate the impact of collisions in the decoding process.

In case that the receiver is not able to decode the GUL transmission due to low post-detection SINR for a given modulation and coding scheme (MCS), hybrid automatic repeat request (HARQ) is supported. Due to the uncertainty of the channel availability as a result of LBT, asynchronous HARQ is assumed. Each UE transmits in the uplink control information (UCI) different HARQ-related parameters such as the HARQ process ID or the redundancy version. Given a failure at decoding process, we assume that the eNB is always able to detect the UCI for each UE, based on the orthogonality achieved by UE-specific demodulation reference signals (DMRS), and schedule the GUL UE on reserved time-frequency resources. Using this approach, the impact of the main source of decoding failures in GUL approach, the collisions due to multiple UEs sharing the same resources, is relaxed since

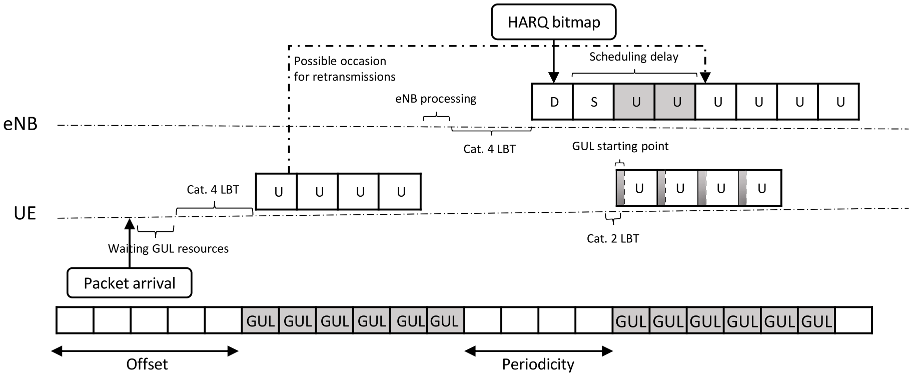


Fig. II.2: Grant-free procedure when there is no ongoing transmissions by the serving eNB. It is assumed that the GUL UEs are already configured.

retransmissions are handled as scheduled transmissions.

Partial bandwidth is the scheme adopted by GUL transmissions. With this scheme, GUL resources in the frequency domain are specified by frequency domain pools (FD-PL). Each FD-PL consists of 5 interlaces and each GUL UE is allowed to transmit in only one FD-PL. The even interlaces that is, interlace 0, 2, 4 and 8, are part of the FD-PL 0. On the other hand, the rest of interlaces compose the FD-PL 1. In GUL resources, the frequency domain scheduler allocates randomly any of the 2 FD-PL to each UE. Thus, every UE is allocated with 5 interlaces, or expressed differently, 50 PRBs. It is assumed that, after the successful LBT, the GUL UEs initiate their transmissions simultaneously at the beginning of the OFDM symbol 1 of the GUL subframe. In order to allow multiplexing capabilities between GUL and SUL UEs, cell-specific timing advance settings can be used to align the GUL and SUL starting points within the first OFDM symbol (see Fig. II.2). It is also possible to prioritize scheduled transmissions by means of defining different starting positions such that GUL UEs sense the channel right after the SUL transmission has started implying that GUL UEs may be blocked due to ongoing SUL transmissions.

3 Simulation methodology

We consider an indoor scenario which follows the guidelines of 3GPP [10]. It consists of 4 eNBs and 50 UEs randomly placed within the scenario. In order to make the system-level simulations in-line with 3GPP NR assumptions for URLLC traffic [11], we assume a packet with a payload of 50 bytes. Different traffic loads are supported during the simulations. To achieve it, several values of the generated traffic are obtained by means of modifying the packet arrival rate of the Poisson process which is used for modelling the packet generation. The eNBs are assumed to be equipped with minimum mean square error-interference rejection combining (MMSE-IRC) receivers. Ideally, it is also considered perfect estimation of the interference covariance matrix used for suppressing both intra-cell and inter-cell interference. Since it is assumed each eNB contains 2 receiving antennas, the interference contribution of one potential collision can be removed from the desired signal. The decoding probability is based on the signal-to-interference-plus-noise ratio (SINR) after the detection process. Exponential effective SINR mapping (EESM) [12] is the method adopted in the link to system mapping to estimate the effective SINR for a given MCS and a given set of per-subcarrier SINR values. Based on this set-up, detailed system level simulations with subcarrier-OFDM symbol resolution inlined with 3GPP and MulteFire assumptions are carried out. The simulator is able to model the majority of the radio resource management techniques such as time-frequency domain scheduling, HARQ, link adaptation or power control. It is also capable of emulating all the specific features related to unlicensed spectrum such as listen before talk or frequency allocation for uplink transmissions in interlaces basis. Specific simulations assumptions regarding both uplink transmissions schemes are listed bellow whereas general assumptions are specified in Table II.1.

3.1 Scheduling based approach

Cat.2 LBT is considered as the type of LBT used by the UEs to send the SR to the serving eNB. It is also assumed that the SR opportunities periodicity is 1 ms and the scheduling delay is set to 4 ms. The minimum allocation size in the frequency domain for uplink transmission is 1 interlace. We assume channel access priority class 3 parameters for the Cat.4 LBT performed by the eNBs. Thus, the MCOT has a duration of 8 ms and the contention window upper bound can take the values of 15, 31 and 63 CCA slots.

3.2 Grant-free uplink approach

GUL resources periodicity is set to 1 ms and it is assumed the MCOT for GUL UEs has a duration of 4 ms in case of absence of ongoing transmission by the serving eNB. Moreover, we adopt that the MCOT cannot be shared by the serving eNB so the GUL frame configuration only contains 4 uplink subframes. Due to shorter MCOT, Cat.4 LBT with high channel access priority class is considered. With this aggressive LBT, the contention window upper bounds can be 7 and 15 CCA slots. On condition that the eNB has previously acquired the channel, the GUL UEs can initiate the uplink transmissions by means of a Cat.2 LBT. In this case and in order to provide higher preference

Table II.1: Simulation details

Parameter	**Assumption**
Layout	Indoor model (3GPP TR 36.889, Annex A.1.1 [10])
Users	Uniformly distributed and static
Channel model	ITU Indoor Propagation Model
Frame configuration	TDD
TTI	14 OFDM symbols 15 kHz sub-carrier spacing
Scheduling metric	Proportional Fair max. scheduled transmissions per TTI: 10
Scheduling type	DL: physical resource block based UL: interlace based
Grant-free uplink	max. GUL UEs per TTI: 10; FD-PL: 2
HARQ	Asynchronous HARQ, chase combining 6 retransmissions at maximum
Link adaptation	Outer loop link adaptation: enabled Block error rate target: 1 %
Power control	$P_0 = -60\,\text{dBm}$, $\alpha = 0.8$
Traffic model	B = 50 bytes, $\lambda = \{50, 125, 250\}$ packets/s
Channel access	LBT energy threshold: $-72\,\text{dBm}$

to scheduled transmissions, e.g. retransmissions of failed GUL transmissions, an offset in the starting point of the GUL transmissions is assumed. Hence, scheduled UEs avoid the possible interference caused by GUL UEs and, therefore, the probability of success in the Cat.2 LBT is higher. It is considered that the GUL transmissions are deferred for 3 CCA slots, i.e. 27 µs, with respect to scheduled based transmissions. Each eNB assigns randomly a frequency domain pool for each of the GUL UEs and it remains fixed throughout the entire simulation. A robust modulation, i.e. QPSK with a coding rate of 1/10, is also considered as fixed.

4 System-level performance

Throughout this section, the results provided by the system-level simulator are presented and analyzed. A comparison in the latency performance between both uplink approaches is presented below. In the conducted simulations, all the UEs deployed in the scenario are using either scheduling based procedure, referred to as SUL, or GUL approach. The complementary commutative distribution function (CCDF) of the delay per packet for both approaches in different load conditions are shown in Fig. II.3. It can be observed that the performance achieved by GUL approach outperforms SUL when the uplink load is lower than 2.5 Mbit/s. This behaviour is due to the skipping of the scheduling process, which at least reduces the latency in 4 ms, and the relatively low scenario load that reduces decoding failures in the serving eNB due to multiple collisions. To a lesser extent, the usage of a high priority class for LBT is also impacting the delay. However, when the load is increased to a certain point, 5 Mbit/s in our case, the GUL scheme performs worse than the SUL approach as can be seen in the tail of the yellow curve in Fig. II.3. The same trend can be observed for the case of 2.5 Mbit/s, however, in this case, the GUL performance still exceeds the SUL approach.

This behaviour is mainly due to 2 reasons. Firstly, as the load increases, the number of GUL UEs competing for the channel increases. Therefore, the delay associated with the mandatory Cat.4 LBT (τ_{LBT}) increases accordingly with the load as can be noted in Fig. II.4. On it, the contribution of the Cat.4 LBT (Δ_{LBT}) to the overall packet delay (τ_{Total}) for different reliability values have been included. The LBT contribution is defined as the ratio of the n^{th} percentile of the time spent doing Cat.4 LBT to the n^{th} percentile of the overall packet delay. It can be noted that, with the load equal to 1 Mbit/s, 29 % and 48 % of the overall delay is due to the Cat.4 LBT for a reliability of $1-10^{-3}$ and $1-10^{-4}$, respectively. This implies that LBT heavily impacts the GUL delay performance in low load cases. On highly loaded scenarios, even though the highest delay corresponds to the highest load, the contribution of LBT only counts for 16 % and 12 % of the overall packet delay.

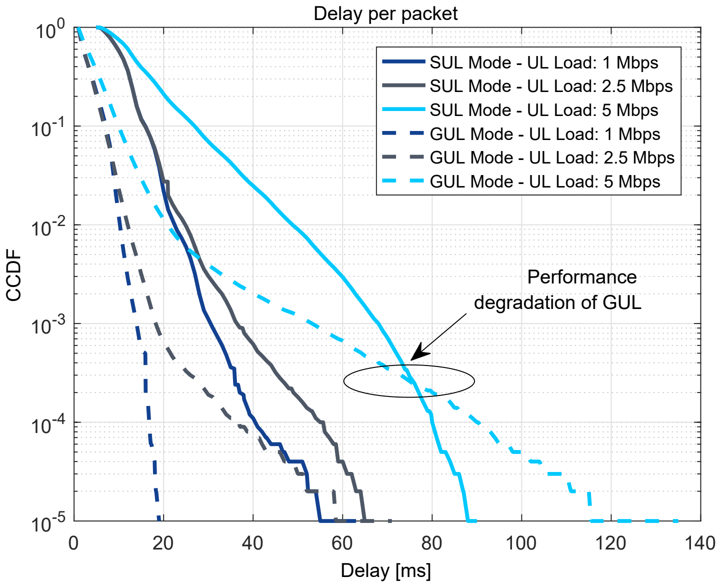


Fig. II.3: CCDF of the delay per packet for different scenario's loads. Solid lines represents the results provided by scheduling-based simulations whereas the dashed lines contains the results for GUL simulations.

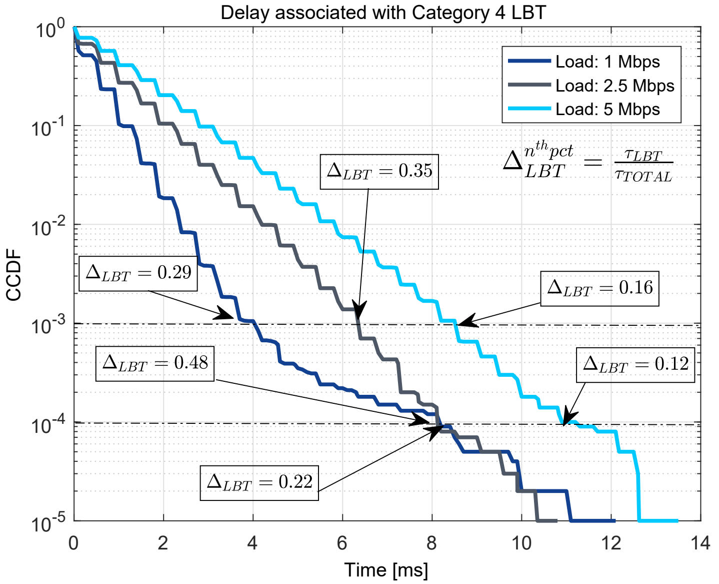


Fig. II.4: CCDF of the delay spent while performing the Cat.4 LBT by the GUL UEs for different loads. The contribution of Cat.4 LBT to the overall delay per packet for different reliability requirements is included.

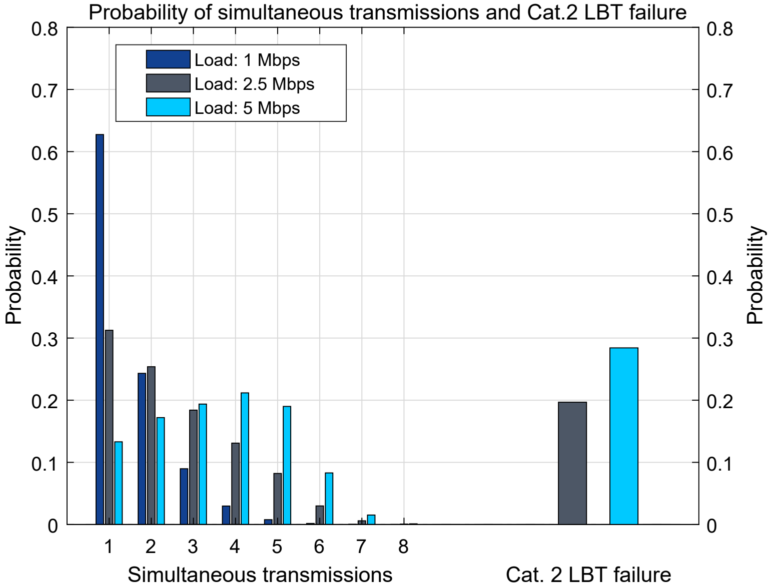


Fig. II.5: Probability of simultaneous transmissions within the same FD-PL (left side) and probability of failure at performing Cat.2 LBT by the GUL UEs (right side) for different scenario loads.

At high load scenarios, the aspect that mainly limits the GUL performance is the fact that multiple UEs are transmitting at the same time using the same resources. In such case, the serving eNB may not be able to decode the uplink information transmitted by the GUL UEs in the same frequency resources. In Fig. II.5, at the left-hand side, the probability of having k GUL UEs transmitting in the same FD-PL in the same TTI is included. It can be noted that the probability of having multiple UEs transmitting in the same GUL resources achieves a more evenly distribution when the load is increased. Therefore, even though we consider a robust MCS, the eNB may fail at decoding and it will schedule the GUL UEs for a retransmission. In that case, the GUL UEs need to process the received grant and prepare the retransmission which lasts 4 ms. Moreover, the UEs must perform a Cat.2 LBT before initiating the retransmission. An unsuccessful Cat. 2 LBT may happen due to other GUL UEs simultaneous transmitting while the sensing is performed. It will cause an inefficient usage of the uplink resources since the UL transmission cannot be carried out. Moreover, it will trigger an additional retransmission that will increase the packet delay. In Fig. II.5, at right-hand side, the probability of failure for the Cat.2 LBT performed by the GUL UEs when needs to send a retransmission. At load equal to 1 Mbit/s, the eNBs are able to handle the collisions in most of the cases and, in case of needed retransmissions, the UEs always sense the channel as free prior the starting point. For 2.5 Mbit/s and 5 Mbit/s, the probability of failure is not negligible which impacts in the overall packet delay.

5 Conclusions

Throughout the paper, we have analysed two different schemes for uplink transmissions under the perspective of latency and reliability. It has been shown by means of detailed system level simulations that, under certain conditions, the grant-less uplink approach provides lower delays as compared to scheduling uplink. Specifically, for uplink loads up to 2.5 Mbit/s, GUL approach turns to be the preferable scheme to use for uplink transmissions. The skipping of the control exchange before the uplink data transmission, the reduction of the number of LBTs and the ability to handle possible collisions by the eNBs make the difference between GUL and SUL. At certain load, 5 Mbit/s in our scenario, the eNBs are not capable to handle the numerous simultaneous transmissions from the GUL UEs, making SUL approach outperforms GUL.

References

[1] B. Chen, J. Wan, L. Shu, P. Li, M. Mukherjee, and B. Yin, “Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges,” *IEEE Access*, vol. 6, pp. 6505–6519, 2018.

[2] M. Alliance, “Multefire release 1.0 technical paper: A new way to wireless,” Tech. Rep. [Online]. Available: https://www.multefire.org/

[3] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Study on NR-based Access to Unlicensed Spectrum (Release 15),” 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.889, 2018.

[4] R. Karaki, A. Mukherjee, and J. Cheng, “Performance of Autonomous Uplink Transmissions in Unlicensed Spectrum LTE,” in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6.

[5] J. Zhang, W. Chang, H. Niu, S. Talarico, and H. Yang, “Grant-less uplink transmission for LTE operated in unlicensed spectrum,” in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2017, pp. 1–6.

[6] T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, P. Mogensen, I. Z. Kovacs, and T. K. Madsen, “System Level Analysis of Uplink Grant-

Free Transmission for URLLC," in *2017 IEEE Globecom Workshops (GC Wkshps)*, 2017, pp. 1–6.

[7] R. Abreu, T. Jacobsen, G. Berardinelli, K. Pedersen, I. Z. Kovács, and P. Mogensen, "Power control optimization for uplink grant-free URLLC," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, 2018, pp. 1–6.

[8] T. Svensson, T. Frank, D. Falconer, M. Sternad, E. Costa, and A. Klein, "B-IFDMA - A Power Efficient Multiple Access Scheme for Non-frequency-adaptive Transmission," in *2007 16th IST Mobile and Wireless Communications Summit*, 2007, pp. 1–5.

[9] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Physical Layer Procedures," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.213, 2018.

[10] *Study on Licensed-Assisted Access to Unlicensed Spectrum; (Release 13) - 3GPP TR 36.889 V13.0.0*, 3rd Generation Partnership Project Std.

[11] *Study on Scenarios and Requirements for Next Generation Access Technologies (Release 14) - 3GPP TR 38.913 V14.3.0*, ETSI Std.

[12] R. Giuliano and F. Mazzenga, "Exponential effective SINR approximations for OFDM/OFDMA-based cellular system planning," *IEEE Transactions on Wireless Communications*, September 2009.

Paper C

Latency and Reliability Analysis of Cellular Networks in Unlicensed Spectrum

Roberto Maldonado, Claudio Rosa, Klaus I. Pedersen

The paper has been published in the
IEEE Access, 2019.

Abstract

In this paper, the achievable latency-reliability performance of a standalone cellular network over the 5 GHz *unlicensed spectrum is analysed. Fulfilling strict latency-reliability requirements comes with significant challenges for unlicensed operation, especially due to mandatory channel access procedures. Using MulteFire as the reference system-model, an analysis of a highly realistic multi-cell network with bidirectional traffic shows that latency of* 23 ms *with a reliability level of* 99.99 % *is achievable for low-loads, while latency is increased to* 79 ms *at high-loads. Different techniques are described to improve the system performance. First, a pre-emptive scheme to cope with continuous uplink listen before talk (LBT) failures for uplink control transmissions is proposed. It provides a latency reduction of* 24 % *at low-loads with two transmission opportunities and* 11 % *for high-loads with three opportunities. Secondly, the possibility of skipping LBT performance under given conditions is evaluated. This results in a lower uplink LBT failure rate which translates to a latency reduction of* 8 % *for low-loads and up to* 14 % *for high-loads, at* 99.99 % *reliability. Thirdly, as an alternative to grant-based uplink, grant-free uplink is evaluated. Grant-free uplink achieves better performance than grant-based uplink at low-loads, offering* 50 % *lower uplink latency. At high-loads, the gain of grant-free uplink decreases due to the high number of simultaneous transmissions.*

1 Introduction

With the arrival of 5G New Radio (5G NR), plenty of novel use cases with diverse requirements are envisioned to be supported. The 3rd Generation Partnership Project (3GPP) has defined use cases and the corresponding requirements for enhanced mobile broadband (eMBB) in [1], mission-critical communications in [2] and massive Internet of Things (IoT) in [3]. Among the use cases defined in the mission-critical communications domain, industrial automation is foreseen as one of the most relevant use cases for private networks. Industrial automation relies on reliable and broadband connectivity to lead to the next stage in the industrial revolution, commonly known as Industry 4.0. The next industrial revolution aims to improve factory plants and production lines in four main aspects: efficiency, flexibility, usability and versatility. In [4], 3GPP describes several vertical domains use-cases including the Factories of the Future. The 5G Alliance for Connected Industries and Automation (5G ACIA) proposes industrial-specific use-cases and its corresponding requirements in [5]. Both entities agree on the need for supporting stringent requirements in latency and reliability for communications between machines, objects and people as a key enabler for the Industry 4.0.

Unlicensed spectrum is considered a valuable asset to be used by cellular technologies, especially for private network deployments. Unlicensed bands are characterised by being a global solution with low cost of operation and larger available bandwidth at below 7 GHz as compared to licensed spectrum. The 3GPP's interest in unlicensed band started with Licensed-Assisted Access (LAA). In LAA, unlicensed bands are jointly operated with an anchor cell deployed in a licensed band offering the possibility of offloading traffic to unlicensed bands [6] [7]. An independent organisation from 3GPP, the MulteFire Alliance, recently designed a system capable of operating in standalone mode in the unlicensed spectrum. This technology is bsaed the Long Term Evolution (LTE) specifications and it is known as MulteFire [8]. Research on unlicensed spectrum operation has continued by 3GPP during the 5G-NR development. As part of the current Release 16, 3GPP aims to design New Radio-Unlicensed (NR-U) as a single and global solution for unlicensed spectrum access based on NR specifications [9]. The considered unlicensed frequencies for LAA, MulteFire and NR-U are located in the 5 GHz frequency band. Currently, there are discussions about extending the frequency ranges for unlicensed operation. Regulatory entities in the United States, Europe and other parts of the world are considering the possibility of opening the 6 GHz band (5.925 GHz to 7.125 GHz) [10]. Frequency bands between 57 GHz and 71 GHz, i.e. millimetre waves, are also candidates [11] [12]. However, none of these bands are yet available for real deployments. The focus of LAA and MulteFire technologies and the first design of NR-U is to support communications with non-critical quality-of-service (QoS) requirements. Further optimisations enabling the support of demanding QoS, such as low-latency and high-reliable communications in unlicensed bands, are expected to be included in future 3GPP releases, i.e Release 17.

As compared to licensed solutions, meeting tight latency-reliability requirements in unlicensed bands brings additional non-trivial problems that need to be properly addressed. Only few available studies address the performance of latency-critical traffic with high-reliability constrains for unlicensed bands in standalone mode. Examples include [13], where the impact of listen before talk (LBT) for downlink-only traffic is analysed based on extensive system-level simulations following the MulteFire design. Similarly, the LBT influence on the uplink-only latency performance is studied in [14], including a latency-reliability comparison between grant-based and grant-free uplink schemes. As a non-standalone alternative, authors in [15] propose tight cooperation between unlicensed and licensed bands to meet the latency-reliability targets. Licensed spectrum is only used under conditions in which unlicensed spectrum represents a bottleneck for the latency-reliability performance. All these studies share a common message: channel access procedures are found to cause increased latencies as transmissions are frequently postponed due to high measured interference-levels. Therefore, supporting tight latency-reliability requirements for standalone unlicensed spectrum systems remains a challenge that calls for more research and development.

In this paper, we further study the latency-reliability performance of a private network deployment operating over the 5 GHz unlicensed band in standalone mode following the MulteFire model. We study advanced cases with time-variant bursty bi-directional traffic. A performance analysis is conducted under realistic conditions for a multi-cell and multi-user system-level. Results are obtained from a highly detailed state-of-the-art system-level simulator. We present a solid baseline performance for latency-critical and high reliable traffic for the 5 GHz unlicensed spectrum that goes beyond results available in the existing open literature. This includes analysis of how the performance varies with different parameters such as the offered traffic, as well as the impact of LBT over the total packet latency. Based on the established baseline performance, and the identified bottlenecks for achieving high reliable and low latency performance, we present multiple promising guidelines and enhancements to further optimise the performance. In particular, we show that by providing users with additional occasions for hybrid automatic request and repeat (HARQ) transmissions achieves significant latency reductions. Secondly, great latency gains at high offered load are obtained by avoiding the LBT performance during the downlink-to-uplink transitions. Finally, attractive uplink latency reductions are achieved by using grant-free as compared to scheduled transmission, especially at low to medium loads.

The paper is organised as follows: the 5 GHz regulatory aspects for unlicensed operation are described in Section 2. The system model definition is included in Section 3, while Section 4 outlines the suggested proposals for achieving improved latency-reliability performance. Performance evaluation is provided in Section 5. Finally, concluding remarks appear in Section 7.

2 Regulatory requirements over 5 GHz band

The regulatory requirements for the 5 GHz unlicensed band vary depending on the region and the specific sub-band. Therefore, to be worldwide deployable, a radio access technology operating in unlicensed spectrum needs to fulfil the most stringent regulatory requirements among those standardised in various regions of the world. Besides, it needs to ensure fairness towards other co-existing radio access technologies deployed in the same frequency band. Consequently, the harmonized standard developed by the European Telecommunications Standard Institute (ETSI) is used to define the minimum requirements that 3GPP-based radio access technologies need to follow to operate in the 5 GHz band. These requirements include limitations on the transmitted power and power spectral density. There are also restrictions on the occupied channel bandwidth. The occupied channel bandwidth should be at least 80 % of the nominal channel bandwidth. Additional information about ETSI regulations can be found in [16].

2.1 Channel access procedures

In order to guarantee a fair coexistence among the different radio access technologies deployed over the 5 GHz band, each node is mandated to assess the availability of the channel before any transmission. The channel access mechanism adopted by 3GPP is based on a clear channel assessment (CCA) procedure that uses LBT in compliance with the ETSI regulations. LBT is a contention-based protocol that allows devices to use the same radio channel without pre-coordination. It is based on an energy detection (ED) threshold mechanism performed over intervals of 9 µs, also known as CCA slots (T_s). During each CCA slot, a node detects the channel activity based on power measurement and posterior comparison with a predefined ED threshold. The medium is declared as idle during a CCA slot if, at least, the measured power is lower than the threshold for 4 µs. Transmission is conditional on the device sensing the channel as idle for a certain number of CCA slots. Upon the idleness declaration, a device is only allowed to occupy the channel for a limited duration of time. LBT is also used by other radio access technologies deployed in the 5 GHz unlicensed band such as the IEEE 802.11 standard (Wi-Fi) which ensures fair coexistence among them [17] [18].

Two types of LBT procedures are standardized. The so-called Category 4 (Cat4) LBT implements a random backoff and a variable contention window size algorithms. Cat4 LBT consists of 1) an initial CCA (iCCA) where the channel is sensed during a defer period ($T_d = 16\,\mu\text{s} + m_p \cdot T_s$) and, upon the success of the iCCA, 2) an extended CCA (eCCA). During the eCCA, the transmitter generates a random number from a uniform distribution defined over the contention window size. This number represents the minimum number of CCA slots the channel needs to be sensed as idle before transmitting. The transmitter can subsequently use the channel for a maximum time, also known as the maximum channel occupancy time (MCOT). The contention window size varies based on the number of unsuccessful and successful transmissions on the medium. Different values for the contention window sizes, the MCOT durations and m_p are classified into channel access priority classes (CAPC) [19]. The so-called Category 2 (Cat2) LBT, also known as single-shot LBT, defines a type of channel access procedure in which nodes can initiate a transmission after sensing the channel to be idle for a fixed duration of 25 µs. The 25 µs interval is split into a 16 µs interval, which contains a CCA slot and an idle slot of 7 µs, and an additional CCA slot. Channel is declared as free if it is sensed as idle during both CCA slots.

Cat4 LBT is used by initiating nodes to gain access to the channel. Cat2 LBT can be used by responding nodes to initiate transmissions within the previously acquired channel occupancy time (COT) by an initiating device.

This is also known as COT sharing. The duration of the transmission after a successful Cat2 LBT is defined by the initiating node and it is limited by the MCOT. Furthermore, within the COT, the regulations also allow responding devices to initiate a transmission without performing LBT. The condition to skip LBT is fulfilled when the gap between the end of the transmission by the initiating device and the start of the transmission by the responding device is shorter than 16 µs. This is known as Category 1 (Cat1) LBT. Further details about the channel access mechanism considered by 3GPP as assumed in our analysis can be found in [19] [13].

3 System model

The system model assumed throughout the paper comprises a single-operator indoor scenario with M evolved Nodes-B (eNBs) and K user equipments (UEs) uniformly distributed within the building. Each node operates over the 5 GHz frequency band with a bandwidth of 20 MHz. Bi-directional traffic is generated following a homogeneous Poisson point process with an average packet arrival rate of λ_T expressed in packets/s/UE. Payloads in downlink are assumed to be generated with an average rate of λ_{DL}, while UEs generate uplink packets with an average arrival rate of λ_{UL}. Both packet arrival rates constitute the overall packet arrival rate per UE (λ_T):

$$\lambda_T = \lambda_{DL} + \lambda_{UL} \tag{C.1}$$

Given the model and assuming a fixed payload size of B bytes, the offered load (L) expressed in bit/s is defined as:

$$L = \lambda_T \cdot K \cdot B \cdot 8 \tag{C.2}$$

Dynamic time domain duplexing (TDD) is assumed. The frame configuration is dynamically updated and its downlink-uplink ratio, in terms of subframes, is adjusted based on the buffers' status at the nodes, i.e. the instantaneous traffic variation. Slot-level synchronization among the nodes is assumed whereas the frame configuration is node specific. Unless explicitly mentioned, it is always assumed that eNBs are the only node capable of starting a channel occupancy time. A single switching point between downlink and uplink subframes within the COT is supported. The transition subframe between downlink and uplink, also referred to special subframe, contains a partial ending subframe, a guard period and a short uplink subframe. The short uplink subframe, referred throughout the text as short-physical uplink control channel (sPUCCH), comprises the last 4 OFDM symbols of the special subframe and it is used for uplink control signalling transmissions, such as scheduling request or HARQ feedback [18].

3.1 Downlink operation

For downlink (DL) initiated transmissions, eNBs are considered as initiating devices. Therefore, a successful Cat4 LBT must be performed before transmitting. Due to the uncertainty about when LBT finishes, each eNB is configured to have two opportunities within a subframe to start the transmission. Specifically, OFDM symbols 0 and 7 are the candidates' starting positions. As compared to licensed operation, the additional starting position at the OFDM symbol 7th reduces the time between LBT finishes and the transmission starts, lowering the probability of channel access by a neighbour node. Upon the reception of the downlink data, the UE processes it and prepares the HARQ feedback that must be sent back to the eNB. HARQ feedback transmissions can be performed over sPUCCH or granted uplink resources.

Based on this model, the end-to-end delay of a downlink packet correctly received at first transmission can be expressed as:

$$\Psi_{DL} = \max\left[\Psi_{eNB}^{prep.}, \Psi_{Cat4}\right] + \Psi_{FA} + \Psi_{TTI} + \Psi_{UE}^{decod.}, \tag{C.3}$$

where Ψ_{Cat4} corresponds to the delay associated with the Cat4 LBT performed by the eNB and Ψ_{FA} accounts for time the eNB needs to wait until the next starting position, also known as frame alignment. Ψ_{TTI} models the transmission time interval (TTI) duration and $\Psi_{eNB}^{prep.}$ and $\Psi_{UE}^{decod.}$ corresponds to the processing time at the transmitter and receiver to prepare and decode the packet, respectively. It is assumed that eNBs prepare the data and perform Cat4 LBT in parallel.

Once the UE decodes the downlink packet, it needs to report the HARQ feedback to the eNB. The delay associated with this process equals:

$$\Psi_{HARQ} = \Psi_{HARQ}^{occ.} + \Psi_{Cat2} + \Psi_{TTI} + \Psi_{eNB}^{decod.}, \tag{C.4}$$

where $\Psi_{HARQ}^{occ.}$ corresponds to the time spent by the UE while waiting for a HARQ feedback occasion, and Ψ_{Cat2} models the fixed delay associated with the Cat2 LBT, i.e. 25 µs. The processing delay at the eNB side is represented by $\Psi_{eNB}^{decod.}$.

Assuming that q retransmissions are needed before the packet is received correctly by the UE, the overall packet delay equals to:

$$\Psi_{total} = \Psi_{DL} + q\left[\Psi_{HARQ} + \Psi_{DL}\right] \tag{C.5}$$

Due to the dependency between LBT and the instantaneous measured interference, uncertainty over whether transmissions will be carried out is added. As a baseline, it is assumed that eNBs only provide UEs with one opportunity for transmitting the HARQ feedback, that is, the sPUCCH resources or dynamically scheduled uplink resources. In case of Cat2 LBT blocking at

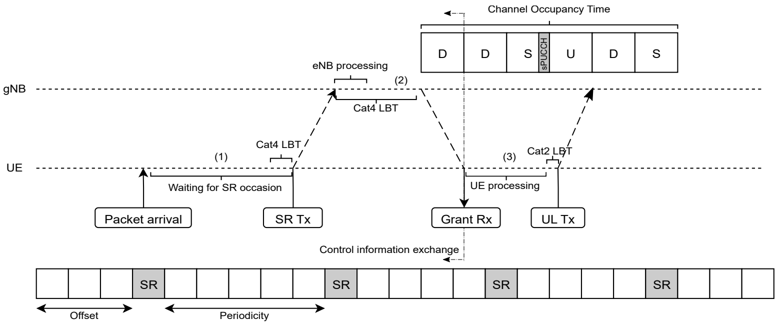


Fig. II.1: Grant-based uplink operation in unlicensed spectrum.

UE side, the control information transmission is blocked. eNB assumes the absence of HARQ feedback as negative feedback which automatically triggers a retransmission. In such a case, the round trip time (RTT), i.e. the time from the packet transmission until the acknowledgement is received at the transmitter side, is defined as follows:

$$\Psi_{RTT} = (r+1)\Big[\Psi_{DL} + \Psi_{HARQ}\Big], \tag{C.6}$$

where r accounts for the number of missing HARQ opportunities due to Cat2 LBT blocking. Note that, in these equations, the contribution of the queuing delay is neglected and equal processing times for new data transmissions and retransmissions are assumed.

3.2 Uplink operation: grant-based uplink

Uplink (UL) transmissions can be performed using grant-based (GB) scheduling. By means of GB uplink, a UE is capable of transmitting data over a specific set of resources granted by its serving eNB. Operation according to GB uplink is depicted in Figure II.1. Before any uplink data transmission takes place, UEs and eNBs need to go through a control information exchange to agree on which time-frequency resources will be used. This handshake process is as follows:

1. First, the UE transmits a scheduling request (SR) message in which it requests resources to be used for a new uplink data transmission. SR transmissions are performed over specific physical uplink control channel (PUCCH) resources, i.e. SR-PUCCH, which are configured by higher layers with certain periodicity and offset. The specific resources are used, provided that they do not collide with physical downlink shared channel (PDSCH) or physical uplink shared channel (PUSCH) transmissions. In such cases, the UE attempts the SR transmission in the next SR occasion.

2. Upon SR message reception and after a eNB-specific processing time, the eNB sends a grant in which it dynamically schedules the UE in uplink.

3. The grant is decoded by the UE and the PUSCH preparation starts. The time between the grant is received and the UE is capable of transmitting the PUSCH is known as scheduling delay. Assuming LTE processing capabilities, the scheduling delay is 4 ms.

As shown in Figure II.1, applying GB uplink for unlicensed operation implies the performance of multiple channel access procedures. The type of LBT used depends on the current COT conditions. For SR transmissions, UEs can either use Cat4 LBT with high CAPC or Cat2 LBT depending on the COT sharing conditions. For uplink grant transmissions, the eNB is considered as an initiating node and, therefore, it needs to perform a Cat4 LBT. The eNB grants permission to the UE to use COT sharing, and after the corresponding UE processing time, the PUSCH transmission occurs upon a successful Cat2 LBT. Furthermore, to fulfil the bandwidth regulations defined in Section 2, a new waveform is designed as alternative to single-carrier frequency domain multiplexing access (SC-FDMA) used for GB uplink in licensed bands. The adopted solution, known as block interleaved frequency division multiple access (B-IFDMA) [20], spans the frequency domain allocation of each UE over the available transmission bandwidth. Each UE is assigned with one interlace as minimum frequency domain allocation. An interlace is a set of M frequency equidistant physical resource blocks (PRBs).

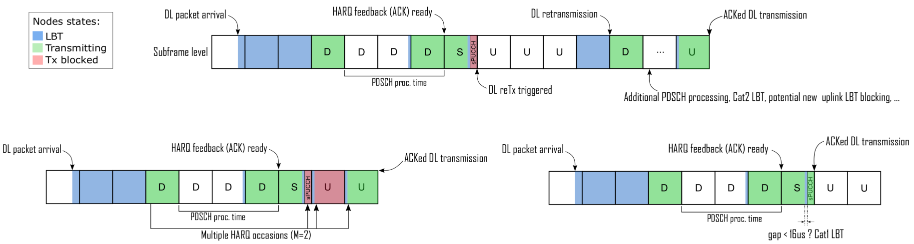
Regarding the delays involved in GB uplink, the time spent in the control information exchange can be expressed as:

$$\Psi_{SR} = \max\left[\Psi_{UE}^{prep.}, \Psi_{Cat4}, \Psi_{SR}^{occ.}\right] + \Psi_{TTI} + \Psi_{eNB}^{decod.} + \max\left[\Psi_{eNB}^{prep.}, \Psi_{Cat4}\right] + \Psi_{FA} + \Psi_{TTI}, \tag{C.7}$$

where $\Psi_{SR}^{occ.}$ defines the time spent while waiting for SR-PUCCH resources. In (C.7), it is assumed that UE is not in COT sharing conditions and, therefore, Cat4 LBT is performed prior to the SR transmission. Given (C.7), the end-to-end delay for an uplink packet correctly received at first transmission equals:

$$\Psi_{UL} = \Psi_{SR} + \Psi_{UE}^{prep.} + \Psi_{Cat2} + \Psi_{TTI} + \Psi_{eNB}^{decod.}, \tag{C.8}$$

where $\Psi_{UE}^{prep.}$ represents the preparation time of the uplink transmission after the grant reception at the UE side, i.e. the scheduling delay and $\Psi_{eNB}^{decod.}$ models the eNB processing time of the uplink packet at the eNB side. In this equation, successful Cat2 LBT has been assumed. Note that throughout these equations the contribution of the queuing delay is neglected.



(a) Baseline mode of operation

(b) Multiple HARQ transmission opportunities

(c) Category 1 LBT

Fig. II.2: Baseline mode of operation in (a) and latency reduction proposals. Multiple HARQ feedback opportunities in (b) and Cat1 LBT in (c).

4 Latency reduction proposals

Three different latency reduction solutions are presented in this section. The first two methods aim to mitigate the impact of Cat2 LBT failure when transmitting HARQ feedback for previously received downlink transmissions. The latter proposal presents a technique to reduce the uplink delay by means of alleviating the two main problems related to GB uplink: the multiple channel accesses needed in the control information exchange and the intrinsic scheduling delay.

4.1 Multiple HARQ feedback opportunities

As introduced in Section 3.1 and shown in Figure II.2(a), a failure in the Cat2 LBT prior to a HARQ feedback transmission triggers additional retransmissions. Additional downlink transmissions might be unnecessary as the UE could have decoded correctly the packet, i.e ACK message is ready to be transmitted, but it was not able to access the channel within the specific resources due to LBT blocking. Unnecessary retransmissions will cause: 1) additional interference in the system which may delay neighbour's transmissions since they could be blocked by LBT, 2) a reduction in the system resource efficiency as no additional information is transmitted in subsequent transmissions and 3) the performance of multiple successful Cat4 LBT. Moreover, it increases the queuing delay for new incoming packets, since they will not be served until previous transmissions are correctly acknowledged. As expressed in (C.6), the RTT delay is highly impacted by the number of missed HARQ feedback opportunities. Additionally, a discontinuous transmission (DTX) detection by the eNB can potentially increase the Cat4 LBT contention window size which, in turn, implies larger channel access delays [19].

To cope with this problem, a solution in which UEs are provided with multiple and consecutive opportunities for transmitting HARQ feedback is proposed. Thereby, each eNB is in charge of signalling, via the downlink control channel, the resources over which UEs can transmit the HARQ feedback. As depicted in Figure II.2(b), a UE using this scheme will be allowed to transmit ACK/NACK feedback in $M + 1$ occasions during the following uplink burst. Through this procedure, the probability of being blocked by Cat2 LBT failure is reduced and, therefore, the number of unnecessary retransmissions, i.e. the component r in (C.6), is also decreased. By comparing Figures II.2(a) and II.2(b), a shorter RTT is achieved with the proposed solution. It is worth noticing that this scheme offers a trade-off between Cat2 LBT reduction and resource efficiency as pre-emptive resources are reserved for potential HARQ feedback transmissions that may not be used if Cat2 LBT succeeds earlier.

4.2 Category 1 LBT

This approach aims to reduce the Cat2 LBT blocking probability, and consequently the potential downlink retransmissions, by avoiding the LBT performance. As defined in Section 2.1 and depicted in Figure II.2(c), a UE can start an uplink transmission without performing Cat2 LBT if the gap between the last downlink transmission and the start of uplink transmission is shorter than 16 µs. Given the system model assumptions, the only transmission in which UEs can leverage from this advantage is the HARQ feedback transmission performed over sPUCCH resources. If an eNB is capable of transmitting during the partial ending downlink of the special subframe, the UEs which have HARQ feedback from previous downlink receptions ready for transmission can access the channel without performing LBT.

4.3 Grant-free uplink

GB presents two main drawbacks from a delay perspective. Firstly, it needs to go through a lengthy control information exchange prior to the actual uplink data transmission, as shown in Figure II.1 and (C.7). In addition to this, uplink transmissions are constrained by a fixed delay upon the reception of the grant, i.e. the scheduling delay. Moreover, employing GB uplink in the unlicensed spectrum could be seen as less appropriate, due to the performance of multiple LBTs which adds uncertainty over the transmission performance. For instance, a UE may fail at accessing the channel due to a Cat2 LBT blocking prior to the PUSCH transmission. This will lead to a decrease in resource efficiency as the granted resources will not be used. In such a case, the eNB will reschedule the UE for a second attempt transmission. This situation will lead to higher delays for uplink packets and it also may block the transmission of neighbour cells since the channel needs to be occupied for longer periods.

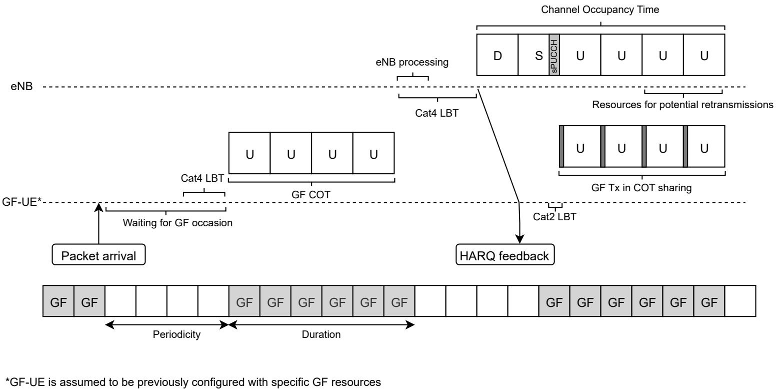


Fig. II.3: Grant-free uplink operation in unlicensed spectrum.

In order to mitigate the disadvantages of GB uplink, grant-free (GF) uplink is proposed. A description of the GF mode of operation is shown in Figure II.3. As compared with GB uplink, GF transmissions allow UEs to perform uplink transmissions avoiding the handshake process used in GB uplink. To achieve that, grant-free UEs (GF-UEs) are configured in a way that uplink transmissions occur over predefined and, potentially, shared resources among multiple UEs. In the time domain, the specific resources are given by the serving eNB based on periodicity and a duration. For frequency domain, GF-UEs are assumed to be configured by the serving eNB with a combination of interlaces, also known as frequency domain pool (FD-PL).

Analytically, the configuration of the GF resources substitutes the delay-prone SR procedure defined in (C.7) by:

$$\Psi_{GF}^{config} = \max\left[\Psi_{eNB}^{prep.} + \Psi_{Cat4}\right] + \Psi_{FA} + \Psi_{TTI} + \Psi_{UE}^{decod.} \tag{C.9}$$

where $\Psi_{eNB}^{prep.}$ and $\Psi_{UE}^{decod.}$ model the delay for preparing and decoding the GF configuration at eNB and UE, respectively. Once the UE is configured with GF resources, this process can be avoided for subsequent uplink transmissions as the configuration remains valid for certain interval of time. This differs from GB uplink in which each uplink transmission needs to go through the SR procedure.

Given the prior knowledge over the grant-free resources, GF-UEs can prepare and start their uplink transmissions without requesting a specific grant. This speeds up the uplink transmissions as the scheduling delay, i.e. $\Psi_{UE}^{prep.}$ in (C.8), is neglected. Moreover, the number of channel access involved in the process is reduced by a factor of 3 as it can be noted by comparing Figures II.1 and II.3. This reduction in the number of required LBTs will lower the transmission delays. GF-UEs support the possibility of starting a grant-free transmission without sharing the COT with the serving eNB. In this case, Cat4 LBT must be performed as the GF-UE is the initiating device. In the case of COT sharing between the eNB and GF-UEs, Cat2 LBT is supported.

The end-to-end delay of an uplink packet correctly received at first transmission when using GF approach can be expressed as:

$$\Psi_{GF} = \max\left[\Psi_{UE}^{prep.}, \Psi_{Cat4}, \Psi_{GF}^{occ.}\right] + \Psi_{TTI} + \Psi_{eNB}^{decod.}, \tag{C.10}$$

where $\Psi_{GF}^{occ.}$ models the time spent while waiting for GF resources. Here, the contribution of Ψ_{GF}^{config} is neglected as it is assumed that the GF-UEs were previously configured. The reduction in the number of channel access between GB and GF is observed by comparing (C.7)-(C.8) and (C.10).

The main disadvantage of GF-uplink is the lack of coordination inherited from a non-grant based approach. Therefore, collisions between several UEs, i.e. multiple transmissions using the same time-frequency resources, may occur. This will potentially have an impact on the receiver side as it may not always be able to decode correctly the uplink information coming from multiple sources. In order to cope with failures in the decoding, two mechanisms are proposed. The first one relies on HARQ protocol to request retransmissions to those UEs that the eNB was not able to decode. Since multiple collisions between GF-UEs are the main source of failure at GF-transmissions decoding, eNBs will send specific grants providing dedicated resources to UEs. Additionally, different starting transmission points can be defined at the transmitter side to reduce the collisions between grant-based UEs and GF-UEs. This is achieved by introducing an eNB-configured offset. By applying that, GF transmissions can be deferred for a duration up to 1 OFDM symbol, assuming an offset duration of 34 µs in our system model. By using this approach, and considering that GB transmissions start at the slot boundary, a higher priority is given to scheduled transmissions as potential LBT blocking by GF-transmissions is avoided. This mode of operation reduces the multiplexing of GB and GF transmissions within the same TTI.

5 Performance evaluation

5.1 Simulation methodology

Our indoor scenario follows the 3GPP guidelines for LAA simulation studies [6], which consists of a single-floor indoor office with an area of 120 m × 50 m and several walls separated by 15 m as depicted in Figure II.4. In order to emulate a private deployment, single operator conditions are assumed. A total of 4 eNBs are deployed over the scenario with a separation equal to 30 m. 50 UEs are randomly placed in the scenario. Each UE selects its serving

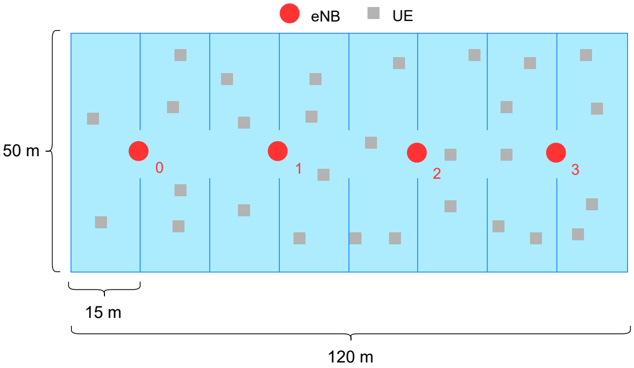


Fig. II.4: Scenario layout compliant with 3GPP guidelines for LAA performance evaluation.

eNB based on the strongest reference signal received power (RSRP) criteria. Performance evaluations are carried out using different offered traffic loads. A fixed packet size is assumed and the packet arrival rate (λ_T) is modified. The payload size is set to 50 B and λ_T takes values from the set: {25, 50, 125, 250} packets/s/UE, which corresponds to the offered loads of: {0.5, 1, 2.5, 5} Mbit/s, respectively. Bidirectional traffic is assumed, 80 % of the overall traffic is generated at eNBs side ($\lambda_{DL} = 0.8\lambda_T$), whereas the remaining 20 % is generated at the UEs ($\lambda_{UL} = 0.2\lambda_T$).

For GB uplink operation, it is assumed that SR opportunities occur every subframe, i.e. 1 ms periodicity. It also assumed that the eNB delays for decoding the SR message and generating the grant are neglected. Grant transmissions are only performed over downlink subframes upon the acquisition of the channel. The PUSCH preparation time equals to 4 TTIs, i.e. 4 ms. Frequency domain scheduling is performed in an interlace basis. Based on the defined bandwidth of 20 MHz and the 15 kHz sub-carrier spacing, each interlace is formed by 10 PRBs which results in a total of 10 available interlaces per TTI. For GF uplink operation, a GF resources periodicity of 1 ms is assumed. Frequency resources are configured in advance by the serving eNB, where each UE is signalled with a frequency domain pool (FD) consisting of 5 interlaces. Frequency domain pool-1 (FD-1) includes the interlaces {0, 2, 4, 6 and 8} whereas FD-2 consists of the interlaces {1, 3, 5, 7 and 9}. Additional simulation assumptions can be found in Table VI.2.

The simulator models with high-level of details the majority of the PHY and MAC functionalities and procedures in line with 3GPP guidelines. It dynamically schedules users in time and frequency domains, uses HARQ in case of decoding failures or performs link adaptation based of channel quality indicator (CQI) reports to fulfil the target block error rate (BLER). Moreover, all the regulatory channel access aspects for operating on the 5 GHz unlicensed band are carefully modelled. The simulator operates on symbol-

Table II.1: Simulation assumptions

Parameter	**Assumption**
Layout	Indoor single floor 3GPP TR 36.889, Annex A.1.1 [6]
Channel model	ITU Indoor Hotspot
Duplexing mode	Dynamic TDD
Bandwidth	20 MHz, 15 kHz sub-carrier spacing
TTI	14 OFDM symbols
Scheduling metric	Proportional fair Max. scheduled UEs per TTI: 10
Scheduling type	DL: physical resource block based UL: interlace based
HARQ	Asynchronous HARQ, incremental redundancy 6 retransmissions at maximum Processing delay at eNB: 2 TTIs Processing delay at UE: 4 TTIs
Link adaptation	Outer link loop adaptation: enabled [21] Block error rate target: 1 %
Receiver type	LMMSE-IRC [22]
MIMO	2×2 configuration DL: Rank-2 SU-MIMO UL: Rank-1 SU-MIMO, receiver diversity
Traffic model	B = 50 bytes $\lambda_T = \{25, 50, 125, 250\}$ packets/s/UE
Channel Access (Initiating node: eNB)	Cat4 LBT with CAPC 3 MCOT: 8 ms Contention window sizes = $\{15, 31, 63\}$
Channel Access (Initiating node: GF-UE)	Cat4 LBT with CAPC 1 MCOT: 4 ms Contention window sizes = $\{7, 15\}$
LBT ED threshold	−72 dBm

level and subcarrier resolution. For each transmission, the SINR at the receiver is calculated for each subcarrier symbol, assuming a linear minimum mean square error with interference rejection combining (LMMSE-IRC) receiver [22]. Inspired by the model in [23] [24], the SINR values are mapped to the mutual information domain, taking the applied modulation scheme into account. The mean mutual information per coded bit (MMIB) is calculated as the arithmetic mean of the values for the subcarrier symbols of the transmission [24]. Given the MMIB and the used modulation and coding rate of the transmission, the error probability is determined from look-up tables that are obtained from extensive link level simulations. Furthermore, it includes proven stochastic models for radio propagation, calibrated against alike models used in 3GPP system level simulations. In order to get statistically stable and reliable results, multiple realizations of the scenario are simulated. For each realization, the UE locations are selected independently and sufficient samples are collected. Results from each realization are combined afterwards.

5.2 Simulation results

In this section, we highlight the improvement in latency-reliability that the aforementioned proposals can provide to an LTE-like standalone system in unlicensed spectrum. Proposals performance are compared against MulteFire baseline approach. MulteFire baseline assumes the usage of single HARQ feedback opportunity, Cat2 LBT in any conditions and grant-based uplink. In order to improve the readability at very high percentile, such as 99.99^{th} percentile, the main key performance indicators (KPI) are represented using empirical complementary commutative distribution functions (CCDF).

Downlink

Figure II.5 shows the CCDF of the downlink delay per packet when using multiple HARQ feedback opportunities. Three different cases are compared. For the baseline case, i.e. with $M = 0$, it is assumed that a Cat2 LBT failure in sPUCCH or granted resources for HARQ transmissions it is translated into a downlink retransmission. $M = 1$ and $M = 2$ refer to the cases where the serving eNB provides 1 or 2 additional opportunities to transmit the ACK/NACK feedback, respectively. It is noted that an improvement is achieved when using this scheme for loads from 1 Mbit/s to 5 Mbit/s. At 0.5 Mbit/s load, the generated traffic is low enough that the downlink delay performance is not impacted by the additional retransmissions. This is shown in the CCDF, as $M = 1$ and $M = 2$ do not provide better performance as compared to baseline assumptions. In fact, the scheme is performing slightly worse from a downlink delay point of view. This is due to the fact that when providing

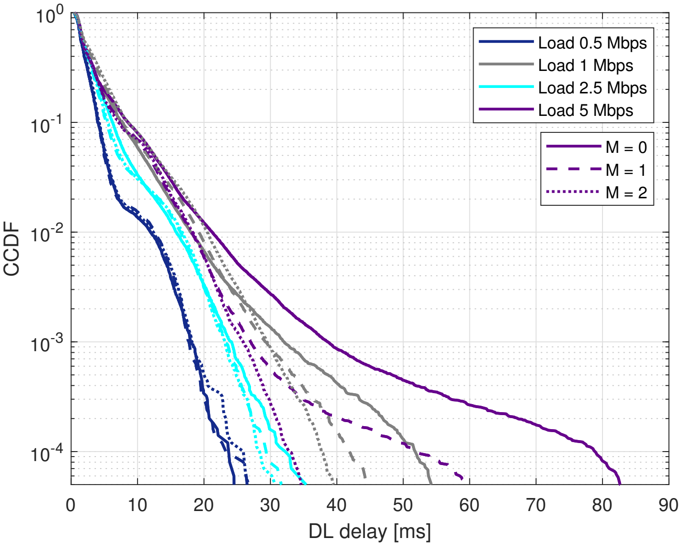


Fig. II.5: CCDF of the delay per downlink packet when multiple HARQ opportunities scheme is used. Solid lines represent baseline assumptions, i.e. $M = 0$, dashed lines and dotted lines refer to $M = 1$ and $M = 2$, respectively.

additional feedback opportunities, UL bursts extend their duration during the next TTIs which, in turns, delays the starting of the channel access procedure for the next COT. For $M = 2$, it provides reasonable improvement for 2.5 Mbit/s and 5 Mbit/s, i.e. the high load cases, whereas for 1 Mbit/s the latency reduction is minimum as compared to $M = 1$.

In Figure II.6, the NACK ratio, that is the number of PDUs considered as NACK due to Cat2 LBT divided by the total number of downlink PDUs received by each UE, is plotted. It is noted that, for all the considered loads, a reduction in NACK ratio is achieved when using multiple occasions for signalling the HARQ feedback. Specifically, providing an additional HARQ transmission opportunity ($M = 1$) highly reduces the NACK ratio at low-load cases, i.e. 0.5 Mbit/s and 1 Mbit/s, whereas at high-load cases, i.e. 2.5 Mbit/s and 5 Mbit/s, the improvement is limited. This is due to the fact that, at high-load cases, the interference which is blocking UEs at the first ACK/NACK transmission attempt is likely to continue in next subframes as compared to the low-load cases. Therefore, additional opportunities are required in these cases. $M = 2$ is needed to reach NACK ratios bellow 10 % for the high-load cases.

The fact that less retransmissions are triggered by the serving eNBs has an immediate effect in the duration of the downlink burst during the COT and, in turn, in the interference level. In Figure II.7, a comparison between

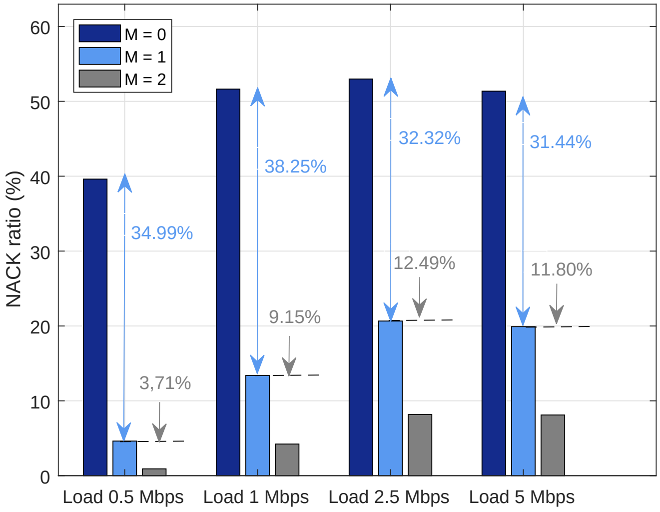


Fig. II.6: Negative acknowledgement ratio for $M = 0$, i.e. baseline assumptions, $M = 1$ and $M = 2$ for 4 different system loads.

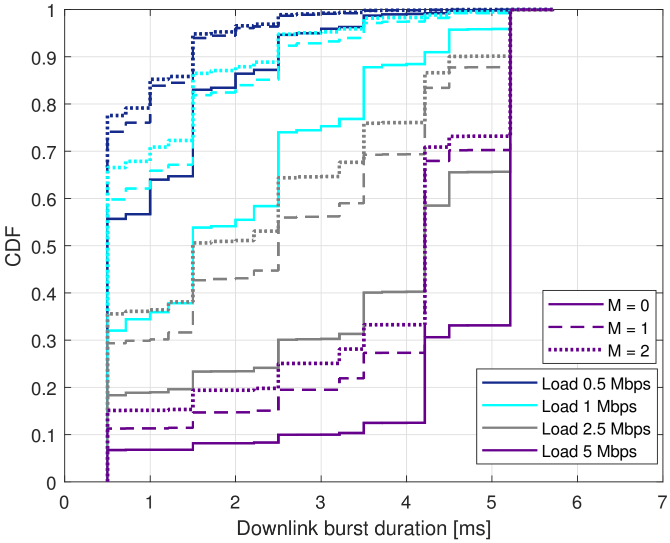


Fig. II.7: CDF of the duration of the downlink burst for the different loads. Solid lines represent baseline assumptions, i.e. $M = 0$, dashed lines and dotted lines refer to $M = 1$ and $M = 2$, respectively.

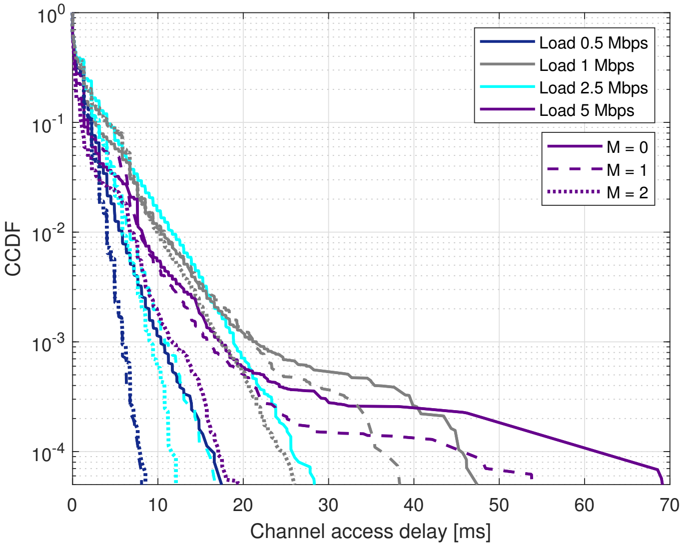


Fig. II.8: CCDF of the time spent performing Cat4 LBT by the eNBs. Solid lines represent baseline assumptions, i.e. $M = 0$, dashed lines and dotted lines refer to $M = 1$ and $M = 2$, respectively.

the duration of the downlink burst for the three different cases is shown. It is noted that the scheme brings a reduction in the downlink burst duration for all the considered loads. The reduction in the interference level highly impacts both neighbour eNBs and UEs channel access procedures. In Figure II.8, a comparison among the eNBs channel access delays is presented. A reduction in the time spent performing the Cat4 LBT is acquired in all the considered loads. Moreover, it reduces the RTT delay, as well as, the queuing delay as new packet transmissions can be served in shorter time.

Figure II.9 provides a latency-reliability comparison between baseline simulations, i.e. UEs are always mandated to perform Cat2 LBT prior to any transmissions, and simulations in which UEs are configured to skip the Cat2 LBT if the gap with the last downlink transmission and the next uplink transmission is lower than 16 μs. It can be noted that the enhancement is highly impacting the high-load cases, whereas low-load cases do not experience such benefit. This is because at low-load regime, the serving eNB rarely can extend its transmission during the partial ending subframe of the special subframe and, thus, the gap is higher than 16 μs. On the other hand, in the high-load cases, it is more likely that eNBs have enough downlink data to keep transmitting until reaching the DL-UL gap in the special subframe. As shown in Figure II.10, the probability of being blocked while performing Cat2 LBT is significantly reduced when implementing the proposal reaching

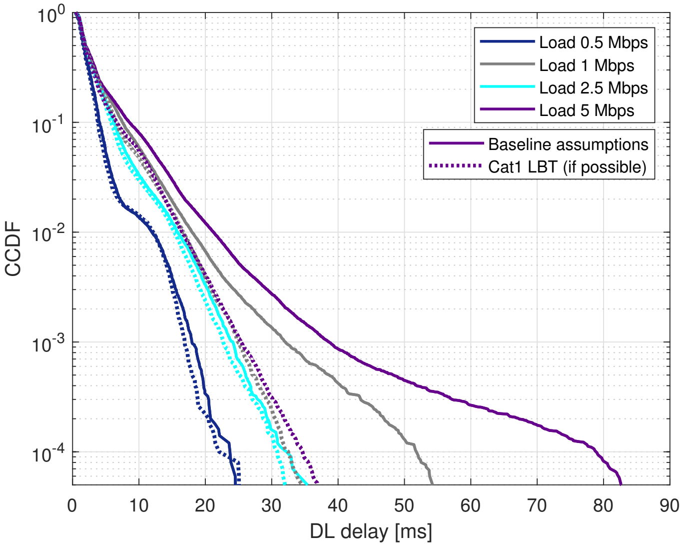


Fig. II.9: CCDF of the delay per packet in downlink. Solid lines represent baseline assumptions while dotted lines show the delay when Cat1 LBT is considered.

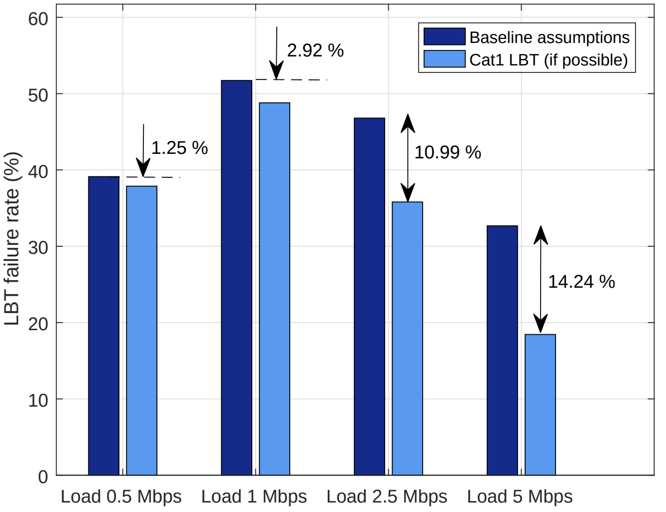


Fig. II.10: Probability of being blocked while accessing the channel over sPUCCH resources by a Cat2 LBT.

the lowest probability at 18 % when having 5 Mbit/s offered load. Looking at the achieved blocking probabilities, it is noted that it is highly likely that a UE is blocked on the sPUCCH resources at any of the considered loads. The main contributor to this is the well-known hidden node problem. This happens when an eNB performs a successful Cat4 LBT and allows COT sharing with its serving UEs. Although the eNB initially sensed the channel as clear, the UEs are exposed to interference coming from neighbour nodes that are out of the range of the eNB. This undiscovered interference by the eNB is preventing UEs to use the channel for potential transmissions. The fact that the blocking probability is lower for the high-load cases compared to the low-load cases is due to an unpredictable coordination in the frame selection is achieved. By aligning the uplink transmissions, the LBT instances are also aligned, which makes the UEs sense the channel as free more frequently.

Even though both previously mentioned solutions are directly aiming at reducing the impact of Cat2 LBT when HARQ feedback is transmitted, they also indirectly impact the latency for uplink packets. This is due to the fact that, if the number of downlink retransmissions is reduced, uplink packets can be served faster in a TDD system. Moreover, since interference in the system is also reduced, the LBT blocking probability is lowered. Figure II.11 shows the reduction in the uplink delay per packet when using Cat1 LBT. Reduction of the uplink latency is achieved for every offered loads. For further description of the delay improvement provided by the aforementioned techniques, see Table II.2.

Uplink

A reliability-latency performance comparison between grant-based uplink and grant-free uplink is provided in Figure II.12. It is observed that for 0.5 Mbit/s and 1 Mbit/s cases, the GF scheme outperforms the GB uplink. Specifically a latency reduction of 58 % and 44 % is achieved at 99.99 % reliability for 0.5 Mbit/s and 1 Mbit/s, respectively. Benefits are obtained due to the skipping of the SR procedure and the scheduling delay. Moreover, these offered loads maintain the number of collisions at a considerable low rate, i.e. the likelihood that 2 UEs transmit over the same shared resources is low. However, when the load is increased up to a certain point the eNB is not capable to decode the multiple transmissions due to the high number of collisions. This is observed when the load is increased up to 5 Mbit/s. At this offered load, grant-based uplink outperforms grant-free uplink, obtaining 23 % lower latency performance at 99.99 % reliability.

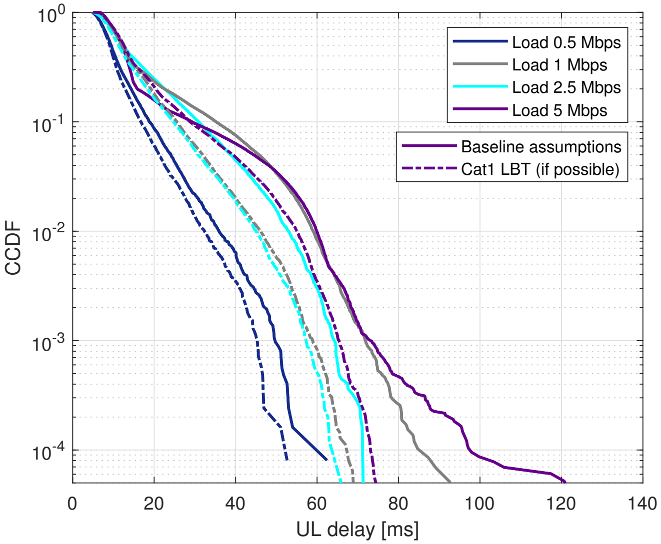


Fig. II.11: CCDF of the delay per packet in uplink. Solid lines represent baseline assumptions while dotted lines show the delay when Cat1 LBT is considered.

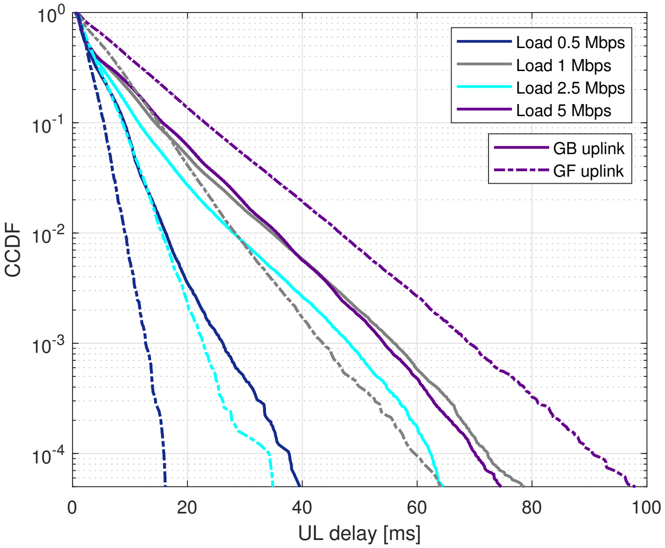


Fig. II.12: CCDF of the delay per packet in uplink. Solid lines represent grant-based uplink performance while dotted lines show grant-free uplink.

Table II.2: Summary results: Delay reduction in percentage compared with baseline assumptions.

Reliability	**Load**	**Multiple HARQ opportunities**	**Cat1 LBT**
99.9^{th} percentile	1 Mbit/s	DL: 3.5 %; UL: 22.1 %; ($M = 1$)	DL: 6.2 %; UL: 10.6 %
	5 Mbit/s	DL: 28.5 %; UL: 3.1 %; ($M = 2$)	DL: 34.3 %; UL: 10.2 %
99.99^{th} percentile	1 Mbit/s	DL: 8.8 %; UL: 10.8 %; ($M = 1$)	DL: 4.2 %; UL: 11.4 %
	5 Mbit/s	DL: 55.2 %; UL: 18.3 %; ($M = 2$)	DL: 55.2 %; UL: 24.8 %

6 Conclusions

A detailed system-level latency-reliability analysis of standalone operation in unlicensed spectrum has been presented for a multi-cell/multi-user scenario with dynamic bi-directional traffic. In line with our initial hypothesis, it is found that the latency-reliability performance is severely limited by the channel access procedures. Specifically, it accounts for an average over the considered loads of 78 % and 44 % of the total one-way packet latency budget for downlink and uplink respectively at 99.99 % reliability. Several latency-reduction solutions have been presented and evaluated. Using multiple HARQ occasions have been shown to achieve a significant reduction in the NACK ratio. On average for all the considered offered traffic load, a 34 % reduction in the NACK ratio is achieved when providing an additional ACK/NACK transmission opportunity. This translates to latency reduction of 26 % of the downlink delay at 99.99 % reliability for highly loaded cases. An additional NACK ratio reduction of 9 % is achieved when $M = 2$. This is especially noticeable at high-offered traffic loads. Providing multiple HARQ opportunities allows the system to serve new transmissions faster and, thereby, reducing the queuing delay. Additionally, it is shown that Cat1 LBT provides substantial advantages as compared to baseline simulations. Especially at high-loads, it achieves a 14 % LBT failure probability reduction as compared to baseline at 5 Mbit/s load. It has been verified that this reduction impacts the delay per packet in both downlink and uplink transmissions. For downlink delay, reductions of 5 % and 46 % have been achieved at 99.99 % reliability for low-loads and high-loads, respectively. The uplink delay impact shows a reduction of 13 % and 17 % at the same reliability level for low and high-loads, respectively. As an uplink specific enhancement, the grant-free scheme was studied and compared with grant-based operation for different loads. It was shown that the system load plays an important role for the performance benefits of GF. We have shown that a latency of reduction of 52 % has been achieved with a reliability level of 99.99 % for low-loads. At the maximum considered offered load, the latency achieved by GF exceeds the latency provided by GB by 23 %.

There is now ongoing research to further improved the latency and reliability performance for standalone unlicensed band operation by the introduction of NR-U. Latency-reliability analysis of NR-U and related enhancements are therefore currently an active research area. Further technology enhancements for supporting stringent latency-reliability requirements are expected as part of future 3GPP releases.

References

[1] 3GPP, "Feasibility Study on New Services and Markets Technology Enablers for Enhanced Mobile Broadband," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.213, 2018.

[2] ——, "Feasibility Study on New Services and Markets Technology Enablers for Massive Internet of Things," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 22.861, 2016.

[3] ——, "Feasibility Study on New Services and Markets Technology Enablers for Critical Communications," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.213, 2018.

[4] ——, "Study on Communication for Automation in Vertical domains (CAV)," 3rd Generation Partnership Project (3GPP), Technical Report (TR) 22.804, 2018.

[5] 5G-ACIA, "5G for Connected Industries and Automation - White Paper, 2nd Edition," 2018.

[6] *Feasibility Study on Licensed-Assisted Access to Unlicensed Spectrum; (Release 13) - 3GPP TR 36.889 V13.0.0*, 3rd Generation Partnership Project Std.

[7] 3GPP, "New Work Item on enhanced LAA for LTE," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.213, 2015.

[8] M. Alliance, "MulteFire Release 1.0 Technical Paper: A New Way to Wireless," Tech. Rep. [Online]. Available: https://www.multefire.org/

[9] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Study on NR-based Access to Unlicensed Spectrum (Release 15)," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.889, 2018.

[10] ——, "Feasibility Study on 6 GHz for LTE and NR in Licensed and Unlicensed Operations," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.805, 2019.

[11] ——, "Study on New Radio access technology; 60 GHz unlicensed spectrum," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 37.890, 2017.

[12] S. Lagen, L. Giupponi, S. Goyal, N. Patriciello, B. Bojovic, A. Demir, and M. Beluri, "New radio beam-based access to unlicensed spectrum: Design challenges and solutions," 2018.

[13] R. M. Cuevas, C. Rosa, F. Frederiksen, and K. I. Pedersen, “On the Impact of Listen-Before-Talk on Ultra-Reliable Low-Latency Communications,” in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec 2018, pp. 1–6.

[14] ——, “Uplink Ultra-Reliable Low Latency Communications Assessment in Unlicensed Spectrum,” in *2018 IEEE Globecom Workshops (GC Wkshps)*, Dec 2018, pp. 1–6.

[15] A. Z. Hindi, S. Elayoubi, and T. Chahed, “Performance evaluation of ultra-reliable low-latency communication over unlicensed spectrum,” in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, May 2019, pp. 1–7.

[16] ETSI, *5 GHz RLAN; Harmonised Standard covering the essential requirements of article 3.2 of Directive 2014/53/EU*, ETSI EN 301 893 V2.1.1 (2017-05) Std.

[17] “IEEE Standard for Information technology—Telecommunications and information exchange between systems Local and metropolitan area networks—Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,” *IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012)*, pp. 1–3534, Dec 2016.

[18] 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.

[19] 3GPP, “Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Physical Layer Procedures,” 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 36.213, 2018.

[20] T. Svensson, T. Frank, D. Falconer, M. Sternad, E. Costa, and A. Klein, “B-IFDMA - A Power Efficient Multiple Access Scheme for Non-frequency-adaptive Transmission.”

[21] K. I. Pedersen, G. Monghal, I. Z. Kovacs, T. E. Kolding, A. Pokhariyal, F. Frederiksen, and P. Mogensen, “Frequency Domain Scheduling for OFDMA with Limited and Noisy Channel Feedback,” in *2007 IEEE 66th Vehicular Technology Conference*.

[22] K. Pietikainen, F. Del Carpio, H. Maattanen, M. Lampinen, T. Koivisto, and M. Enescu, “System-Level Performance of Interference Suppression Receivers in LTE System,” in *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*, May 2012, pp. 1–5.

[23] K. Brueninghaus, D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G. . Seraji, “Link performance models for system level simulations of broadband radio access systems,” in *2005 IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 4, Sep. 2005, pp. 2306–2311 Vol. 4.

[24] R. Srinivasan, J. Zhuang, L. Jalloul, R. Novak, and J. Park, “IEEE 802.16m Evaluation Methodology Document (EMD),” 2008.

Part III

URLLC over unlicensed spectrum in controlled environments

From MulteFire to optimized NR-U for ultra-reliable low-latency communications

This part guides the reader through a transition in the adopted standalone unlicensed technology, from MulteFire to New Radio-Unlicensed. Key technology components, initially designed for fulfilling URLLC requirements in licensed spectrum, are applied and evaluated for unlicensed operation. A comprehensive analysis of the techniques and their additional benefits from an unlicensed spectrum perspective is conducted. Although these enhancements facilitate the achievement of demanding latency and reliability QoS, the channel access mechanisms still play an important role in the overall unlicensed system performance. In this part, we exploit the possibilities of the spectrum regulations together with new RRM techniques to reduce, and in some cases fully mitigate, the impact of the channel access mechanisms. System-level simulations show that, under single network conditions, unlicensed deployments with a proper design achieve promising latency-reliability results.

1 Problem Description

MulteFire is designed based on LTE specifications and therefore it inherits the LTE physical layer design and processing capabilities. In LTE, the minimum time between a DL packet transmission and a retransmission is 8 ms. This is also known as round-trip time (RTT) and accounts for the BS and UE processing times, as well as the transmission times for the DL data packet and the UL control packet. In cases where the packet decoding is not successful in the first attempt, the packet delay further increases. The high LTE RTT prevents the support of sub-10 ms latency performance when only at least one packet retransmission is required. Even in cases where transmissions are correctly received in the first attempt, i.e. no retransmissions are needed, the LTE physical layer design already limits the supported applications. Thus, the achievement of the stringent latency and reliability requirement requires the adoption of certain optimizations. From the analysis of the DL and UL packet delay in Part II, it is noted that several of the delay components contributing to the overall latency are fixed and independent of the type of spectrum used, i.e. licensed or unlicensed. One of these components is the over-the-air delay, i.e. the time spent in the transmission of a packet of the air. MulteFire adopts a minimum allocation size of 14 OFDM symbols with a sub-carrier spacing (SCS) of 15 kHz. In certain cases, the transmission duration can be reduced to 7 OFDM symbols as an optimization to cope with large differences between the LBT finish time and the subframe boundary. These configurations correspond to a over-the-air delay of 1 ms and 0.5 ms, respectively. Using such large time-frequency allocations seems inefficient based on the expected characteristics of URLLC traffic, defined by small payload sizes and with sporadic arrival. On top of that, the capabilities of the nodes for processing the received data and prepare the next transmission are also important from a latency perspective. LTE assumes that UEs require 4 ms for decoding a DL packet. With this being said, a need for improvement in the physical layer design and processing times is required.

Although these enhancements aim to reduce certain latency components of the overall packet delay, the contribution of the channel access mechanisms is still present in unlicensed operation. From Part II, our observation is that the channel access delay represents a large part of total DL packet latency. Therefore, reducing the packet latency requires a review and optimization of the channel access mechanisms at the BS side. Additionally, the UL LBT is still the critical step due to its major impact in the achievable latency. A failure before a DL HARQ feedback transmission implies the retransmission of a packet, even in cases in which the DL packet is already correctly received. In case of failure before an UL data transmission, it requires the issue of a new grant as the reserved time-frequency resources are not utilized. In both cases, the competitiveness in the channel access is increased as nodes need to contend for the channel more frequently.

2 Objectives

The key objectives for this part of the PhD are described in the following:

- Evaluate the latency-reliability performance of key NR technology components when applied to unlicensed spectrum, highlighting additional benefits as compared to operation in licensed spectrum

- Contribute to the reduction of the channel access impact on the system performance by introducing novel RRM techniques.

- Establish a performance comparison among the asynchronous and synchronous channel access designs described in the spectrum regulations and 3GPP for operation on the 5 GHz band.
- Propose coordination schemes as enablers for achieving ultra-reliable low-latency communications in unlicensed deployments.

3 Included Articles

This part of the thesis includes the following articles:

Paper D. Analysis of High-Reliable and Low-Latency Communication Enablers for New Radio Unlicensed

Relevant technology enablers for low-latency and high-reliable communications are individually analyzed. Specifically, the flexible NR time-frequency design that allows for higher SCS and shorter time allocations (also known as mini-slots) and the reduced processing times for preparing and decoding transmissions. The implications and the added benefits of applying these techniques to the unlicensed spectrum are studied and evaluated. Additionally, the possibility of having multiple transitions between DL and UL slots during a TDD frame is also analyzed. Furthermore, a new technique to cope with the potential failures in the UL LBT is presented. The focus is on UL data transmissions, i.e. physical uplink shared channel (PUSCH) transmissions, in deployments using grant-based UL. The proposed solution leverages from the time variability of the system interference and preemptively allocate UEs with multiple and consecutive resources. The technique enhances the information included in the UL grant to indicate the UL resources to be used in a certain transmission time interval (TTI) (as in baseline grant-based) and, additionally, in the K next UL TTIs. The pre-allocated UL resources in the k^{th} TTI are only used in case of experiencing consecutive UL LBT blockage in the previous $k-1$ TTIs. By providing multiple UL resources with a single grant increases the probability of successfully accessing the channel. Moreover, it decreases the number of needed grants per UL packet.

Paper E. A Fully Coordinated New Radio-Unlicensed System for Ultra-Reliable Low-Latency Applications

A performance comparison between the two channel access designs defined by the european regulator for the 5 GHz band is conducted. Under the assumptions of single RAT conditions, the paper motivates the usage of the synchronous channel access rather than the asynchronous channel access for latency and reliability enhancement purposes. The main contribution of this paper is the proposal of a centralized and fully coordinated NR-U deployment that completely mitigates the effect of the channel access. Coordination is ensured in 2 domains: the channel access and the frame selection. Firstly, channel access coordination among nodes is reached by sharing a common timing reference for sensing and occupying the channel. i.e. adopting FBE channel access design. Additionally, nodes are coordinated in terms of the selected frame configuration every time the channel is accessed. A central node is proposed as the enabler for this functionality. Bi-directional information exchanges between the central entity and the nodes ensure an optimal frame configuration selection according to the specific traffic needs. The fully coordinated approach brings the latency and reliability performance of an NR-U system closer to the expected behaviour of an equivalent licensed NR TDD deployment.

4 Main Findings

Technology enablers for low-latency and high-reliability

In Paper D, the importance of the subcarrier spacing, the nodes processing capabilities and the TTI size on the achievement of low-latency at high-reliability levels is analyzed. Switching from 15 kHz to 30 kHz subcarrier spacing and from 42 OFDM symbols to 5.5 OFDM symbols in the UE decoding capabilities brings extraordinary latency benefits. A reduction in the packet delay of 75 % for an offered network load of 1 Mbit/s is observed at 99.99 % reliability. Especially interesting is the duration of the transmission time interval. In unlicensed operation, apart from decreasing a fixed delay component of the overall packet delay, it also reduces the time that a node occupies the channel. This has a direct impact on the channel access performance, as the channel is seen as idle more frequently. Both gNBs and UEs benefit from it. At the UE side, a reduction in the UL blocking probability of 6.7 % and 18.7 % for network loads of 1 Mbit/s and 2.5 Mbit/s is observed when switching from 14 to 7 OFDM symbols TTI. Additionally, allowing multiple transitions (also known as switching points) between DL and UL within a TDD frame are shown as beneficial as it can reduce the gap between transmissions. This reduces the possibility for neighbour nodes to access the channel during gaps within a channel occupancy time (COT). Moreover, it also lowers the number of initiating device LBT attempts in the system since transmissions, and potential retransmissions, can be confined within a single COT. The aforementioned techniques enhance the latency and reliability performance, however, the system is still impacted by the channel access mechanisms. As unlicensed-specific enhancements, in Paper D, it shows that

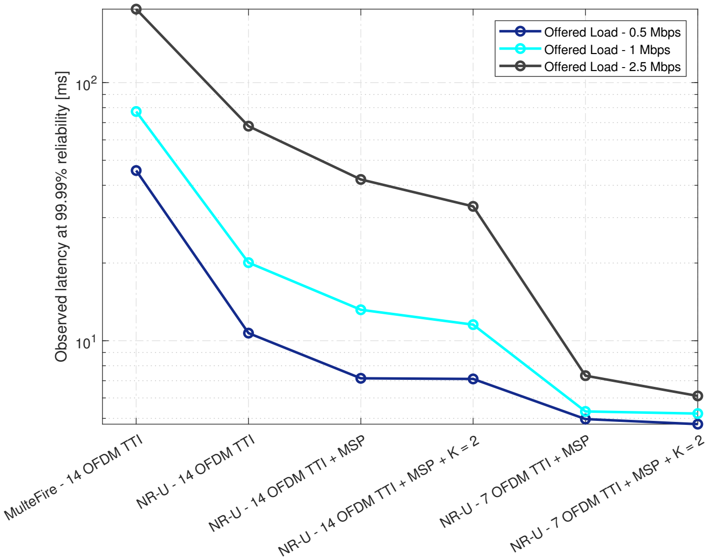


Fig. III.1: Summary of the achieved packet delay for the different technology enablers. TTI stands for tranmission time interval, MSP for multiple switching points and K refers to the additional UL resources signaled in a single grant transmission.

providing multiple resources for UL data transmission as a potential candidate for latency reduction in deployments following grant-based UL guidelines. Despite the focus of the technique on UL transmissions, a reduction in the combined DL + UL delay is achieved. This is because fewer nodes are contending for the channel as the overall number of needed grants per transmission is reduced. The proposed mechanism is shown to be suitable for scenarios with high UL LBT blocking probability. In such scenarios and at 99.99 % reliability, the technique can reduce approximately 23 % the achieved latency as compared to baseline grant-based UL. However, in scenarios with low LBT blocking probability, the benefits of the proposed technique are very limited. A summary of the obtained latency at 99.99 % reliability for each of the technology enablers is shown in Fig. III.1.

Asynchronous vs synchronous channel access

In deployments in which the presence of a single RAT is guaranteed, Paper E promotes the adoption of the synchronous LBT design as it achieves faster channel access delay as compared to the asynchronous LBT. In synchronous LBT, the usage of a common timing reference ensures that all the gNBs can access the channel in the next sensing interval. It is worth mentioning that, since the channel is only accessible during the sensing interval, packet trans-

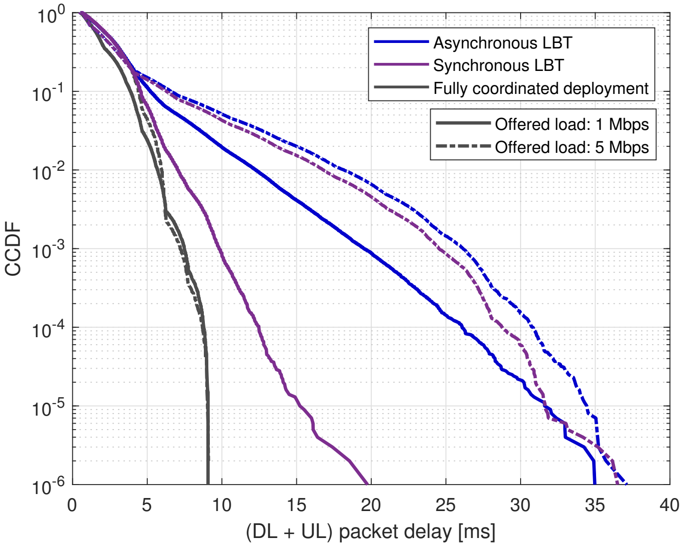


Fig. III.2: Overall packet delay for asynchronous LBT, synchronous LBT and fully coordinated deployment. (Source: Paper E)

missions are susceptible to be delayed while waiting for the next sensing period, i.e. the frame alignment needs to be considered. On the other hand, in scenarios with asynchronous channel access, the observed behaviour confirms the conclusions of Part II. The channel access delays represents a large quota of the overall packet delay, e.g at 99.99 % reliability, 61 % and 49 % for loads of 1 Mbit/s and 5 Mbit/s, respectively. Therefore, under single RAT conditions, our recommendation is to use synchronous channel access for low-latency and ultra-reliable communications. Although synchronous channel access reduces the DL channel access delay, UL transmissions may still experience LBT blockage with consequent impacts on the UL (and DL) latency and reliability performance. This occurs when the channel is simultaneously accessed by multiple gNBs after a successful LBT during the synchronous idle period. Assuming differences in the instantaneous traffic conditions, each gNB selects its optimal TDD frame for the next COT. This heterogeneous decision leads to potential overlappings of DL transmissions with UL sensing intervals among the cells. Neighbours cells transmissions increase the received interference levels, reducing the chances of successful channel access by certain UEs. High UL LBT blocking probabilities, above 30 % in highly loaded scenarios, are observed. This behaviour motivates the search for additional solutions.

Fully coordinated deployment

To solve the mentioned UL LBT problem, a central entity that coordinates the selection of the TDD frame among the gNBs is proposed. Introducing the central entity while assuming synchronous LBT, leads to a fully coordinated unlicensed deployment. This approach provides fast channel access at the gNB side, by leveraging from the synchronous LBT, and ensures no UL LBT blocking, by agreeing on the adopted TDD frame by the active gNBs. As compared to synchronous deployments without the central node, a latency reduction of 29 % and 70 % for loads of 1 Mbit/s and 5 Mbit/s at 99.99 % reliability (see Fig. III.2) is achieved. This scheme is shown as a valuable solution for achieving ultra-reliable and low-latency communications in unlicensed spectrum as packet delays bellow 10 ms are observed at any reliability level. The achieved latency can be further decreased by selecting a proper configuration of the duration of sensing and transmitting periods. Under the boundaries of the spectrum regulations, the sensing period can be reduced to 100 µs and the transmission period to 900 µs. This increases the channel access periodicity and reduces the frame alignment since a new opportunity for the channel access is presented every millisecond.

5 Key recommendations

Based on this study, the following guidelines are advised to be followed to achieve stringent latency and reliability targets:

- Adopt key URLLC technology components such as higher SCS, faster processing times and shorter TTI size. These optimizations reduce key delay components of the overall packet delay as well as reduce the impact of the channel access mechanisms.
- Enable multiple DL and UL transitions within the COT. This flexibility in the TDD pattern leads to higher COT utilization. It allows for faster HARQ feedback and, in case of unsuccessful packet decoding at first transmission, it reduces the probability of using multiple COTs for a single packet. The queuing delay for new packet transmissions is also reduced.
- In asynchronous deployments where UL LBT blocking is impacting the overall performance, consider the multiple PUSCH occasions per UL grant as a solution to mitigate the persistent blocking of UL data transmissions.
- Under single RAT conditions, design the NR-U deployment according to the synchronous channel access rules. Coordination in the TDD selection among the nodes is also advised. With this configuration, the

channel access impact is minimized and the NR-U performance resembles obtained with TDD NR-licensed.

Paper D

Analysis of High-Reliable and Low-Latency Communication Enablers for New Radio Unlicensed

Roberto Maldonado, Claudio Rosa, Klaus I. Pedersen

The paper has been published in the
IEEE Wireless Communications and Networking Conference (WCNC), 2020.

Abstract

In this paper, the performance impact of several high-reliable and low-latency communications technology enablers in the unlicensed spectrum for standalone operation is evaluated. Firstly, a comparison between MulteFire (MF) and New Radio-Unlicensed (NR-U) is established. It is shown that higher sub-carrier spacings and shorter processing times provide clear latency reduction benefits. Additionally, different transmission time intervals (TTI) durations are evaluated. Shortening the TTI duration decreases the latency by a factor of 5.75 *at* 99.99 % *reliability and reduces the uplink listen before talk (LBT) blocking probability by* 18 %. *The possibility of having multiple switching points during the frame is also evaluated. It is concluded that having multiple switching points provides a latency reduction mainly due to the reduction of the number of channel accesses and the reduced gaps within the frame. A time-diversity technique to cope with high uplink LBT blocking probability is also evaluated. By combining the aforementioned features, a latency reduction factor of* 31 *is achieved when optimized NR-U and MulteFire performances are compared.*

1 Introduction

Operation in the unlicensed spectrum by a 3$^{\text{rd}}$ Generation Partnership Project (3GPP) technology started with Licensed-Assisted Access (LAA) [1]. LAA is a Long Term Evolution (LTE) enhancement designed for supporting capacity demanding applications. It uses unlicensed spectrum as a supplementary resource to offload part of the traffic via carrier aggregation. The feasibility of LAA depends on the presence of an anchor base station deployed in the licensed bands. As an alternative, MulteFire [2] emerged as a fully standalone technology deployed on unlicensed bands based on LTE specifications. Nowadays, within the fifth generation (5G) era, 3GPP's interest in unlicensed spectrum is maintained. New Radio (NR), as well as, its unlicensed alternative, New Radio-Unlicensed (NR-U), are currently being designed and standardized. Licensed-unlicensed dual operation and standalone unlicensed are considered as potential scenarios for NR-U [3]. NR is envisioned to support three different categories of use cases: enhanced mobile broadband (eMBB), ultra-reliable and low-latency communications (URLLC) and massive machine-type communications (mMTC). URLLC applications demand very stringent requirements in terms of latency and reliability. Applications with more relaxed latency-reliability requirements that URLLC can be classified as high-reliable low-latency communications (HRLLC).

Supporting URLLC use-cases in the unlicensed spectrum is more challenging. Unlicensed frequency bands are open for use by any radio access technology (RAT). Each node needs to be compliant with a strict regulations in order to ensure fair channel occupancy among RATs. This impacts the channel availability and, consequently, guaranteeing a given quality of service becomes more difficult. However, due to its global availability and simplicity in the deployments, verticals consider unlicensed spectrum as a valuable asset. Industry verticals can make use of NR-U to start a transition towards the Industries of the Future supporting applications such as motion control or inventory management [4].

Multiple examples of latency-reliability enhancements for the support of HRLLC/URLLC in the licensed spectrum can be found in the literature. For instance, in [5] authors propose and analyse grant-free uplink as an enabler for URLLC applications. For unlicensed operation, authors in [6] analysed several enhancements from a latency-reliability perspective. Authors in [7] propose a fully coordinated approach to reduce the impact of the regulatory requirements in latency and reliability. This paper contributes with an analysis of latency-reliability performance impact of different technology enablers on a system fully deployed in the unlicensed spectrum. The goal is to, by means of extensive system-level simulations, highlight not only the intrinsic benefits of the enhancements but also the additional gains that can be achieved when applying these enablers to the unlicensed operation. The remaining of the paper is organized as follows. Sections 2 and 3 describe the regulatory requirements for unlicensed spectrum operation and the assumed system model, respectively. In Section 4 a review of the evaluated technology enablers is included. Simulation assumptions and discussion about the simulation results are included in Sections 5 and 6. Final remarks are drawn in Section 7.

2 Unlicensed spectrum: regulatory requirements

In unlicensed operation, nodes must behave according to specific regulatory requirements to guarantee a fair coexistence among multiple RATs. The requirements may differ depending on the region and the considered sub-band. In order to be globally deployable, NR-U and MulteFire, adopt the most stringent guidelines, i.e. the regulatory requirements provided by the European Telecommunications Standard Institute (ETSI) in its harmonized standard for the 5 GHz band [8]. The requirements include constraints on the transmitted power and the power spectral density. Furthermore, ETSI defines the channel access procedures that nodes must follow before any transmission. Two different channel access designs are specified: load-based equipment (LBE) and frame-based equipment (FBE). In LBE, nodes start the channel access procedure immediately after the data is ready for transmission. However, in FBE nodes follow a periodic sensing/transmit structure that defines when nodes can access the channel. In both cases, after a successful channel access procedure, nodes are allowed to continuously use the channel for a maximum

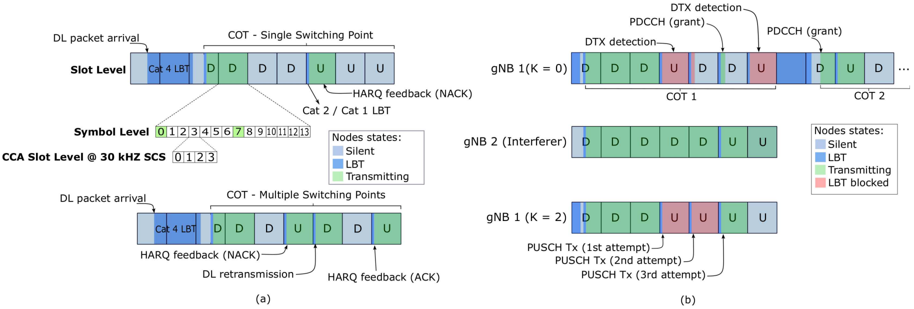


Fig. III.1: LBE operation and comparison between single and multiple switching points in (a) and multiple PUSCH occasions scheme in (b).

duration known as maximum channel occupancy time (MCOT). Initiating nodes, i.e. the nodes starting the channel access procedure, can grant access to other nodes. These nodes, known as responding nodes, are allowed to use the previously acquired channel occupancy time (COT) by the initiating nodes, i.e. applying the COT sharing principle. Channel access procedures are based on the principle of listen before talk (LBT). LBT decides the channel activity based on an interference measurement and a posterior comparison with a predefined energy detection (ED) threshold. The channel is declared as busy if the measured interference is higher than the ED threshold. In LBE, initiating nodes acquire the channel by a Category 4 (Cat4) LBT. Cat4 LBT is a contention window based mechanism with exponential backoff performed in intervals of 9 µs, known as clear channel assessment (CCA) slots. Nodes need to sense the channel free for, at least, the selected number of CCA slots before accessing the channel. The number of CCA slots is given by a uniform distribution whose boundaries are defined by the contention window size. The contention window size is exponentially increased in case of collisions. Responding nodes access the channel with a Category 2 (Cat2) LBT in COT sharing conditions. The sensing interval duration for Cat2 LBT is fixed to 25 µs. This type of LBT is also known as single-shot LBT as only provides one opportunity to access the channel. Thus, if the channel is declared as busy, the opportunity for performing COT sharing is missed. Further details about the channel access procedures for the 5 GHz band, both LBE and FBE, can be found in [8] and [3].

3 System model

A single-operator indoor office scenario with M next generation nodes-B (gNBs) and K user equipments (UEs) uniformly distributed around the layout is assumed. A channel bandwidth of 20 MHz in the 5 GHz frequency band is available. Bi-directional traffic with payloads of B bytes in a dynamic time domain duplexing (TDD) system is assumed. The TDD frame configuration, i.e. the number of downlink and uplink slots, is selected based on the buffer status at the nodes. Each node selects its optimal TDD frame configuration independently. Slot-level synchronization among the nodes is assumed. LBE is the adopted channel access design. gNBs are considered as initiating nodes whereas UEs are responding nodes. Therefore, as shown in Fig III.1(a), gNBs perform the procedure to acquire the channel with a Cat4 LBT as soon as data is ready for transmission. Additionally, gNBs offer the possibility of sharing the COT to their connected UEs. While gNBs can only start the COT at specific symbols during the slot, Cat4 LBT can finish at any time during the slot. Two starting positions per slot are defined, i.e. OFDM symbols 0th and 7th. In case of time-misalignment, gNBs remain silent after a successful Cat4 LBT waiting for the next transmission starting symbol. Before initiating the transmission, an additional Cat2 LBT needs to be performed. In downlink, gNBs schedule downlink transmissions in the physical control downlink channel (PDCCH). UEs are configured with two different PDCCH monitoring occasions depending on the COT conditions. Out of the COT, UEs monitor the PDCCH occasions at OFDM symbol 0th and 7th, whereas, in gNB-acquired COT conditions, the PDCCH monitoring is performed at the slot boundaries. In uplink, UEs must receive a dedicated grant before they can proceed with the transmissions, i.e. grant-based uplink is assumed. Therefore, UEs need to signal that they have data ready to transmit through a scheduling request (SR) procedure. This consists of a message exchange between UEs and gNBs, which is subjected to the LBT outcome in unlicensed operation, and whose ultimate goal is to provide UEs with dedicated resources.

4 Latency and reliability enablers

An overview of the different technology enablers for achieving low-latency and high-reliable communications are presented in this section. Their benefits from an unlicensed operation perspective are also highlighted. First, a description of the physical layer design for NR is discussed. Unlicensed-specific enhancements such as TDD frame structure optimizations and multiple PUSCH occasions scheme are also presented.

4.1 Flexible time-frequency design

NR offers the possibility of multiple sub-carrier spacing (SCS) designs. Apart from the 15 kHz SCS supported by LTE, NR supports 30 kHz, 60 kHz, 120 kHz and 240 kHz as possible SCS values [9]. In frequency ranges below 7 GHz, SCSs up to 60 kHz are supported. Higher SCS imply shorter OFDM symbol duration. Reducing the symbol duration is considered as an enabler for latency-sensitive applications. For example, a reduction by a factor of 2 in the transmission time interval (TTI) duration is achieved when switching from 15 kHz to 30 kHz SCS. Moreover, flexibility in the time domain is achieved by supporting the possibility of using smaller TTI sizes than the slot duration, e.g. 14 OFDM symbols. Smaller TTI sizes, also known as mini-slots, can consist of two, four, or seven OFDM symbols. Supporting higher SCS and shorter TTI sizes not only provide gains in terms of latency performance. It also brings further benefits when it is applied to unlicensed spectrum. As previously explained, any transmission in the unlicensed spectrum is subjected to the LBT outcome, and therefore, to the interference measured during a sensing interval. Consequently, adopting higher SCS and shorter TTI sizes implies a reduction in the time that nodes are occupying the channel which increases the probability of sensing the channel as idle when performing LBT.

4.2 Reduced processing times

At the UE side, faster processing times for decoding a recevied PDSCH and preparing a PUSCH transmission after a grant reception are essential for latency reduction purposes. According to the 3GPP specifications, a UE is capable of transmitting the hybrid automatic request (HARQ) feedback from a previously received downlink data transmission T_{proc} ms after the end of the last symbol of the PDSCH. The UE processing delay (T_{proc}) is defined as follows [10]:

$$T_{proc} = (N_1 + d_{1,1})(2048 + 144) \cdot \kappa 2^{-\mu} \cdot T_c, \tag{D.1}$$

where N_1 is the PDSCH decoding time in OFDM symbols and depends on the assumed UE processing capability. Two types of processing capabilities are defined by 3GPP, being the processing capability 2 the more suitable for latency-critical applications. In that case, N_1 is equal to 4.5 OFDM symbols. $d_{1,1}$ depends on the demodulation reference signal (DMRS) position and κ is defined as the ratio between T_s and T_c, being T_s the basic time unit for LTE and T_c the basic time unit in NR [9]. μ defines the adopted sub-carrier spacing for the downlink transmission. For a system with 30 kHz SCS, processing capabilities 2 and DMRS position 0, the PDSCH decoding time at the UE side is 0.15 ms. Eq. D.1 is also employed for the calculation of the PUSCH preparing time at the UE. In this case, T_{proc} defines the minimum time between the reception of the grant and the uplink transmission. The PUSCH preparing time in symbols (equivalent to N_1 in Eq. D.1) takes the value of 5.5 OFDM symbols for UE capability 2 and 30 kHz SCS. Assuming the first symbol of the PUSCH is dedicated to DMRS, the PUSCH preparing time is 0.19 ms. These values clearly differ from the assumed for LTE/MulteFire systems. For example, in MulteFire, the PUSCH preparation time is in the order of 3 ms. Uplink data is ready for transmission in the slot $n + 4$, being n the slot in which the grant was received. At the gNB side, processing times for decoding PUSCH and preparing PDSCH transmissions are also accordingly reduced for NR operation. It is generally assumed that gNBs processing capabilities are twice faster than the UEs capabilities.

4.3 Multiple switching points

One of the differences between MulteFire and NR-U is the support of multiple switching points within the frame structure. NR-U offers the possibility of having several downlink-to-uplink and uplink-to-downlink transitions during the COT as depicted in Fig. III.1(a). This is identified as beneficial since, together with NR-specific processing capabilities, it provides faster HARQ feedback, allowing the possibility of having multiple retransmissions during a single COT, and faster and more accurate link adaptation. A dynamic TDD frame adaptation is supported in NR based on the slot-format indicator (SFI). The SFI indicates the number of OFDM symbols that are dedicated downlink and uplink transmissions in the following slots. SFI is included as part of the control information signalled by the gNBs in the PDCCH. Adapting the frame structure to the instantaneous traffic demands is positive from a channel access perspective as the gap between transmissions within the COT is reduced which increases the likelihood of performing a successful channel access.

4.4 Multiple PUSCH occasions

As shown in [11] and [7], uplink latency is heavily impacted by high uplink LBT blocking probability, especially at high loads. A detailed description of the delay components of an uplink transmission and the impact of LBT on it can be found in [11]. As a baseline, in grant-based uplink, the dedicated resources are reserved for a single time transmission interval, therefore, if a UE fails the uplink LBT, the granted resources are not used. After a discontinuous transmission (DTX) detection at the gNB side, a new grant needs to be forwarded to the UE for a new uplink transmission attempt. This process implies the performance of additional channel access procedures at gNB side, to send a new grant, and at UE side, to perform the uplink transmission. As shown in Fig. III.1(b) a new COT might need to be acquired to finally perform the uplink transmission. To cope with this issue, a time diversity technique is proposed. This technique consists of signalling the UEs with multiple occasions (K) for the uplink transmissions during the SR procedure. As shown in Fig. III.1(b), by providing additional resources, the probability of uplink LBT failure in consecutive slots is lowered, and therefore, the packet delay is reduced. This scheme is similar to mechanisms such as TTI bundling [12] and

Table III.1: NR-U Simulation assumptions

Parameter	**Assumption**
Layout	Indoor mixed office scenario [13]
Channel model	NR InH Mixed Office [13]
Duplexing mode	dynamic TDD
Bandwidth	20 MHz @ 5 GHz band
Sub-carrier spacing	30 kHz
TTI sizes	14 / 7 OFDM symbols
Scheduling metric	Proportional fair
HARQ	Asynchronous HARQ, incremental redundancy
UEs processing times*	PDSCH: 4.5 OFDM symbols [10] PUSCH: 5.5 OFDM symbols [10]
Link adaptation	Outer link loop adaptation: enabled [14] BLER target: 1 %
Receiver type	LMMSE-IRC [15]
MIMO	2×2 configuration DL: Rank-2 SU-MIMO UL: Rank-1 SU-MIMO, receiver diversity
Traffic model	FTP Model 3; B = 50 bytes $\lambda_T = \{25, 50, 125\}$ packets/s/UE
gNBs channel access	Cat4 LBT for COT acquisition Cat2 LBT for UL-to-DL transitions inside COT
UEs channel access	Cat2 LBT for PUSCH and PUCCH
ED threshold	-72 dBm
MCOT duration	3 ms

* gNBs processing times are assumed 2 times faster than UEs capabilities

K-repetition [10]. However, the main difference is that in this case, UEs only use the additional uplink resources if the Cat2 LBT has previously failed. The remaining uplink resources are not used to avoid an additional increase in the system interference which might impact other nodes channel access procedures.

5 Simulation assumptions

An indoor office scenario with 12 gNBs is assumed. A total of 60 UEs are uniformly distributed in the scenario. Each gNB is assumed to have 5 UEs connected during the simulation time. UEs select its serving gNB based on the strongest reference signal received power (RSRP) criteria. The traffic generation follows a Poisson process according to the FTP Model 3 guide-

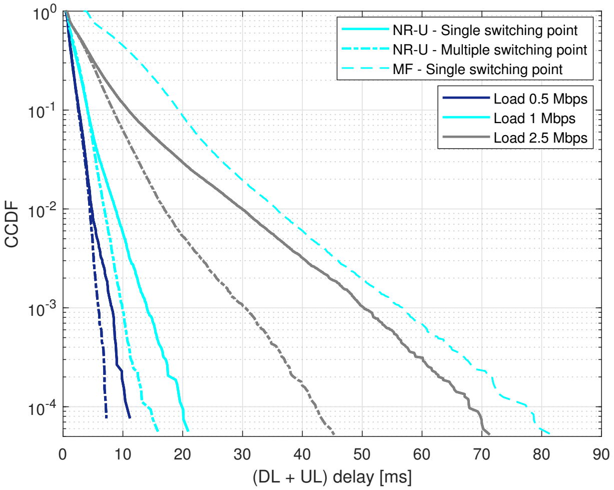
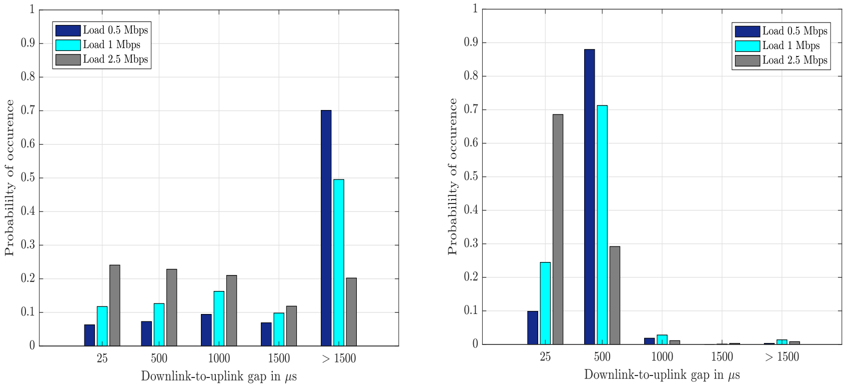


Fig. III.2: Overall packet latency performance comparison for MF and NR-U. For figure readability porpuses, MF performance is only depicted for 1 Mbit/s

lines [16] in which a fixed packet size of 50 bytes is simulated. Several packet arrival rates (λ_T) are evaluated to generate different offered loads in the scenario. The downlink/uplink ratio in the traffic generation is 80 % / 20 %. Scheduling request (SR) transmissions are subjected to LBT outcome. An SR periodicity of 1 slot is simulated. It is assumed that UEs transmit the SR message in uplink slots inside the COT. Out of the COT, UEs can also transmit the SR by accessing the channel by themselves with a Cat2 LBT. Processing times are considered in the packet latency statistics. Therefore, for a 14 TTI OFDM symbol TTI with 30 kHz SCS, the minimum downlink delay is 0.65 ms (0.5 ms TTI + 0.15 ms UE proc. time). The simulator works on an OFDM symbol-subcarrier resolution. It models the majority of the PHY and MAC layer procedures in line with 3GPP guidelines including unlicensed-specific mechanisms. Regarding the methodology, multiple realizations of the scenario with sufficient samples are simulated. In each realization, the UEs' location change. Results from each realization are combined afterwards. A summary of the NR-U simulation assumptions is provided in Table VI.2. MulteFire simulations assumptions and further details about the simulation methodology can be found in [11].



(a)

(b)

Fig. III.3: Downlink-to-uplink gap comparative when (a) having single switching point and (b) having multiple switching points within the COT.

6 Latency-reliability evaluation

In order to evaluate the performance gain of the HRLLC enablers described in Section 4, each of the features is compared against MulteFire and NR-U baseline assumptions. MulteFire assumptions assume 15 kHz SCS, 14 OFDM symbols TTI and LTE processing times, i.e. at UE side, 42 OFDM symbols for both decoding PDSCH and preparing PUSCH transmissions. Processing times at the base stations is assumed to be half the UEs processing times. On the other hand, NR-U baseline assumptions imply the usage of 30 kHz SCS, 14 OFDM TTI and single switching point during the COT.

Firstly, insights into the gains provided by selecting higher SCS and shorter processing times, i.e. switching from MulteFire to NR-U, are presented in Fig III.2. Moreover, the impact of the TDD frame structure in the latency-reliability performance is also addressed. Two TDD schemes are distinguished: a) single switching point and b) multiple switching points. The former refers to the case in which the TDD configuration has only one downlink-to-uplink transition within the COT. The frame configuration is initially selected after a successful Cat4 LBT and maintained during the rest of the COT. The latter case defines the possibility of dynamically adapt the TDD frame configuration to the instantaneous buffer status through periodic SFI signalling. Comparing the latency performance of a MulteFire system with an NR-U baseline, a latency reduction of $\approx 75\,\%$ in 10^{-4} is achieved for 1 Mbit/s offered load. A further improve of $\approx 7\,\%$ is obtained if multiple switching points capabilities are considered. By analysing both NR-U designs, it is noted that providing multiple switching points within the COT is beneficial at any considered load. For instance, at high offered load, i.e. 2.5 Mbit/s, and 99.99 % reliability, i.e. 10^{-4}, a latency reduction of $\approx 38\,\%$ is experienced. The performance boost is achieved due to the possibility of having multiple HARQ retransmissions during the COT, a reduced queuing delay and a shorter DL to UL gap. Reducing the gap duration is positive from a channel access perspective as it increases the likelihood of performing a successful channel access.

In Fig. III.3, a histogram comparing the gap duration for single switching points (a) and multiple switching points (b) is shown. It can be noted that for single switching point assumptions, the gap distances are equally distributed between values ranging from 25 µs, i.e. the Cat2 LBT duration, to 1500 µs or higher, i.e at least 3 NR slots at 30 kHz SCS. On the other hand, by supporting multiple switching points, the downlink-to-uplink gap has been clearly reduced, having more than 90 % of the gaps distributed between 25 µs and 500 µs.

In Fig. III.4, the performance gain achieved by providing multiple PUSCH occasions is depicted. Three different K values are shown, where K defines the number of additional PUSCH transmissions. It is shown that the gain of the scheme is especially noticeable at high loads. At 2.5 Mbit/s offered load, providing an additional opportunity to the PUSCH transmissions gives $\approx 11\,\%$ gain at 99.99 % reliability as compared to baseline grant-based, i.e. $K = 0$. $K = 2$ brings an additional benefit of $\approx 12\,\%$, i.e. an overall benefit of approximately 23 % is achieved. At 0.5 Mbit/s, the gain is limited showing similar performance for any K. This is due to the fact the uplink LBT blocking probability is quite small for low loads as it will be later shown in Fig. III.6. When applying this scheme it should be taken into account that, although it shows lower overall packet latencies, it harms the downlink delay performance as less resources are dedicated during the COT for those transmissions. In any case, the trade-off is positive from an overall packet delay performance.

In Fig. III.5, a latency comparison of an NR-U deployment with different TTI sizes is shown. It compares 14-OFDM symbols TTI against 7-OFDM symbols TTI. Switching from baseline to 7 OFDM symbol TTI provides gains at any considered load. Specifically, at 99.99 % reliability 7-OFDM symbols TTI outperforms 14-OFDM symbols TTI by a factor of 30 %, 63 % and 82 % at 0.5 Mbit/s, 1 Mbit/s and 2.5 Mbit/s, respectively. Lower delays than 10 ms are achieved at any considered load with 7-OFDM TTI. The fact that the transmission time is shortened by half implies a reduction in the time the channel is occupied. This has a direct impact on the channel access. In Fig. III.6, a comparison of the average uplink LBT blocking probability based on the different TTI sizes is depicted. An LBT blocking probability reduction when switching to 7-OFDM TTI is achieved at any offered load being more noticeable at high offered loads. A reduction of 18.7 % is observed at

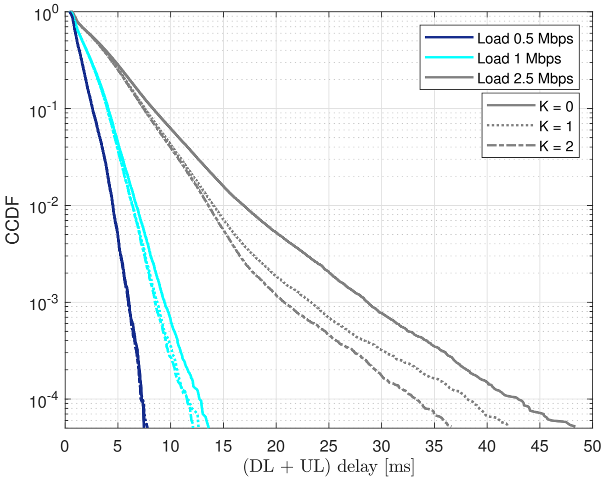


Fig. III.4: Overall packet delay performance with multiple PUSCH opportunities.

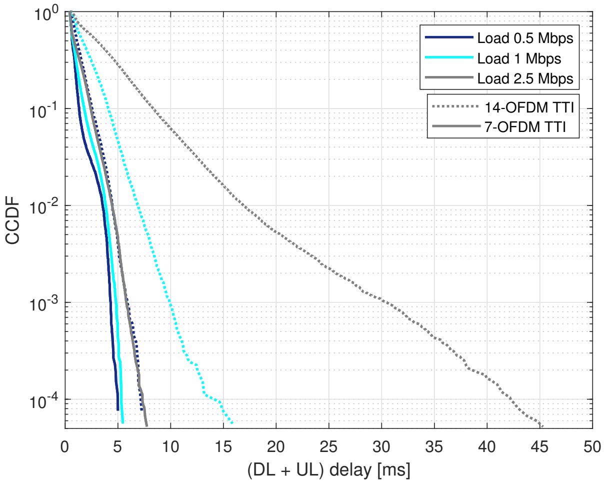


Fig. III.5: Overall packet latency performance comparing different TTI sizes.

Table III.2: Performance summary at 99.99 % reliability

Offered load	**MulteFire** 15 kHz, 14 OFDM	**NR-U** 30 kHz, 14 OFDM	**NR-U** 30 kHz, 14 OFDM $K = 2$	**NR-U** 30 kHz, 7 OFDM	**NR-U** 30 kHz, 7 OFDM $K = 2$
0.5 Mbit/s	45.64 ms	7.15 ms	7.11 ms	4.97 ms	4.75 ms
1 Mbit/s	77.29 ms	13.18 ms	11.54 ms	5.35 ms	5.22 ms
2.5 Mbit/s	192.6 ms	42.15 ms	33.15 ms	7.32 ms	6.11 ms

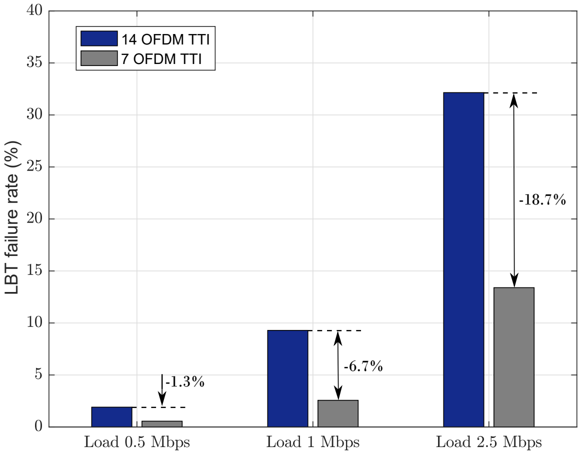


Fig. III.6: Average uplink LBT blocking probability for different TTI sizes.

2.5 Mbit/s offered load. As summary, the latency values achieved for each of the analysed technology enablers at 99.99 % reliability are included in Table III.2.

7 Conclusions

Through this paper different technology enablers for achieving high-reliable and low-latency communications in unlicensed standalone operation have been described and evaluated. As noted in Table III.2, a transition from the support of non-critical delay applications with MulteFire to the support of HRLLC/URLLC applications with optimized NR-U is shown. Comparing MulteFire with NR-U with 7-OFDM symbols TTI and multiple PUSCH occasions with $K = 2$ achieves, at 99.99-th percentile, a latency reduction factor of 9.6, 14.8 and 31.5 for 0.5 Mbit/s, 1 Mbit/s and 2.5 Mbit/s offered loads, respectively. Starting from MulteFire performance as a baseline, it has been described the substantial benefits that the flexible NR time-frequency design provides. Switching to higher SCS, which implies a shorten in the transmission duration, and using reduced processing times decreases the overall packet delay. In unlicensed bands, the fact that the channel is occupied for less amount of time is beneficial for coexistence with neighbour nodes. Supporting dynamic TDD with multiple switching points further improves the performance as multiple HARQ retransmission can be contained within a COT and shorter downlink-to-uplink gaps are achieved. Additionally, a time-diversity scheme for reducing the uplink LBT failure probability has been analysed. Multiple PUSCH occasions provide a reasonable improvement, especially at high offered load, where the uplink LBT failure rate impacts the performance. Finally, a shortening of the TTI is performed by reducing the TTI size from 14 OFDM symbols to 7 OFDM symbols. Significant latency reductions and decrease on uplink LBT blocking probability in all the considered loads are achieved.

References

[1] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Physical Layer Procedures," Technical Specification (TS) 36.213, 2018.

[2] M. Alliance, "MulteFire Release 1.0 Technical Paper: A New Way to Wireless," Tech. Rep. [Online]. Available: https://www.multefire.org/

[3] 3GPP, "Study on NR-based Access to Unlicensed Spectrum (Rel. 15)," Technical Report (TR) 38.889, 2018.

[4] ——, "Service requirements for cyber-physical control applications in vertical domains," Technical Specification (TS) 22.104, 2019.

[5] T. H. Jacobsen, R. Abreu, G. Berardinelli, K. I. Pedersen, I. Z. Kovács, and P. Mogensen, "Multi-cell reception for uplink grant-free ultra-reliable low-latency communications," *IEEE Access*, 2019.

[6] G. J. Sutton et al., "Enabling Ultra-Reliable and Low-Latency Communications through Unlicensed Spectrum," *IEEE Network*, 2018.

[7] R. Maldonado, C. Rosa, and K. I. Pedersen, "A Fully Coordinated New Radio-Unlicensed System for Ultra-Reliable Low-Latency Applications," in *IEEE Wireless Communications and Networking Conference*, 2019, manuscript submitted for publication.

[8] ETSI, *5 GHz RLAN; Harmonised Standard covering the essential requirements of article 3.2 of Directive 2014/53/EU*, ETSI EN 301 893 V2.1.1 (2017-05) Std.

[9] 3GPP, "Physical channels and modulation (Rel. 15)," Technical Specification (TS) 38.211, 2019.

[10] ——, "Physical Layer Procedures for Data (Rel. 15)," Technical Specification (TS) 38.214, 2019.

[11] R. Maldonado, C. Rosa, and K. I. Pedersen, "Latency and Reliability Analysis of Cellular Networks in Unlicensed Spectrum," in *IEEE Access*, 2019, Manuscript submitted for publication.

[12] 3GPP, “Physical layer procedures (Release 8),” Technical Specification (TS) 36.213, 2009.

[13] ——, “Study on channel model for frequencies from 0.5 to 100 GHz,” Technical Report (TR) 38.901, 2018.

[14] K. I. Pedersen et al., “Frequency Domain Scheduling for OFDMA with Limited and Noisy Channel Feedback,” in *2007 IEEE Vehicular Technology Conference*.

[15] K. Pietikainen et al., “System-Level Performance of Interference Suppression Receivers in LTE System,” in *2012 IEEE Vehicular Technology Conference*.

[16] 3GPP, “Feasibility Study on Licensed-Assisted Access to Unlicensed Spectrum,” Technical Specification (TS) 36.889, 2015.

Paper E

A Fully Coordinated New Radio-Unlicensed System for Ultra-Reliable Low-Latency Applications

Roberto Maldonado, Claudio Rosa, Klaus I. Pedersen

The paper has been published in the
IEEE Wireless Communications and Networking Conference (WCNC), 2020.

Abstract

Communications over the unlicensed spectrum are susceptible to be delayed by mandatory channel access mechanisms. Based on the need for improving the latency-reliability performance of New Radio-Unlicensed for supporting new use-cases, such as industrial applications, different types of channel access are evaluated in this paper. By using asynchronous and demand-driven channel access, it is shown that approximately 50 % *of the delay experienced by a downlink packet is due to listen before talk (LBT). Furthermore, UEs might be blocked by an unsuccessful uplink LBT losing the opportunity to transmit their previously scheduled data. As an alternative, synchronous channel access is evaluated. By using a coordinated LBT among the nodes, the channel access delay is reduced to a constant value. However, UEs can still be blocked when initiating their uplink transmissions due to neighbours transmissions. Motivated by this fact, an approach in which a central node is in charge of the frame selection is proposed. Therefore, all the nodes in the system are coordinated in both the channel access and the frame configuration. In this case, a latency reduction of* 70 % *as compared to the previously mentioned alternatives is achieved at high loads and* 99.9999 % *reliability.*

1 Introduction

Industry 4.0 strives to improve the efficiency, flexibility, versatility and usability of the factories of the future. In this context, wireless connectivity in the industries is one of the key enablers. Multiple promising applications within the domain of logistics, inventory management or robot and motion control are foreseen [1]. Many of the tasks supported in an industrial scenario need to fulfil very challenging quality of service requirements in terms of latency and reliability. For instance, motion control applications require a maximum delay of 0.5 ms with 99.999 % reliability. Other applications relax their delay constrains to 10 ms [2]. Depending on the requirements, applications can be classified as high-reliable/ultra-reliable and low-latency communications (HRLLC/URLLC).

Connectivity in the factories is currently dominated by wired solutions. Consequently, a transition towards a wireless approach is required. Licensed spectrum and unlicensed spectrum based wireless solutions can be adopted. The unlicensed alternative is gaining more interest mainly due to its global availability and simplicity of deployments as compared to licensed solutions. The usage of the unlicensed spectrum by a 3rd Generation Partnership Project (3GPP) compliant technology started with Licensed Assisted Access (LAA) [3]. Targeted for supporting enhanced mobile broadband applications, LAA uses the unlicensed spectrum as a supplementary resource over which part of the traffic carried in the licensed band can be offloaded. Additionally, MulteFire emerged as an alternative technology fully deployed in the unlicensed band, i.e. in standalone mode [4]. New Radio-Unlicensed (NR-U) is currently being designed and it is expected to support non-standalone mode of operation, as in LAA, and standalone mode, as in MulteFire [5].

Unlicensed spectrum is defined as a range of frequencies in which multiple radio access technologies (RATs) coexist. The channel usage must be equally shared among them, and therefore, regulatory bodies mandate nodes to behave according to specific requirements. For instance, limits on the power spectral density and the occupied channel bandwidth are specified. Furthermore, regulatory entities impose devices to detect other RATs activity before their transmissions and back-off in case of on-going channel activity detection. Once the channel access is gained, the time the channel can be continuously occupied is also restricted. This overall process is known as channel access mechanism and must be performed by any node prior to a transmission in the unlicensed spectrum. The European Telecommunication Standard Institute (ETSI) defines two channel mechanism designs: load based equipment (LBE) and frame based equipment (FBE) [6]. The main difference is that nodes perform the channel access asynchronously in LBE while synchronously in FBE. LBE provides fairer coexistence in presence of other RATs as compared to FBE, whereas FBE is considered more suitable for industrial applications under the assumption of single RAT operation.

Achieving the latency-reliability requirements for industrial applications in standalone mode over the unlicensed spectrum presents a challenge. Channel access mechanisms impose an additional delay to the transmissions. In [7], an evaluation of an opportunistic multi-channel access mechanism is presented. Authors propose a unlicensed-licensed dual band approach to redude the impact of channel access in the delay. For LBE type of operation, [8] and [9] analyse the impact of the channel access in the overall packet delay in downlink and uplink performance, respectively. Focusing on LBE and Multefire as baseline, [10] proposes different techniques to mitigate the contribution of the channel access to the overall packet delay. For FBE, the channel access delay is reduced due to the synchronicity in the LBT. In order to further improve the FBE latency-reliability performance, we propose a fully coordinated system. By using a central entity, nodes have a common LBT-transmission structure, i.e. FBE is adopted, and moreover, share the same frame configuration. A latency-reliability performance comparison between LBE, FBE and the fully coordinated approach is provided in this text. The remainder of the paper is structured as follows. Section 2 introduces the system model definition. Section 3 includes a description of LBE and FBE channel access mechanisms, while Section 4 describes our fully coordinated proposal. Simulation assumption and simulation results can be found in Sections 5 and 6, respectively. Final remarks are drawn in Section 7.

2 System model

A single-operator indoor office scenario with M next generation nodes-B (gNBs) and K user equipments (UEs) is assumed. Each node operates in standalone mode at the 5 GHz unlicensed band with 20 MHz bandwidth. The traffic model acts in accordance with the FTP Model 3 defined by 3GPP [11]. The packet generation follows a Poisson arrival process with an average packet arrival rate of λ_T measured in packets/s/UE. Bi-directional traffic with a fixed packet size of B bytes is assumed. Payloads in downlink are generated with a rate of λ_{DL} while UEs generate uplink packets with an average arrival rate of λ_{UL}. Both packet arrival rates compose the overall packet arrival rate λ_T. The average offered load in bit/s can be obtained as $B \cdot 8 \cdot \lambda_T \cdot K$

Dynamic time domain duplexing (TDD) with 14 OFDM transmission time interval (TTI) is adopted. gNBs perform a frame selection algorithm based on the current buffer status of its connected devices. Depending on the ratio between the number of buffered downlink and uplink packets, the downlink/uplink frame ratio in terms of slots for the next TDD frame is chosen. Dynamic slot configuration and multiple switching points during the TDD frame are allowed. Having multiple downlink-to-uplink and uplink-to-downlink transitions provides more opportunities to, for instance, send hybrid automatic request (HARQ) feedback and scheduling request (SR) messages. A transition slot is needed to switch between downlink and uplink slots, referred throughout the text as special slot. The special slot consists of a 9-OFDM symbol downlink TTI followed by a 1-OFDM symbol gap for RF switching and a 4-OFDM symbol uplink TTI. The uplink slot is used for short uplink control signalling transmissions such as HARQ feedback or SR messages and it is refereed throughout the text as short physical uplink control channel (sPUCCH). An uplink-to-downlink transition during the COT is considered to be feasible without a special slot. An example of the operation in NR-U is pictured in Fig. III.1.

3 Channel access mechanisms

3GPP adopts Listen Before Talk (LBT) as the CCA mechanism for LAA and NR-U technologies. LBT is an algorithm performed in intervals of 9 µs, commonly known as CCA slots. During each sensing interval, the node decides about the channel activity based on a power measurement and a posterior comparison with a predefined energy detection (ED) threshold. The channel is assumed to be occupied if the detected interference is above the ED threshold. ETSI defines two types of channel access mechanisms [6]. Both are based on the LBT principle but its application is differently implemented.

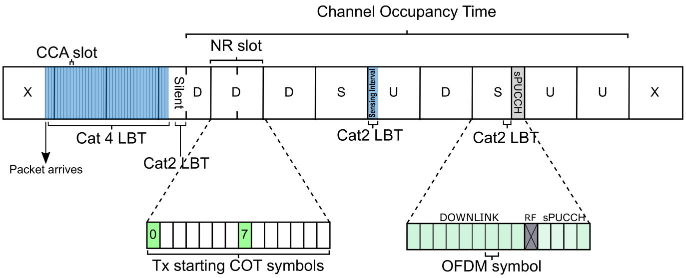


Fig. III.1: New Radio-Unlicensed mode of operation following LBE approach.

3.1 Load Based Equipment

On one hand, nodes can act following the LBE guidelines. The LBE approach follows an asynchronous and demand-driven design. Nodes execute the channel access procedure as soon as there is data available to transmit. LBE nodes perform a random back-off algorithm with variable contention window size. LBT is iteratively performed over a period of at least N consecutive CCA slots, also known as the contention window. Each node independently generates N from a uniform distribution defined between 0 and a maximum contention window size. This process is known as category 4 (Cat4) LBT. Once Cat4 LBT finished, the transmission can start and the channel can be constantly occupied for a maximum interval of time defined as maximum channel occupancy time (MCOT). Different configuration of maximum contention window sizes and MCOT durations are defined by channel access priority classes (CAPCs) [3]. Nodes performing Cat4 LBT are defined as initiating devices and can grant access to other nodes, known as responding devices, to transmit in their previously acquired channel occupancy time (COT). This is denoted as COT sharing and responding devices are mandated to perform a single-shot LBT procedure over a fixed sensing interval of 25 µs. This is known as category 2 (Cat2) LBT. Responding devices need to convey its transmission within the initiating devices MCOT limits. Cat2 LBT can be avoided under the condition of having a gap between the end of the initiating device transmission and the start of the responding transmission is shorter than 16 µs. This type of LBT is known as category 1 (Cat1) LBT. Whereas LBT is performed over CCA slots, NR-U has OFDM symbol granularity. Moreover, NR-U transmissions may only start in certain OFDM symbols within a slot. Therefore potential misalignment between the end of the LBT and the start of the transmission may occur resulting in a gap. During this period, initiating nodes remain silent waiting for a starting Tx OFDM symbol to start the COT. Before start transmitting they need to perform an additional Cat2 LBT as shown in Fig. III.1.

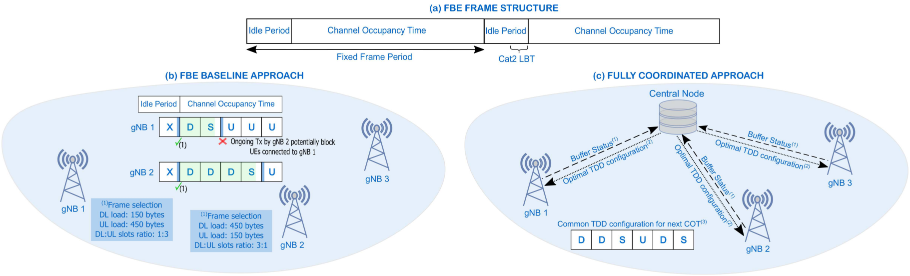


Fig. III.2: Frame-based equipment: Frame structure in (a), FBE baseline mode of operation in (b) and fully coordinated mode of operation in (c).

3.2 Frame Based Equipment

Nodes can also act according to the FBE guidelines and follow a common sense/transmit structure with a predefined periodicity. The FBE structure is depicted in Fig. III.2(a) and it is defined by:

- Fixed frame period (FFP): describes the total interval of time including the channel occupancy time and the idle period. Its duration is restricted to the range from 1 ms to 10 ms. Transmissions must start at the beginning of the FFP.
- Channel occupancy time (COT): defines the interval of time which a node can continuously transmit on a given channel without re-evaluating the channel availability. Its duration is limited, at maximum, to the 95 % of the FFP and it must be followed by an idle period.
- Idle period (IP): contains the single observation slot where CCA is performed and its duration must be, at least, 5 % of the COT with a minimum of 100 µs.

The main difference between FBE and LBE is the channel access for the initiating devices. For FBE, Cat2 LBT is used by the initiating devices and it is performed during the IP. An unsuccessful LBT produces a blockage in the transmission and nodes need to wait until the next FFP to perform the next channel access attempt. Responding devices follow the same guidelines as in LBE. COT sharing among FBE devices is possible with Cat2 LBT. Additionally, Cat1 LBT is allowed if the gap condition is met.

4 A fully coordinated approach

After a successful channel access, the frame configuration for the next COT is selected based on the current traffic demands. In dynamic TDD systems, each initiating device, i.e. the gNBs following our assumptions, chooses its optimal frame configuration independently. Especially in FBE deployments, where multiple gNBs start their transmissions simultaneously, uncoordinated frame selection impacts the system performance. Due to the adaptation of the frame configuration based on the traffic conditions, different nodes of the same network may apply DL and UL slots at the same time. This increases the probability that UEs sense the channel as busy during the Cat2 LBT due to neighbour gNBs transmissions. This supposes a lost opportunity for transmitting over the previously granted uplink resources as it is noted in Fig. III.2(b). In absence of expected uplink transmission, the gNB assumes discontinuous transmission (DTX) as a negative acknowledgement (NACK) and sends a new grant to the UE. This process is repeated until the uplink transmission is successful, which ultimately degrades the uplink latency performance. As a further enhancement to the FBE approach, we propose a fully coordinated and centralized system in which both LBT and frame configuration are common to every deployed node in the system. By having a common TDD frame configuration, mutual blocking in the Cat2 LBT among neighbour nodes is avoided as the sensing intervals are time-synchronized. This is achieved by agreeing in the TDD frame configuration for the COT. It is assumed that a central node is in charge of providing the selected frame configuration for the following FFP to every gNB. The proposed procedure is as follows: during each IP, each gNB computes its downlink-uplink ratio from the data in the buffer and feedbacks it to the central node. Based on the inputs from the different gNBs, the central node decides the frame structure that suits best for the current traffic conditions of all the gNBs. The decision is signalled to the corresponding gNBs and it is applied during the next COT. This mode of operation is depicted in Fig. III.2(c). Removing potential blocking transmissions at the UE side brings benefits in terms of latency performance, however the central node proposes a sub-optimal frame configuration that might not perfectly fit the traffic requirements of certain gNBs. Moreover, the complexity of the decision algorithm increases with the number of nodes.

5 Simulation assumptions and methodology

A set of 12 deployed gNBs is assumed in our industrial indoor scenario. They are organized in 2 rows of 6 gNBs with a separation of 20 m between gNBs, which follows the guidelines for indoor-office scenarios defined in [12]. In order to approach a private deployment, single operator conditions are assumed. 50 UEs are distributed around the scenario and they are connected to the optimal gNB based on the strongest reference signal received power (RSRP) criteria. A maximum of 5 UEs can be connected to each gNB.

Table III.1: Simulation assumptions

Parameter	**Assumption**
Layout	Indoor mixed office scenario [12]
Channel model	NR InH Mixed Office [12]
Duplexing mode	dynamic TDD
Bandwidth	20 MHz @ 5 GHz band
Sub-carrier spacing	30 kHz
TTI	14 OFDM symbols
Scheduling metric	Proportional fair
HARQ	Asynchronous HARQ, incremental redundancy 6 retransmissions at maximum
UEs processing times*	PDSCH: 4.5 OFDM symbols [13] PUSCH: 5.5 OFDM symbols [13]
Link adaptation	Outer link loop adaptation: enabled [14] BLER target: 1 %
Receiver type	LMMSE-IRC [15]
MIMO	2×2 configuration DL: Rank-2 SU-MIMO UL: Rank-1 SU-MIMO, receiver diversity
Traffic model	FTP Model 3; B = 50 bytes $\lambda_T = \{25, 50, 125, 250\}$ packets/s/UE
gNBs channel access	Cat4 LBT for acquiring COT Cat2 / Cat1 LBT for reaccesing during COT
UEs channel access	Cat2 / Cat1 LBT for sPUCCH Cat2 LBT for PUSCH
ED threshold	−72 dBm

* gNBs processing times are assumed 2 times faster than UEs capabilities

UE re-dropping is performed in case of reaching the maximum number of connected UEs per gNB. The radio channel propagation model follows the guidelines for indoor mixed office scenarios defined by 3GPP for NR simulations [12].

Channel access mechanisms are heavily impacted by the scenario interference and, thus, by the traffic model. In order to measure its impact on the system performance, a fixed packet size of 50 bytes (B = 50) and a variable packet arrival rate (λ_T) are assumed. Four different offered loads are simulated, specifically, 0.5 Mbit/s, 1 Mbit/s, 2.5 Mbit/s and 5 Mbit/s. These traffic loads correspond to a packet arrival rate of 25, 50, 125 and 250 packets/s/UE, respectively. The traffic generation follows a ratio of 80 %-20 % for downlink and uplink packets, respectively.

To establish a fair comparison between LBE and FBE, both schemes support the same MCOT duration. An MCOT of 3 ms is assumed. For LBE deployments, nodes perform LBT following the CAPC 2 conditions, which assumes 7 and 15 CCA slots as possible maximum contention window sizes [3]. For FBE, an FFP of 3.5 ms and an IP duration of 0.5 ms is simulated. The remaining 3 ms of the FFP are dedicated for transmissions. In both types of channel access, gNBs act as initiating devices whereas UEs can only transmit based on COT sharing approach, i.e. they are responding devices. Simulations assume that Cat2 LBT can be avoided under the conditions explained in Section 3, i.e. Cat1 LBT is allowed. Separate uplink LBT procedures are needed for sPUCCH and uplink slot resources. A successful LBT before sPUCCH resources allows UEs to occupy the channel for 4 OFDM symbols, i.e. the sPUCCH duration. A successful LBT prior to uplink transmissions allows UEs to continuously access the channel for one or more consecutive slots. gNBs can initiate the transmissions at OFDM symbols 0 and 7 of each slot. In case of Cat4 LBT being successful on a different symbol within the slot, the gNBs remain silent until the next starting symbol. gNBs can also re-access the channel within the previously acquired COT after a successful Cat2 LBT or Cat1 LBT if gap condition is met.

Grant-based uplink is chosen as the scheme for uplink transmissions. Therefore, the scheduling request (SR) procedure is performed before any physical uplink shared channel (PUSCH) transmission. SR opportunities are configured to occur in every slot. Each transmission involved in SR procedure is subjected to LBT outcome and TDD constrains. Thus, given a successful LBT, SR transmissions are performed in special slots whereas grant transmissions occur in downlink slots. gNBs processing times for decoding a scheduling request message and preparing a grant are neglected. To fulfil occupied channel bandwidth restrictions defined by ETSI [6], the minimum frequency allocation is an interlace. Each interlace consists of a set of physical resource blocks (PRBs) equally distant in frequency. Among the multiple interlace designs specified by 3GPP in [5], a configuration with 5 interlaces of 10 PRBs each is assumed. Upon the grant reception, UEs process it and prepare the uplink transmission during a time interval denoted as PUSCH processing time. For downlink operation, UEs need to decode the data transmitted in the physical downlink share channel (PDSCH) for a interval of time known as PDSCH processing time. Thereafter, the packet delay statistics are collected. Processing times at the gNB side are also taken into account. A summary of the simulation assumptions is included in Table VI.2.

The simulator models with high level of details the majority of the PHY and MAC layer procedures in line with 3GPP guidelines while operating on an OFDM symbol-subcarrier resolution. It is capable of, for instance, dynamically schedule UEs in the time-frequency domain, use HARQ in case of decoding failures and perform link adaptation to fulfil a target block error rate (BLER). Additionally, it includes proven stochastic models for radio propagation calibrated against similar models used in 3GPP. To get statistically stable and reliable results, multiple realizations of the scenario with sufficient samples are simulated. For each realization, the UEs locations are independently selected over the layout. Results from each realization are combined afterwards. Simulations times are adjusted based on λ_T to provide a constant number of generated packet for each considered traffic load. 1.000.000 packets are generated for every offered load. For better showing the rare-occurrence events that are evaluated, complementary cumulative distribution functions (CCDF) are used.

6 Performance Results

Firstly, a comparison between the channel access delay of LBE and FBE deployments is established. The channel access delay for FBE deployments provides a constant delay of 25 µs, i.e. the duration of Cat2 LBT, for every considered load and at any reliability level. This is due to the fixed structure defined for FBE deployments. However, frame alignment delay, i.e. the interval of time between the packet arrival and the next IP, needs to be also considered. Since the MCOT duration is assumed to be 3 ms, the frame alignment can be modelled as a uniform distribution between 0 and 3. Therefore, in the worst-case scenario, the contribution of channel access and frame alignment is 3.025 ms. In contrast, as depicted Fig. III.3, the channel access delay for LBE deployments shows a different behaviour. It is noted that the time spent in performing Cat4 LBT is dependant with the considered offered load and higher than the delay experienced in FBE. One fact worth to mention is the lower access delay shown at 5 Mbit/s as compared to the rest of the loads in approximately 80 % of the cases. This seems counter-intuitive since the higher is the load, the higher is the likelihood of finding the channel busy, thus, the time performing Cat4 LBT increases. However, faster channel access is achieved by unplanned coordination of the gNBs. Synchronicity is introduced by the fact that transmissions are deferred until OFDM symbols #0 or #7 are reached.

The channel access mechanism is one of the contributors to the overall latency. To measure how much the channel access contributes to the packet delay, we obtain the percentage of time that gNBs spend accessing the channel as compared to the overall packet delay. In Fig. III.4 the accumulated channel access delay and the downlink latency is plotted for 2 sets of offered loads: 1 Mbit/s and 5 Mbit/s. It is noted that a large portion of the downlink delay is caused by the LBT. Specifically, for a reliability level of $1 - 10^{-3}$, the channel access delay contributes with approximately 54 % and 49 % of the packet delay for 1 Mbit/s and 5 Mbit/s, respectively. At $1 - 10^{-4}$, the contribution increases up to 61 % for 1 Mbit/s and keeps constant as compared to

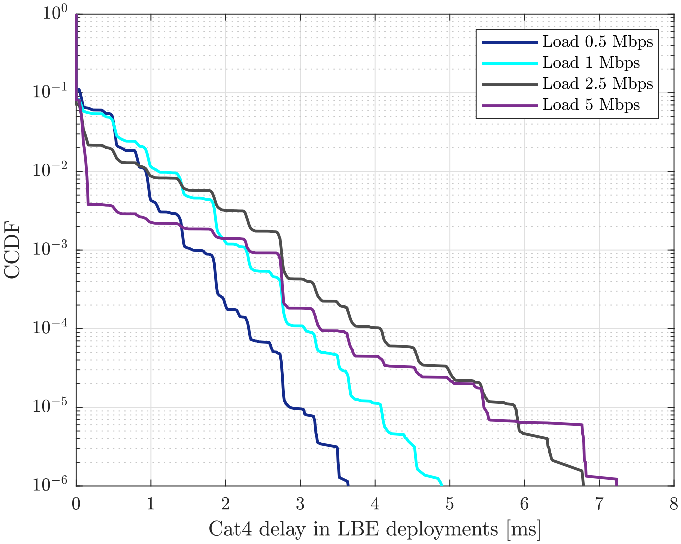


Fig. III.3: Delay associated with a single Cat4 LBT for LBE deployments.

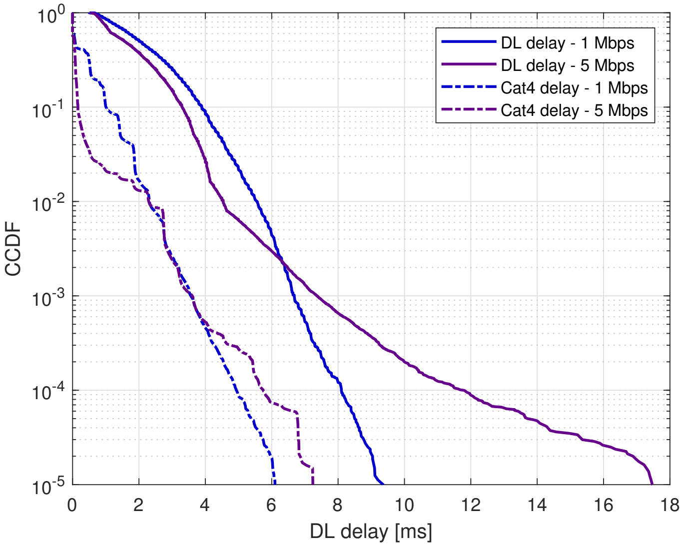


Fig. III.4: Overall downlink packet delay and Cat4 LBT channel access delay.

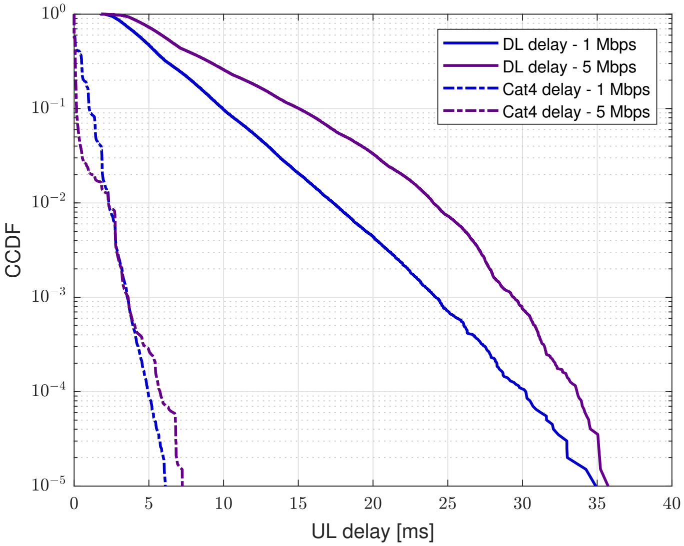


Fig. III.5: Overall uplink packet delay and Cat4 LBT channel access delay.

previous reliability for 5 Mbit/s.

On the other hand, the uplink packet delay is significantly higher than the time consumed by the gNBs accessing the channel as noted in Fig. III.5. The contribution of Cat4 LBT in the uplink latency decays to 15% and 12% for the considered loads at $1 - 10^{-3}$. It is noted by comparing the overall delays depicted in Fig. III.4 and Fig. III.5 that the uplink delay is the main delay component, highly harming the system performance. The uplink delay performance is mainly impacted by two contributors: the SR procedure and the Cat2 LBT. To assess the impact of uplink LBT, the probability of being blocked by Cat2 LBT in uplink slots is plotted in Fig. III.6. For LBE, the LBT failure rate increases as the load increases. Two main factors cause this behaviour. Firstly, the hidden node problem. This problem occurs when a node, in this case, a UE, is blocked by a neighbour node that was not previously detected by its serving gNB when performing Cat4 LBT. Secondly, in some cases, a silent gap between the grant transmission and the actual uplink transmission gives room for neighbour nodes to access the channel. If a neighbour gNB gets access to the channel during this gap, UEs can be potentially blocked. FBE without frame coordination does not suffer from hidden node problem but, as introduced in Section 4, it shows high Cat2 LBT failure rate, especially at high offered loads, due to uncoordinated frame selection. FBE with frame coordination avoids these drawbacks and shows a completely different performance. The coordinated TDD structure makes

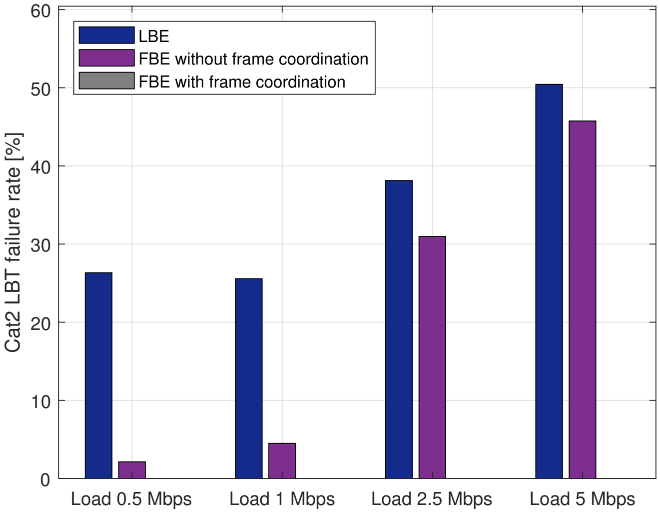


Fig. III.6: Cat2 LBT blocking probability for PUSCH transmissions.

UEs perform the Cat2 LBT simultaneously while the rest of the nodes are silent, achieving 0 % blocking probability at any considered offered load.

During this analysis, the benefits of a fully coordinated and centralized deployment have been highlighted. Before concluding, a comparison in the overall packet delay for both downlink and uplink is presented in Fig. III.7. It is noted that the delay performance of LBE and FBE without frame coordination deployments are dependent on the considered load whereas FBE with frame coordination deployments show approximately the same behaviour for both loads. As shown in previous figures, LBE suffers from non-constant channel Cat4 LBT access delay and high Cat2 LBT blocking probability. Alternatively, the performance of FBE without frame coordination is still heavily impacted by high Cat2 LBT blocking probability, especially at high loads. It is noted that baseline FBE with 1 Mbit/s shows similar performance as FBE with coordination as the Cat2 LBT blocking probability is quite low (see Fig. III.6). Comparing the latency achieved by both schemes we can conclude that FBE with frame coordination outperform the other alternatives. Specifically, for 1 Mbit/s offered load, a reduction of 52 % and 59 % is achieved by switching from LBE to FBE with frame coordination at reliability levels of $1 - 10^{-3}$ and $1 - 10^{-4}$, respectively. As compared to FBE without frame coordination, reported benefits are 25 % and 29 % for 1 Mbit/s. For 5 Mbit/s, the delay is further enhanced as compared to both alternatives, achieving a reduction of approximately 70 % at $1 - 10^{-3}$ and $1 - 10^{-4}$ reliability levels.

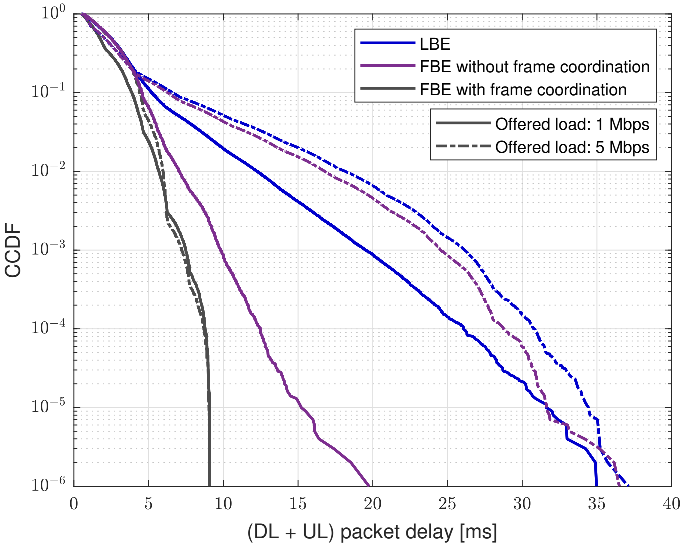


Fig. III.7: Overall packet delay for LBE, FBE without frame coordination and FBE with frame coordination.

7 Conclusions

In this paper, the two types of channel access defined by ETSI has been analysed from a latency-reliability perspective. It has been shown that, under the conditions of single-operator scenario and no inter-RAT interference, FBE outperforms LBE due to its predefined LBT-transmit-receive structure. Additionally, it has been verified that the usage of LBE for supporting HRLLC/URLLC type of applications is quite restricted since the channel access delay already surpass some of the latency requirements defined for industrial automation. Using FBE remarkably reduces channel access delay at the gNB side but still suffers from high uplink LBT blocking probability which impacts on the uplink delay performance. To mitigate this effect, a central-node approach has been proposed. By having a master entity which decides the frame configuration for the next COT, the uplink LBT success rate is increased to 100 %. A design with coordination in LBT and frame configuration has been shown as the best option from a latency-reliability point of view. This approach can serve downlink and uplink traffic with an upper-bound delay lower than 10 ms. A notable delay reduction of 66 % and 30 % is achieved as compared to LBE and FBE without coordination at low loads at 99.9999 % reliability, respectively. At high loads, the improvement increases up to 70 % for both cases.

References

[1] 5G-ACIA, “5G for Connected Industries and Automation - White Paper, 2nd Edition,” 2018.

[2] 3GPP, “Service requirements for cyber-physical control applications in vertical domains,” Technical Specification (TS) 22.104, 2019.

[3] ——, “Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Physical Layer Procedures,” Tech. Rep. 36.213, 2018.

[4] M. Alliance, “MulteFire Release 1.0 Technical Paper: A New Way to Wireless,” Tech. Rep. [Online]. Available: https://www.multefire.org/

[5] 3GPP, “Study on NR-based Access to Unlicensed Spectrum (Rel. 15),” Tech. Rep. 38.889, 2018.

[6] ETSI, *5 GHz RLAN; Harmonised Standard covering the essential requirements of article 3.2 of Directive 2014/53/EU*, ETSI EN 301 893 V2.1.1 (2017-05) Std.

[7] G. J. Sutton et al., “Enabling Ultra-Reliable and Low-Latency Communications through Unlicensed Spectrum,” *IEEE Network*, 2018.

[8] R. Maldonado, C. Rosa, F. Frederiksen, and K. I. Pedersen, “On the Impact of Listen-Before-Talk on Ultra-Reliable Low-Latency Communications,” in *2018 IEEE Global Communications Conference*, 2018.

[9] ——, “Uplink Ultra-Reliable Low Latency Communications Assessment in Unlicensed Spectrum,” in *IEEE Globecom Workshops*, 2018.

[10] ——, “Latency and Reliability Analysis of Cellular Networks in Unlicensed Spectrum,” in *IEEE Access*, 2019, Submitted for publication.

[11] 3GPP, “Feasibility Study on Licensed-Assisted Access to Unlicensed Spectrum,” Tech. Rep. 36.889, 2015.

[12] ——, “Study on channel model for frequencies from 0.5 to 100 GHz,” Tech. Rep. 38.901, 2018.

[13] ——, “Physical Layer Procedures for Data (Rel. 15),” Tech. Rep. 38.214, 2019.

[14] K. I. Pedersen et al., “Frequency Domain Scheduling for OFDMA with Limited and Noisy Channel Feedback,” in *2007 IEEE VTC*.

[15] K. Pietikainen et al., “System-Level Performance of Interference Suppression Receivers in LTE System,” in *2012 IEEE VTC*.

Part IV

URLLC over unlicensed spectrum in hostile environments

Multi-link techniques for performance improvement in hostile environments

The previous studies in Part II and Part III were conducted under controlled environments conditions. Therefore, scenarios with a single radio access technology and fully isolated from any other network were assumed. In such a scenario, the only sources of interference are the transmissions from the same network, i.e. intra-system interference. This type of interference was demonstrated to be sufficient to severely impact the channel access delay, and hence, the URLLC performance. Solutions to strive the performance have been discussed and analyzed in previous chapters. Especially, coordination was proved as an enabler for URLLC under the aforementioned conditions. In this part, we add more complexity to the problem and extend our analysis to more hostile environments by adding the presence of inter-system interference in, at least, part of the available unlicensed spectrum. The goal is to search for mechanisms that ensure stringent latency and reliability in scenarios where the unlicensed bands are partially shared among multiple radio access technologies.

1 Problem Description

In deployments where single-RAT and single network conditions are ensured, i.e. controlled environment conditions, only transmissions from a single technology (MulteFire and NR-U in our studies) can affect the system performance by, for instance, delaying the access to the channel. This assumption is no longer valid in this analysis. Here we assume the presence of multiple radio access technology utilizing the same frequency bands. These scenarios are also referred to as uncontrolled or hostile environments. In this set-up, the channel availability is reduced as the frequencies are shared

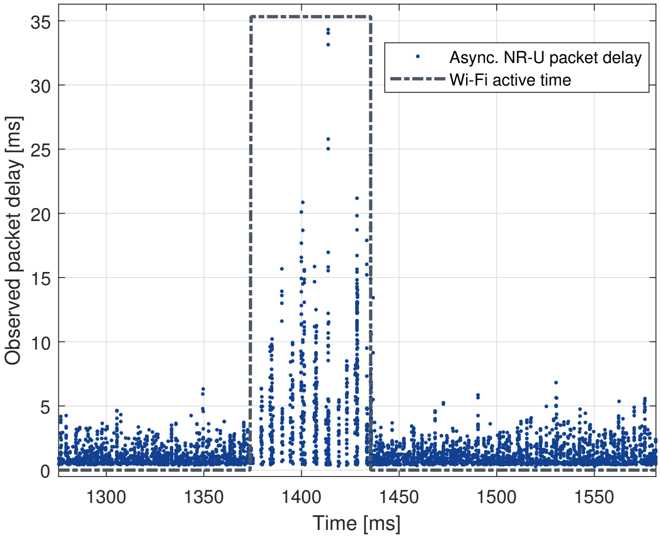


Fig. IV.1: Snapshot of the NR-U packet delay over time in uncontrolled envinroment conditions. Asynchronous channel access is adopted in both technologies. NR-U carries URLLC traffic (5 Mbit/s offered load) whereas IEEE 802.11ax carries eMBB.

among multiple RATs. For these studies and due to its widespread usage Wi-Fi, namely Wi-Fi 6 (IEEE 802.11ax), is the selected radio access technology to coexist with an NR-U network. To highlight the addressed problem, the observed NR-U packet delay versus the time in uncontrolled conditions is depicted in Fig. IV.1. The picture focuses on the time interval in which Wi-Fi is actively accessing the channel. Since both technologies share the same 20 MHz channel, it is noted that the presence of Wi-Fi imposes a clear degradation on the NR-U packet delay. Even though the Wi-Fi transmissions are sporadic and therefore only occur during a very limited time interval, it is enough to limit the supported URLLC QoS.

In the example presented in Fig. IV.1, NR-U and Wi-Fi adopted the asynchronous channel access design. On the other hand, as detailed in Part III, using the synchronous channel access design in NR-U deployments is the prefered option for fulfilling stringent latency and reliability requirements. However, the presence of more than one RAT in the vicinity of the NR-U deployment might lead to a change in the drawn conclusions. The latency-reliability performance of NR-U synchronous channel access in controlled and uncontrolled conditions is shown in Fig. IV.2. The NR-U performance with the asynchronous channel access is also added for comparison. It is observed that due to a channel availability reduction, the presence of Wi-Fi increases the NR-U packet delay for both channel access designs. For the

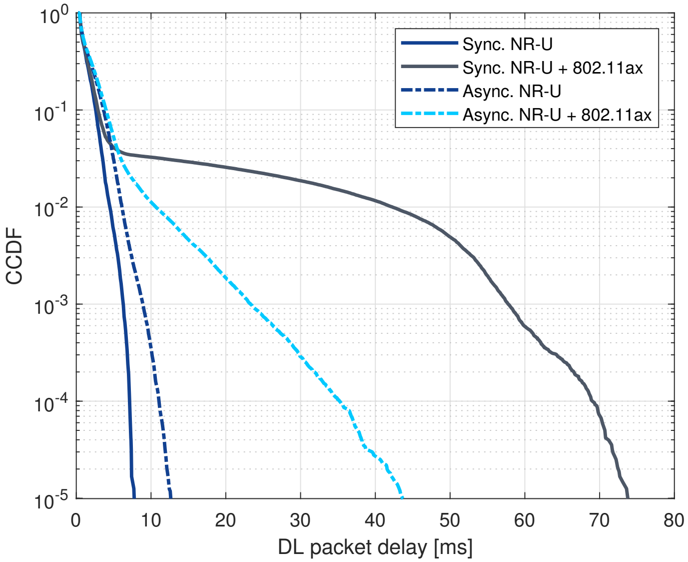


Fig. IV.2: CCDF of the NR-U packet delay in controlled and uncontrolled environments. Synchronous/asynchronous channel access and URLLC traffic (with 5 Mbit/s offered load) are assumed for NR-U. Asynchronous channel access and eMMB traffic for IEEE 802.11ax.

asynchronous channel access, three-fold latency degradation is observed. For the synchronous channel access, it is noted that the presence of Wi-Fi is especially critical. NR-U experiences a latency degradation by a factor of ∼9.7, i.e. approximately 3 times larger than the asynchronous case. The reason for this behaviour is due to differences in the channel access mechanisms among technologies. NR-U implements the synchronous channel access, in which the channel is only acquired on specific time intervals. On the other hand, Wi-Fi adopts the asynchronous approach which allows for channel acquiring at any time. Thus, Wi-Fi potentially acquires the channel during the NR-U reserved periods for sensing the channel. Detecting the presence of Wi-Fi forces NR-U transmissions to be postponed, at least, until the next sensing period. Consecutive unsuccessful channel access during the sensing period lead to large channel access delay and, therefore, high packet delays. In summary, regardless of the adopted channel access strategy for NR-U, a clear degradation on the latency and reliability performance is observed in the presence of Wi-Fi. Given the observed behaviour, the research in this part of the thesis aims to seek solutions to mitigate the effect of Wi-Fi transmissions on the NR-U latency and reliability. Specifically, frequency diversity techniques are investigated as mechanisms to reduce the channel access delay, and in turn, the observed packet delay. For that, a multi-channel NR-U deployment is proposed. It is assumed that, at any time, at least one of the available NR-U channels is completely free of Wi-Fi interference. This research topic aligns with the current 3GPP's views. During the study phase for Rel-17 for IIoT/URLLC [1], 3GPP agreed on the need for supporting URLLC-u in scenarios containing only devices from the same RAT and where unexpected interference from any other wireless technology happens sporadically.

2 Objectives

The research activities conducted in this part have the following main objectives:

- Analyze the impact, from a latency and reliability perspective, of coexisting with multiple radio access technologies operating in the same unlicensed bands.
- Propose frequency diversity techniques to cope with the experienced performance degradation in uncontrolled environments. The target is to provide a solution that ensures a latency-reliability performance comparable with cases with controlled interference conditions. A multi-channel NR-U deployment with at least one of the channels free of Wi-Fi interference is assumed.

3 Included Articles

The following articles form the main body of this part of the thesis:

Paper F. Multi-link techniques for New Radio-Unlicensed URLLC in hostile environments

This paper motivates the importance of seeking for solutions to cope with the presence of inter-RAT interference in a NR-U deployment. It firstly shows that a clear degradation on the latency-reliability is experienced when Wi-Fi is actively transmitting. Under these circumstances, the only possibility to maintain a URLLC QoS in the NR-U network is to use additional frequency resources that are not occupied by Wi-Fi. In order to execute a transmission over multiple-channels LBT is required to be successful in each 20 MHz band. Different multi-link strategies are investigated. The first mechanisms consist of splitting the users among the 2 available channels. Users are only capable of receive data in a single 20 MHz channel. In a further proposal, users can be scheduled on both channels. This resembles the carrier aggregation framework standardized by 3GPP. In both previously described schemes, a packet scheduled in a channel cannot be re-scheduled in a different channel.

Freely rescheduling the packets might be very useful in the sporadic occasions in which Wi-Fi is active. Especially, since it is always assumed that one of the available channels is free of Wi-Fi interference. Therefore, a multi-link scheme with flexible channel switching is also analyzed. This can be achieved by adopting wideband operation or PDCP duplication. All the analyzed proposals are compared against the controlled and uncontrolled 20 MHz baseline NR-U. Moreover, the importance of LBT synchronization boundary (LSB) is evaluated for each of the techniques. LSB is a key parameter configuration in shared RF chain multi-band deployments and defines the amount of time a channel, with a successful LBT, waits for its adjacent channel to finish the channel access and execute a multi-band transmission. If the LSB expires, the node decides not to wait for the channel with on-going Type 1 switch to single-band transmission.

4 Main Findings

NR-U latency performance degradation in presence of Wi-Fi

The impact of unexpected and sporadic interference in an NR-U deployment is evaluated by extensive system-level simulations. In the simulated industrial set-up, a latency degradation by a factor of $\sim$3.2 is observed when NR-U and Wi-Fi fully overlap their channels in the 5 GHz band. Apart from the channel availability reduction, it is worth highlighting that both technologies do not contend the channel with the same conditions. First, there are differences in the adopted energy detection (ED) threshold. The ED threshold defines the maximum allowed interference that a node can sense to declare the channel as idle. Wi-Fi, since it is the incumbent technology in the 5 GHz band, it uses -62 dBm. In contrast, NR-U adopts -72 dBm. This difference makes Wi-Fi less vulnerable to any NR-U transmission, and therefore, it results in a much faster channel access. In order to highlight the importance of the ED threshold, in Fig. IV.3 the packet delay of an NR-U deployment under different ED configurations for Wi-Fi is shown. It can be noted that when all the nodes in the system use the same ED threshold, the impact of the Wi-Fi interference is diminished. A decrease in the latency of $\sim$38 % is observed with a common ED threshold of -72 dBm. The usage of a common LBT threshold for any unlicensed technology has been recently agreed for operation in the 6 GHz band [2]. This ensures a more fair coexistence among the technologies sharing the same frequency bands. Furthermore, NR-U is a synchronous technology as therefore it follows a specific time reference. This might introduce a misalignment between the time the channel access is finished and the first potential start of the transmission. However, Wi-Fi is capable of actively using the channel as soon as the channel access is success-

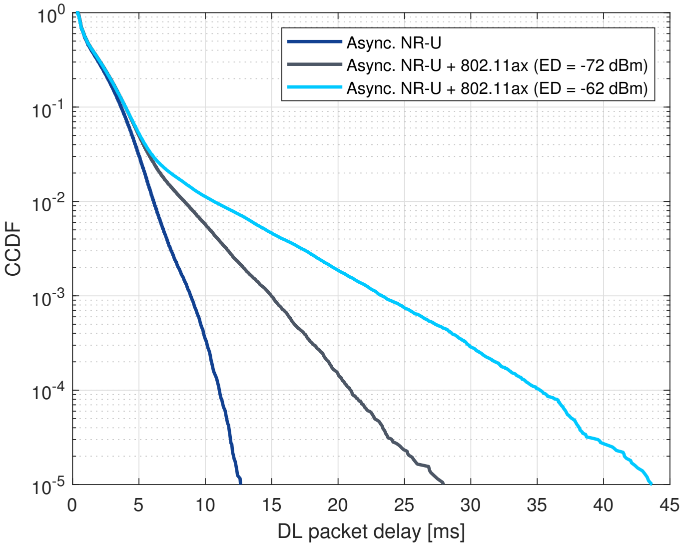


Fig. IV.3: CCDF of the NR-U observed delay in controlled and uncontrolled environments. Different ED threshold configurations are assumed for IEEE 802.11ax.

ful. In summary, Wi-Fi is more aggressive in the channel access which makes the NR-U latency and reliability to be heavily impacted.

Multi-link techniques as a solution to maintain certain URLLC requirements

For multi-link techniques with a tight association between packets and scheduled channel, it has been shown that choosing a very large LSB is advisable. These techniques are not able to re-schedule a packet upon a scheduling decision. Thus, a large LSB gives room for both channels to successfully acquired the channel in presence of uncontrolled interference. Otherwise, the packets scheduled in one of the channels are delayed for, at least, the duration of the transmission in the adjacent channel. This limitation is motivated by the fact, due to impairments in the gNBs RF chains, simultaneous transmissions and sensing in the adjacent channel is not possible. Even under this limitation, the multi-link techniques show better latency-reliability performance as compared to baseline 20 MHz NR-U in uncontrolled conditions. A split of the offered load into the available channels is the motivation for such improvement. However, it is far from fully mitigating the uncontrolled interference since the reported delays are approximately 2.3 times higher than the 20 MHz NR-U controlled case. Frequency diversity techniques with flexible channel re-selection are presented as a better alternative to overcome the addressed problem. It is shown that with such an assumption, it is suitable to choose the most aggressive LSB configuration. Our priority is on latency, and therefore, the optimal strategy is to use the firstly acquired channel and forward (re-select the channel) the packets initially scheduled in the adjacent channel to the active band. By the adoption of this technique, a 2x 20 MHz NR-U deployment exhibits a latency reduction of 83 % as compared to baseline 20 MHz NR-U, under the same levels of uncontrolled interference. Moreover, thanks to the channel access diversity in the multi-channel deployment, a latency decrease of 34 % with respect to the single-RAT single-carrier NR-U is obtained. All the observed trends are obtained by comparing latency values at 99.99 % reliability levels.

5 Key recommendations

The following recommendations are found to be essential to achieve stringent latency and reliability QoS under hostile environments conditions:

- In scenarios with full frequency overlapping with nearby Wi-Fi devices, or any other RAT competing for the same channels as the NR-U deployment, follow the asynchronous channel access guidelines. It provides better resilience to inter-RAT interference as compared to synchronous channel access design. However, the achievement of URLLC performance in this scenario is practically impossible.
- An effective solution to mitigate the effect of Wi-Fi, while still being able to fulfill some URLLC requirements, is to seek for frequency bands free of interference. This can be achieved by carrier aggregation, wideband operation or PDCP duplication or a combination of those.
- Among the possible alternatives, carrier aggregation - as currently specified by 3GPP - is limited by the lack of channel re-selection after a scheduling decision. Therefore, it is recommended to use a technique that allows for fast channel switching during the scheduling procedure. Wideband operation and PDCP duplication can accomplish it.
- Among these 2 solutions, wideband operation is the preferred option due to higher spectral efficiency. PDCP duplication, if no further enhancements are applied in the duplication algorithm, might present a bottleneck in highly loaded scenarios.

References

[1] 3GPP, “Enhanced Industrial Internet of Things (IoT) and URLLC support,” Tech. Rep. RP-193233, 2019.

[2] ETSI, *6 GHz RLAN; Harmonised Standard for access to radio spectrum (draft)*, EN 303 687 V.0.0.7 (2020-09) Std.

Paper F

Multi-link techniques for the support of URLLC New Radio-Unlicensed in hostile environments

Roberto Maldonado, Claudio Rosa, Klaus I. Pedersen

The paper has been submitted to the
IEEE 93rd Vehicular Technology Conference (VTC Spring), 2021

Abstract

Unlicensed spectrum operation mandates the performance of channel access mechanisms to ensure the absence of transmissions by other unlicensed radio access technologies. Packets are prone to be impacted by these mechanisms as transmissions might be delayed. In this paper, we first analyse the impact of unexpected and sporadic IEEE 802.11ax interference on a New Radio-Unlicensed (NR-U) deployment in an industrial environment. Given the need for improvement in such scenarios, frequency diversity techniques are proposed to fulfil tight latency-reliability requirements and combat the impact of 802.11ax interference. A multi-channel NR-U system in which, at least, one of the channels is free of IEEE 802.11ax interference is studied. The paper shows that multi-channel techniques with a lack of flexibility in the channel selection exhibit modest benefits in scenarios with the presence of sporadic interference. A scheme with flexible channel selection is proposed and verified as the preferred option for stringent latency and reliability applications.

1 Introduction

Wireless communications are considered a key enabler for the feasibility of the next industrial revolution. The fourth industrial revolution, also known as Industry 4.0, envisions a shift in the industrial paradigm by relying on the digitization of the manufacturing process and the supply chain. The Industry 4.0 foundations are based on a powerful connectivity infrastructure capable of supporting, among others, tasks requiring very tight latency at extreme availability/reliability levels. For instance, motion control applications demand latencies lower than 2 ms with an availability of 99.999 % or higher [1, 2]. The latest cellular technology standard, the 5G-New Radio (5G-NR), emerges as a candidate to provide a suitable platform to enable Industry 4.0. 5G-NR is designed to support a wide range of service classes: enhanced mobile broadband (eMBB), massive machine-type communications (mMTC) and ultra-reliable and low-latency communications (URLLC) [3]. The latter covers the most demanding use-cases in the context of Industry 4.0, since it aims the support of mission-critical communications with stringent latency-reliability requirements.

The 3rd Generation Partnership Project (3GPP), during the development of the 5G-NR specifications, ensured the applicability of the standard to unlicensed spectrum. New Radio-Unlicensed (NR-U) is the 5G-NR solution for unlicensed spectrum [4]. Unlicensed frequency bands are globally available and deployments are simpler and with reduced cost as compared to licensed spectrum alternatives. Therefore, unlicensed spectrum is considered as an attractive asset for industrial verticals. Especially interesting is the standalone mode, in which transmissions are solely carried over the unlicensed bands without the need of an anchor carrier in the licensed spectrum. On the other hand, due to its open-access nature, unlicensed spectrum operation also presents some challenges. In order to ensure a fair spectrum usage, nodes sharing the unlicensed bands must follow strict regulations. In that sense, spectrum regulators provide technology-agnostic guidelines to govern the frequency usage. For example, they restrict the maximum allowed transmit power and power spectral density. Certain regulators also provide a common framework for the channel access. The channel access guidelines avoid simultaneous transmissions from multiple nodes and limit the duration a node can continuously occupy the channel. Adopting these guidelines implies that the channel availability is reduced as compared to licensed spectrum operation. Therefore the support of URLLC in unlicensed spectrum (URLLC-u) becomes more challenging. Several contributions in the open literature address this topic under single radio access technology (RAT) conditions. Increasing the channel access probability using time diversity techniques is analyzed in [5]. Authors in [6] propose a multi-channel access scheme that, with support on licensed bands, helps in reducing the periods without access to the unlicensed channels. Coordination is exploited in [7] where an NR-U deployment, synchronized in the channel access and the frame selection domains, is shown to reduce the impact of the channel access on the latency.

In this paper, our focus is on achieving URLLC-u requirements in uncontrolled environments. Uncontrolled environments, also referred as hostile environments, define scenarios containing devices from a single-RAT in which unexpected interference from any other RAT happens sporadically. The goal is to evaluate the latency impact of sporadic interference, generated from a wireless local area network (WLAN) following the IEEE 802.11ax standard [8], on an indoor factory NR-U deployment. Frequency diversity techniques are explored as candidates to mitigate the performance degradation due to the presence of unexpected and sporadic interference. For this purpose, a multi-channel NR-U in which at least one of the available channels is free of WLAN interference is assumed. The performance of each mechanism is evaluated, from a latency-reliability perspective, by extensive system-level simulations. The remainder of the paper is structured as follows: Section 2 presents the system model. Section 3 includes a description of 5 GHz channel access mechanisms with focus on multi-band channel access, while Section 4 describes the different mechanisms foreseen as candidates for combating the sporadic WLAN interference. Simulation assumptions and simulation results are found in Sections 5 and 6, respectively. Final remarks are drawn in Section 7.

2 System model

The system model is defined as an indoor factory plant with NR-U and IEEE 802.11ax networks deployed in the 5 GHz band. The NR-U deployment is composed of M next-generation nodes-B (gNBs) and K user equipments (UEs). Nodes are capable of simultaneously transmitting over n adjacent carriers, each with 20 MHz channel bandwidth. The network is assumed to support URLLC-like traffic. The packet generation follows a Poisson distribution with an average arrival rate of λ_{NRU}. Downlink-only traffic with a packet size of B_{NRU} bytes is assumed. In the uplink, control information such as hybrid automatic request repeat (HARQ) messages are sent by UEs in response to previously received downlink packets. A 7-OFDM symbols transmission time interval (TTI) in a dynamic time domain duplexing (dynamic-TDD) system is assumed. Within the factory premises, a WLAN is deployed. The WLAN consists of a single access point (AP) and a single station (STA) deployed over a bandwidth of 20 MHz which overlaps with one of the NR-U channels. This ensures that at least one NR-U channel is free of WLAN interference. As in the NR-U deployment, WLAN sporadically generates downlink-only packets according to a Poisson distribution with an average arrival rate of λ_{11ax}. The occassional WLAN packet generation is ensured by reducing the packet arrival rate with respect to NR-U, i.e. $\lambda_{11ax} \ll \lambda_{NRU}$. eMBB traffic with a payload size of B_{ax} bytes is assumed, where $B_{ax} = 1000 \cdot B_{NRU}$. Nodes from both RATs operate in compliance with the 5 GHz spectrum regulations.

3 Channel access mechanisms

Any 3GPP-based technology operating in the unlicensed spectrum acts according to the European Telecommunications Standards Institute (ETSI) guidelines [9]. For the 5 GHz band, ETSI mandates each node to perform a channel access assessment prior to any transmission. The channel access procedures, performed over a fixed bandwidth of 20 MHz, are based on the listen-before-talk (LBT) mechanism. LBT consists of an interference measurement and posterior comparison against a predefined energy detection (ED) threshold. If the received interference level is higher than the threshold, the node declares the channel as busy. LBT is performed in periods of 9 µs denoted as clear channel assessment (CCA) slots. Nodes acquiring the channel for transmission must 1) sense the channel for a defer duration and 2) complete a back-off procedure where the channel sensing is performed during, at least, a random number of CCA slots obtained from a uniform distribution. This is denoted as Type 1 LBT in NR-U [10] and distributed coordination function (DCF) in IEEE 802.11 [8] and it is usually performed by gNBs and APs,

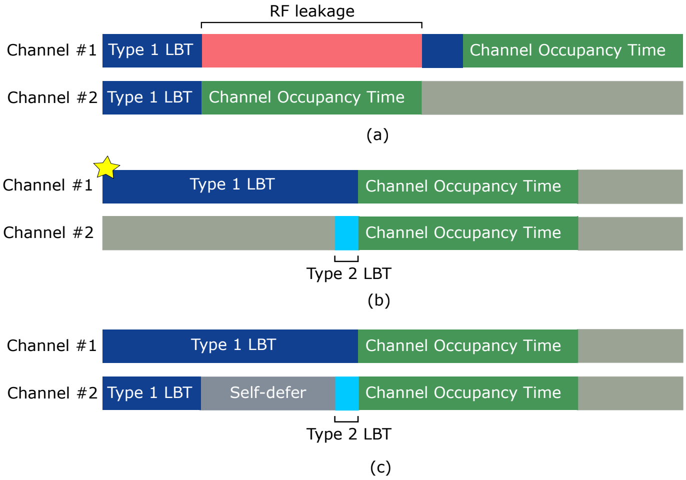


Fig. IV.1: Multi-band LBT options: (a) Type A LBT, (b) Type B LBT with channel #1 as primary and (c) Type A LBT with self-deferal.

respectively. Once the procedure is successful, the node can occupy the channel for a maximum duration known as maximum channel occupancy time (MCOT). During this period, responding devices, typically UEs and STAs, can use the previously gNB/AP-acquired COT and access the channel. In NR-U, UEs perform an LBT with a fixed sensing duration of 25 µs, i.e. Type 2A LBT according to 3GPP. Optionally, under conditions where the gap between the last DL transmission and the next UL transmission is shorter than 16 µs, the performance of this LBT is avoided, i.e. Type 2C LBT according to 3GPP. During the COT, gNBs can also re-access the channel by a successful Type 2A LBT. In 802.11 deployments, STA transmissions avoid the LBT performance as uplink transmissions start within the 16 µs gap.

3.1 Multi-band Listen Before Talk

In case of simultaneous transmissions over multiple channels, nodes must perform independent channel access procedures over bands of 20 MHz. Successful LBT in each channel is required before any multi-band transmission [9]. 3GPP, following the ETSI guidelines, defines different channel access procedures for transmission over multiple channels [10]. As depicted in Fig. IV.1, nodes shall assess the channel availability by employing Type A or Type B channel access. In Type A, independent Type 1 LBT procedures are performed on each channel. In Type B, gNBs access the channel with Type 1 LBT over a selected channel, denoted as the primary channel, and, before starting the transmission, they additionally perform Type 2A LBT over the rest of the available channels. The primary channel is selected either randomly or arbitrarily. In the latter case, the primary channel selection must not be changed more than once per second. The reduced flexibility on the primary channel selection impacts negatively the system performance in presence of sporadic interference on the channel. In such a case, transmissions on the rest of the channels are postponed until a successful Type 1 LBT on the primary channel is obtained. If the primary channel is selected randomly, although a new primary channel can be selected for each multi-band transmission, similar drawback is observed. Therefore, for our URLLC analysis, Type A channel access is assumed. It is also assumed that, due to RF practical impairments, a gNB cannot perform LBT in one of the channels while transmitting on another. When adopting Type A LBT, a solution to perform multi-band transmisisons is to use self-deferral mode [11] in the channels with a successful Type 1 LBT as depicted in Fig. IV.1(c). In such an implementation, gNBs defer their transmission waiting for the rest of the channels to finish their respective Type 1 LBT and, hence, align their transmissions. Nonetheless, waiting for all the active LBT procedures to finish might become unfavourable especially when stringent latency-reliability requirements are targeted. The time spent by a channel waiting for the adjacent channel to declare the medium as idle and start the multi-band transmission is denoted as channel alignment delay. Limiting the self-deferral waiting time, also known as LBT synchronization boundary (LSB), has been studied in [12]. Authors propose an algorithm for LSB adaptation which maximizes the number of active channels, and hence, the network throughput. In our study, the target is to reduce the packet delay and different LSB values are evaluated and results are shown in Section 6.

4 Multi-link techniques

Multi-link techniques are typically designed to fulfil one of the following purposes: a) boost the system throughput by increasing the available transmission bandwidth or b) improve the reliability of a packet decoding at the receiver side by exploiting diversity. Carrier aggregation and wideband operation are classified into the first group, whereas, packet data convergence protocol (PDCP) duplication is in the second group. These features were designed for licensed spectrum deployments and our proposal is to apply them, from a different perspective, to unlicensed spectrum operation. Particularly, we aim to use the multi-link techniques to increase the LBT success probability which, in the presence of unexpected and sporadic WLAN interference, can facilitate the maintenance of a certain level of latency and reliability. A description of each technique is summarized here:

- **Carrier aggregation (CA):** a device with CA capabilities can simultaneously connect to a primary component carrier (PCC) and one or more

Table IV.1: Multi-link techniques comparison in presence of sporadic inter-rat interference

Multi-link technique	Capacity	Latency and reliability
Carrier aggregation	+	−
Wideband operation	+	+
PDCP duplication	−	+

secondary component carriers (SCC). Each component carrier (CC) corresponds to a 20 MHz channel and CA allows simultaneous transmission/reception across the CCs. The logic resides in the MAC layer, which is responsible for scheduling data across the CCs. According to current specifications, once data is scheduled for transmission in a CC, redirection to another carrier is not possible.

- **Wideband operation (WB):** a device with WB capabilities is able to connect to a cell whose bandwidth is larger than 20 MHz and therefore spans over multiple channels. As compared to CA, WB is fully flexible in the scheduling as there is no tight association between transmissions and selected channel. Hence, a packet scheduled in one channel can be re-scheduled in another channel. As a drawback, and according to existing specifications, WB only works over contiguous bands.
- **PDCP duplication:** it consists of systematically transmitting replicas of the same data across multiple wireless links to increase the packet decoding success probability [13].The performance of PDCP duplication is sensitive to the duplication criteria and the network load. Blind duplication in highly loaded scenarios might turn into a performance degradation due to lack of resources and low spectral efficiency [14].

A summary of the capabilities of each of the aforementioned techniques in hostile environments according to supported capacity and latency-reliability is shown in Table IV.1. For stringent latency and reliability requirements, CA is expected to show limited performance due to the lack of capabilities for channel re-selection. WB offers the possibility of flexible channel re-selection and therefore it exhibits better latency-reliability performance than CA. By adopting PDCP duplication, the available channels are fully utilized, achieving similar latency-reliability performance as WB operation. However, due to its lower spectral efficiency, a penalty in the performance is expected from a capacity perspective.

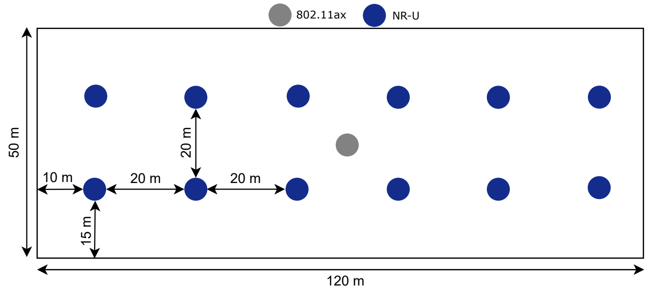


Fig. IV.2: Simulated indoor factory layout [15].

5 Simulation assumptions and methodology

As shown in Fig. IV.2, the NR-U network consists of 12 ceiling-mounted gNBs with an inter-site distance (ISD) of 20 m. 60 UEs are uniformly distributed around the layout. 2 x 20 MHz adjacent channels in the unlicensed band are available for transmissions. UEs select the serving gNB based on the strongest reference signal received power (RSRP) criteria. Load balancing in the scenario is achieved by limiting the number of connected UEs per gNB to 5. The 802.11ax deployment contains a single AP, located in the centre of the layout, and a single STA, uniformly distributed around the scenario. A 20 MHz channel, overlapping with one of the NR-U channels, is available for 802.11ax transmissions. The rest of the assumptions of the simulation are found in Table VI.2. The assumptions are used as the input to our highly-detailed system-level simulator. The simulator is capable of modelling, with OFDM-subcarrier resolution, the majority of the PHY and MAC layer procedures involved in a communication and it is developed in line with the 3GPP and IEEE guidelines. Moreover, it also includes detailed modelling of the unlicensed-specific mechanisms. Interference created by NR-U and 802.11ax transmissions are jointly used in the decoding process and LBT procedures of both technologies. To mimic a realistic industrial deployment, the recently agreed indoor factory channel model is assumed [15]. The methodology consists of the execution of Monte-Carlo simulations. Multiple realizations are simulated and combined afterwards. Sufficient samples are collected in order to extract conclusions on high percentiles, e.g. 99.99 percentile, with enough confidence. For better readability of the statistics at very high percentiles, complementary cumulative distribution functions (CCDF) are used.

Table IV.2: Simulation assumptions

Parameter	**NR-U**	**IEEE 802.11ax**
Layout	Indoor office [15]	
Channel model	3D geometry-based stochastic	
Propagation model	Indoor factory (InF) [15]	
gNB (AP) to UE (STA)	InF-SH	
gNB (AP) to gNB (AP)	InF-HH	
UE (STA) to UE (STA)	InF-SL	
Carrier frequency	5.19 GHz and 5.21 GHz	5.19 GHz
Available bandwidth	2 x 20 MHz	20 MHz
Sub-carrier spacing	30 kHz	312.5 kHz
TTI	7 OFDM symbols	-
Duplexing mode	dynamic TDD	TDD
Transmit power		
gNB/ AP	SC: 23 dBm	SC: 23 dBm
	CA/ WB: 20 dBm	-
UE/ STA	18 dBm	18 dBm
MIMO	2×2 configuration	
Receiver type	LMMSE-IRC [16]	
Traffic model	FTP Model 3	
Packet size	50 B	500 kB
UEs processing times*	4.5 OFDM symbols [17]	-
Max. HARQ reTx	2	-
Link adaptation	Outer link loop adaptation [18]	
BLER target	1 %	5 %
Channel access		
gNB/ AP	Type 1 LBT	
Priority class [10]	2	3
UE/ STA	Type 2A/ 2C LBT	Type 2C LBT
ED threshold	-72 dBm	-62 dBm

* gNBs processing times are assumed 2 times faster

6 Performance Results

6.1 Single carrier

First, the performance analysis of a single carrier (SC) deployment is conducted. In this scenario, the NR-U and 802.11ax frequency bands are fully overlapped. The goal of the analysis is to provide a reference for the latency-reliability performance of NR-U in a hostile environment with no interference-free channel. The latency-reliability comparison of NR-U SC under controlled and uncontrolled conditions is shown in Fig. IV.3. It is noted a clear degradation of the NR-U performance is observed when 802.11ax coexists over the same 20 MHz band. Specifically, an increase in the experienced latency by a factor of $\sim$3.2 is experienced at 99.99 % reliability. This performance degradation is due to the presence of uncontrollable inter-RAT interference, which causes a significant increase in the channel access delay. Moreover, the fact that 802.11ax uses a more aggressive ED threshold as compared to NR-U, allows 802.11ax to acquire the channel faster than NR-U as it is less sensitive to interference. Additionally, 802.11ax channel acquisition is not bounded by the slot-based PHY design of NR. In NR-U, transmissions can only start on specific time instances whereas 802.11ax can use the channel immediately after successful channel access. Fig. IV.3 motivates the need for mechanisms to overcome the delay degradation.

6.2 Multi-channel: single-carrier per UE

The first multi-channel approach consists of evenly distribute the UEs among the 2 available 20 MHz channels. Consequently, the network load is split among the channels. UEs do not support multi-carrier capabilities and therefore are assigned and served from a single carrier. On the other hand, gNBs support multi-link transmissions over 40 MHz. This type of transmissions occurs in cases when, at least, 2 UEs from different carriers are simultaneously scheduled. The combined latency performance between the available channels under different LSB values is shown in Fig. IV.4. The performance of NR-U SC with and without the 802.11ax presence are also included as reference. Observing the figure, the main conclusion is that the experienced packet delay is increased by reducing the LSB. The highest latency is achieved with LSB set to 0. In fact, at 99.99 % reliability, a latency reduction of $\sim$25 % is achieved with the most relaxed LSB value as compared to NR-U SC in hostile conditions. On the other hand, choosing an aggressive LSB experiences a penalty of $\sim$43 % in the packet delay. To explain this behaviour, the observed packet delay per channel is depicted in Fig. IV.5. For a given LSB, differences in the performance among the channels are observed due to the asymmetry in the interference conditions. Starting from the perfor-

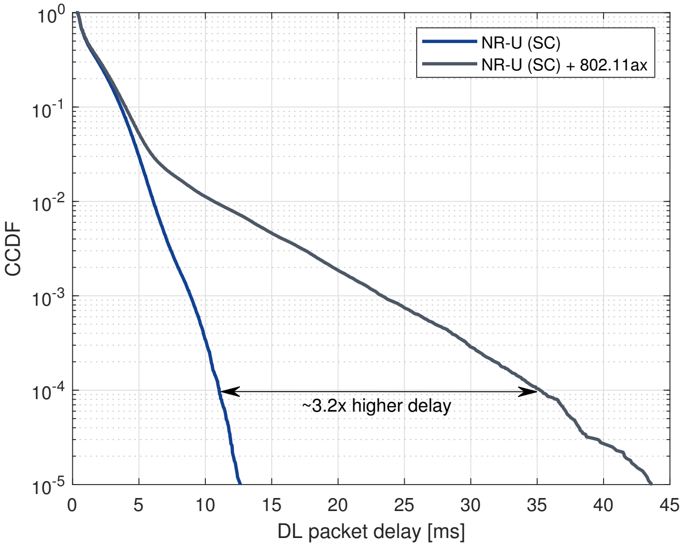


Fig. IV.3: Single carrier. Observed packet delay, for an offered load of 5 Mbit/s, of an NR-U deployment with and without the presence of IEEE 802.11ax.

mance in channel 1 (CH-1), it is noted that a performance gain (∼34 % in latency) is obtained when an aggressive channel access is adopted. CH-1 is the free-of-11ax-interference band and therefore it is likely that transmissions are postponed waiting for CH-0 to successfully finish LBT. Thus, choosing a lower LSB provides latency reduction to the CH-1 as packets are transmitted faster. When LSB expires, the gNB transmits in single-band mode. In such a case, while CH-1 is active, CH-0 transmissions are blocked since RF limitations prevents the gNB to finish LBT in the inactive channel. Consequently, CH-0 packets delays are significantly degraded. Hence, for single-carrier UE scenarios, a high LSB configuration is seen as beneficial. The experienced gain of this scheme as compared to NR-U SC in hostile environments (see Fig. IV.4) is mainly due to the packets scheduled in CH-1. However, the effect of the sporadic inter-RAT interference is still present for those UEs served in CH-0. In fact, the experienced delay in the 99.99 percentile is ∼2.36 times higher than in a NR-U SC controlled deployment.

6.3 Multi-channel: carrier aggregation

In this case, it is assumed that UEs have CA capabilities. Thus, each UE can be scheduled in any of the available 20 MHz channels (PCC and SCC). Given the freedom on the channel selection, the gNBs are able to adapt the scheduling decision based on a specific metric. In our study, the adopted criteria are

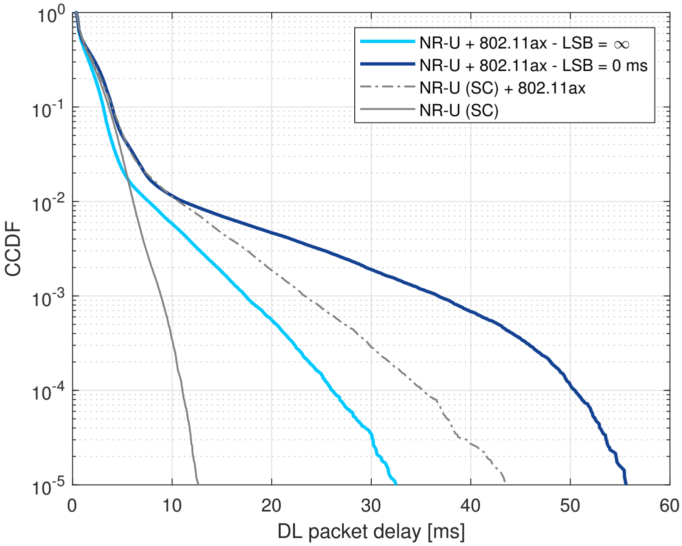


Fig. IV.4: Multi-channel: single-carrier per UE. Observed packet delay for 5 Mbit/s offered load and different LSB values.

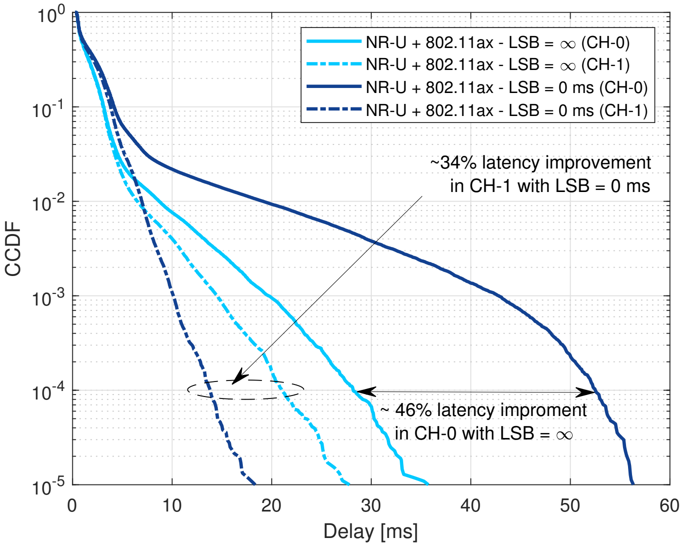


Fig. IV.5: Multi-channel: single-carrier per UE. Observed packet delay per channel (CH) at 5 Mbit/s offered load and different LSB values.

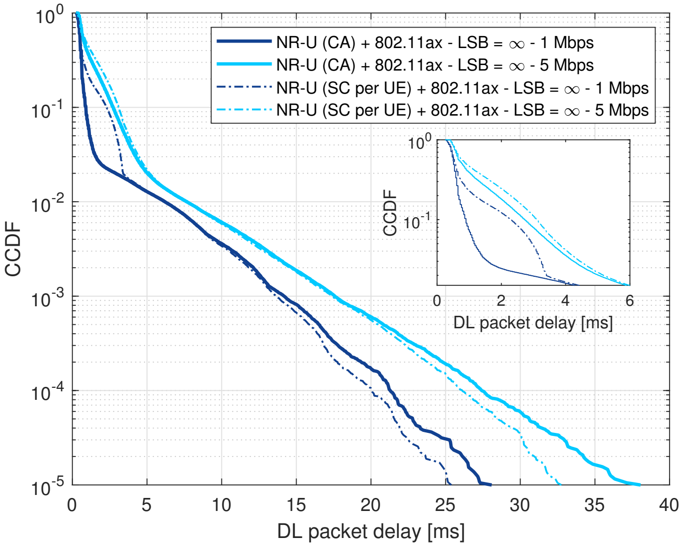


Fig. IV.6: Multi-channel: carrier aggregation vs single-carrier per UE. Observed packet delay comparison at 1 Mbit/s and 5 Mbit/s offered load.

based on the channel occupancy conditions. Therefore, if a gNB has an on-going COT on one of the channels, any new packet is scheduled on that CC. If both CCs have an active COT or both CCs are idle, new packets are always scheduled in PCC. Within a serving gNB, UEs are split in the selected PCC to ensure load balancing. The CA latency-reliability performance as well as the SC per UE performance is shown in Fig. IV.6. Adopting CA under the current standard specifications limits the possibility of packet re-scheduling to different channels. Therefore, CA shows similar behaviour as SC per UE with respect to the LSB values and only the most relaxed LSB configuration is shown here. As compared with SC per UE, a modest packet delay reduction is achieved in cases where 802.11ax is not active (see zoomed part of the CCDF in Fig. IV.6). This behaviour is motivated by the avoidance of Type 1 LBT in cases there is an on-going COT in only one of the CCs. The benefit is more noticeable in the 1 Mbit/s curve since the probability of having a single active COT is higher. However, in cases where 802.11ax is present in the channel, i.e. in the tail of the CCDF, the latency performance is slightly degraded as compared to the SC per UE case (∼5 % latency penalty in both offered loads). This corresponds to 2x 20 MHz transmissions that are delayed by i) 802.11ax interference and ii) alternate NR-U single-band transmissions. The second source of delay is the difference between CA and SC per UE. The possibility of freely scheduling in any CC increases the number of single-

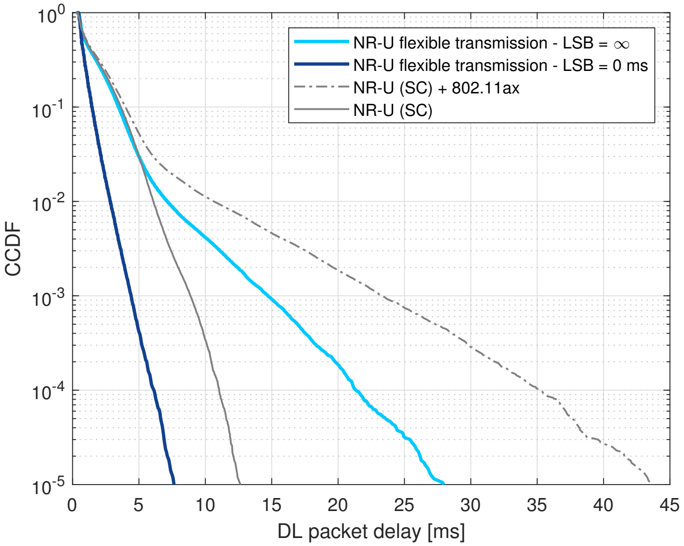


Fig. IV.7: Multi-channel: flexible retransmissions. Observed packet delay for 5 Mbit/s offered load and different LSB values.

band transmissions which, in turn, impact the channel access delay for multi-band transmissions (in which a successful LBT in both channels is required). Thus, from a URLLC perspective, adopting CA does not bring any latency-reliability advantage.

6.4 Multi-channel: flexible transmissions

Based on the previous observations, it is concluded that the lack of flexibility in the selected channel upon a scheduling decision is limiting the performance of SC per UE and CA. A packet scheduled on a channel is tight to that decision and re-scheduling it to a different set of frequencies (potentially free of 802.11ax interference) is not possible. To improve this, we analyze here a multi-channel technique which allows for fast packet re-scheduling between the channels. With the proposed scheme, a packet is duplicated and simultaneously scheduled for transmission on multiple channels. Not transmitted replicas are cancelled when the packet is successfully received at the UE side. While being practically unfeasible, this represents an upper bound of the performance that can be achieved with flexible (re)transmission schemes such as WB operation, PDCP duplication with ideal cancelation and enhanced CA (with cross-channel transmissions capabilities). The latency-reliability performance of the proposal is shown in Fig. IV.7. It is noted that having full flexibility in the scheduling decision provides better latency performance with aggressive LSB configurations. In other words, due to flexible (re)transmission, it is recommended to prioritise faster channel access rather than multi-band transmissions. Setting the LSB to a very high value corresponds to blind packet duplication and it shows benefits as compared to the baseline SC in uncontrolled environment conditions. This is due to i) packet replicas increase the decoding success probability at the UE side and ii) gNBs benefit from channel access diversity during the COT. A successful Type 2A LBT in one of the channel is enough to transmit the packet. In SC, transmissions during the COT are subjected to the LBT outcome of a single channel. However, packets transmitted after Type 1 LBT are still impacted by the inter-RAT interference in one of the channels. To effectively mitigate the impact of 802.11ax interference, an aggressive LSB configuration is recommended. In fact, the achievable latency is improved by $\sim$46 % in the 99.99 percentile as compared to the experienced in controlled environment deployments. The reported gain is caused by channel access diversity. Nodes in the multi-band scenario attempt Type 1 LBT in both channels and only use the fastest for transmissions. Indirectly, this also achieves a split of the carried traffic among the channels that helps to reduce the channel access delay.

7 Conclusions

The paper presents a comprehensive analysis of different multi-link schemes to mitigate the effect of uncontrolled interference on the latency-reliability performance for NR-U deployments. It is shown that techniques with flexible channel (re)selection are the preferred solution to combat the inter-RAT interference. Diversity in the channel access brings the possibility of choosing the less interfered channel for transmissions. On the other hand, techniques with static channel selection are found to not cope well with uncontrolled interference. Packets scheduled on a channel with inter-RAT interference are highly delayed due to reduced channel availability. The importance of a proper LSB configuration is also highlighted. It is outlined that low LSB values report better latency-reliability as they increase the probability of performing multi-band transmissions.

References

[1] 3GPP, "Service requirements for cyber-physical control applications in vertical domains," Tech. Rep. 22.104, 2019.

[2] 5G-ACIA, "Key 5G Use Cases and Requirements," 2020.

[3] International Telecommunication Union, “IMT Vision - Framework and Overall Objectives of the Future development of IMT 2020 and beyond,” ITU-R, Tech. Rep.

[4] 3GPP, “Study on NR-based Access to Unlicensed Spectrum (Rel-15),” Tech. Rep. 38.889, 2018.

[5] R. Maldonado, C. Rosa, and K. I. Pedersen, “Latency and reliability analysis of cellular networks in unlicensed spectrum,” *IEEE Access*, vol. 8, pp. 49 412–49 423, 2020.

[6] G. J. Sutton et al., “Enabling Ultra-Reliable and Low-Latency Communications through Unlicensed Spectrum,” *IEEE Network*, 2018.

[7] R. Maldonado, C. Rosa, and K. I. Pedersen, “A Fully Coordinated New Radio-Unlicensed System for Ultra-Reliable Low-Latency Applications,” in *2020 IEEE WCNC*, 2020.

[8] “IEEE Draft Standard for Information Technology – Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment Enhancements for High Efficiency WLAN,” *IEEE P802.11ax/D6.0, November 2019*, pp. 1–780, 2019.

[9] ETSI, *5 GHz RLAN; Harmonised Standard covering the essential requirements of article 3.2 of Directive 2014/53/EU*, ETSI EN 301 893 V2.1.1 (2017-05) Std.

[10] 3GPP, “Physical layer procedures for shared spectrum channel access (Rel-16),” Tech. Rep. 37.213, 2020.

[11] ——, “Multi-carrier LBT operation for LAA,” Tech. Rep. RP-152784, 2015.

[12] L. H. Vu and J. Yun, “Adaptive Self-Deferral for Carrier Aggregation of LTE-LAA with RF Power Leakage in Unlicensed Spectrum,” *IEEE Access*, vol. 7, pp. 89 292–89 305, 2019.

[13] 3GPP, “Packet Data Convergence Protocol (PDCP) specification, (Rel-16),” Tech. Rep. 38.323, 2020.

[14] M. Centenaro, D. Laselva, J. Steiner, K. Pedersen, and P. Mogensen, “System-Level Study of Data Duplication Enhancements for 5G Downlink URLLC,” *IEEE Access*, vol. 8, pp. 565–578, 2020.

[15] 3GPP, “Study on channel model for frequencies from 0.5 to 100 GHz,” Tech. Rep. 38.901, 2018.

[16] K. Pietikainen et al., “System-Level Performance of Interference Suppression Receivers in LTE System,” in *2012 IEEE VTC*.

[17] 3GPP, “Physical Layer Procedures for Data (Rel. 15),” Tech. Rep. 38.214, 2019.

[18] K. I. Pedersen et al., “Frequency Domain Scheduling for OFDMA with Limited and Noisy Channel Feedback,” in *2007 IEEE VTC*.

Part V

Conclusions

Conclusions

The PhD dissertation proposes and analyzes radio solutions for enabling the support of ultra-reliable and low-latency communications (URLLC) in the unlicensed spectrum, specifically on the 5 GHz band. The focus is on 3$^{\text{rd}}$ Generation Partnership Project (3GPP)-based technologies. The study conducts an assessment of the effect of the unlicensed spectrum regulations on the latency and reliability. The mandatory channel access mechanisms are identified as the main bottleneck that hinders the achievement of URLLC in unlicensed spectrum, even in single-operator deployments with no inter-system interference. Therefore, minimizing the impact of channel access on the latency and reliability is the goal of the research. Multiple radio resource management solutions and channel access optimizations are derived and analyzed to enhance the latency performance. The main contribution of the thesis is that, under specific conditions and design optimizations, the negative impact of mandatory co-existence mechanisms such as listen-before-talk (LBT) can be fully, or at least largely, mitigated. This allows New Radio Unlicensed (NR-U) to achieve promising latency-reliability performance, comparable to the obtained in a 5G New Radio (NR) licensed deployment with time division duplex (TDD). The work motivates the usage of the unlicensed spectrum as a real alternative to licensed-spectrum based solutions for fulfilling URLLC requirements in indoor deployments.

1 Summary of the main findings

A summary of the conducted research activities is found below. As part of the summary, the initial research questions of the PhD are re-formulated and answered.

[Q1] What is the impact of spectrum regulations on the unlicensed cellular latency and reliability performance?

The foundations of our research are built by analyzing the state-of-the-art performance of MulteFire, the only available technology capable of operating exclusively using unlicensed bands by the start of the PhD. Probability-based analysis and system-level simulations indicate that a significant part of the downlink (DL) delay experienced over the radio interface (between 18 % to 42 %) is due to the LBT procedure performed at the base stations (BSs). The thesis also highlights the importance of a high success rate in the uplink (UL) channel access. A failure in the user equipments (UEs) channel access triggers i) potentially unnecessary DL retransmissions if a control transmission is blocked or ii) the need of a new grant after not successfully use the dedicated resources after a successful scheduling request (SR). Based on these results, several technology components were applied for unlicensed operation and studied from a latency and reliability perspective.

[Q2] Which technology components, initially designed for URLLC in licensed spectrum, are suitable for unlicensed operation?

Fundamental design features from NR were integrated into our system model. It represents the transition between MulteFire and NR-U. An enhanced latency and reliability performance is achieved by reducing key components of the overall packet delay. Optimizations of the subcarrier spacing, the UEs and BSs processing times, and the transmission time interval (TTI) durations for URLLC purposes are adopted. It is shown that these enhancements provide additional benefits as compared to licensed spectrum. As an example, the fact that the TTI size is reduced indirectly influence the channel access outcome. Nodes occupy the channel for shorter durations which makes the channel access faster and more reliable. For instance, the outcome of our study is that a $\sim$18 % reduction in the UL LBT failure rate is achieved by simply shorten the TTI from 14 to 7 orthogonal frequency-division multiplexing (OFDM) symbols. Spectrum regulators only allow for continuous usage of the frequency bands for a certain time interval denoted as channel occupancy time (COT). Thus, in order to favour an efficient utilization of the resources, multiple transitions between DL and UL within a COT are introduced. Together with the fast processing times, the multiple transitions allow for faster hybrid automatic repeat request (HARQ) feedback and reduced queuing delay. The combination of these optimizations leads to a latency decrease of $\sim$82 % with respect to the delay experienced in MulteFire.

[Q3] How to mitigate the performance impact of the channel access by means of radio resource management (RRM) techniques?

- **Time diversity techniques for UL transmissions:** In order to mitigate the impact of channel access on UL transmissions, time diversity techniques are proposed. For UL control transmissions, the BSs are aware of potential LBT blockings at the UE side, and therefore, they indicate M additional resources for the transmission of the HARQ feedback in the DL assignment. As compared to the baseline operation, it reduces the number of unnecessary DL retransmissions triggered due to UL LBT failure. This introduces the following advantages: i) it decreases the channel activity and therefore improves the channel access delay of the neighbour nodes and ii) it reduces the experienced queuing delay as new data transmissions are scheduled faster. These benefits are directly reflected in the latency and reliability performance, showing up to 59 % latency reduction at the 99.99 percentile. The avoidance of UL LBT, under given spectrum regulations conditions, is also analyzed. A reduction in the UL LBT failure rate is achieved especially in medium-high load scenarios (e.g. ~14.2 % for 5 Mbit/s offered load). This also brings latency and reliability improvement.

For UL data transmissions, a pre-emptive grant-based scheduling is proposed. The technique consists of signalling the UEs with K occasions for the UL transmissions with a single grant. Thus, the need for several grants due to UL LBT failure is reduced. Our performance analysis concludes that in cases with high LBT failure probability, the proposed scheme can decrease the observed latency up to 23 % with respect to baseline grant-based UL.

- **Grant-free UL:** For UL data transmissions, grant-free (GF) is proposed as an alternative to grant-based (GB). Using GF avoids the time-consuming message exchange involved in the SR procedure. Moreover, when applied to unlicensed spectrum, additional benefits are reported as GF reduces the number of required channel access before the UL data transmissions. Monte-Carlo system-level simulations show that the adoption of GF reports latency benefits in low and medium offered load. However, in high-load cases, the fact that multiple UEs simultaneously use the same time-frequency resources, i.e. collisions, implies a penalty in the latency as compared to GB.

[Q4]: How to efficiently deploy a NR-U network to cope with the reduced channel availability?

Uncoordinated channel access brings large latency penalties as nodes, even from the same network, mutually interfere. In order to provide better channel access latency, coordination schemes are introduced. The first approach is to use a synchronous channel access approach for the initiating devices (typically BS). Nodes follow a common and periodic structure that dictates the time instances for channel assessment and channel occupancy.

This ensures success in the LBT prior to the COT start with bounded delay. However, if no further coordination between nodes is established, our study shows that UEs can still experience failures in the UL LBT. Our proposal is to introduce a coordinated TDD frame selection that guarantees simultaneous UL transmissions in the network. Consequently, the LBT sensing intervals before the UL slots – or mini-slots – are also synchronized for every UE. A central node that dynamically coordinates the frame selection among the BSs is required. This approach is shown to report significant latency reductions (up to ∼70 %) as compared to uncoordinated and BSs-only coordinated channel access approaches. An alternative to this proposal consists of configuring nodes with a silent gap pattern. Nodes following this mechanism periodically stop their transmission for the duration of the UL LBT. The silent gap coordination allows for optimal TDD frame selection in each BS, i.e. no TDD coordination is needed. This is a key aspect, especially in scenarios with very diverse traffic requirements, in which TDD coordination might introduce additional frame alignment delay. Moreover it supports multiple DL/UL transitions withing the COT. Our system-level simulations analysis report that full mitigation of the DL and UL LBT failure rate is achieved with a periodicity equal to the TTI duration. This brings a latency reduction by a factor of ∼3 at the 99.99 percentile as compared to baseline operation. As a drawback, especially in scenarios with a very periodic silencing pattern and short TTI size, a non-negligible reduction of the spectral efficiency is introduced due to frequent silent gaps.

[Q5]: How to achieve URLLC requirements in a scenario with multiple RATs share part of the available spectrum?

Although coordination is an interesting and efficient solution to mitigate the impact of the channel access mechanisms on latency and reliability, their benefits are limited to single-network scenarios. In hostile environments, i.e. scenarios with devices from multiple radio access technologies (RATs) sharing the same unlicensed bands, any coordination approach is affected due to the lack of inter-system cooperation. For instance, the packet delay performance of an NR-U deployment with synchronous channel access is shown to degrade by a factor ∼10 when co-existing with Wi-Fi, as compared to when NR-U is benchmarked in controlled environments, i.e. single-RAT conditions. As a solution to the observed problem, multi-channel techniques are proposed and studied. In hostile conditions, these techniques can provide benefit only under the assumption that unpredictable inter-RAT interference is temporarily observed only on a part of the available spectrum. In other words, at any time, there should always be at least a part of the available spectrum which is free from any unexpected and uncontrolled interference. In such deployments, multi-channel techniques without flexibility in the channel reselection, e.g. carrier aggregation as currently defined by 3GPP, are identified as impractical for ensuring a certain URLLC quality-of-service (QoS). The need for mechanisms that allow for flexible reselection of the channel is emphasized by our system-level simulations analysis. These type of frequency diversity techniques are capable of mitigating the effect of the unexpected and uncontrolled interference, providing a URLLC performance comparable to the observed on a single-channel NR-U deployment in controlled environments. At the same time, these solutions show better spectral efficiency than blind duplication techniques (i.e. better trade-off between capacity and latency-reliability).

2 Recommendations

Based on the acquired knowledge during the PhD, the following recommendations are advisable to be followed:

- A non-negligable latency degradation is expected due to compliance with channel access mechanisms based on LBT. Optimizations in the deployment are needed in order to support strigent latency and reliability requirements.
- The adoption of baseline NR URLLC enablers such as higher subcarrier spacing, fast processing times and mini-slot scheduling are of pivotal importance, even larger than in licensed spectrum, as the channel access delay and success rate are also improved.
- The performance of any type of LBT adds uncertainty to the transmission. Therefore, techniques to reduce the potential number of channel access are advisable. The recommendation is to adopt time-diversity techniques for UL grant-based data and control transmissions. This is especially useful in scenarios with a high UL LBT failure rate. As an alternative for data UL transmissions, GF should be also considered in scenarios supporting a low-to-medium load.
- For deployments with a single NR-U network and no incumbent interference, it is recommended to adopt coordination schemes such as synchronized channel access + TDD coordination or silent-gap coordination.
- In more hostile environments, where the presence of sporadic inter-system interference cannot be avoided, frequency diversity techniques should be adopted. Using multiple channels in the deployments increase the resilience against potential inter-RAT interference. Among the currently available solutions, wideband operation and packet data

convergence protocol (PDCP) duplication are the suggested solutions. They are preferred over carrier aggregation. Currently standardized carrier aggregation lacks the support of channel access reselection.

3 Future Work

The following research topics are considered as potential directions for the future work:

- **Enhancements of RRM techniques:** Optimization of mechanisms such as link adaptation (LA) and power control (PC) are expected to improve the URLLC performance. LA decides the proper modulation and coding scheme (MCS) based on the channel quality indicator (CQI) reported by the UEs. The risk of LBT failure in the CQI transmission hinders the LA functionality, potentially selecting an inaccurate MCS. This results in a performance degradation by either decreasing the reliability, if too optimistic channel quality estimation, or increasing the queuing delay, if too pessimistic channel quality estimation. PC mechanisms and their influence on the LBT outcome is also an interesting reach topic.
- **Channel access schemes optimization for URLLC:** As part of the enhancements for the synchonous channel access, addding the support of lower periodicities is expected to bring additional benefits. Moreover, allowing the UEs to acts as initiating devices can efficiently accommodate stringent latency requirements of UL traffic (in combination with GF). This topic is currently being addressed in 3GPP as part of Release 17. As more creative approach, the design of channel access mecanism not based on LBT but based on, for instance, duty cycles is considered as relevant.
- **Explore frequency-division duplexing mode:** One of the aspects limitings the latency capabilities of NR-U in the 5 GHz is the adoption of TDD. Due to lack of enough frequency separation frequency division duplex (FDD) is not possible. FDD operation opens new research questions and, at the same time, it will position NR-U as a candidate for the support of more demanding URLLC applications.
- **Mobility in NR-U:** In the context of Industry 4.0, logistics within the factories are foreseen as an important application area. This type of applications expects automated guided vehicles (AGV) to move around the industrial facilities, for example, moving materials from a warehouse to a conveyor belt. Seamless, high-reliable and zero interruption

time handovers are required. It is left for future work the evaluation of the support of mobility in NR-U. Potential challenges between channel access and the transmission of BSs reference signals and UEs measurement reports might arise. Advanced mobility support is one of the main advantages of 3GPP-based RAT with respect to IEEE 802.11 technologies (in which mobility is shown to be a bottleneck from a latency perspective [1]).

- **Time-sensitive communications:** In factory automation, the traffic between sensors/actuators and controllers is typically periodic and time-sensitive. These type of communications require, on top of the extreme latency and reliability requirements, demanding constrains for jitter and survival time [2]. It is interesting to identify potential drawbacks and analyze the performance of NR-U with delay-critical and deterministic traffic.

- **Experimiental research on real factories:** Although NR-U is still under development in 3GPP, the first specifications are available in Release 16 [3]. The next step is to have chipset vendors integrating the NR-U capabilities in their devices. Making the technology a reality will allow for on-site performance analysis. This is a very attractive research direction as experimental studies can be a decisive step to show 5G unlicensed operation as a promising alternative to the 5G TDD licensed in the URLLC domain.

- **Network planning optimization:** The channel access latency is highly dependant on the number of initiating devices (typically BSs) contending for the channel. On the other hand, NR is proven to efficiently serve a large number of UEs based on large scheduling capacity, QoS prioritization and advanced receivers. Therefore, an optimal design of the unlicensed deployment with potentially less BSs and higher connected UEs per BS can be considered as direction for future work. As an alternative, the distribution of BSs in orthogonal channels can be also studied.

References

[1] K. Sui, M. Zhou, D. Liu, M. Ma, D. Pei, Y. Zhao, Z. Li, and T. Moscibroda, “Characterizing and improving wifi latency in large-scale operational networks,” in *MobiSys ’16*, 2016.

[2] 3GPP, “Service requirements for cyber-physical control applications in vertical domains (Release 17),” Tech. Rep. 22.104 V17.3.0, Jul. 2020.

[3] ——, “Physical layer procedures for shared spectrum channel access (Release 16),” Tech. Rep. 37.213 V16.2.0, Jun. 2020.

Part VI

Appendix

Paper G

A silent gap coordination for supporting URLLC in unlicensed spectrum deployments

Roberto Maldonado, Claudio Rosa, Klaus I. Pedersen

Abstract

The latency and reliability performance of New Radio-Unlicensed (NR-U) is strongly impacted by the outcome of mandatory clear channel assessment. Due to listen-before-talk (LBT), transmissions can be delayed for large periods, preventing the fulfilment of the ultra-reliable low-latency communications (URLLC) requirements. Especially critical for the latency is the fixed duration listen-before-talk (LBT), also known as Type 2 LBT. For grant-based uplink data transmissions, not succeeding on the Type 2 LBT implies the impossibility of using previously granted resources. In case of uplink control transmissions, it might trigger of potentially unnecessary downlink retransmissions. At the gNB side, downlink transmissions can also be delayed due to failures in the Type 2 LBT. In this paper, we propose a coordination scheme to ensure 100 % *success rate in the performance of Type 2 LBT. NR-U nodes are configured to follow a silencing pattern defined by certain periodicity. The proposal is shown to effectively reduce the Type 2 LBT channel access blocking probability. Moreover, it improves the experience packet delay. For instance, for an offered load of* 5 Mbit/s*, a reduction in the combined downlink-uplink latency by a factor of 3.1 at the* 99.99 percentile *is achieved as compared to baseline NR-U.*

1 Introduction

Ultra-reliable and low-latency communications (URLLC) are considered the most challenging service class to be supported by the 5th generation of cellular technologies (5G). Previous cellular-based generations were designed to optimize the network capacity for the support of ever-increasing broadband data. Thus, the network design was driven by the optimization of the throughput, i.e. the number of packets correctly transferred per time unit. The support of URLLC represents a drastic change in this paradigm. The key parameter indicators (KPIs) determining the design of URLLC are latency and reliability. According to [1], latency is defined as the interval of time that takes to successfully deliver a packet from layer 2/3 at both ends of the communication. Reliability is defined as the success probability of transmitting a packet within a certain time constrain boundary. Latency and reliability are two conflicting metrics, i.e. increasing the reliability often degrades the experienced latency and vice-versa. Therefore, the simultaneous optimization of both KPIs represents a technical challenge. The precise latency-reliability requirements depend on the specific URLLC application. Typical applications require a packet to be correctly delivered over the radio interface (one-way), within a delay between 1 ms and a few tens of ms, and with a reliability of 99.99 % or higher. Designing a 5G system capable of supporting URLLC enables novel use cases in diverse areas, including manufacturing, transportation, healthcare and energy sectors.

Supporting these delay-critical applications exclusively via the unlicensed spectrum is attractive to enterprises. Unlike licensed spectrum, unlicensed bands are defined as free-of-access frequency bands. Therefore, using unlicensed spectrum for industrial applications can provide large cost savings for enterprises. However, unlicensed operation increases the complexity of the achievement of URLLC. In order to ensure fair coexistence between the multiple devices and radio access technologies (RATs) using the same unlicensed bands, spectrum regulators define strict guidelines that every node must follow. Especially critical for the support of the URLLC are the channel access mechanisms. These mechanisms ensure that nodes transmit in periods without any channel activity and limit the duration nodes can continuously transmit on a channel. In other words, transmissions are prone to be delayed since they are always subject to the outcome of channel activity assessment. The 3rd Generation Partnetship Project (3GPP) solution for 5G New Radio (NR)-based access to unlicensed spectrum is known as New Radio-Unlicensed (NR-U) [3]. It is based on Release 15 NR specifications, but with modifications to the PHY/MAC design in order to fulfill the most stringent spectrum regulations worldwide. 3GPP, following the European Telecommunications Standards Institute (ETSI) spectrum regulations [4], designs the NR-U channel access mehcansims based on listen-before-talk (LBT). LBT is a technique that evaluates the channel activity by comparing the measured interference with a predefined energy detection (ED) threshold. A successful LBT occur when the measured interference is lower than the ED threshold. In NR-U, LBT is performed at intervals of 9 µs, also known as clear channel assessment (CCA) slots, and over a bandwidth of 20 MHz. In order to start a transmission in the unlicensed bands as an initiating device, a node must perform a Type 1 LBT. Type 1 LBT is defined as a continuous sensing assessment of the channel activity for, at least, N CCA slots (being N a random number obtained from a uniform distribution). Once the channel is declared as idle, the node can access the channel for a maximum time known as maximum channel occupancy time (MCOT). During that time, the node acquiring the channel (typically the gNB) can grant access to the one or more responding devices (typically the user equipments (UEs)) and make use of channel occupancy time (COT) sharing. In COT sharing conditions, UEs are allowed to access the channel during the gNB-acquired COT, by performing Type 2 LBT. Moreover, within a gNB acquired COT limits, the gNB can re-access the channel after a pause in the DL burst or after a UL burst using Type 2 LBT as well. Type 2 LBT is defined as a CCA check performed over a specific interval of time. Different sensing interval durations are supported by NR-U. Type 2A LBT is adopted when the sensing interval is 25 µs. A reduction in the sensing interval to 16 µs is specified as Type 2B. Additionally, Type 2C defines cases in which nodes are allowed to transmit without assessing the channel activity. Type 2C can only be used by a responding device, following a transmission from the initiating device, if the gap between the transmissions is less than 16 µs. In this case, the duration of the responding device transmission must be constrained to 584 µs.

The constrain of successfully executing LBT before any transmission adds uncertainty and potentially increases the experienced latency. A method to increase the channel access probability has been previously addressed in [5]. Authors propose to use synchronized channel access in combination with TDD frame coordination to improve the Type 2 LBT success rate. In this paper, we describe an alternative that allows for more flexible adaptation to the traffic variations. Our proposal is a silent gap coordination technique that can be applied to synchronous and asynchronous channel access. The solution aims to improve the downlink and uplink latency-reliability performance by avoiding any unexpected delay in the transmissions as a result of a Type 2 LBT failure. The remaining of the paper is structured as follows. Section 2 describes the NR-U system model and highlights the addressed problem. In Section 3 the definition of the proposal is included. The NR-U latency and reliability performance analysis, under the assumptions defined in Section 4, is detailed in Section 5. The main conclusions of the work are summarized in Section 7.

2 System model and problem description

An indoor factory NR-U deployment in the 5 GHz band is assumed. Controlled environment conditions are assumed, therefore, nodes from a single NR-U network are contending for the channel. The NR-U network contains M next generation nodes-B (gNBs) and K user equipments (UEs). The available bandwidth is 20 MHz and dynamic time-division duplexing (TDD) is assumed. A transmission time interval (TTI) duration of 7 OFDM symbols is adopted with 1 OFDM symbol dedicated to control information. A subcarrier spacing (SCS) of 30 kHz is assumed. URLLC-alike traffic is generated in downlink (DL) and uplink (UL) directions with a ratio of 80-20, respectively. The packet generation follows a Poisson arrival distribution with an average of λ packets/s/node. The payload size is assumed to be fixed to B bytes. Uplink data transmissions follow a grant-based approach, and therefore, UEs only transmit on dedicated resources after a successful scheduling request (SR) procedure with their serving gNB. Hybrid automatic request and repeat (HARQ) is adopted as the error correction mechanism to increase the packet decoding probability. With respect to the channel access, asynchronous channel access is assumed. Thus, Type 1 LBT to acquire the channel can be performed at any time. It is assumed that the gNBs act as initiating devices whereas the UEs act as responding devices, i.e. they cannot initiate a channel occupancy time. Multiple transitions between DL and UL TTIs

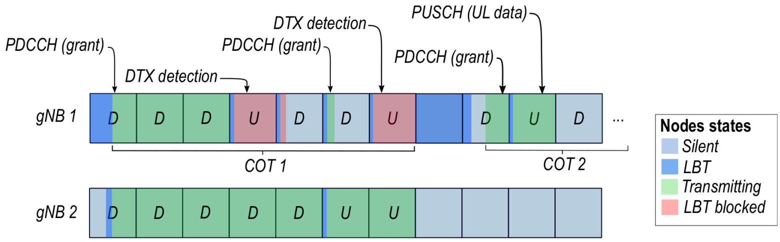


Fig. VI.1: Description of uplink NR-U operation in unlicensed spectrum. It highlights the importance of a high sucess rate in the Type 2 LBT performance.

within the TDD frame are allowed. Each gNB is in charge of selecting the optimal TDD frame configuration based on the current instantaneous traffic conditions. This paper addresses the importance of Type 2 LBT in the latency and reliability performance of NR-U. A declaration of channel busy during a Type 2 LBT has several implications for both uplink and downlink. They are described below based on the NR-U mode of operation depicted in Fig. VI.1.

Uplink The problem is first described from the uplink perspective. It is noted in Fig. VI.1 that, after a successful Type 1 LBT, 2 gNBs acquire the channel at the same time. Due to NR-U transmissions being bounded to start only at the beginning of certain OFDM symbols, e.g. symbols 0 and 7 of each subframe, this can frequently occur. Even in cases where the gNBs are capable start the transmission at any OFDM symbol. With channel access priority class (CAPC) 2 [3], N can take values between 0 and 7. Given that each OFDM symbol for 30 kHz SCS contains 4 CCA slots, nodes drawing $N \leq 4$ will finish Type 1 LBT at the same time (assuming idle channel conditions). This always happens if synchronous channel access is adopted. It can be noted that the gNBs independently select the optimal TDD configuration according to their specific traffic requirements. After the reception of the uplink grant, contained in the control downlink channel (PDCCH), the UE connected to gNB-1 prepares the data for transmission (PUSCH). Before the actual transmission, the UE must ensure that the channel activity is idle via the performance of Type 2 LBT. However, during the sensing interval, the UE is blocked by the gNB-2 DL transmission. Under these circumstances, the time-frequency resources reserved for the UE by gNB-1 are not used. The serving gNB declares discontinuous transmission (DTX) since no uplink transmissions are detected on the given resources. In order to provide a new transmission attempt, the gNB-1 needs to send a new grant. This implies the performance of an additional Type 2 LBT at the gNB side. As shown in the figure, depending on the interference conditions at the gNB side, this process might span over multiple COTs. In that case, the gNB-1 is required to perform an additional Type 1 LBT to acquire a new COT.

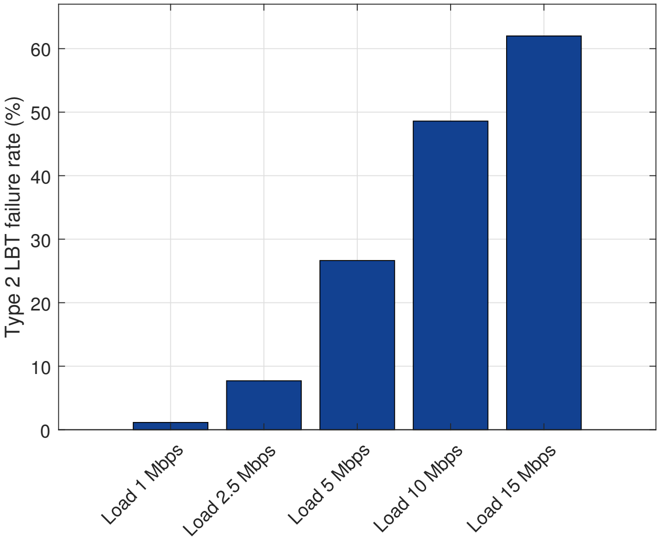


Fig. VI.2: Uplink Type 2 LBT failure probabililty for different offered load conditions. Further details about the simulation assumptions are described in Section 4.

Experiencing a Type 2 LBT blocking during the transmission of the downlink HARQ feedback, i.e. uplink control information, is also critical from the delay perspective. In scenarios targetting high reliability, an LBT blocking triggers a retransmission of the DL packet regardless of the decoding outcome at the UE side. The gNB is uncertain about the UE's decoding outcome as it did not receive a positive or negative acknowledgement on the expected uplink resources. This might force the gNB to retransmit packets already correctly decoded at the UE side, increasing the queuing delay for new transmissions. Moreover, it unnecessary increases the channel activity which directly impacts the channel access delay of the neighbour nodes. These implications clearly show the impact of Type 2 LBT failures on the latency performance, hence the importance of mechanisms that ensure a high Type 2 LBT success probability.

In the example shown in Fig.VI.1, nodes suffer from high Type 2 LBT failure rate as a result of simultaneous channel acquisition by multiple gNBs. Another circumstance in which nodes can be continuously blocked during the Type 2 LBT performance is the hidden node problem. The condition to experience the hidden node is that 2 nodes, e.g. gNB-1 and gNB-2, are out of the LBT range. Consequently, they sense the channel as idle even if any of the two are actively using the channel. However, a UE connected to the gNB-1 is highly interfered by the gNB-2 transmissions. Thus, in absence of other gNBs, the gNB-1 quickly acquires the channel and sends the UL grant. However, during the Type 2 LBT sensing interval, the UE is blocked by gNB-2 transmissions. To highlight the addressed problem, system-level simulations are conducted. The probability of being blocked by Type 2 LBT at the UE side is depicted in Fig. VI.2 under different traffic loads. It can be noted that the failure probability increases with the offered load as the channel activity accordingly increases. The addressed problem becomes significant in highly loaded scenarios, i.e. from 5 Mbit/s and above.

Downlink From the downlink perspective, the gNBs can also suffer from the same problem in a COT with multiple switching points. In this case, a Type 2 LBT blocking has less impact on the latency as compared to the uplink case. A single succesful Type 2 LBT is needed for transmitting the downlink grant (PDCCH) and the downlink data (PDSCH). However, sufficient performance degradation can be observed when consecutive channel accesses are failed due to high measured interference.

3 Proposed solution

Our proposal for reducing the probability of Type 2 LBT blocking probability is to introduce a pre-configured silencing pattern. The silencing pattern is defined by certain periodicity (T_{SG}) specified in transmission time intervals (TTI). Therefore, a node following the scheme and actively occupying the channel, it is required to stop the transmission on the specific TTIs to favour the channel access to neighbour nodes. The time the node must be inactive equals, at least, the duration of the Type 2 LBT sensing interval, i.e. 16 µs in Type 2B or 25 µs in Type 2A, depending on the configuration. For achieving the desired behaviour, it is assumed that all the deployed nodes configured with the same configuration. To better illustrate the proposal several examples are shown in Fig. VI.3. Four cases with different gap periodicities are depicted. For all the set-ups, it is assumed that 2 neighbour gNBs acquire the channel at the same time. However, different traffic conditions are assumed in each gNB. gNB-1 acquires the channel to transmit a full downlink burst during the entire COT, whereas gNB-2 has a more balanced DL/UL traffic and it selects a frame with multiple switching points. Fig. VI.3(a) shows the baseline mode of operation, i.e. no silent gap configured. In such configuration, any potential Type 2 LBT results unsuccessful due to the proximity between the nodes. As explained in Section 2, a latency performance degradation is expected in gNB-2 since all the UL and DL packets cannot be served in the given COT, delaying their transmissions to future COTs. Moreover, it results in a poor utilization of the COT resources, in this case only 2/6 of the available time is utilized. As an alternative in Fig. VI.3(b), the silent gap

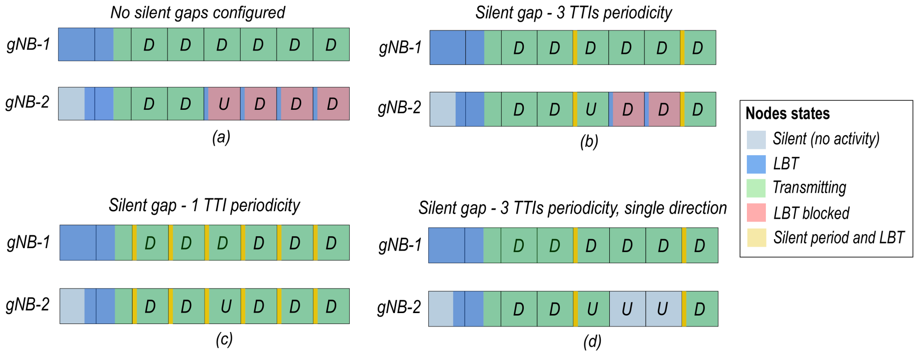


Fig. VI.3: Silent gap coordination scheme. Baseline NR-U operation is shown in (a). Silent gaps periodicities of 3 TTIs and 1 TTI are depicted in (b) and (c), respectively. In (d), silent gaps periodicity of 3 TTIs with single direction TTIs between single gaps.

coordination is introduced. In this case with a periodicity of $T_{SG} = 3$ TTIs. Type 2 LBT blocking probability is reduced as a blockage for transmissions in the slots immediately following the silent gap is avoided. However, subsequent transmissions still suffer from Type 2 LBT failure. To fully reduce the probability of LBT failure within a COT, the silent gap periodicity needs to match the TTI duration (see Fig. VI.3(c)). Alternatively, a complete mitigation of the Type 2 LBT blocking probability is also accomplished by avoiding any direction switching between the TTIs bounded by 2 consecutive silent gaps. As shown in Fig. VI.3(d), with $T_{SG} = 3$ TTIs a 0 % blocking probability is achieved. However, this approach can reduce the COT utilization as some resources might not be used due to lack of new packets for a given direction.

The silent gap coordination scheme represents an alternative to the solution presented in [5]. In that publication, a combined coordination scheme in the channel access and the TDD frame configuration is described. The channel access is coordinated by adopting the synchronous channel access defined in the ETSI regulations. The TDD coordination among nodes is ensured by the presence of a central node. The central node decides the next TDD frame configuration that better suits the traffic requirements of the gNBs. A common TDD frame aligns the sensing intervals of the responding devices, achieving full mitigation of the Type 2 LBT blocking. However, in situations with very uneven traffic conditions, it might introduce large queuing delays. Moreover, due to potential variability in the traffic requirements, the message exchange between gNBs and the central node might represent a significant control overhead. Additionally, the fact the proposal in [5] is centralized also increase the risk of failure since the mechanism essentially depends on the single node. In comparison with this proposal, the silent gap coordination mechanism allows each gNB to select the optimal TDD frame given their specific traffic conditions. It works without the presence of a central node

	Silent gap periodicity (T_{SG})		
TTI size	**3 TTIs**	**2 TTIs**	**1 TTI**
14 OFDM symbols	2.4 %	3.6 %	7.1 %
7 OFDM symbols	4.8 %	7.1 %	14.3 %
4 OFDM symbols	8.3 %	12.5 %	25 %

Table VI.1: Estimated overhead per COT for different silent gaps configurations. A COT duration of 3 ms and a SCS of 30 kHz are assumed.

and does not require message exchange. On the other hand, the silent gap coordination reduces the resource efficiency of the NR-U system due to the introduction of frequent gaps. A summary of the overhead introduced by the silent gap coordination is included in Table VI.1. The table shows that the impact of the overhead must be taken into account especially in cases with short TTI sizes and frequent silent gaps.

4 Simulation assumptions

The adopted layout complies with the 3GPP assumptions for indoor factory simulations [6]. The NR-U network consists of 12 ceiling-mounted gNBs with an inter-site distance (ISD) of 20 m. 60 uniformly distributed UEs are assumed with a total of 5 UEs are connected per gNB. UEs select the best cell based on the strongest reference signal received power (RSRP) criteria. If a UE select a gNB with 5 connected UEs, the UE is re-located in a different position and a new gNB selection procedure is triggered. The packet size is fixed to 50 B. Several traffic conditions are simulated by adopting different packet arrival rates (λ). The ratio 80-20 downlink and uplink traffic is always maintain. The rest of the assumptions of the simulations are listed in the Table VI.2. A highly-detailed system-level simulator in compliance with the latest 3GPP guidelines is used. It is capable to model with an OFDM symbol - subcarrier accuracy the most important PHY and MAC mechanisms involved in a transmission. The recently agreed propagation channel models are also available for simulations. Moreover, it implements the 3GPP guidelines for unlicensed operation in the 5 GHz band.

5 Performance evaluation

The first metric to be analyzed is the blocking probability of UEs during the performance of the Type 2A. As indicated in Table VI.2, UEs are configured to always perform Type 2 LBT when acting as responding devices regardless

Parameter	Assumption
Layout	Indoor office [6]
Channel model	3D geometry-based stochastic
Propagation model	Indoor factory (InF) [6]
gNB to UE	Sparse clutter, High base station
gNB to gNB	High Tx - High Rx
UE to UE	Sparse clutter, Low base station
Bandwidth	20 MHz
Sub-carrier spacing	30 kHz
TTI	7 OFDM symbols
Duplexing mode	dynamic TDD
gNB Tx power	23 dBm
UE Tx power	18 dBm
MIMO	2 × 2 configuration
Receiver type	LMMSE-IRC [7]
Traffic model	FTP Model 3 [8, 9]
Packet size	50 B
UEs processing times*	
DL data reception	4.5 OFDM symbols [10]
UL data preparation	5.5 OFDM symbols [10]
HARQ	Asynchronous, incremental redundancy
Max. retransmissions	2
Link adaptation	Outer link loop adaptation [11]
BLER target	1 %
gNBs channel access	Initiate the COT: Type 1 LBT During the COT: Type 2A LBT
UEs channel access	Initiate the COT: not allowed During the COT: Type 2A LBT
ED threshold	−72 dBm
Channel access priority class [3]	2

* gNBs processing times are assumed 2 times faster

Table VI.2: Simulation assumptions

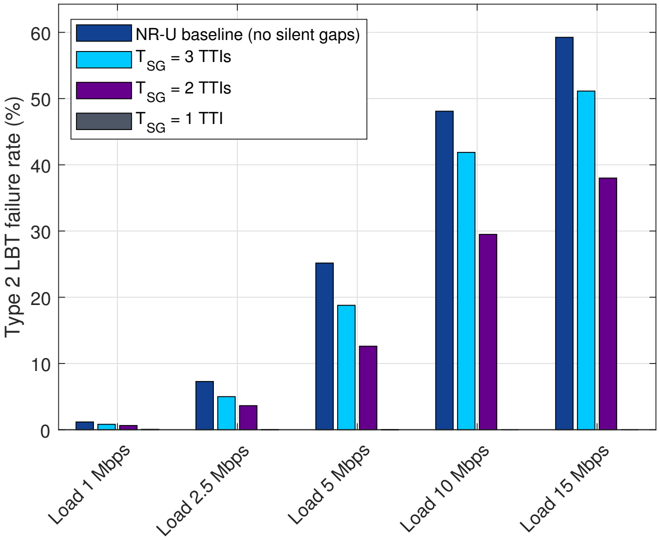


Fig. VI.4: Type 2 LBT blocking probability for different offered loads. It shows the combined Type 2 LBT outcome prior to uplink data and uplink control transmissions.

of the gap duration, i.e. Type 2C is not allowed. Different silent gaps under different offered load conditions are analyzed according to the Type 2 LBT blocking probability and shown in Fig. VI.4. Each bar represents uplink LBT blocking probability averaged over the 60 deployed UEs. Type 2 LBT prior to uplink data and uplink control transmissions are counted. Baseline operation, i.e. nodes are not configured with silent gaps, is also added for comparison. As shown previously, baseline operation exhibits a significant increase in the blocking LBT probability as the channel activity escalates. It is noted that any the adopted silent gaps configurations provide reduced blocking probabilities in comparison with the baseline case. Full mitigation of the Type 2 LBT blocking probability for any of the considered offered loads is achieved when silent gaps are configured with 1 TTI periodicity. It is worth mentioning here that, as discussed previously, the decrease in the Type 2 LBT failure rate is achieved in exchange for a higher overhead.

Reducing the number of transmissions blocked by Type 2 LBT has a straightforward effect on the latency and reliability performance. Under the same assumptions for the silent gaps configurations, the CCDF of the observed (DL + UL) packet delay in the scenario for an offered load of 5 Mbit/s is depicted in Fig. VI.5. As anticipated, the latency and reliability performance of the NR-U deployment is enhanced by any of the proposed configurations. This is due to fewer downlink retransmissions are triggered due to

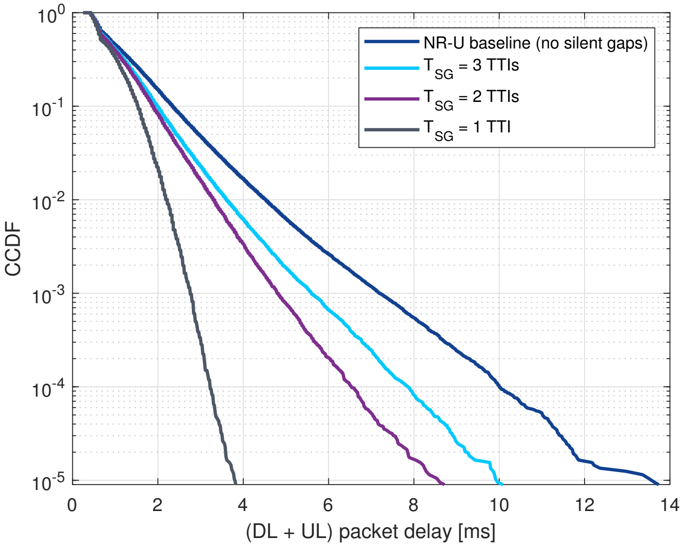


Fig. VI.5: CCDF of the combined downlink and uplink packet delay for different silent gap configurations. The offered load in the scenario is 5 Mbit/s.

absence of HARQ feedback and fewer uplink grants are needed for successfully completing a packet data transmission in uplink. Adopting a T_{SG} of 3 TTIs reports a latency decrease of ∼21.5 % with respect to the baseline configuration. Having more frequent gaps (i.e. $T_{SG} = 2$ TTIs) adds an additional ∼13.2 % on the latency reduction. The best delay performance corresponds to having a silent gap every TTI (i.e. $T_{SG} = 1$ TTI), it provides an extra ∼23.8 % with respect to the previous silent gap configuration. As compared to the baseline case, an overall latency reduction of ∼67.8 % is experienced. Similar observation are experienced for any of the considered offered loads.

6 Conclusions

A mechanism to reduce the probability of Type 2 LBT failure based on coordinated silent gaps is presented. It is shown to provide remarkable benefits as compared to baseline NR-U operation. Among the different configurations, adopting a silent gap periodicity of 1 TTI is the best option to fully remove the Type 2 LBT failure probability. The rest of the silent gap configurations help in decreasing the LBT problem but room for potential LBT failures is left. Adopting the most frequent silencing pattern reduces the spectral efficiency of the deployment due to the introduced overhead. In a 7 OFDM symbols TTI, a reduction of 14 % in the available resources is expected. In shorter

TTIs, e.g. 4 or 2 OFDM symbols TTI duration, the adoption of a frequent silent gap configuration can represent a large loss in the resource utilization. The adoption of this mechanism it is supported with the current NR-U Release 16 specifications. In downlink and uplink grants, the gNB decide, in a fully flexible manner, the starting and the duration of the transmission based on the start and length indicator (SLIV) [10]. Thus, nodes can modify their time allocation according to the presence of a silent gap.

References

[1] 3GPP, "Study on Scenarios and Requirements for Next Generation Access Technologies; (Release 16)," Tech. Rep. 38.913 V16.0.0, Jul. 2020.

[2] ——, "Service requirements for cyber-physical control applications in vertical domains," Technical Specification (TS) 22.104, 2019.

[3] ——, "Physical layer procedures for shared spectrum channel access (Release 16)," Tech. Rep. 37.213 V16.2.0, Jun. 2020.

[4] ETSI, *5 GHz RLAN; Harmonised Standard covering the essential requirements of article 3.2 of Directive 2014/53/EU*, ETSI EN 301 893 V2.1.1 (2017-05) Std.

[5] R. Maldonado, C. Rosa, and K. I. Pedersen, "A Fully Coordinated New Radio-Unlicensed System for Ultra-Reliable Low-Latency Applications," in *2020 IEEE WCNC*, 2020.

[6] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," Tech. Rep. 38.901, 2018.

[7] K. Pietikainen et al., "System-Level Performance of Interference Suppression Receivers in LTE System," in *2012 IEEE VTC*.

[8] 3GPP, "Further advancements for E-UTRA physical layer aspects (Release 9)," Tech. Rep. 36.814 V.9.2.0, Mar. 2017.

[9] ——, "Study on Licensed-Assisted Access to Unlicensed Spectrum (Release 13)," Tech. Rep. 36.889 V.13.0.0, Jun. 2015.

[10] ——, "Physical Layer Procedures for Data (Rel. 15)," Tech. Rep. 38.214, 2019.

[11] K. I. Pedersen et al., "Frequency Domain Scheduling for OFDMA with Limited and Noisy Channel Feedback," in *2007 IEEE VTC*.

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-833-9

AALBORG UNIVERSITY PRESS