eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Departamento de Ingeniería de Comunicaciones
Komunikazioen Ingenieritza Saila
Department of Communications Engineering

Ph.D. Thesis

Power Domain NOMA for 5G Networks and Beyond

Author: Eneko Iradier Gil

Supervisors: Dr. Jon Montalban Sanchez
Dr. Pablo Angueira Buceta

July 2021

This thesis is dedicated to each
of the people who have made
me who I am now.

Abstract

During the last decade, the amount of data carried over wireless networks has grown exponentially. Several reasons have led to this situation, but the most influential ones are the massive deployment of devices connected to the network and the constant evolution in the services offered. In fact, there are several applications that a few years ago were considered futuristic and that have now become a reality, such as augmented reality or virtual reality.

In order to classify the set of applications around 5G networks, the radiocommunications sector of the International Telecommunication Union (ITU-R) has defined three prominent use cases: enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communications (URLLC), and massive Machine Type Communications (mMTC). Each of these use cases is oriented to cover different connectivity needs with different requirements. For example, while eMBB offers high capacity to multimedia entertainment and information services, URLLC gathers critical applications that need high reliability rates and low latency. For its part, mMTC represents, to a large extent, Internet of Thing (IoT) networks and wireless sensor networks where a very high number of user/device connections is the major challenge.

In this context, 5G targets the correct implementation of every application integrated into the use cases. 5G must be understood not only as a standard for mobile communications but as an ecosystem of technologies, techniques, and standards. These latest communication standards present significant advances, such as more efficient coding or flexible waveforms. Nevertheless, the biggest challenge to make ITU-R defined cases (eMBB, URLLC and mMTC) a reality is the improvement in spectral efficiency. In this way, it will be possible to improve both the capacity offered and the number of users connected simultaneously.

In this thesis, a combination of two mechanisms is proposed to improve spectral efficiency: Non-Orthogonal Multiple Access (NOMA) techniques and Radio Resource Management (RRM) schemes. Specifically, NOMA transmits simultaneously several layered data flows so that the whole bandwidth is used throughout the entire time to deliver more than one service simultaneously. Then, RRM schemes provide efficient management and distribution of radio resources among network users. Although NOMA techniques and RRM schemes can be very advantageous in all use cases, this thesis focuses on making contributions in eMBB and URLLC environments and proposing solutions to communications that are expected to be relevant in 6G.

First, NOMA has been proposed as a mechanism that guarantees the transmission of broadcast content in eMBB applications. For this, the necessary architecture has been defined, and a software prototype has been built. Apart from broadcast content, in wireless networks, there is a large volume of unicast content. For this reason, a convergence system of the two services has been designed so that they can be distributed simultaneously. Subsequently, different RRM schemes based on NOMA have been designed, where the convergence has also been managed. Finally, the knowledge obtained has been applied to more specific use cases, such as user subgrouping techniques and wireless communications in the millimeter frequency band.

Regarding URLLC, the gain offered by the services based on NOMA has been oriented to improve different parameters, such as reliability, capacity, and latency. In a first work, a Wi-Fi transceiver prototype based on NOMA has been designed for industrial applications. Likewise, in order to increase the reliability gain, different retransmission schemes have been studied and implemented, in which spatial diversity schemes stand out. Finally, the influence of using NOMA in RRM schemes has also been studied, and it has been found that they improve the overall performance compared to orthogonal techniques, especially when the application and network requirements are high demanding.

Finally, this thesis has contributed to in-band full-duplex communications. It is expected that in-band full-duplex solutions will have great relevance in future communication standards due to their considerable increase in spectral efficiency. In this case, a system that integrates the distribution signal in the same frequency band of the broadcast content has been designed. This scheme is based on the American digital television standard (i.e., ATSC 3.0) and uses LDM, a specific NOMA technique, as a multiplexing mechanism between the traditional content and the distribution signal. Furthermore, as a system evolution, an architecture has been designed to interconnect the transmitters in charge of distributing the television signal. Finally, the most relevant challenges associated with this contribution have been identified, and solutions to be tested in the short term have been proposed.

Laburpena

Azken hamarkadan, haririk gabeko sareetan garraiatzen den datu kopurua esponentzialki hazi da. Hainbat arrazoik eragin dute egoera hori, baina eragin handiena izan dutenak sarera konektatutako gailuen hedapen masiboa eta eskainitako zerbitzu moten etengabeko bilakaera izan dira. Izan ere, gero eta gehiago dira duela urte batzuk futuristiko gisa planteatzen ziren aplikazioak, orain errealitate bihurtu direnak, besteak beste errealitate areagotua edo errealitate birtuala.

Aplikazio multzoa sailkatzeko, International Telecommunication Union (ITU-R) erakundearen irrati-komunikazioen sektoreak hiru erabilera-kasu nagusi definitu ditu aplikazioak taldekatzeko: Enhanced Mobile Broadband (eMBB), Ultra-Reliable low Latency Communications (URLLC) eta massive Machine Tympe Communications (mMTC). Erabilera-kasu horietako bakoitza beharrizan eta baldintza desberdinak betetzera bideratuta dago. Adibidez, eMBB entretenimenduko eta informazioko multimedia-zerbitzuetarako ahalmen handiak eskaintzera bideratuta dagoen bitartean, URLLC sistemetan fidagarritasun-tasa altuak eta latentzia baxuak behar dituzten aplikazio kritikoak daude. Bestalde, mMTCk, neurri handi batean, Internet of Thing (IoT) sareak eta haririk gabeko sentsoreen sareak ordezkatzen ditu, non erabiltzaile-dentsitate oso altuekin lan egin behar den.

Erabilera-kasuetan integratutako aplikazio guztiak behar bezala inplementatzen direla bermatzeko jaio da 5G, komunikazio mugikorren estandar gisa ez ezik, teknologien, tekniken eta estandarren ekosistema gisa ere ulertu behar dena. Nahiz eta azken komunikazio-estandarrek hobekuntza berritzaileak izan, hala nola kodifikazio eraginkorragoa edo uhin-forma malguak, ITU-R-ek aipatutako aplikazioak gauzatzeko erronkarik handiena eraginkortasun espektrala hobetzea da. Horrela, eskainitako gaitasuna eta aldi berean konektatutako erabiltzaileen kopurua hobetu ahal izango da.

Tesi honetan, modu konbinatuan bi mekanismo erabiltzen dira eraginkortasun espektrala hobetzeko: Non-Orthogonal Multiple Access teknikak (NOMA) eta Radio Resource Management eskemak (RRM). Zehazki, lehenengoa datu-fluxuak gainjartzean datza, banda-zabalera osoa erabil dadin aldi berean zerbitzu bat baino gehiago bidaltzeko. Bigarrena, berriz, irrati-baliabideak sareko erabiltzaileen artean eraginkortasunez banatzean eta kudeatzean oinarritzen da. Nahiz eta NOMA tekniken eta RRM eskemen erabilera oso onuragarria izan daitekeen erabileraren kasu guztietan, kontuan izanik tesiaren iraupena mugatua dela, ez da bideragarria hain irismen zabala ezartzea. Hori dela eta, tesi honen ardatza eMBB

eta URLLC inguruneetan ekarpenak egitea da, baita 6Gn garrantzitsuak izatea espero den komunikazioei irtenbideak proposatzea ere.

Lehenik eta behin, NOMA erabiltzea proposatu da, eMBB aplikazioetan broadcast edukien transmisioa bermatzeko mekanismo gisa. Horretarako, beharrezko arkitektura definitu da eta software-prototipo bat eraiki da. Broadcast edukiaz gain, haririk gabeko sareetan unicast eduki-bolumen handia dago, eta, horregatik, bi zerbitzuen konbergentzia-sistema bat diseinatu da, aldi berean bana daitezen. Ondoren, NOMAn oinarritutako zenbait RRM eskema diseinatu dira, non konbergentzia hori ere kudeatu den. Azkenik, lortutako ezagutza erabilera espezifikoagoko kasuetan aplikatu da, hala nola erabiltzaileak taldekatzeko tekniketan eta hari gabeko komunikazioetan frekuentzia milimetrikoen bandan.

URLLCri dagokionez, NOMAn oinarritutako zerbitzuek eskaintzen duten irabazia hainbat parametro hobetzera bideratu da, besteak beste, fidagarritasuna, ahalmena edo latentzia. Lehenengo lanean, aplikazio industrialetarako NOMAn oinarritutako Wi-Fi transtzeptore-prototipo bat diseinatu da, gaur egun aplikazio industrial gehienek beren Wi-Fi sareak erabiltzen baitituzte. Era berean, fidagarritasunean irabazi ahal izateko, birtransmisio-eskema desberdinak aztertu eta ezarri dira, eta espazio aniztasuneko eskemak nabarmentzen dira. Azkenik, RRM eskemetan NOMA erabiltzearen eragina ere aztertu da, eta teknika ortogonalekin alderatuta errendimendua hobetzen dutela zehaztu da, bereziki aplikazioaren eta sarearen eskakizunak zorrotzagoak direnean.

Azkenik, full-duplex komunikazioen esparruan ikertu da; izan ere, etorkizuneko komunikazio-estandarretan garrantzi handia izango dutela espero da, espektro-eraginkortasuna nabarmen handituko delako. Kasu honetan, broadcast edukiaren frekuentzia-banda berean banaketa-seinalea integratzen duen sistema bat diseinatu da. Eskema hau Estatu Batuetako telebista digitaleko estandarrean (ATSC 3.0) oinarrituta dago eta LDM, NOMA teknika espezifikoa, erabiltzen du, banaketa-seinalearen eta eduki tradizionalaren arteko multiplexazio-mekanismo gisa. Era berean, sistemaren bilakaera gisa, telebista-seinalea banatzeaz arduratzen diren transmisoreak elkarrekin konektatzeko arkitektura diseinatu da. Azkenik, ekarpen horri lotutako erronka garrantzitsuenak identifikatu dira, eta epe laburrean frogatuko diren irtenbideak planteatu dira.

Resumen

Durante la última década, la cantidad de datos que se transporta en las redes inalámbricas ha crecido exponencialmente. Son varios los motivos que han propiciado esta situación, pero los que más han influido son el despliegue masivo de dispositivos conectados a la red y la constante evolución en los tipos de servicios ofertados. De hecho, cada vez son más las aplicaciones que hace unos años se planteaban como futurísticas y que ahora se han convertido en realidad, como la realidad aumentada o la realidad virtual.

Con el fin de clasificar el conjunto de aplicaciones, el sector de radiocomunicaciones de la International Telecommunication Union (ITU-R) ha definido tres grandes casos de uso en los que agrupar las aplicaciones: enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communications (URLLC) y massive Machine Type Communications (mMTC). Cada uno de estos casos de uso está orientado a cubrir diferentes necesidades y bajo diferentes requisitos. Por ejemplo, mientras que eMBB está orientado a ofrecer grandes capacidades para servicios multimedia de entretenimiento e información, en URLLC se encuentran las aplicaciones críticas que necesitas altas tasas de fiabilidad y bajas latencias. Por su parte, mMTC representa, en gran medida, a las redes Internet of Thing (IoT) y las redes de sensores inalámbricos donde se debe trabajar con densidades de usuario muy altas.

Para garantizar la correcta implementación de todas y cada una de las aplicaciones integradas en los casos de uso nace 5G, que debe entenderse no solamente como un estándar de comunicaciones móviles, sino como un ecosistema de tecnologías, técnicas y estándares. A pesar de que los últimos estándares de comunicaciones presentan novedosas mejoras, como codificación más eficiente o formas de onda flexibles, el mayor reto para poder hacer realidad las aplicaciones mencionadas por la ITU-R es la mejora en la eficiencia espectral. De esta manera, se podrá mejorar tanto en la capacidad ofrecida como en la cantidad de usuarios conectados simultáneamente.

En esta tesis, se utilizan de manera combinada dos mecanismos para mejorar la eficiencia espectral: técnicas Non-Orthogonal Multiple Access (NOMA) y esquemas Radio Resource Management (RRM). En concreto, el primero consiste en utilizar superposición de flujos de datos para que se aproveche todo el ancho de banda durante todo el tiempo para enviar más de un servicio al mismo tiempo. En cambio, el segundo se basa en la gestión y reparto eficiente de los recursos radio entre los usuarios de la red. A pesar de que el uso de las técnicas NOMA y los esquemas RRM puede ser muy ventajoso en todos los casos de uso, teniendo en cuenta que

la duración de la tesis es limitada no es viable establecer un alcance tan amplio. Por ello, esta tesis se centra en aportar contribuciones en entornos eMBB y URLLC, además de proponer soluciones a comunicaciones que se espera que vaya a ser relevantes en 6G.

En primer lugar, se ha propuesto el uso de NOMA como mecanismo que garantice la transmisión de contenido broadcast en aplicaciones eMBB. Para ello se ha definido la arquitectura necesaria y se ha construido un prototipo en software. Además del contenido broadcast, en las redes inalámbricas existen un amplio volumen de contenido unicast, y por eso, se ha diseñado un sistema de convergencia de los dos servicios para que puedan ser distribuidos al mismo tiempo. Posteriormente, se han diseñado diferentes esquemas de RRM basados en NOMA, donde también se ha gestionado dicha convergencia. Por último, se ha aplicado el conocimiento obtenido a casos de uso más específicos, como son las técnicas de agrupamiento de usuarios y las comunicaciones inalámbricas en la banda de frecuencias milimétricas.

En cuanto a URLLC, la ganancia que ofrecen los servicios basados en NOMA se ha orientado a mejorar diferentes parámetros, entre otros, la fiabilidad, la capacidad o la latencia. En un primer trabajo, se ha diseñado un prototipo de transceptor Wi-Fi basado en NOMA para aplicaciones industriales, ya que a día de hoy la gran mayoría de las aplicaciones industriales utiliza sus redes Wi-Fi. Asimismo, con el fin de aumentar la ganancia en fiabilidad se han estudiado e implementado diferentes esquemas de retransmisiones, en los cuales destacan los esquemas de diversidad espacial. Finalmente, también se ha estudiado la influencia de utilizar NOMA en los esquemas RRM, y se ha determinado que mejoran el rendimiento en comparación con las técnicas ortogonales, especialmente cuando los requisitos de la aplicación y de la red son más exigentes.

Por último, se ha contribuido en el ámbito de las comunicaciones full-duplex ya que se espera que vayan a tener gran relevancia en los futuros estándares de comunicación debido a su considerable incremento en la eficiencia espectral. En este caso, se ha diseñado un sistema que integra en la misma banda de frecuencia del contenido broadcast la señal de distribución. Este esquema está basado en el estándar de televisión digital estadounidense (ATSC 3.0) y utiliza LDM, una técnica NOMA específica, como mecanismo de multiplexación entre el contenido tradicional y la señal de distribución. Igualmente, a modo de evolución del sistema, se ha diseñado una arquitectura para interconectar los transmisores encargados de distribuir la señal de televisión. Por último, se han identificado los retos más relevantes asociados a esta contribución y se han planteado las soluciones que se probaran a corto plazo.

Acknowledgements

Three years after starting the doctorate, it is time to close this stage by dedicating the last lines to all the people who have supported me during this time.

First, I want to thank Pablo and Jon, who introduced me to the world of research. It is their fault that I am here now. Pablo, thank you for offering me this opportunity and for the trust placed in me. Jon, thank you for everything you have taught me over the years and the friendship you have offered me from practically the first day.

In addition, I want to thank my colleagues in the laboratory with whom I have been working this period and have shared successes and failures. Marta, thank you for all the advice you have given me and the doubts that you have clarified during this stage, and for making the hours in the office more enjoyable. Thank you, Lorenzo, for the good times we had had when it was still possible to travel and those I know will come. To Iñigo and Rufino, who have just started their doctorate, cheer up and enjoy this stage; I know you are going to do great! To the "Txispas" (Mikel, Leire, JonGo, and Alex), it is still not very clear to me what you do, but it is a pleasure to share with you the coffee breaks and all kind of conversations outside of work. To Bea, who was one of the first people I met at TSR and with whom it has always been very easy for me to talk. Thank you, Iker, for all the good and funny moments we share inside and outside of work. Teresa, thank you very much for your help and your willingness; you make our work much more manageable. Thanks to Aritz and Aitor for helping us with your work and for the desire to learn that you have shown. Thanks also to the rest of the TSR group teachers since I have learned things from all of you, and you have made the way much easier.

I would also like to remember the CRC colleagues who made me feel at home when I was there. Thank you, Yiyan, for all the trust you have put in me, for all that I was able to learn from you, and for showing me so many places outside of work. To Wei, thank you for keeping an eye on me and making me feel so comfortable. And to Liang, thank you very much for sharing your knowledge with me and for helping me at all times. I hope we meet again in Canada.

Thanks to my former colleagues from Ikerlan, where I began to train as a researcher and have very good memories. In particular, I would like to thank Iñaki for the patience he had with me and all I grew professionally thanks to him. To Óscar, with whom I have shared long talks rambling about different technological concepts (and other non-technological ones) and with whom I am happy to continue working today. Also, thanks to

Cristina, with whom I shared many hours in the car and made the journeys to Arrasate much more enjoyable.

Lastly, I would like to thank my family. To my parents and my sister for having watched me grow up and transmitted the values that define me today. To Patxi and Gemma for treating me always like a son. But mostly thanks to you, Nagore. It is impossible to put into words everything you have meant in this stage and what you mean in my life. Thank you for being the way you are and for always being there. Thanks for everything.

Contents

List of Figures

List of Tables

Chapter 1

Thesis synthesis

1.1 Overview

1.1.1 Introduction

The efficient management of the rapid increase of the amount of traffic that travels through the network has become one of the most significant technological challenges of this decade. In fact, Cisco estimates that global mobile data traffic will increase by 49 Exabytes per month during this year 2021 [1]. Ericsson, for its part, forecasts a 30% annual growth of mobile traffic between 2018 and 2024, and most of this increase will come from video services [2]. One of the reasons for the exponential growth that the world is experiencing is the increase in the number of connected devices. In particular, by 2030, 50 billion devices are expected to be connected to the global network [3] of which 1.4 billion are concretely smartphones, and tablets [4]. On the other hand, another reason for the massive data increase is the rise in the demand for multimedia content such as video streaming or Augmented Reality (AR) and Virtual Reality (VR) services, which will require up to 100 Mbps per user [5]. What is more, the global demand of IP video services is expected to quadruple the demand from 2017 to 2022 [6].

Trying to respond to the increase in data demand, in 2015, the Radiocommunication Sector of the International Telecommunication Union (ITU-R) proposed a thematic group to analyze the role of emerging 5G technologies in future networks. This group is the so-called International Mobile Telecommunications-2020 (IMT-2020) and, among other things, classified the potential current and future applications into three use case families: enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communications (URLLC), and massive Machine Type Communications (mMTC) [7]. In Figure 1.1, a summary of the use cases and the applications conforming with each family is shown. On the one hand, the objective of

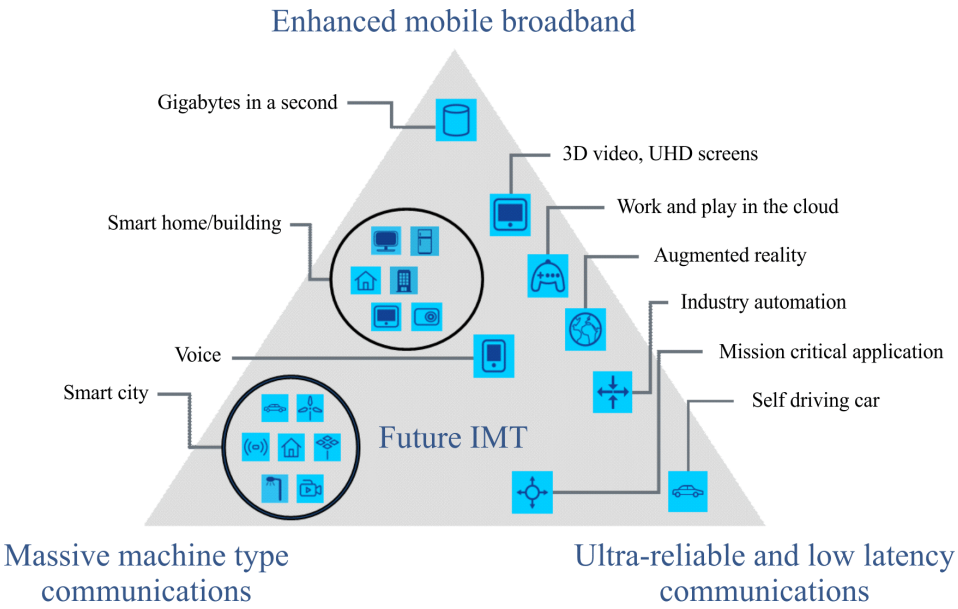


FIGURE 1.1: Use case and application classification in ITU-R M. 2083-0 [7]

eMBB is to provide seamless high data rate coverage, including high mobility scenarios. Some of the representative applications of this use case are 3D videos, 4k/8k contents, and AR/VR services. Then, mMTC is defined for Internet of Things (IoT) related networks, where low power consumption and low data rates for vast numbers of nodes are required. Some well-known examples of mMTC applications are Smart Cities and Wireless Sensor Networks (WSNs). Finally, URLLC is oriented to safety-critical and mission-critical applications, where the application environments are very relevant such as industrial automation or self-driving cars.

Due to the different philosophies of each of these use cases, the requirements associated with the correct implementation of each use case are different. Table 1.1 shows a summary of the most relevant parameters and their reference values as established in [8]. As can be seen, the critical parameters for each use case are different. For example, the most representative Key Performance Indicators (KPIs) of eMBB are related to data rate such as peak data rate, where up to 20 Gbps and 10 Gbps have to be guaranteed for the downlink and uplink, respectively. Then, the main requirements of URLLC are oriented to provide robust (i.e., error rate below 10^{-5}) and low latency (i.e., up to 1 ms in the user plane) services. Finally, concerning mMTC, the most relevant aspect is the connection density, expected around one million devices per km^2.

TABLE 1.1: Minimum requirements related to the ITU-R M. 2083-0 use case classification [8]

Parameter	Value		
	eMBB	URLLC	mMTC
Peak data rate	20 Gpbs in downlink 10 Gpbs in uplink	N/A	N/A
Peak spectral efficiency	30 bps/Hz in downlink 10 bps/Hz in uplink	N/A	N/A
User experienced data rate	100 Mpbs in downlink 50 Mpbs in uplink	N/A	N/A
Area traffic capacity	10 Mbit/s/m^2	N/A	N/A
User plane latency	4 ms	1 ms	N/A
Control plane latency	20 ms	20 ms	N/A
Connection density	N/A	N/A	1 000 000 devices per km^2
Reliability	N/A	10^{-5}	N/A
Mobility	Depending on the scenario up to 500 km/h	N/A	N/A
Mobility interruption time	0 ms	0 ms	N/A
Bandwidth	At least 100 MHz and up to 1 GHz in above 6 GHz bands		

5G was born to cover all the aforementioned use cases. 5G is characterized by a flexible physical layer (PHY) that covers the needs of different use cases [9]. Particularly, the first Release of 5G (i.e., Rel-15) was published in 2018 in [10]. This version focused on eMBB and included New Radio (NR) technologies. Then, Rel-16 was targeted to cover the rest of the use cases (i.e., URLLC and mMTC), as well as other aspects such as Vehicle-to-Everything (V2X) and coexistence with non-3GPP systems [11]. On the other hand, it is expected that Rel-17 and releases beyond (the last release will be Rel-20) will provide the 5G network of efficient communication tools. Some relevant examples are wireless and wired convergence, multicast and broadcast architecture, proximity services, multi-access edge computing, and network automation [12].

5G has triggered a whole ecosystem of technologies, applications, and opportunities around it, which aspire to lead the future of wireless communications. It is expected that 5G networks will take advantage of all these applications, technologies, and standards to become much more than a single standard for wireless communications. In the mid-term, 5G will be a whole technological ecosystem adaptable to almost every use case. In fact, this is the perspective that the 5G term will have in this Ph.D.

One of the technological benefits that 5G is expected to cover is the convergence of different wireless communication systems [13]. For example, the integration of satellite communication systems in the 5G ecosystem is proposed to improve coverage [14], or the convergence of different heterogeneous networks such as broadcast, broadband, and cellular services [15], or fixed and mobile services [16].

However, although 5G is still in its initial roll-out stages and a massive deployment is not expected in short/medium term, some limitations are already noticeable. According to recent literature, some of the most evident limitations are system coverage in ultra-dense environments, the lack of interconnection of everything (IoE), and non-zero-latency systems for industrial critical applications [17]. For this reason, the evolution of the technology will be likely oriented in two directions: 1) design and validate spectral improvement techniques that, in combination with existing technologies, improve the quality of service, and 2) pave the ground to define the requirements and use cases that 6G must cover.

Looking at spectrum efficiency and flexibility, this area covers the research of implementable solutions based on the current 5G ecosystem. To this end, a series of techniques with different characteristics could be used. Specifically, Table 1.2 shows a summary of the most relevant candidates. In particular, T1 and T7 are two techniques closely related to future 5G releases. The use of mmWave provides a considerable increase of the available bandwidth and presents an ideal situation for the deployment of a large number of transmission/reception antennas. Then, NOMA techniques are considered a promising candidate to enhance the spectrum efficiency by splitting the transmitted power to simultaneously serve a high number of users. Another suitable technique is the use high order constellations that would increase the transmitted throughput. This technique is being standardized in some other well-known communication standards such as Wi-Fi or broadcasting technologies. On the other hand, the efficient use and design of several network aspects could be critical for the spectrum efficiency, such as T4 and T5, representing different management aspects. Remarkably, while T4 is based on the intelligent design of virtualized systems such as VNF and SDN, T5 is oriented to the efficient management of the available radio resources. Moreover, coding systems have been traditionally very relevant for enhancing communication robustness, and today's coding schemes are very close to Shannon's limit. The last technique (i.e., T8) is oriented to detect underutilized frequency bands and manage their use for other applications to enhance the communication capacity and decrease network congestion. Finally, it should be highlighted that the listed techniques can be related to different levels of the Open Systems Interconnection (OSI) layer model, which can enhance the communication efficiency at different levels. Specifically, while T2, T3, T6, and T7 are PHY techniques, the rest are oriented to higher layers such as the network and transport layers.

On the contrary, for the second research avenue (path towards 6G), although there is no clear description of what 6G is, it is expected to have enhanced capabilities to broaden the applicable use cases. In fact, recent

TABLE 1.2: Summary of current most popular spectrum efficiency techniques

ID	Technique	Description	References
T1	mmWave bands	Enabling mmWave bands allows higher bandwidths to be used to offer higher capacity services	[18] [19]
T2	Non-Orthogonal Multiple Access (NOMA)	Applying power splitting, it is possible to serve a large number of users within the same resource block	[20] [21]
T3	Massive constellations	Constellation schemes higher than 256 symbols	[22] [23]
T4	Intelligent network management	Distributed network design based on Virtualized Network Function (VNF) and Software Defined Networking (SDN)	[24] [25]
T5	Radio Resource Management (RRM)	Optimization of the radio resource allocation stage	[26] [27]
T6	Advanced coding	Modern coding techniques are approximating Shannon's limit. For example, LDPC, turbo codes, or polar codes	[28] [29] [30]
T7	Massive MIMO	Communication systems with a high number of antennas. Testbeds of up to 128 x 128	[31] [32]
T8	Cognitive Radio	Dynamically lease underutilized frequency bands	[33] [34]

works envision the use of 6G for novel applications such as, in [35], where holographic communications are proposed; in [36], where the combination of Machine Learning (ML) and Quantum Computing (QC) seems to be near or in [37], where authors suggest that 6G will reinforce mobile ultra-broadband communications, super IoT and artificial intelligence. Another aspect of support is to develop a combination of use cases like massive ultra-reliable low latency communications (mURLLC) [38, 39]. In order to cover the new requirements and the proposed use cases, a group of disruptive technologies will be considered for their integration in 6G. Table 1.3 shows a summary of the technologies that are supposed to be key technology enablers in 6G. E1 is the straightforward step of implementing communications in the mmWave bands. However, challenges in spectrum management and network planning on the mmWave bands and the lack of equipment require further research efforts. Then, VLC is considered a promising complement to provide enhanced coverage or backhaul transmission. Considering that energy-friendly communications are required, the correct management of the energy sources and harvesting is considered a relevant topic of future sensor networks. Concerning E4, in-band full-duplex communications can considerably improve the network efficiency by transmitting and receiving simultaneously. Nevertheless, self-interference challenges should be previously overcome. Furthermore, network-based localization is considered a key aspect in industrial environments where the continuous localization of the workers is

TABLE 1.3: Overview of the 6G enabling technologies

ID	Technology	Description	Ref.
E1	THz communications	Frequency bands above mmWave. Higher bandwidths and higher capacity	[40]
E2	Visible Light Communications (VLC)	Complement RF transmission systems with light-emitting mechanisms	[41]
E3	Energy transfer and harvesting	Novel energy sources and management to enhance the energy efficiency	[42]
E4	Full-duplex communications	Receive a signal while also transmitting. Requires accurate interference cancellation	[43]
E5	Network-based localization	Using RF signals for user/device location and mapping	[44]
E6	3D network architecture	Using non-terrestrial transmission as complement. For example, drones or satellites	[45]
E7	Efficient backhaul solutions	Efficient backhauling is required to compensate for the network access increase	[46]
E8	AI-based learning	ML techniques to maximize the value of the information gathered	[47]

critical to guarantee their safety. Novel solutions based on the transmitted/received RF signals are foreseen in 6G. E6 represents the integration of non-terrestrial networks (e.g., drones, balloons, or satellites) in terrestrial network planning. Afterward, due to the exponential increase expected in network connection requests, efficient management and transmission of the backhaul signal is required. Finally, it is expected to use Artificial Intelligence (AI) techniques such as Machine Learning (ML) to manage all the data gathered in the network and make the network intelligent to learn from the information.

In general, it should be highlighted that both research avenues have different timing and expected impact. On the one hand, spectrum efficiency techniques for 5G are the set of solutions that provide network performance in today's systems. On the other hand, 6G-based solutions are long-term developments that cannot be implemented in short/mid-terms, but they can provide enormous performance gains. Therefore, future research lines should cover both proposed avenues in order to provide solutions for the present and the future of wireless communication systems.

1.1.2 Motivation

As described in the previous section, there is an urgent need to improve the spectrum efficiency of current communication systems to face the most demanding and challenging use cases. This thesis is oriented to enhance spectrum efficiency in current communication systems. Specifically, taking as a reference the list of potential techniques for improving the spectrum

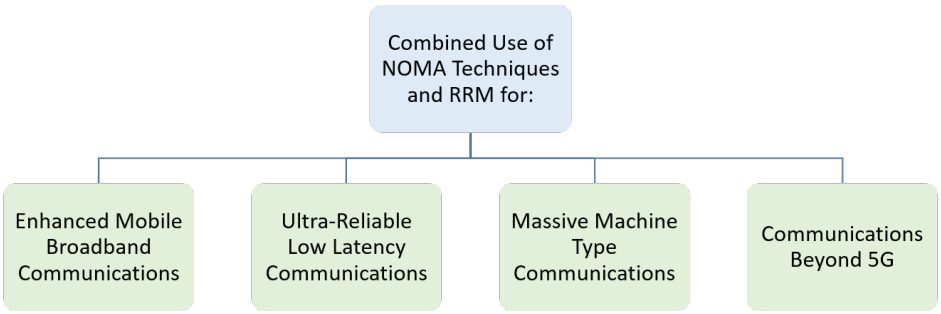


FIGURE 1.2: Main application scenarios of the combined use of NOMA and RRM techniques.

efficiency shown in Table 1.2, this thesis proposes the combined use of T2 (i.e., NOMA) and T5 (i.e., RRM) techniques with technologies of the 5G ecosystem (NR).

The first reason to support the direction of the thesis is the fact that these two technologies concern different OSI layers, which allows for improving spectral efficiency from two different perspectives. Another reason for selecting this research avenue lies in the short margin of PHY designs in current communication standards, which already perform very close to Shannon's limit. Therefore, new spectrum efficiency techniques not exclusive to the PHY have to be proposed.

Taking into account the existing use cases that may be relevant in short term, and taking as a reference the classification made in [7], four possible scenarios have been defined where the 5G results and future 6G networks can be noticeably improved by applying and testing the combination of NOMA and RRM. A summary of the possible scenarios is shown in Figure 1.2, and each one is detailed below:

- Enhanced Mobile Broadband: NOMA and RRM could be used in eMBB for better use of the spectrum to increase the overall network throughput or improve service quality in high-density environments.
- Ultra-Reliable Low Latency Communications: The spectrum efficiency improvement achieved with the combined use of techniques could be used to improve the reliability of the communications by offering more robust services. In this way, the strict requirements would be closer.
- Massive Machine Type Communications: The main difficulty of mMTC environments is to simultaneously guarantee the network access and the service to the high number of devices that comprise these networks. The joint use of NOMA and RRM could provide more

flexibility in the service design and, consequently, a higher number of devices served.

- Communications beyond 5G: Future communication standards are still unknown, whereas the weaknesses of current communication standards and the use cases that cannot be considered due to their strict requirements are reality. Therefore, the improvement in the spectrum efficiency obtained by the combination of techniques could be oriented to support the use cases currently out of the scope of current communication standards.

1.1.3 Document structure

This thesis is organized into three chapters. The content of each chapter is distributed as follows:

Chapter 1: This chapter provides a synthesis of the thesis. First, an overview is presented where the background of this thesis, and description of the motivations are introduced. Then, the technical context of the work is described based on a scientific literature review of the studies related to the topic. Afterward, the hypothesis and objectives are described. Finally, the next section discusses the results obtained in each contribution, while the bibliography closes the chapter.

Chapter 2: A summary of the main conclusions obtained from the work and the description of the future lines associated with each contribution are presented in this last chapter.

Chapter 3: This chapter gathers the papers related to each contribution. For each paper the main quality indicators are provided.

Finally, the appendix section lists other publications of the author, the resume, and the glossary.

1.2 Theoretical framework. State of the art

As shown in the previous section, the increasing demand for new applications and higher data rates is a reality that keeps growing. For this reason, the research community must provide communication systems with the most innovative techniques to improve the use of the spectrum. This Ph.D. thesis aims at designing and testing novel spectrum-friendly techniques. Figure 1.3 illustrates the technological context of this Ph.D. thesis, based on three pillars: NOMA, RRM, and 5G verticals.

On the one hand, NOMA-based techniques are considered promising multiplexing technologies for their considerable improvement in spectral efficiency compared to traditional orthogonal techniques. Therefore, they

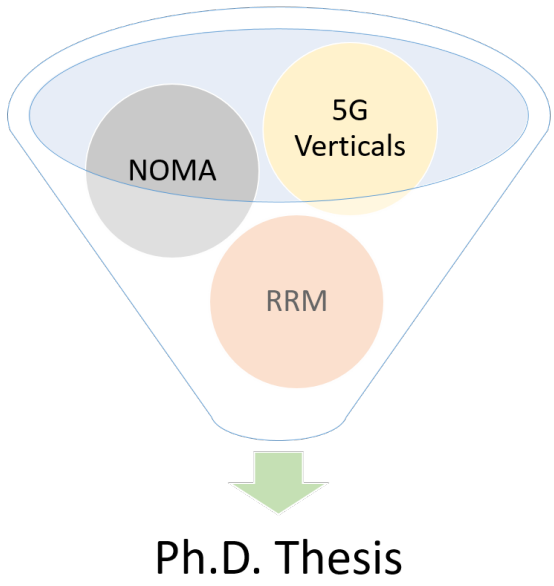


FIGURE 1.3: Main related work areas of this Ph.D. Thesis

can be one of the main evolutions of today's wireless communication standards. Section 1.2.1 presents an overview of NOMA techniques as well as an analysis of the main contributions in this area available in the literature.

On the other hand, the management of radioelectric resources is a fundamental piece in today's landscape, considering the scarcity of existing resources due to the increase in challenging and resource-demanding applications. Therefore, the optimized use of resources is critical in today's communication systems. Furthermore, the combination of techniques of different OSI layers, such as NOMA and RRM, can take the spectral efficiency to higher levels. Section 1.2.2 focuses on offering a detailed explanation of RRM techniques and analyzing the most relevant contributions of the current literature.

Finally, it should be noted that before developing any technology, the use cases where it will be applied must be defined. This step is a key requirement to define the features provided by the technology and expected KPIs. In this case, different use cases and applications within the so-called 5G verticals are evaluated, including also use cases outside the 5G ecosystem expected to be relevant in future wireless communications. The analysis of both 5G and beyond use cases has been presented in Section 1.1.1.

The following subsections provide a detailed tutorial and survey analysis of the two spectral efficiency techniques that have been jointly evaluated in this Ph.D. thesis: NOMA and RRM.

TABLE 1.4: Summary of OMA techniques used for mobile communications

Technique	Generation	Decade	Description
FDMA	1G	1980's	Exclusive frequency bands are allocated for each user
TDMA	2G	1990's	Exclusive time slot are allocated for each user
CDMA	3G	2000's	The transmitted information is mapped to orthogonal spreading sequences
OFDMA	4G	2010's	Allocation is carried out in time and frequency domain

1.2.1 NOMA

1.2.1.1 General concepts

Orthogonal Multiple Access (OMA) techniques have been widely used for service and user multiplexing in broadband communications. OMA techniques exclusively allocate time, frequency, and code resources to each user. Hence there is no interference among users, but the maximum number of users being supported is limited due to exclusive resource allocation. Table 1.4 shows a summary of the evolution of OMA techniques during the last decades and their use in mobile communication standards. The first implemented OMA technique was Frequency Division Multiple Access (FDMA) in 1G. In FDMA, each user transmits its particular signal in a dedicated frequency band, so the receiver can detect the signals from all the users by switching the frequency channel. Then, in 2G, Time Division Multiple Access (TDMA) was introduced to allocate exclusive time slots to each user to distinguish the different receptions by listening to the corresponding time slots. In the 2000's decade, with the introduction of 3G, a new concept was introduced: Code Division Multiple Access (CDMA). In this case, the users do share the same time-frequency resources, whereas the transmitted signal is mapped to orthogonal spreading sequences [48]. One of the most popular spreading sequences used in 3G was the Walsh-Hadamard coding scheme [49]. Finally, the last evolution of OMA techniques for mobile communications systems was Orthogonal Frequency Division Multiple Access (OFDMA). OFDMA could be considered as the combination of FDMA and TDMA techniques since the resources are allocated in a two-dimension grid (i.e., time and frequency) [50].

One of the main benefits of OMA techniques is that the users' signals do not interfere with each other. Therefore, the signals from each user can be obtained with low-complexity detectors. However, the number of orthogonal resources limits the number of users served with OMA techniques. Doubtlessly, this effect becomes a significant drawback for the massive communications considered in every vertical of 5G. That is why NOMA is presented as the enabling technology that can overcome the limitations of OMA techniques [51].

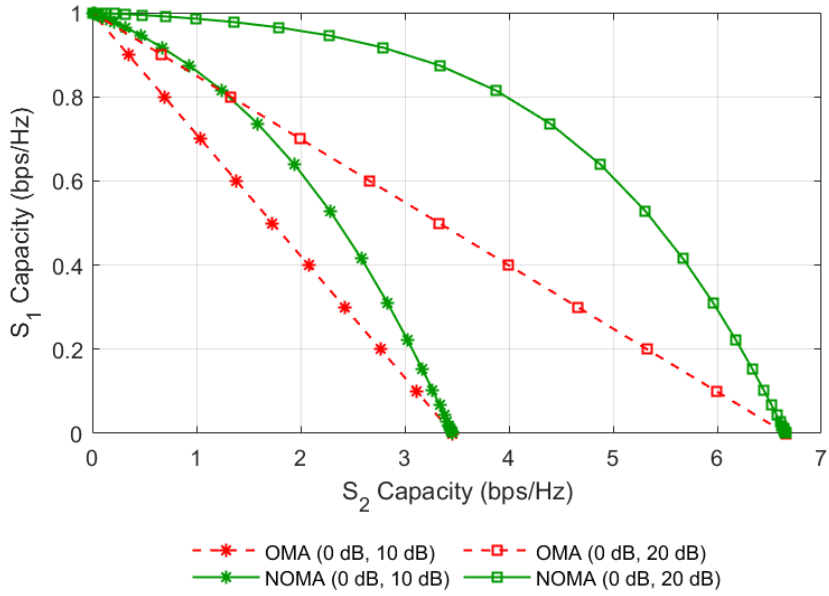


FIGURE 1.4: Service capacity comparison of OMA and NOMA techniques for different receiving conditions

The main distinguishing characteristic of NOMA techniques is the allocation of the same multi-dimensional resources (i.e., time, frequency, or code) to multiple users. Therefore, the whole channel bandwidth is used 100% of the time to deliver different services simultaneously. Hence NOMA techniques can provide service to a higher number of users than OMA techniques due to the non-orthogonal resource allocation system. Following the information theoretic analysis carried out in [52], NOMA presents a better use of the spectrum and, therefore, higher spectral efficiency. In Figure 1.4, a channel capacity comparison between NOMA and OMA is presented, where a pair of services transmitted to a Base Station (BS) is assumed. Two different Signal-to-Noise Ratio (SNR) combinations have been considered. It should be remarked that NOMA outperforms OMA in every case. For example, in services with 0 dB and 10 dB of SNR, a gain close to 0.2 bps/Hz and 0.5 bps/Hz can be obtained for S1 and S2, respectively. In addition, when 0 dB and 20 dB of SNR are assumed a gain higher than 0.4 bps/Hz, and two bps/Hz can be obtained for S1 and S2, respectively. In conclusion, the higher is the asymmetry between transmitted services, the higher is the gain obtained by NOMA in comparison with OMA systems.

In summary, NOMA techniques pose several advantages in comparison with OMA. A brief description of the benefits is introduced below:

- *Exploitation of channel capacity*: As shown above, NOMA techniques considerably improve spectral efficiency when compared to OMA techniques. The limit of the offerable capacity of NOMA is higher than that of OMA for all cases. For this reason, NOMA offers better exploitation of spectral resources.

- *Massive connectivity*: The introduction of NOMA techniques in resource management implies that the number of users served or the number of services offered is not limited by the number of orthogonal resources in the channel. Therefore, NOMA can offer simultaneous access to more users than OMA. Therefore, NOMA is better suited to the massive connections expected for 5G services.

- *Low medium access latency*: In traditional communication systems with OMA access techniques, such as 4G, each user must send an access request to the BS. The BS then manages each user's connection based on the available resources and sends the user a clear-to-send message to start the connection. Undoubtedly, these medium access mechanisms imply a considerable increase in the latency associated with transmission and an increase in overhead. Specifically, according to [53], access mechanisms based on Long Term Evolution (LTE) assume a latency of about 15.5 ms, which is well above the limits described in Table 1.1. On the contrary, some media access techniques based on NOMA (e.g., grant-free multiple access) do not require a connection request. In fact, with proper detection techniques at the receiver, it is possible to differentiate the information of each user, thus reducing the latency and overhead associated with the medium [54].

1.2.1.2 Taxonomy of NOMA techniques

NOMA is a vast concept that can be divided into different more specific techniques. Precisely, the first taxonomy, widespread throughout the literature, consists of dividing the NOMA techniques into two large families: power-domain NOMA (P-NOMA) and code-domain NOMA (C-NOMA) [21, 55, 56, 57]. As shown in Figure 1.5, there are some other NOMA techniques out of this generic classification. This section aims to describe the most relevant existing NOMA techniques.

C-NOMA techniques realize the service/user multiplexing on the code domain. In particular, the concept of C-NOMA is inspired by the classical CDMA schemes used in 3G, where multiple users share the entire available resources (i.e., in the time and frequency domain). However, the main difference between CDMA and C-NOMA schemes is that C-NOMA restricts

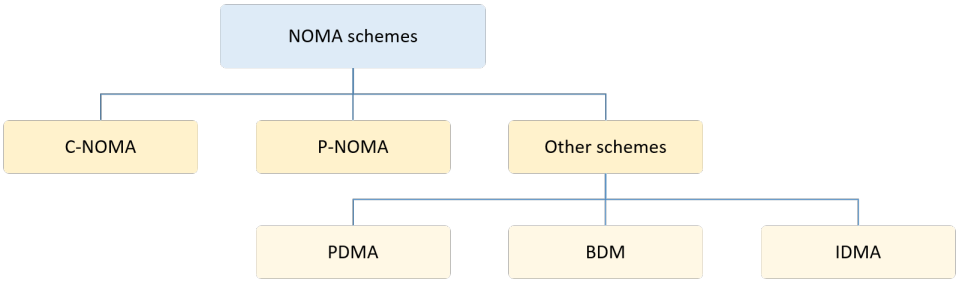


FIGURE 1.5: Classification of the main NOMA schemes

the spreading sequences to sparse sequences or non-orthogonal low cross-correlation sequences [21].

C-NOMA techniques can be classified according to different particular cases. One of the first C-NOMA techniques in the literature is Low-Density Spreading CDMA (LDS-CDMA) [58]. This technique is the first presentation of NOMA based on spread coding sequences. Its main objective was to decrease the effect of interference on each chip of basic CDMA systems [59]. Then, Low-Density Spreading-based OFDM (LDS-OFDM) was presented as an evolution of LDS-CDMA by combining it with OFDM systems [60]. In this case, the information symbols are first spread across low-density spreading sequences, and the resultant chips are then transmitted by a set of subcarriers [61, 62]. Finally, one of the major recent advances in C-NOMA techniques is Sparse Code Multiple Access (SCMA) [63]. Although SCMA is based on LDS-CDMA, the information bits can be directly mapped to different sparse codewords because both bit mapping and bit spreading are combined [64]. The main advantage of SCMA with respect to LDS-CDMA is that it implies lower complexity in the receiver side and improves the performance [65, 66].

The first P-NOMA approach was introduced by Cover and Bergmans in [67, 68] with the Superposition Coding (SC) concept. SC is a technique where the information for several receivers is simultaneously transmitted in the same resource. Then, on the receiver side, a Successive Interference Cancellation (SIC) module is required to decode the superposed information at each receiver. The basic principle of the SIC module is that user signals are successively decoded. Once the first signal is decoded, it is subtracted from the received signal, where all the signals are combined. Then, the next signal is decoded and subtracted. The iterative process is repeated until all the signals are correctly decoded. It is important to highlight that when a specific signal is decoded in the SIC process, the rest of

the signals are treated as interferers. Generally speaking, P-NOMA consists of different signals organized in several layers, where each layer takes some part of the total power transmitted by the transmitter. One characteristic of P-NOMA is that each layer can be independently configured in order to address different targets. Therefore, the configuration depends on the modulation, coding, and injection level, which is defined as the power splitting among layers, and it is described by g (or Δ if measured in dB). The generic P-NOMA signal ensemble can be expressed as:

$$x_{P-NOMA}(k) = x_{S1}(k) + g_1 \cdot x_{S2}(k) + ... + g_{N-1} \cdot x_{SN}(k) \tag{1.1}$$

where $x_{P-NOMA}(k)$ is the P-NOMA signal and $x_{S1}(k)$, $x_{S2}(k)$ and $x_{SN}(k)$ are the independent signals transmitted for users $S1$, $S2$ and SN. Moreover, k is the subchannel index. By applying injection levels to generate the P-NOMA signal, the total transmission power is distributed among the services to be transmitted as follows:

$$\sigma_i = \frac{10^{\frac{\Delta_{i-1}}{10}}}{\sum_{i=1}^{N} 10^{\frac{\Delta_{i-1}}{10}}} \tag{1.2}$$

where σ_i is the power allocated to the i^{th} layer, which is normalized (i.e., $\sum \sigma_i = 1$), and for the upper layer the corresponding injection level is $\Delta_0 = 0$ dB.

Although C-NOMA and P-NOMA are the main families of NOMA schemes, there are a few other NOMA schemes that are currently being investigated [69]. For example, Pattern Division Multiple Access (PDMA) [70] is a promising NOMA class that can be implemented in several domains. First, in the transmitter chain, PDMA is based on non-orthogonal patterns designed by maximizing the diversity and minimizing the overlaps among multiple users. Then, the effective multiplexing stage can be applied in the code domain, spatial domain, power domain, or combined. Another alternative is Bit-Division Multiplexing (BDM) [71]. BDM is oriented to downlink transmissions, and the main concept relies on hierarchical modulation, and the resources of multiplexed users are partitioned at the bit level. It should be highlighted that even if resource allocation in BDM schemes is orthogonal at the bit domain, the signals from the different users share the same constellation, which generates superpositioning in the symbol modulation domain. Finally, Interleave Division Multiple Access (IDMA) [72] is a special NOMA technique that performs chip interleaving after multiplying the symbols by spreading sequences. According to [72], IDMA has demonstrated superior perfomance if compared with CDMA.

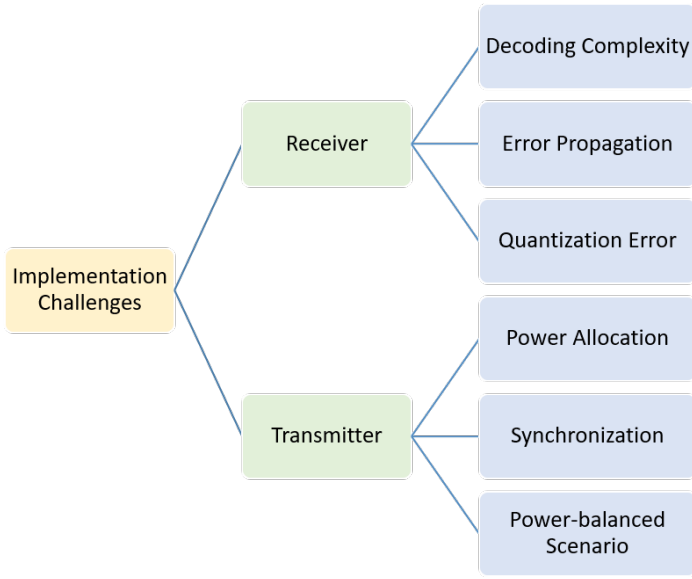


FIGURE 1.6: Main implementation challenges of NOMA techniques in the receiver and transmitter sides

Amongst all NOMA alternatives introduced in this section, this thesis is focused on low complexity P-NOMA systems. The selection criteria are based on their better tradeoff between complexity and performance [73]. Therefore, after this point, NOMA and P-NOMA will be used interchangeably, and where ever NOMA is used, it always refers to P-NOMA.

1.2.1.3 Implementation challenges

Although there are several advantages to use NOMA instead of OMA in wireless communications systems, its implementation brings several difficulties.. This subsection describes these challenges. Figure 1.6 shows a summary of the main implementation challenges that will be presented in this section classified into transmitter and receiver sides.

Decoding complexity. The first and the generally most known challenge is the decoding complexity increase. Compared to orthogonal multiplexing schemes, the decoding stage used in NOMA schemes requires additional processing for layer demapping. One of the most extended techniques is the SIC module. The SIC module decodes first all the data layers above the desired one [74]. Another alternative is presented in [75], which is based on the knowledge of the constellation of underlying signals. However, the complexity of the demapper is still high, and, consequently,

the overall complexity increases as the number of users in the network increases. In fact, in communications where latency is a concern, solutions that reduce the complexity associated with the NOMA decoding process are required. A possible solution used in cellular communications is to group users based on their channel condition and use multicast transmissions so that the SIC module process of each particular group has to decode only up to the layer oriented to themselves. User grouping algorithms such as [76] can be used to create groups of users. Then, in [20] it is highlighted that sequential decoding is only beneficial if there is a high probability of successful decoding. Considering that motto, the authors developed a decoding scheme based on the estimated successful decoding probability.

Error propagation. Then, another relevant challenge is related to the error propagation that occurs in the SIC module. In particular, when an error occurs in the SIC, this error is carried over to the other layers affecting the decoding of the rest of the layers and causing them not to be obtained correctly. To compensate the error propagation, the highest layers should be coded with more robust codes. In fact, authors in [77] demonstrated that the impact of the error propagation in the decoding is almost negligible when users with good channel conditions are scheduled above the users with bad channel conditions. Therefore, a tradeoff between robustness and throughput has to be assumed.

Quantization error. Quantization error is another problem related to the SIC module of NOMA systems. When the power of the received signals is very unbalanced, the analog-to-digital converter (ADC) needs to support an extensive dynamic range in order to obtain high resolution. However, as the implemented number of quantization levels is limited, the quantization noise in the bad channel condition signals (i.e., lower layer signals) could be relevant. Therefore, a tradeoff has to be assumed between the SIC gain and the cost of the quantization process.

Power allocation. In NOMA-based systems, the power allocation of each layer determines the throughput of each service, and consequently, the capacity of the network. If a very high injection level (IL) is used, the robustness of the upper layer (UL) is very high. Still, the lower layer (LL) requires considerably higher SNR to be retrieved. In contrast, low ILs make LL more feasible to decode, but UL is compromised. For this reason, the complexity associated with the choice of the optimum power allocation is critical in NOMA systems. Typically, brute force algorithms are used to determine the best combination of IL and each layer's layer configuration and users. However, this approach entails a tremendous computational cost. Therefore, algorithms such as [78] are required to estimate the best configuration of the NOMA parameters smartly.

Synchronization.Another widely investigated challenge is the layer

synchronization required in NOMA systems. In downlink communications, the BS manages all the transmissions of each user, making it possible to apply perfect synchronization. However, this accurate synchronization cannot be assumed in uplink communications because the UE-BS distance is not constant for all users. Moreover, the users may be continuously changing their position. This effect causes the misalignment of the overlapping codes in the received symbols, and the performance of NOMA gets compromised by the time-offset between the users participating in uplink communications [79]. One of the possible solutions is to increase the detection capacity of the receivers to be able to realign each one of the layers with the corresponding symbols [80].

Power-balanced scenario. Finally, there is a challenge associated with a specific type of scenario: the power-balanced scenario. As seen in Figure 1.4, the greater the unbalance between the services, the greater the NOMA gain. However, when the services are symmetric, the performance of NOMA is reduced, and in some cases, it can be even worse than that of OMA systems. Therefore, it is essential to plan the service configuration to maximize performance of NOMA properly.

1.2.1.4 NOMA applications

NOMA for broadcasting systems

The use of NOMA has been proposed for many applications. However, one of the most relevant milestones is linked to the area of broadcast communications. In fact, a relevant example of the success of NOMA techniques happened in 2017, when the North American ATSC 3.0 (Advanced Television Systems Committee) digital terrestrial television (DTT) standard [81] included a low-complexity NOMA solution, known as Layered Division Multiplexing (LDM) [82]. The key aspect of that success was improving key performance indicators (reliability and capacity, simultaneously) with an acceptable system complexity increase.

The idea of LDM was born from the proposal refered to as Cloud Transmission (Cloud Txn) [83]. Cloud Txn was a flexible multi-layer system that uses spectrum overlay technology to simultaneously deliver multiple program streams with different characteristics and robustness for different services in one RF channel. From there, the concept of Cloud Txn evolved into what is now known as LDM. In particular, LDM, similarly to Cloud Txn, is a two-layered transmission structure that simultaneously delivers several data streams with different power levels and robustness for different target users (e.g., mobile and fixed receivers). Specifically, LDM was designed to configure the UL with higher power allocation to transmit mobile services

and the LL to deliver high capacity services related to high SNR requirements for fixed receivers [82].

LDM was incorporated into the ATSC 3.0 standard for its higher capacity and low complexity requirements. Specifically, the low complexity was obtained by designing adequately several parameters. Firstly, the possible transmission power, signal coding, and modulation configurations were limited to a set of values [81]. Then, the LDM transmitter was designed to perfectly time-align the signal structure and the successive layers' baseband signals to enable a simplified successive signal cancellation decoding algorithm at the receiver side. In addition, signal decoding with very low SNR rates was enabled by Low-Density Parity-Check (LDPC) codes as inner codes. In fact, the latest LDPC codes present performance values very close to Shannon's Limit, around half a dB away from it [84]. Then, three possibilities were considered for the outer coding module: Bose-Chaudhuri-Hocquenghem (BCH) [85], 32 bit Cyclic Redundancy Check (CRC), and absence of outer coding.

The complexity of LDM has been widely studied. The authors of [86, 87] demonstrated that LDM involves a 20% of additional complexity in the decoding stage for an LDM ensemble with 16-NUC UL and a 256-NUC in the LL (NUC= Non Uniform Constellation mapping). This calculation is based on the number of iterations required by the LDPC decoding algorithm. The SNR threshold requirement for a typical LDM UL service is below 5 dB, and the received SNR in locations where the LL decoding is an achievable target should be around 15-20 dB. Consequently, the UL LDPC decoding in those situations (high SNR range) will require less than ten iterations. It should be noted that the typical maximum iteration number on LDPC DTV receivers is 50. Therefore, the pure complexity increase of the LL decoding process is up to 20% (i.e., 10/50). In practice, while the LDPC decoder of the single-layer mode (i.e., LL only) needs to carry out a maximum of 50 iterations, the LDPC decoder in a two-layer LDM receiver should perform a total of 60 iterations. Finally, it should also be remarked that the decoding process of the UL does not imply any additional complexity increase.

Along with the reduced cost of complexity, the capacity gain is the other fundamental reason to include LDM in ATSC 3.0. Specifically, the capacity gain of LDM in comparison with OMA techniques has been demonstrated on multiple occasions. The literature on the topic contains different approaches: theoretically [82, 88], through simulations [89, 90], and through measurements, both in the laboratory and the field [91, 92]. LDM has been proposed for other communication standards out of the broadcasting area. For example, LDM was proposed to be combined with 4G in [93], and [94], where the benefits of LDM for multicast resource allocation and service

TABLE 1.5: Summary of the studies that have been carried out by 3GPP concerning NOMA schemes

Release	Year	TR	Technology	Target description	Decision
Rel-12	2014	36.866	NAICS	Develop an interfering cell cancellation scheme	Accept
Rel-13/14	2016	36.859	MUST	Provide an analysis of downlink multiuser schemes	Accept
Rel-16	2018	38.812	Uplink NOMA	Provide an analysis of potential candidates for 5G NR uplink NOMA	Discard

convergence were highlighted, respectively. Then, in more recent works, the combined use of LDM and 5G NR is proposed in [95, 96, 97], where the enhancement in the overall offered capacity and the flexibility to cover the wide range of potential use cases of the 5G ecosystem are the main pillars of the proposals.

Finally, it should also be mentioned that there was another initiative from the European DTT area (i.e., Digital Video Broadcasting, DVB) to include in their standards multi-layer signal structure based on LDM. This technology was called Wideband reuse-1 (WiB) [98] and consisted of using all the available UHF band channels from all the transmitter sites (i.e., reuse-1) and spread out the transmitted power equally across these channels. Using this transmission planning and robust coding configurations, such as Quadrature Phase Shift Keying (QPSK) 1/2, the WiB signal would be delivered several dB below the principal signal transmitted in each UHF channel. Although high capacity gain rates were foreseen at the expense of assumable complexity/cost increases, WiB was finally not considered for approval.

NOMA in cellular networks

During the last few years, NOMA techniques have also been attractive for the 3rd Generation Partnership Project (3GPP) due to their performance benefits. Therefore, this section aims to provide an overview of the evolution that NOMA techniques have had inside 3GPP Work Items (WI). Table 1.5 shows an up to date publication summary of the 3GPP Technical Reports (TR) concerning NOMA.

The first approach of 3GPP concerning NOMA was in Release 12 (Rel-12) when they proposed Network Assisted Interference Cancellation and Suppression (NAICS) [99]. NAICS was a universal system not only for LTE, but also applicable in other interference cancellation schemes. The main idea of NAICS lay in using the signaling and blind estimation information to obtain the necessary parameters concerning cell interferences and, then, cancel the interference in advanced receivers [100].

Then, 3GPP approved a new study item for Rel-13 regarding downlink

multiuser superposition transmission (MUST) [101]. The main goal of the group was to provide an analysis of the potential candidates for efficient downlink multiuser transmission schemes. The primary KPIs that were evaluated during the work were the system-level gain and the complexity/performance tradeoff. Specifically, the proposed MUST solution was based on SC and could achieve the capacity of multiple access channels due to the SIC module [102]. The technical report was successful and approved during Rel-14.

Finally, concerning 5G NR, two research directions have been considered: downlink NOMA and uplink NOMA. Regarding downlink NOMA, the direct application of MUST to 5G NR looked straightforward. Nevertheless, the interest of 3GPP in NOMA has decreased since the downlink massive MIMO technology already offered significant gains, and the integration of NOMA would not provide enough benefit to overcome the capacity/complexity tradeoff [103]. That is why downlink NOMA is not considered for 5G NR. On the contrary, uplink NOMA schemes were more attractive for 3GPP, and a study on uplink grant-free NOMA transmission was approved [104]. The objective of the project was to provide an analysis of the benefits of uplink NOMA techniques over OMA schemes and decide whether it would be considered for 5G NR. Finally, the performance degradation that NOMA showed in some cases was the main reason for not considering it as a work item in Release 17 [20, 105].

NOMA for other recent applications

After analyzing the evolution of NOMA schemes both in broadcast environments and in broadband communications (i.e., 3GPP), this section analyzes some of the most recent publications that consider the combination of NOMA with other more specific research areas. Table 1.6 shows a summary of the technical works analyzed in this section.

Massive MIMO. The first technology is massive MIMO, which provides a drastic increase in network spectral efficiency and user connectivity. Massive MIMO has been extensively studied in combination with OMA, whereas joint use with NOMA could provide better performance rates. One of the first analyses is included in [106], where it is demonstrated that NOMA-based massive MIMO outperforms OMA. In addition, the sum rate values can be achieved when highly accurate Channel State Information (CSI) is obtained, and no inter-cluster interference is provided. Then, authors in [107] study the performance of MIMO-NOMA in multicell environments, and they propose a user classification and pilot allocation scheme based on the channel covariance information. On the other hand, the work in [108] is oriented to analyze the impact of the resource

TABLE 1.6: Summary of the main contributions on NOMA for different applications

Research area	Benefit	[Ref] - Description
MIMO-NOMA	Massive connectivity, High capacity, Spectral efficiency	[106] - Achievable sum-rate definition under high accurate CSI and without cluster interference
		[107] - Definition of user classification and pilot allocation based on channel covariance
		[108] - 3D resource allocation design for NOMA-aided massive MIMO
		[109] - Determination of the spectral efficiency under specific MIMO-NOMA parameters
		[110] - Achievable data rate in the case of beamforming uncertainty and residual SIC and channel estimation interference
mmWave NOMA	High capacity	[111] - Definition of the achievable rate bounds of mmWave MIMO-NOMA
		[112] - Proposal a low-feedback NOMA design based on multiple SISO-NOMA channels
		[113] - Analog precoder with a finite resolution to reduce the hardware cost
		[114] - Using NOMA to compensate the analog finite resolution of the beamforming
		[115] - Determination of the impact of beam misalignment in mmWave MIMO-NOMA
Cooperative NOMA	Reliability, Security, Spectral efficiency	[116] - Cooperative NOMA based on full-duplex relays
		[117] - Definition of a half-duplex/full-duplex switching algorithm
		[118] - Definition of a relay-aided transmission scheme based on user distance
		[119] - Definition of CoMP transmissions
		[120] - Coordinated beamforming scheme based on CoMP for cell-edge users
		[121] - Coordinated beamforming schemes to decrease the inter-cell interference in multi-cell MIMO-NOMA
NOMA in cognitive communications	Spectral efficiency, High capacity	[122] - Definition of a stochastic model to use NOMA in large scale CR networks
		[123] - Combination of CR-NOMA for relay communications
		[124] - Definition of NOMA-CR model for spectrum efficiency and user fairness
		[125] - Estimation of the impact of user pairing over the performance
PHY security with NOMA	Security, Secrecy, Anti-jamming	[126] - Definition of a BS communication system based on AN transmission in multi-antenna mode
		[127] - Analytical definition of the secrecy using SISO-NOMA under the presence of an eavesdropper
		[128] - Combined legitime and jamming transmission in a two-way NOMA-relay network under the presence of eavesdroppers
		[129] - Cooperative relaying scheme based on alternating jamming and forward signals
		[130] - Propose TAS a security increasing mechanism for MIMO-NOMA systems
NOMA-VLC	High capacity	[131] - NOMA-VLC and OFDMA comparison for two downlink users
		[132] - Definition of an algorithm to set the NOMA power allocation based on user channel quality
		[133] - Definition of QoS guaranteeing probability expression in NOMA downlink
		[134] - Proposal of a Chebyshev precoder to enhance the performance over non-linear channels
		[135] - Calculation of achievable sum-rate expression for NOMA-VLC assuming user mobility
		[136] - Experimental demonstration of bidirectional NOMA-OFDMA-VLC system

allocation in MIMO-NOMA. In particular, they define a three-dimensional resource allocation scheme (i.e., frequency, space, and power domain) that shows promising performance improvements. Moreover, quantifying the impact of the network/system parameters over the MIMO-NOMA communication systems is the main target of [109]. In this case, MIMO-NOMA is tested for relay communications. The parameters under study are the number of antennas at the BS, the number of relay nodes, and the transmitted power. Finally, in [110], the achievable rate values are estimated under a non-ideal situation. In particular, the authors assume that the orientation of the beamforming is unknown, and the SIC module performs imperfectly.

Massive MIMO in mmWave bands. Taking into account the extremely high data rates required by new use cases, the use of mmWave bands has become a need.This direction is addressed by research works that have been oriented to analyze the performance of MIMO-NOMA in the mmWave bands. It should be noted that specific MIMO techniques for mmWave bands have to be developed due to the particular challenges of these frequency bands. Firstly, the achievable data rate limits are described in [111]. Then, authors in [112] are interested in reducing the complexity and the network congestion, and, therefore, they propose a NOMA communication architecture based on low feedback. In this case, a tradeoff has to be assumed between the performance gains and the complexity. Concerning the hardware costs that suppose MIMO communications, it is widely assumed that analog precoders with finite resolution have to be used [113]. Taking this issue into consideration, authors in [114] propose to use MIMO-NOMA in the mmWave bands to compensate for the analog beamforming finite resolution. Finally, in [115] the impact of the beamform misalignment is analyzed in the mmWave environment for Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS) conditions.

Cooperative communications. Another alternative to improve the network capacity and also communication reliability is to use cooperative communications. The main idea of this technology is to enable cooperation between the network nodes and share the available resources, such as power-sharing or relaying schemes. For example, the work in [116] is oriented to enhance the performance of the communications by using cooperative NOMA communications in full-duplex relays. Authors in [117] propose a hybrid algorithm that smartly switches between half-duplex and full-duplex in randomly distributed spatial relay communications. Then, authors in [118] present a two-stage relay-aided NOMA scheme in order to improve the communication reliability in the far users. In particular, if the target user is considered to be in a near position, the BS directly establishes

the communications. For the users located in far positions, a relay node assumes the role of the transmitter. Concerning as well the serving quality of far users, Coordinated Multi-Point (CoMP) cooperative transmissions are designed [119]. Precisely, by using CoMP, multiple BS are coordinated to simultaneously implement user-oriented beamforming to enhance the cell-edge user performance [120]. Similar to CoMP, in [121], the beamforming coordination in multi-cell environments is oriented to decrease the inter-cell interference.

Cognitive Radio. Cognitive Radio (CR)-based communications is a well-known technology for better use of the spectrum. However, the combination of CR techniques with NOMA could provide an improvement in the overall performance. First, in [122], large-scale CR networks are covered using NOMA. Specifically, a stochastic geometry model is proposed in combination with a secondary transmitter to communicate with NOMA users. As an extension of the previous work, authors in [123] design a cooperative NOMA-CR network that implements several relays at the same time. Then, the same authors analyze more in-depth the schemes for user scheduling, and they propose in [124] two new alternatives with different purposes, one to increase the efficiency and the other one to provide fairness among users. Finally, the impact of user pairing algorithms is evaluated in [125] for CR-NOMA systems. In particular, they study the impact on the performance of fixed power receivers and CR-NOMA.

PHY security. Nowadays, wireless communications are exposed to cyberattacks, and, therefore, providing accurate PHY security is becoming more relevant. In particular, the use of NOMA in PHY security techniques could provide additional security to the traditional cryptographic approaches. A well-known technique to provide PHY security is to use Artificial Noise (AN). For example, in [126], a network is proposed where the BS can have single or multiple antennas, and in the case of multiple antennas, AN is generated to avoid eavesdroppers. An analytical evaluation of the secrecy probability is provided. A similar study is provided in [127], where a primary communication network is considered: cooperative SISO-NOMA implementing a single relay with an eavesdropper. The authors established the secrecy outage probability of the network for both Amplify-and-Forward (AF) and Decode-and-Forward (DF) protocols. Then, the work in [128] provides a broader analysis of the network since two NOMA-based relays are used, and a different number of eavesdroppers are assumed. To enhance the secrecy of the communication, the relay switches between confidential forwarding information and jamming signals. Authors in [129] consider as well the use of jamming signals. In fact, two roles are assigned to the network nodes. Source nodes have to transmit jamming signals, while the relay nodes are in charge of delivering the

forwarding message. Finally, another technique combined with NOMA to provide PHY security is Transmit Antenna Selection (TAS), which consists of using only one of the MIMO antennas to deliver the information and reduce the cost and complexity of the communication while the diversity is maintained. A relevant example can be found in [130], where the evaluation of TAS in combination with MIMO-NOMA shows promising results.

Visible Light Communications. Visible Light Communications (VLC) has to be also considered since they are a promising alternative for enhancing the network capacity and reducing the spectrum congestion. One of the first works considering NOMA in combination with VLC technology is [131], where a two-user downlink system is designed and compared with traditional OFDMA schemes. Results show better performance in the NOMA case. Then, in order to provide acceptable quality to the users, authors in [132] propose a Gain Ratio Power Allocation (GRPA) technique that takes into account the instantaneous channel quality of each user. In [133], the target is also to guarantee a minimum Quality of Service (QoS) in the NOMA downlink. Therefore, the QoS achieving outage probabilities are shown, and the data rate expression of opportunistic services. In addition, in [134], Chebyshev precoder is presented to enhance the performance of MIMO-NOMA in non-linear channels. Another aspect that has also been studied in VLC is user mobility. In particular, authors in [135] consider random user mobility and define the QoS outage probability and data rate expressions. Finally, as a proof of concept, in [136], the use of NOMA-VLC is experimentally tested using a bidirectional OFDMA network.

1.2.1.5 Open issues

Previous sections have described the considerable effort that has been invested in developing NOMA-based systems in order to improve the performance of current communication systems. Nevertheless, there are still various research lines associated with the areas studied. In particular, Figure 1.7, shows a summary of the leading research directions of NOMA organized by research areas and following the guidelines provided in several works [55, 56, 57].

Firstly, although one of the most significant landmarks of NOMA occurred in broadcast communications technologies (i.e., ATSC 3.0), there are new areas of research based on NOMA associated with the constant evolution in the services requested by users. For example, to improve the QoS and Quality of Experience (QoE) of the users, Hybrid Broadcast Broadband TV (HbbTV) is being implemented, where an uplink channel is necessary

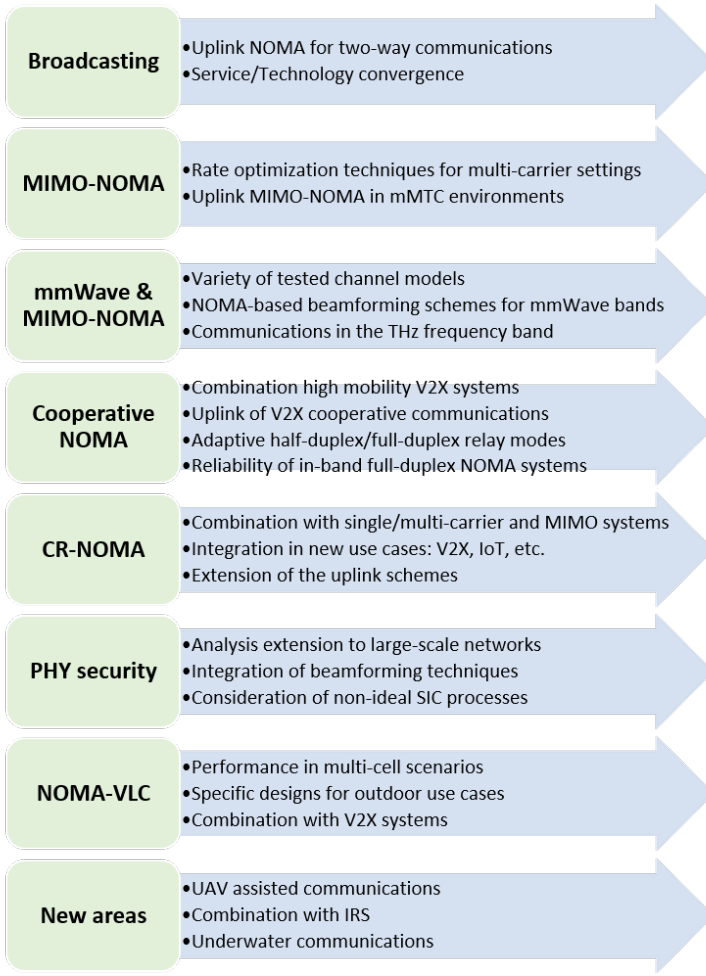


FIGURE 1.7: Future research lines of NOMA according to the presented areas

for users. However, the performance of these NOMA-based uplink channels has not been studied. Furthermore, the broadcast communications of the future will need to guarantee the convergence of services and technologies, for which NOMA may be a critical enabling technology.

Concerning more specific research areas, in MIMO-NOMA, most of the research works are oriented to single-carrier networks. Therefore, the rate optimization and data rate bounds should be more in-depth analyzed in multi-carrier environments. In addition, the uplink channel of MIMO-NOMA has to be exploited in ultra-dense environments (i.e., mMTC).

In the case of mmWave band NOMA, the majority of the works assumed LOS channel conditions due to the proximity to the transmitter.

However, in reality, several situations could lead to more challenging channel conditions that could be studied. Then, an interesting future research work could be to combine or design specific NOMA-based beamforming schemes for mmWave bands applications. Finally, as the experts foresee it, mid-term communications could be implemented in the THz frequency bands, and, therefore, the challenges and implementation issues of those bands should be investigated.

Regarding cooperative NOMA communications, two main research directions can be highlighted. The first is to develop cooperative NOMA systems for new use cases, such as Vehicle-to-Vehicle (V2V) or V2X. Within this line, the definition and performance of the NOMA-based uplink channel have to be studied as well. Then, the other direction is oriented to the full-duplex communications since few works have designed dynamic switching between full-duplex and half-duplex depending on the network conditions. Moreover, the research of the reliability and capacity increase of developing optimum in-band NOMA-based full-duplex communications should be attractive.

As in cooperative NOMA, CR-NOMA systems need to be combined with other technologies such as multi-carrier communications or massive MIMO. Concerning the target use case, the research lines could be broadened in order to cover high demanded applications like V2X or IoT networks. Finally, the uplink schemes of CR-NOMA should also be studied in-depth.

For NOMA based PHY security, the majority of the works are restricted to small networks. Therefore it would be interesting to apply PHY security techniques in large-scale environments. Furthermore, to improve the security rate, NOMA-based PHY security could be combined with beamforming techniques. Finally, almost all the works assume ideal conditions, whereas communications with imperfect SIC have not been studied.

In general, NOMA-VLC has been tested in limited scenarios, and testing the current solutions in multi-cell scenarios will be worthwhile. Following this idea, most of the research considered indoor environments, so the communications conditions are not as harsh as they could be in outdoor application. In addition, the application scope could also be broadened in order to study the integration of NOMA-VLC with other use cases such as V2X systems.

Finally, it should be noted that although they have not been mentioned previously, there some other demanding and novel areas where NOMA has not been applied, or at least very few works have been presented. For example, Unmanned Aerial Vehicle (UAV) applications represent one of the most promising communications areas for the future [137]. Although there are some preliminary works about the integration of NOMA in UAV

systems [138], there is still a crucial unexplored area. On the other hand, Intelligent Reflecting Surfaces (IRS) are becoming more popular during the last years [139, 140], and the joint used with NOMA could provide promising reliability gains [141]. Finally, underwater communications should be pointed out as an attractive research area for future communication systems [142, 143]. This research line should address solutions to acoustic NOMA systems.

1.2.2 RRM

1.2.2.1 General concepts

Considering the increasing demand for new multimedia services and stringent requirements, RRM techniques are essential to manage all the traffic that will be generated correctly. Specifically, to guarantee the challenging demands for capacity, data rate, or user density proposed by the new use cases, the management of resources has to be optimal. In this sense, according to several experts [26, 144], RRM techniques will be critical in the short term due to the following reasons:

- *Traffic volume*: The continuous demand of access to the services (i.e., everything, everywhere, and every time) and the increase in consumption of multimedia content, the volume of traffic that travels through the network is increasing exponentially. Therefore, the requirements demand a scalable and self-configuring network architecture and the need to interact with other Radio Access Technologies (RATs). Resource management under these circumstances must be optimal.

- *Connected equipments*: As with the traffic volume, the number of devices connected to the network in 5G environments is expected to multiply by 100 [7]. For this reason, the medium access mechanisms must be capable of managing both the access and the guarantee of minimum resources for all users. For this, RRM techniques are presented as a critical module that must be adapted to new needs.

- *Range of requirements and characteristics*: Another consequence of the continuous access to the services and the increase in available use cases is the wide range of QoS requirements. That is, the network must simultaneously handle both services with very low data rates (i.e., tens of kbps for text and voice applications) and applications oriented to extremely high throughput values (i.e., tens of Mbps for applications such as AR/VR). In addition, each type of service will be associated with specific characteristics such as maximum latency,

mobility model, or signal robustness. For this reason, RRM techniques will be necessary to correctly manage high volumes of traffic and with very different characteristics.

Traditionally, RRM techniques have been associated with the maximization of the amount of successfully transmitted information. However, due to the reasons shown above, that is not enough in future networks. Particularly, RRM techniques are focused on managing the Radio Access Network (RAN) resources. However, in mobile standards (i.e., 4G/5G), the RRM module is responsible for several tasks [26]. The most relevant functions of RRM techniques follow:

- *Resource allocation*: The RRM module has to share the available radio resources among the users or applications asking for access. In addition, the allocation process is not exclusively for the direct link between users and transmitter node; it has to be also managed for generic network access and backhaul signals. The resource allocation stage has to be adaptable depending on several network parameters such as the number of users, available resources, or channel conditions.

- *Packet scheduling*: The overall network capacity and the spectral efficiency are directly related to the correct packet scheduling carried out at each Transmission Time Interval (TTI) [145]. In particular, the scheduling process is carried out according to several indicators such as CSI, QoS requirements for the applications requested by the user, or Channel Quality Indicator (CQI).

- *Link adaptation*: In order to enhance the received throughput and the robustness of the services, link adaptation implements power control mechanisms and Adaptive Modulation Coding (AMC). What is more, the combined management of power and AMC control with the resource allocation can decrease the impact of external interferences or fading channel effects on the received signal.

- *Admision control and handover*: On the one hand, the radio admission control decides to provide network access to each requester user depending on the number of resources asked and the availability of the moment. When a user asks for network access, it is also checked if it is a new user or belongs to a handover from a neighboring cell. In this case, the RRM module decides whether to give network access and decide the number of resources required according to the QoS requirements of the user.

Even though the RRM concept includes various tasks, this Ph.D. thesis focuses on the resource allocation phase. On the one hand, because it is the one that is most closely related to the continuous improvement of the spectral efficiency and, on the other hand, because the rest of the tasks derive, finally, in the management of resources. Therefore, after this point, RRM and resource allocation will be used interchangeably, and where ever RRM is used, it always refers to resource allocation.

As mentioned before, when the transmitter node addresses the resource allocation, several parameters, such as the required bandwidth and rate or the mobility model, have to be taken into account and directly or indirectly affect the final decision. Therefore, to facilitate the resource allocation process, the whole framework can be divided into two sub-tasks: spectrum allocation and determination of transmission parameters [146].

On the one hand, the first task has to apply efficiently different spectrum management techniques to satisfy the QoS requirements of the users and the offered services. In particular, the spectrum allocation task has several inherent challenges. For example, it has to know at each moment which is the amount of available spectrum. Then, the resource allocation task has to decide which is the optimum spectrum portion at each moment for each user and synchronize the access and the spectrum sharing among all the requesters. In fact, sometimes (e.g., in high concurrence environments), prioritization decisions have to be taken to limit the network access or share the whole available spectrum. Finally, to improve the spectrum management and the response velocity, learning capabilities could be introduced [147].

On the other hand, to efficiently serve a user according to its corresponding QoS request, the RRM stage has to determine the value of several transmission parameters (i.e., transmission power, MCS value, rate). As in the previous case, this task also has its own challenges. First, the RRM module has to know and carry out the resource allocation depending on the maximum tolerable interference of each user. This parameter is usually related to the CQI feedback, which determines the maximum MCS value that the user can decode according to the channel conditions. Then, the transmission parameters have to be determined to provide a minimum QoS, and those parameters can reflect SNR values, minimum data rates, or maximum tolerable interference, among others. Therefore, the RRM has to handle all the requirements simultaneously, even if they are related to different KPIs.

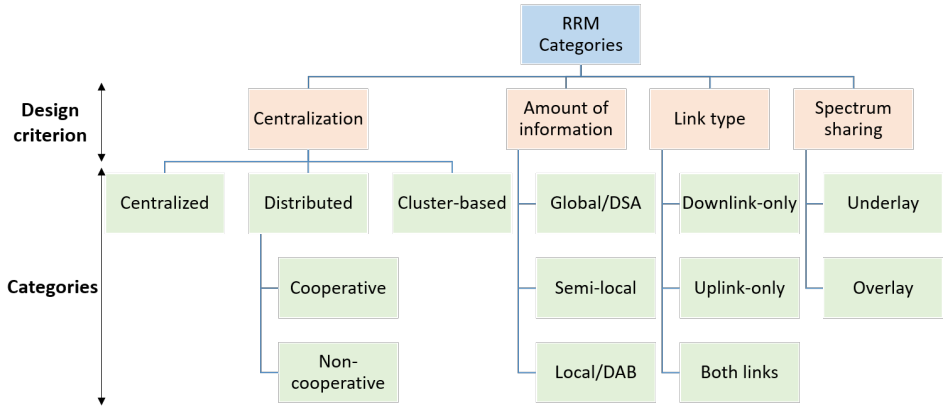


FIGURE 1.8: Classification of RRM techniques according to different criteria

1.2.2.2 Taxonomy of RRM techniques

Many resource allocation techniques found in the literature are based on different optimization criteria and oriented to different system settings. The criteria to classify allocation techniques does not follow a single criterion [146, 148]. This section presents the most widely used RRM technique classifications and a summary of the categories and criteria is shown in Fig. 1.8. As each division corresponds to a particular criterion, each RRM scheme can be classified in various ways.

Centralization degree. One of the most typical classification criteria is the centralization degree from which three categories arise: centralized, distributed, and cluster-based. Firstly, the centralized radio resource allocation scheme is based on a network where the transmitter node is the central node and is in charge of power control, and spectrum allocation decisions [149, 150]. Generally, these schemes implement a two-phase architecture. In the first phase, the spectrum state and the users' requests are collected. Then, in the second phase, the resource allocation decisions are made according to the results obtained in the first phase. The main advantage of this scheme is that the central node has a global view of the entire network, and, therefore, it can achieve the optimization goals easily. On the other hand, the principal drawback is the high signaling overhead caused by the message interchange between users and transmitter node. Furthermore, in the centralized architecture, the central node broadcasts the resource allocation decision to all the nodes included in the network. In extensive networks, high power allocation is required to communicate with the cell-edge user. In addition, if the broadcast communication fails, non of the network user receives the allocation update. In particular, this

category is very typical in mobile networks and in some sensor networks, such as monitoring networks [151].

On the contrary, distributed resource allocation schemes are not based on any central node. In this case, each user is responsible for its transmission decisions [152, 153]. In particular, the decisions can be taken by cooperating with the rest of the user in the network (i.e., cooperative) or following an autonomous way (i.e., non-cooperative). In both cases, a quicker adaptation to changes than in the centralized scheme is obtained because no feedback is required. In addition, due to the same reason, the overhead and network traffic are decreased. However, the main disadvantage of distributed schemes is that each user decides its own resource allocation according to just local information, which can negatively affect the rest of the users by decrease their throughput or reducing the fairness among users. Generally, distributed networks have been oriented to specific cellular networks such as ad-hoc networks, relay-based networks, or cognitive-radio networks.

Finally, the last category determined by the centralization degree is the so-called cluster-based architecture. In this category, the network is subdivided into different clusters, where each cluster is led by a Cluster Head (CH) that collects the information and manages the allocation of the cluster [154, 155]. Among the advantages, low power is required because the CH has to handle a portion of the whole network and, consequently, lower overhead has to be assumed. In addition, when a failure occurs in one CH, the user of that cluster can be assigned to the nearest one. Concerning the drawbacks of cluster-based schemes, the most representative one is the increase in the number of broadcast signals in the entire network. In particular, this architecture is suitable for large-scale multi-hop networks.

Amount of information. Then, if the amount of information shared in the network is taken into account, the possible schemes can be categorized as global (i.e., Dynamic Spectrum Access, DSA), local (i.e., Direct Access-Based, DAB), or semi-local. DSA schemes require information from the entire network to determine an allocation scheme that improves the overall performance [156, 157]. Doubtlessly, these schemes require a high amount of information, and, therefore, high computational cost and high latency have to be assumed to manage all the information. However, the performance of the network is close to the optimum working point. These schemes are practical in networks where a high synchronization rate must be obtained between the resource allocation entity and the users.

Unlike DSA, the DAB scheme implies that the information used in the allocation process is limited to the allocator-receiver pair. In this case, the

resources allocated to each user are negotiated individually with the central entity [158, 159]. This scheme needs low complexity, whereas the network efficiency is also low, and, therefore, the allocation decision might be far away from the optimum solution. DAB is relevant for developing local allocation schemes and for on-demand transmissions.

In order to offer a balanced solution between DSA and DAB, semi-local schemes should be considered. In semi-local allocation schemes, the information is only available in small areas of the network [160, 161]. Therefore, the complexity requirements and efficiency are balanced. This solution is suitable for heterogeneous networks, where users with very different requirements and applications have to co-exist in the same network.

Type of link. If the type of link or the communication direction is taken as a reference, it is possible to distinguish between uplink, downlink, and mixed schemes. In the first case, uplink-only allocation schemes are required when the information only flows from the users to the gNB [162, 163]. This structure implies low overhead and low latency, especially in asynchronous communications. Nevertheless, the main drawback is the transmit power constraint since, generally, mobile users are feed by batteries. A relevant example of uplink-only schemes is sensor networks that transmit their information to a central node.

In the case of downlink-only schemes, users are passive receivers, and the central node takes all the responsibility of the resource allocation. Implementing downlink-only schemes, it is easier to obtain high efficient spectrum use because the allocation process is centralized in the same node [164, 165]. Moreover, the synchronization requirements are more flexible with this scheme. On the contrary, downlink schemes need feedback from the users to carry out the allocation process, which can be based on different parameters such as channel conditions, distances to the gNB, or minimum data rate. A typical example of downlink-only communications are networks implementing master-slave architecture where the slaves execute the orders delivered by the master.

Then, downlink and uplink allocation schemes can be combined and implement the allocation of both links simultaneously [166, 167]. Undoubtedly, these two-plane allocation schemes are much more efficient than the single-plane schemes (i.e., downlink/uplink-only). However, as two planes have to be configured, higher computational cost and complexity must be assumed. These schemes are typical examples of mobile communication systems where users and the central node need to interact. They are also representative of full-duplex communications, which are highly demanded for future communication systems.

Spectrum sharing. Finally, another distinction could be made according to the way of implementing spectrum sharing. This criterion divides

the allocation techniques into underlay access and overlay access. Underlay access scheme implies that part of the spectrum is shared under the noise floor by using spread spectrum techniques [168, 169]. In underlay schemes, strict transmission constraints must not interfere with the rest of the spectrum transmission. Moreover, the complexity of the receiver is increased in order to decode the signals delivered correctly. On the contrary, the benefits of underlay schemes are that low transmission power is required due to the spread spectrum techniques, and the spectrum efficiency is increased. These schemes are suitable for converge priority-based communications where each service has different requirements.

On the other hand, overlay schemes only permit to carry out transmissions when the wireless channel is available [170, 171]. Therefore, it is required to identify the available resources within the entire bandwidth continuously. The main benefit of overlay schemes is the low complexity required to schedule the transmission and the low interference generated to the rest of the transmissions. However, those benefits are obtained due to relaxing the spectrum efficiency obtained with the underlay schemes. The most popular example of overlay allocation technique is OFDMA-based allocation systems.

To conclude, Table 1.7 shows a summary of the RRM categories mentioned above according to the different design criteria. The table highlights the main characteristics of each category and the most representative benefits, drawbacks, and applications.

1.2.2.3 Optimization strategies

Traditionally, resource management algorithms have targeted network capacity maximization to serve higher user data rates. However, with the increasing requirements imposed by new applications and the increase of data transported over the network, the scarcity of resources is getting more evident. For this reason, the most recent research works related to resource management have broadened the scope, and different optimization parameters are proposed. This section presents the most relevant optimization parameters of today, along with some of their most representative works.

Throughput. One of the most used optimization metrics in resource allocation schemes is throughput, which is calculated as the sum of data rates sent to each of the users in the network. An excellent example of throughput maximizing resource allocation technique is shown in [172], where a user association and resource allocation scheme is proposed. The algorithm is based on a Mamdani-type Fuzzy Logic Controller (FLC) that first classifies the users according to the data rate requested and allocates the corresponding resources to each user. The results show better performance

TABLE 1.7: Summary of the RRM technique categories according to different design criterion

Design criterion	Description	Category	Advantages	Disadvantages	Related applications	References
Centralization	Based on the network architecture type and the responsibility for the decision making	Centralized	Complete network view. Easier to achieve optimization goals	High overhead. High power. Failure vulnerability	Mobile and sensor networks	[149] [150]
		Distributed	Low overhead. Quick adaptation	Based on local data. Non-fair	Ad-hoc and relay-based networks	[152] [153]
		Cluster-based	Low power. Low overhead. Robust against failures	Increase in the broadcast signals	Large-scale multi-hop networks	[154] [155]
Amount of information	Classification based on the required quantity of shared information	Global/DSA	Enhancement of the entire network performance	Latency increase. Computational cost	High synchronized networks	[156] [157]
		Semi-local	Low complexity. Balanced solution	Non-optimum	Heterogeneous networks	[160] [161]
		Local/DAB	Low complexity	Low efficiency. Non-optimum	On-demand content transmission	[158] [159]
Type of link	The direction of the data flow determines the category	Uplink-only	Low overhead. Low latency	Transmit power constraint	Sensor networks	[162] [163]
		Downlink-only	High efficiency. Easy synchronization	Need of feedback	Master-slave networks	[164] [165]
		Both links	High spectrum efficiency	Two-plane management. High complexity	Mobile and full-duplex communications	[166] [167]
Spectrum sharing	Determination according to the spectrum organization	Underlay	Low transmission power. High efficiency	High complexity. Transmission power constraint	Priority-based communications	[168] [169]
		Overlay	Low complexity. Low interference	Low efficiency	OFDMA-based systems	[170] [171]

in terms of data rate and bandwidth usage compared to typical greedy-based approaches. The main drawback of this solution is that the network load balance issue is not considered, and an unbalanced situation could lead to a decrease in the offered throughput. Then, the authors in [173] define a complete scheme where the maximization of the offered throughput is the fundamental goal. At the same time, cross-tier interference and the limitation of the transmitted power are considered. The allocation scheme is based on underlay communications, and the optimization problem is solved with the Lagrange dual decomposition method. Results indicate that the optimum configuration is obtained assuming very few iterations. However, the interferences generated inside the network (i.e., co-tier interference) are not considered. In [174], the resource allocation issue is jointly faced with routing. In this case, the resource allocation is solved using a stochastic algorithm, while routing is fulfilled with a linear programming solver. The designed method demonstrated that the collaborative process enhances throughput, spectral efficiency, and energy efficiency. The main disadvantage of this analysis is that the latency involved in the routing decision is not evaluated.

Finally, a family of resource allocation algorithms called subgrouping needs special mention. Subgroupoing algorithms aim at maximizing the throughput of the network by making group allocations [175, 93]. The first work shows an algorithm that distributes the users into groups and decides the configuration of each one (i.e., MCS). In contrast, the second one continues with the same management logic but manages the resources using NOMA. The results indicate that subgrouping techniques considerably improve network throughput. However, the main disadvantage of this technique is that it is based on the analysis of the performance of the possible configurations, which implies a high computational cost.

Spectrum efficiency. Spectrum efficiency is another of the most spread optimization metric, which represents the capacity that can be delivered within a specific amount of spectrum (i.e., bps/Hz). Authors in [176] have developed a relevant allocation scheme based on a clustering game algorithm that also takes into account the generated interferences (co-tier and cross-tier). Two steps are followed; first, the clustering is managed following graph theory games, and then, the resources for each cluster are allocated with an auction game mechanism. The main benefit is that it is compatible with frequency reuse and adaptive small-cell throughputs. On the contrary, the type of service requested by the users is not taken into account. Then, authors in [177] combine the spectral efficiency optimization with the energy efficiency for cooperative small cells. Specifically, two mechanisms are developed, one oriented to interference management and another one to the traffic offloading. The results indicate that both mechanisms outperform non-cooperative schemes in terms of spectral efficiency and energy efficiency. Nevertheless, only co-tier interference issue is taken into account. The last work can be found in [178], where the resource allocation is managed to jointly optimize spectral efficiency and energy efficiency for an entire area. To carry out the joint evaluation, a binary search algorithm is designed. The evaluation shows that even several parameters are constantly changing (e.g., bandwidth, power consumption, or user density), spectral efficiency and energy efficiency can be improved following the designed algorithm. In this case, the overhead that is necessarily generated in the network is the worst aspect.

QoS/QoE. User satisfaction with the received service is one of the most relevant issues in today's wireless communications, and, therefore, QoS/QoE-based resource allocation schemes have to be also taken into account. To cover this issue, in [179], a smart RAT accessing and resource allocation technique is designed for heterogeneous networks that guarantee the QoS fulfillment based on a multiagent reinforcement learning technique. The authors solve first the RAT selection with a Nash Q-learning algorithm, and then, the allocation is carried out with a Monte Carlo Tree

Search (MCTS) algorithm. The results show that the RAT selection is possible with a low number of searches, and the allocation outperforms other popular methods. The main drawback is that high computational cost is required before the stabilization stage is reached. Then, in [180], the QoS problem is understood as guaranteeing a bounded blocking probability of the service. The authors define an adaptive hierarchical resource allocation algorithm that maximizes throughput based on Markov models. According to the results, the adaptive model performs better than static approaches, whereas several iterations and high computational costs are necessary to obtain the final configuration. Finally, in [181], QoE is used as the main optimization criterion for underlay communications. In order to solve the resource allocation issue, a weighted bipartite graph algorithm is used. The results show that while the minimum QoE is achieved, fairness and power allocation metrics can be improved. The work lacks a large-scale analysis since only picocell environments are taken into account.

Fairness. In addition to signal quality indicators, fairness is also an important metric in wireless networks. For example, in [182], the objective is to obtain a balanced fairness/throughput system and to do so, several weighted utility functions are used. The results indicate that positive tradeoffs can be obtained if the feedbacks from the users are considered. Undoubtedly, this condition increases the system overhead. Similar to [182], [183] shows a weighted utility function-based algorithm for fairness and throughput equilibrium. In this case, first, the available bandwidth is fairly divided, and then, the utility functions are applied to enhance the throughput value. The main benefits of this approach are adaptability to different traffic/user models and scalability since it can be used with bigger networks without worsening the results. However, the computational cost necessary to obtain the optimum configurations is quite high. Finally, in [76], a particular subgrouping approach is shown, where multicast subgrouping techniques are combined with NOMA multiplexing with flexible constraints. In this case, one of the proposed algorithms maximizes fairness and user throughput jointly with better performance than OMA-based algorithms. Nevertheless, as in the previous cases, a high computational cost has to be assumed.

Interference management. Another aspect that has to be considered when communication systems are built is interference management. It can refer to the interferences that a user/node generates inside its network (i.e., co-tier) or the interferences generated in the neighboring network (i.e., cross-tier). In this line, [184] proposes to manage interferences jointly with the resource allocation by defining a three-level architecture for different environments (i.e., macrocell, small cell, and Device-to-Device, D2D). The algorithm divides the bandwidth into sub-bands and follows throughput

and interference constraints to carry out the resource allocation. Despite the excellent interference avoidance results, the algorithm processing time (10 seconds) is too high for practical application. Following a similar philosophy, in [185], a fractional frequency reuse (FFR) method that divides the entire network into three sectors and three layers and the whole bandwidth into seven sub-bands is proposed. Then, the allocation of the subbands is carried out according to interference and throughput constraints in different size cells. Results show high performance value with assumable interference. Unfortunately, the computational cost required is considerably high due to the three-level sectoring. Another interference concerning approach can be found in [186] that proposes a three-layer heterogeneous network. In this case, the algorithm is divided into two stages: the subchannel allocation and power control, and the second for interference management. The work combines logarithmic functions with clustering techniques. Although the impact of the interferences is suppressed even in totally distributed environments, the energy consumption performance is worsened. The last work oriented to interference management is presented in [187], where an interference coordination mechanism is introduced in combination with energy-saving supervision. The authors use a Markov model algorithm based on network current and future load situation and QoS feedback. The results indicate that the impact of the interferences can be downgraded and simultaneously obtain QoS improvements.

Energy efficiency. The last optimization metric that is going to be considered is energy efficiency. Energy aspects may be very relevant in future wireless communication networks. On the one hand, in [188], a combination of Hungarian algorithms and multi-objective weighted methods is proposed to optimize the energy efficiency of D2D networks while minimum QoS values are guaranteed. The main conclusion obtained from this work is that although real blockage effects are considered, fairness among users is not considered. On the other hand, a different energy efficiency analysis is carried out in [189]. In this case, the optimization goal is to maximize the weighted sum energy efficiency and is performed following a first-order approximation iterative algorithm. The evaluation shows that the optimization metrics can be enhanced in less than five iterations. The principal drawback of this approach is that the allocation is not entirely fair. Finally, authors in [190] combine energy efficiency and spectral efficiency optimization with NOMA techniques. In this case, a controllable weight system is designed according to the instantaneous network conditions. It should be highlighted that these algorithms outperform OMA techniques in terms of energy efficiency and spectral efficiency with fast convergence. However, the implementation cost might be unaffordable due to the SIC module at the receivers.

TABLE 1.8: Summary of the analyzed works with different RRM optimization metrics

Ref	Primary metric(s)	Secondary metric(s)	Description	Weakness
[172]	Throughput	None	Two-step scheme for user association and resource allocation based on FLC	Network load balance is not considered
[173]		Interference management and transmitting power	Underlay communication system based on the Lagrange dual decomposition method	Co-tier interferences are not considered
[174]		Spectral efficiency and energy efficiency	Routing and resource allocation jointly optimized	Routing latency is not considered
[175]		None	Users are divided into subgroups and the subgroup configuration is optimized	High computational cost
[93]		None	NOMA techniques are used for the subgroup generation and optimization	High computational cost
[176]	Spectral efficiency	Interference management	Clustering-based algorithm following game theory	QoS of the users are not considered
[177]	Spectral efficiency and energy efficiency	Interference management	Two mechanisms are developed following interference alignment and traffic offloading	Cross-tier interference is not considered
[178]		Interference management and fairness	Double optimization based on a binary search algorithm	High overhead generated
[179]	QoS	None	RAT accessing and resource allocation technique based on a multiagent reinforcement learning technique	High computational cost
[180]		Throughput	Adaptive hierarchical resource allocation algorithm based on Markov models	High computational cost
[181]	QoE	Fairness and power allocation	Weighted bipartite graph algorithm for underlay communications	Only picocells are considered
[182]	Fairness and throughput	None	Weighted utility functions to score assumable tradeoffs	High system overhead
[183]		None	Combination of utility functions with previous fair bandwidth división	High computational cost
[76]		None	NOMA-based subgrouping algorithms for multicasting	High computational cost
[184]	Interference management and throughput	None	Bandwidth subdivision and constraint-based allocation	High processing time
[185]		None	FFR method that divides the network into three sectors and three layers and the whole bandwidth into seven subbands	High computational costs
[186]		Energy consumption	Two-stage algorithm with logarithmic functions and clustering techniques	High energy consumption
[187]	Interference management and QoS	Energy savings	Markov model algorithm based on network load and QoS feedback	High system overhead
[188]	Energy efficiency and QoS	None	Combined Hungarian algorithm and multi-objective weighted method for D2D networks	Fairness is not guaranteed
[189]	Energy efficiency	Spectral efficiency	First-order approximation iterative algorithm	Fairness is not guaranteed
[190]	Energy efficiency and spectral efficiency	QoS	Controllable weight system designed according to the instantaneous network conditions	High computational cost

To conclude, Table 1.8 shows a summary of the works presented in this section in the representation of the different optimization goals. In the table, the main characteristics of each reference are highlighted, and the most representative weaknesses.

1.2.2.4 Open issues

Although, as shown in the previous sections, much progress has been made in the area of efficient management of radioelectric resources and very promising proposals have been presented, there are still several areas in which the research community must be focused. Therefore, taking as a reference what was presented above and the analyzes carried out by other experts [26, 144, 146, 148, 191], this section presents the most relevant open research areas of RRM schemes. Figure 1.9 shows a summary of the lines to be highlighted organized by areas.

First, efficient interference management has to be emphasized. Taking into account that the trend of the future wireless networks will involve the implementation of tiny networks (i.e., picocells or femtocells) that have to

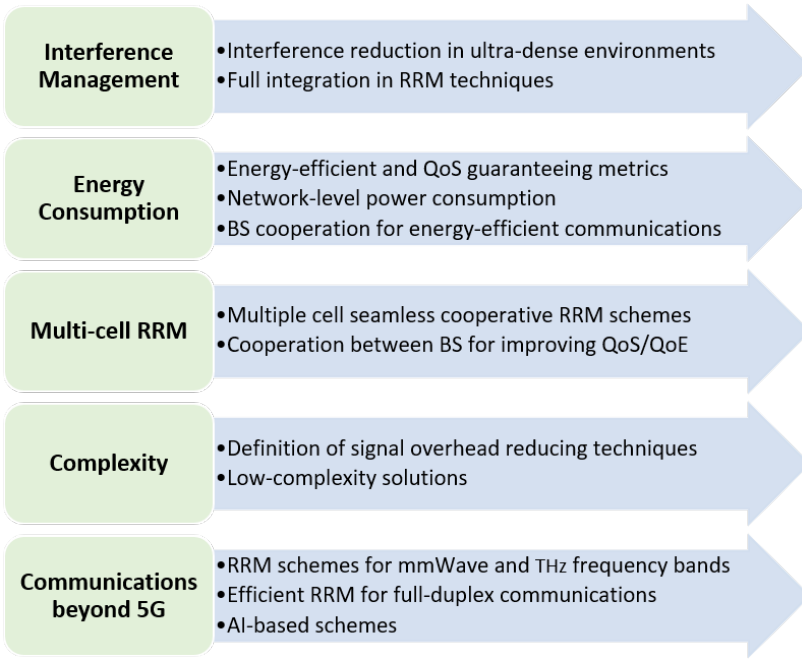


FIGURE 1.9: Future research lines of RRM schemes

assume a high density of users, the potentially dangerous interferences for receivers grows exponentially. Therefore, interference management systems compatible with ultra-dense environments have to be designed. Furthermore, as the services that the transmitter node of each network will have to manage will be very varied, the interference management must be fully integrated with the rest of the RRM schemes. In other words, for future RRM schemes, interference management will be as important as guaranteeing QoS or spectral efficiency.

Concerning the type of devices that are expected to be part of the 5G/6G ecosystem, it is widely assumed that battery devices such as mobile phones, tablets, sensors, or many IoT devices will be very relevant and, the energy-related parameters will be critical. Energy efficiency should be considered in combination with QoS guaranteeing schemes. Moreover, the power consumption of the overall network has to be reduced, and with this aim, several entities are pushing for integrating green technologies in the 5G ecosystem. Then, generally, energy efficiency is analyzed in a single cell, whereas multi-cell schemes are not considered. In particular, an interesting research line will be to design RRM schemes that enable cooperation among BS in order to increase energy efficiency.

In line with this last idea and concerning the relevance that heterogeneous networks will have soon, RRM schemes that consider multiple cells

simultaneously will be essential. In a heterogeneous network environment, the users will be linked to the same network and specific technology for a short time. Therefore, more research activities should be carried out to design new RRM schemes that seamlessly manage multiple cell resources. In fact, the cooperation between BS, even BS that deliver different contents or are based on different technologies (e.g., 5G and WiFi), should be supported as a mechanism to improve the QoS/QoE of the users by using efficient multi-cell RRM schemes.

Another problem that has to be covered in future research lines is complexity. Specifically, two main lines arise from the complexity challenge: system complexity and computational cost. In the first case, reducing signal overhead generated with the feedback from the users will be critical, especially in ultra-dense environments. Therefore, the definition of novel RRM techniques that imply low signal overhead will be interesting to facilitate future implementations. In addition, the reduction of the computational cost is closely related to the type of users that are going to be present in future wireless networks. Future networks will not tolerate high computational costs or will not be compatible with the latency increase derived from that computational effort. That is why further investigation in low-complexity solutions is required.

Finally, it should be mentioned that RRM schemes have to be compliant with the strict use cases that are being proposed for communications beyond 5G. For example, it is expected that 6G will be implemented in the mmWave bands and the THz frequency bands. Doubtlessly, THz communications have some inherent challenges that are not being considered in current RRM schemes, such as high losses, obstacle vulnerability, or high variability. Therefore, the latest RRM proposals have to be in line with the requirements of the highest frequency bands. Another relevant use case is full-duplex communications, which are expected to increase significantly spectral efficiency. However, RRM techniques for full-duplex communications have still not been deeply investigated. Finally, RRM schemes could be reinforced with additional techniques such as Artificial Intelligence (AI). Although some works have indeed proposed IA-based RRM schemes ([192, 193, 194, 195]), there is not an extensive analysis of different optimization goals or different use cases.

1.3 Hypothesis and objectives

1.3.1 Objectives

This thesis focuses on three of the four use cases: eMBB, URLLC, and communications beyond 5G. First of all, it should be noted that eMBB represents the use case for traditional entertainment and distribution services where the PHY of typical communication systems works very close to Shannon's limit. Therefore, the need to design and integrate new upper layer solutions poses an exciting challenge.

Then, URLLC, as it combines the most modern and robust communications with the most traditional industrial sector, implies the defiant challenge of replacing typically wired communications with wireless systems without deteriorating the performance. In addition, URLLC is of specific interest for the geographical area where the thesis is presented, since the Basque Country has historically stood out as an important industrial activity area. Consequently, the analysis can have a relevant impact on surrounding companies. Finally, communications beyond 5G involves working at the cutting edge of research, which is a significant challenge and a personal incentive for motivation when developing the thesis.

Once defined the use cases to be addressed in the thesis, the objectives for each one are defined as follows:

- **Objective 1**: Propose, design, and evaluate new PHY/MAC structures and RRM algorithms based on NOMA to offer broadcast services in eMBB environments to increase the capacity and quality of the services, as well as the number of users served.

- **Objective 2**: Propose, design, and evaluate new PHY/MAC structures and RRM algorithms based on NOMA to improve the performance of critical communications in terms of reliability in URLLC environments without compromising latency.

- **Objective 3**: Propose, design, and evaluate the set of necessary techniques and technologies, as well as the architecture to implement full-duplex communications in traditional broadcasting system environments to incorporate in the same frequency band new services such as inter-tower communications or datacasting.

However, in order to meet each of the objectives mentioned above, it is necessary to face and solve a series of technological challenges. Figure 1.10 shows the summary of challenges associated with each of the use cases integrated into this thesis.

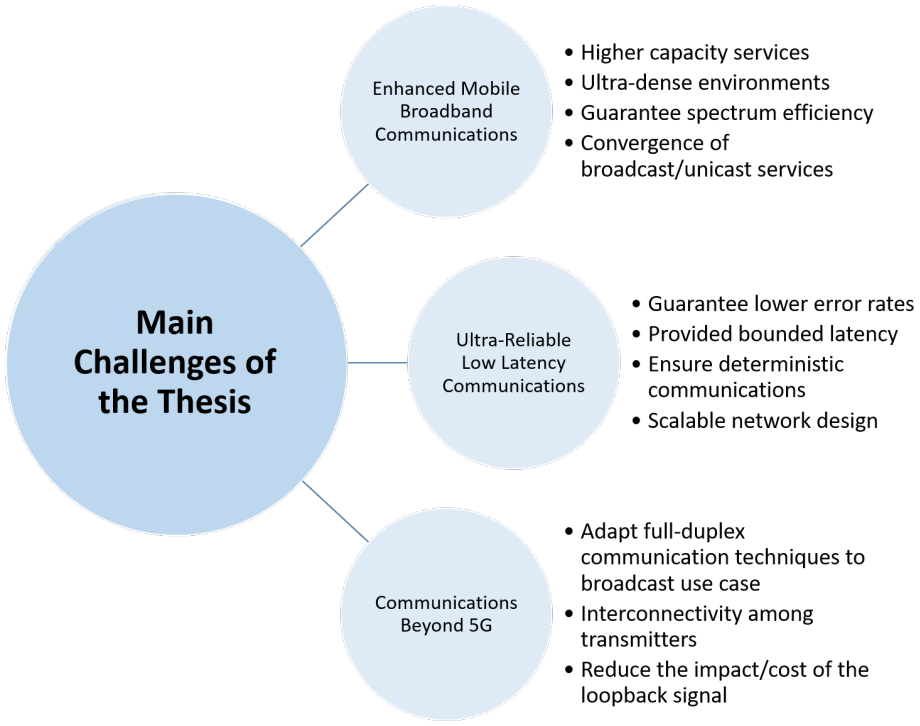


FIGURE 1.10: Main challenges associated with the objectives of the thesis.

On the one hand, the communication systems designed for eMBB environments must improve the spectral efficiency of current communication standards. Due to the latest advances in coding and modulation, current communication systems are very close to Shannon's limit. In addition, they must be relevant for challenging environments where the density of users is very high, which consequently implies that the capacity offered has to be higher. Lastly, due to the characteristics of the services requested by users, broadcast communications cannot wholly replace unicast communications, so the convergence between broadcast and unicast services must be guaranteed in the most efficient possible way.

On the other hand, contributions to URLLC environments face completely different challenges than eMBB. For example, communications in critical environments have to be deterministic, so the proposed MAC structures must guarantee that data packets are received deterministically, as well as they have bounded latency values (i.e., between 1-10 ms, depending on the use case) [196]. Moreover, one of the main KPIs used by the industry is the lost packet ratio, which currently stands at around 10^{-6}-10^{-9}

depending on the application. Therefore, the proposed systems must guarantee error rates within this range to be applicable in the industry. However, most error protection and lost data recovery techniques involve increased latency, which is another critical parameter. So the proposed techniques must guarantee that the maximum latency is limited and within the requirements set by the use case. Finally, considering that URLLC is an environment in constant development, the proposed designs must be scalable.

Finally, designing new techniques that enable full-duplex communications in broadcast environments poses new challenges for future 6G communications. First of all, it should be noted that although full-duplex communications have already been investigated for other applications, such as Wi-Fi or 4G, it is a novelty for 5G/6G broadcast applications. Therefore, the first challenge is to select from the techniques already used for other applications that are relevant for broadcast environments. If necessary, the required modifications to fit with the new use case have to be designed and applied. Afterward, it must be guaranteed that the effect of the loopback signal can be correctly canceled. Therefore, it will be necessary to define the signal cancellation architecture. Finally, the definition of the communication scheme necessary to interconnect the transmission centers is another challenge of the project.

1.3.2 Contributions

The main research area of this thesis is founded on the design and evaluation of new applications of power domain NOMA for the networks that form the 5G ecosystem and beyond. As shown in Figure 1.11, three are the principal research areas that have been addressed in this thesis: enhanced Mobile Broadband Communications, Ultra-Reliable Low Latency Communications, and communications beyond 5G.

The following paragraphs describe the characteristics of the contribution associated with the research challenges of each objective (see Figure 1.11).

Contribution 1: NOMA for Broadcasting in eMBB

eMBB is a 5G use case that combines traditional infotainment applications with innovative multimedia use cases such as AR or 3D video streaming. The new multimedia applications require an increase in the offered capacity, and in order to satisfy the users' requests, NOMA is proposed as a broadcast/multicast enabler. Contribution 1 is oriented to fulfill Objective 1 (see 1.3).

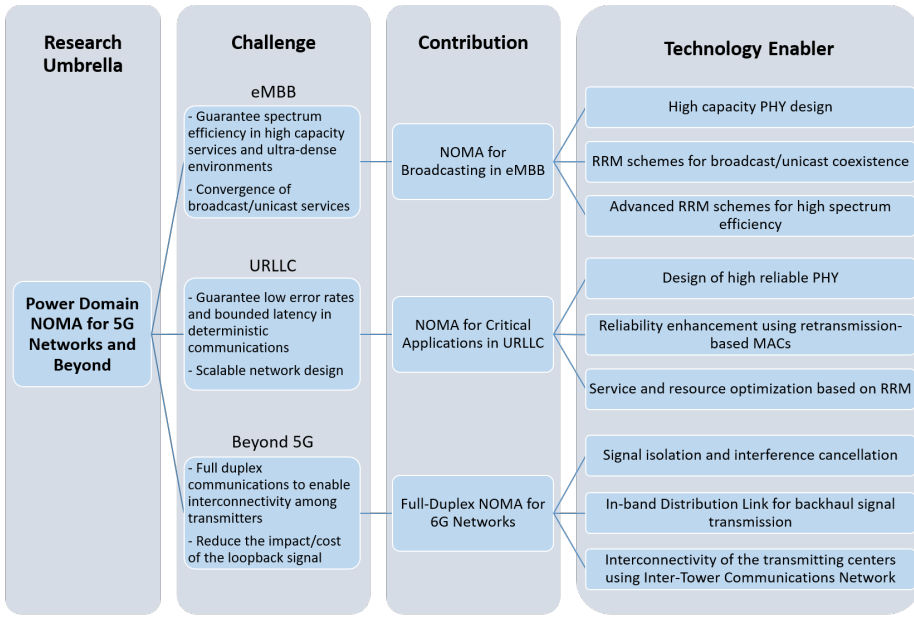


FIGURE 1.11: Graphical representation of the contributions of this thesis.

This thesis targets the design of a NOMA-based 5G PHY layer that provides higher capacity rates than traditional TDMA systems. At the same time, the coexistence of broadcast and unicast services is guaranteed by an efficient multiplexing scheme design. Moreover, to increase the network spectral efficiency, RRM schemes are also considered. In this aspect, different NOMA-based RRM schemes are designed and evaluated. Finally, future use cases are addressed, such as multicast subgrouping techniques and high-density networks in mmWave bands.

Results indicate that introducing NOMA in the current 5G standard provides high spectrum efficiency, which can increase the transmitted capacity, increase the number of users served in the network or provide more robust services.

The publications related to Contribution 1 are:

– E. Iradier, J. Montalban, D. Romero, Y. Wu, L. Zhang and W. Li, "NOMA based 5G NR for PTM Communications," 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Jeju, Korea (South), 2019, pp. 1-6, doi: 10.1109/BMSB47279.2019.8971856.

– E. Iradier et al., "Using NOMA for Enabling Broadcast/Unicast Convergence in 5G Networks," in IEEE Transactions on Broadcasting, vol.

66, no. 2, pp. 503-514, June 2020, doi: 10.1109/TBC.2020.2981759.

- E. Iradier et al., "Broadcast/Unicast Convergence in NOMA-based 5G: a RRM Optimization Algorithm," 2020 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Paris, France, 2020, pp. 1-6, doi: 10.1109/BMSB49480.2020.9379805.
- E. Iradier et al., "Non-Orthogonal Multiple Access and Subgrouping for Improved Resource Allocation in Multicast 5G NR," submitted to Digital Communications and Networks.
- E. Iradier et al., "Advanced NOMA-based RRM Schemes for 5G mmWave Bands," submitted to IEEE Transactions on Broadcasting.

Contribution 2: NOMA for Critical Applications in URLLC

The main characteristics of URLLC are extremely low packet error rate, low and bounded latency values, and deterministic communications. In this case, the spectral efficiency gain that NOMA systems provide could be used combined with current URLLC applications to get the strict requirements closer than ever. Contribution 2 is oriented to fulfill Objective 2 (see 1.3).

On the one hand, this thesis targets the design of particular PHYs that include NOMA in the coding and decoding stage and, on the other hand, the evaluation of retransmission-based MAC layers to increase the overall reliability. In addition, the correct and efficient use of the spectrum resources is also essential in URLLC environments. So, this thesis analyzes different RRM algorithms to enhance the use of radio resources.

Results indicate that using NOMA in URLLC systems improves the reliability of the communications. If combined with different retransmission schemes, the total gain implies reducing the packet error rates several orders of magnitude.

Publications related to Contribution 2 are:

- E. Iradier, J. Montalban, L. Fanari, P. Angueira, O. Seijo and I. Val, "NOMA-based 802.11n for Broadcasting Multimedia Content in Factory Automation Environments," 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Jeju, Korea (South), 2019, pp. 1-6, doi: 10.1109/BMSB47279.2019.8971844.

- E. Iradier, J. Montalban, L. Fanari and P. Angueira, "On the Use of Spatial Diversity under Highly Challenging Channels for Ultra Reliable Communications," 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Zaragoza, Spain, 2019, pp. 200-207, doi: 10.1109/ETFA.2019.8869055.
- E. Iradier et al., "Analysis of NOMA-Based Retransmission Schemes for Factory Automation Applications," in IEEE Access, vol. 9, pp. 29541-29554, 2021, doi: 10.1109/ACCESS.2021.3059069.
- J. Montalban, E. Iradier, P. Angueira, O. Seijo and I. Val, "NOMA-Based 802.11n for Industrial Automation," in IEEE Access, vol. 8, pp. 168546-168557, 2020, doi: 10.1109/ACCESS.2020.3023275.
- E. Iradier et al., "Throughput, Capacity and Latency Analysis of P-NOMA RRM Schemes in 5G URLLC," accepted for publication in Multimedia Tools and Applications.
- E. Iradier and J. Montalban, "Komunikazioak 4.0 Industrian," accepted for publication in Elhuyar Aldizkaria.

Contribution 3: Full Duplex NOMA for 6G Networks

Although 5G has meant an important change in current communication systems, a series of requirements have not been met, such as latency below one ms or improvement in spectral efficiency. In fact, to improve on this last disadvantage, full-duplex communications based on NOMA could provide the necessary improvement in spectral efficiency to implement service beyond 5G. Contribution 3 is oriented to fulfill Objective 3 (see 1.3).

In particular, this thesis analyzes the design and challenges associated with a full-duplex architecture that uses NOMA to combine broadcast content service with backhaul data in the same frequency band. Furthermore, starting from the base application, the architecture of an interconnection system of transmitting stations has been defined using this same technology.

The results show the complete design of a functional architecture and the conditions required for the system to work. In addition, the main challenges associated with implementation (i.e., signal isolation and interference cancellation) have been identified, and the necessary stages to overcome them have been defined.

Publications related to Contribution 3 are:

- L. Zhang et al., "ATSC 3.0 In-band Backhaul for SFN Using LDM with Full Backward Compatibility," 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Jeju, Korea (South), 2019, pp. 1-6, doi: 10.1109/BMSB47279.2019.8971918.
- L. Zhang et al., "Using Layered Division Multiplexing for Wireless In-Band Distribution Links in Next Generation Broadcast Systems," in IEEE Transactions on Broadcasting, vol. 67, no. 1, pp. 68-82, March 2021, doi: 10.1109/TBC.2020.2989638.
- W. Li et al., "Integrated Inter-Tower Wireless Communications Network for Terrestrial Broadcasting and Multicasting Systems," accepted for publication in IEEE Transactions on Broadcasting.

1.4 Summary and results

1.4.1 NOMA for broadcasting in eMBB

Enhanced Mobile Broadband (eMBB) is one of the 5G verticals where traditional infotainment services are combined with innovative multimedia applications. In fact, the characteristics of multimedia applications and the needs of the users have changed radically in recent years. Some clear examples of this change are 3D or 360° video content, 4k/8k quality of service, applications based on AR/VR, or interaction with the cloud. In addition to the list of applications, eMBB has to face another technological challenge: the service has to be guaranteed for everyone and everywhere. Consequently, this includes events and locations with a high density of users such as stadiums, malls, exhibitions, festivals, or concerts. Due to those reasons, the network must be prepared to offer capacities in the range of gigabits per second.

New techniques and technologies, such as NOMA, should be investigated to enhance the overall network spectrum efficiency. However, in order to provide a relevant improvement, a multi-level analysis (i.e., PHY, MAC, and RRM) is mandatory. On the one hand, a PHY/MAC level analysis has to be carried out to evaluate the robustness and capacity of the proposed solutions. On the other hand, the efficiency of the RRM module has to be evaluated to optimize the resource allocation phase. In this contribution, both analyses are carried out.

Concerning the PHY/MAC level analysis, firstly, in paper C1, the first NOMA-based 5G NR transceiver is designed to enable the broadcast content transmission capabilities. The work presents a PHY level architecture based on the 5G NR architecture that includes some modifications in order to include the NOMA coding and decoding processes. The evaluation is

performed over a set of designed use cases related to multimedia content distribution. The results indicate that NOMA achieves the same capacity as traditional TDMA schemes. At the same time, the robustness of the services is around 4 dB higher, representing a significant improvement in service availability.

An additional advance in system design is presented in paper J1. This paper presents a refined NOMA-based 5G NR PHY/MAC architecture that includes HARQ retransmission techniques. Also, the architecture evaluation under more challenging propagation channels (i.e., TDL). The objective of the defined architecture is covering the need to guarantee the convergence between broadcast and unicast services. The use case under study is based on a cellular network where broadcast and unicast services coexist, and the gNB must guarantee the distribution of both contents. The evaluation is carried out using three different KPIs: reliability, latency, and throughput. In general, the results show that NOMA outperforms TDMA in reliability and throughput. In addition, the latency values are similar and acceptable (i.e., below 4ms) in both cases (i.e., NOMA and TDMA). Furthermore, the LL of NOMA has an extra latency component due to the Successive Interference Cancellation (SIC) module, which is taken into account by latency calculations.

Regarding the RRM level analysis, the paper C2 explores resource allocation capabilities in broadcast/unicast convergence environments using TDMA and NOMA. This is achieved with two new RRM algorithms and described to manage resource allocation, one based on NOMA and the other based on TDMA. Based on the results, it should be highlighted that NOMA performs better than TDMA in terms of aggregate unicast throughput and the number of served unicast users. Moreover, the complexity of the RRM algorithms is also evaluated, and NOMA is considered more complex since it requires a higher number of iterations. Aiming at reducing the complexity of the NOMA algorithm, a new approach is proposed based on the algorithm execution frequency. The results obtained with the complexity reducing techniques are presented and organized according to the ranges where NOMA performs better than TDMA. Then, two more contributions are presented, representing two specific and relevant use cases for future wireless communications.

First, in paper J2, a particular application of multicast transmissions is presented: RRM techniques in subgrouping algorithms. The paper describes the better use of the resources by splitting users into subgroups and applying independent and adaptive modulation and coding schemes according to their CQI feedback. The main implications of the combination of subgrouping techniques with NOMA are introduced, and two subgrouping algorithms are designed, one based on NOMA and the other based on

TDMA. Results indicate that if no constraint is applied, both NOMA and TDMA present unfair configurations from the number of users served and the minimum capacity point of view. Then, once constraints are applied, NOMA should be considered a better candidate since it offers better capacity results than TDMA. However, based on a theoretical analysis, it is foreseen that the joint implementation of NOMA and TDMA techniques could provide more efficient algorithms.

Finally, taking as a reference the last conclusion of the previous contribution, in paper J3 the combination of NOMA and TDMA for advanced RRM schemes is analyzed in-depth. This study is defined in another use case that is expected to be relevant for implementing future services: communications in the mmWave bands. To test the different proposed RRM models, a use case based on the on-demand distribution of multimedia content in high-density environments has been used. The evaluation of the models is carried out in terms of throughput, capacity, and availability. The results show that the combined RRM techniques improve the performance of the single technique RRM algorithms by around 50%, and the best performing RRM model out of the presented is indicated.

1.4.2 NOMA for critical applications in URLLC

URLLC gathers a large number of applications with characteristics very different from eMBB applications. In fact, URLLC stands out for having demanding reliability and latency requirements, as specified in Section 1.1. Industry 4.0 [197] wireless links are One of the most representative cases of URLLC, where transmission happens in harsh propagation conditions of industrial environments. The applications concerning Industry 4.0 are related to automated production chains, where machines are fundamental elements of the process. Some representative examples would be assembly, packaging, palletizing, and manufacturing.

Although the ecosystem of applications and technologies represented by 5G is a strong candidate to lead wireless links of URLLC applications, at this moment, the technology on which most industrial systems are based is IEEE 802.11. Even if the 802.11 family of wireless standards was not designed for industrial communications, they have also emerged as candidates for industrial applications because of their wide range of available devices, network deployments, and applications in practically all communication fields. However, the reliability of the 802.11 standards is still limited, and they lack determinism. These two drawbacks have prevented a massive implementation of WIFI in industry. For this reason, industrial communications need solutions that overcome the gap between the industrial requirements and the reliability of the 802.11 standard.

New techniques and technologies, such as NOMA, should be considered to provide higher robustness to industrial communications. Following the approach described in the previous contribution, a two-level analysis is required to guarantee that the proposals and their corresponding evaluations are representative. First, it is necessary to design and test PHY layers that guarantee a minimum acceptable level of reliability and combine them with MAC layers to improve the overall link quality with mechanisms such as retransmissions. Furthermore, secondly, it is necessary to study the possible improvement in spectral efficiency that can be obtained at the RRM level. This contribution presents both types of analyses gathered in six different papers presented during the development of this PhD Thesis.

Concerning the area of PHY/MAC level analysis, paper C3 presents a preliminary evaluation of the performance obtained by combining NOMA with the 802.11n standard. The analysis uses an industrial wireless communication network with variable configuration parameters. The results indicate that NOMA offers better results in critical services and that this improvement increases as the the imbalance between critical and non-critical services gets higher.

Subsequently, the paper J4 proposes a complete PHY level transceiver architecture combined with a superframe scheme that manages medium access and time domain retransmissions. The communication system proposed aims at delivering two different sets of services. The first service class comprises Critical Services (CS) with strict restrictions in both reliability and latency. The same communication system should also convey a second group of services, referred to as Best Effort (BE), with more relaxed requirements. In general, NOMA shows better levels of performance than TDMA, especially for the CS. Regarding latency, both NOMA and TDMA show similar results. However, it should be noted that the deviation from mean latency is lower in NOMA, indicating that communications are more deterministic in this case.

Then, the paper C4 evaluates the impact of spatial diversity-based retransmission schemes. In particular, multiple transmitter networks are tested, where the initial transmission and the retransmissions are carried out by different Access Points (APs) with different physical locations. Results conclude that although a high performance increase is obtained with multiple transmitters, a tradeoff has to be assumed between reliability and offered capacity. The tradeoff is necessary because too many retransmissions would diminish the capacity offered.

In the paper J5, time division-based and multiple transmitter-based retransmission schemes are combined with the NOMA-based 802.11 transceiver presented in J4. In this case, different superframe structures are

proposed where both retransmission schemes are introduced with multiple configuration choices: MCS, number of retransmissions, and IL. The results are analyzed in terms of scalability, reliability, and latency. This architecture provides a considerable reliability improvement for both layers: UL and LL. In addition, low latency values are achieved (the UL does not exceed 250 μs) and with acceptable jitter values (especially from 10 dB).

On the other hand, concerning the RRM level analysis, the main contribution of this thesis is the paper J6. This work proposes to develop specific RRM modules based on NOMA and TDMA for 5G URLLC in industrial environments. A typical industrial use case is defined to test the RRM modules. The study analyzes different configurations of several parameters and their suitability to measure the adaptability of the RRM modules to the network configuration. Both alternatives are deeply analyzed from different perspectives (throughput, capacity, and latency) and optimized to meet different requirements (maximum capacity/throughput and minimum cycle time). In general, results suggest that NOMA techniques are the most suitable option when the optimization goal is to maximize the network capacity or the amount of served unicast users.

Finally, taking as a reference everything learned during the preparation of this contribution, the paper J7 shows a scientific dissemination article. This contribution presents a summary of the current state of communications within Industry 4.0 and a forecast of what industrial communications will look like in the future. It should be noted that this contribution has been published in a scientific dissemination magazine in the Basque Country (Elhuyar Aldizkaria), so the local language has been used for its writing: Basque.

1.4.3 Full duplex NOMA for 6G networks

Despite the constant effort of different wireless communication systems standardization organizations to improve the quality of service, the reality is that the applications and systems requested by the users are more demanding than what current technology can support. A clear example is 5G. Although the reliability of communications has been considerably improved, and a configurable PHY is offered that can be adapted to different scenarios, there are already applications with requirements exceeding the performance of 5G. For this reason, the research community must begin to orient its research towards complementary technologies for future communication standards (i.e., 6G) that match society's demands. For example, in 6G, efforts should focus on zero-latency systems or increasing spectral efficiency.

In particular, a solution that would considerably increase the spectral efficiency of future communication standards is the use of NOMA techniques to achieve in-band full-duplex systems. This challenge requires first the definition of practical use cases that would benefit from a NOMA full-duplex system. In addition, the base technology to test the system should be decided. Moreover, the study requires a transmitter and receiver architecture definition capable of correctly generating and decoding in-band signals based on NOMA. An essential task involves identifying the associated problems and system design challenges of a NOMA full-duplex architecture given the current state-of-the-technology and pointing out possible solutions. Specifically, this contribution covers some of these research challenges.

First, the technology and use case issues are covered in the paper C5. A complete in-band backhaul approach that provides backhaul for both robust mobile and high-data-rate fixed services within the same frequency band as the traditional broadcast content is proposed. This novel solution uses LDM as a particular case of NOMA and a fully backward compatible ATSC 3.0 waveform. The main drawback of the solution is the self-interference signal that generates the transmitter. That is why a preliminary digital cancellation module is introduced to guarantee loopback cancellation values ranging from 30 to 60 dB in a short Typical Urban (TU) channel. The cancellation variation depends on the power difference between the desired signal and the self-interference.

Then, in the paper J8, an extension of the previous work is presented, where a wireless in-band distribution link (WIDL) technology is defined in detail. Firstly, the potential capacity gain is analyzed using next-generation digital TV (DTV) services in a throughput analysis that highlights the benefits of using LDM as the multiplexing technology. Then, the receiver block diagram is defined, and the performance of the cancellation module is analyzed for mobile channels. The results show that the self-interference in LOS conditions can be canceled up to 10 dB more than in other NLOS cases. Finally, the results of a measurement campaign in Ottawa (Canada) are shown in order to determine the signal isolation that could be obtained in real transmitter stations.

Finally, a third contribution is presented in the paper J9, where an evolution of the WIDL system is proposed. In this case, a bi-directional integrated inter-tower communications network (ITCN) is described. The main goal is to provide full-duplex transmission among SFN transmitters to offer enriched data services including IoT, emergency warning, connected car, and other localized data services. The network architecture is presented as well as the possible variations due to its flexibility. Then, a complete system block diagram of an ITCN network node is defined. The

introduction of new modules in the reception chain to improve the cancellation of the self-interference signal stands out. Specifically, the cancellation chain is based on the combination of two different techniques. The first is adaptive antenna techniques, which can maximize the main beam gain in the direction of the remote desired ITCN signal. The second is analog and digital cancellation, which is divided into two stages in order to overcome the limitations of the dynamic range.

1.5 Bibliography

[1] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021*.

[2] Patrik Cerwall et al. "Ericsson mobility report". In: *On the Pulse of the Networked Society. Hg. v. Ericsson* (2015).

[3] Zoran S Bojkovic, Dragorad A Milovanovic, and Tulsi Pawan Fowdur. *5G Multimedia Communication: Technology, Multiservices, and Deployment*. CRC Press, 2020.

[4] *ITU-R, Report ITU-R M.2370-0, "IMT traffic estimates for the years 2020 to 2030," Jul. 2015*.

[5] Mohammed S Elbamby et al. "Toward low-latency and ultra-reliable virtual reality". In: *IEEE Network* 32.2 (2018), pp. 78–84.

[6] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017-2022*.

[7] *ITU-R Rec. ITU-R M. 2083-0, "IMT Vision — Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond," Sept. 2015*.

[8] *ITU-R Rec. ITU-R M. 2410-0, "Minimum requirements related to technical performance for IMT-2020 radio interface(s)," Nov. 2017*.

[9] Mansoor Shafi et al. "5G: A tutorial overview of standards, trials, challenges, deployment, and practice". In: *IEEE journal on selected areas in communications* 35.6 (2017), pp. 1201–1221.

[10] *TS 38.201, Tech. Spec. Group Services and System Aspects, "NR; Physical layer; General description," V15.0.0, January 2018*.

[11] *TR 21.916, Tech. Spec. Group Services and System Aspects, "Release 16 Description; Summary of Rel-16 Work Items (Release 16)," V1.0.0, December 2018*.

[12] Amitabha Gosh et al. "5G Evolution: View on 5G Cellular Technology Beyond 3GPP Release 15". In: *IEEE Access* 7 (2019), pp. 127639–127651.

[13] Raúl Chávez-Santiago et al. "5G: The convergence of wireless communications". In: *Wireless Personal Communications* 83.3 (2015), pp. 1617–1642.

[14] Luca Boero et al. "Satellite networking integration in the 5G ecosystem: Research trends and open challenges". In: *Ieee Network* 32.5 (2018), pp. 9–15.

[15] Jon Montalban, Gabriel-Miro Muntean, and Pablo Angueira. "A utility-based framework for performance and energy-aware convergence in 5G heterogeneous network environments". In: *IEEE Transactions on Broadcasting* 66.2 (2020), pp. 589–599.

[16] Massimo Condoluci et al. "Fixed-mobile convergence in the 5G Era: from hybrid access to converged core". In: *IEEE Network* 33.2 (2019), pp. 138–145.

[17] Shanzhi Chen et al. "Vision, requirements, and technology trend of 6G: how to tackle the challenges of system coverage, capacity, user data-rate and movement speed". In: *IEEE Wireless Communications* 27.2 (2020), pp. 218–228.

[18] Wei Feng et al. "When mmWave communications meet network densification: A scalable interference coordination perspective". In: *IEEE Journal on Selected Areas in Communications* 35.7 (2017), pp. 1459–1471.

[19] Long Zhang et al. "A survey on 5G millimeter wave communications for UAV-assisted wireless networks". In: *IEEE Access* 7 (2019), pp. 117460–117504.

[20] Behrooz Makki et al. "A survey of NOMA: Current status and open research challenges". In: *IEEE Open Journal of the Communications Society* 1 (2020), pp. 179–189.

[21] SM Riazul Islam et al. "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges". In: *IEEE Communications Surveys & Tutorials* 19.2 (2016), pp. 721–742.

[22] Nabil Svenh Loghin et al. "Non-uniform constellations for ATSC 3.0". In: *IEEE Transactions on Broadcasting* 62.1 (2016), pp. 197–203.

[23] Evgeny Khorov et al. "A tutorial on IEEE 802.11 ax high efficiency WLANs". In: *IEEE Communications Surveys & Tutorials* 21.1 (2018), pp. 197–216.

[24] Hsin-Hung Cho et al. "Integration of SDR and SDN for 5G". In: *Ieee Access* 2 (2014), pp. 1196–1204.

[25] Alcardo Alex Barakabitze et al. “5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges”. In: *Computer Networks* 167 (2020), p. 106984.

[26] Sulastri Manap et al. “Survey of radio resource management in 5g heterogeneous networks”. In: *IEEE Access* 8 (2020), pp. 131202–131223.

[27] Nadezhda Chukhno et al. “Efficient Management of Multicast Traffic in Directional mmWave Networks”. In: *IEEE Transactions on Broadcasting* (2021).

[28] Yi Fang et al. “Design guidelines of low-density parity-check codes for magnetic recording systems”. In: *IEEE Communications Surveys & Tutorials* 20.2 (2018), pp. 1574–1606.

[29] Hong Chen, Robert G Maunder, and Lajos Hanzo. “A survey and tutorial on low-complexity turbo coding techniques and a holistic hybrid ARQ design example”. In: *IEEE Communications Surveys & Tutorials* 15.4 (2013), pp. 1546–1566.

[30] Zunaira Babar et al. “Polar codes and their quantum-domain counterparts”. In: *IEEE Communications Surveys & Tutorials* 22.1 (2019), pp. 123–155.

[31] Mahmoud A Albreem, Markku Juntti, and Shahriar Shahabuddin. “Massive MIMO detection techniques: A survey”. In: *IEEE Communications Surveys & Tutorials* 21.4 (2019), pp. 3109–3132.

[32] Emil Björnson and Luca Sanguinetti. “Scalable cell-free massive MIMO systems”. In: *IEEE Transactions on Communications* 68.7 (2020), pp. 4247–4261.

[33] Min Jia et al. “Broadband hybrid satellite-terrestrial communication systems based on cognitive radio toward 5G”. In: *IEEE Wireless Communications* 23.6 (2016), pp. 96–106.

[34] Xuemin Hong et al. “Cognitive radio in 5G: a perspective on energy-spectral efficiency trade-off”. In: *IEEE Communications Magazine* 52.7 (2014), pp. 46–53.

[35] Emilio Calvanese Strinati et al. “6G: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication”. In: *IEEE Vehicular Technology Magazine* 14.3 (2019), pp. 42–50.

[36] Syed Junaid Nawaz et al. “Quantum machine learning for 6G communication networks: State-of-the-art and vision for the future”. In: *IEEE Access* 7 (2019), pp. 46317–46350.

[37] Lin Zhang, Ying-Chang Liang, and Dusit Niyato. "6G Visions: Mobile ultra-broadband, super internet-of-things, and artificial intelligence". In: *China Communications* 16.8 (2019), pp. 1–14.

[38] Walid Saad, Mehdi Bennis, and Mingzhe Chen. "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems". In: *IEEE network* 34.3 (2019), pp. 134–142.

[39] Harish Viswanathan and Preben E Mogensen. "Communications in the 6G era". In: *IEEE Access* 8 (2020), pp. 57063–57074.

[40] Kazi Mohammed Saidul Huq et al. "THz communications for mobile heterogeneous networks". In: *IEEE Communications Magazine* 56.6 (2018), pp. 94–95.

[41] Nan Chi et al. "Visible Light Communication in 6G: Advances, Challenges, and Prospects". In: *IEEE Vehicular Technology Magazine* 15.4 (2020), pp. 93–102.

[42] Bomin Mao, Yuichi Kawamoto, and Nei Kato. "AI-based joint optimization of QoS and security for 6G energy harvesting internet of things". In: *IEEE Internet of Things Journal* 7.8 (2020), pp. 7032–7042.

[43] Sanjay Goyal et al. "Full duplex cellular systems: will doubling interference prevent doubling capacity?" In: *IEEE Communications Magazine* 53.5 (2015), pp. 121–127.

[44] Carlos De Lima et al. "Convergent communication, sensing and localization in 6G systems: An overview of technologies, opportunities and challenges". In: *IEEE Access* (2021).

[45] Emilio Calvanese Strinati et al. "6G in the sky: On-demand intelligence at the edge of 3D networks". In: *ETRI Journal* 42.5 (2020), pp. 643–657.

[46] Sooeun Song et al. "Analysis of wireless backhaul networks based on aerial platform technology for 6g systems". In: *Computers, Materials and Continua* 62.2 (2020), pp. 473–494.

[47] Khaled B Letaief et al. "The roadmap to 6G: AI empowered wireless networks". In: *IEEE Communications Magazine* 57.8 (2019), pp. 84–90.

[48] A. W. Scott and R. Frobenius. "Multiple Access Techniques: FDMA, TDMA, AND CDMA". In: *RF Measurements for Cellular Phones and Wireless Data Systems*. 2008, pp. 413–429.

[49] B Sundar Rajan and Moon Ho Lee. "Quasi-cyclic dyadic codes in the Walsh-Hadamard transform domain". In: *IEEE Transactions on Information Theory* 48.8 (2002), pp. 2406–2412.

[50] Hongxiang Li et al. “OFDMA capacity analysis in MIMO channels”. In: *IEEE transactions on information theory* 56.9 (2010), pp. 4438–4446.

[51] Linglong Dai et al. “A survey of non-orthogonal multiple access for 5G”. In: *IEEE communications surveys & tutorials* 20.3 (2018), pp. 2294–2323.

[52] Kenichi Higuchi and Anass Benjebbour. “Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access”. In: *IEICE Transactions on Communications* 98.3 (2015), pp. 403–414.

[53] Lei Wang et al. “Sparse code multiple access-towards massive connectivity and low latency 5G communications”. In: *Telecommun. Netw. Technol* 5.5 (2015).

[54] Alireza Bayesteh et al. “Blind detection of SCMA for uplink grant-free multiple-access”. In: *2014 11th international symposium on wireless communications systems (ISWCS)*. IEEE. 2014, pp. 853–857.

[55] Mojtaba Vaezi et al. “Interplay between NOMA and other emerging technologies: A survey”. In: *IEEE Transactions on Cognitive Communications and Networking* 5.4 (2019), pp. 900–919.

[56] Omar Maraqa et al. “A survey of rate-optimal power domain NOMA with enabling technologies of future wireless networks”. In: *IEEE Communications Surveys & Tutorials* 22.4 (2020), pp. 2192–2235.

[57] Zhiguo Ding et al. “A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends”. In: *IEEE Journal on Selected Areas in Communications* 35.10 (2017), pp. 2181–2195.

[58] Reza Hoshyar, Ferry P Wathan, and Rahim Tafazolli. “Novel low-density signature for synchronous CDMA systems over AWGN channel”. In: *IEEE Transactions on Signal Processing* 56.4 (2008), pp. 1616–1626.

[59] Dongning Guo and Chih-Chun Wang. “Multiuser detection of sparsely spread CDMA”. In: *IEEE journal on selected areas in communications* 26.3 (2008), pp. 421–431.

[60] Reza Hoshyar, Razieh Razavi, and Mohammad Al-Imari. “LDS-OFDM an efficient multiple access technique”. In: *2010 IEEE 71st Vehicular Technology Conference*. IEEE. 2010, pp. 1–5.

[61] Mohammed Al-Imari and Reza Hoshyar. "Reducing the peak to average power ratio of LDS-OFDM signals". In: *2010 7th International Symposium on Wireless Communication Systems*. IEEE. 2010, pp. 922–926.

[62] Mohammed Al-Imari et al. "Subcarrier and power allocation for LDS-OFDM system". In: *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*. IEEE. 2011, pp. 1–5.

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

[64] Mahmoud Taherzadeh et al. "SCMA codebook design". In: *2014 IEEE 80th Vehicular Technology Conference (VTC2014-Fall)*. IEEE. 2014, pp. 1–5.

[65] Hosein Nikopour et al. "SCMA for downlink multiple access of 5G wireless networks". In: *2014 IEEE global communications conference*. IEEE. 2014, pp. 3940–3945.

[66] Kelvin Au et al. "Uplink contention based SCMA for 5G radio access". In: *2014 IEEE Globecom workshops (GC wkshps)*. IEEE. 2014, pp. 900–905.

[67] Thomas Cover. "Broadcast channels". In: *IEEE Transactions on Information Theory* 18.1 (1972), pp. 2–14.

[68] P Bergmans and Thomas Cover. "Cooperative broadcasting". In: *IEEE Transactions on Information Theory* 20.3 (1974), pp. 317–324.

[69] Linglong Dai et al. "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends". In: *IEEE Communications Magazine* 53.9 (2015), pp. 74–81.

[70] Xiaoming Dai et al. "Pattern division multiple access: A new multiple access technology for 5G". In: *IEEE Wireless Communications* 25.2 (2018), pp. 54–60.

[71] Jiachen Huang et al. "Scalable video broadcasting using bit division multiplexing". In: *IEEE Transactions on Broadcasting* 60.4 (2014), pp. 701–706.

[72] Katsutoshi Kusume, Gerhard Bauch, and Wolfgang Utschick. "IDMA vs. CDMA: Analysis and comparison of two multiple access schemes". In: *IEEE Transactions on wireless communications* 11.1 (2011), pp. 78–87.

[73] Yuanwei Liu et al. "Non-orthogonal multiple access for 5G and beyond". In: *Proceedings of the IEEE* 105.12 (2017), pp. 2347–2381.

[74] David Tse and Pramod Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[75] Eduardo Garro et al. "Information-Theoretic Analysis and Performance Evaluation of Optimal Demappers for Multi-Layer Broadcast Systems". In: *IEEE Transactions on Broadcasting* 64.4 (2018), pp. 781–790.

[76] Jon Montalban et al. "Multimedia multicast services in 5G networks: Subgrouping and non-orthogonal multiple access techniques". In: *IEEE Communications Magazine* 56.3 (2018), pp. 91–95.

[77] Anass Benjebbour et al. "Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access". In: *2013 International Symposium on Intelligent Signal Processing and Communication Systems*. IEEE. 2013, pp. 770–774.

[78] Aritz Abuin et al. "Complexity Reduction Techniques for NOMA-based RRM Algorithms in 5G Networks". In: *2020 IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)*. IEEE. 2020, pp. 86–89.

[79] Huseyin Haci, Huiling Zhu, and Jiangzhou Wang. "Performance of non-orthogonal multiple access with a novel asynchronous interference cancellation technique". In: *IEEE Transactions on Communications* 65.3 (2017), pp. 1319–1335.

[80] Jingmin Liu, Ying Li, and Yue Sun. "Detection of symbol-asynchronous uplink NOMA with imperfect time offsets". In: *IET Communications* 14.21 (2020), pp. 3873–3881.

[81] Luke Fay et al. "An overview of the ATSC 3.0 physical layer specification". In: *IEEE Transactions on Broadcasting* 62.1 (2016), pp. 159–171.

[82] Liang Zhang et al. "Layered-division-multiplexing: Theory and practice". In: *IEEE Transactions on Broadcasting* 62.1 (2016), pp. 216–232.

[83] Jon Montalbán et al. "Cloud transmission: System performance and application scenarios". In: *IEEE Transactions on Broadcasting* 60.2 (2014), pp. 170–184.

[84] Lachlan Michael and David Gómez-Barquero. "Bit-interleaved coded modulation (BICM) for ATSC 3.0". In: *IEEE Transactions on Broadcasting* 62.1 (2016), pp. 181–188.

[85] Kyung-Joong Kim et al. "Low-density parity-check codes for ATSC 3.0". In: *IEEE transactions on broadcasting* 62.1 (2016), pp. 189–196.

[86] Sung Ik Park et al. "Low complexity layered division multiplexing for ATSC 3.0". In: *IEEE Transactions on broadcasting* 62.1 (2015), pp. 233–243.

[87] Sung Ik Park et al. "Low complexity layered division multiplexing system for the next generation terrestrial broadcasting". In: *2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*. IEEE. 2015, pp. 1–3.

[88] David Gómez-Barquero and Osvaldo Simeone. "LDM versus FDM/TDM for unequal error protection in terrestrial broadcasting systems: An information-theoretic view". In: *IEEE Transactions on Broadcasting* 61.4 (2015), pp. 571–579.

[89] Jon Montalban et al. "Performance study of layered division multiplexing based on SDR platform". In: *IEEE Transactions on Broadcasting* 61.3 (2015), pp. 436–444.

[90] Cristina Regueiro et al. "LDM core services performance in ATSC 3.0". In: *IEEE Transactions on Broadcasting* 62.1 (2016), pp. 244–252.

[91] Sung-Ik Park et al. "Field test results of layered division multiplexing for the next generation DTV system". In: *IEEE Transactions on Broadcasting* 63.1 (2016), pp. 260–266.

[92] Sung-Ik Park et al. "Field comparison tests of LDM and TDM in ATSC 3.0". In: *IEEE Transactions on Broadcasting* 64.3 (2017), pp. 637–647.

[93] Eneko Iradier et al. "Adaptive resource allocation in LTE vehicular services using LDM". In: *2016 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE. 2016, pp. 1–6.

[94] Jae-young Lee et al. "ATSC 3.0 and LTE cooperation for LDM and SHVC based mobile broadcasting". In: *2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE. 2019, pp. 1–2.

[95] Liang Zhang et al. "Layered-division multiplexing: An enabling technology for multicast/broadcast service delivery in 5G". In: *IEEE Communications Magazine* 56.3 (2018), pp. 82–90.

[96] Eduardo Garro et al. "LDM integration into 5G NR: Use cases and PHY performance analysis". In: *2020 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE. 2020, pp. 1–5.

[97] Md Shantanu Islam et al. "Layer division multiplexing for 5G DL transmission within ultra-dense heterogeneous networks". In: *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*. IEEE. 2020, pp. 1–7.

[98] Erik Stare, JJ Giménez, and P Klenner. "WIB: a new system concept for digital terrestrial television (DTT)". In: *IBC 2016 Conference*. IET. 2016.

[99] *3GPP TR 36.866, "Study on network-assisted interference cancellation and suppression (NAICS) for LTE v.12.0.1," Mar. 2014.*

[100] Qiangze Chi, Xin Zhang, and Dacheng Yang. "Detection-assisted interference parameter estimation and interference cancellation for LTE-advanced system". In: *2016 IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC)*. IEEE. 2016, pp. 229–233.

[101] *3GPP TR 36.859, "Study on Downlink Multiuser Superposition Transmission (MUST) for LTE v.13.0.0," Jan. 2016.*

[102] Jinho Choi. "On the power allocation for a practical multiuser superposition scheme in NOMA systems". In: *IEEE Communications Letters* 20.3 (2016), pp. 438–441.

[103] Yifei Yuan, Zhifeng Yuan, and Li Tian. "5G non-orthogonal multiple access study in 3GPP". In: *IEEE Communications Magazine* 58.7 (2020), pp. 90–96.

[104] *3GPP TR 38.812, "Study on Non-Orthogonal Multiple Access (NOMA) for NR v.16.0.0," Dec. 2018.*

[105] Mehmet Mert Şahin and Huseyin Arslan. "Waveform-domain noma: The future of multiple access". In: *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE. 2020, pp. 1–6.

[106] Hei Victor Cheng, Emil Björnson, and Erik G Larsson. "Performance analysis of NOMA in training-based multiuser MIMO systems". In: *IEEE Transactions on Wireless communications* 17.1 (2017), pp. 372–385.

[107] Dhanushka Kudathanthirige and Gayan Amarasuriya Aruma Baduge. "NOMA-aided multicell downlink massive MIMO". In: *IEEE Journal of Selected Topics in Signal Processing* 13.3 (2019), pp. 612–627.

[108] Xin Liu et al. "Highly efficient 3-D resource allocation techniques in 5G for NOMA-enabled massive MIMO and relaying systems". In: *IEEE Journal on Selected Areas in Communications* 35.12 (2017), pp. 2785–2797.

[109] Xiaoming Chen, Rundong Jia, and Derrick Wing Kwan Ng. "The application of relay to massive non-orthogonal multiple access". In: *IEEE Transactions on Communications* 66.11 (2018), pp. 5168–5180.

[110] Yikai Li and Gayan Amarasuriya Aruma Baduge. "NOMA-aided cell-free massive MIMO systems". In: *IEEE Wireless Communications Letters* 7.6 (2018), pp. 950–953.

[111] Di Zhang et al. "Capacity analysis of NOMA with mmWave massive MIMO systems". In: *IEEE Journal on Selected Areas in Communications* 35.7 (2017), pp. 1606–1618.

[112] Zhiguo Ding and H Vincent Poor. "Design of massive-MIMO-NOMA with limited feedback". In: *IEEE Signal Processing Letters* 23.5 (2016), pp. 629–633.

[113] Theodore S Rappaport et al. *Millimeter wave wireless communications*. Pearson Education, 2015.

[114] Zhiguo Ding et al. "NOMA meets finite resolution analog beamforming in massive MIMO and millimeter-wave networks". In: *IEEE Communications Letters* 21.8 (2017), pp. 1879–1882.

[115] Mojtaba Ahmadi Almasi, Mojtaba Vaezi, and Hani Mehrpouyan. "Impact of beam misalignment on hybrid beamforming NOMA for mmWave communications". In: *IEEE transactions on communications* 67.6 (2019), pp. 4505–4518.

[116] Caijun Zhong and Zhaoyang Zhang. "Non-orthogonal multiple access with cooperative full-duplex relaying". In: *IEEE Communications Letters* 20.12 (2016), pp. 2478–2481.

[117] Xinwei Yue et al. "Exploiting full/half-duplex user relaying in NOMA systems". In: *IEEE Transactions on Communications* 66.2 (2017), pp. 560–575.

[118] Jung-Bin Kim and In-Ho Lee. "Non-orthogonal multiple access in coordinated direct and relay transmission". In: *IEEE Communications Letters* 19.11 (2015), pp. 2037–2040.

[119] Wonjae Shin et al. "Non-orthogonal multiple access in multi-cell networks: Theory, performance, and practical challenges". In: *IEEE Communications Magazine* 55.10 (2017), pp. 176–183.

[120] Md Shipon Ali, Ekram Hossain, and Dong In Kim. "Coordinated multipoint transmission in downlink multi-cell NOMA systems: Models and spectral efficiency performance". In: *IEEE Wireless Communications* 25.2 (2018), pp. 24–31.

[121] Wonjae Shin et al. "Coordinated beamforming for multi-cell MIMO-NOMA". In: *IEEE Communications Letters* 21.1 (2016), pp. 84–87.

[122] Yuanwei Liu et al. "Nonorthogonal multiple access in large-scale underlay cognitive radio networks". In: *IEEE Transactions on Vehicular Technology* 65.12 (2016), pp. 10152–10157.

[123] Lu Lv et al. "Design of cooperative non-orthogonal multicast cognitive multiple access for 5G systems: User scheduling and performance analysis". In: *IEEE Transactions on Communications* 65.6 (2017), pp. 2641–2656.

[124] Lu Lv et al. "When NOMA meets multiuser cognitive radio: Opportunistic cooperation and user scheduling". In: *IEEE Transactions on Vehicular Technology* 67.7 (2018), pp. 6679–6684.

[125] Zhiguo Ding, Pingzhi Fan, and H Vincent Poor. "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions". In: *IEEE Transactions on Vehicular Technology* 65.8 (2015), pp. 6010–6023.

[126] Yuanwei Liu et al. "Enhancing the physical layer security of non-orthogonal multiple access in large-scale networks". In: *IEEE Transactions on Wireless Communications* 16.3 (2017), pp. 1656–1672.

[127] Jianchao Chen, Liang Yang, and Mohamed-Slim Alouini. "Physical layer security for cooperative NOMA systems". In: *IEEE Transactions on Vehicular Technology* 67.5 (2018), pp. 4645–4649.

[128] Beixiong Zheng et al. "Secure NOMA based two-way relay networks using artificial noise and full duplex". In: *IEEE Journal on Selected Areas in Communications* 36.7 (2018), pp. 1426–1440.

[129] Chaoying Yuan et al. "Analysis on secrecy capacity of cooperative non-orthogonal multiple access with proactive jamming". In: *IEEE Transactions on Vehicular Technology* 68.3 (2019), pp. 2682–2696.

[130] Hongjiang Lei et al. "On secure NOMA systems with transmit antenna selection schemes". In: *IEEE Access* 5 (2017), pp. 17450–17464.

[131] Refik Caglar Kizilirmak, Corbett Ray Rowell, and Murat Uysal. "Non-orthogonal multiple access (NOMA) for indoor visible light communications". In: *2015 4th international workshop on optical wireless communications (IWOW)*. IEEE. 2015, pp. 98–101.

[132] Hanaa Marshoud et al. “Non-orthogonal multiple access for visible light communications”. In: *IEEE photonics technology letters* 28.1 (2015), pp. 51–54.

[133] Liang Yin et al. “Performance evaluation of non-orthogonal multiple access in visible light communication”. In: *IEEE Transactions on Communications* 64.12 (2016), pp. 5162–5175.

[134] Rangeet Mitra and Vimal Bhatia. “Precoded chebyshev-NLMS-based pre-distorter for nonlinear LED compensation in NOMA-VLC”. In: *IEEE Transactions on Communications* 65.11 (2017), pp. 4845–4856.

[135] Yavuz Yapici and Ismail Guvenc. “Non-orthogonal multiple access for mobile VLC networks with random receiver orientation”. In: *2019 IEEE Global Communications Conference (GLOBECOM)*. IEEE. 2019, pp. 1–6.

[136] Bangjiang Lin et al. “Experimental demonstration of bidirectional NOMA-OFDMA visible light communications”. In: *Optics express* 25.4 (2017), pp. 4348–4355.

[137] Mohammad Mozaffari et al. “A tutorial on UAVs for wireless networks: Applications, challenges, and open problems”. In: *IEEE communications surveys & tutorials* 21.3 (2019), pp. 2334–2360.

[138] Amin Farajzadeh, Ozgur Ercetin, and Halim Yanikomeroglu. “UAV data collection over NOMA backscatter networks: UAV altitude and trajectory optimization”. In: *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE. 2019, pp. 1–7.

[139] Qingqing Wu and Rui Zhang. “Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming”. In: *IEEE Transactions on Wireless Communications* 18.11 (2019), pp. 5394–5409.

[140] Qingqing Wu and Rui Zhang. “Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network”. In: *IEEE Communications Magazine* 58.1 (2019), pp. 106–112.

[141] Jiakuo Zuo et al. “Resource allocation in intelligent reflecting surface assisted NOMA systems”. In: *IEEE Transactions on Communications* 68.11 (2020), pp. 7170–7183.

[142] Abdulkadir Celik et al. “A software-defined opto-acoustic network architecture for Internet of underwater things”. In: *IEEE Communications Magazine* 58.4 (2020), pp. 88–94.

[143] Samuel Ryecroft et al. "A first implementation of underwater communications in raw water using the 433 MHz frequency combined with a bowtie antenna". In: *Sensors* 19.8 (2019), p. 1813.

[144] Tafseer Akhtar, Christos Tselios, and Ilias Politis. "Radio resource management: approaches and implementations from 4G to 5G and beyond". In: *Wireless Networks* 27.1 (2021), pp. 693–734.

[145] Bin Liu, Hui Tian, and Lingling Xu. "An efficient downlink packet scheduling algorithm for real time traffics in LTE systems". In: *2013 IEEE 10th Consumer Communications and Networking Conference (CCNC)*. IEEE. 2013, pp. 364–369.

[146] Georgios I Tsiropoulos et al. "Radio resource allocation techniques for efficient spectrum access in cognitive radio networks". In: *IEEE Communications Surveys & Tutorials* 18.1 (2014), pp. 824–847.

[147] Muhammad Anjum Qureshi and Cem Tekin. "Fast learning for dynamic resource allocation in AI-enabled radio networks". In: *IEEE Transactions on Cognitive Communications and Networking* 6.1 (2019), pp. 95–110.

[148] Ayaz Ahmad et al. "A survey on radio resource allocation in cognitive radio sensor networks". In: *IEEE Communications Surveys & Tutorials* 17.2 (2015), pp. 888–917.

[149] Che-Wei Yeh et al. "Centralized interference-aware resource allocation for device-to-device broadcast communications". In: *2014 IEEE International Conference on Internet of Things (iThings), and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom)*. IEEE. 2014, pp. 304–307.

[150] Manli Qian et al. "A super base station based centralized network architecture for 5G mobile communication systems". In: *Digital communications and Networks* 1.2 (2015), pp. 152–159.

[151] Urs Hunkeler et al. "A case for centrally controlled wireless sensor networks". In: *Computer Networks* 57.6 (2013), pp. 1425–1442.

[152] Marco Belleschi, Gábor Fodor, and Andrea Abrardo. "Performance analysis of a distributed resource allocation scheme for D2D communications". In: *2011 ieee globecom workshops (gc wkshps)*. IEEE. 2011, pp. 358–362.

[153] David A Schmidt et al. "Distributed resource allocation schemes". In: *IEEE Signal Processing Magazine* 26.5 (2009), pp. 53–63.

[154] Wenchao Li and Jing Zhang. "Cluster-based resource allocation scheme with QoS guarantee in ultra-dense networks". In: *IET Communications* 12.7 (2018), pp. 861–867.

[155] Rong Wei, Ying Wang, and Yuan Zhang. "A two-stage cluster-based resource management scheme in ultra-dense networks". In: *2014 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE. 2014, pp. 738–742.

[156] Uwe Herzog et al. "Quality of service provision and capacity expansion through extended-DSA for 5G". In: *Transactions on Emerging Telecommunications Technologies* 27.9 (2016), pp. 1250–1261.

[157] Bassem Khalfi et al. "Dynamic power pricing using distributed resource allocation for large-scale DSA systems". In: *2015 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE. 2015, pp. 1090–1094.

[158] Le Liang, Geoffrey Ye Li, and Wei Xu. "Resource allocation for D2D-enabled vehicular communications". In: *IEEE Transactions on Communications* 65.7 (2017), pp. 3186–3197.

[159] Haitao Zheng and Lili Cao. "Device-centric spectrum management". In: *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.* IEEE. 2005, pp. 56–65.

[160] Md Sipon Miah et al. "A cluster-based cooperative spectrum sensing in cognitive radio network using eigenvalue detection technique with superposition approach". In: *International Journal of Distributed Sensor Networks* 11.7 (2015), p. 207935.

[161] Jia Jasmine Meng et al. "Collaborative spectrum sensing from sparse observations in cognitive radio networks". In: *IEEE Journal on Selected Areas in Communications* 29.2 (2011), pp. 327–337.

[162] Jian Sun et al. "Uplink resource allocation for relay-aided device-to-device communication". In: *IEEE Transactions on Intelligent Transportation Systems* 19.12 (2018), pp. 3883–3892.

[163] Syed Tariq Shah et al. "SC-FDMA-based resource allocation and power control scheme for D2D communication using LTE-A uplink resource". In: *EURASIP Journal on Wireless Communications and Networking* 2015.1 (2015), pp. 1–15.

[164] Zhaohui Yang et al. "Downlink resource allocation and power control for device-to-device communication underlaying cellular networks". In: *IEEE Communications Letters* 20.7 (2016), pp. 1449–1452.

[165] Shruti Gupta et al. "Resource allocation for D2D links in the FFR and SFR aided cellular downlink". In: *IEEE Transactions on Communications* 64.10 (2016), pp. 4434–4448.

[166] Lingyang Song, Yonghui Li, and Zhu Han. "Resource allocation in full-duplex communications for future wireless networks". In: *IEEE Wireless Communications* 22.4 (2015), pp. 88–96.

[167] Hyungsik Ju and Rui Zhang. "Optimal resource allocation in full-duplex wireless-powered communication network". In: *IEEE Transactions on Communications* 62.10 (2014), pp. 3528–3540.

[168] Rui Yin et al. "Joint spectrum and power allocation for D2D communications underlaying cellular networks". In: *IEEE Transactions on Vehicular Technology* 65.4 (2015), pp. 2182–2195.

[169] Zhenyu Zhou et al. "Energy-efficient resource allocation for D2D communications underlaying cloud-RAN-based LTE-A networks". In: *IEEE Internet of Things Journal* 3.3 (2015), pp. 428–438.

[170] Xiguang Zhang et al. "Research on overlay D2D resource scheduling algorithms for V2V broadcast service". In: *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*. IEEE. 2016, pp. 1–5.

[171] Yiyang Pei and Ying-Chang Liang. "Resource allocation for device-to-device communications overlaying two-way cellular networks". In: *IEEE Transactions on Wireless Communications* 12.7 (2013), pp. 3611–3621.

[172] Ali Alnoman, Lilatul Ferdouse, and Alagan Anpalagan. "Fuzzy-based joint user association and resource allocation in HetNets". In: *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*. IEEE. 2017, pp. 1–5.

[173] Yongjun Xu et al. "Distributed resource allocation for cognitive HetNets with cross-tier interference constraint". In: *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE. 2017, pp. 1–6.

[174] Pengshuo Ji, Jie Jia, and Jian Chen. "Joint optimization on both routing and resource allocation for millimeter wave cellular networks". In: *IEEE Access* 7 (2019), pp. 93631–93642.

[175] Giuseppe Araniti et al. "A solution to the multicast subgroup formation problem in LTE systems". In: *IEEE Wireless Communications Letters* 4.2 (2015), pp. 149–152.

[176] Fang Ye, Jing Dai, and Yibing Li. "Hybrid-clustering game Algorithm for Resource Allocation in Macro-Femto HetNet." In: *TIIS* 12.4 (2018), pp. 1638–1654.

[177] Chungang Yang et al. "DISCO: Interference-aware distributed cooperation with incentive mechanism for 5G heterogeneous ultra-dense networks". In: *IEEE Communications Magazine* 56.7 (2018), pp. 198–204.

[178] Bei Xie et al. "Joint spectral efficiency and energy efficiency in FFR-based wireless heterogeneous networks". In: *IEEE Transactions on Vehicular technology* 67.9 (2017), pp. 8154–8168.

[179] Mu Yan et al. "Smart multi-RAT access based on multiagent reinforcement learning". In: *IEEE Transactions on Vehicular Technology* 67.5 (2018), pp. 4539–4551.

[180] Lin Li et al. "Traffic-aware resource allocation schemes for HetNet based on CDSA". In: *IET Communications* 11.6 (2017), pp. 942–950.

[181] Niwei Wang, Zesong Fei, and Jingming Kuang. "QoE-aware resource allocation for mixed traffics in heterogeneous networks based on Kuhn-Munkres algorithm". In: *2016 IEEE International Conference on Communication Systems (ICCS)*. IEEE. 2016, pp. 1–6.

[182] Xing Xu et al. "A resource scheduling scheme based on utility function in CoMP environment". In: *2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*. IEEE. 2017, pp. 1–6.

[183] Mikhail Gerasimenko et al. "Adaptive resource management strategy in practical multi-radio heterogeneous networks". In: *IEEE Access* 5 (2016), pp. 219–235.

[184] Abdulkadir Celik et al. "Joint interference management and resource allocation for device-to-device (D2D) communications underlying downlink/uplink decoupled (DUDe) heterogeneous networks". In: *2017 IEEE international conference on communications (ICC)*. IEEE. 2017, pp. 1–6.

[185] Sajjad Ahmad Khan, Adnan Kavak, Kerem Küçük, et al. "A novel fractional frequency reuse scheme for interference management in LTE-A HetNets". In: *IEEE Access* 7 (2019), pp. 109662–109672.

[186] Junwei Huang et al. "Two-stage resource allocation scheme for three-tier ultra-dense network". In: *China Communications* 14.10 (2017), pp. 118–129.

[187] Jose A Ayala-Romero, Juan J Alcaraz, and Javier Vales-Alonso. "Energy saving and interference coordination in HetNets using dynamic programming and CEC". In: *IEEE Access* 6 (2018), pp. 71110–71121.

[188] Syed Ahsan Raza Naqvi et al. "Energy-aware radio resource management in D2D-enabled multi-tier HetNets". In: *IEEE Access* 6 (2018), pp. 16610–16622.

[189] Xiaoming Wang et al. "Resource allocation in OFDMA heterogeneous networks for maximizing weighted sum energy efficiency". In: *Science China Information Sciences* 60.6 (2017), p. 062304.

[190] Shuang Zhang et al. "Energy and spectrum efficient power allocation with NOMA in downlink HetNets". In: *Physical Communication* 31 (2018), pp. 121–132.

[191] Nian Xia, Hsiao-Hwa Chen, and Chu-Sing Yang. "Radio resource management in machine-to-machine communications—A survey". In: *IEEE Communications Surveys & Tutorials* 20.1 (2017), pp. 791–828.

[192] Le Liang et al. "Deep-learning-based wireless resource allocation with application to vehicular networks". In: *Proceedings of the IEEE* 108.2 (2019), pp. 341–356.

[193] Tianyu Yang et al. "Deep reinforcement learning based resource allocation in low latency edge computing networks". In: *2018 15th international symposium on wireless communication systems (ISWCS)*. IEEE. 2018, pp. 1–5.

[194] Peng Yu et al. "Deep learning-based resource allocation for 5G broadband TV service". In: *IEEE Transactions on Broadcasting* 66.4 (2020), pp. 800–813.

[195] Hao Ye, Geoffrey Ye Li, and Biing-Hwang Fred Juang. "Deep reinforcement learning based resource allocation for V2V communications". In: *IEEE Transactions on Vehicular Technology* 68.4 (2019), pp. 3163–3173.

[196] Philipp Schulz et al. "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture". In: *IEEE Communications Magazine* 55.2 (2017), pp. 70–78.

[197] Heiner Lasi et al. "Industry 4.0". In: *Business & information systems engineering* 6.4 (2014), pp. 239–242.

Chapter 2

Conclusions and Future Work

2.1 Conclusions

This Ph.D. thesis has identified the spectral efficiency as one of the areas for improvement of the communication systems integrated into the 5G ecosystem and the future generation designs. Likewise, the thesis has identified the technologies that can provide the necessary improvement in spectral efficiency to satisfy the strict requirements demanded by new applications. The work has focused on the combination of NOMA techniques with RRM schemes. In this way, the improvement in spectral efficiency can be applied at two different levels, providing greater flexibility and benefits.

Considering the broad set of use cases and applications that are integrated within the 5G ecosystem, the characteristics of the solution proposed will depend strongly on their final application. Therefore, taking as reference the possible four families of applications, this Ph.D. thesis has focused on three use cases: eMBB, URLLC, and communications beyond 5G. It should be noted that in all areas, this thesis presents a study of state of the art accompanied by its respective solution design, system evaluation, and analysis of the results.

First, contribution number one ("NOMA for Broadcasting in eMBB") focuses on the joint use of NOMA and RRM in environments related to multimedia applications. The first step has been the development of a 5G-based communications system that incorporates NOMA in the physical layer. After the validation process, the evaluation has shown greater robustness and capacity than OMA-based schemes for the same use case. Later, this work has been extended to include the MAC layer retransmissions and testing the solution under challenging propagation channels. In this case, the evaluation has been based on three KPIs (reliability, latency, and throughput). NOMA offers gains in reliability and throughput at the same time without compromising latency. In addition, the provided solution has been used to enable the convergence between broadcast and unicast services. Different novel approaches have been proposed based on

RRM schemes in combination with PHY/MAC systems. Resource management has been studied in the first work environments where broadcast services and unicast services must coexist. Results obtained using NOMA in resource management are better than those achieved using OMA techniques. Subsequently, two solutions have been developed aimed at more specific use cases. The first is based on using user subgrouping techniques in resource management to transmit multicast content. The relevance of constraints has been highlighted since both NOMA and OMA converge on unfair solutions. However, once the constraints have been determined according to the use case, NOMA has proven to be a better candidate to increase the offered throughput and number of users served. It has also been found that the combined use of NOMA and OMA techniques in the same resource management scheme can provide optimal solutions. For this reason, the last analysis of this contribution consisted of combining RRM techniques to apply them in communications in the mmWave bands. Taking the results obtained as a reference, the combined models improve the single technique models by around 50%.

The research on the second contribution area ("NOMA for Critical Applications in URLLC") has studied advantages of the combined NOMA and RRM optimization for applications related to industrial communications. As in contribution number one, in this case, the analysis of the PHY/MAC layers has been the beginning of the work. Specifically, a transceiver prototype compatible with the 802.11n standard has been developed, and the minimum modifications necessary to be able to use NOMA have been made. For this, the first step has been to design the necessary architecture to design the system and define the superframe structure that manages access to the medium. In addition, a system based on time-retransmissions has been used to increase the robustness of the system. To test the design, the use case of industrial plants with different equipment types with different requirements has been used as a reference. In this case, two types of services have been distinguished, CS and BE, where the first has critical reliability and latency requirements and the second presents more flexibility in the requirements. The results indicate that NOMA provides a considerable gain in reliability and maintains latency at similar values. However, room for improvement has been detected in the structure of the superframe. For this reason, the different retransmission mechanisms compatible with the proposed PHY/MAC model has been further studied. First, different configurations of temporal and spatial diversity transmissions have been proposed in single-layer communications and their impact has been analyzed individually. The main conclusion of the work is that spatial diversity techniques considerably increase reliability. However, a tradeoff must be assumed with the network capacity. Taking the tested

retransmission schemes as a reference, they have been combined with the 802.11n prototype based on NOMA. In this case, the initial superframe has been modified to integrate each of the models. The results presented indicate that a considerable increase in reliability can be obtained in both layers simultaneously, keeping the latency at acceptable values. On the other hand, the influence of the RRM module on 5G industrial communications has also been studied. In this case, it has been used in the same use case where critical services and non-critical services must coexist, and two RRM schemes have been designed, one based on OMA and the other based on NOMA. In order to analyze the performance of the models, ranges of values of the use case parameters have been defined to see how they influence both techniques. In addition, the study has considered multiple variables (throughput, latency, and capacity) to provide a global vision. Overall, NOMA performs better than OMA, especially when the use case requirements are very demanding.

Finally, following the third contribution (Full Duplex NOMA for 6G Networks), in order to improve the spectral efficiency of the communication standards of the future, a combination of NOMA techniques with full-duplex communications has been proposed. A broadcasting ecosystem has been taken as a reference, and a complete in-band backhaul approach that provides backhaul for both robust mobile and high-data-rate fixed services within the same frequency band as the traditional broadcast content has been proposed. This solution, referred to as WIDL,is based on a theoretical throughput analysis and a definition of the required block diagram. WIDL is based on the ATSC 3.0 waveform, and it uses LDM as the mechanism to multiplex the different services. Moreover, it has been identified as the main challenge of this proposal, the self-interference signal generated. To overcome this drawback, as a first approach, digital cancellation has been applied to decrease the effect of the loopback signal. The feasibility study of the implementation of WIDL has been accompanied by a campaign of measurements carried out in Ottawa, Canada. The measuremens aimed at determining the signal isolation that could be had in real environments in order to quantify the impact that the loopback signal would have. Then, ITCN has been described as a bi-directional evolution of the WIDL system to provide full-duplex transmission among SFN transmitters, to offer enriched data services including IoT, emergency warning, connected car, and other localized data services. In this case, the required transceiver architecture has been profoundly defined, and the reception chain has been enhanced in order to decrease the impact of the loopback signal. In particular, the cancellation stage has been divided into three different blocks: adaptive antenna, analog cancellation, and digital cancellation.

2.2 Future Work

During the development of this Ph.D. thesis, various future research lines have been identified that may be of general interest for the communication systems of the future. These lines are presented below as a continuation of the contributions presented in this thesis.

Firstly, regarding the first contribution, these are the future works identified:

- *Energy-efficient NOMA*: Considering that the majority of the network users are consuming multimedia content from battery-powered devices, the energy efficiency of NOMA-based delivery systems should be studied. RRM schemes could be designed to reduce energy consumption in both the transmitter and the receiver.

- *Low-complexity solutions*: Doubtlessly, the computational cost of the SIC module is the main drawback of NOMA systems. Therefore, low-complexity solutions should be developed in order to be possible a massive deployment for any receiver.

- *Multi-cell management*: The majority of the studies related to NOMA techniques are based on single-cell scenarios, and the cooperation between gNBs is not considered. That is why network planning studies considering multiple cells at the same time should be carried out. In this way, RRM schemes for multiple synchronized transmission could be developed to enhance overall performance.

Then, these are the future works identified concerning the second contributions:

- *Power consumption*: One of the main representative use cases of URLLC is industrial sensor networks, where the devices that form the network have to have long-duration batteries. In addition to the reliability and latency analysis, power consumption at the receiver should also be measured to test the impact of NOMA-based techniques from a more global perspective.

- *Network synchronization*: The majority of the works have assumed perfect synchronization between transmitters and receivers, whereas it is not always possible to guarantee this condition. Therefore, it could be interesting to test the proposed NOMA-based solutions under non-perfect synchronization conditions.

- *Spectral efficiency friendly retransmissions*: Generally, although retransmission schemes have arisen as a good candidate to increase the reliability, they imply the redundant transmission of data packets, decreasing the overall spectral efficiency. Consequently, the design of retransmission schemes that do not condition the spectral efficiency would be of high relevance.

Finally, the future lines related to the third contribution are shown below:

- *Analog cancellation*: Although digital cancellation schemes have shown positive performance for decreasing the impact of the loopback signal, the limited dynamic range of professional devices limits a real deployment. Hence, different analog cancellation techniques should be tested in combination with the current digital cancellation solution to improve the overall performance.

- *Digital cancellation*: During the last years, several promising digital signal processing advances have been proposed in the loopback signal cancellation field. Thus, different digital cancellation alternatives could be tested in order to measure the performance under different channel conditions.

- *Adaptive beamforming*: The best use of the antenna beamforming involves not only maximizing the gain in the direction of the desired signal but also minimizing the gain of the noise and the interference signal. In this case, it would be worth implementing adaptive beamforming schemes to reduce the power of the received loopback signal before any cancellation technique is applied.

- *AI-based techniques*: It has been widely demonstrated in the literature that AI-based alternatives can considerably increase the efficiency of current cancellation algorithms. Therefore, it should be analyzed how AI techniques could be combined with the transceiver architecture in order to keep the loopback signal below the noise floor.

Chapter 3

Published or accepted papers

3.1 Publications associated to Contribution 1

3.1.1 Conference paper C1

This subsection presents a conference paper related with Contribution 1. The full reference of the paper is presented below:

- E. Iradier, J. Montalban, D. Romero, Y. Wu, L. Zhang and W. Li, "NOMA based 5G NR for PTM Communications," 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Jeju, Korea (South), 2019, pp. 1-6, doi: 10.1109/BMSB47279.2019.8971856.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Indexed Congress in IEEExplore
- Area: Computer Science and Engineering
- SJR factor: 0.300

NOMA based 5G NR for PTM Communications

E. Iradier, J. Montalban, D. Romero
Department of Communications Engineering
University of the Basque Country (UPV/EHU)
Plaza Torres Quevedo 1, Bilbao (Spain)
{eneko.iradier, jon.montalban}@ehu.eus
dromero007@ikasle.ehu.eus

Y. Wu, L. Zhang, W. Li
Communications Research Centre
CRC
Ottawa, Canada
{yiyan.wu, liang.zhang, wei.li}@canada.ca

***Abstract*— Recent advances in broadband communications have led to the first standardization phase of 5G (Rel' 15). Several works present this novel standard as the main solution in order to cover all the new requirements that have arisen in the multimedia delivery industry. However, the latest 5G release does not include the option of transmitting traditional broadcast services. That is why, in this paper, Non-Orthogonal Multiple Access (NOMA) is presented as a promising candidate in order to boost the implementation of the broadcast/multicast communications within the 5G NR standard. Therefore, a 5G New Radio (NR) fully compliant simulation tool has been developed and NOMA multiplexing schemes have been added, introducing as few modifications as possible. The simulation platform and proposed techniques have been tested within two different use cases. Results present a better overall reliability performance of the communications by using NOMA. In fact, both services performance is improved when compared to traditional TDMA schemes, with a maximum gain around 4 dB.**

***Keywords*— 5G, Broadcast, LDPC, Multicast, NOMA, TDMA**

I. Introduction

The standardization process of the fifth generation of mobile communications (5G), which is considered the key for the new data consumption habits, has already started. In fact, the some of the most innovative NR concepts have been included in Release 15 [1]. One of the main improvements in comparison with previous releases is the use of newly designed LDPC codes (Low-Density Parity-Check). These codes increase the overall spectral efficiency, and so, more challenging wireless scenarios are feasible.

Taking into account this standardization process, the Radiocommunication Sector of the International Telecommunication Union (ITU-R) has defined three scenarios that could be enabled by 5G NR: Enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency Communications (URLLC) and Massive Machine Type Communications (mMTC) [2].

Since the possible applicable scenarios are quite different, specific relevant requirements are expected for each case. The classic information and entertainment media delivery scenario, eMBB, will overcome the barrier of a Gbps per average user inside the coverage area, with possible higher peak data rates. In the case of URLLC systems, latency is considered a critical aspect, this is why, one millisecond of end-to-end latency is established as a reference. Finally, it is expected to increase the connection density up to 10^6/km for mMTC environments [2].

The deployment of 5G communication systems within new environments will enhance the spectral efficiency. However, in order to achieve all the objectives, 5G has to be complemented with other techniques that allow a considerable increase in the spectrum efficiency. Several interesting techniques have been presented in the recent literature and, specifically, NOMA is considered one of the most promising.

As described in [3], NOMA could be based on a layered transmission structure to transmit simultaneously multiple different services with different power and robustness levels. In NOMA, the whole bandwidth during the 100% of the time is used to transmit at least two services. That is why, NOMA is usually presented as a technique that improves the overall spectral efficiency when compared with more classic approaches.

In addition, a more relevant and recent example of the successful implementation of this technology was the acceptance of Layered Division Multiplexing (LDM), a low complexity version of NOMA, as a Physical Layer (PHY) baseline technology for ATSC 3.0 [4]. In LDM, on the one hand, a robust Core Layer (CL) is transmitted for portable and handheld receivers to deliver mobile services in indoor conditions. On the other hand, an Enhance Layer (EL) is designed to deliver high data rate services, such as UHDTV or multiple HDTV services to fix terminals.

In this paper, the authors propose to introduce NOMA as part of the 5G NR Release 15 physical layer. Consequently, services with unbalanced capacity requirements could be simultaneously offered with increased spectrum efficiency. In addition, depending on the use case, this efficiency increase can be turned into a throughput gain, reliability gain or both at the same time. In short, the introduction of NOMA can be turned into a more flexible configuration, which means a wider range of possible implementable PHY configurations.

The paper is organized as follows. Section II describes the related work that has been developed about this topic. Section III provides the definition of the combination of NOMA with eMBMS systems. Moreover, Section IV describes the evaluation system and the proposed use cases. Then, Section V presents the obtained results and, finally, conclusions are described in Section VI.

978-1-7281-2150-5/19/$31.00 ©2019 IEEE

II. Related work

Nowadays, 5G is a hot topic inside every communications forum. This fifth generation is supposed to be the solution to face the latest communication challenges and the final convergence tool for the cooperation of a wide range of communication systems. The first version of 5G, Release 15, corresponds to NR Phase 1, where there are common elements between LTE-A-Pro and NR, such as Orthogonal Frequency Division Multiplexing (OFDM). However, in this phase a very relevant scenario has not been considered yet: the point-to-multipoint (PTM) audiovisual content distribution.

In fact, the European Commission (EC) has recently funded an international research project to work in this area, namely 5G-Xcast [5]. Its main objective is to develop capabilities for 5G broadcast and multicast PTM systems. In order to achieve that goal, as presented in [6], new levels of network management and delivery cost-efficiency are considered. The paper also discusses the implications of PTM for network slicing, to customize and optimize network resources on a common 5G infrastructure to accommodate different use cases and services taking into account the user density. Although this project has presented different alternatives to enable broadcasting in 5G environments, NOMA has not been considered.

In [7], a two-layer architecture is proposed based on a transceiver architecture for LTE-A-Pro. In order to compare both technologies, some results are presented for specific Modulation and Coding Scheme (MCS) values. However, despite the high compatibility that exists between LTE-A-Pro and 5G NR, this approach is not including a fully compatible transceiver with Release 15. For instance, the newly designed LDPC/Polar codes are not implemented, and it must be taken into account, that the efficiency of the Forward Error Codes (FEC) is a critical parameter for the performance of NOMA. Therefore, the performance curves are not completely applicable for Release 15 and the reliability increase that newly designed LDPC/Polar codes offer is not included.

To our best knowledge, this paper is the first comprehensive analysis of the 5G NR in combination with NOMA techniques. Hence, by this analysis, firstly, we are proposing a NOMA-based communication system for enabling broadcasting in 5G that is very close to the specifications of the standard. Secondly, Bit Error Rate (BER) vs. SNR performance curves are presented for two scenarios for both, 5G NR systems and NOMA over 5G NR systems.

III. 5G eMBMS use case

Implementing broadcasting within broadband cellular network technologies (3G/4G) has always been a challenge. Thus, implementing it on 5G platform will not be straightforward. Traditionally, the most popular alternative has been Multimedia Broadcast Multicast Services (MBMS) technology and its subsequent evolutions.

The first standardization phases of MBMS appeared in UMTS Release 6, before any 4G specification. However, it was not included in the first LTE version. In fact, in Release 9 a slightly different version was approved, which was called

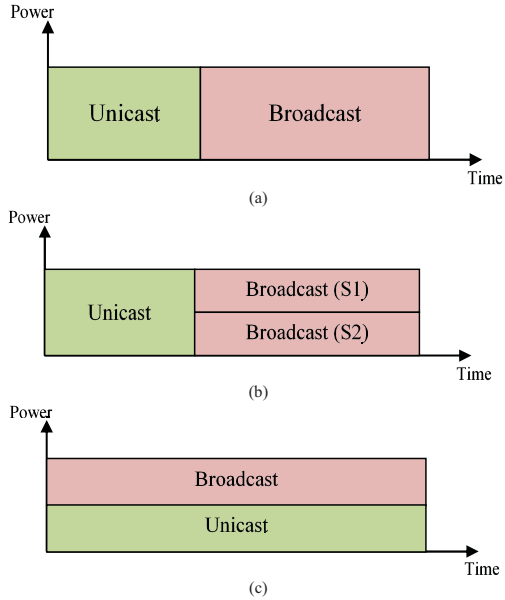


Fig 1. eMBMS representation: (a) only eMBMS, (b) eMBMS and NOMA, (c) NOMA instead of eMBMS

Evolved MBMS (eMBMS). A simplified eMBMS representation is shown in Fig. 1(a). Taking into account this background and the proposed roadmap for 5G NR, a similar version of the actual eMBMS is expected to be implemented in future releases (probably in Rel' 17, since Rel' 16 does not include PTM communications).

Regarding NOMA and its interoperability with eMBMS ecosystems, several works have been carried out based on LTE or theoretical 5G approaches, [8] to [10]. Therefore, it is the appropriate time to integrate NOMA in 5G NR and enable point to multipoint services with better efficiency. The main difference with traditional eMBMS is the multiplexing philosophy, since, while current eMBMS is based on Time Division Multiplexing (TDM) and could be complemented with frequency division; NOMA is based on power division.

If NOMA is adopted as a multiplexing complementary alternative for the existing eMBMS system, a combination of time-power division will be done, as it is represented in Figure 1.b. In this case, firstly, resources are time domain divided for unicast services and for broadcast/multicast services. Hence, in the first part of the frame, a time division will be applied, and unicast services will be transmitted. In addition, in the second part of the frame, time division is not needed again and the broadcast/multicast services will be transmitted overlaid in several layers. In Fig. 1(b), two different services are overlapped, S1 and S2.

However, if NOMA is laid out as a replacement of multiplexing technologies in 5G, exclusively a power domain

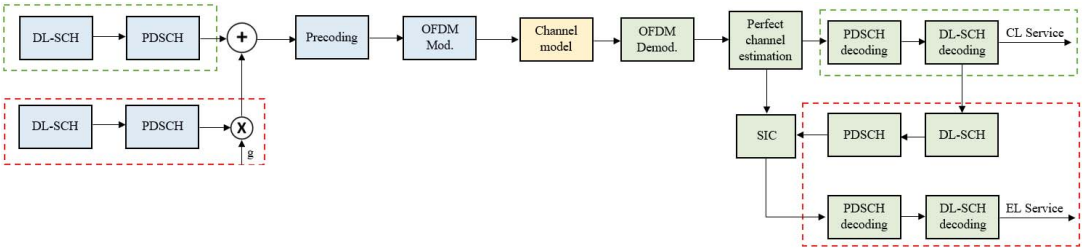


Fig 2. General block diagram of the transceiver architecture

division can be done, and any time or frequency division would be complementary. For instance, a common broadcast service could be delivered in the CL for all the users in the cell, and the rest of the unicast services will be transmitted in the EL, applying time division to share the RF resources among the users. A summarizing diagram is shown in Fig. 1(c) for a simple two-layer scenario.

These proposals can come with an increase in the offered throughput due to the reorganization of the streams and a better exploitation of the resources. Nevertheless, because of the power division that NOMA assumes, the required threshold to decode successfully each of the MCS suffers a penalty.

IV. NOMA-Based 5G NR

In this section, the developed simulation tool is presented. Firstly, the simulation set-up will be described taking into account the implemented modules, and, secondly, the two proposed use cases are presented.

A. Simulation Set-up

A transmitter and receiver chain has been implemented in Matlab. This prototype is fully compliant with the 5G NR Release 15 PHY standard. It is important to highlight that in this work a Physical Downlink Shared Channel (PDSCH) channel has been implemented because 5G NR does not include a specific broadcast channel. However, the proposed transceiver chain is fully compliant with a generic broadcast channel, and, therefore, if broadcast services were included in Release 17 or Release 18, this proposal would be applicable. Afterwards, several modifications have been introduced in the transceiver blocks in order to introduce NOMA multiplexing capabilities. A block diagram of the transceiver architecture is shown in Fig 2, where blue blocks are transmitter modules and green modules are receiver modules.

First, two independent data flows are created, one for the CL (green dashed box on the left) and the other one for the EL (red dashed box on the left). In each of the data flows, Downlink Shared Channel (DL-SCH) transport channel coding is applied and the PDSCH is generated. Then, the EL symbols are attenuated by applying an injection level, g. The injection level defines the power level of the EL signal relative to the CL signal. Since the CL signal is designed to have higher power, g has a real value in [0, 1), where g = 0 results in a single-layer system [3]. After having superposed both signals, the NOMA ensemble is normalized and a precoding is applied before the grid is calculated. Finally, the last step in the transmitter side is the OFDM framing.

Once the transmitter has finished each of the modules, the channel model is applied. In this case, the AWGN channel is proposed, since it is the first approach to evaluate this proposal.

The first step in the receiver side is to demodulate the OFDM signal, in which perfect synchronization is assumed. Concerning the channel estimation, perfect channel estimation is also considered. At this reception point, since the low-power EL service is treated as an additional interference in the CL decoding process, the CL service is obtained just decoding PDSCH and DL-SCH. However, in order to decode the EL service a signal cancelation is applied (SIC module in Fig 2). The input signals for the SIC module are: the received signal and the obtained CL re-modulated again. Finally, the EL service is obtained decoding PDSCH and DL-SCH from the SIC output.

Undoubtedly, the introduction of NOMA schemes increases the receiver complexity. However, drawbacks only affect the EL. In fact, CL is obtained by following a fully complaint receiver chain. Therefore, it is just the EL service the one that is affected by the complexity increase.

For each simulation set, 30 kHz subcarrier spacing with SISO configuration has been considered, within a 10 MHz RF channel. In addition, a normal Cyclic Prefix (CP) length is assumed. Therefore, the CP duration is 2.34 µs per each OFDM symbol, which has a total length of 35.68 µs.

Finally, regarding Medium Access Control (MAC) layer and above layers, Hybrid Automatic Repeat Request (HARQ) retransmission schemes are disabled in this paper. Although HARQ schemes could improve the overall communication system reliability performance, the complexity needed in order to introduce NOMA will be also increased, and so, it is not considered in this first approach. Therefore, the results shown in this work are evaluating the performance of a 5G NR fully compliant PHY layer.

B. Use Case 1: Only Broadcast Mode

In this case, NOMA is configured for conveying two broadcast media contents in the same RF channel. In addition to that, the use of Scalable Video Coding (SVC) techniques is considered.

TABLE I. CONFIGURATION FOR USE CASE 1

NOMA (IL = -5 dB)			TDMA (50%-50%)		
	Configuration	Capacity (Mbps)		Configuration	Capacity (Mbps)
UL	MCS 1 QPSK 193/1024	3.7	Mobile (I)	MCS 2 QPSK 308/1024	2.9
			Mobile (II)	MCS 3 QPSK 449/1024	4.3
LL	MCS 8 16QAM 553/1024	21.0	Fixed (I)	MCS 16 64QAM 719/1024	20.5
			Fixed (II)	MCS 17 64QAM 772/1024	22.0

TABLE II. CONFIGURATION FOR USE CASE 2

NOMA (IL = -14 dB)			TDMA (75%-25%)		
	Configuration	Capacity (Mbps)		Configuration	Capacity (Mbps)
UL	MCS 12 64QAM 517/1024	29.4	Fixed (I)	MCS 15 64QAM 666/1024	28.4
			Fixed (II)	MCS 16 64QAM 719/1024	30.7
LL	MCS 6 16QAM 434/1024	16.5	IoT (I)	MCS 24 256QAM 841/1024	16.0
			IoT (II)	MCS 25 256QAM 885/1024	16.8

Therefore, the first service (CL) should guarantee the delivery of at least one HD service to mobile receivers [11], so the targeting capacity should be between three and five Mbps. For this service, low SNR threshold values are expected. The second service is simulating the transmission of a 4K/8K content to fix receivers. Receivers will enhance the CL service by the use of SVC in order to obtain higher quality content. In challenging reception conditions, fixed receivers will always decode the CL service. In this case, for the EL, the targeting SNR threshold should be higher and the desired data rate is around 20Mbps [12].

Moreover, the exact transmitted capacities have been obtained by implementing real 5G NR MCS values (obtained from Table 5.1.3.1-2 in [13]), they are shown in TABLE I. This capacity values have been calculated for a 10MHz channel bandwidth.

In relation with eMBMS environments, this use case represents an example for the Fig 1.b, where NOMA is applied to deliver two broadcast services. In this case, both broadcast services are not completely uncorrelated, because EL will enhance the CL layer service if the reception is without errors. Moreover, the complexity increment derived from the implementation of NOMA techniques (mentioned in Section IV.A) only affects fixed receivers. In the case of mobile receivers, the existence of the EL is completely transparent.

C. Use Case 2: Mixed Mode

In this use case, NOMA is implemented to deliver broadcast and unicast services, therefore a mixed transmission mode it is assumed. For this case, services are considered independent and they are targeting different receivers.

On the one hand, the CL is configured for addressing the broadcast requirement. Fixed receivers will receive 4K/8K content, so, around 30 Mbps have to be delivered [14]. Correct reception must be guaranteed to 99% of receivers, and, therefore, reliability is considered a critical aspect.

On the other hand, the EL is configured to deliver unicast services based on IoT for smart traffic management. In this case, the EL is used to deliver on demand traffic information to IoT devices. Although a standard quality service is considered for this kind of users, RF resources should be distributed among the users that require EL services. That is why, higher capacity has to be offered in comparison with the service requirement. In this case, around 16 Mbps are going to be transmitted. Therefore, in the case of 160 devices connected simultaneously to the same transmitter, at least, 100 kbps are received in each device.

As in the previous Section, the exact suggested capacities have been obtained by implementing real 5G NR MCS values (obtained from Table 5.1.3.1-2 in [13]). TABLE II summarizes the capacity for each multiplexing technique configuration. This capacity values have been calculated for a 10MHz channel bandwidth.

Comparing this use case with the eMBMS paradigm, this situation should cover the Fig. 1(c) situation, where a common broadcast transmission is delivered in the CL and unicast services are transmitted in the EL. In this case, IoT devices receiving the EL have to cancel the CL service in order to obtain their content. However, for the fixed receivers it does not matter what is transmitted in the EL.

V. Results

A. Use Case 1: Only Broadcast Mode

Results for the only broadcast mode are presented in Fig 3, where CL results are represented in Fig. 3(a) and EL results in Fig. 3(b). In both graphics, a comparison is made between NOMA and TDMA multiplexing schemes performance. In the y-axis the BER value is represented, whereas in the x-axis the SNR value in dBs.

First, in Fig. 3(a), the results for CL services are presented. In a first approach, it seems that NOMA and TDMA I configuration have similar reliability performance. However, it is remarkable that even if MCS 1 by using NOMA and MCS 2 by using TDMA have similar performance (NOMA offers a gain of 0.3 dB for a BER value of 10^{-4}), from the capacity point of view, by using NOMA the transmitted data rate is around 20% higher. Comparing NOMA performance with TDMA II, a gain around 2.1 dB is obtained.

The difference in the capacity values is due to the granularity of the standard configurations for the PHY level.

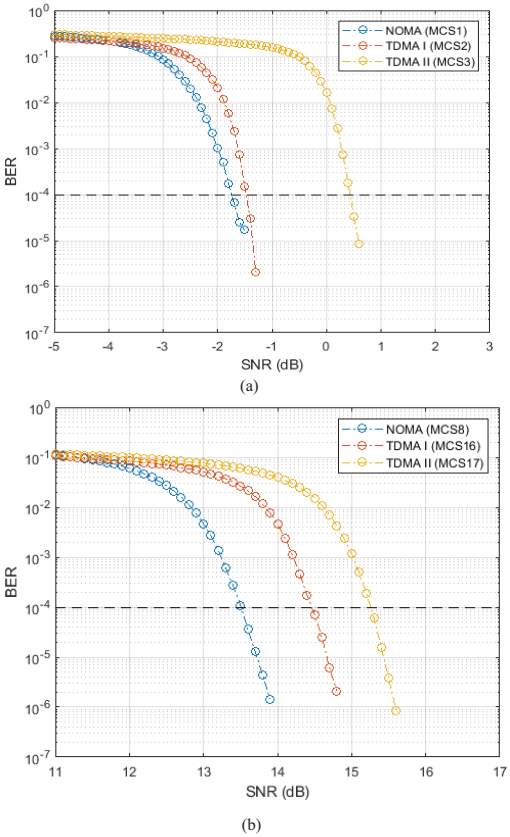


Fig 3. Obtained results for use case 1: (a) NOMA vs. TDMA for mobile service, (b) NOMA vs. TDMA for fix service

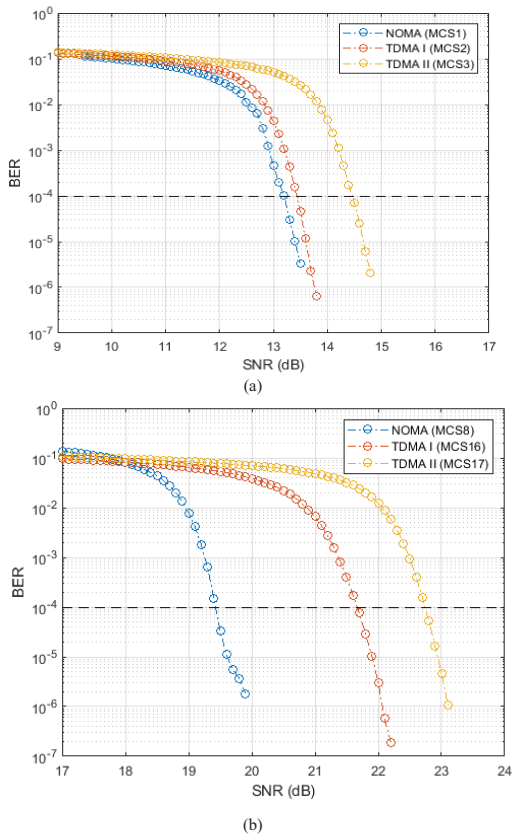


Fig 4. Obtained results for use case 2: (a) NOMA vs. TDMA for fix service, (b) NOMA vs. TDMA for IoT

Since the number of MCS configurations are limited in the 5G NR standard, capacity values are different for each configuration.

Furthermore, in Fig. 3(b), the results for fix receivers are presented for both, NOMA and TDMA multiplexing schemes. In this case, the results show an interesting performance gain of about 1 dB in comparison with TDMA I, and around 2 dB in relation to TDMA II. This gain is due to the asymmetry between the offered services SNR value. In this case, the gain has been invested in improving the performance by reducing the SNR threshold. However, a similar analysis could be carried out, in which the goal will be to increase the capacity of the fix services.

Therefore, it can be assumed that by using NOMA in an only broadcast mode scenario, reliability gains can be considered for both service types. Adding the obtained gains for both services, more than 4 dB can be obtained.

B. Use Case 2: Mixed Mode

Results for the mixed broadcast/unicast mode are presented in Fig 4, where CL results are represented in Fig. 4(a) and EL results in Fig. 4(b). Graphics maintain the format used in the previous Section.

On the one hand, regarding Fig. 4(a), broadcast services have similar performance behavior in terms of reliability. Although, it is true that NOMA offers slightly better results in comparison with the low capacity TDMA approach (TDMA I), around 0.2 dB; a higher gain is obtained in relation to TDMA II case, where around 1.3 dB are obtained.

However, the similarity in the CL results involves a greater unbalance in the EL results, which are shown in Fig. 4(b). In this case, the use of NOMA will benefit IoT receivers from more than 2 dB SNR gain in comparison with TDMA I configuration. Moreover, more than 3 dB of gain are obtained in relation to TDMA II configuration. Undoubtedly, this gain

is really interesting due to the characteristics of the EL receivers, which should be less efficient than the fixed receivers.

Therefore, finally, the possible application of NOMA for a mixed broadcast/unicast mode is also demonstrated. In fact, up to 4.5 dB aggregated gains are obtained. Moreover, gains could be increased implementing more challenging channels, such as Rayleigh, 0dB Echo or TU-6.

VI. Conclusion

This paper presents a novel communication system for broadcasting environments based on the 5G NR Release 15. The theoretical applications obtained including NOMA in a simplified 5G NR architecture are presented in Section III. Afterwards, a fully compliant simulation tool has been developed, presented and evaluated by comprehensively studying two different scenarios.

Results demonstrate that NOMA is a multiplexing scheme that should be taken into account for future releases due to the reliability gain that provides. In the represented use cases, a general behavior has been detected. Firstly, in the CL, slightly better reliability results have been obtained (up to 2.3 dB gain in the first scenario). Secondly, higher gain values are obtained in the EL by using NOMA. In fact, a minimum gain of one dB is presented in the first case, and a maximum gain of three dB in the second case.

To sum up, although a promising technique that offers a wide range of gains has been presented, several improvements should be consider in order to increase the overall performance. That is why, in future works, MAC and upper layers are also going to be taken into account to enhance to whole simulation set up.

Acknowledgments

This work has been partially supported by the Basque Government under the PREDOC grant program (PRE_2018_1_0344) and by the Spanish Government under the grant RTI2018-099162-B-I00 (MCIU/AEI/FEDER, UE).

References

[1] TS 38.201, Tech. Spec. Group Services and System Aspects, “NR; Physical layer; General description,” V15.0.0, January 2018.

[2] ITU-R Rec. ITU-R M. 2083-0, “IMT Vision — Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond,” Sept. 2015.

[3] L. Zhang et al., "Layered-Division-Multiplexing: Theory and Practice," in IEEE Transactions on Broadcasting, vol. 62, no. 1, pp. 216-232, March 2016.

[4] S. Park et al., “Low complexity layered division multiplexing system for ATSC 3.0,” IEEE Transactions on Broadcasting., vol. 62, no. 1, Mar. 2016.

[5] D. Ratkaj and A. Murphy, Eds., “Definition of Use Cases, Requirements and KPIs,” Deliverable D2.1, 5G-PPP 5G-Xcast project, Oct. 2017.

[6] J. J. Gimenez, D. Gomez-Barquero, J. Morgade and E. Stare, "Wideband Broadcasting: A Power-Efficient Approach to 5G Broadcasting," in IEEE Communications Magazine, vol. 56, no. 3, pp. 119-125, March 2018.

[7] D. Vargas and Y. J. D. Kim, “Two-Layered Superposition of Broadcast/Multicast and Unicast Signals in Multiuser OFDMA Systems” https://arxiv.org/abs/1811.00912. Accessed on January 10, 2018.

[8] J. Montalban et al., "Multimedia Multicast Services in 5G Networks: Subgrouping and Non-Orthogonal Multiple Access Techniques," in IEEE Communications Magazine, vol. 56, no. 3, pp. 91-95, March 2018.

[9] L. Zhang et al., "Layered-Division Multiplexing: An Enabling Technology for Multicast/Broadcast Service Delivery in 5G," in IEEE Communications Magazine, vol. 56, no. 3, pp. 82-90, March 2018.

[10] L. Zhang, Y. Wu, W. Li, K. Salehian, A. Florea and G. K. Walker, "Improving LTE eMBMS system spectrum efficiency and service quality using channel bonding, non-orthogonal multiplexing and SFN," 2016 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Nara, 2016, pp. 1-8.

[11] L. Zhang et al., "Using Layered-Division-Multiplexing to Deliver Multi-Layer Mobile Services in ATSC 3.0," in IEEE Transactions on Broadcasting, vol. 65, no. 1, pp. 40-52, March 2019.

[12] J. Lee et al., "Efficient Transmission of Multiple Broadcasting Services Using LDM and SHVC," in IEEE Transactions on Broadcasting, vol. 64, no. 2, pp. 177-187, June 2018.

[13] 3GPP TS 38.214 v15.3.0, Tech. Spec. Group Services and System Aspects, “NR; Physical layer procedures for data (Release 15)”, Sept. 2018

[14] 3GPP TS 22.261 v15.7.0, Tech. Spec. Group Services and System Aspects, “Service Requirements for the 5G System; Stage 1, “Rel. 15; Dec. 2018.

3.1.2 Journal paper J1

This subsection presents a journal paper related with Contribution 1. The full reference of the paper is presented below:

- E. Iradier, J. Montalban, L. Fanari, P. Angueira, L. Zhang, Y. Wu and W. Li, "Using NOMA for Enabling Broadcast/Unicast Convergence in 5G Networks," in IEEE Transactions on Broadcasting, vol. 66, no. 2, pp. 503-514, June 2020, doi: 10.1109/TBC.2020.2981759.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Journal paper indexed in JCR and IEEExplore
- Area: Electrical & Electronic Engineering
- Ranking: 69/266 (Q2) based on JCR 2019
- Impact factor (JCR): 3.419

IEEE TRANSACTIONS ON BROADCASTING, VOL. 66, NO. 2, JUNE 2020 503

Using NOMA for Enabling Broadcast/Unicast Convergence in 5G Networks

Eneko Iradier, *Student Member, IEEE*, Jon Montalban, *Member, IEEE*, Lorenzo Fanari, *Student Member, IEEE*, Pablo Angueira, *Senior Member, IEEE*, Liang Zhang, *Senior Member, IEEE*, Yiyan Wu, *Fellow, IEEE*, and Wei Li, *Member, IEEE*

***Abstract*—This paper addresses the challenge of broadcast and unicast convergence by proposing a PHY/MAC (Physical Layer/Medium Access Control) architecture for 5G New Radio (NR). The solution is based on Power domain Non Orthogonal Multiple Access (P-NOMA). The main PHY/MAC configuration parameters have been analyzed theoretically and their impact on the service configurations is presented in this manuscript. The system concept has been translated into a prototype model and different evaluation tests are presented. First, simulations show that the PHY layer performs better than Time Division Multiplexing/Frequency Division Multiplexing (TDM/FDM) choices of current broadband access systems. Second, performance tests using a network simulation tool are described. The results for capacity, latency and reliability demonstrate that the proposed solution offers an excellent broadcast/unicast convergence choice with significant gain values with respect to legacy PHY/MAC alternatives.**

***Index Terms*—5G, broadcast, convergence, LDM, MAC, NOMA, P-NOMA, unicast.**

I. Introduction

THE STANDARDIZATION process required to create and enhance the fifth generation of mobile communications (5G), which is expected to cover all the increasing connectivity necessities [1], is ongoing. In fact, the first Release (Rel-15) of 5G has already been published, including New Radio (NR) [2], and several improvements in different levels of the protocol stack make this solution a proper alternative for different use cases. One of the most relevant contributions can be considered the development of a flexible Physical Layer (PHY) that can be implemented in several scenarios with completely different requirements. It should also be mentioned that Low-Density Parity-Check (LDPC) codes are introduced to enhance the communication reliability.

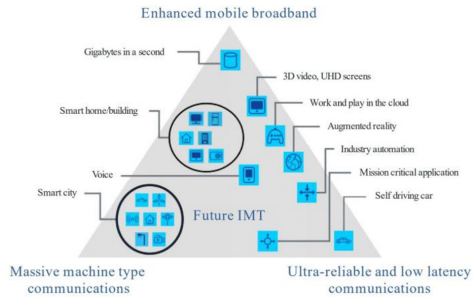


Fig. 1. 5G NR applications and use case families [3].

The Radiocommunication Sector of the International Telecommunication Union (ITU-R) has divided the entire application frame into three use case families: enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency Communications (URLLC) and massive Machine Type Communications (mMTC) [3]. In Fig. 1, a summary of the use case families is shown with some other more specific applications.

Each one of the different use case families shown in Fig. 1 has different requirements. In eMBB, which is related to the classic information and entertainment media delivery scenario, high data rates are required. In fact, it is expected to guarantee peak data rate values close to 10 Gbps for the uplink and 20 Gbps for the downlink and support speeds up to 500 km/h. In the case of URLLC, as it is oriented to mission-critical communications, reliability and latency are the most demanded requirements. In fact, an upper bound of one millisecond of end-to-end latency has been established in the user plane. Finally, concerning mMTC, the network is expected to increase the device density rate to values close to 10^6

Manuscript received December 1, 2019; revised March 2, 2020; accepted March 3, 2020. Date of publication April 14, 2020; date of current version June 5, 2020. This work was supported in part by the Basque Government (Project IOTERRAZ) under Grant KK-2019/00046 ELKARTEK 2019, Grant IT1234-19, and the PREDOC Grant Program PRE_2019_2_0037, and in part by the Spanish Government (Project PHANTOM) under Grant RTI2018-099162-B-I00 (MCIU/AEI/FEDER, UE). *(Corresponding author: Eneko Iradier.)*

Eneko Iradier, Jon Montalban, Lorenzo Fanari, and Pablo Angueira are with the Department of Communications Engineering, University of Basque Country (UPV/EHU), 48013 Bilbao, Spain (e-mail: eneko.iradier@ehu.eus; jon.montalban@ehu.eus; lorenzo.fanari@ehu.eus; pablo.angueira@ehu.eus).

Liang Zhang, Yiyan Wu, and Wei Li are with the Department of Wireless Communications, Communications Research Centre Canada, Ottawa, ON K2H 8S2, Canada (e-mail: liang.zhang@canada.ca; yiyan.wu@canada.ca; wei.li@canada.ca).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2020.2981759

devices per km^2 and increase the battery lifetime to around 10 years.

The combination of the large list of new applications and the configuration flexibility presented in the latest Release gives the opportunity to make 5G the main generic solution for wireless communications. However, a mechanism that could enhance the 5G performance is still missing: Point-to-Multipoint (PTM) communications. Although it is true that broadcast/multicast communications are not considered in this first version, they will probably be integrated in future Releases (Rel-17/18).

In parallel to the integration of PTM communications, the convergence between unicast and broadcast services is still an open issue. In 4G systems, Evolved Multimedia Broadcast Multicast Services (eMBMS) has been implemented, which is based on Time Division Multiplexing (TDM) and is complemented with Frequency Division Multiplexing (FDM).

This paper proposes a different standpoint to replace or to complement eMBMS by means of Non-Orthogonal Multiple Access (NOMA) techniques.

As described in [4], a layered division or power-based NOMA can be implemented in order to deliver simultaneously different services, assuming diverse power levels and configurations. The main advantage of P-NOMA systems is that the 100% of the RF bandwidth is used during the 100% of time to transmit two (or more) services. That is why NOMA is considered a spectral efficiency friendly technique when compared with orthogonal multiplexing access solutions, such as TDM or FDM.

The main example of the success of P-NOMA systems is the acceptance of a low complexity P-NOMA structure, Layered Division Multiplexing (LDM), to be part of the Advanced Television Systems Committee (ATSC 3.0) PHY layer baseline technology [5]. In this case, a robust configuration is implemented in the Upper Layer (UL), oriented to portable and mobile receivers. On the other hand, in the Lower Layer (LL), a high capacity configuration is chosen to deliver high data rate services, such as Ultra High Definition Television (UHDTV) or multiple High Definition Television (HDTV) services, to fixed receivers.

The main objective of this paper is to propose, design and evaluate a novel solution for enabling broadcast/unicast service convergence in 5G networks based on P-NOMA. Afterwards, the main implications of this architecture are presented as well as implementation guidelines based on a prototype. Finally, the prototype is evaluated at PHY level using an upgraded network simulation tool. PHY and network layer results are shown and analyzed, separately.

In summary, the technical contributions of this paper include:

1) Full architecture definition of a P-NOMA solution for 5G NR at PHY/MAC (Medium Access Control) level. Some blocks of the architecture are novel and critical for the overall performance of the proposal.
2) Design and implementation of a frame adaptation layer to combine UL and LL MAC packets.
3) Performance evaluation and gain analysis of the proposed technique over different propagation channels.
4) Presentation of network level simulations and reliability, latency and capacity analysis.

The rest of the paper is organized as follows. The next section describes related work published up to now on this topic. Section III is focused on overview of 5G NR at different levels. In Section IV, NOMA-based 5G model is presented, taking into account the implications and the proposed novelties. Then, in Section V, a PHY level evaluation of the proposed solution is carried out. In Section VI, the proposed solution is tested in a network simulation tool. Finally, Section VII contains the conclusions and the future works.

II. Related Work

A. Broadcast/Unicast Convergence

5G is expected to be the global wireless solution to guarantee convergence and cooperation between different networks, devices and technologies [6]. However, in its first version, Rel-15, a fundamental scenario, especially for eMBB communications, has not been considered: PTM communications.

The broadcast implementation roadmap in 5G has similar milestones when compared to 4G. Although a broadcast traffic distribution alternative was already included in 3G systems within Multimedia Broadcast Multicast Services (MBMS), it was not until Rel-9 that the first broadcast transmission capabilities were introduced. In fact, in 4G, a slightly enhanced solution was presented (evolved MBMS, eMBMS), which brought higher and more flexible data rate configurations, single frequency network (SFN) operations and carrier configuration flexibility [7]. Latest enhancements applied to eMBMS were included in Long Term Evolution (LTE) Rel-14. One of the most relevant differences was that the limitation of the maximum of 60% of radio resources for broadcast transmissions was removed [8]. In addition, support for eMBMS will be introduced for NR mainly for public safety use cases, Vehicle-to-everything (V2X) applications and railways, and therefore, further enhancements will come in Rel-16 [9].

Concerning the future of broadcast in 5G, the most relevant contributions are related to the 5G-Xcast project [10]. 5G-Xcast was an international research project, funded by the European Commission (EC) that started in June 2017. This project presented a cross-layer PTM proposal at different architecture levels. Among other contributions in 5G-Xcast project, in [11], a specific approach of the 5G NR PHY for terrestrial broadcast was presented. In this paper, the critical parameters of the waveform generation are discussed and the NR-based MBMS solution is compared with the latest LTE Rel-14 version. Aiming to cover a different perspective, in [12], the network slicing concept is discussed within the 5G paradigm. The main implications are detailed in order to optimize the network resources of a particular 5G infrastructure over different PTM use cases.

B. NOMA for 5G Systems

P-NOMA-based systems such as LDM, have demonstrated very positive spectral efficiency properties and an excellent performance for broadcast applications [4]. Nevertheless,

LDM has not been so successful for broadband communication systems and although NOMA was adopted by 3GPP Rel-14, it has not been implemented so far. However, several works presented P-NOMA-based solutions as relevant candidates for broadcasting in 5G [13]–[18].

One of the first works that evaluated the use of P-NOMA within eMBMS is reported in [13]. In this paper, several scenarios where LDM could take advantage of its efficiency increase were presented and the theoretical Signal-to-Noise Ratio (SNR) thresholds for different injection level (IL) values were provided. Some other works evaluated the theoretical interoperability of P-NOMA with the eMBMS ecosystem [14], [15]. While [14] studies the use of LDM as a key technology for 5G networks to deliver PTM transmission within the broadband systems, the authors in [15] present NOMA as a candidate to carry out multicast transmissions in subgroup oriented communications.

Concerning non-theoretical contributions, in [16], the transmission of unicast and broadcast services at the same time in a SFN by using LDM were analyzed. In this case, the study was based on measuring the power consumption per base station. The results provide power savings simultaneously for unicast and broadcast transmissions due to the larger bandwidth available. In [17], P-NOMA was proposed in combination with TDM schemes to deliver mixed broadcast and unicast services in a 5G-MBMS environment. The performance and the implications of using P-NOMA for both services, broadcast and unicast, were presented and evaluated. Results indicate that significant capacity gains could be achieved when using P-NOMA in 5G NR. However, the only Key Performance Indicator (KPI) evaluated is capacity, while latency and reliability are not taken into account. Finally, in [18], the first NOMA-based 5G NR PHY layer transceiver was presented. The prototype was tested by implementing a convergence-oriented use case. Results offer significant capacity gains in different configurations. However, the AWGN channel was the only simulation case and the impact of retransmissions (Hybrid Automatic Repeat Request, HARQ) was not evaluated.

III. 5G Overview

As in 4G, 5G NR PHY is based on Orthogonal Frequency Division Multiplexing (OFDM) technique. However, one of the main novelties of this standard is not the base technology, but its configuration flexibility. There are five different numerologies (μ) available with different subcarrier spacing (SCS) values obtained from the following expression:

$$SCS(kHz) = 15 \cdot 2^{\mu} \quad (1)$$

where μ represents the numerology and it can be 0, 1, 2, 3 or 4.

In NR, each radio frame has a fixed length of 10 ms and it is composed of a fixed number of ten subframes. Moreover, each subframe has a duration of one millisecond. The number of slots per subframe varies depending on the numerology that is implemented. In TABLE I, a summary of the possible number of slots per subframes and their duration is presented as a function of the numerology. In comparison with 4G, this aspect is considered a relevant difference. In addition, although the default number of symbols per slot is 14, different configurations can be implemented. On the one hand, in order to enable the transmission of short data packet, a mini-slot configuration has been introduced in the standard, where 2, 4 or 7 symbols can be allocated. On the other hand, just with the opposite aim, slots can be gathered for longer transmissions.

TABLE I
5G NR Subcarrier Spacing Influence

μ	SCS (kHz)	# Slots/Subframe	Slot (ms)
0	15	1	1
1	30	2	0.5
2	60	4	0.25
3	120	8	0.125
4	240	16	0.0625

TABLE II
OFDM Symbols

	$\mu = 0$	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$
OFDM Symbol (μs)	66.67	33.33	16.67	8.33	4.17
CP (μs)	4.69	2.34	1.17	0.57	0.29
Total length (μs)	71.36	35.67	17.84	8.90	4.46

Another variable that can be tuned is the OFDM symbol length. TABLE II presents the OFDM symbol lengths with and without the Cyclic Prefix (CP). As shown, the shortest OFDM symbol lengths allow their transmission in less than 5μs. In consequence, the highest numerologies, especially $\mu = 3$ and $\mu = 4$, are designed to address URLLC by using shorter transmission durations.

Finally, the main PHY level novelty from the reliability point of view is the new channel coding. NR does not include Turbo-codes; LDPC and Polar codes are supported. Polar codes are reserved for control packet transmissions and LDPC codes are used in payload data packets. Indeed, LDPC codes have demonstrated to be a very robust channel coding technique and their reliability is very close to the Shannon limit, just half a dB away [19], [20].

Regarding the MAC layer implemented in NR, several characteristics have been directly inherited from LTE: the mapping process between logical and transport channels, the multiplexing/demultiplexing of MAC Service Data Units (SDU) when required, etc. Moreover, as in LTE, HARQ techniques are used for error correction. 5G NR HARQ are error correction techniques based on packet retransmissions. The main difference when compared with 4G is that the timing between data transmission and the HARQ response is flexible. While in 4G timing was fixed to 4 ms, in 5G the number of time slots between data transmission and the HARQ response can be dynamically adapted for enabling lower latencies.

Another aspect that has been modified in NR is the resource management. As shown in Fig. 2, the smallest unit that appears in the resource grid is the Resource Element (RE), which consists of one subcarrier and one OFDM symbol, in frequency and time domain, respectively. However, the Resource Block (RB) concept is not the same as in 4G. In NR, RBs are defined as 12 consecutive subcarriers in one OFDM symbol. Then, as presented in TABLE III, the minimum and

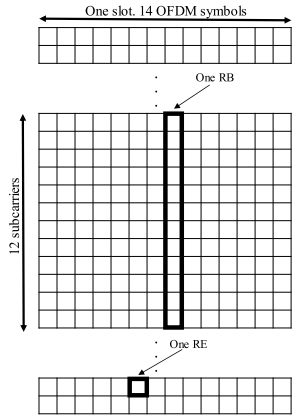


Fig. 2. Basic resource management terminologies.

TABLE III
RESOURCE BLOCKS AND BANDWIDTH SIZES

μ	RB_{MIN}	RB_{MAX}	BW_{MIN} (MHz)	BW_{MAX} (MHz)
0	24	275	4.32	49.5
1	24	275	8.64	99
2	24	275	17.28	198
3	24	275	34.56	396
4	24	138	69.12	397.44

maximum bandwidths were defined for each of the possible numerologies (from 4.32 MHz to 397.44 MHz). Finally, the amount of RBs required to obtain the bandwidth ranges was calculated according to each numerology and bandwidth range by dividing the bandwidth by the size of one RB.

IV. 5G-NOMA

In this section, the implications of introducing NOMA in the 5G PHY are presented. First, the general structure is introduced and then, several issues regarding the implementation are explained.

A. General Structure

In the first phase of this work, a transceiver has been developed in MATLAB. This prototype is partially compliant with the NR Rel-15 standard. A Physical Downlink Shared Channel (PDSCH) has been implemented in the prototype, since 5G NR does not include any specific broadcast channel for data delivery. Nevertheless, the prototype is compatible with any future broadcast channel that could be introduced in the future (Rel-17/18). In addition, in order to introduce NOMA in the 5G NR transceiver some modifications have been done. In Fig. 3, a general block diagram of the overall transceiver is shown. The blocks related to the transmitter are color coded in orange, while receiver blocks are painted in blue.

At the beginning of the transceiver, two independent data flows are created, one per each NOMA layer. The blocks related to UL are located inside the green dashed box, while the blocks related to the LL are inside the red dashed box. A Downlink Shared Channel (DL-SCH) transport channel is generated for each layer, and then, the DL-SCH is encapsulated into PDSCH, where the modulation and the LDPC coding is applied. The LDPC codeword size is adapted depending on the modulation and code rate that is used.

Once PDSCH are generated for both layers, they are combined into a single NOMA signal ensemble. Therefore, symbols corresponding to the LL are attenuated by a predefined IL (g, in Fig. 3, expressed in linear units). The IL indicates the relative power distribution between both layers, so g has a real value within the range [0, 1), where $g = 0$ results in a single-layer system and $g = 1$ results in a two-layer system with equal power distribution [4]. The IL can also be expressed in dBs (IL = 10·log(g)). Then, the UL and the attenuated LL are added creating the definitive NOMA signal ensemble. After superimposing both layers, the output constellation is normalized. Finally, the precoding matrix for the next transmission is calculated using singular value decomposition (SVD). Then, the OFDM signal is generated.

The transmitted signal is filtered through the channel model block. In this case, two channel model families have been integrated: an ideal channel-fading model with Additive White Gaussian Noise (AWGN) and a Tapped Delay Line (TDL) [21]. The first will be used as a reference to test the prototype and define the PHY configurations, while the second one is implemented as a more realistic and challenging channel model. TDL channel models are defined for the full frequency range and for non-MIMO transmissions. They are composed by several taps with different delays and amplitudes. Five different channel profiles are defined, where TDL-A, TDL-B and TDL-C represent channel models for Non-Line of Sight (NLOS) conditions and TDL-D and TDL-E are aimed for Line of Sight (LOS) conditions [21]. More details are presented about the parameters of the implemented channel models in Section V.

At the receiver, the first modules perform OFDM demodulation and channel estimation. The channel estimation block also removes the precoding information form the channel estimation. Then, for UL recovery, demodulation and LDPC decoding is applied at PSDCH and DL-SCH blocks. The UL noise threshold is low and in consequence, the lower layer will be assumed as additional noise for the UL. In order to recover the LL, a Successive Interference Cancellation (SIC) module has been introduced. Since this module takes as input the received signal and the regenerated UL, the DL-SCH and the PDSCH modules have to be applied again to re-create UL bits. The output of the SIC module is the difference between both signals, i.e., the received LL. Hence, to obtain the definitive transmitted LL bits, PDSCH decoding and DL-SCH decoding is applied to the output of the SIC module.

In parallel to the complete 5G-NOMA architecture, HARQ retransmission schemes are also implemented. In this case, when the UL data and the LL data are obtained in the DL-SCH decoding module, received packets are checked. If the packet is received successfully, a new transport block is generated and transmitted. However, if errors occur, a retransmission is scheduled by changing the Redundancy Version (RV) value.

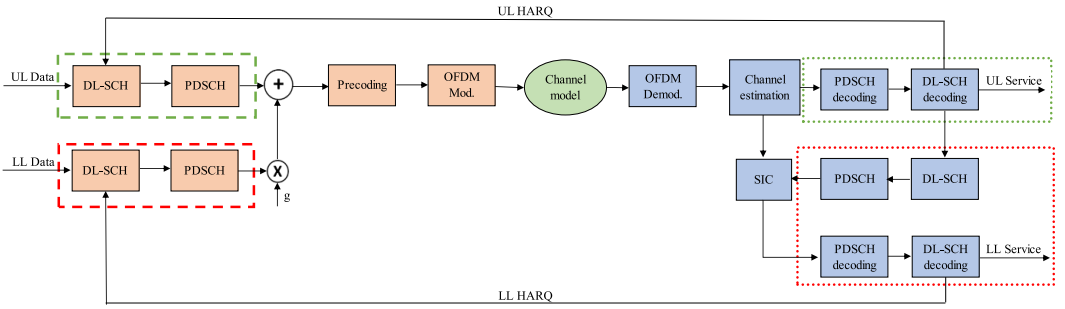


Fig. 3. Simplified general block diagram of the transceiver architecture.

In the proposed 5G NOMA implementation, HARQ processes are independently carried out in each layer.

Finally, it is important to highlight complexity aspects. It is true that introducing NOMA techniques in the 5G NR standard, the complexity of the transceiver will be increased. However, this complexity drawback does not affect the complexity of receivers only intended for the UL decoding. In fact, as it can be seen in Fig. 3, the UL decoding chain follows the traditional modules. Therefore, only the receivers interested on recovering LL suffer from extra complexity increase associated to the SIC module.

B. Frame Adaptation Layer

Although UL and LL data flows are independently configured, a joint framing management is needed. Indeed, when the NOMA signal ensemble is created, the inputs to both layers must have the same length. That is why, a new adaptation layer between the MAC and the PHY is required. The frame adaptation layer is based on the following expression:

$$\#SC_{UL} = \#SC_{LL} \tag{2}$$

where $\#SC_{UL}$ and $\#SC_{LL}$ are the number of subcarriers allocated for the transmission in the UL and in the LL, respectively. This parameter defines the amount of coded bits that can be transmitted:

$$\#\ SC_{UL} = \frac{tx\ bits_{UL}}{m_{UL}} = \frac{tx\ bits_{LL}}{m_{LL}} \tag{3}$$

where *tx bits* is the amount of bits that are introduced in the modulator and m is the modulation index. Finally, the amount of transmitted bits is calculated from the following relation:

$$\#\ SC_{UL} = \frac{info\ bits_{UL} \cdot CR_{UL}}{m_{UL}} = \frac{info\ bits_{LL} \cdot CR_{LL}}{m_{LL}} \tag{4}$$

where *info bits* are the PHY transport block bits and *CR* is the implemented code rate. Assuming a fixed amount of information bits for the UL, the amount of information bits for the LL in each frame is calculated from the following expression:

$$info\ bits_{LL} = \frac{info\ bits_{UL} \cdot CR_{UL} \cdot m_{LL}}{m_{UL} \cdot CR_{LL}} \tag{5}$$

Therefore, the length of the LL packet received from the MAC layer is restricted by the combination of the modulation order and the code rate (Modulation and Coding Scheme, MCS) of each layer and the length of the UL MAC packet.

Simulations use complete time slots so the number of subcarriers allocated for each iteration will be driven by SCS and bandwidth choices. In practice, the length of MAC packets will be then set by defining modulation and code rates of UL and LL.

C. Resource Allocation

As described in Section III, one RB is the minimum data unit that can be allocated in 5G NR. Indeed, this resource allocation technique implies frequency domain multiplexing (subcarriers) and time domain multiplexing if more than one RB is gathered. Therefore, P-NOMA can be considered a two dimensional multiplexing technique. In Fig. 4(a), an example of NR Rel-15 resource allocation pattern is shown. Each color represents the one different service or unicast receiver. However, if P-NOMA is introduced in the 5G NR resource allocation mechanism, another parameter should be taken into account when distributing the resource: the transmitted power. In Fig. 4(b), an example of a resource allocation pattern when P-NOMA is implemented is shown. Again, each color represents a different transmitted service. In this case, inside each RB due to the IL, more than one service can be delivered. Therefore, if P-NOMA is introduced in 5G NR, a three dimensional resource allocation should be carried out: time, frequency and power. Undoubtedly, this third dimension increases the number of parameters and possible pattern combinations, and, therefore, higher flexibility when applying the resource allocation.

D. Broadcast/Unicast Convergence

The resource allocation approach presented in Section IV-C presents a new alternative to cover different use cases. In fact, P-NOMA would enable the convergence between broadcast and unicast services in the same RF channel. On the one hand, Fig. 4(c) shows how the broadcast/unicast convergence issue is solved in eMBMS. As it can be seen, broadcast and unicast services are multiplexed by using frequency domain multiplexing, although a similar allocation pattern can be

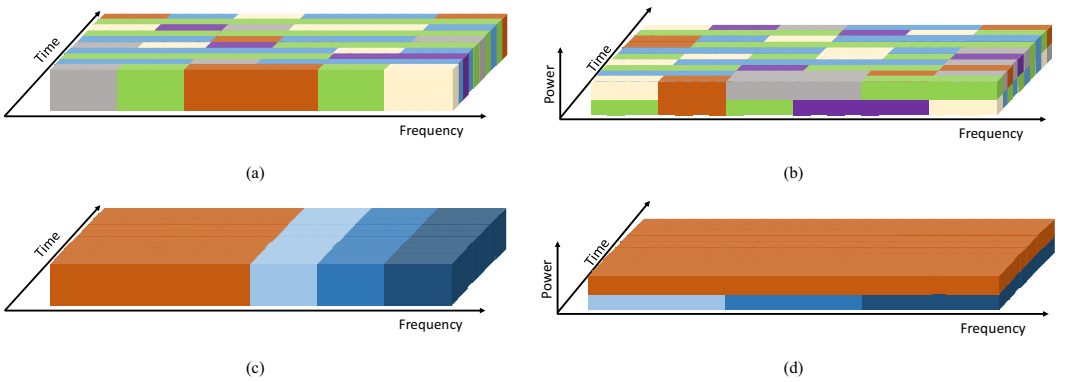


Fig. 4. Resource allocation patterns: (a) generic TDM/FDM, (b) generic P-NOMA, (c) broadcast/unicast convergence solution in TDM/FDM, (d) broadcast/unicast convergence solution in P-NOMA.

obtained if a time domain multiplexing is followed. The dark orange block represents the allocation for the broadcast service, whereas the blue blocks are reserved for different unicast services, where each blue tone represents a specific unicast service. In this case, a 50%-50% time division has been applied between broadcast and unicast services, which means that the available resources are equally shared between broadcast and unicast services. Moreover, concerning the unicast services, three simultaneous services have been assumed with an equal resource sharing among them.

On the other hand, in Fig. 4(d) the broadcast/unicast convergence issue is solved by implementing P-NOMA in the 5G NR resource allocation mechanism. As in Fig. 4(c), dark orange block represents the broadcast service and the blue blocks are different unicast services. In this case, broadcast and unicast services are sharing the same OFDM symbol and the same subcarriers, but they are superimposed with different power levels. The UL is reserved for the broadcast content, whereas the LL is used for delivering unicast services, since traditional broadcast transmissions have to guarantee a high reception rate and so, they have to meet more strict availability requirements. Then, in the LL in order to share the resources among the different unicast services frequency domain multiplexing is used. Therefore, the resource allocation scheme presented in Fig. 4(d), as in Fig. 4(b), implies LDM + FDM (or TDM) multiplexing scheme concatenation. Comparing Fig. 4(c) and 4(d), the latter certainly provides wider available bandwidth for each service, resulting in capacity gain.

When P-NOMA is used to guarantee the convergence between broadcast and unicast services, the main parameter to configure the overall content distribution is the IL. Therefore, the implemented IL value has a direct impact in the delivered services and so, a tradeoff has to be assumed between the characteristics of broadcast and unicast services. For example, if high IL values are selected, higher throughput can be obtained in the UL (or higher reliability) for the broadcast service. However, as the power reserved for the LL is lower, the available capacity will decrease, and less unicast services could be delivered. On the other hand, if low ILs are selected in order to enhance the unicast services, the reliability of the UL will be reduced and the broadcast coverage area could be noticeably impacted.

V. PHY Evaluation

This section describes the evaluation of the PHY P-NOMA solution in different propagation channels.

A. Theoretical Approach

A use case with a mixture of broadcast and unicast services has been designed. The broadcast service is 4K UHDTV content delivered on a 30 Mbps bit stream [22]. Availability will be considered a critical factor because correct reception has to be guaranteed to 99% of the receivers at least. On the other hand, in the case of the unicast services, configurations that allow higher data rates have been used. The reference value for unicast services has been set to 50 Mbps, which has to be shared among all the unicast receivers. For this kind of services, the higher the available data rate is, the higher the amount of users that could simultaneously receive the service can be. Some examples of this type of services are: Internet of Things (IoT) for smart traffic management or on demand TV content for mobile receivers.

From a multiplexing standpoint, P-NOMA offers a new dimension for system flexibility and the capacity and robustness requirements posed by each service can be tailored with different time, frequency and power domain adjustments. Examples shown in TABLE IV explain how broadcast services would be delivered in the UL and unicast services in the LL. The values in TABLE IV provide comparison metrics between the expected capacities and thresholds in P-NOMA and TDMA. Capacity is calculated for a NR signal of 20 MHz bandwidth and SNR thresholds are based on AWGN simulations.

The methodology for comparing P-NOMA and TDMA starts setting the service bitrate requirements in TDMA (30 Mbps UHDTV Broadcast content and 50 Mbps

TABLE IV
USE CASES CONFIGURATION

Service	TDMA (50%-50%)			NOMA (I) (IL = -6 dB)			NOMA (II) (IL = -6 dB)		
	MCS	Cap. (Mbps)	SNR_{TH} (dB)	MCS	Cap. (Mbps)	SNR_{TH} (dB)	MCS	Cap. (Mbps)	SNR_{TH} (dB)
Broadcast	12	30.3	11.1	5	29.5	11	5	29.5	11
Unicast	19	51.2	17.5	10	51.4	15.8	12	60.6	18

unicast capacity). The TDMA choices are MCS 12 (64QAM 517/1024 and 3.0293 bps/Hz) and MCS 19 (64QAM 873/1024 and 5.1152 bps/Hz) (according to [23, Table 5.1.3.1-2]).

Then, P-NOMA is adjusted to provide equivalent UL bitrate and robustness (4K UHDTV content). The UL MCS selection is MCS5 (16QAM 378/1024 and 1.4766 bps/Hz) while two possible variants are proposed for the lower layer. NOMA (I) optimizes the system threshold while maintaining capacity (51.4 Mbps vs. 51.2 Mbps) using a LL MCS10 (16QAM 658/1024 and 2.5703 bps/Hz). NOMA (I) configuration provides a 1.7 dB improvement in the LL with respect to the equivalent TDMA threshold. A second optimization case is NOMA (II) that maintains an equivalent threshold (18 dB) with MCS12 (64QAM 517/1024 and 3.0293 bps/Hz) and varies capacity (60.6 Mbps vs 51.2 Mbps). The injection level in P-NOMA has been set to -6 dB to guarantee that the service area of both systems is equivalent.

B. Simulation Results

Once a preliminary analysis based on the AWGN channel has been carried out, more simulations have been performed over different propagation channels in order to widen the results.

Concerning the wireless propagation channel, TDL-D and TDL-E profiles have been implemented. In both cases, the LOS condition is included. Additionally, three different desired delay spread (DDS) values have been simulated for each channel model: 90 ns, 360 ns and 1100 ns. Those values have been selected because they are aligned with short-, normal- and long-delay profiles for an urban environment [21].

Results obtained with the developed transceiver over the detailed channels are presented in Fig. 5. Following those graphics, a performance comparison can be made between P-NOMA and TDMA multiplexing schemes.

From a general perspective, each transmission configuration shows a similar performance over the different channels. In comparison with the AWGN channel, the performance is deteriorated in the range of 0.3 dB and 2 dB. The short-delay and the normal-delay profiles, present very similar results in both cases. It should be highlighted, however, that the long-delay profile channels have stronger negative effects on the received signal and the required SNR thresholds are higher. In fact, the TDL-E channel model is even more challenging in long-delay conditions than the TDL-D.

As expected, services oriented to fixed receivers (TDMA (I) and NOMA (UL)) show a very similar reliability performance. In fact, the difference between both cases, in terms of required SNR threshold for a BER value of 10^{-6}, is around 0.1 dB. However, the behavior of the BER vs SNR curve is not identical in both cases. While the TDMA case presents a very pronounced decrease, the UL case has a slower decrease. This difference is due to the NOMA configuration. The effective SNR value for the UL is close to 6 dB, due to the IL. This value is very close to the SNR threshold required to decode this configuration in the single layer mode and, in consequence, the UL is not able to offer a fast decrease. Nevertheless, this problem could be solved by increasing the IL value. In that case, the SNR value required to obtain the LL service would be higher.

TABLE V
RELIABILITY/CAPACITY GAIN ANALYSIS OF THE LL (UNICAST SERVICES)

Metric		NOMA (I)	NOMA (II)
SNR_{TH} Gain (dB)	TDL-D 90 ns	1.7	-0.4
	TDL-E 90 ns	1.7	-0.6
	TDL-D 360 ns	1.8	-0.5
	TDL-E 360 ns	1.7	-0.4
	TDL-D 1100 ns	1.8	-0.4
	TDL-E 1100 ns	2.0	-0.2
Capacity Gain (Mbps)		0.2	9.4
Gain in IoT case (Users)		1	47
Gain in Mobile Video case (Users)		0	5

Regarding the performance in the case of mobile receivers (TDMA (II) and LL), the advantages of the different configurations presented in TABLE IV are confirmed. In TABLE V, an in-depth analysis is carried out in terms of capacity and reliability for the unicast services, which are delivered in the LL in the case of P-NOMA. Each of the metrics presented in TABLE V, demonstrates the gain of P-NOMA over the TDMA configuration. The SNR threshold gain has very little variation over the different channels. As expected, the first P-NOMA configuration presents a reliability gain of 1.7 dB for the same capacity value. The second case offers 9.4 Mbps of capacity gain by assuming a similar reliability performance (0.6 dB of deterioration on the worst case). Overall, these P-NOMA cases present a wide range of transmission configurations.

Last rows of TABLE V also present a capacity gain comparison, but in this case in terms of satisfied users for different potential P-NOMA use cases: IoT services and mobile video delivery. For the first one, the characteristics defined in the Rel-13 (Narrow Band IoT LTE) have been taken into account, where a maximum data rate of 200 kbps per user in downlink channels is allowed [24]. For the second use case, where unicast users will receive HD video content in their mobile or portable devices, a minimum data rate of 2 Mbps has been assumed [25]. From the capacity calculations presented in TABLE V, while TDMA supports 256 unicast

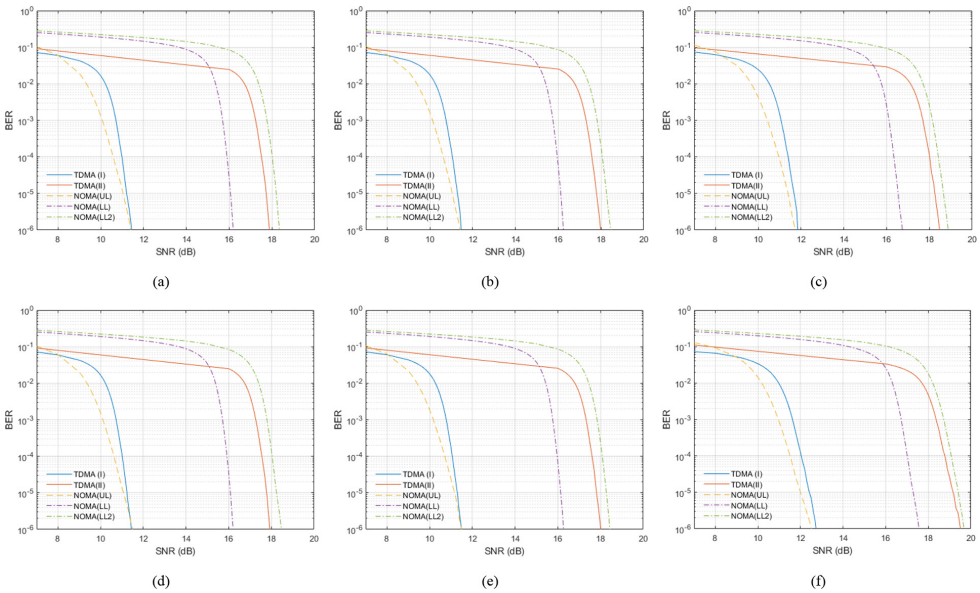


Fig. 5. Performance BER vs. SNR curves for different channel models: (a) TDL-D with 90 ns DDS, (b) TDL-D with 360 ns DDS, (c) TDL-D with 1100 ns DDS, (d) TDL-E with 90 ns DDS, (e) TDL-E with 360 ns DDS, (f) TDL-E with 1100 ns DDS.

users connected in the IoT case and 25 in the mobile video content delivery case, P-NOMA configurations increase considerably the amount of served users with up to 47 IoT devices and 5 mobile video devices.

VI. Network Simulation

The objective of this section is to test the performance of the proposed solution including MAC layer techniques and compare the performance with equivalent traditional TDMA solutions.

A. Simulation Methodology

The system evaluation has been carried out in an upgraded OMNeT++ [26]. This network simulation tool is composed of both PHY and MAC layers for each of the implemented transmitters and receivers. On the one hand, the PHY layer module facilitates the selection of the parameters that define the waveform. In addition, the transmission parameters have to be configured (modulation and code rate). PHY performance curves presented in the previous sections are used to estimate the transmission success. For each one of the received packets, the current SNR value of the receiver is used to determine the packet error rate (PER) according to the performance curves. The conversion from BER to PER is carried out following the technique presented in [27]. On the other hand, concerning the MAC layer, HARQ type retransmission schemes are implemented.

Regarding the network architecture and the implemented scenario, the channel model module and the mobility module are the most relevant parts of the simulated model. The former is in charge of defining the characteristics of the channel model, such as the Doppler frequency, the delay spread value or the implemented channel profile (TDL-D or TDL-E). Indeed, this module is closely related with the user type, since, while the channel for fixed receiver is considered almost constant, the channel affecting the receiver with the highest mobility will show a fastest change. This concept is also taken into account by the mobility model module, which takes care of detailing how the receiver is moving. Different mobility models are available in OMNeT++ for different type of receivers.

B. Network Setup

Three groups of users have been defined: fixed receivers, low-speed pedestrian receivers and car-mounted receivers. The first group of users represents traditional broadcast receivers, based on roof top antenna reception. However, pedestrian and car-mounted receivers will typically request more challenging data applications, e.g., on-demand video content. Therefore, this network architecture emulates a real scenario where a convergence application is required in order to guarantee the correct reception of broadcast and different unicast services.

In each simulation, a total number of 100 devices have been included. Specifically, 75 fixed receivers have been assumed inside each cell, 15 pedestrians and 10 car-mounted user devices. Fixed receivers have a static model, whereas pedestrians have been allocated a concrete walker mobility type referred as Random Way Point (RWP) [28]. In the case of cars, considering the characteristics of city roads, a linear mobility type has been used [29]. Speed is also specified for each

TABLE VI
SIMULATION PARAMETERS

Parameter	Value
Center Frequency	2 GHz
Tx Power	44 dBm
SCS	15 kHz
Distance attenuation	128.1 + 37.6*log(d)
Type of nodes	Fixed, pedestrians, cars
Speed	Fixed: 0 km/h Pedestrians: 3 km/h Cars: 30, 50 km/h
Number of Users	Fixed: 75 Pedestrians: 15 Cars: 10
Mobility type	Fixed: Static Pedestrians: RWP Cars: Linear
Inter-Site Distance (ISD)	500 m
Tx configuration (TDMA)	Broadcast: MCS12 Unicast: MCS19 Time division: 50%-50%
Tx configuration (NOMA)	Broadcast: MCS5 Unicast: MCS10, 12 IL: -6 dB
Desired Delay Spread	90, 360, 1100 ns
Noise Power	-90 dBm
Noise Figure	9 dB
Channel model	TDL-D, TDL-E
Simulated seeds	25
Simulation time	5 min per seed

receiver. Pedestrian receivers walk inside the cell with a mean speed of 3 km/h, and the car receivers vary their speed between 30-50 km/h.

The parameters values are critical when defining the propagation attenuation suffered by the transmitted signal. In this case, all the propagation channels presented in Fig. 5, based on TDL-D and TDL-E, have been considered with specific Doppler frequency values depending on the speed of the receiver.

The rest of the network simulation parameters are gathered in TABLE VI.

C. Results

The evaluation of the proposed NOMA-based 5G model has been based on three KPIs: reliability, latency and throughput. First, to measure the reliability, PER is used as a metric. Second, the user plane latency is evaluated. According to [30], user plane latency is defined as the required time to deliver a data packet between the Next Generation NodeB (gNB) and the user equipment (UE). The user plane latency can be modelled as:

$$T_{UP} = \tau_1 + p(\tau_2 + \tau_3) \tag{6}$$

where, p is the probability of requiring a retransmission, τ_1 is the transmission time, τ_2 is the HARQ request process time and τ_3 is the retransmission time. The complete process to calculate τ_1, τ_2 and τ_3 is obtained from [31]. The user plane latency is defined by the processing time at the UE ($t_{UE,rx}$). P-NOMA increases latency because the UE processing time will be higher (SIC module). In this case, following the complexity analysis of P-NOMA receivers carried out in [5], it is assumed that the $t_{UE,rx}$ increase in a two-layer system is 10% in comparison with a LL only case. Therefore, the minimum user plane latency that can be obtained is lower in the TDMA configuration than in the NOMA configuration, 1.785 ms vs. 1.796 ms, respectively. Moreover, the user plane latency deviation is also used in order to measure the latency variation with respect to the achievable minimum latency value. Finally, for the throughput, two metrics are used: normalized data reception index (NDRI) and effective throughput. The NDRI indicates the relation between the amount of information received and the maximum information rate that could be received in the error free case. Moreover, not only error packets reduce the NDRI value, also the delivered retransmissions, since using retransmissions, the overall received capacity is decreased because the same content is repeatedly transmitted. This parameter is measured in percentage data. The following expression has been used to calculate the NDRI:

$$NDRI\ (\%) = \frac{correct\ packets}{correct\ packets + retx\ attempts} \cdot 100 \tag{7}$$

where, *retx attempts* are the total number of retransmissions that have been carried out for each node. On the other hand, the effective throughput represents the useful received capacity. This parameter also takes into account the number of retransmissions that have been implemented by combing the NDRI with the transmitted capacity:

$$eff.throughput\ (Mbps) = \frac{NDRI \cdot tx\ capacity}{100} \tag{8}$$

In TABLE VII, a dual comparison of reliability and latency is presented between P-NOMA and TDMA configurations. Fixed receivers (UL) show equal PER values and user plane latency is also the same as no retransmissions are available for those users. User plane latency is the same for TDMA and NOMA cases, since being the service a broadcast type service, there is no uplink to transmit the feedback and, so, retransmissions cannot be implemented. In the case of mobile receivers, in general, the best reliability results are obtained in the NOMA (I) case, which is designed to improve the reliability. PER is reduced by 40% with respect to TDMA. As expected, NOMA (II) presents worse PER values than TDMA configuration. If the reliability is compared between different groups of users, the best results are for fixed receivers because they require lower SNR thresholds. In the case of mobile receivers, car-mounted ones show better behavior for their higher Doppler variation, that will make retransmissions more effective.

In TABLE VII, user plane latency values and deviations are presented. Although NOMA-based receivers have to cope with a latency penalization due to the SIC module, the lowest latency values appear in NOMA (I) because the reliability increase that the configuration offers compensates the latency penalization. In addition, user plane latency values are below 4 ms, which is the user plane latency requirement defined in [31] for eMBB applications. Therefore, all the presented cases meet the latency requirements. Moreover, car receivers have lower user plane latency values than pedestrians. High latency values indicate that more retransmissions

TABLE VII
NETWORK LEVEL RELIABILITY/LATENCY ANALYSIS

User Type	PER			User Plane Latency (ms)			User Plane Latency Deviation (μs)		
	TDMA	NOMA (I)	NOMA (II)	TDMA	NOMA (I)	NOMA (II)	TDMA	NOMA (I)	NOMA (II)
Fixed (UL)	$3.75 \cdot 10^{-5}$	$3.75 \cdot 10^{-5}$	$3.75 \cdot 10^{-5}$	1.785	1.785	1.785	-	-	-
Pedestrian (LL)	$6.13 \cdot 10^{-2}$	$3.66 \cdot 10^{-2}$	$7.14 \cdot 10^{-2}$	1.908	1.870	1.940	122.519	73.473	143.477
Car (LL)	$2.98 \cdot 10^{-2}$	$1.75 \cdot 10^{-2}$	$3.47 \cdot 10^{-2}$	1.845	1.831	1.866	59.620	35.193	69.737
Total	$1.22 \cdot 10^{-2}$	$7.26 \cdot 10^{-3}$	$1.42 \cdot 10^{-2}$	1.883	1.854	1.910	97.360	58.161	113.981

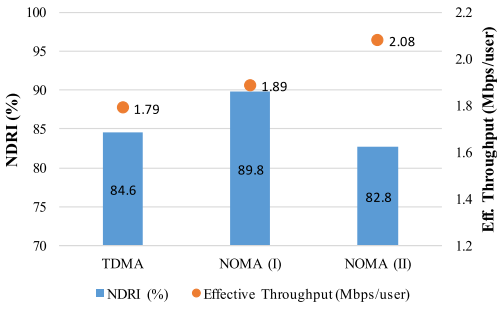


Fig. 6. Network level throughput analysis of the LL (unicast services).

are needed for a proper decoding than in cases with low latency. Therefore, latency results confirm that retransmissions are more effective in the case of cars, while in the case of pedestrians have lower effect. Regarding user plane latency deviation results, NOMA (I) case presents the lowest values, which indicates that the calculated user plane latency is closer to minimum than in any other case. This effect makes NOMA (I) even more interesting as it provides higher degree of determinism.

Fig. 6 shows only mobile receiver throughput (fixed receivers show identical behavior). Looking at the NDRI, the results are equivalent to PER values. NOMA (I) outperforms NOMA (II) and TDMA. If effective throughput is considered, TDMA offers the worst behavior (below 1.8 Mbps/user), while NOMA cases increment the effective throughput when increasing the implemented MCS. The difference between TDMA and the best case (NOMA (II)) is an increment of 14%. These results indicate that the reliability decrease is compensated by transmitting with a higher capacity configuration.

VII. CONCLUSION

This paper proposes a PHY/MAC enabler solution for broadcast and unicast convergence in 5G NR. The solution is based on Power domain Non Orthogonal Multiple Access (P-NOMA). The system configuration parameters have been analyzed, and different configuration options have been discussed. For instance, the IL has been identified as a key configuration parameter for the broadcast/unicast convergence use case since introduces an additional multiplexing variable that provides a wider flexibility range. This flexibility has been used to increase the reliability of the unicast receivers and to deliver higher data rate to the unicast receivers.

A system transceiver prototype has been design and implemented for testing purposes. Two different evaluation procedures have been developed to obtain the system performance from various perspectives. In the PHY layer analysis, different P-NOMA-based configurations have been defined in order to cover a specific service convergence use case, namely NOMA (I) and NOMA (II). Based on these configurations, possible reliability and/or capacity gains were foreseen first, and then, confirmed by using different AWGN and TDL channel models. In the case of the capacity gain, depending on the configuration used, up to 47 IoT receivers and 5 mobile video receivers could be added to the network. The system level performance has been analyzed using a network simulation environment. Reliability, latency and throughput have been evaluated by using different metrics. Reliability and latency measurements indicate that the most appropriate configuration for those terms for the unicast services is NOMA (I). However, from the capacity perspective, NOMA (II) is the best solution because it offers the highest capacity per user rate. Moreover, HARQ retransmission schemes appear to be more effective when the receivers are moving faster, due to the higher channel response variance. All in all, based on the obtained results, NOMA can be considered as a competitive alternative broadcast/unicast convergence.

There are still some future works on the horizon. On the one hand, concerning the stack protocol, upper layers should be taken into account for future developments. For example, if the network layer is included in addition to the proposed service level convergence, IP-based convergence could also be proposed. In this case, 5G NR could be combined with another technology, such as ATSC 3.0 or DVB-T2. On the other hand, as highlighted in Section V-B, one of the proposed NOMA configurations provides a considerable reliability gain that could be transformed in coverage area improvement. Therefore, similar configurations could be used in rural areas or isolated small areas presenting limitations on the service quality. Finally, new resource management strategies could be combined with this solution. Interesting candidate alternatives are subgrouping techniques based on the architecture proposed in [15].

REFERENCES

[1] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021*, Cisco, San Jose, CA, USA, 2017.
[2] *NR; Physical Layer; General Description, Technical Specification Group Services and System Aspects V15.0.0*, Standard TS 38.201, Jan. 2018.
[3] "IMT vision—Framework and overall objectives of the future development of IMT for 2020 and beyond," ITU, Geneva, Switzerland, ITU Recommendation M.2083-0, Sep. 2015.
[4] L. Zhang *et al.*, "Layered-division-multiplexing: Theory and practice," in *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 216–232, Mar. 2016.
[5] S. I. Park *et al.*, "Low complexity layered division multiplexing for ATSC 3.0," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 233–243, Mar. 2016.

[6] L. Fanari *et al.*, "Trends and challenges in broadcast and broadband convergence," in *Proc. IEEE Int. Conf. Elect. Eng. Photon. (EExPolytech)*, St Petersburg, Russia, 2019, pp. 153–156.
[7] D. Lecompte and F. Gabin, "Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: Overview and Rel-11 enhancements," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 68–74, Nov. 2012.
[8] C. Hoymann *et al.*, "LTE release 14 outlook," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 44–49, Jun. 2016.
[9] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5G evolution: A view on 5G cellular technology beyond 3GPP release 15," *IEEE Access*, vol. 7, pp. 127639–127651, 2019.
[10] D. Ratkaj and A. Murphy, Eds., *Definition of Use Cases, Requirements and KPIs, Deliverable D2.1*, 5G-PPP 5G-Xcast project, Oct. 2017.
[11] J. J. Gimenez *et al.*, "5G new radio for terrestrial broadcast: A forward-looking approach for NR-MBMS," *IEEE Trans. Broadcast.*, vol. 65, no. 2, pp. 356–368, Jun. 2019.
[12] D. Gomez-Barquero, D. Navratil, S. Appleby, and M. Stagg, "Point-to-multipoint communication enablers for the fifth generation of wireless systems," *IEEE Commun. Stand. Mag.*, vol. 2, no. 1, pp. 53–59, Mar. 2018.
[13] L. Zhang, Y. Wu, W. Li, K. Salehian, A. Florea, and G. K. Walker, "Improving LTE eMBMS system spectrum efficiency and service quality using channel bonding, non-orthogonal multiplexing and SFN," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Nara, Japan, 2016, pp. 1–8.
[14] L. Zhang *et al.*, "Layered-division multiplexing: An enabling technology for multicast/broadcast service delivery in 5G," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 82–90, Mar. 2018.
[15] J. Montalban *et al.*, "Multimedia multicast services in 5G networks: Subgrouping and non-orthogonal multiple access techniques," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 91–95, Mar. 2018.
[16] J. Zhao, D. Gündüz, O. Simeone, and D. Gómez-Barquero, "Non-orthogonal unicast and broadcast transmission via joint beamforming and LDM in cellular networks," *IEEE Trans. Broadcast.*, submitted for publication.
[17] L. Zhang *et al.*, "Using non-orthogonal multiplexing in 5G-MBMS to achieve broadband-broadcast convergence with high spectral efficiency," *IEEE Trans. Broadcast.*, vol. 66, no. 2, Jun. 2020.
[18] E. Iradier, J. Montalban, D. Romero, Y. Wu, L. Zhang, and W. Li, "NOMA based 5G NR for PTM communications," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, 2019, pp. 1–6.
[19] L. Michael and D. Gómez-Barquero, "Bit-interleaved coded modulation (BICM) for ATSC 3.0," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 181–188, Mar. 2016.
[20] S. Ahn, K. Kim, S. Myung, S. Park, and K. Yang, "Comparison of low-density parity-check codes in ATSC 3.0 and 5G standards," *IEEE Trans. Broadcast.*, vol. 65, no. 3, pp. 489–495, Sep. 2019.
[21] "Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP, Sophia Antipolis, France, Rep. TR 38.901, 2017.
[22] *Service Requirements for the 5G System; Stage 1, V15.7.0, Release. 15, Technical Specification Group Services and System Aspects*, Standard 3GPP TS 22.261, Dec. 2018.
[23] *NR; Physical Layer Procedures for Data (Release 15), V15.3.0, Technical Specification Group Services and System Aspects*, Standard 3GPP TS 38.214, Sep. 2018.
[24] S. Persia and L. Rea, "Next generation M2M cellular networks: LTE-MTC and NB-IoT capacity analysis for smart grids applications," in *Proc. AEIT Int. Annu. Conf. (AEIT)*, 2016, pp. 1–6.
[25] X. Xu, J. Liu, and X. Tao, "Mobile edge computing enhanced adaptive bitrate video delivery with joint cache and radio resource allocation," *IEEE Access*, vol. 5, pp. 16406–16415, 2017.
[26] A. Varga and R. Hornig, "An overview of the OMNET++ simulation environment," in *Proc. 1st Int. Conf. Simulat. Tools Tech. Commun. Netw. Syst. Workshops (ICST)*, 2008, p. 60.
[27] Z. Li *et al.*, "A cross-layer design for wireless VoIP: Playout delay constrained ARQ with ARQ aware adaptive playout buffer," in *Advances in Network and Communications Engineering*, 2004, p. 11.
[28] E. Hyytia, H. Koskinen, P. Lassila, and A. Penttinen, "Random waypoint model in wireless networks," in *Networks and Algorithms: Complexity in Physics and Computer Science*, Jun. 2005.
[29] O. G. Olaleye, A. Ali, D. Perkins, and M. Bayoumi, "Modeling and performance simulation of PULSE and MCMAC protocols in RFID-based IoT network using OMNeT++," in *Proc. IEEE Int. Conf. RFID (RFID)*, Orlando, FL, USA, 2018, pp. 1–5.
[30] "Study on scenarios and requirements for next generation access technologies, v15.0.0," 3GPP, Sophia Antipolis, France, Rep. TR 38.913, Jun. 2018.
[31] E. Garro *et al.*, "5G mixed mode: NR multicast-broadcast services," *IEEE Trans. Broadcast.*, vol. 66, no. 2, Jun. 2020.

Eneko Iradier (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications engineering from the University of the Basque Country (UPV/EHU) in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree. Since 2015, he has been part of the TSR Research Group, UPV/EHU. He was with the Communications Systems Group of IK4-Ikerlan as a Researcher from 2017 to 2018. During his doctoral studies, he did an internship at Communications Research Centre Canada, Ottawa. His current research interests include the design and development of new technologies for the physical layer of communication systems and broadcasting in 5G environments.

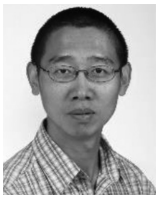
Jon Montalban (Member, IEEE) received the M.S. and Ph.D. degrees in telecommunications engineering from the University of the Basque Country, Spain, in 2009 and 2014, respectively. He is part of the Radiocommunications and Signal Processing Research Group, University of the Basque Country, where he is an Assistant Professor involved in several research projects. He has hold visiting research appointments with the Communication Research Centre, Canada, and Dublin City University, Ireland. His current research interests are in the area of wireless communications and signal processing for reliable industrial communications. He is the co-recipient of several best paper awards including the Scott Helt Memorial Award to Recognize the Best Paper Published in the IEEE TRANSACTIONS ON BROADCASTING in 2019. He has served as a reviewer for several renowned international journals and conferences in the area of wireless communications and currently serves as an Associate Editor for IEEE ACCESS.

Lorenzo Fanari (Student Member, IEEE) received the B.Sc. degree in electrical and electronic engineering and the M.Sc. degree in telecommunication engineering from the University of Cagliari, Italy, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the University of the Basque Country, Spain. His research interests include coding theory and wireless communications.

Pablo Angueira (Senior Member, IEEE), received the M.S. and Ph.D. degrees in telecommunication engineering from the University of the Basque Country, Spain, in 1997 and 2002 respectively.

He joined the Communications Engineering Department, University of the Basque Country in 1998, where he is currently a Full Professor. He is part of the staff of the Signal Processing and Radiocommunication Lab (http://www.ehu.es/tsr), where he has been involved in research on digital broadcasting (DVB-T, DRM, T-DAB, DVB-T2, DVB-NGH, and ATSC 3.0) for more than 20 years. He has coauthored an extensive list of papers in international peer-reviewed journals, and a large number of conference presentations in digital broadcasting. He has also coauthored several contributions to the ITU-R working groups WP6 and WP3. His main research interests are network planning and spectrum management applied to digital terrestrial broadcast technologies. He is currently involved in research activities related to broadcasting in a 5G environment.

Prof. Angueira is an Associate Editor of the IEEE TRANSACTIONS ON BROADCASTING, a member of the IEEE BMSB International Steering Committee and a Distinguished Lecturer of the IEEE Broadcast Technology Society. He serves on the Administrative Committee for the IEEE Broadcast Technology Society.



Liang Zhang (Senior Member, IEEE) received the bachelor's degree from the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China, in 1996, and the M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering, University of Ottawa, Ottawa, ON, Canada, in 1998 and 2002, respectively. He is a Senior Research Scientist with the Communications Research Centre Canada, Ottawa. He has been deeply involved in the ATSC 3.0 standardization activities on developing the layered-division-multiplexing technology, mixed fixed and mobile broadcast service delivery, mobile service detection, co-channel interference mitigation, integrated access and backhaul. He is currently working on technologies for the convergence of future TV broadcast and 5G broadband systems. He has more than 70 peer-reviewed Journal and conference publications and received the Multiple Best Paper Awards from IEEE, IBC, and NAB for his work on the Next Generation ATSC 3.0 and the 5G Broadcasting Systems. He is an Associate Editor of the IEEE TRANSACTIONS ON BROADCASTING and IEEE Broadcast Technology Society Distinguished Lecturer, and the elected member of the IEEE Broadcast Technology Society Administrative Committee.

Yiyan Wu (Fellow, IEEE) received the M.Eng. and Ph.D. degrees in electrical engineering from Carleton University, Ottawa, ON, Canada, in 1986 and 1990, respectively.

He is a Principal Research Scientist with the Communications Research Centre Canada. He is an Adjunct Professor with Carleton University and Western University, Canada. He has more than 400 publications. His research interests include broadband multimedia communications, digital broadcasting, and communication systems engineering. He received many technical awards for his contribution to the research and development of digital broadcasting and broadband multimedia communications. He is a Distinguished Lecturer of the IEEE Broadcast Technology Society, and a member of the ATSC Board of Directors, representing IEEE. He was appointed as a Member of the Order of Canada in 2018. He is a Fellow of the Canadian Academy of Engineering.

Wei Li (Member, IEEE) received the B.E. degree in electrical engineering from Shandong University in 1985, the M.S. degree in electrical engineering from the University of Science and Technology of China, in 1988, and the Ph.D. degree in electrical engineering from the Institut National des Sciences Appliquées of Rennes, France, in 1996. In 2001, he joined Communications Research Centre Canada (CRC), where his major focus is broadband multimedia systems and digital television broadcasting. He is currently a Research Scientist with CRC. He served as the Session Chair for the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting in 2006, 2007, 2015, 2016, and the TPC Chair in 2016. He was the Managing Editor of the IEEE TRANSACTIONS ON BROADCASTING Special Issue on IPTV in Broadcasting Applications in 2009. He also served as a reviewer for several renowned international journals and conferences in the area of broadcasting, multimedia communication, and multimedia processing. He was the BTS IPTV representative at the ITU-T and the Co-Chair of Enhanced TV Planning Team at ATSC. He is the Associate Editor of the IEEE TRANSACTIONS ON BROADCASTING and IEEE ACCESS.

3.1.3 Conference paper C2

This subsection presents a conference paper related with Contribution 1. The full reference of the paper is presented below:

- E. Iradier, A. Abuin, L. Fanari, J. Montalban, P. Angueira, L. Zhang, W. Li and Y. Wu, "Broadcast/Unicast Convergence in NOMA-based 5G: a RRM Optimization Algorithm," 2020 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Paris, France, 2020, pp. 1-6, doi: 10.1109/BMSB49480.2020.9379805.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Indexed Congress in IEEExplore
- Area: Computer Science and Engineering
- SJR factor: 0.206

2020 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB) | 978-1-7281-5784-9/20/$31.00 ©2020 IEEE | DOI: 10.1109/BMSB49480.2020.9379805

Broadcast/Unicast Convergence in NOMA-based 5G: a RRM Optimization Algorithm

E. Iradier, A. Abuin, L. Fanari, J. Montalban, P. Angueira
Department of Communications Engineering
University of the Basque Country (UPV/EHU)
Plaza Torres Quevedo 1, Bilbao (Spain)
{eneko.iradier, lorenzo.fanari, jon.montalban,
pablo.angueira}@ehu.eus, aritzabuin1@gmail.com

L. Zhang, W. Li, Y. Wu
Communications Research Centre Canada (CRC)
Ottawa, Ontario, K2H 8S2, Canada
{liang.zhang, wei.li, yiyan.wu}@canada.ca

Abstract— **This paper presents an optimization algorithm for resource allocation management in a 5G broadcast/unicast convergence scenario. The convergence of broadcast and unicast services has always been a challenge. In this work, Non-Orthogonal Multiple Access (NOMA) techniques are proposed as the enabler of the broadcast/unicast convergence. Hence, the designed and developed algorithm maximizes the data rate offered for unicast services, while for the broadcast users a predefined minimum percentage of served users and a minimum capacity rate is guaranteed. The algorithm can be executed for NOMA and TDMA. Results indicate that the optimization in the resource allocation level leverages the gain of NOMA techniques in comparison to traditional orthogonal techniques and, therefore, more unicast users could be served by following NOMA solutions. Moreover, in order to mitigate the increment in system complexity due to the introduction of NOMA, a complexity reduction technique has been developed and tested. Results indicate that the performance of the technique varies depending on the target users.**

Keywords— **5G, Broadcast, Convergence, LDM, NOMA, NR, P-NOMA, PTM, Resource Allocation, RRM, Unicast.**

I. Introduction

5G is expected to extend the potential application fields of cellular communications to new use cases. According to [1], the application scope has been divided in three different areas: enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency Communications (URLLC) and massive Machine Type Communications (mMTC). Each one of those application areas presents different requirements and 5G New Radio (NR) has been designed to cover all of them. For example, while eMBB is oriented to offer high data rates (peak data rate values close to 10-20 Gbps), the URLLC use case is targeting mission-critical communications, where reliability and latency are crucial parameters. Finally, mMTC requires an exponential increase in the device density rate, around 10^6 devices per km^2.

The already published first 5G NR release (i.e., Rel-15) includes several promising enhancements (e.g., more robust channel coding or a flexible Physical Layer, PHY) [2]. Nevertheless, Point-to-Multipoint (PTM) communications, a mechanism that could enhance the 5G performance, have not been addressed in detail so far and it is expected to be included as part of future releases (Rel-18/19).

In addition to PTM communications, the coexistence of broadcast/multicast services and unicast or Point-to-Point (PTP) services also has to be managed in future 5G releases. Traditionally, Orthogonal Multiple Access (OMA) techniques have been implemented (i.e. Time Division Multiplexing, TDM and Frequency Division Multiplexing, FDM). An example is 4G eMBMS (evolved Multimedia Broadcast Multicast Services), which is based on dividing the resources in the time domain for broadcast/multicast and unicast services [3].

Recent works have demonstrated that the spectral efficiency of NOMA techniques outperform OMA considerably [4]. A good example is the success of Layered Division Multiplexing (LDM), which is a low-complexity power domain NOMA (P-NOMA) solution. LDM was approved in 2016 as part of the ATSC 3.0 PHY layer baseline technology [5].

Following the successful introduction in the ATSC standard, NOMA has also been proposed to be used as a potential enabler for broadcast/multicast and unicast services convergence. In [6], authors propose to use the two-layer architecture provided by NOMA systems to deliver broadcast services in the Upper Layer (UL) and unicast services in the Lower Layer (LL). However, an optimal service configuration in terms of resource allocation and offered capacity values has not been addressed.

The objective of this paper is to design and evaluate a novel algorithm for optimizing the offered capacity and focus on the Radio Resource Management (RRM) on broadcast/unicast service convergence in 5G networks based on P-NOMA. Afterwards, the algorithm is evaluated in terms of capacity and complexity. Finally, some methods are proposed and tested in order to decrease the complexity.

In summary, the technical contributions of this paper include:

1) Design and development of a RRM optimization tool for enabling the convergence of broadcast/unicast services in 5G networks using NOMA.
2) Capacity performance evaluation of the developed algorithm.
3) Complexity evaluation of the developed algorithm and proposals for decreasing complexity.

The rest of the paper is organized as follows. The next section describes the related work published up to now on this topic. Section III is focused on the description of the proposed algorithm. In section IV, a brief description of the simulation methodology and the use case are presented. Then, in section V, the results are presented and analyzed from the capacity and the complexity point of view. Finally, section VI concludes the work.

II. Related Work

In 4G networks, the inclusion of broadcast communications and its co-existence with unicast services was solved by using time and frequency domain multiplexing. Although a first convergence approach was included in 3G systems, the inclusion of eMBMS in 4G, entailed higher and more flexible bit rates, single frequency network (SFN) operations and carrier configuration flexibility [3]. In LTE Release 14, the latest eMBMS version was presented. Release 14 eliminated the restriction of a maximum of 60% of the radio resources allocated for broadcast/multicast [7].

In the case of 5G, a similar roadmap is expected for addressing PTM communications. Some of the most relevant contributions in this area are related to the 5G-Xcast project [8]. This project presented several modifications at different levels of the protocol stack in order to introduce broadcast communications in 5G NR. A good example is found in [9], where a specific approach of the 5G NR PHY to cover the terrestrial broadcast need is presented.

Regarding P-NOMA-based systems (like LDM), although having demonstrated very positive conditions in order to deliver broadcast or multicast contents, they were not included in 4G. However, the integration of NOMA in 5G NR has become again a hot topic inside the debate forums [10]. Previous works have analyzed the possible integration of NOMA techniques in eMBMS environments [10] [11]. Authors in [10] provide a theoretical point of view of the possible use of NOMA as part of 5G capabilities to deliver PTM communications within broadband networks. A second approach [11] demonstrates that NOMA has the potential to become a feasible technique to deliver multicast services in subgroup-oriented communications.

In papers [12]-[14], the use of P-NOMA was analyzed in a 5G-MBMS system for delivering broadcast services with different requirements, and for delivering mixed unicast and broadcast services. Authors provided several results in terms of capacity, where the benefits of using P-NOMA are demonstrated theoretically and carrying out simulations. Broadcast services are configured to be delivered on top of the unicast network with nearly full capacity.

In [15], the first NOMA-based 5G NR PHY layer transceiver was presented. The work addresses simultaneously services convergence and NOMA. Authors provided several results based on specific use cases, where the broadcast/unicast convergence is based on NOMA. Then, in [6], a full PHY/MAC transceiver is presented in order to cope with the service convergence requirements. In this case, network level results demonstrated benefits at different stack levels.

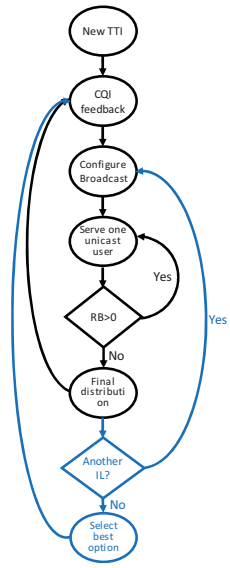


Fig. 1. Proposed algorithm diagram.

Nevertheless, none of those approaches has studied the possibility of optimizing RRM strategy in the broadcast/unicast convergence environment using NOMA.

III. RRM Algorithm for Convergence

In Fig. 1, a diagram of the developed algorithm is presented. The algorithm has two different versions, one for TDMA and another one for NOMA. The blocks in black of Fig. 1 are common blocks for both technologies, whereas blue blocks are specific for NOMA. It is important to highlight that the target of this algorithm is the maximization of the capacity offered to the unicast services once a minimum broadcast capacity is maintained. In this case, the unicast services are considered secondary services (or opportunistic services), so the fairness is less of a concern.

The basis of this algorithm is the set of Channel Quality Indicator (CQI) values received from different user equipment (UEs) in the uplink channel. The algorithm processes the received CQIs to optimize the RRM configuration and it is executed in every Transmission Time Interval (TTI). First, CQI_k values from the k UEs that compose the network are received and the number of users requesting each MCS is calculated as $U = \{u_1, u_2, \dots, u_C\}$, where u_C is the number of terminals having a CQI equal to c. Then, the broadcast service is configured according to the percentage of broadcast users which service has to be guaranteed. To do so, the highest MCS that guarantees that the predefined served user percentage constraint being satisfied is selected. If TDMA is selected, a portion of the total amount of Resource Blocks (RB) is reserved for PTM and the rest is for managing the unicast services. However, in the case of NOMA, the whole bandwidth is used for both type of services, broadcast/multicast and unicast.

Once the PTM configuration is applied, the first unicast user is served. As the main objective of this algorithm is to maximize the unicast service throughput, the user requesting the highest Modulation and Coding Schemes (MCS) is served first, since less RBs are needed to achieve the minimum capacity. The MCS selection is carried out according to the Table 5.1.3.1-2 in [16]. When the required RBs for the first user are reserved, it is checked if there are still enough RBs to continue with the next user. If no more RBs are available, the configuration is finished. On the other hand, if there are still free RBs, the next user with the highest MCS request is selected.

If the algorithm is implemented for NOMA, a longer process has to be applied since several Injection Level (IL) values have to be tested. In fact, this algorithm uses a fixed set of IL values composed by values from -20 dB to -0.5 dB in steps of 0.5 dB (i.e., 40 IL values are tested). When the IL is changed, the required SNR to decode any of the NOMA layers is modified and affects the MCS value that is used to serve the users. If high IL values are used, UL presents similar SNR values in comparison with single-layer cases, whereas the SNR for the LL is considerably increased. In that case, it is easier to guarantee the PTM constraint, but fewer unicast users could be served. However, when low IL values are used, the opposite situation occurs: the difference between PTM configurations and single-layer mode is increased and unicast users are easily satisfied. Therefore, after every IL value has been tested, the combination that maximizes the LL throughput while guaranteeing the minimum percentage of PTM users is selected. Clearly, this process implies an increase in the complexity of the RRM algorithm, since more possible configurations have to be considered. This issue is studied more in detail in section V.B and a solution is proposed and evaluated.

IV. Simulation framework

A. Simulation methodology

In order to test the algorithms for TDMA and NOMA, a two-step methodology is implemented using two different simulators. In the first step, the CQI requests of each user are gathered per TTI. This task is performed in a network simulator, in OMNeT++, specifically. The inputs to this block are the network parameters and the mobility models. The network parameters define the conditions that affect the path loss calculation, including for example, the propagation channel model, the coverage area or the carrier frequency, among others. On the other hand, the mobility model selection is related to the user type. The speed of the receiver is implemented in this block and it has a close relation with the Doppler frequency that the user has to cope with. Taking into account those parameters, the CQI generator collects the CQI request of each user in each TTI and it is introduced in the algorithm block.

The second step is related with the algorithm itself and it is carried out in Matlab. The first task in this step is to gather all the constraints and conditions that need to be defined before running the algorithm. The minimum throughput constraints for broadcast and unicast are some of the most relevant conditions.

TABLE I
Configuration Parameters

Parameter	Value
Bandwidth	20 MHz
Carrier Frequency	2 GHz
SCS	15 kHz
Tx Power	44 dBm
Distance attenuation	128.1 + 37.6*log(d)
ISD	500 m
# of User	100
# of TTI	10000
Minimum PTM data rate	2, 3, 4 Mbps
Minimum Unicast data rate	1, 2, 3, 4 Mbps
Inter-Site Distance	500 m
Channel model	TDL-D, TDL-E
Desired Delay Spread	90, 360, 1100 ns
Noise Power	-90 dBm
Noise Figure	9 dB
Time slot size	1 ms
Mobility type	RWP
User type	Pedestrian, car-mounted
Speed	Pedestrians: 3 km/h Car-mounted: 30 km/h

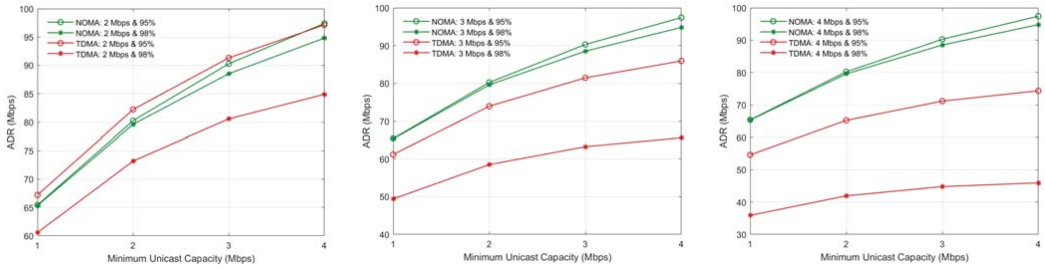
These values are selected depending on the use case being emulated. The algorithm will use them to calculate how many RBs are required to configure each service type. In the case of broadcast services, the minimum percentage of served users has to be configured as well. Meanwhile, it is necessary to indicate if the algorithm has to evaluate the configuration for TDMA or NOMA. Then, the values generated in the CQI generator step are used as input and the algorithm is executed. Once the algorithm has finished, the data rate offered for the PTM communications, the aggregate data rate (ADR) of the unicast services and the number of served unicast users are presented for their evaluation.

B. Use case configuration

The defined use case assumes that different video qualities and contents are delivered to mobile receivers, while the coexistence of broadcast and unicast transmissions is guaranteed. The configuration parameters are listed in TABLE I.

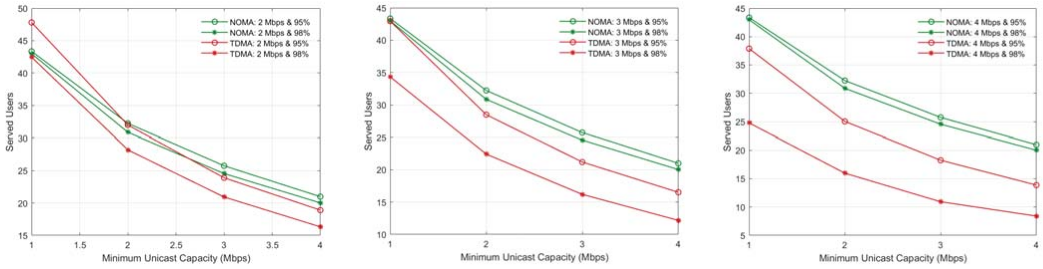
The main objective of the broadcast content is to deliver a basic common quality service for all the users of the network. In this case, different subsets of the same common use case are evaluated, where values from 2 Mbps to 4 Mbps have been considered in order to guarantee different quality expectations. Then, unicast services are delivered for some of the users of the network to increase their Quality of Service (QoS) by increasing the overall data rate. In this case, a similar range of possible data rates has been assumed (from 1 Mbps to 4 Mbps) in order to enhance the received video quality or to increment the number of simultaneously received video contents.

The major driving factors for use case configuration are the propagation channel model and the user definition. The former is based on [17], which uses two different channel profiles (TDL-D or TDL-E) and three different desired delay spread (DDS) values for each channel model: 90 ns, 360 ns and 1100 ns. Those values have been selected for their alignment with



(a) (b) (c)

Fig. 2. Unicast ADR results using different PTM minimum capacities: (a) 2 Mbps, (b) 3 Mbps, (c) 4 Mbps.



(a) (b) (c)

Fig. 3. Number of unicast users served using different PTM minimum capacities: (a) 2 Mbps, (b) 3 Mbps, (c) 4 Mbps.

short-, normal- and long-delay profiles for an urban environment. All the possible DDS and channel profile combination have been simulated jointly in the same cell in order to emulate a realistic scenario where users with different receiving conditions coexist. Actually, the channel model is closely related with the user type, since the channel affecting the receiver with the lowest mobility will vary slowly in comparison with users with higher mobility. This concept is also critical in the mobility model, since it takes care of detailing how the receiver is moving. In this case, the mobility of the receivers is modelled with the Random Way Point (RWP) model [18], which was designed to emulate the mobility of mobile receivers inside the cell. In addition, the speed of the users has been set to 3 km/h for pedestrian receivers and 30 km/h for the car-mounted.

V. Results

Results are divided into two subsections. The first one analyzes capacity results in the case of pedestrian users whereas the second addresses complexity. Finally, a technique for reducing the complexity is proposed and tested with pedestrian and car-mounted users.

A. Capacity

The capacity results of the RRM algorithm are presented in Fig. 2, where ADR values are displayed for different capacity and broadcast served users constraints. First, it should be highlighted that, for most of the cases NOMA offers better capacity performance than TDMA. The performance gain depends on different constraints. The more challenging the configuration is, the higher the gain that NOMA can achieve. That is why, the highest gains are presented in Fig. 2(c), where at least 4 Mbps have to be dedicated to the PTM service. In particular, for the cases where 3 Mbps and 4 Mbps have to be guaranteed, the capacity offered with NOMA duplicates the capacity of TDMA. However, for the configuration where soft constraints are applied (required PTM capacity < 2 Mbps), TDMA seems to be better option. In addition, the results also show that the TDMA-based algorithm is more sensitive to the constraint of the percentage of broadcast users that have to be served, and NOMA-based algorithm offers similar results for different values. The main reason for this is that when the broadcast users' constraints are more restrictive, less RBs are available for the unicast service configuration and, therefore, some users cannot be served. However, NOMA has always the same number of available RBs for the unicast services.

On the other hand, Fig. 3 shows another perspective for the capacity analysis, the number of unicast users served under the same constraints as in Fig. 2. In general, a similar behavior is observed: NOMA offers better performance and the highest gain appear when the applied constraints are more challenging. In this case, the number of served users decreases while the minimum unicast capacity increases. Fig. 3(a) shows very similar results for both technologies. However, in Fig. 3(c), the

TABLE II
ANALYSIS OF THE PROPOSED COMPLEXITY REDUCING TECHNIQUE

User	Served (%)	σ	MCS' = MCS		MCS' = MCS - 1		MCS' = MCS - 2		MCS' = MCS - 3		MCS' = MCS - 4		MCS' = MCS - 5	
			Gain (Mbps)	Success (%)	Gain (Mbps)	Success (%)	Gain (Mbps)	Success (%)	Gain (Mbps)	Success (%)	Gain (Mbps)	Success (%)	Gain (Mbps)	Success (%)
Pedestrian	95	1/8	8.75	73.37	4.45	81.60	3.11	88.63	-1.36	92.98	-5.59	95.92	-10.05	97.48
		1/16	8.7	72.69	4.41	80.54	3.05	87.64	-1.44	92.23	-5.67	95.15	-10.12	96.94
		1/25	8.82	70.81	4.54	78.42	3.32	85.36	-1.09	90.01	-5.31	93.45	-9.78	95.55
		1/40	8.84	68.17	4.56	75.11	3.13	82.50	-1.34	87.70	-5.57	91.37	-10.02	93.67
	98	1/8	25.32	73.92	21.7	81.63	19.07	88.47	13.94	92.73	9.75	95.63	5.37	97.25
		1/16	25.26	73.33	21.65	80.65	19.01	87.42	13.86	91.99	9.67	94.87	5.29	96.58
		1/25	25.29	71.38	21.7	78.52	19.03	85.36	13.92	89.83	9.73	93.20	5.33	95.51
		1/40	25.49	68.68	21.84	75.24	19.18	82.31	14.08	87.40	9.92	91.10	5.53	93.44
Car-mounted	95	1/8	7.67	65.00	3.34	70.35	2.45	75.66	-1.93	79.89	-6.21	83.86	-10.97	86.94
		1/16	7.54	65.48	3.25	70.96	2.37	76.36	-1.98	80.46	-6.28	84.56	-11.04	87.68
		1/25	7.73	65.03	3.43	70.09	2.47	75.67	-1.88	79.99	-6.19	83.92	-10.99	87.29
		1/40	7.65	63.37	3.3	68.96	2.46	74.80	-1.92	79.28	-6.2	82.76	-11	85.74
	98	1/8	23.46	65.35	19.57	70.42	17.85	75.69	12.96	79.90	8.65	83.83	3.95	86.93
		1/16	23.36	65.83	19.51	71.05	17.82	76.41	12.94	80.44	8.63	84.54	3.93	87.70
		1/25	23.59	65.32	19.7	70.16	18	75.71	13.13	79.96	8.82	83.83	4.07	87.15
		1/40	23.59	63.91	19.61	69.02	18.05	74.77	13.24	79.24	8.91	82.68	4.18	85.62

number of unicast users served with NOMA outperforms TDMA considerably. In fact, the maximum gain appears in the last case where 4 Mbps have to be guaranteed for both, PTM and unicast services. In this case, the capacity offered with NOMA is 51.2% and 138% higher than with TDMA for 95% and 98% of broadcast users, respectively.

B. Complexity

When designing a RRM algorithm, complexity is considered one of the main Key Performance Indicators (KPIs). As presented in Fig. 1, TDMA- and NOMA-based algorithms have common steps of the algorithm. However, when NOMA is used, several extra iterations have to be carried out in order to test different IL values. Therefore, the complexity comparison of the NOMA and TDMA can be expressed as:

$$C_{NOMA} = \sum_{\Delta} C_{TDMA} \quad (1)$$

where $C_{NOMA}$and C_{TDMA} are the computational cost of NOMA and TDMA, respectively. As defined in section III, 40 ILs are tested during the execution of the NOMA-based algorithm and, so, the complexity of NOMA is 40 times higher than in TDMA. However, according to Eq. (1), the complexity of the NOMA-based algorithm will decrease if fewer IL values are tested. Nevertheless, reducing the number of ILs may get the RRM configuration further away from the optimum configurations.

In order to decrease the overall computational cost of the NOMA-based algorithm and targeting a real deployment, the execution frequency of the algorithm will be less than one TTI. Therefore, the complexity of the NOMA-based algorithm varying the execution frequency can be expressed as follows:

$$C_{NOMA} = \sigma \cdot \sum_{\Delta} C_{TDMA} \quad (2)$$

where σ represents the execution frequency, (0,1]. In this case, the complexity increases due to the use of NOMA can be compensated by reducing the execution frequency.

However, the main drawback of this solution is that when the configuration is maintained over several TTIs, as the receivers are moving, they may change their requested CQI and they may not receive correctly the information because they are receiving the data with a MCS higher than the maximum MCS they can decode. That is why, the gain that NOMA has demonstrated in the previous section is invested in reducing the MCS used in the delivery, in order to reduce the mismatch probability between the requested and the received MCS, and therefore, improve the performance of the complexity reduction technique.

This approach has been verified with simulations. The execution frequency (σ) has been varied, as well as the MCS used to serve the users (i.e., MCS'). In order to evaluate the proposed solution, a reduced number of subsets have been evaluated: two different served broadcast users constraints (i.e., 95% and 98%) and 3 Mbps as the minimum capacity constraints for both, unicast and broadcast services. Results are presented in Table II. The performance is analyzed using two metrics: the gain of NOMA (i.e., the difference in offered unicast capacity between NOMA and TDMA) and the receiving success (i.e., the percentage of times that users have received a MCS value below or equal to the maximum MCS that can receive).

Concerning the capacity gain, when 95% of the broadcast users have to be guaranteed, up to two MCS can be decreased without perceivable capacity loss. However, if 98% of the broadcast users are guaranteed, even in the last case where five MCS values are decreased, NOMA outperforms TDMA. This effect is supported by the capacity gain that NOMA offers when the broadcast users' constraint is much increased. Therefore, the complexity reducing technique performs better, when the requirement of the served broadcast users is higher.

On the other hand, regarding the receiving success, better results are obtain for the pedestrian users since they change their requested MCS slower. In fact, the best pedestrian results

are close to 97%, while car-mounted users can guarantee 87% of success. Moreover, the relation between the execution frequency and the success is also different depending on the type of user. While differences around 5% between the highest and the lowest frequencies are observed for pedestrian users, in the case of car-mounted users, almost the same success values are obtained. This result indicates that in the case of car-mounted users, less TTIs should be skipped in order to appreciate an evolution on the success rate.

Finally, the case $\sigma = 1/40$ should be highlighted since it makes TDMA and NOMA complexity equal. In this case, positive results are obtained when the percentage of served broadcast users is high. However, the performance is downgraded when the broadcast users' constraint is reduced. The explanation of this difference is the same as in the previous section, where TDMA was more sensitive to the broadcast users' constraint. Therefore, if the complexity due to the introduction of NOMA techniques in the RRM has to be reduced, a tradeoff has to be assumed between the unicast capacity gain and the performance of the algorithm over users with different mobility types.

VI. Conclusions

A NOMA-based RRM algorithm has been designed and evaluated to optimize the resource allocation in a broadcast/unicast convergence environment. In addition to the algorithm specifications, capacity and complexity have been analyzed as the main KPIs and a complexity reduction technique has been proposed and verified.

Capacity results indicate that under the optimum resource allocation conditions, NOMA techniques outperform the performance obtained with OMA techniques. In fact, considerable gains in terms of data rate and served users have been presented for the unicast services. On the other hand, the complexity analysis has demonstrated that NOMA-based algorithm increments the computational cost according to the number of IL values tested. In order to reduce the overall computational cost, a complexity reducing technique has been introduced based on the reduction of the execution frequency. The main problem of this solution is the mismatch generated between the requested and the used MCS. Therefore, the capacity gain that NOMA offers has been invested in serving the users with lower MCS values to allow for a margin that increments the success rate. Results indicate that the performance has a considerable influence on the mobility type and the percentage of served broadcast users. In fact, the technique performs better for pedestrian users than for car-mounted users. Moreover, better success results are obtained, when the percentage of served broadcast users is higher, since the capacity gain that NOMA offers w.r.t TDMA is higher.

Acknowledgments

This work has been partially supported by the Basque Government (project IOTERRAZ under the grant KK-2019/00046 ELKARTEK 2019, the grant IT1234-19 and the PREDOC grant program PRE_2019_2_0037) and by the Spanish Government (project PHANTOM under the grant RTI2018-099162-B-I00 (MCIU/AEI/FEDER, UE)).

References

[1] ITU-R Rec. ITU-R M. 2083-0, "IMT Vision — Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond," Sept. 2015.

[2] TS 38.201, Tech. Spec. Group Services and System Aspects, "NR; Physical layer; General description," V15.0.0, January 2018.

[3] D. Lecompte and F. Gabin, "Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: overview and Rel-11 enhancements," in *IEEE Communications Magazine*, vol. 50, no. 11, pp. 68-74, November *2012.*

[4] L. Zhang et al., "Layered-Division-Multiplexing: Theory and Practice," in *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 216-232, March 2016.

[5] S. Park et al., "Low complexity layered division multiplexing system for ATSC 3.0," in *IEEE Transactions on Broadcasting*, vol. 62, no. 1, Mar. 2016.

[6] E. Iradier, et al., "Using NOMA for Enabling Broadcast/Unicast Convergence in 5G Networks," in *IEEE Transactions on Broadcasting*, vol. 66, no. 2, pp. 503-514, June 2020.

[7] C. Hoymann et al., "LTE release 14 outlook," in *IEEE Communications Magazine*, vol. 54, no. 6, pp. 44-49, June 2016.

[8] D. Ratkaj and A. Murphy, Eds., "Definition of Use Cases, Requirements and KPIs," Deliverable D2.1, 5G-PPP 5G-Xcast project, Oct. 2017.

[9] J. J. Gimenez et al., "5G New Radio for Terrestrial Broadcast: A ForwardLooking Approach for NR-MBMS," in *IEEE Transactions on Broadcasting*, vol. 65, no. 2, pp. 356-368, June 2019.

[10] L. Zhang et al., "Layered-Division Multiplexing: An Enabling Technology for Multicast/Broadcast Service Delivery in 5G," in *IEEE Communications Magazine*, vol. 56, no. 3, pp. 82-90, March 2018.

[11] J. Montalban et al., "Multimedia Multicast Services in 5G Networks: Subgrouping and Non-Orthogonal Multiple Access Techniques," in *IEEE Communications Magazine*, vol. 56, no. 3, pp. 91-95, March 2018.

[12] L. Zhang et al., "Using Layered-Division-Multiplexing to Achieve Enhanced Spectral Efficiency in 5G-MBMS," *2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jeju, Korea (South), 2019, pp. 1-7.

[13] Y. Xue, A. Alsohaily, E. Sousa, W. Li, L. Zhang and Y. Wu, "Using Layered Division Multiplexing for Mixed Unicast-Broadcast Service Delivery in 5G," *2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jeju, Korea (South), 2019, pp. 1-6.

[14] L. Zhang et al., "Using Non-Orthogonal Multiplexing in 5G-MBMS to Achieve Broadband-Broadcast Convergence with High Spectral Efficiency," in *IEEE Transactions on Broadcasting*, vol. 66, no. 2, pp. 490-502, June 2020.

[15] E. Iradier, J. Montalban, D. Romero, Y. Wu, L. Zhang and W. Li, "NOMA based 5G NR for PTM Communications," *2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jeju, Korea (South), 2019, pp. 1-6.

[16] 3GPP TS 38.214 v15.3.0, Tech. Spec. Group Services and System Aspects, "NR; Physical layer procedures for data (Release 15)", Sept. 2018.

[17] 3GPP TR 38.901. "Study on channel model for frequencies from 0.5 to 100 GHz." 3rd Generation Partnership Project; Technical Specification Group Radio Access Network.

[18] E. Hyytia, H. Koskinen, P. Lassila, and A. Penttinen, "Random waypoint model in wireless networks," Networks and Algorithms: Complexity in physics and Computer Science, June 2005

3.1.4 Journal paper J2

This subsection presents a journal paper related with Contribution 1. The full reference of the paper is presented below:

- E. Iradier, M. Fadda, M. Murroni, P. Scopelliti, G. Araniti and J. Montalban, "Non-Orthogonal Multiple Access and Subgrouping for Improved Resource Allocation in Multicast 5G NR," submitted manuscript to Digital Communications and Networks.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Journal paper indexed in JCR and IEEExplore
- Area: Telecommunications
- Ranking: 12/90 (Q1) based on JCR 2019
- Impact factor (JCR): 5.382

Digital Communications and Networks(DCN)

HOSTED BY ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/dcan

Non-Orthogonal Multiple Access and Subgrouping for Improved Resource Allocation in Multicast 5G NR

Eneko Iradier[a,*], Mauro Fadda[b], Maurizio Murroni[b], Pasquale Scopelliti[c], Giuseppe Araniti[c], Jon Montalban[a]

[a]**Dept. of Communications Engineering, University of the Basque Country, Bilbao, Spain**
[b]**Dept. of Electrical and Electronic Engineering - DIEE/UdR CNIT, University of Cagliari, Italy**
[c]**Dept. of Information Engineering, Infrastructure, and Sustainable Energy - DIIES, Mediterranea University of Reggio Calabria, Italy**

Abstract

The effort of the telecom industry to enhance the Quality of Service (QoS) of the cellular networks has been constant during the last decade. To enable new application scenarios, the last generation of mobile networks (i.e., 5G) is being designed to cope with very stringent constraints in terms of device density, latency, user mobility and peak datarate. Nevertheless, as its predecessor (e.g. LTE), the Radio Resource Management (RRM) only implements Orthogonal Multiple Access (OMA) techniques. On the other hand, within the media vertical, vehicular communications represent a demanding and challenging use case. Up to now, solutions based on multicast transmissions have presented considerable efficiency increments by the successful implementation of subgrouping strategies. These techniques enable a more efficient exploitation of multi-user diversity by splitting users into subgroups and by applying independent and adaptive modulation and coding schemes. However, at the RRM, OMA poses a hard limit in the exploitation of the available resources, especially when users' QoS requests are unbalanced, which is a typical feature of vehicular environments. Under these circumstances, this paper proposes to jointly apply the subgrouping and Non-Orthogonal Multiple Access (NOMA) techniques in 5G for mobile and vehicle receivers. This study shows that NOMA is highly spectrum-efficient and in certain conditions could improve the system throughput performance. In the first part of this paper, an in-depth analysis of the implications of the introduction of NOMA techniques in 5G subgrouping at RRM is carried out, and afterwards, the validation is accomplished by applying the proposed approach to different 5G use cases based on vehicular communications. After a comprehensive analysis of the results, the proposed joint solution improves considerably the data rate offered in each use case.

KEYWORDS:
5G, ADR, LDM, NOMA, P-NOMA, Resource Allocation, RRM, Subgrouping, Wireless Communications.

1. Introduction

The fifth generation of cellular networks (5G) is being developed to provide groundbreaking connectivity to everyone and everything, everywhere and everytime. 5G networks are expected to cover a wide range of use cases with customized parameters: devices, type of users, requirements and mobility classes. In 2015, the International Telecommunication Union Radiocommunication Sector (ITU-R) issued the International Mobile Telecommunications-2020 (IMT-2020) standard, which are the requirements for 5G networks, devices and services. The IMT-2020 has put forth some requirements for 5G that focus on fulfilling three Key Performance Indicators (KPIs): $> 10Gb/s$ peak data rates for enhanced Mobile Broadband (eMBB), $> 1M/km^2$ connections for massive Machine-Type Communications (mMTC), $< 1ms$ latency for Ultra-Reliable Low Latency Communications (URLLC) [1].

*Corresponding author Eneko Iradier (email: eneko.iradier@ehu.eus).

The first version of 5G, which is based on the specifications developed by 3GPP in Release 15 (Rel-15), comprises the 5G Core (5GC) and 5G New Radio (NR) with 5G User Equipment (UE) and is currently being deployed commercially throughout the world both at sub-6 GHz and at mmWave frequencies. In Rel-15, there are common elements with 4G Long Term Evolution (LTE) and LTE-A, such as the use of orthogonal frequency division multiplexing (OFDM). The second phase of 5G is being standardized in parallel by 3GPP in the Release 16 (Rel-16). While Rel-15 is focused on eMBB services, Rel-16 includes new features for URLLC, mMTC and Industrial IoT, including Time Sensitive Communication (TSC), Enhanced Location Services (ELS), and support for Non-Public Networks (NPNs). Meanwhile, the feature content of Rel-17 is already underway, targeting specification availability in mid-2021. Rel-17 and beyond is expected to include (but not limited to) enhanced support for wireless and wired convergence, multicast and broadcast architecture, proximity services, multi-access edge computing and network automation [2]. To enable all the potential use cases, high spectral efficiency is one of the main goals to be achieved. As in the previous generation, the radio resource management (RRM) implements orthogonal multiple access (OMA) techniques. Consequently, although 5G has increased the spectral efficiency in comparison with 4G adding flexibility to RRM in the definition of RB allocation, there appears to be still room for improvement.

In multicast transmissions, where users access the same content, RRM is in charge of performing an efficient link adaptation procedure according to the channel conditions experienced by the multicast users. This adaptation has to be accomplished on a per-group basis, taking into account the Channel State Indicator (CSI) of all the users registered to a given multicast service. However, in single rate-transmissions the presence of cell-edge users, which experience poor channel conditions can strongly degrade the Quality of Service (QoS) that the cellular infrastructure could provide. Subgrouping is a solution that has been proposed in literature as an effective technique to overcome those limitations. Based on splitting the multicast users into subgroups and applying subgroup-based adaptive modulation and coding schemes (MCS), it enables more efficient exploitation of multi-user diversity by optimizing opportune cost functions [3]. This solution is very useful for vehicular communications, where the high and rapid variation of the receivers positions lead to unequal reception conditions [4] [5].

In general, vehicular communications imply unbalanced QoS requests from multicast services (e.g., groups with different throughput), and under this situation, OMA could limit the full exploitation of the resources. In such a scenario, Non-Orthogonal Multiple Access (NOMA) techniques have shown the power to allow increasing spectral efficiency [6]. With respect to OMA, where the available frequency/time resources in the network are orthogonally assigned to the different User Equipments (UEs), in NOMA, UEs share the available resources, both in frequency and time. On the one hand, in OMA at the receiver side, under perfect conditions, the desired data can be unequivocally separated from the rest of the information. On the other hand, in NOMA when decoding the desired content, the rest of the signals are considered an additional source of interference.In [7], NOMA techniques have shown higher efficiencies in comparison with OMA, especially when the throughput rate among different users is asymmetric.

In [8], authors explored the joint use of subgrouping multicast techniques and NOMA in an evolved multimedia broadcast multicast service (eMBMS)-like 5G scenario, where different quality video services were delivered to a group of users interested in the same contents. The potentiality of the proposed approach was preliminary investigated in some envisaged 5G mobility environments and the results in terms of QoS indicators such as maximum throughput (MT) and aggregated data rate (ADR) were promising. The objective of this paper is to go deeper into the study of subgrouping with NOMA techniques in 5G scenarios performing. Therefore, an in-depth analysis of the implications of the introduction of NOMA techniques in 5G subgrouping at the resource allocation level (i.e., RRM) is carried out and the approach is formalized and validated. In particular, although [8] is oriented to 5G networks, it makes use of 4G numerology and the data rate calculations are based on the Shannon's capacity approach. Therefore, this work overcomes those limitations with a novel cost-function design and evaluation.

This study shows how, under opportune conditions, NOMA allows improved performance of subgrouping techniques for efficient resource allocation. In summary, the novel technical contribution of this paper is the comprehensive proposal and evaluation of subgrouping techniques combined with 5G and NOMA which includes:

1. Detection of the penalization of using NOMA for subgrouping and provide a solution.
2. Design of a new evaluation cost function for 5G subgrouping based on NOMA.
3. Comprehensive analysis of the results of different techniques based on variable constraints and vehicular environment network parameters.
4. Identification and theoretical demonstration of the system further evolution.

The rest of the paper is organized as follows. The next section describes the basic principles and related

work published in literature both for subgrouping and NOMA techniques. In section III, the system model and problem formulation for the proposed NOMA-based subgrouping technique are detailed. The simulation framework is presented in section IV, whereas section V is dedicated to the analysis of the results. Future works and system evolution are addressed in section VI. Finally, section VII draws the conclusions.

2. Basic Principles and Related Work

Multicast scheduling can be divided into two types of multicast transmissions: single-rate and multi-rate [9]. In the single-rate case, the base station transmits to all users in each multicast group at the same rate. In multi-rate, instead, the base station transmits to each user at different rates exploiting users' frequency diversity, according to the heterogeneity of the wireless channels. This paper focuses on the latter case. Multi-rate multicast schemes allow each user to receive multimedia traffic based on the capabilities of the UE. In particular, subgrouping is one of the multi-rate techniques that exploits the group-splitting approach. Nevertheless, subgrouping techniques introduce new issues related to subgrouping formation: proper resource allocation, MCS selection, etc. Consequently, many schemes have been proposed by the research community to achieve the optimal solution. This section summarizes the main contributions dividing them between the solutions that do not include NOMA (section 2.2.1) and the ones that take advantage of NOMA (section 2.2.2). Priot to that, NOMA is briefly introduced in section 2.1.

2.1. NOMA

NOMA gathers a wide family of medium access techniques where all users share the same frequency and time resources. As described in [10], non-orthogonal access can be classified into code-domain NOMA [11][12] and power-domain NOMA (P-NOMA) [13][14]. In principle, power domain NOMA presents a better trade off between complexity and capacity [10]. During the rest of the paper, NOMA will refer to power domain NOMA.

P-NOMA is based on the idea of overlapping layers, published by Cover and Bergmans in [15][16] and on a technique referred as Successive Interference Cancellation (SIC). A P-NOMA receiver iteratively demodulates, decodes and cancels successive information layers from the received signal ensemble [17][18].

The capacity gain of NOMA has been demonstrated theoretically [14][19], by simulations [20][21] and in both lab tests and field trials [22][23]. However, this gain depends on the robustness/capacity differences between the services conveyed within the NOMA ensemble. This way, the higher the difference between the NOMA layers (services), the higher the gain. On the contrary, multiplexing services of equal capacity in NOMA does not provide any overall capacity gain in comparison to OMA structures.

2.2. Multicast subgrouping

Multi-rate multicast transmissions can be divided into two different categories, stream-splitting and group-splitting. Stream-splitting is based on splitting high-rate multimedia contents in multiple substreams of low-rate [24], while group-splitting techniques divide multicast members into different subgroups, each one formed by users experiencing similar channel conditions [25]. The most suitable subgroup configuration is dynamically selected by the eNodeB based on the minimization (or maximization) of a given cost function, which takes into account the users channel quality indicator (CQI) values and the QoS constraints of the multicast session [26].

The subgrouping techniques consist of three stages:

- *CQI collection:* the eNodeB collects the CQI feedbacks from each UE belonging to the multicast group. In particular, each UE estimates its CQI value based on the Signal to Interference and Noise Ratio (SINR) and the target Block Error Rate (BLER). Afterwards, forwards the CQI back (associated to the highest supported MCS) to the eNodeB.

- *Subgroup creation:* the multicast members are split into subgroups. The subgrouping algorithm determines the subgroup configuration that: *(i)* allows to maximize the system capacity; *(ii)* guarantees that each served UE can successfully demodulate the received signal; *(iii)* optimizes a given cost function.

- *Resource allocation:* depending on the configuration defined in the previous step, all the UEs reporting a CQI value equal to the minimum CQI selected for each subgroup will be served with the closest supported MCS.

The optimization problem can be formulated in order to achieve different goals: the maximization of the system throughput, spectral efficiency, energy efficiency, or the minimization of resource utilization. In addition, it worth noticing that it has been mathematically demonstrated that the optimal subgroup configuration when maximizing the ADR has to be searched between a single subgroup and two-subgroups [3].

2.2.1. OMA Techniques

Several subgrouping algorithms and methods can be found in the literature, differing for adopted strategies, optimization functions, or considered physical characteristics. In [27], the authors presented a two-step procedure that merges areas significantly overlapped in space considering similar users' contents, returning multi-content groups that include cells with very

similar content interests, and eventually, broadcasts the most demanded items thus maximizing the system throughput. Then, in [28] it is described an innovative RRM mechanism developed to increase the eMBMS performance by splitting each multicast group into several subgroups according to the users channel quality.

In [29], a multicast subgroup formation method and an Application-Layer Joint Coding (ALJC) technique are combined to improve the performance of multicast satellite transmission through the Land Mobile Satellite (LMS) channel both in terms of throughput and perceived quality of the video streams. On the other hand, in [30] the authors describe a new approach to separate experts in subgroups assigning properly the importance of such subgroup in the resolution of a large-scale group decision making (LS-GDM) problem, in which experts elicit their preferences by fuzzy preference relations (FPRs) and a fuzzy clustering process that computes the subgrouping.

Recently, the work in [31] focuses on the delivery of augmented reality services in urban scenarios considering the 5G NR standard through eMBMS to achieve both higher capacity and lower latency with respect to LTE. In [32], the design of a heuristic algorithm for dynamic eMBMS over single frequency networks (MBSFN) area formation is presented in order to increase the ADR exploiting the multi-rate transmission typical of subgrouping jointly with the scalable video coding (SVC) technique and radio resource allocation for an efficient spectrum utilization in the configured MBSFN areas. Then, in [33], a precoding for max-min fairness (MMF) for multi-group multicasting is designed and evaluated using a common message. This technique turned into a rate-splitting algorithm that combines private and common information in each group. Results provided by the authors highlight that rate-splitting improves the overall performance especially in overloaded systems. Finally, in [34] authors combine game theory developments with multicast grouping algorithms. In fact, it is focused on an indirect reciprocity game framework that tests the device-to-device multicast (D2MD) stimulation problem. In this case, the request nodes (RNs) are grouped and the broadcast content is transmitted by the head node (HN), which will be assigned different reputation values depending on the delivery.

2.2.2. NOMA Techniques

Iradier et al. in [35] analyzed the benefits of LDM technology with subgrouping algorithms into the resource allocation mechanism of LTE-A standard for multicast vehicular communications. In [36], a cooperative transmission scheme is proposed for unicast/multicast transmissions in a downlink Cognitive Radio (CR)-NOMA system, evaluating the performance improvement when the number of secondary users is increased. In [37], it is presented a multiple access method called power domain sparse code multiple access (PSMA), for heterogeneous networks (HetNets), based on the use of codebooks that can be reused more than one time in the coverage area of a base station to improve the system spectral efficiency. The signal model and detection techniques of PSMA together with power domain (PD)-NOMA and sparse code multiple access (SCMA) were investigated for comparison. In [38], a simple user grouping and pairing scheme for NOMA in a downlink visible light communication system is presented. The proposed scheme is a mix of both conventional NOMA and OMA schemes where every two users are paired using NOMA and all pairs are allocated with conventional OMA. The performance of the proposed scheme is compared to conventional OMA in terms of maximum sum rate.

Recently, [39] investigates the possibility of a more spectrum efficient solution in mixed unicast-broadcast service delivery in 5G-MBMS using a power-based non-orthogonal multiplexing (P-NOM) (i.e., LDM) examining the potential capacity gains from using NOMA over OMA in broadband systems (i.e., LTE and 5G-NR). On the other hand, [40] describes a new method to gather users in P-NOMA and NOMA-2000 [41] over different Rayleigh fading channels. NOMA-2000 combines OFDMA signal and multi carrier-code division multiple access (MC-CDMA) signal and MC-CDMA waveform is spread and superposed to OFDMA waveform.

3. System Model and Problem Formulation

The 5G system model is represented by a next generation base station (gNB), providing services to users with different mobility types. In 5G NR systems, the available radio resources are managed on a RB basis. One RB is the smallest frequency resource that can be assigned to a UE; each RB corresponds to 12 consecutive and equally spaced subcarriers. NR supports multiple numerologies with different Subcarrier Spacing (SCS) values, according to the following equation:

$$\Delta f = 15 * 2^{\mu}(kHz), \tag{1}$$

where μ is the numerology. In this work, $\mu = 0$ is used, which is oriented to eMBB applications. The packet scheduler of the gNB is in charge of properly managing the transmission parameters and the allocation of the RBs according to the CQI feedbacks received from the users. Based on the CQIs, the transmission from the gNB to the users is set with a given MCS. For each MCS, a certain spectral efficiency is achieved in the transmission. The greater is the spectral efficiency, the lower is the number of RBs required to achieve a given datarate.

Concerning spectral efficiency, subgrouping techniques are considered spectrum-efficient techniques, where each subgroup receives a service flow with

a different MCS in an assigned subset of the RBs. While OMA techniques, such as TDMA or FDMA, share the available RF channel resources (i.e., time and frequency) among the subgroups, NOMA-based subgrouping techniques use the transmitted power difference between both layers to multiplex groups.

3.1. SNR adaptation

The Injection Level (IL) is the most relevant parameter of NOMA systems when configuring the overall content distribution. IL defines the power allocation ratio between layers, which indicates how many dB separate the UL from the LL. Therefore, variations in the IL have direct implications on the available SNR values for each layer. This effect is translated in an increase on the required SNR to decode both layers compared with the case where OMA is used. In particular, the SNR required for each layer in linear units can be calculated as:

$$\gamma_i = \frac{1}{\frac{\sigma_i}{\gamma_{SL}} - \sum_{j=i+1}^{N} \sigma_j}, \tag{2}$$

where, N is the total number of layer (i.e., $N = 2$ in this work), i is the layer identifier (i.e., from 1 to N), γ_{SL} is the equivalent SNR value of the corresponding layer configuration in single-layer mode and σ_i is power allocation ratio for layer i. Following the analysis carried out in [14], σ_i can be calculated as:

$$\sigma_i = \frac{10^{\frac{\Delta_{i-1}}{10}}}{\sum_{i=1}^{N} 10^{\frac{\Delta_{i-1}}{10}}}, \tag{3}$$

where, Δ is the IL expressed in dB.

Then, taking Eq. (2) and Eq. (3), the SNR adaptation for a two-layer system in dB units can be mathematically expressed based on the single-layer case, as defined in [42]. First, the required SNR to decode the UL can be calculated as follows:

$$\gamma_{ul} = \gamma_{sl} + 10 * \log_{10}\left(\frac{1 + 10^{(\Delta/10)}}{1 - 10^{((\gamma_{sl}+\Delta)/10)}}\right), \tag{4}$$

where γ_{ul} and γ_{sl} are the required SNR to decode the UL and the equivalent single-layer signal, respectively. Regarding Eq. (4), the argument of the log function could be negative and, consequently, γ_{ul} would not be defined. To avoid this, a validity range of γ_{ul} is defined with following condition: $\gamma_{sl} + \Delta < 0$. Therefore, it is not possible to apply any Δ value. The election of the IL is conditioned by the required SNR to the decode the signal in single-layer mode. Moreover, the more robust γ_{sl} is (i.e., low SNR values), the greater is the range of applicable Δ values. In addition, as expected, the higher the IL (Δ) value, the more similar are the γ_{ul} and γ_{sl}.

Then, the required SNR to decode the LL can be also calculated based on the single-layer SNR [42]:

$$\gamma_{ll} = \gamma_{sl} - \Delta + 10\log_{10}\left(1 + 10^{(\Delta/10)}\right), \tag{5}$$

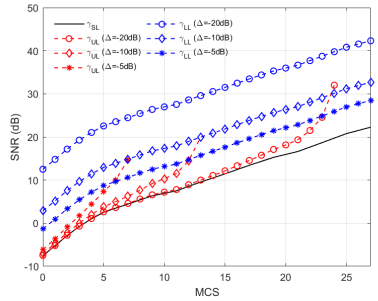


Fig. 1: SNR variation for different MCS cases, layers, and IL values.

where γ_{ll} is the required SNR to decoded the LL.

In Fig. 1, the performance degradation for both layers is shown for different ILs and MCS schemes. The x-axis represents each of the available MCS values in 5G NR according to Table 5.1.3.1-2 in [43], while in the y-axis the required SNR (in dB) value to decode each MCS is shown. The solid black line represents the single-layer case. Regarding UL, it is probed that the higher the IL is, the lower is the difference between single-layer and NOMA case. On the other hand, LL presents the opposite behavior.

3.2. User reorganization

The users' feedback is delivered by the CQI values, which are related to the maximum MCS that the UE is able to decode correctly. The CQI calculation is performed assuming a single-layer transmission mode, and therefore, before applying any NOMA-based subgrouping technique, the SNR adaptation formulas (4) and (5) have to be implemented. However, for the gNB, the CQI feedback represents a range of possible SNR. In fact, when a user requests a specific MCS, the current SNR value that the user has is inside a range of possible SNR values limited by the boundary SNR values of the MCSs. The following expression represents the SNR range: $\gamma_N \leqslant \gamma_{real} < \gamma_{N+1}$, where N is the MCS requested by the user and γ_N and γ_{N+1} are the minimum required SNR values to decode the MCS N and $N + 1$, respectively. Consequently, when the gNB receives the CQI, it cannot guess the exact SNR value that the user has in order to carry out the SNR adaptation in the NOMA case. This uncertainty could downgrade the delivered capacity or, in the worst case, let some users out of the service because they have a real SNR value below the assumed value. The uncertainty cannot be completely solved, but in this paper it is proposed to reorganize the CQIs of users in the subgrouping process to minimize its effect. In particular, three different user reorganization algorithms have been developed and included in the SNR adaptation process together with the cost function evaluation.

3.2.1. Worst case election

The first technique assumes that, within the range of possible SNR values that a user can have before requesting a specific MCS values, the user will have the worst possible. This implies that the capacity that NOMA is going to offer after the subgrouping configuration will be downgraded since not all users will require such strict SNR values within the range of possible SNRs. On the other hand, none of the users will receive a signal with a required SNR higher than their current SNR. In short, it is not a realistic assumption and it is very pessimistic.

3.2.2. Uniform SNR distribution

This algorithm takes for granted that users requesting the same MCS value will have different SNR values. Consequently, the UEs are uniformly distributed among the possible SNR values. It represents a more realistic case and the final capacity will not be severely downgraded. However, it might be the case where a particular user is assigned to a group with higher requirements than its actual SNR value.

3.2.3. Uniform distance distribution

Another alternative is to assume that the position of the users within the coverage ring related to a particular MCS value are uniformly distributed. As in the previous case, it represents a more realistic scenario and, on the other hand, the throughput will not be downgraded. Nevertheless, a particular mismatch could also happen between the real and the assumed SNR after the reorganization.

3.3. Cost function

First, a TDMA-based cost function has been developed as shown in Algorithm 1. The input is the CQI feedback (i.e., CQI_k) perceived by each k^{th} UE connected to the gNB in each Transmission Time Interval (TTI). Using these collected feedback values, the CQI distribution vector $U = \{u_1, u_2, \ldots, u_C\}$ is generated (line 8), where u_c indicates the number of UEs perceiving a CQI equal to c, c varying from 0 to 27, according to Table 5.1.3.1-2 in [43].

Afterwards, the MCS values duple $s_m = [s_{1m}, s_{2m}]$, which provides the maximum cost function value has to be found. The first item, s_{1m}, represents the MCS value assigned to the first group and s_{2m} represents the minimum supported MCS for the second group. Users with lower channel gain are grouped into the s_{1m} subgroup, whereas users under better channel condition are served through higher MCS in the s_{2m} subgroup. For each duple $[s_{1m}, s_{2m}]$, the cost function is iteratively calculated varying the number of resources N_{rb} assigned to each subgroup. Eventually, the configuration that achieves the maximum ADR is the selected (lines 10-14). The ADR is computed as the sum of the datarates of users belonging to both subgroups, according to the MCS values ($[s_{1m}, s_{2m}]$) of the selected

Algorithm 1 TDMA algorithm.

1: **Define:** N_{rb}. The number of RB available in the channel bandwidth;
2: **Define:** $m = 1, ..., M$. Set of possible configurations of couples of MCS values;
3: **Let** t_m be the rate achieved using a single resource block using a MCS value = m;
4: **Define:** $r = 1, ..., N_{rb}$. Set of possible RB assigned to the first subgroup;
5:
6: **INITIALIZATION:** Request the previously obtained CQI_k values for each user;
7:
8: **Create the CQI distribution vector:** $U = \{u_1, u_2, ..., u_C\}$;
9: **Find the best configuration of MCS levels couple and RBs distribution**:
10: **for** $m \leqslant M$ **do**
11: **for** $r \leqslant N_{rb}$ **do**
12: $ADR_{r,m} = r * s_{1_m} * t_{1_m} + (N_{rb} - r) * s_{2_m} * t_{2_m}$
13: **end for**
14: **end for**
15:
16: $[s_{1_{r,m}}, s_{2_{r,m}}] = \text{Arg max } \{ADR_{r,m}\}$;

configuration and to the number of RB assigned to each subgroup:

$$ADR = r * u_{1_m} * t_{1_m} + (N_{rb} - r) * u_{2_m} * t_{2_m}, \quad (6)$$

where r is the amount of RBs assigned to first subgroup, t_{1_m} is the rate achieved with a single resource unit (i.e., 1 RB) by users belonging the first subgroup (s_{1m}), t_{2_m} is the rate achieved with a single resource unit by users belonging the second subgroup (s_{2m}).

In [8], although NOMA was evaluated, the cost function was based on Shannon capacity formula. Therefore, in this work, several modifications have been included to take into account the numerology that imposes 5G NR [43]. The cost function implemented for carrying out 5G subgrouping using NOMA is presented in Algorithm 2. In this case, it is possible to choose one of the three available user reorganization techniques described in subsection 3.2 (i.e., NOMA (I), NOMA (II), and NOMA (III), respectively) as an additional input. The first step, line 7, is the same as in Algorithm 1. In the second step (line 8), the minimum SNR (γ_{min}) required to decode the MCS related to each CQI in single-layer mode is evaluated:

$$\gamma_{min} = (2^{eff} - 1) * (-log(5 * BER)/1.5). \quad (7)$$

The formula is obtained from [44] and according to the authors, the Bit Error Rate (BER) has been set to $5 \cdot 10^{-5}$. Then, a loop is defined for different Δ values (line 10), evaluating the rest of the steps iteratively. In order to set the range of possible injection levels (Δ), the same range as in ATSC 3.0 has been used [45]. It is

Algorithm 2 NOMA algorithm.

1: **Define:** N_{rb}. The number of RB available in the channel bandwidth;
2: **Define:** $m = 1, ..., M$. Set of possible configurations;
3: **Define:** $l = \Delta_1, ..., \Delta_L$. Set of possible injection levels;
4:
5: **INITIALIZATION:** The CQI requested by each user has been previously obtained as CQI_k;
6:
7: **Create the CQI distribution vector:**
$$U = \{u_1, u_2, ..., u_C\};$$
8: **Calculate the SNR related to each CQI according to:**
$$\gamma_{min} = (2^{eff} - 1) * (-log(5 * BER)/1.5);$$
9: **Evaluate different IL values:**
10: **for** $l \leqslant \Delta_L$ **do**
11: **Apply SNR adaptation and calculate** γ_{min-UL} **and** γ_{min-LL}
12: **Apply the corresponding SNR reorganization technique:**
$$U_{UL} = Reorganize(U, \gamma_{min-UL});$$
$$U_{LL} = Reorganize(U, \gamma_{min-LL});$$
13: **Obtain the configuration that maximizes the ADR for different** Δ **and different configurations:**
14: **for** $m \leqslant M$ **do**
$ADR_{m,l} = N_{rb} * (s_{1,m} * U_{UL,m} + s_{2,m} * U_{LL,m})$;
15: **end for**
16: **end for**
17: $[s_{1,m}, s_{2,m}]$ = Arg max $\{ADR_{m,l}\}$;

composed by 31 values, where steps of 1 dB are used for Δ between -25 dB and -5 dB and steps of 0.5 dB between -5 dB and 0 dB.

The first step inside the loop is to calculate the minimum SNR required to decode the MCS related to each CQI in the two-layer mode (γ_{min-UL} and γ_{min-LL}). This evaluation process is carried out according to the formulas (4) and (5). Once the limit SNR values are calculated, one of the three available user reorganization techniques described before is applied and new CQI distribution vectors are obtained, for UL and LL, respectively (line 12). The results obtained for each of the reorganization techniques will be compared in Section 5. Then, similar to the second step in Algorithm 1, different configurations are searched to find the one with the highest ADR. As shown in line 14, in this case the whole bandwidth is used for configuring both groups, that is, each of the NOMA layers uses N_{rb} resources. This search is iteratively carried out by varying Δ. The combination of ($s_m = [s_{1,m}, s_{2,m}]$) and Δ finally provides the output of the algorithm.

3.3.1. Complexity

In addition to the functionality of each cost function, the computational complexity is also considered another KPI in RRM algorithms. In this case, the number of operations that defines the computational burden of each cost function is based on a double iterative process, which carries out different operations inside the loop. In fact, NOMA has to perform more operations as shown in lines 12 and 13 of Algorithm 2. Regarding the iterative processes, while in the TDMA-based cost function the parameters that define the iterative process are m and r, in the case of NOMA are l and m. This means that the overall complexity of both solutions depends firstly on the number of possible solutions and, then, TDMA is conditioned by the amount of RBs that can be shared (i.e., r) and NOMA by the amount of IL values (i.e., l). Therefore, the number of iterations in TDMA is $N_{rb}b * M$, while in NOMA is $L * M$. In this particular case, the NOMA-based cost function has less iterations since the number of IL values is lower than the number of available RBs for typical 10 MHz and 20 MHz channels (31 vs. 50 and 31 vs. 100, respectively). Consequently, the NOMA-based approach presents better computational complexity characteristics than TDMA, especially when the channel bandwidth increases.

4. Simulation Environment

In this section, the simulation methodology and the use cases defined to evaluate the proposed algorithms are shortly described.

4.1. Methodology

The simulation framework is based on a dual simulation tool. On the one hand, a network level simulator is used to emulate a specific network configuration and to obtain the periodical CQI reports from the users. And on the other hand, a mathematical simulator is used to implement the algorithms presented in in section 3.3.

Regarding the network level simulator, OMNeT++ has been used [46]. Using this simulation tool, each user in the cell periodically sends its CQI to the gNB and these feedback values are saved and used as input for the second part of the simulation framework. In particular, the CQI election of the users is conditioned by the propagation channel, which is based on [47], and by the mobility model that each user follows.

Finally, concerning the mathematical simulation tool, Matlab has been used. In this part, the subgrouping algorithms described in section 3.3 are performed. Two different subgrouping algorithms have been evaluated, based on TDMA and NOMA, Algorithm 1 and Algorithm 2, respectively. Both use the CQI values obtained from the mobility model generator module as input to perform their assessment.

Table 1: Simulation parameters

Parameter	Value
Center Frequency	2 GHz
Tx Power	44 dBm
Distance attenuation	128.1 + 37.6*log(d)
Type of nodes	Pedestrians, car-mounted
Speed	Pedestrians: 3 km/h Car-mounted: 30-50 km/h
Number of Receivers	100
Mobility type	RWP
ISD	500 m
Delay Spread	90, 360, 1100 ns
Noise Power	-90 dBm
Noise Figure	9 dB
Time slot size	1 ms
Evaluated time slots	10000
SCS	15 kHz
Bandwidth	10, 20 MHz
Injection levels	As defined for ATSC 3.0 [45]
Channel model	TDL-D, TDL-E

4.2. Scenario

An urban mobile network environment has been emulated as close as possible to the real environments. Two types of users have been defined: low-speed pedestrian and car-mounted receivers. Due to the characteristics of urban environments and the user types, the Random Way Point (RWP) mobility model [48] has been implemented. In addition, the speed is also specified for each receiver type. In this case, pedestrian receivers walk inside the cell with a mean speed of 3 km/h and the car receivers randomly vary their speed between 30 and 50 km/h.

Taking into account that mobile receivers will use small size screens to display the video content, the minimum data rate for delivering High Definition (HD) or Ultra High Definition (UHD) content has been established between 1 and 4 Mbps [49]. The minimum data rate that has to be guaranteed for both groups of users will be discussed and analyzed in detail in the following section.

The CQI feedback is sent from each user to the gNB in each time slot. According to the implemented numerology (i.e., $\mu = 0$, SCS = 15 kHz), this TTI is equal to 1 ms. The rest of the network simulation parameters are summarized in Table 1.

5. Results

In this section, the most relevant results are presented and analyzed considering two approaches. In the first case, neither capacity constraint, nor user constraints have been applied. In the other cases, minimum capacity constraints and minimum percentages of served users have been defined and applied.

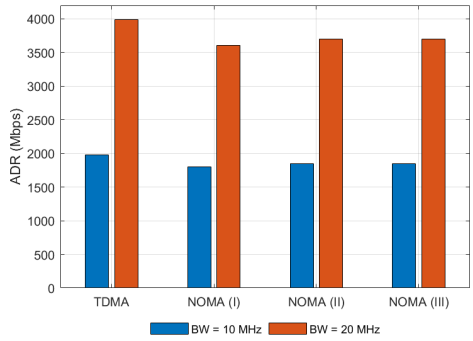


Fig. 2: Case 0 - ADR results.

5.1. Case 0: Without constraints

Firstly,in Fig. 2 ADR results are presented for car-mounted users moving with a speed between 30 km/h and 50 km/h. The vertical and the x-axis represent the ADR in Mbps and the multiplexing configurations, respectively. In the NOMA case, the three reorganization techniques presented in section 3.2 are depicted. The main conclusion is that similar results are obtained for both multiplexing technologies (i.e., TDMA and NOMA). Considering the 10 MHz channel bandwidth, results are almost equal, while in the case of 20 MHz, TDMA offers slightly better results. This is due to the limitation of the IL values, which maximum is set to -25 dB. If more separation between layers is applied, the difference between TDMA and NOMA is reduced.

However, when the mean data rate delivered by each group and the percentage of users served by each group are analyzed, as shown in Table 2, both TDMA and NOMA present unfair configurations. On the one hand, in TDMA, the algorithm assigns almost all the resources to one group and the other group receives just one RB. With this configuration the group with the lowest datarate offers 90 kbps, which is not enough to guarantee mobile video reception. On the other hand, in the NOMA-based solutions, the drawback is that due to the SNR penalization suffered in both layers, the algorithms present groups where users are not served (i.e., 38.03-38.56%) because of the high SNR requirements.

Simulations have also been carried out for pedestrian users moving with a mean speed of 3 km/h. The results have not been included in this work because, without any constraint, are almost the same showed in Fig. 2 and Table 2.

5.2. Case 1: Capacity and users constraints

In order to guarantee that each group is fairly configured, a minimum capacity and percentage of served users constraints have been defined. For the capacity constraint, four different values have been defined:

Table 2: Served users and offered capacity analysis in Case 0 results

BW	Metric	TDMA		NOMA (I)		NOMA (II)		NOMA (III)	
		G1	G2	G1	G2	G1	G2	G1	G2
10 MHz	Served users (%)	42.50	57.50	0.4	61.51	0.97	61.00	0.2	61.24
	Non-served users (%)	0		38.09		38.03		38.56	
	Data rate (Mbps)	0.08	34.88	7.83	29.59	7.75	30.66	7.86	30.42
20 MHz	Served users (%)	42.48	57.52	0.4	61.51	0.97	61.00	0.2	61.24
	Non-served users (%)	0		38.09		38.03		38.56	
	Data rate (Mbps)	0.08	70.45	15.66	59.17	15.50	61.32	15.72	60.84

1, 2, 3, and 4 Mbps. This constraint has to be guaranteed simultaneously in each group and it has been selected according to the potential users requirements described in section 4.2. If it is not possible to guarantee the constraint, a single-group configuration is applied. For the users percentage constraint, two different restrictions have been applied: 90% and 95%. This constraint represents the total number of users of the configuration, which is calculated by adding the number of users served in each group.

Fig. 3 depicts the results obtained for pedestrian and car-mounted users using 10 MHz (light blue box) and 20 MHz (light orange box) channel bandwidths. Overall, NOMA-based configurations outperform the results obtained with TDMA. Nonetheless, the difference between the technologies varies considerably depending on the constraints. If the minimum capacity constraint is used (i.e., 1 Mbps), similar results are obtained for both technologies with the 10 MHz bandwidth (i.e., results inside light blue boxes), while TDMA outperforms NOMA in the 20 MHz case (i.e., results inside light orange boxes). However, when the highest constraint is used (i.e., 4 Mbps), the gain of NOMA is maximized for all the cases. The maximum gain appears for the 10 MHz bandwidth, where the ADR that NOMA offers is more than two times the ADR of TDMA configuration: 120.7% for pedestrian users and 155.9% for car-mounted users. The main reason for these results is that NOMA performs better than TDMA when the most challenging conditions are requested due to the unbalance between the received CQIs.

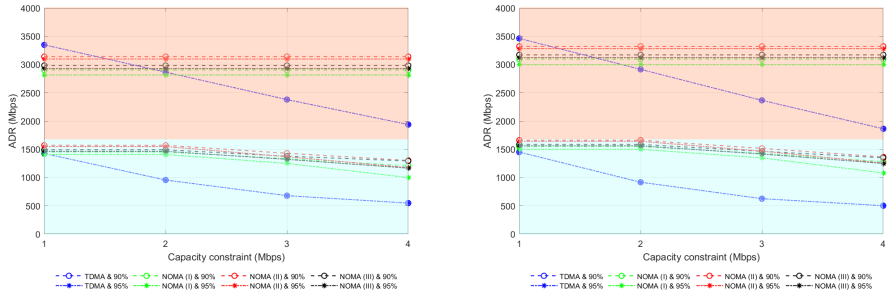
Also, the influence of the channel bandwidth can be analyzed thanks to graphical results summarized in Fig. 3. On the one hand, more extreme values are obtained using a 10 MHz bandwidth and the minimum and the maximum gains of NOMA are maximized in this case. Consequently, the gain that NOMA offers in relative terms is maximized. On the other hand, in the 20 MHz bandwidth case, results are more linear compared with the ones obtained for the 10 MHz case. In fact, results are constant for all the capacity constraints in the NOMA cases. This effect is due to the use of NOMA, which allows each group to exploit the 100% of the bandwidth resources. Therefore, if the transmission is delivered using the lowest MCS (i.e., QPSK 120/1024 and 0.2344 bps/Hz), the obtained capacity is 4.22 Mbps, which is always above the capacity constraints.

Table 3 shows the data rate that is offered by each group for some cases shown in Fig. 3. As in the case of the ADR, the gain obtained by using NOMA is maximized when the most challenging conditions are imposed. In fact, for the 10 MHz channel, the high capacity group (i.e., G2) offers up to three times the capacity offered in TDMA when 4 Mbps are required, while in the case of 1 Mbps, the gain is considerably reduced. On the other hand, the user percentage constraint does not affect TDMA cases since as neither SNR adaptation, nor user reorganization are required, all the user in the network are always served. However, in NOMA cases when the percentage of users is increased, the obtained ADR is reduced. Although the reduction varies depending on the case, it is always below 18%. The worst situation is obtained in the case of 10 MHz bandwidth with the 4 Mbps capacity constraint, where a deterioration of 17.3% is obtained for pedestrians and 15.8% for car-mounted users when increasing from 90% to 95% served users.

Finally, the difference between the proposed NOMA-based algorithms in terms of ADR should be remarked. The algorithm that assumes uniform SNR distribution (i.e., NOMA (II)) is the one that performs better. However, the difference between the NOMA-based solutions is around 10% with respect to the worst case election algorithm (i.e., NOMA (I)) and around 5% with respect to the uniform distance distribution algorithm (i.e., NOMA (III)). These results indicate that a trade-off has to be assumed between the maximum capacity that is going to be offered and the potential mismatch between the requested and the offered MCS.

6. Future work and system evolution

Although NOMA-based subgrouping techniques have better performance in comparison with TDMA-based techniques, they cannot be considered the optimum solution. In 3(a) and Fig. 3(b), the ADR results are almost constant for different capacity constraints. That is why, a solution based on the combination of NOMA and T/FDMA techniques could achieve an optimum exploitation of the resource management. In



(a) Pedestrian users.

(b) Car-mounted users.

Fig. 3: ADR results with minimum capacity and served users constraints for different configurations and mobility types.

Table 3: Example of mean data rate offered to each group for car-mounted users and 95% of served broadcast users

	TDMA		NOMA (I)		NOMA (II)		NOMA (III)	
	G1	G2	G1	G2	G1	G2	G1	G2
10 MHz & 1 Mbps	1.0	25.6	2.1	32.8	2.1	32.4	2.1	32.9
10 MHz & 4 Mbps	4.0	5.6	4.2	9.9	4.2	18.9	4.2	20.0
20 MHz & 1 Mbps	1.0	61.0	4.2	65.51	4.2	64.8	4.2	65.8
20 MHz & 4 Mbps	4.1	32.4	4.2	65.5	4.2	64.8	4.2	65.8

this section, a theoretical approach is provided in order to prove that future T/FDMA+NOMA solutions are viable.

Taking into account the algorithm developed for NOMA (see Algorithm 2), the resource allocation management implemented by NOMA techniques implies that the 100% of the channel bandwidth resources are used to simultaneously deliver the content for both multicast groups. As the low capacity group (i.e., UL - Group 1) uses all the available resources, the minimum capacity offered can be calculated as follows:

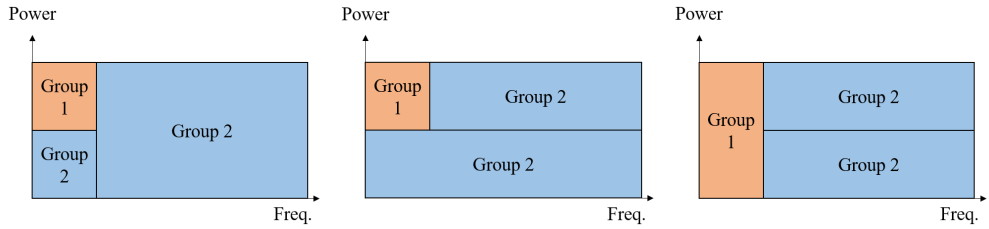
$$C_{G1} = N_{RB} \cdot BW_{RB} \cdot eff \tag{8}$$

where, BW_{RB} is the bandwidth size of a unique RB, in this case 180 kHz (i.e., 12 subcarriers with 15 kHz of subcarrier spacing). Assuming that the lowest MCS value is used for the transmission, the channel bandwidth defines the offered capacity. In the case of 10 MHz bandwidth (i.e., 50 RB), the minimum capacity that is going to be offered using NOMA is 2.11 Mbps, while if a 20 MHz bandwidth (i.e., 100 RB) is used, 4.22 Mbps can be offered. Therefore, in the case of NOMA the capacity constraint is not as relevant as in the case of T/FDMA, the capacity is always above the constraints established.

The solution is to offer a combination of both techniques. In Fig. 4, different alternatives for the combined NOMA + T/FDMA resource allocation are presented. Fig. 4(c) provides the most robust Group 1 services among the three options since single layer mode is used. On the other hand, Fig. 4(b) presents the highest spectral efficiency improvement since all the channel resources are shared in both layers. However, this alternative is the less robust solution since all the services are affected by the injection level. Then, Fig. 4(a) presents an intermediate solution, where services of the Group 2 are provided in single-layer mode and in the LL.

Doubtlessly, the RRM algorithm for each solution is different. For example, concerning 4(a), first, some channel resources are assigned to the NOMA based part to configure the low capacity group (i.e., Group 1) and will vary depending on the capacity constraint that is applied. Then, the high capacity group (i.e., Group 2) is configured in the NOMA part with the same number of RBs as in the case of Group 1. Finally, the remaining RBs are used to maximize the capacity offered to Group 2 and the transmission is carried out in single-layer mode. It is important to highlight that the portion of data delivered in the single-layer mode has to be configured to be decoded with the same SNR value (or lower) than the LL. It is quite evident that these approaches provide a more accurate RRM, in exchange for an increase in the complexity of the algorithm since more SNR and capacity values will have to be handled and, therefore, the possible solutions will considerably increase. That is why, this paper includes a theoretical analysis of the potential capacity gain that combined NOMA+T/FDMA solutions will have in future implementations.

The preliminary evaluation results are summarized in Table 4. For the calculations, a 20 MHz (i.e., 100



(a) Group 1 in UL and Group 2 in LL and single-layer (b) Group 1 in UL and Group 2 in UL and LL (c) Group 1 in single-layer and Group 2 in UL and LL

Fig. 4: Different resource allocation representations of the TDMA+NOMA solution.

Table 4: Analysis of the theoretical offered capacity using a combined T/FDMA + NOMA approach

	NOMA-only		T/FDMA+NOMA (1 Mbps)		T/FDMA+NOMA (2 Mbps)		T/FDMA+NOMA (3 Mbps)		T/FDMA+NOMA (4 Mbps)	
	# RB	Mbps	# RB	Mbps	# RB	Mbps	# RB	Mbps	# RB	Mbps
G1 - UL MCS 0	100	4.22	24	1.01	48	2.03	72	3.04	95	4.01
G2 - LL MCS 4	100	21.16	24	5.08	48	10.16	72	15.24	95	20.11
G2 - SL MCS 10	0	0	76	35.16	52	24.06	28	12.95	5	2.31
Total G2	100	21.16	100	40.24	100	34.22	100	28.19	100	22.42

RB) channel bandwidth is assumed following the approach presented in 4(a). The reference NOMA configuration is a configuration obtained from the simulations carried out in section 5.2, where Group 1 uses the lowest MCS, Group 2 is delivered with MCS 4 and the IL is -5 dB. Then, four different approaches of the T/FDMA+NOMA solutions are presented for four different capacity constraints: 1, 2, 3, and 4 Mbps. The gain obtained in the capacity of Group 2 should be highlighted. In fact, the lower the capacity constraint, the higher the gain obtained with the combined solution. For a minimum capacity equal to 1 Mbps, the overall capacity offered for Group 2 is around the double of the capacity offered in the NOMA-only mode.

7. Conclusions

A NOMA-based 5G NR subgrouping techniques have been designed and tested. Besides the design and implementation of the required algorithms, the impact on the SNR deterioration due to the use of NOMA has been detected and opportune solutions based on user reorganization techniques have been proposed.

First, it has to be highlighted that if no constraint is applied, both proposals, NOMA and TDMA, present unfair configurations from the served users or minimum capacity point of view, respectively. Then, once the fairness constraints are applied, NOMA appears as a better candidate. In general, NOMA outperforms TDMA in terms of ADR in almost all the cases. It has been demonstrated that the variables affecting the configuration (capacity and users constraints and channel bandwidth) are critical to determine the gain of NOMA. In fact, the gain of NOMA is maximized in the most challenging situations. Regarding the proposed algorithms, NOMA (II) presents the highest ADR values, although the difference among the algorithms is not critical (between 5%-10% of deterioration). Finally, both multiplexing technologies, NOMA and TDMA, have been considered as suboptimal solutions since results could be considerably improved combining TDMA and NOMA solutions in the same subgrouping algorithm. Based on the results obtained in section 6, the novel T/FDMA+NOMA solution will considerably outperform any of the proposed solutions that use a single multiplexing technology.

There are still some future works on the horizon. On the one hand, the performance of these subgrouping algorithms should be analyzed from the network layer perspective by using a network simulator. This approach could lead to a novel PHY/MAC analysis of the subgrouping techniques. In addition to the KPIs used in this work (ADR or data rate per group), KPIs related with the PHY and the MAC layers could be used, such as latency, effective throughput or Packet Error Rate (PER). On the other hand, the combination of TDMA and NOMA solutions should be analyzed more in detail and the influence of the rest of the affecting parameters should be studied. Then a simulation platform that integrates a new cost function for

TDMA+NOMA case should be developed.

Acknowledgements

This work was supported by the Italian Ministry of University and Research (MIUR), within the Smart Cities framework, Project Cagliari2020 ID: PON04a2_00381, by the Basque Government (the grant IT1234-19 and the PREDOC grant program PRE_2019_2_0037) and by the Spanish Government (project PHANTOM under the grant RTI2018-099162-B-I00 (MCIU/AEI/FEDER, UE)).

References

[1] 3GPP "ETSI TR 121 915 V15.0.0 – Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; 5G; Release description; Release 15" Oct. 2019.
URL https://www.etsi.org/deliver/etsi_tr/121900_121999/121915/15.00.00_60/tr_121915v150000p.pdf

[2] A. Gosh, A. Maeder, M. Baker, D. Chandramouli, 5G evolution: View on 5G cellular technology beyond 3GPP release 15, IEEE Access 7 (2019) 127639–127651. doi:10.1109/ACCESS.2019.2939938.

[3] G. Araniti, M. Condoluci, M. Cotronei, A. Iera, A. Molinaro, A Solution to the Multicast Subgroup Formation Problem in LTE Systems, IEEE Wireless Commun. Letters 4 (2) (2015) 149–152. doi:10.1109/ACCESS.2019.2939938.

[4] L. Militano, D. Niyato, M. Condoluci, G. Araniti, A. Iera, G. M. Bisci, Radio resource management for group-oriented services in lte-a, IEEE Transactions on Vehicular Technology 64 (8) (2015) 3725–3739.

[5] M. Condoluci, G. Araniti, A. Molinaro, A. Iera, Multicast resource allocation enhanced by channel state feedbacks for multiple scalable video coding streams in lte networks, IEEE Transactions on Vehicular Technology 65 (5) (2016) 2907–2921.

[6] E. Iradier, J. Montalban, L. Fanari, P. Angueira, L. Zhang, Y. Wu, W. Li, Using noma for enabling broadcast/unicast convergence in 5g networks, IEEE Transactions on Broadcasting 66 (2) (2020) 503–514.

[7] S. M. R. Islam, N. Avazov, O. A. Dobre, K.-s. Kwak, Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges, IEEE Commun. Surveys & Tutorials 19 (2) (2017) 721–742. doi:10.1109/COMST.2016.2621116.

[8] J. Montalban, P. Scopelliti, M. Fadda, E. Iradier, C. Desogus, P. Angueira, M. Murroni, G. Araniti, Multimedia Multicast Services in 5G Networks: Subgrouping and Non-Orthogonal Multiple Access Techniques, IEEE Communications Magazine 56 (3) (2018) 91–95. doi:10.1109/MCOM.2018.1700660.

[9] R. O. Afolabi, A. Dadlani, K. Kim, Multicast Scheduling and Resource Allocation Algorithms for OFDMA-Based Systems: A Survey, IEEE Communications Surveys Tutorials 15 (1) (2013) 240–254. doi:10.1109/SURV.2012.013012.00074.

[10] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, L. Hanzo, Nonorthogonal Multiple Access for 5G and Beyond, Proceedings of the IEEE 105 (12) (2017) 2347–2381.

[11] R. Hoshyar, F. P. Wathan, R. Tafazolli, Novel low-density signature for synchronous CDMA systems over AWGN channel, IEEE Transactions on Signal Processing 56 (4) (2008) 1616–1626.

[12] F. Brannstrom, T. M. Aulin, L. K. Rasmussen, Iterative detectors for trellis-code multiple-access, IEEE Transactions on Communications 50 (9) (2002) 1478–1485.

[13] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, H. V. Poor, Application of non-orthogonal multiple access in LTE and 5G networks, IEEE Communications Magazine 55 (2) (2017) 185–191.

[14] L. Zhang, W. Li, Y. Wu, S.-I. Wang, Xianbin Park, H. M. Kim, J.-Y. Lee, P. Angueira, J. Montalban, Layered-Division-Multiplexing: Theory and Practice, IEEE Transactions on Broadcasting 62 (1) (2016) 110964–110971. doi:0.1109/TBC.2015.2505408.

[15] T. Cover, Broadcast channels, IEEE Transactions on Information Theory 18 (1) (1972) 2–14.

[16] P. Bergmans, T. Cover, Cooperative broadcasting, IEEE Transactions on Information Theory 20 (3) (1974) 317–324.

[17] R. Zhang, L. Hanzo, A unified treatment of superposition coding aided communications: Theory and practice, IEEE communications surveys & tutorials 13 (3) (2010) 503–520.

[18] P. Gandotra, R. K. Jha, S. Jain, Green NOMA with multiple interference cancellation (MIC) using sector-based resource allocation, IEEE Transactions on Network and Service Management 15 (3) (2018) 1006–1017.

[19] D. Gómez-Barquero, O. Simeone, LDM versus FDM/TDM for unequal error protection in terrestrial broadcasting systems: An information-theoretic view, IEEE Transactions on Broadcasting 61 (4) (2015) 571–579.

[20] J. Montalban, I. Angulo, C. Regueiro, Y. Wu, L. Zhang, S.-I. Park, J.-Y. Lee, H. M. Kim, M. Velez, P. Angueira, Performance study of layered division multiplexing based on SDR platform, IEEE Transactions on Broadcasting 61 (3) (2015) 436–444.

[21] C. Regueiro, J. Montalban, J. Barrueco, M. Velez, P. Angueira, Y. Wu, L. Zhang, S.-I. Park, J.-Y. Lee, H. M. Kim, LDM core services performance in ATSC 3.0, IEEE Transactions on Broadcasting 62 (1) (2016) 244–252.

[22] S.-I. Park, J.-Y. Lee, B.-M. Lim, Y. Kim, S. Kwon, H. M. Kim, J. Kim, Field test results of layered division multiplexing for the next generation DTV system, IEEE Transactions on Broadcasting 63 (1) (2016) 260–266.

[23] S.-I. Park, J.-Y. Lee, B.-M. Lim, S. Kwon, J.-H. Seo, H. M. Kim, N. Hur, J. Kim, Field comparison tests of LDM and TDM in ATSC 3.0, IEEE Transactions on Broadcasting 64 (3) (2017) 637–647.

[24] C. Suh, J. Mo, Resource allocation for multicast services in multicarrier wireless communications, IEEE Transactions on Wireless Communications 7 (1) (2008) 27–31. doi:10.1109/TWC.2008.060467.

[25] G. Araniti, M. Condoluci, L. Militano, A. Iera, Adaptive Resource Allocation to Multicast Services in LTE Systems, IEEE Transactions on Broadcasting 59 (4) (2013) 658–664. doi:10.1109/TBC.2013.2271387.

[26] G. Araniti, V. Scordamaglia, M. Condoluci, A. Molinaro, A. Iera, Efficient Frequency Domain Packet scheduler for Point-to-Multipoint transmissions in LTE networks, in: 2012 IEEE International Conference on Communications (ICC), 2012, pp. 4405–4409. doi:10.1109/ICC.2012.6364383.

[27] C. Casetti, C.-F. Chiasserini, F. Malandrino, C. Borgiattino, Area formation and content assignment for LTE broadcasting, Computer Networks 126 (2017) 174–186. doi:10.1016/j.comnet.2017.07.006.

[28] M. Algharem, M. H. Omar, R. F. Rahmat, R. Budiarto, A Low-Complexity Subgroup Formation with QoS-Aware for Enhancing Multicast Services in LTE Networks, in: 2nd International Conference on Computing and Applied Informatics 2017. doi:10.1088/1742-6596/978/1/012094.

[29] G. Araniti, I. Bisio, M. De Sanctis, F. Rinaldi, A. Sciarrone, Joint Coding and Multicast Subgrouping Over Satellite-eMBMS Networks, IEEE Journal on Selected Areas and Communications 36 (5) (2018) 1004–1016. doi:10.1109/JSAC.2018.2832818.

[30] A. Labella, R. M. Rodriguez, G. De Tre, L. Martinez, A Cohesion Measure for Improving the Weighting of Experts' subgroups in Large-scale Group Decision Making Clustering Methods, in: 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), New Orleans, LA, USA, 23-26 June

2019. doi:10.1109/FUZZ-IEEE.2019.8858858.

[31] N. Benenati, C. Desogus, P. Scopelliti, E. Iradier, J. Montalban, M. Murroni, G. Araniti, P. Angueira, M. Fadda, Group-oriented broadcasting of augmented reality services over 5g new radio, in: 2019 IEEE Broadcast Symposium (BTS), 2019, pp. 1–7. doi:10.1109/BTS45698.2019.8975412.

[32] G. Araniti, F. Rinaldi, P. Scopelliti, A. Molinaro, A. Iera, A Dynamic MBSFN Area Formation Algorithm for Multicast Service Delivery in 5G NR Networks, IEEE Transactions on Wireless Communications 19 (2) (2020) 808–821. doi:10.1109/TWC.2019.2948846.

[33] A. Z. Yalcin, M. Yuksel, B. Clerckx, Rate splitting for multi-group multicasting with a common message, IEEE Transactions on Vehicular Technology (2020) 1–1.

[34] B. Zhang, Y. Chen, J. Yu, Z. Han, An indirect-reciprocity-based incentive framework for multimedia service through device-to-device multicast, IEEE Transactions on Vehicular Technology 68 (11) (2019) 10969–10980.

[35] E. Iradier, J. Montalban, G. Araniti, M. Fadda, M. Murroni, Adaptive resource allocation in lte vehicular services using ldm, in: 2016 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), 2016, pp. 1–6. doi:10.1109/BMSB.2016.7521985.

[36] L. Lv, J. Chen, Q. Ni, Cooperative Non-Orthogonal Multiple Access in Cognitive Radio, IEEE Communications Letters 20 (10) (2016) 2059–2062. doi:10.1109/LCOMM.2016.2596763.

[37] M. Moltafet, N. Mokari, M. R. Javan, H. Saeedi, H. Pishro-Nik, A New Multiple Access Technique for 5G: Power Domain Sparse Code Multiple Access (PSMA), IEEE Access 6 (2017) 747–759. doi:10.1109/ACCESS.2017.2775338.

[38] E. M. Almohimmah, M. T. Alresheedi, A. F. Abas, J. Elmirghani, A Simple User Grouping and Pairing Scheme for Non-Orthogonal Multiple Access in VLC System, in: 20th International Conference on Transparent Optical Networks (ICTON), 1-5 July 2018, Bucharest, Romania. doi:10.1109/ICTON.2018.8473907.

[39] Y. Xue, A. Alsohaily, E. Sousa, W. Li, L. Zhang, Y. Wu, Using layered division multiplexing for mixed unicast-broadcast service delivery in 5g, in: 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), 2019, pp. 1–6. doi:10.1109/BMSB47279.2019.8971940.

[40] Y. Yin, Y. Peng, M. Liu, J. Yang, G. Gui, Dynamic User Grouping-Based NOMA Over Rayleigh Fading Channels, IEEE Access 7 (2019) 110964–110971. doi:10.1109/ACCESS.2019.2934111.

[41] H. Sari, A. Maatouk, E. Caliskan, M. Assaad, M. Koca, G. Gui, On the foundation of NOMA and its application to 5G cellular networks, in: 2018 IEEE Wireless Communications and Networking Conference (WCNC), 15-18 April 2018, Barcelona, Spain. doi:10.1109/WCNC.2018.8377142.

[42] J. Montalbán, L. Zhang, U. Gil, Y. Wu, I. Angulo, K. Salehian, S.-I. Park, B. Rong, W. Li, H. M. Kim, et al., Cloud transmission: System performance and application scenarios, IEEE Transactions on Broadcasting 60 (2) (2014) 170–184.

[43] 3GPP TS 38.214 v15.3.0, Tech. Spec. Group Services and System Aspects, NR; Physical layer procedures for data (Release 15), Tech. rep. (Sept. 2018).

[44] A. J. Goldsmith, S.-G. Chua, Variable-rate variable-power MQAM for fading channels, IEEE transactions on communications 45 (10) (1997) 1218–1230.

[45] Advanced Television Systems Committee and others, ATSC Standard: Physical Layer Protocol (A/322). Doc. A/322: 2106, Washington, DC, USA.

[46] A. Varga, R. Hornig, An overview of the OMNeT++ simulation environment, in: Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 2008, p. 60.

[47] 3GPP TR 38.901, Radio Access Network Working Group and others, Study on channel model for frequencies from 0.5 to 100 ghz (release 15), Tech. rep. (2018).

[48] E. Hyytiä, H. Koskinen, P. E. Lassila, A. Penttinen, J. T. Virtamo, J. Roszik, Random waypoint model in wireless networks, 2005.

[49] X. Xu, J. Liu, X. Tao, Mobile edge computing enhanced adaptive bitrate video delivery with joint cache and radio resource allocation, IEEE Access 5 (2017) 16406–16415.

3.1.5 Journal paper J3

This subsection presents a journal paper related with Contribution 1. The full reference of the paper is presented below:

- E. Iradier, A. Abuin, R. Cabrera, I. Bilbao, J. Montalban and P. Angueira, "Advanced NOMA-based RRM Schemes for 5G mmWave Bands," submitted manuscript to IEEE Transactions on Broadcasting.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Journal paper indexed in JCR and IEEExplore
- Area: Electrical & Electronic Engineering
- Ranking: 69/266 (Q2) based on JCR 2019
- Impact factor (JCR): 3.419

IEEE TRANSACTIONS ON BROADCASTING, VOL. X, NO. Y, MAY 2021 1

Advanced NOMA-based RRM Schemes for 5G mmWave Bands

Eneko Iradier, *Graduate Student Member, IEEE,* Aritz Abuin, *Graduate Student Member, IEEE,* Rufino Cabrera, *Graduate Student Member, IEEE,* Iñigo Bilbao, *Graduate Student Member, IEEE,* Jon Montalban, *Senior Member, IEEE,* Pablo Angueira, *Senior Member, IEEE,*

***Abstract*—A relevant solution for the high demand for new multimedia applications and services is millimeter frequency bands (mmWave) in 5G. However, in order to face the technological challenges of the present and those that will appear in the short-term future, it is necessary to improve the spectral efficiency of 5G systems. In particular, the Radio Resource Management (RRM) module is considered an essential aspect. However, resource allocation techniques that combine orthogonal multiplexing (OMA) schemes, such as Time Division Multiplexing (TDM), with Non-Orthogonal Multiple Access (NOMA) techniques have not been studied in depth. Therefore, this article designs and evaluates different RRM models that combine TDM with NOMA for innovative applications in the 5G mmWave bands. To this end, the advantages and challenges associated with mmWave bands and potential future applications are introduced. To test the RRM models, an innovative use case has been designed based on the on-demand distribution of multimedia content in high-density environments, such as museum halls. The models have been evaluated in terms of throughput, capacity, and availability. According to the results, the combined RRM techniques offer up to 50% more capacity than the single multiplexing technology models and can provide video-on-demand service to practically the entire cell.**

***Index Terms*—5G, AR, broadcast, LDM, millimeter bands, mmWaves, NOMA, personalization, resource allocation, RRM, unicast.**

I. Introduction

Serving the constant evolution of multimedia services requested by users has become one of the most relevant challenges of today's wireless communication networks. The emergence of new applications such as Augmented Reality (AR) and Virtual Reality (VR) or the increase in the definition of video services (4k/8k) has caused the data rate required by the end-user to grow exponentially during the last years. In fact, according to recent studies, around three-fourths of the data consumed in the world is expected to be video by the end of 2021 [1].

5G is expected to be the technology that leads the implementation of the most challenging and demanded use cases. In fact, the Radiocommunication Sector of the International Telecommunication Union (ITU-R) has divided the entire application frame into three use case families: enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communications (URLLC), and massive Machine Type Communications (mMTC) [2]. Each of these three families has very different requirements. For example, eMBB, which is based on traditional multimedia delivery, requires around 10 Gbps of peak data rate values for the uplink and around 20 Gpbs for the downlink. Moreover, high mobility rates have to be also supported (up to 500 km/h). Then, URLLC is oriented to mission-critical applications, where reliability (loss rates below 10^{-9}) and end-to-end latency (up to 1 ms) are the most critical Key Performance Indicators (KPI). Finally, mMTC is focused on covering extremely crowded networks, such as Internet of Things (IoT) or sensor networks. In this case, density rates above 10^6 devices per km^2 have to be guaranteed.

To serve such a broad set of applications, the first version of 5G New Radio (NR) [3], presents an increase in the robustness of the services thanks to the Low-Density Parity-Check (LDPC) codes and better use of radio resources. Additionally, 5G implements a flexible physical layer (PHY) in which the parameters can be configured to suit the latency or capacity needs required by the application. Another of the most representative novelties is the switch to the millimeter-wave (mmWave) frequency bands, where bandwidths of hundreds of MHz can be allocated [4].

In addition to going up in frequency, an improvement in the spectral efficiency of 5G can facilitate the implementation of the most demanding use cases. In particular, Non-Orthogonal Multiple Access (NOMA) techniques have presented a very promising performance vs. complexity trade-off during the last years. In fact, an excellent example of the success of NOMA happened in 2016, when a low-complexity solution (Layered Division Multiplexing, LDM) was accepted to be part of the PHY of the latest Advanced Television Systems Committee (ATSC) standard [5]. Moreover, NOMA solutions have also been proposed to be combined with broadband standards such as 4G [6] or 5G [7], based on the capacity advantage that they offered in comparison with traditional Time or Frequency Domain Multiplexing (TDM/FDM) solutions. However, recent studies have discovered that although the Radio Resource Management obtained with NOMA techniques outperforms TDM or FDM, it is still a non-optimal solution [8]. As opposed to TDM/FDM, the main reason for this is that NOMA solutions are short of flexibility when Resource Blocks are allocated. Therefore, the combination of NOMA and TDM/FDM techniques in the same RRM scheme could optimize the resource allocation issue.

Eneko Iradier, Aritz Abuin, Rufino Cabrera, Iñigo Bilbao, Jon Montalban, and Pablo Angueira are with the Department of Communications Engineering, University of the Basque Country (UPV/EHU), Torres Quevedo 1, 48012 Bilbao, Spain (email: eneko.iradier@ehu.eus, aabuin003@ikasle.ehu.eus, rufinoreydel.cabrera@ehu.eus, inigo.bilbao@ehu.eus, jon.montalban@ehu.eus, pablo.angueira@ehu.eus).

This paper aims to analyze the possible RRM schemes that combine NOMA and TDM techniques for 5G mmWave bands. The advantages and disadvantages of the implementation of 5G in mmWave bands have to be analyzed. Then, different NOMA+TDM RRM schemes have to be designed and evaluated under novel use cases oriented to potential applications for mmWave bands. In summary, the technical contributions of this paper include:

1) Analysis of the challenges and benefits of using mmWave bands in 5G, as well as a description of the potential applications.
2) Design of different RRM schemes that combine NOMA and TDM multiplexing.
3) Evaluation of the proposed RRM schemes under an innovative use case for future mmWave band applications.

The rest of the paper is organized as follows. The next section describes the related works published up to now on this topic. Section III is focused on the overview of the impact of the mmWave bands in 5G and the potential future applications. In Section IV, the proposed NOMA+TDM RRM schemes are designed and detailed. Then, in Section V the use case is introduced, and Section VI presents the results obtained with the RRM schemes in the proposed use case based on different evaluation metrics. Finally, Section VII contains the conclusions.

II. RRM in mmWave Bands: Related Work

Increasing the center frequency of the services enables the use of bandwidths greater than previous standards. Therefore, the design of RRM algorithms that efficiently manage a large number of radio resources is crucial [9]. This section presents some of the most relevant studies in the literature concerning RRM in the mmWave bands.

One of the main concerns of the applications carried out in the mmWave bands is to coordinate the access of different service requests. For example, in [10], the resource allocation strategy is evaluated in mixed multicast/unicast traffic environments using queuing theory. To do so, a mathematical model is developed to optimize the 5G parameters. The authors define a set of conditions that considerably affect the unicast RRM management, such as significant unicast requests or high spatial arrival intensity. Another alternative to handle users/services with different characteristics is to use heterogeneous networks. In particular, in [11], the authors study the resource allocation in multiple heterogeneous networks in the mmWave bands, where each band has specific propagation conditions. Two different access techniques are proposed: the single-band access scheme and the multi-band access scheme. The first approach is based on the time fraction allocation, whereas the second is based on a Markov model framework. Results indicate that multi-band schemes outperform single-band solutions, especially when the network is in light load mode.

In addition to the networks implemented in the mmWave bands, many applications will continue to be based on sub-6 GHz networks, mainly due to the lower path losses. Therefore, the correct coordination of both frequency bands and managing the resources dedicated to each of the bands is another of the RRM challenges. This issue is discussed in [12], where resource allocation techniques are studied inside a broader framework that also considers relay address, beam selection, and interference coordination. In this case, the sub-6 GHz band and the mmWave band are used simultaneously with different objectives. While the first is used for control and reliable communications, the latter is used for high throughput applications. The authors evaluate the coordinated management of the multi-parameter framework in both frequency bands at the same time. Then, in [13], the authors propose a diverse network, where sub-6 GHz macrocells and mmWave bands small cells co-exist. The study targets to obtain an energy-efficient RRM algorithm using uplink NOMA in ultradense Internet of Thing (UD-IoT) environments. Results indicate that the proposed algorithm outperforms the existing works.

Then, another relevant issue of the mmWave bands is the correct implementation and optimum use of the backhaul signal. Concerning this challenge, in [14], the Radio Access (RA) and the backhaul (BH) are implemented in the same mmWave band, and so, both services are unified in the same network. The authors present three dynamic RRM algorithms to manage the resource allocation between RA and BH for several pico stations together. Algorithms are oriented to enhance the network capacity and guarantee fairness among stations. The results indicate that the network capacity could be increased up to 35% when the proposed dynamic algorithms are used.

Finally, the RRM algorithms oriented to mmWave bands have to be adapted to novel use cases, such as Unmanned Aerial Vehicle (UAV) applications. For example, in [15], the RRM performance is evaluated in downlink-UAV networks that implement antenna sectoring approaches. In this case, the goal is to maximize the network capacity while ensuring a minimum data rate to each user and limiting the transmission power.

Although several works have been presented in the literature about different RRM algorithms in mmWave bands, the potential gain obtained by combining non-orthogonal techniques with orthogonal techniques, such as TDM or FDM, has not been studied. This work studies and evaluates different combined NOMA+TDM RRM alternatives for a use case applicable in networks based on mmWave bands.

III. 5G in mmWave Bands

This section discusses the main implications of the implementation of 5G in mmWave bands. The propagation channel conditions and the potential gains are presented and the description of the potential applications.

A. Configuration

The 5G configuration is different in the millimeter frequency band and the sub-6GHz band. Therefore, this subsection analyzes the main differences and their impact on the system.

Firstly, it should be noted that although technically the radioelectric spectrum defined as millimeter frequency band is between 30 GHz and 300 GHz (wavelengths between 10 mm and one mm), in the 5G ecosystem, mmWave bands are

TABLE I: Influence of the numerology value over different waveform parameters

μ	SCS (kHz)	#Slot/Subframe	Slot (ms)	OFDM Symbol (μs)	CP (μs)	Total Length (μs)	RB_{MAX}	BW_{MAX} (MHz)
0	15	1	1	66.67	4.69	71.36	275	49.5
1	30	2	0.5	33.33	2.34	35.67	275	99
2	60	4	0.25	16.67	1.17	17.84	275	198
3	120	8	0.125	8.33	0.57	8.90	275	396
4	240	16	0.0625	4.17	0.29	4.46	138	397.44

understood as the frequencies above 6 GHz [16]. In particular, the frequency band reserved for mmWaves is defined as the Frequency Range 2 (FR2) and is comprised between 24.25 GHz and 52.6 GHz. According to the latest announcements from governments and world spectrum regulators, the first 5G in mmWave bands systems are expected to be implemented in the 26 GHz (3GPP n258), 28 GHz (3GPP n257 and n261), and 39 GHz (3GPP n260) bands [4].

Moreover, as mentioned before, one of the main advantages of 5G is the PHY layer flexibility, which uses different configurations depending on the scope of the application. One of the key parameters to define the 5G waveform is the subcarrier spacing (SCS). SCS is defined from a set of possible numerologies (μ) as follows:

$$SCS(kHz) = 15 \cdot 2^{\mu} \tag{1}$$

where μ represents numerology and can be 0, 1, 2, 3, or 4. As shown in Table I, the numerology also defines the slot duration and the symbol length. Furthermore, these parameters present the opposite behavior if compared with SCS. That is, the higher the SCS, the shorter the slot and symbol durations.

One of the consequences of using high frequencies is that the Doppler shift is also increased. To avoid a negative impact on the received signal, it is recommended to use high μ configurations (μ = 3 or μ = 4) [17]. These configurations imply that the minimum SCS is 120 kHz and that the symbol time, including the Cyclic Prefix (CP), does not exceed 10 μs.

As shown in Table I, the numerology also determines the maximum bandwidth that can be used in each case. Except for μ = 4, the maximum number of RBs that can be reserved in the rest of the cases is 275 (for μ = 4, it is 138). The following expression is used to calculate the maximum bandwidth in each case:

$$BW = NRB \cdot BW_{RB} = NRB \cdot 12 \cdot SCS \tag{2}$$

where NRB is the number of RB allocated and BW_{RB} is the bandwidth of a single RB, which is equal to the size of 12 subcarriers separated SCS kHz. Therefore, following the recommended configurations for the mmWave bands, the maximum bandwidths offered are close to 400 MHz. Moreover, in FR2, channels can be aggregated, increasing the overall bandwidth.

Lastly, it should be remarked that the 5G waveform configurations for the mmWave bands affect the RRM process. In particular, as high numerologies have to be used, the symbol time is short, and, consequently, the Transmission Time Intervals (TTIs) are shorter. Therefore, RRM algorithms must guarantee a balance between performance and frequency of execution. For this reason, in general, RRM algorithms can be classified into Fast RRM since they are executed in each TTI, and Slow RRM, which are those that are executed with a lower frequency [18].

B. Propagation Challenges

The propagation channel is a challenge in itself for any wireless communication medium. Most communication systems have been implemented below 6 GHz, and, therefore, there is extensive knowledge. However, in the case of the mmWave bands, some behaviors/effects are different from those assumed so far. Therefore, this section summarizes the main characteristics and differences of the propagation channels of the mmWave bands with respect to traditional microwave systems.

One of the main consequences of the increase in frequency is that the path losses grow considerably. Specifically, taking free space losses as a reference, increasing the frequency by an order of magnitude (i.e., x10) implies having an additional 20 dB of losses. However, since path losses do not depend solely on frequency, there are propagation models for different frequency ranges. Regarding the mmWave bands, the 3GPP TR 38.901 [19] channel model models different path losses depending on the application scenario. Among them, the models Urban Microcells (UMi) and Urban Macrocells (UMa) represent traditional cellular networks, and Indoor Hotspot (InH) and Indoor Factory (InF) are used for different indoor facilities. Fig. 1 shows the evolution of path losses as a function of distance for the UMi and UMa models, assuming Line-of-Sight (LOS) and non-LOS (NLOS), for the 30 GHz band (Fig. 1(a)) and the 60 GHz band (Fig. 1(b)). Furthermore, in [19], the probability of having LOS conditions is also modeled as a function of the scenario. In particular, these probability distributions have a negative exponential trend, and the UMa model has a higher probability of having a LOS condition than the UMi model.

Another vital factor in mmWave bands is that as they have smaller wavelengths than microwaves, they are more sensitive to obstacles [20]. Specifically, according to [21], obstacles in indoor environments or walls show higher penetration losses in the mmWave bands. In addition, when placing objects between the transmitter and the receiver in the mmWave bands communications, the characteristics of the channel vary considerably, and the shadowing effects are very damaging. The deterioration of the signal due to meteorological causes is another effect due to the wavelength size. In particular, when exceeding 10 GHz, meteorological effects such as rain, hail, or snow cause attenuations that must be taken into account [22].

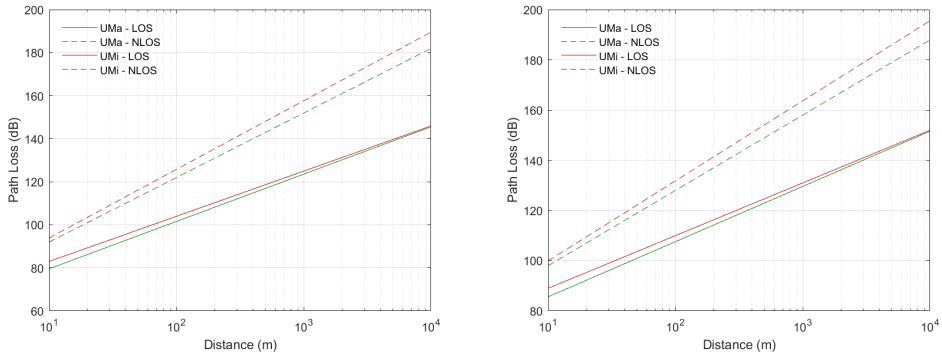


Fig. 1: Path loss attenuation following 3GPP TR 38.901 for different frequency bands: (a) 30 GHz, (b) 60 GHz.

C. Benefits & Drawbacks

One of the main benefits of the mmWave bands is the increase in capacity offered because the available bandwidths are much higher than those of sub-6GHz bands. This effect is reflected in two parameters: the number of users served, and the throughput offered. Specifically, the first parameter is analyzed in more detail in Table II.

To carry out the analysis, two typical bandwidths of 5G communications in the sub-6GHz bands (i.e., 50 and 100 MHz) and two other bandwidths related to the mmWave bands (i.e., 400 MHz and 1 GHz) have been chosen. In addition, since they determine the transmitted throughput, three MCS values have also been selected according to Table 5.1.3.1-2 in [23]: the minimum (i.e., MCS0, QPSK 120/1024), an intermediate one (i.e., MCS14, 64QAM 616/1024) and the maximum (i.e., MCS28, 256QAM 948/1024). Then, in order to contextualize the analysis, three relevant use cases for the future 5G communications with completely different capacity constraints have been used as reference. The first is Internet of Things (IoT), and the characteristics defined in the Rel-13 (Narrow Band IoT) have been taken into account, where a maximum data rate of 200 kbps per user in downlink channels is allowed [24]. The second is the reception of video content in mobile and portable devices, which, according to [25], a minimum data rate of 2 Mbps is required for high-quality video content. Finally, AR/VR applications have been used due to their promising future. In this case, the necessary capacity is usually related to a wide range, where the highest quality applications can reach up to 100 Mbps and the ones of basic qualities between 20 and 30 Mbps [26]. For the analysis, 25 Mbps has been chosen as a potential representative value of the mobile AR/VR applications. As it can be seen in Table II, the increase in bandwidth leads to a proportional increase in the number of users served, which occurs in applications based on mmWave bands. In fact, the number of users served on the channels oriented to the mmWaves bands is considerably higher regardless of the MCS used. In addition, it should be noted that if due to the channel or application conditions, it is necessary to use the most robust MCS, the performance obtained with the channels of the sub-6GHz bands is very limited (e.g., it is not possible to implement AR/VR). In contrast, in the mmWave bands, access to a relevant number of users in each case can be guaranteed.

On the contrary, the main disadvantages of mmWave bands are the consequences derived from the increase in the path losses suffered by the transmissions. Therefore, to compensate more gNBs must be entered in the mmWave bands than in the sub-6GHz bands. For example, taking as a reference a network with a coverage radius of 1km (i.e., an area of 3140 km^2) and a central frequency of 2 GHz, if it is assumed that the path loss model that affects it is the UMa-LOS shown in Section III-B, the path loss at the edge of the cell is 100 dB. However, if the network is based on 30 GHz transmissions and the same path loss value must be guaranteed, the cell edge is located at 85 meters with an area of 22.7 km^2. Consequently, the number of gNBs needed to cover that area would be 138 times greater.

Undoubtedly, due to the presented facts, mmWave bands can be considered ideal for creating small networks (e.g., picocells or femtocells) where additional network access should be guaranteed or additional capacity has to be offered.

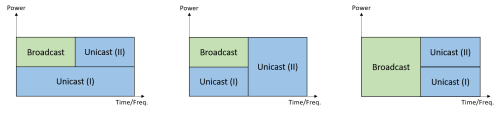
D. Potential Applications

The increase in the frequency of implementation of 5G services and, as a consequence, the considerable increase in the channel bandwidths used allows challenging applications to be possible. This subsection shows some of the applications that will possibly be implemented in the 5G mmWave bands.

One of the straightforward applications of 5G mmWave bands is to offer traditional broadband applications with considerably enhanced qualities [27]. For example, applications that require high data rates such as AR/VR content, 4k/8k video content, or 360° video streaming may be easier to implement on the mmWave bands.

TABLE II: Analysis of the maximum number of users that could be served using different bandwidth sizes for different applications

MCS	Efficiency	Application	Constraint	Users in 50 MHz	Users in 100 MHz	Users in 400 MHz	Users in 1 GHz
MCS0	0.2344 bps/Hz	NB-IoT	200 kbps	58	117	468	1172
		Mobile HD video	2 Mbps	5	11	46	117
		AR/VR	25 Mbps	0	0	3	9
MCS14	3.6094 bps/Hz	NB-IoT	200 kbps	902	1804	7218	18047
		Mobile HD video	2 Mbps	90	180	721	1804
		AR/VR	25 Mbps	7	14	57	144
MCS28	7.4063 bps/Hz	NB-IoT	200 kbps	1851	3703	14812	37031
		Mobile HD video	2 Mbps	185	370	1481	3703
		AR/VR	25 Mbps	14	29	118	296



(a) (b) (c)

Fig. 2: Representation of the proposed RRM schemes: (a) Model A, (b) Model B, (c) Model C.

Another possible application is to use the 5G mmWave bands to guarantee access to content in high-attendance environments [28]. Stadiums, concerts, exhibitions, or festivals are clear examples of locations where the density of users is very high and traditional networks are not always capable of serving all users correctly. Therefore, using the 5G mmWave bands in specific areas with large bandwidths considerably reduces the possibility of network saturation. Furthermore, this alternative could be used in combination with existing small networks or could replace the access networks of certain places.

Finally, the use of the 5G mmWave bands for backhaul applications should be highlighted [29]. Although fiber optics systems have traditionally been used for the backhaul of wireless networks, the reduction in the coverage area in the mmWave bands and the increase in the number of networks generates an increase in the cost of implementation and scalability that may not be assumable. Therefore, using frequency bands above 6 GHz for the backhaul is a real alternative. In particular, the use of 60 GHz unlicensed bands with LOS conditions for backhaul applications is understudy [28].

IV. Proposed RRM Schemes

As mentioned in [8], pure TDMA and pure NOMA RRM schemes can be considered suboptimal solutions, and the design of combined TDMA and NOMA RRM solutions should provide better performance rates. Therefore, the objective of this section is to propose different combined RRM schemes.

In this work, three different RRM schemes are proposed, and Fig. 2 presents the representation of each model. In general, each of the models is based on three different blocks: broadcast, unicast (I), and unicast (II). It should be noted that the unicast (I) and unicast (II) blocks are a set of resources dedicated to the transmission of unicast services, so each of the blocks is divided into multiple individualized unicast services. The main difference between the three models is the combination of the multiplexing schemes. For example, Model A (see Fig. 2(a)) is based on a NOMA scheme where the Upper Layer (UL) is TDMized to distribute the resources between the broadcast service and one of the unicast blocks. The main benefit of this model is that channel resources are distributed twice, once for the UL and once for the Lower Layer (LL). However, the SNR associated with decoding each of the services will assume a penalty for being a NOMA-based transmission. On the other hand, Model B (see Fig. 2(b)) presents a TDMA scheme where in the first part NOMA is applied to multiplex the broadcast service and a block of unicast services with the same number of resources as the broadcast block. In this case, the unicast block (II) is transmitted in single-layer mode, so the required SNR per user will be lower than in Model A, but not all the resources are available. Finally, Model C (see Fig. 2(c)) is another TDMA model where NOMA has been applied to multiplex the two blocks of unicast content. This model presents the most robust broadcast service; since it is configured in single-layer mode, its associated SNR is the lowest. However, the use of the spectrum is lower than in the case of Model A.

Algorithm 1 shows a five-step generalization of the RRM schemes presented above. It should be mentioned that the RRM schemes are designed so that they can be executed in each TTI. The first step consists in obtaining feedback from each user. In this case, the feedback is the CQI (i.e., u_k) perceived by each k^{th} user connected to the gNB. By gathering each feedback, the Channel Quality Indicator (CQI) distribution vector $U = \{u_1, u_2, ..., u_K\}$ is generated, being K the total number of users in the network.

Then, the three steps related to configuring the three service blocks are executed for all the Injection Level (IL) values, where $l = \Delta_1, ..., \Delta_L$ represents the set of possible IL values. In order to set the range of possible ILs, the same range as in ATSC 3.0 has been used [30]. It is composed of 31 values, where steps of 1 dB are used for IL between -25 dB and -5 dB and steps of 0.5 dB between -5 dB and 0 dB. First, the broadcast service is configured in Step 2. To do so, the RRM scheme calculates, according to the minimum broadcast capacity constraint, the required number of resource blocks (BW_{BR_l}). It should be noted that the BW_{BR_l} calculation varies depending on if the broadcast block is configured in single-layer or the UL of a two-layer system. In particular, in the case of Model C, the BW_{BR_l} calculation is directly carried out with the SNR related with the CQI values. In contrast, in Models A and B, the SNR required to obtain each CQI has

Algorithm 1 Common steps of the proposed RRM schemes

1: **Define**: $l = \Delta_1, ..., \Delta_L$. Set of possible ILs;
2: **Define**: $tti = 1, ..., TTI$. Set of evaluated TTIs;
3: **for** $tti \leq TTI$ **do**
4: **Step 1 - CQI adquisition:**
$$U = \{u_1, u_2, ..., u_K\}$$
5: **for** $l \leq \Delta_L$ **do**
6: **Step 2 - Broadcast service configuration:**
$$BW_{BR_l}, eff_{BR_l}, th_{BR_l}$$
7: **Step 3 - Unicast service configuration in block (I):**
$$[BW_{UN_{1,l}}, eff_{UN_{1,l}}, th_{UN_{1,l}}] ...$$
$$[BW_{UN_{N,l}}, eff_{UN_{N,l}}, th_{UN_{N,l}}]$$
8: **Step 4 - Unicast service configuration in block (II):**
$$[BW_{UN_{N+1,l}}, eff_{UN_{N+1,l}}, th_{UN_{N+1,l}}] ...$$
$$[BW_{UN_{N+M,l}}, eff_{UN_{N+M,l}}, th_{UN_{N+M,l}}]$$
9: **end for**
10: **Step 5 - Get the configuration that maximizes unicast aggregated throughput:**
$$\max_l \sum_{i=1}^{N+M} th_{UN_{i,l}}$$
11: **end for**

to be recalculated since they assume a penalization due to the two-layer structure. The SNR adaptation for the UL has been carried out according to [DCAN], and [31]. Then, the throughput of the broadcast service (th_{BR_l}) is calculated as follows:

$$th_{BR_l} = BW_{BR_l} * eff_{BR_l} \tag{3}$$

eff_{BR_l} is the efficiency of the transmitted service (bps/Hz), which is obtained from the selected MCS value according to Table 5.1.3.1-2 in [23].

In Step 3, the first block of the unicast services is configured. In this work, it should be noted that it has been decided to start serving the unicast users through the unicast blocks located at the LL. Through simulations, it has been proven that this structure gives better results than starting serving unicast users through simulations of the blocks in the UL or single-layer. One of the main differences among the RRM models is the RB dedicated to each unicast service configuration block. Concretely, the three models have a different amount of resources ($BW_{UN(I)}$) in Step 3. In the case of Model A, as the unicast block (I) is extended during the whole bandwidth, the amount of resources is equal to the total amount of RB in the channel (BW_{TOT}). On the contrary, both Model B and Model C have a limited amount of resources. In fact, while in Model B $BW_{UN(I)}$ is equal to BW_{BR}, in Model C is calculated as follows:

$$BW_{UN(I)} = BW_{TOT} - BW_{BR} \tag{4}$$

Once the total number of RBs available for the unicast block (I) has been determined, the RRM scheme begins to serve users one by one. In this case, users are served based on the feedback sent so that users with better CQIs are managed first since they require fewer RBs than users with bad CQI, and access to a greater number of users can be guaranteed. In order to configure the service to each user, the requested CQI is taken as a reference to decide the number of RBs required for each user ($BW_{UN_{l,k}}$). Then, similar to Eq. (3) the throughput delivered to the k^{th} user using the l^{th} IL is calculated. As in the case of the broadcast service, an SNR adaptation has to be applied since the SNR associated with the requested CQI value has been calculated in single-layer mode, and the unicast services delivered in the first unicast block are organized in the LL. Also in this case, the LL SNR adaptation is based on the procedure defined in [DCAN], and [31]. Then, in the first unicast block, the CQI value for the LL is selected to guarantee that the SNR required for the LL is equal to or less than the required in single-layer mode. When the configuration for the k^{th} user is managed, the number of RBs available for unicast block (I) is updated by subtracting the resources allocated to the previous user:

$$BW_{UN(I)} = BW_{UN(I)} - BW_{UN_{l,k}} \tag{5}$$

Then, the algorithm checks if there are available RBs to continue with the resource allocation of the first block ($BW_{UN(I)} > 0$). If there are still available RBs, the following user (i.e., $k + 1$) in U is served. If not, the RRM algorithm goes to Step 4.

For its part, configuring the services of the unicast block (II) is similar to the process described in Step 3. In this case, the first user to be served is the user $N+1$, N being the last user to be served entirely in the unicast block (I). If, on the other hand, the user N has not been fully served in the previous block, before starting to serve $N + 1$, the missing throughput of N is completed. Afterward, the same process as shown in Step 3 is repeated, but taking into account in the SNR adaptation process that the unicast block (II) is not located in the LL. In fact, SNR adaptation is necessary only in Models A and C since the unicast service is transmitted on the UL. In Model B, no adaptation is required because the transmission of the unicast services is carried out in single-layer mode. Once all the resources dedicated to the unicast block (II) are finished ($BW_{UN(II)} \leq 0$), all the user configurations are saved, and the following IL is evaluated (l+1). At this point, M represents the number of users served in the second unicast block and $N + M$ the total number of unicast users served.

The resource allocation process is repeated until the calculations are completed for all the ILs described in l, and, in the last step (i.e., Step 5), the best performing configuration is selected. In this case, the throughput of the unicast services is determined as the optimization parameter. Therefore, the best performing configuration is determined by looking for the configuration that has the highest aggregate unicast throughput:

$$\max_l \sum_{i=1}^{N+M} th_{UN_{i,l}} \tag{6}$$

V. Evaluation Use Case

The objective of the selected use case to evaluate the RRM schemes shown in the previous section is to provide service

in a high occupancy environment. A museum room is taken as a reference where many people accumulate in a reduced environment. The service offered is a mixed broadcast/unicast mode. Firstly, the primary content will be transmitted through the broadcast service. Each user in the room can access a common online platform where the physical distribution of the room and the list of additional content that can be requested are detailed. Then, the goal of the unicast service is to customize/personalize the service that the user is receiving. For this, the unicast service may be video content with additional explanations of the works of art, AR content to interact with the exhibition, or interactive games about additional data about the content of the exhibition hall. To guarantee the correct functioning of the application, users must have access to the broadcast service ensured, and the application has to provide on-demand content to the largest possible number of users.

Concerning the services, the needs of broadcast and unicast services are different in terms of various parameters. On the one hand, the broadcast service must guarantee the download of the application to all users in the room so that it will be encoded with the lowest possible Modulation and Code Scheme (MCS), MCS0 (QPSK 120/1024) [23]. In addition, the minimum capacity that must be offered will range between one and ten Mbps, depending on the type of content in the download (data, images, video, etc.) and the type of end-user (mobile, tablet, laptop, etc.). In contrast, the unicast personalization service does not have as strict accessibility requirements as the broadcast service, so the service will be encoded based on the Channel Quality Indicator (CQI) received from the user. Regarding the minimum capacity to offer, a range from three to 30 Mbps will be studied in order to include in the study from services with low requirements (video content to small screen mobile devices) to more demanding services such as AR applications on tables with large high definition screen displays.

In order to carry out the simulations, OMNeT++ has been used [32], which is a packet simulator. In this simulator, a room of 30x30 meters and 3 m height has been defined, representing the room where the works of art and service receiver users are concentrated. A transmitter has been placed on the room ceiling (i.e., 3 m transmitter antenna height), and the receivers have been randomly distributed throughout the room. The receivers are pedestrians that move at an average speed of 3 km/h following the Random Waypoint Point (RWP) mobility model [33], which emulates the movements of pedestrians. The central frequency used to offer the service is 28 GHz, and a bandwidth of 400 MHz is used to provide broadcast and unicast services among all users. To represent propagation, the indications described in [19] for indoor environments have been followed. In this case, the path loss model used is described in Table III and taking into account the typical layout of the showrooms and that the transmitter is located on the ceiling, Line-of-Sight (LOS) condition is assumed. Regarding the propagation channel, following the indications in [19], the TDL-D and TDL-E (Tapped Delay Line) profiles are used to model the channels with a short delay spread of 16 ns and a shadowing effect with a mean zero and σ 3 dB. Finally, it should be noted that the designed RRM algorithms are executed with a frequency of one ms (each subframe). That is, during the resource allocation phase, all the RBs contained in a subframe are distributed.

TABLE III: Use case parameters

Parameter	Value
Center Frequency	28 GHz
Bandwidth	400 MHz
SCS	120 kHz
Tx Power	24 dBm
BS Antenna Height	3 m
RX Antenna Height	1.5 m
LOS/NLOS	LOS
Path Loss	InH model in [19]
User Type	Pedestrians
Number of Users	Up to 320
Speed	3 km/h
Mobility Type	RWP
Hall Size	30m x 30m x 3m
Desired Delay Spread	16 ns
Channel Model	TDL-D, TDL-E
Shadowing	σ = 3 dB
Noise Power	-90 dBm
Noise Figure	9 dB
RRM Frequency	1 ms
Broadcast Service Capacity	1-10 Mbps
Unicast Service Capacity	3-30 Mbps/user

The rest of the use case parameters are shown in Table III.

VI. Results

This section evaluates the performance of the RRM schemes shown in Section IV under the evaluation use case defined in V. First, the evaluation metrics are introduced and, then, the results are discussed.

A. Evaluation Metrics

In this work, the performance of the proposed RRM algorithms will be evaluated from the throughput and number of users served perspective. Next, the characteristics of each parameter are shown.

Regarding capacity, two parameters will be used. The first is the throughput per service, which is a double parameter that defines the datarate dedicated to the broadcast service and the unicast service. The throughput of the broadcast service (th_{BR}) is calculated in Eq. (3). The objective of the th_{BR} is to show if the RRM algorithm is capable of adjusting the offered capacity to the minimum throughput requirement imposed by the application. In contrast, unicast service throughput (th_{UN}) measures the volume of unicast traffic on the network and is calculated as the sum of all the individual unicast services offered:

$$th_{UN} = \sum_{i=1}^{N} th_{UN_i} = \sum_{i=1}^{N} BW_{UN_i} * eff_{UN_i} \qquad (7)$$

where, i is the unicast user index, N is the amount of served unicast users, and th_{UN_i} is the individual throughput value of i^{th} unicast user, which is calculated as in (3), but following the bandwidth allocated for each user (BW_{UN_i}) and the efficiency decided for each user (eff_{UN_i}). Then, the other parameter related to the capacity measurement is the ADR, representing

the sum of all the data received by each user. It is defined as follows:

$$ADR = th_{BR} * \sigma + th_{UN} \quad (8)$$

where, σ represents the number of users able to decode the broadcast service correctly.

To evaluate the algorithm performance in terms of the number of users served, two other metrics will be used. On the one hand, availability will be used as a robustness measure for the broadcast service. Availability measures the number of users who can receive the common service correctly with respect to the total number of users. To ensure that users can access the application fluently, availability must be above 95%. On the other hand, the number of users that can be served simultaneously with the personalized unicast service will also be measured. In this case, the unicast user number served in each configuration will be presented directly. It should be noted that the personalized service is a non-priority service, so the number of users that will be given access will be considerably less than in the case of the broadcast service.

B. NOMA-only and TDMA-only modes

This subsection aims to analyze the performance of RRM schemes based on a single technique (NOMA-only or TDMA-only) in the use case proposed above. For this, Table IV presents the results obtained in terms of throughput for both services (th_{BR} and th_{UN}), the number of unicast users served, and the ADR.

First, it should be noted that the results of NOMA-only do not adapt to variations in the capacity requirements of the broadcast service. This effect occurs because the entire bandwidth (i.e., 400 MHz) is used for the broadcast service, which is much higher than the requirements established by the use case. On the contrary, variations in the unicast service requirement cause a slight reduction in th_{UN} and a considerable decrease in the number of users served. On the other hand, the TDMA-only algorithm does allow the services offered to be adapted according to the application requirements. For this reason, the th_{BR} adapts to the minimum to offer and allows the th_{UN} and the number of unicast users served to be better than those offered by NOMA-only. Finally, it is necessary to emphasize the difference in the ADR results. NOMA-only offers around ten times more ADR than TDMA-only in all cases. This effect is caused because with NOMA-only, it is not possible to adapt the th_{BR} to the imposed requirements, and the capacity offered is well above what is requested, which causes extremely high ADR values. However, this data is not representative since the difference does not indicate an improvement compared to TDMA-only, but rather a misuse of available resources.

Undoubtedly, these results agree with the conclusions described in [8] and [DCAN], where it is shown how the schemes that use NOMA-only are not capable of adapting to the use case requirements. However, in [8] and [DCAN], the mmWave bands are not used, and, therefore, the bandwidths are smaller, which makes the differences between NOMA and TDMA not as evident as in this case. In fact, in sub-6 GHz applications, NOMA still outperforms TDMA, but using larger channels is more detrimental to NOMA-only schemes. Therefore, it seems that the RRM schemes that can approach the optimal distribution of resources are those that combine TDMA and NOMA. In this way, the ability to adapt to the application requirements that TDMA offers can be combined with the improvement in spectral efficiency that NOMA provides.

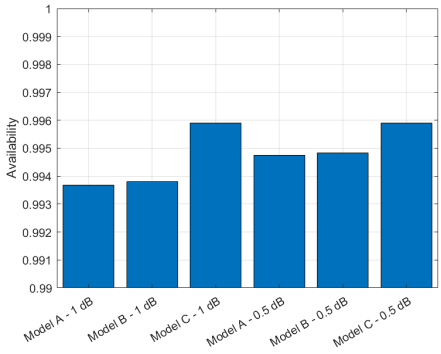


Fig. 3: Availability results of the proposed RRM methods.

C. Combined RRM algorithms

First, the availability values are presented since it is the primary condition to receive the rest of the services correctly. Availability results are shown in Fig. 3 for all models with two different UL SNR penalties (0.5 dB and 1 dB). It should be noted that the variation in the requirements of broadcast and unicast services hardly affects availability, so the values shown in Fig. 3 are average values of all the TTIs. First of all, it must be emphasized that all the values are above 99% so that they are all valid results. Among all the models, as expected, the one that offers the best availability is Model C. The main reason is that the broadcast service is implemented in SL mode, it has lower SNR than Models A and B, and it is not affected by the UL SNR penalty. However, the difference between the other two models is slight and assumable. Also, as expected, on Models A and B, availability is increased by reducing the UL SNR penalty. Based on the above-shown results, a UL SNR penalty of 1 dB will be used in the rest of the results since it has a similar availability rate to 0.5 dB.

The number of unicast users served is shown in Fig. 4 for three different broadcast service requirements: three Mbps (see Fig. 4(a)), six Mbps (see Fig. 4(b)) and nine Mbps (see Fig. 4(c)). As can be seen, Models A and C are once again the ones with the best performance, with very similar results. Also, Model B shows worse performance than the other two models. Regarding the service requirements, the broadcast requirement hardly interferes with the results. In contrast, the minimum unicast service requirement does have an impact on the number of users served, especially in Model B, which is the one that decreases the most rapidly. If the requirements of unicast services are related to real personalization services, services related to video applications would be below ten Mbps, while

TABLE IV: Throughput, served unicast users and ADR results of NOMA-only and TDMA-only RRM schemes for different broadcast and unicast service requirements

Minimum Broadcast Service	Minimum Unicast Service	NOMA-only				TDMA-only			
		th_{BR} (Mbps)	th_{UN} (Mbps)	Unicast Users	ADR (Mbps)	th_{BR} (Mbps)	th_{UN} (Mbps)	Unicast Users	ADR (Mbps)
3 Mbps	6 Mbps	92.82	1276.57	202.50	30822.35	3.04	1914.38	293.72	2882.49
	12 Mbps	92.82	1539.29	122.46	31085.07	3.04	2610.01	200.77	3578.12
	24 Mbps	92.82	1545.31	63.00	31091.09	3.04	2835.50	112.05	3803.61
6 Mbps	6 Mbps	92.82	1276.57	202.50	30822.35	6.03	1893.11	290.24	3815.88
	12 Mbps	92.82	1539.29	122.46	31085.07	6.03	2545.45	195.61	4468.22
	24 Mbps	92.82	1545.31	63.00	31091.09	6.03	2741.47	109.98	4664.24

those oriented to AR applications between 20 and 30 Mbps. In the case of videos with Models A and C, practically all users could receive them (minimum 300 users), while with Model B, the service could be guaranteed to about 250 users. On the contrary, for AR-based applications, Models A and C show that they can serve 50% more users since they guarantee access to around 150 users, while with Model B, it would not reach 100 users.

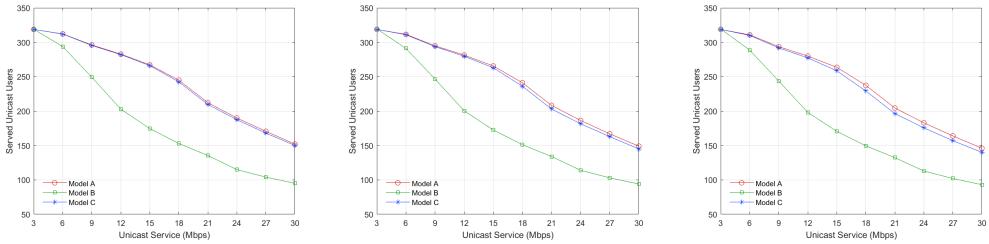
Concerning the throughput results, first, Table V shows the throughput results per service that have been obtained with each of the RRM models for different broadcast and unicast requirements. As can be seen, the th_{BR} obtained in all cases is adjusted to the minimum required by the application, unlike the NOMA cases shown in the previous subsection. Also, the th_{UN} shows a considerable increase when the minimum unicast throughput per user increases. On the contrary, when increasing the broadcast service requirement, the th_{UN} presents a slight reduction. This effect indicates that the RBs reserved to cover the broadcast service are very few compared to those that remain to serve the unicast users and, therefore, it hardly affects the results. When comparing the results based on the models, it is seen that Model B is less efficient since it is capable of offering a much lower th_{UN} than Models A and C. In the configuration with the most significant difference (minimum broadcast service 3 Mbps and minimum unicast service 24 Mbps), models A and C improve the performance of Model B by almost 60%. The best results are found in Model A since it offers values of th_{UN} slightly higher than Model C, up to 4% better in some cases.

Then, the ADR results for each RRM model are shown in Fig. 5, where the minimum broadcast services are 3 Mbps, 6 Mbps, and 9 Mbps in Fig. 5(a), Fig. 5(b), and Fig. 5(c), respectively. As in the throughput case, the ADR results also show that Model B has a significantly lower performance than Models A and C. However, the difference between the models is reduced, since the performance difference between Model A and Model B in Fig. 5 (a) is around 30%, while in the case of Fig. 5 (c) it does not exceed 22%. Likewise, the difference in performance between Models A and C also increases as the requirement for broadcast service grows, indicating that Model A always offers the best ADR result. The different trends shown by the curves should also be highlighted since they have, first, a phase of exponential increase and, then, a saturation phase. The exponential increase occurs because increasing the unicast service requirement does not imply a linear decrease in users number. In fact, as explained in Section IV, by serving the users with the best conditions first, the first users who are left out of the personalization service are the users who require more RBs, since they request low MCS values. Therefore, as the unicast service requirement increases, the efficiency of the services offered increases, and the ADR grows. On the contrary, in the saturation phase, this effect does not occur. In the case of saturation, the users with the worst conditions have already been rejected, and those who are served have good channel conditions, so the increase in the minimum unicast service practically implies a linear reduction in the number of users served. Finally, it is worth highlighting the ADR result obtained for Models A and C between three and six Mbps of unicast service. In those cases, despite increasing the unicast service, the ADR hardly changes. This effect is because in the case of three Mbps, all users are served once and the RRM model performs a second round in serving the majority of users during the same TTI. So as the requirement of the unicast service increases, the users that are left out are those that had been served a second time, that is, those with good channel condition. Consequently, the total spectral efficiency falls, and the ADR does not grow. This effect does not occur in Model B because it presents a less efficient use of resources. It does not serve as many users in the second round as Models A and C. Therefore, the effect is less noticeable and does not present such an abrupt change in the behavior. However, it can be seen that in Model B, the increase from three to six Mbps is less than the increase from six to nine Mbps.

Finally, when comparing the results shown in this subsection with those obtained with the NOMA-only and TDMA-only schemes (see Section VI-C), it can be seen that Models A and C improve all KPIs and that Model B is slightly better. Specifically, both Model A and Model C improve the results shown in Table IV by around 50%. Especially, the difference increases when the requirements of broadcast and unicast services increase. Therefore, as a summary, it should be noted that the RRM schemes that combine TDMA and NOMA considerably improve the performance of the simple models and that, in particular, Models A and C are the best possible options.

VII. Conclusions

This article considers resource allocation techniques that combine TDMA and NOMA to improve spectral efficiency in applications implemented in the mmWave bands. For this, three different RRM models have been designed and evaluated in terms of throughput (th_{BR} and th_{UN}), capacity (number of unicast users served), and availability. In addition, they

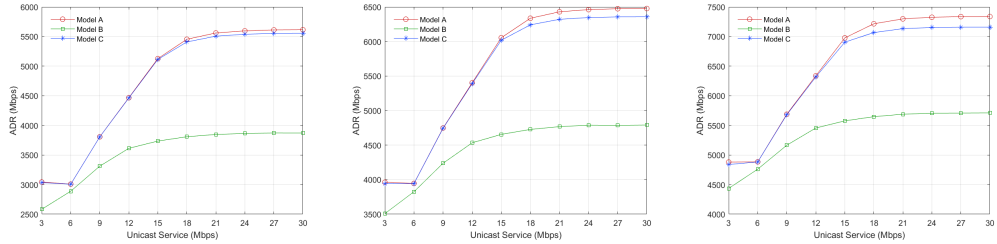


(a) (b) (c)

Fig. 4: Number of served unicast users for different minimum broadast capacity values: (a) 3 Mbps, (b) 6 Mbps, (c) 9 Mbps.

TABLE V: Obtained throughput per service results for different broadcast and unicast requirements

Minimum Broadcast Service	Minimum Unicast Service	Model A		Model B		Model C	
		th_{BR} (Mbps)	th_{UN} (Mbps)	th_{BR} (Mbps)	th_{UN} (Mbps)	th_{BR} (Mbps)	th_{UN} (Mbps)
3 Mbps	6 Mbps	3.04	2041.48	3.04	1920.58	3.04	2038.11
	12 Mbps	3.04	3500.97	3.04	2645.89	3.04	3493.16
	24 Mbps	3.04	4630.19	3.04	2896.99	3.04	4570.80
6 Mbps	6 Mbps	6.03	2025.62	6.03	1904.30	6.03	2020.03
	12 Mbps	6.03	3483.70	6.03	2614.94	6.03	3467.22
	24 Mbps	6.03	4543.06	6.03	2870.12	6.03	4423.98
9 Mbps	6 Mbps	9.03	2011.39	9.03	1887.28	9.03	2004.64
	12 Mbps	9.03	3465.86	9.03	2584.45	9.03	3439.82
	24 Mbps	9.03	4455.26	9.03	2835.19	9.03	4275.34



(a) (b) (c)

Fig. 5: ADR results for different minimum broadast capacity values: (a) 3 Mbps, (b) 6 Mbps, (c) 9 Mbps.

have been compared to schemes that only implement one of the multiplexing mechanisms (only-NOMA and only-TDMA), and the improvement is considerable. Moreover, the main challenges and possible applications of 5G in the mmWave bands have also been highlighted, and an innovative use case has been designed to test the designed RRM schemes.

Regarding the results, it has been verified that the RRM schemes that combine TDMA and NOMA considerably improve the schemes of a single multiplexing mechanism. Especially when the application requirements are high, the difference in performance exceeds 50%. Concerning the RRM models designed, it should be noted that even Model B has acceptable performance, Models A and C outperform Model B in all KPIs. Furthermore, the results indicate that for the proposed use case, using Models A and C, in the case of additional high-quality videos, access to the personalization service can be guaranteed to around 300 users (almost all the users in the cell). In contrast, in AR services, a minimum of 150 users is guaranteed with Models A and C. Finally, it should be highlighted that Model A outperforms Model C slightly since Model A downgrades less the results when the broadcast and unicast requirements are increased.

ACKNOWLEDGMENT

This work has been partially supported by the Basque Government (the grant IT1234-19 and the PREDOC grant program PRE_2020_2_0105) and by the Spanish Government (project PHANTOM under the grant RTI2018-099162-B-I00 (MCIU/AEI/FEDER, UE)).

REFERENCES

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017-2022.

[2] ITU-R Rec. ITU-R M. 2083-0, "IMT Vision — Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond," Sept. 2015.
[3] 3GPP TS 38.201 v15.0.0, Tech. Spec. Group Services and System Aspects, "NR; Physical layer; General description," Tech. Rep., Jan. 2018.
[4] A. V. Lopez, A. Chervyakov, G. Chance, S. Verma, and Y. Tang, "Opportunities and challenges of mmwave nr," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 4–6, 2019.
[5] L. Zhang, W. Li, Y. Wu, X. Wang, S.-I. Park, H. M. Kim, J.-Y. Lee, P. Angueira, and J. Montalban, "Layered-division-multiplexing: Theory and practice," *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 216–232, 2016.
[6] J. Montalban, P. Scopelliti, M. Fadda, E. Iradier, C. Desogus, P. Angueira, M. Murroni, and G. Araniti, "Multimedia multicast services in 5g networks: Subgrouping and non-orthogonal multiple access techniques," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 91–95, 2018.
[7] E. Iradier, J. Montalban, L. Fanari, P. Angueira, L. Zhang, Y. Wu, and W. Li, "Using noma for enabling broadcast/unicast convergence in 5g networks," *IEEE Transactions on Broadcasting*, vol. 66, no. 2, pp. 503–514, 2020.
[8] E. Iradier, A. Abuin, L. Fanari, J. Montalban, P. Angueira, L. Zhang, W. Li, and Y. Wu, "Broadcast/unicast convergence in noma-based 5g: a rrm optimization algorithm," in *2020 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2020, pp. 1–6.
[9] S. Scott-Hayward and E. Garcia-Palacios, "Multimedia resource allocation in mmwave 5g networks," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 240–247, 2015.
[10] A. Samuylov, D. Moltchanov, R. Kovalchukov, R. Pirmagomedov, Y. Gaidamaka, S. Andreev, Y. Koucheryavy, and K. Samouylov, "Characterizing resource allocation trade-offs in 5g nr serving multicast and unicast traffic," *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3421–3434, 2020.
[11] R. Liu, Q. Chen, G. Yu, and G. Y. Li, "Joint user association and resource allocation for multi-band millimeter-wave heterogeneous networks," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8502–8516, 2019.
[12] J. Deng, O. Tirkkonen, R. Freij-Hollanti, T. Chen, and N. Nikaein, "Resource allocation and interference management for opportunistic relaying in integrated mmwave/sub-6 ghz 5g networks," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 94–101, 2017.
[13] N. Nouri, J. Abouei, M. Jaseemuddin, and A. Anpalagan, "Joint access and resource allocation in ultradense mmwave noma networks with mobile edge computing," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1531–1547, 2019.
[14] Z. Shi, Y. Wang, L. Huang, and T. Wang, "Dynamic resource allocation in mmwave unified access and backhaul network," in *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE, 2015, pp. 2260–2264.
[15] S. Kumar, S. Suman, and S. De, "Dynamic resource allocation in uav-enabled mmwave communication networks," *IEEE Internet of Things Journal*, 2020.
[16] C. K. Agubor, I. Akwukwuegbu, M. Olubiwe, C. O. Nosiri, A. Ehinomen, A. A. Olukunle, S. O. Okozi, L. Ezema, and B. C. Okeke, "A comprehensive review on the feasibility and challenges of millimeter wave in emerging 5g mobile communication," *Adv. Sci. Technol. Eng. Syst*, vol. 4, no. 3, pp. 138–144, 2019.
[17] A. A. Zaidi, R. Baldemair, V. Molés-Cases, N. He, K. Werner, and A. Cedergren, "Ofdm numerology design for 5g new radio to support iot, embb, and mbsfn," *IEEE Communications Standards Magazine*, vol. 2, no. 2, pp. 78–83, 2018.
[18] Y. Li, E. Pateromichelakis, N. Vucic, J. Luo, W. Xu, and G. Caire, "Radio resource management considerations for 5g millimeter wave backhaul and access networks," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 86–92, 2017.
[19] 3GPP TR 38.901, Radio Access Network Working Group and others, "Study on channel model for frequencies from 0.5 to 100 ghz (release 15)," Tech. Rep., 2018.
[20] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5g) wireless networks—with a focus on propagation models," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6213–6230, 2017.
[21] S. Sun, T. S. Rappaport, M. Shafi, P. Tang, J. Zhang, and P. J. Smith, "Propagation models and performance evaluation for 5g millimeter-wave bands," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 9, pp. 8422–8439, 2018.
[22] ITU-R, "Specific attenuation model for rain for use in prediction methods, propagation in non-ionized media," Tech. Rep. P.838-3, 2005.
[23] 3GPP TS 38.214 v15.3.0, Tech. Spec. Group Services and System Aspects, "NR; Physical layer procedures for data (Release 15)," Tech. Rep., Sept. 2018.
[24] S. Persia and L. Rea, "Next generation m2m cellular networks: Lte-mtc and nb-iot capacity analysis for smart grids applications," in *2016 AEIT International Annual Conference (AEIT)*. IEEE, 2016, pp. 1–6.
[25] X. Xu, J. Liu, and X. Tao, "Mobile edge computing enhanced adaptive bitrate video delivery with joint cache and radio resource allocation," *IEEE Access*, vol. 5, pp. 16 406–16 415, 2017.
[26] M. S. Elbamby, C. Perfecto, M. Bennis, and K. Doppler, "Toward low-latency and ultra-reliable virtual reality," *IEEE Network*, vol. 32, no. 2, pp. 78–84, 2018.
[27] K. Sakaguchi, T. Haustein, S. Barbarossa, E. C. Strinati, A. Clemente, G. Destino, A. Pärssinen, I. Kim, H. Chung, J. Kim *et al.*, "Where, when, and how mmwave is used in 5g and beyond," *IEICE Transactions on Electronics*, vol. 100, no. 10, pp. 790–808, 2017.
[28] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5g wireless systems," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 136–143, 2014.
[29] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmwave) for 5g: opportunities and challenges," *Wireless networks*, vol. 21, no. 8, pp. 2657–2676, 2015.
[30] Advanced Television Systems Committee and others, "ATSC Standard: Physical Layer Protocol (A/322). Doc. A/322: 2106," *Washington, DC, USA*, 2016.
[31] J. Montalbán, L. Zhang, U. Gil, Y. Wu, I. Angulo, K. Salehian, S.-I. Park, B. Rong, W. Li, H. M. Kim *et al.*, "Cloud transmission: System performance and application scenarios," *IEEE Transactions on Broadcasting*, vol. 60, no. 2, pp. 170–184, 2014.
[32] A. Varga and R. Hornig, "An overview of the OMNeT++ simulation environment," in *Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, 2008, p. 60.
[33] E. Hyytiä, H. Koskinen, P. E. Lassila, A. Penttinen, J. T. Virtamo, and J. Roszik, "Random waypoint model in wireless networks," 2005.

3.2 Publications associated to Contribution 2

3.2.1 Conference paper C3

This subsection presents a conference paper related with Contribution 2. The full reference of the paper is presented below:

- E. Iradier, J. Montalban, L. Fanari, P. Angueira, O. Seijo and I. Val, "NOMA-based 802.11n for Broadcasting Multimedia Content in Factory Automation Environments," 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Jeju, Korea (South), 2019, pp. 1-6, doi: 10.1109/BMSB47279.2019.8971844.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Indexed Congress in IEEExplore
- Area: Computer Science and Engineering
- SJR factor: 0.300

NOMA-based 802.11n for Broadcasting Multimedia Content in Factory Automation Environments

E. Iradier, J. Montalban, L. Fanari, P. Angueira
Department of Communications Engineering
University of the Basque Country (UPV/EHU)
Plaza Torres Quevedo 1, Bilbao, Spain
{eneko.iradier, jon.montalban, pablo.angueira}@ehu.eus

O. Seijo, I. Val
Communication Systems Group
IK4-Ikerlan Technology Research Centre
Pº J.M. Arizmendiarrieta 2, Mondragon, Spain
{oseijo, ival}@ikerlan.es

***Abstract*—Industry 4.0 requirements in reliability and latency are still a challenge for wireless communications systems due to the factory critical propagation scenarios. Additionally, further improvements are required in order to enhance the performance of wireless technologies in terms of reliability and latency before presenting them as a valid alternative to the wired technologies. These challenges are added to the increasing needs of bitrate to deliver multimedia content in Factory Automation (FA) environments. This paper proposes several use cases, which include video transmission and a specific technical solution based on a combination of NOMA and the IEEE 802.11n standard. The PHY/MAC proposal is evaluated on a reference scenario based on the deployment of the typical industrial applications. Results show that NOMA-based systems offer a very promising reliability gain in comparison to classical TDMA-based systems. In fact, for the most unbalanced service combinations a PER (Packet Error Rate) improvement of more than one order of magnitude can be obtained.**

Keywords— 802.11, Broadcast, Factory Automation, Multicast, Multimedia Content, NOMA, Reliability.

I. INTRODUCTION

Multimedia content distribution and, specifically, video consumption will be one of the most relevant factors for the traffic increase in wireless networks in the coming years. According to the Cisco Visual Networking Index (VNI) Global Mobile Data Traffic Forecast Update [1], over three-fourths (78 percent) of the world's mobile data traffic will be video by 2021.

One of the novelties is that the environments where video is gaining momentum are not restricted to traditional video consumption and social networks. A new application area, which is rapidly developing and growing, is the industrial communications. Advanced communications in industry are one of the pillars of a process usually referred as the fourth industrial revolution (or Industry 4.0). This revolution consists of bringing all the benefits of the information and communications technologies (ICT) into the industrial sector in all its methods, tools and infrastructures. In consequence, this trend is expected to redesign the technical characteristics and even the philosophy of the industrial communications [2].

Communications in industry have been traditionally oriented to design control and alarm systems. In consequence, these systems have very strict temporal and reliability requirements. What is more, usually they have been designed to transmit and receive very short and accurate amount of data. Nevertheless, the range of communication use cases in industry is widening and the delivery of video content may be needed sooner than expected [3]. In [4], how graphic and media technologies support operators in Industry 4.0 is presented.

Even if in principle they were not designed for deterministic and reliable communications, the 802.11 family of wireless standards is one of the most spread communication technologies in industry. There is a variety of reasons behind this success: the amount of available devices, the worldwide network deployments and the huge amount of applications that already have it implemented. Another key factor that make the difference between Wi-Fi and the rest of wireless connections (Bluetooth, NFC, LoRa…) is the use of Low Density Parity Check (LDPC) coding in the latest releases of the standard (802.11n and above), a more robust channel coding scheme in comparison with traditional convolutional codes. This improvement has turned 802.11n into a promising candidate for delivering critical information in factory automation environments.

In [5], authors present SHARP (Synchronous and Hybrid Architecture for Real-time Performance): a new communication system based on 802.11g for industrial automation. This design includes a modified 802.11g PHY layer and a deterministic Medium Access Control (MAC) based on Time Division Multiple Access (TDMA). This solution targets very challenging communication scenarios with strict time and reliability requirements. Therefore, determinism is almost guaranteed for critical services (CS), whereas the arrival of any other secondary communication service is not guaranteed.

The main contribution of the present paper is a solution that addresses the reliability requirements of the industry and the efficient transmission of multimedia contents. In particular, the use the Non-Orthogonal Multiple Access (NOMA) technique over the physical layer of the 802.11n standard is proposed. NOMA is a well-known technology that has been analyzed in combination with other technologies in different fields. Some of the most relevant examples of the success of NOMA are [6] and [7]. In [6], NOMA is presented as a candidate to be used in broadband mobile networks in

978-1-7281-2150-5/19/$31.00 ©2019 IEEE

TABLE I. SUMMARY OF INDUSTRIAL USE CASES REQUIRING VIDEO TRANSMISSION

Use Case	ID	Critical Service: Reliability – Latency	Best Effort Service: Bitrate - Capacity
Remote Vehicle Guidance/Operation	1.A	Control/Telemetry data	Video Layer
	1.B	Core Video Layer (Scalable Video Coding)	Enhancement Video Layer (Scalable Video Coding)
Surveillance & Production Control Video Systems	2	Core Video Layer (Reliability, critical)	Enhancement Video Layer
Staff Safety in Industry: Augmented Reality and complementary Aid Systems	3.A	Workers Critical Telemetry Uplink: Bio indicators, position, sensors in hazardous production	Video Service (Uplink video from smart glasses, helmet camera, etc)
	3.B	Critical Warning / Safety Message Data	Augmented Reality Data
General Purpose Automation and Multimedia delivery	4	Automation Control Service	Multimedia Service (safety information to screens, panels, etc)
Advanced Imagery Systems	5.A	Robust Video: Target Tracking, Low Latency Video (low fps, I only)	Enhancement Video Layer: High Resolution thermographic video
	5.B	Control Data	Enhancement Video Layer: High Resolution thermographic video
Condition Monitoring Applications	6	Critical data depending on the application: Stress, temperature, turns per second, speed, etc.	Video monitoring of the application

combination with subgrouping techniques in order to improve the overall network capacity. In [7], a low complexity version of NOMA, Layered Division Multiplexing (LDM), is proposed for the next generation terrestrial broadcasting systems, especially focused on ATSC 3.0.

The main benefit of NOMA is that, under certain conditions, provides higher spectrum efficiency with respect to the classical orthogonal multiplexing techniques. Moreover, NOMA offers an interesting increase in flexibility, which means a wider range of possible transmission configurations. As a matter of fact, in the literature it has been demonstrated that using LDM it is possible to increase the reliability, the transmission capacity or even both simultaneously [8].

To our best knowledge, this paper offers a comprehensive analysis of a video broadcasting/multicasting system within an industrial environment and based on an 802.11n-NOMA physical layer.

The paper is organized as follows. Section II describes the characteristics and requirements of industrial communications and defines different use cases for broadcasting video. Section III provides a realistic application scenario and a particular PHY/MAC solution architecture. Then, Section IV describes the results of the system network performance, and finally, the conclusions are presented in Section V.

II. Environment overview

In this section, the possible future ecosystem for wireless industrial communications is presented. Firstly, the tight requirements that must be fulfilled are shown, and secondly, the most relevant industrial use cases including video content are highlighted.

A. Factory Automation requirements

Factory Automation (FA) scenarios are considered one of the most challenging scenarios where industrial communications have to be deployed. The FA concept is related to the complete production chains where the most important elements are mainly machines. These environments include Device-to-Device (D2D) communications and complete control systems, which require real-time control. Some FA examples are: assembly, packaging, palletizing and manufacturing industries.

These processes communications must be extremely robust, and thus, reliability becomes a critical factor. In addition, the latency value can be as critical as the reliability. In a recent work [9], the maximum affordable latency is considered between 0.25-10 ms, with an update time of 0.5-50 ms. What is more, the Packet Loss Rate (PLR) is also a tight requirement with a goal that can be as low as 10^{-9} (if retransmissions or some other reliability-increasing mechanisms are enabled). Finally, it must be noted that the typical coverage are for these communication systems is up to 50-100 meters.

B. Posible use cases

Apart from the strict requirements that control communications require, there is also a demand for delay and loss tolerant services. These flexible services are referred as Best Effort (BE) services and among other things, they can be related to the media delivery.

Taking into account the quick change that the industry is facing, it is time to consider multimedia content (especially, audiovisual contents) as a real candidate for the BE services. What is more, in some cases they might be considered for the critical services too. This service distinction leads to an interesting unbalance between the capacities of both services. As it has been proved in theory [8], the greater the unbalance, the higher is the achievable gain with non-orthogonal multiplexing techniques.

TABLE II. SERVICES CHARACTERISTICS

ID	Name	Description	Transmitter	Receiver	Transmission method	Critical/BE
S1	Network synchronization	Superframe starting signal, responsible for the synchronization of the nodes and emergency alarm.	AP	Every node	Broadcast	Critical
S2	Sensing instructions	Concrete parameters for configuring the sensing. If the packet is not received, sensing will be done with previous parameters.	AP	One slave per superframe	Unicast	BE
S3	UHD Video	Video content about the overall system performance.	AP	Both displays	Multicast	BE
S4	Sensed parameters	Basic results of the sensing process.	Slave receiving S2	AP	Unicast	Critical
S5	Video data	Video data of the sensing process. Complementary information.	Slave receiving S2	AP	Unicast	BE

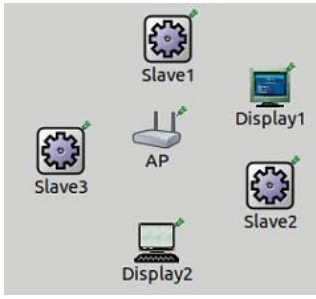


Fig 1. Simulated network in OMNeT++.

Taking this idea as a reference, a summary of different possible industrial use cases (1-6) is shown in TABLE I, where video broadcasting is considered as an enhanced service or even as a critical service. Moreover, for each use case, up to two different examples of critical and enhanced services have been proposed. Each of this use cases involves two services with different requirements: critical service (CS) and BE. Firstly, the critical service is loss and delay intolerant. On the one hand, CS are related with workers safety or with vital data communication for the application. On the other hand, BE services are less strict and so, they are considered loss and delay tolerant. Normally, they are related with supplementary application data, which does not interfere in the overall performance.

Regarding TABLE I, it is important to highlight that depending on the use case, video content can also be delivered as a critical service. For example, in use case 5 (Advanced Imagery Systems), which is a very promising field for surgery applications, a demand for critical video data can be assumed. In fact, a critical basic video is delivered in order to guarantee a pseudo-perfect communication, whereas a best effort video service is also offered for enhancing the video quality or adding some other features (thermography, humidity…).

III. Performance Evaluation Scenario

In order to study the performance of a NOMA-based 802.11n system for broadcasting multimedia content, a complete evaluation tool has been developed. In particular, in this paper, the work is based on the use case 6 from TABLE I. The evaluation performance is simulated in OMNeT++ [10].

TABLE III. AVAILABLE MCS VALUES IN THE 802.11N STANDARD

MCS	Modulation	Coderate	Capacity (Mbps)
0	BPSK	1/2	6
1	QPSK	1/2	12
2	QPSK	3/4	18
3	16-QAM	1/2	24
4	16-QAM	3/4	36
5	64-QAM	2/3	48
6	64-QAM	3/4	54
7	64-QAM	5/6	60

A. General settings of the scenario

In this section, a generic monitoring application for an industrial indoor environment is presented. For that aim, a six-device network topology is included, where each node has a different role. The network is composed of an Access Point (AP), three sensing slaves and two displays. In Fig. 1, the simulated network over OMNeT++ is shown.

The AP node is the master node of the network. One of its main roles is the network synchronization. The AP also requests sensing information to each node in a superframe time spam and distributes the video content that has to be delivered. The sensing slaves have also transceiver capacities. They are able to receive information from the AP, and send their sensed information (data and complementary video).

In TABLE II, a summary of the different offered network services is shown, including a service description as well as the target receivers. In this proposal, the unicast, multicast and broadcast services will share the same superframe structure. In the last column of TABLE II, each service is classified in terms of the loss tolerance level, and so, two categories are differentiated again: critical and BE. As critical services are related to loss intolerant services, they are transmitted with low Modulation and Code Scheme (MCS) value in order to obtain a higher reliability (see TABLE III). A network restriction has been imposed in order to guarantee a minimum quality: the overall critical service traffic must offer a minimum value of 6 Mbps. However, due to the philosophy of the BE services, this traffic is transmitted with higher MCS values to achieve high transmission rates.

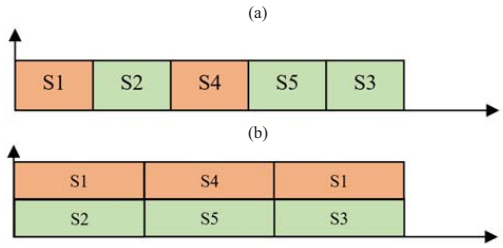


Fig 2. Time schedule representation: (a) TDMA-based, (b) NOMA-based

From the industrial communications requirements point of view, this communication system has to guarantee low latency and low update time values, especially for critical services. Hence, the length of the superframe has a constant value of 3 ms, and the slot size is modified depending on the number of slots, the multiplexing technology and the implemented MCS.

B. PHY/MAC level definition

The management of the medium access is a key factor for the deterministic communications, where reliability is considered a critical factor. Hence, as it is shown in Fig. 2, different time schedules have to be defined depending on the applied multiplexing technique.

Firstly, for the TDMA systems (Fig. 2.a), traffic categories are divided in different slots. A portion of the time is reserved for critical service (orange time slots), and the rest of the time for BE services (green time slots). For example, if each slot has the same time-length, the time division will be: 2/5 for critical services and 3/5 for BE services. That means that to achieve the minimum required value for the critical traffic, the lowest available MCS is number 2, QPSK 3/4, which offers 18 Mbps. The choice for the suitable MCS value for BE is a tradeoff between capacity and maximum allowable error rate.

In the case of the NOMA-based systems, traffic is multiplexed not only in the time domain, but also in the power domain (Fig. 2.b). Due to that, both traffic types can be offered the 100% of the time. Consequently, in this case, the lowest MCS value (BPSK 1/2) can be used without breaking the minimum required capacity. Moreover, for BE services any of the available MCS values can be used.

In order to implement this specific multiplexing scheme, the optimal injection level has to be defined. This parameter divides the overall transmitted power between the Core Layer (CL), which is reserved for Real-Time traffic, and Enhanced Layer (EL), which is used to deliver BE services.

IV. Results and Discussion

In this section, the proposed technical solution is evaluated. First, the simulated cases and propagation scenarios are presented, and then, the obtained results are plotted and discussed.

A. Simulated scenarios

Several configurations have been designed including different parameters and multiplexing techniques. However, capacity constraints are common for both multiplexing techniques. In this work, 6 Mbps are assumed for critical services, which is enough capacity to offer critical service, and two different approaches have been proposed for BE services (Case A and Case B), 24 Mbps and 48 Mbps, respectively. Both of them can accommodate high quality multimedia content, even in UHD format.

TABLE IV. CONFIGURATION PARAMETERS

Scenario parameter	
Service area	30x30, 50x50, 100x100
Fading model	NLOS Industrial Model
Path Loss coefficient	2.5
Transmitted power	10, 20 dBm
Rx noise threshold	90 dBm
Cycle time	3 ms
PHY/MAC parameter	
Critical service throughput	6 Mbps
BE throughput	24, 48 Mbps
TDMA configuration	Case A: 50% - 50% (MCS1-MCS5) Case B: 20% - 80% (MCS4-MCS7)
NOMA configuration	Case A: -2.5 dB (MCS0-MCS3) Case B: -1.5 dB (MCS0-MCS5)

First, in order to configure TDMA services, the time resources are shared between both services, and then, a suitable MCS value is chosen to achieve the expected minimum capacity value. Consequently, as it is shown in TABLE IV, in Case A, a 50% - 50% time division is set and the selected MCS values are MCS1 and MCS5, respectively. Additionally, in the second approach, in Case B, a 20% - 80% time division is fixed and the corresponding MCS values are MCS4 and MCS7, respectively.

Second, in the NOMA case the injection level value has been defined for each capacity combination (see TABLE IV) in order to multiplex the services in the power domain. For Case A, a -2.5 dB injection level has been applied, and for Case B, -1.5 dB. In addition, in the NOMA-based case, lower MCS values are selected because they are transmitted over the whole RF channel the 100% of the time.

Finally, the most relevant scenario parameters have been defined. Concerning the communication channel, an industrial channel model with Non-Line-Of-Sight (NLOS) condition and 29 ns rms delay spread [11] has been implemented. The path loss coefficient is set to 2.5 [12]. Moreover, three different service areas have been evaluated (30x30, 50x50 and 100x100) in order to broaden the study scenarios. In addition, transmitted packets airtime is set to 100µs. Therefore, each packet size varies depending on the MCS value. Eventually, two different transmission power has been tested 10 dBm and 20 dBm (this is the maximum allowed value for the ISM frequency bands in Europe) [13].

B. Results

Results are gathered based on the configurations described in the previous section (Fig 3 and Fig 4). What is more, two subplots are shown in each figure, one for each use case (Case A and Case B). Eventually, the following services are studied for each scenario: NOMA critical, NOMA BE, TDMA critical and TDMA BE. Regarding the plots, the vertical axis describes the PER value and the horizontal axis a certain scenario.

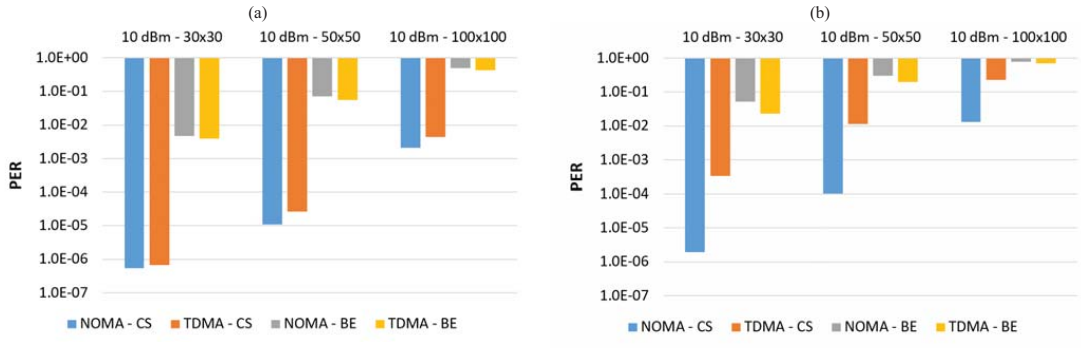


Fig 3. PER comparison for constant transmitted power: (a) Case A, (b) Case B

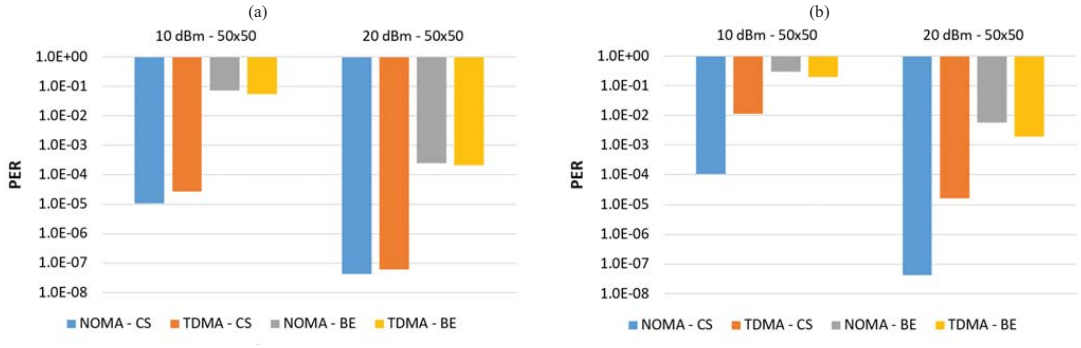


Fig 4. PER comparison for constant service area: (a) Case A, (b) Case B

In Fig 3, it is proven that as expected, the service area size is directly proportional to the obtained PER. The bigger is the factory size area, the higher are the path losses. Secondly, concerning Fig 4, it can be concluded that the transmitted power has just the opposite impact. Therefore, the higher the transmitted power, the lower the packet error values. Moreover, an in-depth analysis can be done involving the service area and the transmitted power. On the one hand, in the case of the smallest service area (30x30), implementable reliability values are obtained by transmitting 10 dBm, especially with NOMA. Consequently, for this case, it is not necessary to increase the transmitted power. On the other hand, the 50x50 case is more dependent on the applied MCS. Therefore, transmitting 10 dBm should be reserved only for very robust MCS values and the rest should be transmitted with a higher transmitted power. Finally, in the biggest service area case, the conclusion is obvious, 10 dBm is not enough power in order to guarantee a reliable service. Hence, using 10 dBm is rejected and higher transmitted power have to be implemented, preferably by using NOMA as the multiplexing technique.

On the other hand, analyzing the different services performance for each scenario it can be seen that although a little reliability loss is obtained for BE services in the NOMA case, a considerably big gain is obtained for critical services. This trend is maintained for both propagations scenarios. In addition, it must be pointed out that the reliability is the key factor mostly for critical services, whereas in the case of BE services, reliability is not a priority, and consequently, a little degradation can be tolerated.

It is also remarkable that the reliability gain is very dependent of the implemented MCS case. In fact, the performance differences between NOMA (blue and orange bars) and TDMA (grey and yellow bars) for Case B (Fig 3.b and Fig 4.b) are higher than the ones obtained in Fig 3.a and Fig 4.a for Case A. The main reason lies in the NOMA information theory basis. As it is shown in [8], the bigger the unbalance between CL and EL capacities, the greater is the obtained gain in comparison with orthogonal techniques.

The maximum gain related to critical services in Case A is obtained in the 10 dBm and 50x50 scenario, where the PER value is reduced from $2.67 \cdot 10^{-5}$ to $1.08 \cdot 10^{-5}$. However, this gain is considerably improved in each of the evaluated configurations in Case B. Furthermore, using NOMA in Case B improves around two magnitude orders in almost each scenario. For instance, the minimum gain is obtained in 10 dBm and 100x100 scenario, where more than a magnitude order gain is obtained (from 0.226 to 0.013), whereas the maximum is obtained for the 20 dBm and 50x50 configuration, where NOMA-based reliability offers a gain close to three magnitude orders (from $1.63 \cdot 10^{-5}$ to $6.03 \cdot 10^{-8}$).

Finally, it is also important to highlight the performance of both multiplexing techniques in the most challenging scenario: Case B, 10 dBm and 100x100 (see Fig 3.b). By using TDMA, a PER value of 0.2 is obtained, whereas NOMA reliability is set close to 0.01. Although both values are quite high, the difference between them is critical. Taking into account that retransmissions are not included in the MAC layer, PER values can be easily improved in the medium access layer. Therefore, the reliability obtained in the worse performing scenario using NOMA could be used in some soft-real-time applications. However, the reliability value obtained with TDMA is not enough for this scenario and, hence, it has no possible application.

V. Conclusion

This paper presents a novel technical solution for including efficiently multimedia content delivery in industrial communications. In order to carry out this work, traditional TDMA multiplexing schemes and the more recent NOMA multiplexing schemes have been combined with the 802.11n standard. To test the performance a realistic scenario has been described and a specific MAC layer has been designed for each technology. The evaluation phase has been completed with different configurations of the common scenario, trying to prove the viability of the proposal in different challenging situations.

The results show that NOMA can be considered as a real candidate to distribute multimedia content within industrial environments due to its great simulation results. In fact, NOMA-based communications have overcome TDMA-based results in each of the scenarios for the critical services, which are the most reliability dependent services. Despite the little losses that are obtained for the BE services, the PER gain for the critical services would recommend its implementation. Actually, a reduction of two order of magnitude is achieved in several cases by implementing NOMA.

Acknowledgments

This work has been partially supported by the Basque Government under the PREDOC grant program (PRE_2018_1_0344) and by the Spanish Government under the grant RTI2018-099162-B-I00 (MCIU/AEI/FEDER, UE).

References

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021.

[2] R. Drath and A. Horch, "Industrie 4.0: Hit or Hype? [Industry Forum]," in *IEEE Industrial Electronics Magazine*, vol. 8, no. 2, pp. 56-58, June 2014.

[3] L. L. Bello, "Novel trends in automotive networks: A perspective on Ethernet and the IEEE Audio Video Bridging," *Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA)*, Barcelona, 2014, pp. 1-8.

[4] J. Posada et al., "Graphics and Media Technologies for Operators in Industry 4.0," in *IEEE Computer Graphics and Applications*, vol. 38, no. 5, pp. 119-132, Sep./Oct. 2018.

[5] Ó. Seijo, Z. Fernández, I. Val and J. A. López-Fernández, "SHARP: A novel hybrid architecture for industrial wireless sensor and actuator networks," *2018 14th IEEE International Workshop on Factory Communication Systems (WFCS)*, Imperia, 2018, pp. 1-10.

[6] J. Montalban et al., "Multimedia Multicast Services in 5G Networks: Subgrouping and Non-Orthogonal Multiple Access Techniques," in *IEEE Communications Magazine*, vol. 56, no. 3, pp. 91-95, March 2018.

[7] S. I. Park et al., "Low complexity layered division multiplexing system for the next generation terrestrial broadcasting," *2015 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, Ghent, 2015, pp. 1-3.

[8] L. Zhang et al., "Layered-Division-Multiplexing: Theory and Practice," in *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 216-232, March 2016.

[9] P. Schulz et al., "Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture," in *IEEE Communications Magazine*, vol. 55, no. 2, pp. 70-78, February 2017.

[10] A. Varga and R. Hornig, "An overview of the OMNeT++ simulation environment," in *Proc. 1st Simutools*, Brussels, Belgium, 2008.

[11] F. Molisch et al., "IEEE 802.15. 4a channel model-final report," *IEEE P802*, vol. 15, no. 04, pp. 1–41, 2004.

[12] J. Miranda et al., "Path loss exponent analysis in Wireless Sensor Networks: Experimental evaluation," *2013 11th IEEE International Conference on Industrial Informatics (INDIN)*, Bochum, 2013, pp. 54-58.

[13] M. Luvisotto, Z. Pang and D. Dzung, "Ultra High Performance Wireless Control for Critical Applications: Challenges and Directions," in *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1448-1459, June 2017.

3.2.2 Journal paper J4

This subsection presents a journal paper related with Contribution 2. The full reference of the paper is presented below:

- J. Montalban, E. Iradier, P. Angueira, O. Seijo and I. Val, "NOMA-Based 802.11n for Industrial Automation," in IEEE Access, vol. 8, pp. 168546-168557, 2020, doi: 10.1109/ACCESS.2020.3023275.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Journal paper indexed in JCR and IEEExplore
- Area: Electrical & Electronic Engineering
- Ranking: 61/266 (Q1) based on JCR 2019
- Impact factor (JCR): 3.745

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 26, 2020, accepted September 8, 2020, date of publication September 10, 2020,
date of current version September 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3023275

NOMA-Based 802.11n for Industrial Automation

JON MONTALBAN[1], (Member, IEEE), ENEKO IRADIER[1], (Graduate Student Member, IEEE), PABLO ANGUEIRA[1], (Senior Member, IEEE), OSCAR SEIJO[2], (Member, IEEE), AND IÑAKI VAL[2], (Senior Member, IEEE)

[1]Department of Communications Engineering, University of the Basque Country (UPV/EHU), 48013 Bilbao, Spain
[2]IKERLAN Technology Research Centre, Basque Research and Technology Alliance (BRTA), 20500 Mondragón, Spain

Corresponding author: Eneko Iradier (eneko.iradier@ehu.eus)

This work was supported in part by the Basque Government under Grant IT1234-19, in part by the PREDOC Grant Program under Grant PRE_2019_2_0037 and project U4INDUSTRY (ELKARTEK), and in part by the Spanish Government (Project PHANTOM) (MCIU/AEI/FEDER, UE) under Grant RTI2018-099162-B-I00.

ABSTRACT Industry 4.0 and Industrial Internet refer to the expected revolution in production, utility management and, in general, fully automated, interconnected and digitally managed industrial ecosystems. One of the key enablers for Industry 4.0 lies on reliable and timely exchange of information and large scale deployment of wireless communications in industry facilities. Wireless will bring solutions to overcome the main drawbacks of the current wired systems: lack of mobility, deployment costs, cable damage dependency and scalability. However, the strict requirements in reliability and latency of use cases such as Factory Automation (FA) and Process Automation (PA) are still a major challenge and a barrier for massive deployment of currently available wireless standards. This paper proposes a PHY/MAC wireless communication solution for FA and PA based on Non-Orthogonal Multiple Access (NOMA) in combination with the 802.11n standard. The communication system proposed aims at delivering two different sets of services. The first service class is composed of Critical Services (CS) with strict restrictions in reliability and latency. The same communication system should convey also a second group of services, referred as Best Effort (BE) with more relaxed boundary conditions. The proposal theoretical background, a detailed transmission-reception architecture, the physical layer performance and the MAC level system reliability are presented in this paper. The solution provides significantly better reliability and higher flexibility than TDMA systems, jointly with a predictable control-cycle latency.

INDEX TERMS 802.11, Factory Automation, IWSN, local area networks, NOMA, P-NOMA Process Automation.

I. INTRODUCTION

The fourth industrial revolution is driving the research and the development in the field of industrial communications. One of the expected trends is the progressive deployment of wireless interconnections as well as global automation for improving efficiency in the use of resources and the integration between processes and factories [1]–[4]. Even though, wired connections are still ruling the majority of the industrial environments. The main reason is the uncertainty of current wireless standards for guaranteeing the tight requirements posed by industrial use cases.

As a paradigmatic example, the 802.11 family of wireless standards was not designed for industrial communications. Nevertheless, they have emerged also as a candidate for industrial applications because of their wide range of available devices, network deployments and applications in practically all communication fields. A major advantage of 802.11 relies on its similarity in OSI layer architecture and compatible protocol structure with Ethernet. This fact provides a high level of interoperability and ensures simple implementation of Ethernet/WLAN internetworking functions. On the weak side, its medium access mechanism, CSMA/CA, does not guarantee deterministic behavior. This fact is a major drawback for real-time services in networks where a massive number of nodes are potentially operating in a crowded spectrum environment [5]. In addition, the standard lacks efficiency for traffic profiles associated to industrial wireless communications (short packets) [6].

A typical Industrial Wireless Sensor Network (IWSN) is depicted in Fig. 1. A generic control system with specific performance and control cycle requirements is assumed.

The associate editor coordinating the review of this manuscript and approving it for publication was Giovanni Pau.

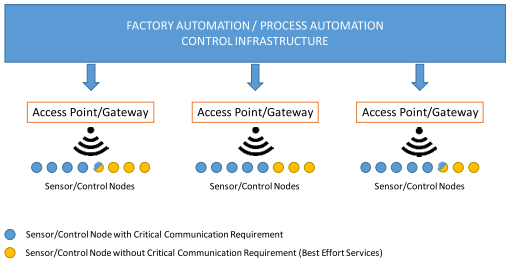


FIGURE 1. Typical industrial wireless sensor network in an industrial environment.

The control system has several access points/gateways (AP) that will serve a set of sensor/control nodes. The distinction of sensor and control functions will be associated to uplink and downlink capabilities of each node. There will be two types of nodes depending on the services associated. Some nodes will be linked to critical services (CS), with strict performance and latency restrictions. Others will not present tight restrictions and will be served on a best effort (BE) basis. Moreover, nodes receiving CS traffic could also be BE receivers. Additionally and in line with performance/latency restrictions, CS and BE service throughput requirements will also be different. In consequence, it is expected that the wireless connection will carry services with orthogonal requirements.

This unbalanced traffic scenario is the perfect operating point for Non-Orthogonal Multiple Access (NOMA) techniques and especially for Layered Division Multiplexing (LDM) [7], a specific power domain NOMA (P-NOMA). LDM has proven better reliability and throughput than the traditional TDM/FDM resource allocation schemes [8]. There are already preliminary proposals in literature for applying NOMA to industrial communications. In [9], authors calculate the robustness and capacity of an ideal NOMA system based on Shannon capacity. However, this work does not provide a complete architecture description in combination with existing standards (PHY and MAC layers) and does not consider specific modulation and coding schemes (MCS), neither provides the combined performance of a PHY and MAC layer. Then in [10], a cross-layer is design to be used between the physical and the data link layers to enhance the energy efficiency in downlink NOMA wireless networks for Visible Light Communication (VLC). Nevertheless, the analysis is carried out adapting the coding and the modulation without considering the integration in any standard and the reliability performance is not studied. In [11], a NOMA-based 802.11n PHY approach is presented and tested in some scenarios. However, no MAC layer is used to provide a realistic industrial communication and the results presented do not reach the low Packet Error Rate (PER) values required the industry.

Aiming at flexibility, capacity and reliability in wireless industrial applications, the present paper proposes the joint use of NOMA/LDM and 802.11n. Although this paper proposes a viability analysis using the well-stablished 802.11n standard, this proposal will be applicable for future versions, such as 802.11ax and 802.11be. The technical contributions of this paper include:

1. To our best knowledge, this paper is the first comprehensive system design and analysis of a NOMA-based 802.11n architecture for industrial communications. A fully detailed explanation of the transmitter and receiver chains is provided, and in addition, the modification that have been assumed in order to make it compatible with the current 802.11n are also included.
2. The proposal gathers also for the first time, in a single framework, the principal steps that any technology has to overcome during the first phase of the standardization process: theoretical background and evaluation, design of the architecture, computer based simulations at PHY and MAC level and network evaluation.
3. A specific MAC layer superframe schedule is designed and evaluated to cope with the specifications of a NOMA-based service delivery.
4. The evaluation of each Key Performance Indicator (KPI) has been carried out according to strict conditions close to real applications. In particular, extensive simulations have been run in order to obtain PER values up to 10^{-3} on the physical layer and 10^{-9} at MAC level. The main reason for this demanding evaluation is to facilitate the possible know-how transfer to industry in the near future.

The paper is organized as follows. The next section describes the characteristics and requirements of industrial communications, as well as the description of two typical industrial scenarios. Section III provides an overview of the system proposal based on 802.11n and NOMA. Section IV discusses necessary modifications in the PHY in order to use NOMA with the 802.11n and the proposal is evaluated in terms of reliability. Section V introduces a deterministic MAC layer based on Time Division Multiple Access (TDMA) to evaluate the PHY proposal in combination with retransmission techniques. Then, in section VI the PHY/MAC proposal is tested in a realistic use case describing the system network performance. Finally, Section VII contains the main conclusions.

II. INDUSTRIAL COMMUNICATION TECHNOLOGIES AND REQUIREMENTS

This section presents the main wireless solutions for industrial communications and the possible use cases.

A. WIRELESS TECHNOLOGIES

According to available literature, such as [12], 802.11 and 802.15 are the most common building blocks in proposals for wireless industrial communications.

802.15 is a working group inside the IEEE 802 specialized in Wireless Personal Area Networks (WPAN) [13]. 802.15 standards are focused on low energy, low-range,

low-data rate, and multi-hop topologies; thus, they are a suitable family for building industrial wireless sensor networks. In addition, the standard supports a deterministic access scheme that provides a bounded latency [14]. However, their usability is commonly limited to monitoring applications due to their low data rate, modest PER, and relatively large latencies [15]. Some wireless networks for industry process monitoring have been effectively implemented using 802.15.4. For instance, [16] presents an 802.15.4-based network for predictive maintenance of manufacturing installations and railway transport. In [16], the authors point out that 802.15.4 can hardly be applied for strict control or safety applications for latency and reliability concerns. Some other works try to improve the feasibility of 802.15.4 for control applications. In [17], a wireless network build with 802.15.4 and a custom TDMA MAC is presented. The TDMA MAC optimizes the network delays and minimizes the cycle time. However, the minimum achievable cycle time was in the 10 ms order, which is out of several industrial applications [18], [19] and they do not mention the achieved PER.

802.11 has been also considered for industrial communication applications. Compared to the 802.15 family, 802.11 provides higher throughput and range. In 2009, the IEEE standardization committees released the standard version IEEE 802.11n [20], which offers up to five times faster transmission modes in comparison with its predecessor, especially due to the introduction of Multiple-Input Multiple-Output (MIMO) systems [21]. Authors in [6] have tested it on real-time industrial communications and the outcome has been a set of guidelines and recommendations in terms of reliability and latency. However, reliability values are only presented up to PER values of 10^{-1}, which is far away from typical industrial performance requirements and it is not realistic to extrapolate those values to very low PER values taking into account the high variance of the industrial wireless propagation channels. Recently, cross layer approaches are a matter of discussion, not only from a PHY-MAC point of view, but also from the network layer perspective. In [22], authors detail different key design aspects in ultra-high-performance wireless networks for critical industrial control applications but unfortunately no simulations results were included. Finally, in [23], SHARP (Synchronous and Hybrid Architecture for Real-time Performance) is presented, a new architecture for industrial automation. Its wireless part includes a novel PHY layer based on 802.11g and integrated in a TDMA MAC structure in order to guarantee deterministic behavior. In this case, convolutional codes are used as channel coding, which are far away from the reliability performance of the LDPC codes introduced in 802.11n.

In summary, the solutions presented so far do not meet the strict industrial requirements and although the 802.11 standard family seems to be more prepared to be integrated in the industry, it cannot be still considered the optimum technology.

B. USE CASES AND SCENARIOS

This paper addresses two industrial communication application use cases: Factory Automation (FA) and Process Automation (PA).

FA refers to industrial communication environments where either individual machines or complete systems require real-time control. These cases are related to production chains, where machines are fundamental elements of the process. Some representative examples would be assembly, packaging, palletizing and manufacturing. Due to the critical role of communications in their respective processes, depending on the application, they require bounded latencies (0.25-10 ms) and Packet Loss Rate (PLR) in the range of 10^{-7}-10^{-9}. In addition, the update time is between 0.5-50 ms and coverage areas are usually well below 100 meters [24], [25].

PA refers to industrial communications for monitoring and diagnosing applications. Examples of PA typical fields are heating, cooling, stirring, and pumping procedures and, in general, machinery condition monitoring. A major difference with FA is the variation speed of measurements and control cycle latency requirements. Most representative values of latency and reliability are within the ranges of 50-100 ms and 10^{-3}-10^{-4} packet loss rate, respectively. Moreover, an update time of 0.1-5 seconds is usually required. Communication ranges are larger than in FA and may vary between 100 and 500 meters [24], [25].

In this work, two deployment scenarios are proposed for system evaluation purposes. In both cases, two services are expected: one critical service (CS), and one non-critical service that will be referred as best effort (BE) service. On the one hand, both PA and FA services will be regarded as critical (CS) even though their requirements in terms of latency and reliability will be different. On the other hand, BE services will not have latency and reliability restrictions and will be applicable in any scenario and use case.

The scenarios provide boundary conditions such as number of nodes, bitrate allocated to each service, service area dimensions and propagation channel characteristics. Scenarios and associated parameters have been based on reference models and experimental data available in [26] and [27] and are gathered in TABLE 1. On the one hand, scenario A represents a relatively large manufacturing hall, with 3200 sq. meters (80×40 m). We assume up to 100 nodes randomly distributed throughout the hall. The access point/gateway will be installed in an optimal position, where LOS and partially obstructed Fresnel zone conditions dominate. On the other hand, Scenario B represents a manufacturing cell, with limited dimensions (10×10 m), up to 20 nodes per cell and NLOS reception conditions. In both cases, a typical manufacturing environment is emulated with plenty of metallic structures and moving machines and staff. The propagation channel (CM 7 or CM 8) choice associated to each environment has been based on [28] and the experiments carried

out on [26]. TABLE 1 summarizes the main parameters and values of each case.

TABLE 1. Use cases and application scenarios.

Parameter	Scenario A Manufacturing Hall	Scenario B Manufacturing Cell
Application	PA	FA
Dimensions (m)	80 x 40 x 10	10 x 10 x10
Number of Nodes	Up to 100	Up to 20
Critical Services (Mbps)	24	12
Non - Critical Services (Mbps)	Up to 48	Up to 48
Channel Model (IEEE 802.15.4)	CM7	CM8

The service capacity requirements associated to the critical services have been differentiated in each scenario. Scenario A has higher number of nodes and lower PER requirement, whereas Scenario B presents less nodes, lower distances and requires a very low PER. Consequently, we have chosen a bitrate capacity close to 24 Mbps for the critical service, whereas this number is reduced to 12 Mbps for the manufacturing cell in scenario B. In both cases, the target throughput of the BE services is 48 Mbps.

III. NOMA-BASED 802.11n

This section introduces the basic principles of NOMA and the requirements for its integration within the 802.11n standard.

A. NON-ORTHOGONAL MULTIPLE ACCESS (NOMA)

Power domain NOMA has recently been successfully applied to broadcast communications [29]. P-NOMA is very efficient in providing different levels of robustness to different services within the same RF channel. This makes the technology an adequate candidate for multiplexing services with unbalanced requirements in industrial wireless applications.

NOMA consists of a signal ensemble composed of several layers, each one taking a portion of the total power delivered by the transmitter. Each layer is configured targeting different robustness levels, decoding thresholds and capacities. The configuration is a function of the modulation/coding choice and the power assignment to each layer. The power splitting is described by Δ (injection level, measured in decibels). The concept of multiplexing two services with NOMA is illustrated in Fig. 2 and the NOMA signal ensemble can be expressed as:

$$x_{NOMA}(k) = x_{CS}(k) + g \cdot x_{BE}(k), \tag{1}$$

where $x_{CS}(k)$ and $x_{BE}(k)$ are the two data streams combined into $x_{NOMA}(k)$ and k is the subchannel index. The injection level g defines the linear power allocation ratio between layers in linear units [7]. This work is based on a two-layer system and, therefore, layers are referred as Upper Layer (UL) and Lower Layer (LL), where CS is transmitted in the UL and BE in the LL, as described in Fig. 2. Due to the power splitting, each service is delivered with less power than in the single-layer case and the power allocated to each

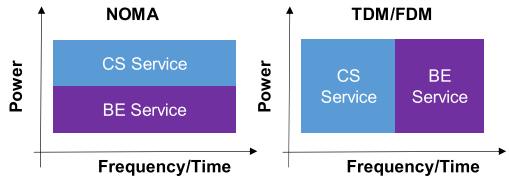


FIGURE 2. Non-Orthogonal and Orthogonal Multiplexing (OM) schemes sharing both the time and the frequency resources.

service can be calculated as:

$$\sigma_i = \frac{10^{\frac{\Delta_{i-1}}{10}}}{\sum_{i=1}^{N} 10^{\frac{\Delta_{i-1}}{10}}}, \tag{2}$$

where, σ_i is the power allocation ratio for layer i (i.e., $i = 1, 2$), N is the total number of layers in the signal ensemble (i.e., $N = 2$) and Δ is the injection level in dB.

The idea of multiple signals on the same channel with unequal error protection was described time ago by some information theory papers such as [30]. The currently available advances in forward error correction make NOMA feasible in real systems. The performance of latest LDPC codes, for instance, is less than half a dB away from the Shannon limit [31]. The second driver for feasibility is the implementation of the signal cancellation structure proposed in [6], which reduces significantly the complexity of the receiver [32]. In [6], a comprehensive comparison of the spectral efficiency of NOMA and traditional TDM/FDM schemes was presented. It is proven that provided the service reception thresholds are unbalanced, the overall gain offered by NOMA can be up to 3 bps/Hz. In practice this gain can be invested in two benefits: 30% higher throughput compared to TDMA solutions [33], [34] or a considerable PER reduction [35]. NOMA will always outperform TDM/FDM, provided an efficient enough channel coding mechanism is used.

Industrial communication systems are typically rolled out for supervision and control purposes, where security, reliability and delay will be critical design constraints. NOMA addresses both security and reliability directly by operating close to zero and even with negative reception thresholds for the most challenging scenarios. Flexible configuration of layers is also an advantage for increasing the security of the PHY. Moreover, processing time and complexity at the receiver are key for system feasibility in industrial environments. The cancellation associated to additional layer decoding involves additional processing latency on the receiver. However, the CS is not affected by this inconvenience as it is transmitted in the UL. Even though, the selected error coding technique for the physical layer waveform will be very important for the Successive Interference Cancellation (SIC) delay [36]. In [32], for instance, the latency associated to large LDPC codes was analyzed. The results show that the number of iterations of the LDPC decoding algorithm remains lower than 5 for Signal-to-Noise Ratio (SNR) values below 4-5 dB,

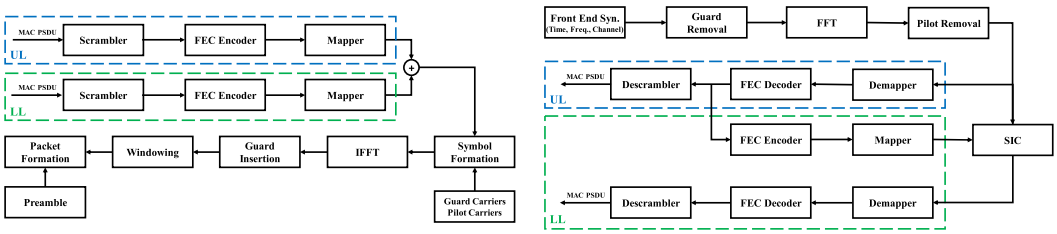


FIGURE 3. General block diagram of the proposed transceiver: (a) NOMA-based 802.11n transmitter, (b) NOMA-based 802.11n receiver.

making the associated decoding time not relevant on the overall OFDM receiver processing time budget, mostly influenced by MAC operations [32]. It should be noted that time requirements cannot be described on a general basis. Each application scenario requires a thorough analysis of the control-cycle vs. receiver processing time.

B. NOMA ON 802.11n

A new communication architecture based on modifications to the PHY of the 802.11n standard in order to include NOMA is presented in this subsection.

First, at the MAC layer, the LL Physical Service Data Unit (PSDU) packet length is adjusted according to the UL length in order to ensure that both components, UL and LL, will make use of the same number of symbols per frame in the physical layer. This adaptation will enable the use of a simplified NOMA mapping block, which will be included immediately after the combination of individual mappers plus an amplitude adaptation operation (injection level). Consequently, the LL receiver complexity will be greatly reduced. Previous work already assessed the relevancy of a proper choice of the data overlapping for efficient and simple receiver implementation [32]. In particular, the length of the BE PSDU can be calculated as:

$$PSDU_{BE} = \left\lfloor \left(Sym_{Num(CS)} * NDBPS_{BE} - N_{Service} \right) / 8 \right\rfloor \quad (3)$$

where, $Sym_{Num(CS)}$ is the number of complex symbols required for the CS, $NDBPS_{BE}$ is the number of coded bits per OFDM symbol of the BE service and $N_{Service}$ is the number of bits in the service field.

In fact, if the system uses OFDM and the framing structure of each one of the layers are synchronized, the receiver complexity and latency are limited. The mapping and framing stage are inherited directly from 802.11n MCS. The only difference is that both services are combined into a single stream after mapping (see Fig. 3 (a)). To do so, the complex symbols corresponding to the LL are attenuated by a predefined injection level. The injection level in linear units (i.e., g) indicates the relative power distribution between both layers, so g has a real value within the range [0, 1), where $g = 0$ results in a single-layer system and $g = 1$ results in a two-layer system with equal power distribution. Then, the UL and the attenuated LL are added creating the definitive NOMA signal ensemble. After superimposing both layers, the output constellation is normalized. In this particular case, the UL is in charge of delivering CS content, while BE content is mapped into the LL. Once both services are combined into a single P-NOMA signal ensemble, both services will undergo the same processing stages. The final physical waveform is then sent in the form of PHY packets composed of a preamble and a data field. The detailed structure of the PHY packet is maintained as defined in 802.11n.

The first modules of the NOMA receiver are identical to those on current 802.11n receivers (See Fig. 3 (b)). After RF conditioning, the carrier and time synch stages prepare the input for channel and equivalent noise estimation.

Then a Least Square (LS)-based equalization is performed and the CS goes through decoding, demapping and descrambling stages. The CS bit stream will be available for all sensor/actuators of the service area without additional complexity since the lower layer will be assumed as additional noise for the UL. If the available SNR permits, the cancellation and lower layer decoding will be the next steps. The cancellation will depend on the injection level, the channel estimation metrics and the CS itself. At the receiver, the NOMA ensemble on the k-th channel can be expressed as:

$$y_{NOMA}(k) = \left(x_{CS}(k) + g \cdot x_{BE}(k) \right) \cdot h(k) + n(k), \quad (4)$$

where $y_{NOMA}(k)$ is the received symbol, $h(k)$ is a static multipath channel and $n(k)$ is the sum of AWGN noise and other additive interference. CS can be accessed as usual without any further requirements. Once the UL has been decoded, a signal cancellation algorithm provides access to the BE. In this work, for the signal cancellation stage the Hard-SIC cancellation scheme is proposed due to its good trade-off between complexity and performance. In particular, the same Hard-SIC cancellation mechanism has been adopted by receivers of latest Advanced Television Systems Committee (ATSC) digital TV standard, ATSC 3.0, showing a successful balance between complexity and performance [29]. The decoded CS layer is coded and modulated again, and then, it is removed from the equalized NOMA signal

(See Fig. 3 (b)). After these stages, the received signal can be expressed as:

$$\tilde{x}_{BE}(k) = g \cdot x_{BE}(k) + i(k) + n(k), \quad (5)$$

where $i(k)$ is the error cancellation due to the channel estimation error and $\tilde{x}_{BE}$ is the obtained BE data [6]. Eventually, the obtained signal goes through the same decoding, demapping and descrambling stages as the CS.

C. CAPACITY COMPARISON

The purpose of this section is to demonstrate theoretically that the inclusion of NOMA in 802.11n standard provides an increase of the system capacity. Fig. 4 shows the capacity gain of NOMA for different configurations. The x-axis represents the percentage of time dedicated to the delivery of CS content, while the y-axis is the capacity gain obtained as the difference between the optimum TDM configuration throughput and the optimum NOMA throughput. The different curves imply different restriction on the maximum SNR value that CS and BE services can assume. Based on those values the optimum MCS configuration is carried out for both technologies.

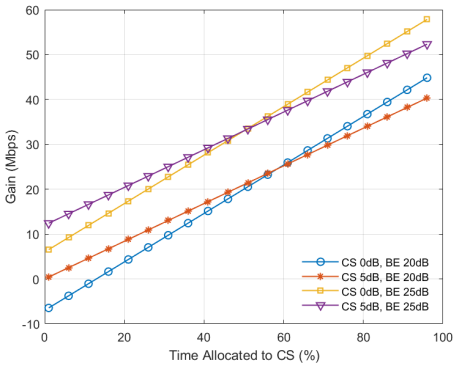


FIGURE 4. Potential gain using NOMA in combination with 802.11n.

Firstly, Fig. 4 shows that in the majority of the cases, NOMA offers a considerable capacity gain with maximum values between 40-60 Mbps. The gain increases linearly proportional to the percentage of time allocated to CS, where the higher are the time requirements of CS, the higher is the gain that NOMA can offer. In addition, the curves present different gradient due to the performance of NOMA. In fact, the higher is the asymmetry between the SNR requirements of the services, the higher is the gradient of the grain that NOMA can offer. Finally, it should also be remarked that for the low values of time percentage allocated to CS, NOMA does not provide enough gain, and so, TDM could be a better solution. However, taking into account the industrial time requirements, cases where the majority of the time are allocated to non-critical services are not very representative. The main reason is that service configuration is carried out to meet the strict requirements of CS and, therefore, a low percentage of time dedicated to CS may indicate that the service configuration is not optimized.

D. PROPOSED CONFIGURATIONS

Based on the scenarios described in previous sections, two different configurations are proposed for each of the multiplexing schemes. On the one hand, in the NOMA-Config. A (PA), the UL is targeting a CS with a throughput of about 24 Mbps (MCS1), whereas the LL is delivering 48 Mbps (MCS3) as a BE service and the injection level is −10.5 dB. In NOMA-Config. B (FA), the UL is decreased to 12 Mbps (MCS0) and the injection level to −10 dB in order to satisfy the more tight requirements of FA, while the EL configuration remains the same. PHY configurations are summarized in TABLE 2.

TABLE 2. PHY configurations.

		NOMA	TDM
PA - Scenario A	Config. A CS (24 Mbps) BE (48 Mbps)	QPSK 1/2 16QAM 1/2	16 QAM 1/2 64QAM 2/3
FA - Scenario B	Config. B CS (12 Mbps) BE (48 Mbps)	BPSK 1/2 16QAM 1/2	QPSK 1/2 64QAM 2/3

On the other hand, the resources are always 50%-50% shared by both service types in TDM/FDM configurations. In the first case, TDM-Config. A, the MCS3 is proposed for the CS, whereas for the second scenario TDM-Config. B the MCS1 is enough to satisfy the minimum CS throughput. Eventually, in both cases the BE data is conveyed with MCS5 and a throughput of 48 Mbps. The packet length of the CS service has been configured to deliver a payload of 18 bytes.

IV. PHYSICAL LAYER (PHY)

This section provides guidelines for the design of the proposed PHY layer and presents the performance analysis.

A. DESIGN

The potential gain of the NOMA proposal has been studied by through simulations of the complete physical layer [37], [38]. A prototype transmitter and receiver chain has been implemented for this purpose. This prototype is compliant with the 802.11 a/g/n standard and includes modifications to enable NOMA multiplexing capabilities as described in Section III (See Fig. 3).

The system evaluation set up has included two industrial propagation models for performance evaluation purposes: the CM 7 (Scenario A) and CM 8 (Scenario B) channel models [28]. Considering this work as a feasibility analysis of the upper limit of the potential gain provided by NOMA to 802.11n, perfect synchronization and channel estimation are assumed. Access the Lower Layer is granted after SIC cancellation as explained in Section III.B. In line with the ideal

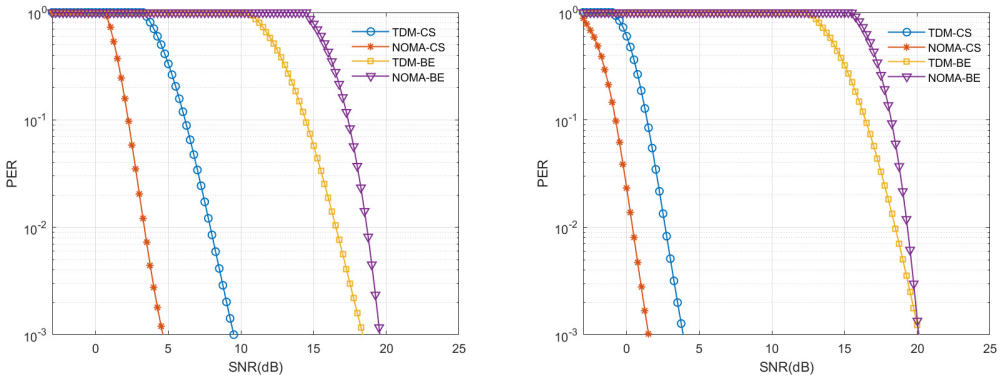


FIGURE 5. Performance results for the PHY layer: (a) Scenario A (b) Scenario B.

synch and estimation considerations, cancellation has been also considered ideal. Previous works have demonstrated that it is possible to keep the cancellation noise below the system AWGN [7], [39].

The PHY performance has been evaluated using. This is one of the key performance indicators in industrial applications (a PSDU packet is considered erroneous if any of the bits are decoded incorrectly). The number of simulated packets and the SNR simulation steps have been adapted to the expected PER values. In particular, the simulated packet number is always at least one order of magnitude higher than the required value for obtaining the desired PER value assuming one single error. In addition, for the best granularity the lower PER values are calculated within 0.25 dB steps. The obtained results are fed to OMNeT++ as the PHY error model.

B. EVALUATION

As part of the proposal study, the PHY performance of configurations proposed in Section III.D are displayed in Fig. 5 (a) and Fig. 5 (b) for Scenario A and Scenario B, respectively. The receiving thresholds at the physical layer are limited to PER $= 10^{-3}$. It must be noted, that the MAC level retransmissions, even limited for latency considerations, will bring PER values down to 10^{-8}.

The UL gain in Fig. 5 (a) is close to 5 dB, whereas the LL performance is within a 1.5 dB margin for both multiplexing schemes. In Fig. 5 (b), however, the NOMA UL gain is reduced to about 2.5 dB and the LL performance is even closer. The relevance of the result lies on the significant gain in PHY performance, close to 5 dB gain for a very challenging scenario, which will provide room for flexible MAC design. The reason for this gain is that in order to compensate the capacity reduction in TDMA systems because of the resource sharing between CS and BE, higher modulation orders have to be used in comparison with NOMA. Therefore, more robust configurations can be used in NOMA, while the same capacity as in TDMA is offered.

V. MAC LAYER DEFINITION AND PERFORMANCE

This section presents the design of the proposed MAC layer and the performance analysis.

A. DESIGN

In this section, a specific MAC layer for implementing the NOMA-based 802.11n PHY proposal is presented. This MAC layer proposal combines in the same time schedule the NOMA signal ensemble with a TDMA medium access technique to guarantee deterministic packet delivery and a packet retransmission scheme in the time domain to increase the reliability. Additionally, it is assumed that the nodes share a common time reference that allows guaranteed access to the medium without any interferences, such as [40], [41]. In order to measure the benefits of introducing NOMA, a single-layer TDMA MAC with packet retransmission schemes has been designed and implemented. Both MAC layer superframe structures, TDMA and NOMA + TDMA, are presented in Fig. 6.

Both superframe structures are composed of four different periods: CS transmission, uplink feedback, on-demand retransmissions and BE period. During the CS transmission period, the AP sends the corresponding CS information to each slave in a dedicated time slot. Then, in the uplink feedback period, each slave sends an ACK packet if the CS has been correctly received or NACK packet if the CS has not been correctly received. The ACK/NACK packet transmission is carried out in single-layer mode in both cases, TDMA and NOMA. In the on-demand retransmission period, each of the requested retransmissions in the previous period is carried out by the AP in another dedicated time slot. To define

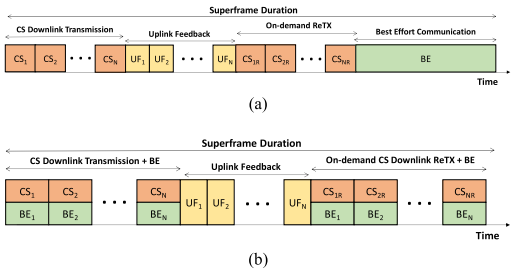


FIGURE 6. Superframe time representation: (a) TDMA, (b) NOMA + TDMA.

the duration of this period, the worst possible case is assumed, that is, one slot per each CS packet retransmission is reserved. Finally, in the BE period non-critical information is delivered. The length of each period depends on the MCS on use. The lower the MCS is, the higher the required transmission time is, and so, the longer the period is. Taking into account the PHY configurations (TABLE 2), since NOMA uses lower MCS values, NOMA periods are longer than TDMA periods for the same packet lengths.

The main difference between the two superframe structures is that by using NOMA, BE services are delivered in the LL. On the one hand, this effect reduces the superframe time duration since only three periods have to be reserved in the time domain. On the other hand, as NOMA techniques are implemented in combination with a deterministic TDMA structure, BE services are transmitted in the lower layer aligned with CS. In consequence, it will be much easier to determine the timing of the transmission and the reception of each BE packet since the determinism obtained by the use of TDMA is also applied to the BE services. The consequence is an easier prediction of latency also for BE services. Another advantage of the simultaneous delivery of CS and BE services is that when a retransmissions are required for CS, the BE packet is also retransmitted, and so, a second chance is obtained for the correct reception of BE traffic.

B. EVALUATION

In order to evaluate the reliability of the proposed MAC layer, the OMNeT++ network simulator has been used [42]. The designed MAC layer has been implemented in combination with the PHY configurations described in the previous section. The scenarios (A and B) described in section II.B have been used (see TABLE 1). The number of nodes in the network has been set initially to 50 in scenario A and 10 in scenario B. Later, the impact of increasing the number of nodes and the maximum slots of a TDM/NOMA superframe as a function of T_{cycle} will be calculated.

Firstly, the superframe length is different for each technology and configuration. Therefore, the minimum T_{Cycle} would also be different, since in this work, the T_{Cycle} is assumed as the minimum required time to carry out all the required communications described in the superframe (i.e., the superframe length), which means that the lower the T_{Cycle}, the faster are the slaves updated. In particular, for the same number of nodes, NOMA superframes are shorter than TDM superframes. In fact, TDM superframe is 22% and 17.5% longer than NOMA superframe in scenario A and scenario B, respectively.

In Fig. 7, CS reliability curves obtained from the MAC evaluation are presented in terms of PER and PLR vs. SNR. This work assumes PER as the error rate without retransmissions and PLR as the error rate after implementing retransmissions. Regarding the performance obtained in scenario A (Fig. 7 (a)), NOMA-based MAC offers better results than TDM for restrictive PER and PLR values (i.e., 10^{-9}). In particular, NOMA performs 2.5 dB better than TDM in terms of PLR and more than 6 dB in terms of PER.

If the impact of the retransmission schemes is individually analyzed, the relative reliability improvement achieved in TDM is higher than in NOMA (i.e., 7dB vs. 4 dB). The difference on the impact of the retransmissions is associated to PHY configurations in each case, as TDM uses a less robust MCS, retransmissions are more effective. Similar reliability results are obtained in Fig. 7 (b) for scenario B. In this case, the gain that NOMA-based system offers in comparison with TDM system is 2.3 dB and 1.7 dB in terms of PLR and PER, respectively. In general, performance values are better in scenario B since more robust configurations are implemented for the CS.

The PER performance of the BE component is presented in Fig. 8. In the case of scenario B, performance results are very similar in both cases, the difference is lower than half dB. However, in scenario A TDM offers slightly higher reliability (around 3 dB). Nevertheless, the reliability difference for both technologies is not as critical as in the case of CS. Moreover, PLR results of the NOMA MAC layer are not included in Fig. 8 since they are very similar to the PER values. The main reason for this effect is the short time interval between the transmission and the retransmission, usually much shorter than channel time coherence.

VI. COMPREHENSIVE NETWORK EVALUATION

In this section, the PHY layer (Section IV) and the MAC (Section V) are combined based on the scenarios detailed in Section II.B. The objective of this study is to test the proposed solution in terms of reliability and latency in a realistic network. Simulations have been carried out in OMNeT++. The system uses a 40 MHz channel bandwidth in the 2.4 GHz ISM band, −90dBm is the thermal noise power and the transmitted power is set to 10 dBm. It is assumed that the nodes are randomly distributed along the network. The path loss model includes the free space path losses plus a shadowing component that is characterized with a log-normal distribution with mean zero and $\sigma = 6$ dB [43]. A summary of the parameters is presented in TABLE 3.

Fig. 9 shows the results of reliability associated to each scenario. Scenario A and Scenario B are presented independently in Fig. 9 (a) and Fig. 9 (b), respectively. In general,

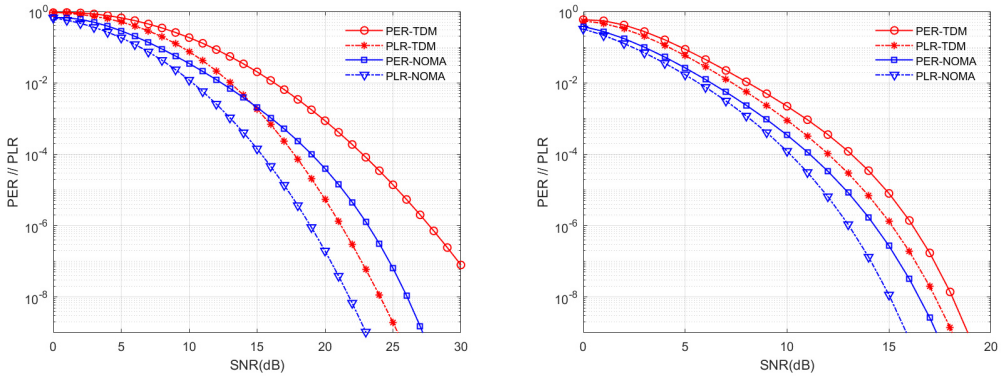


FIGURE 7. MAC layer CS reliability results: (a) Scenario A, (b) Scenario B.

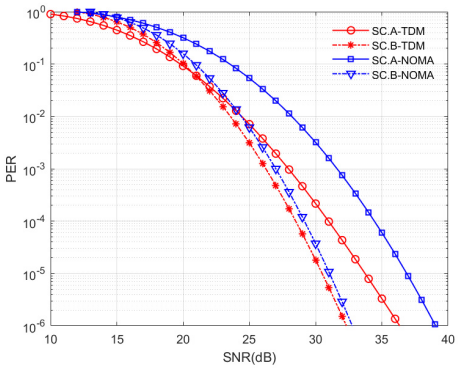


FIGURE 8. MAC layer BE reliability results.

TABLE 3. Simulation parameters.

Parameter	Value
Center Frequency	2.4 GHz
Bandwidth	40 MHz
Noise Power	-90 dBm
Transmitted Power	10 dBm
Shadowing	Mean = 0 dB σ = 6 dB
Number of Nodes	50 in Sc. A 10 in Sc. B
Channel Model	CM7 in Sc. A CM8 in Sc. B
Network Size	80 x 40 m^2 in Sc. A 10 x 10 m^2 in Sc. B
CS data rate	24 Mbps in Sc. A 12 Mbps in Sc. B
BE data rate	48 Mbps

NOMA offers considerable better performance for the CS. In fact, in all CS cases (see Fig. 9), NOMA outperforms TDM in at least one order of magnitude. In particular, PLR values below 10^{-8} are obtained using NOMA. On the other hand, looking at BE services, TDM offers slightly better performance results than NOMA in Scenario A, while in Scenario B, results are very similar for both technologies. In addition, to the reliability results, the effect of the error packets and the retransmissions over the received throughput is presented in TABLE 4. As in the case of reliability results, NOMA provides a better use of system data capacity for CS services as results appear closer to the maximum (i.e., 24 Mbps in scenario A and 12 Mbps in scenario B) results. On the contrary, regarding the received effective throughput related with BE services, TDM-based architecture provides slightly better results, although, in both cases, the throughput decrease is assumable taking into account the characteristics of BE services. In summary, in order to improve the reliability of CS by using NOMA, a tradeoff has to be assumed between the performance of the CS and the BE.

Concerning time requirements, an end-to-end (E2E) latency analysis is presented in TABLE 5. NOMA can assume lower T_{Cycle} values while carrying out the same services as in TDMA. Then, the minimum E2E latency is defined as the time required to deliver a CS data packet. As expected, the minimum latency of NOMA is higher than the minimum latency of TDM because it uses lower MCS in the PHY configuration. The maximum latency is obtained when the CS packet is received in the on-demand retransmission period. Again, the different MCS choices cause different maximum values. Finally, the last column in TABLE 5, presents the mean latency obtained from the simulation of all the scenarios. In general, mean values are very close to the minimum latency values in all cases. This means that few retransmissions have been required to deliver the CS information correctly. However, the difference between the minimum and the mean latency is lower in NOMA than in TDM, because of its higher reliability. Doubtlessly, results

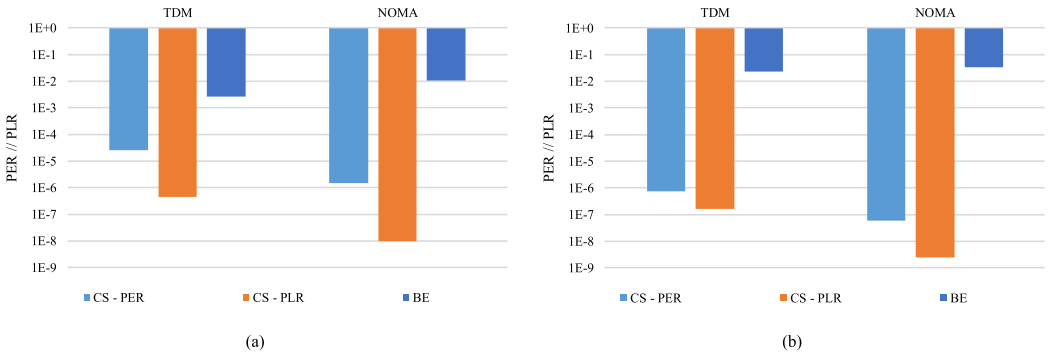


FIGURE 9. Reliability results obtained in: (a) Scenario A, (b) Scenario B.

TABLE 4. Received effective throughput.

	CS (Mbps)	BE (Mbps)
Sc. A - TDM	23.999401	47.869998
Sc. A - NOMA	23.999965	47.478968
Sc. B - TDM	11.999991	46.875871
Sc. B - NOMA	11.999999	46.349872

TABLE 5. E2E latency comparison.

	T_{cycle} (ms)	Min. E2E Latency (µs)	Max. E2E Latency (ms)	Mean E2E Latency (µs)
Sc. A - TDM	10.4	50	5.20	50.13000
Sc. A - NOMA	8.50	58	5.60	58.00500
Sc. B - TDM	2.28	58	1.12	58.00100
Sc. B - NOMA	1.94	70	1.24	70.00003

indicate that NOMA-based systems are more deterministic than TDM systems, since instantaneous latency values are less variable.

The length of the superframe structure is conditioned by the number of nodes that are served in each cycle time. Therefore, depending on the time requirements that have to be guaranteed (maximum cycle time), different number of nodes can be served depending on the multiplexing technology used. TABLE 6 contains the number of nodes that can be included in the designed superframe for different T_{cycle} values. The first two values (100 ms and 50 ms) are related to PA environments, where time requirements are more flexible. Then, the other three values (10 ms, 5 ms and 1 ms) are related to FA environments. In general, by using NOMA, more nodes can be included, specifically, the amount of served nodes can be increased by around 20%. It is worth mentioning the case of a T_{cycle} of 100 ms, where the number of nodes in the superframe is maximum. In this case, by using NOMA 108 extra nodes could be served in scenario A and 77 more in scenario B.

Finally, it is important to highlight that although originally only scenario B is oriented to FA environments (scenario A is oriented to PA), based on the latency and

TABLE 6. Served nodes for different T_{CYCLE} sizes.

	T_{cycle} 100 ms	50 ms	10 ms	5 ms	1 ms
Sc. A - TDM	480	240	48	24	4
Sc. A - NOMA	588	294	58	29	5
Sc. B - TDM	438	219	43	21	4
Sc. B - NOMA	515	257	51	25	5

reliability results obtained in TABLE 5 and Fig. 9, in both scenarios FA applications could be delivered by using NOMA.

VII. CONCLUSION

This article provides a comprehensive proposal of a NOMA-based 802.11n architecture for industrial applications. In addition to the necessary theoretical basis and the specific PHY/MAC NOMA design in combination with 802.11n, realistic requirements and scenarios have been detailed based on PA and FA, where the NOMA-based 802.11n proposal could be implemented. First, a simulation setup has been presented at PHY level in order to test the reliability increase obtained by introducing NOMA in the PHY. Then, a particular MAC layer structure has been presented to manage a deterministic medium access with a useful retransmission scheme. Finally, the performance of the PHY/MAC design has been tested in the described industrial scenarios. Reliability and latency have been the principal KPI.

In general, NOMA offers significantly better reliability results than TDMA. On the one hand, as shown in Section IV, the reliability of the CS is highly increased by using more robust MCS configurations. On the other hand, the robustness increase in the NOMA-based CS is maintained in the MAC layer performance curves. In fact, up to 6 dB of gain can be obtained for PER values of 10^{-9}, since NOMA-based CS can be obtained for a SNR of around 27 dB, while for TDMA around 33 dB of SNR are required. In addition, the effect of the proposed time retransmission scheme is confirmed in both technologies NOMA and TDM, since the required SNR is reduced. It has been observed that in some cases, TDMA-based systems offer slightly better results for BE services, but this fact has not been considered critical for FA or PA scenarios. It is also important to emphasize that

in all the simulated realizations in section VI, an important improvement is obtained PER and PLR metrics for NOMA. In fact, NOMA-based results outperform TDM results in at least one order of magnitude in each of the evaluated cases. In both scenarios, the PLR value obtained by using NOMA is below 10^{-8}. Consequently, a tradeoff has to be assumed between CS and BE, where in order to obtain a considerable gain in the CS, BE services have to assume a little penalization.

Concerning latency analysis, results are more similar between both technologies than in the case of reliability. As presented in section VI, although it is true that minimum and maximum latency values are slightly lower in TDM system, in both cases latency values meet the FA time requirements. However, the mean latency results in NOMA are very close to the minimum latency, which means that the instantaneous latency values are more constant. This effect occurs due to the increase in reliability in NOMA cases and it implies that NOMA-based communications are more deterministic than TDMA-based communications. Moreover, it is important to highlight that due to the superframe structure, more nodes can be included in the NOMA-based superframe than in the TDMA-based superframe for the same T_{cycle}.

Finally, an extra advantage is obtained with NOMA in comparison to TDMA: determinism in BE services. In most cases, absolute latency values are not critical for BE services, but in our proposal, latency of BE services becomes more easily predictable than in TDMA frame structures. Finally, as a future work, it would be interesting to validate this proposal in an industrial environment with real hardware. Hence, how to implement this proposal in commercial devices would be studied for creating compatible transceivers with NOMA technology.

REFERENCES

[1] A. Varghese and D. Tandur, "Wireless requirements and challenges in industry 4.0," in *Proc. Int. Conf. Contemp. Comput. Informat. (IC3I)*, Mysore, India, Nov. 2014, pp. 634–638.

[2] R. Drath and A. Horch, "Industrie 4.0: hit or hype? [Industry Forum]," *IEEE Ind. Electron. Mag.*, vol. 8, no. 2, pp. 56–58, Jun. 2014.

[3] R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems: The next computing revolution," in *Proc. Design Automat. Conf.*, Anaheim, CA, USA, Jun. 2010, pp. 731–736.

[4] M. Annunziata and P. C. Evans, "Industrial Internet: Pushing the boundaries of minds and machine," in *General Electric*. General Electric Reports, 2012.

[5] Z. Fernandez, C. Cruces, I. Val, and M. Mendicute, "Deterministic real-time access point concepts for industrial hybrid Ethernet/IEEE 802.11 networks," in *Proc. IEEE Int. Workshop Electron., Control, Meas., Signals Appl. Mechatronics (ECMSM)*, Donostia-San Sebastian, Spain, May 2017, pp. 1–6.

[6] F. Tramarin, S. Vitturi, M. Luvisotto, and A. Zanella, "On the use of IEEE 802.11n for industrial communications," *IEEE Trans. Ind. Informat.*, vol. 12, no. 5, pp. 1877–1886, Oct. 2016.

[7] L. Zhang, W. Li, Y. Wu, X. Wang, S.-I. Park, H. M. Kim, J.-Y. Lee, P. Angueira, and J. Montalban, "Layered-Division-multiplexing: Theory and practice," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 216–232, Mar. 2016.

[8] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[9] E. Arruti, M. Mendicute, and M. Barrenechea, "Unequal error protection with LDM in inside carriage wireless communications," in *Proc. 15th Int. Conf. ITS Telecommun. (ITST)*, Warsaw, Poland, May 2017, pp. 1–5.

[10] S. Murugaveni and K. Mahalakshmi, "A novel approach for non-orthogonal multiple access for delay sensitive industrial IoT communications for smart autonomous factories," *J. Ambient Intell. Humanized Comput.*, to be published.

[11] E. Iradier, J. Montalban, L. Fanari, P. Angueira, O. Seijo, and I. Val, "NOMA-based 802.11n for broadcasting multimedia content in factory automation environments," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Jeju, South Korea, Jun. 2019, pp. 1–6.

[12] A. Frotzscher, U. Wetzker, M. Bauer, M. Rentschler, M. Beyer, S. Elspass, and H. Klessig, "Requirements and current solutions of wireless communication in industrial automation," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, Sydney, NSW, Australia, Jun. 2014, pp. 67–72.

[13] M. Bal, "Industrial applications of collaborative wireless sensor networks: A survey," in *Proc. IEEE 23rd Int. Symp. Ind. Electron. (ISIE)*, Istanbul, Turkey, Jun. 2014, pp. 1463–1468.

[14] G. Cena, A. Valenzano, and S. Vitturi, "Hybrid wired/wireless networks for real-time communications," *IEEE Ind. Electron. Mag.*, vol. 2, no. 1, pp. 8–20, Mar. 2008.

[15] V. C. Gungor and G. P. Hancke, "Industrial wireless sensor networks: Challenges, design principles, and technical approaches," *IEEE Trans. Ind. Electron.*, vol. 56, no. 10, pp. 4258–4265, Oct. 2009.

[16] K. Das, P. Zand, and P. Havinga, "Industrial wireless monitoring with energy-harvesting devices," *IEEE Internet Comput.*, vol. 21, no. 1, pp. 12–20, Jan. 2017.

[17] M. Anwar, Y. Xia, and Y. Zhan, "TDMA-based IEEE 802.15.4 for low-latency deterministic control applications," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 338–347, Feb. 2016.

[18] M. Luvisotto, Z. Pang, and D. Dzung, "High-performance wireless networks for industrial control applications: New targets and feasibility," *Proc. IEEE*, vol. 107, no. 6, pp. 1074–1093, Jun. 2019.

[19] G. J. Sutton, J. Zeng, R. P. Liu, W. Ni, D. N. Nguyen, B. A. Jayawickrama, X. Huang, M. Abolhasan, Z. Zhang, E. Dutkiewicz, and T. Lv, "Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2488–2524, 3rd Quart., 2019.

[20] *Orthogonal Frequency Division Multiplexing (OFDM) PHY Specification, de Standard—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, Nueva York, Estados Unidos, IEEE Computer Society, Standard 802.11, 2012, pp. 1583–1630.

[21] J. Lorincz and D. Begusic, "Physical layer analysis of emerging IEEE 802.11n WLAN standard," in *Proc. 8th Int. Conf. Adv. Commun. Technol.*, Phoenix Park, South Korea, Feb. 2006, p. 194.

[22] M. Luvisotto, Z. Pang, and D. Dzung, "Ultra high performance wireless control for critical applications: Challenges and directions," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1448–1459, Jun. 2017.

[23] O. Seijo, Z. Fernandez, I. Val, and J. A. Lopez-Fernandez, "SHARP: A novel hybrid architecture for industrial wireless sensor and actuator networks," in *Proc. 14th IEEE Int. Workshop Factory Commun. Syst. (WFCS)*, Imperia, Italy, Jun. 2018, pp. 1–10.

[24] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Müller, T. Elste, and M. Windisch, "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 70–78, Feb. 2017.

[25] S. Dietrich, G. May, O. Wetter, H. Heeren, and G. Fohler, "Performance indicators and use case analysis for wireless networks in factory automation," in *Proc. 22nd IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Limassol, Cyprus, Sep. 2017, pp. 1–8.

[26] R. Croonenbroeck, L. Underberg, A. Wulf, and R. Kays, "Measurements for the development of an enhanced model for wireless channels in industrial environments," in *Proc. IEEE 13th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Rome, Italy, Oct. 2017, pp. 1–8.

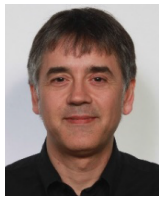
[27] X. Jiang, Z. Pang, M. Luvisotto, F. Pan, R. Candell, and C. Fischione, "Using a large data set to improve industrial wireless communications: Latency, reliability, and security," *IEEE Ind. Electron. Mag.*, vol. 13, no. 1, pp. 6–12, Mar. 2019.

[28] A. F. Molisch, *IEEE 802.15. 4a Channel Model-Final Report*, IEEE Standard P802, vol. 15, no. 4, pp. 1–41, 2004.

[29] S. I. Park, J.-Y. Lee, S. Myoung, L. Zhang, Y. Wu, J. Montalban, S. Kwon, B.-M. Lim, P. Angueira, H. M. Kim, N. Hur, and J. Kim, "Low complexity layered division multiplexing for ATSC 3.0," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 233–243, Mar. 2016.
[30] P. Bergmans and T. Cover, "Cooperative broadcasting," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 317–324, May 1974.
[31] L. Michael and D. Gomez-Barquero, "Bit-interleaved coded modulation (BICM) for ATSC 3.0," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 181–188, Mar. 2016.
[32] S. I. Park, Y. Wu, L. Zhang, J. Montalban, J.-Y. Lee, P. Angueira, S. Kwon, H. M. Kim, N. Hur, and J. Kim, "Low complexity layered division multiplexing system for the next generation terrestrial broadcasting," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.*, Jun. 2015, pp. 1–3.
[33] S.-I. Park, J.-Y. Lee, B.-M. Lim, S. Kwon, J.-H. Seo, H. M. Kim, N. Hur, and J. Kim, "Field comparison tests of LDM and TDM in ATSC 3.0," *IEEE Trans. Broadcast.*, vol. 64, no. 3, pp. 637–647, Sep. 2018.
[34] J. Montalban, P. Scopelliti, M. Fadda, E. Iradier, C. Desogus, P. Angueira, M. Murroni, and G. Araniti, "Multimedia multicast services in 5G networks: Subgrouping and non-orthogonal multiple access techniques," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 91–95, Mar. 2018.
[35] E. Iradier, J. Montalban, L. Fanari, P. Angueira, L. Zhang, Y. Wu, and W. Li, "Using NOMA for enabling Broadcast/Unicast convergence in 5G networks," *IEEE Trans. Broadcast.*, vol. 66, no. 2, pp. 503–514, Jun. 2020.
[36] M. Zhan, Z. Pang, D. Dzung, and M. Xiao, "Channel coding for high performance wireless control in critical applications: Survey and analysis," *IEEE Access*, vol. 6, pp. 29648–29664, 2018.
[37] F. Qin, X. Dai, and J. E. Mitchell, "Effective-SNR estimation for wireless sensor network using Kalman filter," *Ad Hoc Netw.*, vol. 11, no. 3, pp. 944–958, May 2013.
[38] A. A. Kumar, K. Ovsthus, and L. M. Kristensen, "An industrial perspective on wireless sensor networks—A survey of requirements, protocols, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1391–1412, 3rd Quart., 2014.
[39] L. Zhang, Y. Wu, W. Li, H. M. Kim, S.-I. Park, P. Angueira, J. Montalban, and M. Velez, "Application of DFT-based channel estimation for accurate signal cancellation in cloud-txn multi-layer broadcasting system," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.*, Beijing, China, Jun. 2014, pp. 1–6.
[40] A. Mahmood, R. Exel, H. Trsek, and T. Sauter, "Clock synchronization over IEEE 802.11—A survey of methodologies and protocols," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 907–922, Apr. 2017.
[41] O. Seijo, J. A. Lopez-Fernandez, H.-P. Bernhard, and I. Val, "Enhanced timestamping method for subnanosecond time synchronization in IEEE 802.11 over WLAN standard conditions," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5792–5805, Sep. 2020.
[42] A. Varga and R. Hornig, "An overview of the OMNeT++ simulation environment," in *Proc. 1st Int. ICST Conf. Simulation Tools Techn. Commun. Netw. Syst.*, 2008, p. 60.
[43] 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 *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Porto, Portugal, Jun. 2018, pp. 1–6.

JON MONTALBAN (Member, IEEE) received the M.S. and Ph.D. degrees in telecommunications engineering from the University of the Basque Country, Spain, in 2009 and 2014, respectively. He is part of the TSR (Radiocommunications and Signal Processing) research group at the University of the Basque Country, where he is currently an Assistant Professor, involved in several research projects. He has held visiting research appointments with Communication Research Centre (CRC), Canada, and Dublin City University (DCU), Ireland. His current research interests include the area of wireless communications and signal processing for reliable industrial communications. He was a co-recipient of several best paper awards, including the Scott Helt Memorial Award to recognize the best paper published in the IEEE TRANSACTIONS ON BROADCASTING, in 2019. He has served as a reviewer for several renowned international journals and conferences in the area of wireless communications, and he currently serves as an Associate Editor for IEEE ACCESS.

ENEKO IRADIER (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications engineering from the University of the Basque Country (UPV/EHU), in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree. Since 2015, he has been a part of the TSR Research Group, UPV/EHU. He was a Researcher with the Communications Systems Group, IK4-IKERLAN, from 2017 to 2018. During his Ph.D. studies, he did an internship with Communications Research Centre Canada, Ottawa. His current research interests include the design and development of new technologies for the physical layer of communication systems and wireless solutions for Industry 4.0. He has served as a reviewer for several renowned international journals and conferences in the area of wireless communications.



PABLO ANGUEIRA (Senior Member, IEEE) received the M.S. and Ph.D. degrees in telecommunication engineering from the University of the Basque Country, Spain, in 1997 and 2002, respectively.

He joined the Communications Engineering Department, University of the Basque Country, in 1998, where he is currently a Full Professor. He is part of the Staff of the Signal Processing and Radiocommunication Lab, where he has been involved in research on digital broadcasting (DVB-T, DRM, T-DAB, DVB-T2, DVB-NGH, and ATSC 3.0) for more than 20 years. He is coauthor of an extensive list of papers in international peer-reviewed journals, and a large number of conference presentations in digital broadcasting. He has also coauthored several contributions to the ITU-R working groups WP6 and WP3. His main research interests include the design and development of new technologies for the physical layer of communication systems in industrial wireless environments.

Dr. Angueira is a member of the IEEE BMSB International Steering Committee, an Associate Editor of the IEEE TRANSACTIONS ON BROADCASTING, and a Distinguished Lecturer of the IEEE BTS. He serves on the Administrative Committee of the IEEE BTS.

OSCAR SEIJO (Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications engineering from the University of Oviedo, Spain, in 2015 and 2017, respectively. He is currently pursuing the Ph.D. degree with the Communications Systems Group, IKERLAN Technological Research Centre, Mondragón, Spain, in collaboration with the Signal Theory and Communications (TSC) Group, University of Oviedo, Spain. His research interests include wireless high-performance PHY and MAC design for industrial applications, time synchronization over wireless systems, and digital signal processing.

IÑAKI VAL (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the Department of Electronics Engineering, University of Mondragon, Spain, in 1998 and 2001, respectively, and the Ph.D. degree from the Department of Signals, Systems, and Radiocommunication, Polytechnic University of Madrid, Spain, in 2011. Since 2001, he has been with the Communications Systems Group, IKERLAN, Mondragón, Spain. He has been with Fraunhofer IIS of Erlangen, Germany, as an Invited Researcher, from 2005 to 2006. He is currently a Team Leader of the Communication Systems Group. His research interests include the design and implementation of digital wireless communications systems, industrial real-time requirements, communications for distributed control systems, vehicular communications, time synchronization, wireless channel characterization, and digital signal processing. He is currently focused on industrial communication applications.

3.2.3 Conference paper C4

This subsection presents a conference paper related with Contribution 2. The full reference of the paper is presented below:

- E. Iradier, J. Montalban, L. Fanari and P. Angueira, "On the Use of Spatial Diversity under Highly Challenging Channels for Ultra Reliable Communications," 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Zaragoza, Spain, 2019, pp. 200-207, doi: 10.1109/ETFA.2019.8869055.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Indexed Congress in IEEExplore
- Area: Computer Science and Engineering
- SJR factor: 0.365

On the Use of Spatial Diversity under Highly Challenging Channels for Ultra Reliable Communications

E. Iradier, J. Montalban, L. Fanari, P. Angueira
Department of Communication Engineering
University of the Basque Country (UPV/EHU)
Plaza Torres Quevedo 1, Bilbao, Spain
{eneko.iradier, jon.montalban, lorenzo.fanari, pablo.angueira}@ehu.eus

***Abstract*— The Industry 4.0 concept is the result of latest technical advances that have completely changed the existing industrial vision, not only from the mechanical point of view, but also from the Machine-to-Machine (M2M) communication perspective. This revolution encompasses wireless communication technologies that are expected to be the replacement for traditional wired and fixed communication systems. However, industrial propagation channels are still a challenge for wireless industrial network deployments. This paper addresses this problem by proposing spatial diversity based retransmissions. Two general retransmission schemes have been tested: default and smart. Moreover, specific time scheduling choices have been proposed. The impact of the number of transmitters in relation to communication reliability has been analyzed. Results show that the use of spatial diversity based retransmissions makes restrictive reliability values feasible under realistic conditions. In fact, in the case of smart retransmissions, adding a second transmitter the reliability increases around 8 dB. Moreover, whereas the channel coherence time is a critical parameter for time based retransmissions, in the case of spatial diversity, the reliability performance has no dependency on the channel coherence time.**

***Keywords*— 802.11, communications, factory automation, industry 4.0, PER, reliability, retransmissions, spatial diversity**

I. INTRODUCTION

Industrial communications are radically changing the landscape of traditional industry. The introduction of networking and smart devices has led to the fourth industrial revolution, or Industry 4.0 [1]. One of the main challenges is the deployment of wireless network systems able to replace the traditional wired networks. In fact, wireless systems must guarantee the same performance as wired systems within the strictest industrial environments. In [2], the main requirements for Factory Automation (FA) use cases are listed. Certain reliability and latency values can be taken as goal values: 10^{-9} and 0.25-10 ms, respectively. Up to now, wired communication systems provide much better performance indicators. Nevertheless, applications that require mobility and complex deployments would benefit from wireless standard deployments.

A major challenge of current wireless communication systems is the simultaneous improvement of both, reliability and latency.

One of the most extended communication technologies for industrial wireless applications is the 802.11 standard family. Moreover, in 2009, the IEEE standardization committees released the standard IEEE 802.11n version, as an improvement of the existing 802.11g version. Authors in [3] tested for the first time the 802.11n standard for real-time industrial communications and provided a set of recommendations for reliability and latency enhancement. Furthermore, in the very recent literature, two interesting proposals should be highlighted: one providing guidelines for ultra-high-performance wireless networks and the other one proposing modifications to the 802.11g PHY/MAC (Physical Level/Medium Access Control) standard for industrial automation.

First, authors in [4] presented an overview of the different relevant characteristics that have to be taken into account in ultra-high-performance wireless networks for critical industrial control applications.

Second, in [5], SHARP (Synchronous and Hybrid Architecture for Real-time Performance) was presented, which is a novel proposal for wireless industrial automation networks. It included a modification of the 802.11g standard PHY/MAC levels and the determinism is introduced by using a TDMA access control scheme.

Industrial wireless channels present deep fading occurrences. SHARP and similar approaches compensate deep fades with high operation margins, and consequently, high Received Signal Strength Indicator (RSSI) values. The performance of these solutions in lower RSSI values is still a challenge.

Consequently, the difficulties in the physical propagation channel maintain wireless communications at a disadvantage in comparison with other wired communication systems. The reduction of the channel impact on system performance is necessary for real industrial deployments.

Current wireless communication standards, like the 802.11 family [6], are based on a cross-layer design. Typically, they standardized the PHY and the MAC levels. In this way, robust PHY layers can be designed, which allow the reception of frames with low Signal-to-Noise Ratio (SNR), and the overall performance can be improved by adding techniques for increasing the reliability in the MAC layer.

978-1-7281-0303-7/19/$31.00 ©2019 IEEE

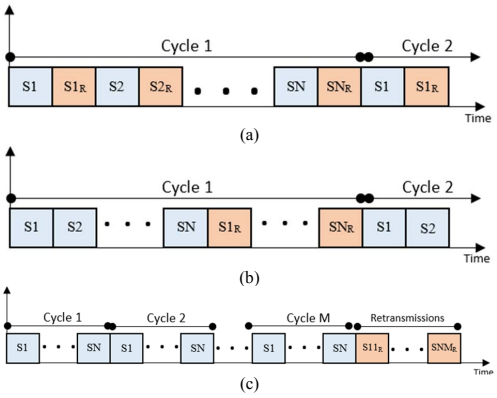


Fig 1. Retransmission schemes: (a) intra-slot, (b) intra-superframe, (c) superframe aggrupation

In this line, a widely extended technique to reduce the amount of lost packets is the use of retransmissions. However, this technique is not always efficient in communications that need to guarantee low cycle times. If the reception of a packet fails due to eventual fading, it is highly probable that the retransmission slot fails again, because the link has not had enough time to recover.

In consequence, the main objective of this paper is to present spatial diversity as a mechanism for increasing the reliability and enhancing the communication success of 802.11 industry deployments. Different diversity schemes are presented here in order to avoid the wireless channel fading effect. For each presented scheme, the performance is evaluated in terms of reliability. Some guidelines obtained from the simulations are presented in order to facilitate the implementation of diversity based schemes and to guarantee its success.

The paper is organized as follows. The next section describes some theoretical work that have been done involving different mechanisms for increasing reliability. Section III provides the definition of the evaluating s. Section IV presents the simulation methodology that has been followed in this work. Then, Section V discusses the results obtained by applying a capacity constraint and Section VI shows the upper reliability bound results when none capacity constraint is applied. Section VII contains the future work and, finally, conclusions are highlighted in Section VIII.

II. MECHANISMS FOR INCREASING RELIABILITY

One of the modules in charge of measuring and guaranteeing robustness for a specific technology is the PHY layer. However, it should not attract all the attention; other techniques have to be introduced in order to obtain a meaningful performance. Many of these techniques are located either in the MAC level or managed at MAC and implemented in the PHY.

Retransmissions are a classical solution implemented in a wide variety of communication systems [7]. Their main drawback is efficiency, which generally is conditioned by the propagation channel characteristics. The channel coherence time is specifically a performance downgrading factor. Depending on the system requirements and the service characteristics, retransmissions can be scheduled in the time domain following diverse criteria. If the design is oriented to control systems, where low latency is needed, retransmissions will be carried out in the same time slot (intra-slot), as it is depicted in Fig 1.a, where retransmissions are represented by the letter R inside orange blocks. As it is designed in [5], retransmissions can also be delivered intra-superframe. In this case, different periods for transmissions and retransmissions are defined inside the superframe (see Fig 1.b). Finally, and especially for non-critical applications, where latency is not a restriction, retransmissions could be considered in groups of superframes (see Fig 1.c). Hence, it should be highlighted that before implementing retransmissions a tradeoff has to be assumed between reliability and latency.

In conclusion, alternative techniques have to be taken into account, in order to reduce the channel condition dependence. A simple but effective technique is spatial diversity. This technique consists of using more than one transmitting point to create several link options between the end device and the network, reducing the probability of a packet loss. Authors in [8] introduce spatial diversity in industrial communications and discuss potential applications. The main conclusion of this work is that it identifies the spatial diversity as a technical solution for next generation wireless industrial systems. However, authors do not provide any measurement to quantify the potential gain that could be achieved by using spatial diversity in comparison with time domain retransmissions. In addition, the use of relay nodes is developed in [9], another form of spatial diversity. For a given source, destination and wireless channel, authors evaluate the position of the relay node together with the most suitable retransmission strategy: retransmissions on demand or after error packet detection.

Wi-Red, Wi-Fi Redundancy, is another interesting approach for industrial reliable wireless communications [10]. Wi-Red aims at providing seamless link-level redundancy in IEEE 802.11 networks. Two general mechanisms are proposed and evaluated: Reactive Duplicate Avoidance (RDA) and Proactive Duplicate Avoidance (PDA). The results demonstrate that both, RDA and PDA, reduce considerably the latency in industrial networks in comparison with other existing solutions. In a more recent work [11], authors present a Wi-Red prototype based on 802.11 commercial modules. The system overcomes the performance of existing mechanisms, such as Parallel Redundancy Protocol (PRP). The same authors have also studied the problems of mutual interference between channels in spatial diversity schemes [12].

Finally, there is a group of proposals based on various Cooperative MAC solutions [13]. The main idea of Cooperative MACs is the help of neighboring nodes in forwarding data packets to specific destination nodes when necessary. An interesting example is shown in [14], where a cooperative MAC

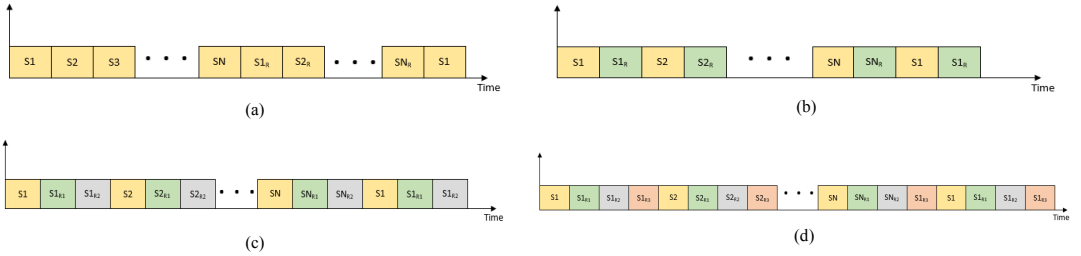


Fig 2. Proposed time and spatial diversity based retransmission schemes: (a) Time division retransmissions, (b) Retransmissions with two APs, (c) Retransmissions with three APs, (d) Retransmissions with four APs

protocol with rapid relay selection (RRS-CMAC) is presented. RRS-MAC selects an optimal relay whenever the data rate between the source and the destination is low or whenever the link performance shows a degraded behavior. The main benefit of this protocol is to improve the cooperation efficiency and multiple access performance in wireless ad hoc networks. In addition to reliability, cooperative MAC solutions have also been proposed for addressing other challenges, such as energy consumption in sensor networks. An example can be found in [15], where authors propose a distributed energy adaptive location-based CMAC protocol, namely DEL-CMAC, for Mobile Ad-hoc Networks (MANET) with optimized energy consumption.

In conclusion, although enhancements in reliability have been addressed by several authors, there are still challenges for applying wireless communication standards in industrial propagation channels. This is why, this paper analyzes the feasibility and performance of spatial diversity techniques under highly challenging channel conditions.

III. Proposed Diversity Schemes

A. *Time division retransmissions*

The first mechanism that has been taken into account is time redundancy: retransmissions on a time division based scheme. As described in Section II, there are several options to handle retransmissions. In this work, intra-superframe retransmissions have been used, since they are considered more balanced than the other schemes in terms of maximum manageable latency. Therefore, the transmitter node sends the information required for each receiver within a control cycle. Then each communication is retransmitted redundantly (see Fig 2.a).

On the one hand, it is important to highlight the low complexity level of this scheme, without any extra cost nor synchronization penalties.

On the other hand, just one transmitter is used for both, transmissions and retransmissions, which means that the same link path will be applied to transmissions and retransmissions. In consequence, this scheme is vulnerable to long fading effects. Additionally, there is not any intelligence or feedback (no uplink is implemented), and so, retransmissions are redundant no matter the standard packet reception success. From the capacity point of view, this scheme divides the overall capacity offered by the transmitter by two. So, higher MCS (Modulation and Coding Scheme) would be required to maintain the original throughput.

TABLE I. Simulated Configurations

Retransmissions	# of Transmitters	Redundancy Level (R)	MCS
No	1	0	BPSK 1/2
Time based	1	1	QPSK 1/2
Spatial diversity	2	1	QPSK 1/2
Spatial diversity	3	2	QPSK 3/4
Spatial diversity	4	3	16QAM 1/2

B. *Spatial diversity based retransmissions*

This second approach can be considered a complementary mechanism for the previous proposal, especially for highly challenging propagation channels. In this case, spatial diversity is applied on the transmitter side. Several transmitters will be introduced to increase path diversity. Ideally, N number of transmitters could be introduced. In this work up to four transmitters have been considered.

The main benefit of this technique lies in the uncorrelation of paths between each transmitter and the receiver, which will minimize simultaneous fading occurrences. Moreover, the impact of individual fading events on one link will be less critical. Furthermore, the higher the number of transmitters, the better performance is expected.

The maximum latency that a receiver will assume in a correct communication is very close to N number of packet air times. In comparison with the time division retransmissions solution, an interesting latency decrease is obtained.

However, the main drawback of this mechanism is that retransmissions are again redundant no matter the packet delivery success. So, the overall network capacity is proportionally reduced as a function of the number of transmitters. Therefore, a tradeoff has to be assume between the offered capacity and the desired reliability.

IV. Simulation Methodology

In order to validate each of the proposed schemes, a dual simulator has been developed, consisting of a network simulator and a mathematical simulator. The first is used to

implement a complete communication network (OMNET ++), while the second is used to evaluate the system robustness under specific channel conditions (Matlab).

A transmitter-receiver prototype chain has been implemented in Matlab, which is fully compliant with the 802.11 a/g/n standard. Based on this prototype, PER vs. SNR curves are obtained for each MCS. This process takes into account all the effects of the channel and the characteristics of both transmission and reception. Therefore, the Matlab output for the 802.11 simulator could be considered a key performance indicator. The channel model is the so called "industrial channel" with Non-Line-Of-Sight (NLOS) condition and 29 ns rms delay spread defined in [16].

The network simulation tool (OMNeT++) is composed of both PHY and MAC layers. It is worth mentioning that in this prototype, the exponential backoff has been disabled in order to send transmission and retransmission frames inside controlled time slots. The PHY module facilitates the 802.11 MCS choice, as well as the length of each transmission slot and the airtime of each data packet. In addition, the MAC layer includes the implementation of the proposed access time schemes (see Fig. 2) and the access control to the transmission medium. Finally, the 802.11 performance curves obtained from Matlab as well as the channel impulse responses are loaded in OMNeT++ for full system simulation.

The simulations cover a range of mean received SNR values from 0 dB to 24 dB. For each simulation, a realization of the impulse channel response is assumed. Both effects, the channel impulse response and the mean SNR value, are combined to obtain the instantaneous SNR value. This instantaneous SNR value is used to evaluate packet reception success (one SNR value per packet). Finally, PER values are calculated independently in each receiver, and then, the mean PER value is calculated among all the receivers within the network. In fact, this is the performance evaluation metric that is going to be used to evaluate the reliability of each proposed and simulated scheme.

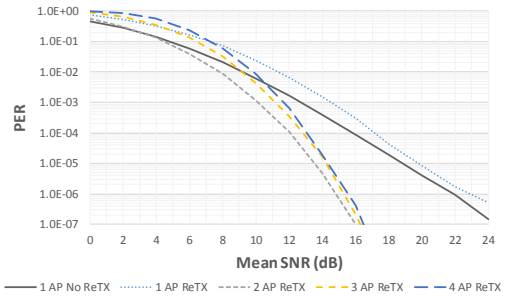
V. Performance Evaluation

The performance results of the schemes proposed in Section III are presented here.

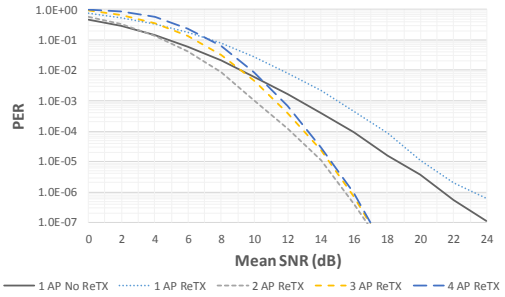
A. Simulated configurations

Each simulation represents a sensor network spatial deployment within a manufacturing plant. The network is composed by two device classes: Access Points (APs) and Slaves. On the one hand, Slaves represent typical industrial sensors or actuators. On the other hand, the AP is the master node of both the architecture and the synchronization. The AP distributes information to all the devices connected to the network. The number of Slaves is fixed to 20 in each simulation, whereas the number of APs changes depending on the case.

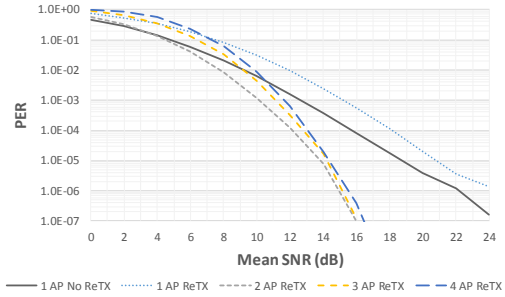
For each mean SNR value, $2 \cdot 10^6$ superframes have been simulated. Furthermore, taking into account that the network is composed by 20 devices and that the PER value is obtained by averaging the results in each slave, for each configuration $4 \cdot 10^7$ superframes have been evaluated. This methodology leads to a minimum PER value close to 10^{-8}. Therefore, in order to show reliable results, each graphic will show PER values up to 10^{-7}.



(a)



(b)



(c)

Fig 3. Performance PER vs. Mean SNR curves evaluation with redundant retransmissions for: (a) 30ms coherence time, (b) 50ms coherence time, (c) 100ms coherence time

Different diversity and retransmission schemes have been analyzed. First, a communication scheme without retransmissions is simulated. Undoubtedly, this is the simplest scenario and it will be the reference point for comparison purposes. After that, a time division retransmission scheme is implemented (see Fig 2.a). In this scheme, retransmissions are carried steadily and after having served every slave. Finally,

TABLE II. SNR THRESHOLDS FOR 10^{-5} PER

Retransmissions	# of Transmitters	T_{coh} = 30ms	T_{coh} = 50ms	T_{coh} = 100ms
No	1	18.8 dB	18.7 dB	18.8 dB
Time based	1	19.8 dB	20.1 dB	20.9 dB
Spatial diversity	2	13.5 dB	13.9 dB	13.8 dB
Spatial diversity	3	14.2 dB	14.5 dB	14.2 dB
Spatial diversity	4	14.3 dB	14.6 dB	14.3 dB

spatial diversity based retransmissions are simulated based on Fig 2.b, Fig 2.c and Fig 2.d, where two, three and four APs are needed. Retransmissions are redundant information for the slaves no matter the communication success, and they are carried out by the extra APs in order to obtain independent paths. This case assumes that the distance between APs is large enough to provide uncorrelated propagation AP-Slave paths.

Redundancy provokes an overall capacity decrease. For example, in the time division retransmission scheme and in the spatial diversity based scheme with two APs, the overall capacity is divided by two. Generalizing this approach, it can be said that:

$$Useful\ capacity = \frac{Tx\ capacity}{R+1} \quad (1)$$

where R is the redundancy level (see Table I) and Tx capacity is the capacity obtained by implementing the 802.11 standard. If the same throughput has to be offered to the network devices, transmitters have to transmit using higher MCS values, so, useful capacity decrease needs to be compensated. In this work, a constraint has been assumed before defining the offered services: the transmitters have to offer a minimum useful capacity of 6 Mbps. The effect of the constraint is shown in Table I in the MCS column. As shown, the higher the redundancy level, the higher is the required MCS value and the lower the modulation and coding robustness. Anyway, the reliability gain expected from spatial diversity based retransmissions should be high enough to compensate this issue.

Finally, the channel coherence time, a specific feature of industrial channels has been also considered. This parameter shapes the channel impulse response. In fact, the lower the coherence time, the faster the change of the channel impulse response. The coherence time is expected to have a significant impact on the time domain retransmission schemes, whereas, its effect should be almost negligible on spatial diversity based schemes. These assumptions have been evaluated with tests in three channel conditions, associated to channel coherence time values of 30, 50 and 100ms.

B. Results

All the results obtained from simulations are gathered in Fig 3. Three curves are presented, one for each channel coherence time. Each curve shows the reliability performance of each mechanism by plotting the obtained PER for a specific mean SNR value.

Regarding the effect of the channel coherence time, it has a remarkable effect over time division retransmission schemes (light dotted blue line). Comparing it with the no retransmission scheme (dark solid grey line), the higher is the coherence time; the worse is the performance of the time division based retransmissions scheme. In each case, for 30ms, 50ms and 100ms of channel coherence time (Fig 3.a, Fig 3.b and Fig 3.c, respectively), time division based retransmissions are unable to offer better performance than the scheme with no retransmissions. This makes time based solution no longer feasible. This effect appears because of the constraint that states that all services have to guarantee at least 6 Mbps. On the other hand, spatial diversity solutions seem to be independent from the effect of the channel coherence time, since they offer in the three cases very similar results.

The difference between the performance tendency for time and spatial diversity technologies is also remarkable. The first group has a very flat PER decrease, which means that very high SNR values are required to guarantee reliable PER values. In other words, the performance of time division based schemes does not converge fastly. However, spatial diversity based schemes present faster reliability decrease while increasing the SNR. The overall tendency is more similar to waterfall performance curves, with fast convergence.

Analyzing in detail spatial diversity based mechanisms, a point, which should particularly be noted, is that using two AP (light dashed grey line) offers always the best performance for low SNR values. Nevertheless, three AP schemes (yellow dashed line) and four AP schemes (dark dashed blue line) curves converge slightly faster than two AP case. This effect is denoted in very low PER values where the difference among mechanisms is smaller than in higher error rates.

Another interesting point is the similar performance of spatial diversity based schemes. Despite the fact that increasing the number of transmitters involves a relevant change on network architecture, the spatial diversity based results are very similar. The increase in reliability derived from using higher number of independent paths is compensated with the increase needed in the MCS value in order to maintain the minimum throughput.

Another perspective of the reliability is shown in Table II. In this table, the SNR thresholds for a PER value of 10^{-5} are shown for each simulated case. Generally, a remarkable difference is found between time and spatial diversity schemes. Time based mechanisms require SNRs around 20 dB, with a maximum value of 20.9 dB. However, required SNR for spatial diversity schemes is always below 15 dB. All in all, SNR values around 14 dB are still a challenge in many harsh industrial environments.

As regards time requirements, each of the proposed and simulated solutions leads to a determined maximum latency

value. As packets are not retransmitted in other cycle times, latency can be predicted in this way:

$$\max latency = (R + 1) * packet\ airtime \qquad (2)$$

where each packet airtime varies depending on the used MCS. Hence, the lowest maximum latency values are obtained by using the two AP schemes. As it is obvious, propagation time has not been included because it is almost negligible in comparison with packet airtime values.

On the whole, taking into account each of the presented discussion topics, the mechanism that includes two APs with spatial diversity based retransmissions is the most suitable option, since it offers the lowest error rate, a fast convergence and the lowest maximum latency with minimum complexity.

VI. Upper Reliability Bound

Although the proposed schemes offer remarkable gains, they are a suboptimal solution, since reliability gain is partially sacrificed for maintaining throughput. Doubtlessly, this drawback is still compensable by avoiding the use of default retransmissions and replacing them with smart and/or efficient retransmission schemes.

Some reference works already available in the literature include smart retransmission schemes. The proposal referred as SHARP [5], proposes an ACK based retransmissions system to avoid the retransmission of each packet if that is not required. Moreover, in [10] several approaches are presented based on Duplicate Avoidance (DA) mechanisms, again to avoid redundancy in cases where is not required.

Furthermore, the success key for a smart retransmission scheme is to obtain independent physical channels on each path. In that way, the fading effect that transmissions and retransmissions suffer are uncorrelated. Consequently, theoretically, the relation between the PER value and the number of paths can be defined as:

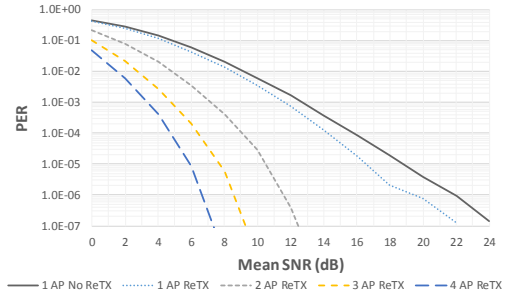
$$PER_{N\ AP} = (PER_{1\ AP})^N \qquad (3)$$

where N is the number of APs in the network, and so, the number of independent paths.

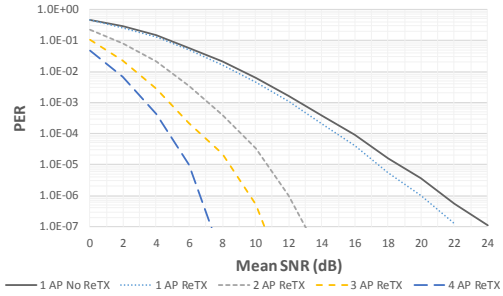
The implementation of a DA mechanism that enables transmission and re-transmissions without any capacity constraint is analyzed in this section. Specifically, based on [10], we are assuming the use of PDA mechanisms that aim at preventing unnecessary transmissions.

Thus, no matter how many APs are used for transmitting, the most robust MCS value has been used in each simulation (MCS 0, BPSK 1/2). Regarding the simulated mechanisms and the channel coherence time, the same combinations as in Table I and Fig 3 have been used. All the results obtained from simulations are gathered in Fig 4.

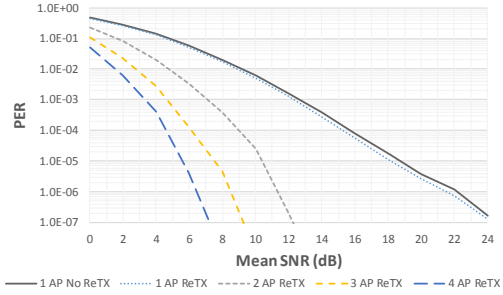
On the one hand, time division based retransmission schemes are more vulnerable to the influence of the propagation channel coherence time. In fact, these curves do not converge except for high SNR values. Moreover, as it was expected, the gain obtained in reliability by applying time division based retransmissions is higher in the 30ms coherence time case (see Fig 4.a), and it is almost negligible in the 100ms case (see Fig 4.c).



(a)



(b)



(c)

Fig 4. Performance PER vs. Mean SNR curves evaluation with smart retransmissions schemes for: (a) 30ms coherence time, (b) 50ms coherence time, (c) 100ms coherence time

On the other hand, spatial diversity techniques show independence among the number of transmitters and the channel coherence time. In fact, results for spatial diversity cases are almost constant in relation to the coherence time. Furthermore, these curves have a clear converging tendency.

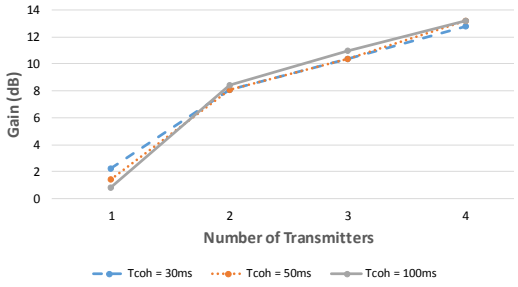


Fig 5. Spatial diversity gain with smart retransmission schemes

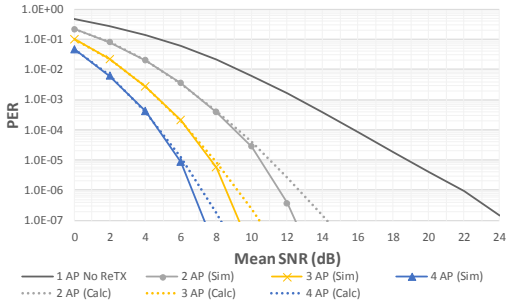


Fig 6. Simulated and calculated PER comparison for 30ms coherence time

The main difference between curves presented in Fig 3 and the ones shown in Fig 4 lies in the relevance of the number of transmitters. Fig 4 shows that the higher the number of transmitters (dark dashed blue line, 4AP), the more reliable the communication, whereas using two AP (light dashed grey line) is the case that provides less gain.

In order to have an in-depth point of view of the gain obtained by applying smart or efficient retransmission systems, Fig 5 is added. In this figure, the x-axis of the graphic represents the number of transmitters working in the network and the y-axis represents the gain by using the current retransmission scheme in relation to not retransmitting at all. The gain is calculated at a SNR threshold equivalent to a 10^{-5} PER value.

The relevance of the channel coherence time is demonstrated again. In the case of a single transmitter, the highest gain is obtained for 30ms (dashed blue line), which is around 2.2 dB, whereas the lowest gain is obtained for 100ms (solid grey line), 0.8 dB. This effect demonstrates the channel state dependence for the time division retransmission schemes. However, in the cases involving spatial diversity, channel dependence disappears and almost the same gain is obtained for each number of transmitters, no matter the channel coherence time.

Another interesting aspect in Fig 5 is the change in the overall gain if additional transmitters are included. In this case, the biggest difference appears when the second AP is introduced. In that situation, up to 7.6 dB are obtained in the 100ms case. If more APs are added to the network, the gain continues increasing but not as fast as in the first case. Moving from two AP to three, around 2.3-2.6 dB are obtained, and changing from three to four, the gain is below 2.5 dB. This behavior means that the gain obtained by adding transmitters to the network and to the retransmission schemes is logarithmic. That is why, the curve grows fast at the beginning, but it slows down at the same time that the number of transmitters is increased. In the case this experiment could be done for any transmitter number, it could be demonstrated that the curve will saturate at a certain value. Generally, when a smart retransmission scheme is implemented, a tradeoff must be assumed between the required reliability, and the complexity and the costs derived from enlarging the network.

Moreover, based on Eq. (3), the theoretical PER curves for two, three and four independent paths have been calculated taking as a reference the simulations that do not include retransmissions. Results for a channel coherence time of 30 ms are presented in Fig 6. Calculated PER curves are represented with the same color as its simulated pair, whereas they are characterized with dotted lines. Several comments can be done based on Fig 6. Firstly, it is remarkable the high similarity rate between calculated and simulated curves at least up to PER 10^{-5}. This means that the simulated conditions are similar to the ideal case where physical channels are uncorrelated. Therefore, the proposed situations are very close to the existing upper reliability bound, where no improvement can be achieved because the optimum configuration is already in place. It is also important to highlight the different behavior that curves have for PER values below 10^{-5}. This mismatch is derived from the tendency of the original curve, the scheme that does not include retransmissions (dark solid grey). This curve does not converge, even if high SNR values are assumed and that effect is indirectly included as well in the calculated curves.

VII. Future work

This paper has analyzed comprehensively the use of spatial diversity techniques with the PHY level of the 802.11n standard, whereas there are still several open issues.

Based on this study, the first proposal will be to enhance diversity. A first approach could be to introduce more than four APs, test how the reliability increases and obtain a generic expression that estimates the gain associated to each hypothetical situation. Moreover, still with low number of APs, some intelligence could be applied to diversity algorithms. For example, the time for switching between APs could be directly related with the deep fading events appearing frequency of the channel, instead of using one AP just to retransmit.

A second area is the combination of these approaches with more sophisticated cooperative MAC solutions. In this case, new relay paths will be presented by using neighbors, other networks or even different technologies. While implementing this wider concept, different parameters can be taken into account, such as energy consumption.

Finally, frequency hopping could be also introduced as a complementary technique. Even if interferences and noise will not be fully eliminated, better performance is expected, due to the change of phase and amplitude of the reflected paths.

VIII. Conclusions

In this article, a comprehensive proposal of several techniques to increase communication reliability has been presented. The focus has been put on spatial diversity based schemes, which aim to fight against challenging channel time coherence conditions. Those techniques have been integrated in the 802.11n standard and the performance results have been presented. Results have been organized in two main groups: plain redundant retransmission schemes and smart retransmission schemes.

In the first approach, a relevant gain is easily obtained by retransmitting from different transmitters in order to create independent gains. However, transmitting redundant information affects directly the overall offered capacities. If the same throughput has to be offered to the network devices, transmitters have to transmit using higher MCS values. Consequently, the gain obtained by introducing more than two independent paths is almost cancelled with the MCS increase. This effect automatically presents the two AP spatial diversity based retransmissions case as the best alternative to guarantee low error rate, assuming the lowest complexity.

On the other hand, in the second approach, redundancy is avoided if not necessary by using smart retransmission schemes. In this case, the higher the number of transmitters, the lower the error rate. The increase has an almost logarithmic shape. Thus, the most beneficial situation is the introduction of the second transmitter in order to switch to spatial diversity based mechanisms.

In parallel, it has also been demonstrated that time channel coherence time does only affect time division retransmission schemes. The introduction of independent paths cancels the effect of the coherence time.

Acknowledgements

This work has been partially supported by the Spanish Government (project PHANTOM under the grant RTI2018-099162-B-I00 (MCIU/AEI/FEDER, UE)) and by the Basque Government under the PREDOC grant program (PRE_2018_1_0344), the project IOTERRAZ (under the grant KK-2019/00046, ELKARTEK 2019) and the Grant IT1234-19.

References

[1] R. Drath and A. Horch, "Industrie 4.0: Hit or Hype? [Industry Forum]," in IEEE Industrial Electronics Magazine, vol. 8, no. 2, pp. 56-58, June 2014.

[2] P. Schulz et al., "Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture," in IEEE Communications Magazine, vol. 55, no. 2, pp. 70-78, February 2017.

[3] F. Tramarin, S. Vitturi, M. Luvisotto and A. Zanella, "On the Use of IEEE 802.11n for Industrial Communications," in IEEE Transactions on Industrial Informatics, vol. 12, no. 5, pp. 1877-1886, Oct. 2016.

[4] M. Luvisotto, Z. Pang and D. Dzung, "Ultra High Performance Wireless Control for Critical Applications: Challenges and Directions," in IEEE Transactions on Industrial Informatics, vol. 13, no. 3, pp. 1448-1459, June 2017.

[5] Ó. Seijo, Z. Fernández, I. Val and J. A. López-Fernández, "SHARP: A novel hybrid architecture for industrial wireless sensor and actuator networks," 2018 14th IEEE International Workshop on Factory Communication Systems (WFCS), Imperia, 2018, pp. 1-10.

[6] IEEE, «Orthogonal frequency division multiplexing (OFDM) PHY specification,» de 802.11 Standard - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, New York, United States of America, IEEE Computer Society, 2012, pp. 1583-1630.

[7] M. Jonsson and K. Kunert, "Towards Reliable Wireless Industrial Communication With Real-Time Guarantees," in IEEE Transactions on Industrial Informatics, vol. 5, no. 4, pp. 429-442, Nov. 2009.

[8] A. Willig, "How to exploit spatial diversity in wireless industrial networks," Annu. Rev. Control, vol. 32, no. 1, pp. 49–57, Apr. 2008.

[9] S. Girs, E. Uhlemann and M. Björkman, "The effects of relay behavior and position in wireless industrial networks," 2012 9th IEEE International Workshop on Factory Communication Systems, Lemgo, 2012, pp. 183-190.

[10] G. Cena, S. Scanzio and A. Valenzano, "Seamless Link-Level Redundancy to Improve Reliability of Industrial Wi-Fi Networks," in IEEE Transactions on Industrial Informatics, vol. 12, no. 2, pp. 608-620, April 2016.

[11] G. Cena, S. Scanzio and A. Valenzano, "A Prototype Implementation of Wi-Fi Seamless Redundancy with Reactive Duplication Avoidance," 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), Turin, 2018, pp. 179-186.

[12] G. Cena, S. Scanzio and A. Valenzano, "Improving Effectiveness of Seamless Redundancy in Real Industrial Wi-Fi Networks," in IEEE Transactions on Industrial Informatics, vol. 14, no. 5, pp. 2095-2107, May 2018.

[13] D. O. Akande, M. F. M. Salleh, and F. K. Ojo, "MAC protocol for cooperative networks, design challenges, and implementations: a survey," in Telecommunication Systems, 1-17, 2018.

[14] K. Liu, X. Chang, F. Liu, X. Wang, and A. V. Vasilakos, "A cooperative MAC protocol with rapid relay selection for wireless ad hoc networks" in Computer Networks, vol. 91, 262-282, 2015.

[15] X. Wang and J. Li, "Improving the Network Lifetime of MANETs through Cooperative MAC Protocol Design," in IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 4, pp. 1010-1020, April 2015.

[16] A. F. Molisch et al., "IEEE 802.15. 4a channel model-final report," IEEE P802, vol. 15, no. 04, pp. 1–41, 2004.

3.2.4 Journal paper J5

This subsection presents a journal paper related with Contribution 2. The full reference of the paper is presented below:

- E. Iradier, L. Fanari, I. Bilbao, J. Montalban, P. Angueira, O. Seijo, I. Val, "Analysis of NOMA-Based Retransmission Schemes for Factory Automation Applications," in IEEE Access, vol. 9, pp. 29541-29554, 2021, doi: 10.1109/ACCESS.2021.3059069.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Journal paper indexed in JCR and IEEExplore
- Area: Electrical & Electronic Engineering
- Ranking: 61/266 (Q1) based on JCR 2019
- Impact factor (JCR): 3.745

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 2, 2021, accepted February 10, 2021, date of publication February 12, 2021, date of current version February 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3059069

Analysis of NOMA-Based Retransmission Schemes for Factory Automation Applications

ENEKO IRADIER[1], (Graduate Student Member, IEEE), LORENZO FANARI[1], (Graduate Student Member, IEEE), IÑIGO BILBAO[1], (Graduate Student Member, IEEE), JON MONTALBAN[1], (Senior Member, IEEE), PABLO ANGUEIRA[1], (Senior Member, IEEE), OSCAR SEIJO[2], (Member, IEEE), AND IÑAKI VAL[2], (Senior Member, IEEE)

[1]Department of Communication Engineering, University of the Basque Country (UPV/EHU), 48013 Bilbao, Spain
[2]IKERLAN Technology Research Centre, Basque Research and Technology Alliance (BRTA), 20500 Mondragón, Spain

Corresponding author: Eneko Iradier (eneko.iradier@ehu.eus)

This work was supported in part by the Basque Government under Grant IT1234-19, in part by the PREDOC under Grant PRE_2020_2_0105, and in part by the Spanish Government through project PHANTOM (MCIU/AEI/FEDER, UE) under Grant RTI2018-099162-B-I00.

ABSTRACT New use cases and applications in factory automation scenarios impose demanding requirements for traditional industrial communications. In particular, latency and reliability are considered as some of the most representative Key Performance Indicators (KPI) that limit the technological choices addressing wireless communications. Indeed, there is a considerable research effort ongoing in the area of wireless systems, not only from academia, but also from companies, towards novel solutions that fit Industry 4.0 KPIs. A major limitation for traditional wireless architectures is related to the harsh nature of the industrial propagation channel. Accordingly, this paper addresses these challenges by studying the reliability and latency performance of the joint use of different retransmission schemes in combination with Non-Orthogonal Multiple Access (NOMA) techniques. Two general retransmission schemes have been tested: time-based and spatial diversity-based retransmissions. An adaptive injection level NOMA solution has been combined with the retransmission schemes to improve the reliability of critical information. In all cases, a particular set of simulations has been carried out varying the main parameters, such as modulation, code rate and the injection level. Moreover, the impact of the number of transmitters in relation to the communication reliability has been analyzed. Results show that spatial diversity-based retransmissions overcome considerably the reliability obtained with time-domain retransmissions while maintaining assumable latency rates.

INDEX TERMS 802.11, factory automation, industry 4.0, LDM, NOMA, P-NOMA, retransmissions, spatial diversity, wireless communications.

I. INTRODUCTION

Wireless communications are considered one of the most challenging and promising research areas of the so-called Industry 4.0 [1], [2]. In fact, the high costs and the lack of scalability and mobility that traditional automated processes composed by wired control systems have to assume make this solution less efficient than wireless systems. Moreover, the deployment of wireless systems provides the opportunity to enhance traditional systems by introducing new capabilities.

The associate editor coordinating the review of this manuscript and approving it for publication was Biju Issac.

However, the main challenge is the deployment scenario. Industrial environments entail strict requirements that are still hard to meet by wireless technologies, especially in the Factory Automation (FA) cases. For instance, very low latency values (0.25-10 *ms*) and ultra-high reliability (error rates below 10^{-9}) are required [3], [4]. Determinism is another mandatory condition that is difficult to achieve with current wireless standards. In recent literature, in order to enable a massive deployment of industrial wireless networks, several technologies and standards have been proposed: IEEE 802.11 [5], [6], Bluetooth [7], 802.15.4 [8], LTE [9], and 5G [10], [11]. However, the latest versions of Wi-Fi and 5G stand out among all of them for a considerable improvement in critical

applications. On the one hand, the next generation Wi-Fi standard (i.e., IEEE 802.11be) will offer, among other things, full-duplex multi-channel communications, which makes it easier to meet the reliability and latency requirements established by the industry [12]. On the other hand, 5G also presents conditions to meet industrial requirements due to its PHY/MAC layer with a wide range of possible configuration parameters. Among them, 5G allows to use small OFDM symbols (i.e., below 10 μs) with very robust modulation and code rates [13].

Since current standard wireless technologies do not cope with the demanding requirements of industrial applications and future standards such as 5G or IEEE 802.11be are still far from industrial deployment, some researchers have proposed proprietary solutions in order to get close to those barriers. For example, in [14], [15], Luvisotto *et al.* proposed the joint optimization of the PHY and MAC layers, improving the PHY reliability in several orders of magnitude and reducing the system latency by introducing a reliable MAC layer. In [5], [16], the authors follow the same joint PHY and MAC optimization approach in the solution referred to as SHARP (Synchronous and Hybrid Architecture for Real-time Performance). SHARP is based on the 802.11g standard and adds determinism to the PHY and MAC layers. In order to obtain competitive reliability and latency values, on the one hand, PHY frame aggregation techniques are implemented. On the other hand, a TDMA-based MAC structure with controlled time-division retransmissions is designed to guarantee a deterministic behavior.

A different strategy, based on improving the architecture of the access network is explored in [17], where several MAC level techniques in combination with the Wi-Fi PHY layer in challenging wireless channels are tested. Introducing multiple transmitters provided a reliability gain range of 8-12 dB. Specifically, the tests evaluated the use of two, three and four transmitters and they showed a non-linear behavior of the reliability increase while increasing the number of transmitters. Recently, in [18], the performance of spatial diversity with multiple redundant transmitters for mobile industrial communications is experimentally evaluated in a testbed with mobile robots. Authors implemented diversity using the Multipath TCP (MPTCP) protocol, which allows establishing different TCP connections between a transmitter and a receiver using multiple disjoint paths. The results indicate that redundant retransmissions improve reliability without compromising latency.

In [19], the use of Non-Orthogonal Multiple Access (NOMA) in combination with the IEEE 802.11n standard to meet the strict industrial requirements is considered. These results show that NOMA has the potential for enabling highly robust communications for industrial environments. However, the combination of NOMA and retransmission techniques has not been evaluated as a potential way to get closer to the performance requirements of current industry communications.

Therefore, in this paper, we propose and evaluate the joint use of NOMA and different diversity schemes to enhance the latency and reliability in industrial wireless networks. The diversity schemes are based on temporal and spatial diversity-based retransmissions using NOMA. Besides, some guidelines and recommendations are presented for real hardware implementation. In summary, the technical contributions of this paper include:

1) A comprehensive analysis of the PHY performance of a combined 802.11n and NOMA transceiver.
2) An analysis of the Injection Level (IL) capability to enhance the reliability.
3) A detailed analysis and performance evaluation of time-domain retransmissions.
4) Proposal and evaluation of combined multiple transmitter retransmissions with NOMA techniques.
5) Proposal and evaluation of a MAC layer that includes time, spatial and layer division.

The rest of the paper is organized as follows. The next section describes the related work, including the basis of NOMA techniques and state-of-the-art of retransmission and diversity schemes. Section III is focused on the presentation of the industrial use case and the simulation methodology. In Section IV, the PHY/MAC layers of the NOMA-based 802.11n transceiver are presented and evaluated. Then, in Section V, time-domain retransmissions are tested. Section VI is focused on the design and evaluation of MAC schemes that include multiple synchronized transmitters. Afterward, in Section VII a MAC layer that includes time, space and power multiplexing is designed and evaluated. Then, the latency and reliability results of the three different solutions are discussed and compared in Section VIII. Finally, Section IX contains the conclusions of the article.

II. RELATED WORK

This section contains, first, an overview of NOMA techniques and, then, a state-of-the-art of retransmission schemes including spatial diversity.

A. NOMA: GENERAL CONCEPTS AND INDUSTRIAL APPLICABILITY USING 802.11

NOMA represents a set of different medium access techniques where the receivers use the resource (either space, frequency, and time) in a non-orthogonal way. According to [20], NOMA techniques can be divided into two main families: code-domain NOMA (C-NOMA) [21], [22] and power-domain NOMA (P-NOMA) [23], [24]. In particular, although both NOMA families have been repeatedly proposed as alternatives to be included in 3GPP standards [25], P-NOMA techniques have shown a better complexity/performance tradeoff than C-NOMA [20]. During the rest of the paper, NOMA will refer to P-NOMA.

NOMA consists of different signals organized in several layers, where each layer takes some part of the transmitted total power. Each layer of the NOMA ensemble can be independently configured in order to address different reception

targets. Each NOMA configuration depends on the modulation and the coding choices assigned to each layer, and on the injection level (Δ, measured in dB), defined as the power splitting ratio among layers.

The main benefit of NOMA in comparison with classical TDM/FDM systems is the increase in spectral efficiency [26].

In [19], the first approach of a NOMA-based 802.11n system was proposed for FA environments. In this work, the communication architecture (PHY/MAC) of the 802.11n standard to include NOMA features was redesigned. Then, the proposed architecture was evaluated and the results showed a considerable better performance than the 802.11n standard PHY on several industrial use cases. In [27], a specific application of the NOMA-based solution was proposed for an industrial multimedia content broadcasting environment.

However, the reliability obtained in the previous works is still not enough for the most challenging industrial applications. Following the approach taken in previous works [14], advances in PHY reliability can be complemented with different MAC level techniques. Specifically, it is expected that MAC layer enhancements can improve the 10^{-4} Packet Error Rate (PER) provided by the physical layer down to 10^{-8}. That is why, in this paper, different MAC level techniques are proposed and implemented in combination with the NOMA-based 802.11n prototype.

B. RETRANSMISSIONS AND DIVERSITY SCHEMES

In communication systems, especially in industrial environments, the design of a robust PHY layer goes together with an efficient MAC layer. Among others, the main functionalities of the MAC layer are: the management of the transmission/reception instants based on the traffic requirements and on the wireless medium conditions, the configuration of the PHY layer, and the implementation of diversity mechanisms. One of the most widespread techniques is the use of retransmissions [28], with a variety of different configurations.

First, the authors in [17] evaluate time-domain retransmissions over 802.11n jointly with a TDMA scheme for different time-spacings between the first transmissions and the concurrent retransmissions. They determine that the effectiveness of these methods and the latency associated with them are closely related, since the more time that passes between transmission and retransmission, the greater the probability of success, but the greater the latency. In [29], the authors point out that wireless interferences can appear with different characteristics of intensity and duration. Therefore, they propose a MAC layer with temporal diversity that adapts the retransmissions to the characteristics of the interferences. The results show that their proposal can be applied for different types of information, meeting the requirements of real-time communications.

The retransmission of erroneous packets is also considered in cellular network technologies (i.e., LTE and 5G) and that is performed through Hybrid Automatic Repeat reQuest (HARQ) techniques. In particular, HARQs imply that instead of retransmitting the entire packet, only a portion of the packet or extra redundant information is retransmitted to demodulate the original packet. In [30], the authors propose the use of a new early HARQ scheme based on LDPC subcodes (SC E-HARQ). This technique decreases the overall latency since enables to start with the feedback calculation before the entire codeword is received. Results indicate that latency values below one millisecond can be obtained for Block Error Rate (BLER) values of less than 10^{-4}. Then, in [31], authors focus their study on delay-constrained scenarios and propose a fast HARQ protocol where in order to decrease the latency some feedback signals and successive messages are disabled. Different transmission rounds are tested and the latency results are reduced up to 60% in the case with five transmissions. Finally, in [32], HARQ techniques are combined with NOMA for the uplink of short packet communications. Simulations include one retransmission per user and coordinated and uncoordinated transmissions. Results show that NOMA-HARQ systems outperform orthogonal multiple access techniques for specific power and latency constraints.

Other authors have also investigated different retransmission schemes, such as [33], where a series of reliability tests using retransmissions are carried out in a laboratory environment. A combination of spatial and frequency diversity over four different RF channels is tested. The results indicate that under certain channel conditions, PER values close to 10^{-7} can be obtained. Later, in [34], different retransmission schemes based on the IEEE 802.15.4 standard are evaluated and compared. The authors conclude that PER values below 10^{-7} are possible with a maximum latency of 3 ms over a one-to-two transmission network with a 30 MHz frequency hop. Finally, in [35], the use of a backup wireless network is proposed to deal with errors. In particular, the authors propose two different models, one based on static redundancy and the other dynamic. The results show that both solutions significantly improve the reliability of the system.

Redundant and spatial diversity based transmissions have led to the development of new communication architectures. One approach is based on ''Parallel Redundancy Protocol'' (PRP) [36], which was originally developed to achieve seamless redundancy for Ethernet networks requiring high availability. Different works have proposed wireless PRP-like approaches such as [37], where PRP is used as diversity method on the wired Ethernet interfaces of two independent IEEE 802.11 WLAN channels that operate in parallel links. In [38] and [39], the same authors extended their study including the use of PRP for WLAN networks from different points of view such as reliability or latency. Specifically, reliability is measured for different noise levels, different jamming situations, and different packet lengths. In addition, they offer latency and jitter measurements in which they show that the use of PRP considerably improves its performance. Finally, in [40], the use of PRP for sensor networks in a star network configuration is proposed. The MAC level simulations show that when one of the two channels used for redundancy is

interfered with, the latency and jitter are improved by 121% and 376%. However, when both channels suffer interference, the improvement is drastically reduced.

Redundancy over Wi-Fi has been recently investigated under the name of Wi-Red (Wi-Fi Redundancy) [41]. Wi-Red aims at providing seamless link-level redundancy in IEEE 802.11 networks for industrial reliable wireless communications. The authors propose two working mechanisms depending on the complexity/reliability tradeoff: Reactive Duplicate Avoidance (RDA) and Proactive Duplicate Avoidance (PDA). From a latency perspective, the authors carried out several tests, where they conclude that both, RDA and PDA, reduce considerably the latency in industrial networks in comparison with other existing solutions. Then, in [42], the authors implemented a Wi-Red prototype using commercial 802.11 devices and they proved with a set of experiments that Wi-Red was able to enhance the latency and reliability in real industrial scenarios.

III. FA USE CASE AND EVALUATION PROCEDURE

This section contains the description of the use case implemented in this work and the evaluation procedure followed to tested the proposed solutions.

A. USE CASE

In this work, the use case emulates a small manufacturing cell with FA requirements [43]. Given that it is a reduced space (i.e., 10 × 10 m), it is assumed that the maximum number of nodes is 20. The network is made up of two types of equipment: Access Points (AP) and nodes. The former is in charge of distributing the information and of the synchronization tasks. On the contrary, the nodes are mainly receivers of the information or the commands sent by the APs. APs and nodes are organized in a centralized topology, where all the nodes receive data from the AP and send back their feedback in a timely organized manner. The diagram of the implemented architecture is shown in Fig. 1, where each node is connected to each AP in the network. Since industrial applications require deterministic communications, the wireless devices are preconfigured to transmit at specific time instants. As depicted in the figure, M indicates the number of nodes, N the number of APs, and γ the period of time between the communication of a node with two different APs. The information transmitted on the network is classified into two types: Critical Service (CS) and Best Effort (BE) service. CS is critical information that requires high reliability rates and low latency. Instead, BE is non-critical information, where the reliability and latency requirements are more flexible. Therefore, CS is configured with the lowest Modulation and Coding Scheme (MCS) (i.e., BPSK 1/2) and BE is configured with higher MCS that provide higher data rates (i.e., QPSK 1/2 - 16QAM 1/2). Regarding the size of the packets, as in [19], the size of the CS information packet has been set at 18 bytes, while for the BE the size is flexible.

Industrial scenarios are characterized by harsh wireless propagation conditions. To perform a realistic evaluation,

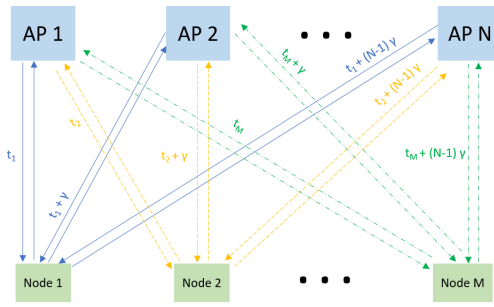


FIGURE 1. Network architecture.

two standard industrial wireless channels have been used, namely, CM7 and CM8 [44], which represents two industrial scenarios with different propagation properties. Although the CM7 and CM8 channels were initially generated based on IEEE 802.15.4a networks, the fact that there are no specific IEEE 802.11n models for industrial environments makes them a suitable candidate for our research purposes. For example, [45] describes several standard channel models for 802.11, but none of them reflects the characteristics of industrial environments. Concerning particular features of CM7 and CM8 channel models, the former represents line-of-sight (LOS) conditions, while CM8 is oriented to non-LOS (NLOS) conditions. As described in [44], channel measurements were carried out in larger enclosures (i.e., factory halls), filled with a large number of metallic reflectors, which implies a severe multipath effect. Moreover, the model was obtained following different measurements that cover a range from 2 to 8 m and frequency band from 2 to 10 GHz. As an example, Fig. 2 shows the channel gain time variability of three uncorrelated realizations of the CM7 channel.

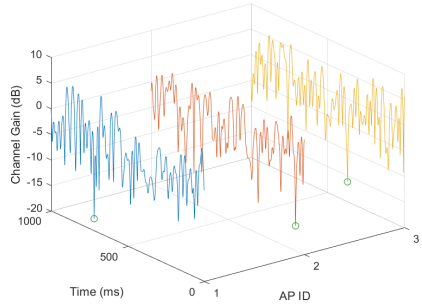


FIGURE 2. CM7 channels obtained from different APs.

The rest of the parameters related to the use case are shown in Table 1.

B. EVALUATION PROCEDURE

The evaluation procedure described in this subsection has been applied to the technical solutions proposed in the

TABLE 1. Use case parameters.

Parameter	Value
Application	FA
Center Frequency	2.4 GHz
Bandwidth	40 MHz
Channel model	CM7 - CM8
Type of Devices	AP and Nodes
Number of Nodes	Up to 20
Cell Size	10 x 10 m
Critical Service (Mbps)	12 Mbps
Non-critical Service (Mbps)	Up to 48 Mbps
Critical Service Payload	18 Byte

following sections. The simulations are based on a software tool designed on purpose for this work. It comprises a mathematical simulator (Matlab) and a network simulator (OMNeT++ [46]). The objective of the first one is the characterization of the reliability under a certain propagation channel while the second one emulates the communication network.

At the beginning of this work, different versions of the 802.11 standard were considered. In the first place, versions older than 802.11n were discarded due to the lack of LDPC codes. Afterwards, the 802.11ac standard was also discarded due to its higher overhead when compared with 802.11n. Finally, the 802.11ax standard presents configurations with similar or even better overhead than 802.11n since it incorporates several novelties, such as the trigger-frame based OFDMA medium access [47]. However, the 802.11ax transceiver presents a much higher complexity than 802.11n and its combination with NOMA may present several challenges. That is why it has been opted to use 802.11n and maintain a tradeoff between complexity and the performance parameters of industrial environments (i.e., reliability and latency) [6].

Concerning the implementation part, in a first step, a transmitter-receiver 802.11n standard chain has been implemented in Matlab [48]. Based on this implementation, several modifications have been carried out in the PHY level to introduce NOMA as a multiplexing option [19]. In addition, the transmitter-receiver chain also implements channel phenomena simulations, including fast-fading, free space loss, etc. To carry out the simulations, packets are sent assuming different channel realizations, and then, in reception, it is evaluated whether the packet has been correctly received or not. Each packet transmission set is performed for different Signal-to-Noise Ratio (SNR) values, with steps of 0.25 dB. The number of simulated packets and the SNR simulation steps have been adapted to the expected PER values. In particular, the simulated packet number is always at least one order of magnitude higher than the required value for obtaining the desired PER value assuming one single error.

In a second step, the network simulation tool is fed with the results obtained with Matlab. In particular, the PHY level performance measurements for different SNR values are introduced in OMNeT++ in order to use them in the error calculation block. Then, the network simulation tool comprises the network level simulation of both PHY and MAC layers. On the one hand, the PHY module facilitates parameters closely related to the physical layer, such as the MCS choice, the length of each transmission slot and the airtime of each data packet. On the other hand, the MAC layer implements the diversity schemes and the medium access control mechanisms. Since OMNeT++ does not support NOMA communications, NOMA has been modelled in OMNeT++ through two independent data flows in the same transmission period. Then, in the reception stage, both data flows are individually and orderly managed (i.e., first data in the UL, then, the data in the LL if UL is successfully recovered).

To decide if a packet is erroneous, this methodology uses the receiver instantaneous SNR. As shown in [49], it is difficult to quantify the uncorrelation grade of two channels with different paths within the same environment, and it is not possible to completely uncorrelate those channels. Therefore, in this work, a case of partial uncorrelation is assumed, where the instantaneous SNR of each node has been modeled as a combination of two components: the mean SNR, which is a static value that depends on the reception characteristics, and the variable attenuation of the channel which varies with time:

$$SNR_{i,k} = \theta_i + \alpha_{i,j,k}, \tag{1}$$

where i identifies the node, j the AP, and k is the time when the transmission takes place. On the other hand, θ is the mean SNR and the variable attenuation of the channel is α. To carry out the simulations with this model, for each path, an independent evolution of the channel variability is assumed (α), where a new channel is generated by varying the generation seed. In Fig. 2, three examples of three different CM7 channels are shown. It should be noted that the three examples have similar behavior since the channel varies with similar rate. However, the minimum fading of each channel (i.e., green circle in the figure) occurs at different moments, which indicates that despite having a similar long-term trend, the instantaneous attenuation at each time point is uncorrelated.

IV. PHY/MAC DESIGN

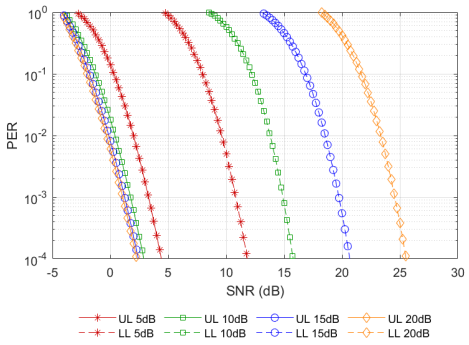
This section shows, on the one hand, the reliability performance that could be obtained solely with the PHY layer, and, on the other hand, the general architecture of the MAC layer and the retransmission techniques presented afterwards.

A. PHY LAYER

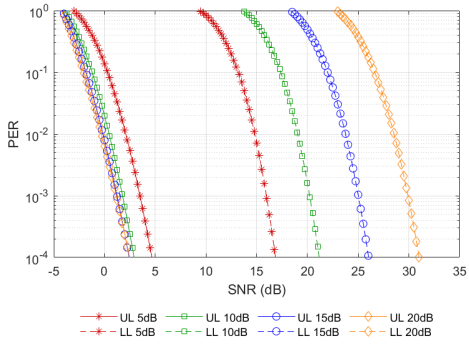
The development of this section is based on [19], and therefore, the transmitter-receiver chain, as well as the NOMA signal generation and cancellation have not been modified. However, this subsection extends the previous work by focusing on the effect of the IL and the MCS of the Lower Layer (LL). The IL ranges from a case where the LL is injected a few decibels below the Upper Layer (UL) (i.e., 5 dB) to cases where LL is deeply buried (i.e., 20 dB). The LL signal

configuration choices are, first, a robust configuration (i.e., MCS1, QPSK 1/2) and, then, a case with doubled capacity and reduced robustness (i.e., MCS3, 16QAM 1/2). On the other hand, the MCS0 is always selected for the UL in order to guarantee high reliability cases. The channel model used in the simulations is the CM7.

Fig. 3, contains the UL performance curves for different injection values (from 5 to 20 dB range). The best result occurs for an IL of 20 dB and provides a SNR requirement of 2.1 dB for a PER value of 10^{-4}. Additionally, UL curves show that the SNR difference in the range [10,20] is insignificant (<1 dB). In consequence, following analyses of the UL behaviour will focus only on the IL range from 5 to 10 dB.



(a) MCS0 in UL and MCS1 in LL.



(b) MCS0 in UL and MCS3 in LL.

FIGURE 3. PHY layer performance for different injection level values and MCS configurations under CM7 propagation channel.

The impact of IL is a two-fold problem because the performance of the LL has to be taken into account as well. The difference among the cases for the LL is more evident. The gap between LL curves is the IL value. Also, for a given IL, the difference between MCS1 and MCS3 is close to 5 dB, which is approximately the performance difference between both MCS configurations. Looking at the LL SNR thresholds for a PER value of 10^{-4}, the results obtained with both 15 and 20 dB lead to SNR thresholds well above 20 dB and look unrealistic for industrial application. IL values of 5 and 10 dB are the most suitable ones since they provide assumable SNR values for the LL.

The same simulations for the CM8 channel model have been carried out and the results are very similar (see Fig. 3). In general, when the CM8 channel model is used, the reliability performance shows a degradation between 0.1 dB and 0.3 dB. Given the similar performance trend in both channel model conditions (CM7 and CM8), the rest of the results presented in the following sections are based only on the CM7 channel model.

B. MAC LAYER

To meet the strict requirements associated with FA environments, PHY techniques have to be combined with efficient MAC level tools. In particular, an adequate MAC layer can potentially improve the reliability, guarantee bounded latency and increase the determinism of the overall communication system. In this section, the description of a MAC layer proposal and the different techniques described in the following sections are introduced.

Fig. 4 shows a time diagram of the superframe structure, which is based on TDMA in order to guarantee deterministic medium access to all the network devices. Three type of blocks can be identified: Downlink CS + BE transmission, Uplink Feedback (UF) and On-demand retransmissions. The first block is the initial transmission of both the critical and best-effort services. This transmission occurs in the downlink, from the AP to each of the nodes. Each downlink slot has a duration of 70 μs because is the minimum time required to transmit the 18 bytes of the CS (see Table 1) by using the lowest and most robust MCS (i.e., MCS0, BPSK 1/2). Furthermore, as the network is made up of 10 nodes the total length of the first block is 700 μs. Then, the uplink feedback phase begins, in which each node informs individually the AP (using each node a dedicated time slot) whether or not the CS and BE services have been correctly received (i.e., ACK/NACK). This information is sent in single-layer mode using the most robust configuration (i.e., MCS0, BPSK 1/2). The duration of each slot is 54 μs since, in this case, only one OFDM symbol is required to transmit the ACK/NACK information. Once the feedback from the nodes is received, the AP retransmits only the packets reported as incorrect. That is, in case the node reports that both services, CS and BE, have been incorrectly received, the AP would retransmit CS and BE in two-layer mode. However, if the node reports that only the BE service is erroneous, the AP retransmits only the BE content in single-layer mode. This block has the same duration as the initial block. Finally, since the number of retransmissions is configurable, blocks two and three would be repeated until the configured number of retransmissions is reached.

Taking into account the superframe structure, the minimum and maximum latency values can be calculated. In the case of the minimum latency, this is obtained when the packet is delivered in the initial transmission block and, therefore, the 70 μs correspond to the duration of a transmission slot.

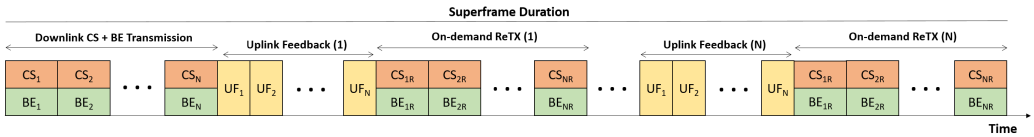


FIGURE 4. Superframe time representation.

On the contrary, the maximum latency value is obtained when the packet is received in the last on-demand retransmissions block. Therefore, the higher number of allowed retransmissions, the higher the maximum latency. That is why, up to three retransmission attempts are evaluated in order to keep assumable maximum latency values. Specifically, the maximum latency values are 1.24 ms, 2.48 ms and 3.72 ms for the one, two and three retransmissions, respectively.

Besides the flexibility in the number of retransmissions, the main difference with the superframe shown in [19] relays on the techniques used to carry out the retransmissions. In this case, in addition to the time-domain (see Section V), retransmissions with spatial diversity will also be implemented (see Section VI), where each one of the retransmissions is transmitted by a different AP. Finally, an additional retransmission mode is proposed. This alternative consists of using a higher injection level in the retransmissions to favor the correct reception of the CS packet (see Section VII). The following sections detail the design of each of the retransmission schemes and evaluate them from the point of view of reliability.

Although the MAC layer configuration presented in this section is specifically designed for the use case introduced in Section III-A, it would be possible to scale the MAC layer configuration to support more complex scenarios. For example, in a more heterogeneous case, there could be nodes that require only the CS, only the BE service, or both. In that case, each node should access the corresponding NOMA layer. If only the CS is needed, once the desired information has been decoded, the node would not access the LL layer. On the contrary, if only the BE service is required, the node should follow the entire decoding process associated with NOMA. However, note that the use of more nodes or with other configurations would not affect the overall reliability performance. Another variation could involve an industrial application built on top of the network architecture that requires a higher uplink traffic. In that case, the superframe could be modified to introduce more specific blocks for the uplink traffic. Furthermore, to introduce more uplink traffic, the medium could be managed through NOMA so that two nodes transmit the information simultaneously. However, this alternative would considerably increase the complexity of the receiver, since the IL between the layers would not be predetermined and would vary in each time slot and, therefore, at the receiver side, a block for estimating the IL would be necessary and/or transmitters should arrange their timing in advance [50].

V. TIME DOMAIN RETRANSMISSIONS

This section contains the design and evaluation of the combination between time-domain retransmissions with the NOMA-based 802.11n communications system.

A. DESIGN

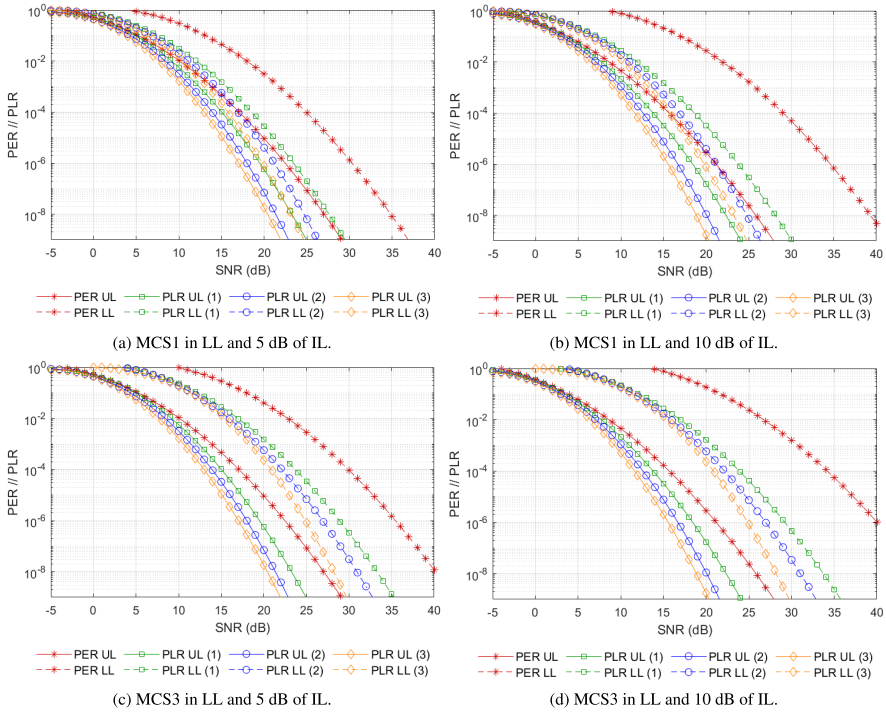
This scheme follows the time diagram presented in Fig. 4, where the blocks of on-demand retransmissions are carried out by the same AP that made the initial attempt (i.e, single transmitter). Consequently, each node SNR for the retransmitted signal is highly correlated with the SNR received in the previous transmission. In particular, following Eq. (1), the only difference between the initial transmission and the retransmission is the temporal evolution of the channel (k), while the origin (j) and the destination (i) is the same.

These types of retransmissions offer better results under channels with high variability and lower time coherence. However, although they are not offering a meaningful gain in all cases in terms of reliability, it should be noted that they entail a very small increase in complexity and implementation cost.

B. EVALUATION

The reliability results obtained with time-domain retransmissions are presented in Fig. 5, which is based on PER and Packet Loss Rate (PLR) measurements. In this case, PER represents the performance when no retransmissions are used, while PLR is for cases with retransmissions. Firstly, the best reliability performance values for the UL are obtained in the 10 dB case, which presents a difference close to 2 dB in comparison with the worst case (i.e., IL 5 dB). When the number of retransmissions is increased, the gap is reduced. In this case, gains close to 4 dB, 6 dB and 7 dB are obtained for one, two and three retransmissions, respectively. Although the reliability performance is improved, the best case SNRs are still around 20 dB, which is considered too high for the critical services. LL shows a different behavior. Although different ILs have been used, the results obtained are very similar, since the retransmissions are done in single-layer mode.

In general, time-domain retransmissions improve reliability but not significantly since the coherence time of the channel is higher than the time between retransmissions. Furthermore, it is worth noting the gain provided by the case of one retransmission, since introducing the first retransmission improves reliability more than when introducing the second or third retransmission. However, it would be necessary to



(a) MCS1 in LL and 5 dB of IL.

(b) MCS1 in LL and 10 dB of IL.

(c) MCS3 in LL and 5 dB of IL.

(d) MCS3 in LL and 10 dB of IL.

FIGURE 5. Reliability results obtained using a single transmitter with time-domain retransmissions under CM7 channel model and MCS0 in UL for different configurations.

evaluate depending on the target application which case best meets the reliability vs latency/complexity relationship.

VI. RETRANSMISSIONS FROM MULTIPLE TRANSMITTERS

This section shows the design and evaluation of combining retransmissions from multiple transmitters with NOMA.

A. DESIGN

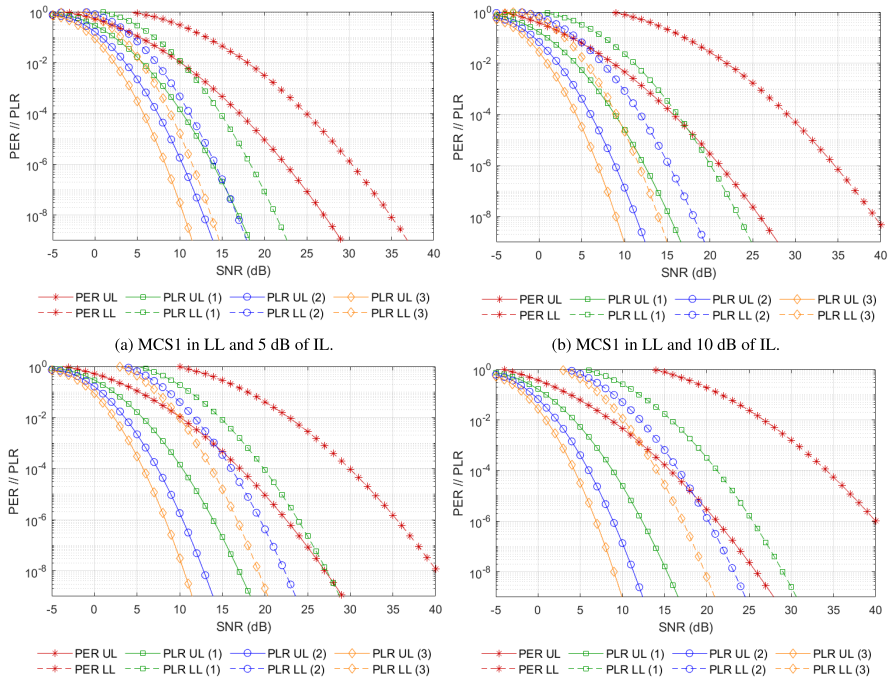
The second technique belongs to the family of spatial diversity and consists of using multiple transmitters. Specifically, taking into account the structure of the superframe (see Fig. 4), each block of on-demand retransmissions will be delivered by a different AP. So if the number of retransmissions configured is N, the number of APs that make up the network has to be $N + 1$. In addition, in the UF period, each node will send its feedback information to the AP that sent it the data in the previous block. It is assumed that all the APs are connected and synchronized so that they all receive the information that the UF contains. Additionally, nodes are suppposed to share a common time reference that grants interference free access to the medium as described in [51], [52].

The retransmissions are conveyed from a different AP, and thus, the initial transmission and the retransmission do not follow the same propagation path. Specifically, according to Eq. (1), since the mean SNR (θ) is affected by the receiver and the implementation environment, it will remain constant for the different retransmissions. On the other hand, when varying the retransmission path using another AP, the characteristics of the variable attenuation will be quite uncorrelated (i.e., j).

Time/space retransmissions from multiple transmitters will improve the reliability offered by time-domain retransmissions. However, introducing this scheme in a network includes higher implementation and synchronization costs. Therefore, depending on the gain obtained in reliability and the type of target application, they could be decisive.

B. EVALUATION

The results are gathered in Fig. 6. First, it is observed that the UL results are considerably better than those obtained using time-domain retransmissions only. In particular, the best results obtained using three retransmissions are around 10 dB. The gain associated to one, two and three retransmission is close to 11, 15 and 18 dB, respectively and for any IL value. The case that implements single retransmission is already better than any of the cases presented in the previous section (see Section V).



(a) MCS1 in LL and 5 dB of IL.

(b) MCS1 in LL and 10 dB of IL.

(c) MCS3 in LL and 5 dB of IL.

(d) MCS3 in LL and 10 dB of IL.

FIGURE 6. Reliability results obtained using retransmissions with multiple transmitters under CM7 channel model and MCS0 in UL for different configurations.

On the other hand, the results obtained for LL present a trend similar to that obtained in the previous section (see Section V). The case with the lowest reliability (i.e., MCS1 and 5 dB of IL) has gains of 14, 19 and 22 dB for one, two and three retransmissions, respectively. The case with the highest gain is that of the MCS3 with an IL of 10 dB, where 18, 24 and 28 dB are achieved for one, two and three retransmissions, respectively.

In general, the reliability obtained using multiple transmitters is greatly improved and the PLR rates obtained are applicable in mission-critical environments. However, an increase in complexity associated with the synchronization and deployment of APs has to be assumed. It is important to note that introducing a second AP is critical, since it provides a considerable increase in reliability. On the contrary, the differential gain provided by a third and fourth AP decreases significantly. Therefore, its implementation will depend on the use case and the reliability that needs to be addressed.

VII. RETRANSMISSIONS WITH VARIABLE IL

This section shows the design and evaluation of the combination of the retransmission techniques shown in the previous sections with adaptive injection levels.

A. DESIGN

This section presents a complementary technique to the previous time/space retransmissions. In this case, an adaptive injection level is proposed to increase the reliability of the UL. The adaptive IL provides a tool to make the UL more robust. If one of the nodes reports in the UF period that both services have been erroneously received, the AP retransmits the packets encoded with a larger IL, so that the LL impact is lower and the SNR required to decode the CS is lower also. However, in case the node has correctly received the CS and only needs the retransmissions of the BE service, the retransmission is carried out in single-layer mode. On the other hand, this technique does not present any additional complexity for the receiver nodes, since as in the UF period each node has requested the corresponding retransmissions if required, each node knows which IL is going to use the AP to encode the information. It should be noted that from now on the acronyms IL1 and IL2 represent the injection level used in the initial transmission and the one used for the retransmissions, respectively.

The adaptive injection level is a complementary technique to those schemes already evaluated in Section V and VI. Although the gain that can be obtained with this technique

is limited, the complexity involved is very low, so the reliability/complexity ratio offers an interesting alternative.

B. EVALUATION

Fig. 7 shows the results obtained using the adaptive IL method for all retransmission cases (i.e., time and space) using a different number of retransmissions (i.e., one, two and three). The gain represents the variation of the SNR required to achieve a PLR value of 10^{-9} when introducing the adaptive IL. Given that the gain obtained in all cases is similar and that it does not show a straightforward behavior, all cases have been grouped according to layer and MCS. In Fig. 7 the black whiskers represent the maximum and minimum values of each case, the blue lines represent the 25^{th} and 75^{th} percentiles and the red line the median value. Concerning the IL, for this case, IL1 has been set to 5 dB and IL2 to 10 dB.

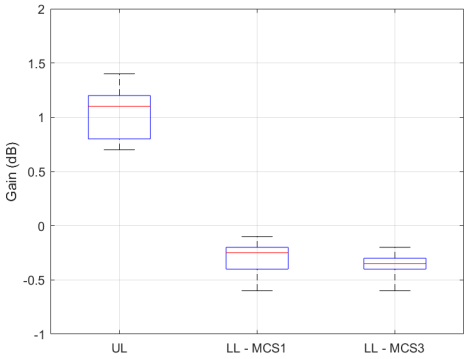


FIGURE 7. Improvement obtained with adaptive IL.

Fig. 7, proves that the performance of the UL layer is improved, assuming small losses in the LL layer. Specifically, the UL shows a median gain of 1.1 dB and taking into account that time retransmissions only provide around 4 dB of improvement with a single retransmission, it is a considerable gain. On the contrary, from the point of view of retransmissions using multiple transmitters, the gain is small compared to the values shown in Section VI-B. On the other hand, the LL reliability degrades. However, taking into account that LL is a BE service, and that degradation is lower than 1 dB, it is considered an acceptable performance loss.

VIII. COMPARISON OF PROPOSED RETRANSMISSION SCHEMES

This section compares all the solutions proposed in this paper in terms of the superframe and network size, reliability, and latency.

A. SUPERFRAME AND NETWORK SIZE

The number of nodes in the network or the duration of the superframe are also parameters to identify different industrial environments. FA, for instance, is more restrictive than Process Automation (PA) in terms of time related parameters such as delay, jitter or update time [43]. Therefore, this section presents the estimation of superframe lengths and the number of nodes in the network using the superframe structure shown in Fig. 4.

First, Table 2 shows the necessary superframe duration for networks in which the number of nodes is pre-established. It should be noted that all cases with a single retransmission have a superframe duration below 10 ms. However, as the number of retransmissions increases, the length of the superframe increases considerably and exceeds FA limits.

TABLE 2. Superframe size for different number of nodes.

Number of Nodes	1 ReTX (ms)	2 ReTX (ms)	3 ReTX (ms)
2	0.39	0.64	0.88
5	0.97	1.59	2.21
15	2.91	4.77	6.63
20	3.88	6.36	8.84
30	5.82	9.54	13.26
40	7.76	12.72	17.68
50	9.70	15.90	22.10

Table 3 displays results from a different approach. We assume that the maximum duration of the superframe is fixed by the use case and limited by the application cycle period. In this case, the table shows the maximum number of nodes that can be included in the superframe. For example, the first case, where the superframe cannot exceed one millisecond, is an FA case where a very small cycle time is required. In that case, the superframe supports up to five, three, and two nodes using one, two, and three retransmissions, respectively. On the other hand, if the limit value of FA environments (i.e., 10 ms) is taken into account, the number of possible nodes is multiplied by ten. Finally, the last two cases are oriented to PA environments and the number of admissible nodes is an order of magnitude higher than in the case of 10 ms.

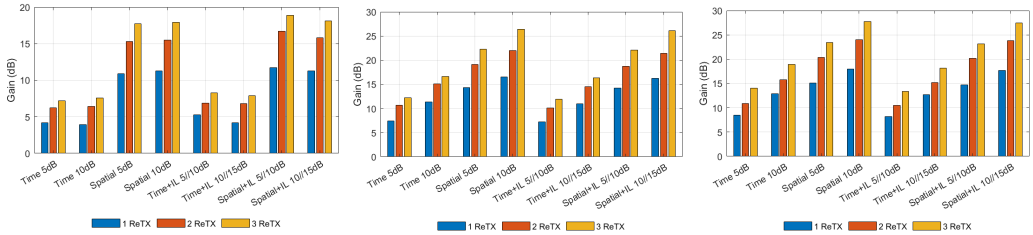
TABLE 3. Maximum number of nodes for different superframe sizes.

Superframe Length	1 ReTX	2 ReTX	3 ReTX
1 ms	5	3	2
5 ms	25	15	11
10 ms	51	31	22
50 ms	257	157	112
100 ms	515	314	226

B. RELIABILITY

In Fig. 8, the gain is calculated as the SNR difference in dB between the PER and the PLR required to have an error rate of 10^{-9}. The analysis is divided into three parts: UL, LL with MCS1 and LL with MCS3, which are represented in Fig. 8(a), Fig. 8(b) and Fig. 8(c), respectively.

In Fig. 8(a), the first conclusion is that retransmissions with multiple transmitters have a greater impact on the UL case. Regarding the adaptive injection level, it is observed that the reliability improves when using it but in a moderate



(a) Gain of the UL. (b) Gain of the LL using MCS1. (c) Gain of the LL using MCS3.

FIGURE 8. Comparison of the gain obtained by the different MAC techniques and for different configurations.

TABLE 4. Mean E2E latency analysis.

SNR	1 ReTX (μs)				2 ReTX (μs)				3 ReTX (μs)			
	Time	Sp.	Time + IL	Sp. + IL	Time	Sp.	Time + IL	Sp. + IL	Time	Sp.	Time + IL	Sp. + IL
UL												
5 dB	118	207	157	216	172	241	200	234	235	248	245	237
10 dB	76	83	80	83	83	83	84	83	88	83	86	83
15 dB	70	70	70	70	71	70	70	70	71	70	71	70
LL using MCS1												
5 dB	960	950	962	977	1048	1281	1045	1218	1145	1424	1139	1311
10 dB	451	466	453	468	476	492	475	488	498	494	496	489
15 dB	132	133	132	133	134	133	134	133	135	133	135	133
LL using MCS3												
5 dB	1226	1224	1226	1226	1406	1800	1407	1751	1605	2293	1605	2213
10 dB	970	983	970	985	1057	1242	1057	1237	1155	1348	1155	1339
15 dB	461	478	462	478	485	500	486	500	507	501	508	501

way. Finally, increasing the number of retransmissions is more effective over retransmissions with spatial diversity, with overall gain values close to 20 dB in some cases.

On the other hand, in Fig. 8(b) and Fig. 8(c), the LL shows higher gains than the UL, due to the configuration of retransmissions in single-layer mode. The performance behavior depending on the configuration is quite similar, although in general, the LL configured with MCS3 shows a slightly higher gain than MCS1. Furthermore, in this case, although the difference in performance between time-domain retransmissions and with multiple transmitters is still evident, it has diminished considerably. On the other hand, it can be said that the use of the adaptive injection level does not affect the LL, since the losses are negligible. Finally, it is worth noting that there are several cases in both MCS1 and MCS3 that exceed 25 dB of gain.

C. E2E LATENCY

Latency is another critical parameter in FA environments, which in general is related to reliability since many of the techniques to improve reliability involve an increase in latency. Therefore, in this work latency and reliability are analyzed to give a global vision of performance. In this case, End-to-End (E2E) latency is used to represent the time that elapses since the transmission of a packet begins until it is finally received, including retransmissions if necessary. The most representative E2E latency values have been gathered in Table 4. It shows the mean E2E latency values for different SNR values and each of the retransmission techniques. Specifically, the values shown for retransmissions in the time domain and with multiple transmitters have been obtained using an IL of 5 dB and those that use the adaptive injection level vary between 5 dB and 10 dB. Those cases have been chosen because they present the most balanced reliability results between UL and LL in the previous subsection. To calculate the mean latencies, three specific SNR values have been selected (i.e., 5, 10 and 15 dB). On average, the UL presents low latency values in which 250 μs are not exceeded in any case. In fact, the latency values with 15 dB SNR are practically the same as the minimum value, which indicates a high reliability rate. In turn, as expected, the latency values for LL are higher than for UL. However, despite the low SNR values used, in no case do they exceed 2 ms and taking into account that it is a BE service, it is an acceptable rate. Furthermore, it should be noted that for the 10 dB and 15 dB SNR cases using MCS1 and the 15 dB case using MCS3, the maximum mean latency values are around 500 μs.

Finally, if the latency values are analyzed as a function of the retransmission technique, it can be seen that the higher the reliability rate, the higher the average latency that must be assumed. That is why, the highest values appear in the case of multiple transmitters combined with adaptive IL since it is the case in which retransmissions are more efficient. The main reason for this is the success rate of the retransmission

TABLE 5. Analysis of the jitter results obtained for the UL.

SNR	1 ReTX (μs)				2 ReTX (μs)				3 ReTX (μs)			
	Time	Sp.	Time + IL	Sp. + IL	Time	Sp.	Time + IL	Sp. + IL	Time	Sp.	Time + IL	Sp. + IL
5 dB	114.37	114.06	114.43	114.26	194.57	168.27	194.16	168.41	226.62	203.50	226.21	203.54
10 dB	16.24	17.84	16.19	17.88	24.59	21.58	24.41	21.51	30.81	23.49	30.89	23.54
15 dB	0.81	0.82	0.77	0.85	0.98	0.88	0.96	0.89	1.12	0.87	1.06	0.89

schemes since error packets are not taken into account for the latency calculation. However, taking into account the gains in reliability that have been obtained in the previous section, the increase in latency is acceptable. Similarly, the more retransmissions that are used, the greater the latency that is obtained, since the size of the superframe is also lengthened.

To complement the E2E latency measurements, Table 5 shows the jitter values obtained for the UL with the same configurations as in Table 4. First of all, it should be noted that with an SNR of 5 dB the jitter values are quite high and even in some cases very close to the average latency. However, the jitter is greatly reduced when the SNR is increased. In particular, with 10 dB of SNR the jitter does not exceed 30 μs and with 15 dB the values are around one μs, which indicates that hardly any retransmissions are needed. Furthermore, in general, retransmissions based on multiple transmitters have better jitter values, since as they are more efficient, they require fewer retransmissions to guarantee the correct reception of the packets.

In summary, depending on the requirements of the final application, a compromise between latency and reliability has to be made.

IX. CONCLUSION

This paper presents and evaluates a set of techniques to increase communication reliability over the 802.11n standard. These techniques include the integration of NOMA within the 802.11n PHY layer, the use of a TDMA-based MAC scheme, and the use of different retransmission techniques in different diversity domains. To our best knowledge, this paper is the first to consider different retransmission schemes for a NOMA-based 802.11n communication system. In particular, time-domain and multiple transmitter based retransmissions have been combined with adaptive injection level and evaluated from the reliability and latency point of view.

The main conclusion is that each one of the evaluated retransmission techniques has its advantages and disadvantages. In particular, time-domain retransmissions are the simplest to implement, but also the ones that offer lower reliability due to channel dependence. However, it has been shown that by increasing the number of retransmissions, up to 7 dB of gain can be obtained for the CS. In contrast, retransmission schemes that include multiple transmitters have demonstrated much higher reliability rates. The main disadvantages of these techniques are the synchronization that is required and the deployment costs. The specific NOMA technique aimed at adapting the injection level to favor the reception of the UL in retransmissions has shown moderate gain values, but it can be useful in cases where the highest reliability rate is needed. Moreover, the latency analysis indicates that the higher the reliability offered by the retransmission schemes, the higher the latency that the system has to assume. However, taking into account the type of services proposed and the requirements of industrial environments, the latency values obtained are below the limits. Therefore, depending on the final application, in order to decide the optimal solution, a tradeoff has to be evaluated and assumed between reliability and latency.

This paper is oriented to the combination of NOMA techniques with TDMA-based medium access, therefore, the next step is to integrate NOMA in OFDMA-based communication schemes such as 802.11ax/be standards. To do this, on the one hand, the necessary modifications in the current 802.11ax/be transceiver architecture will be studied, as well as the potential gains that could be obtained. In addition, the complexity implications of the introduction of NOMA should be measured and low-complexity alternatives should be proposed to improve the complexity/performance tradeoff.

REFERENCES

[1] R. Drath and A. Horch, "Industrie 4.0: Hit or hype? [industry forum]," *IEEE Ind. Electron. Mag.*, vol. 8, no. 2, pp. 56–58, Jun. 2014.

[2] A. Varghese and D. Tandur, "Wireless requirements and challenges in industry 4.0," in *Proc. Int. Conf. Contemp. Comput. Informat. (IC I)*, Nov. 2014, pp. 634–638.

[3] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Müller, T. Elste, and M. Windisch, "Latency critical IoT applications in 5G: Perspective on the design of radio interface and network architecture," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 70–78, Feb. 2017.

[4] K. Montgomery, R. Candell, Y. Liu, and M. Hany, *Wireless User Requirements for the Factory Workcell* (Advanced Manufacturing Series (NIST AMS)), vol. 8. Gaithersburg, MD, USA: NIST, 2019.

[5] O. Seijo, I. Val, and J. A. Lopez-Fernandez, "w-SHARP: Implementation of a high-performance wireless time-sensitive network for low latency and ultra-low cycle time industrial applications," *IEEE Trans. Ind. Informat.*, early access, Jul. 7, 2020, doi: 10.1109/TII.2020.3007323.

[6] F. Tramarin, S. Vitturi, M. Luvisotto, and A. Zanella, "On the use of IEEE 802.11n for industrial communications," *IEEE Trans. Ind. Informat.*, vol. 12, no. 5, pp. 1877–1886, Oct. 2016.

[7] R. Rondón, M. Gidlund, and K. Landernäs, "Evaluating Bluetooth low energy suitability for time-critical industrial IoT applications," *Int. J. Wireless Inf. Netw.*, vol. 24, no. 3, pp. 278–290, Sep. 2017.

[8] M. Anwar, Y. Xia, and Y. Zhan, "TDMA-based IEEE 802.15.4 for low-latency deterministic control applications," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 338–347, Feb. 2016.

[9] I. Aktas, M. H. Jafari, J. Ansari, T. Dudda, S. A. Ashraf, and J. C. S. Arenas, "LTE evolution—Latency reduction and reliability enhancements for wireless industrial automation," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–7.

[10] X. Jiang, M. Luvisotto, Z. Pang, and C. Fischione, "Latency performance of 5G new radio for critical industrial control systems," in *Proc. 24th IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2019, pp. 1135–1142.

[11] G. Hampel, C. Li, and J. Li, "5G ultra-reliable low-latency communications in factory automation leveraging licensed and unlicensed bands," *IEEE Commun. Mag.*, vol. 57, no. 5, pp. 117–123, May 2019.

[12] D. Lopez-Perez, A. Garcia-Rodriguez, L. Galati-Giordano, M. Kasslin, and K. Doppler, "IEEE 802.11be extremely high throughput: The next generation of Wi-Fi technology beyond 802.11ax," *IEEE Commun. Mag.*, vol. 57, no. 9, pp. 113–119, Sep. 2019.

[13] A. Slalmi, R. Saadane, A. Chehri, and H. Kharraz, "How will 5G transform industrial IoT: Latency and reliability analysis," in *Human Centred Intelligent Systems*. Singapore: Springer, 2020, pp. 335–345.

[14] M. Luvisotto, Z. Pang, and D. Dzung, "Ultra high performance wireless control for critical applications: Challenges and directions," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1448–1459, Jun. 2017.

[15] M. Luvisotto, Z. Pang, and D. Dzung, "High-performance wireless networks for industrial control applications: New targets and feasibility," *Proc. IEEE*, vol. 107, no. 6, pp. 1074–1093, Jun. 2019.

[16] O. Seijo, Z. Fernandez, I. Val, and J. A. Lopez-Fernandez, "SHARP: A novel hybrid architecture for industrial wireless sensor and actuator networks," in *Proc. 14th IEEE Int. Workshop Factory Commun. Syst. (WFCS)*, Jun. 2018, pp. 1–10.

[17] E. Iradier, J. Montalban, L. Fanari, and P. Angueira, "On the use of spatial diversity under highly challenging channels for ultra reliable communications," in *Proc. 24th IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2019, pp. 200–207.

[18] M. C. Lucas-Estan, B. Coll-Perales, and J. Gozalvez, "Redundancy and diversity in wireless networks to support mobile industrial applications in industry 4.0," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 311–320, Jan. 2021.

[19] J. Montalban, E. Iradier, P. Angueira, O. Seijo, and I. Val, "NOMA-based 802.11n for industrial automation," *IEEE Access*, vol. 8, pp. 168546–168557, 2020.

[20] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[21] R. Hoshyar, F. P. Wathan, and R. Tafazolli, "Novel low-density signature for synchronous CDMA systems over AWGN channel," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1616–1626, Apr. 2008.

[22] F. Brannstrom, T. M. Aulin, and L. K. Rasmussen, "Iterative detectors for trellis-code multiple-access," *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 1478–1485, Sep. 2002.

[23] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, I. Chih-Lin, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[24] E. Iradier, J. Montalban, L. Fanari, P. Angueira, L. Zhang, Y. Wu, and W. Li, "Using NOMA for enabling broadcast/unicast convergence in 5G networks," *IEEE Trans. Broadcast.*, vol. 66, no. 2, pp. 503–514, Jun. 2020.

[25] N. M. Balasubramanya, A. Gupta, and M. Sellathurai, "Combining code-domain and power-domain NOMA for supporting higher number of users," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[26] L. Zhang, W. Li, Y. Wu, X. Wang, S.-I. Park, H. M. Kim, J.-Y. Lee, P. Angueira, and J. Montalban, "Layered-division-multiplexing: Theory and practice," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 216–232, Mar. 2016.

[27] E. Iradier, J. Montalban, L. Fanari, P. Angueira, O. Seijo, and I. Val, "NOMA-based 802.11n for broadcasting multimedia content in factory automation environments," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Jun. 2019, pp. 1–6.

[28] M. Jonsson and K. Kunert, "Towards reliable wireless industrial communication with real-time guarantees," *IEEE Trans. Ind. Informat.*, vol. 5, no. 4, pp. 429–442, Nov. 2009.

[29] P. G. Peon, E. Uhlemann, W. Steiner, and M. Bjorkman, "Applying time diversity for improved reliability in a real-time heterogeneous MAC protocol," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Jun. 2017, pp. 1–7.

[30] B. Goektepe, S. Faehse, L. Thiele, T. Schierl, and C. Hellge, "Subcode-based early HARQ for 5G," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[31] B. Makki, T. Svensson, G. Caire, and M. Zorzi, "Fast HARQ over finite blocklength codes: A technique for low-latency reliable communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 194–209, Jan. 2019.

[32] F. Ghanami, G. A. Hodtani, B. Vucetic, and M. Shirvanimoghaddam, "Performance analysis and optimization of NOMA with HARQ for short packet communications in massive IoT," *IEEE Internet Things J.*, early access, Oct. 2, 2020, doi: 10.1109/JIOT.2020.3028434.

[33] H. Beikirch, M. Voss, and A. Fink, "Redundancy approach to increase the availability and reliability of radio communication in industrial automation," in *Proc. IEEE Conf. Emerg. Technol. Factory Autom.*, Sep. 2009, pp. 1–4.

[34] B. Kilberg, C. B. Schindler, A. Sundararajan, A. Yang, and K. S. J. Pister, "Experimental evaluation of low-latency diversity modes in IEEE 802.15.4 networks," in *Proc. IEEE 23rd Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, vol. 1, Sep. 2018, pp. 211–218.

[35] P. G. Peon, W. Steiner, and E. Uhlemann, "Network fault tolerance by means of diverse physical layers," in *Proc. 25th IEEE Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, vol. 1, Sep. 2020, pp. 1697–1704.

[36] H. Kirrmann, M. Hansson, and P. Muri, "IEC 62439 PRP: Bumpless recovery for highly available, hard real-time industrial networks," in *Proc. IEEE Conf. Emerg. Technol. Factory Autom. (EFTA)*, Sep. 2007, pp. 1396–1399.

[37] M. Rentschler and P. Laukemann, "Towards a reliable parallel redundant WLAN black channel," in *Proc. 9th IEEE Int. Workshop Factory Commun. Syst.*, May 2012, pp. 255–264.

[38] M. Rentschler, O. A. Mady, M. T. Kassis, H. H. Halawa, T. K. Refaat, R. M. Daoud, H. H. Amer, and H. M. ElSayed, "Simulation of parallel redundant WLAN with OPNET," in *Proc. IEEE 18th Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2013, pp. 1–8.

[39] M. Rentschler and P. Laukemann, "Performance analysis of parallel redundant WLAN," in *Proc. IEEE 17th Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2012, pp. 1–8.

[40] M. Hendawy, M. ElMansoury, K. N. Tawfik, M. M. ElShenawy, A. H. Nagui, A. T. ElSayed, H. H. Halawa, R. M. Daoud, H. H. Amer, M. Rentschler, and H. M. ElSayed, "Application of parallel redundancy in a Wi-Fi-based WNCS using OPNET," in *Proc. IEEE 27th Can. Conf. Electr. Comput. Eng. (CCECE)*, May 2014, pp. 1–6.

[41] G. Cena, S. Scanzio, and A. Valenzano, "Seamless link-level redundancy to improve reliability of industrial Wi-Fi networks," *IEEE Trans. Ind. Informat.*, vol. 12, no. 2, pp. 608–620, Apr. 2016.

[42] G. Cena, S. Scanzio, and A. Valenzano, "A prototype implementation of Wi-Fi seamless redundancy with reactive duplication avoidance," in *Proc. IEEE 23rd Int. Conf. Emerg. Technol. Factory Autom. (ETFA)*, Sep. 2018, pp. 179–186.

[43] J. F. Coll, J. Chilo, and B. Slimane, "Radio-frequency electromagnetic characterization in factory infrastructures," *IEEE Trans. Electromagn. Compat.*, vol. 54, no. 3, pp. 708–711, Jun. 2012.

[44] A. F. Molisch, K. Balakrishnan, C.-C. Chong, S. Emami, A. Fort, J. Karedal, J. Kunisch, H. Schantz, U. Schuster, and K. Siwiak, "IEEE 802.15. 4a channel model-final report," *IEEE P802*, vol. 15, no. 4, p. 0662, Nov. 2004.

[45] V. Erceg, L. Schumacher, and P. Kyritsi, *TGN Channel Models*, document IEEE 802.11-03/940r1, Garden Grove, CA, USA, 2004.

[46] A. Varga and R. Hornig, "An overview of the OMNeT++ simulation environment," in *Proc. 1st Int. ICST Conf. Simulation Tools Techn. Commun. Netw. Syst.*, 2008, p. 60.

[47] B. Bellalta, "IEEE 802.11ax: High-efficiency WLANS," *IEEE Wireless Commun.*, vol. 23, no. 1, pp. 38–46, Feb. 2016.

[48] *Orthogonal Frequency Division Multiplexing (OFDM) PHY Specification, de Standard—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Nueva York, Estados Unidos*, Standard 802.11, IEEE Computer Society, 2012, pp. 1583–1630.

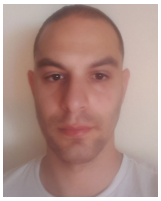
[49] V. N. Swamy, P. Rigge, G. Ranade, B. Nikolic, and A. Sahai, "Wireless channel dynamics and robustness for ultra-reliable low-latency communications," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 705–720, Apr. 2019.

[50] S. Abeywickrama, L. Liu, Y. Chi, and C. Yuen, "Over-the-air implementation of uplink NOMA," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

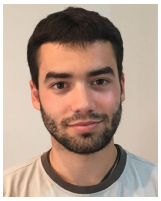
[51] O. Seijo, J. A. Lopez-Fernandez, H.-P. Bernhard, and I. Val, "Enhanced timestamping method for subnanosecond time synchronization in IEEE 802.11 over WLAN standard conditions," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 5792–5805, Sep. 2020.

[52] A. Mahmood, R. Exel, H. Trsek, and T. Sauter, "Clock synchronization over IEEE 802.11—A survey of methodologies and protocols," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 907–922, Apr. 2017.

ENEKO IRADIER (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications engineering from the University of the Basque Country (UPV/EHU), in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree with the TSR Research Group. Since 2015, he is part of the TSR Research Group, UPV/EHU. He was with the Communications Systems Group, IK4-Ikerlan, as a Researcher, from 2017 to 2018. During his doctoral studies, he did an internship with the Communications Research Centre Canada, Ottawa. His current research interests include design and development of new technologies for the physical layer of communication systems and wireless solutions for Industry 4.0. He has served as a reviewer for several renowned international journals and conferences in the area of wireless communications.



LORENZO FANARI (Graduate Student Member, IEEE) received the B.Sc. degree in electrical and electronic engineering and the M.Sc. degree in telecommunication engineering from the University of Cagliari, Italy, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the University of Basque Country, Spain. His research interests include coding theory and wireless communications.



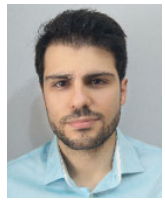
IÑIGO BILBAO (Graduate Student Member, IEEE) received the B.S. degree in physics and the M.S. in space science and technology from the University of the Basque Country (UPV/EHU), in 2019 and 2020, respectively, where he is currently pursuing the Ph.D. degree with the TSR Research Group. Since 2020, he is part of the TSR Research Group, UPV/EHU. His current research interests include design and development of new technologies for the physical layer of wireless solutions for Industry 4.0.

JON MONTALBAN (Senior Member, IEEE) received the M.S. and Ph.D. degrees in telecommunications engineering from the University of the Basque Country, Spain, in 2009 and 2014, respectively. He is part of the TSR (Radiocommunications and Signal Processing) Research Group with the University of the Basque Country, where he is an currently an Assistant Professor involved in several research projects. He has hold visiting research appointments with the Communication Research Centre (CRC), Canada, and Dublin City University (DCU), Ireland. His current research interests include wireless communications and signal processing for reliable industrial communications. He was a co-recipient of several best paper awards, including the Scott Helt Memorial Award to recognize the best paper published in the IEEE TRANSACTIONS ON BROADCASTING, in 2019. He has served as a reviewer for several renowned international journals and conferences in the area of wireless communications. He is currently serves as an Associate Editor for IEEE ACCESS.

PABLO ANGUEIRA (Senior Member, IEEE) received the M.S. and Ph.D. degrees in telecommunication engineering from the University of the Basque Country, Spain, in 1997 and 2002, respectively.

In 1998, he joined the Department of Communications Engineering, University of the Basque Country, where he is currently a Full Professor. He is part of the staff of the Signal Processing and Radiocommunication Laboratory, where he has been involved in research on digital broadcasting (DVB-T, DRM, T-DAB, DVB-T2, DVB-NGH, and ATSC 3.0) for more than 20 years. He is a coauthor of an extensive list of articles in international peer-reviewed journals, and a large number of conference presentations in digital broadcasting. He has also coauthored several contributions to the ITU-R working groups WP6 and WP3. His main research interest includes design and development of new technologies for the physical layer of communication systems in industrial wireless environments.

Dr. Angueira is an Associate Editor of the IEEE TRANSACTIONS ON BROADCASTING, a member of the IEEE BMSB International Steering Committee, and a Distinguished Lecturer of the IEEE BTS. He serves on the Administrative Committee for the IEEE BTS.



OSCAR SEIJO (Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications engineering from the University of Oviedo, Spain, in 2015 and 2017, respectively. He is currently pursuing the Ph.D. degree with the Communications Systems Group, IKERLAN Technological Research Centre, Mondragón, Spain, in collaboration with the Signal Theory and Communications (TSC) Group, University of Oviedo. His research interests include wireless high-performance PHY and MAC design for industrial applications, time synchronization over wireless systems, and digital signal processing.

IÑAKI VAL (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees from the Department of Electronics Engineering, University of Mondragon, Spain, in 1998 and 2001, respectively, and the Ph.D. degree from the Department of Signals, Systems, and Radiocommunication, Polytechnic University of Madrid, Spain, in 2011. Since 2001, he has been with the Communications Systems Group, IKERLAN, Mondragón, Spain. He was with the Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany, as an Invited Researcher, from 2005 to 2006. He is currently the Team Leader of Communication Systems Group. His research interests include design and implementation of digital wireless communications systems, industrial real-time requirements, communications for distributed control systems, vehicular communications, time synchronization, wireless channel characterization, and digital signal processing. He is currently focused on industrial communication applications.

3.2.5 Journal paper J6

This subsection presents a journal paper related with Contribution 2. The full reference of the paper is presented below:

- E. Iradier, A. Abuin, L. Fanari, J. Montalban and P. Angueira, "Throughput, Capacity and Latency Analysis of P-NOMA RRM Schemes in 5G URLLC," accepted manuscript for publication in Multimedia Tools and Applications.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Journal paper indexed in JCR
- Area: Computer Science, Software Engineering
- Ranking: 34/108 (Q2) based on JCR 2019
- Impact factor (JCR): 2.313

Throughput, Capacity and Latency Analysis of P-NOMA RRM Schemes in 5G URLLC

Eneko Iradier[1], Aritz Abuin[2], Lorenzo Fanari[3], Jon Montalban[4], Pablo Angueira[5]

Abstract – 5G is expected to cover a wide range of potential use cases due to its flexible and configurable physical layer waveform. One of the use cases proposed is the application of 5G on Ultra-Reliable Low Latency Communication (URLLC), which are characterized by very challenging reliability/availability and latency requirements. In addition, multimedia applications are being consolidated as a relevant aspect of current industrial environments. In order to meet those strict requirements, Radio Resource Management (RRM) becomes a critical phase of any wireless communication system. This work proposes the use of power domain Non-Orthogonal Multiple Access (P-NOMA) techniques in 5G RRM for factory automation environments. The research presented in this paper includes the design and evaluation of different RRM algorithms based on P-NOMA and on traditional Orthogonal Multiple Access (OMA) techniques, such as Time/Frequency Division Multiplexing Access (T/FDMA). Those algorithms are comprehensively explained and oriented to optimize the resource allocation based on different metrics (i.e., capacity, number of users, or cycle time). Moreover, extensive results are presented, where the performance of NOMA and OMA techniques is compared in terms of different metrics under the influence of several parameters, such as payload, bandwidth, or the number of users. Results indicate that although both NOMA and OMA provide positive aspects, eventually NOMA-based RRM algorithms are the solutions that enhance considerably the spectral efficiency.

Keywords – 5G, Factory Automation, NOMA, P-NOMA, Resource Allocation, Resource Block, RRM, TDMA, URLLC, Wireless Communications.

Declarations:

Funding - This work was supported in part by the Basque Government (Project IOTERRAZ) under Grant KK-2019/00046 ELKARTEK 2019, Grant IT1234-19 and the PREDOC Grant Program PRE_2020_2_0105, and in part by the Spanish Government (Project PHANTOM) under Grant RTI2018-099162-B-I00 (MCIU/AEI/FEDER, UE).

Conflicts of interest/Competing interests - Not applicable

Availability of data and material - Not applicable

Code availability - Not applicable

[1] Department of Communications Engineering, University of the Basque Country, Bilbao, Spain. Email: eneko.iradier@ehu.eus. ORCID ID: 0000-0002-0424-3857
[2] Department of Communications Engineering, University of the Basque Country, Bilbao, Spain. Email: aabuin003@ikasle.ehu.eus. ORCID ID: 0000-0003-4868-4029
[3] Department of Communications Engineering, University of the Basque Country, Bilbao, Spain. Email: lorenzo.fanari@ehu.eus. ORCID ID: 0000-0001-8023-1973
[4] Department of Electronic Technology, University of the Basque Country, Eibar, Spain. Email: jon.montalban@ehu.eus. ORCID ID: 0000-0003-0309-3401
[5] Department of Communications Engineering, University of the Basque Country, Bilbao, Spain. Email: pablo.angueira@ehu.eus. ORCID ID: 0000-0002-5188-8412

1. Introduction

5G is expected to radically change wireless communications, not only from a qualitative point of view but also by widening the range of potential applications and use cases. The first step has been the standardization of 5G New Radio (NR) in Release 15 (Rel-15) [2]. 5G NR introduces several novelties in comparison with its predecessor (i.e., Long Term Evolution, LTE) at different levels. For example, it presents a configurable physical layer (PHY) that makes 5G a real candidate for applications with completely different requirements. In this release, there is also a considerable increase of the physical layer reliability due to the use of Low-Density Parity-Check (LDPC) codes for data content and Polar codes for control packets. The International Telecommunication Union Radiocommunication sector (ITU-R) has highlighted three main use cases (See Fig. 1) within the 5G framework [31]. The first, enhanced Mobile Broadband (eMBB) is oriented to data-driven use cases that require high throughput within a large coverage area. This use case is related to classical infotainment applications and extended to Augmented Reality (AR) and Virtual Reality (VR) applications. It is expected to overcome the Gpbs barrier by offering up to 20 Gbps of peak data rate for the downlink and up to 10 Gbps for the uplink. The second use case is Ultra Reliable Low Latency Communications (URLLC), which targets mission-critical applications, such as remote surgery or autonomous vehicles [5]. These use cases imply strict requirements on latency and reliability [29]. In fact, a maximum end-to-end latency in the user plane of one millisecond has been established. Finally, the third use case is massive Machine Type Communications (mMTC), which focuses on Internet of Things (IoT) applications. In this case, the simultaneous connection of a very large number of devices has to be supported in small areas. The target device density is around 10^6 devices per square kilometer.

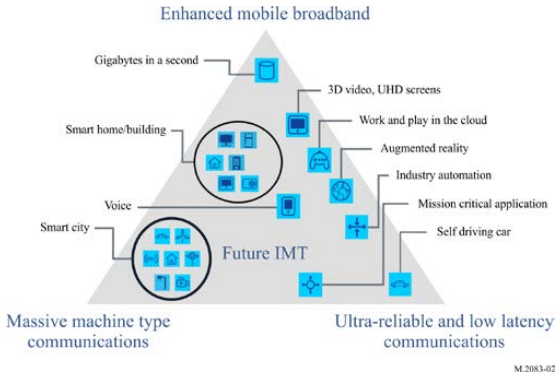


Fig. 1 5G NR applications and use case families [1]

Meanwhile, the second phase of 5G is being standardized in parallel by 3GPP in the Release 16 (Rel-16) and it is expected to be completed during 2020. Although Rel-15 was oriented to eMBB services, Rel-16 goes beyond and targets URLLC and mMTC services by introducing new features, such as Time Sensitive Communication (TSC) or Enhanced Location Services (ELS). Finally, the standardization process of Rel-17 has also taken off. In this case, it is in an initial phase and its first outcomes are expected for 2021. Rel-17 and further releases are oriented to support solutions to more global features, such as the convergence of wireless and wired systems, multicast and broadcast architectures, or multi-access edge computing [18].

URLLC is a wide use case closely related to the industry, where several applications can be included with considerably different requirements. In particular, in the last decade, there has been a considerable increase in the inclusion of multimedia applications within the industry. On the one hand, the use of AR is being promoted within different industrial applications, such as shipyards [9] [15]. Moreover, in [14], AR is used to deliver assembly instructions to workers. Finally, the authors in [17] propose and validate an extended reality (XR)-based framework that executes five typical iterative phases of industrial environments: requirements analysis, solution selection, data preparation, system implementation, and system evaluation. On the other hand, another classic aspect of the industry that is reinforced with the improvement of multimedia applications is video surveillance. Some representative examples are [21] and [36], where the use of video applications is proposed to automatically identify products in the steel industry and for

acquisition and self-learning autonomous surveillance services, respectively. Furthermore, in [7], a video-monitoring system with a distributed architecture is proposed. Finally, the authors in [16], through a more specific proposal, propose and design a multi-dimensional architecture based on video monitoring on MPEG-DASH (Dynamic Adaptive Streaming over HTTP).

In order to support the combination of multimedia applications with industrial characteristics, the network should support the simultaneous delivery of services with different requirements, such as unicast and broadcast services, loss tolerant and loss intolerant services, or downlink and uplink connections. Doubtlessly, the increase in the number of potential applications within a common cell makes the provision of high spectral efficiency a critical metric. As in LTE, Radio Resource Management (RRM) is based on orthogonal multiple access (OMA) techniques that share the available resource blocks (RB) among the users. However, although, if compared with LTE, the spectral efficiency has been improved considerably in 5G, there is still room for improvement in the RRM module.

This work proposes the introduction of non-orthogonal multiple access (NOMA) techniques in current RRM modules. NOMA techniques have demonstrated better spectral efficiency results than OMA [28]. In the literature, there are several proposals to use power domain NOMA (P-NOMA) to guarantee convergence of different services, while increasing the spectral efficiency [22] [47]. Indeed, both approaches have been focused on the study and the benefits of a P-NOMA-based PHY without taking into account the RRM level. However, this work is oriented to provide and evaluate different RRM algorithms.

The objective of this paper is to propose and evaluate the use of NOMA in RRM techniques for 5G URLLC in industrial environments. Two main multiplexing access techniques are considered in this paper: Time Division Multiplexing Access (TDMA) as a reference of OMA, and P-NOMA in the representation of NOMA techniques. Both alternatives are deeply analyzed from different perspectives (throughput, capacity, and latency) and optimized to meet different requirements (maximum capacity/throughput and minimum cycle time). Then, the parameters that affect the performance are identified and some guidelines are provided in order to decide the best performing multiplexing technique according to different parameter ranges.

In summary, the technical contributions of this paper include:

- Propose and evaluate TDMA for RRM in 5G URLLC and design several broadcast/unicast convergence algorithms.
- Propose and evaluate P-NOMA for RRM in 5G URLLC and design several broadcast/unicast convergence algorithms.
- Provide a detailed analysis of both alternatives in terms of throughput (ADR), capacity (served users), and latency (cycle time).
- Determine the best performing multiplexing technique according to a specific set of parameter ranges (payload size, maximum cycle time, bandwidth, and number of users).

The rest of the paper is organized as follows. Section 2 presents the basic principles of NOMA. An overview of the RRM in 5G for URLLC is also provided in that section. The industrial use case and the designed algorithms are introduced in section 3 and section 4, respectively. In section 5, the results obtained from evaluating the proposed algorithms in the industrial use case are presented and a summary is provided, where the best performing multiplexing technique is selected according to some specific parameter range. Finally, the conclusions and future works are highlighted in section 6.

2. Basic Principles

2.1. Basic principles of NOMA

NOMA represents a wide family of medium access techniques where all users share the same frequency and time resources. In fact, NOMA can be classified into code-domain NOMA [20] and power-domain NOMA [45]. This paper is focused on low complexity P-NOMA systems since they are considered less complex solutions [30]. During the rest of the paper, NOMA will refer to power domain NOMA.

NOMA refers to the signal ensemble composed of several superposed layers, where each layer takes a portion of the total power delivered by the transmitter. Generally, modulation and coding of each layer are

independently configured in order to offer different robustness and capacities. The injection level (Δ) describes the power splitting among layers in dB. A two-layer NOMA signal can be expressed as:

$$x_{NOMA}(k) = x_{s1}(k) + g \cdot x_{s2}(k), \quad (1)$$

where $x_{s1}(k)$ and $x_{s2}(k)$ are the two data streams combined into $x_{NOMA}(k)$, k is the subchannel index and g is the injection level expressed in linear units.

NOMA was firstly introduced by Cover and Bergmans in [8][11]. It was based on a technique called Successive Interference Cancellation (SIC), in which the receiver demodulates, decodes, and cancels successive information layers from the received signal ensemble. Layered Division Multiplexing (LDM) is a particular low-complexity case of NOMA [46], which presents an architecture that facilitates its implementation in real hardware elements. In addition, LDM also provided a considerable increase in spectral efficiency in comparison with OMA techniques. In fact, due to both reasons, complexity, and capacity, LDM was accepted as part of the latest digital television standard (ATSC 3.0) [37].

Power domain NOMA techniques have also been proposed to be used in broadband technologies. First, in [24] LDM was proposed in combination with LTE-Advanced (LTE-A) for multicast vehicular communications. Then, in [33], a more in-depth analysis of the use case was presented. In addition, the use of NOMA for enabling multicast subgrouping techniques in an evolved multimedia broadcast multicast service (eMBMS)-like 5G scenario was proposed in [33]. The latest works, such as [27], [22], and [47], are oriented to integrate NOMA in the PHY of 5G NR providing different capacity, reliability, and latency analysis where NOMA demonstrates that outperforms classical OMA techniques.

As an example of the advantages of NOMA, in Fig. 2, a representation of the spectral efficiency offered by several NOMA and T/FDMA configurations is shown for different Signal to Noise Ratio (SNR) requirements. It is important to highlight the effect of the service asymmetry: the higher the difference between the service requirements (i.e., the difference between the required SNR for each service), the higher the gain of NOMA. Specifically, the greatest gains are depicted for the services that combine SNRs of 0 dB and 20 dB. What is more, for a hypothetical value of 0.6 bps/Hz for service one, more than 2 bps/Hz of gain is achieved for service two by using NOMA.

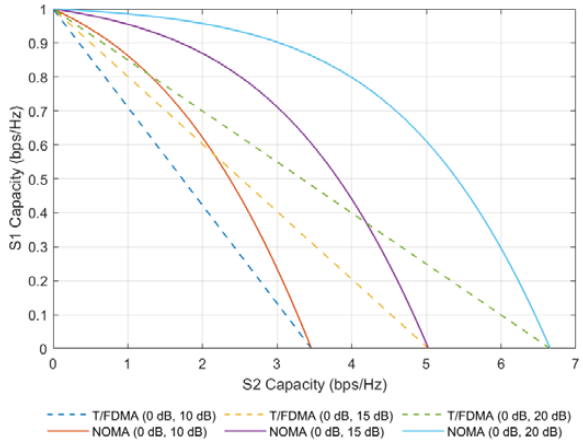


Fig. 2 Spectral efficiency comparison between NOMA and T/FDMA

2.2. RRM in 5G

The physical layer waveform of LTE is based on Orthogonal Frequency Division Multiplexing (OFDM). 5G follows the same principle. Nevertheless, one of the novelties of this standard is the flexibility of the PHY configuration. Five numerologies (μ) have been designed by varying the subcarrier spacing (SCS). The SCS value of each numerology can be obtained as follows:

$$SCS(kHz) = 15 \cdot 2^{\mu}, \quad (2)$$

where μ is the numerology and it can be 0, 1, 2, 3, or 4.

Concerning the frame structure, the length of the frames in 5G NR is set to 10 ms and they are composed of 10 consecutive subframes of one millisecond. Then, the slot number within a subframe varies depending on the numerology that is implemented. Table 1 describes the number of slots per subframe and their duration is described as a function of the numerology. These configuration choices of 5G NR are one of the main differences with LTE since those five different numerologies enable applications with different requirements. Moreover, depending on the used carrier frequency, the possible numerologies are limited: SCS of 15, 30, and 60 kHz are used below 6 GHz (i.e., frequency range 1, FR1), and 60 and 120 kHz SCSs are used above 6 GHz (i.e., frequency range 2, FR2) [40]. As described in Table 1, numerologies with high SCS imply short slot durations and, therefore, lower latencies could be achieved. This is the reason why, the highest numerologies (i.e., μ = 2 below 6 GHz and μ = 3 and μ = 4 above 6 GHz) are more suitable for URLLC.

Table 1 5G NR frame components based on different numerologies [1]

μ	SCS (kHz)	# Slots/Subframe	Slot (ms)
0	15	1	1
1	30	2	0.5
2	60	4	0.25
3	120	8	0.125
4	240	16	0.0625

Regarding the number of OFDM symbols per slot, the same strategy as in LTE has been followed. In general, 14 symbols per slot is the default value in 5G NR, whereas some flexibility is also introduced in the symbol number. On the one hand, mini-slot configurations are available, where 2, 4, or 7 symbols can be used per slot. This configuration is designed to satisfy the requirements of applications where short data packets have to be delivered like in URLLC. On the contrary, especially for eMBB environments, several slots can be merged for long packet transmissions.

Another parameter affected by the flexibility of the numerologies is the OFDM symbol length. As presented in Table 2, the length of the OFDM symbol decreases while the SCS is increased. For the two highest numerologies, the duration of the OFDM symbol including the Cyclic Prefix (CP) is less than 10 μs, and thus, it is a perfect fit for URLLC applications.

Table 2 OFDM symbol sizes in 5G [1]

	μ = 0	μ = 1	μ = 2	μ = 3	μ = 4
OFDM Symbol (μs)	66.67	33.33	16.67	8.33	4.17
CP (μs)	4.69	2.34	1.17	0.57	0.29
Total length (μs)	71.36	35.67	17.84	8.90	4.46

The resource allocation in 5G NR has also been modified in comparison with LTE. In this case, the smallest unit defined in the resource grid is the Resource Element (RE). RE is defined by one subcarrier and one OFDM symbol, in frequency and time domain, respectively [1]. Nevertheless, it is not possible to directly allocate RE, as in LTE, Resource Block (RB) are allocated, although there are not as in 4G. While in LTE one resource block was defined by 12 consecutive subcarriers in one slot (i.e., 7 OFDM symbols), in 5G NR, RBs are composed by one OFDM symbol in the time domain and 12 consecutive subcarriers in the frequency domain, corresponding to 180 kHz for 15 kHz SCS, 360 kHz for 30 kHz, and so forth [38]. Finally, in 5G NR, the maximum channel bandwidth is up to 100 MHz in FR1 and up to 400 MHz in FR2. In the second case, channels can be aggregated with a maximum bandwidth of 800 MHz.

3. Industrial use case

URLLC is a wide use case, where several types of applications are involved. One of the most relevant concepts and considered as the fourth industrial revolution is the so-called Industry 4.0 [13]. Industry 4.0 is leading the research and the advances to define and design the future industrial systems, including industrial communications. One of the main characteristics is the progressive deployment of wireless communications. Although wired connections are still commonly used in the industry due to their high

reliability, wireless systems offer an interesting improvement in scalability, cost reduction, and mobility that make them an ideal candidate for future industrial communication systems [44].

Several technologies and standards have been proposed towards a massive deployment of industrial wireless networks, such as IEEE 802.11 [43][42], Bluetooth [39], 802.15.4 [6], LTE [4], or WirelessHart [19]. However, none of those technologies can be considered the optimum solution for implementation in the industrial environment due to the strict characteristics and requirements involved. A detailed explanation of the strict requirements is provided in the following paragraphs.

One of the main challenges of industrial environments is the propagation channel. Metallic objects, machines in the industrial manufacturing plants, moving robots, etc. create a wireless propagation channel with a harsh multipath effect and a high access concurrence [10], which is hard to overcome. Such challenging channels cannot be overcome solely by physical layer techniques. Different MAC level techniques are required to complement the PHY performance, such as time division retransmissions [42][34] or spatial diversity techniques [25][23].

Industrial communications comprise different use cases with different requirements and characteristics. This paper addresses Factory Automation (FA), in which machines or complete systems require real-time control. Generally, FA represents production chains, where machines are the main elements of the process. Some examples of typical FA applications are assembly, packaging, palletizing, and manufacturing. These processes are critical and, therefore, require strict latency and reliability rates. In particular, cycle times are normally bounded between 0.25-10 ms, and a Packet Loss Rate (PLR) of 10^{-7} - 10^{-9} is required. Moreover, the update time for the devices of the network is between 0.5-50 ms. The size of the cells in FA implies a coverage range up to 100 meters and it is composed of up to 100 nodes. The data packet size is small in comparison with other use cases, normally, up to 300 Bytes, according to [41].

In a typical Industrial Wireless Sensor Network (IWSN), the nodes can be split into two groups: access points/gateways (AP) and sensor/control nodes. The AP node is the master node of the network. Among others, the main roles of the AP are network synchronization, collection of the sensed information by the sensors, and distribution of information and commands to the sensor nodes. On the other hand, sensor/control nodes are considered slave nodes that could have transceiver capacities. In general, they receive the information and the commands from the AP and they send back their sensed data within the cycle time [26].

The rest of the paper will consider a specific FA use case that involves three types of services. The first one is the transmission of Channel Quality Indicators (CQI)-based feedback from the nodes to the AP in an exclusive uplink channel. The second one is a broadcast service implemented from the AP to all the nodes in the network to deliver the critical information, such as synchronization, superframe starting signaling, and emergency alarms. It is important to highlight that the broadcast service is also used to deliver one of the widespread multimedia industrial applications: streaming video surveillance service. In this case, the objective of the broadcast service is to guarantee the safety of the workers and the live video-monitoring of the machines. The variable payload size of this service contains mission-critical information for a large number of nodes and it can be up to 300 Bytes. Then, the third one is a unicast service where the AP transmits a dedicated packet to each slave with the specific sensing commands and/or high-quality images/videos of the task that is carrying out. These packets (up to 64 Byte long) are also variable in payload size but contain less information than the broadcast packets. The main difference between the broadcast and the unicast services is the reliability requirement. While the proposed broadcast service is considered mission-critical, the unicast service is a non-critical and loss and delay-tolerant service (i.e., also known as Best Effort service). In order to deploy both multimedia video services, broadcast and unicast, a capacity of hundreds of kbps or a few Mbps per device has to be guaranteed according to the Monitoring & Diagnostic use case defined in [35].

The wireless modeling will be based on the proposals made by the 802.15.4 standardization groups [32] and the experiments carried out on [12]. Specifically, the so-called Channel Model 7 (CM7), with a coherence time (T_{Coh}) of 30 ms has been selected. Up to 100 nodes are assumed randomly distributed through the cell, with the AP installed in an optimal position in the middle of the hall, where Line-of-Sight (LOS) conditions dominate. The system is based on a 10/20 MHz channel bandwidth in the 2.4 GHz ISM band, -90dBm is the thermal noise power and the transmitted power is set to 10 dBm. The rest of the network parameters are summarized in Table 3.

Table 3 Network parameters of the industrial use case

Parameter	Value
Application	FA
Dimensions	80 x 40 m
# Nodes	Up to 100
Broadcast payload	Up to 300 Byte
Unicast payload	Up to 64 Byte
Channel Model	CM7
Shadowing	μ = 0 σ = 6 dB
Bandwidth	10/20 MHz
Frequency	2.4 GHz
SCS	60 kHz
Thermal Noise	-90 dBm
T_{Cycle}	Up to 5 ms
Tx Power	10 dBm
T_{Coh}	30 ms

4. RRM algorithm

In this scenario, the AP performs the role of a Next Generation NodeB (gNB) and delivers the information with different requirements and characteristics to the nodes integrated into the network. The information flow goes from the AP to each of the nodes in the network. Therefore, as it is a downlink communication network, the Physical Downlink Shared Channel (PDSCH) has been used. In this section, there are presented different RRM algorithms with different optimization goals for both TDMA and NOMA. The algorithms presented in this section distribute the available RBs between the broadcast services and the dedicated unicast services.

4.1. RRM management using TDMA

In this first approach, TDMA is used as the multiplexing technology for conveying broadcast and unicast services simultaneously. A summary of the designed RRM algorithm for TDMA is shown in Algorithm 1 and a detailed explanation of each step is presented below.

Algorithm 1: TDMA-based RRM algorithm (I)
Step 1: Receive CQI feedback from the users. Step 2: Calculate the available RBs. Step 3: Configure the broadcast service. for r > 0   Step 4: Configure the unicast service of one user and update the amount of available RBs. end for

First, in Step 1, the AP collects the CQIs that each device has transmitted in the PUCCH (Physical Uplink Control Channel). Each user informs the AP with the CQI, which is the maximum Modulation and Code Scheme (MCS) that could correctly decode from the possible values presented in Table 5.1.3.1-2 in [3] according to its instantaneous SNR value.

Then, Step 2 is in charge of calculating the amount of RBs that are available for sharing among the users. Eq. (3) will be used to obtain N_{RB}:

$$N_{RB} = BW \cdot N_{Symbols} \cdot N_{Slots} \quad (3)$$

where N_{RB} is the total amount of available RBs, BW is the number of RBs forming the channel bandwidth (e.g., 25 RBs for a 20 MHz channel), $N_{Symbols}$ is the number of symbols in a time slot (i.e., 14 symbols) and N_{Slots} is the number of slots gathered for the RRM management. The calculation of the total number of slots is based on the acceptable maximum cycle time (T_{Cycle}). Taking into account that using 60 kHz as SCS the duration of a time slot is 0.25 ms (see Table 1), the number of time slots used for each iteration of the RRM algorithm is calculated as follows:

$$N_{Slots} = \frac{T_{Cycle}}{0.25} \tag{4}$$

where T_{Cycle} is expressed in ms.

Once, the amount of available RBs is calculated, the broadcast service is configured in Step 3. Taking into account that the broadcast service is critical with mandatory reliability requirements, the lowest MCS in Table 5.1.3.1-2 in [3] is used. Then, the amount of RBs that have to be allocated for the broadcast service is calculated in two steps. First, the required number of subcarriers is calculated as follows:

$$Subcarriers_{Br} = \frac{Payload\ Size_{Br}}{Efficiency\ MCS0} \tag{5}$$

where the $Payload\ Size_{Br}$ is expressed in bits and the $Efficiency\ MCS0$, as defined in Table 5.1.3.1-2 in [3], is 0.2344 bps/Hz. Then, the number of RBs is calculated based on $Subcarriers_{Br}$ as follows:

$$RB_{Br} = round\left(\frac{Subcarriers_{Br}}{12}\right) \tag{6}$$

where 12 represents the number of subcarriers within an RB and the round function is used to obtain the next integer value of the division result. After obtaining the number of RBs allocated to the broadcast service, the value is subtracted from the total number of available RBs:

$$r = N_{RB} - RB_{Br} \tag{7}$$

Then, before configuring any unicast services, it has to be checked if there are still available RBs to allocate (i.e., $r > 0$). If the statement is true, the first unicast user is configured in Step 4 by calculating the RBs allocated to the first unicast user. The set of unicast users can be expressed as $U = \{u_1, u_2, \dots, u_N\}$. This algorithm is oriented to maximize the network capacity or the number of served users, so, the users with the highest requested MCS values are served first since they will require a small portion of the available bandwidth. Therefore, as in Eq. (5), first, the number of subcarriers allocated to the first user is calculated:

$$Subcarriers_{u_1} = \frac{Payload\ Size_{u_1}}{Efficiency_{u_1}} \tag{8}$$

In this case, $Efficiency_{u_1}$ is the efficiency related to the MCS requested by u_1. Then, the amount of RBs dedicated to u_1 are calculated as in Eq. (6):

$$RB_{u_1} = round\left(\frac{Subcarriers_{u_1}}{12}\right) \tag{9}$$

After that, the amount of available RBs is updated by subtracting RB_{u_1}:

$$r = r - RB_{u_1} \tag{10}$$

If the amount of allocable RBs is positive (i.e., $r > 0$), the next unicast user is configured following Eq. (8) and Eq. (9) to calculate the amount of dedicated RBs and the available RBs (i.e., r) is updated again as in Eq. (10). This looping process is executed until all the RBs in N_{RB} are allocated. If all the users forming the network have been served and there are still RBs to allocate, users are served again in order to increase the network capacity and reliability.

Algorithm 1 implies that the T_{Cycle} is fixed and the number of users served varies depending on the input parameters. In order to provide a different perspective, another algorithm has been developed, where the number of users that have to be served is fixed and the required T_{Cycle} to serve all of them is calculated. A summary of the second RRM algorithm is presented in Algorithm 2 and an in-depth analysis of each step is provided below.

Algorithm 2: TDMA-based RRM algorithm (II)
Step 1: Receive CQI feedback from the users. Step 2: Configure the broadcast service and calculate the required RBs. for n ≤ N Step 3: Configure the unicast service of the user u_N and update the amount of required RBs. end for Step 4: Calculate the required T_{Cycle}.

The starting point is again the reception of the CQI feedback of each user from the PUCCH channel. In this case, the AP has to manage the RBs that are necessary to offer the whole set of users (i.e., $U = \{u_1, u_2, \dots, u_N\}$) the critical broadcast service and the non-critical unicast services. Then, in Step 2, the broadcast service is configured by using the most robust MCS value available and the required amount of RBs (i.e., RB_{Br}) is calculated following Eq. (5) and Eq. (6).

When the broadcast service is configured, unicast services are individually configured. In this case, it is mandatory that all the users have to be served. In consequence, there is no preference in the serving order for users with high MCS. To configure each unicast service and calculate the required amount of RBs (i.e., RB_{u_n}), as in Algorithm 1, Eq. (8) and Eq. (9) are followed.

Once the allocation of RBs for each unicast user is carried out, the required T_{Cycle} is estimated. First, the total amount of RBs required to offer broadcast and unicast services (i.e., RB_{Total}) is calculated as follows:

$$RB_{Total} = RB_{Br} + RB_{u_1} + RB_{u_2} + \cdots + RB_{u_N} \tag{11}$$

Then, the number of slots required to organize RB_{Total} is calculated from Eq. (3) and, finally, taking into account the duration of each slot (see Table 1), the required T_{Cycle} is calculated from Eq. (4).

4.2. RRM management using NOMA

In this section, two algorithms are presented in order to manage the RRM of 5G NR using NOMA. In this case, the broadcast service is delivered in the Upper Layer (UL) of the NOMA signal, in order to provide high reliability, and the unicast services, as they are considered non-critical, are transmitted in the Lower Layer (LL). A summary of the first algorithm is presented in Algorithm 3, where as in Algorithm 1, the maximum T_{Cycle} is fixed.

Algorithm 3: NOMA-based RRM algorithm (I)
Step 1: Receive CQI feedback from the users. Step 2: Calculate the available RBs. Step 3: Configure the set of possible Δ values. for m ≤ M Step 4: Configure the broadcast service for r > 0 Step 5: Configure the unicast services of one user and update the amount of available RBs. end for end for Step 6: Select the optimum solution

The first two steps of the NOMA-based algorithm are the same as in Algorithm 1, that is, the CQI feedback from the users is received in the PUCCH channel and the amount of RBs that are available for sharing among the users (i.e., N_{RB}) is calculated.

Then, in Step 3 the set of possible Δ values is defined as $\Delta = \{\Delta_1, \Delta_2, \dots, \Delta_M\}$. In order to define the set of values, the maximum assumable degradation in the UL of NOMA (γ) has to be managed. γ represents the maximum SNR difference between the most robust configuration using and not using NOMA and it can be calculated with the following expression:

$$\gamma < SNR_{MCS0_{UL}} - SNR_{MCS0_{SL}} \quad (12)$$

γ implies a tradeoff between reliability and capacity. In fact, the closer is γ to zero, the more similar are the performance of UL and broadcast service in TDMA (see the previous section). However, in order to achieve γ close to zero, high Δ values are required and those cases suppose a considerable reduction in the capacity delivered by the LL. Therefore, a balance has to be sought between the acceptable reliability losses in the UL in exchange for the capacity performance of the LL. Once, γ and $SNR_{MCS0_{UL}}$ are defined, the minimum value of Δ that guarantees that γ is accomplished is searched starting from Δ = 0 dB and with steps of 0.5 dB. Then, the set of possible Δ values is completed with values up to Δ = -20 dB and steps of 0.5 dB.

Afterward, an iterative process starts in order to test the RRM performance of each value in Δ. First, the broadcast service is configured, which is delivered in the UL 100% of the time in the whole RF bandwidth. The broadcast service is configured with the lowest MCS value available in Table 5.1.3.1-2 in [3].

Once the broadcast service is defined, the same process as in the case of Algorithm 1 is performed, where unicast users are individually served until all the available RBs (i.e., $r = 0$) are allocated. Then, the configuration of the users is saved and the next value in the set of possible injection levels is tested.

When all the possible values in Δ are evaluated the best solution has to be selected. To do so, different rules could be applied, in this work, two are going to be evaluated: the solution that maximizes the capacity of the unicast services and the solution that maximizes the number of served unicast users. This is one of the main differences in comparison with the TDMA-based algorithm. While the RB management cannot be optimized in TDMA since, there is one possible configuration under the predefined conditions, using NOMA several configurations are possible under the same conditions and the best performing one can be selected.

As in TDMA, another version of the NOMA-based RRM algorithm has been developed, where the goal of the algorithm is to find the minimum T_{Cycle} for a fixed number of unicast users. A summary of the second algorithm is presented in Algorithm 4 and an in-depth analysis of each step is provided below.

Algorithm 4: NOMA-based RRM algorithm (II)

Step 1: Receive CQI feedback from the users.
Step 2: Configure the set of possible Δ values.
for m ≤ M
  Step 3: Configure the broadcast service and calculate the required RBs.
  for n ≤ N
  Step 4: Configure the unicast service of the user u_N and update the amount of required RBs.
  end for
  Step 5: Calculate the required T_{Cycle}.
end for
Step 6: Select the optimum solution

This second algorithm starts similar to Algorithm 1 since first, the CQI feedbacks are collected forming the set of users ($U = \{u_1, u_2, \dots, u_N\}$) and in a second step the set of possible injection levels ($\Delta = \{\Delta_1, \Delta_2, \dots, \Delta_M\}$) is calculated according to the maximum assumable degradation in the UL of NOMA (γ). Then, an iterative loop is started to evaluate each possible injection level value.

The first step in the iterative loop is to configure the critical broadcast service and calculate the required amount of RBs (i.e., RB_{Br}). To do so, as in TDMA, Eq. (5) and Eq. (6) are followed by using the most robust configuration in order to ensure a high-reliability rate. After that, each user in the set of users is managed and the required amount of RBs (i.e., RB_{u_n}) is calculated following Eq. (8) and Eq. (9). In contrast to Algorithm 2, in this case, the number of required RBs are separately aggregated depending on the type of service. Therefore, the total amount of RBs required to manage all the unicast services is calculated:

$$RB_{Uni} = RB_{u_1} + RB_{u_2} + \cdots + RB_{u_N} \quad (13)$$

Then, the corresponding T_{Cycle} for each value in the set of possible injection levels is calculated in Step 5. In this case, taking into account the characteristics of NOMA, the number of RBs allocated for the broadcast services (i.e., RB_{Br}) are exclusively used in the UL, while the number of RBs allocated for the unicast services (i.e., RB_{Uni}) are used in the LL. Therefore, the total amount of required RBs is no longer obtained by adding both values, but it is defined by the maximum of both values:

$$RB_{Total} = \max(RB_{Br}, RB_{Uni}) \quad (14)$$

Once the required amount of RBs is calculated, the process as in TDMA is followed: first, the number of slots is calculated from Eq. (3) and, second, the required T_{Cycle} is calculated from Eq. (4).

Finally, when all the values in Δ are analyzed, the best solution is chosen in Step 6 by selecting the one implying the lowest T_{Cycle} value.

5. Results

This section presents and compares the results obtained by executing the algorithms of section 4 in the use case described in section 3.

5.1. Capacity analysis

First, the capacity obtained by using Algorithm 1 and Algorithm 3 is compared. The comparison will be performed by first finding the configuration that maximizes the capacity of the unicast services. The evaluation will be based on three capacity metrics: the throughput of the broadcast service, the Aggregate Data Rate of the unicast services, and the total ADR.

Firstly, the throughput obtained for the broadcast service (Th_{Br}) is shown in Fig. 3 for different T_{Cycle} values. In this case, the payload of the broadcast packet is 100 Bytes and the network is composed by 100 devices. It is important to highlight, that NOMA-based RRM outperforms TDMA with higher capacity values. What is more, the difference between technologies increases as the T_{Cycle} increases. In fact, the maximum gain case appears for one millisecond T_{Cycle}, where NOMA offers a gain of 3.36 Mbps in the 20 MHz channel (1.33 Mbps in the 10 MHz channel). On the other hand, the maximum assumable SNR degradation (γ) does not modify the results of the broadcast capacity since the parameter affected is the availability of the broadcast service, which represents the ratio between the number of correctly served users and the total number of users. In this case, very similar availability values are obtained in all the cases. In fact, although the highest availability value is obtained in the TDMA case with 0.999812, NOMA configurations offer very similar performance: 0.999707 for γ = 0.5 dB and 0.999611 for γ = 1 dB. Finally, it should also be highlighted that NOMA offers a constant capacity value even if the T_{Cycle} value is modified, whereas TDMA offers variable capacity. In fact, the constant capacity achieved by NOMA is the maximum achievable capacity for the MCS0. Moreover, the capacity variation of TDMA is due to the RB sharing between broadcast and unicast services. Once the predefined broadcast payload has been managed, RBs are allocated for unicast services, and, therefore, the more RBs are allocated for unicast services, the lower is the broadcast capacity. Based on the broadcast service throughput results obtained and taking into account that the throughput ranges between 1.43 Mbps and 4.22 Mbps, the broadcast service could correctly address the transmission of video streaming content, as established in [35]. Furthermore, based on Fig. 3, the throughput obtained when NOMA is used in the 20 MHz channel could be used for higher quality services such as 4k content to receivers on small screens (displays, mobiles devices).

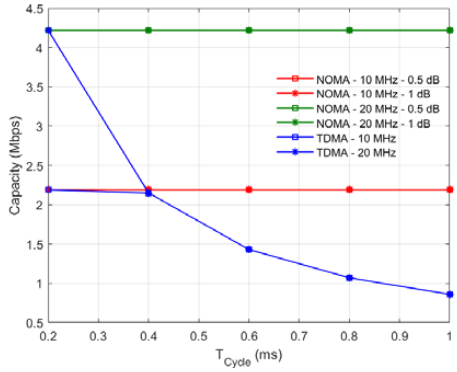


Fig. 3 Broadcast service throughput (Th_{Br}) for different T_{Cycle} values

Afterward, the ADR of the unicast services (ADR_{Un}) is shown in Fig. 4, calculated as the sum of the capacities obtained by serving all the possible unicast communications in the network within a specific T_{Cycle} (i.e., up to 100 devices). In this case, in order to consider a device served, 32 Bytes have to be delivered. Unicast level results are more variable than the results obtained for the broadcast services. First, the best technology for the proposed cases is affected by the channel bandwidth. In general, NOMA is better in the small bandwidth channel (i.e., 10 MHz) and TDMA performs better when the bandwidth is increased (i.e., 20 MHz). However, a positive aspect of NOMA configurations is that always is able to offer the capacity to the unicast devices, while, on the contrary, TDMA does not, as it can be seen in the lowest T_{Cycle} values. As opposed to the broadcast throughput, variations on the maximum assumable SNR degradation (γ) have direct effect on the unicast service capacity. In particular, the capacity performance when γ = 0.5 dB is assumed is considerably worsen in comparison with γ = 1 dB. Based on the unicast ADR results shown in Fig. 4, although high throughput values are obtained (close to 90 Mbps), the throughput per node ratio ranges between 1-0.5 Mbps/node since as will be shown in the next subsection, not all the unicast devices are served. This throughput rate will be enough to offer a standard quality video streaming service or high definition images, as established in [35]. In addition, it should be noted that this service will have assumable latency values (cycle time values below one ms). Then, if images or video content with higher definition are required (high challenging multimedia services), some network parameters could be modified such as reducing the number of nodes or increasing the bandwidth without any negative impact on the overall performance.

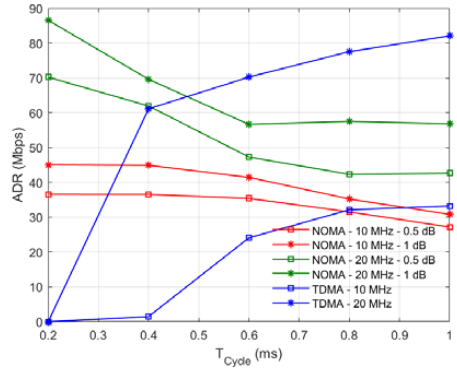


Fig. 4 Unicast service ADR (ADR_{Un}) for different T_{Cycle} values

Once broadcast and unicast services capacity performance has been evaluated, the combined ADR is presented in Fig. 5, as the sum of all the received capacities, broadcast and unicast, within a specific T_{Cycle} value. The same parameters as in the previous case have been used. These results present NOMA as a relevant candidate for RRM managing algorithms since it provides a considerable increase in the total ADR

of the network. In fact, again, the best cases are located in T_{Cycle} = 1 ms, where NOMA is close to triplicate and to duplicate the ADR of TDMA in the 20 MHz and 10 MHz channel, respectively. In consequence, taking into account that the same number of resources has been share in NOMA and TDMA, NOMA demonstrates that is a highly spectrum efficient technique.

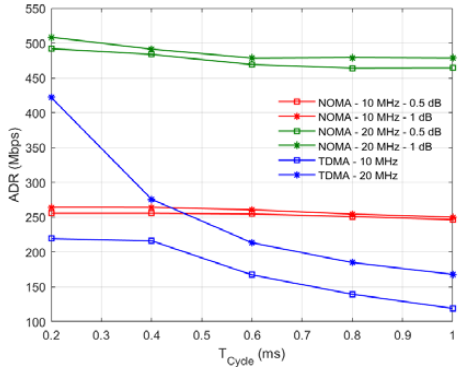


Fig. 5 Total ADR for different T_{Cycle} values

In addition to the T_{Cycle}-based analysis, capacity performance can also be analyzed from the payload size perspective. First, the impact of the broadcast service packet size is summarized in Table 4. In this case, an intermediate case has been selected: T_{Cycle} of 0.6 ms and up to 1 dB of SNR degradation for NOMA configurations. Results show that the payload of the broadcast service does not affect the RRM management of NOMA since the same results are obtained for each metric. However, TDMA is affected by the broadcast service packet size. When the payload of the broadcast services increases, the throughput of the broadcast content increases, and the ADR of the unicast services decreases. In particular, when the 10 MHz channel is used and the broadcast payload is incremented to 200 Bytes, there are no available RBs to manage the unicast services and, therefore, no capacity can be achieved for the unicast devices.

Table 4 Impact of the broadcast service packet size on the capacity

	10 MHz				20 MHz			
	NOMA		TDMA		NOMA		TDMA	
	100 B	200 B	100 B	200 B	100 B	200 B	100 B	200 B
Th_{Br} (Mbps)	2.19	2.19	1.43	2.19	4.22	4.22	1.43	2.86
ADR_{Un} (Mbps)	41.44	41.44	24.09	0.00	56.64	56.64	70.24	40.35
ADR (Mbps)	260.76	260.76	167.21	219.36	478.40	478.40	213.37	326.09

Finally, by using the same parameter configuration, the impact of the unicast service packet size on the capacity is shown in Table 5. From the capacity perspective, an increment of the unicast packet size involves an increment on the unicast ADR for both technologies, while the throughput of the broadcast services remains constant since unicast services are configured after the broadcast content configuration. Although the increase is more evident in the 20 MHz channel, in the case of the 10 MHz channel the capacity also increases. In particular, the performance of TDMA in the 10 MHz channel is almost constant since the increment in the payload size is compensated with an equivalent decrease in the number of served unicast users.

Table 5 Impact of the unicast service packet size on the capacity

	10 MHz				20 MHz			
	NOMA		TDMA		NOMA		TDMA	
	32 B	64 B	32 B	64 B	32 B	64 B	32 B	64 B
Th_{Br} (Mbps)	2.19	2.19	1.43	1.43	4.22	4.22	1.43	1.43
ADR_{Un} (Mbps)	41.44	45.07	24.09	24.12	56.64	77.65	70.24	81.61
ADR (Mbps)	260.76	264.38	167.21	167.24	478.40	499.41	213.37	224.74

5.2. Analysis of the number of users

This subsection compares the performance of TDMA and NOMA in the RRM layer in terms of the number of unicast users served. First, the evolution of the served unicast users for different T_{Cycle} values is shown

in Fig. 6. In general, these curves show two parts: the fast increment and the stabilization. As it is shown in the figure, TDMA shows a fast behavior and arrives in the stabilization phase faster than NOMA. However, as in Fig. 4, not all the configurations are useful since in some cases, TDMA is not able to offer unicast services. Concerning NOMA, the maximum assumable SNR degradation (γ) is also critical in this analysis since the higher is γ, the higher is the number of served users. In order to enhance the comparison, if the benchmark is set in 90% of unicast users served, the best RRM algorithm is NOMA with γ = 1 dB (i.e., 0.4 ms and 0.76 ms are required for 20 MHz and 10 MHz channel bandwidths, respectively) and TDMA performs better than NOMA with γ = 0.5 dB. However, if more strict benchmarks are set, results may vary in favor of TDMA.

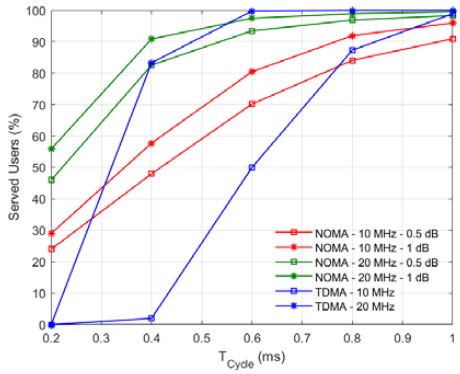


Fig. 6 Served unicast users for different T_{Cycle} values

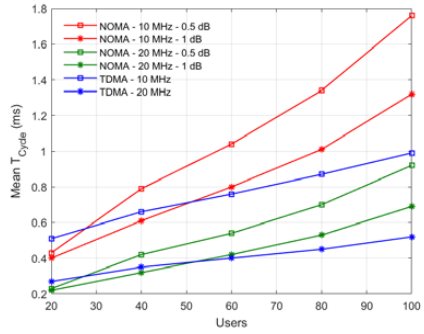
Then, the analysis of the number of users served can go deeper by taking into account the influence of the payload of the broadcast and unicast services (See Table 6). To do so, an intermediate set of parameters has been selected, where T_{Cycle} is 0.6 ms and up to 1 dB of SNR degradation for NOMA is acceptable. In general, the number of users served with TDMA is more dependent on the payload sizes than in the case of NOMA. In particular, the results obtained with NOMA are only affected by the payload of the unicast content. Table 6 shows that the performance obtained with NOMA remains constant no matter the size of the broadcast service packet size. On the other hand, TDMA results are affected by both payload sizes, broadcast, and unicast payloads. In addition, the number of unicast users served decreases considerably fast, while the packet sizes are increased. In fact, in half of the cases, there are not enough resources to offer unicast services, making NOMA a better candidate to guarantee a minimum number of unicast services.

Table 6 Impact of the packet size on the number of served users (%)

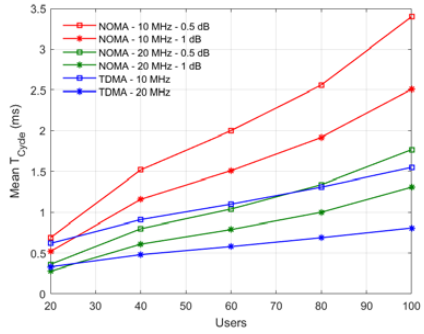
Broadcast packet size	100 B		200 B		300 B	
Unicast packet size	32 B	64 B	32 B	64 B	32 B	64 B
NOMA - 10 MHz	80.45	47.98	80.45	47.98	80.45	47.98
TDMA - 10 MHz	49.98	25.00	0.00	0.00	0.00	0.00
NOMA - 20 MHz	97.47	82.85	97.47	82.85	97.47	82.85
TDMA - 20 MHz	99.76	84.83	82.46	44.99	0.00	0.00

5.3. Cycle time duration analysis

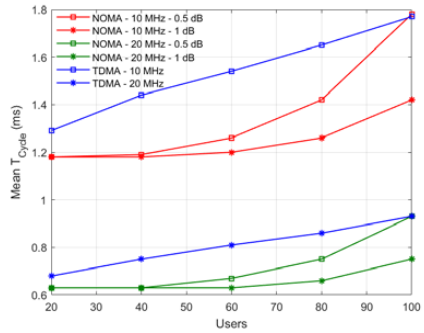
Finally, the T_{Cycle} is used in this subsection as a metric to describe how fast can Algorithm 2 and Algorithm 4 manage all the users in the network. The number of users that compound the network and that have to be mandatorily served is varied to obtain cycle time measurements under different conditions. Fig. 7 shows the mean minimum T_{Cycle} for different broadcast and unicast payload sizes. In general, there is not a clear winning technology and the selection of the best performing multiplexing technology depends on the parameter configuration. For example, an increment in the broadcast payload involves a better performance of NOMA, while increasing the unicast packet size TDMA outperforms NOMA. Regarding the number of users in the network, TDMA performs better in very dense configurations, while NOMA provides lower T_{Cycle} values when the network is not so crowded. In any case, all the simulations carried out in this section with different parameter configurations guarantee that the minimum T_{Cycle} value specified for FA environments (i.e., below 10 ms) is accomplished.



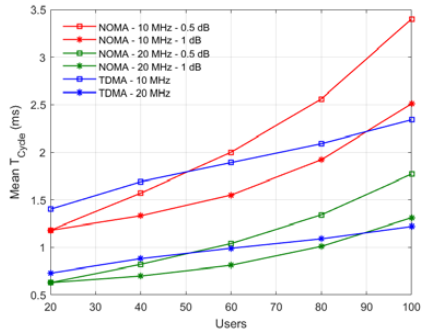
(a) Broadcast packet size 100 B and unicast packet size 32 B



(b) Broadcast packet size 100 B and unicast packet size 64 B



(c) Broadcast packet size 300 B and unicast packet size 32 B



(d) Broadcast packet size 300 B and unicast packet size 64 B

Fig. 7 Mean T_{Cycle} values varying the number of users for different configurations.

5.4. Result summary

The objective of this subsection is to summarize the obtained results in previous sections and to select the most suitable multiplexing technique according to the already evaluated parameters. Table 7 summarizes results by highlighting the best performing multiplexing technique in relation to different parameter ranges. In general, although there are several ranges where similar performance is obtained with both techniques, NOMA is the best technique in more ranges. In particular, NOMA offers better performance for achieving the first two goals (i.e., maximize capacity and maximize served users), while in the case of minimizing the T_{Cycle} there is not a clear winning technique. It is also important to point out that there are cases, where it is not evident to decide which multiplexing technique performs better because there is another more critical parameter. For example, in the case of minimizing the T_{Cycle} the channel bandwidth is not a critical factor since the payload of the broadcast service is more critical and determines which the best performing alternative is. On the other hand, it is also hard to determine the best multiplexing technique when the goal is to maximize the capacity since, as shown in section 5.1, there is not a single metric. For instance, in the range of broadcast payload $\leq$ 100 Byte both technologies are selected because NOMA performs better in the broadcast throughput and the total ADR, whereas TDMA offers better results in the unicast ADR metric.

Table 7 Selection of the best performing multiplexing technique according to different parameters and optimization goals

Goal	Parameter	Range	Best Performant Technique
Maximize capacity	Broadcast payload	≤ 100 Byte	TDMA/NOMA
		> 100 Byte	NOMA
	Unicast payload	≤ 32 Byte	TDMA/NOMA
		> 32 Byte	TDMA/NOMA
	T_{Cycle}	≤ 0.5 ms	TDMA/NOMA
		> 0.5 ms	NOMA
	BW	≤ 10 MHz	NOMA
		> 10 MHz	TDMA/NOMA
Maximize served users	Broadcast payload	≤ 100 Byte	TDMA/NOMA
		> 100 Byte	NOMA
	Unicast payload	≤ 32 Byte	TDMA/NOMA
		> 32 Byte	TDMA/NOMA
	T_{Cycle}	≤ 0.5 ms	NOMA
		> 0.5 ms	TDMA
	BW	≤ 10 MHz	NOMA
		> 10 MHz	TDMA/NOMA
Minimize T_{Cycle}	Broadcast payload	≤ 100 Byte	TDMA
		> 100 Byte	NOMA
	Unicast payload	≤ 32 Byte	TDMA/NOMA
		> 32 Byte	TDMA/NOMA
	Number of users	≤ 40	NOMA
		> 40	TDMA
	BW	≤ 10 MHz	TDMA/NOMA
		> 10 MHz	TDMA/NOMA

6. Conclusions

In this paper, the use of P-NOMA to manage the resource allocation layer has been proposed and evaluated using different RRM algorithms. In addition, the resource allocation and the obtained results have been compared with TDMA. In order to make a fair comparison, different affecting parameters have been identified and several RRM algorithms have been designed according to different optimization goals. Results indicate that there is not a clear winning multiplexing technology and the most suitable election totally depends on the network parameters. However, NOMA techniques are the most suitable option when the optimization goal is to maximize the network capacity or the amount of served unicast users.

This work opens the door to several future research avenues. On the one hand, the analysis presented in this paper could go deeper by carrying out more simulations in order to widen the ranges of each parameter or in different FA environments to confirm or adapt the parameter ranges provided in this paper. Moreover, different RRM algorithms could be designed with combined optimization goals, such as availability and capacity simultaneously. On the other hand, a solution based on the combination of NOMA and T/FDMA techniques could be sought in order to improve the spectral efficiency and to distribute better the gain of NOMA between broadcast and unicast services. This solution, theoretically, should increment the granularity of the configuring options and, therefore, the best performing solution for each case could be achieved. Finally, this kind of RRM solution could be combined with a 5G NR PHY/MAC level transceiver to simulate other aspects of a 5G-based communication system. In addition to the metrics used in this work (i.e., ADR or T_{Cycle}), metrics related to the PHY and the MAC layers could be used, such as user plane latency or Packet Error Rate (PER).

Acknowledgments

This work was supported in part by the Basque Government (Project IOTERRAZ) under Grant KK-2019/00046 ELKARTEK 2019, Grant IT1234-19, and the PREDOC Grant Program PRE_2020_2_0105, and in part by the Spanish Government (Project PHANTOM) under Grant RTI2018-099162-B-I00 (MCIU/AEI/FEDER, UE).

References

[1] 3GPP TR 38.802 V14.2.0 (2017) Study on New Radio (NR) Access Technology Physical Layer Aspects, 3rd Generation Partnership Project. Tech Rep.

[2] 3GPP TS 38.201 V15.0.0 (2018) NR; Physical Layer; General Description, 3rd Generation Partnership Project. Tech Spec.

[3] 3GPP TS 38.214 v15.3.0 (2018) NR; Physical layer procedures for data (Release15), 3rd Generation Partnership Project. Tech Rep.

[4] Aktas I et al (2017) LTE evolution—Latency reduction and reliability enhancements for wireless industrial automation. 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), IEEE, Montreal, pp. 1-7. https://doi.org/10.1109/PIMRC.2017.8292603

[5] Al-Turjman, F., Alturjman, S. 5G/IoT-enabled UAVs for multimedia delivery in industry-oriented applications. Multimed Tools Appl 79, 8627–8648 (2020). https://doi.org/10.1007/s11042-018-6288-7.

[6] Anwar M, Xia Y, Zhan Y (2016) TDMA-Based IEEE 802.15.4 for Low-Latency Deterministic Control Applications. IEEE Transactions on Industrial Informatics 12(1): 338-347. https://doi.org/10.1109/TII.2015.2508719

[7] Ao W, Cui Y, Yao Q, Li X (2019) Design and Implementation of Distributed Video Equipment Monitoring and Management System. IEEE International Conference on Industrial Internet (ICII), IEEE, Orlando, pp. 423-427. https://doi.org/10.1109/ICII.2019.00078

[8] Bergmans P, Cover T (1974) Cooperative broadcasting. IEEE Transactions on Information Theory 20(3): 317–324. https://doi.org/10.1109/TIT.1974.1055232

[9] Blanco-Novoa O, Fernández-Caramés TM, Fraga-Lamas P, Vilar-Montesinos MA (2018) A Practical Evaluation of Commercial Industrial Augmented Reality Systems in an Industry 4.0 Shipyard. IEEE Access 6:8201-8218. https://doi.org/10.1109/ACCESS.2018.2802699

[10] Coll JF, Chilo J, Slimane B (2012). Radio-frequency electromagnetic characterization in factory infrastructures. IEEE Transactions on Electromagnetic Compatibility 54(3): 708-711. https://doi.org/10.1109/TEMC.2012.2197753

[11] Cover T (1972) Broadcast channels. IEEE Transactions on Information Theory 18(1): 2–14. https://doi.org/10.1109/TIT.1972.1054727

[12] Croonenbroeck R, Underberg L, Wulf A, Kays R (2017) Measurements for the development of an enhanced model for wireless channels in industrial environments. 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), IEEE, Rome, pp. 1-8. https://doi.org/10.1109/WiMOB.2017.8115764

[13] Drath R, Horch A (2014) Industrie 4.0: Hit or Hype? [Industry Forum]. IEEE Industrial Electronics Magazine 8(2): 56-58. https://doi.org/10.1109/MIE.2014.2312079

[14] Egger-Lampl S, Gerdenitsch C, Deinhard L, Schatz R, Hold P (2019) Assembly Instructions with AR: Towards measuring Interactive Assistance Experience in an Industry 4.0 Context. Eleventh International Conference on Quality of Multimedia Experience (QoMEX), IEEE, Berlin, pp. 1-3. https://doi.org/10.1109/QoMEX.2019.8743266

[15] Fraga-Lamas P, Fernández-Caramés TM, Blanco-Novoa O, Vilar-Montesinos MA (2018) A Review on Industrial Augmented Reality Systems for the Industry 4.0 Shipyard. IEEE Access 6:13358-13375. https://doi.org/10.1109/ACCESS.2018.2808326

[16] Ghini V, Casadei M, Borgo FD, Vincenzi N, Prandi C, Mirri S (2019) Industry 4.0 and Video Monitoring: a Multidimensional Approach Based on MPEG-DASH. 16th IEEE Annual Consumer Communications & Networking Conference (CCNC), IEEE, Las Vegas, pp. 1-6. https://doi.org/10.1109/CCNC.2019.8651683

[17] Gong L, Fast-Berglund A, Johansson B (2021) A Framework for Extended Reality System Development in Manufacturing. IEEE Access 9:24796-24813. https://doi.org/10.1109/ACCESS.2021.3056752

[18] Gosh A, Maeder A, Baker M, Chandramouli D (2019) 5G evolution: View on 5G cellular technology beyond 3GPP release 15. IEEE Access 7:127639 - 127651. https://doi.org/10.1109/ACCESS.2019.2939938

[19] Hassan SM, Ibrahim R, Bingi K, Chung TD, Saad N (2017) Application of wireless technology for control: A WirelessHART perspective. Procedia Computer Science *105*(supplement C): 240-247. https://doi.org/10.1016/j.procs.2017.01.217

[20] Hoshyar R, Wathan FP, Tafazolli R (2008) Novel low-density signature for synchronous CDMA systems over AWGN channel. IEEE Transactions on Signal Processing 56(4):1616–1626. https://doi.org/10.1109/TSP.2007.909320

[21] Hsu C, Kang L, You T, Jhong W (2017) Vision-Based Automatic Identification Tracking of Steel Products for Intelligent Steel Manufacturing. IEEE International Symposium on Multimedia (ISM), IEEE, Taichung, pp. 376-377. https://doi.org/10.1109/ISM.2017.75

[22] Iradier E et al (2020) Using NOMA for Enabling Broadcast/Unicast Convergence in 5G Networks. IEEE Transactions on Broadcasting. https://doi.org/10.1109/TBC.2020.2981759

[23] Iradier E et al (2021) Analysis of NOMA-Based Retransmission Schemes for Factory Automation Applications. IEEE Access 9:29541-29554. https://doi.org/10.1109/ACCESS.2021.3059069
[24] Iradier E, Montalban J, Araniti G, Fadda M, Murroni M (2016) Adaptive resource allocation in LTE vehicular services using LDM. International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), IEEE, Nara, pp. 1-3. https://doi.org/10.1109/BMSB.2016.7521985
[25] Iradier E, Montalban J, Fanari L, Angueira P (2019) On the Use of Spatial Diversity under Highly Challenging Channels for Ultra Reliable Communications. 24th International Conference on Emerging Technologies and Factory Automation (ETFA), IEEE, Zaragoza, pp. 200-207. https://doi.org/10.1109/ETFA.2019.8869055
[26] Iradier E, Montalban J, Fanari L, Angueira P, Seijo O, Val I (2019) NOMA-based 802.11n for Broadcasting Multimedia Content in Factory Automation Environments. International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), IEEE, Jeju, pp. 1-6. https://doi.org/10.1109/10.1109/BMSB47279.2019.8971844
[27] Iradier E, Montalban J, Romero D, Wu Y, Zhang L, Li W (2019) NOMA based 5G NR for PTM Communications. International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), IEEE, Jeju, pp. 1-6. https://doi.org/10.1109/BMSB47279.2019.8971856
[28] Islam SMR, Avazov N, Dobre OA, Kwak K (2017) Power Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges. IEEE Communications Surveys & Tutorials 19(2): 721-742. https://doi.org/10.1109/COMST.2016.2621116
[29] Kumar, M., Tripathi, R. & Tiwari, S. QoS guarantee towards reliability and timeliness in industrial wireless sensor networks. Multimed Tools Appl 77, 4491–4508 (2018). https://doi.org/10.1007/s11042-017-4832-5
[30] Liu Y, Qin Z, Elkashlan M, Ding Z, Nallanathan A, Hanzo L (2017) Nonorthogonal Multiple Access for 5G and Beyond. Proceedings of the IEEE 105(12):2347–2381. https://doi.org/10.1109/JPROC.2017.2768666
[31] M Series (2015) IMT vision—Framework and overall objectives of the future development of IMT for 2020 and beyond. Recommendation ITU-R M.2083-0
[32] Molisch AF et al (2004) IEEE 802.15. 4a channel model-final report. IEEE P802 15(4): 1–41
[33] Montalban J et al (2018) Multimedia Multicast Services in 5G Networks: Subgrouping and Non-Orthogonal Multiple Access Techniques. IEEE Communications Magazine 56(3): 91-95. https://doi.org/10.1109/MCOM.2018.1700660
[34] Montalban J et al (2020) NOMA-Based 802.11n for Industrial Automation. IEEE Access 8:168546-168557. https://doi.org/10.1109/ACCESS.2020.3023275
[35] Montgomery K, Candell R, Liu Y, Hany M (2020) Wireless user requirements for the factory workcell. US Department of Commerce, National Institute of Standards and Technology.
[36] Nawaratne R, Bandaragoda T, Adikari A, Alahakoon D, De Silva D, Yu X (2017) Incremental knowledge acquisition and self-learning for autonomous video surveillance. IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society, IEEE, Beijing, pp. 4790-4795. https://doi.org/10.1109/IECON.2017.8216826
[37] Park SI et al (2016) Low Complexity Layered Division Multiplexing for ATSC 3.0. IEEE Transactions on Broadcasting 62(1): 233-243. https://doi.org/10.1109/TBC.2015.2492459
[38] Pedersen K, Pocovi G, Steiner J, Maeder A (2018) Agile 5G scheduler for improved E2E performance and flexibility for different network implementations. IEEE Communications Magazine 56(3): 210-217. https://doi.org/10.1109/MCOM.2017.1700517
[39] Rondón R, Gidlund M, Landernäs K (2017) Evaluating bluetooth low energy suitability for time-critical industrial iot applications. International Journal of Wireless Information Networks 24(3): 278-290. https://doi.org/10.1007/s10776-017-0357-0
[40] Sachs J, Wikstrom G, Dudda T, Baldemair R, Kittichokechai K (2018) 5G radio network design for ultra-reliable low-latency communication. IEEE network 32(2): 24-31. https://doi.org/10.1109/MNET.2018.1700232
[41] Schulz P et al (2017) Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture. IEEE Communications Magazine 55(2): 70-78. https://doi.org/10.1109/MCOM.2017.1600435CM
[42] Seijo O, Fernández Z, Val I, López-Fernández JA (2018) SHARP: A novel hybrid architecture for industrial wireless sensor and actuator networks. 14th International Workshop on Factory Communication Systems (WFCS), IEEE, Imperia, pp. 1-10. https://doi.org/10.1109/WFCS.2018.8402358
[43] Tramarin F, Vitturi S, Luvisotto M, Zanella A (2016) On the Use of IEEE 802.11n for Industrial Communications. IEEE Transactions on Industrial Informatics 12(5): 1877-1886. https://doi.org/10.1109/TII.2015.2504872
[44] Varghese A and Tandur D (2014) Wireless requirements and challenges in Industry 4.0. International Conference on Contemporary Computing and Informatics (IC3I), Mysore, pp. 634-638. https://doi.org/10.1109/IC3I.2014.7019732

[45] Zhang L et al (2016) Layered-Division-Multiplexing: Theory and Practice. IEEE Transactions on Broadcasting 62(1):216-232. https://doi.org/ 10.1109/TBC.2015.2505408

[46] Zhang L et al (2019) Layered-Division-Multiplexing for High Spectrum Efficiency and Service Flexibility in Next Generation ATSC 3.0 Broadcast System. IEEE Wireless Communications 26(2): 116–123. https://doi.org/10.1109/MWC.2018.1800092

[47] Zhang L et al (2020) Using non-orthogonal multiplexing in 5G-MBMS to achieve broadband-broadcast convergence with high spectral efficiency. IEEE Transactions on Broadcasting. https://doi.org/10.1109/TBC.2020.2983563

3.2.6 Journal paper J7

This subsection presents a journal paper related with Contribution 2. The full reference of the paper is presented below:

- E. Iradier and J. Montalban, "Komunikazioak 4.0 Industrian," accepted manuscript for publication in Elhuyar Aldizkaria.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Online journal paper indexed in Elhuyar Aldizkaria
- Area: Multidisciplinary

elhuyar
zientzia eta teknologia

Komunikazioak 4.0 Industrian

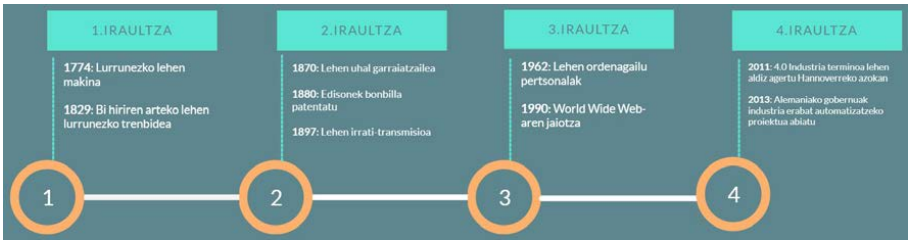
Eneko Iradier, doktoretza aurreko ikertzailea Bilboko Ingeniaritza Eskolan (EHU)

Jon Montalban, irakaslea Gipuzkoako Ingeniaritza Eskolan (EHU)

Azken mendeetan, industria-iraultzek markatu dute garapen teknologikoaren erritmoa. Une honetan, laugarren industria-iraultza (4.0 Industria), garatzen ari da. Bereziki, komunikazio-sistemak dira prozesu honen gidariak. Baina, nola eraldatu beharko dira komunikazio-sistema tradizionalak? Nolakoa izango da etorkizuneko industria?

Iraganera begira

Egokitzeko eta eboluzionatzeko gaitasuna da gizakia bereizten duen ezaugarrietako bat. Hori dela eta, historian zehar hiru industria-iraultza bizi izan ditugu sektore espezifikoen garapenaren ondorioz. Zehazki, lehenengo industria-iraultza mekanizazioaren garapenak bultzatu zuen; bigarrena, energia elektrikoaren ekoizpen masiboak eta industria kimiko eta automobilistikoen garapenak eta hirugarrena, mikroelektronika eta informazioaren teknologiak ekoizpen-kateetan sartzeak. Gaur egun, aurrerapen informatiko ugariek, tresna digital berrien garapenak eta, batez ere, makinen eta prozesuen automatizazioak laugarren industria-iraultzaren hasieran jartzen gaitu, hots, 4.0 Industria izenez ezagunagoa den iraultzaren hasieran. Laugarren industria iraultza honen helburua lantegi adimendunen edo 'smart-industry' delakoen garapena da, modu autonomo eta adimentsuan planifikatu, aurreikusi, kontrolatu eta ekoizteko gaitasuna izango lukeena, eta hortaz, funtzionamendu eta ekoizpen-kateari balio handiagoa emango liokeena.



1. Industria-iraultzen denbora-lerroa

Zer da 4.0 Industria?

Komunikazioak 4.0 Industriaren funtsezko ezaugarria izango dira. Izan ere, fabrika adimendunak ahalbidetzeko, industria osatzen duten makina bakoitzari elkarren artean komunikatzeko eta elkarrekin konektatuta egoteko gaitasuna eman beharko zaio. Hala ere, horrelako komunikazioak ezartzea zailtzen duten zenbait oztopo daude.

Lehenik eta behin, komunikazio kableatuetatik haririk gabeko komunikazioetarako trantsizioak oraindik ez ditu beharrezko bermeak eskaintzen. Tradizionalki, industria-komunikazioak sistema kableatuetan oinarritu izan dira; alabaina, horrelako komunikazioak ez dira egokiak sareen tamaina handitzeko edo ekipoei mugikortasuna emateko. Horregatik, haririk gabeko teknologiak dira hautagai egokienak, nahiz eta oraindik sistema kableatuen sendotasuna ezin duten bermatu.

elhuyar
zientzia eta teknologia

Bestalde, aplikazio-ingurunea beste erronka bat da, berez. Aplikazio-lekuak (lantegiak, ekoizpen-plantak, e.a.) ingurune itxiak izan ohi dira, ekipoen kontzentrazio handiarekin eta islapenak sortzen dituzten ekipo metaliko askorekin. Faktoreen konbinazio horren ondorioz, jasotako seinalea oihartzun ugariz osatuta dago eta hedapen-kanal bortitz batez eraginda. Zalantzarik gabe, harrera baldintza horiek komunikazioak zailtzen dituzte, eta erronka bat planteatzen dute: ingurune horietarako egokiak diren sistema ultra-sendoak diseinatzea.

Azkenik, haririk gabeko komunikazioetan erabiltzen diren frekuentzia-bandak beste erronka nagusietako bat dira. Haririk gabeko sistema gehienek, ISM (Industrial, Scientific and Medical) banda libreak erabiltzen dituzte. Banda horiek konkurrentzia handia dute, eta, beraz, sareak kongestionatu egin daitezke, paketeen arteko talkara iritsiz edo ekipo jakin batzuetarako sarbidea galaraziz.

Gaur egungo eskaintza teknologikoa

Lehenengo eta behin, momentuz industria-komunikazioak unibertsalki definitzen dituen estandar edo teknologiarik ez dagoela aipatu beharra dago. Horregatik, amaierako aplikazioaren edo ingurunearen arabera, hainbat aukera teknologiko daude.

WiFi-a oso onartuta dagoen irtenbidea da. Zalantzarik gabe, teknologia hori bultzatzeko argudio nagusia gaur egungo hedapena da; izan ere, egungo enpresa guztiek edo ia guztiek kudeatzen baitute beren WiFi sare propioa. Hori dela eta, ekipo automatizatuak sare horretan sartzeak kostu txikiagoa dakar sare berri esklusibo baten hedapenarekin alderatuta. Alabaina, teknologia honek ere baditu bere desabantailak, ez baitu komunikazioen fidagarritasuna bermatzen duen mekanismorik inplementatzen, ezta determinismorik ere. Horregatik, ikerketa asko bideratu dira bi puntu ahul horiek sendotzera, ingurunerako sarbidea modu ordenatuan koordinatzen duten eta komunikazioen sendotasuna hobetzeko mekanismoak ezartzen dituzten denbora-eskemak garatuz, besteak beste, paketeen berbidalketa edo erredundantzia.

Hautagaien rankingaren goiko postuetan Bluetooth-a aurki daiteke, bereziki, Bluetooth Low Energy (BLE) soluzioa. BLE ekipoen interkonexiora bideratutako teknologia da, non nodoak maisu-rola betetzen duten ekipoekin konektatzen diren. Irtenbide horren sendotasuna energia-kontsumo baxuan datza, eta horrek eragiten du, hain zuzen, oso teknologia interesgarria izatea bateria duten ekipoen aplikazioetarako, hala nola, sentsore-sareetarako. WiFi-arekin alderatuta, ordea, BLE-k datu-tasa txikiagoa du (2 Mbps) eta horrek informazio gutxi transferitzea eskatzen duten aplikazioetara mugatzen du.

Bukatzeko, banda zabaleko komunikazio-sistema mugikorrak ere proposatu dira zenbait behar asetzeko. Horien artean, aipatzekoa da komunikazio mugikorren laugarren belaunaldiaren erabilera, Long Term Evolution (LTE), esaterako, Internet of Things-era (IoT) bideratutako sare eta aplikazioei euskarria emateko erabili izan dena. Irtenbide horri Narrow Band IoT deritzo eta, BLE bezala, datu tasa txikien transmisioan oinarritzen da erabiltzaile kopuru handi bati zerbitzua emateko.

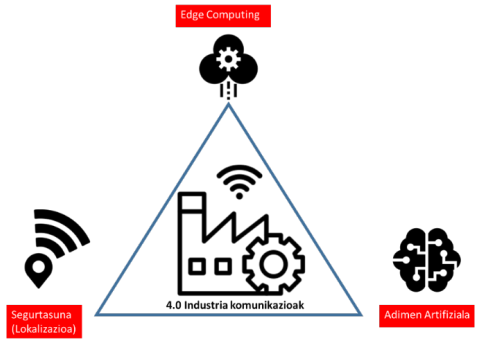
Une honetan, ikerketa 5G aldera bideratzen ari da, komunikazio mugikorren bosgarren belaunaldiak industria-betekizunen antzeko modalitate bat baitu. Hala ere, oraindik garapen-fasean dagoenez, ez dago arrakastaren bermea ematen duen lan esperimentalik.

elhuyar
zientzia eta teknologia

Etorkizunari begira

Gaur egun, irrati-komunikazioen mundua geldiezinezko abiaduran doa aurrera, batez ere, telefonia mugikorraren bultzadari esker. Izan ere, orain dela gutxi hedabideetara heldu den 5G teknologiaren hedapenaren hastapenean gauden honetan, ikerkuntzaren munduan 6G ezizenaz ezagutzen den hurrengo fasearen gaineko lanak gero eta nabariagoak dira. Etengabeko bilakaera honen erritmoa ezin da esparru guztietara hedatu eta 4.0 Industriaren aplikazioek, esaterako, beren ibilera propioa markatu behar dute. Horrenbestez, aurretik aipatu den bezala gaur egun 5Gren barruan jada onartuak eta lege diren hainbat teknologia oraindik probatzeke daude industria-komunikazioen alorrean.

Aipatzekoak dira seinalearen atzerapena (edo prozesaketa denbora) murrizten duten teknikak. Hauen artean, nabarmentzekoa da *Edge Computing* esparru honetan inportantea izan daitekeen aldaketaren eredu gisa. Laburki, honen helburua sarearen lanak muturrera eramatean datza, hots, datuak sortu diren iturritik gertu; modu horretan, muturretik muturrera informazioa garraiatzeko beharrezkoa den denbora gutxituz. Bestetik, beste alor batzuetan oso ezaguna den *Cloud Computing*-ak ere garrantzia izan lezake industria-komunikazioetan, baina horretarako honi lotutako atzerapen denborak murriztu beharko ziratekeen.



2. 4.0 Industriaren etorkizuneko ardatzak

Arestian aipatutako, eta komunikazio mugikorretatik oinordetzan hartutako irtenbideez gain, haririk gabeko industria-komunikazioetarako aipagarriak izan daitezkeen teknologia handizaleagoak ere badira. Azken urtean, Nature aldizkariaren arabera, *Adimen Artifiziala (AA)* laugarren postura igo da gehien bilatutako ikerkuntza terminoen artean. Ez da harritzekoa, beraz, laugarren industria iraultza honetan ere hainbat alorretan eragina aurreikustea: energia sorkuntza eta banaketa sistemetan, IoT eta sentsore sistemetan, kontrolaren monitorizazioan e.a. Zerrenda horretan haririk gabeko irrati-komunikazioak ere koka daitezke. AA oso erabilgarria izan daiteke lantegi berrien komunikazio sareen plangintzarako. Orain arte, plangintza tradizionalean transmisoreak estatikotzat jotzen ziren, hartzaileak baitziren mugikortasuna zuten gailu bakarrak. Alabaina, lantegi-adimendu batean konektatuta egon daitezkeen gailu kopurua eta tipologia handitzen doa etengabe, igorleak barne eta horrenbestez, etorkizuneko lantegi baten estaldurarako planifikazio estatikoak muga nabariak ditu. Bestalde, AAk ere ekarpena egin dezake eraso baten aurrean haririk gabeko komunikazioek daukaten babes eskasaren aurrean. Adibidez, AA teknikak erabiliz sareko trafikoa azter daiteke, eta, aurretiaz ondo entrenatutako ereduen bidez, gure sarean gerta daitezkeen eraso maleziatsuak identifika daitezke trafikoan aldaera susmagarriak aztertuz.

elhuyar
zientzia eta teknologia

Bestetik, 4.0 Industriaren alorrean irrati-komunikazioak segurtasun fisikoaren ardatz berri gisa ere uler daitezke. Azken urteetan argitaratutako ikerketek erakutsi dute haririk gabeko sistemek informazioa garraiatzeaz aparte, beste hainbat xede ere bete ditzaketela, esaterako, gela baten barruan dauden pertsonak zenbatzeko edota pertsona horiek egiten ari diren jarduera zehazteko. Hitz gutxitan, hori lortzeko pertsonen presentziak igorritako seinalean eta zeharkatutako kanalean izandako aldaketak estatistikoki aztertu behar dira. GPS teknologia erabili ezin den esparru batean, irtenbide hobeezina izan daiteke edozein arrisku edo istripuren aurrean langileak lokalizatuta eta zenbatuta izateko. Zalantzarik gabe, horrelako identifikazio sistemak inplementatzea, eta hauek aurretiaz azaldu diren AA teknikoekin bateratzea, ezinbesteko ikerkuntza lerroa da etorkizuneko lantegi adimentsuen garapen seguruentzako.

Azken gogoetak

4.0 Industria iritsi da jada eta haririk gabeko komunikazioen eraldaketa sakona eskatzen du. Egungo eskaintza teknologikoak hainbat hautagai aurkezten ditu, baina batek ere ez ditu jarraitu beharreko eredua izateko ezaugarri guztiak betetzen. Horregatik, ikerketa- eta garapen-prozesuak jarraitu egin behar du, eta dagoeneko dauden teknologiak errendimendua hobetuko duten teknika disruptiboekin konbinatu beharko dira etorkizuna orainaldira ekartzeko.

Bibliografia

R. Drath and A. Horch, "Industrie 4.0: Hit or Hype? [Industry Forum]," in IEEE Industrial Electronics Magazine, vol. 8, no. 2, pp. 56-58, June 2014.

A. Varghese and D. Tandur, "Wireless requirements and challenges in Industry 4.0," 2014 International Conference on Contemporary Computing and Informatics (IC3I), Mysore, 2014, pp. 634-638.

K. Montgomery, R. Candell, Y. Liu, and M. Hany, “Wireless user requirements for the factory workcell,”NIST Report, Adv. Manuf. Ser.(NISTAMS)-300, vol. 8, 2019.

3.3 Publications associated to Contribution 3

3.3.1 Conference paper C5

This subsection presents a conference paper related with Contribution 3. The full reference of the paper is presented below:

- L. Zhang, Y. Wu, Wei Li, S-I. Park, J-Y. Lee, H-M. Kim, N. Hur, E. Iradier, P. Angueira and J. Montalban, "ATSC 3.0 In-band Backhaul for SFN Using LDM with Full Backward Compatibility," 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Jeju, Korea (South), 2019, pp. 1-6, doi: 10.1109/BMSB47279.2019.8971918.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Indexed Congress in IEEExplore
- Area: Computer Science and Engineering
- SJR factor: 0.300

ATSC 3.0 In-band Backhaul for SFN Using LDM with Full Backward Compatibility

Liang Zhang, Yiyan Wu, Wei Li
Communications Research Centre, Canada

Sung-Ik Park, Jae-young Lee, Heug-Mook Kim, Namho Hur
Electronics and Telecommunications Research Institute, Korea

Eneko Iradier, Pablo Angueira and Jon Montalban
University of Basque Country, Spain

***Abstract*— In BMSB'2018, the authors presented a paper on using Layered-Division-Multiplexing (LDM) to deliver in-band backhaul of robust mobile services for ATSC 3.0 Single-Frequency-Netowrk (SFN) transmitters and gap-fillers [1]. Only robust mobile services are backhauled to form an SFN in indoor or isolated environments to extend and improve reception Quality of Experience (QoE) for mobile devices. This paper proposes a complete in-band backhaul approach that provides backhaul for both robust mobile and high-data-rate fixed services. The proposed solution uses LDM to carry backhaul data alongside broadcast data intended for reception by the public. The system is fully backward compatible with the ATSC 3.0 Next Gen TV service, causing no degradation on consumer service reception. The scheme can greatly reduce broadcasters' operating costs, while improving service quality – especially for mobile, handheld and indoor reception. In comparison to On-Channel-Repeater (OCR), the proposed system can provide SFN service with flexible emission timing control. The regenerated signals at each transmitter can have high SNR to meet RF spectrum mask requirement and support multi-hop operation. A multi-hop backhaul system can form a mash network for reliability (re-route in case of one transmitter failure), and scalability to follow the population and market growth. The proposed system can deliver backhaul not only for SFN broadcast services, but also for non-broadcast data distribution, such as IoT and connected cars.**

Keywords*— *LDM, SFN, STL, In-band Backhaul, Mesh Network, Full-Duplex Transmission,

I. INTRODUCTION

With the rapidly increasing number of mobile and handheld devices, delivering robust mobile services becomes a necessary capability for the next generation TV broadcasting system, also known as ATSC 3.0 [2]. Many earlier studies have shown, however, that it is quite challenging to provide good mobile service coverage using the existing high-tower-high-power (HTHP) single transmitter approach. Single Frequency Networks (SFN) with multiple transmitters is a preferred infrastructure that can significantly improve quality of experience (QoE), especially for small-size handheld and portable receivers in challenging environments, such as indoor locations. Implementing SFN requires deploying multiple low-power SFN transmitters (SFN-Txs) in addition to the existing HTHP facilities.

An example of the current approach to implement an ATSC 3.0 SFN is illustrated in Figure 1. In this example SFN, mixed mobile and fixed services are delivered using Layered Division Multiplexing (LDM) technology [3].

As shown in Figure 1, deploying each additional SFN-Tx requires a distribution link, or backhaul, also called a Studio-to-Transmitter Link (STL), to deliver the service data and control signaling from a Broadcast Gateway (BGW). In current systems, STLs are implemented using either fiber circuits or dedicated microwave links. Fiber links are not always available at desired locations and are expensive to install or to rent. Microwave spectrum often is scarce, and installation is also very costly. Consequently, a new solution is required for SFN distribution links that offers both low installation and low operation costs.

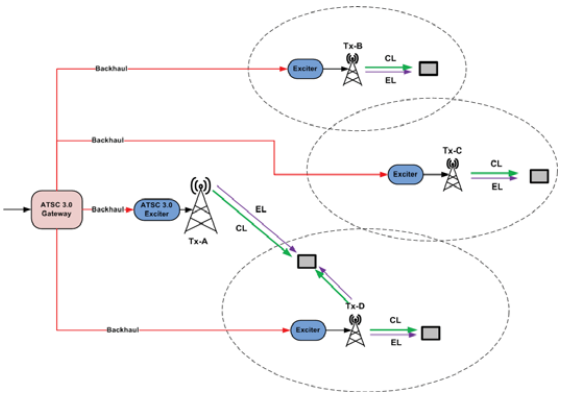


Figure 1: A 2-Layer ATSC 3.0 Single Frequency Network

II. BACKWARD-COMPATIBLE WIRELESS IN-BAND BACKHAUL USING LDM

The use of wireless in-band backhaul distribution links is a promising alternative solution, offering both low infrastructure and operational costs, and high spectrum efficiency for ATSC 3.0 SFN deployments.

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2017-0-00081, Development of Transmission Technology for Ultra High Quality UHD). This work has been partially supported by the Basque Government under the PREDOC grant program (PRE_2018_1_0344), and Project 61420106008.

978-1-7281-2150-5/19/$31.00 ©2019 IEEE

During transition period, ATSC 1.0 and ATSC 3.0 programs must co-exist. Since there is no new spectrum allocated for transition, broadcasters need to collaborate so that more than one ATSC 3.0 programs be transmitted in one 6MHz RF channel. Meanwhile, ATSC 3.0 with HEVC and LDM technologies can accommodate many programs in a 6 MHz channel. For example, LDM Core Layer (CL) with data rate of 4 Mbps can deliver 2 x 720p, or 6 x 480p programs for robust mobile service, while Enhanced Layer (EL) with data rate of 18 Mbps can provide 4 x 1080p, or 8 x 720p programs for fixed service. Statistical multiplexing can further increase the number of programs by 20-30%. Scalable video coding can also reduce the video data rate, or improve the video quality. Therefore, using part of the ATSC 3.0 data capacity for in-band backhaul is viable.

Figure 2 shows the basic concept of using wireless in-band backhaul in an ATSC 3.0 SFN. The main transmitter, Tx-A, is an existing high-power transmitter that has a dedicated STL connection to the BGW. Tx-A receives the STL data for each new secondary SFN-Tx (Tx-B/C/D) and modulates it onto part of the ATSC 3.0 signal for over-the-air transmission. This yields an in-band solution because the backhaul data distribution shares the same spectrum as the broadcast services.

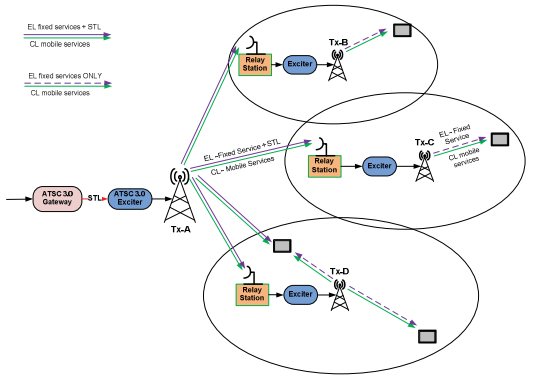


Figure 2: Using LDM to implement in-band wireless backhaul

To achieve higher combined backhaul and broadcast-service throughput, LDM is used to combine backhaul data and broadcast-service data within one RF TV channel. In this case, the main transmitter, Tx-A, transmits a 2-layer LDM signal wherein CL delivers robust mobile services and part of EL capacity delivers backhaul data for the secondary SFN-Tx's. At each secondary SFN-Tx, an SFN relay receiver (RL-Rx) is implemented to decode the backhaul data from the received LDM signal. The decoded data is then fed to an ATSC 3.0 Exciter to generate the SFN broadcast-service signal for emission.

III. SFN and Backhaul Data Timing and Synchronization

To achieve SFN operation, the re-transmitted signals from the secondary SFN-Tx's (Tx-B/C/D) need to be frequency- and time-synchronized with the signal emitted from the main transmitter, Tx-A. Such synchronization requires the backhaul data from Tx-A to have a time advance relative to the broadcast service signal. This time-advance needs to be sufficiently long to allow the RL-Rx's to decode the backhaul data and the related transmitters to re-encode/re-modulate it into the SFN broadcast-service signal. Consequently, a timing control mechanism is required. For example, if T is the required time advance and $X(t)$ is the broadcast service data, backhaul data should be transmitted as $X(t–T)$. In case of multiple-hop backhaul with n-hop, the backhaul data from different RL-Tx should be $X(t–nT)$, $X(t–nT+T)$, $X(t–nT+2T)$,… $X(t–T)$, respectively.

When using part of the EL signal for backhaul data distribution, the fixed broadcast service (FBS) delivery capacity of the EL is reduced. This EL capacity loss can be mitigated by installing high-gain antennas at high elevation for the RL-Rx's, which provides high reception signal-to-noise ratio (SNR) and a stationary close to line-of-sight (LOS) channel for backhaul signals. These conditions allow the use of high-order modulation (e.g., 1024QAM or 4096QAM) and high channel coding rate, thereby reducing the percentage of transmission capacity required for backhaul transmission.

IV. Multiple Scenarios of In-band Backhaul Solutions Using LDM

There are different implementation scenarios for LDM in-band backhaul, which are shown in Figure 3. Figure 3(a) and 3(b) present scenarios with 2-layer LDM, which are fully backward compatible with the current ATSC 3.0 standard. There is no performance impact to consumer receivers. In Figure 3(a), the backhaul data of both mobile and fixed services are distributed/backhauled in EL. There is no capacity loss in CL for robust mobile services. In Figure 3(b), a dedicated single-layer portion is allocated for backhaul data delivery for both CL mobile and EL fixed services, which is essentially a time-division-multiplexing (TDM) approach for in-band backhaul signal transmission. While this approach allocates more transmission power to backhaul signal and removes the need for CL cancellation at the RL-Rx, it suffers capacity loss in both the CL and EL signals.

A third LDM signal layer can be added to the current 2-layer ATSC 3.0 signal to deliver the backhaul data of the CL mobile services. Figure 3(c) and Figure 3(d) shows two possible signal configurations. In this solution, the third signal layer has very low power and is configured with strong coding and modulation that is similar to those of the CL mobile services. In order to decode the backhaul data in the third signal layer and limit its impact to the EL fixed services, the SFN Relay Station (SFN-RS) needs very high SNR. Therefore, it is necessary for the SFN-RS to have high dynamic range tuner and directional antenna at high elevation. Adding the third layer could reduce the broadcast service capacity loss.

The power of this third layer needs to be carefully designed to have negligible impact on the broadcast service data reception.

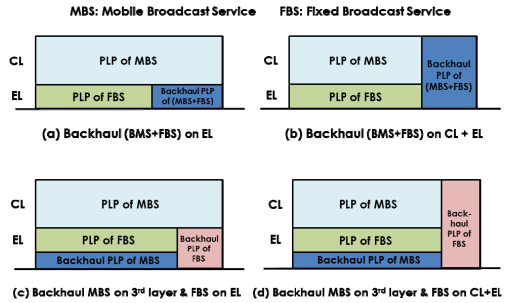


Figure 3: LDM signal structure for wireless In-band backhaul

In both scenarios shown in Figure 3(c) and (d), the EL backhaul data are transmitted using separate PLPs, in order to reduce the spectrum consumption and, therefore, the impact to fixed 4k-UHD service delivered in EL. In addition, the CL backhaul signal in the third layer does not overlap with the EL backhaul signal. This helps to increase the EL backhaul signal SNR (without the third layer as interference), and allow it to use higher order modulation and coding parameters.

Since backhaul links generally have stationary near-LOS channel and operate at high SNR conditions, high order modulation and high coding rate could be used, which results in less than 1/3 of the RF channel capacity required for backhaul data for all the broadcast services, i.e., both CL mobile and EL fixed services. When MIMO and other advanced technologies are used for backhaul transmission, the total channel capacity loss can be reduced to 10% to 12%.

V. In-band Backhaul Signal Detection at Relay Stations

A. Backhaul Signal Detection

A simple block diagram of an SFN-Tx is shown in Figure 4, where a high-gain directional antenna is installed at the RS-Rx pointing to the main transmitter, Tx-A. This antenna provides high SNR and near LOS channel condition for backhaul signal contained in the 2-layer LDM signal emitted by Tx-A, which is referred to as forward signal (FWS).

Figure 4 also shows a major challenge for full-duplex in-band backhaul system, also known as self-interference. The RS-Rx antenna not only receives the FWS from Tx-A, but also the re-transmission signal from the SFN-Tx transmitter (Tx-B) antenna, which is referred to as loopback signal (LBS).

Assuming a 2-layer LDM transmission with in-band backhaul structure as shown in Figure 3B, the signal models of the FWS and LBS are plotted in Figure 5. In the FWS, 2-layer LDM is deployed to deliver mobile broadcast services (MBS) in CL and fixed broadcast services (FBS) in EL. A dedicated single-layer time-slot is assigned to deliver STL data of both MBS and FBS.

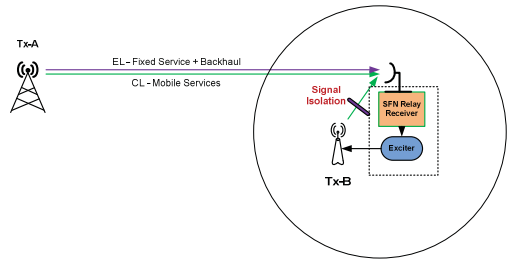


Figure 4: Backhaul signal detection at RS-Rx

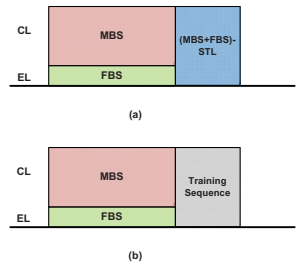


Figure 5: Signal model at SFN-Tx relay receivers;
(a) Received signal from Tx-A with backhaul data;
(b) Transmission signal of RS-Rx

For broadcast services, a careful timing adjustment is required so that the FWS and LBS deliver time and frequency-synchronized SFN services. On the other hand, as shown in Figure 5, the re-transmission signal from Tx-B during the backhaul time slot becomes an interference for backhaul data detection.

The received signal OFDM symbols in the kth subchannel at the RS-Rx during the backhaul time slot can be expressed as,

$$Y_{RL}(k) = X_{BH}(k) \cdot H_{FW}(k) + \sqrt{\kappa_{LB}} \cdot X_{LB}(k) \cdot H_{LB}(k) + N_0(k) \quad (1)$$

where X_{BH} is the backhaul signal transmitted from Tx-A; X_{LB} is the transmission symbols from Tx-B (the relay transmitter); H_{FW} is the normalized forward channel response from Tx-A to RS-Rx; H_{LB} is the normalized loopback channelresponse, κ_{LB} is the power ratio of the LB signal over the FW signal, and N_0 is the thermal noise power.

Ideally, if Tx-B has the option of transmitting a "Null" signal with zero-power, i.e., $X_{LB}(k) = 0$, there would be no self-interference. This, however, is not possbile in ATSC 3.0. Even when Tx-B does not use this part of time allocation for service/backhaul data transmission, it needs to fill dummy

symbols in order to ensure a constant transmission power. These dummy symbols are known to the RS-Rx receiver and can be used as a training sequence for loopback channel estimation and self-interference cancellation.

Since the transmitter antenna and the RS-Rx receiver antenna are closely located at the SFN-Tx, the loopback signal power could be significantly higher than the forward signal. This could impact the reception performance of the backhaul signal. This is the same case as full duplex transmission, where data are transmitted-received in both directions on the same spectrum at the same time [4].

To reduce the impact of the loopback signal, a signal isolation (SI) module could be implemented at the SFN-Tx, which attenuates the received loopback signal power at the RS-Rx, as shown in Figure 4. Good isolation between the SFN Tx antenna and the backhaul receiving antenna can be achieved by using antenna distance, blocking, antenna directivity and polarization, and other isolation schemes. Tests showed that 70-80 dB isolation can be achieved in practical conditions.

However, for medium-power SFN re-transmitter, even with a good SI, the loopback signal power could still be ten to hundrent times (10-20 dB) higher than that of the forward signal. Loop back signal cancellation needs to be applied,

$$\begin{aligned} Y_{BH}(k) &= X_{BH}(k)\cdot H_{FW}(k) \\ &= Y_{RL}(k) - \sqrt{\kappa_{LB}}\cdot X_{LB}(k)\cdot \hat{H}_{LB}(k) \end{aligned} \quad (2)$$

where $\hat{H}_{LB}$ is the loopback channel estimate.

The signal to noise ratio (SNR) of the backhaul signal after the loopback signal cancellation can be calculated as,

$$\gamma_{BH}(k) = \frac{1}{\kappa_{LB}\cdot \nu_H^2 + N_0} \quad (3)$$

where ν_H^2 is the mean square error (MSE) of the loopback channel estimate. Therefore, the key technology for good loopback signal cancellation is the channel estimation module.

Since the backhaul signal takes away part of the transmission resource for service delivery, it is desirable that the backhaul signal be configured using high modulation order and high coding rate. This configuration, however, requires high SNR for signal detection. From Eq-3, a high SNR for backhaul signal requires low residual loopback signal power after the loopback cancellation. This translates into low channel estimation MSE, especially for high loopback signal power.

The loopback channel could be estimated using decision-directed channel estimation algorithms, since the loopback signal is known to the RS-Rx. The Least Square (LS) channel estimates in each subchannel is first obtained as,

$$\tilde{H}_{LB}(k) = \frac{Y_{RL}(k)}{X_{LB}(k)} = H_{LB}(k) + \frac{Y_{FW}(k) + N_0(k)}{X_{LB}(k)} \quad (4)$$

where the received forward signal from Tx-A, Y_{FW}(k), is treated as an additional noise.

Both frequency-domain and time-domain filtering can be applied to the LBS estimate to enhance the channel estimation accuracy.

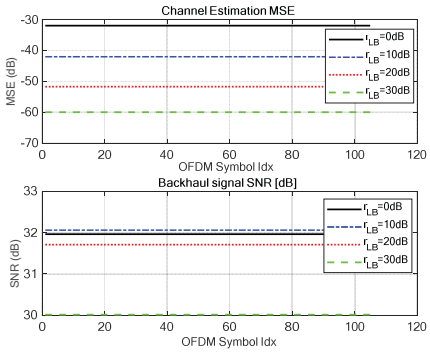


Figure 6: Loopback signal cancellation performance

B. *Loopback Signal Cancellation Simulation Results*

Matlab simulations are conducted to evaluate the performance of loopback signal cancellation with a practical two-dimentional channel estimator (2D-CE), where Wiener filtering is applied in both frequency-domain and time-domain.

Due to the short distance between the RS-Rx receive antenna and the SFN transmitter antenna, the loopback channel should be characterized as either a near-LOS channel or a multipath channel with short delay spread. In simulations, a shortened multipath channel is used following the characteristics of Typical Urban (TU) channel [5], where the mean delay spread is defined as 0.1 μsec, and the maximum delay spread is 0.7 μsec.

In simulations, an ATSC 3.0 system is assumed with transmission mode of 16k FFT. The forward signal has an SNR of 25 dB. For good understanding of the impact from the loopback signal power, four loopback over forward signal power ratios are considered: [0 10 20 30] dB.

Figure 6 summarizes the simulation results of the loopback signal cancellation. In the upper subplot, the average powers of the residual loopback signal after cancellation are plotted; the lower subplot shows the SNR of the backhaul signal taking into account of only the residual loopback signal power.

For higher loopback signal power, the channel estimate is more accurate, due to the lower interference level from the forword signal, Y_{FW}(k), in Eq-2. For loopback signal power of [0 10 20 30] dB, the 2D-Wiener channel estimator can achieve MSE of [-30 -41 -51 -60] dB.

For all tested scenarios, after cancellation, the residual loopback signals are all 30+ dB lower than the backhaul signal received from the main transmitter. For backhaul signal with an SNR threshold of 25 dB, this residual loopback signal introduces very minor impact on the backhaul data detection.

It needs to be pointed out that the loopback signal only impacts the backhaul signal reception. It has absolutely no impact on the broadcast service reception by the ATSC 3.0 consumer receivers. The backhaul signals are simply ignored by the consumer receivers.

VI. Multi-Hop In-band Backhaul and Mesh Network

Because each backhaul relay station regenerates the broadcast service signal, multi-hop backhaul is viable. Figure 7(a) demonstrates a 3-hop STL mesh network. The network is scalable that more Txs can be added to extend or shape the coverage to meet the market growth. If one SFN transmitter fails, the backhaul network can be re-routed to recover the service, Figure 7(b). The coverage radiuses in Figures 4 indicate fixed EL service. The robust CL service should fully recover the failed Tx service area.

VII. On-Channel Repeaters vs. In-band Backhaul

A. *On-Channel Repeater*

On-channel repeater (OCR) has been available for many years as a low-cost solution to extend broadcast service coverage areas and to fill coverage holes [6]-[9]. An OCR basically receives the broadcast signal from a main transmitter (MTx), amplifies the signal, and re-transsmit on the same RF channel. An OCR output signal SNR is always lower than the input signal SNR. It is difficult to implement a multi-hop OCR system, while still meets the RF spectrum mask requirement. The emission time of the OCR cannot be adjusted.

The main issue with OCR is the feedback signal from the re-transmission antenna to the receiving antenna, which can have significantly higher feedback power than the received signal from the MTx. This feedback signal can cause system oscillations and instability that disable the effective operation of the OCR [8].

Isolation mechanism becomes the critical component for OCR to reduce the feedback signal power. Most existing OCR solutions use both physical isolation and echo cancellation (EC) to achieve sufficient isolation performance. Physical isolation by antenna placement and antenna pattern design can usually reduce the loopback signal power by 30 to 50 dB, which might not be sufficient. The additional isolations are usually achieved by EC, where decision-feedback EC based EC (DF-EC) and FIR-filter-based EC (FIR-EC) are the two most popular solutions.

The DF-EC is able to provide up to 30 dB feedback signal cancellation for a single carrier modulation system, such as ATSC 1.0 [9]. It is, however, not applicable to the ATSC 3.0 system that uses OFDM modulation. For SFN, the emitted signal from OCR should fall within the equalization range of consumer receivers. In the case of OFDM systems, all signals from different transmitters must fall within the Cyclic Prefix (CP), or Guard Interval (GI). Using DF-EC requires the demodulation of the frequency-domain signal symbols, and therefore a delay of at least one FFT duration, without even counting the delays for time interleaving and re-modulation. The GI in ATSC 3.0 is at most ¼ of FFT duration. Hence, OCR output signal will be emitted outside the GI of the signal from the main transmitter. This will result in significant co-channel interference.

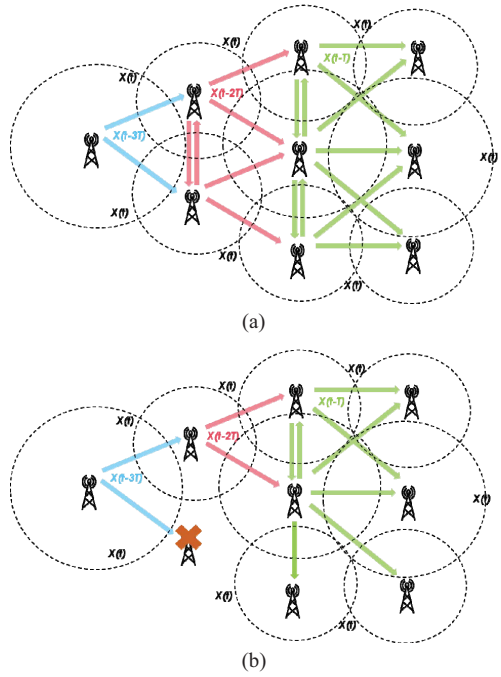


Figure 7: Multi-hop in-band backhaul SFN network:
(a) SFN Mesh Network;
(b) SFN mesh network with Tx failure and re-route.

The FIR-EC solutions could be implemented to achieve low delay. However, since linear equalization causes noise enhancement, received signal at OCR from the main transmitter must have very high SNR, or the received signal strength must be very high.

In summary, OCR has certain operating restrictions:

- OCR requires a good quality input signal with very high SNR and low multipath distortion to operate. The OCR signal output SNR is always worse than the input SNR.
- While existing EC techniques could achieve good cancellation performance, their stable operation and effectiveness are not guaranteed, due to the dynamic multipath characteristics of the feedback channel conditions. For challenging scenarios, existing OCR solutions could end-up in oscillation or instable conditions.
- OCR re-transmission power might be limited, and signal emission time cannot be controlled.

- It would be difficult to cascade the OCR operation due to signal SNR limitations.
- OCR only retransmits the received signal, no local data insertion of non-broadcast service data, such as IoT and connected car services are allowed.

Threfore, OCR is best used for isolated or enclosed areas.

B. *In-band backhaul*

For In-band backhaul based SFN, the SFN-RS receives the time-advanced backhaul data with equalization/demod/decode to remove all distortions (noise and interference). The regenerated signal is re-encoded and re-modulated to generate a clean/flat SFN re-transmission signal/spectrum. There is a total isolation between re-transmitted and the received signals. The emission timing can be controlled to be within the OFDM system GI range for SFN operation.

The backhaul and loopback signal isolation is also much relaxed. The loopback signal power can be much higher than that of the received signal, the system is still operational. Signal cancellation technology, similar to the one used in the LDM system, can reliably retrieve the backhaul data. Therefore, SFN can transmit a re-generated clean signal that fully complies with the RF spectrum mask regulation, at much higher power level than that of the OCR.

In-band Backhaul Link (IBL) SFN Relay Station behavior:

- IBL is designed for general SFN deployment with full SFN Tx timing control.
- The SFN-RS output SNR is always better than that of the input SNR, due to signal re-generation.
- While IBL SFN-RS requires a good quality input signal to retrieve backhaul data, its requirement is less demanding in comparison to the OCR, which is based on the non-error-corrected feedbacks causing noise enhancement and possible error propagation.
- It is possible to cascade the IBL operation or forming a multi-hop mesh network.
- Backhaul can also carry non-broadcast service data, such as IoT, connected car.
- It is viable to combine the IBL SFN with OCR as the last hop. IBL and OCR can co-exist and complement each other.

In addition, a fundamental advantage of IBL over OCR is that IBL is not necessarily limited to SFN backhaul only. IBL can be used for distributing general-purpose non-broadcast IP-data services, such as IoT, connected car, etc. The only negative impact of IBL the consumption of a small portion of the in-band broadcast data capacity

VIII. Conclusions

A wireless in-band backhaul system for ATSC 3.0 SFN operation was presented. The system uses LDM transmission to carry backhaul data alongside broadcast service data. The system is fully compatible with the ATSC 3.0 Next Gen TV service, causing no degradation of consumer reception. It can reduce broadcasters' operating costs, while improving service. It is designed for general SFN deployment with full SFN timing control. Due to signal re-generation, each SFN Txs has high SNR output that multi-hop mesh network can be constructed, which has high service reliability and scalable to the market growth. Backhaul can also carry non-broadcast service data, such as IoT, connected car. The only negative impact of in-band backhaul is the consumption of a small portion of the in-band broadcast data capacity. With the application of MIMO and other advanced signal processing technologies, the percentage of broadcast data capacity reduction could be limited to 10 – 15%.

References

[1] L. Zhang, Y. Wu, W. Li, K. Salehian, S. Laflèche, Z. Hong, S.-I. Park, J.-Y. Lee, H. M. Kim, N. Hur, M. Simon and D. G.-Barquero, "Using Layered-Division-Multiplexing for In-Band Backhaul for ATSC 3.0 SFN and Gapfillers," Proceedings of IEEE Int'l Symposium on Broadband Multimedia Systems and Broadcasting 2018 (BMSB2018), June 6-8, 2018, Valencia, Spain.

[2] L. Fay, L. Michael, D. Gomez-Barquero, N. Ammar and M.W. Caldwell, "An overview of the ATSC 3.0 Physical Layer Specification," IEEE Trans. Broadcasting, vol. 62, no. 1, part II, pp.154-158, Mar. 2016.

[3] L. Zhang, W. Li, Y. Wu, X. Wang, S.-I. Park, H.-M. Kim, J.-Y. Lee, P. Angueira, and J. Montalban, "Layered division multiplexing: theory and practice," IEEE Trans. on Broadcasting, vol. 62, no. 1, part II, pp. 216–232, Mar. 2016.

[4] D. Kim, H. Li, and D. Hong, "A Survey of In-Band Full-Duplex Transmission: From the Perspective of PHY and MAC Layers," IEEE Comunications Survey & Tutorial, vol. 17, no.4, 4th Quarter, 2015.

[5] COST 207 Report, Digital land mobile Radio Communications, Commission of European Communities, Directorate General, Telecommunications, Information Industries and Innovation, Luxembourg, 1989.

[6] L. Zhang, et. al., "Wireless In-band Distribution Link using LDM for SFN Transmitters and Gapfillers in ATSC 3.0," Proceedings of IEEE BTS Annual Broadcast Symposium," Washington DC, USA, Oct. 9-11, 2018.

[7] M. M. Wang, "Dynamic Gain Management for On-Channel Repeaters," IEEE Trans. Broadcasting, vol. 59, no. 4, pp. 685-692, Dec. 2013.

[8] K. M. Nasr, J. P. Cosmas, M. Bard, and J. Gledhill, "Performance of an Echo Canceller and Channel Estimator for On-Channel Repeaters in DVB-T/H Networks," IEEE Trans. Broadcasting, vol. 53, no. 3, pp. 609-618, Sept. 2007.

[9] S. W. Kim, Y. T Lee, S. I. Park, H. M. Eum, J. H. Seo, and H. M. Kim, "Equalization Digital On-Channel Repeater in the Single Frequency Network," IEEE Trans. Broadcasting, vol. 52, no. 2, pp. 137-146, June 2006.

3.3.2 Journal paper J8

This subsection presents a journal paper related with Contribution 3. The full reference of the paper is presented below:

- L. Zhang, W. Li, Y. Wu, S. Lafleche, Z. Hong, S-I. Park, J-Y. Lee, H-M. Kim, N. Hur, E. Iradier, P. Angueira and J. Montalban, "Using Layered Division Multiplexing for Wireless In-Band Distribution Links in Next Generation Broadcast Systems," in IEEE Transactions on Broadcasting, vol. 67, no. 1, pp. 68-82, March 2021, doi: 10.1109/TBC.2020.2989638.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Journal paper indexed in JCR and IEEExplore
- Area: Electrical & Electronic Engineering
- Ranking: 69/266 (Q2) based on JCR 2019
- Impact factor (JCR): 3.419

Using Layered Division Multiplexing for Wireless In-Band Distribution Links in Next Generation Broadcast Systems

Liang Zhang, *Senior Member, IEEE*, Wei Li, *Member, IEEE*, Yiyan Wu, *Fellow, IEEE*, Sébastien Lafleche, Zhihong Hong, Sung-Ik Park, *Senior Member, IEEE*, Jae-Young Lee, *Senior Member, IEEE*, Heung-Mook Kim, *Member, IEEE*, Namho Hur, *Member, IEEE*, Eneko Iradier, *Graduate Student Member, IEEE*, Pablo Angueira, *Senior Member, IEEE*, and Jon Montalban, *Member, IEEE*

***Abstract*—To meet the ever increasing demand for better service quality and availability, the next generation digital TV (DTV) broadcast, a.k.a., the ATSC 3.0, is developed with new capabilities to deliver 4K ultra-high-definition (4K-UHD) services to fixed receivers, robust HD-quality services to mobile receivers, as well as non-broadcast services (e.g., broadband), including Internet of Things (IoT), Connected Vehicles, etc. To realize these capabilities, it becomes necessary to evolve the existing single-transmitter high-power-high-tower (HPHT) network to multi-transmitter low-power-low-tower single-frequency-networks (SFN), which requires installing new transmitters. Each new transmitter needs a feederlink to receive the service data from the broadcast gateway (BG) for transmission. This feederlink contributes to a significant portion of installation and operational costs of the new transmitters. This paper proposes a wireless in-band distribution link (WIDL) technology, where the distribution data is delivered to the newly deployed transmitters from the existing DTV towers wirelessly sharing the same TV band with the broadcast services. The distribution signal is multiplexed with service signal in the same ATSC 3.0 waveform. Layered division multiplexing (LDM) technology is proposed to achieve more efficient transmission of the distribution signal. The WIDL offers better performance and more robust operation than the on-channel repeater (OCR) technologies. More importantly, it offers the possibility of delivering backhaul data for future applications over the DTV infrastructure, such as IoT and connected vehicles. Therefore, the proposed WIDL is one enabling technology to achieve convergence of broadcast services with broadband and other wireless services on the DTV spectrum.**

***Index Terms*—In-band distribution, in-band backhaul, layered division multiplexing, LDM, Digital TV, ATSC 3.0, relay, mesh network, full-duplex transmission.**

Manuscript received February 7, 2020; revised April 4, 2020; accepted April 10, 2020. Date of publication May 14, 2020; date of current version March 4, 2021. This work was supported by the Institute of Information and Communications Technology Planning and Evaluation grant Funded by the Korea Government (MSIT) (Development of Transmission Technology for Ultra High Quality UHD) under Grant 2017-0-00081. *(Corresponding author: Liang Zhang.)*

Liang Zhang, Wei Li, Yiyan Wu, Sébastien Lafleche, and Zhihong Hong are with the Department of Wireless Communications, Communications Research Centre Canada, Ottawa, ON K2H 8S2, Canada (e-mail: liang.zhang@canada.ca; wei.li@canada.ca; yiyan.wu@ieee.org; sebastien.lafleche@canada.ca; zhihonghunter.hong@canada.ca).

Sung-Ik Park, Jae-Young Lee, Heung-Mook Kim, and Namho Hur are with the Broadcasting System Research Department, Electronics and Telecommunications Research Institute, Daejeon 305-700, South Korea (e-mail: psi76@etri.re.kr; jaeyl@etri.re.kr; hmkim@etri.re.kr; namho@etri.re.kr).

Eneko Iradier, Pablo Angueira, and Jon Montalban are with the Department of Communications Engineering, University of Basque Country, 48940 Bilbao, Spain (e-mail: eneko.iradier@ehu.eus; pablo.angueira@ehu.eus; jon.montalban@ehu.eus).

Digital Object Identifier 10.1109/TBC.2020.2989638

I. Introduction

TO MEET the fast increasing demand for wireless communication services, a majority of efforts in 5G development have been put on developing technologies to be used at millimeter-wave (mmWave) frequencies, where a vast amount of spectrum is available. On the other hand, sub-1GHz bands, including the UHF spectrum currently allocated for TV broadcasting services, offers many advantages over the mmWave frequencies due to the low propagation loss, high penetration capability, as well as low device costs [1]. These benefits make the sub-1GHz bands better choices to deliver some key future wireless applications including the Internet of Things (IoT), the connected vehicles, the public safety and emergency alert systems, the ultra-robust communications, the multicast/broadcast services, etc.

This offers new business opportunities for the broadcast industry if they can use the UHF bands to deliver more diversified services in addition to TV broadcasting. However, to materialize the business opportunities, the existing DTV technologies and DTV network infrastructures need to be upgraded, and the spectrum regulation and policies also need to be revised.

Currently, the TV broadcasting industry is facing multiple challenges: the reduced spectrum, emerging of many other media delivery platforms over the Internet, shifting of consumer viewing patterns to more on-demand services, etc. To address these challenges and to benefit from new emerging business cases, the next generation DTV system has been developed, which is also called the ATSC 3.0 system [2]. Taking advantage of the latest transmission and video technologies, with Internet Protocol (IP) based transport layer, the ATSC 3.0 system is capable of delivering high-definition (HD)

mobile services, ultra-HD (UHD) fixed services, interactive services, datacasting services, and can be readily integrated with 4G/5G broadband systems at the network level. More specifically, the high-throughput, flexible program channel configuration, and the IP-based transport layer makes the ATSC 3.0 an excellent candidate to deliver applications such as IoT and connected vehicles over large coverage areas.

The current generation DTV system in North America, the ATSC 1.0 system, uses single-transmitter high-power-high-tower (HPHT) network structure, where a single high-power transmitter is deployed at a high elevation to cover the target coverage area. This infrastructure, while works well for receivers with rooftop antennas, cannot deliver reliable services to mobile receivers that are frequently facing challenging reception conditions, including high-speed vehicle receivers, handheld devices, and receivers in indoor environments.

A good alternative to deliver reliable mobile services is the single-frequency-network (SFN) based on low-power-low-tower (LPLT) infrastructure [3], [4]. In an SFN, multiple transmitters emit the same signal in a time-synchronized and frequency-locked manner. Receivers, especially those close to the boundaries among adjacent coverage areas, can effectively combine the multiple signals received from adjacent transmitters. This results in a better reception performance due to the spatial diversity, which is also called an SFN gain.

Implementing an SFN requires deploying new transmitters, where each transmitter needs a distribution link, also called studio to transmitter link (STL), to obtain the service data from a Broadcast Gateway (BG) for transmission. The STL is also responsible for sending control signal to each SFN transmitter (SFN-Tx) to apply specific timing and frequency offsets to optimize the coverage performance.

Until now, the STL links are implemented using either fiber links or dedicated microwave links. Fiber links are not always available for locations to deploy SFN transmitters. Even if it is available, the cost to rent a fiber link is quite expensive. The dedicated microwave link is also quite expensive to install. In addition, with the current trend of allocating more spectrum for consumer broadband systems, the microwave spectrum becomes less available for distribution data transmission. Therefore, using either of these conventional STL methods would introduce significant cost from both initial installation and monthly operation. This cost quickly becomes unaffordable for SFNs with large number of low-power transmitters. Consequently, a cost-efficient solution is needed for the SFN deployment of the emerging ATSC 3.0 system.

Another important scenario is to deliver services to indoor or enclosed places within a coverage area that are difficult for wireless signals to reach, such as shopping malls, train stations, airport, stadiums, etc. These areas, however, provide good commercial opportunities for broadcasters due to the high density of headcounts. An efficient solution to provide robust services to these enclosed areas is to deploy gap-fillers, i.e., very low-power re-transmitters. To be economically viable, the gap-fillers need to be very low-cost, since they are used to cover only a small area. Neither fiber link nor the microwave link is suitable solution due to their high cost. This scenario also calls for low-cost STL solutions.

Thanks to the latest video coding and transmission technologies, the newly developed ATSC 3.0 system could deliver significantly high throughput. This offers the opportunity to use part of the TV channel to transmit the distribution data while still delivering same or more mobile and fixed broadcast services than the current ATSC 1.0 system.

In this paper, a wireless in-band distribution link (WIDL) solution is proposed, where the distribution data is transmitted from an existing transmitter, also called Master Transmitter (MTx), to the newly installed SFN-Tx's, using the same TV band for broadcast service transmission. In the proposed approach, the STL is integrated into the ATSC 3.0 signal emitted from the MTx as a special non-broadcasting service, which can only be decoded by the SFN-Txs. Therefore, the STL and the broadcast services are sharing the same TV band, and hence called an in-band distribution solution.

In the proposed WIDL, since the STL data is multiplexed with the service data over the same time-frequency resource, the achievable throughput of broadcast services is reduced. To minimize the service throughput loss, the non-orthogonal layered-division-multiplexing (LDM) technology is proposed to multiplex the STL data and the broadcast service data over a single TV channel, due to its significant capacity advantage over the traditional orthogonal time-domain multiplexing (TDM) and frequency-domain multiplexing (FDM) [5].

Since the proposed technology delivers the STL in a special logical channel, also called physical layer pipe (PLP) in ATSC 3.0, the WIDL equipment shares many system modules with conventional ATSC 3.0 equipment. This significantly reduces the engineering effort and implementation cost.

A similar concept has been studied for mobile broadband systems in [6], where the distribution data is called backhaul data. The 4G Long Term Evolution Advanced (LTE-A) system adopted integrated access and backhaul (IAB) technology to implement low-cost backhaul links to relay nodes (RNs) [7]. The backhaul transmission can operate in either in-band or out-of-band modes. In in-band mode, the backhaul transmission and service transmission share the same spectrum; while in out-of-band mode, the backhaul data is transmitted using a separate spectrum. Apparently, the in-band mode is more attractive since there is no need to apply for dedicated backhaul channels. However, it also means that the service throughput will be reduced to accommodate the backhaul transmission.

The RNs could operate in either half-duplex mode or full-duplex mode. In half-duplex mode, at any given time, an RN either receives the backhaul signal or transmits the service signal; while in full-duplex mode, the RN performs these two operations simultaneously. It becomes obvious that full-duplex mode offers a clear throughput advantage over the half-duplex mode. However, full-duplex mode creates self-interference issue that requires careful design and advanced

signal processing algorithms to obtain good backhaul signal detection. This results in increased implementation complexity.

Therefore, to achieve the highest system throughput, the in-band full-duplex is the most attractive solution. Up until now, all the in-band full-duplex IAB solutions adopted by LTE use orthogonal multiplexing (OM) methods, such as TDM and FDM to combine the service and backhaul data within the same channel [8]. Spatial domain multiplexing (SDM) was proposed in some recent works to reduce the interference between the service and backhaul signals [9].

OM technologies have been the preferred multiplexing technologies by earlier systems due to the simplicity in the receiver implementation. When using OM, there is little interference between the service and backhaul signals. On the other hand, it has been proved that OM technologies are far from achieving channel capacity for multiple signal transmissions over the same frequency resource [10].

In [11], a wireless distribution link was first proposed for the newly developed ATSC 3.0 system. The proposed technology targeted low-power gap-fillers to provide mobile broadcast services to large public enclosed areas, such as airports, train stations, stadiums, etc. This concept is later expanded to a fully backward compatible WIDL solution in [12] that can deliver distribution data to general high-power SFN transmitters. In [13], this technology is adapted for the deployment of a 5G-Broadcast network.

To the best knowledge of the authors, using wireless distribution links for broadcast services has never been studied prior to the work presented in [11].

This paper presents further development of the WIDL technology in ATSC 3.0 system. A complete distribution link solution is proposed, which uses LDM, multiple-input-multiple-output (MIMO) and other latest technologies to minimize the capacity loss of the service transmission. Different WIDL implementation options are discussed with their respective capacity-complexity trade-offs. Achievable system capacities are analyzed for different system configurations. Feasibility study is conducted at the SFN transmitters and implementation challenges are addressed. Finally, a future extension to a mesh distribution network is also discussed that could provide more network fault tolerance for large-area SFN with many transmitters.

There is currently a global trend on convergence of different wireless communications systems to achieve more efficient use of the spectrum and the infrastructure. The deployment of the IP-based ATSC 3.0 using medium-size cells is the approach for convergence from the broadcast industry. The proposed WIDL technology could play a critical role to enable quick deployment of new ATSC 3.0 systems and to deliver new types of services.

It needs to be pointed out that the proposed WIDL technology can deliver broadcast and different non-broadcast distribution data to each SFN transmitter. This is the key difference between the WIDL and the existing on-channel repeaters (OCR), which are designed to solely expand the coverage area. Therefore, using WIDL would offer broadcasters business opportunities to use ATSC 3.0 infrastructure and the TV spectrum to deliver broadband and other wireless services.

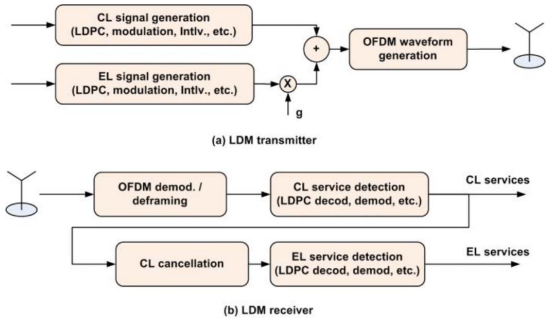


Fig. 1. ATSC 3.0 with a 2-layer LDM.

This paper is organized as follows: Section II briefly describes the concept of the WIDL solution in ATSC 3.0 systems with LDM, different signal configurations, and the achievable transmission capacities when different technologies are used. Section III presents an analysis on the achievable throughput of WIDL. In Section IV, challenges for detecting WIDL signal at the SFN transmitters are discussed and solutions are proposed. Section V describes a case study in Ottawa area to demonstrate the feasibility of the proposed WIDL technology for a wide range of ATSC 3.0 system configurations. Section VI proposes potential future evolutions of the WIDL technology. Section VII focuses on a comparison between the WIDL solution and the OCR technology. Finally, Section VIII gives the conclusions.

II. Wireless in-Band Distribution Links in ATSC 3.0

A. ATSC 3.0 and LDM

As a new generation DTV system, ATSC 3.0 uses the latest high efficiency video coding (HEVC) and advanced transmission technologies, including improved low-density-parity-check (LDPC) code [14], non-uniform QAM (nuQAM) modulation [15], MIMO [16], and LDM [5]. While most of these technologies provides only marginal performance improvement, LDM has been shown to offer significant service capacity gain.

LDM is a non-orthogonal multiplexing (NOM) technique that was adopted by ATSC 3.0, to provide significantly higher throughput when delivering services with different quality of service (QoS) requirements over the same channel [5]. In LDM, multiple signal layers are transmitted over the same time and frequency radio resource, where each signal layer delivers services with a specific QoS requirement.

The block diagrams of ATSC 3.0 transmitter and receiver with a 2-layer LDM are shown in Figure 1. At the transmitter, two signal layers are first generated separately, which are subsequently combined with a pre-defined power ratio, determined by the coefficient g. The higher-powered layer is called Core Layer (CL), while the lower-powered layer the Enhanced Layer (EL). At the receiver, conventional signal

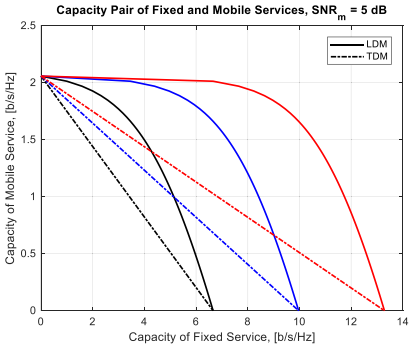


Fig. 2. LDM capacity gain over TDM.

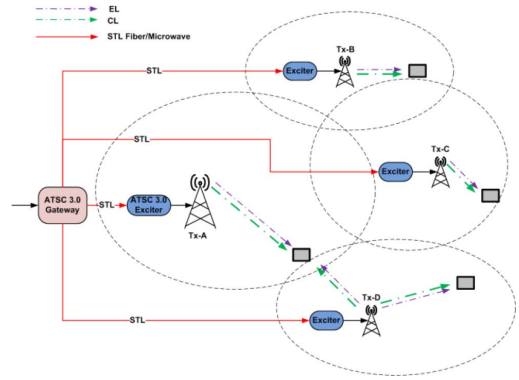


Fig. 3. Traditional ATSC 3.0 SFN.

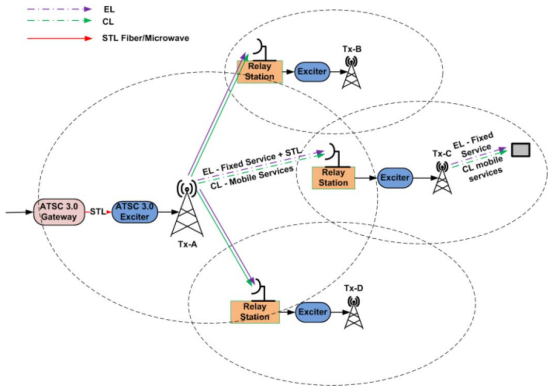


Fig. 4. ATSC 3.0 SFN with wireless in-band distribution.

detection is carried out to decode the CL services, treating the EL as interference. To decode the EL services, the receiver first decodes and cancels the CL signal, and then performs conventional signal detection. Canceling CL signal is called successive signal cancellation (SSC), which is the main additional signal processing required at an LDM receiver [5].

Figure 2 compares the achievable service capacities of a 2-layer LDM system and a TDM system, where both systems deliver two services: a robust service requiring a low signal to noise ratio (SNR) of 5 dB and a fixed service requiring a high SNR of 20, 30, or 40 dB. The LDM system uses CL for mobile service delivery and EL for fixed service delivery, while the TDM counterpart uses different time slots for the two services.

It is clearly shown that LDM offers significantly higher capacity than the TDM system. This capacity gain becomes higher for fixed services with higher SNR requirements, i.e., higher fixed service throughput.

The achievable capacities of ATSC 3.0 with LDM has been demonstrated by theoretical analysis, computer simulations and field tests [17]. With LDM, an ATSC 3.0 system could simultaneously deliver four to six 1080p high-definition (HD) fixed TV programs and one to two 720p HD mobile programs in one TV channel of 6 MHz, in comparison to only one 1080p HD fixed service delivered by the current ATSC 1.0.

Assuming an ATSC 3.0 system is deployed to deliver two 1080p fixed and two 720p mobile services, which is more than twice the throughput of an ATSC 1.0 system, it still has an additional throughput for 2∼4 fixed services in EL. This extra capacity could be used to deliver the distribution data.

Distribution data is delivered between transmission towers, where both the transmitter and the receiver antennas are located on towers at high elevations. This enables very high signal SNR at the distribution data receiver, especially high-gain directional receiving antenna can be installed. As shown in Figure 2, this is an ideal application scenario for using LDM EL, where high SNR brings a high capacity benefit.

B. Wireless in-Band Distribution Links in ATSC 3.0

A traditional ATSC 3.0 network with four transmitters is illustrated in Figure 3. The BG is responsible for sending distribution data to each transmitter via STLs, implemented by fiber links or dedicated microwave links. The distribution data for each transmitter includes both the broadcast service data and specific control signaling that configures the time and frequency offsets for signal emission.

In SFN, all transmitters emit the same signal and the transmissions are time-synchronized and frequency-locked. Therefore, the STLs for different transmitters deliver the same service data. On the other hand, the control signaling is unique for each transmitter, where different time and frequency offsets are applied to different transmitters to achieve a more uniform coverage performance.

The traditional structure suffers not only high installation and operation costs, but also very low efficiency since all the four STLs mostly carry the same service data.

The proposed WIDL solution is shown in Figure 4 for the same ATSC 3.0 SFN with four transmitters. There is one main transmitter with a traditional STL by fiber or microwave link, which will be referred as anchor transmitter (A-Tx) hereinafter. The A-Tx receives its own STL data and the STL data for the other three transmitters, which will be referred to as remote transmitters (R-Tx) in the rest of this paper.

The A-Tx uses its own STL data to generate the broadcast service signal for transmission within its own coverage area. At the same time, the A-Tx will encode the STL data for the three R-Tx's into the same ATSC 3.0 waveform. The STL data would be encoded into PLPs with special service identifications that would only be decoded by the distribution data receivers at the R-Tx's. The consumer receivers would simply ignore the STL PLPs due to the unrecognized service identification. Therefore, the proposed WIDL is fully backward compatible and has no impact on existing services for consumer receivers.

The A-Tx only needs to encode the broadcast service data once in the transmission signal for all three R-Tx's, because the broadcast service data is the same for different transmitters in an SFN. This results in high spectral efficiency. On the other hand, A-Tx needs to transmit all the specific control signaling for all R-Tx's. Fortunately, this won't require much extra capacity since the control signaling requires very little data.

Most likely, the A-Tx would be an existing ATSC 1.0 high transmitter tower, capable of emitting high power. The BG to the A-Tx link is the existing STL for ATSC 1.0 installation.

At an R-Tx, a Relay Station (RS) is implemented to receive and decode the distribution data in the signal from A-Tx. The decoded distribution data is then fed into an Exciter which generates the service signal for emission.

At the R-Tx, the RS receive antenna can be installed at a high elevation on the transmission tower, which results in a line-of-sight (LOS) propagation channel for the distribution signal. In addition, a highly directional receiving antenna could be installed so that the received distribution signal could have very high SNR. These favorable reception conditions make the LDM EL the ideal choice to deliver the STL data. Furthermore, with high received SNR, the in-band distribution signal can be configured using spectral efficient modes, i.e., with higher-order modulations and higher coding rates. Therefore, the spectrum resource required for STL data transmission could be very small and there is no significant reduction in service delivery capacity.

C. Transmitter Timing Control for SFN Operation

In an SFN as shown in Figure 4, all the transmitters need to deliver the same service signal, and the emissions from different transmitters need to be time-synchronized. When using WIDL, this requires that the STL data embedded in the A-Tx transmission signal has a time advance with respect to the service data. This time advance is needed for the R-Tx to receive and decode the STL signal, and to generate the service signal for re-emission. Therefore, a timing control mechanism needs to be designed to configure the relative timing between the STL data and the service data to align the operations of the different transmitters.

The timing control for SFN is illustrated in Figure 5 and explained as follows:

- In the transmission signal from the A-Tx, the STL data, $d(t-T)$, is transmitted with a time advance of T than the service data, $s(t)$.

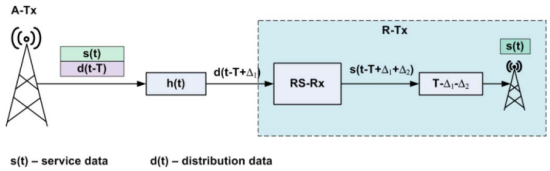


Fig. 5. ATSC 3.0 SFN with wireless in-band distribution.

- At the R-Tx, the RS receives a delayed STL data $d(t-T+\Delta_1)$, where the delay, Δ_1, is introduced by the signal propagation.
- The RS performs necessary decoding signal processing to recover the STL data and subsequently encoding signal processing to generate the ATSC 3.0 SFN service signal. The output is an ATSC 3.0 signal, $s(t-T+\Delta_1+\Delta_2)$, which contains the same SFN service signal as that from the A-Tx, with a relative delay of $-T+\Delta_1+\Delta_2$, where Δ_2 is the time required to decode the STL data and encode the service data.
- To align the service signal from A-Tx and R-Tx, an additional emission adjustment delay of $T-\Delta_1-\Delta_2$ is further introduced so the emission service signal from the R-Tx is synchronized to that from A-Tx, i.e., $s(t)$.

It becomes obvious that the time-advance of the STL data, T, needs to be at least as long as the sum of the two delays, $(\Delta_1+\Delta_2)$. While the propagation delay, Δ_1, is usually quite small, the signal processing delay, Δ_2, could be fairly long since it involves two interleaving/deinterleaving processes. Δ_2 could be expressed as,

$$\Delta_2 = \Delta_{STL-DeIntlv} + \Delta_{SER-Intlv} + \Delta_{res} \tag{1}$$

where $\Delta_{STL-DeIntlv}$ and $\Delta_{SER-Intlv}$ are the delays introduced by the deinterleaving for STL decoding and the interleaving for service signal encoding, respectively, and Δ_{res} is the delay introduced by all the other signal processing components.

Since STL signal transmission faces near LOS channel and high SNR conditions, it is possible to use a short time-domain interleaver, or even disable it. Therefore, $\Delta_{STL-DeIntlv}$ could be a small delay as well.

The length of $\Delta_{SER-Intlv}$, however, could vary significantly depending on what type of services the STL data is used for. For fixed services, this delay could be a fraction of an ATSC 3.0 frame length. For extremely robust mobile services, the time-domain interleaver could cover multiple frames and the delay could be multiple ATSC 3.0 frame durations. This is, however, a rare case in actual deployments.

In typical service scenarios, a time-advance of one ATSC 3.0 frame duration is sufficient for SFN operation.

III. WIDL Throughput Analysis

There are different signal structures to integrate the STL signal and the service signals into one ATSC 3.0 waveform. In this section, the WIDL throughput requirement is analyzed for different broadcast service scenarios.

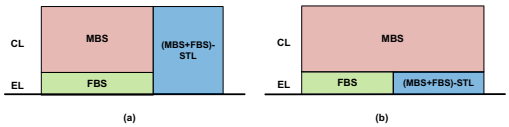


Fig. 6. WIDL signal models for both mobile broadcast service (MBS) and fixed broadcast service (FBS) distribution: (a) TDM; (b) LDM.

TABLE I
EXAMPLES OF WIDL SCENARIOS WITH TDM AND LDM

		WIDL (%)	MBS [Mbps]	FBS [Mbps]	STL [Mbps]
STL-SISO	**TDM**	33%	2.6	10.7	13.4
	LDM	41%	4.0	9.5	13.5
STL-MIMO	**TDM**	20%	3.2	12.8	16.0
	LDM	24%	4.0	12.1	16.1
No WIDL		0%	4.0	16.0	0

A. Multiplexing of Distribution and Service Signals

Two WIDL signal configurations are plotted in Figure 6, where the STL data is multiplexed with the service data using TDM (STL-TDM) and LDM (STL-LDM), respectively. Figure 6(a) shows the TDM configuration where the STL signal is assigned to a single-layer PLP. For the LDM configuration in Figure 6(b), the STL data is delivered in an LDM-EL PLP. The STL data contains service data for both mobile broadcast services (MBS) and fixed broadcast services (FBS).

Comparing these two configurations, the STL-TDM uses a higher power to transmit the STL data and therefore will have a higher SNR at the R-Tx, making it easier to decode. In addition, the STL-TDM requires low-complexity receivers at the R-Tx, since no SSC is needed to decode the STL data.

In comparison, the STL-LDM provides a higher mobile service capacity in the CL, as shown in Figure 6. Since the mobile services need to use strong coding and modulation to cope with the various fading conditions, they usually have lower bit rate. To keep the good QoE on mobile service, it is therefore desirable not to lose any mobile capacity.

The STL PLP throughput should be higher than the combined throughput of all service PLPs. This can be expressed as,

$$\begin{aligned} R_{STL} &= R_{MBS} + R_{FBS} + R_{ctrl} \\ &\cong R_{MBS} + R_{FBS} \end{aligned} \tag{2}$$

where R_{ctrl} is required for SFN control signaling.

Let's assume an original ATSC 3.0 system delivering two 720p-HD mobile service in CL and four 1080p-HD fixed services in EL. In TABLE I, the STL throughput and the service throughput are listed for both STL-TDM and STL-LDM, assuming a mobile service PLP with QPSK and rate-6/15 LDPC, a fixed service with 64QAM and rate-8/16 LDPC, and the STL PLP with 1k-QAM and rate-12/15 LDPC.

It is shown that using STL-LDM could keep the full MBS capacity of 4.0 Mbps, where using STL-TDM results in a 35% MBS capacity loss. On the other hand, for FBS services, using STL-LDM suffers a capacity loss of 11%.

MIMO technology could be applied to increase the spectral efficiency of the STL link [16]. Since the propagation model for the STL link is close to LOS channel, the most effective option is to use dual-polarized 2x2 MIMO. With the favorable SNR condition of the STL link, it is possible to achieve doubled capacity and significantly reduce the required radio resource for STL transmission.

TABLE I. also shows the transmission throughputs when MIMO is applied to the STL link. With MIMO, much smaller time percentage needs to be allocated for STL transmission. It is shown that using STL-LDM has a mobile service capacity gain of 25% over STL-TDM, and a fixed service capacity loss of 5.5%.

In TABLE II, some practical use cases for the STL-LDM signal configuration are listed. In all scenarios, a mobile service of HD-1080p is always delivered in CL. In EL, different scenarios are considered that deliver multiple full-HD services, a 4k-UHD, or the enhancement layer for a service with scalable-video coding, each requiring different throughput. For service delivery, it is assumed both the transmitter and the receiver have only one antenna, i.e., single-input-single-output (SISO), which is the typical commercial case. For STL link, both SISO and 2x2 MIMO are considered.

The first three scenarios in TABLE II. clearly show that the high transmission throughput available from ATSC 3.0 makes it feasible to transmit many mobile and fixed broadcast programs, and their STL data, in a single 6 MHz TV channel. These scenarios are also important for transition from ATSC 1.0 to ATSC 3.0, where one tower needs to transmit multiple programs from different broadcasters in one ATSC 3.0 signal.

It is shown that, even for delivering two (2) 720p-HD mobile service and one 4k-UHD fixed service (or four to six1080p-HD fixed services simultaneously), it is possible to use 40% of the time allocation to transmit the distribution data within the same ATSC 3.0 signal, with a single antenna on both A-Tx and R-Tx. However, high SNR is required at R-Tx since large constellation and high coding rate need to be used.

With 2x2 MIMO, a time allocation of less than 30% could be used to deliver the distribution data, for the same scenarios. The SNR requirements becomes much lower since 256QAM and rate-10/15 LDPC code could be used.

For a second scenario where only the STL data of mobile services is delivered using WIDL to gap-fillers for indoor public areas or R-Tx's in remote areas, only a small percentage of time is required for the STL transmission. The STL-TDM and STL-LDM solutions for this scenario are plotted in Figure 7; while the required STL signal configurations for different time allocations are calculated, for two mobile service qualities of 720p (2.7 Mbps) and 1080p (4.1 Mbps).

It is observed that, to deliver STL data for mobile service only, a time percentage of only 5 $\sim$ 10% is needed.

IV. IN-BAND DISTRIBUTION SIGNAL DETECTION

Figure 8 shows a block diagram of the STL signal detection at an R-Tx. The received signal from the A-Tx contains both the SFN service component (MBS+FBS) and the STL component. This signal will be referred to as forward signal (FWS)

TABLE II
IN-BAND BACKHAUL SCENARIOS FOR MIXED MOBILE AND FIXED SERVICES

Sce.	Mobile Service (SISO)	Fixed Services (SISO)		STL Configuration		
				SISO **TDM-40%**	2x2 MIMO **TDM-30%**	2x2 MIMO **TDM-20%**
1*	4.1 Mbps **(2~3) x 720p/30fps** or **(6~8) x 480p/30fps**	8.2 Mbps		12.3 Mbps 1k-QAM, L-9/15	12.3 Mbps, 256QAM, L-8/15	12.3 Mbps 1k-QAM, L-9/15
		(4~6) x 720p/60fps	**(2~3) x 1080p/60fps**			
2*		12.3 Mbps		16.4 Mbps 1k-QAM, L-12/15	16.4 Mbps 256QAM, L-10/15	16.4 Mbps 1k-QAM, L-12/15
		6~9) x 720p/60fps	**(3~5) x 1080p/60fps**			
3*	1.8 Mbps **1 x 720p/30fps** or **3 x 480p/30fps**	16.4 Mbps		20.5 Mbps 4k-QAM, L-11/15	20.5 Mbps 1k-QAM, L-9/15	20.5 Mbps 4k-QAM, L-11/15
		(8~12) x 720p/60fps	**(4~6)x1080p/60fps**			
4	3.4 Mbps **1080p SHVC-BL**	10.1 Mbps **4k-UHD SHVC-EL**		13.5 Mbps 1k-QAM, L-10/15	13.5 Mbps 256-QAM, L-9/15	13.5 Mbps 1k-QAM, L-10/15

* A 30% statistical multiplexing gain is assumed.

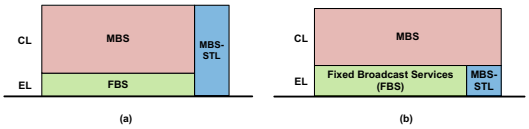


Fig. 7. WIDL signal model for mobile service distribution only: (a) TDM; (b) LDM.

TABLE III
WIDL SCENARIOS FOR MOBILE SFN ONLY

Sce.	R_{mob} (Mbps)	STL Signal Configuration				
		50%	25%	15%	10%	5%
1	2.7	16Q 4/15	16Q 7/15	64Q 9/15	256Q 8/15	1k-Q 13/15
2	4.1	16Q 6/15	64Q 8/15	256Q 10/15	1k-Q 12/15	NA

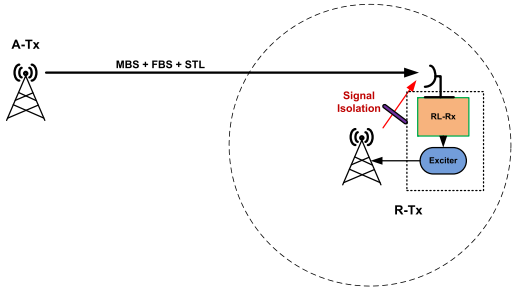


Fig. 8. Backhaul signal detection at RS-Rx.

for the rest of this paper. At the R-Tx, the RS-Rx decodes the STL data, from which the SFN service signal is generated and fed into the Exciter for emission. To achieve high SNR condition for the STL data detection, a powerful directional antenna is usually installed at the RS-Rx.

Figure 8 also shows a design challenge at the RS-Rx for STL detection. The RS-Rx receiver not only receives the FWS from the A-Tx, but also picks up the emission signal from the R-Tx transmission antenna, which is called loopback signal (LBS). Since the RS-Rx receive antenna and the R-Tx

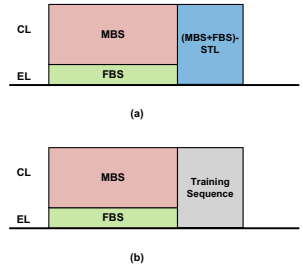


Fig. 9. Signal model at RS-Rx; (a) FWS; (b) LBS.

transmission antenna are usually installed on the same tower and both at high elevations, they are closely located. This results in a very high LBS signal power received at the RS-Rx receive antenna.

In Figure 9, the structures of the FWS and LBS signals received at the RS-Rx antenna are illustrated for an ATSC 3.0 SFN system with WIDL using STL-TDM. In the FWS, MBS and FBS are delivered in a two-layer LDM configuration in a specific time slot, while the STL data for both services is delivered in a different time slot.

During the time slot allocated for broadcast services, the LBS is the same as the FWS, because A-Tx and R-Tx delivers synchronized SFN services. However, during STL time slot, the FWS and LBS are different. Therefore, the LBS becomes a strong interference to the STL detection at the RS-Rx. This interference is also called self-interference in the in-band full-duplex relay for LTE/5G [18].

For OFDM-based ATSC 3.0 system, the received signal at the RS-Rx during the STL time slot can be expressed as,

$$\begin{aligned} Y_{RL}(k) = & X_{STL}(k) \cdot H_{FWS}(k) \\ & + \sqrt{\kappa_{LBS}} \cdot X_{LBS}(k) \cdot H_{LBS}(k) \\ & + N_0(k) \end{aligned} \tag{3}$$

where $X_{STL}(k)$ and $X_{LBS}(k)$ are the FWS and LBS symbols in the k^{th} subchannel, $H_{FWS}(k)$ and $H_{LBS}(k)$ are the channel responses of the forward and loopback channels, κ_{LBS} is the power ratio of the LBS over FWS, and $N_0(k)$ is the thermal noise.

Ideally, the self-interference could be removed if the R-Tx could transmit zero power during the STL time slot. This, however, is not allowed in the ATSC 3.0 standard. Even if allowed, this option is not desirable since part of the transmission capacity is wasted at the R-Tx, which can be used to deliver local services or STL to a subsequent R-Tx. When the R-Tx does not transmit service within the STL time slot, it needs to insert dummy symbols. Therefore, the RS-Rx needs to cope with the LBS self-interference to decode the STL data.

The impact of the LBS self-interference directly depends on the relative power ratio of the LBS over the FWS, κ_{LBS}, which depends on the A-Tx transmission power, the propagation distance from the A-Tx to the R-Tx and the R-Tx emission power. This results in a very large range of κ_{LBS} value. For a low-power SFN gap-filler (R-Tx) and a high-power A-Tx, κ_{LBS} could be quite small (close to 0 dB); while for a full-power R-Tx, κ_{LBS} could be well over 120 dB.

When the LB signal power is very high, it could easily saturate the RS-Rx's radio frequency (RF) front end (RFFE) and completely disable the RS-Rx operation.

To reduce the impact of the LBS on the STL signal detection, the RS-Rx could implement signal isolation mechanism between the transmit and receive antennas, followed by signal cancellation algorithm in the RS-Rx receiver.

A signal isolation (SI) mechanism could be designed to reduce the LBS power received by the RS-Rx antenna, as shown in Figure 8. Effective signal isolation could be achieved by antenna spacing, blocking, receive antenna shielding, and antenna directivity design. It has been shown that signal isolation of 70-100 dB could be achieved in practice [19], [20].

With the SI, the received signal at the RS-Rx becomes,

$$Y_{RL}(k) = X_{STL}(k) \cdot H_{FWS}(k) + X_{LBS}(k) \cdot H_{LBS}(k) + N_0(k) \quad (4)$$

For an R-Tx that emits medium to high power levels, even with a significant isolation, the residual LBS power could still be 10 to 20 dB higher than the received FWS, especially when the A-Tx is far away from the R-Tx. To further reduce the LBS component in (4), signal cancellation algorithm is carried out in the RS-Rx receiver as,

$$\begin{aligned} Y_{STL}(k) &= Y_{RL}(k) - X_{LBS}(k) \cdot \hat{H}_{LBS}(k) \\ &= X_{STL}(k) \cdot H_{FWS}(k) + \gamma_{LBS} + N_0(k) \end{aligned} \quad (5)$$

where $\hat{H}_{LBS}$ is the loopback channel (LBCH) estimate.

In (5), γ_{LBS} is the residual LBS signal after the signal isolation and signal cancellation, whose variation is calculated as,

$$\sigma^2_{LBS} = E\left\{\left|H_{LBS} - \hat{H}_{LBS}\right|^2\right\} = \nu^2_{H-LBS} \quad (6)$$

where ν^2_{H-LBS} is the mean square error (MSE) of the LBS channel estimate.

Therefore, the critical system components at the RS-Rx are the signal isolation and the signal cancellation modules.

A. Loopback Signal Isolation

The objective of the SI is to minimize the LBS power arriving at the RS-Rx receive antenna. This can be achieved by several approaches:

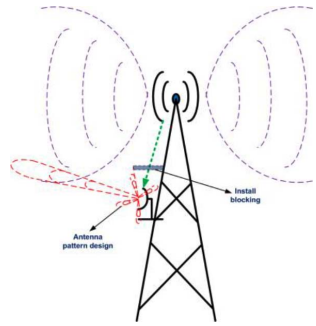


Fig. 10. Loopback signal isolation methods at R-Tx.

1) *Increasing the Antenna Spacing:* At the R-Tx, both the RS-Rx receive antenna and the transmit antenna are installed on the same tower at high elevations. The loopback signal propagation channel between these two antennas should be a LOS channel. The distance between them is at most a few hundred meters. For this distance, the propagation loss could be well modeled as free space path loss (FSPL), which is calculated as,

$$FSPL(dB) = 20\log(d) + 20\log\ (f) + 32.44 \quad (7)$$

where d is the distance in km, f is the frequency in MHz. Therefore, an antenna distance of 100 meters gives a LBS power 20 dB lower than that from a distance of 10 meters.
2) *Loopback Signal Blocking:* A metal shielding could be installed above the RS-Rx receive antenna, as shown in Figure 10. This is especially useful for some antennas with small form factors, such as the panel antennas. In [20], a metal mesh installed on top of a panel antenna is shown to be quite effective to block the loopback signal.
3) *Receive Antenna Directivity:* Modern antenna design could be applied on the receive antenna to have a Null towards the transmission antenna, further reducing the LBS power. However, this may require installation of multiple antenna elements and more engineering effort and larger space on the tower.

B. Loopback Signal Cancellation

To cancel the loopback signal as (5), the RS-Rx first needs to estimate the LBCH, H_{LBS}. To reduce the amount of radio resources required for STL data delivery, it is desirable to use high-throughput signal configuration for the STL signal, i.e., high-order modulation, high coding rate, MIMO, etc. This requires high SNR to decode and puts a high requirement on the LBCH estimation accuracy.

Because the RS-Rx knows exactly what signal is being transmitted during the STL time slot, as shown in Figure 9, the loopback channel can be estimated using decision-directed channel estimation (DD-CE) algorithms.

To perform DD-CE, the frequency-domain (FD) least square (LS) channel estimation is first obtained as,

$$\tilde{H}_{LBS}(k) = \frac{Y_{RL}(k)}{X_{LB}(k)} = H_{LBS}(k) + \frac{X_{STL}(k) \cdot H_{FWS}(k) + N_0(k)}{X_{LB}(k)} \quad (8)$$

where the received forward signal from A-Tx.

A two-dimensional filtering (2D-Filt) could be used to enhance the channel estimation accuracy. A frequency-domain filter (FD-Filt) is first applied on the LS estimates,

$$\hat{H}_{LBS-FD} = \Psi_F\{\tilde{H}_{LBS}\} \quad (9)$$

where Ψ_F could be a minimum mean square error (MMSE) filter [21], a singular value decomposition (SVD)-based filter [22], a DFT-Filter [23], a Wiener filter, or simply a smooth windowing function.

A time-domain (TD) filter is subsequently applied to further improve the accuracy of the channel estimate:

$$\hat{H}_{LBS-2D} = \Psi_T\{\hat{H}_{LBS-FD}\} \quad (10)$$

where Ψ_T is usually a Wiener filter or a smooth windowing.

Computer simulations were conducted to evaluate the achievable LBS cancellation performance, assuming a low-complexity DD-CE with 2D-Filt, which consists of a FD Wiener filter followed by a TD average windowing.

Since the STL receive antenna and the R-Tx transmit antenna are installed on the same tower and are closely located, the LBCH could be well approximated by a LOS channel. However, for cases where there are some obstacles surrounding the R-Tx tower, the LBCH could be modeled as a multipath channel with very short delay spread. Therefore, two channel models are tested in simulations, a typical LOS channel and a rare multipath channel, which is modeled as a Typical Urban [24] channel with a mean delay spread (DS) of 0.1 μsec, and a maximum DS of 0.7 μsec.

An ATSC 3.0 system with 16k transmission mode is used in simulations. It is assumed that the forward signal has an SNR of 25 dB at the RS-Rx receiver. Four LBS/FWS power ratios, [0 10 20 30] dB, are tested to evaluate the LBS cancellation performance under a wide range of operational conditions. Since the power ratio is for LBS after the signal isolation, a 30 dB of LBS/FWS is a rather pessimistic condition.

Figure 11 shows the LBS cancellation performance for a LOS channel. In this case, the Wiener filter in frequency-domain becomes an averaging window, where a window size of 500 taps is used. In time-domain, an averaging window of 40-taps is used. The upper subplot shows the MSE of the channel estimation after the 2D-Filt, while the lower subplot shows the SNR of the STL signal after the LBS cancellation.

For LOS channels, for all LBS/FWS power ratios, using this simple 2D channel estimator could achieve a residual loopback signal power of more than 40 dB lower than the FWS signal. Considering that the required SNR for high-throughput STL signal in the FWS is usually from 25 to 30 dB, this LBS cancellation performance is more than sufficient.

Figure 12 shows the LBS cancellation performance in a more challenging short TU channel, which could serve as a worse-case scenario. Due to the frequency selectivity, a Wiener filter of 50-tap is used in frequency-domain and a large time-domain windowing of 100 taps is used.

Even for this multipath channel, for all scenarios, the loop-back signal cancellation can reduce the LBS signal power to be 30 dB lower than the FWS signal, which is still enough for STL detection requiring an SNR of 25 dB.

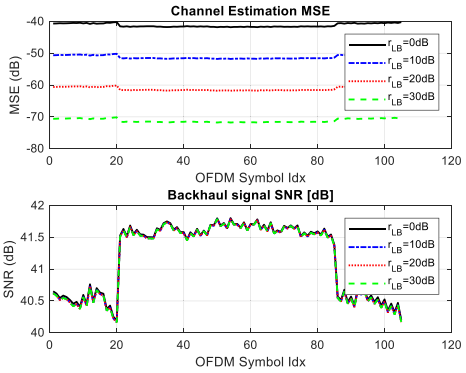


Fig. 11. LBS cancellation performance, LOS channel.

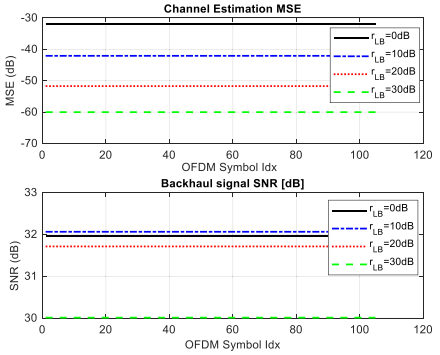


Fig. 12. LBS cancellation performance, short TU channel.

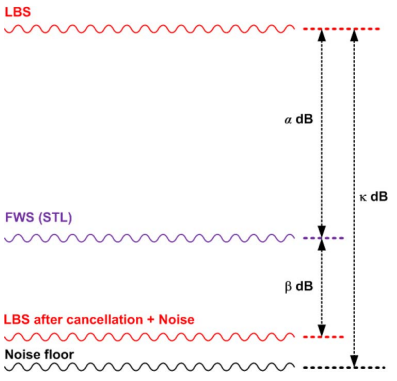


Fig. 13. RS-Rx dynamic range requirement.

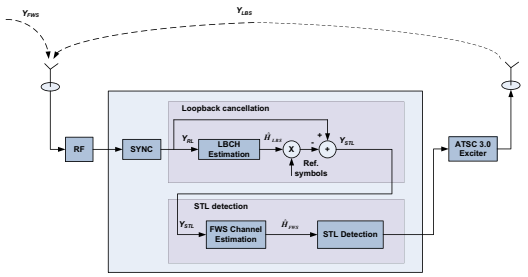


Fig. 14. RS-Rx receiver block diagram, WIDL-TDM.

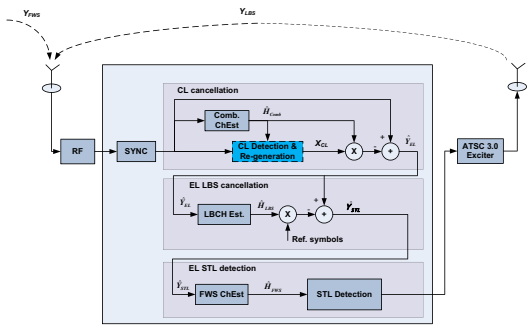


Fig. 15. RS-Rx receiver block diagram, WIDL-LDM.

It needs to be pointed out that the loopback signal only impacts the backhaul signal reception. It has absolutely no impact on the broadcast service reception by the ATSC 3.0 consumer receivers. The backhaul signals are simply ignored by the consumer receivers.

C. RS-Rx Dynamic Range Requirement

The previous analysis assumes that the RS-Rx has sufficient dynamic range to receive a signal consists of both the FWS and the LBS. When the power of LBS is much higher than that of FWS, the required dynamic range is also increased.

The dynamic range requirement of the RS-Rx is illustrated in Figure 13. Assuming the LBS is α dB higher than the FWS and the required SNR to decode the STL signal is β dB, the dynamic range of the RS-Rx, κ, should be larger than $\alpha + \beta$.

For example, for an LBS/FWS power ratio of 30 dB and a required SNR of 25 dB for STL detection, the receiver needs to have a dynamic range of at least 55 to 60 dB. While this is not impossible, it is quite challenging to implement.

To reduce the dynamic range requirement, the LBS power level needs to be reduced. This can be achieved by designing a more effective SI module using the technologies mentioned in Section IV-A. If an advanced SI module could lower the LBS power by 15 dB, the required receiver dynamic range becomes 45 dB for the above example. This is quite achievable for professional equipment.

A second solution to lower the LBS power at the receiver tuner input is to implement an analog LBS signal cancellation module, as proposed for on-channel repeaters (OCR) in [25]. The proposed method can reduce the LBS signal strength by up to 50 dB. The cost is the additional complexity.

It should be pointed out that for most multi-cell deployment scenarios, the emission power from the R-Tx is much lower than the A-Tx. With a good SI mechanism, an LBS/FWS power ratio of 30 dB is highly unlikely to occur. A reasonble range is from 0 to 15 dB.

D. STL Receiver Structure at R-Tx

A detailed receiver block diagram of the RS-Rx is plotted in Figure 14 for WIDL using TDM, which was shown in Figure 9. There are two major components in this receiver: the LBS cancellation module and the STL signal detection module. Two channel estimation (ChEst) modules are needed: a first ChEst for LBS cancellation and the 2nd ChEst to estimate the FWS channel for STL signal detection.

For WIDL-LDM, a more complicate receiver is needed which is plotted in Figure 15. An additional signal processing module is needed to first cancel the CL broadcast service signal. The CL signal cancellation does not pose a challenge, since the A-Tx and R-Tx transmit the same CL signal for SFN operation. The receiver simply performs the conventional successive signal cancellation, treating the combined FWS and LBS as a single 2-layer LDM signal.

TABLE IV
DTV CHANNELS FOR MEASUREMENT

Ch.	Ch. #	Fc [MHz]	H_{Tx}* [m]	Tx ERP [dBm]
Global Toronto	14	473	169.1	81.6
TVO	24	533	130.9	79.8
CBC Ottawa	25	539	197.8	84.9
Tele-Quebec	30	569	141.3	84.8
Radio Canada	33	587	197.8	83.8
V-Tele	34	593	141.3	74.8

* H_{Tx} is the height of the antenna relative to the bottom of tower.

V. A CASE STUDY IN OTTAWA

A measurement campaign was carried out in Ottawa area to evaluate the feasibility of implementing WIDL in an ATSC 3.0 SFN using an existing DTV tower as A-Tx. The DTV tower is a 229-meter tower located on top of mountain Camp Fortune, Chelsea, Quebec, with a sea level of 360 meters. Six DTV channels are evaluated, for which the parameters are listed in TABLE IV.

It is shown in Figure 17 that for all measurement points, the received FWS signal power is higher than -60 dBm.

Next, the loopback power levels for different TV stations are measured at the bottom of the transmission tower. The loopback power levels could be scaled to calculate the actual loopback power level as,

$$P_{LBS}(d_1) = P_{LBS}(d_0) + 20\log_{10}\left(\frac{d_0}{d_1}\right) \tag{11}$$

where d_0 is the height of the transmission antenna, d_1 is the distance of the RS-Rx antenna to the R-Tx transmission antenna, assuming d_1 is at least 10 meters.

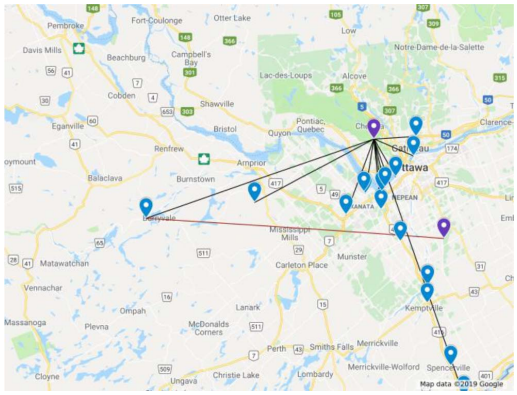


Fig. 16. Measurement campaign map in Ottawa.

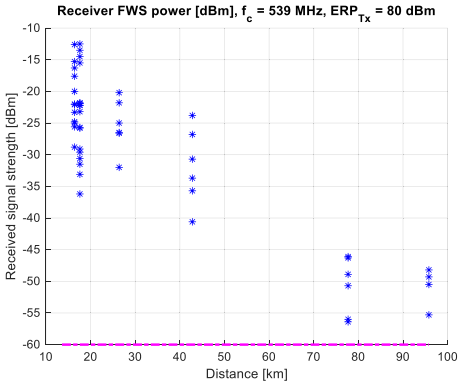


Fig. 17. Received FWS signal power.

Based on the measurement of the received signal power and the LBS power estimation from (11), the power ratio of LBS/FWS for different channels at different locations are calculated and plotted in Figure 18, for an R-Tx emission power of 70 dBm and a Tx-Rx antenna distance of 30 m.

It is shown that, even for a distance close to 100 km, the LBS over FWS power ratio is always less than 30 dB. From Figure 12, it is clearly shown that a low-complexity LBS cancellation module could achieve a residual LBS signal (after cancellation) 30 dB lower than the FWS signal.

This Case Study shows that the WIDL could be realized at all testing points, even those with elevations lower than 10 meters. The actual implementation with a high tower should have much better performance.

VI. Future Evolutions

There are many possible extensions to the proposed WIDL technology, to support more application scenarios and to achieve higher spectral efficiency.

A. Multi-Hop Distribution and Mesh Network

The concept of WIDL could be extended to deliver the distribution data over multiple hops, which could eventually

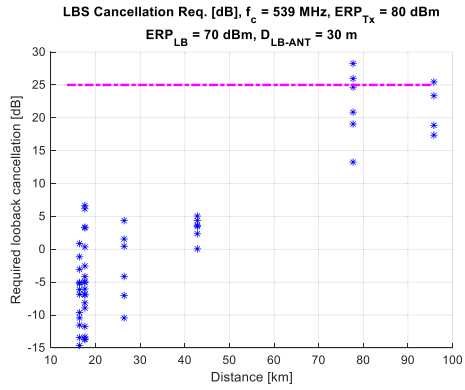


Fig. 18. P_{LBS}/P_{FWS} power ratio.

lead to a mesh distribution network in an SFN as shown in Figure 19(a). The A-Tx delivers the STL data to the two closest R-Tx in one hop, where the 1st hop R-Tx's then delivers the STL data to R-Tx's farther away. In this case, the longest multi-hop distribution link consists of three hops. To accomplish SFN transmission from the A-Tx and all the R-Tx's, the STL data delivered from the A-Tx needs to have a time advance of 3T over the service data emitted in the same ATSC 3.0 signal; in the signals emitted from the 1st hop R-Tx's, the STL data has a time advance of 2T; and so on for the R-Tx's in the following hops. The R-Tx's in the last hop delivers only service data.

In this mesh network, one R-Tx has multiple paths to obtain the STL data. A network management entity (NME) could help to organize the propagation path of the multi-hop distribution link for each R-Tx. In the case of one R-Tx fails, as shown in Figure 19(b), the NME could re-route the STL link for the subsequent R-Tx's original receiving distribution data from the failed R-Tx. Therefore, this multi-hop structure provides more failure tolerance in the entire SFN.

B. Three-Layer LDM

A third layer could be introduced to deliver the STL data for the CL mobile service only. This could reduce the capacity loss of the EL fixed services. Two configurations of using 3-layer LDM to deliver in-band STL data are plotted in Figure 20, where Figure 20(a) shows an LDM-STL solution and Figure 20(b) shows the STL-TDM solution.

The required STL time percentage and the transmission throughput of the mobile/fixed service and STL link are calculated and summarized in TABLE V. Comparing to the capacities for 2-layer solutions in TABLE I, one can see some saving in the fixed service capacity by using the 3-layer WIDL solutions.

C. Application of VVC

The next-generation video coding after HEVC (H.265) is called Versatile Video Coding (VVC) and will be standardized in 2020. The target of VVC is to achieve another 50% gain

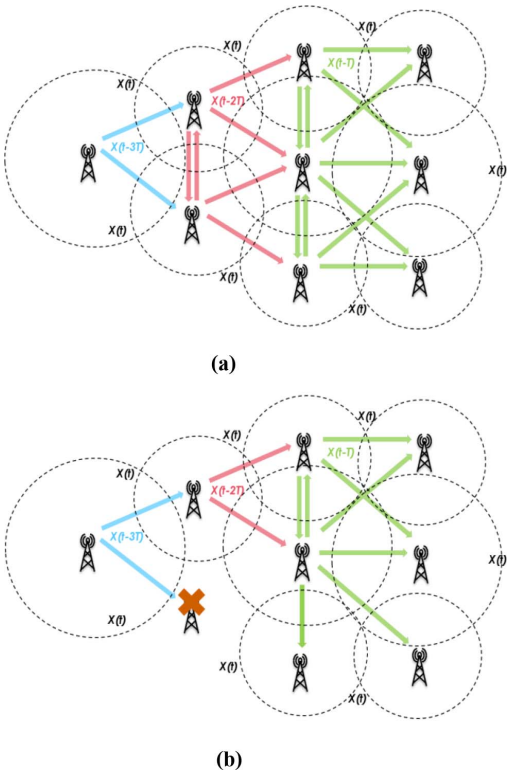


Fig. 19. Multi-hop WIDL SFN network; (a) SFN mesh network; (b) SFN mesh network with Tx failure and re-route.

in compression radio. Therefore, using VVC could reduce the required data rate by almost half for the same broadcast service video quality.

VVC offers a solution to further reduce the amount of radio resource to deliver the in-band distribution data. Without changing the video decoder at the consumer receivers, the video services could be encoded using VVC and distributed from the BGW to the remote broadcast transmitters. This would reduce the amount of required radio resource for STL transmission by almost half and leave more capacity for service delivery.

In order to use VVC to achieve higher spectral efficiency for STL delivery, a VVC to HEVC transcoder needs to be implemented at the R-Tx's. This would guarantee backward compatibility, i.e., no impact to consumer receivers.

VII. WIDL Versus On-Channel Repeater

On-channel repeater (OCR) is an existing cost-effective solution to provide terrestrial TV coverage area extension and to fill coverage holes [26]–[28]. In this section, the proposed WIDL technology is compared to existing OCR solutions to demonstrate its performance advantages and additional capabilities in the next generation DTV system.

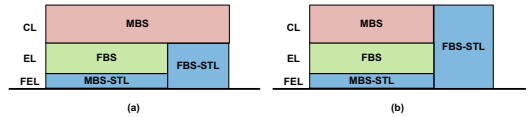


Fig. 20. WIDL signal model for both mobile and fixed services distribution only: (a) TDM; (b) LDM.

TABLE V
Examples of WIDL Scenarios With TDM and LDM

		WIDL (%)	MBS [Mbps]	FBS [Mbps]	STL [Mbps]
STL-SISO	**TDM**	33%	2.6	10.7	13.4
	LDM	41%	4.0	9.5	13.5
STL-MIMO	**TDM**	20%	3.2	12.8	16.0
	LDM	24%	4.0	12.1	16.1
STL-MIMO 3-Layer	**TDM**	17%	3.3	13.3	13.3/3.3
	LDM	20%	4.0	12.9	12.9/4.0
No WIDL		0%	4.0	16.0	0

A. On-Channel Repeater

An OCR usually receives a broadcast signal from a TV transmitter, amplifies it, and re-transmits the signal on the same TV channel. Since the power amplifier usually introduces additional noise into the received signal, the OCR output signal always suffers a lower SNR as comparing to the input signal. This makes it difficult to use OCR to cover large areas.

A significant challenge for OCR is the impact of the feedback signal from the re-transmission antenna to the receive antenna. Because the feedback signal is usually much stronger than the received signal, it could cause system oscillations and disable the operation of the OCR [27].

Therefore, reducing the feedback signal power is critical for proper operation of an OCR. Existing solutions in the literature include physical isolation methods and echo cancellation (EC). Physical isolation methods include antenna placement and antenna directivity, which has been shown to achieve a feedback signal attenuation of 30-50 dB. Further feedback signal isolation could be achieved by EC.

Two commonly used EC approaches are the decision-feedback based EC (DF-EC) and FIR-filter-based EC (FIR-EC). It is shown in [28] that using DF-EC could achieve 30 dB cancellation of the feedback signal. However, the DF-EC is designed for a single-carrier transmission such as ATSC 1.0, and cannot be directly applied to the OFDM-based ATSC 3.0 system [12]. The FIR-EC methods, although they are applicable to ATSC 3.0, could suffer significant noise enhancement due to the linear equalization process. This requires the input signal at the OCR to have very high SNR to achieve high-quality re-transmission.

All these limitations suggest that OCR should be used for isolated small areas, with also limitations on service qualities.

B. Wireless in-Band Distribution Link

In the proposed WIDL, an R-Tx decodes the received STL signal emitted from the A-Tx and re-transmits a clean ATSC

TABLE VI
WIDL VS OCR

WIDL	OCR
- Output clean signal	- Output SNR lower than input SNR
- Does not require high-SNR input	- Requires high SNR input
- Always provides stable operation	- Even with EC, stable operation is not guaranteed
- Could support medium to high output power, i.e., medium to large areas	- Limited output power, i.e., small areas
- Easy to implement multi-hop	- Difficult to support multi-hop links
- Supports STL/backhaul of local services and non-broadcast services, e.g., IoT, connected car	- No support for distribution of different services
- Some service capacity loss	- No service capacity loss

3.0 signal. The output signal quality is much better than that from an OCR.

Similar to OCR, a practical realization of R-Tx also faces the challenge of loopback signal from the transmit antenna to the RS-Rx receive antenna. However, because high-performance signal cancellation could be realized by digital signal processing, the loopback signal isolation requirement is more relaxed as comparing to the OCR solutions.

Therefore, WIDL could be used to deploy R-Tx for large coverage areas.

Furthermore, in WIDL-based systems, the STL link could be used to deliver distribution data for other types of services, in addition to the SFN broadcast services. In future DTV networks, different R-Tx could deliver location-based services to receivers within their own coverage area. These could include local news/commercials, IoT services, connected vehicles, etc., which cannot be realized by OCR.

C. OCR vs WIDL

In TABLE VI, the benefits and additional capabilities of the proposed WIDL technology over OCR are summarized.

It needs to be pointed out that, while OCR lacks many capabilities of WIDL, it serves as a cost-efficient solution to provide broadcast service coverage extensions over areas with favorable conditions. In addition, OCR can be combined with WIDL as the last hop solution in a multi-hop SFN scenario.

The advantages of WIDL are achieved at the cost of more signal processing in the RS-Rx for loopback signal cancellation and an RF tuner with high dynamic range. This added cost is insignificant considering the RS-Rx receiver being a professional equipment.

In addition, using WIDL results in less radio resources for service delivery. However, with the much higher transmission capacity offered by ATSC 3.0, there is plenty of capacity for both STL and services using one RF channel. This will become even less an issue when the next generation video coding (VVC) is adopted.

VIII. Conclusion

A wireless in-band distribution (WIDL) technology is proposed to enable fast and cost-efficient deployment of the next generation ATSC 3.0 single-frequency-network. In WIDL, the service distribution data is transmitted from an existing high transmission tower to the newly deployed SFN transmitters, sharing the same TV band with the broadcast services. To achieve higher spectral efficiency, LDM technology is proposed to multiplex the distribution data with the service data in the same ATSC 3.0 signal. Different signal configurations to realize the WIDL are presented with different capacity vs complexity tradeoffs. It is shown that the significantly higher capacity of ATSC 3.0 allows the transmission of distribution data over the same TV channel with the service data and still provides both mobile and fixed 4k-UHD broadcast services, i.e., more services than the existing ATSC 1.0 system at much higher qualities. A detailed study on the implementation of WIDL at remote SFN sites shows that a practical WIDL receiver could be designed using conventional signal isolation approaches and low complexity signal processing algorithms, i.e., no new algorithms need to be designed. A case study in Ottawa area is also reported to demonstrate the feasibility of the proposed WIDL technology.

The proposed WIDL technology is essentially a wireless backhaul solution which could play a key role to enable the success of the ATSC 3.0 system to deliver location-based services, IoT, connected vehicles, and others.

References

[1] *NGMN 5G White Paper*, NGMN Alliance, Frankfurt, Germany, Feb. 2015. Accessed: Nov. 25, 2017. [Online]. Available: https://www.ngmn.org/fileadmin/ngmn/content/downloads/Technical/2015/ NGMN_5G_White_Paper_V1_0.pdf

[2] L. Fay, L. Michael, D. Gómez-Barquero, N. Ammar, and M. W. Caldwell, "An overview of the ATSC 3.0 physical layer specification," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 159–171, Mar. 2016.

[3] A. Mattsson, "Single frequency networks in DTV," *IEEE Trans. Broadcast.*, vol. 51, no. 4, pp. 413–422, Dec. 2005.

[4] U. Meabe, X. Gil, C. Li, M. Velez, and P. Angueira, "On the coverage and cost of HPHT versus LPLT networks for rooftop, portable, and mobile broadcast services delivery," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 133–141, Jun. 2015.

[5] L. Zhang *et al.*, "Layered-division-multiplexing: Theory and practice," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 216–232, Mar. 2016.

[6] O. Bulakci, "On backhauling of relay enhanced network sin LTE-advanced," Licentiate Seminar, Dept. Commun. Netw., Aalto Univ., Espoo, Finland, 2010.

[7] "Further advancements for E-UTRA, physical layer aspects, V1.5.1," 3GPP, Sophia Antipolis, France, Rep. TR-36.814, Dec. 2009,

[8] C. Hoymann, W. Chen, J. Montojo, A. Golitschek, C. Koutsimanis, and X. Shen, "Relaying operation in 3GPP LTE: Challenges and solutions," *IEEE Commu. Mag.*, vol. 50, no. 2, pp. 156–162, Feb. 2012.

[9] C. Zarbouti, G. Tsoulos, and G. Athanasiadou, "4G multicell systems with in-band full duplex relays: Using beamforming to lower self-interference and/or transmitted powers," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC'2015)*, Glasgow, U.K., May 2015, pp. 1–6.

[10] P. P. Bergmans and T. M. Cover, "Cooperative broadcasting," *IEEE Trans. Inform. Theory*, vol. 20, no. 3, pp. 317–324, May 1974.

[11] L. Zhang *et al.*, "Using layered-division-multiplexing for in-band backhaul for ATSC 3.0 SFN and gapfillers," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Valencia, Spain, Jun. 2018, pp. 1–6.

[12] L. Zhang *et al.*, "ATSC 3.0 in-band backhaul for SFN using LDM with full backward compatibility," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Jeju-do, South Korea, Jun. 2019, pp. 1–6.

[13] L. Zhang *et al.*, "Using non-orthogonal multiplexing for in-band full-duplex backhaul for 5G broadcasting," in *Proc. IEEE 5G World Forum (5GWF)*, Silicon Valley, CA, USA, Jul. 2018, pp. 1–6.

[14] K.-J. Kim *et al.*, "Low-density parity-check codes for ATSC 3.0," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 189–196, Mar. 2016.

[15] N. Loghin, J. Zöllner, B. Mouhouche, D. Ansorregui, J. Kim, and S. Park, "Non-uniform constellations for ATSC 3.0," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 197–203, Mar. 2016.

[16] D. Gómez-Barquero *et al.*, "MIMO for ATSC 3.0," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 298–305, Mar. 2016.

[17] L. Zhang *et al.*, "Layered-division-multiplexing for high spectrum efficiency and service flexibility in next generation ATSC 3.0 broadcast system," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 116–123, Apr. 2019.

[18] D. Kim, H. Li, and D. Hong, "A survey of in-band full-duplex transmission: From the perspective of PHY and MAC layers," *IEEE Comun. Survey Tuts.*, vol. 17, no. 4, pp. 2017–2046, 4th Quart. 2015.

[19] *Technical Report: Informe Pruebas Aislamiento En KONTRASTA*, ITELAZPI, Zamudio, Spain, Dec. 2012.

[20] *ATSC Recommended Practice: Design of Multiple Transmitter Networks*, document A/111:2009, Adv. TV Syst. Committee, Washington, DC, USA, Sep. 2009.

[21] J.-J. Van De Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *Proc. 45th Veh. Technol. Conf. (VTC'95)*, vol. 2. Chicago, IL, USA, Jul. 1995, pp. 715–719.

[22] O. Edfors, M. Sandell, J.-J. Van De Beek, S. K. Wilson, and P. O. Borjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. Commun.*, vol. 46, no. 7, pp. 931–939, Jul. 1998.

[23] L. Zhang, Z. Hong, and L. Thibault, "Improved DFT-based channel estimation for OFDM systems with null subcarriers," in *Proc. Veh. Technol. Conf. (VTC)*, Anchorage, AK, USA, Sep. 2009, pp. 1–5.

[24] Commission of European Communities, Directorate General, Telecommunications, and Information Industries and Innovation, *Digital Land Mobile Radio Communications. COST 207 Report*. Luxembourg, Europe: OPOCE, 1989.

[25] S. W. Kim, Y. T. Lee, S. I. Park, H. M. Eum, J. H. Seo, and H. M. Kim, "Equalization digital on-channel repeater in the single frequency network," *IEEE Trans. Broadcast.*, vol. 52, no. 2, pp. 137–146, Jun. 2006.

[26] M. Wang, "Dynamic gain management for on-channel repeaters," *IEEE Trans. Broadcast.*, vol. 59, no. 4, pp. 685–692, Dec. 2013.

[27] K. M. Nasr, J. P. Cosmas, M. Bard, and J. Gledhill, "Performance of an echo canceller and channel estimator for on-channel repeaters in DVB-T/H networks," *IEEE Trans. Broadcast.*, vol. 53, no. 3, pp. 609–618, Sep. 2007.

[28] S. W. Kim, Y. T. Lee, S. I. Park, H. M. Eum, J. H. Seo, and H. M. Kim, "Equalization digital on-channel repeater in the single frequency networks," *IEEE Trans. Broadcast.*, vol. 52, no. 2, pp. 137–146, Jun. 2006.

Liang Zhang (Senior Member, IEEE) received the bachelor's degree from the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China, in 1996, and the M.S. and Ph.D. degrees from the Department of Electrical and Computer Engineering, University of Ottawa, Ottawa, Canada, in 1998 and 2002, respectively. He is a Senior Research Scientist with the Communications Research Centre Canada, Ottawa. He has been deeply involved in the ATSC 3.0 standardization activities on developing the layered-division-multiplexing technology, mixed fixed and mobile broadcast service delivery, mobile service detection, co-channel interference mitigation, integrated access and backhaul. He is currently working on technologies for the convergence of future TV broadcast and 5G broadband systems. He has more than 70 peer-reviewed journal and conference publications and received multiple Best Paper Awards from IEEE, IBC, and NAB for his work on the next generation ATSC 3.0 and the 5G broadcasting systems. He is an Associate Editor of the IEEE TRANSACTIONS ON BROADCASTING, a IEEE BTS Distinguished Lecturer, and an Elected Member of the IEEE BTS Administrative Committee.

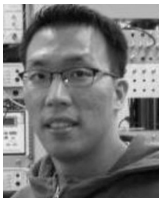
Wei Li (Member, IEEE) received the B.E. degree in electrical engineering from Shandong University in 1985, the M.S. degree in electrical engineering from the University of Science and Technology of China, in 1988, and the Ph.D. degree in electrical engineering from the Institut National des Sciences Appliquées of Rennes, France, in 1996. In 2001, he joined Communications Research Centre Canada (CRC), where his major focus is broadband multimedia systems and digital television broadcasting. He is currently a Research Scientist with CRC. He served as the Session Chair for the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting in 2006, 2007, 2015, and 2016, and a TPC Chair in 2016. He was the Managing Editor of the IEEE TRANSACTIONS ON BROADCASTING Special Issue on IPTV in Broadcasting Applications in 2009. He also served as a reviewer for several renowned international journals and conferences in the area of broadcasting, multimedia communication, and multimedia processing. He was the BTS IPTV representative at the ITU-T and the Co-Chair of Enhanced TV Planning Team at ATSC. He is an Associate Editor for the IEEE TRANSACTIONS ON BROADCASTING and IEEE ACCESS.

Yiyan Wu (Fellow, IEEE) received the M.Eng. and Ph.D. degrees in electrical engineering from Carleton University, Ottawa, Canada, in 1986 and 1990, respectively. He is a Principal Research Scientist of the Communications Research Centre Canada. He has more than 400 publications and received many technical awards for his contribution to the research and development of digital broadcasting and broadband multimedia communications. His research interests include broadband multimedia communications, digital broadcasting, and communication systems engineering. He is a Fellow of the Canadian Academy of Engineering, an Adjunct Professor of Carleton University, and Western University in Canada. He is a Distinguished Lecturer of the IEEE Broadcast Technology Society, and a member of the ATSC Board of Directors, representing IEEE. He was appointed as a Member of the Order of Canada in 2018.

Sébastien Laflèche received the B.A.Sc. degree in electrical engineering from the University of Ottawa in 1999 and upon graduation has joined the Communications Research Centre Canada. For over 15 years, he has been working in the evaluation and testing of advanced digital broadcasting systems. His experience includes single frequency network, mobile digital television, coverage simulation, and new broadcasting services. He expanded his knowledge by conducting research on 3GPP LTE. His research interests include digital broadcasting, spectrum sharing, LTE communications, and quality of service.

Zhihong Hong received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 1994, and the M.S. and Ph.D. degrees in electrical engineering from North Carolina State University, Raleigh, in 1998 and 2002, respectively. From 1999 to 2002, he was a Research Assistant with the Center for Advanced Computing and Communication, North Carolina State University. From 2002 to 2003, he was with the University of Wisconsin at Madison, as a Postdoctoral Research Associate. Since August 2003, he has been a Research Scientist with Communications Research Centre Canada. His research interests include advanced coding and modulation in wireless communications.

Sung-Ik Park (Senior Member, IEEE) received the B.S.E.E. degree from Hanyang University, Seoul, South Korea, in 2000, the M.S.E.E. degree from POSTECH, Pohang, South Korea, in 2002, and the Ph.D. degree from Chungnam National University, Daejeon, South Korea, in 2011. Since 2002, he has been with the Broadcasting System Research Group, Electronics and Telecommunications Research Institute (ETRI), where he is the Project Leader and the Principal Member of Research Staff. He has over 200 peer-reviewed journal and conference publications, and multiple best paper and contribution awards for his work on broadcasting technologies. His research interests are in the area of error correction codes and digital communications, in particular, signal processing for digital television. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON BROADCASTING and *ETRI Journal*, and a Distinguished Lecturer of IEEE Broadcasting Technology Society.



Jae-young Lee (Senior Member, IEEE) received the B.S. degree (High Hons.) in electrical and computer engineering from Rutgers University in 2001, the M.S. degree in electrical and computer engineering from the University of Wisconsin at Madison in 2003, and the Ph.D. degree in engineering science from Simon Fraser University in 2013. He joined Electronics and Telecommunications Research Institute in 2003, where he is currently a Principal Research Associate with Broadcasting Systems Research Group. His research interests are in the areas of digital signal processing for various applications, including digital broadcasting, telecommunications, and human–computer interaction systems.

Heung Mook Kim (Member, IEEE) received the B.S. and M.S. degrees in electronics and electrical engineering from POSTECH, Pohang, South Korea, in 1993 and 1995, respectively, and the Ph.D. degree from KAIST, Daejeon, South Korea, in 2013. He was with POSCO Technology Research Laboratory, Pohang, from 1995 to 2001, and Maxwave, Inc., Daejeon, from 2002 to 2003. Since 2004, he has been with Electronics and Telecommunications Research Institute, Daejeon, where he is the Managing Director of Media Transmission Technology Research Group. He is currently a Vice-Chair of Broadcasting Technical Committee (TC8) of TTA, South Korea, and Managing Committee of Future Broadcast Media Standard Forum, South Korea, and the Chair of Physical Layer Group of FOBTV Technical Committee. His research interests are in the areas of digital and RF signal processing and multi-platform contents delivery for digital broadcasting and broadband multimedia communications.

Namho Hur (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from the Pohang University of Science and Technology, Pohang, South Korea, in 1992, 1994, and 2000, respectively. He is currently with the Broadcasting and Media Research Laboratory, Electronics and Telecommunications Research Institute, Daejeon, South Korea, where he is the Project Leader of "Development of Transmission Technology for Ultra High Quality UHD." He was the Chair of Future of Broadcast Television Initiative Management Committee from April 2018 to April 2019. His research interest includes realistic digital broadcasting systems.

Eneko Iradier (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications engineering from the University of the Basque Country (UPV/EHU) in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree. Since 2015, he has been part of the TSR Research Group, UPV/EHU. He was with the Communications Systems Group of IK4-Ikerlan as a Researcher from 2017 to 2018. During his doctoral studies, he did an internship with Communications Research Centre Canada, Ottawa. His current research interests include the design and development of new technologies for the physical layer of communication systems and broadcasting in 5G environments.

Pablo Angueira (Senior Member, IEEE), received the M.S. and Ph.D. degrees in telecommunication engineering from the University of the Basque Country, Spain, in 1997 and 2002, respectively. He joined the Communications Engineering Department, University of the Basque Country in 1998, where he is currently a Full Professor. He is part of the staff of the Signal Processing and Radiocommunication Lab, where he has been involved in research on digital broadcasting (DVB-T, DRM, T-DAB, DVB-T2, DVB-NGH, and ATSC 3.0) for more than 20 years. He has coauthored an extensive list of papers in international peer-reviewed journals, and a large number of conference presentations in digital broadcasting. He has also coauthored several contributions to the ITU-R working groups WP6 and WP3. His main research interests are network planning and spectrum management applied to digital terrestrial broadcast technologies. He is currently involved in research activities related to broadcasting in a 5G environment. He is an Associate Editor of the IEEE TRANSACTIONS ON BROADCASTING, a member of the IEEE BMSB International Steering Committee, and a Distinguished Lecturer of the IEEE Broadcast Technology Society. He serves on the Administrative Committee for the IEEE Broadcast Technology Society.

Jon Montalban (Member, IEEE) received the M.S. and Ph.D. degrees in telecommunications engineering from the University of the Basque Country, Spain, in 2009 and 2014, respectively. He is part of the Radiocommunications and Signal Processing Research Group, University of the Basque Country, where he is an Assistant Professor involved in several research projects. He has hold visiting research appointments with the Communication Research Centre, Canada, and Dublin City University, Ireland. His current research interests are in the area of wireless communications and signal processing for reliable industrial communications. He is the co-recipient of several best paper awards, including the Scott Helt Memorial Award to Recognize the Best Paper Published in the IEEE TRANSACTIONS ON BROADCASTING in 2019. He has served as a reviewer for several renowned international journals and conferences in the area of wireless communications. He is currently serves as an Associate Editor for IEEE ACCESS.

3.3.3 Journal paper J9

This subsection presents a journal paper related with Contribution 3. The full reference of the paper is presented below:

- W. Li, L. Zhang, Y. Wu, Z. Hong, S. Lafleche, S-I. Park, S. Kwon, S. Ahn, N. Hur, E. Iradier, I. Bilbao, J. Montalban and P. Angueira, "Integrated Inter-Tower Wireless Communications Network for Terrestrial Broadcasting and Multicasting Systems," accepted for publication in IEEE Transactions on Broadcasting.

Then, the most representative quality indicator concerning this paper are listed below:

- Type of publication: Journal paper indexed in JCR and IEEExplore
- Area: Electrical & Electronic Engineering
- Ranking: 69/266 (Q2) based on JCR 2019
- Impact factor (JCR): 3.419

Integrated Inter-Tower Wireless Communications Network for Terrestrial Broadcasting and Multicasting Systems

Wei Li, *Member IEEE*, Liang Zhang, *Senior Member, IEEE*, Yiyan Wu, *Fellow, IEEE*, Zhihong Hong, Sébastien Laflèche, Sung-Ik Park, *Senior Member, IEEE*, Sunhyoung Kown, Sungjun Ahn, *Member*, Namho Hur, *Member, IEEE*, Eneko Iradier, *Graduate Student Member, IEEE*, Iñigo Bilbao, *Graduate Student Member, IEEE*, Jon Montalban, *Senior Member, IEEE*, and Pablo Angueira, *Senior Member, IEEE*

***Abstract*—This paper describes systems, devices, and methods to implement a bi-directional integrated inter-tower wireless communications network (IITWCN). The described technology can be implemented in combination with the Broadcast Core Network (BCN) in next generation broadcast eco-system and, therefore, support new business cases for broadcast operators such as the delivery of flexible datacasting services or support broadcast or point-to-point internet services. The introduced bi-directional inter-tower communications network (ITCN) extends the previous unidirectional in-band distribution links (IDL) and adopts the on-channel repeater (OCR) as a simplified backhaul solution in single frequency networks (SFN). The concept of the coordinated ITCN is also presented, aiming at future broadcast internet services. The ITCN provides a scalable and configurable network solution embedded in a broadcast system, which becomes independent from any non-broadcasting telecommunication infrastructure. The described technology partially relies on the infrastructure of the underlying broadcast/multicast network, using the allocated service channels without requiring additional frequency bands or a separate frequency band. The bi-directional inter-transmitter communication links are therefore referred to as integrated transmission links and the corresponding network as an integrated network.**

***Index Terms*—LDM, SFN, STL, In-band Backhaul, Wireless Inter-Tower Communications Network, OCR, Full Duplex Transmission, Broadband and Broadcasting Convergence, Broadcast Core Network**

I. Introduction

THE next generation digital TV (NextGen TV) broadcasting system (ATSC 3.0) is being deployed in North America. To provide robust mobile services to portable and handheld devices, advanced transmission technologies such as layered-division multiplexing (LDM) [1], low-density-parity-check (LDPC) coding, non-uniform constellation (NUC) modulation, up to 32k FFT OFDM and MIMO, are adopted in ATSC 3.0 [2].

LDM is a power-based multi-layer signal multiplexing technology that achieves higher cumulative transmission capacity when delivering multiple services with different quality requirements [1] [3]-[11]. In a typical application scenario for a two-layer LDM, the first signal layer (L1 or Core Layer in ATSC 3.0 [12]-[14]) is configured with stronger power to deliver robust signals, targeting mobile, handheld, and indoor receivers. The second layer (L2 or Enhanced Layer in ATSC 3.0 [12]) is configured with lower power to deliver high-throughput signals. The high-throughput of the L2 signal is achieved by using fixed receivers with rooftop or other powerful high-gain antennas, which can provide high signal-to-noise ratio (SNR) reception. The signal model of a two-layer ATSC 3.0 system is shown in Fig. 1. A typical application scenario is to simultaneously deliver time and frequency synchronized L1 and L2 signals super-imposed within the same frequency band. At the receiver, L1 signal is decoded using a normal single layer receiver, considering L2 as an interference. L2 signal is detected using successive signal cancellation (SSC), which decodes, re-encodes the L1 signal, and cancels it from the received LDM signal. After the SSC, the L2 signal detection follows the conventional signal detection process.

Single frequency network (SFN) [15] has been considered as an effective broadcast network deployment solution since it provides better spectrum efficiency, less power consumption, higher coverage reliability, and much reduced adjacent-cell

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2017-0-00081, Development of Transmission Technology for Ultra High Quality UHD), and in part by the Basque Government (the grant IT1234-19 and the PREDOC grant program PRE_2020_2_0105) and by the Spanish Government (project PHANTOM under the grant RTI2018-099162-B-I00 (MCIU/AEI/FEDER, UE)). (*Corresponding author: Wei Li*).

W. Li, L. Zhang, Y. Wu, Z. Hong, and S. Laflèche are with the Communications Research Centre Canada, Ottawa, ON K2H 8S2, Canada (e-mail: wei.li@canada.ca; liang.zhang@canada.ca; yiyan.wu@canada.ca; zhihonghunter.hong@canada.ca; sebastien.lafleche@canada.ca).

S.-I. Park, S.. Kown, S. Ahn, and N. Hur are with the Broadcasting System Research Department, Electronics and Telecommunications Research Institute, Daejeon 305-700, South Korea (e-mail: psi76@etri.re.kr; shkwon@etri.re.kr; sjahn@etri.re.kr; namho@etri.re.kr).

E. Iradier I. Bilbao, J. Montalban and P. Angueira are with the Department of Communications Engineering, University of Basque Country, 48013 Bilbao, Spain (e-mail: eneko.iradier@ehu.eus; inigo.bilbao@ehu.eus; jon.montalban@ehu.eus; pablo.angueira@ehu.eus).

interference. Varied ranges of operation modes in ATSC 3.0 are defined with different cyclic prefix (CP) lengths, allowing flexible SFN deployments, to offer broadcast services with various quality requirements over explicit terrestrial areas. Instead of high tower high power (HTHP) transmitter deployment, the SFN provides the low tower low power (LTLP) solution, making the most of the service coverage and quality [16]. A studio-to-transmitter link (STL) connection is in need for each new SFN transmitter to transfer data from the broadcast gateway (BG) [17]. The number of required STL connections will grow proportionally as more SFN transmitters are being deployed. The current STL solutions relying on fiber links or dedicated microwave links suffer from both poor accessibility and high infrastructure and operation costs. New STL solutions that can offer both low installation and operation costs are required.

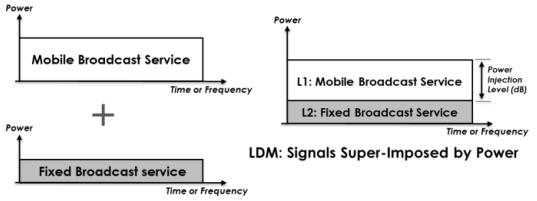


Fig. 1. A diagram illustrating a signal model of a 2-layer LDM system

An effective alternative using in-band backhaul (wireless in-band distribution link (IDL) technology) has been proposed in [18] [19]. It uses LDM to transmit the STL data via wireless links from the BG to the SFN transmitters within the same spectrum as the broadcast services. IDL guarantees full backward compatible ATSC 3.0 backhaul for SFN operation. Moreover, this solution is a complete in-band approach that provides backhaul for both robust mobile and high-data-rate fixed services. In addition, it has no impact on existing services for consumer TV receivers and it is a spectrum reuse technology that makes better use of the spectrum. The cost of applying IDL as backhaul in SFN is the potential partial occupancy of the transmitting spectrum, as well as a series of challenges such as loopback signal cancellation/isolation, transmitter timing control, etc.

Continuous efforts are carrying out on new solutions of fering low installation and low operation costs for SFN distribution links, as well as more controllable, scalable, and cost-effective STL technologies. In Section II, on-channel repeaters (OCR) are proposed to be used within ATSC 3.0 SFN, which provide simplified low cost solutions. In Section III, the bi-directional IDL solution is proposed, which is named bi-directional wireless Inter-Tower Communications Network (ITCN). The ITCN provides full duplex transmission among SFN transmitters, aiming at enriched data services including IoT, emergency warning, connected car, and other localized data services. In Section IV, the coordinated ITCN concept is introduced, which applies for multiple operators sharing the same SFN or multiple SFNs coordination. The ITCN node at a broadcast transmitter site is also described in this section. Integrated Inter-Tower Wireless Communications Network (IITWCN) indicates a large-scale wireless IP-network that inter-connects multiple SFN/ITCNs and Multi-Frequency Networks (MFN). In the last section, Section V, conclusions and discussions are presented.

II. Using OCR in ATSC 3.0 SFN

OCRs have been used in terrestrial broadcasting systems [20] to provide additional coverage or to fill coverage holes. OCRs are classified into analog types, which include radio frequency (RF) and intermediate frequency (IF) processing OCRs, and digital types, which include OCRs with feedback interference canceler (FIC) and OCRs with equalizer. An OCR may also combine analog and digital signal processing.

An OCR is a low-cost implementation of SFN, where no inter-tower distribution link is required. Fig. 2 shows functional blocks of a digital OCR example with an equalizer [20] configured for inter-tower communications (ITC) signal injection, e.g. at an added LDM layer. An OCR, in its simplest form, may include a receiving antenna, a band pass filter, low noise amplifier(s), and a power amplifier connected to the transmitting antenna. It is also possible to down-convert the input signal to IF. The same local oscillator can be used to up-convert the IF to maintain the synchronization between the transmitted and received signals. An optional module for down converting the signal to baseband gives the possibility of making error corrections before retransmission.

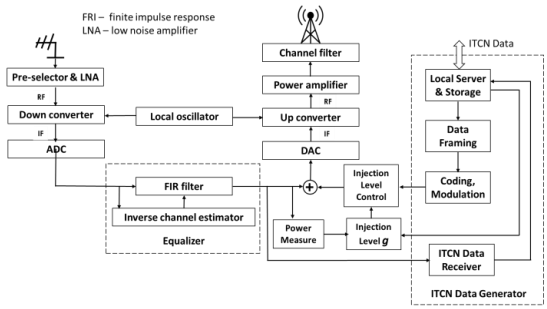


Fig. 2. A functional block diagram of an OCR with an ITCN signal injection.

In comparison to in-band backhaul solutions for a relay transmitter station (TS), the OCR has the advantage of simplicity and spectrum saving. It does not need to use service spectrum for backhaul signal delivery. An illustration of the OCR, rather than in-band backhaul link, is shown in Fig. 3. A targeted deployment is an SFN of broadcast services carried in the L1 and L2. In Fig. 3, the OCR transmitter Tx-C may be in a remote area or is located in a coverage hole of the SFN. Instead of deploying a dedicated fiber link or microwave link to carry the STL data, the broadcast L1 and L2 signals are picked up over-the-air from the other SFN transmitters, i.e., Tx-A or Tx-B, for processing and re-transmission by OCR. The operation of the OCR relay (Fig. 3) may be explained in detail as follows:

- The targeted SFN broadcast service to end users may be delivered in L1 and L2 from all transmitters: Tx-A, Tx-

B, and Tx-C.

- A directional high-gain antenna may be installed at a high location at the Tx-C, to achieve a good channel condition from Tx-A (or Tx-B), ideally a line of sight (LOS) channel.
- Since the receiving antenna of the OCR may be very close to the transmission antenna of Tx-C, sufficient signal isolation may be needed to prevent the Tx-C transmission signal from interfering with the signal detection process. If isolation is not high enough, it might also cause system instability;
- OCR may simply pick up the over-the-air signal and use signal processing in the analog [21] and/or digital domain [20] to eliminate the multipath distortion and loopback signal from the re-transmission antenna. The advantage of OCR is its simplicity since it does not need to decode the backhaul signal and generate the service signal.

The weak points may be: (1) since OCR cannot control the emission time, the signal processing delay needs to be relatively small, e.g., in tens of micro-second level, to reduce the multipath delay spread in the service area; (2) OCR may not be allowed to transmit very high power, e.g. under 500W, to prevent the system stability problem caused by loopback signal; (3) since there is no demodulation/decoding and remodulation/reencoding to clear up the signal, each OCR may cause some signal deterioration, which will accumulate in multi-hop OCR relay cases. The number of OCR relay stages may therefore be limited.

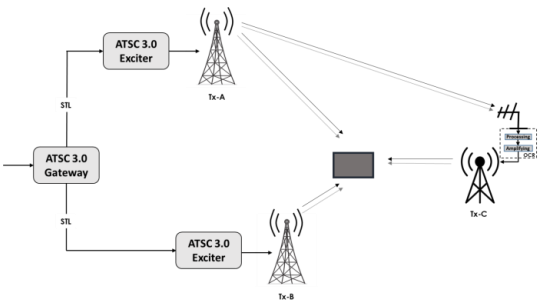


Fig. 3. A schematic diagram of an example ATC 3.0 Single Frequency Network with OCR as in-band relay.

An example scenario of ATSC 3.0 SFN with OCRs is illustrated in Fig. 4. This situation may happen when an SFN service area is not covered by the main transmitter Tx-A. Three OCRs receiving LDM L1 and L2 data from Tx-A form an SFN for a cost-effective solution.

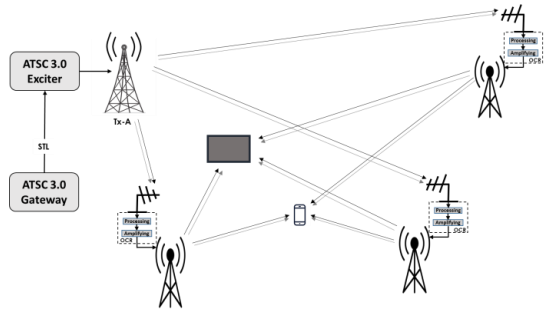


Fig. 4. A schematic diagram of an example ATC 3.0 SFN with three OCR relays.

III. Bi-Directional Inter-Tower Wireless Communications Network

The IDL technology [18] [19] provides a one-way in-band backhaul data transmission solution. In this section, it will be shown that sending the signal back in the same time frame is also possible. Remote SFN towers may be much higher than the 10-meter home reception antennas, so that there may be LOS paths among SFN towers, making it possible to implement a two-way inter-tower network. In particular, an in-band full-duplex transmission among towers can be implemented.

A. Two-Way Inter-Tower Communications Network

In Fig. 5, a simple ATSC 3.0 LDM based SFN with ITCN using bi-directional inter-tower communication (ITC) signals is illustrated. The ITCN shown may also support local datacasting. The operation of the in-band backhaul using two-layer LDM may be the same as explained in [18], or the backhaul may be provided via an STL link in some implementations. In the meantime, Tx-A can transmit ITC signals to Tx-B and vice versa, e.g. through LOS paths, which are multiplexed with conventional broadcast/multicast services (MS) signals, e.g. TV broadcast, and possibly in-band backhaul signals, using TDM, FDM and/or LDM. Each SFN tower can emit different ITCN data, transmitting different ITC signals on the same frequency band. Different types of services can be delivered embedded in ITC signals and can be received within each tower's broadcast coverage. These can include broadcast network cue and control data for network operation and monitoring, which are not intended for consumer services. ITCN data can also include consumer or professional service data, such as IoT, emergency warning, software download, connected car, and other localized data services or advertisement. The combination of TDM+LDM, or FDM+LDM, different modulation/coding, and reception conditions can provide tiered services for different robustness, data rates, and reception conditions.

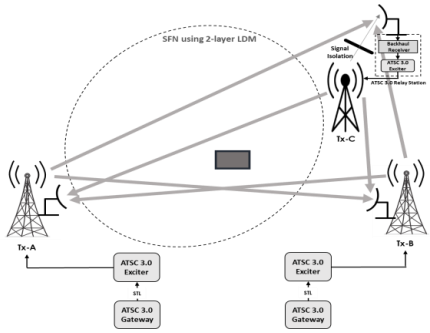


Fig. 5. A schematic diagram of ATSC 3.0 SFN with ITCN.

Different reception antennas might be used for over the air (OTA) ITCN signal reception to limit the co-channel interference from undesired transmission towers during the ITC transmission. Some towers can operate as SFN with main transmitters, if desired. In some scenarios, an ITCN can also connect transmission towers that are not part of the SFN, but nearby towers transmitting on a different frequency. This scenario makes signal reception easier since there is no loopback signal cancellation needed like for SFN towers. The in-band backhaul and ITCN may be mutually independent. They can co-exist, or they can be implemented independently. In-band backhaul can be replaced with other communication methods, such as OCR. The ITCN is an embedded network in the broadcast infrastructure and can operate independently from any other telecommunications infrastructure, which makes it extremely robust.

Most of the datacasting content/information is non-real-time, which may be desirable to have a data server at each participating tower to store the datacasting content/information and broadcast out locally, e.g., as data-carousel. The timing and content can be controlled by an ITCN administrator via an ITCN server. This approach can also reduce the data rate requirement on the ITCN.

For the Tx-A to receive the ITCN data from Tx-B, since the receiving antenna may be very close to its broadcast transmission antenna, sufficient signal isolation may be needed to prevent the Tx-A transmission signal, also called the loopback signal, from interfering with the ITCN signal emitted from Tx-B [19].

Fig. 6 shows an example of two ATSC 3.0 LDM-based OCR SFN interconnected with ITCN. In this example, each OCR SFN (circled separately) is composed of three OCR relay stations, and the two SFNs are working generally independently, e.g., serving different geographical areas, and may be operating on different frequencies. All the towers in the Tx-A SFN are connected in star type of configuration [22]. The Tx-B SFN uses a ring topology. There are inter-SFN ITCN connections to link the two SFN networks, which could be between two main transmitters or between two OCRs. All main transmitters (i.e., Tx-A and Tx-B) and OCR-SFN transmitters can be connected via multi-hop transmission. The network topology is reconfigurable and scalable, if one tower fails the data can be re-routed to reach all towers.

ITCN services using the same frequency band can be established among the towers in each OCR SFN. To limit the co-channel interference during the ITCN period, different reception antennas may be used for over the air ITCN signal reception from different adjacent transmission towers. In this example, the ITCN transmission connections between Tx-A and Tx-B might be on different transmission frequencies. Tx-A and Tx-B may each combine an in-band transmission with out-of-band reception when communicating with each other.

Each transmitter site can also have a local server that may store data for control and local datacasting. All OCR transmitters can be replaced by SFN transmitters (non-OCR) with in-band backhaul or other backhaul methods. But the ITC network remains the same.

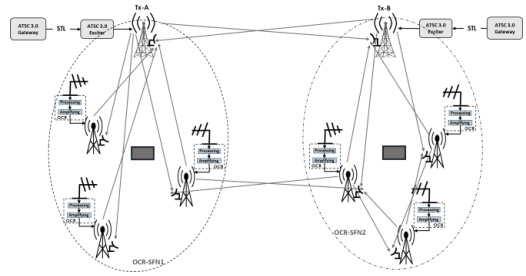


Fig. 6. A schematic diagram illustrating two interconnected ATSC 3.0 SFNs with OCR relays and ITCN.

B. *ITCN and Backhaul Signal Structures*

Two examples of a two-layer LDM signal structure for inter-tower network and datacasting (ITND) are illustrated in Fig. 7. Since the L2 signal is usually configured to have much higher throughput than the L1 signal, delivering the ITCN takes only part of the L2 signal capacity. The signal structure in Fig. 7(a) uses only one ATSC 3.0 physical layer pipe (PLP) [12], or time-frequency resource block, in the L2 to deliver the ITND, while the signal structure in Fig. 7(b) may use multiple PLPs, or time-frequency resource blocks, in the L2 to deliver the ITND. The ITND signal in structure in Fig. 7(b) may be more robust due to the better time-frequency diversity.

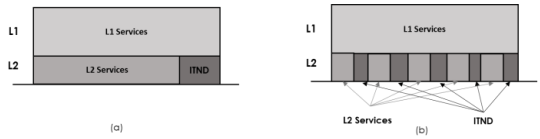


Fig. 7. Two-Layer LDM signal structure with ITCN in L2. (a) ITND in one PLP. (b) ITND in multiple PLPs.

The additional flexibility of layered structures of delivering the combined mobile, fixed, backhaul, and ITND services in two-layer LDM systems are shown in Fig. 8. There are three different services illustrated:

- SFN or non-SFN broadcasting services to mobile/fixed

terminals using LDM;

- SFN backhaul link (i.e., IDL): one-way, high data rate, high SNR for spectrum efficiency;
- ITCN: two-way communications among towers, and LDM datacasting in each tower coverage area, while each tower can transmit different data. Inter-tower networking can operate as SFN or as non-SFN multi-frequency network. ITCN could be used, for example, for datacasting.

In Fig. 8, these three services can be TDM, FDM or LDM multiplexed depending on application scenarios for efficient use of spectrum. Fig. 8(a) shows ITND TDM with LDM of L1 and L2 data. Fig. 8(b) shows LDM of two ITNDs TDM with LDM of L1 and L2. Fig. 8(c) depicts TDM of ITND and fixed broadcast services (FBS) as L2 data. Fig. 8(d) illustrates ITND TDM with LDM of mobile broadcast services (MBS) and fixed broadcast service (FBS)/in-band backhaul. Fig. 8(e) shows the TDM of three services, including LDM of mobile broadcast service (MBS)/FBS, ITND, and in-band backhaul. Fig. 8(f) depicts LDM of L1 (which is composed of MBS and ITND) and L2 (which is composed of FBS and in-band backhaul data). Fig. 8(g) illustrates LDM of MBS and L2 (which is composed of FBS, in-band backhaul, and ITND services).

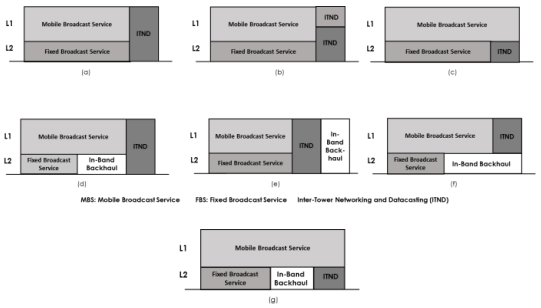


Fig. 8. Examples of two-layer LDM signal structures for delivering mobile, fixed, backhaul, and ITCN services.

In the system design, using antenna diversity allows communication with different towers (possibly at reduced SNR or data rate due to co-channel interference). ITND services can also run on multi-frequency towers, and under this circumstance, no signal cancellation is required. The system may be much simplified and with reduced system delay.

C. ITCN Interference and Control

In some scenarios, ITCN data may include three parts: (1) Network control, monitoring, and signaling data to operate the ITCN network and distributed data servers; (2) Backhaul data among transmission towers; (3) Datacasting service, where each tower can emit different data service to users. ITCN receiving antenna might receive multiple signals from more than one Tx tower, which may lead to co-channel interference. A smart adaptive antenna may be used to reduce the co-channel interference levels and multipath distortions. It will increase the ITCN data transmission capacity.

Another approach is that an ITCN network supporting datacasting may have the other towers' transmission data stored at the receiving tower's ITCN server, which can be used to generate an interference waveform that can be used to cancel or reduce the impact of the co-channel interference signal. More discussions on signal processing and cancellation issues can be found in [23].

Most of the datacasting content/information is non-real-time, and it may be desirable to have a data server at a participating tower to store the datacasting content/information and broadcast out locally as a data-carousel. The timing and content can be controlled via the ITCN network by an ITCN server. This can also reduce the data rate requirement on the ITCN.

D. Using Three-layer LDM for ITCN and/or In-band Backhaul

When using L2 of a two-layer LDM for in-band backhaul and ITCN, there is a penalty on L2 service throughput reduction, as shown in Fig. 7. One solution to avoid the reduction of L2 throughput is to use a three-layer LDM structure, where the third layer (L3) is added for ITCN for data backhaul. This solution is shown in Fig. 9, where Tx-B/C are relay stations.

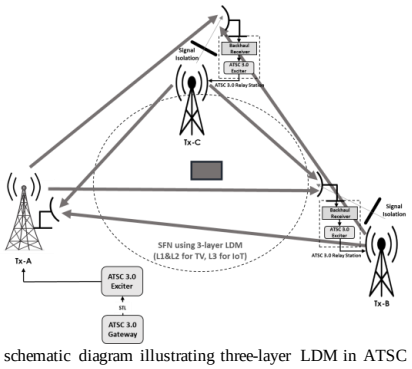


Fig. 9. A schematic diagram illustrating three-layer LDM in ATSC 3.0 SFN with ITCN.

The operation of an ITCN and in-band backhaul using 3-layer LDM in some Multicast/Broadcast Communications System (MBCS) scenarios may be as follows:

- The targeted SFN broadcast service is delivered in L1 and L2 from all transmitters: Tx-A, Tx-B, Tx-C;
- The L1 data backhaul or datacasting is delivered using L3 to an ATSC 3.0 In-Band Relay Station (IBRS) mounted on Tx-B/C. The IBRS would decode the backhauled data and re-transmit in broadcast SFN signal format and synchronize it with other SFN transmitters;
- To synchronize the emission of the L1 and L2 broadcast signals from the three transmitters, the backhaul in L3 needs to be sent to Tx-B/C earlier than the L1 and L2 data transmission time. The timing difference is easier to be set at one ATSC 3.0 frame, or any other time advance values that allows the IBRS the time needed to decode the L3 backhaul signal, re-encode and modulate the L1 signal for SFN emission;
- The L1 and L2 backhaul may transmit both the:

- data for L1 and L2 SFN services;
- control signaling to configure the transmission parameters of Tx-B/C, for SFN transmitter coordination, including the timing offset of Tx-B/C to optimize the SFN coverage.

- A directional high-gain antenna may be installed at a high altitude location at the IBRS, to achieve good channel condition from Tx-A/B/C, ideally a LOS channel;
- Since the receiving antenna of the IBRS may be very close to the transmission antenna of Tx-B/C, sufficient signal isolation may be needed to prevent the high-power Tx-B/C transmission signal from interfering the signal detection at the IBRS;
- More accurate L1 and L2 signal cancellation and equalization algorithms at the IBRS may be used to help to decode the backhaul signal under the interference from the Tx-B/C transmission signal;
- Two-way full-duplex ITCN services are carried out among transmitters Tx-A/B/C, through LOS paths, multiplexed either TDM or LDM with backhaul signal. Tx-A/B/C can emit different data that operating like 4G/5G networks [24]. The combination of TDM+LDM, different modulation/coding and reception conditions can provide tiered services for different robustness and reception conditions;
- L3 is not in the ATSC 3.0 standard, but it will be backwards compatible with all TV sets sold on the market, which are only required to detect L1 and L2 signals.

Examples of three-layer LDM signal structure with ITCN and in-band backhaul are illustrated in Fig. 10. In these cases, the L3 signal is a low power signal. Its data rate depends on the L1 and L2 signal cancellation accuracy and system linearity.

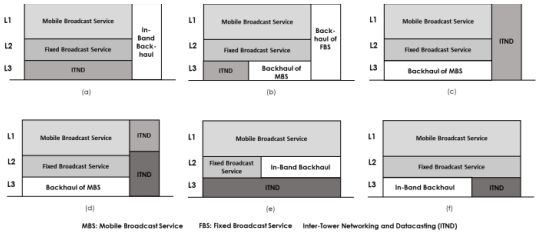


Fig. 10. Examples of three-layer LDM signal structure with in-band backhaul and ITCN in L3.

When using OCR, there is no need for in-band backhaul, Fig. 10 can be much simplified. One easy implementation for ITCN with OCR is to insert ITND signal as the L3 layer. Fig. 11 illustrates the signal structure of L3 ITCN. In some implementations, the L3 can be time-frame interlaced for two-way ITCN in a half-duplex mode, rather than full-duplex mode. The time and which transmitter to emit the L3 ITND in a data frame can be pre-arranged or re-configured by the ITCN network servers. Each transmitter emits three-layer LDM or two-layer LDM over the time to form an ITCN network.

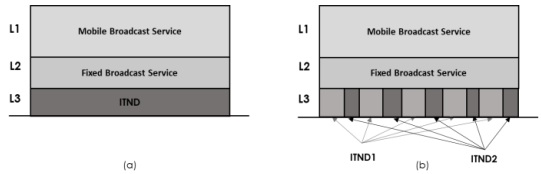


Fig. 11. Three-layer LDM signal structure for OCR-ITND service in L3: (a) one ITND. (b) multiple ITNDs.

L3 ITCN can also operate in full-duplex scheme, e.g., if the OCR inserted L3 ITCN data are using a very robust coding and modulation mode, where the SNR threshold of the L3 layer is a negative dB value, e.g., -3dB. OCR may simply insert the L3 ITCN signal as superimposed on the received and processed broadcast signal at its L3 power level. In this case, the L3 data rate may be low due to the robust coding and modulation.

E. Multi-Hop Relay

Multi-hop relay can also be realized, where an example of a two-hop OCR relay system is shown in Fig. 12. The OCRs on Tx-B/D receive the data and control signals from Tx-A, which are processed and re-transmitted to provide broadcast services. Tx-B relays the broadcast signal as the second hop to Tx-C. The consumer receiver receives broadcast signals from all transmitters Tx-A/B/C/D. Tx-B/D communicate with Tx-A over the ITCN channel, i.e., forming a looped network infrastructure.

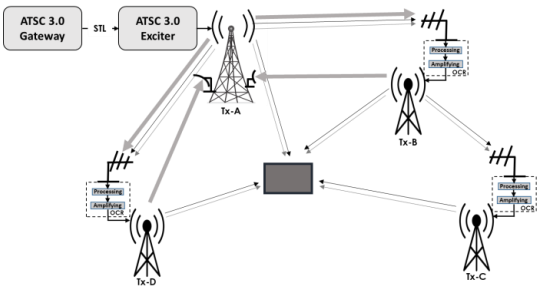


Fig. 12. An example ATSC 3.0 SFN with two-hop OCR relay using LDM.

IV. Coordinated Inter-Tower Wireless Communications Network

For a single broadcast operator, the in-band ITCN solution could be used where the ITCN links share the same TV channel with the broadcast services. This enables each broadcast operator to implement its own ITCN within its own channel(s), even with a single TV channel. However, for in-band solutions, the remote SFN transmitters may be affected by the loopback signal, which is the leakage signal from the broadcast transmission antenna into the ITCN receive antenna since the broadcast transmission signal may be continuously present in the time domain. With the high capacity offered by ATSC 3.0, especially with the two-layer LDM, multiple broadcast operators can multiplex their programs in fewer channels. Furthermore, shared SFN infrastructure may be more feasible

for different operators in the same area.

An alternative solution is to implement a combined ITCN using a separate TV channel for multiple broadcast operators. Each operator shares the ITCN capacity by TDM and/or FDM. In this case, time or frequency division duplex modes could be used for the bi-directional transmissions, which removes loopback signals. For the full-duplex transmission mode, highly directional antennas could be used for ITCN links that can significantly reduce the loopback signal power, potentially removing the requirement for loopback signal cancellation.

An example of the coordinated ITCN with two operators sharing the same SFN infrastructure is shown in Fig. 13.

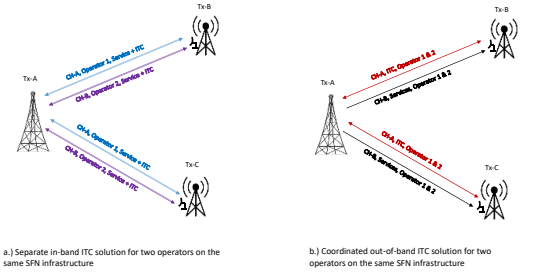


Fig. 13. Coordinated ITCN with two broadcast operators sharing the same SFN transmitters.

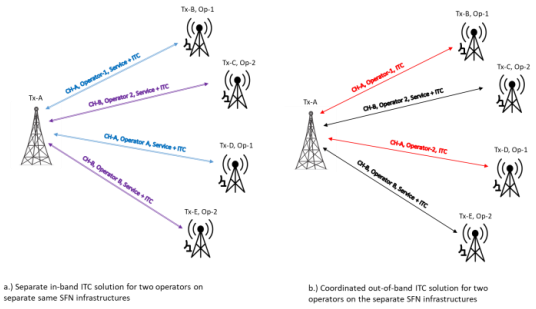


Fig. 14. Coordinated ITCN with two broadcast operators with separate SFN transmitters.

A second example of the coordinated ITCN solutions with two operators having separate SFN infrastructures is shown in Fig. 14. In this case, the ITCN links for the two broadcasters from remote SFN transmitters (Tx-B and Tx-D) to the anchor transmitter Tx-A need to be time and frequency locked. This can be achieved by using GPS time reference to control the transmission timing at Tx-B and Tx-D, which already exist in SFN transmitters.

Fig. 15 illustrates a MBCS example including local ITCN data servers at each participating TS. One or several ITCN data servers may include ITCN data storage for storing various types of data that may be transmitted to users in the TS broadcast area and another TS for re-transmitting. The stored data may include datacasting data, local advertisement, ITCN data such as ITCN backhaul among towers, local IoT data to consumers, and others. The ITCN data server may be configured to implement datacasting and/or a data carousel. Datacasting examples may include datacasting different software upgrades, local emergency alerts, weather updates, local news, etc. It takes advantage of broadcast systems one-to-many distribution capacity. Non-real-time (NRT) data, which may include TV program-related data and other data, e.g., for datacasting services, that are not TV program-related.

In some scenarios, TS participating in an ITCN network may coordinate their operations to conduct datacasting at a certain time frame. Different towers can emit different data content at the same time in an SFN environment. Towers that are not operating in a SFN may also be connected by ITCN to other towers for performing control, monitor, diagnosis, data backhaul functions. An ITCN network may be reconfigurable and scalable to extend the network or re-route ITCN data, if some network node/tower is out of service. Broadcasters may also coordinate to use a dedicated broadcast RF channel for ITCN only (without conventional broadcast service). For example, after midnight or in early morning, when traditional broadcast programs are not broadcasting (after-hour period), the entire radio frequency channel can be used for data communication/distribution. Participating broadcasters can share the network resource of the ITCN, for example using network slicing. Large and scalable inter-connected ITCN/SFN and MFN form the IITWCN, being part of the broadcast core network [25] for future broadcast internet applications.

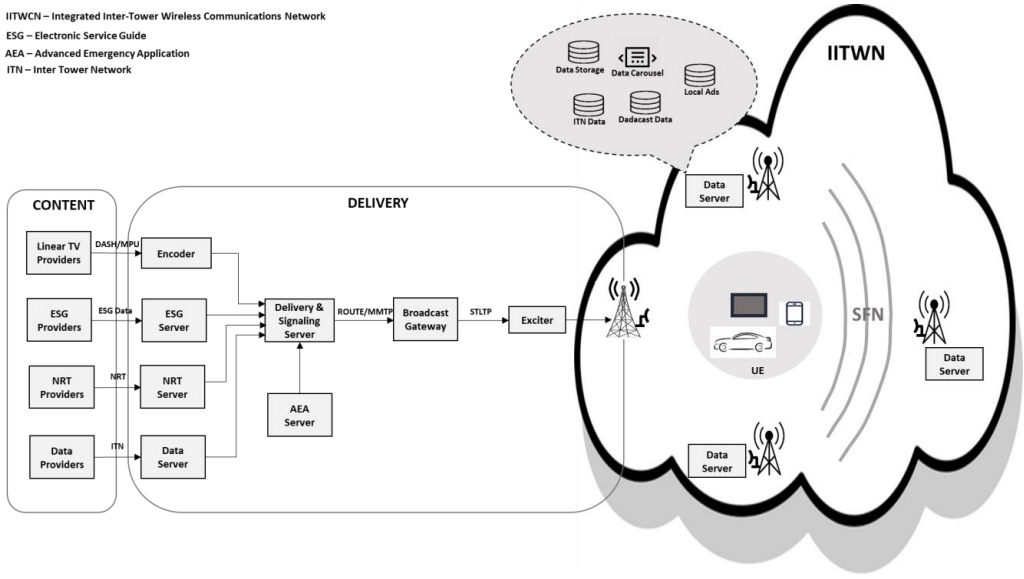


Fig. 15. Block diagram of a next generation TV station headend with MBCS containing an integrated ITCN data server.

Fig. 16 shows an example of an ITCN node at a broadcast transmitter site. An adaptive antenna can be used to maximize the main beam gain in the direction of the remote desired ITCN signal. The antenna adaptive system can also use the antenna nulls to reduce the multipath distortion of the desired signal, and more importantly to reduce the loopback signal and its multipath reflections from nearby structures. At the antenna array adaptive system block, artificial intelligence (AI) based machine learning (ML) and neural networks (NN) can be used to optimize the antenna array pattern by controlling each antenna elements' signal amplitude and phase using RF elements w1, w2, … wN. These elements may also perform RF filtering to reduce the adjacent channel interferences from other broadcast RF signals. The adaptive antenna array output is fed to loopback signal cancellation, where an RF analog feedback signal is used to cancel the loopback signal and related interferences and distortions.

The analog RF signal is converted to baseband for physical (PHY) layer processing, i.e., synchronization, channel estimation, digital loopback signal cancellation, demodulation, and decoding. A baseband signal feedback from the exciter is used to further suppress or cancel the loopback signal. The PHY processing output will go to the link layer processing. The output may be IP packets for the broadcast core network local server. The local server will store and control the broadcast program, backhaul, and ITCN data for relay to next node(s), datacasting data, local data, and other control data. It can also connect to a local gateway to other IP-based networks. The local network server is also fed by the broadcast scheduler and exciter to generate the RF broadcast signal waveform, followed by an RF power amplifier to generate the transmission signal to feed the broadcast transmission antennas.

An ITCN node is similar to gNB in 5G. It provides the link between UEs and Broadcast Core Network, interfaces with broadcast PHY, and communicates with other ITCN nodes, therefore it can be seen as Broadcast Node (bNB or bcastNode [25]) - to be equivalent as gNB in 5G.

The MBCSs example shown in Fig. 16 differs from the existing wireless broadcast systems in some aspects, including:

- providing a two-way bidirectional ITCN that is integrated with a wireless broadcast infrastructure;
- having one or more embedded ITCN network servers, for example, an ITCN network server may be provided at each broadcast tower that is part of the ITCN, and may be configured to perform node and/or ITCN control functions, storage, and communicate with other ITCN towers, as well as coordinate with other towers to data carousel and datacasting local contents;
- using adaptive smart antenna arrays to improve the system performance and capacity;
- using analog interference pre-cancellation in an integrated ITCN system to improve system performance;
- using OCRs as an option to provide a physical link for ITCN and to replace in-band backhaul, thereby saving spectrum for broadcast services.

Although the example ITCNs described here primarily concern SFN operation, an ITCN may also be implemented in a multi-frequency environment, where loopback signal cancellation function is not required. Fig. 17 depicts an out-of-band ITCN network node scenario, where transmitter and receiver are in two different frequencies, suitable for non-SFN applications. In this situation, no loopback signal estimation and cancellation are necessary, which saves the signal processing efforts quite a lot at the receiver side.

An ITCN may connect multiple SFNs using the same

frequency or different frequencies, and using the infrastructures of the respective MBCS networks. In some implementations, an ITCN may be implemented using a frequency band that is separate from the frequency band of the underlying SFN or SFNs (off-band). When implemented in an SFN environment, an ITCN may use one or more TV channels of the underlying TV broadcast network to transmit ITCN data between towers. This more complicated large-scale network is called IITWCN.

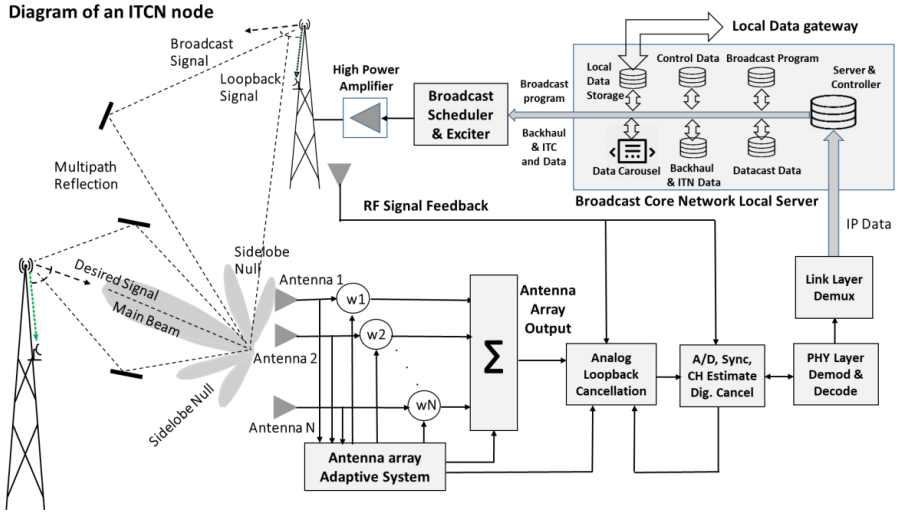


Fig. 16. A system block diagram of an in-band ITCN network node.

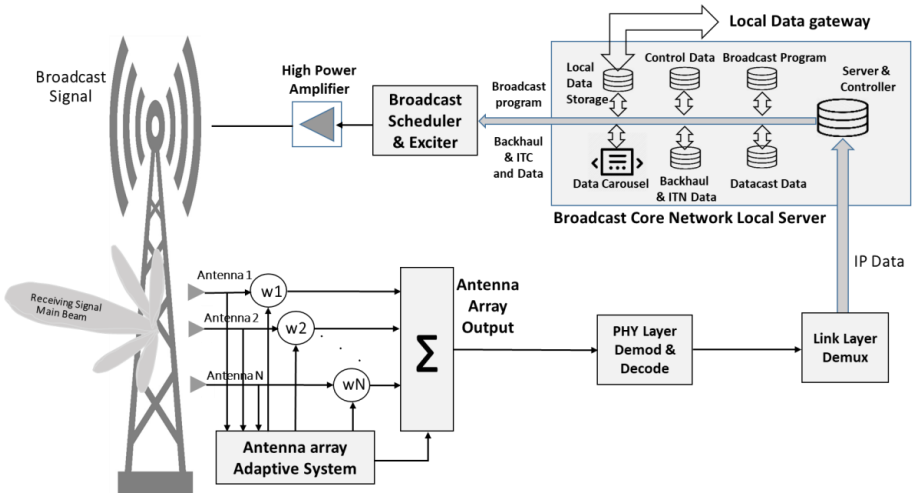


Fig. 17. A system block diagram of an out-of-band ITCN network node.

In some scenarios, one or more TV channels used for TV broadcasting during the day may be switched to ITCN communications during the TV off-hours of the TV broadcast, e.g. at nights or early mornings. In other scenarios, all or most TV channels could be switched to ITCN communications during the TV off-hours.

V. Conclusions

The concept of bi-directional integrated inter-tower wireless communications network (IITWCN) for terrestrial broadcasting and multicasting systems has been introduced. The system is backward compatible with the ATSC 3.0 NextGen TV services, causing no degradation of consumer reception. The bi-directional IITWCN is designed for inter-

tower signal transmissions of datacasting, broadcast internet, network cue, and control, etc., which specifies a large-scale wireless IP-network that inter-connects multiple SFN/ITCNs and MFNs. The wireless IITWCN extends the IDL into a scalable and configurable network embedded in a broadcast system, which becomes independent of any non-broadcasting communications infrastructure. Various IITWCN and backhaul signal structures are discussed, including the use of a third layer of LDM signal. The coordinated IITWCN was presented in the MBCS for full-fledged broadcast internet applications.

To summarize, IDL and IITWCN are part of the broadcast core-network. IDL distributes broadcast programs to transmit towers (SFN or non-SFN). It reduces broadcast operating costs by using part of the broadcast spectrum for one-way program distribution to transmission towers. Each tower will broadcast TV programs while receiving future program data at the same time on the same frequency. IITWCN is a two-way extension of IDL. It connects all broadcast towers (SFN or non-SFN) to form an IP-network for datacasting, network control, etc. Each tower can broadcast localized content, e.g., different SFN towers can emit different contents that operate like 4G/5G. All broadcast network towers form a scalable and reconfigurable IP-network embedded in a broadcast infrastructure independent to any telecommunication infrastructure.

TABLE I shows some salient aspects of SFN, IDL and IITWCN in ATSC 3.0 evolution on broadcast and broadband convergences.

TABLE I
COMPARISON OF SFN, IDL AND ITC NETWORKS

SFN broadcast	OCR or in-band distribution Link	ITCN
• Improve service quality for mobile, handheld and indoor receptions • Allow new service: IoT, connected car, datacasting • One-to-many timely services for large rural areas for traffic map update, weather forecast, emergency warning	• Eliminate studio-to-transmitter link and inter-SFN tower feeder link spectrum requirement • Reduce broadcast operating costs • Spectrum sharing and reuse	• Scalable & configurable network embedded in a broadcast system • Broadcast network cue & control that don't rely on other telecom infrastructure – surviving emergency and nature disaster • Backhaul data services among towers: IP based IoT, and wide area datacasting • Each tower can broadcast localized content • Full duplex transmission: transmission & reception on the same frequency – improving spectrum efficiency • Dynamic spectrum re-use & sharing + LDM: converging broadcast and wireless services • Inter-tower network can work under SFN, OCR, or multi-frequency network environments

REFERENCES

[1] L. Zhang, W. Li, Y. Wu, X. Wang, S.-I. Park, H.-M. Kim, J.-Y. Lee, P. Angueira, and J. Montalban, "Layered division multiplexing: theory and practice," *IEEE Trans. on Broadcasting*, vol. 62, no. 1, part II, pp. 216-232, Mar. 2016.

[2] L. Fay, L. Michael, D. Gomez-Barquero, N. Ammar and M.W. Caldwell, "An overview of the ATSC 3.0 Physical Layer Specification," *IEEE Trans. Broadcasting*, vol. 62, no. 1, part II, pp. 159-171, Mar. 2016.

[3] "Method and System for Wireless Data Communications", by Y. Wu et.al. (October 25, 2016). US Patent 9,479,826.

[4] J. Lee et al., "IP-Based Cooperative Services Using ATSC 3.0 Broadcast and Broadband," IEEE Transactions on Broadcasting, vol. 66, no. 2, pp. 440-448, June 2020.

[5] E. Iradier et al., "Using NOMA for Enabling Broadcast/Unicast Convergence in 5G Networks," IEEE Transactions on Broadcasting, vol. 66, no. 2, pp. 503-514, June 2020.

[6] J. Montalban et al., "Improved Semi-Blind Channel Estimation With Time Domain Cancellation for LDM-LSI," IEEE Transactions on Broadcasting, vol. 66, no. 3, pp. 613-619, Sept. 2020.

[7] S. Ahn et al., "Mobile Performance Evaluation of ATSC 3.0 Physical Layer Modulation and Code Combination under TU-6 Channel," IEEE Trans. Broadcasting, vol. 66, no. 4, pp.751-769, Dec. 2020.

[8] S. Ahn et al., "Large-Scale Network Analysis in NOMA-Aided Broadcast/Unicast Joint Transmission Scenarios Considering Content Popularity, IEEE Trans. Broadcasting, vol. 66, no. 4, pp.770-785, Dec. 2020.

[9] D. Gomaz et al., "Convergence of Broadcast and Broadband in the 5G Era," IEEE Trans. Broadcasting, vol. 66, no. 2, part II, pp.383-389, June. 2020.

[10] W. Li et al., "Coverage Study of ATSC 3.0 under Strong Co-Channel Interference Environment," IEEE Trans. Broadcasting, vol. 65, no. 1, pp.73-82, Mar. 2019.

[11] L. Zhang et al., "Layered-Division-Multiplexing for High Spectrum Efficiency and Service Flexibility in Next Generation ATSC 3.0 Broadcast System," IEEE Wireless Communications, vol. 26, no.2, pp.116-123, Apr. 2019.

[12] ATSC Standard, Doc. A/322, "Physical Layer Protocol", January 23, 2020.

[13] S-I Park et al., "Performance Analysis of All Modulation and Code Combinations in ATSC 3.0 Physical Layer Protocol," IEEE Trans. Broadcasting, vol. 65, no. 2, pp.197-210, June 2019.

[14] L. Zhang et al., "Using Layered-Division-Multiplexing to Deliver Multi-Layer Mobile Services in ATSC 3.0," IEEE Trans. Broadcasting, vol. 65, no. 1, pp.40-52, Mar. 2019.

[15] A. Mattsson, "Single Frequency Networks in DTV," *IEEE Trans. on Broadcasting*, vol. 51, no. 4, pp. 413–422, Dec. 2005.

[16] S. Ahn et al., "Mobile Performance Evaluation for ATSC 3.0 Physical Layer Modulation and Code Combinations Under TU-6 Channel," IEEE Transactions on Broadcasting, vol. 66, no. 4, pp. 752-769, Dec. 2020.

[17] ATSC Standard, Doc. A/324, "ATSC Standard: Scheduler Studio to Transmitter Link", January 5, 2018.

[18] W. Li, L. Zhang, Y. Wu, S. Park, N. Hur, J. Lee, "LDM in Wireless In-Band Distribution Link and In-Band Inter-Tower Communication Networks for Backhaul, IoT and Datacasting", in Proceedings of IEEE Int'l Symposium on Broadband Multimedia Systems and Broadcasting 2020 (BMSB2020), October 27-29. 2020, Paris, France.

[19] L. Zhang et al., "Using Layered Division Multiplexing for Wireless In-Band Distribution Links in Next Generation Broadcast Systems," IEEE Transactions on Broadcasting, vol. 67, no. 1, Mar. 2021.

[20] S-I. Park, H. Eum, S. Park, G. Kim, Y-T. Lee, H. Kim, and W. Oh, "Novel Equalization On-Channel Repeater with Feedback Interference Canceller in Terrestrial Digital Multimedia Broadcasting System", ETRI Journal, vol. 31, no.4, pp. 357-364, August 2009.

[21] K. Salehian, Y. Wu and B. Caron, "Design Procedures and Field Test of a Distributed-Translator Network, and a Case Study for an Application of Distributed-Transmission," IEEE Transactions on Broadcasting, vol 52, no. 3, pp 281-291, Sept. 2006.

[22] Grant, T. J., ed. Network Topology in Command and Control. Advances in Information Security, Privacy, and Ethics. IGI Global. pp. xvii, 228, 250. ISBN 9781466660595, 2014.

[23] Zhihong Hong et al, "In-Band Distribution Link Signal Detection", submitted to IEEE Int'l Symposium on Broadband Multimedia Systems and Broadcasting (BMSB'2021), Aug. 4-6, 2021, Chengdu, China.

[24] D. He et al., "Overview of Physical Layer Enhancement for 5G Broadcast in Release 16," IEEE Transactions on Broadcasting, vol. 66, no. 2, pp. 471-480, June 2020.

[25] J. Montalban et al., “Design of an IP-Based Broadcast Core-Network: Converging Broadcasters with the Connected World,” submitted to IEEE Transactions on Broadcasting.

Appendix A

Other Publications

Apart from the publications directly related with this thesis, the following list shows the publications carried out during this period.

A.1 International Journals

OJ1. Improved Semi-Blind Channel Estimation With Time Domain Cancellation for LDM-LSI

- Authors: Jon Montalban; Eneko Iradier; Lorenzo Fanari; Pablo Angueira; Sung-Ik Park; Sunhyoung Kwon; Namho Hur
- Journal: IEEE Transactions on Broadcasting
- Publisher: IEEE
- Year: 2020
- DOI: 10.1109/TBC.2020.2984995
- Abstract: The adoption of OFDM as the physical layer waveform in ATSC 3.0 leaves the door open for the implementation of Single Frequency Networks (SFN). In addition, Layered Division Multiplexing (LDM), which can be understood as a low complexity approach for Non-Orthogonal Multiple Access (NOMA) techniques, has also been included as a baseline technology. These two new features can be combined to offer LDM based Local Contents (LC) in SFNs with unprecedented spectrum efficiency. Up to now, the design of a reliable channel estimation algorithm that does not require any modification of the standard has been the main challenge for the successful development of this approach. This paper aims at providing tools for satisfactory and efficient application of LDM-LSI (Local Service Insertion). In particular, in this paper, three major contributions are presented: a new semi-blind channel estimation algorithm that takes

advantage of the LDM structure of the network, a time domain cancellation technique for NOMA, and the implementation of the proposed solution in a fully compliant ATSC 3.0 transceiver.

A.2 International Conferences

OC1. On the use of White Rabbit for Precise Time Transfer in 5G URLLC Networks for Factory Automation Applications

- Authors: Óscar Seijo; Iñaki Val; Jesus Alberto López-Fernández; Jon Montalban; Eneko Iradier
- Proceedings: IEEE International Conference on Industrial Cyber Physical Systems (ICPS)
- Publisher: IEEE
- Year: 2019
- DOI: 10.1109/ICPHYS.2019.8780178
- Abstract: The implementation of 5G networks has several technological challenges, which are mainly related to the UltraReliable and Low Latency Communications (URLLC). One of these challenges is precise time transfer. Precision Time Protocol (PTP) is the most widely used protocol for precise time transfer. However, PTP performance may not be sufficient in large 5G networks.In this paper, we propose the use of white rabbit as the main technology for frequency and time transfer in 5G transport networks. However, white rabbit can only be implemented in wired links, thus the time transfer performance in wireless links may not be enough for the most challenging URLLC use cases. Therefore, we propose in this paper a modification of the white rabbit protocol that can be used in wireless communications. The proposed wireless white rabbit protocol has been tested by simulations over a single carrier point to point communication system, and different channel models. The results show significant improvements of the time transfer accuracy in Line Of Sight channels compared to other common wireless time transfer approaches and a similar performance to white rabbit over wired links.

OC2. New Semi-Blind Channel Estimation for LDM-LSI

- Authors: Jon Montalban; Eneko Iradier; David Romero; Pablo Angueira; Sung Ik Park; Sunhyoung Kwon; Namho Hur

- Proceedings: IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)
- Publisher: IEEE
- Year: 2019
- DOI: DOI: 10.1109/BMSB47279.2019.8971906
- Abstract: Abstract: Local content insertion and viewer segmentation with proximity service information is a critical business case for broadcasting operators. This personalized service delivery will open up an opportunity for new use cases, such as local service delivery, local advertisement insertion and local emergency warning. In practice, single frequency networks (SFNs) offer a better spectrum efficiency when compared with traditional Multiple Frequency Network (MFN) approaches. Nevertheless, the addition of local content delivery can dramatically reduce that efficiency. Recently, the Layered Division Multiplexing (LDM) based Local Service Insertion (LSI) solution has been proposed as an interesting alternative for including local content in single frequency networks. Nonetheless, the required channel estimation procedure has been considered a major drawback for its successful implementation. In this work, a new semi-blind channel estimation algorithm is proposed. The algorithm takes advantage of the LDM structure and it is based on the statistical properties of the time domain Orthogonal Frequency Division Multiplexing (OFDM) frame. It proposes a solution for the major challenges that have remain unsolved so far. The paper also includes the first simulation results for this new approach.

OC3. Group-Oriented Broadcasting of Augmented Reality Services over 5G New Radio

- Authors: Nicola Benenati; Cristina Desogus; Pasquale Scopelliti; Eneko Iradier; Jon Montalban; Maurizio Murroni; Giuseppe Araniti; Pablo Angueira; Mauro Fadda
- Proceedings: IEEE Broadcast Symposium (BTS)
- Publisher: IEEE
- Year: 2019
- DOI: DOI: 10.1109/BTS45698.2019.8975412

– Abstract: Augmented Reality applications represent the current emerging trend of broadcast services whereby users require the same data content. Broadcast applications in smart environments ask for low-latency data transmission, low-energy communication, and location- and customer-based group creation procedures. The standardization efforts done from the 3rd Generation Partnership Project (3GPP) to make the 5th generation (5G) a reality have involved also the existing Long Term Evolution (LTE) radio access technology leading to the 5G New Radio (5G-NR) standard. One of the main innovation is the definition of three different Modulation and Coding Scheme (MCS) tables to allow a differentiation according to the class of devices, grouped into five different categories based on their basic characteristics. 5G broadcast/multicast is one of the topics that is under discussion at 3GPP for 5G phase II (release 17). In this paper, authors focused on a 5G-ready LTE system, referring to real broadcast mobile urban scenarios where users are interested in Augmented Reality (AR) applications. The authors analyzed the Subgrouping Optimal Aggregate Data Rate (SubOptADR) algorithm whereby broadcast destinations are grouped into different subgroups depending on the perceived user equipment (UE) channel quality in order to maximize the Aggregate Data Rate (ADR), which is the sum of data rate values obtained by all the broadcast members. The capabilities of the algorithm are evaluated comparing LTE and 5G-NR networks, focusing on different user classes and type of devices. The paper details three envisaged AR application scenarios, describing the subgrouping optimization algorithm in 5G-NR and showing how group-oriented communications can improve spectrum efficiency in the broadcast of AR services.

OC4. Trends and Challenges in Broadcast and Broadband Convergence

– Authors: Authors: Lorenzo Fanari; Eneko Iradier; Jon Montalban; Pablo Angueira; Sung-Ik Park; Namho Hur; Sunhyoung Kwon

– Proceedings: IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)

– Publisher: IEEE

– Year: 2019

– DOI: 10.1109/EExPolytech.2019.8906883

– Abstract: The convergence between broadcast and broadband networks has drawn significant attention since the very beginning of their coexistence. Nonetheless, a real cooperation framework has not been successfully implemented yet due to the limits of the contemporary technologies. However, the newly proposed architectures for both broadband and broadcast allow the development of new solutions for multimedia and entertainment content delivery in multi-tier radio access networks. In fact, the convergence of broadcast and broadband technologies is one of the solutions that has been already proposed. Convergence would offer a better spectrum efficiency, improving the quality of service of the broadcasted services to mobile receivers. The main contribution of this paper is a comprehensive state of the arte of the past efforts and on top of that, the definition of the most relevant use cases for several future scenarios.

OC5. PEG-LDPC Coding for Critical Communications in Factory Automation

– Authors: Lorenzo Fanari; Eneko Iradier; Jon Montalban; Pablo Angueira; Óscar Seijo; Iñaki Val

– Proceedings: IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)

– Publisher: IEEE

– Year: 2020

– DOI: 10.1109/ETFA46521.2020.9211899

– Abstract: Abstract: Critical communication requirements included in Factory Automation applications are complex to implement due to the difficulties encountered in guaranteeing high reliability and ultra-low latencies at the same time. In this work-in-progress, a technical solution for the physical layer is proposed: the Quasi-Cyclic LDPC of the Progressive Edge Growth family (QC-PEGLDPC). This coding scheme is considered as a promising candidate due to two main factors: the good decoding performance for short packet transmissions and the low latency that can be obtained by using full parallel decoding architectures. The obtained results are compared with the 5G New Radio coding scheme, which includes LDPCs as part of the solution for Ultra Reliable Low Latency (URLLC) use cases. In these first results, QC-PEG-LDPC shows a performance improvement of 1 dB when compared with the 5G LDPC codes for a message length of 128 bits. Latency analysis indicate that QC-PEG-LDPC could allow

decoding latencies of 0.13 μs providing that the full parallel decoding architecture is enabled.

OC6. LDPC Matrix Analysis for Short Packet Transmission in Factory Automation Scenarios

- Authors: Lorenzo Fanari; Eneko Iradier; Jon Montalban; Pablo Angueira; Óscar Seijo; Iñaki Val
- Proceedings: IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)
- Publisher: IEEE
- Year: 2020
- DOI: 10.1109/EExPolytech50912.2020.9243861
- Abstract: Critical application requirements proposed for Factory Automation scenarios involve demanding and simultaneous restrictions in reliability and latency. Channel coding is a basic tool to guarantee reliability but it usually involves high computation and memory resources at both transmitting and receiving sides. Among all the possible choices, Low Density Parity Check codes provide outstanding correction capacity. This paper studies two different LDPC coding approaches: 5G New Radio (5G NR) and IEEE 802.11 (WLAN). One of the drawbacks of LDPCs is their degraded performance and complex matrix adaptation for short information packets. Both 5G and 802.11 LDPCs use Quasi-Cyclic LDPC as the adaptation technique for matching the data packet size and coding matrix dimension. This paper evaluates 5G and 802.11 LDPC techniques, analyzing their structure and reliability performance for very short information messages by means of simulations. The results show that 5G NR LDPC is 1 dB closer to the Shannon limit than WLAN LDPCs, but this gain is lost for short message lengths as the ones used in Factory Automation environments. However, latency results are promising due to the opportunity of combining the analyzed techniques with new MAC techniques in order to achieve the desired reliability while not affecting latency dramatically.

OC7. Complexity Reduction Techniques for NOMA-based RRM Algorithms in 5G Networks

- Authors: Aritz Abuin; Eneko Iradier; Lorenzo Fanari; Jon Montalban; Pablo Angueira

- Proceedings: IEEE International Conference on Electrical Engineering and Photonics (EExPolytech)
- Publisher: IEEE
- Year: 2020
- DOI: 10.1109/EExPolytech50912.2020.9243860
- Abstract: 5G technology is expected to be a revolution in mobile communications due to its flexible physical layer (PHY). It is expected to cover a wide range of applications such as industry, virtual reality (VR) and autonomous vehicles. In particular, this work is focused in one of the main use cases of 5G: enhanced Mobile Broadband (eMBB), which targets large transmission capacities. This paper proposes algorithms that optimize the distribution of radio resources in broadcast/unicast scenarios in 5G networks using Non-Orthogonal Multiple Access (NOMA) techniques. First, a methodology to measure the complexity of those algorithms is presented and compared with Time Division Multiplexing (TDM) solutions. In addition, a specific algorithm that reaches optimum solution with less computational cost is proposed here. Results indicate that the overall complexity can be reduced more than 80% in comparison with current solutions.

OC8. Design and Performance Analysis of an ATSC 3.0 Model in NS-3

- Authors: Jon Montalban; Eneko Iradier; Lorenzo Fanari; Pablo Angueira; Sung-Ik. Park; Sunhyoung Kwon; Namho Hur
- Proceedings: IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)
- Publisher: IEEE
- Year: 2020
- DOI: 10.1109/BMSB49480.2020.9379699
- Abstract: ATSC 3.0 has been conceived and defined as the first IP-based digital terrestrial television standard. The viewers will have access to more advanced and immersive experiences, while the broadcasters will have a new way to interact with the audience. What is more, this unique feature also paves the path to the definition of a real convergent architecture with other broadband IP networks and the

specification of new cooperative and inter-working use cases. Nevertheless, this new approach demands a thorough evaluation of the ATSC 3.0 features, not only as a point-to-multipoint (PTM) tool for media delivery, but also as an interactive player within the Internet framework. So far, there is not a tool to accomplish adequately the evaluation of ATSC 3.0 performance from a data network perspective. This work aims at filling this gap, and consequently, presents an ATSC 3.0 model that can be integrated in NS-3, a widely employed and validated discrete-event network simulator for internet systems. This tool can test the capability of ATSC 3.0 to offer new services in this interconnected heterogeneous scenarios. In the first part of the work, the paper presents the different layers that are part of this software module and how they are connected to the different parts of the network stack. Eventually, in the second part, the model is evaluated for two different scenarios and the most important parameters are studied and discussed.

OC9. Multiple Layer P-NOMA in 5G

- Authors: Eneko Iradier; Iñigo Bilbao; Gorka Pujana; Lorenzo Fanari; Jon Montalban; Pablo Angueira
- Proceedings: IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)
- Publisher: IEEE
- Year: 2020
- DOI: 10.1109/BMSB49480.2020.9379385
- Abstract: The PHY layer presented by 3GPP in Rel-15 as 5G New Radio (NR) has considerably increased the spectrum efficiency of broadband systems, especially, due to the inclusion of Low-Density Parity-Check (LDPC) codes. Non-Orthogonal Multiple Access (NOMA) techniques can take efficiency a step further and they are currently considered a promising complementary technology for future releases. This paper addresses this technology and presents a performance evaluation of the use of multiple NOMA layers in 5G networks on a network simulator, which provides a set of new evaluation metrics. The design of the required architecture for introducing multiple layer communications in 5G is presented as the main contribution of the paper. In addition, two 5G use cases have been designed. In one of the cases, an innovative priority-based access communication architecture is proposed. Finally, the use cases are tested

in a discrete-event network level simulator and their performance is evaluated in terms of reliability, capacity and latency.

OC10. PHY Layer Performance of N-LDM for Broadcast and IoT Use Cases

- Authors: Gorka Pujana; Eneko Iradier; Jon Montalban; Lorenzo Fanari; Pablo Angueira
- Proceedings: IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)
- Publisher: IEEE
- Year: 2020
- DOI: 10.1109/BMSB49480.2020.9379456
- Abstract: This paper presents a comprehensive performance analysis of a multiple layer LDM (N-LDM) system. Although previous works have proposed the use of N-LDM for delivering different services within a RF channel, none of them has presented results for more than three layers. What is more, a detailed analysis where the system efficiency is taken to its limit is still missing. Hence, in this work a new simulation tool has been developed, where any number of layers can be transmitted and received by using multiple layer frequency domain cancellation. On the one hand, the results address the number of services that can be simultaneously transmitted for specific user characteristics and propagation channels. On the other hand, the minimum interlayer injection level (IL) that has to be applied for correct reception has been defined and calculated for different configurations.

A.3 Book Chapters

B1. Digital Terrestrial Broadcast Standards

- Authors: Jon Montalban; Eneko Iradier; Lorenzo Fanari; Peter Siebert
- Book: Wiley Encyclopedia of Electrical and Electronics Engineering
- Publisher: Wiley
- Year: 2021

– Abstract: The European Digital Video Broadcasting (DVB) project started in 1993 as an industry forum providing the specifications for digital broadcast. Then, the first Digital Terrestrial Television (DTT) system, DVB-Terrestrial (DVB-T), was agreed in 1997, and one year later, the first DVB-T broadcast began in Sweden and the UK. However, due to the second generation standard (DVB-T2), the data carrying capacity of a single Radio Frequency (RF) broadcast channel was really maximized. Specifically, a minimum capacity increase of 30% was achieved as well as improved SFN performances, a mechanism to reduce the peak-to-average-power ratio (PAPR) and a flexible system layer for transport streams. This article presents a system overview of the DVB-T2 standard. In particular, the general architecture of the DVB-T2 transceiver is described module by module. Moreover, several implementations and evaluation results of the standard are presented.

Appendix B

Resume

Eneko Iradier received his B.Sc. degree and M.Sc. in Telecommunications Engineering from the University of the Basque Country (UPV/EHU) in 2016 and 2018, respectively. Since 2015, he is part of the TSR Research Group, in UPV/EHU, where he is currently a PhD student. He was with the Communications Systems Group of IK4-Ikerlan as researcher from 2017 to 2018. During his doctoral studies, he did an internship at Communications Research Centre Canada in Ottawa. He has served as a reviewer for several renowned international journals and conferences in the area of wireless communications. His current research interests include the design and development of new technologies for the physical layer of communication systems, broadcasting solutions for 5G and wireless solutions for Industry 4.0.

Appendix C

Acronyms

1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
6G	Sixth Generation
ADC	Analog-to-Digital Converter
AF	Amplify-and-Forward
AI	Artificial Intelligence
AMC	Adaptive Modulation Coding
AN	Artificial Noise
AP	Access Point
AR	Augmented Reality
ATSC	Advanced Television Systems Committee
BCH	Bose-Chaudhuri-Hocquenghem
BDM	Bit-Division Multiplexing
BE	Best Effort
BS	Base Station
CDMA	Code Division Multiple Access
CH	Cluster Head
Cloud Txn	Cloud Transmission
C-NOMA	Code-domain Non-Orthogonal Multiple Access
CoMP	Coordinated Multi-Point
CQI	Channel Quality Indicator
CS	Critical Service
CSI	Channel State Information
D2D	Device-to-Device
DAB	Direct Access-Based
DF	Decode-and-Forward

DSA	Dynamic Spectrum Access
DTT	Digital Terrestrial Television
DTV	Digital TV
DVB	Digital Video Broadcasting
eMBB	enhanced Mobile Broadband
FA	Factory Automation
FFR	Fractional Frequency Reuse
FLC	Fuzzy Logic Controller
gNB	gNodeB
GRPA	Gain Ratio Power Allocation
HARQ	Hybrid Automatic Repeat reQuest
HbbTV	Hybrid Broadcast Broadband TV
IDL	In-band Distribution Link
IDMA	Interleave Division Multiple Access
IEEE	Institute of Electrical and Electronics Engineers
IL	Injection Level
IoE	Interconnection of Everything
IRS	Intelligent Reflecting Surfaces
ITCN	Inter-Tower Communications Network
ITU	International Telecommunication Union
ITU-R	International Telecommunication Union - Radiocommunication
KPI	Key Performance Indicator
LDM	Layered Division Multiplexing
LDS-CDMA	Low-Density Spreading CDMA
LDS-OFDM	Low-Density Spreading-based OFDM
LL	Lower Layer
LOS	Line-of-Sight
LTE	Long Term Evolution
MAC	Medium Access Control
MCS	Modulation and Code Scheme
MCTS	Monte Carlo Tree Search
MIMO	Multiple Input Multiple Output
ML	Machine Learning
mmWaves	millimeter frequency bands
mURLLC	massive ultra-reliable low latency communications
MUST	Multiuser Superposition Transmission
NAICS	Network Assisted Interference Cancellation and Suppression
NLOS	Non-Line-of-Sight
NOMA	Non-Orthogonal Multiple Access
NR	New Radio

NUC	Non Uniform Constellation
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Orthogonal Multiple Access
OSI	Open Systems Interconnection
PA	Proccess Automation
PDMA	Pattern Division Multiple Access
PER	Packet Error Rate
PHY	Physical Layer
PLR	Packet Loss Rate
P-NOMA	Power-domain Non-Orthogonal Multiple Access
PTM	Point-to-Multipoint
QC	Quantum Computing
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RAN	Radio Access Network
RAT	Radio Access Technology
RB	Resource Block
RE	Resource Element
RRM	Radio Resource Management
SC	Superposition Coding
SCMA	Sparse Code Multiple Acces
SDN	Software Defined Networking
SFN	Single Frequency Network
SIC	Successive Interference Cancellation
SISO	Single Input Single Output
SNR	Signal-to-Noise Ratio
TAS	Transmit Antenna Selection
TDL	Tapped Delay Line
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
TDMA	Time Division Multiple Access
TR	Technical Report
TTI	Transmission Time Interval
UAV	Unmanned Aerial Vehicle
UHF	Ultra High Frequency
UL	Upper Layer
URLLC	Ultra Reliable Low Latency Communications
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-Everything
VLC	Visible Light Communications
VNF	Virtualized Network Function
VR	Virtual Reality

WI	Work Item
WiB	Wideband reuse-1
WSN	Wireless Sensor Network