5G Mixed Mode:
NR Multicast-Broadcast Services

E. Garro, M. Fuentes, J.L. Carcel, H. Chen, D. Mi, F. Tesema, J.J. Gimenez and D. Gomez-Barquero

Abstract—**This work presents a potential solution for enabling the use of multicast in the 5G New Radio Release 17, called 5G NR Mixed Mode. The proposed multicast/broadcast mode follows one of the two approaches envisaged in 3GPP, which enables a dynamic and seamless switching between unicast and multicast, both in the downlink and the uplink. This paper also provides a performance evaluation of several IMT-2020 KPIs, including available data rate and spectral efficiency, user and control plane latencies, energy efficiency, mobility highlighting the potential advantages of this solution over unicast in relevant scenarios. Finally, other multipoint–based KPIs such as coverage or packet loss rate are also evaluated by means of system level simulations.**

Index Terms—**5G, NR, Mixed Mode, multicast, KPIs, IMT-2020**

I. INTRODUCTION

5G New Radio (NR) is the fifth generation of mobile communications systems defined by 3GPP (3rd Generation Partnership Project) in Release 15 (Rel-15) and successive releases. Designed towards the framework of International Mobile Telecommunications at 2020, i.e. IMT-2020, NR covers an extensive number of use cases for the digitization of new verticals in three main usage scenarios: enhance Mobile BroadBand (eMBB), Ultra-Reliable and Low Latency Communications (URLLC) and massive Machine-Type Communications (mMTC) [1][2]. The normative work of NR in 3GPP toward the IMT-2020 submission started in March 2016 with Rel-15. To introduce the aforementioned functionalities, NR Rel-15 last drop standardized a complete NR solution with full user and control plane capabilities in the Radio Access Network (RAN) and the System Architecture (SA). The new 5G system continued to be developed during Rel-16 aiming at meeting the requirements defined in [3] [4].

The NR Rel-15 and Rel-16 air interface brings a large number of improvements compared to Long Term Evolution (LTE). Some of the most important improvements in the air interface are more efficient Forward Error Correction (FEC) coding schemes [5][6], larger bandwidths, new OFDM waveform numerologies adapted to the 5G spectrum band allocation or massive Multiple-Input Multiple-Output (MIMO).

NR has also included new functionalities into the upper architectural layers. Among others, the layer 2 protocols include Service Data Adaptation Protocol (SDAP), Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC), and Medium Access Control (MAC). A major change is that the concatenation of packets no longer takes place in RLC layer, but is moved to the MAC layer. A completely new element is the SDAP layer, which is used for packet marking with Quality of Service (QoS) flow ID (QFI) and mapping of QFI to radio bearers.

Up to Rel-16, NR has been only focused on unicast Point-To-Point (PTP) transmissions, which may not support those scenarios where a very large number of users consume the same data, such as popular media content, emergency messages or software updates [7][8]. In those scenarios, the use of multicast/broadcast Point-To-Multipoint (PTM) schemes can offer huge capacity gains, ensuring a cost-effective high-quality delivery mechanism [9][10]. The lack of PTM capabilities may imply a future limitation of 5G networks, leading to inefficient service provisioning and utilization of the network and spectrum resources. In order to address these limitations, 3GPP identified a flexible multicast service as a basic feature to be used in 5G [2][4]. A Study Item (SI) on NR mixed mode broadcast/multicast [11] was drafted in June 2018, which aimed at studying different aspects to enable broadcast/multicast over NR. The discussions in last 3GPP RAN plenary meetings have been based on the assignment of tasks according to two priority levels. The specification of a group scheduling mechanism to allow UEs to receive broadcast and/or multicast services or to specify the required changes to improve the reliability of broadcast and/or multicast services by uplink feedback mechanisms are some of the minimum set of objectives that are being currently considered [12]. Following the aforementioned devised modifications, this paper introduces a PTM design that enables a dynamic allocation of unicast and multicast resources over NR following also the principle designs of [12]. This extension, named 5G NR Mixed Mode, reuses as much as possible the existing NR radio components with some

E. Garro, M. Fuentes and D. Gomez-Barquero are with the Universitat Politècnica de Valencia, Valencia, 46022, Spain (e-mail: {edgarcre, mafuemue,dagobar}@iteam.upv.es).

J.L. Carcel is with the Fundación ValenciaPort, Valencia, 46024, Spain (e-mail: jlcarcel@fundacion.valenciaport.com).

H. Chen and D. Mi are with the University of Surrey, Guildford, GU2 7XH, United Kingdom (e-mail: {d.mi,hongzhi.chen}@surrey.ac.uk)

F. Tesema is with Nomor Research GmbH, Munich, 81541, Germany (e-mail: tesema@nomor.de).

J.J. Gimenez is with the Institut für RundfunktTechnik GmbH, Floriansmuhlstrasse 60, 80939 Munich, Germany (e-mail: jordi.gimenez@irt.de).
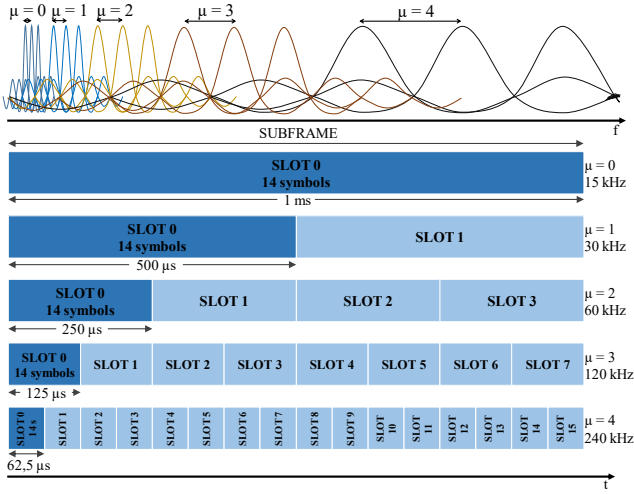
Fig. 1. NR numerologies (μ) (top), and associated slots duration time (bottom).

modifications to enable the efficient transmission over several users requesting the same data content. Two operational deployments have been envisaged to ensure the maximum compatibility with the current NR Rel-15. The required modifications in the air interface include the introduction of a Group Radio Network Identifier, multiple cell coordination mechanisms for supporting Single Frequency Networks, and narrower subcarrier spacings in order to allow larger inter-site distances. The proposed solution for the RAN upper layers involves the introduction of two new logical control and traffic channels, and new RAN procedures that will enable the seamless switching between unicast and multicast control and data radio bearers. While this paper is focused on the necessary changes to enable multicast and broadcast services through NR, [13] and [14] design the Next-Generation RAN to enable SFN and 5G Core Network modifications needed to support a holistic 5G multicast – broadcast system, respectively.

The key enablers of 5G NR Mixed Mode has been evaluated across several Key Performance Indicators following the requirement criteria stated by the International Telecommunications Union within the IMT-2020 framework. It has been demonstrated that the additional modifications contemplated for the 5G NR Mixed Mode designs still fulfill the data rate, spectral efficiency, mobility, latency and energy efficiency IMT-2020 requirements. Moreover, it has been estimated that 5G NR Mixed Mode could provide up to 0.5 Gbps data rate gains compared to current NR Rel-15 thanks of alleviating the physical signaling overhead. The support of larger Single Frequency Network deployments has been validated by coverage studies, where it has been highlighted that the numerology $\mu = -2$ could provide up to 5 dB SINR gain at cell edge users. Finally, system level simulations have been carried out to analyze the 2nd layer FEC and eOLLA Precoding Matrix Indicator and Rank Indicator feedback schemes benefits. It has been shown that while the 2nd layer FEC exhibits further improved packet loss rate performance as compared to conventional AL-FEC, the fixed and cyclic PMI and RI performances depend on the targeted PDU loss rate.

The rest of the paper is structured as follows: Section II briefly describes the main new features of the NR air interface

and radio resource management [7]. Section III highlights the NR modifications needed for the two proposed mixed mode solutions, the so-called Single Cell Mixed Mode (SC-MM) and Multiple Cell Mixed Mode (MC-MM) solutions. They are next evaluated and compared with current NR system analytically and by means of link level, coverage and system-level simulations in Sections IV and V. Finally, the conclusions are summarized in Section VI.

II. 5G NR

This section provides a brief description about the first 3GPP specification of NR, i.e. Rel-15, which presents a more flexible design than LTE to fulfill a wider set of heterogeneous requirements.

A. Physical layer

This section provides a brief description about the physical layer of the NR. Waveform numerologies, physical downlink and uplink channels and signals are described in the following subsections

1) Waveforms and numerologies

As LTE, NR has adopted Cyclic Prefix - Orthogonal Frequency Division Multiplexing (CP-OFDM) waveform for both downlink and uplink. However, the biggest difference between both specifications is the use of new numerologies (subcarrier spacing and symbol length) in NR. Whereas LTE has a fixed subcarrier spacing of $\Delta f = 15$ kHz, NR introduces and scalable approach with multiple types according to the parameter μ as $\Delta f = 2^\mu \times 15 = \{15, 30, 60, 120, 240\}$ kHz. In time domain, the higher the numerology factor (μ), the shorter the OFDM symbol length, and consequently the shorter the slot duration. NR reduces the delivery of low latency applications, thanks to the shorter slot durations, as well as to the introduction of mini-slots, which are also useful in massive MIMO beamforming procedures. Fig. 1 depicts the different numerologies included in NR specification. It should also be remarked that the new NR slot structure allows for a dynamic assignment of the link direction in each OFDM symbol within the slot, minimizing the potential uplink and downlink traffic congestions.

2) Physical channels and signals

Channels are known as flows of information transmitted between the different protocol layers. Thanks to them, different types of data are separated and transported across different layers. In particular, physical channels carry MAC layer information, and they are differentiated between downlink and uplink transmissions.

i. Physical downlink channels and signals

Three physical downlink channels and five physical downlink signals are defined. They differ on their functionality:

- **Physical Broadcast Channel (PBCH):** transmits the static part of the System Information, known as Master Information Block (MIB), to any User Equipment (UE) requiring attaching to the network.
- **Physical Downlink Control Channel (PDCCH):** Specifies the scheduling and allocation of the data content for every UE by means of the Downlink Control
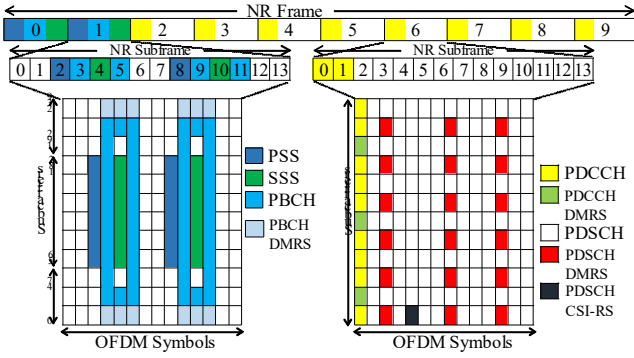
Fig. 2. Framing structure for a NR frame with all slots configured as downlink.

Fig. 3. User plane data flow across NR radio protocols [13].

Information (DCI). As PBCH, PDCCH uses CRC + Polar coding schemes and QPSK constellations. The encoded information is mapped within the PDCCH in Control Resource Sets (CORESETs), whose size depends on the Aggregation Level (AL).

- **Physical Downlink Shared Channel (PDSCH):** Transmits the data content to the UE and the System Information Blocks (SIBs). PDSCH uses CRC + LDPC coding schemes and from QPSK up to 256QAM modulation orders.
- **Primary and Secondary Synchronization Signals (PSS, SSS):** Combined with the PBCH they allow the UE to access the NR network. Specifically, they provide frame timing information and cell ID at the initial search, as well as beam management in IDLE state.
- **Demodulation Reference Signals (DMRS):** They are used for channel estimation in order to retrieve the information in PBCH, PDCCH and PDSCH.
- **Phase Tracking Reference Signals (PTRS):** They are only used at Frequency Range 2 (FR2) {24.25 GHz - 52.6 GHz} for phase noise estimation in the PDSCH.
- **Channel State Information Reference Signals (CSI-RS):** Used to provide CSI, needed for link adaptation and for beam management in CONNECTED state.

Fig. 2 shows where physical downlink channels and signals are allocated, when all the slots of a frame are configured as downlink (Slot Format Indicator = 0).

ii. Physical uplink channels and signals

Three physical uplink channels and three physical uplink signals are defined in NR. Their names and functionalities are listed below.

- **Physical Random Access Channel (PRACH):** Used by the UE to request the uplink initial access, as well as for beam management processing.
- **Physical Uplink Control Channel (PUCCH):** It carries the Uplink Control Information (UCI) that contains different information such as CSI, HARQ retransmission and scheduling requests.
- **Physical Uplink Shared Channel (PUSCH):** Transmits the data content to the Next Generation Node B (gNB), and it can optionally convey UCI transmissions.
- **Demodulation Reference Signals (DMRS):** Used for channel estimation in order to allow the proper demodulation of PUCCH and PUSCH.
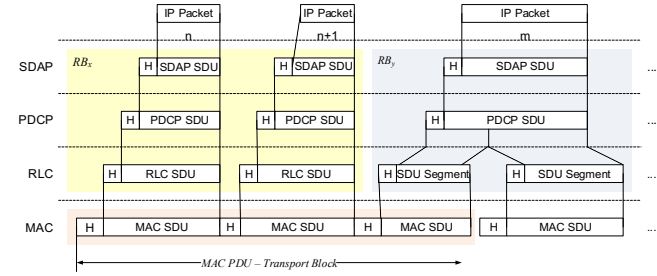- **Phase Tracking Reference Signals (PTRS):** Used for the

same functionalities than in downlink.
- **Sounding Reference Signals (SRS):** Equivalent to CSI-RS for uplink, providing CSI to the gNB for link adaptation and scheduling configurations.

B. Upper layers of NR

This subsection highlights the main functionalities and major changes of NR radio protocols [15], with special focus on layer 2 that will require the main modifications for accommodating the proposed 5G NR Mixed Mode design into the current RAN upper layers.

1) NR protocol stack

Fig. 3 describes the architecture of the radio protocol function pertinent to the communication between an NR gNodeB (gNB) and a UE. A short description of SDAP, PDCP, RLC and MAC in accordance with 3GPP's general description of NR is presented below:

- **SDAP:** This new layer introduced in NR takes care of QoS flow handling, e.g., mapping between a QoS flow and a data radio bearer, and marking QFI in both downlink and uplink packets.
- **PDCP:** Header compression and decompression, Robust Header Compression (ROHC) only; Reordering and duplicate detection; PDCP Protocol Data Unit (PDU) routing (in case of split bearers); Retransmission of PDCP SDUs; Ciphering, deciphering and integrity protection; PDCP Service Data Unit (SDU) discard; PDCP re-establishment and data recovery for RLC Acknowledged Mode (AM); Duplication of PDCP PDU.
- **RLC:** This layer supports three transmission modes, which are Transparent Mode (TM), Unacknowledged Mode (UM) and Acknowledged Mode (AM). The supported services and functions are segmentation (AM and UM) and re-segmentation (AM only) of RLC SDUs; reassembly of SDU (AM and UM); RLC SDU discarding (AM and UM); error correction through ARQ (AM only); duplicate Detection (AM only); and protocol error detection (AM only).
- **MAC:** This layer provides mapping between logical channels and transport channels. It multiplexes/ demultiplexes MAC SDUs belonging to one or different logical channels into/from transport blocks (TB) delivered to/from the physical layer on transport channels. The MAC layer is also responsible of scheduling information reporting, error correction through HARQ, priority handling between logical channels of one UE, priority
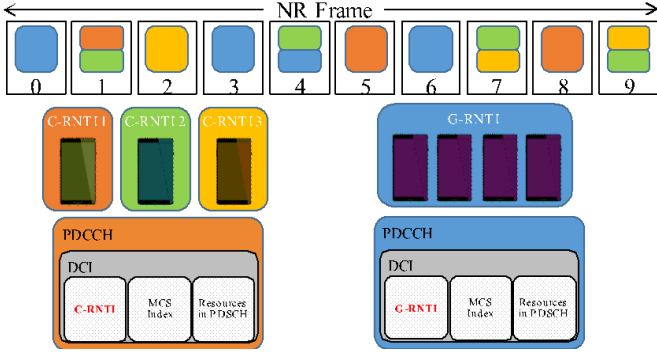
Fig. 4. C-RNTI for NR and G-RNTI for 5G MM.

Fig. 5. Cell scrambling sequence initialization N_{ID}^{nSCID} according to the physical layer cell ID (a), and given by the high-layer parameter *DL-DMRS-Scrambling-ID* (b).

TABLE I
ISD FOR EXTENDED CP WITH NEGATIVE μ

μ	Δf (kHz)	T_U (μs)	T_{CP} (μs)	ISD (km)	Overhead (%)
0	15	66.6	16.6	5	
-1	7.5	133.3	33.3	10	20
-2	3.75	266.6	66.6	20	

handling between UEs, and packet re-ordering with retransmissions with HARQ.

2) NR procedures, format headers and data flows

The data flow across NR radio protocols can also be seen in Fig. 3, which takes downlink user plane as an example. First, higher layer Internet Protocol (IP) packets are marked with QFI and mapped to radio bearers. Then, the PDCP layer performs header compression and security (e.g., ciphering and integrity protection) and forwards PDCP PDUs to the RLC layer. After that, the RLC layer wraps RLC SDUs or segments thereof into RLC PDUs based on the available MAC layer transport block size. The 5G NR can operate in UM or AM mode where retransmissions of lost packets can be performed via ARQ procedures. Following the RLC functions, the MAC layer multiplexes RLC PDUs, which may come from the same or different sources, e.g. different radio bearers, into the available MAC transport block.
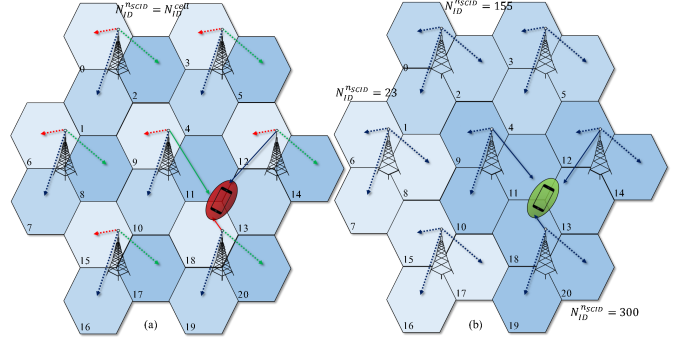
III. 5G NR MIXED MODE

Although the Rel-16 air-interface will improve among others V2X communications, the access to unlicensed spectrum, power saving and positioning support with respect to previous releases, PTM is not being included as an enhanced feature. This limitation can be solved with the proposed design for Rel-17. One of the key principles adopted for the proposed design is to limit investment costs and implementation complexity over the existing PTP infrastructure by minimizing the added footprint for delivering PTM services [16][17].

A. Modifications in physical layer

The proposed 5G NR Mixed Mode introduces light modifications in the physical layer, also known as air interface, in order to be as similar as current NR specifications, but enabling the use of multipoint transmissions. The proposed modifications are widely described below.

1) Modifications in PDCCH

As explained previously, the PDCCH plays a key role in the reception of scheduling information and de-scrambling of data by means of the Cell-Radio Network Temporary Identifier (C-RNTI). Proceeding in the same manner as the Single-Cell Point-to-Multipoint solution of LTE [18], it is possible to define a common identifier, Group-RNTI (G-RNTI), so that several UEs camping on a single cell can be easily grouped and access the same radio resources conveying the desired content. The

introduction of the G-RNTI enables a dynamic, flexible and scalable scheduling and multiplexing between unicast and multicast data within the PDSCH channel. The basic mechanism behind this process is illustrated in Fig. 4. As it can be observed, thanks to the introduction of G-RNTI, a single DCI could be transmitted for a group of UEs interested in the same content. This solution avoids the transmission of several CORESETs announcing the same data to all users and reduces considerably the PDCCH overhead within a NR frame.

One of the main drawbacks of SC-MM is the inefficiency to cover large areas due to inter-cell interferences. The coverage can be extended by means of multiple cell coordination mechanisms, like Single Frequency Network (SFN) deployments. The following air interface modifications are proposed to enable SFNs, resulting in the defined MC-MM design.

2) Common Cell Scrambling Sequence

In SFN deployments, the same content should be transmitted from different sites. This requires not only to transmit the same data content but also in the same resource elements, as well as the same control content, including DMRS values. All this information is determined with the cell specific scrambling sequence, which its initialization depends on N_{ID}^{nSCID} parameter. Instead of performing the widespread physical layer cell ID procedure that would lead to a different scrambling sequence per cell, MC-MM forces it to a certain value, $N_{ID}^{nSCID} \in \{0, 1, ..., 65535\}$ for all the coordinated cells, which will be given by the higher-layer parameter *DL-DMRS-Scrambling-ID*. Fig. 5 illustrates the N_{ID}^{nSCID} traditionally used (left) and with the proposed parameter for MC-MM (right).

3) Negative numerologies and extended CP

The maximum Inter-Site Distance (ISD) between SFN transmitters is limited by the CP length, which in turn depends on the NR numerology. The maximum ISD with current Rel-15 is 1.4 km. This ISD may be suitable for some limited scenarios such as stadium, campus or malls, but urban or rural
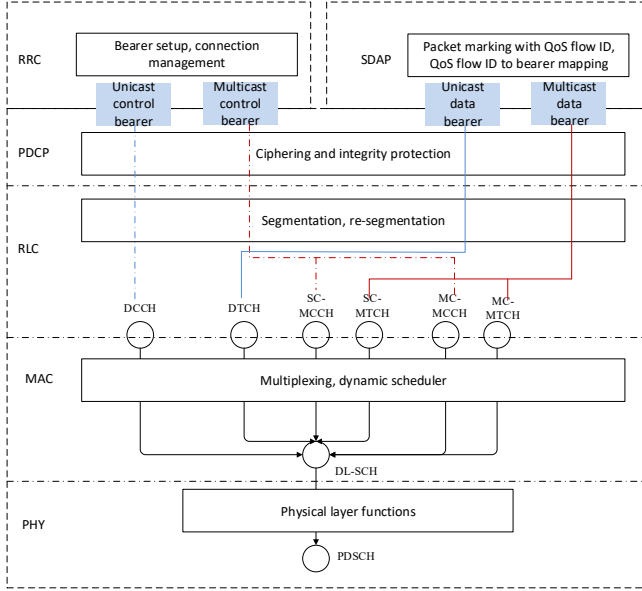
Fig. 6. Radio protocol enhancement to support seamless switching between unicast RB and multicast RB.

environments are characterized by longer distances. To support these scenarios, a set of different enhancements may be introduced. Longer CP lengths can be obtained by employing narrower Δ_f and by making use of the extended CP type adopted in NR[1]. In particular, two negative numerologies in conjunction with Extended CP are proposed for MC-MM (see Table I).

Note that these solutions will lead to greater vulnerability in high-speed conditions and will require larger and more complexity demanding FFT sizes. In addition, since the associated slots of negative numerologies will span multiple OFDM symbols of the regular numerology, over more than one subframe, the use of mini-slots (e.g. slots of 2, 4, or 7 OFDM symbols) is proposed for compatibility reasons with NR framing structure [19].

B. Modifications in upper layers

The proposed solution for the Mixed Mode in RAN upper layers aims at providing radio access technology (RAT) protocol and radio resource management (RRM) functionalities that being able to support flexibility and efficiency of new radio required by existing and future PTP/PTM services [20]. Such solution uses the NR Rel-15/16 as baseline for enhancement, and its major contributions are presented as follows.

1) Seamless transition between NR and 5G NR Mixed Mode schemes

To optimize utilization of radio resources in scenarios that have a considerable diversity of traffic volume, the proposed RAN-level system should be able to switch between unicast and multicast/broadcast transmission modes in a flexible manner.

i.   New logical multicast channels

Fig. 6 shows the radio protocol modifications to support seamless switching between unicast RB and multicast RB. The unicast control and data bearers use the logical channels Dedicated Control Channel (DCCH) and Dedicated Traffic Channel (DTCH), respectively. On the other hand, the multicast
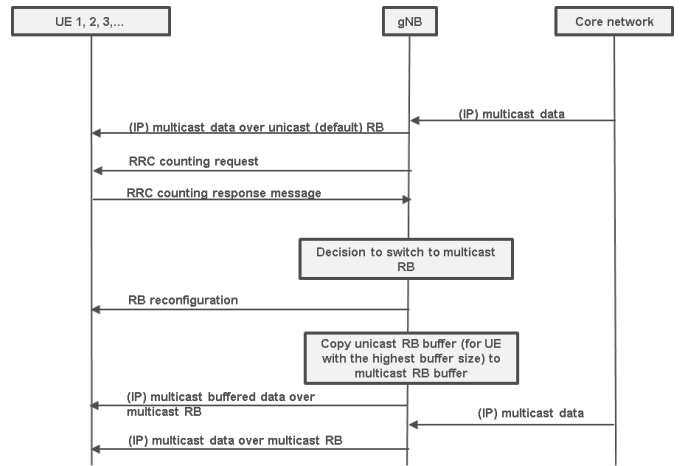

Fig. 7. High-level RAN procedures to switch from unicast RB to multicast RB.

control bearer uses Single Cell Multicast Control Channels (SC-MCCH) and Multi-Cell Multicast Control Channels (MC-MCCH) for SC-MM and MC-MM respectively, while multicast data bearer uses Single Cell Multicast Traffic Channel (SC-MTCH) and Multi-Cell Multicast Traffic Channel (MC-MTCH)[2]. Seamless transition of unicast and multicast radio bearers is facilitated by mapping the relevant logical channels, DCCH, DTCH, SC-MCCH, SC-MTCH, MM-MC-MCCH, MM-MC-MTCH, onto shared transport channel, i.e., Downlink Shared Channel (DL-SCH).

ii.   New RAN procedures for seamless switching

In 3GPP, the flexible delivery of content as unicast or broadcast service was included in eMBMS as MBMS Operation on Demand (MooD) feature [21]. In MooD, the decision for transition between unicast and broadcast/multicast was carried out by the Broadcast Multicast Service Center (BM-SC) core network entity. BM-SC utilized user-service consumption report from UEs to make the switching decision. Following this approach, but in order to have optimized utilization of radio resources the 5G NR Mixed Mode solution proposes to move it to the RAN-level for a seamless transition between unicast and broadcast/multicast modes.

For delivery of Internet Protocol (IP) multicast/broadcast content, RAN-level transmission modes should include unicast or multicast/broadcast radio transmission in association with various Radio Resource Control (RRC) states as described in [22]. In that context, while unicast utilizes UE-specific data and control dedicated RBs, the multicast/broadcast transmission modes cannot be dedicated to a specific UE. Accordingly, the delivery of IP multicast/broadcast data to UEs for the 5G NR Mixed Mode can be realized by mapping multicast RBs to unicast RBs for the RRC_CONNECTED state, and keeping the multicast RB for the delivery of IP multicast/broadcast data to UEs in RRC_INACTIVE or RRC_IDLE states. Thanks to the introduction of the UE's RRC_INACTIVE state in NR, that maintains UE's connection of the RAN to the core network, the transition to RRC_CONNECTED state can be performed with extremely low latency.

---

[1] While Normal CP length is obtained as $T_{CP} \simeq \Delta f /16$, Extended CP length represents $T_{CP} \simeq \Delta f /4$.

[2] The logical channels MC-MCCH and MC-MTCH support the new physical layer numerologies specified in Section III.A.3.

The criteria for switching between unicast RB and multicast RB can be the number of UEs demanding multicast service and/or UEs' QoS requirement. For example, the criteria on the number of UEs can be implemented by re-using LTE counting procedure as standardized in [23]. However, unlike LTE, which uses the counting function to disable (suspend) or enable (resume) multicast RB transmission, the counting function for the Mixed Mode can be used to make the decision of RB switching between unicast and multicast RBs.

Fig. 7 describes a high-level RAN procedure to switch from unicast RB to multicast RB to deliver IP multicast data. First, information on the number of multicast UEs is collected by using counting functions. Then, decision for switching bearer is made based on a threshold configured by the network operator or network planner. Following bearer-switching decision, the new RB re-configuration is sent to the UEs that consume multicast service. Moreover, the buffered data in the unicast RB is copied to the multicast RB. Accordingly, the buffered data and newly arriving IP multicast data are transmitted over the multicast RB. Note that with unicast RB transmission the RLC SDU buffer can be different for various UEs since UEs have independent dedicated radio link. Hence, copying the buffer for the UE with the highest buffer size is more crucial in avoiding loss of packets. In addition to the above procedures that perform RAN-level seamless switching between unicast or multicast RB across all UEs being served by the gNB, UE-specific switching between unicast and multicast RB can be done depending the channel condition of the user. For example, if the UE is experiencing severe degradation of received signal quality while being served by a multicast RB, it can request the gNB to switch transmission from multicast RB to unicast RB. To assist the gNB for UE-specific switching decisions, the RRC_CONNECTED UE has to provide signal measurement to the gNB. Those UEs in RRC_IDLE or RRC_INACTIVE states, should first transition to RRC_CONNECTED state.

For the case where multicast RB is used, the decision on using SC-MM or MC-MM can rely on the geographical distribution of multicast UEs, which may be taken from GPS location services. However, location services may have privacy constraints if users do not consent to provide their location. In such case, UE measurement should be used to measure interference levels in order to assist switching decisions. If a considerable number of UEs is receiving multicast data via SC-MM at the border of two or more cells, coordinating the cell for multi-cell transmission avoids the interference between the cells, which in turn improves the spectral efficiency.

2) Prospect of feedback for Mixed Mode

The proposed scheme allows for service and media specific extensions, which could consist of methods for user grouping, multiplexing the feedback or channel user selection for feedback.

i. Feedback schemes with QoS

The system efficiency depends on the amount of data transmitted in a period. NR follows same enabler as LTE,

namely HARQ, which can significantly increase the system efficiency when an appropriate Modulation and Coding Scheme (MCS) is selected. On the other hand, in eMBMS, the UE is not aware of QoS associated with MBMS bearer that is determined at the Multicast Coordination Entity (MCE). Nevertheless, to improve the feedback in terms of received QoS, the availability of the QoS parameters at the UE receiver is proposed for the Mixed Mode.

In a conventional unicast HARQ scheme, the receiver sends back an Acknowledgment (ACK) or Negative Acknowledgement (NACK), depending on the successful packet reception. When the conventional HARQ method is applied to multicast, multiple ACK should be assigned for the serving multicast users. This may result in unacceptable uplink feedback overhead for a significant number of multicast users. In addition, if the transmitter selects the same MCS as in the case of unicast transmission but the receivers are in similar radio condition, then the probability of the transmitter receiving NACK increases in the proportion of the number of receivers[3]. Using more robust MCS reduces the probability of NACKs but also decreases the spectral efficiency. To cope with these limitations, the packet loss or delay rate QoS parameters as well as the observed QoS at the receiver are proposed to be used in Mixed Mode to decide whether ACK, NACK or no feedback shall be sent for an erroneously received transport block. Among the different solutions to optimize HARQ for multicast, [24] studies an optimal number of autonomous retransmissions based on limited feedback from the UEs, so that the automatic retransmission of a packet is done in a predetermined number without ACK/NACK feedback. Furthermore, assuming that the users receiving IP multicast traffic are mainly interested in the QoS, the receiver NACK can be sent only if the received data is not going to meet the target QoS. Another proposed alternative is to send no feedback if the HARQ process at the transmitter operates with implicit ACKs where ACK is assumed if NACK is not received. In 5G, the SDAP layer maps the QoS flows to radio bearers and it is possible to continuously monitor QoS and adjust the MAC configuration accordingly.
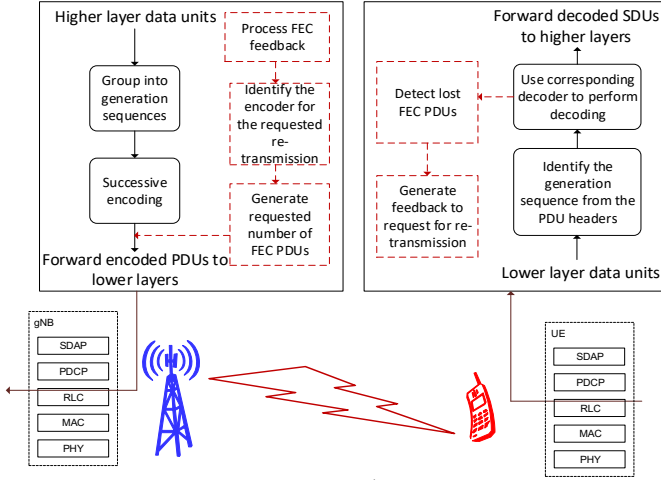
According to the proposed feedback schemes previously exposed, the capacity of the communication channel can be improved when the QoS parameters indicate the user satisfaction. The radio performance and user perceived QoS can be aligned by the type of delivered service, since different performance situations (to retransmit or not) will have different impacts into QoS. This approach is suitable for different RLC modes and is especially suitable for streaming multimedia traffic in PTM with users in different conditions [25].

ii. Link adaptation for Mixed Mode

With the feedback procedures being enabled thanks to the use of the uplink, the introduction of Adaptive Modulation and Coding (AMC) link adaptation schemes are proposed for the 5G NR Mixed Mode.

In existing NR specifications, link adaptations are typically applied for unicast communication and their suitability with

---

[3] For example, for two independent receiver probability reception of 70% (Pc = 0.7), the cumulative probability is only 49%. For three receivers, the probability of correct reception by all receivers is 34.3%.

Fig. 8. Simplified functional diagram for 2nd layer of FEC in RAN.

Fig. 9. 5G MM peak data rate gain with respect 5G NR for different number of users and carrier aggregation values.

TABLE II
NR REL-15 DOWNLINK PEAK DATA RATE (FDD, FR1, MIMO 8X8)

μ	BW_{max}(MHz)	Data Rate (Gbit/s)	CA	Data Rate (Gbit/s)
0	50	2.40		38.54
1	100	4.87	16	78.05
2	100	4.78		76.62

broadcast and multicast communication is quite complex. Moreover, NR is expected to work with significantly large antenna arrays, hence, increasing the problem complexity with respect to conventionally simple antenna configurations. The use of AMC schemes, such as Outer/Inner Loop Link Adaptation – (OLLA/ILLA), are seen as one of the most efficient feedback solutions for the Mixed Mode [26]. On the one hand, ILLA takes into account that the performance of mobile broadband systems depends on the conditions of the radio links, so that instant modification of MCS following changes of radio channel (due to fast fading, etc.) are typically applied in the following transmission time intervals. On the other hand, when the modified MCS does not improve the performance, MCS modification via OLLA techniques and HARQ should be supported.

With the aim of reducing the uplink signaling overhead, and based on [27], which proposed for Single Cell PTM (SC-PTM), a dynamic AMC scheme with only Channel Quality Indicator (CQI) feedback is proposed. This mechanism called enhanced-OLLA (eOLLA) biases the received CQI values by an offset (Δ) dependent on the number of UEs (k) as $CQI_{biased} = CQI - \Delta(k)$. From the simulations results provided in [27], it was shown that this basic and biased solution provides a similar performance as transmissions with both CQI and HARQ feedback, but simplifying receiver's design and implementation complexities.

iii.   2nd Layer FEC

The previously proposed solution although improves current state-of-the-art AMC solutions for PTM transmissions, it limits high spectral efficiency with very high reliability transmissions, as there are no means to reliably fix packet losses by the CQI reports. An alternative error correction scheme with minimal overhead of feedback messages that at the same time provides high reliability is proposed next. It is based on Random Linear Network Coding (RLNC), which is selected due to its suitability for radio channels that induce packet losses [28], and the flexibility of decoding with or without packet re-ordering as long as the required number of network coding PDUs is available at the receiver.
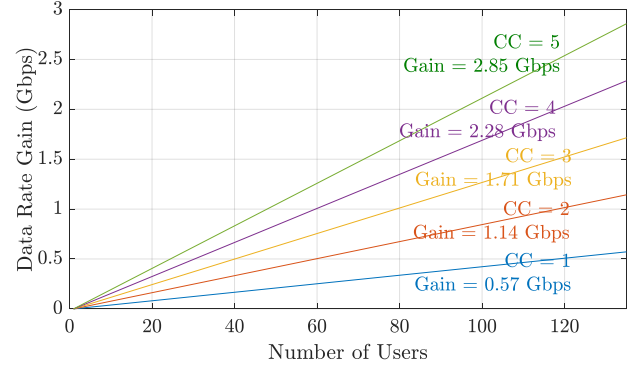
The main requirement for a UE to decode RLNC encoded data is to receive at least as many FEC PDUs as the number of encoded FEC SDUs. However, some FEC PDUs can be lost due to lossy wireless transmission channel. Hence, a certain number of extra FEC PDUs will have to be sent to the UE to compensate for the loss of packets. Existing approaches like Application Layer FEC (AL-FEC) standardized for LTE-A do this only in a pre-emptive manner, which may transmit more than needed in some situations and still not be sufficient in others. Hence, the proposal of this work is to use feedback from the UEs to signal how many more PDUs would be required. Fig. 8 depicts a simplified functional diagram for the proposed 2nd layer of FEC in RAN. Herein, higher layer data units are grouped into generation sequences upon which successive encoding is performed. Then, the encoded FEC PDUs are forwarded to the lower layers to be transmitted to UEs via the gNB multicast/broadcast channel. After UEs successfully receive the FEC PDUs from lower layer, the generation sequence of the FEC PDU is read from the PDU header and the corresponding decoder is used to perform decoding in order to extract service data units, which are forwarded to higher layers. If the UE is unable to decode all FEC SDUs of a certain generation after reception of a given number of FEC PDUs, it can use uplink feedback to signal to the network the number of extra FEC PDUs required for that generation. Then, the network transmits additional FEC PDUs from the notified generation, doing so again over the multicast/broadcast channel, which is a clear improvement over the conventional packet specific HARQ considered in [16]. Unlike HARQ feedback messages that are triggered with every reception of a packet, the FEC uplink feedback is triggered only if the UE is unable to decode after reception of FEC PDUs that are outputs of a successive encoding of SDUs from a certain generation sequence. For efficiency, this checking and reporting can be restricted to be performed only with a certain periodicity depending on the latency requirements of the service, e.g. 50ms.

TABLE III
NR Rel-15 and 5G MM Peak Spectral Efficiency

Numbers of users	NR Rel-15 (bit/s/Hz)	5G MM (bit/s/Hz)	Spectral efficiency Gain
1	48.17		0%
5	47.93		0.48%
10	47.51		1.37%
25	46.23	48.17	4.03%
50	44.09		8.45%
100	39.83		17.31%
135	36.84		23.51%

IV. NR Mixed Mode Performance Evaluation: IMT-2020 KPIs

This section evaluates the performance of the 5G NR Mixed Mode for the peak data rate, peak spectral efficiency, mobility, average and user spectral efficiency, user and control plane latency and energy efficiency KPIs, according to the evaluation methodology defined within the IMT-2020 framework [29]. The 5G NR Mixed Mode performance is compared against the current NR unicast standardized system in Rel-15.

A. Peak data rate

Following the analytical formula described in [30], the maximum achievable peak data rate for NR transmissions has been calculated for different numerology and bandwidth options in Frequency Range 1 (FR1), i.e. 450 MHz - 6 GHz. The overhead introduced by downlink control channels and signals is close to 10.5% for the evaluated configurations. The highest MIMO configuration, i.e. 8 layers, and modulation orders, i.e. 256QAM, are both considered. As shown in Table II, NR is able to provide peak data rate values up to 4.9 Gbps with current Rel-15 unicast transmissions. In 5G NR Mixed Mode, the transmission of a common CORESET avoids the transmission of the same information to all users, thus leading to an overhead reduction of 72 Resource Elements (REs) per user with respect to NR Rel-15. Fig. 9 shows the 5G NR Mixed Mode data rate gain when different number of users request the same data. For simplicity sake, ideal channel and reception conditions to reach the maximum throughput are assumed in all users. As depicted, the transmission of a single CORESET may lead up to data rate increases of 570 Mbps when 135 users are involved. It should be noticed that in spite of the data rate gain, the PDCCH overhead is still limited by the number of users that can be allocated in a PTM CORESET.

B. Peak spectral efficiency

The peak spectral efficiency is defined as the peak data rate normalized by the system bandwidth when excluding radio resources used for physical layer synchronization, reference signals, guard bands and guard times. Peak spectral efficiency is calculated per each component carrier aggregated and for the maximum MIMO configuration. For that purpose, same peak data rate values estimated above are assumed. Regarding 5G NR Mixed Mode spectral efficiency, it has been calculated considering the data rate overhead reductions. As shown in Table III, one 5G NR Mixed Mode component carrier is able to
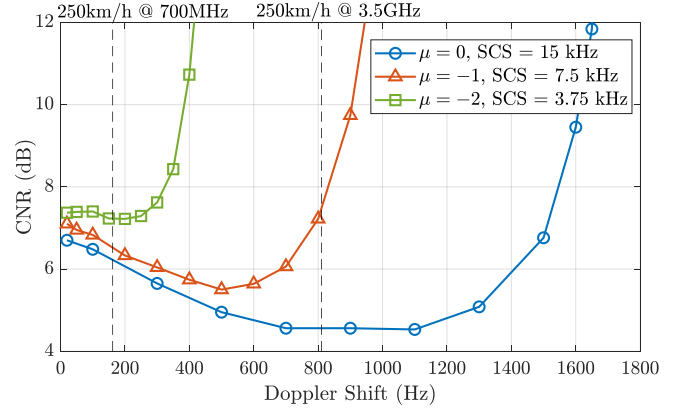

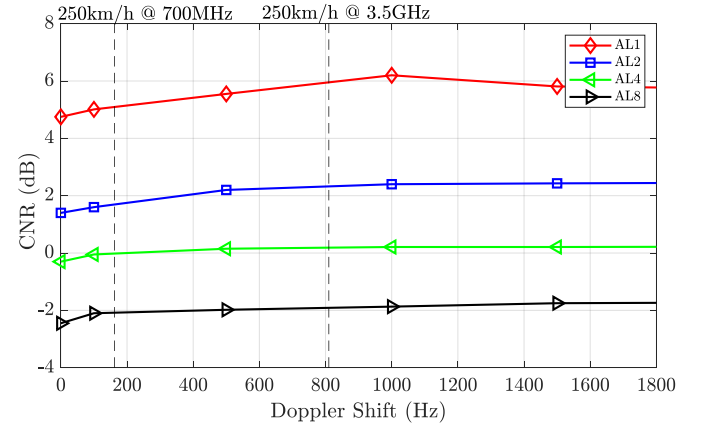Fig. 10. 5G MM PDSCH mobility performance for the proposed numerologies.

Fig. 11. CNR against Doppler Shifts for the PDCCH with TU-6 channel.

offer a constant spectral efficiency of 48.17 bit/s/Hz when different number of users request the same content thanks to the transmission of a single CORESET. Consequently, spectral efficiency gains of up to the 23.51% are introduced with respect to NR Rel-15.

C. Mobility

The mobility performance of PDSCH and PDCCH has been evaluated for different Doppler shifts by link level simulations. A Typical Urban (TU-6) has been assumed. The receiver has been configured with a Least Square real channel estimator followed by a linear time interpolation and FFT frequency interpolation.

1) PDSCH

Results in Fig. 10 show that the use of a different numerology has a great impact on the mobility tolerance. The mixed mode allows Doppler shifts of up to 1600 Hz with $\mu = 0$, equivalent to 2470 km/h and 500 km/h in the UHF band (700 MHz) or at 3.5 GHz, respectively. The other two numerologies ($\mu = -1$ and $\mu = -2$), reduce the maximum tolerated Doppler to 900 and 400 Hz, equal to 1380 and 620 km/h at 700 MHz band, and 270 and 120 km/h at 3.5 GHz band, respectively. Therefore, the 5G NR Mixed Mode negative numerologies can address the IMT-2020 speed requirement of speed tolerance $v \geq 250$ km/h in the UHF band, but only $\mu = -1$ fulfils this KPI requirement at 3.5 GHz.

2) PDCCH

In contrast with PDSCH, based on the current PDCCH pilot distribution provided in the frame structure of Fig. 2, it can be
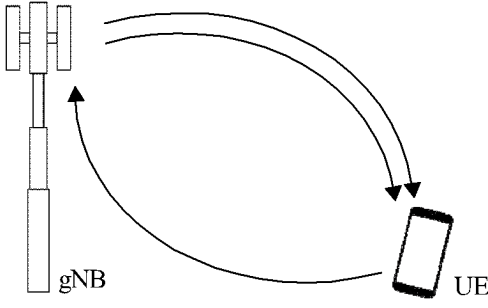
Fig. 12. User plane latency basic scheme.

TABLE IV
USER PLANE LATENCY (MS) OF THE 5G NR MIXED MODE.

Slot configuration	HARQ probability	$\mu = -2$	$\mu = -1$	$\mu = 0$
2 symbols	$p = 0$	1,78	0,93	0,56
	$p = 0,1$	2,08	1,09	0,66
	$p = 1$	4,78	2,60	1,56
4 symbols	$p = 0$	-	1,43	0,82
	$p = 0,1$	-	1,66	0,95
	$p = 1$	-	3,76	2,14
7 symbols	$p = 0$	-	-	1,19
	$p = 0,1$	-	-	1,34
	$p = 1$	-	-	2,77

seen that the PDCCH DMRS are spread throughout all the REs in time domain of the selected subcarriers. Thus, it is expected that the number of PDCCH DMRS symbols should be sufficiently dense to capture the time variation of the channel, and hence, it is expected that control information can be reliably transmitted with a very wide range of user speed tolerance under typical channels.

In [32], the PDCCH performance for the TDL-A channel with AL = {1, 2} shown that the required CNR to achieve BLER<10^{-3} it was just slightly right shifted from the static case to the moving scenario and without error floor. In order to evaluate the supported Doppler tolerance, similar analysis has been extended to Doppler shifts up to 2000 Hz. Results are shown in Fig. 11, where it can be seen that for all ALs, the CNR requirement is slightly increased but stays flat in high Doppler shift regimes. Hence, the results match our hypothesis on the considered PDCCH configurations can support all required user speed tolerance for all the NR frequency bands. Also, it can be seen that the higher the AL, the half the required CNR, due to the better coding rate used.

D. Average and 5th percentile user spectral efficiency

In [20] and [34], a performance comparison between NR and the proposed 5G NR Mixed Mode in terms of average user spectral efficiency and 5th percentile user spectral efficiency was provided. The extracted key findings were:

a) In urban eMBB with 100% penetration of indoor UEs scenario, if the number of UEs per cell is relatively low, e.g., 10 to 15 5G NR in general outperforms 5G NR Mixed Mode. When the number of UEs per cell increases, e.g., 15 to 30, the average spectral efficiency provided by 5G NR is higher than for 5G NR Mixed Mode, while the latter outperforms the former in 5th percentile user spectral efficiency for cell-edge UEs. Finally, in very dense scenarios, e.g., more than 30 UEs per cell

5G NR Mixed Mode performs better than 5G NR in terms of both average and 5th percentile user spectral efficiency.

b) In urban eMBB with 100% penetration of outdoor UEs scenario, 5G NR outperforms 5G NR Mixed Mode when the number of UEs are below 17. When the number of UEs per cell is greater than 17 but lower than 30, 5G NR Mixed Mode outperforms 5G NR for the 5th percentile user spectral efficiency for cell-edge UEs, while the average spectral efficiency provided by 5G NR Mixed Mode is higher than 5G NR when the number of serving UEs is larger than 30.

c) In eMBB indoor office hot-spot scenarios, again 5G NR outperforms 5G NR Mixed Mode if the number of serving UEs per cell is below 100. However, in very dense hot-spot scenarios with more than 230 UEs per cell, 5G NR Mixed Mode performs better than 5G NR for both average and 5th percentile user spectral efficiency.

E. User plane latency

The user plane (UP) latency is defined as the delay needed to transmit data between the gNB and the UE. It consists of the transmission (τ_1), HARQ request (τ_2) and retransmission (τ_3) between both entities as shown in Fig. 12. The transmission process can be modelled as follows:

$$T_{UP} = \tau_1 + p(\tau_2 + \tau_3) \tag{1}$$

where p is the probability of a retransmission. The specific formulas for calculating τ_1, τ_2 and τ_3 can be found in [16]. This analysis presents first an example for numerology $\mu = -2$, extended CP of 3 symbols and non-slot based scheduling of 2 symbols and probability of retransmission $p = 0.1$. It is also necessary to define values for N_1 and N_2, since they are only defined in 3GPP for positive numerologies. Values of $N_1 = 2$ and $N_2 = 2.5$ for $\mu = -2$, as well as $N_1 = 2.5$ and $N_2 = 4$ for $\mu = -1$ has been assumed.

1) First transmission: The processing time in the gNB, $t_{BS,tx}$, in this case is 356.8 μs [35]. To start transmitting the content, the gNB needs to be aligned with the first possible symbol to transmit. The gNB waits a minimum time t_{FA1} of 309.9 μs and a maximum time of 643.2 μs. On average, the time needed is 476.6 μs. Then, the TTI is transmitted in 666.6 μs. Finally, the processing time in the UE, $t_{UE,rx}$, is 285.4 μs. In summary, the total time needed for the data transmission without a HARQ retransmission is thus $\tau_1 = 1.78$ ms.

2) HARQ petition: If the data is not correctly received, then the UE sends the HARQ petition. The UE then needs 285.4 μs to process the petition, and waits an average time t_{FA1} of 95,8 μs. The time for the HARQ petition is 1 OFDM symbol, having that t_{HARQ} is 333.3 μs. The gNB then processes the request in 356.8 μs. In total, the HARQ petition needs a time $\tau_2 = 1.1$ ms.

3) HARQ retransmission: The gNB processes the retransmission in 356.8 μs and, on average, needs 286.5 μs to be aligned with the TTI and retransmit. Then, the data is retransmitted in 666.6 μs, and the UE processes the data again in 285.4 μs. The total time of retransmission $\tau_3 = 1.6$ ms. The total UP latency with a probability of retransmission $p = 0.1$ is therefore $T_{UP} = 1.78 + 0.1 \cdot (1.1 + 1.6) = 2.05$ ms.

The process can be easily extrapolated to all numerologies, as well as different slot configurations and retransmission probabilities of 0, 0.1 and 1. Table IV shows the results for all possible configurations. The results on UP latency for the 5G

TABLE V
CONTROL PLANE LATENCY (MS) OF THE 5G NR MIXED MODE.
PRACH = 1 MS.

Mapping	Symbols per slot	$\mu = -2$	$\mu = -1$	$\mu = 0$
Type A	4	-	17,6	16,3
	7	-	-	16,5
Type B	2	18,3	15,6	14
	4	-	17,6	14,6
	7	-	-	16,3

TABLE VI
CONTROL PLANE LATENCY (MS) OF THE 5G NR MIXED MODE.
PRACH = 2 SYMBOLS.

Mapping	Symbols per slot	$\mu = -2$	$\mu = -1$	$\mu = 0$
Type A	4	-	16,6	15,3
	7	-	-	15,5
Type B	2	18,3	15	13,2
	4	-	17	13,7
	7	-	-	15,1

NR Mixed Mode show that the minimum UP latency is 0.56 ms, with $\mu = 0$. This number increases to 0.66 ms if $p = 0.1$. With $p = 1$ complete retransmission, the UP latency goes up to 1.56 ms. Therefore, the IMT-2020 requirement of 1 ms would only be fulfilled by 5G NR Mixed Mode if no retransmission takes place.

F. Control plane latency

The control plane latency in 5G NR refers to the UE transition time required from idle to active state. The UE will naturally require some time to go from a battery efficient state to a starting point with continuous data transfer. The IMT-2020 requirement is 20 ms. Equation (2) describes the control plane latency calculation:

$$T = \sum_{i=1}^{10} T_i + \sum_{j=1}^{3} T_{FA,j} \qquad (2)$$

As done with the UP latency, the procedure is divided into several stages. In particular, the control plane latency calculation process is divided into 10 steps related to scheduling, processing and transmissions aspects, plus 3 additional steps related to additional frame alignments [35]. The transmission process can only start in OFDM symbols where a PRACH preamble is used.

1. The first step is related to the delay due to the RACH scheduling period. Since the transition from idle to a different state does not start until the transmission of the RACH preamble, this step is considered irrelevant and therefore $T_1 = 0$ ms.
2. The RACH preamble delay T_2 logically depends on the preamble length as specified in [36].
3. The third step is the preamble detection and processing in the gNB. The delay is calculated as $T_3 = t_{gNB}/2$, with t_{gNB} as calculated in UP latency [16].
4. After time for frame alignment, $T_{FA,1}$, the gNB sends the RA response, which takes the length of 1 slot or non-slot,

depending on the configuration used, which includes PDCCH and PDSCH. Hence, $T_4 = T_s$.
5. The UE processing delay comprehends the decoding of scheduling grant, timing alignment, C-RNTI assignment and L1 encoding of the RRC resume request. It is calculated as $T_5 = N_{T,1} + N_{T,2} + T_{wait}$, where $N_{T,1}$ is the time to transmit N_1 symbols for PDSCH reception with processing capability 1 and additional DMRS configuration. $N_{T,2}$ is the time to transmit N_2 symbols for PUSCH reception with processing capability 1. T_{wait} is the average waiting time between the reception and transmission of data, which is assumed 0.5 ms. These values can be obtained from tables 5.3-1 and 6.4-1 in [37] respectively.
6. After frame alignment $T_{FA,2}$, the next step is the transmission of the RRC resume request, which takes $T_6 = T_s$.
7. The gNB then processes the L2 and RRC request. Following the procedure given in [35], it is assumed that $T_7 = 3$ ms.
8. The following step is the slot alignment $T_{FA,3}$ and transmission of RRC resume, which takes $T_8 = T_s$.
9. Finally, the UE processes the RRC. Following [35], we assume $T_9 = 7$ ms.
10. The step 10 is considered to be the start of the data transfer, since it includes the transmission of RRC resume complete signal but user plane data. Therefore, it is assumed that $T_{10} = 0$ ms.

Tables V and VI present the control plane latency for the two negative numerologies adopted in the MC-MM. Following the same values that in the UP latency calculation, we assume $N_1 = 2$ and $N_2 = 2.5$ for $\mu = -2$, and $N_1 = 2.5$ and $N_2 = 4$ for $\mu = -1$ with UE capability 2. Note that extended CP is used with Mixed Mode and therefore some of the possibilities, initially designed for NR, are not available in this case. As expected, the 5G NR Mixed Mode introduces some additional latency compared to unicast. While in $\mu = 0$ is due to the use of extended CP, in $\mu = -1$ and $\mu = -2$ is because of having even large symbol durations. Nevertheless, the control plane latency is below the required 20 ms for IMT-2020.

G. Energy efficiency

The energy efficiency can be evaluated from both network and device perspectives. It is considered in IMT-2020 as a qualitative measure [2]. The network energy efficiency is defined in [35] as the capability of a Radio Interface Technology (RIT) or a set of RITs (SRIT) to minimize the RAN energy consumption in relation to the provided traffic capacity. On the other hand, the device energy efficiency is the capability of the RIT/SRIT to minimize the device power consumption in relation to the traffic characteristics. Note that this KPI can be in turn calculated as the sleep ratio, i.e. fraction of unoccupied resources in a period of time (%); or the sleep duration, i.e. the absolute value or continuous period of time with no transmission and reception. In this paper, network energy efficiency has been evaluated, since it demands more stringent requirements than user energy efficiency [35].
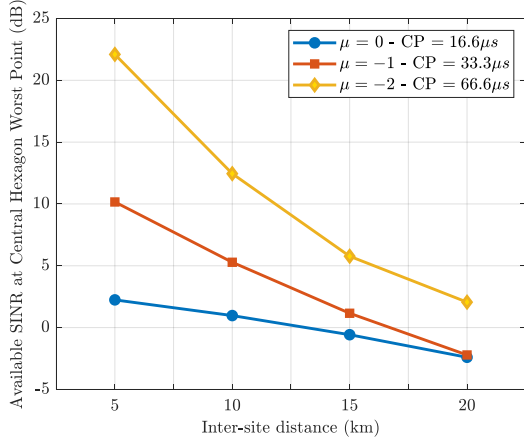
Fig. 13. Network simulations results for LPLT SFN deployments (inter-site distances 5, 10, 15 and 20 km) with negative numerologies and extended CP.
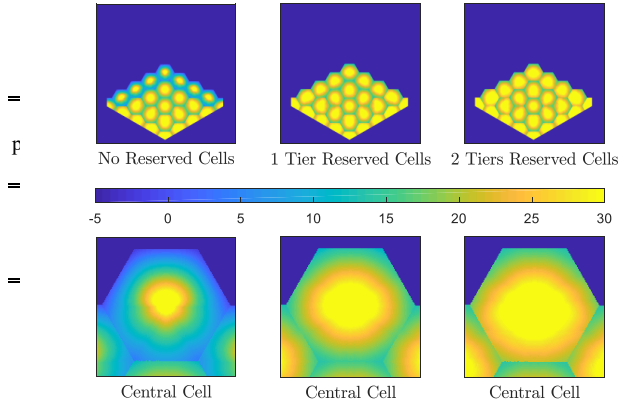

Fig. 14. SINR at the central cell. By deactivating tiers of adjacent cells, larger data rates will be possible at the expense of limiting cell coverage and disabling transmission of content in those cells.

When no data transfer takes place (unloaded case), NR performs periodical transmission of SS/PBCH blocks and Remaining Minimum System Information (RMSI) paging signals, so that UEs can access the RAN. The sleep ratio per slot is calculates as follows:

$$E_{slot} = 1 - \frac{\lceil L/2 \rceil}{2^{\mu} \cdot P_{SSB}} \qquad (3)$$

where L is the number of SS/PBCH blocks in a SS Burst set, and P_{SSB} is the SSB periodicity. The sleep duration can be easily derived by just multiplying the sleep ratio, E_{slot}, by P_{SSB}. Additionally, the sleep ratio can be calculated per symbol as:

$$E_{symbol} = 1 - \frac{L(2/7)}{2^{\mu} \cdot P_{SSB}} - \beta \frac{L/7}{2^{\mu} \cdot P_{RMSI}} \qquad (6)$$

where β is a flag variable ($\beta = 1$ for FR1 and $\beta = 0$ for FR2) and P_{RMSI} is the RMSI periodicity. In this analysis, we assume FR1 and L=1. Each SS/PBCH block occupies 4 OFDM symbols and 20 RBs. One or multiple SS/PBCH blocks compose an SSB, which is confined in half radio frame window, i.e. 5 ms. The SSB periodicity (P_{SSB}) can be 5, 10, 20, 40, 80 or 160 ms [37]. RMSI is usually time-division multiplexed with SS/PBCH and therefore it can be transmitted in the same slot. For the sake of simplicity, we assume that SS/PBCH blocks are always transmitted using numerology 0, as it is the most optimistic study case. Therefore, the energy efficiency from 5G NR is maintained, as Table VII shows. As done in NR, the 5G NR Mixed Mode performs periodical transmission of SS/PBCH blocks and RMSI paging signals, so that UEs can access the RAN. Assuming that SS/PBCH blocks are transmitted keeping $\mu = 0$, the energy efficiency from NR is maintained. A maximum energy efficiency of 99.73% has been obtained with the 5G NR Mixed Mode. Another option would be the adoption of SS/PBCH blocks to the new numerologies. In such case, block would expand $4 \times 2^{\mu}$ OFDM symbols, fitting the current window for a slot. Since the periodicity would still be the same, the energy efficiency would be also maintained.

V. NR MIXED MODE PERFORMANCE EVALUATION: PTM KPIs

In addition to the IMT-2020 KPIs evaluated in the last section, we refine some performance metrics as PTM scenario-performance, including a coverage analysis, as well as

Cumulative Distribution Functions (CDF) and Packet Loss Rate (PLR) for multicast link adaptation and 2nd layer FEC by means of system level simulations. These evaluations are aimed for providing a better knowledge of the potential of the proposed solution.

A. Coverage analysis and extensions

A set of network simulations have been conducted in order to evaluate the possibility of using the 5G NR Mixed Mode approach for distribution content over larger coverage areas. The methodology employed for the analysis is described in [38] and adapted to LPLT networks in the range 5 to 20 km inter-site distance and portable reception. The receiving environment parameters used in the simulations are in-line with [39]. These larger coverage areas can be achieved by the SFN extensions as introduced in Section III.A, or with other alternatives such as reserving cells without transmission or the use of inter-cell coordination creating multiple frequency networks (MFN).

Fig. 13 presents the available SINR at cell edge when using the negative numerologies proposed for MC-MM. It can be seen that the use of SFN modes is beneficial in dense deployments (with short inter-site distance) when, otherwise, the low SINR may result in an impractical operation for high data rates. In general, although the theoretical inter-site distance derived from the CP duration may provide a rough estimate of the feasibility of the deployment, the resultant SINR is, in the three cases studied, not higher than 5 dB.

In order to increase the SINR at cell border, it may be convenient to deactivate (applying reserved cells) adjacent gNBs. Fig. 14 shows the effect under the central cell of a hexagonal grid. By eliminating tiers of adjacent cells the available SINR increases resulting in the possibility of transmitting large data rates at the expense of not being able to extend the coverage using the same frequency. These adjacent cells may be using a different carrier frequency in order to complement the coverage area by effectively applying an MFN. This approach would provide a comparably lower spectral efficiency with respect to an SFN deployment and a more complex network planning. Note that the deployment of a SC-MM cell might also be affected by unicast transmissions on adjacent cells, therefore requiring an orthogonal resource allocation of unicast resources with respect to the content of the
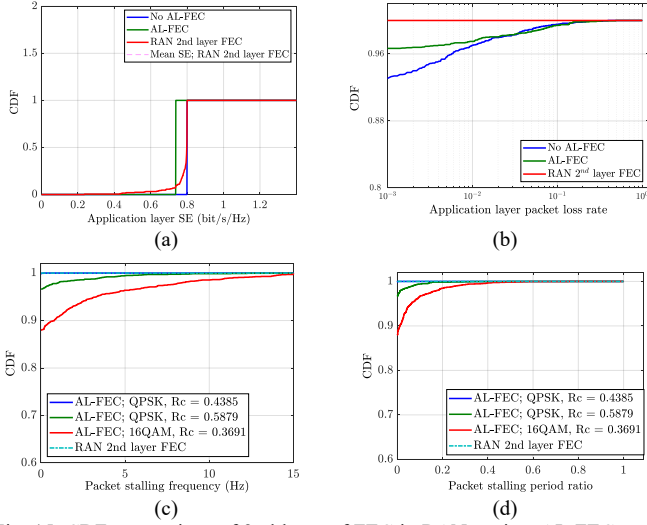
Fig. 15. CDF comparison of 2nd layer of FEC in RAN against AL-FEC and no AL-FEC, in terms of (a) application layer spectral efficiencies, (b) packet loss rates, (c) packet stalling frequency and (d) packet stalling period ratio.

SC-MM cell or the total deactivation of the carrier.

B. 2nd Layer FEC System Level Simulation

Of the various test environments defined for IMT-2020 evaluations [29], urban dense test environment is used for performance evaluation, where the detailed simulation parameters can be found in [39]. In this paper, we consider that 10 UEs per cell are dropped random uniformly. Multiple drops are considered to collect stable statistics for performance evaluation. The 2nd layer FEC mechanisms all operate in the Galois Fields (GF) 256 and with a generation size of 100 symbols, i.e., over 1 s. The proposed feedback-based 2nd layer FEC scheme is compared against two reference schemes: operation with LTE-like AL-FEC employing a systematic RLNC code with optimal decoding is used, and operation without AL-FEC.

In Fig. 15 (a) and (b), we compare the MCS = 2 CDF of 2nd layer of FEC in RAN against AL-FEC and no AL-FEC, in terms of application layer spectral efficiencies and PLR, respectively. Since the 2nd layer of FEC in RAN utilizes periodic feedback (50ms) for triggering transmission of appropriate numbers of additional RLNC PDUs to compensate for lost packets, it exhibits further improved PLR performance as compared to conventional AL-FEC. At the same time, the spectral efficiency for 2nd layer of FEC in RAN is higher than AL-FEC because additional RLNC PDUs are not sent pre-emptively, but are generated and sent only based on request. Accordingly, with the current configuration, in approximately 60% of all drops no additional RLNC PDUs are required for decoding, and in less than 10% of all drops, the overall spectral efficiency is lower than that of conventional AL-FEC, but with the benefit of reducing PLR $< 10^{-3}$. In order to achieve a high Quality of Experience (QoE), the target is to minimize the buffering delay while keeping the frequency of stalling events and the total relative time of stalls, i.e., the aggregated stalling time normalized by the observation window length, low. Fig. 15 (c) and (d) show the CDFs of packet stalling frequency and relative packet stalling period, respectively, assuming a play-out buffer
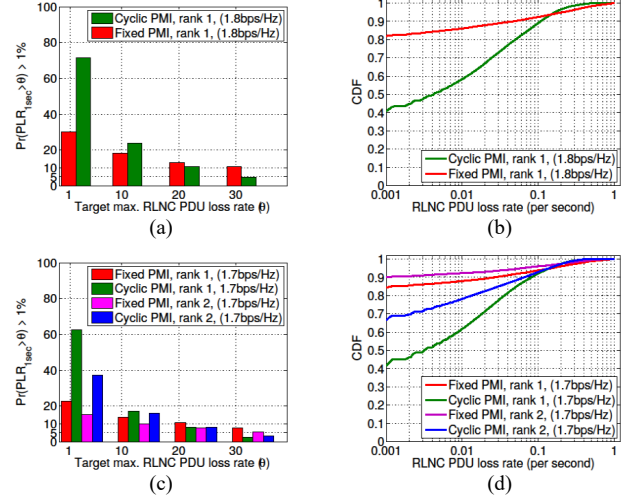

Fig. 16. System level simulation results of link adaptation for PTM: (a) The percentage of UEs that violate the QoE threshold $\epsilon = 1\%$ as a function of targeted PLR θ for fixed and cyclic PMI; (b) The overall PLR in the system for fixed and cyclic PMI; (c) The comparison in (a) with consideration of rank 1 and 2; (d) The comparison in (b) with consideration of rank 1 and 2.

size / stalling threshold of 1.1s. This value is slightly larger than what is covered by one generation of RLNC SDUs to allow the repaired packets of the systematic AL-FEC code sent at the end of the generation to repair also losses on all packets of the generation. In this case, the 2nd layer of FEC provides a better performance in terms of packet stalling frequency and packet stalling period as compared to AL-FEC. Furthermore, it can be observed that it exhibits delay characteristics very similar to those of conventional AL-FEC with MCS = 3, while the overall spectral efficiency is about 30% higher compared to this reference scheme.

C. Link Adaptation for PTM

With the same parameters setting as for 2nd layer FEC evaluation, multicast link adaptation system level simulations are performed, including investigation of MCS modifications (with heuristic fixed offsets as well as adaptive MCS via CQI report from a UE that has the worst radio link), Precoding MIMO Matrix Indicator (PMI) and Rank Indicator (RI). The fixed PMI setting refers to use of fixed precoder to PRB association throughout the simulation. On the other hand, the cyclic PMI refers to adaptive use of a precoder via cyclic access to the PMI codebook, leading to diversity benefits.

The evaluation targets achieving optimal trade-off between coverage and spectral efficiency. Herein, coverage refers to the percentage of UEs for which the probability that PLR greater than the minimum allowed loss rate θ is lower than a certain QoE threshold ϵ, configured by the network operator or planner. In this paper, the threshold ϵ is assumed to be 1%. The minimum allowed targeted PLR θ is a design parameter in combination with higher layer FEC schemes such as AL-FEC and 2nd layer of FEC in RAN. For example, for the 2nd layer of FEC in RAN, which uses RLNC, the PLRs are measured on RLNC PDUs over 1 second interval, which is the higher layer FEC interval. The targeted coverage is 95 % or above.

Fig. 16(a) demonstrates the percentage of UEs that violate the QoE threshold $\epsilon = 1\%$, i.e, $\Pr(\text{PLR}_{1\text{sec}} > \theta) > 1\%$. The

MIMO rank is configured to 1. Various targeted PLR θ are analysed for fixed and cyclic PMI settings. For the same performance in spectral efficiency (1.8bps/Hz) and at lower target RLNC PLR (e.g., $\theta = 1, 10$), the cyclic PMI has higher percentage of UEs that violate the QoE threshold. The reason is that cyclic PMI adaptively changes beams and affect PLR of most of the UEs at a lower targeted PLR θ. On the other hand, for higher target RLNC PLR (e.g., $\theta = 20, 30$), the cyclic PMI has lower violation of QoE criteria as compared to fixed PMI setting due to diversity benefits. Fig.16 (b) shows the CDF of the overall PLR in the system. Herein, it is shown that the cyclic PMI has overall lower probability of high PLR as compared to fixed PMI. When comparing the overall QoE threshold at various PLR per second, it is observed that cyclic PMI provides >95% coverage than fixed PMI.

In Fig. 16 (c) and (d), we add with respect to previous comparison the consideration of MIMO rank 1 and 2. At lower targeted PLR θ, rank 2 has lower percentage of UEs that violate the QoE threshold as compared to rank 1 for the same PMI setting. The main reason is the improved diversity benefits from rank 2. At higher targeted PLR θ, the diversity benefits from rank 2 saturates since most of the diversity benefits are already exploited by using cyclic PMI. Accordingly, cyclic PMI with both rank settings 1 and 2 provide coverage >95 %. Fig. 16(d) shows the overall PLR in the system for fixed and cyclic PMI with consideration of rank 1 and 2. Similarly, the diversity benefits of rank 2 are exhibited at lower packet loss rates for the same PMI settings. On the other hand, as it was previously stated, the diversity benefits saturate at higher packet loss rates.

VI. CONCLUSIONS

This paper extends the air interface of 3GPP 5G New Radio Release 15 to point-to-multipoint communications. The proposed solution, called 5G NR Mixed Mode, enables a flexible, dynamic and seamless switching between unicast and multicast or broadcast transmissions and the multiplexing of traffic under the same radio structures. Two 5G NR Mixed Mode operational deployments, Single-Cell Mixed Mode and Multiple-Cell Mixed Mode, have been envisaged for fulfilling the different 5G IMT-2020 usage scenarios, i.e. eMBB, URLLC and mMTC. The key principle design is to ensure the maximum compatibility with the current NR Rel-15 by reusing the original air interface and RAN upper layers as much as possible. The required modifications in the air interface include the introduction of a Group Radio Network Identifier, a multiple cell coordination for supporting Single Frequency Networks, and narrower subcarrier spacings in order to allow larger inter-site distances. The proposed solution for the RAN upper layers involves the introduction of two new logical control channels (SC-MCCH and MC-MCCH) and two new logical traffic channels (SC-MTCH and MC-MTCH) as well as new RAN procedures that will enable the seamless switching between unicast and multicast control and data radio bearers. Furthermore, an enhanced Outer/Inner Loop Link Adaptation and 2nd Layer FEC feedback schemes are proposed for the 5G NR Mixed Mode with the aim of reducing the uplink signaling overhead that may be unacceptable with current solutions for a significant number of multicast users.

The key enablers of 5G NR Mixed Mode has been evaluated across several Key Performance Indicators following the requirement criteria stated by the International Telecommunications Union within the IMT-2020 framework. It has been demonstrated that the additional modifications contemplated for the 5G NR Mixed Mode designs still fulfill the data rate, spectral efficiency, mobility, latency and energy efficiency IMT-2020 requirements. Moreover, it has been estimated that 5G NR Mixed Mode could provide up to 0.5 Gbps data rate gains compared to current NR Rel-15 thanks of alleviating the physical signaling overhead. The support of larger Single Frequency Network deployments has been validated by coverage studies, where it has been highlighted that the numerology $\mu = -2$ could provide up to 5 dB SINR gain at cell edge users. Finally, system level simulations have been carried out to analyze the 2nd layer FEC and eOLLA Precoding Matrix Indicator and Rank Indicator feedback schemes benefits. It has been shown that while the 2nd layer FEC exhibits further improved packet loss rate performance as compared to conventional AL-FEC, the fixed and cyclic PMI and RI performances depend on the targeted PDU loss rate.

REFERENCES

[1] ITU-R M.2083., "IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond," Sept. 2015.

[2] ITU-R M.2410-0, "Minimum Requirements related to Technical Performance for IMT-2020 Radio Interface(s)," November 2017.

[3] 3GPP TS 22.261 v16.3.0, "Service Requirements for Next Generation New Services and Markets," April 2018.

[4] 3GPP TR 38.913 v14.3.0, "Study on Scenarios and Requirements for Next Generation Access Technologies," August 2017.

[5] E. Arikan, "Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels," IEEE Trans. on Information Theory, vol. 55, no. 7, July 2009.

[6] D. Hui, S. Sandberg, Y. Blankenship, M. Andersson, and L. Grosjean, "Channel Coding in 5G New Radio: A Tutorial Overview and Performance Comparison with 4G LTE," IEEE Vehicular Technology Magazine, vol. 13, no. 4, pp. 60–69, December 2018.

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

[8] M. Fallgren et al.,"Multicast and Broadcast Enablers for High-Performing Cellular V2X Systems", IEEE Trans. Broadcasting, vol. 65, no. 2, pp. 454 - 463, June 2019..

[9] D. Gomez-Barquero, D. Navratil, S. Appleby, and M. Stagg, "Point-to-Multipoint Communication Enablers for the Fifth Generation of Wireless Systems," IEEE Communications Standards Magazine, vol. 2, no. 1, pp. 53–59, March 2018.

[10] D. Gomez-Barquero et. al,, "5G for Broadband Multimedia Systems and Broadcasting," IEEE Trans. Broadcasting vol.65, no.2 Part II, pp.351-355, June 2019.

[11] 3GPP RP-180669, "New SID on NR Mixed Mode Broadcast/Multicast," June 2018.

[12] 3GPP RP-19XXX, "New Work Item on NR Multicast and Broadcast Services", December 2019.

[13] M. Säilly et al., "5G Radio Access Network Architecture for Terrestrial Broadcast Services," IEEE Trans. Broadcasting, vol. 66, no. 2, Part II, June 2020.

[14] T. Tran et al., "Enabling Multicast and Broadcast in the 5G Core for Converged Fixed and Mobile Networks," IEEE Trans. Broadcasting, vol. 66, no. 2, Part II, June 2020.

[15] 3GPP TS 38.300 v15.2.0, "NR and NG-RAN Overall Description; Stage 2 (Release 15)," June 2018.

[16] E. Garro, M. Fuentes, J.J. Gimenez and J.L Carcel, Eds., "Air Interface," Deliverable D3.2, 5G-PPP 5G-Xcast Project, November 2018.

[17] E. Garro, M. Fuentes, D. Gomez-Barquero, J.L. Carcel, "5G NR Mixed Mode: An Innovative Point-to-Multipoint Solution for New Radio", in 2019 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), June 2019, pp. 1–5.

[18] 3GPP TR 36.890 v13.0.0, "Study on Single-Cell Point-to-Multipoint Transmission for E-UTRA (Release 13)," July 2015.

[19] 3GPP R1-1712879, "Forward Compatibility Consideration for ECP Design and NR-MBMS," August 2017.

[20] F. Tesema *et al.*, Eds., "RAT Protocols and Radio Resource Management in 5G-Xcast," *Deliverable D3.4*, 5G-PPP 5G-Xcast Project, May 2019.

[21] 3GPP TS 26.346, "Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs," September 2018.

[22] M. Saily *et al.*, Eds., "RAN Logical Architecture and Interfaces for 5G-Xcast," *Deliverable D3.3*, 5G-PPP 5G-Xcast Project, February 2019.

[23] 3GPP TS 36.300 v14.6.0, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); E-UTRAN; Overall description; Stage 2 (Release 14)," April 2018.

[24] Y. Jung, and J. Choi, "Hybrid ARQ Scheme with Autonomous Retransmission for Multicasting in Wireless Sensor Networks." Sensors (Basel, Switzerland) vol. 17, no. 3, February 2017.

[25] J. Nightingale *et. al.,* "5G-QoE: QoE Modelling for Ultra-HD Video Streaming in 5G Networks," IEEE Trans. Broadcasting, vol. 64, no. 2, pp. 621 - 634, June 2018.

[26] X. Zhu, D. Yang, J. Wang, and X. Zhang, "Small Area Broadcast in LTE," in 2015 International Conference on Wireless Communications Signal Processing (WCSP), October 2015, pp. 1–5.

[27] A. Awada, D. Navratil, and M. Saily, "A Study on Single-Cell Point-to-Multipoint Transmission for Public Safety Communications with eMBMS LTE Networks," in 2016 IEEE Wireless Communications and Networking Conference (WCNC), April 2016, pp. 1–6.

[28] T. Ho, M. Médard, R. Koetter, D.R. Karger, M. Effros, J. Shi and B. Leong "A Random Linear Network Coding Approach to Multicast", IEEE Transactions on Information Theory, Oct 2016.

[29] ITU-R M2412-0: "Guidelines for Evaluation of Radio Interface Technologies for IMT-2020", October 2017.

[30] J.L. Carcel, B. Mouhouche, M. Fuentes, E. Garro, D. Gomez-Barquero, "IMT-2020 Key Performance Indicators: Evaluation and Extension towards 5G New Radio Point-to-Multipoint*",* in *2019 IEEE BMSB,* June 2019, pp. 1–5.

[31] International Telecommunications Union (ITU), Draft New Report ITU-R M., "Guidelines for evaluation of radio interface technologies for IMT-2020", Oct. 2017.

[32] H. Chen, D. Mi, B. Mouhouche, P. Xiao, and R. Tafazolli, "Mobility and Coverage Evaluation of New Radio PDCCH for Point-to-Multipoint Scenario", in *2019 IEEE BMSB,* June 2019, pp. 1–5

[33] 3GPP TS 38.211 v15.1.0, "NR, Physical channels and modulation," April, 2018.

[34] M. Saily, C. Barjau, D. Navratil, A. Prasad, D. Gomez-Barquero and F. B. Tesema, "5G Radio Access Networks: Enabling Efficient Point-to-Multipoint Transmissions," IEEE Vehicular Technology Magazine, vol. 14, no. 4, pp. 29-37, Dec. 2019.

[35] 3GPP TR 37.910 v1.0.0, "Study on Self Evaluation towards IMT-2020 Submission (Release 15)", September 2018.

[36] 3GPP TS 38.211 v15.1.0: "NR; Physical Channels and Modulation", April 2018.

[37] 3GPP TS 38.214 v15.1.0: "NR; Physical Layer Procedures for Data", April 2018.

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

[39] "Simulation parameters for theoretical LTE eMBMS network studies," EBU, Geneva, Switzerland, Rep. 034, 2015

[40] F. Tesema, V. Pauli, "Layer 2 FEC in 5G Broadcast / Multicast Networks," *Nomor Whitepaper*, September 2018.