IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 21 September 2023, accepted 6 October 2023, date of publication 11 October 2023, date of current version 18 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3323815

RESEARCH ARTICLE

Optimizing Mixed Capacity of Extended Reality and Mobile Broadband Services in 5G-Advanced Networks

POURIA PAYMARD[1], (Graduate Student Member, IEEE), ABOLFAZL AMIRI[2], TROELS E. KOLDING[2], AND KLAUS I. PEDERSEN[1,2], (Senior Member, IEEE)

[1]Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark
[2]Nokia Standards, 9220 Aalborg, Denmark

Corresponding author: Pouria Paymard (pouriap@es.aau.dk)

ABSTRACT Extended reality (XR) is an emerging technology that has gained significant attention in the context of fifth-generation (5G) and 5G-Advanced cellular networks and beyond. One of the less explored areas for practical XR service deployments is the study of its interaction with the existing traffic such as enhanced mobile broadband (eMBB). This study explores the performance of having both XR and eMBB users simultaneously in a multi-cell network for two different indoor and outdoor deployment scenarios. We show that the main limitation to maximizing XR capacity in the mixed scenario is inter-cell interference (ICI) generated by eMBB users. ICI from eMBB results in a loss of about 80% in XR capacity when an XR source data rate of 45 Mbps and a strict packet delay budget (PDB) of 10 ms is enforced. To mitigate this, we propose new radio resource management enhancements that apply restrictions on eMBB radio resource usage to balance between eMBB and XR simultaneous capacity. With the proposed enhancements, maximum XR capacity can be maintained for the case with an XR source data rate of 45 Mbps and a PDB of 20 ms while restricting eMBB throughput by about 50%. The impact on eMBB throughput performance from adding XR users depends on the XR PDB, deployment environment, and the eMBB radio resource usage restriction. The results demonstrate that the eMBB throughput declines with a factor of 1 to 4 of the XR sum rate.

INDEX TERMS eXtended reality (XR), mixed traffic, inter-cell interference (ICI), interference management, 5G-advanced, system-level simulations (SLS).

I. INTRODUCTION

Extended reality (XR) is a universal term used to describe immersive technologies such as augmented reality (AR), virtual reality (VR), and mixed reality (MR). XR services are rapidly growing within a wide range of applications such as entertainment, gaming, marketing, training, and remote industrial work [1]. As XR services require high data rates, reliability, and low latency, support of such cases requires additional research and standardization actions. Among others, 5G e-health use cases are emerging rapidly that also call for XR-alike features with high data rates under bounded latency constraints [2]. For fifth-generation (5G)-Advanced wireless networks, there is ongoing work for Release-18 to offer improved XR service delivery [3] over that of Release-17 [4]. As part of the 3rd generation partnership project (3GPP) Release-17, the baseline new radio (NR) XR system-level evaluation methodology was agreed and several XR-specific key performance indicators (KPIs) were defined. Among others, this included XR performance evaluation at 28 GHz, where configurations with multi-user multiple-input multiple-output (MU-MIMO) and high sub-carrier spacing of 120 kHz were also studied. The study concluded that the current 5G networks can support XR services. In addition, several types of enhancements to potentially further boost the XR capacity were also found [4]. Recently, the 3GPP technical report in [3] identified several research directions such as enhanced channel

The associate editor coordinating the review of this manuscript and approving it for publication was Olutayo O. Oyerinde.

quality indicator (CQI) for code block groups (CBG)-based transmissions and enhanced packet scheduling to improve the XR capacity. 3GPP is now pursuing the introduction of further XR enhancements such as multiple configured grant transmission occasions, buffer status report, XR packet discarding operation, and XR traffic assistance information in the radio access network (RAN) as outlined in [5].

In addition to the 3GPP standards, the study in [6] focuses on the design and optimization of medium access control (MAC) scheduling to improve the system performance for XR services. A resource allocation problem is formulated to maximize the number of satisfied user equipment (UEs) under the data rate, reliability, and latency requirements for each UE. Moreover, in [7], the authors propose a MAC scheduling scheme based on maximum aggregate delay-capacity utility and link adaptation (LA) with a dynamic block error rate (BLER) target, which can maximize the number of satisfied XR UEs in the 5G network. LA enhancements, including new CQI schemes and the enhanced outer loop LA (OLLA), are proposed in [8] and [9], respectively, tailored to XR transmissions with CBG.

It is noteworthy that the majority of the XR studies in the literature consider XR-only traffic cases [6], [7], [8], [9], [10], [11], [12], while a common scenario in real wide-area networks is when XR services coexist together with existing services such as best-effort enhanced mobile broadband (eMBB). The paper in [13] evaluates the performance of a 5G system with mixed XR and eMBB traffic. System-level simulations (SLS) show that adding eMBB traffic to an XR-only network significantly drops the XR capacity, where this degradation is mainly contributed to the increased co-channel inter-cell interference (ICI) from the eMBB traffic.

One approach to serving users with diverse quality of service (QoS) requirements, such as XR and eMBB, is to split the wireless network infrastructure into isolated virtual slices under their own management, requirements, and characteristics [14]. However, in this study, we consider a fully flexible solution without the use of slicing, where the MAC scheduler dynamically applies QoS-aware scheduling decisions when it on a per sub-frame basis assigns transmission resources to its users. Another proper alternative for improving the XR/eMBB performance is to introduce ICI management techniques [12], [15]. The impact of ICI varies depending on the deployment environment such as indoor hotspots (InH) and dense urban areas (DU), as these environments have different characteristics; e.g. different next-generation node B (gNB) locations, deployment density, inter-site distance (ISD), line of sight (LOS) probability, and channel models [16], [17], [18], [19]. Hence, in a multi-cell and multi-user scenario, ICI coordination (ICIC) and proper radio resource management (RRM) play a key role in meeting the UEs' QoS requirements [12]. In [12], the authors highlight the ICI challenges for the XR-only studies, where the interference from neighbor cells results in large fluctuations of the signal-to-interference-plus-noise ratio (SINR) depending on the neighbor cell traffic patterns. It is concluded that coordination across gNBs to reduce ICI at XR UEs could prove crucial in improving XR performance. They observe that the number of satisfied XR UEs can be improved from 2 to 3 UEs per cell by applying ICIC. Although this study shows the importance of ICIC in an XR-only network, adding eMBB traffic to that network significantly affects the XR capacity due to the increased ICI from the eMBB traffic.

A. PRIOR ARTS ON URLLC AND eMBB CO-EXISTENCE

The design of 5G networks was made to support a wide range of services, including eMBB and ultra-reliable low-latency communications (URLLC). Numerous studies have investigated the joint performance of eMBB and URLLC in 5G networks; see e.g. [20], [21], [22], [23], and [24]. The authors in [22] study the co-scheduling problem of eMBB and URLLC traffic based on the puncturing technique. They formulate an optimization problem aiming to maximize the minimum expected achieved rate of eMBB UE while fulfilling the URLLC traffic constraints. The study in [23] investigates a resource slicing problem for a scenario of eMBB and URLLC services. The resource slicing problem is formulated as an optimization problem that aims at maximizing the eMBB data rate subject to a URLLC reliability constraint, the variance of the eMBB data rate to reduce its impact on scheduled URLLC traffic. Finally, a deep reinforcement learning algorithm is applied to puncture the incoming URLLC traffic across the allocated eMBB radio resources in the URLLC scheduling phase.

URLLC traffic with typical payload sizes of 20-50 bytes and a latency constraint of 1 ms with 99.999% reliability, is, however, significantly different from XR traffic with source data rates (SDRs) of 30-45 Mbps at 60 frames per second (fps) with 10-15 ms latency constraints at 99% reliability. In addition, multiplexing of URLLC and eMBB traffic can benefit from preemption scheduling [24], while this mechanism is not applicable for XR traffic with the transmission of large payloads, typically using the same transmission time interval (TTI) sizes as eMBB.

B. CONTRIBUTIONS

Despite the importance of joint performance evaluation in the co-existence of eMBB and XR, there is very little research on the topic [13]. In this study, we further analyze the co-existence of XR services with strict QoS requirements in combination with best-effort eMBB UEs on the same 5G-Advanced carrier. The main contributions of the paper are:

- We present state-of-the-art extensive 5G-Advanced system level performance for the co-existence of XR and eMBB under realistic conditions for different deployments.
- We show how the real-time XR capacity is influenced by also carrying eMBB traffic, and vice versa, as well as associated tradeoffs to leverage the performance between the two service classes.
- We show how the XR capacity can be protected by limiting the radio resource usage for eMBB to limit the

generated ICI, which is shown to be a major limitation for obtaining attractive XR capacity figures.
- We evaluate our methods for the two most relevant deployments for XR and eMBB coexistence, namely DU and InH scenarios.

C. PAPER STRUCTURE

The paper is organized as follows. In Section II we introduce the system model and performance indicators. We focus on two different deployments, a traditional dense urban macro-cellular case as well as an indoor hotspot scenario to cover both cases of nomadic XR/eMBB UEs and indoor UEs (XR events, VR). After introducing the simulation methodology in Section III, we show the performance results of mixed traffic in Section IV. In Section V we propose a simple, but effective, radio resource allocation and scheduling approach, where the eMBB resource utilization is limited to protect the XR UEs' performance. After proving its performance benefits in Section VI, we conclude the paper in Section VII. The abbreviations used throughout the paper are summarized in Appendix A.

II. SETTING THE SCENE

A. NETWORK DEPLOYMENT MODEL

We study two rather different network deployments in this paper that are both relevant for XR and eMBB use cases as a simple means to understand how sensitive the conclusions on joint XR and eMBB performance are to the actual deployment assumptions. Note that the DU and the InH scenarios were also in focus for the Release-17 (5G) and Release-18 (5G-Advanced) XR study items in 3GPP [3], [4], where complementary details for those scenarios can be found. Both deployments are multi-cell and multi-user scenarios with a focus on the downlink performance. We study the InH scenario as pictured in Figure 1 and the DU deployment as illustrated in Figure 2.

The InH scenario covers a 120 $\times$ 50 meters indoor open space with 12 ceiling-mounted single-cell low-power gNBs positioned in two rows with 20 meters ISD. That is each cell on average covers 500 m^2. The radio propagation channel model for InH is specified in [25] and is briefly explained in Subsection II-B. UEs are assumed to be spatially uniformly distributed within the InH scenario (i.e., all UEs are indoor), subject to the so-called even load constraint, where an equal number of UEs are served by each cell [4]. The material assumed for the walls of the InH scenario is made of concrete. We do not study cases with explicit user mobility, where a UE moves between cells and undergoes primary cell handover procedures. As background information, a survey of 5G handovers can be found in [26]. Instead, we adopted a simplified approach by assuming a constant local speed of 3 km/h to account for variations of small-scale radio propagation effects such as fast fading (the same assumptions are applied also to the DU scenario). This is well-known approach used in 3GPP simulations [4]. The InH scenario is e.g. relevant for XR gaming scenarios with indoor UEs having a gaming party, and for use cases where service agent persons are sitting in an indoor ''call center'' offering remote expert assistance through virtual reality connections.

The DU scenario consists of high-power outdoor three-sector gNBs that form a traditional hexagonal cell grid with 200 meters ISD. The DU scenario includes buildings of 6 floors. In the horizontal plane, UEs are uniformly spatially distributed, subject to having an even load per cell in terms of the number of connected UEs. A Bernoulli random variable with $q = 0.8$ probability is used to determine if the UE is inside a building Indoor UEs are uniformly distributed between floors 1 to 6, i.e. at different heights. Outdoor UEs (with a probability of 0.2) are always at a height of 1.5 meters. A 12-degree antenna downtilt is applied at the gNBs to limit the ICI, while still offering good cell coverage. The average coverage area of each cell is on the order of 13600 m^2. The radio propagation model for this scenario also appears in [25] as described in Subsection II-B. As will be demonstrated in Section IV the DU scenario has less ICI coupling as compared to the denser InH cases with fewer obstructions between cells. In an XR context, the DU scenario is relevant for XR UEs on the move, where it is desirable to offer reliable wide area XR service connectivity from macro sites to UEs independent of whether they are outdoors in streets or inside buildings without being connected to low power indoor access nodes. E.g. to facilitate wide area augmented reality for persons on the move doing sightseeing in new cities.

B. 3GPP-DEFINED RADIO PROPAGATION MODELS

We adopt the 3GPP-defined radio propagation models that have been developed over many years as defined in [25]. The 3GPP propagation is jointly developed by many partners from academia and industry, based on a large number of field measurement campaigns. It includes generic stochastic models for effects such as large-scale, small-scale, and directional (spatial) characteristics of radio propagation. The article in [27] from the special interest group initiative, summarizes conclusions and observations from measurement campaigns as input to the 3GPP propagation model. Commonly the 3GPP radio propagation models, all include a mixture of line-of-sight (LOS) and non-LOS (NLOS) conditions. Cross-correlation between different propagation metrics such as cross-correlation between polarizations, between the shadow fading and the root mean square (RMS) delay spread and the angular spreads. The effects of spatial dispersion and spatial correlation properties are part of the model as well as being important when conducting simulations with MIMO. For easy reference, a subset of the main large-scale propagation parameters for the assumed DU and InH environments are summarized in Table 1. Furthermore, the radio propagation model between a gNB and a UE takes into account the height of the transmitter and receiver, as well as additional outdoor to indoor (O2I) propagation loss for indoor UEs. NLOS propagation conditions are more frequently observed for the DU scenario (assuming a dense area with buildings).

TABLE 1. Summary of selected large-scale propagation parameters for DU and InH, extracted from [25].

	DU		InH	
	LOS	NLOS	LOS	NLOS
Pathloss exponent	2.1	3.5	1.7	3.8
Shadow fading standard deviation	4 dB	7.8 dB	3 dB	8 dB
O2I penetration loss	20 dB + inside building loss		Not relevant as all gNBs and UEs are indoor	

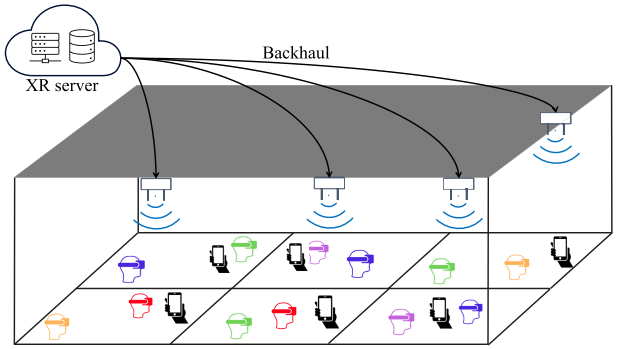


FIGURE 1. InH deployment with one eMBB and several XR UEs per cell in the downlink scenario.

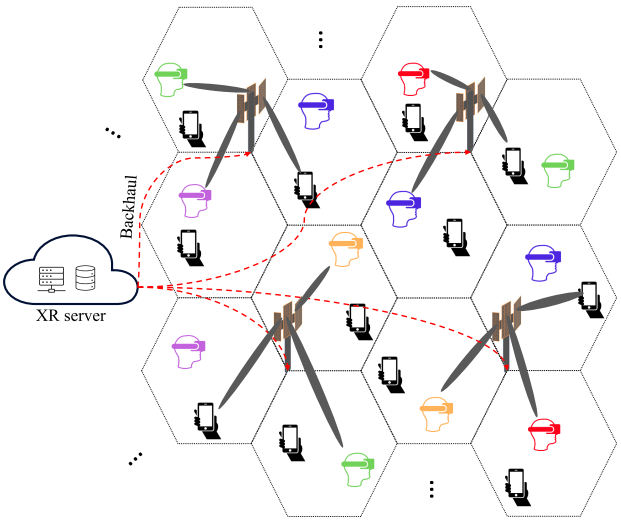


FIGURE 2. DU deployment with the existence of eMBB and XR UEs in the DL multi-cell network.

C. TRAFFIC MODELS

For XR traffic, we assume a cloud XR server that generates/renders the XR video frames with an average of 60 fps and transmits the rendered XR video frames to the corresponding gNB in line with [4]. This results in an average inter-frame arrival time of approximately 16.67 ms. The arrival time of the generated XR video frames in the gNB is affected by random jitter that is modeled with the truncated Gaussian distribution $J_f \sim \mathcal{TN}(\mu_j, \sigma_j, a_j, b_j)$, where μ_j and σ_j express the mean and the standard deviation of random jitter distribution. It is assumed that $\mu_j = 0$ ms and $\sigma_j = 2$ ms. The distribution is truncated to an interval from $a_j = -4$ ms to $b_j = 4$ ms. The XR video frame size also adheres to the truncated Gaussian distribution as $\Gamma_f \sim \mathcal{TN}(\mu_\Gamma, \sigma_\Gamma, a_\Gamma, b_\Gamma)$, where μ_Γ and σ_Γ express the average and the standard deviation of video frame size [4]. We assume 45 Mbps XR video feeds (corresponding to up-scaled 4K video quality), which means that the $\mu_\Gamma = \frac{45\text{Mbps}}{60\text{fps}} = 750$ kbits/frame. In line with [4], the truncation limits are $a_\Gamma = 0.5\sigma_\Gamma$ to $b_\Gamma = 1.5\sigma_\Gamma$.

The assumed traffic model for eMBB is commonly known as the full-buffer traffic model. It means that there is always available data in the gNB for transmission to the eMBB UEs, i.e., equivalent to assuming an infinite amount of buffered data. In each cell, there is only one eMBB UE with a full-buffer traffic model. Despite its simplicity, the assumed full-buffer traffic model for eMBB is considered useful for studying massive eMBB data downloads that will impact XR traffic, and how growing XR traffic will influence eMBB UE-experienced data rates. The same simple eMBB full-buffer traffic model was also used in past studies of the joint performance of the URLLC and eMBB [21], as well as in the recent study of XR and eMBB performance for InH cases [13].

D. FRAME STRUCTURE AND NUMEROLOGY

The network operates in a time-division duplex (TDD) mode with a slotted pattern of DDDSU, where D, S, and U represent DL, special, and UL slots respectively. The number of orthogonal frequency division multiplexing (OFDM) symbols in a slot is 14. The S slot includes ten downlink symbols, two guard symbols, and two uplink symbols. The physical downlink control channel (PDCCH) signaling is carried on the first OFDM symbol in the DL slots, and the physical downlink shared channel (PDSCH) data are transmitted over the remaining downlink symbols. We assume the default configuration for 5G deployments at carrier frequencies of 4 GHz in line with the 3GPP endorsed assumptions for XR performance studies as defined in [3] and [4]. Therefore, a sub-carrier spacing (SCS) of 30 kHz is assumed. This means that the slot time equals 0.5 ms. A 100 MHz carrier bandwidth is assumed that consists of 272 physical resource blocks (PRBs) each counting 12 subcarriers. UEs are dynamically multiplexed over the grid of time-frequency orthogonal frequency-division multiple access (OFDMA) radio resources.

E. LINK AND RANK ADAPTATION

Dynamic LA is applied to adjust the adaptive modulation and coding scheme (MCS) used for transmitting data to match the channel conditions. We assume the enhanced CQI scheme in [8], where the CQI expresses the highest supported MCS by the UE while at most N out of M CBGs in the first transmission are in error with probability P. This allows controlling the radio resources used for hybrid automatic repeat request (HARQ) retransmissions while using the highest possible MCS under such constraints. An OLLA

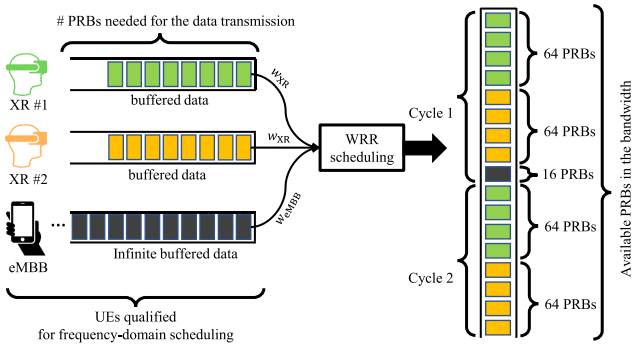


FIGURE 3. Example of WRR scheduling for two different UE groups (eMBB and XR), assuming that $w_{\text{eMBB}} = 16$ and $w_{\text{XR}} = 64$.

algorithm is used in the gNB to offset the received CQI to account for inaccuracies in UE measurement imperfections, delays in CQI reporting, and other imperfections. The complete description of the assumed OLLA scheme for the case with CBG-based transmissions can be found in [9].

All transmissions are with single-user MIMO (SU-MIMO), using either single-stream or dual-stream transmission. Dynamic MIMO rank adaptation between single- and dual-stream transmissions is conducted independently per UE for each transmission interval where it is scheduled [28]. The UE feeds back the CQI to the gNB for both Rank-1 and Rank-2 transmissions, enabling the gNB to perform rank adaptation for each scheduled transmission.

F. PACKET SCHEDULING

A weighted round-robin (WRR) scheduler is used to schedule different UEs [13], [29]. Different weights are assigned to the UEs depending on the traffic class (XR or eMBB) such that XR traffic is prioritized over eMBB. An example of the scheduler is illustrated in Figure 3. The basic principles of the scheduler are as follows: First, the scheduling of pending HARQ retransmissions is prioritized. Afterward, UEs with the buffered data are scheduled in a priority-based manner according to the UEs' weights. The scheduler allocates PRBs in a cyclic manner, where each UE, in each cycle, is assigned PRBs in proportion to its scheduling weight. Higher scheduling weight means that the UE will be assigned more PRBs. As seen in the exemplified figure, a scenario with three UEs (2 XR + 1 eMBB) in each cell and a total number of 272 PRBs in the carrier bandwidth is assumed. In each TTI, these available PRBs are assigned to UEs. The scheduler allocates PRBs in a cyclic manner, where each UE is assigned PRBs in proportion to its scheduling weight. In the first cycle, the WRR scheduler starts with XR #1 and allocates 64 PRBs (since the scheduling weight of XR UEs is 64), then it moves to the second XR UE and similarly allocates 64 PRBs. Lastly, it selects the third UE (the eMBB UE), which has a weight equal to 16 and accordingly allocates 16 PRBs. Immediately after the end of Cycle 1, the second cycle starts, and each of the XR UEs (XR #1 and XR #2) is allocated 64 PRBs based on their weights. Since there are no more available PRBs left in the bandwidth ($64 + 64 + 16 + 64 + 64 = 272$), the eMBB UE is simply not scheduled in Cycle 2. Basically, the scheduler continues to allocate PRBs for a candidate UE until there are no more buffered data for the UE at the gNB or the scheduler runs out of available PRBs. It is worth mentioning that w_{eMBB} is chosen intentionally high for this example for more visible illustration purposes.

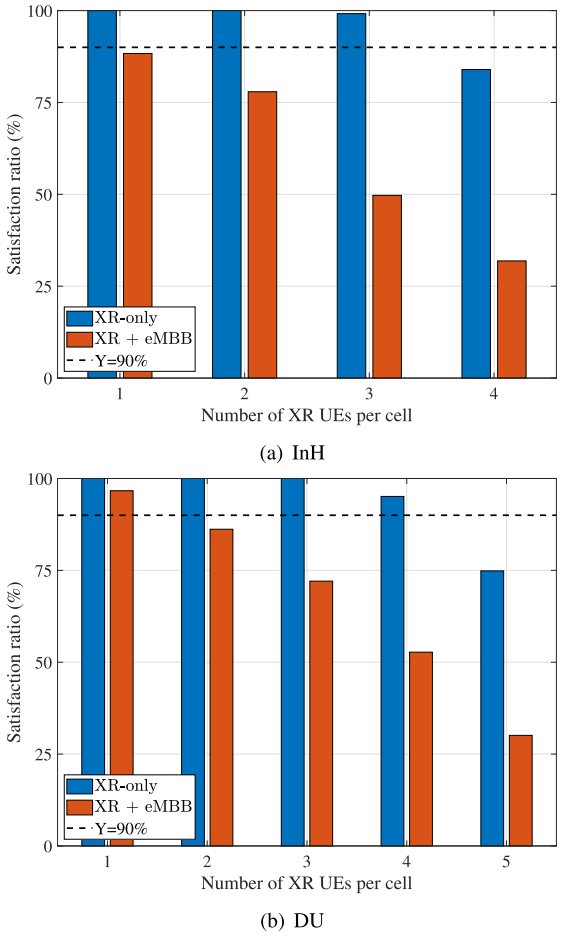
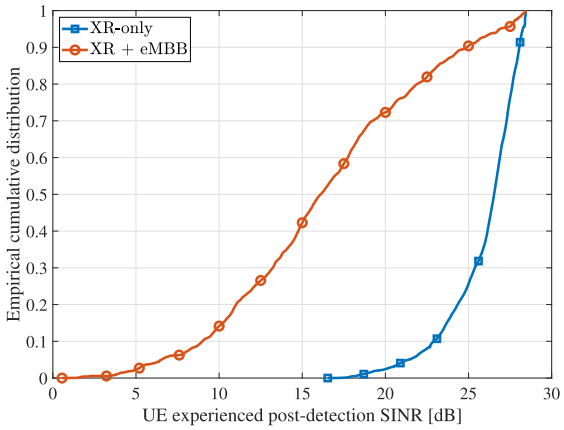


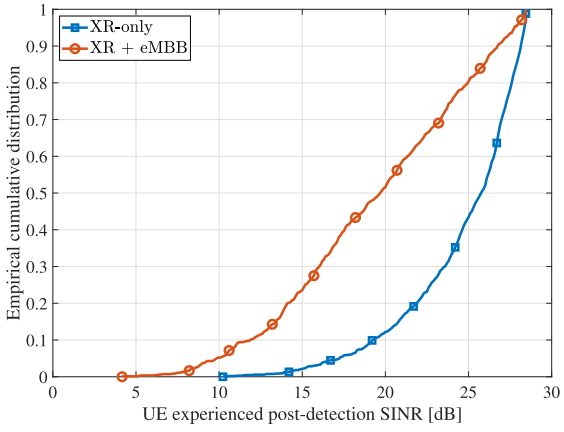
FIGURE 4. UE satisfaction ratio versus the number of connected UEs per cell for schemes with/without eMBB traffic in different deployment scenarios.

G. KEY PERFORMANCE INDICATORS (KPIs)

In line with the 3GPP evaluation methodology [4], the main KPI for XR cases is the supported XR capacity, which is defined as the maximum number of XR UEs where at least 90% of them are satisfied. An XR UE is satisfied if at least 99% of its video frames are received successfully within the packet delay budget (PDB). The value of PDB varies for different applications and traffic types. As per [3], [4], the PDB of cloud gaming use cases is assumed to be 10, 15, or 30 ms, while VR applications need the PDB of 5, 10, or 20 ms. As will be shown in the results sections, the XR capacity naturally varies for the considered deployments depending on the presence of eMBB UEs and the assumed



(a) InH, 3 XR UEs per cell



(b) DU, 4 XR UEs per cell

FIGURE 5. eCDF of the XR UE experienced post-detection SINR after interference rejection for the InH and DU scenarios.

PDB, among other factors. In order to gain a deeper insight into the XR performance, we also inspect other statistics such as experienced XR packet latency, accounting for various effects such as potential gNB queuing, processing times at the gNB and UE sides, potential HARQ retransmissions, etc. The delay is derived by collecting statistics for each XR frame transmission. For the eMBB UEs, the primary KPI is the average aggregated throughput (TP) per cell. This eMBB KPI is naturally influenced by the number of XR UEs per cell. This KPI is calculated from the average UE TP samples within each cell per simulation run. We also consider standard radio performance KPIs such as the average percentage of utilized PRBs per cell as measured by the radio load utilization and UE-experienced post-detection SINR after interference rejection combining at the UEs [30], [31]. The SINR statistics are built on the collection of experienced post-detection SINR of all transport block (TB) receptions for both first transmissions and HARQ retransmissions. Finally, we define the fairness score as the ratio of the average eMBB cell TP degradation to the sum rate of satisfied XR users. The expression for this KPI is given in Section VI-C and it quantifies the extent to which the eMBB TP decreases when XR UEs are introduced to a cell.

III. EVALUATION METHODOLOGY

Simulations are carried out by an advanced dynamic system-level simulator in which the simulation assumptions and methodology are aligned and calibrated against the 3GPP Release-18 NR evaluation guidelines outlined in [3]. The simulations include modeling of physical (PHY), MAC, radio link control (RLC), packet data convergence protocol (PDCP), internet protocol (IP), transport, and application layers. The simulator includes RRM-related functionalities such as CQI measurements, reporting, inner loop LA, OLLA, CBG-based HARQ transmissions, packet scheduling, etc.

For each scheduled transmission, the SINR per resource element (RE) at the receiver end is calculated. The SINR calculation takes the effect of the SU-MIMO into account with the used transmitter precoder and minimum mean-square-error interference rejection combining (MMSE-IRC) receiver. Following this, the per-RE SINR values for the transmission are mapped to an effective SINR per CBG and TB and, subsequently, the mean mutual information per bit (MMIB) is computed [32]. Given the effective SINR, MMIB, and the used MCS for the transmission, the block error probability (BLEP) of the transmission is obtained from a lookup table. That look-up table is obtained from extensive link-level simulations. A failure to correctly decode a transmission will trigger an asynchronous adaptive HARQ retransmission. The simulation assumptions are summarized in Table 2.

To ensure statistically reliable simulation results for the KPIs of interest, each simulation is run for at least 10 seconds, corresponding to 600 simulated video frames per UE. Each simulated configuration is repeated 10 times to have sufficient statistical reliability for estimating the XR capacity. The confidence interval calculations for the simulations are addressed in Appendix B.

IV. PERFORMANCE RESULTS WITH/WITHOUT eMBB TRAFFIC

Figure 4 shows the XR UE satisfaction ratio versus the number of XR UEs per cell for the two considered scenarios (InH and DU) for 10 ms PDB for XR. Throughout the paper, we assume that there exists only one eMBB UE per cell for the cases of mixed traffic (*XR+eMBB* scheme). It is clear from this plot that the co-existence of eMBB traffic has a significant impact on the XR performance in both the InH and DU scenarios. As an example, for the DU scenario, the XR system capacity is reduced to only one satisfied XR UE per cell when eMBB traffic is present. Without eMBB traffic, the XR system capacity equals 3 satisfied XR UEs per cell in the InH scenario and 4 UEs for DU which implies XR cell TP of $3 \times 45 = 135$ Mbps and $4 \times 45 = 180$ Mbps, respectively. Although the supported XR capacity in terms of the number of UEs per cell looks similar, there are differences worth noticing. Especially when taking into consideration that the area supported by each cell is significantly different for InH and DU. Taking this into account, one happy XR UE can be supported per 167 m^2 for the InH scenario and 2891 m^2 for

TABLE 2. Summary of system-level evaluation parameters.

Parameter	Setting	
Deployment	InH	DU
World Area	120m × 50m	528 m × 460 m
Layout	12 cells	21 cells
Inter-site Distance	20 m	200 m
gNB height	3 m	25 m
gNB Tx power	31 dBm	51 dBm
gNB down tilt	90°	12°
Indoor UE probability	1	0.8
Number of building floors	1	6
UE distribution per floor		uniform(1,6)
Simulation time	10 sec per run	
Simulation runs	10 runs	
TDD Frame structure	DDDSU	
TTI length	14 OFDM symbols	
PDCCH overhead	1 OFDM symbol	
Carrier frequency	4 GHz	
Bandwidth	100 MHz	
SCS	30 kHz	
MIMO scheme	SU-MIMO with rank adaptation	
Modulation	QPSK to 256QAM	
gNB Tx processing delay	2.75 OFDM symbols	
gNB antenna	1 panel with 32 elements (4×4 and 2 polarization)	
UE speed	3 km/h	
UE height	1.5 m	
UE Rx processing delay	6 OFDM symbols	
UE receiver	LMMSE-IRC	
UE antenna polarization	dual-polarized	
Number of UE antennas	2 Rx antennas	
eMBB Traffic model	full-buffer best effort	
number of eMBB UEs/cell	1	
XR Traffic model	quasi-periodic	
XR random Jitter	$\mathcal{TN}(0, 2, -4, 4)$ ms	
XR frame rate	60 fps	
XR SDR	45 Mbps	
XR frame size (45Mbps)	$\mathcal{TN}(93, 10, 46, 141)$ kB	
Scheduler	M-WRR	
w_{XR}	1000	
w_{eMBB}	1	
HARQ scheme	CBG-based HARQ retransmissions	
HARQ combining method	Chase combining	
Channel estimation	Realistic	
CQI	Periodic CQI every 2.5 ms	
Cell Selection	RSRP Slow Fading	
Target CBG error rate	2 out 8 CBGs with probability of 50%	

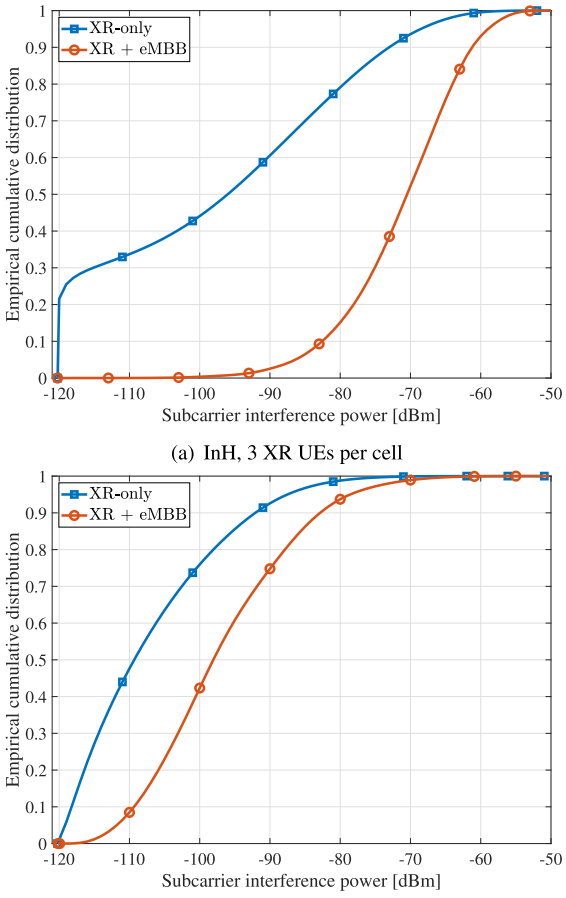


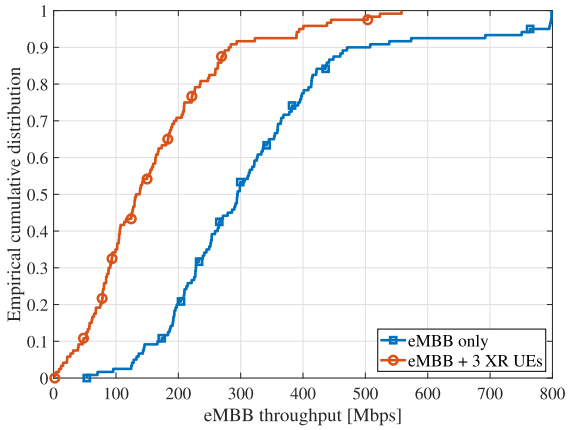
FIGURE 6. eCDF of the subcarrier interference power in the receiver side for InH and DU scenarios.

the DU scenario, when XR-only traffic is assumed. Thus, the density of supported satisfied XR UEs is a factor 17 higher for the InH scenario, as compared to DU, due to its much higher cell density.

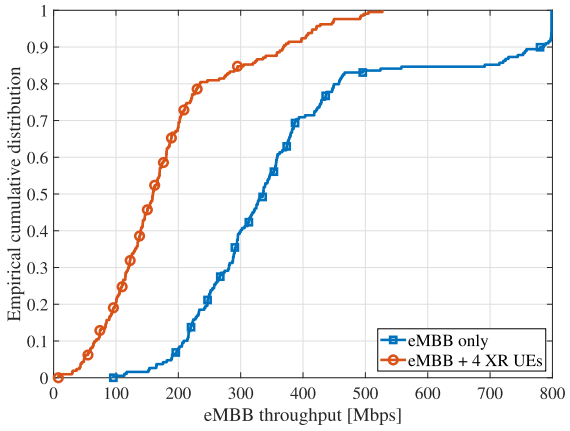
Figure 5 shows the empirical cumulative distribution function (eCDF) of UE experienced post-detection SINR after interference rejection for the InH and DU scenarios. The eCDF is built on the collection of SINR samples of both first transmissions and HARQ retransmissions for XR UEs. These results are plotted for 3 and 4 XR UEs per cell for the InH and DU scenarios, respectively. The highest SINR level is experienced when there is only XR traffic in the network as the network operates at fractional PRB load under such conditions. For XR-only cases, the PRB utilization is only on the order of 40-45% in the DU and InH scenarios, respectively. For the cases with both XR and eMBB traffic, the PRB utilization grows to 100%, resulting in the lower experienced SINR due to higher ICI. At the 50$^{\text{th}}$ percentile of the eCDF, the SINR is degraded by about 10 dB and 7 dB in the InH and DU scenarios, respectively, when the traffic goes from XR-only to mixed XR-eMBB.

Figure 6 illustrates the eCDF of per-subcarrier received ICI power at the XR UEs receiver. Comparing such results for the InH and DU scenarios, it is evident that the InH scenario exhibits higher levels of ICI. On average, the ICI is 26 dB higher for the InH scenario, as compared to DU, for the cases with eMBB. Taking into account that the gNB transmit power is 20dB higher for the DU scenario (51 dBm), as compared to InH (31 dBm), this translates to effectively 6dB higher ICI as compared to desired signal levels.

Figure 7 presents the eCDF of average eMBB cell TP. This eCDF is constructed from average cell TP samples from each of the eMBB UEs per simulation run. With no XR UEs, the average eMBB cell TP approximately equals 325 Mbps for the InH scenario and 383 Mbps for the DU scenario. With XR UEs in the system, a significant reduction in the eMBB TP is observed, since scheduling the XR UEs is prioritized over eMBB UEs. For the InH scenario with 3 XR UEs at 45 Mbps,



(a) InH



(b) DU

FIGURE 7. eCDF of the average eMBB cell TP for both InH and DU scenarios.

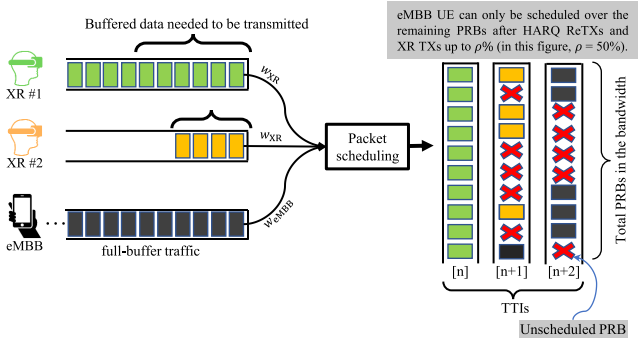


FIGURE 8. Example of M-WRR scheduling for two UE groups, assuming that $w_{\text{eMBB}} = 1$ and a large weight for XR, e.g., $w_{\text{XR}} = 1000$.

the eMBB TP drops by approximately 135 Mbps. Similarly, when 4 XR UEs are added in the DU scenario, the eMBB TP drops by roughly 180 Mbps.

V. EMULATING ICIC VIA eMBB

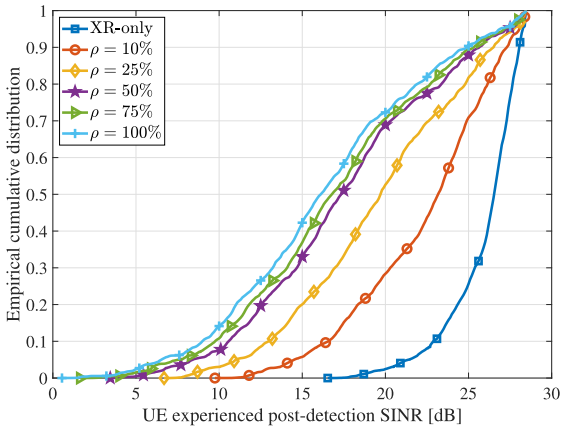
PRB Utilization Limit Based on the observations in Section IV, we investigate the potential performance tradeoffs by imposing restrictions on eMBB radio resource usage to limit the ICI generated from eMBB transmissions. These

Algorithm 1 Radio Resource Allocation per TTI for Mixed XR and eMBB Traffic

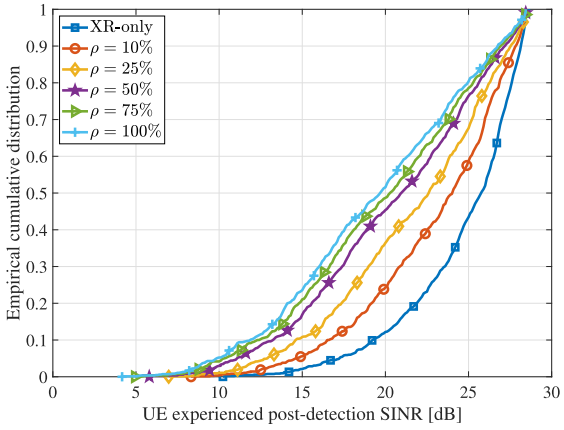
Inputs & Parameters:
$\mathbb{C}_{\text{UE}}$: candidate UEs,
L_{UE}: A Boolean parameter represents whether there is still data in the gNB buffer to be transmitted to the UE.
$\mathbb{V}_{\text{PRB}}$: set of available PRBs in a TTI,
β_{PRB}: A counter for the allocated PRBs.
S_{PRB}: the maximum number of PRBs in the bandwidth.
$w_{\text{UE}} = \{w_{\text{eMBB}}, w_{\text{XR}}\}$: UEs' class weights
ρ: The upper band portion for the eMBB allocation.
Output: A vector that indicates the allocated PRBs for the corresponding UEs at the serving cell.

1: Order candidate UEs based on their priority. XR always has a higher priority than eMBB.
2: Schedule the HARQ retransmissions.
3: Update $\mathbb{V}_{\text{PRB}}$.
4: **while** $(\mathbb{V}_{\text{PRB}} \neq \emptyset)$ **do**
5: **for** UEs $\in \mathbb{C}_{\text{UE}}$ **do**
6: $\beta_{\text{PRB}} = 0$.
7: **while** $(L_{\text{UE}}$ & $\beta_{\text{PRB}} < w_{\text{UE}}$ & $|\mathbb{V}_{\text{PRB}}| > 0)$ **do**
8: **if** UE class == XR **then**
9: Assign one PRB
10: $++\beta_{\text{PRB}}$
11: **else if** UE class == eMBB **then**
12: **if** $|\mathbb{V}_{\text{PRB}}| - (1-\rho) \times S_{\text{PRB}} > 0$ **then**
13: Assign one PRB
14: $++\beta_{\text{PRB}}$
15: **else**
16: Stop allocating PRBs to the UE.
17: **end if**
18: **end if**
19: Update $\mathbb{V}_{\text{PRB}}$.
20: **end while**
21: **end for**
22: **end while**

restrictions are placed to keep the XR UEs' experience unaffected. We choose to mimic the ICIC by modifying the scheduler and limiting the eMBB resource usage. This is described in Algorithm 1. The main body of the scheduler is similar to the WRR method described in Section II-F, where the scheduler starts with allocating resources to the retransmissions and then continues to allocate PRBs for a candidate UE until either no buffered data is left or all PRBs are used. The scheduling order is based on pre-configured weights for each UE class, e.g., XR and eMBB. UEs with XR service are freely scheduled, while the eMBB UE's allocation is limited to reach at most $\rho\%$ of the PRBs used for the transmission in the carrier frequency bandwidth. This restriction is expressed in Step 12 of Algorithm 1. Hence, the eMBB UE can only be scheduled over the remaining PRBs after HARQ retransmissions and XR transmissions up to $\rho\%$. A detailed analysis of the complexity of Algorithm 1 is presented in Appendix C. As can be

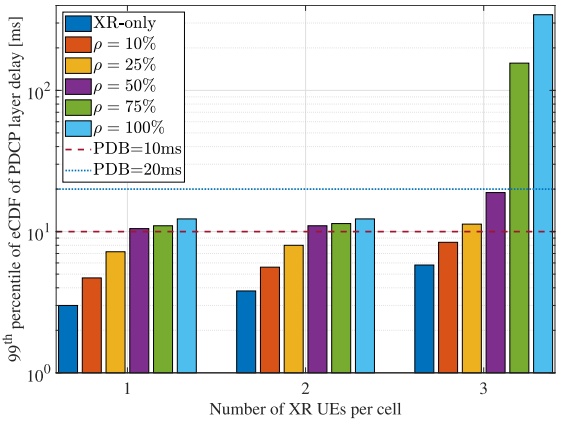


(a) InH, 3 XR UEs per cell

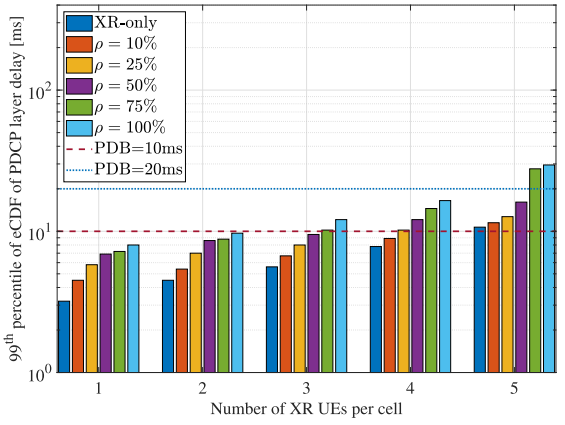


(b) DU, 4 XR UEs per cell

FIGURE 9. eCDF of the XR UE experienced post-detection SINR after interference rejection in the receiver side for different values of ρ in both InH and DU scenarios.



(a) InH



(b) DU

FIGURE 10. 99th percentile of eCDF of PDCP layer delay versus the number of XR UEs per cell for different values of ρ.

seen, the complexity of the Algorithm is very similar to the WRR scheduler (except for step 12 of Algorithm 1), and thus it is affordable for state-of-the-art base station implementations.

A simple example is illustrated in Figure 8. Here, the assumption is that there are no queued retransmissions, i.e. pending HARQ retransmissions are transmitted/scheduled already. Afterward, XR UEs have the priority to fill out the remaining PRBs for the transmissions. The UEs are scheduled in a priority-based manner according to their weights. In this example, the assumed weights are $w_{\text{XR}} = 1000$ and $w_{\text{eMBB}} = 1$ meaning that XR UEs are always prioritized over eMBB UEs to take up whole PRBs in the bandwidth. If there are still more than ρ% of the total number of PRBs unused (In Figure 8, ρ is 50 %), those are assigned to the eMBB UEs.

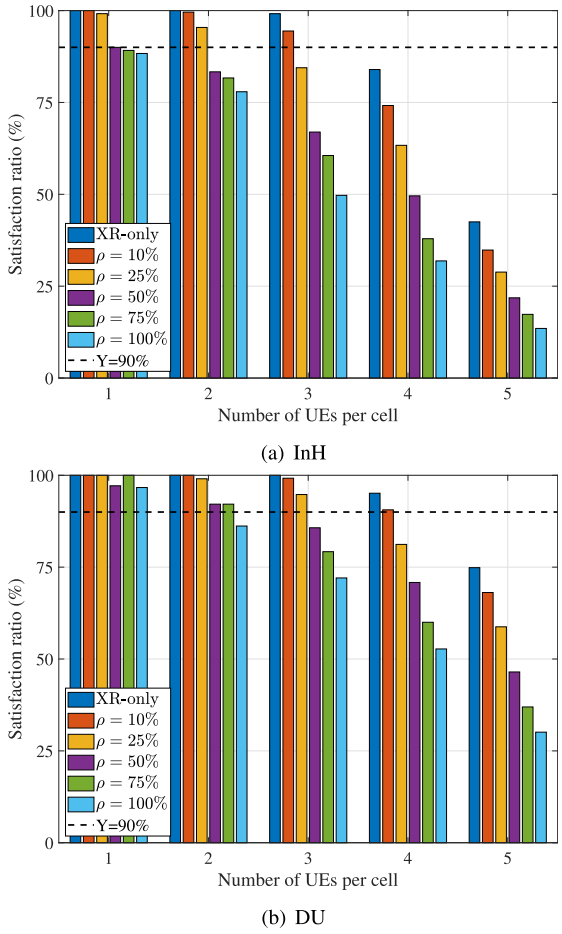
VI. PERFORMANCE RESULTS FOR THE eMBB PRB UTILIZATION LIMIT

We evaluate the performance of our new scheduling scheme with restrictions for eMBB PRB usage with the following scenarios:

1) *XR-only* case to create a baseline for the best XR experience.
2) Mixed traffic case with $\rho \in \{10\%, 25\%, 50\%, 75\%, 100\%\}$ to analyze the impact of the eMBB PRB restriction on XR and eMBB KPIs.

A. THE IMPACT OF eMBB PRB RESTRICTION ON THE XR PERFORMANCE

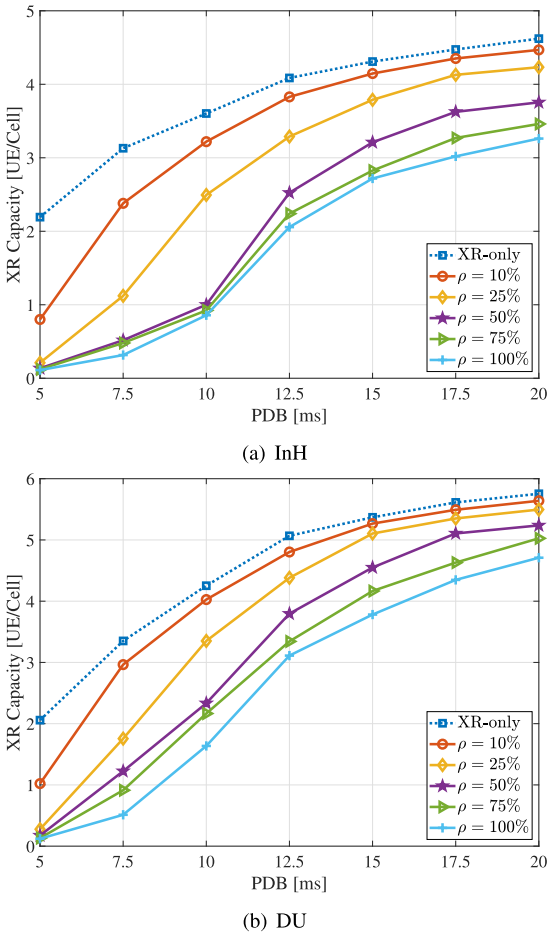
The eCDF of XR UEs' post-detection SINR after interference rejection is shown in Figure 9 for different values of ρ, where there are 1 eMBB UE and 3 (or 4) XR UEs for the InH (or DU) scenarios in each cell, in sub-figures (a) and (b), respectively. The number of XR UEs is fixed based on the maximum XR capacity observed in Figure 4. As expected, experienced SINR for the XR UEs is improved when reducing the value of ρ as it reduces the overall ICI level in the system. However, SINR improvement is not a linear function of ρ as larger ρ values (e.g., $\rho \geq 50\%$) do not improve the SINR significantly. In addition, it is seen that XR UEs experience higher SINR levels in the DU scenario than in the InH scenario in the mixed traffic scenario with large ρ values, thanks to the environmental isolation in the former case.



(a) InH

(b) DU

FIGURE 11. UE satisfaction ratio versus the number of connected UEs per cell for different values of ρ, considering a PDB of 10 ms.



(a) InH

(b) DU

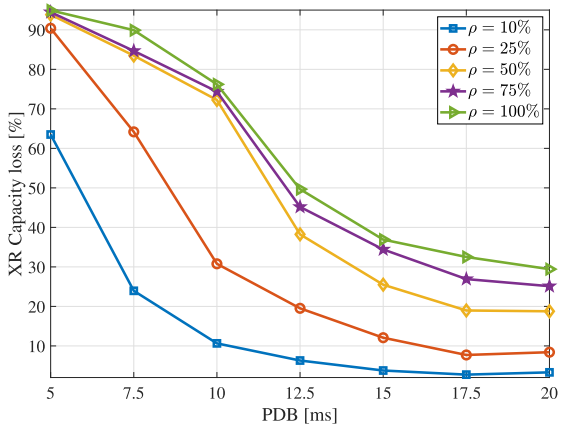
FIGURE 12. XR capacity versus the PDB for different values of ρ.

The impact of the number of XR UEs per cell and ρ on the 99th percentile of eCDF of PDCP layer delay, which should be less than PDB for satisfied XR UEs, is illustrated in Figure 10. The packet latency accounts for the duration of the radio access part of the data transmission. It is calculated from the moment an XR payload arrives at the PDCP layer of the gNB until it is successfully received at the PDCP layer of the UE. The latency includes the queuing delay of the payload at the gNB, the transmission time of the payload, and the processing time at both the gNB and the UE. If the UE fails to decode the initial transmission, the latency calculation also considers the round-trip time of the HARQ process. As expected, the delay grows with more XR UEs. Similarly, larger ρ values result in higher application layer delays. Consequently, lowering the ρ value helps improve the delay and keep it under the PDB.

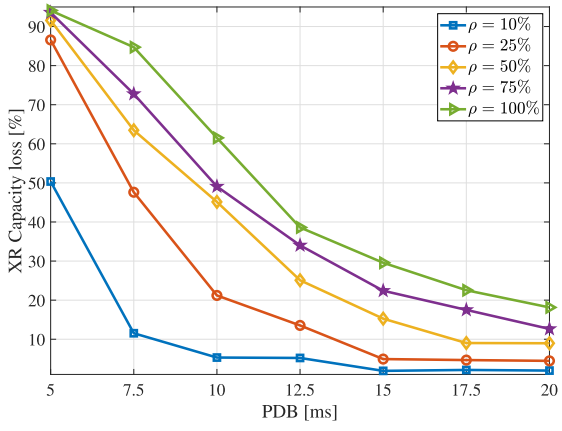
In addition, Figure 10 shows an improved delay for the DU scenario as compared to the InH scenario which is the consequence of higher UE SINR levels that lead to the selection of higher MCS indexes, and accordingly, higher TP and less transmission times. As a comparison among different ρ values, the delay of $\rho = 10\%$ scheme is below 10 milliseconds target, while $\rho \geq 50\%$ cannot meet 10 ms PDB in the InH scenario with 3 XR UEs per cell.

Figure 11 shows the XR satisfaction ratio versus the number of connected UEs per cell for different values of ρ. By limiting PRB utilization for eMBB UEs, the XR system capacity comes close to the XR-only scheme which is the result of the eMBB ICI mitigation technique. For instance, with $\rho = 25\%$ in the DU scenario, the XR system capacity stays unchanged with respect to the XR-only baseline. In other words, not only the XR services can run unaffected, but also additional eMBB traffic can be carried by the network. To keep the XR system capacity unchanged in the InH scenario, the ICIC limit must be set to stricter values such as $\rho = 10\%$ due to a higher ICI level.

Figure 12 shows the impact of PDB on the XR system capacity for different values of ρ. As expected, services with larger PDB can tolerate longer successful packet reception times, especially when the network is heavily loaded or multiple HARQ retransmissions are required. Consequently, larger PDB cases, e.g., 20 ms, result in a higher number of satisfied UEs compared to using a tighter PDB duration, such as 5 ms or 10 ms. In the DU scenario, if the XR service requires small PDB values, maintaining the XR system capacity necessitates putting very strict limitations,

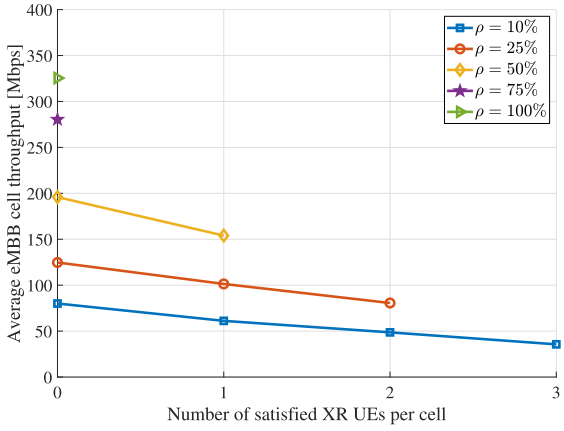


(a) InH

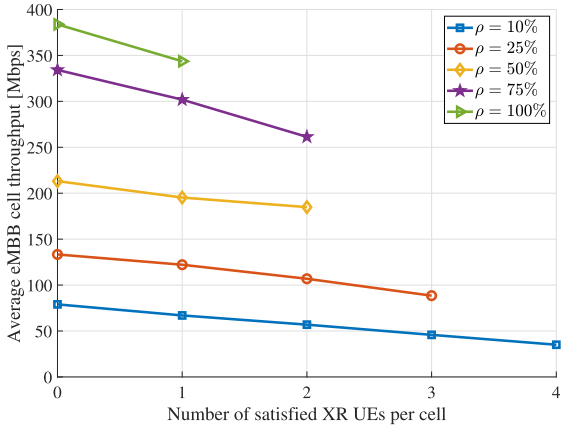


(b) DU

FIGURE 13. XR capacity loss versus PDB for different values of ρ.



(a) InH



(b) DU

FIGURE 14. Average eMBB cell TP versus the number of XR UEs per cell for different values of ρ. when XR PDB is 10ms.

e.g., $\rho = 10\%$, on the eMBB traffic. For relaxed PDB cases with 20ms, there is no need to impose hard restrictions on the eMBB PRB utilization. On the other hand, In the InH scenario, even for relaxed PDB cases, having a strict limit on the eMBB traffic (e.g., $\rho = 25\%$) is essential to ensure that the XR capacity remains intact. This observation indicates that the XR capacity remains unaffected by eMBB traffic as long as appropriate limitations are placed based on the PDB duration.

The XR capacity loss for different ρ and PDB values is illustrated in Figure 13. The relative loss is defined as the degradation of the XR capacity from adding eMBB compared to the XR-only case. It is formulated as:

$$\text{CL} = 1 - \frac{\text{XR capacity for mixed traffic case}}{\text{XR capacity for XR-only case}}. \quad (1)$$

As illustrated, the capacity loss is significant for small PDBs and when the eMBB UE has the freedom to utilize more than half of the whole bandwidth i.e., $\rho \geq 50\%$. Services with relaxed PDBs have lower XR capacity loss. For instance, for the scheme of $\rho = 50\%$ and comparing the PDB of 10 ms with 20 ms, the XR capacity loss difference is around 53 % and 36 % in the InH and DU scenarios, respectively. On the other hand, with a more restricted eMBB PRB limit (e.g., $\rho = 10\%$), the XR capacity loss is near 10 % and 5 % in the InH and DU scenarios, respectively, for PDB of 10 ms.

B. THE IMPACT OF eMBB PRB RESTRICTION ON THE eMBB PERFORMANCE

Figure 14 presents the impact of serving XR UEs and different values of ρ on the average eMBB cell TP. As expected, the XR UEs cause a big reduction in the average eMBB cell TP. For instance, in Figure 13(b), the average eMBB cell TP with $\rho = 100\%$ is close to 383 Mbps, while it is declined to about 80 Mbps, i.e. 80 % reduction, for $\rho = 10\%$.

Figure 15 shows the impact of different ρ values on the XR capacity for different PDBs and the eMBB TP in different environments. The figure illustrates the trade-off between the XR system capacity and eMBB TP. As seen, the slope of the XR capacity curve for PDB of 10 ms is steeper than the 20 ms case. This is due to the higher sensitivity of the services with smaller PDB to eMBB traffic. In addition, the degradation of XR capacity for PDB of 10 ms in the InH scenario is more significant compared to the DU scenario. This implies that the

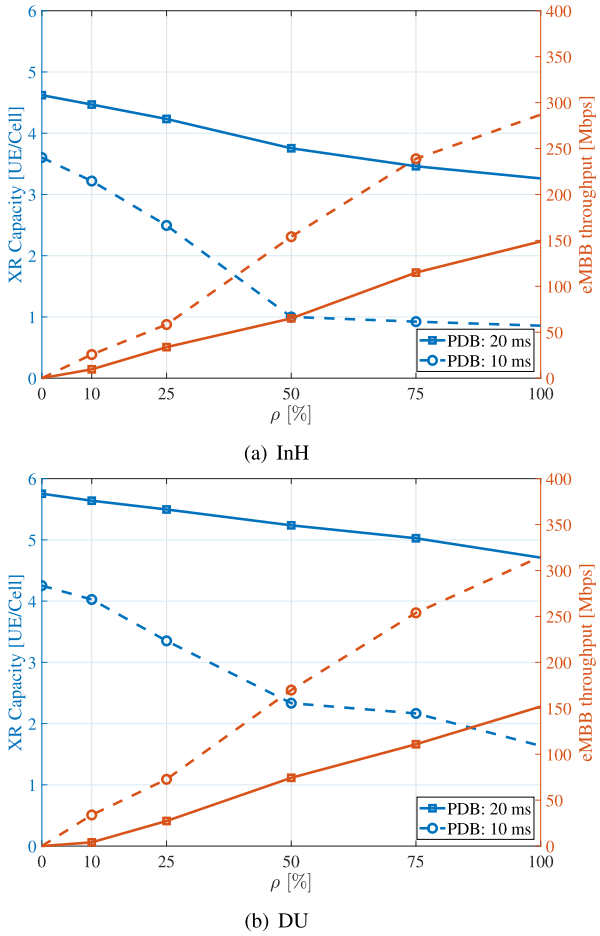


FIGURE 15. The impact of different ρ values on the XR capacity and the average eMBB cell TP.

XR capacity with PDB of 10 ms or less in the InH scenario is more sensitive than in the DU scenario.

C. THE FAIRNESS SCORE

Fairness metrics are used in wireless networks to determine whether UEs are receiving a fair share of radio resources. There are several mathematical definitions of fairness. Using well-known fairness metrics, like Jain's fairness index [33], is challenging for this scenario where one group of UEs is of the fixed data rate type (i.e., XR). So, we have defined a simple fairness score for eMBB UEs, where we assume the system is fair if eMBB UEs see a throughput decrease that fits what throughput increase is gained by added XR UEs. We consider the balance between the two UE groups and score it against the different radio resource utilization schemes introduced in the paper. The fairness score can be expressed as a function of ρ as follows:

$$\zeta(\rho) = \frac{R_{\text{eMBB-only}} - R_{\text{eMBB-mixed}}(\rho)}{R_{\text{XR}} \times C_{\text{XR}}^{\text{PDB}}(\rho)}, \tag{2}$$

where $R_{\text{eMBB-only}}$ and $R_{\text{eMBB-mixed}}(\rho)$ are the eMBB cell TP for the eMBB only and mixed traffic scenario with ρ, respectively. R_{XR} is the XR user TP. Additionally, $C_{\text{XR}}^{\text{PDB}}(\rho)$ represents the XR capacity as a function of ρ and PDB. ζ is the fairness score. A totally fair system has a score of 1 and an unfair system has scores of larger than 1. A higher score tends to result in lower fairness. For instance, $\zeta = 1$ means that adding an XR UE results in a reduction of the eMBB cell TP of 1×45 Mbps and this is the most fair system.

TABLE 3. The fairness score for different XR PDB cases in both InH and DU scenarios.

			ρ				
			0%	10%	25%	50%	75%
Deployment	InH	10 ms	2.41	2.22	2.97	3.81	-
		20 ms	1.80	1.79	1.62	1.93	1.56
	DU	10 ms	2.13	1.94	2.31	2.38	1.44
		20 ms	1.71	1.71	1.59	1.37	1.21

We summarize the results of the fairness score for different XR PDB cases and different values of ρ in both InH and DU scenarios in Table 3. As seen, fairness ζ varies depending on the environment, XR PDB, and the value of ρ. The largest ζ is for the case with the stricter PDB requirement for the InH scenario, where the ICI is most dominant. On the other hand, ζ is less when the PDB requirement is relaxed and the deployment DU with lower ICI values. The results reveal low fairness (i.e. high value of ζ) in scenarios where there is a large PRB restriction on eMBB UEs (e.g., $\rho = 10\%$). In addition, we observe that the fairness index reaches close to 1 when the network operates with ρ set to larger than 75% and a related PDB of 20 ms.

VII. CONCLUSION

In this paper we have shown the following:

- Both the InH and DU scenarios support demanding XR UEs with 45 Mbps and PDB of 10 ms. The spatial density of supported satisfied XR UEs is a factor 17 higher for the InH scenario, as compared to DU, due to its much higher cell density. The InH scenario exhibits approximately 6 dB higher ICI, making the XR performance much more sensitive to the addition of eMBB background traffic.
- With a tight XR PDB value of 10 ms, the XR capacity drops to nearly zero satisfied UEs when fully loading the system with eMBB background traffic, while few satisfied XR UEs can be supported for PDB values of 20ms, or higher.
- Restricting the PRB utilization for eMBB traffic is a powerful method for optimizing the mixed XR and eMBB capacity. The level of recommended PRB utilization restriction depends on the XR UEs' PDB constraints and environment scenarios.
- For 10 ms PDB, it is recommended to set the eMBB PRB utilization restriction to 30% ($\rho = 30\%$) in both InH and DU scenarios. For 20 ms PDB, setting ρ to 75% is recommended for DU, while a stricter ρ value of 40% is suggested for InH due to the higher ICI levels. For these

TABLE 4. Complexity analyses of Algorithm 1.

Step number	Operation	Total complexity of the step	Remarks
1	Sorting	$C_{\text{UE}} \times \mathcal{O}(\log_2(C_{\text{UE}}))$	-
4	Boolean comparison	$\mathcal{O}(n^2) \times S_{\text{PRB}}$	The condition is checked S_{PRB} times
5	Boolean comparison	$\mathcal{O}(n^2) \times C_{\text{UE}} \times S_{\text{PRB}}$	The condition is checked for each of the UEs per each PRB
6	Variable assignment	$[2n + \mathcal{O}(n^{\frac{1}{2}})] \times C_{\text{UE}}$	The assignment is done once per UE
7 & 8	Boolean comparison	$6 \times \mathcal{O}(n^2) \times S_{\text{PRB}}$	6 Boolean operations per PRB
9 & 10	Variable assignment	$2 \times [2n + \mathcal{O}(n^{\frac{1}{2}})] \times S_{\text{PRB}}$	Two variable assignments per PRB
11	Boolean comparison	0	Is calculated in step 8
12	Boolean comparison	$3 \times \mathcal{O}(n^2) \times S_{\text{PRB}}$	6 Boolean operations per PRB (worst case scenario for no XR UE in the queue)
13 & 14	Variable assignment	0	Is calculated in steps 9 & 10 (it is included when multiplying to S_{PRB})

recommended values of ρ, the network can support 5 XR UEs with 45Mbps and 20 ms PDB, plus 112 Mbps eMBB cell TP for the DU case. For InH it equals 4 XR UEs and 72 Mbps eMBB TP.

- The effect of adding XR UEs on eMBB TP performance is influenced by factors such as the XR PDB, deployment environment (InH/DU), and the value of ρ. Our findings show that the eMBB TP declines with a factor of 1 to 4 of the XR sum rate, depending on the ρ setting.

The findings of this study can help cellular operators configure their networks when introducing XR traffic and determine how many carriers are needed to accommodate a certain number of satisfied XR users and eMBB cell throughput. The reported findings also aspire to further develop enhanced interference mitigation schemes to handle the interference originating from eMBB transmissions in one cell to XR UEs in a neighboring cell. Furthermore, as includes relaxing the currently assumed full-buffer traffic model for eMBB and evolving to more realistic models (e.g., web browsing or emailing).

APPENDIX A
LIST OF ABBREVIATIONS

3GPP	3rd Generation Partnership Project.
5G	Fifth Generation.
ACK	Acknowledgment.
AR	Augmented Reality.
BLER	Block Error Rate.
CBG	Code Block Group.
CI	Confidence Interval.
CQI	Channel Quality Indicator.
CSI	Channel State Information.
DL	Downlink.
DU	Dense Urban.
eCDF	Empirical Cumulative Distribution. Function.
eMBB	Enhanced Mobile Broadband.
fps	frames per second.
gNB	Next Generation Nodeb.
HARQ	Hybrid Automatic Repeat Request.
InH	Indoor Hotspot.
ICI	Inter Cell Interference
ICIC	Inter Cell Interference Coordination.
IP	Internet Protocol.
KPI	Key Performance Indicators.
LA	Link Adaptation.
LOS	Line Of Sight
MAC	medium access control.
MMSE-IRC	Minimum Mean Square Error Interference Rejection Combining.
MCS	Modulation and Coding Scheme.
MIMO	Multiple Input Multiple Output.
MMIB	Mean Mutual Information per Bit.
MR	Mixed Reality.
MU-MIMO	Multi User Multiple Input Multiple. Output.
NACK	Negative Acknowledgment.
NLOS	Non Line Of Sight.
NR	New Radio.
OFDMA	Orthogonal Frequency division. Multiple Access.
OFDM	Orthogonal Frequency division Multiplexing.
O2I	Outdoor to Indoor.
OLLA	Outer Loop Link Adaptation.
PCDP	Packet Data Convergence Protocol.
PDB	Packet Delay Budget.
PHY	physical.
PDCCH	physical downlink control channel.
PDSCH	physical downlink shared channel.
PRB	Physical Resource Block.
QoS	Quality of Service.
RAN	Radio Access Network.
RLC	Radio Link Control.
RRM	Radio Resource Management.
RE	Resource Element.
SINR	Signal to Interference plus Noise. Ratio.
SU-MIMO	Single User Multiple Input Multiple. Output.
SDR	Source Data Rate.
SCS	Sub Carrier Spacing.

SLS	System Level Simulations.
TP	Throughput.
TB	Transport Block.
TDD	Time Division Duplex.
TTI	Transmission Time Interval.
URLLC	Ultra Reliable Low Latency. Communications.
UE	User Equipment.
VR	Virtual Reality.
XR	Extended Reality.

APPENDIX B STATISTICAL CONFIDENCE CONSIDERATIONS

The primary KPI is the XR capacity as mentioned in Section II-G. To determine the necessary number of generated packets per UE for estimating if 99% of the packets are received correctly by the UE within the PDB, careful consideration to guarantee a certain confidence interval (CI) is needed. In this regard, we use the binomial proportion CI method. The binomial proportion CI method is a way to estimate the probability of success based on the results of a series of success-failure experiments (also known as Bernoulli trials) [34]. It provides an interval estimate for the success probability, denoted as p, when we only know the number of experiments conducted (n) and the number of successful outcomes (n_s). To calculate the CI, a commonly used formula approximates the distribution of error around the observed proportion of successes ($\hat{p}$) as a normal distribution [35], where it is given by

$$\text{CI} = \hat{p} \pm z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \tag{3}$$

where $\hat{p}$ is the proportion of successes in the trials (i.e., n_s/n) and z represents the $(1-\frac{\alpha}{2})$-th percentile of a standard normal distribution for a given error level α. For example, with a confidence level of 95%, α is 0.05, so z corresponds to 1.96.

For the XR traffic evaluations in this study, we assume uncorrelated samples and a confidence level of 95% with a $\pm 1\%$ margin of error. Based on these assumptions, (3) determines that the required number of generated packets per UE is 384. To include some safety margin, we generate at least 600 uncorrelated samples for each XR UE which results in a simulation time of 10 seconds.

For each simulation, we consider N XR UEs per cell, and B cells, collecting statistics from the $B \times N$ calls. Typical values of N are between 1 and 7 depending on the XR QoS requirements. B is assumed to be 12 or 21 for InH and DU scenarios, respectively. We repeat each simulation M times, resulting in statistics from $B \times N \times M$ calls. With $B = 21$, $M = 10$, and $N = 5$, we obtain statistics from 1050 XR calls. According to (3), this is sufficient to estimate with a CI of $95\% \pm 1.8\%$ if 90% of the XR UEs are satisfied as per the XR capacity definition metric.

The calculated simulation time also allows us to collect reliable statistics for the eMBB average cell TP by monitoring the TP samples in 10 seconds for each of the 12 cells, repeating it 10 times.

APPENDIX C COMPLEXITY ANALYSES OF ALGORITHM 1

The detailed complexity analyses of Algorithm 1 are listed in Table 4, where n is the total number of bits for each of the variables and C_{UE} expresses the maximum number of UEs in the cell. More details on each step's complexity calculation can be found in [36]. The total complexity of the algorithm, i.e. sum of the complexities, is:

$$\begin{aligned}\mathcal{C}_{\text{Alg1}} = {} & C_{\text{UE}} \times \mathcal{O}(\log_2(C_{\text{UE}})) + S_{\text{PRB}} \times (10 + C_{\text{UE}}) \\ & \times \mathcal{O}(n^2) + (C_{\text{UE}} + 2S_{\text{PRB}}) \times (2n + \mathcal{O}(n^{\frac{1}{2}})).\end{aligned} \tag{4}$$

The Complexity of the legacy WRR scheduling method which has a simpler step 12 operation consisting of only a logical comparison, can be calculated similarly. As expected, the complexity is very similar to $\mathcal{C}_{\text{Alg1}}$ except for the multiplier of term $\mathcal{O}(n^2)$ which is $S_{\text{PRB}} \times (8 + C_{\text{UE}})$ for the WRR method. To give a numerical example, for the case of $C_{\text{UE}} = 6$ the ratio of dominant components, i.e. $\mathcal{O}(n^2)$, of complexity between the WRR and Algorithm 1 is $\frac{S_{\text{PRB}} \times (10+C_{\text{UE}})}{S_{\text{PRB}} \times (8+C_{\text{UE}})} = \frac{16}{14} \approx 1.14$.

REFERENCES

[1] *Extended Reality (XR) in 5G (Release 17)*, document 26.928, V17.0.0, 3GPP, Mar. 2022.

[2] D. Zhang, J. J. P. C. Rodrigues, Y. Zhai, and T. Sato, "Design and implementation of 5G e-health systems: Technologies, use cases, and future challenges," *IEEE Commun. Mag.*, vol. 59, no. 9, pp. 80–85, Sep. 2021.

[3] *Study on XR Enhancements for NR (Release 18)*, document 38.835, V1.0.1, 3GPP, Feb. 2023.

[4] *Study on XR (Extended Reality) Evaluations for NR (Release 17)*, document 38.838, V17.0.0, 3GPP, Dec. 2021.

[5] *XR Enhancements for NR*, document RP-223502, 3GPP, Dec. 2022.

[6] E. Chen, S. Dou, S. Wang, Y. Cao, and S. Liao, "Frame-level integrated transmission for extended reality over 5G and beyond," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2021, pp. 1–6.

[7] M. Huang and X. Zhang, "MAC scheduling for multiuser wireless virtual reality in 5G MIMO-OFDM systems," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[8] P. Paymard, A. Amiri, T. E. Kolding, and K. I. Pedersen, "Extended reality over 3GPP 5G-advanced new radio: Link adaptation enhancements," 2022, *arXiv:2210.14578*.

[9] P. Paymard, A. Amiri, T. E. Kolding, and K. I. Pedersen, "Enhanced link adaptation for extended reality code block group based HARQ transmissions," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2022, pp. 711–716.

[10] M. Gapeyenko, V. Petrov, S. Paris, A. Marcano, and K. I. Pedersen, "Standardization of extended reality (XR) over 5G and 5G-advanced 3GPP new radio," 2022, *arXiv:2203.02242*.

[11] S. Dou, S. Liao, J. Wu, K. Wu, E. Chen, W. Chen, H. Shen, and N. Li, "XR quality index: Evaluating RAN transmission quality for XR services over 5G and beyond," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2021, pp. 1–6.

[12] J. K. Sundararajan, H.-J. Kwon, O. Awoniyi-Oteri, Y. Kim, C.-P. Li, J. Damnjanovic, S. Zhou, R. Ma, Y. Tokgoz, P. Hande, T. Luo, K. Mukkavilli, and T. Ji, "Performance evaluation of extended reality applications in 5G NR system," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2021, pp. 1–7.

[13] P. Paymard, A. Amiri, T. E. Kolding, and K. I. Pedersen, "Performance of joint XR and best-effort eMBB traffic in 5G-advanced networks," in *Proc. IEEE 97th Veh. Technol. Conf. (VTC-Spring)*, Jun. 2023, pp. 1–5.

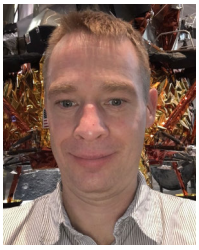
[14] M. Richart, J. Baliosian, J. Serrat, and J.-L. Gorricho, "Resource slicing in virtual wireless networks: A survey," *IEEE Trans. Netw. Service Manag.*, vol. 13, no. 3, pp. 462–476, Sep. 2016.

[15] B. Soret and K. I. Pedersen, "On-demand power boost and cell muting for high reliability and low latency in 5G," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Jun. 2017, pp. 1–5.
[16] A. A. Ateya, A. Muthanna, M. Makolkina, and A. Koucheryavy, "Study of 5G services standardization: Specifications and requirements," in *Proc. 10th Int. Congr. Ultra Modern Telecommun. Control Syst. Workshops (ICUMT)*, Nov. 2018, pp. 1–6.
[17] W. Li, L. Tian, J. Zhang, and Y. Cheng, "Analysis of base station deployment impact on LOS probability model for 5G indoor scenario," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Oct. 2017, pp. 1–5.
[18] J. Lorincz, Z. Klarin, and D. Begusic, "Modeling and analysis of data and coverage energy efficiency for different demographic areas in 5G networks," *IEEE Syst. J.*, vol. 16, no. 1, pp. 1056–1067, Mar. 2022.
[19] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A survey on 5G usage scenarios and traffic models," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 905–929, 2nd Quart., 2020.
[20] G. Pocovi, K. I. Pedersen, and P. Mogensen, "Joint link adaptation and scheduling for 5G ultra-reliable low-latency communications," *IEEE Access*, vol. 6, pp. 28912–28922, 2018.
[21] G. Pocovi, T. Kolding, and K. I. Pedersen, "On the cost of achieving downlink ultra-reliable low-latency communications in 5G networks," *IEEE Access*, vol. 10, pp. 29506–29513, 2022.
[22] A. K. Bairagi, Md. S. Munir, M. Alsenwi, N. H. Tran, S. S. Alshamrani, M. Masud, Z. Han, and C. S. Hong, "Coexistence mechanism between eMBB and uRLLC in 5G wireless networks," *IEEE Trans. Commun.*, vol. 69, no. 3, pp. 1736–1749, Mar. 2021.
[23] M. Alsenwi, N. H. Tran, M. Bennis, S. R. Pandey, A. K. Bairagi, and C. S. Hong, "Intelligent resource slicing for eMBB and URLLC coexistence in 5G and beyond: A deep reinforcement learning based approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4585–4600, Jul. 2021.
[24] K. I. Pedersen, G. Pocovi, and J. Steiner, "Preemptive scheduling of latency critical traffic and its impact on mobile broadband performance," in *Proc. IEEE 87th Veh. Technol. Conf.*, Jun. 2018, pp. 1–6.
[25] *Study on Channel Model for Frequencies From 0.5 to 100 GHz*, document 38.901, V17.0.0, 3GPP, Mar. 2022.
[26] I. Shayea, M. Ergen, M. Hadri Azmi, S. Aldirmaz Çolak, R. Nordin, and Y. I. Daradkeh, "Key challenges, drivers and solutions for mobility management in 5G networks: A survey," *IEEE Access*, vol. 8, pp. 172534–172552, 2020.
[27] S. Sun, T. S. Rappaport, T. A. Thomas, A. Ghosh, H. C. Nguyen, I. Z. Kovács, I. Rodriguez, O. Koymen, and A. Partyka, "Investigation of prediction accuracy, sensitivity, and parameter stability of large-scale propagation path loss models for 5G wireless communications," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 2843–2860, May 2016.
[28] I. Z. Kovacs, K. I. Pedersen, T. E. Kolding, A. Pokhariyal, and M. Kuusela, "Effects of non-ideal channel feedback on dual-stream MIMO-OFDMA system performance," in *Proc. IEEE 66th Veh. Technol. Conf.*, Sep. 2007, pp. 1852–1856.
[29] H. M. Chaskar and U. Madhow, "Fair scheduling with tunable latency: A round-robin approach," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 592–601, Aug. 2003.
[30] M. Lampinen, F. Del Carpio, T. Kuosmanen, T. Koivisto, and M. Enescu, "System-level modeling and evaluation of interference suppression receivers in LTE system," in *Proc. IEEE 75th Veh. Technol. Conf. (VTC Spring)*, May 2012, pp. 1–5.
[31] K. Pietikainen, F. Del Carpio, H.-L. Maattanen, M. Lampinen, T. Koivisto, and M. Enescu, "System-level performance of interference suppression receivers in LTE system," in *Proc. IEEE 75th Veh. Technol. Conf. (VTC Spring)*, May 2012, pp. 1–5.
[32] T. Tang, R. Doostnejad, and T. J. Lim, "Mean mutual information per coded bit based precoding in MIMO-OFDM systems," in *Proc. IEEE 72nd Veh. Technol. Conf.*, Sep. 2010, pp. 1–5.
[33] R. K. Jain, D. M. W. Chiu, and W. R. Hawe, "A quantitative measure of fairness and discrimination," Eastern Res. Lab., Digit. Equip. Corp., Hudson, MA, USA, Tech. Rep., 21, 1984.
[34] S. Wallis, "Binomial confidence intervals and contingency tests: Mathematical fundamentals and the evaluation of alternative methods," *J. Quant. Linguistics*, vol. 20, no. 3, pp. 178–208, Aug. 2013, doi: 10.1080/09296174.2013.799918.
[35] L. D. Brown, T. T. Cai, and A. DasGupta, "Confidence intervals for a binomial proportion and asymptotic expansions," *Ann. Statist.*, vol. 30, no. 1, pp. 160–201, Feb. 2002.
[36] I. Wegener, *The Complexity of Boolean Functions*. Hoboken, NJ, USA: Wiley, 1987.

POURIA PAYMARD (Graduate Student Member, IEEE) received the B.Sc. degree in electrical engineering from Shiraz University, Shiraz, Iran, in 2016, and the M.Sc. degree in communication systems from Tarbiat Modares University, Tehran, Iran, in 2019. He is currently pursuing the Ph.D. degree with the Department of Electronic Systems, Aalborg University, in collaboration with Nokia Standard, Aalborg, Denmark. His research interests include network optimization, eXtended reality communications, time-sensitive communications, and 5G/6G radio resource management.

ABOLFAZL AMIRI received the M.S. degree in electrical and communication systems engineering from the University of Tehran, Tehran, Iran, in 2016, and the Ph.D. degree from the Department of Electronic Systems, The Technical Faculty of IT and Design, Aalborg University, Denmark, in 2022. From 2016 to 2017, he was with Huawei as an RF Engineer and a Cellular Network Optimizer. He was a Research Assistant on multi-antenna signal processing with Aalborg University, from 2017 to 2018. He is currently a Senior Standardization Research Specialist with Nokia Standard. His research interests include applications of signal processing and machine learning in wireless communications, radio resource management, and multi-antenna systems.

TROELS E. KOLDING received the M.Sc. and Ph.D. degrees from Aalborg University, Denmark, in 1996 and 2000, respectively. His M.Sc. was achieved in collaboration with the Wireless Information Network Laboratory (WINLAB), New Brunswick, NJ, USA. Since joining Nokia Standards, in 2001, he has been active in research and management for standardization, network architecture, and portfolio management. He is currently a Distinguished Member of Technical Staff. He is the author of more than 100 scientific publications. He holds more than 50 granted U.S. patents. His current research interests include 6G radio architecture, cloud RAN, the industrial IoT, time-sensitive communications, time-synchronization, and radio resource management.



KLAUS I. PEDERSEN (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from Aalborg University, Aalborg, Denmark, in 1996 and 2000, respectively. He is currently a Bell Labs Fellow with Nokia Standards, leading the Radio Access Research Team, Aalborg, and a part-time External Professor with Aalborg University. He has authored publications on a wide range of topics, as well as an inventor on several patents. His current research interests include access protocols and radio resource management enhancements for 5G new radio and its evolution to 5G-advanced.