Enhancing URLLC Uplink Configured-grant Transmissions

Trung-Kien Le†, Umer Salim*, Florian Kaltenberger†
†EURECOM, Sophia-Antipolis, France
*TCL Mobile, Sophia-Antipolis, France
Trung-Kien.Le@eurecom.fr

***Abstract*—The 3rd Generation Partnership Project (3GPP) has defined Ultra-Reliable Low-Latency Communication (URLLC) as one of the main objectives of 5G development to satisfy the applications with stringent requirements of latency and reliability. Uplink (UL) configured-grant (CG) transmission where the user equipment (UE) transmits a packet without scheduling request (SR) and UL grant is standardized by 3GPP Release 15 to reduce latency. The UE also transmits automatically a configured number of repetitions without feedback from the base station (gNB) to increase reliability.**

Nevertheless, the repetitions are not allowed to transmit outside the hybrid automatic repeat request (HARQ) process containing the first repetition. It might cause a smaller number of transmitted repetitions than configuration that is harmful to the performance of URLLC.

This paper uses the reserved resources where the UEs can transmit the repetitions outside the original HARQ process until the configured number is reached. The scheme is developed further when the gNB is equipped with a successive interference cancellation (SIC) receiver so it can decode multiple repetitions of the different UEs in the same reserved resource. The use of reserved resources ensures the performance of UL CG transmission while SIC receiver minimizes the reserved resource's consumption. The numerical results show a higher transmission reliability and lower reserved resource consumption of the proposed scheme compared to the related works.

***Index Terms*—5G, URLLC, uplink configured-grant transmission, K repetitions, reserved resources, SIC receiver**

I. Introduction

The development of technology paves the way for the emergence of new applications such as virtual reality, autonomous vehicles, real time remote surgery, to name but a few. To cover such wide variety of use cases with diverse requirements, the 3rd Generation Partnership Project (3GPP) defined three service paradigms for 5G: Enhanced Mobile Broadband (eMBB), Massive Machine-Type Communication (mMTC) and Ultra-Reliable Low-Latency Communication (URLLC). Among these three service categories, URLLC raises the most challenge because it has to deal with two conflicting factors at the same time: reliability and latency.

For 3GPP Release 15, which is the first release of 5G New Radio (NR), the URLLC requirements are specified in [1]: "A general URLLC reliability requirement for one transmission of a packet is 10^{-5} for 32 bytes with a user plane latency of 1 ms". In Release 16, the more stringent requirements are targeted: "Higher reliability (up to 10^{-6}), higher availability, short latency in the order of 0.5 to 1 ms, depending on the use cases (factory automation, transport industry and electrical power distribution)" [2].

In order to achieve URLLC requirements, some principal techniques are specified in 3GPP Release 15 and Release 16. Firstly, flexible sub-carrier spacing (SCS) of 15 kHz, 30 kHz, 60 kHz, 120 kHz and 240 kHz is applied instead of only 15 kHz as long-term evolution (LTE) [3] that results in shorter symbols and slots.

Second, mini-slot based transmission is supported so that a packet can be scheduled to be transmitted in downlink (DL) or uplink (UL) in the interval of one or several symbols instead of the whole slot as LTE [4].

Third, the user equipment (UE) can be configured with the periodic transmission resources by the base station (gNB) so it can transmit data in these resources without the presence of scheduling request (SR) and UL grant [5]. This is called the UL configured-grant (CG) transmission.

Fourth, the automatic repetition in UL CG transmission is allowed so that the UE can transmit the repetitions in the consecutive transmission occasions without waiting hybrid automatic repeat request (HARQ) feedback [5]. The UE is also permitted to transmit the repetitions in the consecutive transmission occasions in mini-slot level instead of the consecutive slots [6].

This paper focuses on the enhancements in design of UL CG transmission to meet URLLC requirements. The problem of UL CG transmission with a configured number of repetitions and prior art to solve this problem are explained in Section II. Section III introduces the proposed scheme of reserved resources allocated to the UE and a successive interference cancellation (SIC) receiver in the gNB. The novelty in the proposed scheme is the procedure to optimize the sizes of reserved resources so that resource consumption is minimized while CG transmission's performance is guarantee. Section IV shows the numerical results of the proposed scheme in comparison to the prior art. Section V concludes this paper.

II. Automatic repetitions in UL CG transmissions

A. K-repetition problem

In the CG resources, the UEs can transmit automatically K repetitions (1, 2, 4 or 8 repetitions) without waiting feedback from the gNB as configured by parameter $repK$ from higher layer. However, the UEs are only allowed to carry out repetitions of a packet in an interval with periodicity P from several symbols to several slots [7] to avoid a confusion of HARQ identities (IDs) of the repetitions in different HARQ process at the gNB. If the gNB misses the first transmission then the second and third transmissions have different IDs, the gNB will not know the ID of the original transmission to send an UL grant for a retransmission and recognize the order of the repetitions to do soft combining.

Therefore, in case the UE reaches the boundary of an interval, it must stop the repetitions of the corresponding packet starting in that interval even if the maximum number of repetitions have not been attained. As a result, the number of repetitions transmitted might be smaller than the configured number K as shown in Fig. 1 where K is configured to 4. An interval of a HARQ process with time length P contains 4 CG occasions. For the first packet, it arrives at the beginning of an interval and before all 4 CG occasions so the UE can carry out 4 repetitions as configured. Nevertheless, the second and third packet comes after the first CG occasion, thus, the UE is only able to do 3 and 2 repetitions, respectively, instead of 4 repetitions as configured.

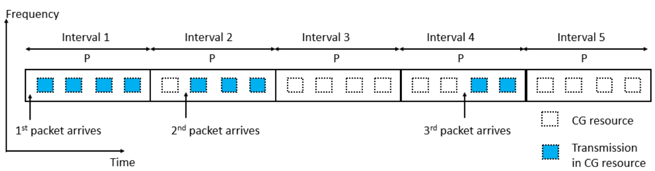

Fig. 1. Less than K repetitions in CG UL transmission.

Reliability of UL CG transmission decreases when the number of actual repetitions are smaller than the configured number. This situation also increases latency of the transmission because the gNB would need to reschedule the packet and waits for the next round of retransmission to decode the packet.

B. Related works

In 3GPP Release 15, the UE only can wait until the next interval to transmit all K repetitions if data arrives lately. The waiting time might be large if SCS is small or the packet arrives only after some CG occasions.

In [8], 3GPP agreed that multiple configurations are used to enhance reliability and reduce latency including ensuring K repetitions. The UE can choose the configuration with the closest starting point to transmit all K repetitions as shown in Fig. 2. Two drawbacks of this scheme are overhead of signal to schedule multiple configurations and resource consumption of multiple configurations.

In [9] and [10], the UE is able to transmit the repetitions across the consecutive HARQ intervals. It requires lots of effort in standardization to avoid the confusion between HARQ IDs at the gNB such as a mechanism to communicate HARQ IDs to the gNB or different DMRS sequences in the repetitions.

In [11], [12] and [13], the UEs transmits the repetitions in the shared resource. However, the constraint of HARQ

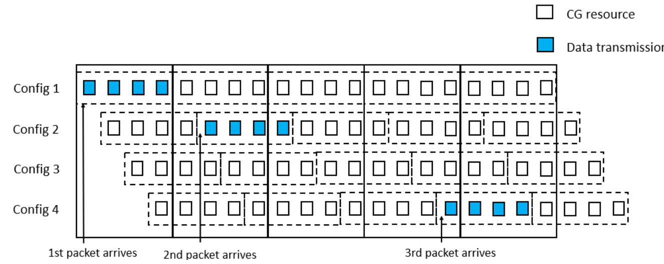

Fig. 2. Multiple configurations to ensure K repetitions.

process boundary is not considered. Moreover, the size of all resources for repetitions are the same and not optimized for each location in a HARQ interval. Both factors cause a degradation of reliability, latency and resource consumption.

In [14], reserved resources is used for repetitions outside the HARQ process of the first transmission. The size of reserved resources is optimized based on the their location but the gNB has no capacity to decode the repetitions in case of collision between several UE's repetitions in the reserved resources.

III. Ensuring Latency and Reliability in UL CG transmissions with reserved resources for the UEs and a SIC receiver at the gNB

A. Reserved resources

The configured number of repetitions can be guaranteed by generating some transmission resources which are dedicated for the CG transmissions in case the repetitions can not be completed within the CG resources. The reserved resources are configured with the same period as CG occasions. These reserved resources are shared among the UEs following random access, resulting in lower overhead of resource creation. If the UE reaches the boundary of a HARQ process while not carrying out K repetitions, it will use the reserved resources configured in the next transmission interval (reserved for a different HARQ process) to continue to transmit until attaining K repetitions as shown in Fig. 3.

The optimal sizes of reserved resources are calculated in the next sections in order to minimize resource consumption.

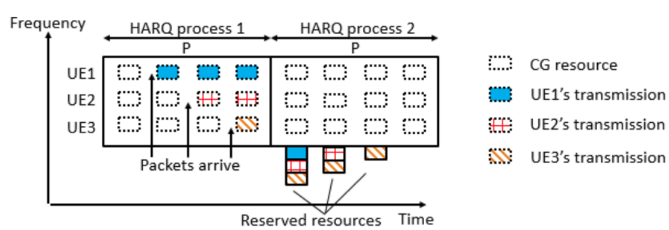

Fig. 3. Reserved resources for repetitions.

B. System model

In Fig. 4, N UEs are configured to transmit K repetitions in the shared consecutive CG resources. A HARQ process with time interval P contains K CG transmission occasions. The results derived below are also valid if the number of

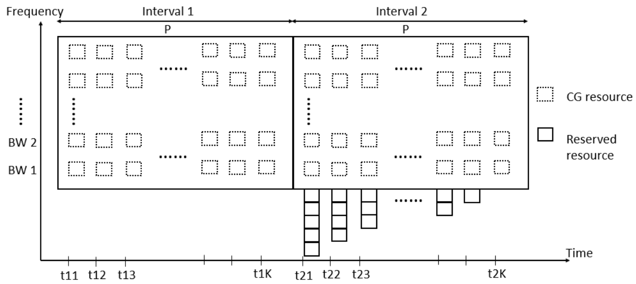

Fig. 4. UL transmission resources' distribution.

CG transmission occasions in an interval P are bigger than K. $K-1$ reserved resources are configured with the same period as the CG resources in each period P. All N UEs can use the reserved resources to attain the configured number of repetitions. The reserved resource at the ith transmission occasion in a period has M_i blocks where each block has the size of CG resource in one transmission occasion. Packet arrival follows a Poisson process. In an interval of T between two consecutive CG resources, there are λ packets arriving at a UE.

The gNB is equipped with a SIC receiver to decode the repetitions in the reserved resources so even if there is a collision between the repetitions of the UEs in the system, the gNB still can decode correctly the repetitions. When the gNB decodes correctly a packet in the CG resources or the previous reserved resources, it stores that packet. After that, if the gNB encounters a collision between the successful packet and another packet in the reserved resources, it can cancel the successful packet from the received signal to remove the interference. Thereby, the gNB decodes the other packet without interference and has higher successful probability. With big number of repetitions (for example, 4 or 8 repetitions), there is low probability that the gNB has a collision among all non-decoded packets so the SIC receiver is useful in improving performance of repetitions in the reserved resources.

The successful probability of a packet in the transmission with SIC receiver in the gNB is complex to calculate. It depends on the channel condition of other UEs, the successful probability of other packets competing for the resources and the time arrival of data. Therefore, a model of SIC receiver in physical layer called K-multipacket reception (K-MPR) is used as in [15]. In this model, the gNB is assumed to be able to decode correctly all the UE transmissions in the same block of reserved resources if the number of the UEs in that block are smaller than a threshold L. On the contrary, if the number of the collided UEs are bigger than L, all the packets in the collided resources cannot be decoded.

C. Error probability due to collisions in reserved resources for UL CG repetitions

The collision probability of a UE of interest and $N-1$ UEs in the reserved resource at the first transmission occasion of a period (at t21 in Fig. 4) is calculated below. The calculations only focus on the error due to collision and do not count the radio errors.

The probability of UE transmission in an interval of T is

$$P_{data} = 1 - e^{-\lambda}, \tag{1}$$

A UE transmits in the reserved resource at t21 if a packet arrives after the first CG transmission occasion at t11. A UE transmits after the first CG transmission occasion in a period P with a probability

$$P_{d1} = (1 - P_{data})(1 - (1 - P_{data})^{K-1}). \tag{2}$$

The probability that no UE out of $N-1$ UEs transmits after the first CG occasion is

$$P_0 = (1 - P_{d1})^{N-1}. \tag{3}$$

The probability that n UEs out of $N-1$ UEs transmit after the first CG occasion is

$$P_{n1} = \binom{N-1}{n} P_{d1}^{n} (1 - P_{d1})^{N-1-n}. \tag{4}$$

The probability that l UEs in these n UEs access the same block in the first reserved resource as the UE of interest is

$$P_{al\_n} = \binom{n}{l} \left(\frac{1}{M_1}\right)^{l} \left(\frac{M_1 - 1}{M_1}\right)^{n-l}. \tag{5}$$

If the value of l is smaller than $L-1$, the gNB still can decode all the packets in that block. The probability that the UE of interest collides with any other UEs ($l < L-1$) at t21 but the packet is still decodable is

$$\begin{aligned} P_s &= \sum_{n=1}^{N-1} P_{n1} \sum_{l=0}^{L-1} P_{al\_n} \\ &= \sum_{n=1}^{N-1} P_{n1} \sum_{l=0}^{L-1} \binom{n}{l} \left(\frac{1}{M_1}\right)^{l} \left(\frac{M_1 - 1}{M_1}\right)^{n-l}. \end{aligned} \tag{6}$$

From (3) and (6), the error probability of a packet from the UE of interest due to collision in the first reserved resource is

$$\begin{aligned} P_{col\_SIC\_1} &= 1 - P_0 - P_s \\ &= 1 - (1 - P_{d1})^{N-1} - \\ &- \sum_{n=1}^{N-1} P_{n1} \sum_{l=0}^{L-1} \binom{n}{l} \left(\frac{1}{M_1}\right)^{l} \left(\frac{M_1 - 1}{M_1}\right)^{n-l}. \end{aligned} \tag{7}$$

The model in [14] provides a case of (7) as shown below where L is 1 meaning that the packet cannot be decoded if there is a collision

$$\begin{aligned} P_{c1} &= 1 - P_0 - P_s \\ &= 1 - \left(\frac{M_1 - e^{-\lambda} + e^{-K\lambda}}{M_1}\right)^{N-1}. \end{aligned} \tag{8}$$

Similarly, collision probability for the reserved resource at any transmission occasion in a period can be derived as

$$P_{di} = (1 - P_{data})^{i}(1 - (1 - P_{data})^{K-i}). \tag{9}$$

$$P_{ni} = \binom{N-1}{n} P_{di}^{n} (1 - P_{di})^{N-1-n}. \tag{10}$$

$$P_{col\_SIC\_i} = 1 - (1 - P_{di})^{N-1} - \\ - \sum_{n=1}^{N-1} P_{ni} \sum_{l=0}^{L-1} \binom{n}{l} \left(\frac{1}{M_i}\right)^l \left(\frac{M_i - 1}{M_i}\right)^{n-l}. \tag{11}$$

where $i \in [1, K-1]$ is index indicating the position of the reserved resource based on the position of transmission occasion in a period.

Based on (11), the optimal size of reserved resources can be calculated with a target collision probability. The presence of SIC receiver reduces resource consumption of the reserved resources and makes the system support more UEs with higher data arrival rate than the model in [14].

D. Explicit HARQ-ACK feedback

In UL CG transmission, the mechanism used to terminate a transmission is time-based structure. The UEs carry out the repetitions automatically as configured and data is considered to be transmitted successfully after a certain timer configured by a parameter $ConfiguredGrantTimer$ expires. With this mechanism, there might be a waste of time and frequency resources for the unnecessary retransmissions. Moreover, if the UEs need to use the reserved resources to attain the maximum number of repetitions, the redundant data transmitted might cause a collision with data of other UEs that really need to be transmitted in the reserved resource to achieve reliability.

Therefore, when the UEs transmit less than the configured repetitions in a period and need to use the reserved resources, an explicit HARQ-ACK feedback is expected from the gNB to prevent the UEs from doing the unnecessary transmissions in the reserved resources.

IV. Numerical results

In Fig. 5, the simulation is done to compare the resource consumption of the proposed scheme and prior art that are used to guarantee the configured number of repetitions. The set of the parameters in the first scenario is: $N = 28, M_1 = 10, K = 4, \lambda = 1.25 \times 10^{-4}, P_{c1} = P_{c2} = P_{c3} = 10^{-3}$.

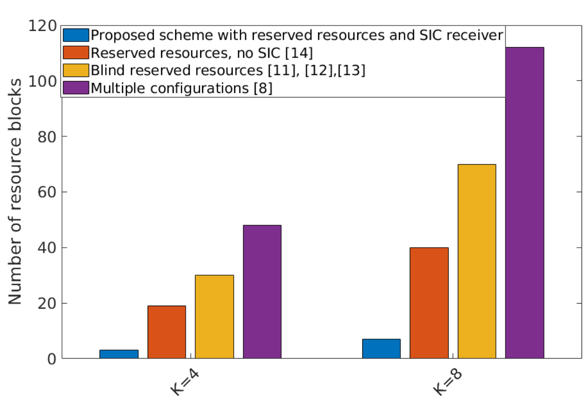

Fig. 5. Comparison of resource consumption in different schemes.

The proposed scheme also consumes much less resources than the scheme of multiple configurations in [8]. As shown in Fig. 2, if 4 repetitions are configured by the gNB, 4 configurations must be configured to ensure that the UEs always can transmit at the beginning of a period and reach 4 repetitions as configured. For a group of the UEs sharing the CG resources, 4 configurations are needed. Each configuration consists of 4 CG resources in one period. Thus, one group of the UEs requires $4 \times 4 = 16$ resource blocks in a period. There are 4 groups of the UEs so in total, $16 \times 4 = 64$ resource blocks are demanded in a period. While the scheme with reserved resources and SIC receiver only requires 16 CG resource blocks and 3 reserved resource blocks in a period that are 19 resource blocks in total. Reserved (additional) resource consumption decreases by $(1 - 3/48) \times 100\% = 93.75\%$ and total resource consumption decreases by $(1 - 19/64) \times 100\% = 70.31\%$.

One more factor taken into account when multiple configurations is applied is an increase of DMRS port. The distinction of configurations at the gNB is based on DMRS detection. Each UE transmits a specific DMRS sequence when using a configuration. Therefore, if 4 configurations are used, the number of orthogonal DMRS ports required are 4 instead of one port in single configuration with reserved resources.

The proposed scheme also consumes 84.21% and 90% less reserved resources than the scheme in [14] and the scheme in [11], [12] and [13], respectively.

The second scenario considered in Fig. 5 has a higher configured number of repetitions where the UEs are configured to transmit 8 repetitions. The set of the parameters is: $M_1 = 10, K = 8, \lambda = 1.25 \times 10^{-4}$ and $P_{c1} = 10^{-3}$. The result also shows a significant decrease of resource consumption of the proposed scheme compared to prior art.

Fig. 6 shows the error probability due to collision in the first reserved resource ($P_{col\_SIC\_1}$ in (7)) in terms of the average number of random access events λ in an interval of T between two consecutive CG resources in licensed spectrum. When the gNB can decode more packets in the same block (L increases), the system can support much higher data rates while still achieving the same target reliability of 10^{-3} due to packet collision. In [14], with $L = 1$, the system only can support $\lambda = 1.25 \times 10^{-4}$. But with $L = 2$ or $L = 3$ in the proposed scheme, the system can support $\lambda = 5.8 \times 10^{-3}$ or $\lambda = 2.53 \times 10^{-2}$, respectively.

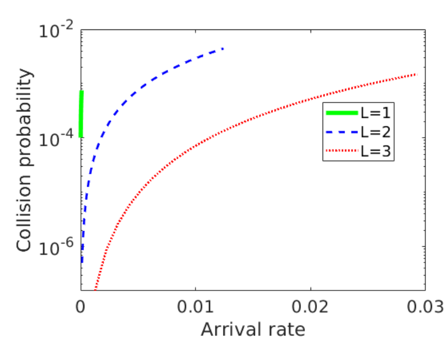

Fig. 6. The arrival rate vs collision probability.

TABLE I
PERFORMANCE COMPARISON OF DIFFERENT SCHEMES AT $SNR = -3.9dB$

Case	**Scheme**	**Starting time offset (ms)**	**Number of repetitions**	**Error probability**
Packet comes between the 1 st CG occasion and the 2 nd CG occasion	Conventional transmission	0	3	$10^{-4.5}$
	Conventional transmission with the UE waiting the next period	0.75	1	$10^{-1.5}$
	Transmission with reserved resources	0	4	10^{-6}
Packet comes between the 2 nd CG occasion and the 3 rd CG occasion	Conventional transmission	0	2	10^{-3}
	Conventional transmission with the UE waiting the next period	0.5	2	10^{-3}
	Transmission with reserved resources	0	4	10^{-6}
Packet comes between the 3 rd CG occasion and the 4 th CG occasion	Conventional transmission	0	1	$10^{-1.5}$
	Conventional transmission with the UE waiting the next period	0.25	3	$10^{-4.5}$
	Transmission with reserved resources	0	4	10^{-6}

TABLE II
SIMULATION PARAMETERS

Parameters	**Values**
Waveform	CP-OFDM
Subcarrier spacing	60kHz
Channel model	Additive white Gaussian noise (AWGN)
Channel coding	Low-density parity-check (LDPC) code
TB length	160 bits
Number of repetitions/TB	4
Number of repetitions/slot	1
MCS Index	1

Table I shows the reliability of an UL CG transmission depending on whether the configured number of repetitions is ensured or not in the simulations done with the parameters in Table II. As can be seen, due to the constraint of HARQ process boundary, the UE might not transmit all 4 repetitions as configured if data comes late that leads to an increase of packet loss. The utilization of reserved resources always guarantees the number of repetitions so the target reliability is ensured.

V. CONCLUSION

This paper introduces a scheme where reserved resources are configured to the UE and a SIC receiver is equipped at the gNB to resolve the collision of repetitions of the different UEs in the same reserved resources so that the configured number of repetitions in UL CG transmission is guaranteed. The scheme optimizes an amount of resources to be reserved while assuring the reliability of URLLC transmission.

ACKNOWLEDGMENTS

The work has been partially supported by TCL and the H2020 project 5GENESIS (5genesis.eu).

REFERENCES

[1] 3GPP TR 38.913 v15.0.0, "Study on scenarios and requirements for next generation access technologies."

[2] Huawei, HiSilicon, Nokia, Nokia Shanghai Bell, "New SID on Physical Layer Enhancements for NR URLLC". 3GPP RP-182089, TSG-RAN#81, Gold Coast, Australia, Sept 10–13, 2018.

[3] 3GPP TS 38.211 v16.0.0, "Physical channels and modulation."

[4] 3GPP TR 38.802 v14.2.0, "Study on new radio access technology physical layer aspects."

[5] 3GPP TS 38.214 v16.0.0, "Physical layer procedures for data."

[6] 3GPP TS 38.213 v16.0.0, "Physical layer procedures for control."

[7] 3GPP TS 38.331 v15.3.0, "Radio Resource Control (RRC) protocol specification."

[8] Chairman's notes RAN1 #95, Spokane, USA, Nov 12–16, 2018.

[9] Ericsson, "Enhancement of Configured Grant for NR URLLC", 3GPP R1-1812162, RAN1#95, Spokane, USA, November 12–16, 2018.

[10] Huawei, HiSilicon, "Enhanced UL configured grant transmissions", 3GPP R1-1812226, RAN1#95, Spokane, USA, November 12–16, 2018.

[11] R. breu, G. Berardinelli, T. Jacobsen, K. Pedersen and P. Mogensen, "A Blind Retransmission Scheme for Ultra-Reliable and Low Latency Communications", 2018 IEEE 87th Vehicular Technology Conference (VTC Spring), June 2018.

[12] Z. Zhou, R. Ratasuk, N. Mangalvedhe and A. Ghosh, "Resource Allocation for Uplink Grant-Free Ultra-Reliable and Low Latency Communications", 2018 IEEE 87th Vehicular Technology Conference (VTC Spring), June 2018.

[13] N. H. Mahmood, R. Abreu, R. Böhnke, M. Schubert, G. Berardinelli and T. H. Jacobsen, "Uplink Grant-Free Access Solutions for URLLC services in 5G New Radio," 2019 16th International Symposium on Wireless Communication Systems (ISWCS), Oulu, Finland, 2019.

[14] Trung-Kien Le, Umer Salim, Florian Kaltenberger, "Optimal reserved resources to ensure the repetitions in Ultra-Reliable Low-Latency Communication Uplink Grant-free transmission", 2019 European Conference on Networks and Communications (EuCNC), June 2019.

[15] C. Stefanovic, E. Paolini, and G. Liva, "Asymptotic performance of coded slotted aloha with multipacket reception," IEEE Communications Letters, vol. 22, no. 1, pp. 105–108, 2018.

[16] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," in IEEE Journal on Selected Areas in Communications, vol. 18, no. 3, pp. 535-547, March 2000.

[17] 3GPP TS 37.213 v16.0.0, "Physical layer procedures for shared spectrum channel access."

[18] Trung-Kien Le, Umer Salim, Florian Kaltenberger, "Strategies to meet the configured repetitions in URLLC Uplink Grant-Free transmission," 2019 16th International Symposium on Wireless Communication Systems (ISWCS), Oulu, Finland, August 2019.

[19] Trung-Kien Le, Umer Salim, Florian Kaltenberger, "An Overview of Physical Layer Design for Ultra-Reliable Low-Latency Communications in 3GPP Releases 15, 16, and 17," in IEEE Access, vol. 9, pp. 433-444, 2021.