Improving Ultra-Reliable Low-Latency Communication in multiplexing with Enhanced Mobile Broadband in grant-free resources

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

Abstract—In uplink (UL) transmission, the Ultra-Reliable Low-Latency Communication (URLLC) users (UEs) might be assigned the grant-free (GF)/configured-grant (CG) periodic resources to transmit data straight-away instead of sending scheduling request (SR) and receiving UL grant. However, when these resources are not in use by the URLLC traffic, the base station (called gNB) can dynamically schedule the Enhanced Mobile Broadband (eMBB) UEs to transmit in the GF resources to increase the resource efficiency. This may lead to potential collision and detrimental QoS for URLLC as some of the URLLC UEs may become active and try to use the same resource assigned to an eMBB UE. In this paper, a two-step strategy containing an overlap indication and explicit Hybrid automatic repeat request (HARQ) feedback is proposed to improve URLLC performance in multiplexing with eMBB. Besides the explicit HARQ feedback structure, a scheme with an additional SR is also presented. Simulation results show that these two schemes help achieving URLLC requirements by reducing error probability due to Demodulation Reference Signal (DMRS) miss-detection while allowing better resource efficiency.

Index Terms—5G, URLLC, uplink scheduling scheme, eMBB and URLLC multiplexing, grant-free resources

I. INTRODUCTION

One main target of the 5G New Radio (NR) standard is to support URLLC for devices requiring low latency and high link reliability. In [1], The 3rd Generation Partnership Project (3GPP) defines targets for the URLLC scenario: “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”. The next release of 3GPP will have higher requirements of URLLC: “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].

A. Techniques accepted in 3GPP Release 15

3GPP Release 15 specified new features in physical layer design to help URLLC achieve its requirements.

Subcarrier spacings (SCS) has a flexible range from 15 kHz to 240 kHz in 5G that results in very short symbol and slot timings [3]. In addition, the transmission is scheduled in mini-slot level in both UL and downlink (DL) [4]. Due to these features, the network becomes very reactive to UEs’ UL and DL traffic demands and the response time to accommodate DL or UL traffic is very small compared to Long-Term Evolution (LTE) and LTE-Advanced.

Another feature is the standardization of the GF/CG UL transmissions which allows the UEs to transmit data in UL without having to transmit an explicit SR and receiving an UL grant to reduce latency [5].

B. Problem of multiplexing URLLC and eMBB in GF resources

In 5G, as mentioned in Section I-A, the gNB can schedule GF resources for URLLC UEs but it does not have any prior information which of these GF resources will actually be used by URLLC UEs or which of the UEs in the group configured to the resources will use a specific resource. If the cell is loaded and the gNB schedules some eMBB UEs on the resource overlapping with CG occasion, as shown in Fig. 1, there is going to be transmission collision of dynamically scheduled eMBB and URLLC GF transmissions.

![Fig. 1. A collision of UL URLLC GF transmission with eMBB transmission in case of Frequency Division Duplex (FDD).](image)

When the transmissions from eMBB and URLLC UEs in GF resources overlap, it results in lower decoding probability due to lower resulting SINR for both UEs. This can be a serious problem for URLLC UEs in particular due to their tight latency and reliability targets.

In case the gNB is able to identify the URLLC UE from its DMRS sequence, it may try to quickly reschedule the UE over non-overlapping resources if an error happens.
The increased interference due to overlapping transmissions of eMBB and URLLC UEs may lead to a catastrophic situation when the gNB may not even identify the URLLC UE (DMRS miss-detection). The current HARQ structure in UL transmission for NR is timer based, which means that upon transmission of packet, the UE will start the HARQ timer. If it receives an UL grant for the re-transmission of the same transport block (TB) from the gNB, it does the retransmission over the resources scheduled in the UL grant. If it receives no UL grant from the gNB and the HARQ timer expires, it considers that the TB was successfully decoded at the gNB and discards the data in the buffer.

The timer-based HARQ feedback and UL GB retransmission are standardized because this minimizes the control overhead for sending HARQ feedback. This is reasonable in general but in the cases of dynamic UL multiplexing, giving rise to overlapping transmissions with GF UEs, when the gNB will not be able to identify the UEs transmitting on GF resources, the UEs will discard their packets and consider the successful detection that leads to serious performance degradation for URLLC UEs.

C. Prior art

In [6] and [7], when there is an overlap, the gNB asks the URLLC UE to apply a different pre-configured transmission power that is higher than power level used in case of no overlap. However, an increase of power causes an interference among the neighboring cells. Secondly the cell-edge UEs may be power limited and cannot raise their transmission power. This is also the problems of [8] when the gNB assigns URLLC physical uplink shared channel (PUSCH) with updated transmission parameters such as resource, MCS, transmit power, etc. on GF resources that are occupied by eMBB PUSCH.

In [9], the group common control channel reveals resource range allocated to GB eMBB UE so the GF UE can exclude the occupied resource. Nevertheless, the GF UE has less resources left to transmit than the original configured resources so this partial transmission may not be decodable at the gNB.

In [10], the gNB informs the URLLC UE which of the GF resource set has overlapping eMBB transmission such that the URLLC UE can initiate the GF transmission over resources not occupied by the ongoing eMBBB transmission. It might result in high latency if all resources in the current occasion are full and the UE must wait until the next transmission occasion.

A preemptive scheme in [11] cannot be applied in URLLC GF transmission because the gNB does not know the URLLC transmission in advance to preempt eMBB transmission.

A successive interference cancellation (SIC) receiver in [12] only benefits eMBB UE rather than URLLC UE because URLLC data is decoded first due to latency requirement.

In this work, a multiplexing scheme with two-step strategy is presented in Section II. The first step comprises of the gNB transmitting an overlap indication whenever it schedules an UL transmission having an overlap with the resources configured to the GF transmissions. The second step of the proposed strategy comprises of making the overlapped transmissions to use explicit HARQ feedback structure. An alternative way of the second step is to make the gNB indicate the UEs to send SR in parallel to data transmitted on the GF resources to improve the reliability. Section III shows numerical results and performance evaluation. Finally, Section IV is conclusion.

II. Strategy to multiplex the eMBB and URLLC UEs

A. Overlap indication and explicit HARQ acknowledgement (ACK) feedback

This paper proposes two-step strategy to overcome the problem of the multiplexed UL eMBB and URLLC transmissions. In the first step, upon scheduling a GB transmission of the eMBB UE over the GF resources, the gNB sends an indication of resource overlap to the URLLC UEs. As the gNB does not know which of the URLLC UEs configured for GF resources may become active in the current interval, this indication needs to be sent to all UEs who have been configured with the GF resources in the overlapping interval as illustrated in Fig. 2. Upon receiving this overlap indication, the URLLC UEs are aware of the resources which have been dynamically scheduled for other UEs and in case of transmission, their transmissions will be received with an increased interference.

![Fig. 2. Signalling the URLLC UEs about an overlap with the eMBB UEs in GF regions.](image-url)
successful detection of their TB or UL grant for re-
transmission in case the gNB failed to decode the TB
but was able to identify the transmitting UE. For the
third case, if the gNB even fails to identify the UE due
to high interference (DMRS miss-detection), it cannot
schedule the UE for a re-transmission in the conventional
timer-based feedback so the UE assumes that a packet
is decoded correctly and drops it from buffer to transmit
the next packet. This is the scenario where our proposed
explicit HARQ feedback becomes the most promising.
If the GF UE receives neither an ACK nor an UL grant
within a configured time, the UE does not consider that
its data is successful, rather it considers that the gNB
failed to identify its identity (ID) due to high interference
in the overlapping transmissions and retransmits the TB
on the subsequent GF resources. Table I summarizes the
operation of conventional timer-based feedback structure
and explicit feedback structure.

### Table I
Comparison of Different Feedback Structures

<table>
<thead>
<tr>
<th>Case</th>
<th>Timer-based feedback</th>
<th>Explicit feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMRS: detected TB: decoded</td>
<td>No Ack/UL grant</td>
<td>Ack</td>
</tr>
<tr>
<td>DMRS: detected TB: failed</td>
<td>UL grant: reschedule</td>
<td>No Ack/UL grant: packet lost</td>
</tr>
<tr>
<td>DMRS: failed</td>
<td>No Ack/UL grant: packet lost</td>
<td>No Ack/UL grant: packet transmitted automatically</td>
</tr>
</tbody>
</table>

Block error rate (BLER) of a packet in timer-based
feedback structure and explicit feedback structure is
shown in (1) and (2) in Section III.

The timer value, for which the UE should wait to
receive ACK or UL grant, can be configured by radio
resource control (RRC) parameters. This value can be
selected as a function of reliability and latency targets
of the UEs. For some UEs with extremely high latency
and reliability targets, this timer value can be put to zero
which means to do the automatic transmission in case of
overlapping transmissions to maximize the chance of
correct data detection at the gNB.

### B. Overlap indication and additional SR

In a variation of the proposed scheme in Section II-A,
the second step of making the transmissions explicit
HARQ feedback-based can be replaced to use an addi-
tional SR. The gNB can indicate the URLLC GF UEs by
overlap indication in the first step described in Section
II-A to send a SR in parallel for the TB transmitted over
the overlapping GF resources. The SR sent to the gNB
will provide a further means besides DMRS detection
to detect the ID of the UE transmitting in the interfered
GF transmission. When the gNB is unable to identify
the UE making the GF transmission because of DMRS
miss-detection, the gNB still can identify that UE by
decoding the additional SR and react fast to the received
SR by sending an UL grant to this UE. Thanks to the
UL grant, the UE is likely to retransmit the packet and
has a successful transmission in latency budget instead
of assuming a successful transmission and dropping the
packet as in the conventional scheme. When the gNB is
able to decode the data successfully, it still reacts to SR
to send some indications to the UE about the successful
detection.

Physical uplink control channel (PUCCH) and
PUSCH are not allowed to be transmitted simultaneously.
In case the UE needs to transmit uplink control in-
formation (UCI) while transmitting UL data on PUSCH,
it sends the UCI on PUSCH. When the UE is not trans-
mitting PUSCH, it sends the UCI carrying SR, feedback
for DL data etc, on the PUCCH resources using appro-
priate PUCCH format as a function of UCI content and
the PUCCH configurations. The simplest strategy to send
SR (which is UCI) when the URLLC UE is transmitting
over overlapping GF resources would be to transmit it
over PUSCH GF resources along with the transmission
of the TB. This can be simple from implementation
perspective but from the performance point of view, it
does not help to improve significantly the performance of
the UE ID detection. If DMRS of PUSCH is not detected
by the gNB because of bad channel and interference
from eMBB transmission, there is high chance that the
gNB also cannot decode the SR to obtain the UE ID
multiplexed with PUSCH. For this reason, to achieve
a better UE ID detection, the SR is proposed to be
transmitted using PUCCH configuration on the specified
PUCCH resources separated from PUSCH. As PUCCH
resources are dedicated resources on different frequency
physical resource blocks (PRBs) and OFDM symbols,
this provides additional diversity advantage to the SR
transmitted in these resources compared to multiplexing
and transmitting it over UL GF resources along with
the TB. Therefore, the UE is configured to transmit this SR
on PUCCH resources in parallel to the transmission
of TB on the overlapping resources.

The overlap indication may have an explicit indica-
tion, in the form of a single bit flag, which may require
the UEs transmitting over overlapping GF occasions to
send SR. In fact, more flexibility and better performance
can be achieved by having the flexible control of the
explicit HARQ feedback structure and the transmission
of SR. The gNB can then choose which strategy to
choose in different situations. As an example, if the
periodicity of current GF occasions is not very fast, it
may make sense to indicate the overlapping GF UEs to
send an SR. And in the cases, if the resources configured
for SR are not sufficient or not in close proximity
compared to the subsequent GF occasions, it may be
suitable to prioritize the explicit HARQ feedback and
automatic retransmission in case the UEs receive no ACK/UL grant indication for the transmitted TB.

C. Configuration and Signalling for the Overlap Indication

One very important feature of the proposed scheme is that the overlap indication is sent to the UEs who have been pre-configured for the GF resources. As GF resources may be shared by multiple UEs, and the gNB has no idea a priori which of these UEs may transmit their data over GF resources, this overlap indication is sent in a group-common manner. Thus, the proposal is to send the overlap indication in a group-common downlink control information (DCI).

For UL overlap indication sent in group-common DCI, a DCI format similar to the DCI format 2_1 in [13] can be used. DCI 2_1 is also used for DL pre-emption indication. The size of DCI format 2_1 is configurable by higher layers up to 126 bits and each indication has 14 bits.

The UL overlap indication sent to the URLLC UEs comprises of the indication of UL GF resources typically scheduled for the eMBB UEs. The eMBB UEs would be normally scheduled for the slot duration or most part of the valid UL symbols so it would be judicious to have more bits of UL reference resource field defining the frequency granularity. To keep a format close to the DL pre-emption indication and adapted to indicate the UL overlap resource, there are two possibilities of UL overlap indication. In the first design, all 14 bits are used to indicate the frequency PRB region for the whole slot. Thus, each bit in the 14-bit long bitmap indicates 1/14 of the frequency PRBs of the carrier as shown in left part of Fig. 3. The second option is to split the time-frequency grid of the slot in 7 frequency zones each spanning one half slot. Thus, each bit may indicate an overlap over 1/7 of the frequency PRBs for a half slot time duration as illustrated in right part of Fig. 3.

The UEs configured with GF resources can be configured to listen and decode the overlap indication as part of GF configuration. The activation and de-activation of this operation can be done in different ways for Type 1 and Type 2 GF as specified in [5]. For RRC configured Type 1, the activation and de-activation can be done by RRC signalling. For configured grant Type 2, where some parameters of GF can be updated by DCI, the activation or de-activation of overlap indication can be made through DCI signalling which is used to update other GF parameters.

The overlap indication can be sent at the same time when the gNB sends the dynamic UL grant scheduling a UE over the GF resources as shown in the left part of Fig. 4.

When the gNB sends an UL grant scheduling an eMBB UE, the scheduled resources are not necessarily in the same slot where UL grant is sent. Rather typically the UL grant will be for the resources located in one of the subsequent slots. If the gNB transmits the overlap indication along with the UL grant, it needs to indicate the slot where this overlap will occur. To avoid this additional signalling and to keep the treatment of UL grant simple, the overlap indication can be transmitted in the slot where overlap occurs as shown in the right part of Fig. 4. However, it may not be preferable to send and receive at the same time, so transmitting the overlap indication in the DL direction in the same slot as of the overlapping CG occasions may not be very interesting. In some other cases like time division duplex (TDD) operation mode, it may be completely impossible. Thus, based on the system design, one of these two transmission schemes is determined.

One big advantage of sending the overlap indication to URLLC UEs is that the overlap indication is primarily indicating the overlap caused by dynamic scheduling of eMBB UEs over the CG resources. As eMBB UEs may be scheduled only once during a slot, the indication periodicity can be kept to be once per slot and no mini-slot monitoring is needed to receive the overlap indication. This is advantageous in the sense that it does not overload the UEs with additional DCI monitoring and decoding burden.

D. Design of Explicit HARQ Feedback

In the proposed strategy, the gNB shifts the GF transmissions in the overlap region from timer-based approach to explicit-HARQ-feedback approach. Thus, a design for the explicit HARQ feedback in general may be needed. The proposal is to use DCI as an explicit HARQ feedback sent with UE specific configured scheduling-radio network temporary identifier (CS-RNTI) which is
used with configured grant-based transmissions. If the gNB is able to successfully decode the data despite the overlap, it can send an UL grant to this UE with the same HARQ process number (HARQ ID) as of the successfully received TB, and the UE upon receiving this UL grant would know that this is in fact not a retransmission request but an explicit ACK for the previously transmitted TB. To avoid any confusion, new-data-indicator (NDI) field can be set to zero. Further, some of the fields in the DCI which are actually not needed, such as the time and frequency resource assignment fields, may be sent with fixed known values which can be pre-decided to be used in the ACK indication.

III. NUMERICAL RESULTS AND PERFORMANCE EVALUATION

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waveform</td>
<td>CP-OFDM</td>
</tr>
<tr>
<td>Subcarrier spacing</td>
<td>60kHz</td>
</tr>
<tr>
<td>Channel model</td>
<td>Rician</td>
</tr>
<tr>
<td>K factor</td>
<td>1</td>
</tr>
<tr>
<td>Number of allocated PRB</td>
<td>8</td>
</tr>
<tr>
<td>DMRS detection mechanism</td>
<td>Time-domain correlation</td>
</tr>
</tbody>
</table>

Fig. 5. DMRS detection performance.

Simulation parameters are shown in Table II. Fig. 5 illustrates the performance of DMRS detection in three cases: the URLLC UE transmits in the GF resources without any collision with another eMBB UE, the URLLC UE has a collision with an eMBB UE at the same power level and the URLLC UE has a collision when increasing power 1dB higher than an eMBB UE. For each DMRS detection, the correlation result is compared with a threshold to determine whether DMRS exists or not. This threshold is chosen according to a target false alarm rate (FAR) indicating the cases that the gNB determines the existence of DMRS while in reality there is no DMRS transmitted. A higher threshold is required for a lower FAR but also results in more missed detection.

Due to a collision between DMRS of the URLLC UE of interest and data/DMRS of another eMBB UE, the performance of DMRS detection of the URLLC UE degrades significantly as can be seen in Fig. 5 and cannot achieve the miss detection probability $10^{-5}$ at the same SNR level of the case without collision. At FAR of 0.001 and SNR of -1.4dB, the miss detection probability increases from $10^{-5}$ of the case without collision to $3.4 \times 10^{-4}$ of the case with a collision with an eMBB UE at the same power. Fig. 5 also shows that even with power control scheme in [6] and [7] when the URLLC UE’s power increases 1dB higher than the eMBB UE’s power, miss detection probability is still $2.44 \times 10^{-4}$ that is higher than $10^{-5}$ of the case without collision. As DMRS detection is mandatory for channel estimation to decode data as well as for recognizing UE ID to reschedule a retransmission if necessary in conventional scheme, a degradation of DMRS detection makes the system unable to support reliability URLLC requirement. BLER of an UL transmission with one potential retransmission in the conventional scheme or power control scheme ($P_{d1}^e$) is calculated as

$$P_d = P_{DMRS}^e + (1 - P_{DMRS}^e)P_{d1}^e + (1 - P_{DMRS}^e)P_{d2}^e,$$

where $P_{DMRS}^e$, $P_{DMRS}^d$ are the miss detection probabilities of the initial (with collision) and retransmitted (without collision) DMRS (see Figure 5), and $P_{d1}^e$, $P_{d2}^e$ are the error probabilities of the initial (with collision) and retransmitted (without collision) PUSCH.

In (1), the first term is the error probability when the gNB cannot detect DMRS to decode or reschedule data so the UE does not retransmit data and data is lost. The second term is the error probability when the gNB detects DMRS and identifies UE ID but fails to decode data so it reschedules data. However, it cannot decode the retransmission and an error still occurs.

The usage of an explicit HARQ feedback as explained in Section II-A solves the problem of DMRS miss-detection in the overlapping region because it allows the UE to carry out the retransmissions in the interference-free regions even if DMRS is not detected by the gNB. BLER of an UL transmission with one potential retransmission in the proposed scheme with explicit feedback ($P_{d}^e$) is calculated as

$$P_d = P_{DMRS}^e + (1 - P_{DMRS}^e)P_{d1}^e + (1 - P_{DMRS}^e)P_{d2}^e,$$

Compared to (1), the first term in (2) is enhanced because of a retransmission while the second terms are the same.

The final column in Table III shows a remarkable enhancement of the first term of error probability in
UL transmission with explicit HARQ feedback in case of URLLC and eMBB multiplexing in comparison to the conventional scheme with timer-based feedback and power control scheme in [6] and [7]. Moreover, the proposed scheme can be applied to all UEs in a cell while the power control scheme to increase URLLC UEs’ power cannot be applied to the cell-edge UEs because of power limitation.

Table III also shows an improvement of UE ID detection (the first term in (3)) when an additional SR is transmitted in the separate PUCCH in parallel with data (PUSCH) in GF resources. As can be seen in (3), the SR provides another chance for the gNB to detect UE ID. Therefore, the error probability because of DMRS miss-detection decreases. BLER of an UL transmission with one potential retransmission in the proposed scheme with an additional SR ($P^e_{SR}$) is calculated as

$$P^e_{SR} = P^e_{DMRS1} \cdot P^e_{SR} + \left(1 - P^e_{DMRS1} \cdot P^e_{SR}\right) \cdot P_{d1} \times (P^e_{DMRS2} + \left(1 - P^e_{DMRS2}\right) \cdot P_{d2}),$$

where $P^e_{SR}$: the error probability of SR.

The selection between two proposed schemes is explained in Section II-B. The presence of retransmission in explicit feedback leads to latency and resource consumption but guarantees target reliability in case of DMRS miss-detection, while the conventional scheme stops the transmission straightforwardly and causes packet loss so latency and resource consumption have no meaning when a packet is already failed to be decoded correctly. Moreover, with SCS 60kHz and the decoding time of one transmission being 0.1ms for a packet spreading in 4 OFDM symbols, even with one retransmission, the system consumes 0.5ms in total and still satisfies the latency requirement of 1ms.

Compared to conventional scheme, overhead of explicit feedback structure is higher due to ACK signal but is limited because the explicit feedback structure is only used when there is an overlap in GF resources (a cause of high interference) and an overlap indication triggers this structure. In addition, as can be seen in Table III, the error probability in the overlapping transmissions increases so the number of ACK feedback decrease while the number of the cases without any feedback because of DMRS miss-detection increase. For this reason, a mechanism to trigger a retransmission as the proposed explicit feedback structure becomes imperative. It is the same reason that the overhead of SR transmitted in parallel with PUSCH is acceptable in eMBB and URLLC multiplexing.

### IV. Conclusion

This paper presents a strategy to multiplex the GB eMBB and GF URLLC transmission while guaranteeing the strict requirements of URLLC. An overlap indication is used and combined with an explicit HARQ feedback or an additional SR to help reduce the error probability of URLLC UE in case of DMRS miss-detection. The proposed scheme provides a mechanism to meet very stringent URLLC constraints with minimal additional control overhead.

**ACKNOWLEDGMENTS**

The work described in this paper has been partially supported by TCL and the H2020 project 5GENESIS.

**REFERENCES**

[1] 3GPP TR 38.913 v15.0.0, “Study on scenarios and requirements for next generation access technologies.”
[3] 3GPP TS 38.211 v15.3.0, “Physical channels and modulation.”
[8] Sony, “Inter-UE uplink multiplexing of eMBB and URLLC”, 3GPP R1-1812745, RAN1#95, Spokane, USA, November 12–16, 2018.
[13] 3GPP TS 38.212 v15.3.0, “Multiplexing and channel coding.”

**TABLE III**

<table>
<thead>
<tr>
<th>Case</th>
<th>URLLC UE ID miss detection probability</th>
<th>Retransmission in UE ID miss detection</th>
<th>URLLC UL transmission’s BLER due to the first UE ID miss detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>No collision</td>
<td>$10^{-5}$</td>
<td>No</td>
<td>$10^{-7}$</td>
</tr>
</tbody>
</table>