

Enhancing URLLC Performance in 5G Networks Through Hybrid ARQ Feedback for eMBB and URLLC Multiplexing in Grant-Free Resources

Sneha Bhambure
snehabhambure@gmail.com

Abstract—In 5G uplink (UL) transmission, Ultra-Reliable Low-Latency Communication (URLLC) and Enhanced Mobile Broadband (eMBB) users may simultaneously utilize the same grant-free (GF) resources, leading to potential collisions and performance degradation for URLLC. This paper introduces a two-step strategy to mitigate this issue. First, an overlap indication is employed to notify URLLC users of potential resource conflicts with eMBB transmissions. Second, explicit Hybrid Automatic Repeat Request (HARQ) feedback or an additional scheduling request (SR) is proposed to improve URLLC performance. Simulation results demonstrate a significant reduction in error probability caused by Demodulation Reference Signal (DMRS) miss-detection, ensuring URLLC’s stringent reliability and latency requirements are met while optimizing resource efficiency. These findings highlight the potential of combining explicit HARQ feedback and overlap indicators to enhance multiplexing in GF resources.

I. INTRODUCTION

The ongoing evolution of advanced wireless technologies continues to address the increasing need for high-speed, ultra-reliable, and low-latency communications (URLLC). A critical objective of emerging network standards, such as the 6G initiative, is to enable devices requiring exceptional reliability and minimal delay to operate at their fullest potential. As highlighted in [1], the international standardization body, the International Telecommunication Union (ITU), proposes a target reliability of 10^{-6} for packet transmission of 64 bytes, with a target user plane latency under 0.5 ms. Future versions of this framework are anticipated to push the envelope even further, aiming for near-zero latency and higher reliability targets, particularly for applications in autonomous vehicles, healthcare systems, and smart cities [2].

A. Enhancements in 6G Specification

In the ongoing development of the 6G specification, several breakthroughs at the physical layer are being considered to address the critical demands of URLLC. A standout feature in 6G is the adaptable subcarrier spacing (SCS) ranging from 20 kHz to 500 kHz, which facilitates much shorter transmission durations [3]. Furthermore, the introduction of micro-slot scheduling for both uplink (UL) and downlink (DL) communications [4] allows the system to adjust dynamically to the varying data needs of user devices (UEs). These innovations allow the network to rapidly adapt to fluctuations in traffic, providing much faster response times compared to older generations like LTE.

A key addition is the introduction of grant-free (GF) uplink transmission, eliminating the need for UEs to send scheduling requests (SRs) and await network grants, thus enhancing efficiency and reducing latency [5].

B. Challenges in Coordinating URLLC and eMBB with GF Resources

Despite these advances, one of the ongoing challenges in modern networks lies in efficiently multiplexing URLLC and enhanced mobile broadband (eMBB) on shared GF resources. As described in Section I-A, while the next-generation base station (gNB) can allocate GF resources to URLLC devices, it lacks precise information about which devices will access these resources and when. This uncertainty leads to potential interference when eMBB devices are assigned overlapping resources with URLLC transmissions, as shown in Fig. 1, which can result in transmission collisions and substantial signal degradation.

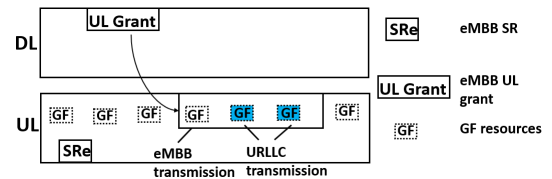


Fig. 1. Interference between URLLC and eMBB transmissions in overlapping GF regions of frequency division duplex (FDD) systems.

Such overlap results in reduced signal-to-noise-plus-interference ratios (SINR) for both eMBB and URLLC UEs, leading to an increased likelihood of transmission failure. This problem is particularly problematic for URLLC UEs, which rely on strict latency and reliability standards.

If the gNB can detect the URLLC device via its reference signal, it may attempt to reschedule transmissions to non-overlapping resources. However, if the interference is significant or the gNB fails to detect the URLLC device (due to missed reference signal detection), the gNB cannot decode the transmission successfully. This results in data loss, as the UE will discard the data based on a hybrid automatic repeat request (HARQ) feedback mechanism driven by timers, creating a gap in communication.

C. Existing Solutions and Their Drawbacks

Various solutions have been proposed to mitigate interference resulting from overlapping transmissions. For instance, [6] and [7] suggest that the gNB instructs URLLC UEs to boost transmission power to combat interference. While this approach can reduce interference, it often leads to additional issues, such as interference with neighboring cells and power limitations for devices on the edge of the cell. Similarly, [9] addresses resource conflict by adjusting the physical uplink shared channel (PUSCH) parameters for URLLC transmissions that overlap with eMBB resources, but this does not fully solve the interference problem.

In [10], the use of a common control channel enables the gNB to notify URLLC devices about the presence of eMBB transmissions in the same resource blocks, allowing devices to avoid using these resources. However, this can lead to inefficient use of resources, as fewer transmission opportunities are left for URLLC devices, which can compromise reliability.

Another solution discussed in [11] proposes signaling resource overlap directly to the URLLC devices, allowing them to avoid conflict. However, this could increase latency if the available transmission slots are already occupied, forcing devices to wait for the next available slot.

Preemptive scheduling schemes, as noted in [12], are unsuitable for URLLC due to the unpredictable nature of its transmission timing. Additionally, successive interference cancellation (SIC), explored in [13], mainly benefits eMBB UEs and does not address the time-sensitive needs of URLLC UEs.

This work introduces a novel two-phase approach to addressing resource multiplexing challenges between URLLC and eMBB transmissions. The first phase involves signaling resource overlap, and the second introduces an explicit HARQ feedback mechanism to improve transmission reliability and reduce packet loss. The proposed strategy is detailed in Section II, followed by performance analysis in Section III, with conclusions in Section IV.

II. PROPOSED APPROACH FOR COORDINATING EMBB AND URLLC TRANSMISSIONS

A. Resource Overlap Detection and Feedback Strategy

To mitigate the impact of overlapping resources between uplink (UL) transmissions from eMBB and URLLC UEs, this paper propose a dual-phase strategy. In the first phase, the gNB detects potential overlaps and sends a resource conflict indication whenever eMBB transmissions are scheduled on GF resources that might coincide with URLLC transmissions. This indication is broadcasted to all URLLC UEs using the same resources, as shown in Fig. 2. Once the URLLC devices receive this indication, they can adjust their transmission strategy to avoid interference.

The second phase of the strategy introduces an explicit HARQ feedback mechanism, replacing the conventional timer-based feedback. The overlap notification from the gNB includes a flag indicating whether explicit feedback is required. Upon receiving the flag, URLLC UEs using shared resources

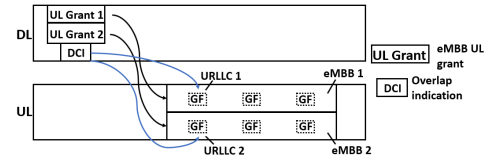


Fig. 2. Notification of resource overlap between eMBB and URLLC transmissions in GF regions.

will expect explicit acknowledgments (ACKs) or uplink grants for retransmission, ensuring they can quickly identify transmission failures due to interference and reschedule their transmissions accordingly.

If the gNB is unable to detect the URLLC device due to severe interference, the device will assume that its transmission was successful and will proceed with retransmission. This method minimizes the chances of premature data discarding, enhancing reliability.

TABLE I
COMPARISON OF FEEDBACK STRATEGIES IN DIFFERENT TRANSMISSION SCENARIOS

Transmission Scenario	Timer-based Feedback	Explicit Feedback
DMRS detected	No acknowledgment or uplink grant received	Acknowledgment (ACK) received for the transport block (TB)
DMRS detected	Uplink grant: No retransmission needed	Uplink grant: Reschedule the transmission
DMRS failed	No acknowledgment or uplink grant: Packet lost	No acknowledgment or uplink grant: Packet is automatically retransmitted

The block error rate (BLER) for both the timer-based and explicit feedback mechanisms is outlined in equations (1) and (2) in Section III.

Radio resource control (RRC) settings determine the timer duration for waiting for feedback. For highly reliable and low-latency applications, the timer may be set to zero, prompting automatic retransmission if overlaps are detected, thus increasing the likelihood of successful data delivery at the gNB.

B. Overlap Signaling and Additional Scheduling Request

An alternative approach to the method outlined in Section II-A introduces the use of a supplementary Scheduling Request (SR) in place of the second phase of feedback signaling based on HARQ. In this configuration, the gNB notifies UEs engaged in ultra-reliable low-latency communications (URLLC) with a specific overlap indication in the first step. This signals the UEs to transmit an SR in parallel with the transmission of the transport block (TB) over the shared GF resources. The SR is crucial for identifying the transmitting UE, particularly in scenarios where the DMRS detection fails. Even if the gNB is unable to identify the UE due to missed DMRS, the SR still allows the gNB to recognize the UE and issue an uplink grant, facilitating the retransmission of packets and ensuring

that the transmission meets the latency requirements. Unlike traditional methods that assume packet loss in the absence of an ACK or UL grant, this approach enhances reliability by supporting automatic retransmissions.

When the UE needs to send uplink control information (UCI) while simultaneously transmitting data on PUSCH, the UCI, such as SR or feedback related to downlink data, is transmitted using the appropriate PUCCH format. The PUCCH resources are allocated based on the type of UCI and the configured PUCCH parameters, ensuring efficient communication. While sending the SR over PUSCH alongside the TB transmission is simpler, it may not always provide optimal UE identification. In severe channel conditions or high interference, where DMRS detection on PUSCH fails, the gNB may also fail to decode the multiplexed SR with the PUSCH. To improve UE identification, it is proposed that the SR be transmitted over PUCCH resources, separate from PUSCH. These PUCCH resources, isolated on distinct frequency PRBs and OFDM symbols, offer better detection performance due to their diversity advantages compared to multiplexing the SR with PUSCH.

Additionally, the overlap indication can be transmitted with an explicit flag, usually represented by a single bit, which instructs the UEs to send SR during the overlapping GF transmission windows. This explicit HARQ feedback approach offers enhanced flexibility and performance, as it allows the gNB to adapt based on the scenario. For example, if the overlap period between GF occasions is longer, the gNB may prioritize SR transmission. Conversely, when SR resources are scarce, the gNB can opt for automatic retransmission based on explicit HARQ feedback.

C. Configuration and Signaling for Overlap Indication

A significant feature of the proposed approach is the group-common nature of the overlap indication. Since multiple UEs may share GF resources, and the gNB is unaware of which UE will transmit over those resources, the overlap indication is broadcast to all relevant UEs using a group-common downlink control information (DCI) format. This DCI format is similar to DCI format 2_1, as discussed in [14], which is also used for downlink pre-emption indication. The DCI size can be configured by higher layers to support up to 126 bits, with each overlap indication occupying 14 bits.

For uplink overlap indication, the gNB transmits a group-common DCI to URLLC UEs, identifying the UL GF resources generally allocated to eMBB UEs. The format is flexible, allowing adjustment of the bit field to indicate frequency regions. There are two primary configurations for uplink overlap indication: In the first configuration, all 14 bits of the bitmap correspond to the entire frequency PRB region for the slot. In the second configuration, the slot is divided into seven frequency zones, each occupying half a slot, with each bit representing one of these zones, offering finer granularity as illustrated in Figure 3.

Devices that utilize grant-free (GF) resources are capable of detecting and processing overlap signals as part of the GF

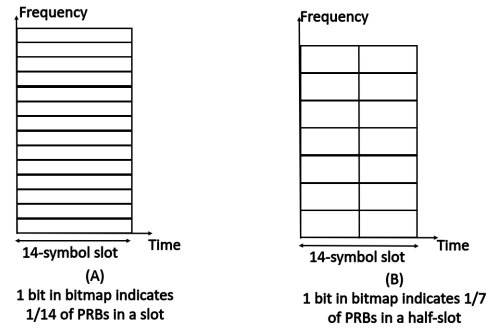


Fig. 3. Resource Indication in Uplink Overlap Indication.

setup. The activation and deactivation of this feature can vary between Type 1 and Type 2 GF, as described in [5]. In Type 1 systems, this control is managed through Radio Resource Control (RRC) signaling, while in Type 2 systems, the overlap signal is controlled via Downlink Control Information (DCI), which also handles other GF configuration updates.

Typically, the overlap indication is transmitted along with the uplink (UL) grant, which allocates resources for a user equipment (UE) within the GF framework, as shown on the left side of Fig. 4.

In cases where a UL grant is issued for scheduling an enhanced Mobile Broadband (eMBB) UE, the assigned resources may not align with the time slot of the UL grant. Normally, the resources allocated in the UL grant are used in a subsequent time slot. If the overlap indication is sent with the UL grant, it is essential to specify the exact time slot where the overlap will occur. To simplify signaling and optimize UL grant management, the overlap indication can be incorporated within the same time slot as the overlap, as depicted on the right side of Fig. 4. However, simultaneous transmission might not always be feasible, particularly in Time Division Duplex (TDD) systems, where such an approach may be challenging to implement. Therefore, system design will determine whether to use one of these two transmission methods.

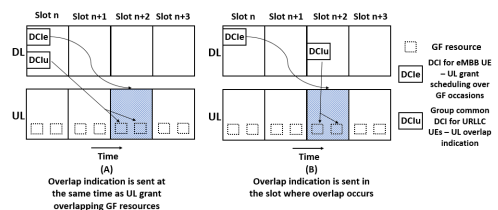


Fig. 4. Timing to send the UL Overlap Indication.

One key advantage of sending the overlap indication to URLLC UEs is that it primarily represents the overlap created by the dynamic scheduling of eMBB UEs within the GF resources. Since eMBB UEs are generally scheduled once per slot, the overlap indication can be sent once per slot, removing the need for monitoring mini-slots. This strategy reduces the UEs' workload of extra DCI monitoring and decoding, thus improving overall system efficiency.

D. Enhanced Hybrid Feedback Mechanism for Uplink Transmission

In this novel approach, the gNB moves away from using traditional timing-based systems and adopts an advanced Hybrid Automatic Repeat reQuest (HARQ) feedback method, specifically optimized for interference-prone regions. This requires the development of a strong mechanism for explicit HARQ feedback. The solution entails leveraging Downlink Control Information (DCI) as the explicit HARQ feedback signal, transmitted through a UE-specific Scheduling-Radio Network Temporary Identifier (CS-RNTI), already configured for grant-based transmissions. When the gNB successfully decodes data amidst overlapping signals, it sends an uplink (UL) grant to the appropriate UE, linking it with the same HARQ process number as the received transport block (TB). Upon receiving the UL grant, the UE interprets this as an explicit acknowledgment (ACK) for the previous TB transmission, not a retransmission request. To avoid any ambiguity, the New Data Indicator (NDI) is set to zero. Additionally, any unnecessary fields in the DCI, like those pertaining to time and frequency resource allocation, are assigned default values that are pre-set and understood in the ACK signal.

III. PERFORMANCE ANALYSIS AND EVALUATION

TABLE II
SIMULATION PARAMETERS

Parameters	Values
Waveform	CP-OFDM
Subcarrier spacing	60 kHz
Channel model	Rician
K factor	1
Number of allocated PRB	8
DMRS detection mechanism	Time-domain correlation

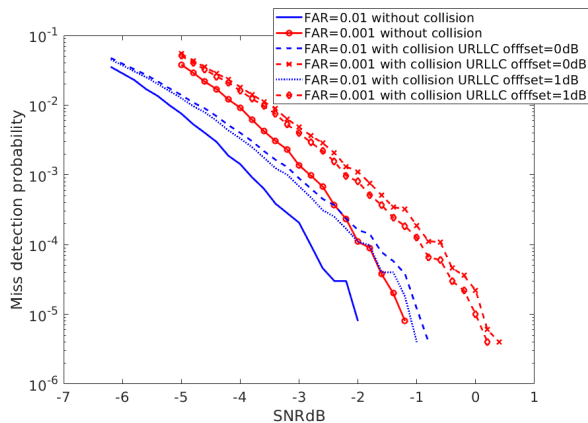


Fig. 5. Performance of CSI detection in various scenarios.

The simulation parameters used are outlined in Table II. As illustrated in Fig. 5, CSI detection is analyzed across three different situations: the URLLC UE transmits in reserved resources without interference from another eMBB UE, the

URLLC UE faces interference from an eMBB UE transmitting at the same power level, and the URLLC UE experiences interference while boosting its transmission power by 1 dB over the eMBB UE's level. For each CSI detection, the correlation result is compared against a predefined threshold to ascertain the presence of CSI. This threshold is designed to maintain a specific false alarm rate (FAR), which quantifies the likelihood of the gNB mistakenly detecting CSI when none is transmitted. A higher threshold decreases the FAR but increases the probability of missing detection.

In scenarios where the URLLC UE's CSI overlaps with that of an eMBB UE, the CSI detection performance of the URLLC UE significantly deteriorates, as shown in Fig. 5. It cannot achieve the desired missed detection rate of 10^{-5} at the same SNR as in the collision-free case. At a FAR of 0.001 and an SNR of -1.4 dB, the missed detection probability increases from 10^{-5} in the no-interference case to 3.4×10^{-4} with interference from an eMBB UE. Even when a power control strategy, as referenced in [6] and [7], is applied, the missed detection probability remains at 2.44×10^{-4} , even with the URLLC UE raising its transmission power by 1 dB over the eMBB UE, still above the 10^{-5} threshold in the interference-free case. Since CSI detection is vital for channel estimation and UE identification in conventional systems, failing to detect CSI hampers the ability to meet URLLC reliability requirements. The Block Error Rate (BLER) for an uplink (UL) transmission, assuming one possible retransmission in either conventional or power control schemes, is given by

$$C_1^p = C_{CSI1}^p + (1 - C_{CSI1}^p)C_{d1}^p(C_{CSI2}^p + (1 - C_{CSI2}^p)C_{d2}^p), \quad (1)$$

where C_{CSI1}^p, C_{CSI2}^p represent the missed detection probabilities for the initial and retransmitted CSI signals (see Figure 5), and C_{d1}^p, C_{d2}^p represent the error probabilities for the initial and retransmitted data transmission.

In (1), the first term represents the error probability when the gNB fails to detect CSI, preventing the UE from retransmitting, leading to data loss. The second term accounts for the situation when the gNB detects CSI but fails to decode the data, prompting a retransmission that also ends in failure.

The challenge of missed CSI detection in high-interference regions is addressed by the explicit HARQ feedback mechanism, as detailed in Section II-A. This feedback allows the User Equipment (UE) to initiate retransmissions in areas with reduced interference, even when the gNB fails to detect CSI. The BLER for an uplink (UL) transmission with one retransmission under the explicit HARQ feedback scheme (C_2^p) is given by

$$C_2^p = C_{CSI1}^p(C_{CSI2}^p + (1 - C_{CSI2}^p)C_{d2}^p) + (1 - C_{CSI1}^p)C_{d1}^p(C_{CSI2}^p + (1 - C_{CSI2}^p)C_{d2}^p). \quad (2)$$

Here, the first term in (2) benefits from retransmission, while the second term remains unchanged.

As shown in the final column of Table III, the first error term for UL transmissions employing explicit HARQ feedback improves considerably, especially when URLLC and

TABLE III
EVALUATION OF TRANSMISSION SCENARIOS AT SNR = -1.4 dB, FAR = 0.001, $P_{d2}^e = P_{SR}^e = 0.01$

Scenario	Missed ID Probability for URLLC UE	Retransmission Due to Missed ID	BLER for URLLC UL Transmission from First Missed ID
No interference	5×10^{-6}	No	5×10^{-6}
Traditional method with interference	2.7×10^{-3}	No	2.7×10^{-3}
Power control with interference ([6] and [7])	1.8×10^{-3}	No	1.8×10^{-3}
Interference with explicit feedback	2.9×10^{-3}	Yes	1.2×10^{-7}
Interference with additional SR (proposed method)	1.1×10^{-7}	No	1.1×10^{-7}

eMBB transmissions are multiplexed. This is a clear advantage over conventional methods that rely on timer-based feedback and power control strategies, as described in [6] and [7]. Furthermore, the proposed method is applicable to all UEs within a cell, while power control methods, which increase the transmission power for URLLC UEs, are ineffective for cell-edge UEs due to power limitations.

Table III also shows a reduction in UE ID detection errors when an extra Scheduling Request (SR) is sent over a separate Physical Uplink Control Channel (PUCCH), operating concurrently with data transmission (PUSCH) in the shared spectrum. This SR gives the gNB an additional chance to detect the UE ID, reducing missed CSI detection errors. The BLER for a UL transmission, considering one retransmission with the updated scheme incorporating SR (C_3^p), is written as

$$C_3^p = C_{CSI1}^p \times C_{SR}^p + (1 - C_{CSI1}^p \times C_{SR}^p) C_{d1}^p \times (C_{CSI2}^p + (1 - C_{CSI2}^p) C_{d2}^p), \quad (3)$$

where C_{SR}^p is the error probability associated with SR transmission.

The selection between these proposed schemes is addressed in Section II-B.

While retransmissions in explicit feedback mechanisms increase latency and resource usage, they ensure reliable performance even with missed CSI detection. In contrast, the conventional method halts transmission after a failed detection, leading to packet loss without considering resource or latency constraints. Moreover, with a subcarrier spacing (SCS) of 60 kHz and a decoding time of 0.1 ms per transmission across 4 OFDM symbols, the total system time remains 0.5 ms, thus meeting the 1 ms latency requirement.

Although the explicit feedback scheme introduces overhead due to the ACK signal, it remains manageable, especially as it is only triggered during interference in shared resources. As seen in Table III, overlapping transmissions result in higher error probabilities, leading to fewer ACKs and increased cases of missed CSI detection. This emphasizes the importance of an explicit retransmission feedback mechanism. Similarly, the overhead of SR transmission alongside PUSCH is acceptable for multiplexed eMBB and URLLC cases.

IV. CONCLUSION

This study introduces a novel approach for efficiently managing the simultaneous transmission of GB eMBB and GF

URLLC data, ensuring that the critical demands of URLLC are met. By incorporating an overlap signaling method and coupling it with either explicit HARQ feedback or an extra SR, the approach effectively mitigates error rates for URLLC UEs, particularly in scenarios involving DMRS miss-detection. The proposed technique ensures compliance with rigorous URLLC requirements while minimizing the extra control signaling overhead.

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 v15.3.0, "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 v15.3.0, "Physical layer procedures for data."
- [6] Huawei, HiSilicon, "UL inter-UE transmission prioritization and multiplexing", 3GPP R1-1810158, RAN1#94bis, Spokane, Chengdu, China, 8–12 October, 2018.
- [7] ZTE, "UL inter-UE multiplexing between eMBB and URLLC", 3GPP R1-1900074, RAN1 AH 1901, Taipei, 21–25 January, 2019.
- [8] Sony, "On Enhancement of UL grant-free transmissions", 3GPP R1-1808345, RAN1#94, Gothenburg, Sweden, August 20–24, 2018.
- [9] Sony, "Inter-UE uplink multiplexing of URLLC & eMBB traffics", 3GPP R1-1812745, RAN1#95, Spokane, USA, November 12–16, 2018.
- [10] Institute for Information Industry (III), "UL Inter UE Tx prioritization/multiplexing", 3GPP R1-1813481, RAN1#95, Spokane, USA, November 12–16, 2018.
- [11] NEC, "UL inter-UE multiplexing of eMBB and URLLC", 3GPP R1-1812419, RAN1#95, Spokane, USA, November 12–16, 2018.
- [12] C.-P. Li, J. Jiang, W. Chen, T. Ji, and J. Smees, "5G ultra-reliable and low-latency systems design," in 2017 European Conference on Networks and Communications (EuCNC), Jun. 2017, pp. 1-5.
- [13] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge University Press, 2005.
- [14] 3GPP TS 38.212 v15.3.0, "Multiplexing and channel coding."