# Control and data channel combining in Ultra-Reliable Low-Latency Communication

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

***Abstract*—5G will be supporting new services that have remarkably higher requirements than LTE 4G and Ultra-reliable and low-latency communication (URLLC) is one of those emerged categories. Although various techniques have been proposed to improve the data reliability, there has been a gap in how to improve the reliability of control/scheduling information pointing to the scheduled data. In this paper, we propose an intelligent combining of retransmissions of physical downlink control channel (PDCCH) and the physical downlink data channel (PDSCH). In the proposed scheme, the downlink control information (DCI) on PDCCH already indicates the location of a potential retransmission of the corresponding PDSCH. Moreover, the retransmitted DCI can be combined with the first transmission so that resource consumption and latency are reduced compared to the conventional scheme. Theoretical calculations and simulation results show a decrease of resource consumption.**

***Index Terms*—5G, URLLC, PDCCH blocking, PDCCH combining, downlink scheduling scheme**

## I. Introduction

The advent of new paradigms like connected self-driving cars and virtual reality led the wireless standardization bodies to take them into account. To that respect, 3GPP has defined three service categories 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 requirements: reliability and latency. Basically, one of two factors must be sacrificed to attain the other factor.

### A. Techniques accepted in 3GPP Release 15

In Release 15, some modifications in frame structure of physical layer have been agreed to support URLLC.

In LTE, subcarrier spacing (SCS) is fixed at 15kHz while in 5G, SCS can be selected flexibly among 30kHz, 60 kHz, 120kHz or even 240kHz [12]. For this reason, the symbol length drops substantially and transmission latency is reduced.

Another option to reduce transmission latency is to do the transmission in mini-slot level instead of slot level [10]. Therefore, the waiting time of an incoming packet falls rapidly.

### B. Techniques under research for next 3GPP releases

New techniques in transmission scheme to attain the strict requirements focus on repetition-based transmission aiming to relax the complexity of control and data channel design. A repetition in frequency domain is proposed in [4] that is equivalent to a utility of a higher aggregation level (AL). It may cause PDCCH blocking due to a shortage of resources in a control resource set (CORESET). Besides, a PDCCH can be repeated in time domain without waiting for hybrid automatic repeat request (HARQ) feedback considered in [5]. PDSCH is delayed until PDCCH is decoded successfully in the UE and the base station (gNB) receives acknowledgement (ACK) of PDCCH. This method may cause a waste of resources if ACK feedback is not fast enough to stop the repetition process of the entire PDCCHs.

[6] considers a PDCCH repetition with asymmetric HARQ where the retransmission has a lower code rate than the initial transmission to increase reliability so the requirement of block error rate (BLER) target for the initial transmission is relaxed. However, in case the second transmission becomes necessary, it requires a much bigger resource than the first one.

PDCCH feedback is proposed in [7] to speed up PDCCH repetition. A PDCCH feedback is transmitted from the UE to the gNB if PDCCH is decoded correctly. If PDCCH fails, there is no feedback to the gNB and it detects this discontinuous signal to trigger PDCCH repetition without waiting for HARQ feedback. Another proposal is to do back indication of PDSCH. This means that if the first PDCCH fails, this slot containing PDSCH is stored and back indicated by the second PDCCH. However, the second transmission with both PDCCH and PDSCH might cause resource blocking.

In [8], PDCCH repetitions are generated by puncturing PDSCH of the eMBB UEs. Nevertheless, the performance of the eMBB UEs drops substantially due to the loss of the punctured bits' information.

In this work, we propose a repetition-based downlink transmission scheme with a smart combining of physical downlink control channel (PDCCH) and physical downlink shared channel (PDSCH) repetitions. The novelty in this scheme is the repetition and combining of PDCCH. Due to the concern about UE memory size and the number of blind decodes, PDCCH repetition and combining is still not standardized by 3GPP. The proposed scheme solves these concerns and enhances the performance of URLLC downlink transmission. Firstly, latency and reliability requirements of URLLC are analyzed in Section II. Section III explains in detail the scheme and the techniques to implement it. The simulation results are illustrated in Section IV. Finally, Section V is the concluding remarks.

## II. URLLC REQUIREMENTS

The requirement for URLLC is specified in [11]: "A general URLLC reliability requirement for one transmission of a packet is $10^{-5}$ for 32 bytes with a user plane latency of 1ms".

It is also noted in [11] that spectral efficiency and energy consumption also should be considered when trying to achieve a reliability target.

### A. Analysis of URLLC latency

Latency in physical layer can be expressed by the sum of 4 terms [1]:

$$T_{\mathrm{L}} = T_{\mathrm{ttt}} + T_{\mathrm{procTx}} + T_{\mathrm{prop}} + T_{\mathrm{procRx}} + T_{\mathrm{reTx}}, \tag{1}$$

- $T_{\mathrm{ttt}}$: time-to-transmit latency, is the time required to transmit a packet
- $T_{\mathrm{prop}}$: propagation latency, is the time for a signal travel from the transmitter to the receiver
- $T_{\mathrm{procTx}}$, $T_{\mathrm{procRx}}$: processing time for channel estimation, encoding and decoding of the first transmission
- $T_{\mathrm{reTx}}$: time for the retransmission (containing $T_{\mathrm{ttt}}$, $T_{\mathrm{prop}}$, $T_{\mathrm{proc}}$ and feedback time of the retransmission)

In LTE, $T_{\mathrm{ttt}}$ is fixed to 1ms while $T_{\mathrm{proc}}$ and $T_{\mathrm{reTx}}$ also contribute significantly to latency so URLLC latency requirement cannot be fulfilled.

### B. Analysis of URLLC reliability

In one-shot transmission, the error probability of a transmission is:

$$P^e = 1 - (1 - P_c^e)(1 - P_d^e), \tag{2}$$

- $P^e$ : the error probability of a transmission
- $P_c^e$ : the error probability of PDCCH
- $P_d^e$ : the error probability of PDSCH

From (2), in order to achieve probability $10^{-5}$ of URLLC , it requires that the error probabilities of PDCCH and PDSCH are required to be below $10^{-6}$. The design of control and data channel to attain this value is sophisticated.

The complexity of channel design can be relaxed by one retransmission in [9]. The successful probability is:

$$P = P_c P_{d1} + (1 - P_c) P_{DTX} P_c P_{d1} + P_c (1 - P_{d1}) P_N P_c P_{d2}, \tag{3}$$

- $P_c$: the successful probability of PDCCH
- $P_{d1}$: the successful probability of a single PDSCH
- $P_{d2}$: the successful probability of a retransmitted PDSCH
- $P_{DTX}$: the successful probability of DTX or NAK detection if no ACK/NAK is sent by the UE
- $P_N$: the successful probability of DTX or NAK detection if NAK is sent by the UE

From (3), the reliability of URLLC can be achieved with $P_c$, $P_{d1}$, $P_{d2}$, $P_{DTX}$ and $P_N$ being around 0.999. However, these values are still higher than the LTE probability of 0.99 and new techniques are necessary to enhance the performance of the channels. This work focuses on increasing of $P_c$, $P_{d1}$ and $P_{d2}$ while $P_{DTX}$ and $P_N$ are assumed to be the same as LTE.

## III. PDCCH AND PDSCH COMBINING IN DOWNLINK TRANSMISSION OF URLLC

### A. Transmission scheme

In the transmission scheme with HARQ process for repetitions as [6] and [7], the gNB transmits PDCCH and PDSCH in the downlink at the beginning of the transmission. If the UE fails to decode PDCCH, PDSCH also cannot be decoded because its location is unknown to the UE. Thereby, the gNB will not receive ACK/NACK for PDSCH from the UE. After detecting the missing HARQ response, the gNB knows that PDCCH is not detected by the UE and reschedules both PDCCH and PDSCH. It causes very sub-optimal resource usage because both control and data need to be transmitted again with a lower code rate. Moreover, it leads to increase of latency when the UE needs to wait for the full retransmission.

The retransmission of complete PDCCH and PDSCH as if it is a fresh transmission can be avoided by repeating PDCCH and PDSCH partially and combining these multiple transmitted PDCCHs and PDSCHs. In this scheme as shown in Fig. 1, for the original transmission, the gNB transmits PDCCH (C1) and PDSCH (D1). When no HARQ response is received at the gNB for PDSCH at the expected time, it retransmits special versions of PDCCH and PDSCH that have shorter length in the next transmission occasion while the UE stores the current transmission occasion (containing C1 and D1). The two transmissions of PDCCH (C1 and C2) are designed such that they can be combined for joint detection (explained in III-B). Additionally to this, the combined PDCCH enables the UE to locate the two transmissions of PDSCH, the original transmission (D1) and the retransmission (D2) in order to combine them. This allows the network to enhance the reliability of both PDCCH and PDSCH for URLLC users in a very timely manner.

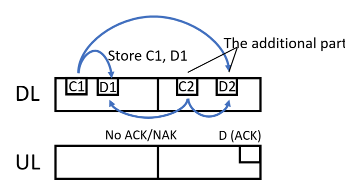



Fig. 1. Downlink transmission with PDCCH and PDSCH comnination.

To allow the combination of PDCCHs, the content of DCI on PDCCH must be the same for both the initial transmission and the retransmission. One way is to make the initial and retransmitted PDCCHs indicate the resource allocation information for both two transmissions as illustrated in Fig. 1. This does not necessarily mean that all the fields in the DCI relevant to scheduling of resources, precoding, layers, MCS etc are doubled up. The resources for the retransmission PDSCH (D2) can be pre-defined as a function of resources indicated for the original transmission (D1). The other transmission parameters, like MCS, RV etc can remain the same.

In case the UE can decode PDCCH (C1) but encounters an error in decoding PDSCH (D1), PDSCH (D2) can be retransmitted independently in the next transmission occasion without a request of another PDCCH because the resource allocation information of this repetition is known by the UE by decoding the initial PDCCH. After that, the UE combines these two transmissions to decode data. It avoids the blockage of PDSCH repetition due to the shortage of resource for PDCCH.

There is no retransmission if the UE is able to correctly decode D1 based upon C1 and sends ACK to the gNB.

The operations of the gNB and the UE in different scenarios are shown by the flow charts in Fig. 2 and Fig. 3.

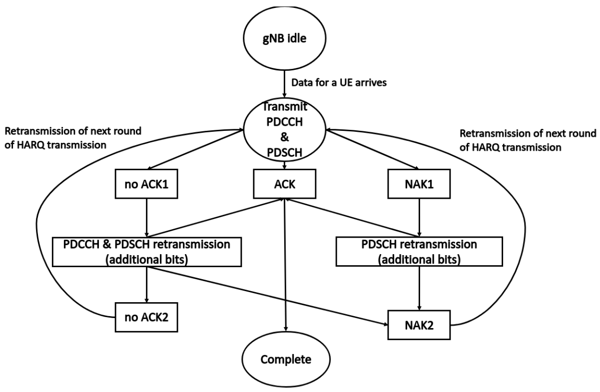



Fig. 2. The gNB's view of the transmission process.

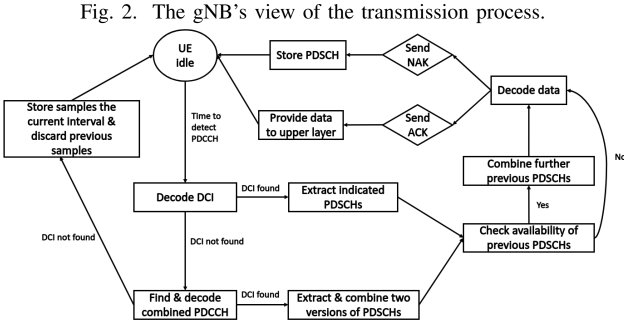



Fig. 3. The UE's view of the transmission process.

When the first DCI on PDCCH (C1) indicates locations of data transmissions on PDSCHs (D1 and D2) in the current slot and the next transmission slot, it means that the resource for the data retransmission is configured. It can lead to a waste of resources in case PDCCH and PDSCH are decoded successfully by the UE and there is no need for a retransmission. Therefore, this configured resource can be shared by a group of UEs and dynamically allocated to the UE needing a retransmission as shown in Fig. 4. There is a need to retransmit data for the UE2 when it returns NACK while the other UEs return ACK. Therefore, the gNB retransmits data for the UE2 in $D_{rep}$. The starting time of the initial transmissions is not required to be synchronized, the gNB only needs to receive HARQ feedback from the UEs before the starting time of the configured resource for the retransmission.

The same principle can be applied to PDCCH repetition. A resource in the next transmission occasion is allocated to PDCCH repetition of a group of the UEs. If the first PDCCH is decoded and the retransmission is not necessary, the configured resource will be allocated to PDCCH of other UEs in the group.

One disadvantage of this method is that it only allows one UE with the first coming NACK an opportunity to receive

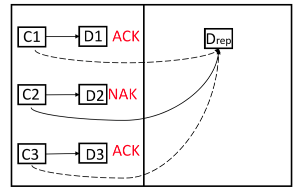



Fig. 4. Reserved resource for PDSCH repetitions.

a second transmission during $D_{rep}$. Other UEs reporting a NACK are blocked from receiving a second transmission. However, due to a high reliability of URLLC transmission, the probability that two or more UEs encounter the errors at the same time is small (below $10^{-4}$) so this method allows a good use of the configured resource.

## *B. Design of PDCCH and PDSCH repetions*

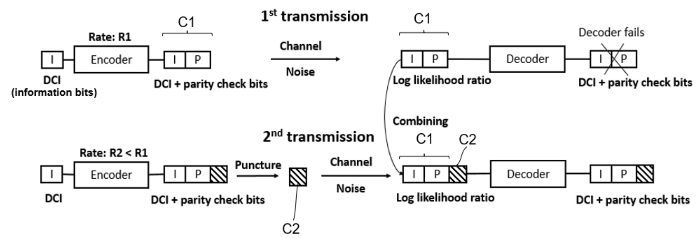



Fig. 5. Processing a DCI to form and combine two PDCCHs.

Fig. 5 shows in detail the contents of the multiple transmissions on PDCCH. When the transmission starts, the gNB encodes the control signal (DCI) using Polar code with the code rate $R_1$ to generate PDCCH containing DCI and parity check bits. PDCCH is transmitted from the gNB to the UE. The UE calculates the log likelihood ratios (LLRs) of the incoming codeword. The UE attempts to decode the incoming codeword and if the decoder fails to find the correct codeword, the UE stores the current slot containing PDCCH. This allows the UE to combine with PDCCH repetition in the next transmission occasion, in case the gNB retransmits the DCI. The buffered slot also comprises PDSCH so that the UE can retrieve the data on PDSCH if the PDCCH repetition is decoded successfully.

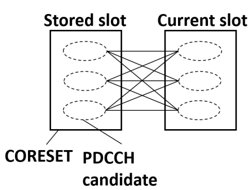
On the transmitter side, after not receiving ACK or NACK of the initial transmission from the UE, the gNB initiates the second transmission of DCI that has the same content with the first DCI because both DCIs indicate the same PDSCHs. In this second transmission, a lower code rate $R_2$ ($R_2 < R_1$) is used in order to increase the reliability of PDCCH. There will be an overlap between the encoded bits at code rates $R_1$ and $R_2$. The bits that have already been transmitted in the first transmission are punctured from the second transmission. This leaves only the additional bits due to the lower code rate are transmitted to the UE. The UE receives the second transmission and combines the additional bits (C2) with the stored bits (C1) of the first transmission in Fig. 5 to build a codeword with lower code rate $R_2$. Due to the lower code rate, BLER for the combination ($P_c^e$ in (2)) is smaller than the BLER of the first transmission so the UE has a better chance to decode the transmitted DCI.

The content of DCI is the same for the initial transmission and the retransmission so the gNB generates a mother codeword of PDCCH with low rate. It selects suitable bits to be transmitted in the original PDCCH transmission and the repetition by applying puncturing. In this way, it does not need to encode PDCCH multiple times for multiple repetitions.

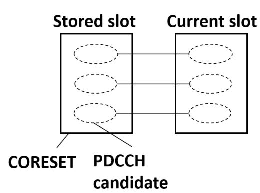
The same principle is applied to PDSCH repetition to reduce resource consumption and increase reliability.

### *C. Containing the complexity of blind decode for PDCCH combining*

The scheme in Section III-A and III-B requires the UE to combine multiple PDCCHs from different control occasions. In principle, the location of PDCCH is unknown to a UE so each UE must perform blind decode until it finds a DCI that passes the CRC. In the PDCCH combining, the complexity of blind decode becomes more severe when the number of blind decodes grows remarkably. The UE monitors the current control resource to determine by blind decoding if a new DCI for the UE is received in the slot. If the UE fails to detect a PDCCH intended for it, it combines the current monitored occasion with the previous stored occasion for a potential combining. The number of stored occasion is limited to 1 to reduce memory and energy consumption. Fig. 6(a) shows this results in a large number of blind decode combinations of the different PDCCH candidates.



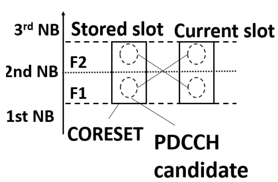

(a) No mapping.



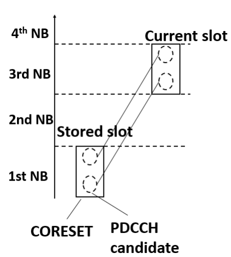

(b) One-to-one mapping.

Fig. 6. Combining of blind decode candidates.

The number of blind decodes in PDCCH combining decreases dramatically if the gNB is configured to transmit PDCCH and the repetition in the candidates with a one-to-one correspondence mapping as shown in Fig. 6(b). The total number of combinations is significantly lower than that of Fig. 6(a). Other mapping possibilities are possible. The mapping rules may be defined a-priori in standards specifications or may be configured between the gNB and the UE.




(a) Same band's CORESETs.



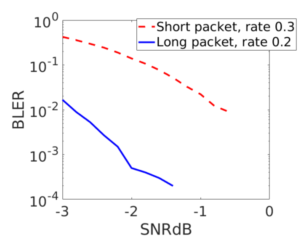

(b) Different bands' CORESETs.

Fig. 7. Frequency hopping in PDCCH combining.

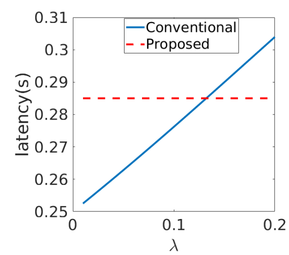
Based on the same principle of one-to-one correspondence of PDCCH combining, the improvement of URLLC's service quality through interference cancellation and frequency diversity in PDCCH transmission can be achieved by frequency hopping. In one method, PDCCH candidates can be in different CORESETs in the same band but these candidates are hopped over different frequencies and have an association between locations in the two occasions, such as a one-to-one correspondence between the candidates in Fig. 7(a). In another method to allow additional frequency diversity advantage in PDCCH combining, the initial transmission and repetition are allocated to the disjoint frequency bands with large frequency distance. As NR does not allow a single CORESET to change its frequency allocation in time, one way to achieve this is that the gNB configures two CORESETs at disjoint frequency locations as illustrated in Fig. 7(b).

## IV. Simulation results

The simulation is performed in link level. Polar code is used to encode the input codeword (DCI) with 48 bits consisting of 24 information bits and 24 CRC bits. The encoding mechanism of the encoder follows 3GPP standard [13]. The code rates 0.3 and 0.2 are used in different scenarios. The output codeword is modulated in QPSK and transmitted over AWGN channel. The decoder calculates LLRs and uses them to decode the receiving signal. The decoder is min-sum Successive cancellation list (SCL) decoder with list size 8 [2].




(a) BLER of conventional and proposed schemes.




(b) Latency in 2-UE system.

Fig. 8. Simulation results.

Fig. 8(a) shows a comparison between the initial transmissions of two schemes. The first scheme is a conventional one where a long packet with low rate 0.2 containing 240 bits is used for the initial transmission. The second one is the proposed scheme where a short packet with higher rate 0.3 containing 160 bits is used in the initial transmission. Subsequently, if the initial packet is not decoded correctly, the gNB will only sends 80 additional bits so that the UE can combine these 80 bits and the initial stored 160 bits to create a packet with 240 bits as the conventional scheme.

As can be seen at SNR -2dB, BLERs of the conventional scheme with long packet and the combining scheme with short packet are 0.001 and 0.14, respectively. This means that in 1000 times, the conventional scheme succeeds 999 times while the combining scheme succeeds 860 times without a need of retransmisison and combining. Thereby, by using the combining scheme, the system can save time and frequency resources in 860 times with 80 bits equal to 40 REs in QPSK modulation saved each time so blocking probability decreases.

TABLE I

RESOURCE AND LATENCY COMPARISON OVER 1000 PACKET TRANSMITTED OVER AWGN CHANNEL WITH SNR -2$dB$ AND SCS 60KHZ

| | $1^{st}$ **transmission's success** | | | $2^{nd}$ **transmission's success** | | | **Average expected resource (REs)** | **Average expected latency with 1-UE system (ms)** |
|---|---|---|---|---|---|---|---|---|
| | **Percent** | **Res (REs)** | **Lat (ms)** | **Percent** | **Res (REs)** | **Lat (ms)** | | |
| Conventional | 0.999 | 120 | 0.25 | 0 | 0 | 0 | 120 | 0.250 |
| Prior art [6] and [7] | 0.860 | 80 | 0.25 | 0.139 | 120 | 0.5 | 96.70 | 0.285 |
| Proposed scheme | 0.860 | 80 | 0.25 | 0.139 | 40 | 0.5 | 85.57 | 0.285 |

Fig. 8(b) shows latency when the proposed scheme is used in a system with 2 UEs and resource pool (CORESET) of 160 REs in each slot for PDCCH. The arrival of data for each UE follows a Poisson process with the average number of random access events in a slot $\lambda$. The probability that one UE has one or more random transmissions is: $P_{data} = 1 - e^{-\lambda}$. In the conventional scheme, a packet on PDCCH has size of 120 REs so only PDCCH of a UE can be transmitted in CORESET of a slot and PDCCH of the other UE must wait for the next slots. With SCS of 60kHZ, the average expected latency of a UE in this system is:

$$\begin{aligned} E[T] = &0.25(1-P_{data}) + 0.5P_{data}(1-P_{data}) \\ &+ 0.75P_{data}^2(1-P_{data}) + ... \end{aligned} \tag{4}$$

In (4), the first term presents the case that the gNB can transmit PDCCH to the UE of interest and succeeds in the first slot. The subsequent terms are when the gNB must wait one or several slots before transmitting to the UE of interest.

In contrast, a packet's size on PDCCH is only 80 REs in the proposed scheme so 2 UEs can transmit at the same time in a CORESET of a slot. Therefore, when the arrival rate of data is high as can be seen in Fig. 8(b) at $\lambda$ bigger than 0.14, the proposed scheme helps to reduce latency because two UEs can transmit at the same rather than one UE waiting for the next slots and the blocking probability decreases.

The proposed scheme also brings an efficiency of resource consumption compared to other schemes applying retransmission and combining in [6] and [7]. In the proposed scheme, in total there are 240 bits transmitted in case of the initial transmission's failure (140 times in 1000 times). While in other schemes with conventional HARQ process, an entire packet with lower rate 0.2 containing 240 bits is transmitted in the retransmission after the initial packet's failure with rate 0.3 and 160 bits. In total, there are 400 bits transmitted.

In 140 times, the conventional scheme succeeds in the first transmission while the combining scheme fails and a retransmission is required so that the packets can be combined to have rate 0.2 as the conventional scheme and achieve target reliability. It leads to latency in the proposed scheme. HARQ feedback consumes 1 OFDM symbol. If PDCCH is not decoded, the gNB waits after the normal time for an appearance of feedback to decide PDCCH failure and retransmits the additional bits of both PDCCH and PDSCH. These PDCCH and PDSCH are short versions as mentioned in Section III-B and only consumes 3 OFDM symbols. The maximum processing time of the decoder to decode the combined PDCCH and PDSCH is 0.125ms equivalent to 7 OFDM symbols with SCS 60kHz (one slot with SCS 60kHz has 14 OFDM symbols spreading in 0.25ms). Therefore, latency in 140 times over 1000 times of the proposed scheme is at least 11 OFDM symbols more than the conventional schemes with 1 UE. This amount is approximate to 1 slot with 0.25 ms that is still acceptable to achieve the latency requirement 1ms of URLLC.

Table I summarizes the above calculation and shows the expected resource consumption as well as latency over 1000 packets of three schemes at SNR of -2dB and SCS of 60kHz to achieve the same reliability.

## V. CONCLUSION

A downlink transmission scheme applying an intelligent combining based approach to get improved PDCCH and PDSCH detection while reducing latency and resource consumption is presented. The designs of PDCCH and PDSCH and the techniques to implement the combining are explained.

## ACKNOWLEDGMENTS

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

## REFERENCES

[1] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra Reliable and Low Latency Communications in 5G Downlink: Physical Layer Aspects," IEEE Wireless Commun, June 2018.

[2] A. Balatsoukas-Stimming, M. Bastani Parizi, and A. Burg, "LLR-Based Successive Cancellation List Decoding of Polar Codes," IEEE Trans. Signal Process., vol. 63, no. 19, pp. 51655179, Oct 2015.

[3] R. Abreu, P. Mogensen and K. I. Pedersen, "Pre-Scheduled Resources for Retransmissions in Ultra-Reliable and Low Latency Communication," IEEE Wireless Communications and Networking Conference (WCNC), March, 2017.

[4] Intel Corporation, "Ultra-reliability for NR PDCCH", 3GPP R1-1717381, RAN1#90bis, Prague, Czech Republic, Oct 9–13, 2017.

[5] Huawei, HiSilicon, "PDCCH reliability for URLLC", 3GPP R1-1719406, RAN1#91, Reno, USA, Nov 27–Dec1, 2017.

[6] MediaTek Inc, "HARQ design and PDCCH reliability for URLLC", 3GPP R1-1716620, RAN1 Ad Hoc#3, Nagoya, Japan, Sept 18–21, 2017.

[7] Sony, "Reliable PDCCH operation for NR", 3GPP R1-1716248, RAN1 Ad Hoc#3, Nagoya, Japan, Sept 18–21, 2017.

[8] Sequans Communications, "On PDCCH repetition for NR URLLC", 3GPP R1-1802534, RAN1#92, Athens, Greece, Feb 26–March 2, 2018.

[9] ZTE, ZTE Microelectronics, "Analysis of URLLC reliability for HARQ", 3GPP R1-1701595, RAN1#88, Athens, Greece, Feb 13–17, 2017.

[10] 3GPP TR 38.802 v14.1.0, "Study on new radio access technology physical layer aspects."

[11] 3GPP TR 38.913 v14.3.0, "Study on scenarios and requirements for next generation access technologies."

[12] 3GPP TR 38.211 v15.0.0, "Physical channels and modulation."

[13] 3GPP TR 38.212 v15.0.0, "Multiplexing and channel coding."