Reduced CBG HARQ Feedback for Efficient Multimedia Transmissions in 5G for Coexistence with URLLC Traffic

Baris Göktepe, Thomas Fehrenbach, Thomas Schierl and Cornelius Hellge

May 24, 2018

Reduced CBG HARQ Feedback for Efficient Multimedia Transmissions in 5G for Coexistence with URLLC Traffic

Barış Göktepe, Thomas Fehrenbach, Thomas Schierl and Cornelius Hellge
Fraunhofer Heinrich Hertz Institute
Einsteinufer 37, 10587 Berlin
Email: [baris.goektepe, thomas.fehrenbach, thomas.schierl, cornelius.hellge]@hhi.fraunhofer.de

***Abstract*—In recent 5G standardization activities, finely granular HARQ feedback (CBG HARQ) for partial retransmissions was introduced to support high data rates, while enabling coexistence with ultra-reliable low latency communication (URLLC). However, the CBG HARQ scheme comes with the drawback of increased feedback payload. In this paper, we discuss two reduced HARQ feedback schemes for using partial retransmissions, as used in CBG HARQ, while decreasing the feedback overhead implicated by finely granular feedback. These schemes exploit information on the location of puncturing by URLLC traffic, designated as Preemption Indication (PI). In this work, we show that these schemes achieve a similar performance to the regular CBG HARQ scheme, while significantly lowering the feedback overhead burden on the communication system.**

I. Introduction

In the last decade, the requirements on wireless mobile networks have reached a new level of diversification. With emerging new use cases, such as Ultra-Reliable Low Latency Communication (URLLC) and massive Machine Type Communications (mMTC), the focus in Fifth Generation (5G) has moved from only increasing data rates to enabling a variety of applications with utmost differing requirements. The IMT-2020 requirements foresee three principal dimensions of performance [1], [2]. These include supporting massive amounts of densely distributed devices, extreme low-latencies, and very high data rates. At the same time, multimedia transmissions have increased dramatically and now account for the largest part of cellular traffic, which is covered by the enhanced Mobile BroadBand (eMBB) use case in 5G standardization. The eMBB use case requires enabling high data rates and serving many users per cell [3]. However, due to the bandwidth shortage, multimedia traffic has to share the same resources with other traffic types, such as latency critical traffic, designated as URLLC. This is a challenge for the design of a cellular communication system since both use cases come with contradicting requirements [4]. Whereas high data rates and efficient transmissions play a key role for multimedia transmissions, ultra-high reliability and low latency in the sub-millisecond range characterize latency critical traffic [2]. The efficient coexistence of both traffic types is a key parameter of 5G and is still an important topic in the standardization progress in 3rd Generation Partnership Project (3GPP).
Section II provides an overview on the state-of-the-art feedback scheme adopted by 5G standardization. Section III introduces two new improved Code Block Group (CBG) Hybrid Automatic Repeat reQuest (HARQ) feedback schemes and discusses their advantages and drawbacks compared to state-of-the-art schemes. Section IV presents the simulation environment and assumptions. Then, in Section V the results of the simulations are discussed and evaluated.

II. State-of-the-art Coexistence of eMBB and URLLC Traffic

Coexistence issues are a highly discussed topic in 5G standardization [4]. Overprovisioning solutions have been considered, which pre-reserve resources for sporadically occurring URLLC transmissions. However, these solutions come at the high cost of unused resources and do not provide an efficient way of solving the coexistence issues. Thus, preemption of eMBB transmissions in downlink by URLLC traffic was introduced [5]. Preemption is performed by partly puncturing an ongoing eMBB transmission with URLLC traffic. This often causes a failure of the eMBB transmission since the User Equipment (UE) is not aware of the puncturing. The package sizes of eMBB transmissions are expected to be relatively large compared to the URLLC traffic, thus preemption affects only a part of the eMBB transmission. However in Long Term Evolution (LTE), even the failure of a single codeword causes a retransmission of the whole block since feedback is provided only for the Transport Block (TB) [6]. This drawback

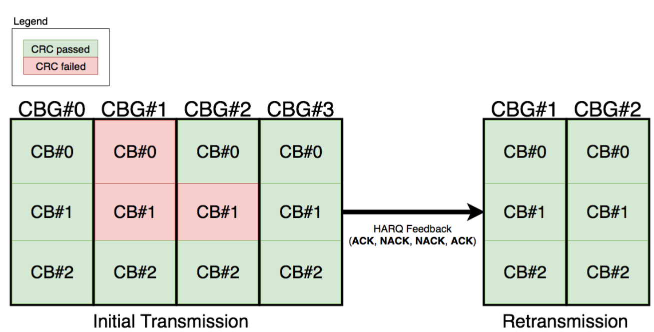

Fig. 1. Illustration of CBG HARQ feedback.

is solved in 5G by introducing so-called CBGs, consisting each of at least one Code Block (CB), to provide a finely granular feedback for eMBB transmissions, as depicted in Fig. 1. CBG feedback enables an efficient retransmission of only the corrupted CBGs [7]. Nevertheless, as illustrated in Fig. 1, some correctly received CBs may be retransmitted unnecessarily. Still, the size of the feedback payload grows significantly compared to regular 1-bit feedback putting burden on transmitting the feedback, especially in the uplink, the required Signal-to-Noise Ratio (SNR) for transmitting the feedback in the short Physical Uplink Control Channel (PUCCH) format increases for 10-bits CBG instead of 1-bit TB feedback. For cell edge UEs, this is a challenge, which only is solved by reserving more resources and using a lower code rate in the PUCCH region (long PUCCH format) for transmitting the feedback. Apart from the high resource usage of this approach, the multiplexing capabilities are reduced [8], which is a critical requirement for supporting a large number of users.

III. Reduced CBG HARQ Feedback

In this paper, we discuss a scheme to reduce the amount of feedback in uplink by exploiting information on the location of the puncturing. The transmission of this information is already a part of the 5G specification in form of the, so-called, Preemption Indication (PI). However, the PI, as it is foreseen in the current specification, is only used for the decoding process to avoid misinterpretation in the decoder. This information can be used as a basis for reduced HARQ feedback with finely granular retransmissions, since it represents the common understanding of the Base Station (BS) and the UE. The transmission of a TB, which is divided into independently decodable CBGs, is designed to achieve a target Block Error Rate (BLER) of approximately 3 - 10%. Preemption increases the decoding failure probability for parts of the transmission significantly and thus also the failure probability of the whole TB. Since the retransmission of the affected parts by preemption is required with high probability, it is unnecessary that the UE reports these as corrupted to the BS, which is always aware of the preempted parts. The finely granular feedback only provides in the case of an independent failure, a more efficient retransmission that equals to approx. 3 - 10 % depending on the designed target BLER.

A. Fixed-size Feedback

In this section, a fixed size 3-state feedback, designated as Reduced Code Block Group (R-CBG), based on the PI for CBG retransmissions, is discussed. A fixed-size feedback has the advantage of enabling simple mechanisms to multiplex the feedback. Also in the case of feedback bundling, splitting the feedback information of the different transmissions is accomplished very easily. The meaning of the states is described as follows:

1) ACK - all CBGs have been received correctly
   - No retransmission required
2) Partial ACK - all CBGs have been received correctly except the CBG/CBGs affected by puncturing
   - CBGs affected by puncturing are retransmitted
3) NACK - other CBGs have failed
   - The BS retransmits all CBGs of the TB

After receiving the PI, which includes the puncturing information, the UE generates the feedback for previous transmissions and transmits these on pre-assigned resources to the BS. This avoids redundant information in feedback signaling. Depending on this feedback the BS ends the transmission, and retransmits the affected CBGs or retransmits the whole TB. Since multimedia traffic is much more delay tolerant than URLLC traffic, the additional delay does not have a critical impact. In case the UE misses the PI, it uses the first and third state (ACK and NACK) only. This avoids misalignment between the UE and the BS. Fixed-size feedback increases the coverage and the energy efficiency of the UEs, since less feedback is transmitted.

B. Adaptive-size Feedback

The fixed-size feedback scheme comes with the drawback of retransmitting more than necessary, if some but not all non-punctured CBGs fail at the receiver. This flaw is overcome by utilizing the NACK state for assigning a new PUCCH resource for the detailed CBG HARQ feedback, which is called Adaptive Reduced Code Block Group (AR-CBG) in the following. This brings the advantage of operating as efficiently as full CBG feedback in terms of retransmissions but increases the overhead caused by feedback. The states of the feedback are described as follows:

1) ACK - all CBGs have been received correctly
   - No retransmission required
2) Partial ACK - all CBGs have been received correctly except the CBG/CBGs affected by puncturing
   - CBGs affected by puncturing are retransmitted
3) NACK - other CBGs have failed
   - The BS assigns a PUCCH resource for transmitting detailed CBG HARQ feedback

In case the UE misses the PI, it falls again back on the first and third state, such that in case an ACK is sent, the BS ends the transmission and in case of a NACK, it assigns a resource for detailed CBG feedback.

IV. Simulation Setup

For evaluating the performance of the described schemes compared to TB and CBG HARQ feedback, 3GPP-compliant link level simulations have been performed over a range of Modulation and Coding Scheme (MCS) levels. Each CBG spans over 14 Orthogonal Frequency Division Multiplexing (OFDM) symbols corresponding to 1 ms in time. One transmission (TB) is divided into four CBGs, each consisting of one CB. Depending on the the data size including a 24-bit Cyclic Redundancy Check (CRC) either Low-Density Parity-Check (LDPC) Base Graph 1 (BG1) for sizes larger than 3840 bits or LDPC Base Graph 2 (BG2) for smaller ones from [9] is used for the encoding and decoding. The mapping of CBGs in the time-frequency space is depicted in Fig. 2. URLLC

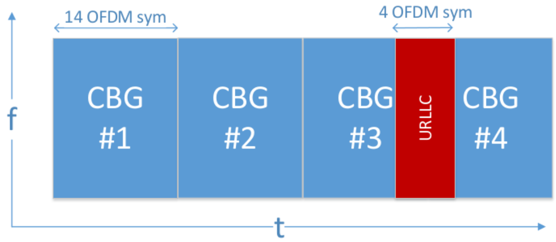

Fig. 2. Mapping of CBGs and URLLC puncturing.

puncturing is assumed to be four consecutive OFDM symbols over the whole frequency range and random in the time range, where it can also span over CBG boundaries. The PI includes only the information which CBG/CBGs are affected by the puncturing regardless of the amount. Thus, a CBG is marked as potentially corrupted also if the puncturing only affects a small part of it. The simulations parameters are summarized in Table I.

The whole MCS range foreseen for eMBB is used for the simulations. Since the size of the allocated resources is assumed to be fixed, the packet size is varying over the MCS level. Table II shows the corresponding configurations of the MCS levels and the resulting spectral efficiencies. The feedback transmission is assumed to be ideal without any errors, which is valid if the transmission is sufficiently reliable. Naturally, the coding overhead is larger if smaller packets are transmitted. However, considering feedback bundling, the feedback data can be transmitted in groups such that the packet size does not depend on the size of a single feedback packet.

TABLE I
LINK-LEVEL SIMULATION PARAMETERS

Channel type	1Tx 1Rx, TDL-C 100ns, 2.9 GHz, 3.0 km/h
Waveform	3GPP OFDM, 1.4 MHz, normal cyclic-prefix
Modulation	Quadrature Amplitude Modulation (QAM)
Equalizer	Frequency domain MMSE
MCS levels	0 - 27
Number of CBGs	4
Puncturing	4 consecutive OFDM symbols
Channel Code	LDPC
Parity-check matrix	see BG1 and BG2 in [9]
Decoder type	Min-Sum
Decoding iterations	50

V. RESULTS

A. Target BLER

In this section, we present the link-level simulation results of the adaptive-size AR-CBG and fixed-size R-CBG feedback compared to full CBG HARQ feedback. As mentioned previously, the switching SNRs are determined by the target BLER for the first transmission. Naturally, only unpunctured transmissions are considered for the target BLER. Fig. 3 shows

TABLE II
MCS LEVEL PARAMETERS [10]

MCS	Modulation order	Code rate	Spectral efficiency in bit/symbol
0	4	0.1172	0.2344
1	4	0.1885	0.3770
2	4	0.3008	0.6016
3	4	0.4385	0.8770
4	4	0.5879	1.1758
5	16	0.3691	1.4764
6	16	0.4238	1.6952
7	16	0.4785	1.9140
8	16	0.5400	2.1600
9	16	0.6016	2.4064
10	16	0.6426	2.5704
11	64	0.4551	2.7306
12	64	0.5049	3.0294
13	64	0.5537	3.3222
14	64	0.6016	3.6096
15	64	0.6504	3.9024
16	64	0.7021	4.2126
17	64	0.7539	4.5234
18	64	0.8027	4.8162
19	64	0.8525	5.1150
20	256	0.6665	5.3320
21	256	0.6943	5.5544
22	256	0.7363	5.8904
23	256	0.7783	6.2264
24	256	0.8213	6.5703
25	256	0.8643	6.9141
26	256	0.8950	7.1602
27	256	0.9258	7.4062

the initial BLER of unpunctured CBs for the MCS range. As expected, higher MCS require a higher SNR. Small differences in the waterfall curve behavior are noted for the switching points of different modulation orders. Especially, the change from MCS 19 to 20, which corresponds to a change from a 64-QAM to a 256-QAM is notable.

The switching points are chosen at the SNR at which the next higher MCS achieves a BLER of 10^{-1}. This ensures that the working range for the target BLER is always less than 10 %. The results are summarized in Table III. The MCS levels 0 - 27 cover the SNR range from -3.75 dB to approx. 30 dB. As visible in Fig. 3 and Table III, the SNR differences between the switching points become smaller with increasing MCS. This effect is explained by the growing packet size which results in increased codeword length. Due to this effect in combination with statistical inaccuracy MCS 22 has a slightly higher switching SNR than MCS 23.

B. Gain Evaluation

An important key metric for evaluating enhancements of CBG HARQ feedback is the gain over TB feedback, which was the motivation for introducing CBG feedback. The gain is defined as the relative amount of CBG retransmissions, which are saved by the extended feedback. TB feedback always retransmits four CBGs if any error occurs. The other feedback schemes (CBG, AR-CBG and R-CBG) avoid some unnecessary retransmissions. Hence, the gain is stated as

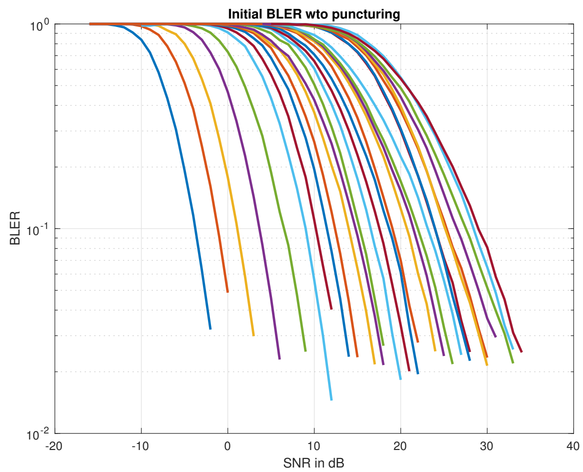

Fig. 3. Initial BLER over SNR in TDL-C channel. Markers: black - 4-QAM, red - 16-QAM, blue - 64-QAM, green - 256-QAM.

TABLE III
SWITCHING SNRS FOR MCS LEVELS

MCS	SNR in dB	MCS	SNR in dB
0	-3.75	14	17.58
1	-1.15	15	18.66
2	1.1	16	19.05
3	3.67	17	20.76
4	6.48	18	21.46
5	9.0	19	23.13
6	10.18	20	23.96
7	11.47	21	23.97
8	12.41	22	25.28
9	13.98	23	25.14
10	14.77	24	26.51
11	14.79	25	26.51
12	15.27	26	27.56
13	16.47	27	28.96

follows:

$$g = \frac{N_{CBG,Scheme}}{N_{CBG,TB}}, \tag{1}$$

where $N_{CBG,Scheme}$ is the number of CBGs retransmitted by the corresponding scheme and $N_{CBG,TB}$ is the number of CBGs retransmitted by the TB feedback scheme. Fig. 4 shows the gain of full CBG, AR-CBG and R-CBG feedback in means of CBG retransmissions. Discontinuities in the gain are due to the switching to the next MCS level. It is notable that AR-CBG achieves the same gain as the full CBG feedback. This is easily explained by the fact that AR-CBG uses the full CBG feedback if required and otherwise stays with the reduced feedback. R-CBG feedback achieves a slightly less gain over the TB feedback. Still, the gain of R-CBG feedback is in the range of full CBG and AR-CBG feedback. However, it is visible that the gain of CBG feedback is less than 20 % anyway for the 4-QAM MCSs (0 - 4). Another observable effect are the discontinuities of the overall behavior, between MCS 4 and 5, 10 and 11, 19 and 20. Looking at Table II reveals that this phenomenon arises because of the changing of the modulation order, i.e., 4-QAM to 16-QAM, 16-QAM to 64-QAM and 64-QAM to 256-QAM.

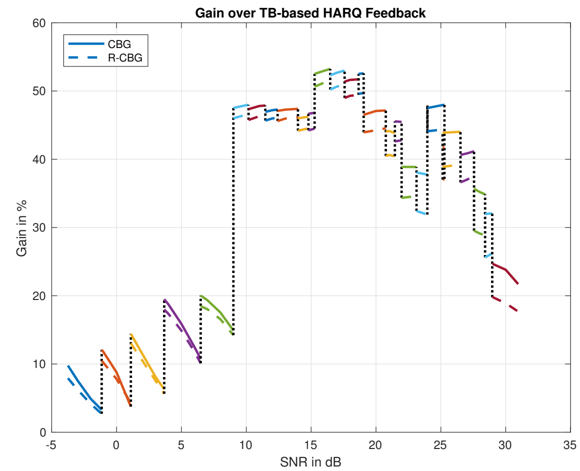

Fig. 4. Gain over TB-based HARQ feedback in means of reduced retransmissions. Solid line - regular CBG and AR-CBG, dashed line - R-CBG.

C. Feedback Overhead

As discussed in Section III, the better performance of CBG and AR-CBG feedback comes at the cost of increased feedback. Whereas CBG feedback has a fixed overhead as it is with R-CBG feedback, the overhead of AR-CBG feedback is varying based on the channel conditions. The feedback overhead compared to the TB-based feedback is shown in Fig. 5. Notably, fixed-size R-CBG feedback has the lowest overhead, i.e., 1.5 bits instead of 1 bit per HARQ feedback. The AR-CBG feedback has a slightly higher but comparable overhead relative to the R-CBG feedback, whereas regular CBG feedback has a significantly higher overhead.

D. Average Performance

Under the simulation assumptions the packet size varies over the MCS levels, the SNR gap between the switching points is decreasing with higher MCS levels. However, the MCS levels given a fixed packet size are designed to operate between in the target BLER region between 3 % and 10 %. Hence, the average gain of the schemes over the whole relevant SNR regions has been evaluated. Since the AR-CBG scheme achieves the same gain as the CBG scheme, only the latter one has been used for comparison. In Fig. 6 the absolute and relative gain loss of R-CBG compared to CBG is presented. Although the absolute loss is increasing with the MCS level, the relative loss has two peaks at high and low MCS levels and achieves a minimum between MCS 5 and MCS 10. Nevertheless, the absolute loss is less than 6 % and the relative loss less than 18 % over the whole range.

Fig. 5 shows the average feedback overhead of the HARQ feedback schemes over the TB-based feedback for the relevant SNR regions (initial BLER between 3 % and 10 %). As

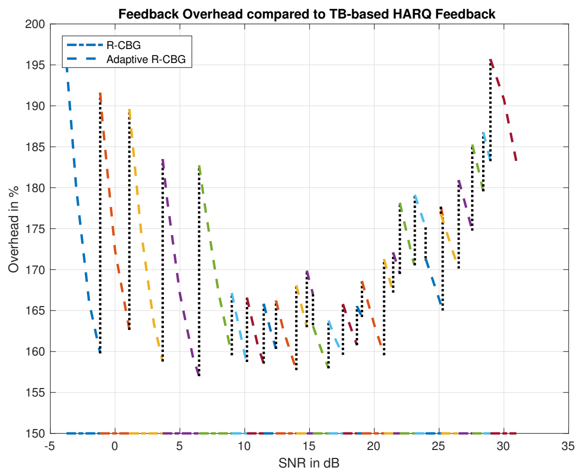

Fig. 5. Feedback overhead compared to TB-based HARQ feedback in TDL-C channel. Solid line - regular CBG, dotted line - AR-CBG, dashed line - R-CBG.

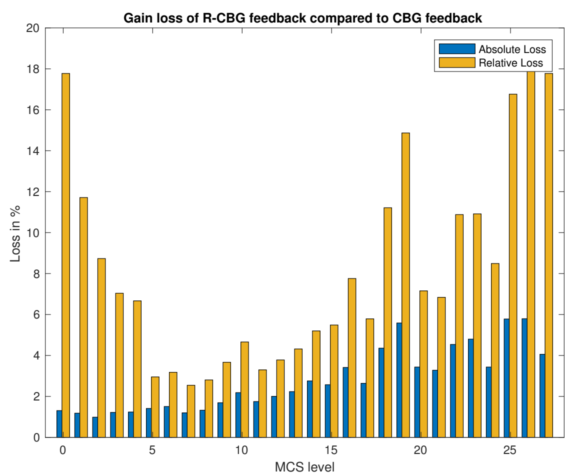

Fig. 6. Average gain loss of R-CBG feedback compared to CBG HARQ feedback in TDL-C channel.

described in Section III, the feedback overhead of CBG and R-CBG feedback schemes is constant for all MCS levels. The AR-CBG feedback scheme has a slightly higher overhead than the R-CBG feedback. Compared to CBG feedback, AR-CBG feedback reduces the feedback overhead significantly from 400 % to around 170 %. There is again a slight increase in feedback overhead for very low and very high MCS levels to approx. 180 %.

VI. Conclusions

In this work, we present two new feedback schemes, R-CBG and AR-CBG, and show that they provide significant improvements on the current CBG HARQ feedback scheme. Whereas R-CBG has the advantage of fixed-size feedback, which simplifies the multiplexing of the feedback and has only a slight increased feedback overhead compared to TB

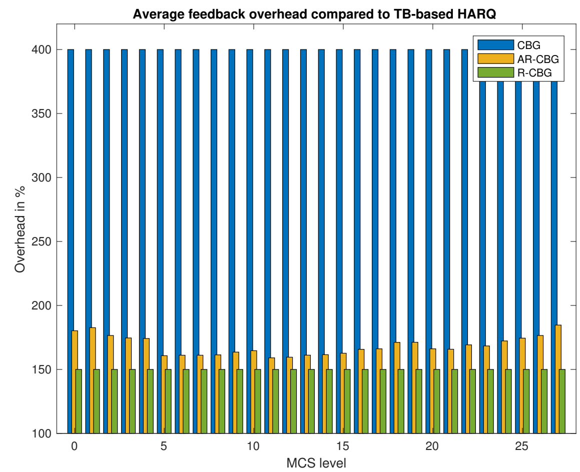

Fig. 7. Average feedback overhead compared to TB-based HARQ feedback in TDL-C channel.

feedback, it performs slightly less efficiently than CBG HARQ feedback. In contrast, AR-CBG feedback achieves the same performance of CBG feedback with a slightly increased feedback overhead compared to the fixed-size R-CBG scheme. However, performance of the presented schemes also depends on the amount and location of the puncturing, as well as the channel model. Further studies on how these parameters affect the performance of the evaluated schemes have to be performed.

References

[1] ITU, "IMT Vision Framework and overall objectives of the future development of IMT for 2020 and beyond ," tech. rep., ITU, 2015.

[2] R. El-Hattachi and J. Erfanian, "NGMN 5G White Paper," tech. rep., Next Generation Mobile Networks (NGMN), 02 2015.

[3] 3GPP SA1, "Feasibility Study on New Services and Markets Technology Enablers - Enhanced Mobile Broadband," Tech. Rep. TR 22.863 v2.0.0, 3GPP, 06 2016.

[4] C.-P. Li, J. Jiang, W. Chen, T. Ji, and J. Smee, "5g ultra-reliable and low-latency systems design," in *2017 European Conference on Networks and Communications (EuCNC)*, pp. 1–5, June 2017.

[5] P. Popovski, J. J. Nielsen, C. Stefanovic, E. de Carvalho, E. G. Ström, K. F. Trillingsgaard, A. Bana, D. Kim, R. Kotaba, J. Park, and R. B. Sørensen, "Ultra-reliable low-latency communication (URLLC): principles and building blocks," *CoRR*, vol. abs/1708.07862, 2017.

[6] L. Jiang and F. Jiang, "A comprehensive and practical model for harq in lte system," in *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 901–905, April 2013.

[7] S. R. Khosravirad, L. Mudolo, and K. I. Pedersen, "Flexible multi-bit feedback design for harq operation of large-size data packets in 5g," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, June 2017.

[8] Fraunhofer HHI, "On CBG-based (re-)transmission," Tech. Rep. R1-1718313, 3GPP, 10 2017.

[9] MCC Support, "3GPP TS 38.212 v1.0.1," tech. rep., 3GPP, 09 2017.

[10] 3GPP, "NR; Physical layer procedures for data," Technical Specification (TS) 38.214, 3rd Generation Partnership Project (3GPP), 01 2018. Version 15.0.0.