# Subcode-based Early HARQ for 5G

Barış Göktepe, Stephan Fähse, Lars Thiele, Thomas Schierl and Cornelius Hellge
Fraunhofer Heinrich Hertz Institute
Einsteinufer 37, 10587 Berlin
Email: [baris.goektepe, stephan.faehse, lars.thiele, thomas.schierl, cornelius.hellge]@hhi.fraunhofer.de

***Abstract*—Tactile Internet (TI) constitutes one of the major use cases for the development of the fifth generation (5G) mobile specification. TI services put high demand on the latency and reliability requirement, which is covered in the ultra-reliable low-latency communication (URLLC) discussion in 5G standardization. Hybrid Automatic Repeat Request (HARQ) is used in LTE to achieve high robustness in an efficient way with the cost of introducing additional latency. In this paper, we propose a new early HARQ scheme based on LDPC subcodes (SC E-HARQ), which enables to provide faster feedback and thus an earlier retransmission. The SC E-HARQ technique makes use of substructures in LDPC codes to start feedback calculation already on partially received codewords. This paper investigates the performance of SC E-HARQ in comparison with a second E-HARQ scheme based on log-likelihood ratio (LLR) estimation. The results show that SC E-HARQ achieves a comparable reliability to regular HARQ. In SNR regions relevant for URLLC, it clearly outperforms also the LLR-based E-HARQ in means of reliability as well as latency. Sub-millisecond latency with a total block error rate (BLER) of less than $10^{-4}$ is attained in TDL-C by allowing 1% false negative retransmissions.**

## I. INTRODUCTION

With emerging new applications, fifth generation (5G) mobile networks are expected to tackle three major use cases, enhanced mobile broadband (eMBB), ultra-low latency communication (URLLC) and massive machine-type communication (mMTC). Especially URLLC, puts high demands on the underlying physical layer procedures. Next Generation Mobile Networks Alliance (NGMN) has proposed a requirement of one-millisecond end-to-end latency and reliability of 99.999% [1]. These contradicting requirements have major impacts, especially on the hybrid automatic repeat request (HARQ), which is used in the current mobile network standard, Long Term Evolution (LTE), and in the currently specified 5G standard. In the past of mobile networks, HARQ has proven as an essential mechanism to trade latency for spectral efficiency while maintaining the reliability [2]. In case of decoding failure on the receiver side, the transmitter sends a retransmission that is combined with previous transmissions by utilizing chase combining (CC) or incremental redundancy (IR). The so-called HARQ round trip time (RTT), which is the time interval between receiving the initial transmission and the retransmission, poses a bottleneck for URLLC transmissions. Hence, early HARQ schemes based on early stopping criteria for the decoding process have been proposed and studied for Turbo codes extensively [3] [4]. Less work has been published for LDPC codes although there are many studies on stopping criteria for LDPC decoders [5] [6]. Nevertheless, these works focus on shortening the processing time after full reception of the codeword.

After IEEE 802.11n [7], also 3GPP has adopted Low-Density Parity-Check (LDPC) Codes as channel coding scheme for fifth-generation (5G) mobile networks at least for eMBB use case [8] [9], which also makes LDPC a natural candidate for URLLC. In this work, we present an early HARQ scheme based on LDPC subcodes (SC E-HARQ). Since LDPC codes offer excellent opportunities to exploit substructures, a scheme providing feedback before full reception of a codeword is proposed. In section II the proposed approach for SC E-HARQ and another E-HARQ scheme based on LLR estimation (LLR E-HARQ) is described. In section III, we present how SC E-HARQ could be applied to URLLC and discuss the latency of the different HARQ approaches. In section IV, the performance of the two early HARQ schemes, SC E-HARQ and LLR E-HARQ proposed in [10], is evaluated and compared to regular HARQ.

## II. EARLY HARQ FEEDBACK

The physical layer retransmission scheme HARQ introduces additional latency to the overall transmission due to the round-trip time (RTT). HARQ RTT arises of feedback reporting from the receiver, the retransmission from the transmitter and associated processing durations. First, the receiver has to process the transmission and reports the decoding outcome behind the transmitter. The transmitter has to process the feedback before issuing a retransmission. RTT is composed of the following components:

- $\tau$ - propagation delay
- $T_{TTI}$ - transmission time interval (TTI) duration
- $T_{RX}$ - processing time at the receiver
- $T_{A/N}$ - transmission time for ACK/NACK feedback
- $T_{TX}$ - processing time of the feedback at the transmitter

The propagation delay $\tau$ is caused by the physical length of the propagation path and is the only component which cannot be manipulated. Also, the transmission and processing time of the feedback, $T_{A/N}$ and $T_{TX}$, are assumed to be fixed.

Since a reduction of these results in a decreased reliability of the correct reception of the feedback, which is essential for HARQ. $T_{TTI}$ can be shortened by reducing the OFDM symbol duration and increasing subcarrier spacing. However, although this is a valid option, it is limited by the coherence bandwidth of the channel [11]. A similar approach has been followed in the 3GPP latency reduction study item for LTE [12] by composing the TTIs of a smaller number of orthogonal frequency duplexing modulation (OFDM) symbols. Although the study concluded that processing time $T_{RX}$ could be linearly scaled down with the TTI length, the processing time at the receiver is still the main component of HARQ RTT. The processing time $T_{RX}$ composes of processing the received signal to log-likelihood ratios (LLRs) $T_{LLR}$ as well as the time for feedback generation $T_{FB}$ involving full decoding, which is represented as:

$$T_{RX} = T_{LLR} + T_{FB}. \tag{1}$$

$T_{LLR}$ depends on the hardware implementation and is not in the scope of this work. To reduce $T_{FB}$, early stopping criteria for LDPC decoding have been proposed in [5] and [6]. These approaches allow an earlier detection of decoding failure, thus stopping the decoding process before reaching the maximum iterations. However, for reliable and efficient feedback transmission, practical communication systems, such as LTE, have fixed time slots for HARQ feedback transmission. Hence, the time between the transmission itself and the feedback slot has to be calculated based on the worst-case scenario that equals the maximum number of decoding iterations. So, under these constraints early stopping criteria cannot achieve any latency reduction.

Another way to reduce $T_{FB}$ is by replacing the decoder in the feedback generation chain by a predictor. The predictor analyzes the actual transmission and estimates the outcome of the decoding process ahead of the actual decoding start. A scheme based on a log-likelihood ratio (LLR) estimation, designated as LLR E-HARQ, proposed by Berardinelli et al. in [10] uses this approach by reducing the computational complexity of feedback generation, thus minimizing $T_{FB}$. The feedback generation is described more in detail in section II-A. In contrast to LLR E-HARQ, a second scheme, designated as Subcode E-HARQ (SC E-HARQ), proposed by this work utilizes only a part of the transmission which enables even earlier feedback reporting. The SC E-HARQ makes use of substructures in LDPC codes as presented in section II-B.

### A. LLR estimation based on whole codeword transmission

The LLR gives information on the likelihood of a bit being either $1$ or $0$. Denoting $\boldsymbol{y}$ as the observed sequence at the receiver, the LLR of the $k^{th}$ bit $b_k$ is defined as:

$$L(b_k) = \log \frac{P(b_k = 1|\boldsymbol{y})}{P(b_k = 0|\boldsymbol{y})}. \tag{2}$$

Using the definition of LLRs in 2, the probability of wrong bit estimates can be expressed as shown in [10]:

$$P_W(b_k) = \frac{1}{1 + |L(b_k)|}. \tag{3}$$

Furthermore, authors of [3] state that an estimation of the bit error rate (BER) is obtained by averaging over all $P_W$:

$$B\hat{E}R = \frac{1}{M} \sum_k P_W(b_k), \tag{4}$$

where $M$ is the length of the LLR vector of the whole transmission. The calculated $B\hat{E}R$ metric is employed to perform a prediction on the reliability of the transmission by using a threshold $th_{LLR}$ such that the feedback $FB_{LLR}$ is expressed as:

$$FB_{LLR} = \begin{cases} ACK & \text{if } B\hat{E}R \leq th_{LLR} \\ NACK & \text{if } B\hat{E}R > th_{LLR} \end{cases}. \tag{5}$$

Finally, the feedback generation for LLR E-HARQ is done in two steps. First, all received LLRs are combined to the estimated bit error rate $B\hat{E}R$. In the second step, the calculated $B\hat{E}R$ is mapped to ACK or NACK feedback based on a threshold $th_{LLR}$ which has to be obtained empirically.

### B. LDPC Subcode based estimation

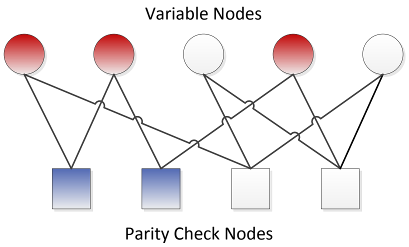

Fig. 1: Substructures of Tanner graphs

Low-density parity-check codes are represented by parity-check matrices with columns corresponding to variable nodes and rows corresponding to check nodes. In general, belief-propagation based decoders which pass messages with amplitudes corresponding to their reliability between variable nodes and check nodes iteration by iteration, are employed for decoding LDPC codes. This specific structure can be used to define subcodes and use them for prediction of the decoding outcome, as shown in Figure 1. Choosing a subset of check nodes (blue check nodes), i.e. rows of the parity-check matrix, from the mother code together with all associated variable nodes (red variable nodes) gives a subcode with a higher or equal code rate than the mother code. Similar as to the LLR estimation, the received LLRs are used for prediction whereas the bits belonging to the subcode are transmitted first in time before the residual bits of the transmission. So, calculation of the estimated bit error rate is started after the subcode has arrived but residual parts of the transmission are still being received in parallel. Applying LLR estimation only on this subcode, as described in the previous section, degrades the performance of the prediction [3]. For subcode based estimation, instead of utilizing the LLRs directly, the a posteriori LLRs $\Lambda_k$ after few belief propagation iterations are employed. Here, $\Lambda_k$ is expressed as [13]:

$$\Lambda_k = L(b_k) + \sum_{m \in \mathcal{M}(k)} \beta_{m,k}, \tag{6}$$

where $\mathcal{M}(k)$ is the set of check nodes which are associated to the variable node of $k$ and $\beta_{m,k}$ is the check-to-variable node message from check node $m$ to variable $k$. The estimated a posteriori bit error rate $B\hat{E}R_{ap}$ is calculated analogously to the LLR estimation from the a posteriori LLRs as presented:

$$B\hat{E}R_{ap} = \frac{1}{M} \sum_{k} \frac{1}{1 + |\Lambda_i|}. \tag{7}$$

The corresponding feedback $FB_{SC}$ is generated by applying a threshold $th_{SC}$ as:

$$FB_{SC} = \begin{cases} ACK & \text{if } B\hat{E}R_{ap} \leq th_{SC} \\ NACK & \text{if } B\hat{E}R_{ap} > th_{SC} \end{cases}, \tag{8}$$

where the threshold $th_{SC}$ has to be obtained empirically. In addition to the LLR E-HARQ, a further step for generating the a posteriori LLRs from the initial LLRs is introduced. In section III-A, the impact of this additional step on the latency is investigated.

## III. USING EARLY HARQ IN 5G

E-HARQ enables receivers to provide feedback at an earlier stage. However, the reduced computational complexity and using only parts of the transmission for SC E-HARQ comes at the cost of false prediction. Though, the prediction errors have to be distinguished to false positive and false negative predictions. A false positive prediction is defined as the event that the codeword is undecodable but has been predicted decodable. A false negative prediction is defined as the event that the codeword is decodable but has been predicted undecodable. Let $FB \in \{ACK, NACK\}$ be the feedback sent to the transmitter and $d \in \{0, 1\}$ the actual decoding success. Then, the corresponding false negative probability is expressed as:

$$\begin{aligned} fn &= Pr(d = 1 | FB = NACK) \\ &= \frac{Pr(d = 1, FB = NACK)}{Pr(FB = NACK)}. \end{aligned} \tag{9}$$

False negative predictions result in unnecessary transmissions which degrade spectral efficiency of the HARQ scheme but have no direct impact on latency and reliability, i.e., block error rate (BLER). Hence, false negative predictions can be tolerated up to a certain limit. Analogously, the false positive probability is shown as:

$$\begin{aligned} fp &= Pr(d = 0 | FB = ACK) \\ &= \frac{Pr(d = 0, FB = ACK)}{Pr(FB = ACK)}. \end{aligned} \tag{10}$$

False positive prediction corresponds to failure since the HARQ was not able to deliver corresponding transmission successfully within the given latency constraint. Hence, selecting the thresholds appropriately is critical for the performance of E-HARQ since the achieved BLER (together with the retransmission) is the measure for reliability.

### A. Latency Analysis of early HARQ

The latency of a HARQ system is mainly dominated by the HARQ round-trip time (RTT), which is comprised of the time required for generating the feedback and the time from transmitting the feedback till receiving the retransmission. Each HARQ feedback level requires an extra RTT, which limits the number of HARQ levels that are feasible for an ultra-low latency system. Since 1 ms latency is targeted, only a single HARQ step is used to achieve the required reliability. Thus, the maximum latency of the schemes corresponds to one HARQ RTT, as depicted in Figure 2. RTT is composed of two components, the time from starting the transmission till the feedback computation, designated as $T_1$, and the time from sending the feedback till receiving the whole retransmission, which is called $T_2$. The time to feedback generation $T_1$ is the parameter on which we focus in this work. $T_1$ is comprised of:

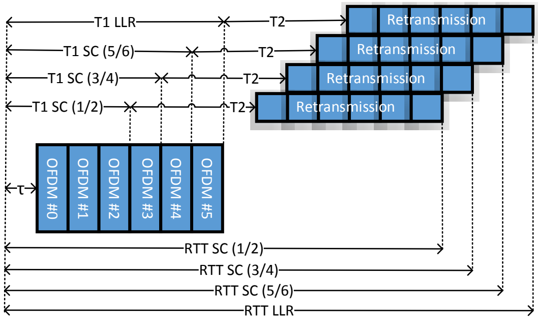

Fig. 2: Latency comparison of different E-HARQ schemes

$$T_1 = \tau + r_{SC} \cdot T_{TTI} + T_{LLR} + T_{FB}, \tag{11}$$

where $r_{SC}$ is the ratio between the subcode size and the total transmission size. The second timing parameter $T_2$ includes processing time and the time required for transmission of the feedback and the retransmission itself. The constant term $T_2$ looks as following:

$$T_2 = 2\tau + T_{A/N} + T_{TX} + T_{TTI}. \tag{12}$$

The assumed parameters are shown in Table I. Due to the very low computational complexity, $T_{FB}$ has been assumed 0 for LLR E-HARQ. For LDPC decoders the latency is given in [14]:

$$T_{FB,LDPC} = \frac{N d_v}{Z f_C} \cdot I, \tag{13}$$

where $N$ is the number of variable nodes, $d_v$ is the average variable node degree, $Z$ is the lifting size, $f_C$ is the clock frequency assumed to be 1 GHz and $I$ is the number of maximum iterations.

For simplicity reasons, we assume the number of variable nodes $N$ to be 1872, the number of variables of the whole codeword. Furthermore, $d_v$ equals 3.79, $Z$ is 36 and 50 decoding iterations are assumed for regular HARQ and 5 for SC E-HARQ. Under these assumptions, Table II shows the maximum latencies of the different schemes. Only the SC E-

TABLE I: Timing assumptions for latency evaluation

| Timing parameter | Regular HARQ | LLR E-HARQ | SC (1/2) E-HARQ | SC (2/3) E-HARQ | SC (3/4) E-HARQ | SC (5/6) E-HARQ |
|---|---|---|---|---|---|---|
| $\tau(1km)$ | 0.003 ms | 0.003 ms | 0.003 ms | 0.003 ms | 0.003 ms | 0.003 ms |
| $r_{SC}$ | 1 | 1 | $\frac{1}{2}$ | $\frac{2}{3}$ | $\frac{3}{4}$ | $\frac{5}{6}$ |
| $T_{TTI}$ | 0.4 ms | 0.4 ms | 0.4 ms | 0.4 ms | 0.4 ms | 0.4 ms |
| $T_{LLR}$ | 0.1 ms | 0.1 ms | 0.1 ms | 0.1 ms | 0.1 ms | 0.1 ms |
| $T_{FB}$ | 0.01 ms | 0 ms | 0.001 ms | 0.001 ms | 0.001 ms | 0.001 ms |
| $T_2$ | 0.6 ms | 0.6 ms | 0.6 ms | 0.6 ms | 0.6 ms | 0.6 ms |

TABLE II: RTT of E-HARQ and regular HARQ

| Latency | Regular HARQ | LLR E-HARQ | SC (1/2) E-HARQ | SC (2/3) E-HARQ | SC (3/4) E-HARQ | SC (5/6) E-HARQ |
|---|---|---|---|---|---|---|
| $T_1$ | 0.51 ms | 0.5 ms | 0.3 ms | 0.37 ms | 0.4 ms | 0.44 ms |
| RTT | 1.11 ms | 1.1 ms | 0.9 ms | 0.97 ms | 1.0 ms | 1.04 ms |

HARQ schemes are able to achieve a latency less than 1 ms, which is proposed as latency requirement for URLLC in [1]. The significant gain in latency for SC E-HARQ is achieved by the approach to calculate the feedback before the whole transmission is received. Due to the low latency of the LDPC decoder, the gain of LLR E-HARQ over regular HARQ is apparently very small and cannot achieve the same latency reduction which is realized by SC E-HARQ.

## IV. PERFORMANCE EVALUATION

In order to compare the prediction accuracy of both E-HARQ schemes link-level simulations have been performed. As explained in the previous sections, the false positive rate is critical for URLLC. Hence, false negative predictions have been fixed and the false positive rates observed, as presented in section IV-A and IV-B. Additionally, system performance, i.e. total BLER, of both E-HARQ schemes have been compared to regular HARQ with perfect feedback. The system setup and results are shown in section IV-C.

Table III shows the simulation assumptions for link-level simulations. The simulations have been performed with the rate-1/5 LDPC base graph 2 code, which was agreed recently for 5G in [15]. For SC E-HARQ, the subcodes are constructed based on a subset of rows of the parity-check matrix as documented in Table III. The associated variable nodes constitute the subcodeword for prediction. The fractions behind the row numbers represent the ratio between subcodeword size and codeword size.

TABLE III: Link-level simulation parameters

| | |
|---|---|
| Transport block size | 360 |
| Channel Code | LDPC |
| Rate | 1/5 |
| Parity-check matrix | see BG2 in [15] |
| Lifting size Z | 36 |
| Check constraints for prediction | 1-600 (1/2), 1-800 (2/3),<br>1-1000 (3/4), 1-1200 (5/6) |
| Modulation order and algorithm | QPSK, Approximated LLRs |
| Waveform | 3GPP OFDM, 1.4 MHz,<br>normal cyclic-prefix |
| Channel type | 1Tx 1Rx, TDL-C 100ns, 2.9 GHz |
| Equalizer | Frequency domain MMSE |
| Decoder type | Min-Sum |
| Decoding iterations | 50 |
| Iterations for estimation | 5 |
| Max. false negatives | 0.1, 0.05, 0.01 |

### A. Threshold evaluation

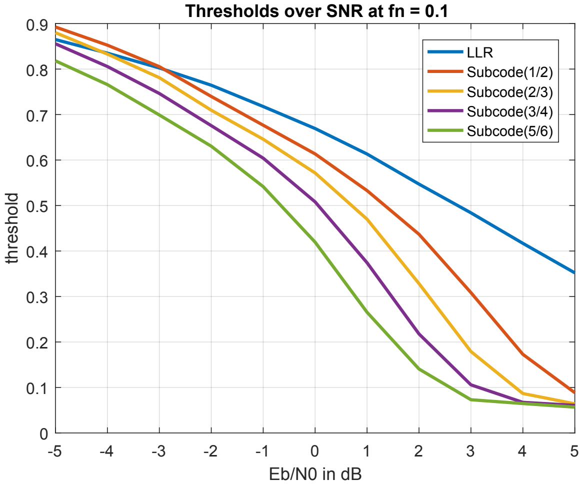

Fig. 3: Threshold over SNR in TDL-C for a false negative $fn = 0.1$

Choosing the thresholds impacts the performance of E-HARQ schemes significantly. In this section, the threshold for a false negative rate $fn = 0.1$ over the SNR is shown in Figure 3. With increasing SNR the threshold is decreasing for all schemes. However, the thresholds of SC E-HARQ schemes decrease faster than the threshold of LLR scheme. This is due to the higher SNR, which increases the probability that a received codeword converges in the belief propagation algorithm. At the highest SNRs a saturation is reached such that the decreasing of thresholds stops immediately. This effect can be explained by the fact that a $\hat{BER}_{ap}$ of $0$ would correspond to LLRs going to infinity which is not achievable by the limited number of decoder iterations.

### B. Prediction performance evaluation

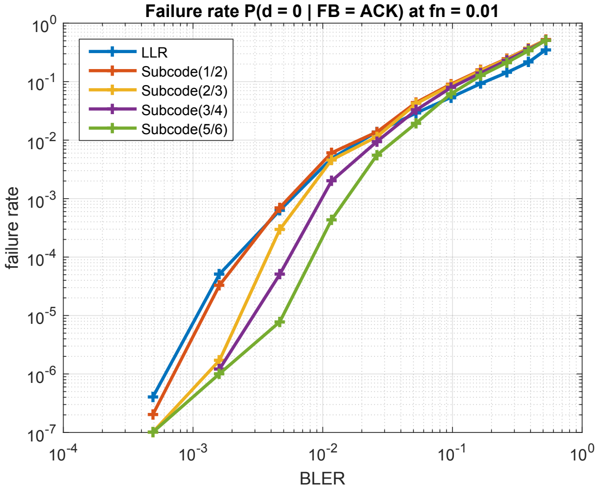

Fig. 4: False positive (failure) rate over BLER for LLR and SC E-HARQ in TDL-C

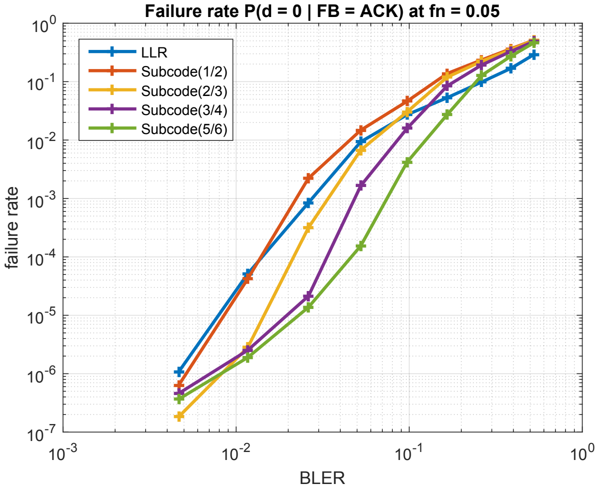

Fig. 5: False positive (failure) rate over BLER for LLR and SC E-HARQ in TDL-C

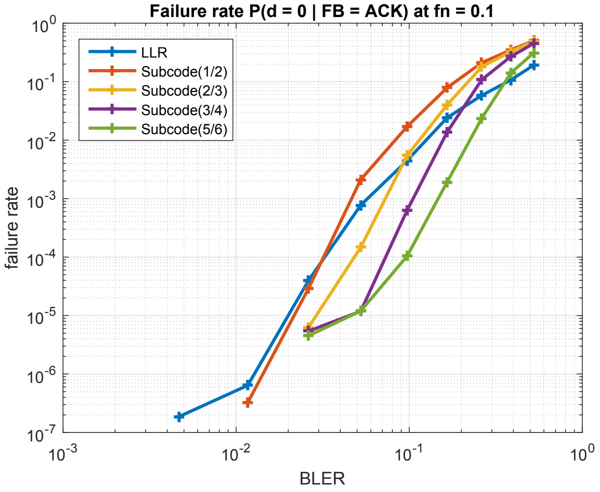

Fig. 6: False positive (failure) rate over BLER for LLR and SC E-HARQ in TDL-C

In Figures 4, 5 and 6, the false positives (failures) in TDL-C are shown over the BLER for different thresholds which have been calculated for $fn = 0.01$, $fn = 0.05$ and $fn = 0.1$, respectively, as decribed in section III in (9) and (10). As expected, allowing a higher false negative rate decreases the failure rate of all schemes. At high BLERs, larger than $10^{-1}$, the LLR based estimation achieves better results than all Subcode based schemes. This can be explained by the strong noise at low SNRs which prohibits codewords from converging. Nevertheless, this region is rather unfeasible for URLLC since the BLER is too large and also false positives lie in the region larger than $10^{-1}$ for all false negative assumptions. However, with decreasing BLER the failure rates are also decreasing. Especially, the Subcodes (2/3), (3/4) and (5/6) achieve significantly lower failure rates than the Subcode (1/2) and the LLR estimation. Comparing the figures over the different threshold assumptions shows that allowing a higher false negative minimizes the failure rates at the same BLERs. This can be noted in Figure 4 at $fn = 0.01$ where a failure rate of $fp = 10^{-4}$ is achieved by all schemes in the BLER region between $2 \cdot 10^{-3}$ and $9 \cdot 10^{-3}$ whereas in Figure 6 at $fn = 0.1$ the same failure rate is achieved between $3 \cdot 10^{-2}$ and $1.5 \cdot 10^{1}$. This result means that in these BLER regions approx. every $10,000^{th}$ ACK is a wrong one. Table IV summarizes the results for a failure rate of $10^{-5}$, which is a feasible assumption if a BLER larger than $10^{-3}$ is assumed for the initial transmission. In the evaluated BLER region all SC E-HARQ schemes outperform the LLR E-HARQ. However, a significant gain is observed for the Subcodes larger than (2/3). The SC E-HARQ achieves almost 1 dB gain over the LLR E-HARQ for all threshold assumptions, which is even increased for Subcodes (3/4) and (5/6).

TABLE IV: Required SNR for false positive of $10^{-5}$

| Scheme fn | LLR | SC (1/2) | SC (2/3) | SC (3/4) | SC (5/6) |
|---|---|---|---|---|---|
| 0.01 | 4.82 dB | 4.71 dB | 3.98 dB | 3.83 dB | 3.0 dB |
| 0.05 | 2.83 dB | 2.78 dB | 1.98 dB | 1.61 dB | 1.32 dB |
| 0.1 | 1.77 dB | 1.67 dB | 0.98 dB | 0.29 dB | 0.27 dB |

### C. System performance

The prediction accuracy evaluated in the previous section is critical for the performance in a communication system. URLLC requires a total BLER of $10^{-5}$ within 1 ms latency. Hence, as shown in section III-A only a single retransmission can be performed to achieve the required BLER. For evaluation purposes the retransmission is composed of the same codeword such that the receiver combines both transmission by using chase combining. Figure 7 and 8 show the achieved BLER with false negative rates of $fn = 0.01$ and $fn = 0.05$, respectively. Additionally, the BLER of a single transmission (BLER wo RTx) and the BLER of regular HARQ with perfect

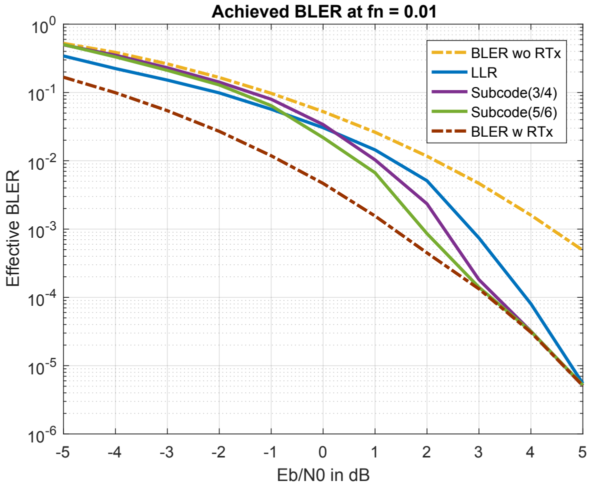

Fig. 7: Achieved effective BLER over SNR in TDL-C for a false negative $fn = 0.01$

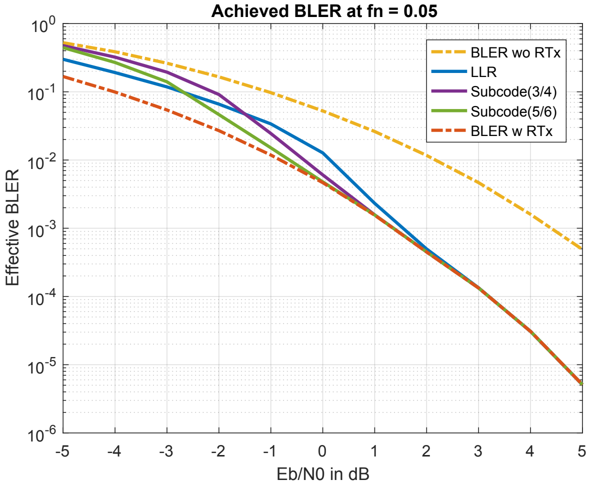

Fig. 8: Achieved effective BLER over SNR in TDL-C for a false negative $fn = 0.05$

feedback (BLER w RTx) is depicted for comparison. At low SNRs the LLR scheme can achieve a better performance than the Subcode based schemes. However, a BLER of more than $10^{-1}$ is not acceptable for reliable communication. Hence, the performance at higher SNRs plays a more important role. Here, as it can be seen clearly in Figure 7 and Figure 8, the Subcode based scheme achieves a significantly lower BLER at these SNR values. This is due to the more accurate prediction at the same false negative rate, as shown in the previous section. For $fn = 0.01$ in Figure 7, the Subcode (5/6) can achieve a gain of approximately $1\,\mathrm{dB}$ at a BLER of $10^{-4}$. At a BLER of $10^{-5}$ both schemes achieve the performance of regular HARQ which the limiting bound here. A gain of approx. $1\,\mathrm{dB}$ over the LLR based scheme can also be seen for $fn = 0.05$ at $10^{-2}$ in Figure 8. The Subcode (5/6) reaches the limiting bound of regular HARQ at $0\,\mathrm{dB}$ whereas the LLR E-HARQ requires a SNR of $2\,\mathrm{dB}$. However, the corresponding BLER at $0\,\mathrm{dB}$ of approx. $5 \cdot 10^{-3}$ is too high for URLLC. Anyway, a lower coding rate has to be used to achieve the required $10^{-5}$. However for larger SNRs, the performance of the early schemes with $fn = 0.05$ is limited by the BLER curve with regular HARQ (BLER w RTx) at too BLERs. Hence, the assumption of $fn = 0.05$ is too loose for URLLC for this SNR range. The threshold assumption of $fn = 0.01$ is a feasible operation point for reliable low-latency communication. Taking an unnecessary retransmission rate of $1\%$ into account which decreases spectral efficiency slightly, both Subcode based schemes can totally replace the regular HARQ mechanism for a total BLER of lower than $10^{-4}$.

## V. CONCLUSION

In this paper, we propose and evaluate an E-HARQ scheme based on LDPC subcodes compared to another E-HARQ scheme and regular HARQ. We show that the proposed SC E-HARQ scheme achieves a better false positive performance than the LLR based scheme at regions of total BLER smaller than $10^{-5}$ which are relevant for URLLC transmission. For Subcodes (3/4) and (5/6) this results to a superior reliability compared to LLR E-HARQ allowing a false negative rate of $1\%$ while decreasing the latency by up to $10\%$ compared to LLR E-HARQ as well as regular HARQ. In the low BLER regions, SC E-HARQ is even able to achieve a comparable reliability to regular HARQ with a gain of approx. $2\,\mathrm{dB}$ over the LLR E-HARQ. In future works, further investigations for different block sizes and rates have to be performed to evaluate the performance of the SC E-HARQ.

## REFERENCES

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

[2] F. Frederiksen and T. E. Kolding, “Performance and modeling of wcdma/hsdpa transmission/h-arq schemes,” in *Proceedings IEEE 56th Vehicular Technology Conference*, vol. 1, pp. 472–476 vol.1, 2002.

[3] G. Berardinelli, S. R. Khosravirad, K. I. Pedersen, F. Frederiksen, and P. Mogensen, “On the benefits of early harq feedback with non-ideal prediction in 5g networks,” in *2016 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 11–15, Sept 2016.

[4] D. Lee, K. Ko, C. H. Nam, and H. S. Lee, “A hybrid arq scheme using packet error prediction on ofdm based hsdpa systems,” in *2007 16th IST Mobile and Wireless Communications Summit*, pp. 1–5, July 2007.

[5] F. Kienle and N. Wehn, “Low complexity stopping criterion for ldpc code decoders,” in *2005 IEEE 61st Vehicular Technology Conference*, vol. 1, pp. 606–609 Vol. 1, May 2005.

[6] T.-C. Chen, “An early stopping criterion for ldpc decoding based on average weighted reliability measure,” in *CSQRWC 2012*, pp. 123–126, July 2012.

[7] M. Peyic, H. Baba, E. Guleyuboglu, I. Hamzaoglu, and M. Keskinoz, “A low power multi-rate decoder hardware for ieee 802.11n ldpc codes,” *Microprocessors and Microsystems*, vol. 36, no. 3, pp. 159 – 166, 2012.

[8] MCC Support, “Final Report of 3GPP TSG RAN WG1 #86bis v1.0.0,” Tech. Rep. R1-1611081, 3GPP, 10 2016.

[9] MCC Support, “Draft Report of 3GPP TSG RAN WG1 #87 v0.1.0,” tech. rep., 3GPP, 11 2016.

[10] G. Berardinelli, S. R. Khosravirad, K. I. Pedersen, F. Frederiksen, and P. Mogensen, “Enabling early harq feedback in 5g networks,” in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, pp. 1–5, May 2016.

[11] N. Marchetti, M. I. Rahman, S. Kumar, and R. Prasad, *OFDM: Principles and Challenges*, pp. 29–62. Boston, MA: Springer US, 2009.

[12] Ericsson, “Way forward on processing timing reduction for sTTI,” Tech. Rep. R1-165854, 3GPP, 05 2016.

[13] X. Wu, Y. Song, M. Jiang, and C. Zhao, “Adaptive-normalized/offset min-sum algorithm,” *IEEE Communications Letters*, vol. 14, pp. 667–669, July 2010.

[14] Qualcomm, “Efficient Channel Coding Implementations for EMBB,” Tech. Rep. R1-1610139, 3GPP, 10 2016.

[15] MCC Support, “3GPP TS 38.212 v1.0.1,” tech. rep., 3GPP, 09 2017.