Flexible Multi-Bit Feedback Design for HARQ Operation of Large-Size Data Packets in 5G

Khosravirad, Saeed; Mudolo, Luke; Pedersen, Klaus I.

Published in:
IEEE Proceedings of VTC-2017 spring, June 2017

DOI (link to publication from Publisher):
10.1109/VTCSpring.2017.8108610

Publication date:
2017

Document Version
Accepted author manuscript, peer reviewed version

Flexible Multi-Bit Feedback Design for HARQ Operation of Large-Size Data Packets in 5G

Saeed R. Khosravirad†, Luke Mudolo†, Klaus I. Pedersen†§
†Nokia - Bell Labs
§ Aalborg University, Denmark

***Abstract*—A reliable feedback channel is vital to report decoding acknowledgments in retransmission mechanisms such as the hybrid automatic repeat request (HARQ). While the feedback bits are known to be costly for the wireless link, a feedback message more informative than the conventional single-bit feedback can increase resource utilization efficiency. Considering the practical limitations for increasing feedback message size, this paper proposes a framework for the design of flexible-content multi-bit feedback. The proposed design is capable of efficiently indicating the faulty segments of a failed large-size data packet thanks to which the transmitter node can reduce the retransmission size to only include the initially failed segments of the packet. We study the effect of feedback size on retransmission efficiency through extensive link-level simulations over realistic channel models. Numerical result present significant savings in retransmission resources offered by the proposed flexible-content feedback design.**

I. Introduction

Research and development for the design of the 5th generation mobile networks (5G) is currently attracting significant effort from both industry and academia with the vision of improving overall performance of mobile networks as compared to the existing technologies by providing higher data rate, lower end-to-end latency and increased reliability among other key performance indicators (KPIs) [1]. New service types are envisioned for 5G to cover the challenging and highly diverse requirements of the massive machine type of communication (mMTC) and the ultra-reliable low latency communication (URLLC) [2]. In particular, the emerging 5G technology is expected to operate on comparably wider carrier bandwidths with respect to the existing technologies in order to offer high data rate radio links to the users over carrier bandwidth of multiple hundred MHz [3], [4]. Thus, for the next generation enhanced mobile broadband (eMBB) communication the scheduler entity will expectedly deliver larger data packets to the lower layer HARQ, compared to the existing long term evolution (LTE) technology. This calls for enhancements in retransmission operation of large-size data packets to increase resource utilization efficiency.

A retransmission attempt for an initially failed packet is best to only include the faulty segments of the packet [5], [6]. Such refined retransmission will require extra information as compared to the conventional acknowledgement (ACK)/negative acknowledgement (NACK) over the feedback channel calling for an increased number of feedback bits. The feedback bits are however costly due to the high order repetition coding that is typically used in physical layer to provide high reliability to the feedback message delivery. E.g., in LTE the resources used to report a single-bit feedback can span over multiple resource element (RE) up to a physical resource block (PRB) in up-link (UL) and down-link (DL) HARQ respectively [7]. Hence, a new design for HARQ multi-bit feedback must carefully acknowledge the crucial trade-off between the cost of increasing the number of feedback bits and the resulting retransmission resource savings.

The use of enriched multi-bit feedback in optimization of HARQ performance and the trade-off between throughput, outage probability and packet delay have been studied extensively in the literature [8]–[12]. For instance, a multi-bit feedback enriched with partial channel state information (CSI) in [13] or conveying outdated CSI in [6] have shown to be beneficial for an optimal adaptation of transmission power and rate. In [5], the authors have shown that decoder state information (DSI) measured in accumulated mutual information (ACMI) offers attractive throughput gain for the cost of a single extra feedback bit per HARQ process.

In this paper we propose to use multi-bit HARQ feedback to indicate the decoding status of code block (CB) segments in a large-size transport block (TB). We propose to indicate the decoding error with higher resolution as compared to the bundled acknowledgment provided by a single-bit feedback. In particular we study the decoding error performance of large-size TB in different multi-path channel models and propose several approaches of reporting CB decoding status in a limited-size multi-bit feedback. The proposed solutions are compared by the provided resource saving gains against the required control channel overhead.

The rest of the paper is organized as follows: in Sec. II we lay out the setup for the simulation analysis in this paper and study the block error performance a TB in different channel models; in Sec. III the proposed design for flexible-content multi-bit feedback is presented; Sec. IV presents the numerical simulation results; finally, Sec. V covers the concluding remarks.

II. System Model & Problem Formulation

In this paper we aim to study the resource saving gains offered by deployment of multi-bit HARQ feedback in practical channel models. We adopt the physical layer numerology assumptions for LTE and focus on link-level error performance of data channel conveying large-size TB using an orthogonal frequency-division multiple access (OFDMA) system. We assume the use of 15 kHz subcarrier spacing with 14 OFDM symbols per 1 ms subframe. All the simulations in this paper are performed assuming 20 MHz carrier bandwidth with 100 PRB and resolution of 12 subcarriers per PRB. We further assume transmission time interval (TTI) duration of multiple 1 ms subframes when it is needed to accommodate a large size TB. The simulations use Turbo coding with basic code rate of 1/3 for channel coding with QPSK, 16-QAM and 64-QAM as the available modulation orders. The modulation and coding scheme (MCS) list in [7] is used as reference for generating different code rates by puncturing or repeating coded bits from the encoder output. The layout of the CBs over physical resources is assumed to follow LTE specification [14], meaning that CBs will be laid out one-by-one over the allocated resources. Each CB will occupy resource elements (REs) along the frequency axis first and then moves on to the next OFDM symbol.

We further adopt several time-variant multi-path fading channel models for this study namely, the extended pedestrian A (EPA), the extended vehicular A (EVA) and the extended typical urban (ETU) channel models while trying out different maximum Doppler frequencies. For further details regarding the description of these channel models we refer to [15]. For simplicity reasons we will denote the channel models together with the assumed maximum Doppler frequency (e.g., ETU300 denotes the ETU multi-path model with 300 Hz maximum Doppler frequency). The simulation parameters are summarized in Table II.

The bandwidth per 5G new radio (NR) carrier is estimated to increase as compared to LTE carrier to up to 400 MHz [4]. Therefore, scheduling an eMBB user over the full transmission bandwidth could result in very large TBS. Using LTE terminology, a data packet is typically formed as a TB which, prior to channel encoding, will be segmented into multiple smaller size CBs with maximum CB size of 6144 bits [14]. Each TB is typically associated with one HARQ process in the Media Access Control (MAC) layer where feedback acknowledgment corresponds to TB decoding status. Thus, receiver node will only report an ACK if all the CBs in the corresponding TB are correctly decoded and will report NACK otherwise.

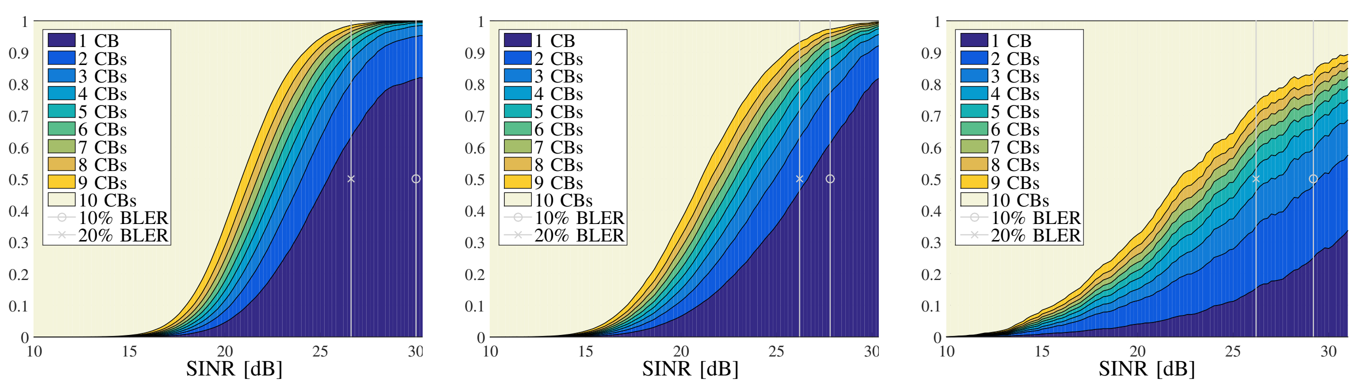
The TB segmentation process is mainly deployed in LTE to reduce the complexity of the encoding/decoding operations, which for similar reason is expected to be used in 5G technology too. The hidden gain of the segmentation process which is not exploited in LTE technology is the chance to indicate the erroneous CBs in a multi-bit feedback report. Therefore, less resources will be used by only retransmitting the failed CBs as opposed to retransmission of the whole TB offered by the traditional single-bit bundled feedback. In other words, by using e.g., cyclic redundancy check (CRC) error detecting code for each of the CBs the decoding failure can be detected separately. Reducing the retransmission to only convey failed CBs will result in saving retransmission physical resources and transmit energy.

The block error rate (BLER) performance of CBs in a large-size TB is a function of the experienced channel, coded block length and channel coding. In LTE, the interleaver matrix sizes impose extra limitation on the CB segmentation size which may result in non-uniform segmentation of a given TB. Therefore, as also noticed in [16] the CBs in a large TB are not bound to the similar BLER performance. Assuming independent decoding output for the CBs in a large TB has been suggested as a simplified model in the literature (e.g., see [16]). The model assumes an independent and identically distributed (i.i.d.) block fading channel model where each CB experiences one i.i.d. flat-fading channel block. As a result of such simplified model the BLER for TB denoted as $\mathrm{BLER_{TB}}$ will be as follows,

$$\mathrm{BLER_{TB}} = 1 - \prod_{i}(1 - \mathrm{BLER}_{\mathrm{CB}_i}), \tag{1}$$

where $\mathrm{BLER}_{\mathrm{CB}_i}$ is the BLER of the ith CB. In more realistic channel situations the error performance of CBs are expected to have dependency e.g., due to correlated fading or the coherence time duration [17].

As an example, in Fig. 1 the chances of different number of failed CBs causing a bundled NACK for a TB with 10 CBs is compared in different multi-path channel models. It is



(a) (b) (c)

Figure 1: The chances of different number of erroneous CBs resulting in a NACKed TB. The results are shown for a range of multi-path channel models as follows: (a) ETU300, (b) EVA70, (c) EPA5. We assume packet size of transport block size (TBS) = 60536 bits (10 CBs) while transmitting using MCS22.

Table I: Chances of different events (in percentage) causing a bundled TB NACK.

	TB BLER	ETU300	EVA70	EVA5	EPA5	i.i.d.
One failed CB	10%	82	61	62	25	95.3
	20%	63	46	47	15	90.2
One or two failed CBs	10%	95	77	77	47	99.8
	20%	80	63	64	35	99.4

shown that over typical operating signal to interference and noise ratio (SINR) values the likelihood of having different number of failed CBs in a failed large-size TB varies for different fading scenarios. For the SINR points corresponding to $\text{BLER}_{\text{TB}} = 10\%$ and 20%, in Table I a summary of the results from Fig. 1 are compared to the case of i.i.d. block fading channel model. As shown, for fast varying channel conditions the chances of having only one or two failed CBs causing a TB NACK is much higher compared to the case of slowly varying channel. For instance, in $\text{BLER}_{\text{TB}} = 10\%$ SINR point the single-bit HARQ NACK feedback is cause by one or two failed CBs in 95% of the cases in the ETU300 channel model while for the case of EPA5 in similar target SINR point the chances of only one or two failed CBs reduces to 47%. Moreover, this observation offers significant potential resource saving gain by omitting the successfully decoded CBs in the HARQ retransmission in all tested channel models.

III. Flexible Content Multi-Bit HARQ Feedback

An efficiently set up HARQ retransmission will require to exclude the correctly decoded segments of an initially failed packet thus, requiring extra bits of feedback information. As mentioned, a practical design for multi-bit feedback must consider the overhead expenses of each extra feedback bit. In this study we adopt the assumption of an error-free feedback channel with given capacity of m information bits and propose three different approaches of using such feedback channel for efficient retransmission setup.

CB-indexing feedback: The indexing approach (*INDXfb*) reports the indexes of the failed CBs. We assume that the transmitter and receiver nodes have shared a table with $2^{\mathit{IndLen}(N)} = 2^N$ rows, where each row conveys one of the possible ACK/NACK patterns for a TB with N CBs. The rows are sorted as follows: the first row corresponds to zero failed CBs (i.e., TB ACK); the following immediate rows will correspond to the cases of one failed CB, then comes the cases of two failed CBs, and so on. Thus, the last row corresponds to the event where all CBs have failed in decoding. After decoding a given TB the receiver node will realize the CB failure pattern and then finds the corresponding row from such table and sends the row number as feedback message denoted in this paper by *RowIndex*$\in [0 : 2^N - 1]$. Using such sorting method the row number corresponding to l failed CB case will require not more than *IndLen*(l) bits to be reported where,

$$\mathit{IndLen}(l) = \left\lceil \log_2 \sum_{i=0}^{l} \binom{N}{i} \right\rceil, \qquad (2)$$

and $\binom{c}{k}$ denotes the k-combination of c and $\lceil . \rceil$ is the ceiling function. This approach will help with retransmission resource savings only if $m \geq \mathit{IndLen}(1)$. Algorithm 1 summarizes the *INDXfb* approach.

Algorithm 1: CB-indexing feedback

Input : number of feedback bits, m and the number of failed CBs, l
Output: CB-indexing feedback message, *INDXfb*(m)
if $l == 0$ **then**
  return zero;
else if $m \geq \mathit{IndLen}(l)$ **then**
  return *RowIndex*;
else
  return $2^N - 1$;
end

CBG-based feedback: It has been agreed for the HARQ operation in 5G NR to allow for code block group (CBG)-based retransmission with configurable granularity of the CBGs [4]. In this proposed feedback model we assume that CBs are grouped into m separate CBGs where each CBG will be acknowledged separately. We further assume that a CBG acknowledgment is ACK if all of the CBs in the CBG are successfully decoded, and it is NACK otherwise. The size of CBGs are assumed to be configured based on TBS and m where each CBG includes minimum $\lfloor N/m \rfloor$ and maximum $\lceil N/m \rceil$ adjacent CBs over physical resources. This approach is summarized in Algorithm 2.

Algorithm 2: CBG-based feedback

Input : number of feedback bits m
Output: CBG-based feedback message, *CBGfb*(m)
for $i \in [1 : m]$ **do**
  CBGfbvector(i) = bundled feedback of the ith CBG;
end
return *CBGfbvector*;

flexible content feedback: The flexible content feedback (*FCfb*(m)) approach reports the message created by one or the other of the above approaches. The receiver will reserve one bit of the feedback message as header bit which will be used to indicate which of the *INDXfb* or *CBGfb* feedback messages will be conveyed over the remaining $m - 1$ bits. Next, the receiver will evaluate which of the two approaches will trigger a smaller retransmission size for the given CB decoding failure pattern (we denote this evaluation by function *Best*$\{.\}$). For the cases where $m \leq \mathit{IndLen}(1)$ the *FCfb*(m) does not reserve any header bits and is equivalent to the *CBGfb*(m) feedback content. Algorithm 3 summarizes the flexible content feedback approach.

Algorithm 3: flexible content feedback

FCfb(m);
Input : number of feedback bits m
Output: feedback message
if $m \in [1 : \text{IndLen}(1)]$ **then**
  return *CBGfb*(m);
else
  return 1 bit *header* + *Best*$\{$*CBGfb*$(m-1)$, *INDXfb*$(m-1)\}$;
end

IV. Numerical results

In this section we present link-level simulation results for the proposed muti-bit HARQ feedback approaches. We focus on link-level analysis of data channel where large-size TBs are generated at the base station (BS) and transmitted to the user equipment (UE) with the given attributes in Table II. We analyze three multi-path channel models to capture the effect of CB decoding performance on resource saving gains offered by the proposed multi-bit feedback models. To acknowledge LTE link adaptation over the range of SINR values we try three different MCSs as shown in Table II together with the corresponding TBS with $N = 50$ CBs as follows: MCS5 with TBS$=$ 270200 bits; MCS13 with TBS$=$ 276960 bits; and, MCS22 with TBS$=$ 302680 bits.

In Fig. 2 the three different approaches of multi-bit feedback are compared by the normalized retransmission resource savings against the number of available feedback bits m for the case of $N = 50$ CBs. We assume that retransmission of a failed CB will utilize the same amount of physical resources as the initial transmission. The normalized retransmission ratio is defined as the number of CBs in the retransmission triggered by a given multi-bit feedback approach, normalize by N (i.e., the number of CBs in a retransmission after single-bit NACK). The results confirm that for the range of feedback length where $m \in [1 : \textit{IndLen}(1) = 6]$ the $\textit{CBGfb}(m)$ approach provides the HARQ operation with the best resource savings among the proposed approaches while for larger number of feedback bits it is suggested to reserve one header bit in the feedback message and use the $\textit{FCfb}(m)$ approach. In particular with $m = 10$ bits of feedback the proposed $\textit{FCfb}(m)$ can provide ∼90% savings in retransmission resources as compared to single-bit bundled feedback per TB.

In Fig. 3 the retransmission resource saving ratio, calculated as one minus the normalized retransmission ratio, is shown against SINR for the ETU300 channel model and the potential savings offered by different number of feedback bits are compared for the *FCfb* approach proposed in Sec. III. Particularly, using 10 bits of feedback in Fig. 3(c) 90% retransmission resource saving is offered at SINR$=$ 30 dB. Similar performance metric is illustrated in Fig. 4 and Fig. 5 respectively for the EVA70 and the EPA5 channel models. The retransmission resource savings using 10 bits of feedback at the the target SINR point with $\text{BLER}_{\text{TB}} = 10\%$ in Fig. 4(c) and Fig. 5(c) are as high as 90% and 58% respectively.

The attractive retransmission resource savings explained above are thanks to considering a single header bit in the *FCfb* approach which makes it possible to take advantage of the benefits of both *INDXfb* and *CBGfb* approaches. Specifically, in the *CBGfb* approach the retransmission size is on average a multiple of the CBG size causing it to have limited resource saving gains in the case where only one or two of the CBs have failed in decoding, e.g., in fast varying channel conditions.

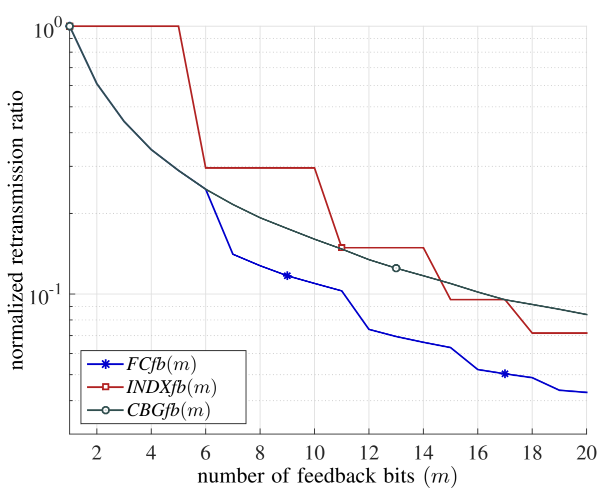


Figure 2: Average normalized retransmission ratio is compared against feedback size m: the results are for the EVA70 channel model with TBS= 302680 bits (50 CBs) using MCS22 at transmit SINR = 30 dB.

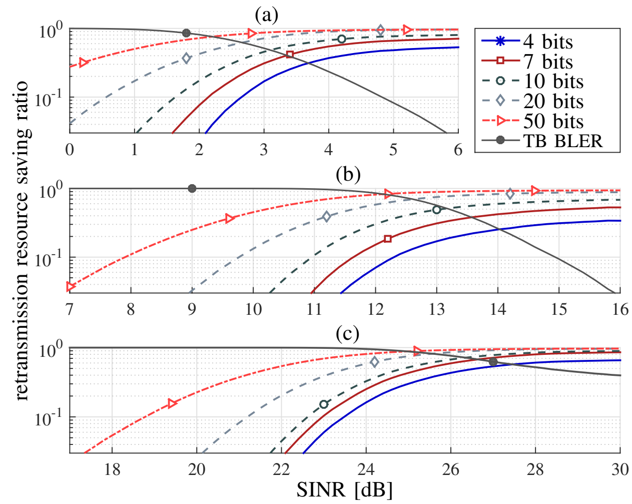


Figure 3: The average retransmission resource saving ratio offered by the proposed *FCfb* approach for (a) MCS5, (b) MCS13 and (c) MCS22 over ETU300 channel model.

Table II: Link-level analysis parameter setup

Parameter	Value
Carrier frequency	2.6 GHz
Carrier bandwidth	20 MHz
Subcarrier spacing	15 kHz
Number of allocated PRBs	100
Number of subcarriers per PRB	12
FFT size	2048
TTI interval	multiples of 1 ms
Number of useful OFDM symbols per 1 ms	14
Power delay profile	ETU300, EVA70, EPA5
Channel estimation	Ideal
MCS formats	MCS5, MCS13, MCS22
Antenna scheme	SISO
Channel coding	Turbo coding with basic rate 1/3

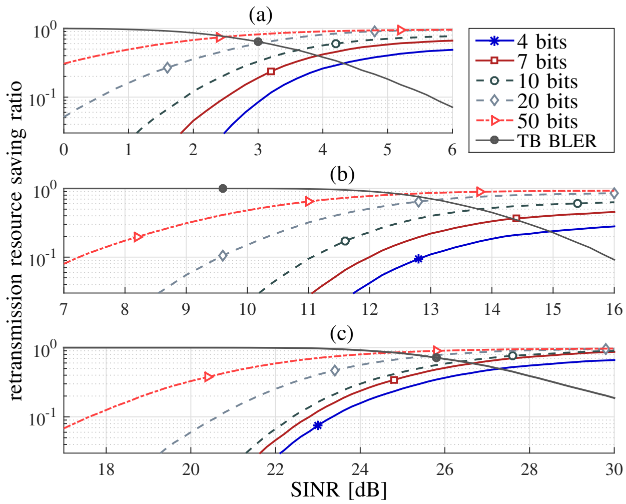


Figure 4: The average retransmission resource saving ratio offered by the proposed *FCfb* approach for (a) MCS5, (b) MCS13 and (c) MCS22 over EVA70 channel model.

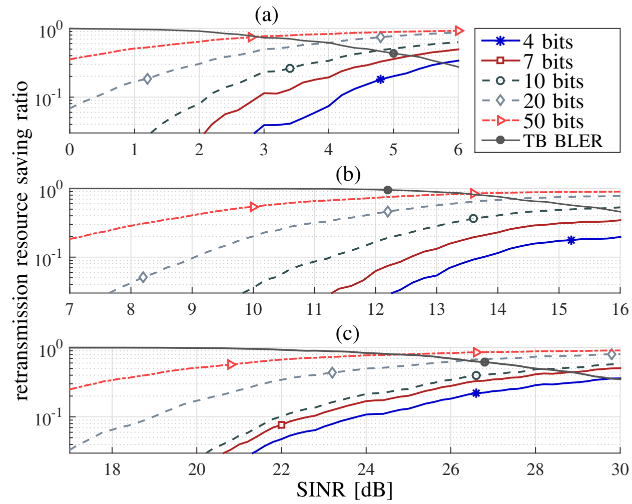


Figure 5: The average retransmission resource saving ratio offered by the proposed *FCfb* approach for (a) MCS5, (b) MCS13 and (c) MCS22 over EPA5 channel model.

On the other hand, the *INDXfb* approach can well reduce the retransmission size to only cover the failed CBs when the number of failed CB indexes is small enough to fit in the feedback bits. However, in scenarios with high dependency between decoding failure of the CBs it is likely to have more CBs fail in decoding where as a result the *INDXfb* approach will trigger retransmission of the whole TB and thus fail to offer resource saving. Thus the *FCfb* chooses the optimal content of the multi-bit feedback depending on the decoding output of the CBs and notifies the transmitter node about this choice over a single-bit header.

V. Conclusions

5G NR is expected to support increasingly larger transport block sizes as compared to LTE. This will result in a larger maximum number of code block segments per transport block. Using a configurable multi-bit feedback, HARQ process can perform more efficiently both in the utilized physical resources and consumed transmission energy by skipping the correctly decoded segments of the transport block in the retransmission. In this study we proposed several multi-bit feedback approaches and studied the offered resource saving gains against the range of SINR in different channel models. The results show that the approach of reporting one or a few indexes of failed CBs as HARQ feedback and the CBG-based acknowledgment approach can both offer resource saving gains where, the former is more suitable to fast varying channel conditions and the latter is a better choice in slowly varying channels. A flexible multi-bit feedback design was proposed that is able to offer the best of the two mentioned feedback contents and provide the HARQ process with attractive retransmission resource savings of more than 90% in some scenarios at the expense of one feedback bit per five CB segments.

References

[1] IMT Vision, "Framework and overall objectives of the future development of IMT for 2020 and beyond," *ITU, Feb.*, 2014.

[2] K. I. Pedersen *et al.*, "A flexible frame structure for 5G wide area," in *IEEE Vehicular Technology Conference (VTC'15)*, Sep. 2015.

[3] "3GPP TR 36.913; study on scenarios and requirements for next generation access technologies," 3GPP, Tech. Rep., Mar. 2016.

[4] "3GPP TR 38.802; study on new radio access technology physical layer aspects," 3GPP, Tech. Rep., Mar. 2017.

[5] S. R. Khosravirad *et al.*, "HARQ enriched feedback design for 5G technology," in *IEEE Vehicular Technology Conference (VTC'16)*, Sep. 2016.

[6] L. Szczecinski *et al.*, "Rate allocation and adaptation for incremental redundancy truncated HARQ," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2580–2590, Jun. 2013.

[7] "3GPP TS 36.213; evolved universal terrestrial radio access (E-UTRA); physical layer procedures," 3GPP, Tech. Spec., Jan. 2015, release 10.

[8] E. Uhlemann *et al.*, "Optimal incremental-redundancy strategy for type-II hybrid-ARQ," in *IEEE 2003 International Symposium on Information Theory*, Jun. 2003, p. 448.

[9] J.-F. Cheng *et al.*, "Adaptive incremental redundancy [WCDMA systems]," in *IEEE VTC 2003 (Fall)*, vol. 2, Oct. 2003, pp. 737–741.

[10] D. Djonin *et al.*, "Joint rate and power adaptation for type-I hybrid-ARQ systems over correlated fading channels under different buffer-cost constraints," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 421 –435, Jan. 2008.

[11] B. Makki and T. Eriksson, "On the performance of MIMO-ARQ systems with channel state information at the receiver," *IEEE Trans. Commun.*, vol. 62, no. 5, pp. 1588–1603, May 2014.

[12] M. Jabi *et al.*, "Outage minimization via power adaptation and allocation in truncated hybrid ARQ," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 711–723, Mar. 2015.

[13] D. Tuninetti, "On the benefits of partial channel state information for repetition protocols in block fading channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5036 –5053, Aug. 2011.

[14] "3GPP TS 36.213; evolved universal terrestrial radio access (E-UTRA); multiplexing and channel coding," 3GPP, Tech. Spec., Dec. 2016, release 10.

[15] "3GPP TS 36.104; evolved universal terrestrial radio access (E-UTRA); base station (BS) radio transmission and reception," 3GPP, Tech. Spec., Jan. 2017, release 14.

[16] J. C. Ikuno *et al.*, "LTE rate matching performance with code block balancing," in *17th European Wireless Conference (EW2011)*, Apr. 2011, pp. 2306–2311.

[17] A. Goldsmith, *Wireless Communications*. Cambridge University Press, Aug. 2007.